

Maestría en Inteligencia Artificial

Inteligencia Artificial y Computación Cognitiva

Índice

Esquema. Tema 1	3
Ideas clave. Tema 1	4
1.1. Introducción y objetivos	4
1.2. Definición y origen del concepto inteligencia artificial	4
1.3. Fundamentos de la inteligencia artificial	8
1.4. Historia de la inteligencia artificial	12
1.5. Inteligencia artificial y conceptos relacionados	16
1.6. Referencias bibliográficas	24
Esquema. Tema 2	27
Ideas clave. Tema 2	28
2.1. Introducción y objetivos	28
2.2. La inteligencia artificial en la empresa	28
2.3. Herramientas comerciales para inteligencia artificial	37
2.4. Tecnologías comerciales de inteligencia artificial	41
2.5. El futuro de la inteligencia artificial	51
Esquema. Tema 3	55
Ideas clave. Tema 3	56
3.1. Introducción y objetivos	56
3.2. Contexto legal aplicable a proyectos de inteligencia artificial	56
3.3. Sesgos, legalidad y responsabilidad	68
3.4. Seguridad y tolerancia ante ataques	75
3.5. Explicabilidad de algoritmos	78
3.6. Referencias bibliográficas	80
Esquema. Tema 4	83
Ideas clave. Tema 4	84
4.1. Introducción y objetivos	84

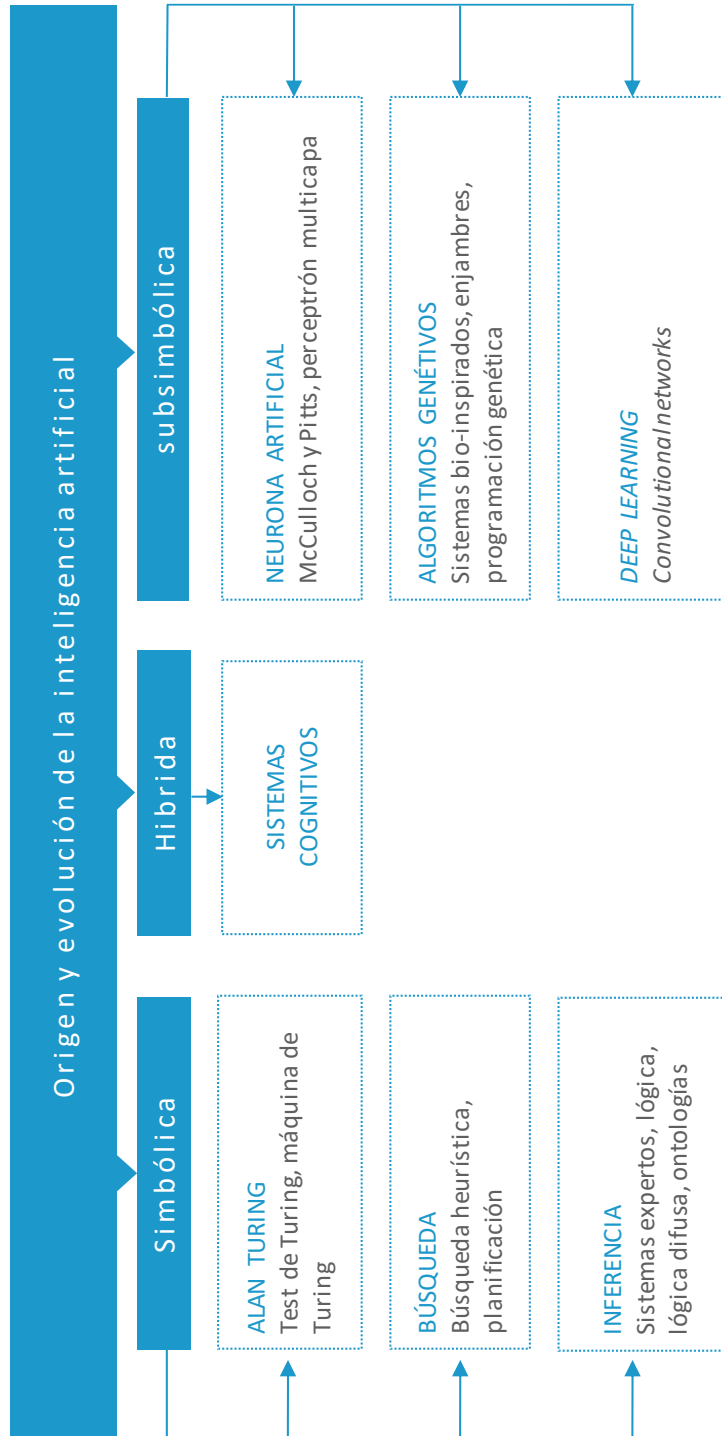
4.2. Introducción a la computación cognitiva	85
4.3. Elementos de un sistema cognitivo	92
4.4. Aplicaciones de la computación cognitiva	96
4.5. Referencias bibliográficas	97
Esquema. Tema 5	100
Ideas clave. Tema 5	101
5.1. Introducción y objetivos	101
5.2. Introducción a la cognición y la ciencia cognitiva	102
5.3. Inteligencia artificial en la ciencia cognitiva	110
5.4. Procesos básicos de la cognición humana y funciones cognitivas superiores	114
5.5. Modelos computacionales basados en la cognición humana	115
5.6. Referencias bibliográficas	119
Esquema. Tema 6	123
Ideas clave. Tema 6	124
6.1. Introducción y objetivos	124
6.2. Mecanismos perceptivos	125
6.3. Modelos computacionales de la visión	127
6.4. Organización del sistema motor	131
6.5. Referencias bibliográficas	133
Esquema. Tema 7	136
Ideas clave. Tema 7	137
7.1. Introducción y objetivos	137
7.2. Psicología de la emoción	138
7.3. Modelos computacionales de la emoción	142
7.4. Psicología de la motivación	144
7.5. Modelos computacionales de la motivación	147
Esquema. Tema 8	150
Ideas clave. Tema 8	151
8.1. Introducción y objetivos	151

8.2. Organización de la memoria	152
8.3. Modelos de memoria operativa	153
8.4. Psicología del lenguaje	160
8.5. Modelos computacionales del lenguaje	163
Esquema. Tema 9	167
Ideas clave. Tema 9	168
9.1. Introducción y objetivos	168
9.2. Componentes de las funciones ejecutivas	169
9.3. Supervisión de la conducta	171
9.4. Modelos computacionales de las funciones ejecutivas	173
9.5. Referencias bibliográficas	174

Inteligencia Artificial y Computación Cognitiva

Origen y evolución de la inteligencia artificial

Esquema. Tema 1



Ideas clave. Tema 1

1.1. Introducción y objetivos



Accede al vídeo «Introducción. Origen y evolución de la inteligencia artificial» a través del aula virtual

Este tema supone la primera introducción del alumno al concepto de inteligencia artificial, se trata, por tanto, de un tema introductorio con el objetivo genérico de adentrarse en la disciplina. Objetivos más concretos son:

- ▶ Ser capaz de enmarcar la evolución de la inteligencia artificial en su contexto histórico.
- ▶ Conocer las principales disciplinas o escuelas de la inteligencia artificial.
- ▶ Conocer algunos de los términos principales en el ámbito.

1.2. Definición y origen del concepto inteligencia artificial



Accede al vídeo «Actualidad de la inteligencia artificial» a través del aula virtual



Accede al vídeo «¿Qué es la inteligencia artificial?» a través del aula virtual

La inteligencia artificial está presente en todos los ámbitos de nuestra vida. La presencia de este término en los medios de comunicación es frecuente. La herramienta Google Trends (<https://trends.google.es/trends/>) permite analizar el interés a lo largo del tiempo de un término concreto a partir de las búsquedas realizadas en Google. Con el término en castellano «inteligencia artificial» se aprecia un interés creciente en el último año (figura 1).

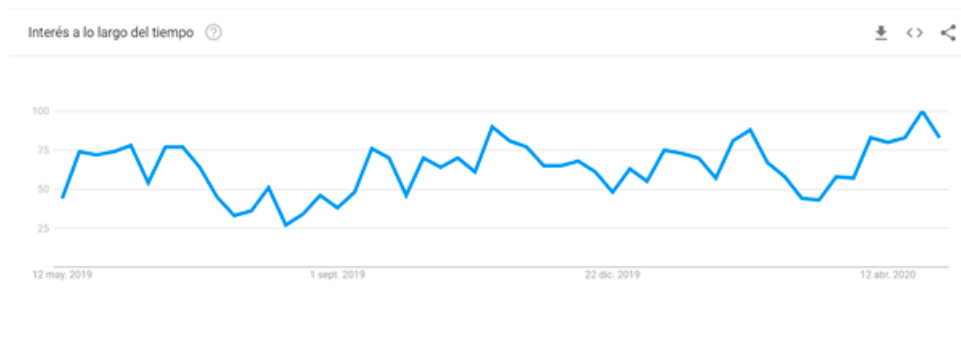


Figura 1. Interés a lo largo del tiempo del término «inteligencia artificial».
Fuente: <https://trends.google.es/trends/explore?date=2010-10-05%202020-05-10&geo=ES&q=Inteligencia%20Artificial>

Este interés es mucho más evidente si analizamos el término en su versión anglosajona (*artificial intelligence*). Como se ha explicado anteriormente, el inglés es la lengua franca de la ciencia y la técnica, por lo que todos los análisis de este tipo adquieren mayor relevancia si se hacen en este idioma. La figura 2 evidencia de forma muy clara como este interés se ha incrementado en los últimos años.

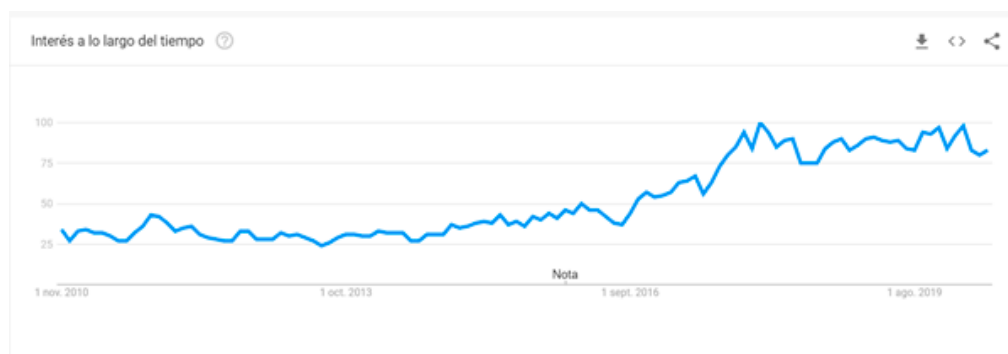


Figura 2. Interés a lo largo del tiempo del término *artificial intelligence*.
Fuente: <https://trends.google.es/trends/explore?date=2010-10-05%202020-05-10&q=artificial%20intelligence>

Pero ¿qué es la inteligencia artificial?, ¿para qué sirve? Se intentará dar respuesta a estas preguntas en los siguientes apartados.

¿Qué es la inteligencia artificial?

Intuitivamente, el término «inteligencia artificial» hace referencia a la **capacidad que tienen ciertas máquinas de comportarse como humanos**. Sin embargo, esta descripción no puede considerarse una definición, ya que es sumamente ambigua y difusa.

Para llegar a una buena comprensión del término deberíamos primero entender el **significado de inteligencia**. Según la Real Academia Española este término hace referencia a:

1. Capacidad de entender o comprender.
2. Capacidad de resolver problemas.
3. Conocimiento, comprensión, acto de entender.

Aunque las definiciones parecen claras y evidentes, el espectro se complica según nos adentramos en la interpretación del término. ¿Cuántos tipos de inteligencia hay?, ¿uno, varios?

Howard Gardner (1983) en su famosa obra *Frames of Mind: The Theory of Multiple Intelligences* se atrevió a formular diez tipos de inteligencia:

- ▶ Musical.
- ▶ Espacial o visual.
- ▶ Lingüística-verbal.
- ▶ Lógico-matemática.
- ▶ Corporal-cinestésica.
- ▶ Interpersonal.
- ▶ Intrapersonal.
- ▶ Naturalista.
- ▶ Existencial.
- ▶ Otras: categoría en la que entra incluso una inteligencia sexual.

La visión de Gardner no es plenamente aceptada, ya que algunos autores critican que este autor contempla como inteligencias lo que comúnmente se han denominado habilidades o competencias (Geake, 2008). Aun así, **esta teoría ha tenido gran impacto en ámbitos como el educativo**, donde tradicionalmente las pruebas de evaluación se enfocaban a un enfoque puramente lógico matemático y lingüístico.

Son numerosas las instituciones educativas que tratan hoy en día de ampliar la evaluación del estudiante incorporando factores asociados a sus capacidades para la expresión artística o corporal, por poner algunos ejemplos.

Otra discusión asociada al concepto de inteligencia es la exclusividad o no de la **inteligencia como atributo humano**. Sin embargo, los resultados de la experimentación científica han conseguido apartar esta discusión al poner en evidencia que son numerosas las especies animales que consiguen generar nuevas estrategias para adaptarse mejor a las necesidades del entorno y resolver los problemas asociados (Rayner, 1887), y todo ello a pesar de que la experimentación con otras especies animales en este ámbito no es trivial (MacLean, 2014).

Comentado el concepto de inteligencia, queda profundizar en el término que nos ocupaba al inicio. Según la RAE (Real Academia Española), la **inteligencia artificial** es:

«La disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico».

Otras definiciones del término muy extendidas son las siguientes:

«La automatización de actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje...» (Bellman, 1978).

«El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia» (Kurzweil, Richter, Kurzweil y Schneider, 1990).

Como se puede apreciar, la automatización o realización de estas actividades por máquinas (artificial) es una característica común a estas definiciones.

1.3. Fundamentos de la inteligencia artificial



Accede al vídeo «La inteligencia como atributo humano» a través del aula virtual



Accede al vídeo «Disciplinas relacionadas con la inteligencia» a través del aula virtual

El objetivo de este apartado es identificar algunas disciplinas y conceptos que han contribuido al desarrollo de la inteligencia artificial.

Sin duda alguna, la **filosofía** es una de las disciplinas que más ha contribuido al desarrollo de la inteligencia artificial y lo ha hecho desde diversos enfoques. Las ideas filosóficas han sido básicas para definir los conceptos de «inteligencia», «racional», etc.

Aristóteles (384-322 a. C.) fue el primer autor que intentó definir el conjunto de leyes que rigen el razonamiento. El mallorquín Ramón Llull (1232-1315), patrón de los ingenieros informáticos, dedicó parte de su vida a diseñar una máquina lógica capaz de demostrar una verdad o refutar una mentira. Thomas Hobbes (1588-1679) contemplaba amplias analogías entre el proceso de razonamiento y el proceso de computación numérica.

A partir del siglo XVI aparecen corrientes filosóficas que marcan el debate de la época. Por un lado, **nos encontramos el dualismo frente al materialismo**. El dualismo hace referencia a la existencia de dos componentes independientes entre sí y que marcan nuestras elecciones. Uno de esos componentes de la mente sería el asociado al alma y funciona al margen de las leyes físicas, mientras que el otro se basa exclusivamente

en la aplicación de los principios físicos universales. René Descartes (1596-1650) era uno de los firmes defensores de esta corriente.

El materialismo surge como contraposición al dualismo al considerar que todas las operaciones del cerebro se realizan de acuerdo con las leyes de la física que constituyen la mente. Según los materialistas, la conciencia es fruto de la materia. Francis Bacon (1561-1626) y Thomas Hobbes (1588-1679) fueron dos de los máximos exponentes de esta corriente.

Respecto a la procedencia del conocimiento, el **movimiento empírico** iniciado por Bacon establece que todo conocimiento tiene que pasar por la percepción de los sentidos. Como decía John Locke (1632-1704) «nada existe en la mente que no haya pasado antes por los sentidos».

La relación entre conocimiento y acción busca explicar los desencadenantes de las acciones y constituye una cuestión de especial relevancia en la inteligencia artificial. Los **postulados de Aristóteles** son, todavía hoy, una fuente imprescindible a la hora de entender esta relación. Según el filósofo griego las acciones pueden justificarse en base a la conexión lógica entre los objetivos planteados y el conocimiento de la operativa y consecuencia de las acciones aprendidas.

Obviamente, la filosofía no es la única disciplina que ha aportado valor a la evolución de la inteligencia artificial. Las matemáticas, economía, neurociencia, psicología e ingeniería computacional también han realizado indispensables contribuciones a este campo.

Otras contribuciones de la inteligencia artificial

La **matemática** es una de las ciencias que más ha contribuido al desarrollo de la inteligencia artificial. Hoy en día es complicado considerarse experto en inteligencia artificial sin poseer una elevada competencia en el manejo y comprensión de los fundamentos matemáticos relacionados con la lógica, álgebra, cálculo, etc.

Además del concepto de algoritmo trabajado desde la Antigüedad por un amplio abanico de matemáticos, la **lógica matemática** es esencial para permitir el modelado de ciertas conductas. Dos contribuciones fundamentales son el trabajo de **George Boole** (1815-1864) definiendo la lógica proposicional o booleana (se mencionará con más detalle más adelante), y las contribuciones de **Gottlob Frege** (1848-1925), extendiendo la lógica de Boole para incluir objetos y relaciones y dando lugar a la lógica de primer orden que permite la representación del conocimiento.

En 1930, **Kurt Gödel** (1906-1978) da lugar a uno de los mayores hitos de la historia de las matemáticas y la ciencia en sí mismo formulando los famosos **teoremas de incompletitud**. De forma muy resumida y empleando términos sencillos, tomando como base los dos teoremas formulados por Gödel se puede demostrar que existen afirmaciones que no se pueden demostrar verdaderas o falsas empleando la sintaxis y notación algorítmica.

Para muchos autores, el principio de incompletitud pone límite al conocimiento humano al evidenciar la existencia de postulados no demostrables mediante la argumentación matemática.

Alan Turing (1912-1954) centró sus investigaciones en lo que sí podía ser demostrado y dio pie a la **máquina de Turing**. El trabajo de Turing fue absolutamente esencial para el desarrollo de la teoría de la computación tal y como lo conocemos actualmente.

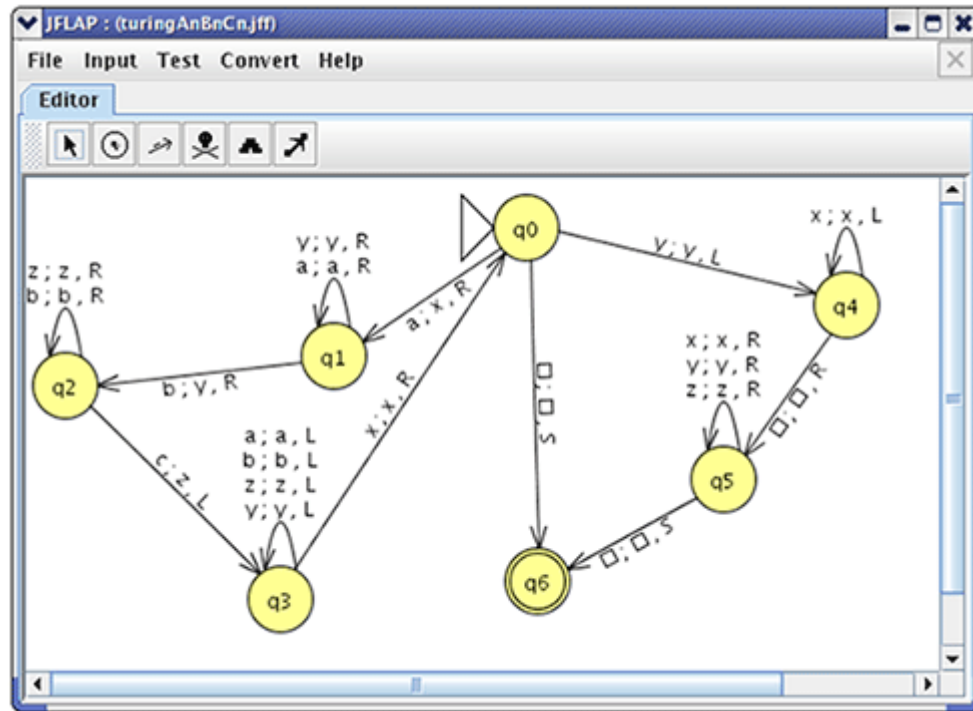


Figura 3. Ejemplo de máquina de Turing empleando herramienta JFlap.

Fuente: <http://www.jflap.org/tutorial/turing/one/index.html>

Por último, los trabajos matemáticos relacionados con el estudio de la complejidad algorítmica y el desarrollo de los algoritmos probabilistas también merecen ser mencionados en este resumen.

Respecto a la **economía** pura y las matemáticas aplicadas a la economía, las investigaciones centradas en teoría de la decisión, teoría de juegos e investigación operativa permitieron entender y modelar mejor el concepto de agente inteligente.

La **neurociencia**, en todo lo referente a la estructura neuronal, y la **psicología**, en todo lo referente al análisis del comportamiento, realizan aportaciones destacadas a la evolución de los métodos empleados en la disciplina.

Finalizaremos este apartado hablando de las contribuciones de la **ingeniería informática**, y lo haremos centrándonos exclusivamente en las aportaciones más recientes. En un contexto de gran disposición de datos para entrenar los sistemas inteligentes y mejorar su desempeño, las aportaciones de la ingeniería informática al

sector han sido esenciales permitiendo el diseño e implementación de sistemas escalables que permiten realizar en paralelo miles de millones de operaciones posibilitando la creación de aplicaciones que, en muchas ocasiones, compiten de tú a tú con las capacidades humanas (cuando no las superan).

1.4. Historia de la inteligencia artificial



Accede al vídeo «Hitos históricos de la inteligencia artificial» a través del aula virtual

Se resumirá de forma básica en este apartado la evolución de la inteligencia artificial. Se resaltarán algunos hitos destacados en la disciplina.

Aunque el trabajo puro relacionado con la inteligencia artificial comenzó poco después de la Segunda Guerra Mundial, y el término fue formalmente fijado en 1956, existen algunos antecedentes que merece la pena recordar.

En 1912, el genio español **Leonardo Torres Quevedo** (llegó a diseñar el funicular del Niágara) creó una máquina autónoma capaz de jugar al ajedrez.



Figura 4. Autómata para jugar al ajedrez diseñado por Torres Quevedo.

Fuente: <http://quhist.com/wp-content/uploads/2012/07/Ajedrecista-de-Torres-Quevedo-300x203.jpg>

Constituyó este un hito importante en la época, dando una idea del potencial que podía llegar a adquirir una máquina.

En 1943, **McCulloch y Pitts** proponen una unidad de cálculo que intenta modelar el comportamiento de una neurona «natural», similar a las que constituyen el cerebro humano.

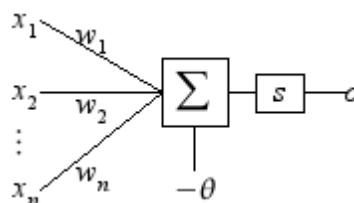


Figura 5. Unidad de cálculo propuesta por McCulloch y Pitts.

En 1949, **Donald Hebb** formula la regla de Hebb.

En 1950, el gran científico **Alan Turing** publica su obra *Computing Machinery and Intelligence* donde expone su famosa prueba de Turing. Esta prueba se diseñó para proporcionar un mecanismo que permitiese decidir si una máquina tenía un comportamiento «inteligente» o no. La prueba consiste en un humano que contempla dos terminales. En estos terminales hay un humano y un programa que contestan y dialogan con el humano o humanos entrevistadores. Si el humano no es capaz de distinguir entre el humano real y el *software*, entonces podemos decir que este *software* ha superado la prueba y le podemos considerar inteligente.

En **1950** también aparecen programas informáticos que aplican técnicas de inteligencia artificial como el programa de ajedrez de Arthur Samuel, el solucionador de problemas lógicos de Newell y Simon y la máquina de Gelernter, capaz de deducir problemas de geometría euclídea.

De 1952 a 1969 se da el período conocido como «Look, Ma, ¡no hands!». Esta época se caracteriza por un gran entusiasmo y expectativa en las aplicaciones de la inteligencia artificial.

En **1956** tiene lugar la Conferencia de Dartmouth (en el Dartmouth College, Estados Unidos). En este encuentro de muchos de los investigadores más influyentes del momento en el área se adoptó oficialmente el término «inteligencia artificial».

En **1965 Robinson** propone un algoritmo capaz de aplicar criterios de razonamiento lógico. Al mismo tiempo, aparece en el MIT el primer *chatbot* denominado Eliza.

En el período comprendido **entre 1966 y 1974** se profundiza en el concepto de la complejidad computacional, algo esencial para las futuras evoluciones de la disciplina.

Entre 1969 y 1979 se desarrollan los primeros sistemas basados en el conocimiento.

En **1974** aparecen los primeros trabajos con coches autónomos, aunque todavía muy primitivos y no muy alejados de coches a control remoto.

Entre 1980 y 1988 aparecen en el mercado diversos sistemas expertos.

Entre los años **1985 y 1995** se retoma el interés por las redes neuronales.

A partir de 1988 nuevas soluciones como algoritmos genéticos o *soft computing* contribuyen al enriquecimiento de la disciplina.

A partir de 1995 se produce la gran explosión de los agentes inteligentes.

En **1997** el ordenador de IBM (Deep Blue) derrota a Gary Kasparov en uno de los duelos de ajedrez que más expectación han generado.

En el año **2011**, el programa de IBM Watson derrota al campeón (humano) del conocido juego *Jeopardy*.

En el año **2015**, el *software* desarrollado por Google, DeepMind AlphaGo, derrota al campeón mundial de *Go*.

En el año **2019**, DeepMind AlphaStar consigue ser gran maestro en la liga mundial de jugadores de *Starcraft II*, ganando el 98 % de las partidas.

No todo han sido éxitos en la historia de la inteligencia artificial, **los períodos que abarcan los años 1974-1980 y 1987-1993 son conocidos como los inviernos de la inteligencia artificial**. Estas franjas de tiempo se caracterizan por una gran desilusión en las posibilidades de esta disciplina y descenso del número de aportaciones.

Aunque no constituye un hito puramente científico, merece la pena mencionar la **aparición de las tres leyes de la robótica de Asimov** (aunque también hay quien asocia la coautoría de estas leyes con John W. Campbell). La aparición de estas leyes propició interesantes debates sobre la evolución de la inteligencia artificial. Las tres leyes de la robótica de Asimov establecen los ejes fundamentales que deberían guiar la conducta de un robot. Son las siguientes:

1.ª Ley	2.ª Ley	3.ª Ley
Un robot no hará daño a un ser humano o, por inacción, permitir que un ser humano sufra daño	Un robot debe hacer o realizar las órdenes dadas por los seres humanos, excepto si estas órdenes entrasen en conflicto con la primera ley	Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o la segunda ley

Figura 6. Leyes de la robótica de Asimov.

Las **leyes de Asimov** guardan relación con los retos éticos que afronta la inteligencia artificial y que serán tratados en próximos temas.

1.5. Inteligencia artificial y conceptos relacionados



Accede al vídeo «Áreas destacadas de la inteligencia artificial» a través del aula virtual

Inteligencia artificial es un concepto muy amplio y genérico. Por ello es útil y necesario precisar con mayor exactitud algunas de las aplicaciones o campos de actuación de la inteligencia artificial más referenciados.

Robótica

Comencemos con el concepto de robótica. Según la RAE, un robot es una «máquina o ingenio electrónico programable, capaz de manipular objetos y realizar operaciones antes reservadas solo a las personas». El término robot proviene de su correspondiente anglosajón (con la misma grafía), que proviene a su vez de la voz checa *robota*, que significa ‘trabajo, prestación personal’.

El significado original del término tiene sentido porque los robots se concibieron como máquinas destinadas a ayudar al ser humano a realizar las tareas más pesadas, estando ahora totalmente introducidos en casi la totalidad de las cadenas de producción.

Robótica, por tanto, es la ciencia que estudia el diseño y construcción de máquinas autónomas capaz de realizar tareas de forma inteligente, resolviendo problemas y adaptándose a los cambios que suceden en el entorno.

A la hora de hablar de robótica, haremos referencia con relativa frecuencia a Sophia por el nivel de interlocución con los humanos que demuestra. Sophia es producto de Hanson Robotics.



Figura 7. Sophia, robot creado por Hanson Robotics.
Fuente: <http://www.hansonrobotics.com/robot/sophia/>

Sin embargo, no es Sophia el tipo de robot susceptible de estar más presente en nuestras casas, los robots aspiradoras o robots juguetes son ejemplos más cercanos a nuestros hogares.

Sistemas expertos

Una de las ramas más conocidas de la inteligencia artificial es la de los sistemas expertos. Un sistema experto **intenta acumular el conocimiento existente en un ámbito concreto (por ejemplo, el área de atención primaria de medicina) y aplicarlo a la toma de decisiones empleando procesos de razonamiento lógico**. El desarrollo de sistemas expertos se potenció especialmente a partir de los años 70. Hoy en día disponemos de sistemas expertos que ayudan a tomar mejores decisiones en el ámbito clínico, transporte aéreo, seguridad, gestión de empresas, etc.

El desarrollo de un sistema experto suele exigir el modelado del conocimiento asociado para que la máquina pueda tenerlo en consideración. Para esta tarea son especialmente indicadas las **ontologías**. Una ontología pretende representar el conocimiento existente, así como las propiedades y relaciones entre los distintos

conceptos de forma consistente. La figura 8 muestra un sencillo ejemplo de una ontología.

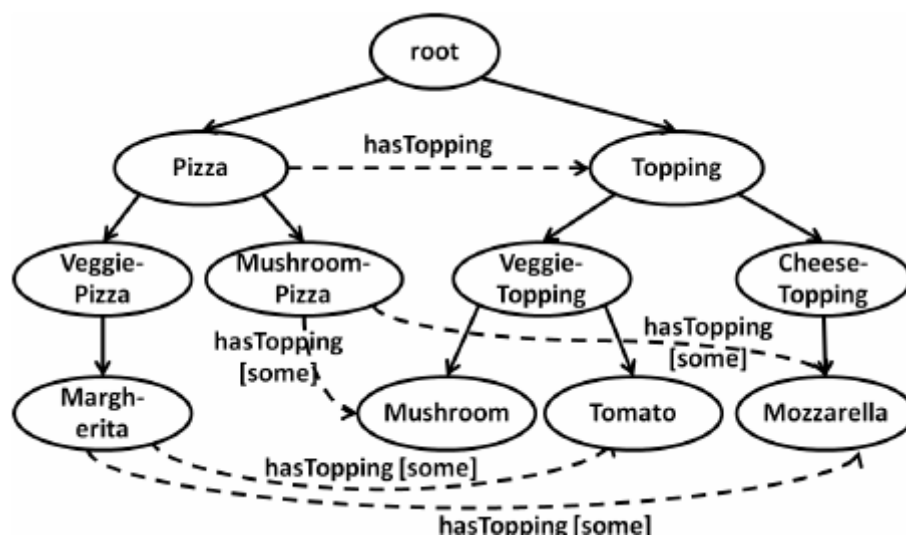


Figura 8. Ejemplo de una ontología para modelar la disponibilidad de pizzas existentes.

Fuente: https://www.researchgate.net/figure/236842047_fig1_Figure-1-Example-pizza-ontology-represented-as-a-graph-G-a-and-a-changed-version-of

Procesamiento del lenguaje natural

El procesamiento del lenguaje natural (PLN) estudia las **interacciones entre las computadoras y el lenguaje humano**. Es este uno de los ámbitos de trabajo más fructíferos de los últimos años y ha dado lugar a conocidos productos como Siri de Apple o Alexa de Amazon.

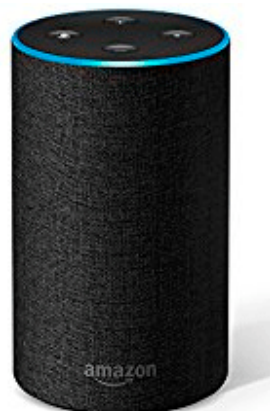


Figura 9. Echo & Alexa.

Fuente: https://images-na.ssl-images-amazon.com/images/I/51TFnR7AtGL.AC_US218.jpg

Estos productos permiten al usuario comunicarse con un dispositivo concreto, como puede ser el teléfono móvil, a través del lenguaje natural.

Las técnicas de procesamiento del lenguaje natural se aplican también al diseño de *chatbots*. Los *chatbot* son piezas de *software* diseñada para aplicaciones de mensajería que interactúan con el usuario intentando comprender y satisfacer sus necesidades proveyendo acceso al servicio más adecuado en cada momento.

Productos como Siri o Alexa se enfrentan a multitud de retos a la hora de interpretar el lenguaje hablado. La gran diversidad de idiomas y acentos, así como la necesidad de proporcionar una respuesta coherente en tiempo y forma adecuado hace que este tipo de soluciones no sean triviales.

Algoritmos genéticos

Una parte de la inteligencia artificial se ocupa de lo que algunos autores vienen a denominar **pautas para la vida artificial**. Este tipo de técnicas simulan realidades virtuales que evolucionan en función de un conjunto de reglas previamente definidas. Uno de los ejemplos más conocidos de este paradigma son los algoritmos genéticos.

Los algoritmos genéticos **suelen emplearse para encontrar la solución a un problema concreto simulando las leyes evolutivas básicas**. En primer lugar, se parte de una inicialización, frecuentemente aleatoria, de la población. Cada elemento de la población está compuesto de una serie de genes que se interpretan según una función de fitness previamente definida por el usuario. Esta función de fitness permite comparar los distintos individuos entre sí y saber cuáles están más cerca del objetivo buscado.

Por ejemplo, supongamos que buscamos maximizar la función $f(x, y) = x^2 + 5y^3$. Cada individuo podría estar compuesto de dos genes representando a x e y . Según esto, los mejores individuos considerando el objetivo inicialmente planteados serían los que tuviesen un valor más alto a la hora de calcular $f(x, y)$.

Ciertos elementos de la población (por ejemplo, los más aptos al entorno según el criterio elitista) se cruzan entre sí (*crossover*) generando una descendencia y compartiendo genes. Algunos estos genes podrían sufrir una mutación. A continuación, se evalúa cada individuo de la población según la función de fitness determinando, según el criterio seguido, quiénes sobreviven y pasan a la siguiente iteración y quiénes fallecen. La figura 10 muestra el esquema general de un algoritmo genético.

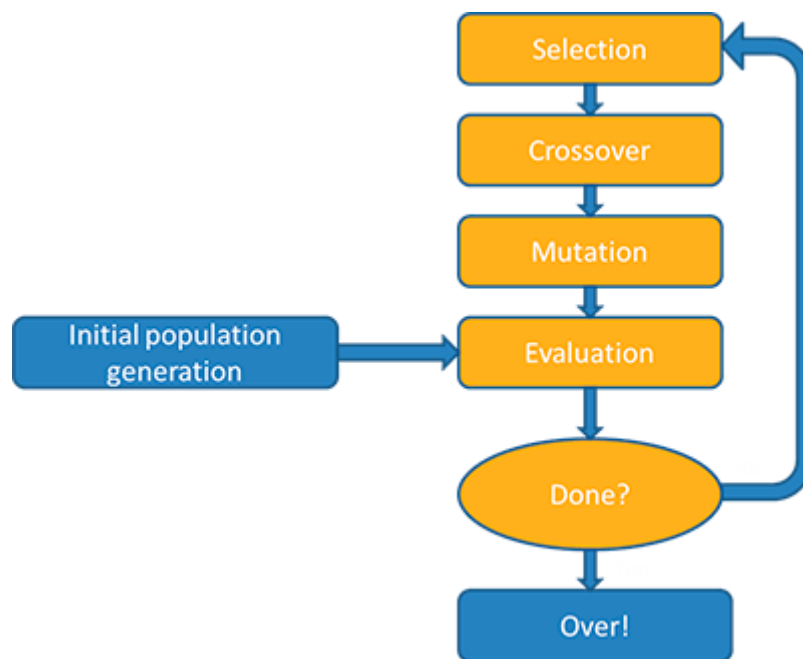


Figura 10. Esquema general de un algoritmo genético.

Fuente: https://genetic.io/wp-content/uploads/2016/07/processus_en.png

Machine learning (aprendizaje automático)

Las técnicas de *machine learning* o aprendizaje automático pretenden **generalizar comportamientos y encontrar patrones** en función de los ejemplos proporcionados de antemano. Existen dos grandes tipologías a la hora de afrontar un problema de aprendizaje automático: aprendizaje supervisado y aprendizaje no supervisado.

Con las técnicas de **aprendizaje supervisado** se entrena al algoritmo con un conjunto de datos previamente etiquetado (conjunto de entrenamiento). En función de ese

entrenamiento previo el algoritmo generará un modelo. La aplicación de dicho modelo a nuevos datos permitirá predecir entradas futuras.

Por ejemplo, supongamos que queremos generar un modelo capaz de discernir si una imagen representa una cara humana o no. Para ello alimentaremos al algoritmo con imágenes de caras humanas variadas (y las catalogaremos como «caras») e imágenes de otros objetos (y las catalogaremos como «no caras»). El algoritmo tratará de encontrar patrones comunes de las «caras» frente a las «no caras» para generar un modelo. Posteriormente, probaremos dicho modelo con un conjunto nuevo de imágenes (conjunto de prueba). En este caso, no se proporciona información al modelo sobre si está leyendo una cara o no. Será el modelo el que haga la predicción oportuna. Comparando los resultados del modelo en este conjunto de prueba con la realidad podremos obtener una idea de su fiabilidad. La figura 10 muestra el esquema asociado a la fase de entrenamiento.

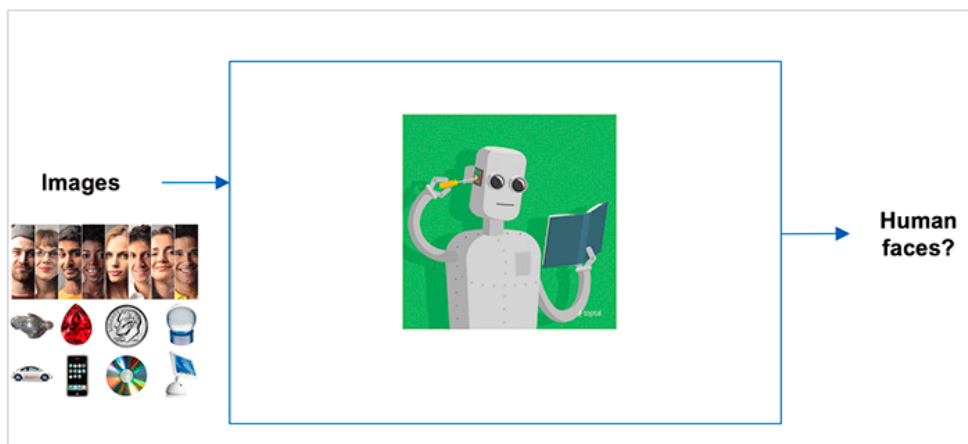


Figura 11. Ejemplo de aprendizaje supervisado.

Fuente: <https://uploads.toptal.io/blog/image/443/toptal-blog-image-1407508081138.png>

En el caso del **aprendizaje no supervisado**, no se proporciona al algoritmo información previa de etiquetado. Por ejemplo, supongamos que queremos clasificar a los clientes de una empresa en determinados grupos en función de su actividad (*clustering*). En este caso no sabemos de forma previa el grupo al que pertenece el cliente, por tanto, no se puede proporcionar un etiquetado previo. En función de los

parámetros de cada cliente, propiedades y relaciones numéricas entre los distintos individuos el algoritmo propondrá distintas tipologías o grupos de clientes.

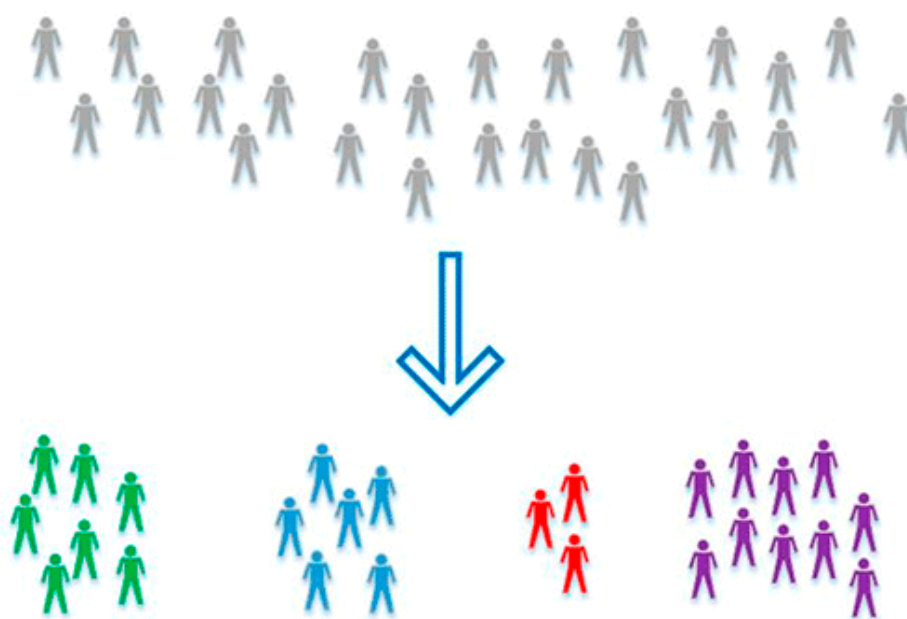


Figura 12. Ejemplo de aprendizaje no supervisado.

Fuente: <https://insidebigdata.com/>

Redes neuronales y *deep learning*

Las redes neuronales representan un modelo computacional de las conexiones entre neuronas que se dan en el cerebro. Cada neurona recibe información de entrada procedente de varias fuentes y emite una salida concreta en función de su configuración.

Asociado al concepto de redes neuronales, un término en auge es el concepto de *deep learning*. Las técnicas de *deep learning* constituyen una familia específica de **algoritmos de aprendizaje automático centrados en aprender representaciones de datos**, por ejemplo, la estructura de una imagen que represente una cara humana. El ejemplo más común de *deep learning* es una red neuronal de varias capas ocultas. La figura 13 muestra un conjunto de neuronas interconectadas entre sí. Las capas del medio (marcadas en el gráfico como *hidden layer 1, 2 y 3*) reciben datos de las capas

previas, procesan el resultado según un procedimiento previamente estipulado antes de compartir la salida con otras neuronas.

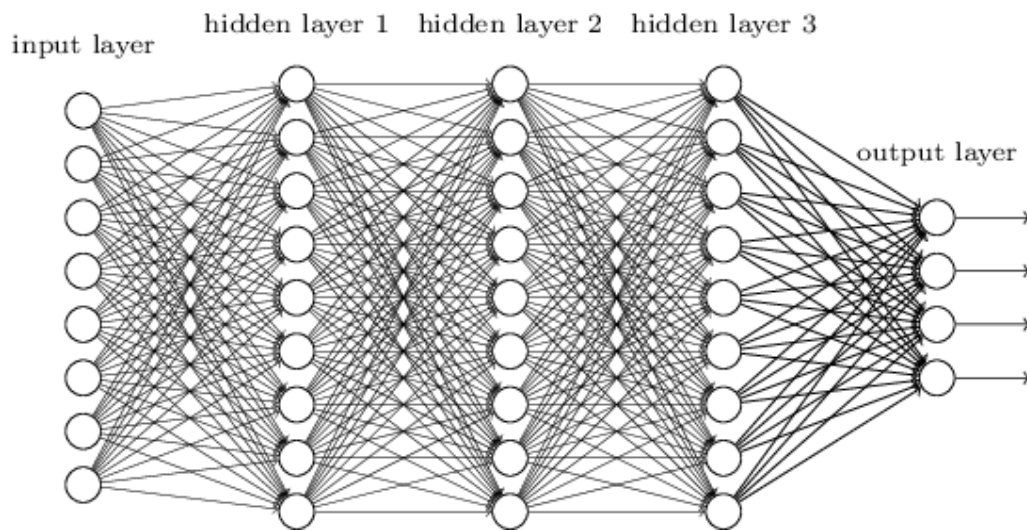


Figura 13. Red neuronal de varias capas ocultas.

Fuente: https://cdn-images-1.medium.com/max/1600/1*Pn42wNB6_HBKgpvxRzGBWw.png

Las aplicaciones de las técnicas de *deep learning* son varias y diversas, desde visión por computador hasta reconocimiento automático del habla pasando por sistemas de recomendación. En este punto, el alumno no debe preocuparse por que le resulten vagos y poco concretos los contenidos sobre redes neuronales y *deep learning* expuestos en este tema. Pretendemos aquí simplemente introducir estos términos de forma muy básica.

Computación cognitiva

El último de los conceptos que vamos a estudiar en este tema es el concepto de «computación cognitiva», concepto sumamente amplio que está centrado en **intentar que las máquinas piensen de la forma más parecida posible a como lo haría un ser humano**. La computación cognitiva exige aprovechar todos los datos que tenemos a nuestra disposición tanto estructurados (bases de datos) como no estructurados (imágenes, vídeos, sonidos...). La computación cognitiva obliga a usar estos datos en contexto. Pongamos un ejemplo de esto: supongamos que están hablando dos personas y una de ellas dice a la otra: «¡Claro!, me parece perfecto».

¿Qué ha querido decir? ¿Está de acuerdo? La respuesta es: depende. El tono del mensaje, la expresión facial a la hora de transmitir el mensaje, todo es relevante. Si un programa simple de procesamiento del lenguaje natural analizase la frase, lo normal es que interpretase que sí, que efectivamente está de acuerdo. Una aplicación cognitiva, sin embargo, sería capaz de contextualizar el mensaje en línea con el tono de voz y la expresión facial a la hora de tomar una decisión.

Las aplicaciones cognitivas están en auge en la industria, pero debemos ser especialmente cuidadosos porque, en ocasiones, el término se emplea con demasiada libertad.

1.6. Referencias bibliográficas

Bellman, R. (1978). *An introduction to artificial intelligence: Can computers think?* Thomson Course Technology.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. NY: Basics.

Geake, J. (2008). Neuromythologies in education. *Educational Research*, 50(2), 123-133.

Kurzweil, R., Richter, R., Kurzweil, R. y Schneider, M. L. (1990). *The age of intelligent machines* (Vol. 579). Cambridge: MIT press.

MacLean, E. L. et al. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences*, 111(20), E2140-E2148.

Rayner, H. (1887). Breeding for Intelligence in Animals. *Nature*, 36, 246-246.

Russell, S. y Norving, P. (2004). *Inteligencia Artificial: un enfoque moderno* (pp. 1-35). Madrid: Pearson Educación.



Accede a los ejercicios de autoevaluación a través del aula virtual



Accede al vídeo «Resumen del tema» a través del aula virtual

Inteligencia Artificial y Computación Cognitiva

Soluciones empresariales de inteligencia artificial

Esquema. Tema 2

Aplicación de la inteligencia artificial al ámbito empresarial		
Tipos de datos		Visualización
Google		Marketing, ventas, CRM
Facebook		Automoción
DeepMind		<i>Business Intelligence</i>
OpenAI		ChatBots
Baidu		Plataformas
Microsoft Research		Ciberseguridad
Apple		<i>Fintech</i>
IBM		Salud y sanidad
AlBrain		Internet de las cosas
Amazon		Robótica
Netflix		Narrativa
		Visión artificial
		Biología y medio ambiente

Ideas clave. Tema 2

2.1. Introducción y objetivos



Accede al vídeo «Introducción. Soluciones empresariales de inteligencia artificial» a través del aula virtual

En este tema se transmite una visión amplia de los principales proveedores y soluciones basadas en *machine learning* disponibles en el mercado. Paralelamente, el alumno recibirá información sobre el estado del arte de este tipo de soluciones en las diferentes industrias.

Por último, se pretende que el alumno sea capaz de detectar los ámbitos en los que los principales actores del área están centrando sus esfuerzos.

2.2. La inteligencia artificial en la empresa



Accede al vídeo «Presencia de la inteligencia artificial en negocio» a través del aula virtual



Accede al vídeo «Principales desarrolladores de inteligencia artificial» a través del aula virtual

Las **soluciones basadas en inteligencia artificial están a nuestro alrededor a día de hoy más de lo que creemos**. Según un estudio realizado por Microsoft en quince países europeos, se estima que el porcentaje de empresas a nivel europeo que usan o están en desarrollo de proyectos de inteligencia artificial actualmente es del 28 %. Y se estima que un 51 % de ellas están planificando su implantación en el futuro. Por lo tanto, la inteligencia artificial está empezando a penetrar poco a poco en todo el tejido empresarial, tanto en las grandes empresas como en las pequeñas y medianas.

Según el mismo estudio el 89 % creen que la IA (inteligencia artificial) generará beneficios comerciales al optimizar las operaciones de sus compañías y el 74 % espera que la IA sea clave para atraer a más clientes.

Informe del estado de la inteligencia artificial en las empresas europeas. Puedes acceder a este recurso pinchando en el siguiente enlace.

<https://pulse.microsoft.com/es-es/business-leadership-es-es/na/fa1-artificial-intelligence-report-at-a-glance/>

Aunque son datos extraídos de este informe en concreto pueden no ser al 100 % representativos de la realidad. Sí que **denotan una tendencia clara a que el peso de la IA en el futuro será muy grande en el ámbito empresarial.** Pero hoy, ¿cuál sería la penetración de la IA en las empresas actuales? Podríamos dar muchos ejemplos desde medición de cosechas usando internet de las cosas y *machine learning*, hasta asistentes virtuales en diferentes compañías, o el uso de la IA para mejorar la obtención de imágenes en las cámaras de fotos o en la calidad de imagen de nuestros televisores. Pero nos vamos a centrar principalmente en los grandes actores que muchas veces están detrás de estas implementaciones concretas de IA en las empresas. Siempre teniendo en cuenta que es un entorno muy dinámico y cambiante y las tendencias pueden cambiar en muy poco tiempo. La lista de actores y herramientas que enunciamos, por tanto, en este tema no pretende ser una lista exhaustiva pero sí incluir empresas y soluciones más importantes y que deberían ser conocidas por todos los profesionales del sector para ser tenidas en cuenta.

Líderes en investigación y productos basados en inteligencia artificial

Comenzaremos desglosando algunas de las compañías que se están posicionando como líderes en temas de investigación en inteligencia artificial. Destacaremos a Google, Facebook, Deepmind, OpenAI, Baidu, Microsoft Research. Muchos de los productos que estas empresas ofrecen están basados en la inteligencia artificial, aprendizaje automático o subdominio del área.

Alphabet

Alphabet es la **matriz de conglomerados de Google que engloba a una gran cantidad de empresas bajo el mismo paraguas corporativo**. Muchas de ellas relacionadas directa o indirectamente con la IA. Google es una de las empresas más conocidas del mundo y también una de las más prolíficas en lo que se refiere a la generación de soluciones basadas en inteligencia artificial. Su funcionamiento interno, orientación hacia el mundo digital y la gran cantidad de datos de los que dispone facilita el desarrollo de este tipo de productos.

Una de las herramientas más destacados de Google en el campo de la inteligencia artificial es **Google Brain**. Este proyecto combina técnicas de aprendizaje automático (haciendo un uso más intensivo de las técnicas de *deep learning*) con el desarrollo de sistemas computacionales escalables y eficientes. Su objetivo es **contribuir a que las máquinas puedan aprender por sí mismas a realizar tareas cada vez más complejas**. Y todo ello haciendo uso eficaz y eficiente de los recursos computacionales puestos a su disposición. Sus investigadores han publicado varios resultados relacionados con la detección de imágenes, traducción automática, criptografía, etc. Son habituales las colaboraciones de esta ramificación de Google con la comunidad académica y con otras empresas del área.

Alguno de los proyectos más interesantes que ha generado Google Brain es Deep Dream, un sistema que genera arte usando *machine learning*.

Deep Dream o cómo generar arte con *machine learning*. Puedes acceder a este recurso pinchando en el siguiente enlace.

<https://deepdreamgenerator.com/>

Otro de los productos destacados que ha generado Google Brain ha sido TensorFlow. **TensorFlow** es un *framework* que permite, esencialmente, el diseño y programación de redes de neuronas artificiales de distintas tipologías y complejidades, con soporte

para compilar álgebra lineal, *open source* y que cada vez tiene mayor número de usos y algoritmos implementados.

TensorFlow una herramienta *open source* que permite ejecutar modelos de *deep learning* usando la GPU. Puedes acceder a este recurso pinchando en el siguiente enlace.

<https://www.tensorflow.org>

DeepMind es una de las empresas más activas en los últimos años en el campo de la IA y con los resultados más impresionantes. Es otra de las patas de Alphabet, tras ser adquirida por Google en el año 2014. La empresa está ubicada en Londres y son los creadores de **AlphaGo y AlphaStar**, ambos proyectos los primeros capaces de ganar a los mejores jugadores de *Go* y del videojuego de estrategia en tiempo real *Starcraft 2*.

Sus investigadores son realmente prolíficos y respetados. Realizan contribuciones a diversos campos dentro de la inteligencia artificial, campos como *deep reinforcement learning*, redes neuronales bayesianas, robótica, *transfer learning*, etc. Se atreven con los temas más diversos llegando incluso a abarcar la relación entre ecología e inteligencia.

DeepMind, creadores de AlphaGo y AlphaStar. Puedes acceder a este recurso pinchando en el siguiente enlace.

<https://deepmind.com/>

Microsoft Research

Es la **división de Microsoft dedicada a la investigación en general y en IA en particular**. Actualmente es una de las empresas más activas en esta tecnología junto con IBM y Alphabet, y se está potenciando su peso en ella a través de diferentes servicios agrupados en lo que denominan Azure AI. Su objetivo es **tener diferentes**

plantillas para crear proyectos de forma rápida. Microsoft ofrece módulos para diferentes soluciones con muchas de las funcionalidades necesarias ya implementadas. A esto lo llama **Azure Cognitive Services**, que es una completa familia de servicios de inteligencia artificial y API (application programming interface) cognitivas que le ayudan a crear aplicaciones inteligentes. También dispone de **Azure Machine Learning** que es un servicio de aprendizaje automático para empresas, que permite aprender modelos usando Azure. También existen otros servicios como chatbots o plataformas de análisis de datos usando Apache Spark.

Microsoft Azure Machine Learning y Cognitive Services son los dos pilares principales de la IA de Microsoft destinada para empresas. Puedes acceder a este recurso pinchando en el siguiente enlace.

<https://azure.microsoft.com/es-es/services/machine-learning/>
<https://azure.microsoft.com/es-es/services/cognitive-services/>

Facebook

Facebook es otro de los grandes actores de esta industria. A lo largo de su historia ha producido conocimientos diversos relacionados con el procesamiento de texto, gestión de recursos computacionales, evaluación de máquinas inteligentes, etc. La división de investigación de la compañía se denomina **Facebook AI Research (FAIR)**.

Una de las aportaciones más destacadas de Facebook a la comunidad es el *framework* **Pytorch**, al que se puede considerar competidor de TensorFlow como recurso para generar modelos basados en *deep learning*.

Facebook y su librería *open source* Pytorch. Puedes acceder a este recurso pinchando en el siguiente enlace.

<http://pytorch.org>

OpenAI

OpenAI es una **pequeña compañía sin ánimo de lucro liderada por conocidos actores del mundo de la inteligencia artificial y de la que es fundador Elon Musk entre otros**. Tiene el objetivo de lograr una inteligencia artificial ligada a las necesidades reales de la sociedad. Suele liberar los resultados de sus investigaciones y patentes. Uno de sus productos más conocidos es **Universe**, una plataforma de *software* para medir y entrenar soluciones basadas en inteligencia artificial. Open AI ha tenido importantes éxitos en robótica, logrando vencer a jugadores humanos del videojuego *Dota 2* así como la creación de una IA que es capaz de jugar a juegos retro (Retro Contest)

Open AI y su proyecto Retro Contest donde una IA aprende a jugar a juegos retro usando aprendizaje por refuerzo profundo. Puedes acceder a este recurso pinchando en el siguiente enlace.

<https://openai.com/blog/retro-contest/>

Baidu

Baidu, el motor de búsqueda por excelencia en el idioma chino, es uno de los grandes gigantes del mercado. Esta compañía nacida en Pekín **lidera el mercado chino abarcando una gran diversidad de productos** que van desde asistentes para el hogar a coches autónomos, pasando por distintas soluciones de ayudas a discapacitados.

Apple

Apple es una de las marcas más conocidas del panorama empresarial. Como compañía innovadora no podría permanecer ajena a la innovación basada en inteligencia artificial. En este ámbito, uno de sus productos más conocidos es **Siri**. Esta aplicación basada en procesamiento del lenguaje natural es capaz de interpretar las instrucciones (orales) del usuario para realizar diversas acciones como programar una alarma o realizar una llamada. Apple también trabaja intensamente en técnicas

de reconocimiento facial u optimización del *hardware* de los diferentes dispositivos móviles para ejecución óptima de algoritmos basados en inteligencia artificial.

IBM

Sin duda, **Watson** es uno de los productos estrella de IBM (relacionado además con la inteligencia artificial). Watson es una solución orientada a dar respuesta a preguntas formuladas mediante lenguaje natural. Es capaz de acumular, ordenar y clasificar información proveniente de diversas fuentes de información. En su diseño y desarrollo se aplican técnicas y metodologías procedentes de varios subdominios, procesamiento del lenguaje natural, representación del conocimiento, razonamiento automático, aprendizaje automático, etc. Está diseñado para usar IA a gran escala (similar a AWS, Azure AI, Google Cloud AI) y tiene diferentes divisiones especializadas en características concretas como Watson for business (la más importante para lo que ocupa este tema) Watson assistant o for medical donde ha tenido éxitos importantes.

IBM Watson, la solución de IA y sistemas cognitivos de IBM. Puedes acceder a este recurso pinchando en el siguiente enlace.

<https://www.ibm.com/watson/>

AlBrain

AlBrain es una empresa californiana **centrada en el desarrollo de soluciones basadas en inteligencia artificial para la resolución de problemáticas de diverso tipo**. Destacan sus soluciones para telefonía móvil de última generación y robótica. AlBrain está especializada en el desarrollo de soluciones cognitivas y con capacidad autónoma de aprendizaje.

Amazon

Amazon es uno de los grandes líderes tecnológicos mundiales. Sus algoritmos de recomendación son foco de atención continua por parte de la comunidad especializada y la competencia. Actualmente, **está centrado en ofrecer servicios relacionados con la inteligencia artificial a través de su oferta de servicios en la nube**. También prevé extender servicios de consultoría digital en este ámbito. Algunos de sus principales productos son:

- ▶ Amazon SageMaker: entrenamiento de modelos de *machine learning* (integración con Jupyter Notebook).
- ▶ Amazon Recognition: reconocimiento de imágenes.
- ▶ Amazon Polly: reconocimiento de voz.
- ▶ Amazon Lex: asistentes virtuales (Alexa).

También tiene una tienda que permite comprar modelos ya entrenados.

Amazon web services for machine learning. Puedes acceder a este recurso pinchando en el siguiente enlace.

<https://aws.amazon.com/es/machine-learning/>

Netflix

A la hora de hablar de algoritmos de recomendación, si existe una empresa que venga a la mente de los especialistas en este ámbito es Netflix. Netflix es una **empresa estadounidense de entretenimiento que oferta series, películas y documentales a través de una plataforma de *streaming***. Los algoritmos de recomendación implementados por Netflix ayudan al usuario a explorar la amplia diversidad de oferta existente sugiriendo productos acordes a sus preferencias y gustos.

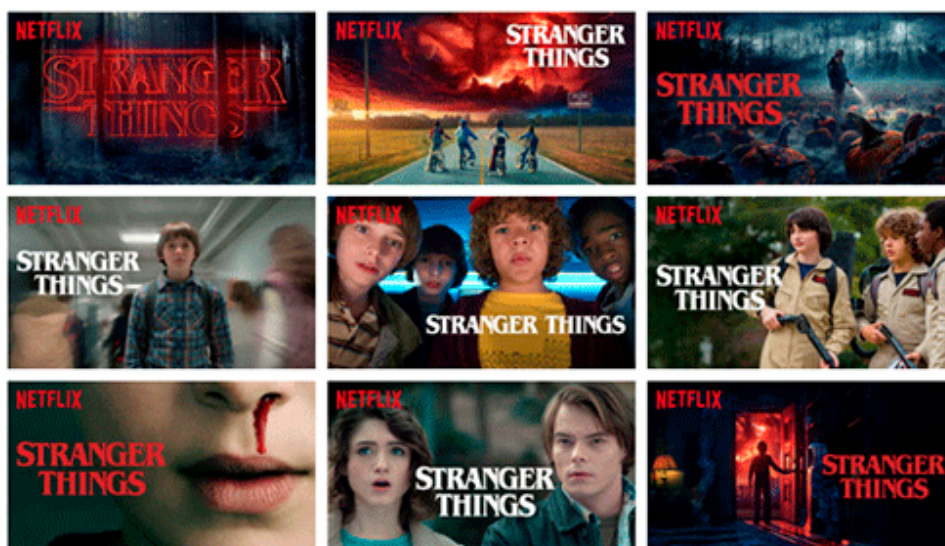


Figura 1. Recomendaciones en Netflix.

Fuente: <https://medium.com/netflix-techblog/artwork-personalization-c589f074ad76>

Otros actores importantes

Como se comentaba inicialmente, aunque se ha intentado dibujar un espectro amplio de actores en el ámbito de la inteligencia artificial identificando a los principales agentes, el listado no es exhaustivo. Un mapeo absolutamente detallado de este mercado no es objetivo de esta unidad, simplemente se pretende que el alumno esté informado sobre algunas de las que están consideradas como compañías de referencia que están desarrollando productos basados en inteligencia artificial.

El listado anterior puede enriquecerse mencionando a otras compañías que, aunque quizá no hayan alcanzado un grado de innovación tan destacado como las anteriores, se posicionan como **empresas punteras** que están realizando valiosas contribuciones al estado de la disciplina.

En el campo financiero, **la banca de inversión es gran consumidora y productora de soluciones basadas en inteligencia artificial**. Siendo más concretos y dentro del sector bancario español, destaca el **caso de BBVA**, que ha sido premiado por varias iniciativas realizadas en el campo del aprendizaje automático y ha sido reconocido por el Massachusetts Institute of Technology (MIT) como una de las entidades más innovadoras de la industria.

La compañía norteamericana **Salesforce** está ampliando su equipo tanto mediante la contratación externa como mediante la adquisición de pequeñas compañías innovadoras en el área. El objetivo de sus proyectos va desde mejorar la eficiencia de los procesos hasta mejorar las herramientas de ayuda automatizada para sus clientes.

Spotify, la empresa sueca que centra su actividad en la distribución de contenido musical vía *streaming*, es otra de las compañías innovadoras especialmente en el campo del aprendizaje automático y sistemas de recomendación. Debido al gran volumen de datos y usuarios también se posiciona como una compañía innovadora en lo que se refiere al tratamiento de grandes volúmenes de datos y optimización de los sistemas de información.

Por último, no podríamos dejar de mencionar a **Twitter**. La compañía norteamericana ha invertido ingentes cantidades económicas en la adquisición de varias compañías especializadas en inteligencia artificial. Su ámbito de actuación es amplio: desde proponer al usuario los tuits más relevantes en función de sus gustos, hasta mejorar sus algoritmos de recomendación de anuncios publicitarios.

2.3. Herramientas comerciales para inteligencia artificial



Accede al vídeo «Principales herramientas de desarrollo» a través del aula virtual

En este apartado pretendemos destacar las herramientas *software* más importantes que podemos utilizar para IA. De algunas ya hemos hecho mención en el apartado anterior, pero aquí nos centraremos en sus principales bondades y añadimos algunas otras que no hemos citado anteriormente.

TensorFlow

Es una **biblioteca de código abierto para aprendizaje automático** y que dispone de multitud de modelos de redes profundas. Está desarrollado por Google para satisfacer sus necesidades de sistemas capaces de construir y entrenar redes neuronales para detectar y descifrar patrones y correlaciones, análogos al aprendizaje y razonamiento usados por los humanos. Esta biblioteca se centra en *deep learning*, pero ha ido incluyendo otros algoritmos con el tiempo. Está integrada con los servidores de Google, con lo cual podemos de forma fácil lanzar los algoritmos en la nube de Google. Así que una de sus principales bondades es que permite un fácil despliegue de las aplicaciones. También tiene acceso a las GPU (*graphics processing unit*) instaladas a nivel local, con lo que podemos aprovecharnos de las capacidades de computación de estas tarjetas para acelerar los cálculos. Tiene una interfaz de uso en Python, aunque muchos de los algoritmos internamente están desarrollados en C++.

Scikit-learn

Scikit-learn **lidera las librerías de *machine learning* en Python** al estar construido a su vez sobre las mejores librerías para matemáticas y ciencias (NumPy, SciPy y Matplotlib). La librería está disponible bajo licencia BSD, por lo tanto, es completamente abierto y reutilizable e implementa un gran conjunto de algoritmos de *machine learning*. Casi cualquier algoritmo conocido está disponible en la librería y no se centra solo en *deep learning* como TensorFlow. Es además muy fácil de usar e incorpora algunos *datasets* para poder hacer pruebas con los diferentes algoritmos.

Universe

Universe es una **API open source de OpenAI**. Su principal reclamo es que es totalmente accesible al público. Sirve para medir y entrenar algoritmos de *machine learning* de propósito general. Y proporciona una sencilla interfaz de comunicación con la plataforma Gym de OpenAI (<https://github.com/openai/gym>) Dicha plataforma es un conjunto de herramientas para desarrollar y comparar algoritmos de aprendizaje de refuerzo.

Universe permite a cualquiera entrenar y evaluar agentes de IA en un rango extremadamente amplio de entornos complejos en tiempo real, haciendo posible que cualquier programa existente se convierta en un entorno válido para OpenAI Gym, sin necesidad de un acceso a dicho entorno. Lo consigue empaquetando el programa en un contenedor Docker, y presentando la IA con la misma interfaz que usa un humano, es decir, enviando eventos de teclado y ratón, y recibiendo píxeles de pantalla: <https://github.com/openai/universe>

H2O

Plataforma de aprendizaje automático de código abierto para el análisis predictivo de *big data*, scoring de datos y modelado de datos. Tiene diferentes integraciones con otras librerías como Spark y es muy usada para análisis de datos. Aseguran que buscan una IA responsable y fomentan las buenas prácticas favoreciendo cosas como la explicabilidad, la comprensibilidad o la IA centrada en las personas.

Apache Spark

Apache Spark es un *framework* de programación para procesamiento de datos distribuidos, tiene muchos algoritmos y tipos de datos útiles (basado en Apache Hadoop). Diseñado para ser rápido y escalable. Aunque Java es el primer lenguaje

para trabajar con ella, los usuarios de Python pueden conectar con NumPy. Está destinada a *big data* principalmente.

Accord.NET

Librería para *machine learning* y procesamiento de señales hecha para .NET, es una extensión del proyecto AForget.NET. Accord incluye un conjunto de librerías para procesamiento de audio y flujo de imágenes (vídeos). Este algoritmo puede ser utilizado para detección de rostros o para seguimiento del movimiento de objetos.

Keras

Biblioteca de red neuronal de código abierto, escrita en Python, que admite redes recurrentes y redes convolucionales. Está construido sobre TensorFlow y su objetivo es hacerlo más usable. Presume de ser usado en algunas de las principales organizaciones científicas de todo el mundo como el CERN, NASA, NIH. Lo cual la hace una librería con mucho apoyo a nivel académico.

NGC

La plataforma acelerada por GPU **permite a los científicos de datos administrar datos**, así como también diseñar y capacitar modelos de redes neuronales. Está desarrollado por Nvidia para su uso en sus propias gpus para abstraer a los desarrolladores de la complejidad de acceso a la misma.

Caffe

Marco de aprendizaje profundo de código abierto que brinda asistencia para GPU y CPU y que está desarrollado en la universidad de Berkeley en la división de AI Research. Perfecto para investigaciones académicas y aplicaciones industriales.

Presume de poder configurarse sin programar mediante ficheros de configuración, así como ser muy extensible y rápido.

2.4. Tecnologías comerciales de inteligencia artificial



Accede al vídeo «Soluciones basadas en inteligencia artificial» a través del aula virtual



Accede al vídeo «Sectores y proveedores de servicios de inteligencia artificial» a través del aula virtual

En este apartado pretendemos destacar **otras soluciones comerciales basadas en inteligencia artificial**, y que son menos conocidas que los productos ofrecidos por las grandes compañías o, al menos, menos conocidas comercialmente que las comentadas en el apartado anterior. Enumeraremos ante todo productos ofrecidos por *startups* o pequeñas empresas. Dividiremos las distintas propuestas teniendo en cuenta la industria en la que operan. De nuevo, no se pretende crear un listado exhaustivo sino, simplemente, destacar casos de uso que se han considerado relevantes. Según el caso, se detallará de forma individual cada compañía o por el contrario se proporcionará una descripción genérica del sector enumerando destacadas compañías de este.

Marketing, ventas y CRM

En este apartado, se hace referencia a compañías que principalmente operan en el sector de la publicidad, gestión de ventas, *marketing* o *customer relationship management* (CRM).

- ▶ DrawBrigde: compañía californiana fundada en el año 2011. Permite conectar a personas operando a través de dispositivos diferentes para ofrecer anuncios personalizados.
- ▶ Appier: compañía con sede central en Taipéi (Taiwán). Su producto estrella es CrossXAI y, al igual que el caso anterior, uno de sus objetivos es enlazar los dispositivos asociados al mismo usuario.
- ▶ Persado: compañía con localizaciones en Estados Unidos y Europa, que centra su trabajo en la personalización del lenguaje empleado en las campañas comerciales con el fin de obtener la mejor reacción por parte del potencial cliente.
- ▶ InsideSales.com: compañía estadounidense especializada en la creación de distintas soluciones tecnológicas con la misión de apoyar la actividad de los equipos de venta comercial.
- ▶ BloomReach: ofrece el producto BloomReach DXP, consistente en una plataforma abierta e inteligente focalizada en el desarrollo, personalización y pruebas de experiencias digitales, ayudando a sus clientes a identificar en tiempo real la estrategia digital más adecuada a cada momento. Está presente en EE. UU., Europa y Asia.

Tecnología para el sector automovilístico

En lo que respecta a compañías centradas en la producción de soluciones tecnológicas para el sector de la automoción, podemos citar las siguientes:

- ▶ Nutonomy: empresa nacida en el seno del MIT, centrada en el diseño de coches autónomos.
- ▶ Nauto: compañía radicada en Silicon Valley centrada en mejorar la seguridad en el transporte rodado. Sus soluciones suelen basarse en el empleo de cámaras y análisis de imágenes para evitar accidentes de tráfico.

Finalizamos este apartado destacando el trabajo de la compañía automovilística **Tesla**. Esta compañía norteamericana está en la lista de las empresas llamadas a liderar la producción de los futuros coches autónomos.

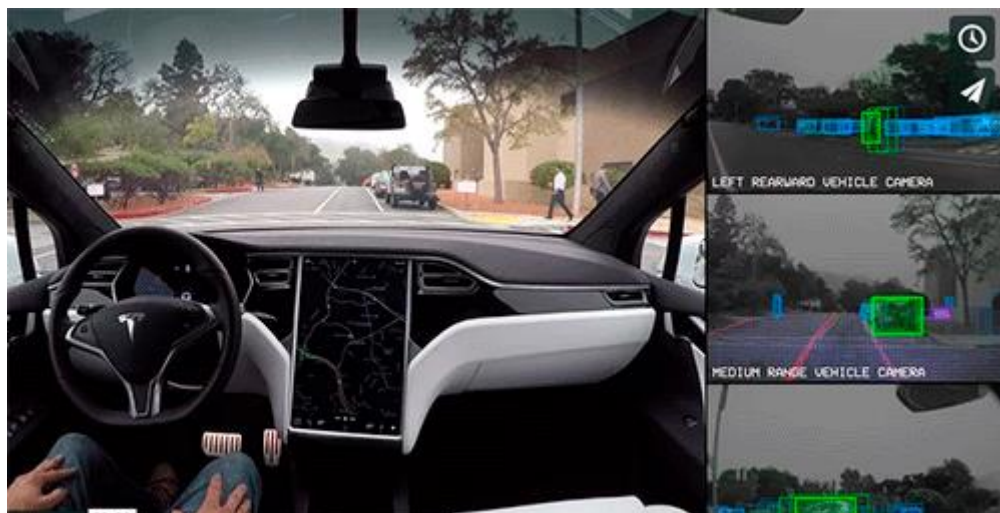


Figura 2. Coche autónomo Tesla.

Fuente: <https://www.tesla.com/autopilot?redirect=no>

Tecnología de *business intelligence* y analítica genérica

Los sistemas de *business intelligence* se apoyan en la tecnología para analizar la información de que dispone la compañía con vistas a apoyar la toma de decisiones. Se listan a continuación algunas de las compañías destacadas:

- ▶ Logz.io: compañía israelí que se dedica a analizar el inmenso volumen de información generado por los sistemas informacionales (*logs*). A partir de dichos análisis se ofrecen soluciones de cara a mejorar la eficiencia de los procesos y evitar problemas en los sistemas.
- ▶ CrowdFlower: conocida compañía ubicada en San Francisco (EE. UU.). Ofrece soluciones que permiten enriquecer con información adicional los datos a nuestra disposición para dotarlos de mayores posibilidades.
- ▶ RapidMiner: compañía con sedes en Europa y EE. UU. RapidMiner ofrece una plataforma del mismo nombre que permite ejecutar el flujo completo de análisis de datos y aplicación de técnicas de aprendizaje automático, *deep learning*,

análisis de texto o análisis predictivo con fines varios. Esta solución se ofrece a varias industrias y sectores.

- ▶ Tamr: compañía con delegaciones en EE. UU., Europa y Asia, centrada en lograr la integración de datos procedentes de varias fuentes para dotar así al análisis de mayor potencia.
- ▶ Versive: pequeña compañía afincada en Seattle desde el año 2012. Versive tiene en su foco la automatización de procesos de negocio, captura de conocimiento experto para la toma de decisiones y el desarrollo de modelos basados en aprendizaje automático.
- ▶ Datarobot: compañía radicada en Boston que ofrece a sus clientes una plataforma que permite cargar los datos del cliente y aplicar algoritmos de aprendizaje automático obteniendo modelos analíticos que pueden aplicarse en el día a día.
- ▶ Paxata: empresa californiana nacida en el año 2012. Paxata ofrece un servicio que permite agregar, limpiar y transformar datos procedentes de diversas fuentes con el fin de facilitar el tratamiento posterior de estos para la aplicación de técnicas de aprendizaje automático.
- ▶ Trifacta: compañía establecida en San Francisco con un objetivo parecido a Paxata. Esta empresa ofrece un producto que permite explorar, transformar y enriquecer los datos para su posterior análisis.
- ▶ Dataminr: se trata de una empresa neoyorquina centrada en la monitorización de redes sociales y otras fuentes de datos con el fin de proporcionar las alertas adecuadas en tiempo real a negocios de diversos sectores.

Soluciones conversacionales, *bots*, lenguaje natural

En el mundo de los asistentes personales que intentan interactuar con el usuario empleando técnicas de procesamiento del lenguaje natural encontramos diversas opciones.

Mindmeld, Nice, Nuance, OpenText, Mobvoi y X.ai son conocidos casos de este ámbito. Mobvoi está centrada en el idioma chino. Nuance y OpenText abarcan

también el procesamiento de información documental y en formatos variados. Por el contrario, X.ai se adentra en el terreno de los gestores personales.



Figura 3. Ejemplo chatbot para la adquisición de una pizza.

Fuente: <https://www.theverge.com/2016/7/13/12170682/pizza-hut-chatbot-catch-up>

Plataformas y soluciones de *hardware*

Diversas empresas, entre ellas algunas de las ya mencionadas, **se dedican a proporcionar algoritmos, herramientas, datos, utilidades y acceso a sistemas e infraestructuras de *hardware* de distinto tipo** con el fin de que otras compañías (sus clientes) puedan emplear estos recursos para analizar sus datos, añadir fuentes de datos adicionales y desarrollar modelos de distintos tipos que aporten valor a las empresas usuarias. Entre las compañías que ofrecen este tipo de servicios encontramos a Amazon, Fractal Analytics, Google, H2O.ai, Microsoft, SAS, Skytree, etc.

De forma más concreta, debemos citar también a empresas que participan en la fabricación y diseño de componentes de *hardware* centrados en ejecutar

aplicaciones que requieren alta capacidad de cómputo, muchos de ellos pensados para entrenar algoritmos de aprendizaje automático. Un ejemplo clásico de este tipo de piezas son las tarjetas gráficas (GPU por sus siglas en inglés: Graphics Processing Unit) de altas prestaciones. Ejemplos de compañías que abarcan este tipo de productos son Alluviate, Cray, Google, IBM, Intel y, como no, Nvidia.

Dentro de las soluciones basadas en reglas y razonamiento lógico para la gestión de procesos de negocio y apoyo a la toma de decisiones, destacan las soluciones propuestas por compañías como Advanced Systems Concepts, Informatica, Maana, Pegasystems o UiPath.

Un sector que genera una cifra de negocio importante es el **sector de la biometría**, centrado en la identificación de las personas basándose en atributos diferenciales. Empresas destacadas del ámbito son: 3VR, Afectiva, Agnitio, FaceFirst, Sensory, Synqera oTahzoo.

Core AI-Aplicaciones genéricas

Existen compañías difíciles de etiquetar dentro de un sector concreto debido a la diversidad de servicios que ofrecen. Se proporciona a continuación un listado de las compañías susceptibles de ser agrupadas en esta categoría.

Numenta

Empresa afincada en California que cuenta entre sus filas con expertos muy reconocidos en el ámbito de la inteligencia artificial. Produce tanto soluciones propietarias como soluciones abiertas a la comunidad. Sus productos abarcan diversos casos de uso, desde análisis de valores bursátiles hasta procesamiento del lenguaje natural.

Cognitivescale

Reconocida compañía estadounidense que ofrece soluciones en todos los ámbitos bajo el concepto de inteligencia aumentada. Su concepto principal es crear productos no orientados a reemplazar el personal humano sino a dotarlo de herramientas que le permitan hacer mejor su trabajo.

Afectiva

Esta compañía ramifica de trabajos previos realizados en el MIT (EE. UU.). Su foco principal es la medición y reconocimiento de las emociones humanas para su explotación con fines comerciales.

Sentient Technologies

Compañía norteamericana que ha generado gran interés entre fondos de inversión varios. Actúa principalmente en dos ámbitos que no parecen tener relación directa, el comercio electrónico y operaciones de inversión en los mercados financieros. Sentient centra su actividad en la creación de sistemas de recomendación que tienen en cuenta el perfil del usuario para proponer la mejor recomendación o producto. Dicho producto varía dependiendo del sector, puede ser un producto financiero o no.

Digital Reasoning

Esta compañía estadounidense centra su actividad principal en la aplicación de la inteligencia artificial para apoyar la labor de servicios de inteligencia varios e instituciones financieras.

Voyager Labs

Compañía con sedes en EE. UU. e Israel. Se trata de una compañía pionera en soluciones basadas en inteligencia cognitiva. Aunque opera en diversos sectores, es quizá en el sector comercial (incluyendo comercio electrónico) donde sus soluciones están más extendidas.

Ciberseguridad

El sector de la seguridad informática o ciberseguridad está llamado a experimentar una gran transformación en los próximos años. El contexto político y tecnológico obliga a ello. Las **estructuras informáticas son, desde hace ya muchos años, parte esencial de la infraestructura de un país**. La inteligencia artificial es un pilar esencial para proteger de forma eficiente dichos recursos. Compañías conocidas que operan en este ámbito son Sift Science, Darktrace o Cylance entre otras.

Fintech

Fintech (Financial Technology) **hace referencia a un paradigma de negocio centrado en la aplicación de la tecnología para ofrecer y gestionar productos financieros de forma innovadora**, ágil y rompiendo las barreras y los procedimientos estandarizados típicos de las grandes corporaciones bancarias. Son múltiples y diversas las opciones en este ámbito. Algunas que se pueden citar a modo de ejemplo son: Alphasense, Kensho, Due.com, SoFi, LendUp, LendFriend, WePay, Lending Club, etc.

Salud y sanidad

Las posibilidades que se abren en el sector sanitario en relación con la explotación de grandes volúmenes de datos y aplicación de técnicas de inteligencia artificial son infinitas. Los focos principales de actuación son la personalización de los tratamientos clínicos, soporte para la toma de decisiones, optimización de los procesos

hospitalarios, análisis predictivo de la evolución de la enfermedad, diagnóstico por imagen, soporte clínico virtual, etc. Algunas de las compañías que están liderando este sector son: Zebra Medical Vision, Babylon Health, Benevolent.ai o Icarbonx.

Internet de las cosas

El internet de las cosas (Internet of Things o IoT) hace referencia al **ecosistema que se genera con la interconexión de dispositivos electrónicos varios a través de las redes de comunicaciones**. Dicha interconexión permite integrar fuentes de datos diversas, contextualizar la información y tomar decisiones en tiempo real.

Este paradigma se puede aplicar a una gran diversidad de ámbitos, desde la gestión de las ciudades impulsando el concepto de «ciudad inteligente», hasta la agricultura, pasando por sectores tan diversos como la gestión integral de fábricas, gestión y optimización de los dispositivos del hogar, eficiencia energética, logística y un largo etcétera. Compañías con aportaciones interesantes en este ámbito son Verdigris Technologies y Sight Machine.

Robótica

El sector industrial y empresarial asociado al mundo de la robótica está ofreciendo importantes avances en los últimos años. A modo de ejemplo, se ha hecho referencia con anterioridad a Sophia, producto de la compañía Hanson Robotics.

La **industria contempla el uso de robots en multitud de contextos y ámbitos**, asistencia en el hogar, asistencia a personas enfermas o con algún tipo de discapacidad, producción industrial, exploración espacial, actuación ante emergencias (como en un accidente nuclear), intervención policial o militar, entretenimiento, etc. Algunas empresas destacadas en este ámbito son: Rokid, Ubtechrobotics, Anki, Vicarious...

Narrativa y análisis semántico

Otro de los campos en auge dentro de la inteligencia artificial, **intenta transformar los datos en una narración con sentido para comunicar de forma automática ideas, noticias o hechos relevantes**. Este tipo de soluciones se aplican en diversos sectores, algunos ejemplos son el sector editorial, periódicos, noticias..., pero también puede aplicarse al ámbito empresarial transmitiendo historias fáciles de entender y compartir en lugar de una avalancha de datos de complicada interpretación.

Narrative Science es una de las compañías más activas en este campo.

Cambridge Semantic es otra de las compañías que podemos citar en este sector. Esta compañía norteamericana está focalizada en añadir a la capa de datos una capa adicional de conocimiento que permite establecer relaciones entre datos y dota de mayor valor a estos.



Figura 4. Página web de Narrative Science.

Fuente: <https://narrativescience.com/>

Visión artificial

La visión artificial **hace referencia a la disciplina científica que desarrolla metodologías y algoritmos que permiten adquirir, procesar, analizar e identificar imágenes de distinto tipo con fines varios.** Este tipo de conceptos pueden aplicarse a gran diversidad de campos, tales como seguridad, diagnóstico por imagen, coches autónomos, robótica, etc. A las empresas que hayan podido ser citadas en algún apartado anterior podemos añadir otras como Chronocam, Orbital insight, Clarifai o Captricity.

Biología, agricultura y medioambiente

Finalizamos este apartado haciendo referencia a casos de usos particulares que quizá resulten menos conocidos para la comunidad no especializada. **Las técnicas de inteligencia artificial, y más concretamente aprendizaje automático, pueden ser aplicadas al ámbito de la biología y análisis de moléculas,** virus o bacterias para conseguir procesos más eficientes en la fabricación de medicamentos, fertilizantes o productos empleados en nuestra vida cotidiana. **Zymergen** es una compañía estadounidense volcada en el desarrollo de este tipo de productos.

Haciendo hincapié en el sector de la agricultura, citaremos a **Blue River Technology**. Esta compañía emplea robótica y visión artificial para controlar y optimizar la gestión de explotaciones agrícolas. Al tratarse de un campo de actuación particular se ha creído oportuno incluir a Blue River en este apartado en lugar de otros previos.

2.5. El futuro de la inteligencia artificial



Accede al vídeo «El futuro de la inteligencia artificial» a través del aula virtual

Llegados a este punto, tenemos una visión amplia de sectores en los que opera la inteligencia artificial y tipos de soluciones que se están ofreciendo. Resta ahora centrarnos en las posibilidades de futuro.

Antes de nada, y a modo de «renuncia de responsabilidad» (*disclaimer*), debemos insistir en que no es posible determinar con certeza (y en opinión del autor ni siquiera con una precisión elevada) el futuro de la inteligencia artificial.

Cualquier descubrimiento puntual puede abrir un espectro totalmente inesperado o generar una nueva línea de productos. Fuera de este apartado queda el contexto regulatorio y legal que, especialmente en Europa, puede influir en el desarrollo de nuevos productos basados en inteligencia artificial en general y aprendizaje automático en particular.

Respecto a los sectores donde la industria centrará sus esfuerzos de cara al futuro, sin duda, la **generación de vehículos autónomos** de todo tipo será uno de los grandes caballos de batalla para las próximas décadas.

Otro interesante campo de acción es la traducción automática y en tiempo real. Google ha lanzado los auriculares inalámbricos Pixel Buds, producto que pretende convertirse en un referente de la **traducción instantánea**.

La fusión de *big data*, internet de las cosas e inteligencia artificial permitirá evolucionar la gestión de las ciudades, permitiendo una gestión ágil de los recursos y espacios públicos en función de la actividad diaria de la ciudad.

Como se ha comentado en el apartado anterior, también se abre un abanico sin fin de posibilidades en el ámbito sanitario. Lograr una atención clínica auténticamente personalizada al paciente es un reto por afrontar. La **medicina personalizada** considerará datos comportamentales, genéticos, sanitarios, etc., con el fin de

proporcionar el mejor servicio que optimice el sistema preventivo, minimice la posibilidad de reincidencia y maximice las probabilidades de sanación.

La educación no puede permanecer ajena a la evolución de la inteligencia artificial, disciplina que posibilita mejorar brindar una auténtica **educación personalizada** para todos, de forma que pueda conseguir un aprendizaje efectivo ajustado a su perfil.

El sector comercial con sistemas de recomendación que se ajustan a nuestras preferencias, **supermercados inteligentes** que nos facilitarán la compra empleando el mínimo de personal, integración de todo tipo de dispositivo en la cadena de atención al cliente..., sufrirá una transformación difícil de predecir.

La industria financiera y aseguradora, el sector energético, minería, exploración espacial, seguridad, agricultura, ecología y medio ambiente, logística, etc., no hay campo del dominio humano que no se vaya a ver afectado por la evolución que marcará la inteligencia artificial.



Accede a los ejercicios de autoevaluación a través del aula virtual

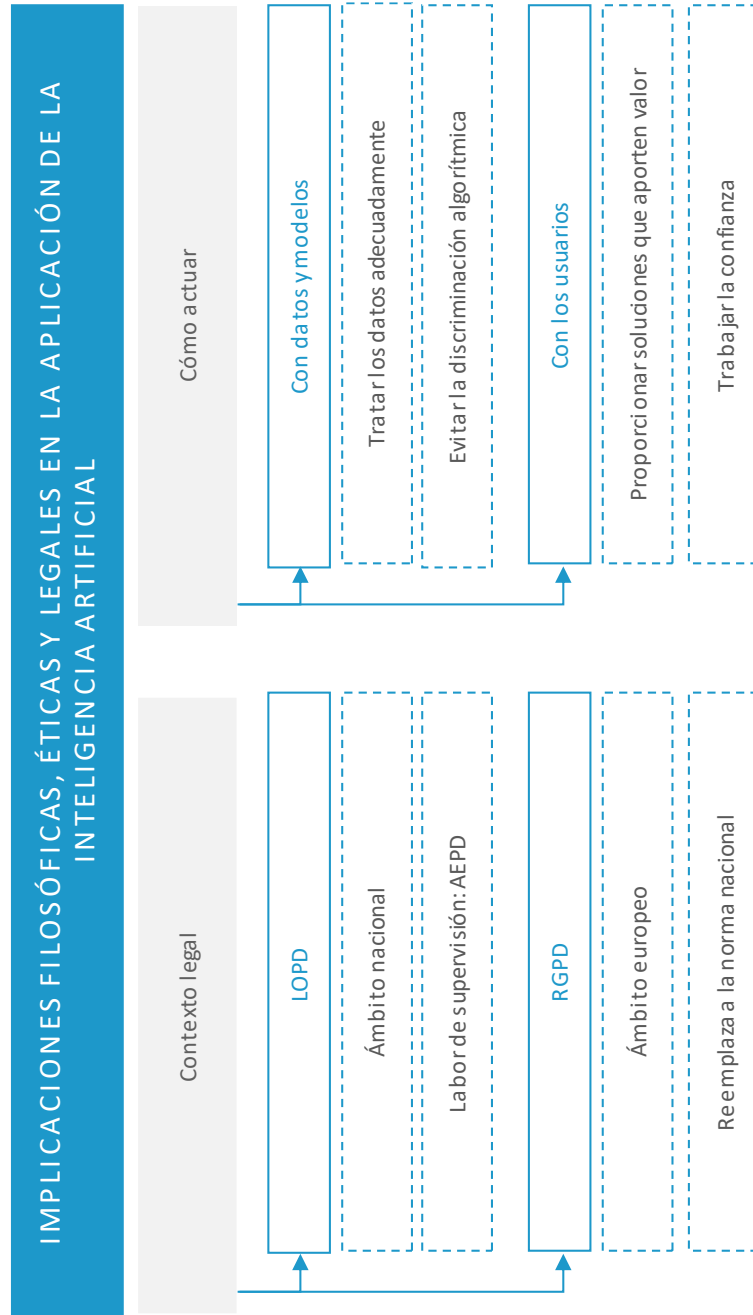


Accede al vídeo «Resumen del tema» a través del aula virtual

Inteligencia Artificial y Computación Cognitiva

Implicaciones filosóficas, éticas y legales en la aplicación de la inteligencia artificial

Esquema. Tema 3



Ideas clave. Tema 3

3.1. Introducción y objetivos



Accede al vídeo «Introducción. Implicaciones filosóficas, éticas y legales en la aplicación de la inteligencia artificial» a través del aula virtual

La aplicación de la inteligencia artificial está unida no solo a retos técnicos o científicos sino también a **retos éticos y legales** propios de una disciplina que está llamada a cambiar notablemente la forma en la que interactuamos con los demás, los métodos de trabajo y toda la actividad económica. En esta unidad nos planteamos los objetivos siguientes:

- ▶ Identificar retos éticos y filosóficos en el empleo de la IA (inteligencia artificial).
- ▶ Conocer las peculiaridades legales del trabajo basado en datos.
- ▶ Identificar nuevas tendencias regulatorias que implican cambios a la hora de trabajar con datos.

3.2. Contexto legal aplicable a proyectos de inteligencia artificial



Accede al vídeo «Contexto legal en inteligencia artificial» a través del aula virtual



Accede al vídeo «Protección de datos en proyectos de inteligencia artificial» a través del aula virtual

Los proyectos de inteligencia artificial **están basados en datos**. Necesitamos datos para entrenar y para validar los modelos. También necesitamos almacenar y analizar

los datos que genera la interacción con el modelo, solo así seremos capaces de determinar si la puesta en práctica de la solución cumple con los objetivos previstos. La evaluación constante de los modelos es lo que permite su mejora y evolución.

Por tanto, los proyectos de inteligencia artificial se ven claramente influenciados por la legislación vigente en materia de protección de datos.

En México, el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI) tiene como misión garantizar en el Estado mexicano los derechos de las personas a la información pública y a la protección de sus datos personales, así como promover una cultura de transparencia, rendición de cuentas y debido tratamiento de datos personales para el fortalecimiento de una sociedad incluyente y participativa.

El Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI) es el organismo constitucional autónomo garante del cumplimiento de dos derechos fundamentales: el de acceso a la información pública y el de protección de datos personales.

En relación al segundo derecho, es **garantizar el uso adecuado de los datos personales, así como el ejercicio y tutela de los derechos de acceso, rectificación, cancelación y oposición** que toda persona tiene con respecto a su información.



Para dar cumplimiento al derecho fundamental de protección de los datos personales surgen tres grandes leyes:

- ▶ Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPO), publicada en el Diario Oficial de la Federación el 26 de enero de 2017.
- ▶ Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP), publicada en el Diario Oficial de la Federación el 05 de julio de 2010.
- ▶ Reglamento de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (RLFPDPPP) publicado en el Diario Oficial de la Federación el 21 de diciembre del 2011

A continuación, se desglosarán **algunas de las características más destacadas de estas normativas**, así como sus implicaciones para la tipología de los proyectos que nos ocupa.

Protección de datos de carácter personal

Se considera un dato de carácter personal **cualquier información concerniente a una persona física identificada o identificable**. Por tanto, la legislación no solo contempla como datos de carácter personal aquellos casos en los que es posible identificar unívocamente al individuo en base a un atributo exclusivo como la credencial de elector que emite el Instituto Nacional Electoral (INE). También contempla la posibilidad de que con dicha identificación pueda producirse en el futuro vía cruce de datos de distintas tipologías.

Por ejemplo, supongamos que tenemos almacenada en un archivo o base de datos información sobre los clientes de una determinada compañía. Supongamos ahora que contamos con campos poco relevantes y genéricos. En esta situación, la compañía decide añadir la edad de cada usuario a la información existente. Si los datos anteriores realmente son irrelevantes (siempre referido a cuestiones de identificación personal), la edad por sí sola

no podría servir para identificar a un individuo, puesto que, ¿cuántas personas hay de 20 años?, ¿y de 35? Las opciones son muy diversas. En estas condiciones, la compañía decide incorporar a la base de datos el género. Ahora las opciones se reducen, evidentemente, hay más personas que tienen 35 años a personas que tienen 35 años y son mujeres (por ejemplo). No obstante, parece poco creíble pensar que solo existe una única persona que tenga 35 años y sea mujer. Animados por lo útil que es insertar este tipo de atributos en el estudio, el responsable del proyecto decide ahora incorporar el código postal del usuario. En esta situación, y si la compañía dispone de especialistas en seguridad y normativa de protección de datos, se levantará una señal de alarma. Con la edad, género y código postal sí que existe una mínima posibilidad de que se pueda identificar de forma unívoca a una persona (especialmente en grupos de edades menos frecuente). En diversos entornos de la materia, a la combinación de los atributos edad, género y código postal se le conoce con el nombre de triada.

Antes de continuar, queremos dejar patente que no se prohíbe el almacenamiento de información personal. Si así fuese, ninguna empresa o institución pública podría funcionar. El fin de la ley es regular su tratamiento legítimo controlado e informado, a efecto de garantizar la privacidad y el derecho a la autodeterminación informática de las personas, aplicando las medidas de seguridad para la protección de los datos.

En la legislación mexicana, se identifican **datos personales** y **datos personales sensibles**. Estos últimos son aquellos datos personales que afectan a la esfera más íntima de su titular, o cuya utilización indebida pueda dar origen a discriminación o conlleve un riesgo grave para este. Por ejemplo, aspectos como origen racial o étnico, estado de salud presente y futuro, información genética, creencias religiosas, filosóficas y morales, afiliación sindical, opiniones políticas, preferencia sexual, entre otros.

Las distintas iniciativas regulatorias que van surgiendo se encaminan a posibilitar a los usuarios finales o clientes un mayor control sobre los datos que ceden a las compañías y el uso que estas hacen de estos datos. En el artículo 6 de la LFPDPPP en materia de tratamiento de datos personales, se abordan los principios que sirven de inspiración para el manejo legal de los datos:

«Los responsables en el tratamiento de datos personales deberán observar los principios de licitud, consentimiento, información, calidad, finalidad, lealtad, proporcionalidad y responsabilidad».

En el reglamento explica los principios definidos en la ley y son los siguientes:

»Principio de licitud

»Artículo 10. El principio de licitud obliga al responsable a que el tratamiento sea con apego y cumplimiento a lo dispuesto por la legislación mexicana y el derecho internacional.

»Principio de consentimiento

»Artículo 11. El responsable deberá obtener el consentimiento para el tratamiento de los datos personales, a menos que no sea exigible con arreglo a lo previsto en el artículo 10 de la Ley. La solicitud del consentimiento deberá ir referida a una finalidad o finalidades determinadas, previstas en el aviso de privacidad. Este consentimiento puede ser tácito o expreso.

»Principio de información

»Artículo 23. El responsable deberá dar a conocer al titular la información relativa a la existencia y características principales del tratamiento a que serán sometidos sus datos personales a través del aviso de privacidad, de conformidad con lo previsto en la Ley y el presente Reglamento.

»Principio de calidad

»Artículo 36. Se cumple con el principio de calidad cuando los datos personales tratados sean exactos, completos, pertinentes, correctos y actualizados según se requiera para el cumplimiento de la finalidad para la cual son tratados.

»Principio de finalidad

»Artículo 40. Los datos personales solo podrán ser tratados para el cumplimiento de la finalidad o finalidades establecidas en el aviso de privacidad, en términos del artículo 12 de la Ley.

»Principio de lealtad

»Artículo 44. El principio de lealtad establece la obligación de tratar los datos personales privilegiando la protección de los intereses del titular y la expectativa razonable de privacidad, en los términos establecidos en el artículo 7 de la Ley.

»Principio de proporcionalidad

»Artículo 45. Solo podrán ser objeto de tratamiento los datos personales que resulten necesarios, adecuados y relevantes en relación con las finalidades para las que se hayan obtenido.

»**Principio de responsabilidad**

»Artículo 47. En términos de los artículos 6 y 14 de la Ley, el responsable tiene la obligación de velar y responder por el tratamiento de los datos personales que se encuentren bajo su custodia o posesión, o por aquellos que haya comunicado a un encargado, ya sea que este último se encuentre o no en territorio mexicano».

Es preciso, que **el propietario de los datos conozca y autorice el aviso de privacidad para su almacenamiento y tratamiento de su información personal**. Así mismo, los responsables del manejo de los datos deben garantizar a los usuarios los métodos adecuados para la rectificación, modificación, acceso y cancelación de la información y contar con las medidas de seguridad para la protección de datos personales.

Aviso de privacidad

De conformidad con la LFPDPPP, señala que el **aviso de privacidad** es el documento físico, electrónico o en cualquier otro formato generado por el responsable que es puesto a disposición del titular, previo al tratamiento de sus datos personales, de conformidad con el artículo 15 de dicha Ley:

«El responsable tendrá la obligación de informar a los titulares de los datos, la información que se recaba de ellos y con qué fines, a través del aviso de privacidad».

Es importante reiterar que el objetivo principal del aviso de privacidad es dar a conocer al titular de los datos cuáles son los datos que se solicitan, el tratamiento que se le darán, el tiempo de conservación, el procedimiento para el ejercicio de sus derechos ARCO, disponibilidad del aviso de privacidad, entre otros.

Confianza

Dentro de ciertos sectores sociales existe una preocupación relevante por el uso que determinadas compañías están haciendo con sus datos, y quizá esta preocupación esté justificada si nos basamos en ciertos titulares periodísticos que se han publicado en los últimos años. En este tipo de cuestiones entraremos más en detalle en el próximo apartado. Ahora intentaremos explicar cómo se puede establecer una relación de confianza entre el usuario y la organización para que este ceda sus datos.

Supongamos que un paciente estándar de 50 años está en la revisión anual con su médico. El hospital está interesado en crear un banco de datos y compartirlos con personal externo. Dado que es obligatorio otorgar permiso explícito para ceder datos de carácter personal, el facultativo nos pone delante un formulario donde simplemente se indica el texto: «autorizo la cesión de mis datos ...».

Sin disponer de más información, ¿qué haría el lector? Cambiemos la situación yendo a un caso algo más extremo.

Supongamos que ese mismo paciente ha sido diagnosticado recientemente de una enfermedad que tiene una incidencia muy baja, pero, desgraciadamente, sus efectos en el enfermo son importantes. Textualmente, lo que nos comenta ahora el médico es:

«Nos gustaría que nos cediese sus datos personales y clínicos para incorporarlos a nuestro banco de datos. Lo que pretendemos con este proyecto es compartir información con especialistas de todo el país e investigadores muy destacados en el área. Dado que su enfermedad es extremadamente rara, consideramos que sería muy útil compartir sus datos con la comunidad científica. Seguro que así podremos recibir recomendaciones que nos ayudarían a personalizar el tratamiento a sus características. Además, no compartiremos sus datos personales más íntimos, toda la información estará anonimizada. Ante cualquier cuestión, seremos nosotros quienes nos comunicaremos con usted, como venimos haciendo hasta ahora».

¿Qué haría el lector en esta hipotética situación?

Es probable que las respuestas ante ambos casos de uso sean muy distintas. Y sería normal, puesto que se trata de dos casos muy diferentes. En la primera situación, el

paciente no aprecia el retorno de valor que le produce la cesión de sus datos. Además, al recibir poca información sobre el proyecto es difícil generar una relación de confianza. En el segundo caso, el beneficio que el paciente percibe como consecuencia de la cesión de datos es evidente y relevante. La comunidad podría aportar un tratamiento que retarde la enfermedad o incluso la cure. Pero, para que esto sea una realidad, es preciso que especialistas compartan información sobre qué tratamientos han funcionado, cuáles no, cuáles han sido los efectos secundarios, qué peculiaridades tenían los pacientes... Además, el facultativo ha trabajado la confianza al solicitar la cesión de los datos, ha explicado la utilidad de la cesión de los datos y ha establecido unas garantías que invitan al paciente a estar tranquilo sobre quién podrá contactarle. ¿Cuántas veces hemos recibido llamadas publicitarias ofreciendo productos sin ningún tipo de interés para nosotros y sin saber cómo habían accedido a nuestros datos?

Lo que se ha pretendido explicar en los párrafos anteriores se puede resumir en los siguientes puntos:

- ▶ Los algoritmos de inteligencia artificial (por ejemplo, los algoritmos de aprendizaje automático) emplean datos como materia prima.
- ▶ Muchas veces estos datos incluyen información personal.
- ▶ Para manipular información personal (aunque esté anonimizada) el usuario debe haber emitido su autorización.
- ▶ Para que el usuario ceda sus datos personales es preciso trabajar un entorno de confianza. Este entorno de confianza se basa en los siguientes puntos:
 - El usuario tiene claro a quién (a nivel de institución) y para qué va a acceder los datos.
 - El usuario tiene claro qué se va a hacer con sus datos.
 - El usuario percibe un retorno de valor evidente que le invita a permitir la cesión de datos.
 - El usuario es informado de que sus datos van a ser tratados de forma adecuada y segura.

Establecer mecanismos que cultiven e implementen esta relación de confianza entre el usuario o cliente y la empresa o la institución es uno de los grandes retos a los que se enfrenta la industria hoy en día.

Disociación

Dentro de la legislación mexicana en su Ley Federal de Protección de Datos Personales en Posesión de Particulares (LFPDPPP) en su artículo 3, fracción VIII señala:

«Disociación es el procedimiento mediante el cual los datos personales no pueden asociarse al titular ni permitir, por su estructura, contenido o grado de desagregación, la identificación del mismo».

Es importante señalar que cuando existe el proceso de disociación, no es necesario el consentimiento de titular para el tratamiento de los datos personales, de conformidad con el artículo 10, fracción III de LFPDPPP.

También dicho procedimiento se realiza en otros países como España, en donde se le denomina anonimización.

Anonimización de datos personales

El tratamiento, análisis y explotación de grandes volúmenes de datos puede producir infinitud de beneficios a la sociedad, pero es preciso compatibilizar dicho almacenamiento y manipulación con el respeto a la protección de datos personales.

El mecanismo principal que permite alinear la generación de beneficios tangibles con el respeto a la privacidad es **la anonimización de la información**. Según la RAE, anonimizar es «expresar un dato relativo a entidades o personas, eliminando la referencia a su identidad». Por otro lado, y según la AEPD, «un proceso de anonimización es aquel que elimina o reduce al mínimo los riesgos de identificación

de los datos anonimizados manteniendo la veracidad de los resultados tras el tratamiento de los mismos».

Para garantizar su efectividad, **este proceso de anonimización debe ser irreversible**. Es decir, no debe existir posibilidad de recuperar el dato adicional empleando solo los datos anonimizados. Es necesario precisar que, debido a los avances tecnológicos y algorítmicos, no es posible garantizar de forma absoluta y completa la anonimización de la información. El objetivo realizable es proporcionar las mayores garantías posibles de cara a asegurar el respeto a la privacidad de las personas.

Según la AEPD, el proceso de anonimización debe atender a los siguientes principios:

- ▶ Principio proactivo. La privacidad se debe garantizar de forma proactiva y no de forma reactiva y una vez que se haya producido alguna fuga de información.
- ▶ Principio de veracidad por defecto. Se debe considerar la granularidad o grado de detalle final que deben tener los datos anonimizados. Esto lleva a que, en ocasiones, se exija la eliminación de ciertos datos para garantizar la anonimización del conjunto.
- ▶ Principio de privacidad objetiva. Siempre existirá un error residual de riesgo de reidentificación que deberá ser aceptable en función de la información anonimizada, conocido por el usuario y asumido por el responsable del fichero.
- ▶ Principio de plena funcionalidad. El proceso de anonimización debe garantizar la utilidad de los datos anonimizados en base a los objetivos inicialmente establecidos.
- ▶ Principio de privacidad en el ciclo de vida de la información. El respeto a la privacidad de los usuarios debe garantizarse durante todo el proceso de anonimización. Por ejemplo, realizando el proceso de anonimización en los sistemas preparados y autorizados a almacenar la información sin anonimizar.
- ▶ Principio de información y formación. Todo el personal con acceso a datos anonimizados o no deben estar correctamente formados e informados acerca de sus obligaciones.

Técnicamente hablando, existen varios **procedimientos para asegurar la anonimización de la información**. Algunos de los más destacados son:

- ▶ **Desnaturalizar:** consistente en transformar la naturaleza del dato. Por ejemplo, en lugar de representar la edad (edad = 42), podemos indicar el rango de edad al que pertenece en base a alguna división previamente establecida (edad = 5 donde 5 hace referencia al intervalo [40, 50]).
- ▶ **Cifrar:** consiste en hacer ilegible un mensaje concreto en base a la aplicación de un algoritmo que precisa de un conjunto de claves. En este caso, el descifrado de la información es posible siempre que se disponga del algoritmo de las claves necesarias.
- ▶ **Tokenizar:** se reemplaza el valor a anonimizar por un valor distinto (*token*) que no suele respetar la naturaleza del dato. Por ejemplo, se *tokeniza* el DNI 04345566D cambiándolo por el valor YID884S3VVQW4ZZY1. Se puede observar cómo no se respeta el formato estándar de un DNI. Para que el proceso mantenga la coherencia, al mismo valor le debe siempre corresponder el mismo *token*. La reversibilidad es posible siempre que se disponga del *token* correspondiente a cada valor.
- ▶ **Funciones hash:** es un método parecido a la *tokenización*. En este caso se aplica una función matemática al valor a anonimizar. Dicho valor reemplaza al valor original. En este caso, y por la propia naturaleza del proceso, se garantiza la irreversibilidad del proceso. No obstante, esta técnica podría, ocasionalmente, generar el mismo valor *hash* para distintos valores de entrada.
- ▶ **Disociar:** eliminar parte de la información para evitar la identificación personal. Por ejemplo, pongamos que disponemos de la siguiente información sobre un paciente: fecha de la consulta, hora de la consulta, código postal, edad, sexo y síntomas. Con esa información encima de la mesa existiría un riesgo de identificar al paciente, ya que solo la fecha y hora de la consulta proporcionan bastante información. La disociación consistiría en eliminar ciertos campos quedándonos solo (por ejemplo), con código postal, sexo y síntomas. De esta forma se reduce el riesgo de identificación.

Reglamento General de Protección de Datos (RGPD)

Aunque diseñada en el año 2016, el RGPD de la Unión Europea pasó a ser de obligado cumplimiento en el año 2018. La aplicación de esta normativa supone un gran reto para muchas compañías en las que el dato y la analítica sobre el mismo constituye un eje esencial de su actividad. El incumplimiento de la normativa implica cuantiosas multas para los infractores.

El RGPD **exige medidas adicionales que garanticen la transparencia en el tratamiento de datos personales**. Se pretende así empoderar al ciudadano para que tome las decisiones más adecuadas. Las organizaciones deben informar al usuario sobre el tipo de perfilado o modelización que realizan con base en sus datos personales facilitándole la denegación del permiso para realizar dicho tratamiento si así es solicitado.

Especialmente en el entorno financiero y asegurador, una de las medidas que más impacto causará es el **derecho a explicación**. Esta medida obliga, por ejemplo, a las entidades financieras a explicar las razones por las que un crédito ha sido denegado. De esta forma, el usuario puede actuar en consecuencia buscando alternativas para obtener una mejor clasificación en el futuro.

Además, los algoritmos implementados deben asegurar que se garantiza la no discriminación a la hora de tomar decisiones basadas en datos. No podrán tomarse decisiones basadas en criterios como la raza, la edad, el sexo, la religión, etc.

La Unión Europea ha habilitado el portal <https://www.eugdpr.org/> donde se desglosa todo tipo de información sobre esta regulación.



Figura 2. Portal de la RGPD.
Fuente: <https://www.eugdpr.org/>



En el período 2018-2020, el despliegue de GDPR en la UE (Unión Europea) está teniendo un impacto en las empresas de todo el mundo, gracias a su alcance legislativo más allá de las fronteras de la UE. Pero las empresas también tienen que contentarse con una ola de nueva legislación sobre privacidad de datos en los Estados Unidos, el Reino Unido, China y muchos otros países. Algunas leyes están inspiradas en el GDPR, mientras que otras adoptan un enfoque único que satisface las necesidades de su país.

Este resumen de las leyes de privacidad de datos aborda siete mercados principales, que cubren la legislación vigente, cómo afecta a las compañías nacionales e internacionales y cualquier característica única que pueda causar problemas para las empresas que manejan datos personales.

Descargar: 7 Pasos para el cumplimiento de la legislación sobre la protección de datos



Para conocer más de la normativa en cuanto a la privacidad y protección de datos personales en otros países visita el siguiente link:

<https://blog.ipswitch.com/es/leyes-globales-de-privacidad-de-datos-usa-ue-china-y-m%C3%A1s>

3.3. Sesgos, legalidad y responsabilidad



Accede al vídeo «Sesgos, legalidad y responsabilidad» a través del aula virtual



Accede al vídeo «Propuestas de regulación de la inteligencia artificial» a través del aula virtual

Lo que hace unos años era una cuestión ética, evitar la discriminación a la hora de emplear datos y algoritmos en la toma de decisiones, ahora se ha convertido en ley. En un entorno tecnológico y científico con una evolución tan rápida e impredecible, el entorno regulatorio no siempre es capaz de llegar a tiempo para proteger los derechos de los ciudadanos, por lo que el especialista, científico o técnico debe mantener altos estándares éticos en el día a día de su trabajo.

Hay una serie de puntos para tener en cuenta en este sentido:

1. Tener siempre presente la normativa de protección de datos.
2. Asegurarse que los algoritmos que empleamos no conllevan la toma de decisiones implicando la discriminación de algún colectivo por edad, sexo, raza, religión o cualquier otro aspecto.
3. Comprobar que los datos empleados no contienen sesgos que puedan llevar a tomar decisiones equivocadas.
4. Interpretar los resultados de los modelos científicamente, evitando interpretaciones interesadas y no ajustadas a la realidad.
5. Emplear los métodos de trabajo adecuados que garanticen la fiabilidad de los resultados.

Conceptos como inteligencia artificial y aprendizaje automático forman parte del vocabulario diario de comités de empresa y gobiernos de todo el mundo. Las decisiones que se toman en base a estos modelos afectan a millones de personas. Por tanto, es responsabilidad de los expertos asegurar que el objetivo final es aportar beneficio a la sociedad en general o los clientes de la compañía en particular, respetando los derechos fundamentales de la ciudadanía.

Sesgos: el motivo por el que los algoritmos aprenden y su principal punto débil

Como ya vimos anteriormente, **los sesgos son imposibles de eliminar ya que es el mecanismo por el que los algoritmos aprenden**. Crear sesgos y hacer suposiciones globales despreciando los detalles concretos es la base del aprendizaje y, por tanto, las excepciones siempre van a estar ahí. Lo importante en este caso es minimizarlos y que estos sesgos no aprendidos no sean sesgos inducidos por el entrenamiento debido a una mala elección de los datos de entrenamiento.

Es por tanto muy importante tener en cuenta este tipo de cuestiones en entornos sensibles y con sesgos que pueden incurrir en discriminaciones de cualquier tipo. Si no se puede discriminar a un individuo por razones ideológicas, de sexo, étnicos, etc., los algoritmos no pueden hacerlo tampoco.

¿Cómo evitar este hecho?

- ▶ Seleccionar cuidadosamente los datos de entrenamiento.
- ▶ Validar los algoritmos no solo con los datos provenientes del primer mundo o de nuestra área de influencia sino de otras partes del mundo con rasgos, culturas o éticas diferentes.
- ▶ Mantener una vigilancia continua de las decisiones que estos toman, para intervenir lo antes posible si se detectan estos sesgos.
- ▶ Tener evaluadores humanos que confirmen las decisiones tomadas por los algoritmos o al menos que los usuarios que se vean afectados por ellos puedan acudir para que se revise su caso particular.

Este va a ser uno de los retos más importante de aquí a unos años cuando los algoritmos de IA sean los que vayan controlando cada vez más y más procesos en los que la vida de los ciudadanos se vea afectada.

En cuanto a la **responsabilidad de los errores de estos algoritmos**, pues es bastante complicado establecerla y es algo en lo que debemos esforzarnos como sociedad. Obviamente, la responsabilidad siempre debe ser de la organización que use estos algoritmos que deben probarlos correctamente. Pero nada está exento de errores al 100 %. Así que hay que establecer cuál es el margen de error admisible para cada tarea que se delegue a la IA. Al fin y al cabo, los humanos también nos equivocamos. La diferencia aquí es que la cadena de responsabilidad entre los humanos es trazable y queda más o menos clara. Cuando entra dentro un algoritmo de IA ya no está tan clara, ¿es de los diseñadores del algoritmo?, ¿es de quien seleccionó los datos de entrenamiento? ¿Es de quien no revisó correctamente dicha decisión? Se entra en un terreno desde nuestro punto de vista pantanoso que tendremos que ir valorando con el tiempo y con leyes nuevas que probablemente lleguen tarde para algunos casos. Este es un reto importante a tener en cuenta en el futuro.

Implicaciones de la inteligencia artificial

La explosión de la inteligencia artificial ha marcado o marcará el inicio de una nueva revolución de dimensiones equivalentes a la Revolución Industrial. Como en toda gran transformación, y a pesar de las nuevas posibilidades que deslumbran, surgen también dudas y temores. Por ejemplo:

- ▶ ¿Perderán las personas su trabajo siendo reemplazadas por máquinas?
- ▶ ¿En qué trabajarán los humanos? ¿Cuánto tiempo?
- ▶ ¿Cómo se transformará nuestro tiempo de ocio?
- ▶ ¿Qué sucederá con nuestro derecho a la privacidad?
- ▶ ¿Cómo interactuaremos con los sistemas basados en inteligencia artificial? ¿Quién liderará la relación?
- ▶ ¿Podrán las máquinas sentir? ¿Cómo actuaremos entonces?
- ▶ ¿Dominarán los robots a la raza humana?

A la mayoría de estas preguntas no es posible responderla en estos momentos. Sin embargo, el debate sobre las implicaciones laborales de la robótica y la inteligencia artificial ya ha empezado.

En primer lugar, la gran mayoría de flujos de trabajo serán automatizados y no requerirán intervención humana. Esto permitirá mejorar la eficiencia de los procesos y, quizá, reducir los costes.

Los **algoritmos se convertirán en los principales agentes en ciertos escenarios**. Por ejemplo, muchas de las operaciones bursátiles que se llevan a cabo hoy en día son ejecutadas de forma automática mediante algoritmos que evalúan la evolución del mercado e incluso el contexto informativo analizando la información que circula sobre el contexto económico y social.

Los robots no solo han transformado las fábricas del siglo XXI, también están introduciéndose en el terreno sanitario e incluso militar.

El reto que tenemos por delante es **convertir los riesgos en oportunidades reales**.

Igual que sucedió con la Revolución Industrial, muchos trabajos desaparecerán porque ya no serán necesarios al poder realizarse de forma automática. Sin embargo, surgirán otros muchos nuevos trabajos y, por supuesto, los métodos de trabajo cambiarán radicalmente. La **educación adquiere una relevancia especial** para dirigir la transformación global de habilidades y conocimientos que permita la inmersión en el nuevo ecosistema. Algunos autores sugieren gravar mediante nuevos impuestos ciertos productos basados en inteligencia artificial, como los robots industriales para que el Estado reciba, al menos temporalmente, fuentes adicionales de financiación que permitan ayudar a aquellas personas que se queden sin trabajo debido a la automatización de sus posiciones y, además, pueda impulsar un nuevo sistema educativo que promueva la creatividad y genere el ecosistema humano que este nuevo paradigma necesita. A esto se le ha denominado que los **robots paguen impuestos**. Otros autores comentan que se debe crear una renta básica que permita

la subsistencia mínima individual. En cualquier caso, es un reto muy importante que debemos abordar en los próximos años.

La inteligencia artificial redefinirá la economía del futuro impulsando la productividad y generando una oleada de productos personalizados basados en las necesidades del cliente. Este impacto se extenderá por todos los ámbitos de actividad.

Regulando la inteligencia artificial

En los últimos años, destacados y muy relevantes personajes públicos han planteado la necesidad de regular de forma más exigente el ámbito de la inteligencia artificial de cara a evitar un mal uso de esta.

Entre estos actores destaca especialmente la organización Future of Life (<https://futureoflife.org/>). Esta fundación propone medidas para mitigar los riesgos de un mal uso no solo de la inteligencia artificial, sino también de la biotecnología o de las armas nucleares. Además, contemplan acciones para mitigar el cambio climático.

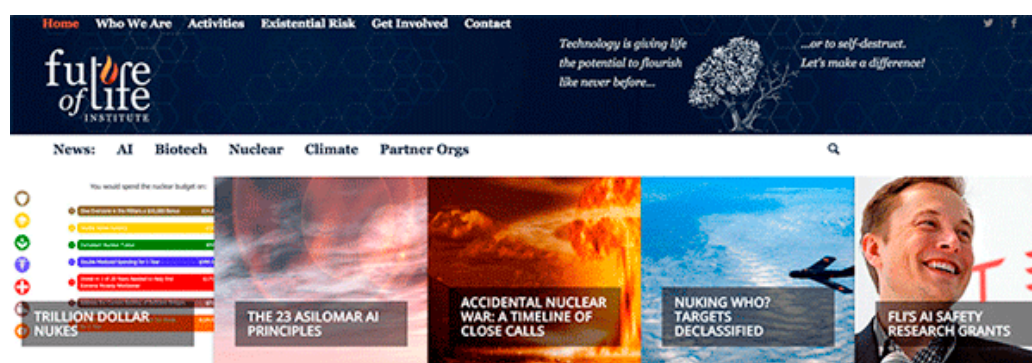


Figura 3. Página web de Future of Life.

Fuente: <https://futureoflife.org/>

La lista de personalidades detrás de este proyecto es absolutamente espectacular. Entre sus fundadores podemos encontrar destacados científicos procedentes del MIT, DeepMind, Universidad de Santa Cruz o Boston, además de a Jaan Tallinn, cofundador de Skype. La labor de la institución está apoyada públicamente por

figuras como Alan Alda, Erik Brynjolfsson (director del MIT Center for Digital Business), Morgan Freeman, el conocidísimo Stephen Hawking, Elon Musk (fundador de Tesla), Stuart Russell (autor del libro de referencia en esta asignatura) y un largo etcétera.

Una de las grandes preocupaciones de este grupo es **la aplicación de la inteligencia artificial con fines militares**. Atributos como la empatía, la justicia, la responsabilidad, la compasión..., son de momento atributos puramente humanos que las máquinas no tienen ocasión de contemplar.

Future of Life también se plantea romper con muchos de los mitos ligados a la inteligencia artificial:

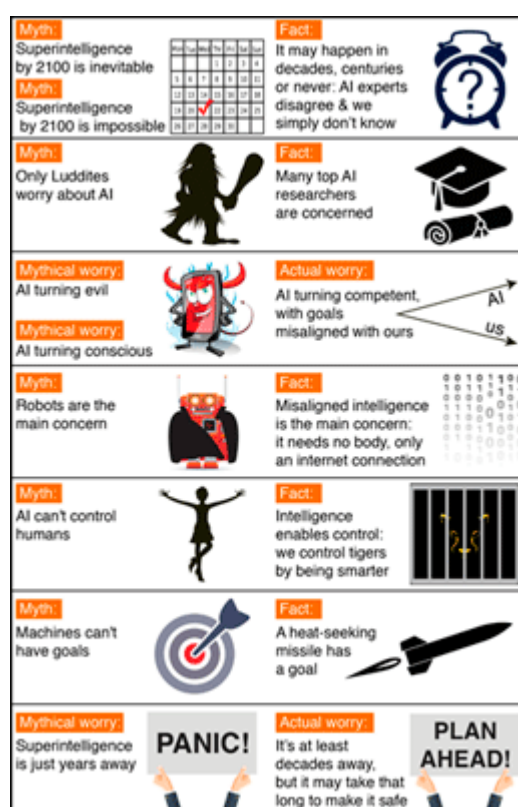


Figura 4. Mitos asociados a la inteligencia artificial según Future of Life.

Fuente: <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/>

La imagen anterior podría resumirse en los siguientes puntos:

- ▶ No es preciso definir qué es lo que seremos capaces de conseguir y cuándo.
- ▶ Muchos y grandes expertos están muy preocupados sobre el uso que se podría hacer de la inteligencia artificial.
- ▶ En ocasiones la inteligencia artificial puede producir soluciones no alienadas con los objetivos adecuados.
- ▶ No es solo cuestión de preocuparse por los robots, los algoritmos autónomos son también una cuestión que hay que vigilar.
- ▶ Ciertas soluciones de inteligencia artificial podrían condicionar y controlar el comportamiento humano.
- ▶ Las máquinas pueden plantearse un objetivo de forma predefinida.
- ▶ Es mejor prevenir que curar. Actuemos antes de que sea tarde estableciendo las bases que permitan el desarrollo de una inteligencia artificial segura y alineada con las necesidades humanas.

Hay que destacar que este grupo no está en contra de la inteligencia artificial. De hecho, mucho de sus colaboradores forman parte de la élite investigadora de la materia. Lo que se pide no es abandonar el desarrollo de la inteligencia artificial, sino **establecer unas bases regulatorias adecuadas** que permitan asegurar que su desarrollo se produce de forma segura.

3.4. Seguridad y tolerancia ante ataques



Accede al vídeo «Seguridad y robustez de la inteligencia artificial» a través del aula virtual

Con el uso cada vez mayor de la inteligencia artificial y de la compartición de datos, las posibilidades de ataques en este tipo de sistemas están aumentando considerablemente. El uso intensivo de datos en los sistemas de IA hace que estos mejoren y se puedan aplicar a múltiples entornos que antes no eran posibles. Pero

nos surge una duda, ¿cómo queda la seguridad y la privacidad de mis datos al adoptarla?

Ahora mismo muchas compañías están utilizando sistemas y modelos que van aprendiendo sobre la marcha y estos modelos no solo son usados por una compañía. Muchas veces, distintas compañías hacen uso de los mismos modelos debido a que la mayoría de estos modelos se enfocan al uso del *software* como servicio. De forma que, de una u otra forma, tus datos forman parte de una red que usa más gente. Por esta razón, ¿están seguros esos datos?

Por otro lado, con la creciente digitalización cada vez habrá más datos conectados a internet y, por tanto, más puertas abiertas que pueden ser aprovechadas por los cibercatacantes. Por eso, es muy importante que estos datos estén bien protegidos; sobre todo, si son datos sensibles.

También comienzan a verse ataques a los propios algoritmos de IA. Como todo sistema informático, estos sistemas pueden tener agujeros de seguridad. Pero también podemos ir más allá. Los sistemas de IA y de *machine learning* como sabemos tienen sesgos y los atacantes pueden utilizar esos sesgos a su favor. Por ejemplo, falseando los datos de entrada, aprovechando de alguna vulnerabilidad del algoritmo para que este falle en su beneficio. Por ejemplo, para conseguir que una IA descarte o acepte un préstamo en un banco.

Los sistemas de reconocimiento de texto y de lenguaje natural, por ejemplo, han avanzado enormemente en los últimos años, pero los programas de IA que analizan el texto pueden ser engañados por frases cuidadosamente elaboradas. Una frase que parece sencilla para un humano puede engañar a un algoritmo de IA. Esto puede implicar, por ejemplo, que un sistema que filtre noticias falsas sea engañado colando una noticia falsa sutilmente camuflada o que rechace o clasifique correctamente un posible candidato a un empleo. A este tipo de ejemplos que se le conoce con el nombre de adversarial examples.

Podemos definir, por tanto, un **adversarial example** como un caso con pequeñas perturbaciones intencionales de características que hacen que un modelo de aprendizaje de una máquina haga una predicción falsa.

Algunos ejemplos de este tipo de adversarial examples pueden ser:

- ▶ Un auto que se conduce a sí mismo choca con otro coche porque ignora una señal modificada para que el algoritmo lo identificara erróneamente.
- ▶ Armas diseñadas para engañar los sistemas de escaneo de un aeropuerto.
- ▶ Engañamos a un sistema recomendador para que recomiende nuestros productos cuando los usuarios buscan el de la competencia.

En la siguiente figura podemos ver un ejemplo de cómo se ha conseguido engañar a AlexNet, un conocido clasificador de imágenes. En las imágenes de la izquierda el algoritmo los reconoce, pero al aplicar el ruido de la imagen central, el algoritmo clasifica todas las imágenes como avestruces.

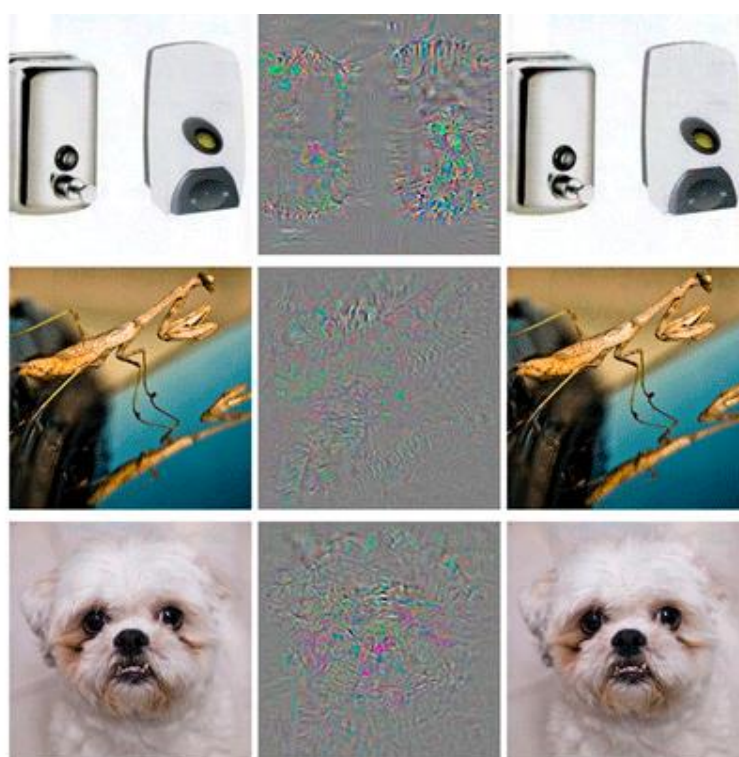


Figura 5. Algunas imágenes adversarias que engañan a Alexnet.

Fuente: <https://christophm.github.io/interpretable-ml-book/adversarial.html>

¿Cómo podemos defendernos de estos ataques? Es complicado, pero debemos ser conscientes de ello para intentar remediarlo. Podemos mitigar el riesgo con las **siguientes medidas**:

1. Conocer a tu adversario. Conocer cuáles son las motivaciones que un atacante tiene para hacerlo.
2. Ser proactivo, esto es, intentar engañar constantemente a los sistemas con tus propios ejemplos de adversario para intentar detectar los posibles errores.
3. Protegerte a ti mismo con reentrenamientos activos con adversarios o utilizar varios clasificadores y decidir en mayoría.

3.5. Explicabilidad de algoritmos



Accede al vídeo «Explicabilidad y transparencia en inteligencia artificial» a través del aula virtual

Como hemos visto en el apartado anterior, es posible engañar a los algoritmos de *machine learning* entregando ejemplos preparados para que estos se equivoquen. Si podemos explicar el modelo que se genera en el proceso de aprendizaje es más fácil evitar o predecir estos errores. Pero esto no siempre es posible, debido a que no todos los algoritmos son explicables o fácilmente explicables.

Los **modelos que generan los algoritmos de *machine learning* podemos clasificarlos en función de su explicabilidad**: en modelos de caja negra y modelos de caja blanca.

Los **modelos de caja negra** son aquellos en los que conseguir entender el modelo para poder analizarlo es muy complejo o casi imposible. Potencialmente todos pueden ser de caja negra si el número de parámetros de configuración es muy alto. Pero, en general, son propensos a ser algoritmos que generan modelos de caja negra las redes de neuronas, los random forest, el razonamiento basado en casos con medidas de distancia compleja, etc.

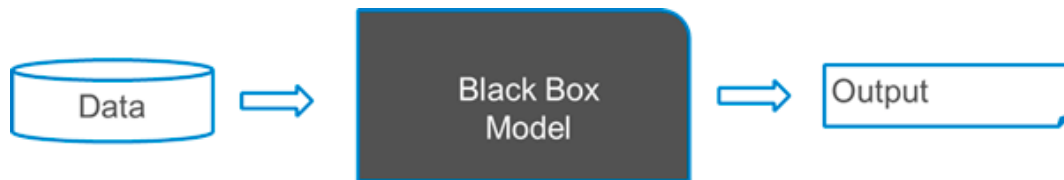


Figura 6. Modelo de caja negra.

Los **modelos de caja blanca**, por el contrario, son más sencillos de explicar y, por tanto, de analizar. Los árboles de decisión simples son algoritmos que generan modelos de cajas blancas. También las redes bayesianas o cualquier sistema que utilice lógica de predicados, con encadenamiento hacia delante o hacia atrás.

La **explicabilidad** nos da una serie de ventajas que debemos tener en cuenta como, por ejemplo:

- ▶ **Confiabilidad. Es muy importante poder confiar en las decisiones de un algoritmo**, sobre todo si este está a cargo de algo importante. Por ejemplo, está conduciendo un vehículo por ti, tomando decisiones de compra en bolsa o manejando una central nuclear. Pero también para otros temas menores es importante saber qué decisión está tomando el algoritmo debido a que este puede incurrir en, como hemos visto durante este tema, sesgos, discriminaciones etc. Saber cómo llega a esa decisión puede ayudarnos a prevenir estos malos comportamientos de los algoritmos.
- ▶ **Adquirir nuevo conocimiento. Los algoritmos a veces son capaces de resolver problemas o descubrir nuevas soluciones a problemas que antes no se conocían.** Pero estos problemas muchas veces no pueden ser analizados correctamente debido a que no sabemos cómo el algoritmo ha llegado a esa conclusión. Por lo tanto, perdemos los detalles de ese nuevo conocimiento adquirido.
- ▶ **Detección de fallos. Si el modelo tiene fallos y conocemos el modelo, podremos predecirlos, mitigarlos o reentrenarlo.** Hasta ahora, solo podemos saber si un algoritmo de caja negra tiene fallos probándolo exhaustivamente. Pero siempre puede haber casos que no se han contemplado en los que el algoritmo falle. Durante el tema hemos visto varios de ellos.

Pero eso no significa que dejemos de usar algoritmos de caja negra. Hay ciertos entornos donde no hay problema por usar algoritmos de caja negra debido a que la necesidad de verificación del modelo no es crítica. Pensemos en algoritmos que controlan la calidad de imagen en un videojuego (DLSS), el algoritmo de YouTube o el recomendador de Netflix, o en algoritmos que ayuden a los científicos de la NASA a encontrar nuevos exoplanetas. Si hay algún error en estos algoritmos que clasifica un exoplaneta como planeta que no lo es, el error no es de vital importancia. Con el tiempo se descubrirá que no lo es, pero no tiene un coste de error alto y es mayor el beneficio de poder haber encontrado cientos de exoplanetas que sin la IA no hubieran podido ser encontrados.

En resumen, hay que ser conscientes en qué dominios que el modelo no sea explicable no es un problema y que los beneficios de estos modelos compensen la falta de explicabilidad, y en qué dominios es recomendable usar algoritmos explicables, aunque obtengan peor rendimiento.

La aproximación, normalmente, a este tipo de dominios sensible es **tener soluciones híbridas entre algoritmos de caja negra y algoritmos de caja blanca**. Es decir, que parte del razonamiento esté inducido por algoritmos de caja blanca para que el cuerpo principal de la solución sea explicable y se use los algoritmos de caja negra en aquellas tareas donde la explicabilidad sea menos crítica. O también se busca intentar explicar los algoritmos de caja negra con otros métodos que ayuden a entenderlos.

3.6. Referencias bibliográficas

Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP), publicada en el Diario Oficial de Federal el 05 de julio de 2010.

Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPSO), publicada en el Diario Oficial de Federal el 26 de enero de 2017.

Reglamento de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (RLFPDPPP) publicado en el Diario Oficial de la Federación el 21 de diciembre del 2011

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.



Accede a los ejercicios de autoevaluación a través del aula virtual

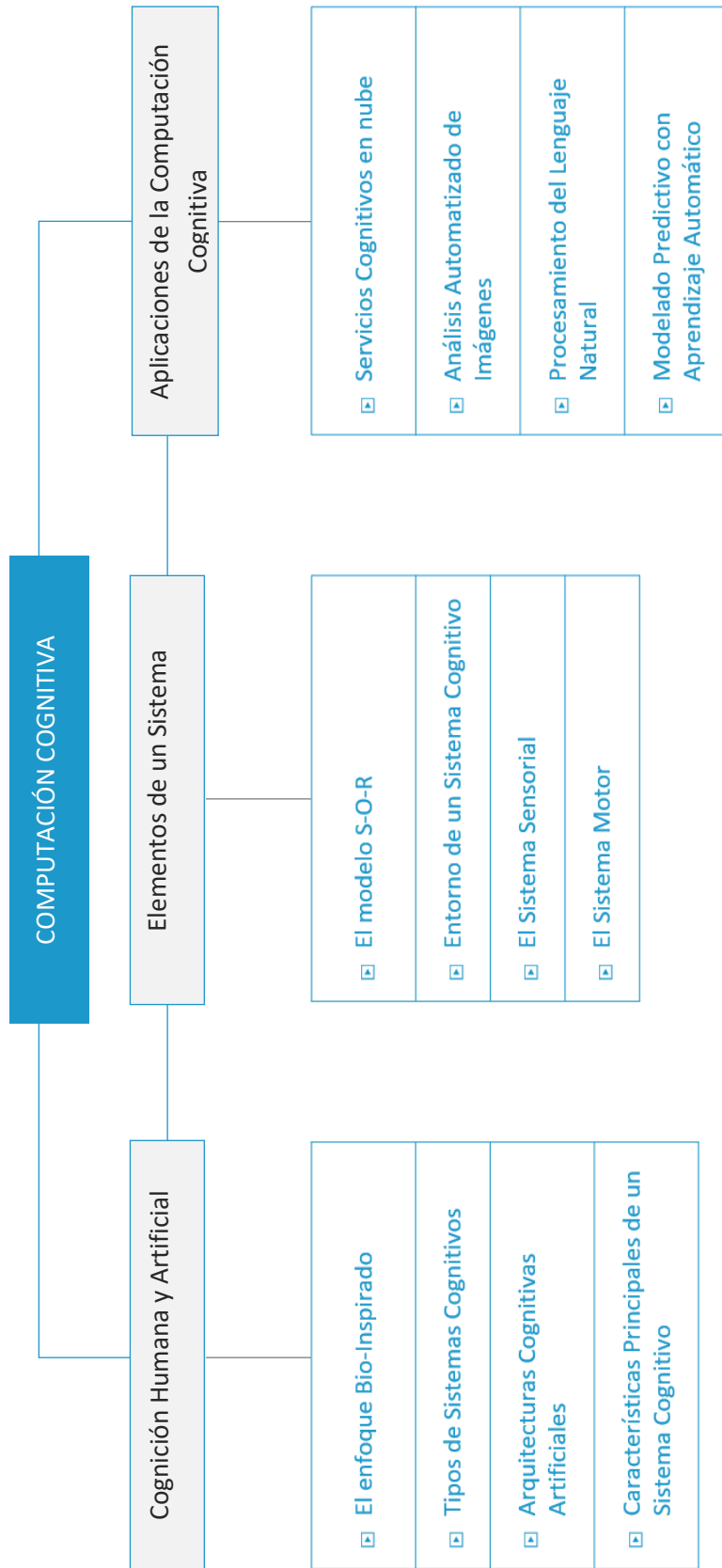


Accede al vídeo «Resumen del tema» a través del aula virtual

Inteligencia Artificial y Computación Cognitiva

Computación Cognitiva

Esquema. Tema 4



Ideas clave. Tema 4

4.1. Introducción y objetivos



Accede al vídeo «Introducción. Computación cognitiva» a través del aula virtual

Para estudiar este tema lee las Ideas clave disponibles a continuación y el siguiente artículo:

Martínez, E.C. y Méndez, M.R. (2013). El sujeto desde la neurociencia y la inteligencia artificial. *Revista de estudios de juventud*, 103, 9-19.

Accede al artículo a través del aula virtual o de la siguiente dirección web:

<http://www.injuve.es/sites/default/files/1%20El%20sujeto%20desde%20la%20neurociencia%20y%20la%20inteligencia%20artificial.pdf>

En este tema analizaremos tres puntos fundamentales para poder entender la importancia que tiene la ciencia cognitiva tanto en la concepción y diseño, como en la aplicación práctica de la inteligencia artificial:

- ▶ Qué es la cognición y el procesamiento de la información en los sistemas cognitivos.
- ▶ Cuáles son los principales componentes de un sistema cognitivo, independientemente de que sea biológico o artificial.
- ▶ Cuáles son las aplicaciones prácticas de los sistemas cognitivos artificiales contemporáneos y futuros.

4.2. Introducción a la computación cognitiva



Accede al vídeo «¿Qué es la computación cognitiva?» a través del aula virtual



Accede al vídeo «Tipos de sistemas cognitivos» a través del aula virtual



Accede al vídeo «Características de los sistemas cognitivos» a través del aula virtual

El concepto de **computación cognitiva** hace referencia al diseño, construcción y uso de sistemas informáticos que demuestran capacidades y características similares a los humanos. En este sentido, los sistemas cognitivos artificiales se enmarcan en el área de la Inteligencia Artificial y su objetivo es **realizar un procesamiento de la información que permita obtener un conocimiento útil para la acción**.

Cognición, en este contexto, es equivalente a la **adquisición de conocimiento**. Como veremos más adelante este concepto se traduce en la transformación de datos en información, para posteriormente convertir esa información en **conocimiento útil para guiar la conducta** y, por lo tanto, conseguir un comportamiento adaptado al medio e inteligente.

La computación cognitiva se basa en que, tanto los humanos (y otros animales) como las computadoras, realizan un procesamiento de la información. Aunque lo hacen utilizando medios físicos y mecanismos diferentes, existen claros paralelismos que pueden estudiarse con el objeto de **mejorar el diseño de los sistemas inteligentes artificiales**.

De una máquina inteligente se espera que sea capaz de procesar los datos de entrada de tal forma que pueda **resolver problemas**, de forma análoga a como lo hace un humano. Por ello, en enfoque de **la computación cognitiva busca a menudo**

inspirarse en el funcionamiento del sistema nervioso para encontrar nuevos mecanismos de procesamiento de la información.

La **bio-inspiración** es el proceso de diseño que habitualmente se usa en computación cognitiva para guiar el diseño de sistemas computacionales inteligentes. **Comprender cómo funciona el cerebro**, descubriendo la relación existente entre los circuitos neuronales y la generación de percepciones, emociones, pensamientos y conducta, permite acelerar la construcción de sistemas inteligentes cada vez más avanzados.

El gran reto actual se centra en **descubrir y comprender los procesos que dan lugar a la extraordinaria complejidad mental y conductual** que exhiben los humanos y otros animales. Desde el punto de vista de la inteligencia artificial, este es un tema importante, pues se cree que el conocimiento del funcionamiento de los sistemas inteligentes naturales puede aplicarse más o menos directamente al diseño y la construcción de máquinas inteligentes. De hecho, es muy habitual que se usen en inteligencia artificial **técnicas claramente bio-inspiradas, como las redes de neuronas artificiales**.

Intuitivamente todos sabemos lo que es un **sistema cognitivo**, porque **el propio ser humano es uno**. Sin embargo, es preciso definir claramente a qué nos referimos, en general, cuando hablamos de cognición. En primer lugar, debido al avance de la inteligencia artificial, debemos distinguir entre varios tipos de sistemas cognitivos:

- ▶ **Sistemas cognitivos naturales.** Los organismos biológicos dotados de sistema nervioso son en general sistemas cognitivos. Aunque habitualmente pensamos específicamente en los humanos, muchas otras especies de animales también muestran capacidades avanzadas de procesamiento de la información y funciones cognitivas como el aprendizaje, la memoria o la atención.
- ▶ **Sistemas cognitivos artificiales.** En el ámbito de la inteligencia artificial, construimos sistemas computacionales capaces de realizar complejos

procesamientos de la información, que también incluyen mecanismos de aprendizaje, de memoria, etc.

- ▶ **Sistemas cognitivos híbridos.** Cada vez es más común pensar en la combinación sinérgica de humanos y máquinas en la resolución de tareas. Este es un escenario en el que podemos considerar no solo que hay dos sistemas cognitivos colaborando, sino que se puede llegar a producir una integración entre ambos (llegándose incluso a considerar el concepto de cibernético u organismo cibernético).

Desde el punto de vista de los sistemas artificiales, cuando tenemos algoritmos que únicamente realizan una función cognitiva, como por ejemplo los algoritmos de aprendizaje automático, que se centran exclusivamente en el aprendizaje, no lo consideramos habitualmente como un sistema cognitivo. Simplemente, sería un algoritmo o módulo que realiza dicha función. En cambio, como veremos a continuación, **cuando combinamos en una misma arquitectura diferentes funciones cognitivas**, entonces sí consideramos que se trata de un **sistema cognitivo artificial**.

Es importante destacar que las arquitecturas cognitivas artificiales se pueden presentar o diseñar tomando diferentes formas o modalidades. Algunos de los formatos típicos en los que nos podemos encontrar un sistema cognitivo artificial son:

- ▶ **Robots de servicio.** Robots físicos destinados a la realización de tareas en entornos no industriales, que despliegan ciertas funciones de forma autónoma, como por ejemplo realizar tareas de limpieza doméstica.
- ▶ **Agentes conversacionales.** Robots *software* (o *bots*) que a veces toman la forma de «altavoces inteligentes» o servicios de atención al cliente integrados en un sitio web, que son capaces de interactuar con el usuario a través de conversaciones en lenguaje natural.

- ▶ **Sistemas embebidos o embarcados y robots industriales.** Sistemas operativos inteligentes que controlan en mayor o menor medida medios de transporte, fábricas u otras infraestructuras con capacidad autónoma. Ejemplos típicos de este tipo de sistemas son los vehículos y los drones autónomos, aunque también existen en muchos otros formatos, como los vehículos submarinos autónomos no tripulados o las sondas espaciales.
- ▶ **Gestores y generadores de contenidos online.** Sistemas automáticos que son capaces de comprender el significado de contenidos online (redes sociales, páginas web, aplicaciones móviles, etc.) y también son capaces de generar contenidos de forma autónoma en forma de imágenes y/o lenguaje. Normalmente se usan estos sistemas para optimizar los procesos de *marketing* y publicidad online.
- ▶ **Sistemas de aumento o apoyo a las capacidades humanas.** Más recientemente también se están empezando a desarrollar sistemas inteligentes artificiales para compensar o rehabilitar la pérdida de funciones cognitivas de pacientes, o incluso para aumentar las capacidades cognitivas de personas sin ninguna patología o discapacidad.

Como vemos, estos sistemas pueden tomar muchas formas y estar destinados a la realización de múltiples y diferentes tareas. Aunque cada línea de aplicación específica tiene sus propias consideraciones y requisitos, las principales características que solemos perseguir en el diseño de sistemas cognitivos artificiales son:

- ▶ **Trabajo en colaboración con humanos.** Una de las grandes diferencias entre la forma de funcionar de un sistema informático tradicional y un ser humano es que las personas somos capaces de **manejar situaciones con alto grado de incertidumbre y definiciones ambiguas**. Sin embargo, la programación clásica de un computador necesita todos los parámetros exactamente determinados para funcionar. En el diseño de los sistemas cognitivos artificiales se busca que el

proceso de la información tenga estas mismas características, de forma que pueda ocurrir una interacción natural con las personas.

- ▶ **Adaptación a diferentes entornos.** De nuevo, los sistemas informáticos tradicionales, e incluso la mayoría de las aplicaciones de inteligencia artificial, están programados para funcionar en entornos claramente definidos, muy estables y cuyas reglas no cambian a lo largo del tiempo. Un ejemplo clásico sería el juego del ajedrez, donde un sistema automático se puede desenvolver sin problemas porque aplica un único algoritmo sobre un conjunto de reglas de juego preestablecidas y fijas. Sin embargo, la computación cognitiva debe hacer honor a su nombre, es decir, **conocer cuál es el entorno en que se está ejecutando en cada momento**, para así **adecuarse de forma efectiva a las circunstancias**, de igual forma que hacen los humanos constantemente en su vida diaria.
- ▶ **Anticipación de la necesidad de actuar.** Una forma de entender el cerebro humano es como una máquina biológica diseñada para predecir. La mente humana está constantemente analizando (consciente o inconscientemente) las posibles consecuencias de las acciones propias y las de los demás. Esta capacidad de predicción le permite anticiparse a las circunstancias que van apareciendo en un mundo cambiante y adaptarse con éxito a su entorno. En general, buscamos que los sistemas cognitivos artificiales también tengan esta capacidad de anticipación, incluso **detectando a tiempo cuando los aprendizajes anteriores ya no funcionan** y por lo tanto es necesario flexibilizar la conducta para dar respuesta a un entorno complejo e incierto.
- ▶ **Aprendizaje a partir de la experiencia.** Aunque contamos en la actualidad con gran variedad de técnicas de aprendizaje automático, dichos algoritmos se centran en reconocer patrones a partir de un conjunto de datos previamente preprocesado, preparado o etiquetado. Es decir, para que un algoritmo de aprendizaje automático pueda funcionar bien suele ser necesario aplicar una fase previa de ingeniería y extracción de características. Esta labor la suele desempeñar un humano (habitualmente el científico de datos o el ingeniero de aprendizaje

automático), pero no es algo que los sistemas actuales puedan automatizar para adaptarse a diferentes tareas. En el ámbito de los sistemas cognitivos artificiales sí que se busca esta automatización, que el propio sistema sea capaz de transformar los datos «en crudo» que llegan por los sensores y obtener una serie de características que sean efectivas a la hora de comprender, aprender y actuar en respuesta a cada situación. Por lo tanto, el concepto de aprender a partir de la experiencia se refiere a **dar sentido a las experiencias sensoriales anteriores para poder hacer un uso adecuado de dicho conocimiento.**

- **Manejo de lo inesperado.** Una de las características principales de los sistemas cognitivos artificiales es que **deben ser robustos**. En este contexto robustez se refiere a la capacidad de un sistema de gestionar situaciones no contempladas previamente, pero que en el mundo real se producen con relativa frecuencia. Los humanos somos capaces de reaccionar ante una gran variedad de posibles situaciones inesperadas, explorando las mejores vías de acción en cada caso, y finalmente aprendiendo de la experiencia. También es importante considerar una **transferencia del aprendizaje** desde unas situaciones a otras, de forma que el sistema sea capaz de identificar elementos comunes a una situación pasada e intentar la estrategia que en aquel momento fue efectiva. Por supuesto, ningún sistema autónomo, natural o artificial, puede solventar absolutamente todas las situaciones que se presenten, pero la robustez se refiere también a poder **recuperarse ante un mal funcionamiento, un error o un fallo.**
- **Tolerancia a fallos.** En relación con la robustez, como hemos visto, es importante que un sistema cognitivo artificial se pueda recuperar de una situación de fallo. Esto se puede lograr mediante la **implementación de subsistemas redundantes o subsistemas complementarios**. Además, incluso aunque no se cuente con subsistemas redundantes, la robustez también implica que se pueda seguir la operación en un modo degradado. Por ejemplo, en un robot de 4 patas en el que se haya estropeado una de ellas, el sistema de control debería ser capaz de adaptar la estrategia de locomoción y seguir trasladándose a 3 patas, aunque sea de forma menos eficiente.

► **Autogestión, autodiagnóstico y autoreparación.** Idealmente, un sistema cognitivo artificial, dada su complejidad, debería minimizar el mantenimiento requerido por parte de operadores humanos. En este sentido, sería deseable contar con sistemas integrados de **autogestión de los propios recursos** y la propia capacidad. Para realizar esta gestión sería preciso que el sistema contara con **mecanismos de autodiagnóstico**, es decir, **sensores interoceptivos** (que proporcionan información sobre el propio cuerpo y los propios mecanismos, a diferencia de los sensores exteroceptivos, que reciben información sobre el medio). Siguiendo en esta línea de autosuficiencia, también existe investigación en el ámbito de la autoreparación, donde se busca conseguir sistemas que no solo detecten el daño, sino que también puedan tomar medidas para mitigarlo, reducirlo o incluso solucionarlo. Por supuesto, estas capacidades son mucho más difíciles de conseguir y hoy en día se sitúan en los contextos de investigación básica.

► **Metacognición.** La metacognición se refiere al conocimiento sobre el conocimiento. En el contexto de los sistemas cognitivos artificiales se refiere a la capacidad de un sistema para disponer de un modelo interno que represente sus propios mecanismos operativos. La metacognición es una característica típica de los humanos, que somos capaces de **manejar conocimiento sobre nuestra propia forma de procesar la información**. Ejemplos típicos de metacognición son la meta-memoria y el meta-aprendizaje. La meta-memoria se refiere al conocimiento sobre el número y características de los contenidos que tenemos almacenados en nuestra memoria. Si a alguien le piden que enumere todas las especies de ave que conoce estaría haciendo una tarea de memoria, pero si le pedimos que nos diga cuántas especies de ave cree que conoce sería un ejercicio de meta-memoria. Es decir, el número de especies de ave que creo que conozco es un metadato. Análogamente, en el ámbito del meta-aprendizaje, los humanos podemos conocer los mecanismos que usamos para aprender y saber (meta-saber) qué estrategia nos funciona mejor en cada situación. Este tipo de mecanismos auto-reflexivos

implican el uso de funciones cognitivas de nivel superior que, de hecho, diferencian a los humanos de otros animales.

- **Comunicación de contenidos complejos.** Finalmente, bien sea porque el sistema ha de comunicarse con los humanos o porque necesite establecer colaboraciones con otros sistemas cognitivos artificiales, es muy importante que cuente con un sistema avanzado de comunicación, probablemente en forma de lenguaje con una gramática, para poder expresar sus estados internos, intenciones, acciones, etc. La coordinación entre diversos sistemas y por lo tanto la formación de grupos colaborativos (que pueden incluir también a seres humanos) no sería posible sin la capacidad de comunicarse y mantener un intercambio efectivo de información con otros sistemas cognitivos. Podemos distinguir entre formas de comunicación naturales, como el lenguaje humano, y protocolos de comunicación tipo M2M (*machine to machine*, de máquina a máquina).

4.3. Elementos de un sistema cognitivo



Accede al vídeo «Elementos de un sistema cognitivo» a través del aula virtual

Las arquitecturas cognitivas artificiales son sistemas diseñados para **integrar diferentes funciones de procesamiento de la información** equivalentes a las que realizan los humanos (u otros seres inteligentes). El propio concepto de arquitectura sugiere que estos sistemas están organizados en diferentes módulos o componentes.

Un esquema básico e inicial de una arquitectura cognitiva implica considerar los siguientes elementos:

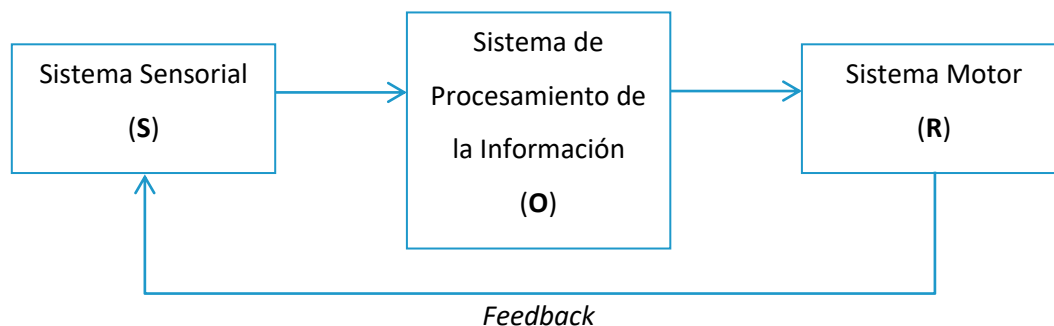
- **Entorno.** Se refiere al medioambiente en el que desarrolla sus operaciones el sistema. Si se trata de un agente físico (robot) el entorno es el medio físico en el que el robot se encuentra. Si, por el contrario, consideramos un agente software,

el entorno quedará definido por las comunicaciones e interacciones que pueda realizar el agente a través de sensores software (entradas de datos) y actuadores software (ejecución de acciones). En muchas ocasiones, cuando el sistema esté diseñado para interactuar con humanos, las personas serán también parte del entorno de la arquitectura cognitiva artificial. Aunque típicamente se podría considerar el entorno como una entidad externa a la arquitectura cognitiva, es muy importante tenerlo en cuenta como parte integral de los mecanismos de cognición. De igual forma que ocurre con los humanos, si no tenemos en cuenta el medioambiente no podemos explicar ni los pensamientos, ni las conductas, ni las emociones de una persona.

- ▶ **Sistema Sensorial.** Compuesto por todos los sensores del sistema sería el equivalente a los sentidos en un ser humano. De igual forma que una persona obtiene información del medioambiente a través de la vista, el oído, el tacto, etc. las arquitecturas cognitivas artificiales también necesitan tener una capa dedicada a la **adquisición de datos del medioambiente**. Recordemos que también será necesario obtener datos internos, por lo que en esta capa podríamos incluir los sensores interoceptivos. Un ejemplo de sentido interoceptivo en los humanos es el nociceptivo (el sistema que nos permite percibir el dolor y que proporciona información vital sobre posibles daños en los tejidos).
- ▶ **Sistema Motor.** La implementación de un sistema cognitivo artificial sin capacidad para actuar en el mundo no tendría sentido. En el caso de los robots físicos es muy fácil identificar los componentes del sistema motor, ya que a menudo están compuestos por motores, servos y otros tipos de actuadores electrónicos. Sin embargo, al pensar en agentes software debemos considerar componentes del sistema motor a todas las funciones que permiten emitir o proporcionar contenido hacia el exterior del sistema (y que por lo tanto puede ser recibido por otros sistemas o personas).
- ▶ **Sistema de Procesamiento de la Información.** A este nivel general, podríamos incluso considerar a este sistema como una “caja negra” que se encuentra entre

los sistemas sensoriales y los sistemas motores. Este sistema central sería el equivalente al sistema nervioso central de los humanos (que principalmente reside en el encéfalo) y es donde se realiza el procesamiento de la información que da lugar a numerosos subsistemas. Normalmente nos referimos a estos **subsistemas como funciones cognitivas y afectivas**. Como veremos más adelante, entre estas funciones se encuentran la memoria, la atención, el aprendizaje, la motivación, etc. En los humanos también los conocemos como los procesos psicológicos básicos o procesos mentales.

Los módulos descritos anteriormente corresponden al modelo S-O-R (Estímulo – Organismo – Respuesta), que considera que cualquier sistema cognitivo se autorregula a la vez que también influye en su medioambiente y detecta las consecuencias de sus acciones. En este sentido es primordial considerar que existe un **bucle de retroalimentación (o *feedback*) a través del medioambiente**.



Más adelante, veremos en detalle los procesos o funciones que se realizan como parte del módulo de Procesamiento de la Información, aunque de momento podemos enumerar algunos de los elementos que forman parte de estos módulos o capas.

El entorno de un sistema cognitivo artificial puede estar compuesto, entre otros elementos, de:

- ▶ Usuarios (humanos).
- ▶ Otros sistemas cognitivos artificiales.
- ▶ Otros tipos de agentes, como sistemas informáticos tradicionales.

- ▶ Servicios de Internet o redes privadas.
- ▶ Un mundo virtual simulado.
- ▶ El mundo real.
- ▶ Etc.

En el sistema sensorial podemos encontrarnos, entre otros, con sistemas de percepción computacional especializados en las diferentes modalidades sensoriales:

- ▶ Procesamiento del audio (incluyendo el procesamiento de la señal de voz).
- ▶ Visión artificial (incluyendo diferentes tareas de reconocimiento de objetos, clasificación de imágenes, etc.).
- ▶ Reconocimiento de patrones en otras fuentes de datos no visuales o auditivas, como archivos de log, lecturas de otros sensores, etc.
- ▶ Subsistemas de fusión sensorial multimodal, que combinan datos provenientes de diferentes sensores.
- ▶ Etc.

En el sistema motor podemos encontrarnos, entre otros, con sistemas de locomoción o de actuación:

- ▶ Generación del lenguaje y síntesis de voz.
- ▶ Subsistemas de comunicación e integración con diferentes APIs (interfaces de aplicación de programa).
- ▶ Subsistemas de control de extremidades, ruedas, brazos robóticos, manipuladores, etc.
- ▶ Control de la postura, la posición y la navegación.
- ▶ Control motor fino y coordinación de diferentes actuadores.
- ▶ Etc.

4.4. Aplicaciones de la computación cognitiva



Accede al vídeo «Aplicaciones de la computación cognitiva» a través del aula virtual

El uso de aplicaciones de computación cognitiva se ha popularizado en los últimos años porque el rendimiento de estos sistemas es cada vez superior. De hecho, ya existen muchas tareas en las que **el rendimiento de un sistema artificial es mejor que el de los humanos**. Es común que los grandes proveedores de soluciones tecnológicas, como por ejemplo IBM, Microsoft, Google o Amazon, proporcionen **servicios cognitivos a través de su infraestructura de nube**.

Es importante destacar varios aspectos primordiales sobre los servicios cognitivos que ofrecen estas compañías a través de la nube:

- ▶ Es común que, dependiendo de los términos más populares en cada momento, y por una motivación comercial, cambien los nombres de estos servicios, aunque el funcionamiento básico sea el mismo. Algunas de las denominaciones que suelen usar son: «servicios cognitivos», «servicios de inteligencia artificial», «servicios de analítica avanzada», «servicios de aprendizaje automático y *deep learning*», etc. En todos los casos, suelen referirse a servicios que, parcialmente, **automatizan tareas que típicamente serían realizadas por un humano**, es decir, tareas que necesitan cognición.
- ▶ Es preciso tener en cuenta que la descripción que hemos hecho anteriormente de un sistema cognitivo artificial es mucho más completa y ambiciosa que los servicios existentes actualmente el mercado, por lo que **ninguno cumplirá con todas las características deseables de un sistema cognitivo**.
- ▶ Hoy en día podemos distinguir claramente tres dominios de aplicación de los servicios cognitivos en función del tipo de datos que toman como entrada:

- Servicios relacionados con el análisis automatizado de imágenes (visión artificial).
- Servicios relacionados con el análisis del lenguaje (procesamiento del lenguaje natural).
- Servicios para la construcción de modelos predictivos a partir de datos estructurados (aprendizaje automático).

Aunque muchos de estos servicios cognitivos han nacido cumpliendo solo una pequeña parte de los que sería el esquema de un sistema cognitivo completo, su desarrollo es cada vez más sofisticado y se van incluyendo nuevas funcionalidades constantemente. Por ejemplo, mientras que al inicio los sistemas de visión artificial solo cumplían con una función rudimentaria de detección y localización de objetos en las imágenes, hoy en día están incorporando sofisticadas funciones de atención visual que permiten aumentar su rendimiento. Análogamente, en el mundo del procesamiento del lenguaje natural, se ha pasado de solo hacer reconocimiento del lenguaje a hacer también generación de este (es decir, el sistema perceptivo y también el sistema motor).

4.5. Referencias bibliográficas

Goertzel, B., Lian, R., Arel, I., De Garis, H., & Chen, S. (2010). A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures. *Neurocomputing*, 74(1-3), 30-49.

High, R. (2012). The era of cognitive systems: An inside look at IBM Watson and how it works. IBM Corporation, Redbooks, 1-16.

Haines, D. E. (2014). *Principios de Neurociencia. Aplicaciones básicas y clínicas*. España: Elsevier.

Hollnagel, E., & Woods, D. D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering. CRC press.

Kosslyn, S. M. (2010). Wet mind: The new cognitive neuroscience. Nueva York: Simon and Schuster.

Lieto, A., Bhatt, M., Oltramari, A., & Vernon, D. (2018). The role of cognitive architectures in general artificial intelligence.

Martínez, E. C. y Méndez, M. R. (2013). El sujeto desde la neurociencia y la inteligencia artificial. *Revista de estudios de juventud*, 103, 9-19.



Accede a los ejercicios de autoevaluación a través del aula virtual

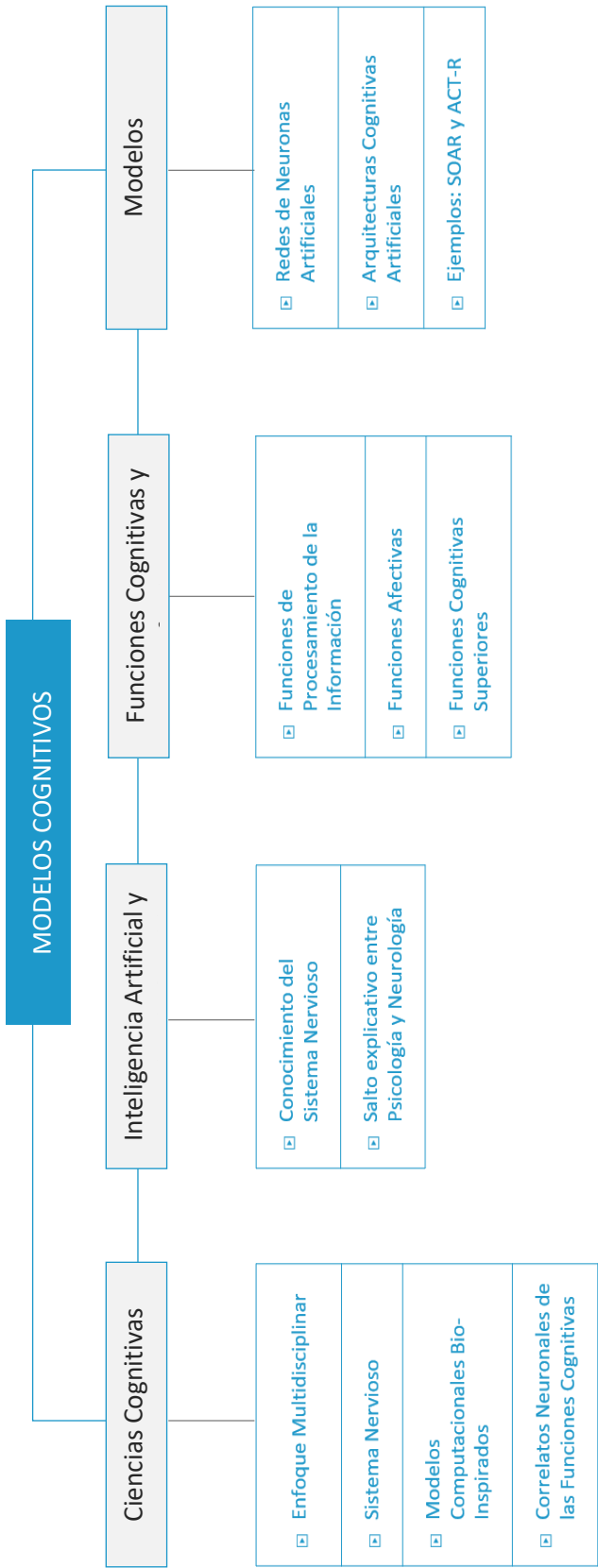


Accede al vídeo «Resumen del tema» a través del aula virtual

Inteligencia Artificial y Computación Cognitiva

Modelos cognitivos

Esquema. Tema 5



Ideas clave. Tema 5

5.1. Introducción y objetivos



Accede al vídeo «Introducción. Modelos cognitivos» a través del aula virtual

Para estudiar este tema lee las Ideas clave disponibles a continuación y el siguiente artículo:

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). *Neuroscience-inspired artificial intelligence*. *Neuron*, 95(2), 245-258.

Accede al artículo a través del aula virtual o de la siguiente dirección web:

https://sociales.uexternado.edu.co/wp-content/uploads/sites/11/2019/07/S-2-Ai_Neuroscience.pdf

En este tema analizaremos en más detalle la interrelación entre neurociencias e inteligencia artificial, comprendiendo cómo se construyen las arquitecturas cognitivas artificiales a partir de los hallazgos neurocientíficos.

Principalmente veremos:

- ▶ Qué son los correlatos neurobiológicos de las funciones mentales.
- ▶ El salto explicativo entre procesos neurológicos y procesos psicológicos.
- ▶ Qué funciones cognitivas y afectivas se identifican en la mente.
- ▶ Cuáles son los enfoques clásicos para la construcción de arquitecturas cognitivas artificiales.

5.2. Introducción a la cognición y la ciencia cognitiva



Accede al vídeo «Multidisciplinariedad en ciencia cognitiva» a través del aula virtual



Accede al vídeo «Cómo funciona la mente» a través del aula virtual



Accede al vídeo «Los modelos bio-inspirados» a través del aula virtual



Accede al vídeo «Correlatos neuronales de la cognición» a través del aula virtual

La principal característica de la **ciencia cognitiva** es su marcado carácter **multidisciplinar**. Hoy en día no se concibe que se pueda obtener una comprensión profunda de la mente humana sin realizar un esfuerzo que involucre a diferentes campos científicos. Es más, la concepción actual de la mente implica que el comportamiento inteligente no nace únicamente del sistema nervioso, sino que lo hace en constante interacción con otros factores biológicos, psicológicos y sociales (enfoque **psicobiosocial**).

Algunas de las disciplinas científicas, muy relacionadas entre ellas, que suelen contribuir a la configuración actual de la investigación en ciencia cognitiva son:

- ▶ Las **neurociencias**, que también es un área multidisciplinar, cuya área de interés específico es la comprensión detallada del funcionamiento del sistema nervioso, incluyendo el análisis a nivel fisiológico, anatómico y neurobiológico.
- ▶ La **psicología**, que se centra en el estudio de la mente y el comportamiento, dando sentido al humano en su compleja interacción a niveles biológicos, mentales y sociales.

- ▶ La **filosofía de la mente**, que trata de responder a las grandes preguntas que la ciencia no ha podido aún resolver acerca de la naturaleza de la mente y su interacción con el cuerpo.
- ▶ La **antropología**, que estudia los patrones que ha emergido en la especie humana y sus manifestaciones culturales y sociales.
- ▶ La **lingüística**, que se centra en el estudio del lenguaje, su uso, el manejo del significado y su implicación en el desarrollo humano y su inteligencia.
- ▶ La **inteligencia artificial**, que supone un nuevo dominio (no biológico) donde aparecen procesos análogos a los humanos, de procesamiento de la información y resolución de problemas.

El objetivo último de la ciencia cognitiva es comprender **cómo funciona la mente**, analizando todos sus procesos internos. Específicamente, se aplica el método científico para estudiar:

- ▶ El **sustrato o base física** que hace posible el procesamiento de la información asociado al funcionamiento de la mente.
- ▶ El conjunto de diversas **tareas que puede realizar un ser inteligente** gracias a la aplicación del procesamiento de la información.
- ▶ Las funciones específicas (**funciones cognitivas**) que permiten que la mente desarrolle todo su potencial y haga posible la aparición de comportamientos inteligentes.

En este contexto, es de especial interés **poner foco en la neurociencia**, ya que los enfoques bio-inspirados están teniendo mucho éxito en la inteligencia artificial (el uso de **las redes de neuronas artificiales** y los enfoques de **aprendizaje profundo**). Por ese motivo, es interesante comprender cómo podemos hacer **ingeniería inversa de los sistemas cognitivos naturales**.

La neurociencia se preocupa por los procesos y las características que se pueden observar a nivel de **anatomía, biología molecular y fisiología** del tejido nervioso. El tejido nervioso está constituido por multitud de diferentes tipos de células, aunque

es habitual poner el foco en las neuronas (y las redes de interconexión que estas forman en estructuras nerviosas complejas como, por ejemplo, el cerebro).

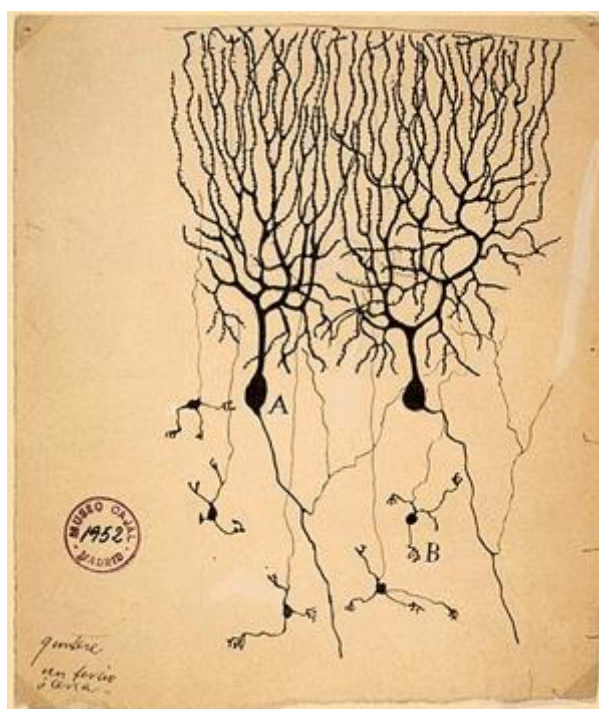


Figura 1. Reproducción de dibujo de las células del cerebelo de una paloma.
Fuente: Ramón y Cajal; S. (1899). Madrid: Instituto Cajal.

La neurociencia se centra en el estudio a diversos niveles del funcionamiento del sistema nervioso. A lo largo de la historia **el foco de la neurociencia se ha ido ampliando**. En los últimos años los objetivos giran en torno a **comprender cómo funciona el cerebro**, comprendiendo la relación existente entre los circuitos neuronales y la generación de percepciones, emociones, pensamientos y conducta.

En el siglo XIX, gracias a los trabajos pioneros de científicos como Ramón y Cajal o Golgi se empezó a **conocer la extrema complejidad que encierra el sistema nervioso**, descubriendo la estructura de las neuronas y las vastas redes de interconexión que forman en el cerebro. La neurociencia moderna hace uso de múltiples técnicas avanzadas de estudio del tejido nervioso, tanto *in vivo* como *post mortem*, posibilitando que cada vez conozcamos mejor cómo funciona nuestro propio cerebro.

El gran reto actual se centra en **descubrir y comprender los procesos que dan lugar a la extraordinaria complejidad mental y conductual** que exhiben los humanos y otros animales. Desde el punto de vista de la inteligencia artificial, este es un tema importante, pues se cree que el conocimiento del funcionamiento de los sistemas inteligentes naturales puede aplicarse, más o menos directamente, al **diseño y la construcción de máquinas inteligentes**. De hecho, es muy habitual que se usen en inteligencia artificial técnicas claramente bio-inspiradas, como las redes de neuronas artificiales.

Es importante tener presente que el conocimiento que tenemos hoy en día del funcionamiento del sistema nervioso es aún limitado. Todavía queda mucho por avanzar en la comprensión de los procesos que dan lugar a la generación del comportamiento inteligente y adaptativo.

Los **modelos computacionales bio-inspirados** con los que contamos hoy en día tienen principalmente dos limitaciones:

- ▶ Realizan **procesos mucho menos complejos** que los procesos reales que se dan en un sistema nervioso natural (las células nerviosas realizan procesos bioeléctricos y bioquímicos órdenes de magnitud más complejos que las simulaciones computacionales).
- ▶ Además de una simplificación algorítmica, también existe una **limitación en cuanto al modo de procesamiento**, ya que los sistemas artificiales actuales no alcanzan los niveles de paralelismo y distribución presente en los sistemas naturales.

El sistema nervioso de un humano cuenta aproximadamente con unos 86 000 millones de neuronas, estando cada una de ellas conectada con decenas de miles de otras neuronas, intercambiando mensajes de diversas características a una frecuencia de cientos de veces por segundo.

Se dice que el sistema nervioso es un sistema masivo de **Procesamiento Distribuido Paralelo** (PDP). Los sistemas artificiales **conexionistas** (redes de neuronas artificiales) tratan de imitar estos esquemas PDP, sin embargo, aún no se ha llegado a construir ningún sistema artificial con una capacidad de cómputo equivalente (en términos de paralelismo y número de conexiones).

La complejidad del sistema nervioso es tan grande que la tarea de estudiarlo se ha segmentado tradicionalmente entre diversas disciplinas. En otras palabras, el funcionamiento del cerebro es tan complejo que se ha ido abordando históricamente desde **diferentes campos científicos** y enfocando **diferentes niveles de descripción**.

La ciencia cognitiva moderna trata de integrar los diferentes campos de investigación y los diferentes niveles de descripción para proporcionar una visión más completa y explicativa de la mente humana.

En el nivel más fundamental contamos con los conocimientos de la **física para describir fenómenos relacionados con la interacción de la materia, el espacio, el tiempo y la energía**. Sin duda, todos estos fenómenos están también presentes en el sistema nervioso, pero **es necesario seguir ascendiendo en niveles de abstracción** que contemplen la interacción bioquímica para empezar a atisbar su funcionamiento. Por lo tanto, también recurrimos a la bioquímica y la biología molecular para ir descubriendo los procesos básicos que tienen lugar en el cerebro (así como en otros órganos y diferentes especies de seres vivos).

Sin embargo, los estudios de biología básica tampoco permiten comprender los procesos que ocurren en la interacción entre los miles de millones de neuronas, siendo necesario recurrir a la **neurobiología** y la neurología. La neurobiología se centra en el **estudio de las células nerviosas y la interacción que se produce entre ellas**. Concretamente, se analiza la organización de las neuronas en circuitos específicos que cumplen funciones de procesamiento de la información y generación de la conducta.

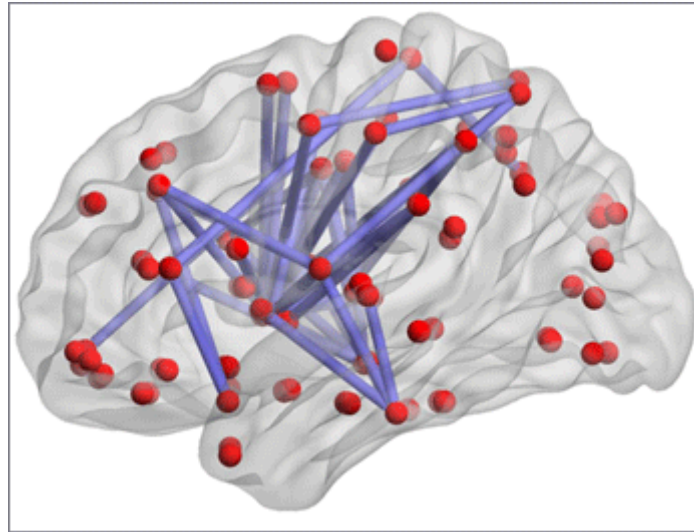


Figura 2. Representación simplificada de un circuito neuronal en el cerebro humano.

Fuente: Hong, S. B., Zalesky, A., Cocchi, L., Fornito, A., Choi, E. J., Kim, H. H. y Yi, S. H. (2013). Decreased functional brain connectivity in adolescents with internet addiction. *PloS one*, 8(2).

Aunque a nivel neurobiológico podamos comprender el funcionamiento detallado de ciertos mecanismos, seguimos teniendo problemas para establecer cómo se generan los pensamientos y los comportamientos complejos a partir de estas estructuras neuronales. Es decir, existe un nivel de descripción superior, relativo al ámbito de la **psicología**, que se ocupa de **estudiar la mente y el comportamiento**.

Como el resto de las disciplinas científicas que estamos repasando, la psicología cuenta con multitud de ramas y especializaciones, que a menudo tienen que ver con aspectos sociales, culturales, sanitarios, educativos, etc. Sin embargo, desde el punto de vista de la comprensión del funcionamiento del sistema nervioso, nos interesa centrarnos en la **psicología cognitiva**. Concretamente, desde el punto de vista de la construcción de sistemas de inteligencia artificial, es muy interesante analizar los procesos psicológicos básicos, como la **percepción, la atención, la emoción, la motivación, el aprendizaje**, etc. Así como otras funciones cognitivas superiores como **el pensamiento o el lenguaje**.

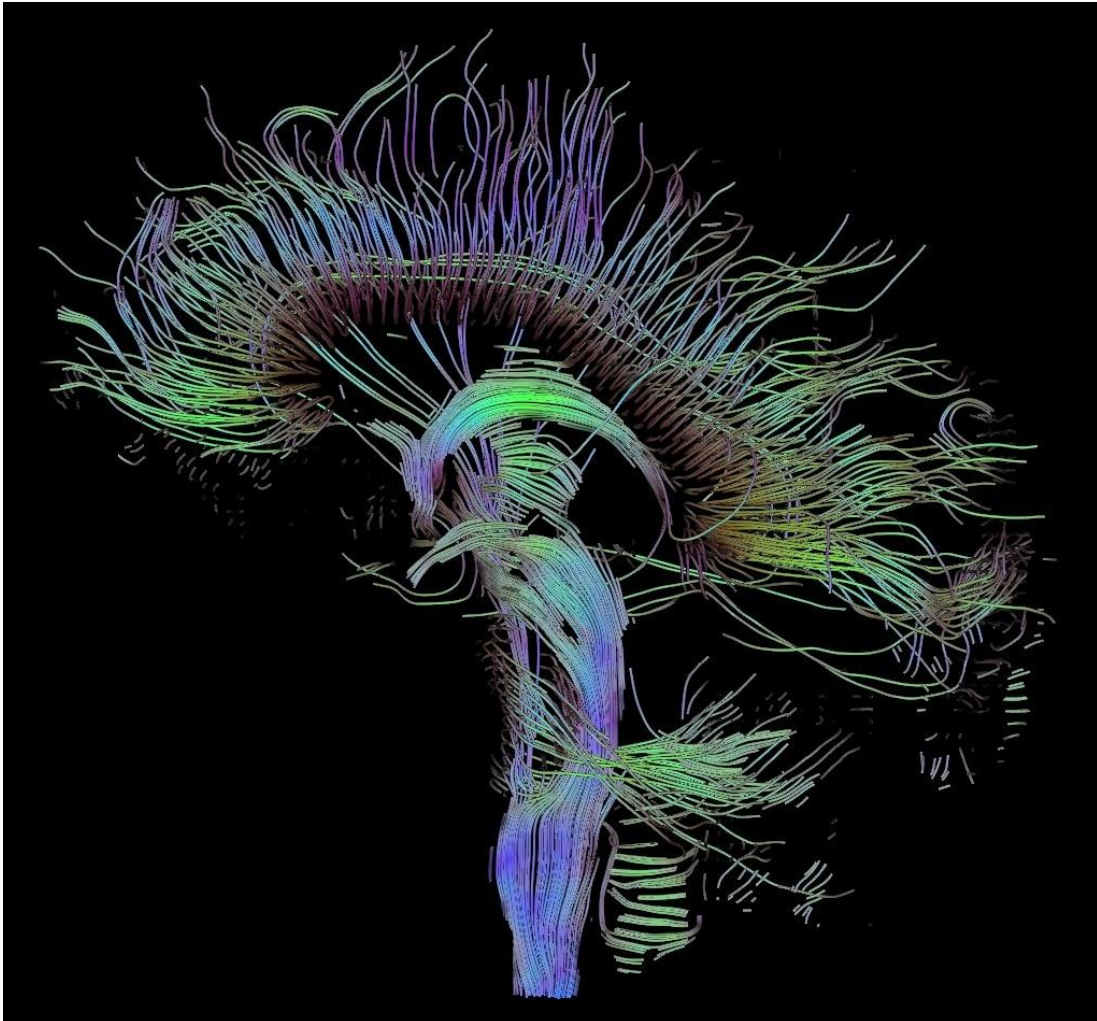
La psicología cognitiva se ocupa del estudio de estos procesos, proponiendo teorías y modelos que se ajusten en lo posible a la realidad empírica. Por otro lado, también es interesante **integrar estos modelos psicológicos con los conocimientos neurobiológicos** provenientes de otras disciplinas más cercanas a la realidad biológica del sistema nervioso. Por ese motivo, aparece la disciplina de la **psicobiología**, que pretende establecer puentes entre los procesos mentales y el estudio del cerebro, proporcionando explicaciones fundadas en la realidad neurobiológica.

Conocer la configuración específica de los circuitos neuronales humanos, comprendiendo cuál es su funcionamiento y el papel que desempeñan en los procesos mentales y la conducta, nos sirve para diseñar redes neuronales artificiales capaces de resolver problemas más complejos.

El estudio del sistema nervioso humano (y el de otros animales) nos permite descubrir qué circuitos neuronales intervienen en el desempeño de diferentes funciones mentales. Concretamente, los estudios neurocientíficos suelen identificar:

- ▶ **Áreas o zonas del cerebro** que están asociadas con funciones cognitivas o afectivas específicas. Esta asociación se establece en función de la activación observada en estas áreas cuando un sujeto está realizando una tarea. La información sobre la activación de las áreas se suele obtener a través de experimentos que usan técnicas de neuroimagen o analizando los síntomas de pacientes con daño cerebral localizado. De este modo sabemos, por ejemplo, que la amígdala está directamente involucrada en el procesamiento emocional de los estímulos y que el hipocampo se encarga de fijar los contenidos en la memoria a largo plazo.
- ▶ **Circuitos o vías neuronales** que establecen conexiones entre diferentes zonas del cerebro, identificando una arquitectura determinada de integración de las diferentes áreas especializadas. Estos circuitos neuronales se pueden identificar utilizando diversas técnicas de tractografía, como las imágenes de tensor de difusión. Por ejemplo, sabemos que la vía dopaminérgica mesolímbica (el llamado

“circuito de la recompensa”) está involucrada en la sensación de placer que se obtiene al obtener un refuerzo o saciar una necesidad.



Fuente: https://en.wikipedia.org/wiki/Diffusion_MRI#/media/File:DTI-sagittal-fibers.jpg

Reconstrucción tractográfica de las conexiones neuronales usando imágenes de tensor de difusión.

En general, tanto si identificamos áreas como circuitos neuronales, hablamos de **correlatos neuronales de una función cognitiva** para referirnos al sustrato físico existente en el sistema nervioso que hace posible que se produzca un proceso mental. Por ejemplo, la corteza prefrontal sería el correlato neuronal de la función de inhibición de la conducta, es decir, la capacidad de poder refrenar otros impulsos cuando la situación lo requiera.

5.3. Inteligencia artificial en la ciencia cognitiva



Accede al vídeo «Inteligencia artificial y neurociencia cognitiva» a través del aula virtual

No existe un enfoque único al estudio del sistema nervioso, sino todo lo contrario. Lidar con la complejidad tan grande del cerebro humano, y en general, el de cualquier otra especie, implica combinar efectivamente conocimientos y técnicas provenientes de múltiples disciplinas. Sin embargo, casi todas las líneas de investigación en ciencia cognitiva se basan en el análisis de los sujetos biológicos (humanos u otras especies de animales). Desde el punto de vista de la inteligencia artificial es muy interesante comprender y destacar el papel que juega la investigación en sistemas inteligentes artificiales en la comprensión de la mente.

Dentro del ámbito de la ciencia cognitiva, cabe destacar la línea de investigación en **neurociencia cognitiva**, una combinación de la inteligencia artificial, la psicología cognitiva y la neurociencia.

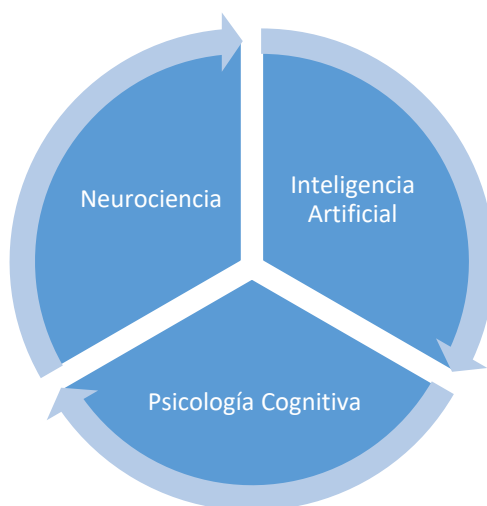


Figura 3. Neurociencia Cognitiva.

La neurociencia cognitiva es de especial interés porque se enfoca en la contribución y el papel que juega la construcción de sistemas artificiales. Es decir, no es un estudio

de la mente cuyas disciplinas son todas relacionadas con el dominio biológico, sino que **se incorpora el dominio de las máquinas inteligentes**.

Como hemos visto, comprender el sistema nervioso y cómo este es capaz de desarrollar funciones tan avanzadas como la inteligencia, el lenguaje o la conciencia requiere un esfuerzo conjunto de multitud de expertos en diferentes campos. Los diferentes niveles de descripción con los que se aborda el problema pueden finalmente proporcionar un camino útil a la comprensión si se combinan adecuadamente. Este apasionante camino de **descubrimiento de lo que es la mente no se puede completar sin tener en cuenta los desarrollos en inteligencia artificial**.

Por ejemplo, en el ámbito de la neurociencia cognitiva, los experimentos realizados con un sistema de inteligencia artificial basado en una teoría neurobiológica podrían contribuir al refinamiento de los modelos neurocientíficos considerados en dicha teoría.

Existe una **interacción constante entre las diversas disciplinas** que se engloban dentro del ámbito de las **ciencias cognitivas**. En este entorno multidisciplinar, disciplinas ingenieriles como la inteligencia artificial **son a la vez beneficiarias y contribuyentes** del avance de otros campos como la psicología o la neurobiología.

El estudio científico del sistema nervioso se ha intensificado en las últimas décadas. Existen varios factores que han contribuido a este auge, entre ellos el reconocimiento de la importancia de la salud mental y psicológica en todos los ámbitos de la vida de las personas. Sin embargo, quizás han sido los avances tecnológicos los que han hecho posible un resurgimiento contundente de las neurociencias. El estudio del sistema nervioso es ahora posible a un nivel que antes no podíamos ni tan siquiera soñar. Especialmente, gracias a las diversas **técnicas de neuroimagen** que han proliferado desde los años 80 y 90 del siglo pasado.

Desde el punto de vista de la inteligencia artificial es importante conocer en detalle el funcionamiento del sistema nervioso por varios motivos:

- ▶ Por un lado, los sistemas inteligentes naturales constituyen un elemento claro de **inspiración para el diseño de sistemas inteligentes artificiales** (bio-inspiración).
- ▶ Por otro lado, existe una tendencia actual a la combinación de las capacidades mentales humanas y artificiales, con lo que se empieza a considerar a la **inteligencia artificial como un elemento amplificador de las capacidades humanas**.

En general, podríamos decir que ya hay aspectos en los que las máquinas superan ampliamente las capacidades mentales humanas como, por ejemplo, jugar al ajedrez. Por otro lado, hay otros aspectos, como el aprendizaje orientado a tareas de propósito general, en el que los humanos superan claramente a las máquinas. En este contexto, cada vez más se plantea **la combinación y la colaboración humano-máquina**.

Tanto en el sentido de la inspiración como en el sentido de la integración de capacidades, es preciso conocer en detalle cómo funciona el sistema nervioso. Específicamente, nos interesa conocer **cómo emergen de la actividad neuronal funciones complejas como el aprendizaje** o la memoria semántica.

En definitiva, queremos conocer cuáles son las **bases neurobiológicas de los procesos mentales**. Necesitamos comprender qué procesos específicos tienen lugar en el sistema nervioso para producir percepciones adaptativas, aprendizajes efectivos y memorias semánticas virtualmente infinitas.

En este contexto es importante recordar el concepto de **correlatos neuronales**, que se refiere a la actividad subyacente del sistema nervioso que se detecta cuando un sujeto está realizando una determinada tarea mental o proceso físico.

Para avanzar en el desarrollo de la inteligencia artificial necesitamos disminuir o cerrar el **salto explicativo** que actualmente existe entre los diferentes niveles de análisis del sistema nervioso y de la mente.

Por un lado, contamos con modelos que tratan de explicar **cómo se producen fenómenos mentales** (provenientes mayoritariamente de la psicología) y, por otro lado, contamos con modelos que tratan de **explicar cómo se organizan y funcionan las redes de neuronas** (provenientes de disciplinas neurobiológicas).

La **psicobiología** (o neurociencia del comportamiento) busca aplicar los principios de la neurobiología al estudio de los mecanismos fisiológicos, genéticos y del desarrollo implicados en la generación del comportamiento en humanos y en otros animales. Como ocurre en general en el ámbito de las ciencias cognitivas, aquí también nos encontramos con orígenes provenientes de escuelas o disciplinas tradicionalmente dispares como la neuroanatomía, la medicina, la neuropsicología, etc. En general, todas las ciencias «neuro» están interrelacionadas y lo que aparecen son líneas de especialización que tratan de combinar aspectos específicos (por ejemplo, la neuropsicología se centra en comprender las funciones cerebrales para diagnosticar y tratar enfermedades neurológicas).

Para la investigación en inteligencia artificial es interesante trabajar con enfermedades mentales de origen neurológico pues constituyen una oportunidad para comprender el **funcionamiento anómalo del sistema nervioso**. De hecho, los sistemas actuales de inteligencia artificial se podrían describir como «mentes de funcionamiento anómalo», al ser comparadas con el funcionamiento natural del cerebro humano.

En general, la **neurociencia cognitiva**, agrupando e integrando los estudios en psicología, inteligencia artificial y neurociencia pretende disminuir el salto explicativo que hay entre el funcionamiento neuronal y el funcionamiento mental observado. Esta comprensión integradora del sistema nervioso es de vital importancia para el desarrollo de la inteligencia artificial, ya que permite replicar procesos análogos a los naturales en arquitecturas cognitivas artificiales.

5.4. Procesos básicos de la cognición humana y funciones cognitivas superiores



Accede al vídeo «Tipos de procesos cognitivos» a través del aula virtual

La cognición humana se basa en un proceso que transforma los datos de entrada (estímulos obtenidos a partir de los sentidos) en información, y posteriormente en conocimiento útil para la acción. En este apartado simplemente haremos una breve definición y clasificaremos las distintas funciones que dan lugar a ese complejo proceso. En los siguientes temas abordaremos en detalle cómo son y qué características tienen cada una de estas funciones.

Aunque la mente podría considerarse como un proceso unitario, es cierto que, para lidiar con la gran complejidad que presenta, se suele analizar desde **las diferentes funciones que realiza**. Es decir, podemos analizar la mente identificando **los procesos psicológicos** que forman parte de ella misma. En este sentido, se pueden distinguir diferentes tipos de procesos o funciones:

- ▶ **Funciones cognitivas.** Las funciones cognitivas son las que se asocian directamente con el procesamiento de la información y la manipulación de contenidos mentales. Ejemplos típicos de estos procesos mentales son la memoria, el pensamiento y el lenguaje.
- ▶ **Funciones afectivas o de activación.** Las funciones afectivas son las que se asocian directamente con el estado fisiológico y la activación corporal, es decir, las emociones y la motivación. Ejemplos típicos de este tipo de procesos mentales son el miedo, el asco, el hambre, la sed o la curiosidad.
- ▶ **Funciones Ejecutivas o Funciones Cognitivas Superiores.** Tanto las funciones cognitivas como las funciones afectivas necesitan ser gestionadas

adecuadamente, de forma que se orqueste en la mente una sintonía general y exista una sincronización de procesos. Las funciones ejecutivas y la metacognición cumplen este papel. Ejemplos típicos de funciones psicológicas superiores son el control de la atención y el cambio de contexto, la inhibición, la planificación a largo plazo.

5.5. Modelos computacionales basados en la cognición humana



Accede al vídeo «Arquitecturas cognitivas artificiales» a través del aula virtual

Dado que **la inteligencia artificial se interesa por desarrollar funciones cognitivas y afectivas** como las presentes en los humanos, los estudios de ciencia cognitiva son interesantes como fuentes de inspiración y comprensión de los mecanismos subyacentes a estas funciones. El conocimiento actual del funcionamiento del sistema nervioso se basa principalmente en un tipo específico de célula nerviosa: la neurona. Sin embargo, es preciso recordar que existen muchos más tipos de células en el sistema nervioso y que también necesitamos entender la dinámica de los circuitos neuronales que se forman (no solo comprender cómo funciona una única célula nerviosa).

Existe una gran variedad de neuronas, como las motoneuronas que se encargan de inervar los músculos o las neuronas localizadas en los sentidos, que traducen diversas magnitudes físicas en impulsos nerviosos. Se considera que **la neurona es la unidad mínima de procesamiento** de la información en el sistema nervioso y que las funcionalidades complejas del cerebro derivan de la organización de vastas redes de neuronas.

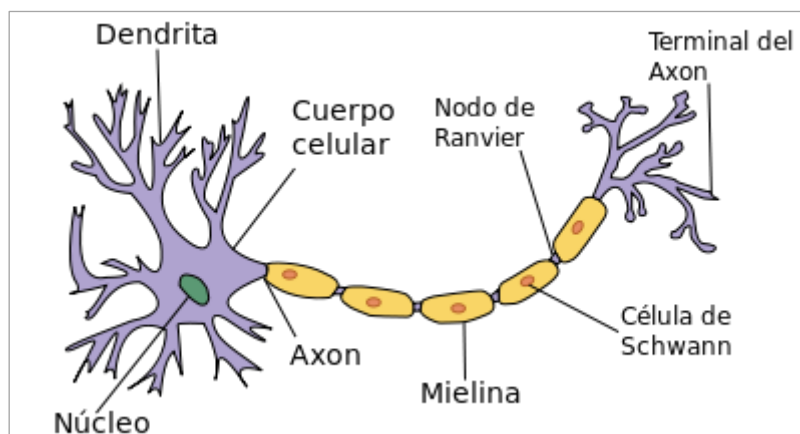


Figura 4. Esquema de una neurona.

Fuente: <https://commons.wikimedia.org/wiki/File:Neuron.jpg>

Una neurona típica se compone de un **cuerpo** (o soma), donde se encuentra el **núcleo**, **dendritas** y un **axón**. Las dendritas son estructuras finas que pueden extenderse cientos de micrómetros ramificándose hasta constituir complejas estructuras arbóreas (lo que se conoce como árbol dendrítico). El axón, al que también se le suele denominar fibra nerviosa (cuando está mielinizado), es una extensión que sale del cuerpo celular y puede llegar a longitudes de un metro en humanos.

La mayoría de las neuronas reciben señales a través de las dendritas y envían señales de salir a través del axón. Lo que comúnmente denominamos **nervios son en realidad haces de axones** agrupados que constituyen una especie de hebra constituida por múltiples «cables». Normalmente una neurona tiene muchas dendritas y un único axón (aunque el axón puede ramificarse también en su extremo final).

Las neuronas se comunican entre ellas a través de diversos mecanismos, pero el más conocido es la **sinapsis**. Existen a su vez diferentes tipos de sinapsis (eléctricas, químicas, etc.). En la mayoría de las sinapsis, las señales se envían por el axón de una neurona y llegan a las dendritas de otra neurona.

Todas las neuronas son eléctricamente excitables, ya que mantienen una diferencia de potencial a lo largo de sus membranas gracias a la acción de diferentes bombas

de iones. Los mecanismos electroquímicos y biológicos que dan lugar a la transmisión de los **impulsos nerviosos** son extremadamente complejos y variados, y aún no se conocen en todo detalle. No obstante, contamos con modelos que parecen ajustarse bien a la realidad y sirven para explicar gran cantidad de procesos. La forma de organización de estas neuronas y cómo forman **circuitos son la base, en última instancia, de los procesos cognitivos y afectivos.**

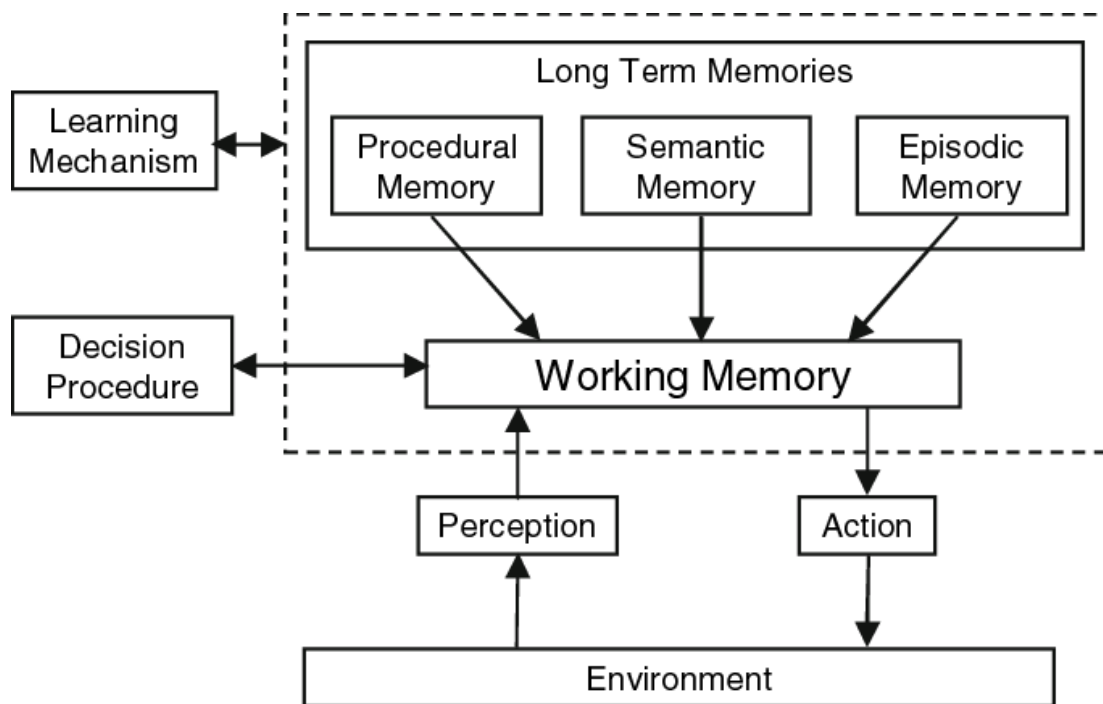
Además de los modelos conexionistas, basados exclusivamente en la simulación de los procesos neuronales, también contamos con **modelos de más alto nivel**. Estos modelos suelen denominarse **arquitecturas cognitivas artificiales**. Decimos que son modelos computacionales de más alto nivel porque no se inspiran exclusivamente en los mecanismos neurobiológicos, sino que pretenden emular también procesos psicológicos básicos, es decir, **funciones cognitivas y afectivas.**

Normalmente, las arquitecturas cognitivas artificiales permiten reproducir el procesamiento de la información asociado con:

- ▶ Adquisición de datos externos a través de **sensores de múltiples modalidades**, que transducen determinadas magnitudes físicas y las codifican como un patrón de datos. En el caso de los agentes software, que no tienen un cuerpo físico, los sensores pueden estar constituidos por sondas software, que también permiten obtener medidas y datos del entorno.
- ▶ **Sistemas actores, motores y/o de locomoción**, que permiten al sistema efectuar acciones que tienen un impacto en su entorno. En el caso de robots físicos, se trata de motores y actuadores, en el caso de agentes software pueden ser acciones en forma de comandos o mensajes.
- ▶ Mecanismos de **aprendizaje automático**, que permiten la adquisición de nuevo conocimiento, patrones o asociaciones a partir del análisis de los datos recibidos a través de los sensores.
- ▶ Gestión de la **memoria operativa**, manteniendo una dinámica de memoria a corto plazo y memoria a largo plazo gracias a una infraestructura basada en diferentes tipos de almacenes.

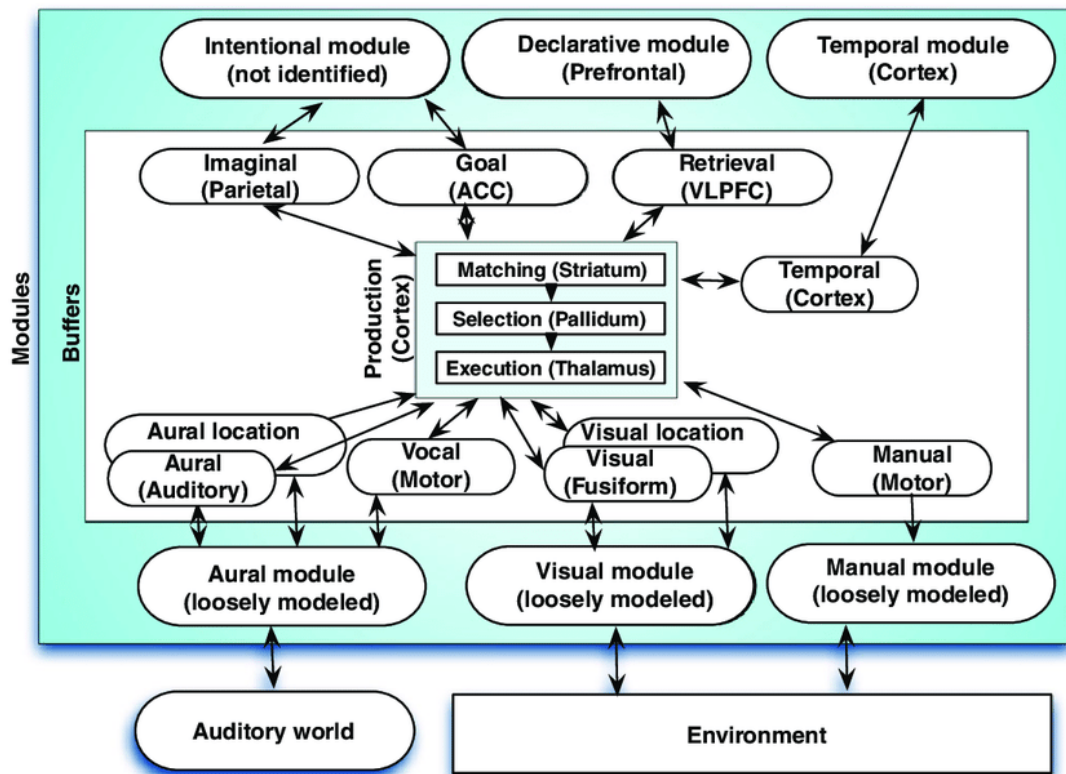
- ▶ Gestión de la **memoria semántica**, pudiendo administrar la codificación y la recuperación de información en forma de redes asociativas.
- ▶ Gestión de los **mecanismos de atención**, permitiendo que no todos los datos de entrada se procesen con la misma profundidad y decidiendo también qué estímulos o contenidos son más relevantes en cada caso.
- ▶ Mecanismos de evaluación de la situación, asociados o inspirados en los **procesos afectivos y emocionales** presentes en los organismos biológicos.

Dos arquitecturas cognitivas artificiales ampliamente usadas en la investigación son SOAR y ACT-R, aunque existen muchas otras propuestas más recientes, estas dos arquitecturas clásicas incorporan modelos computacionales correspondientes a los procesos cognitivos que se estudiarán en el resto de la asignatura.



[https://www.researchgate.net/publication/225257926 Integrated cognitive architectures A survey/figures?lo=1](https://www.researchgate.net/publication/225257926_Integrated_cognitive_architectures_A_survey/figures?lo=1)

Esquema de alto nivel de la arquitectura SOAR



https://www.researchgate.net/publication/329493100_ACT-R_A_cognitive_architecture_for_modeling_cognition/figures?lo=1

Esquema de alto nivel de la arquitectura ACT-R

5.6. Referencias bibliográficas

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American psychologist*, 51(4), 355.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4), 439-462.

Chong, H. Q., Tan, A. H., & Ng, G. W. (2007). Integrated cognitive architectures: a survey. *Artificial Intelligence Review*, 28(2), 103-130.

Goertzel, B., Lian, R., Arel, I., De Garis, H., & Chen, S. (2010). A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures. *Neurocomputing*, 74(1-3), 30-49.

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245-258.

Haines, D. E. (2014). *Principios de Neurociencia. Aplicaciones básicas y clínicas*. España: Elsevier.

Hong, S. B., Zalesky, A., Cocchi, L., Fornito, A., Choi, E. J., Kim, H. H. y Yi, S. H. (2013). Decreased functional brain connectivity in adolescents with internet addiction. *PloS one*, 8(2).

Kosslyn, S. M. (2010). *Wet mind: The new cognitive neuroscience*. Nueva York: Simon and Schuster.

Laird, J. E. (2012). *The Soar cognitive architecture*. MIT press.

Martínez, E. C. y Méndez, M. R. (2013). El sujeto desde la neurociencia y la inteligencia artificial. *Revista de estudios de juventud*, 103, 9-19.

Morgado, I. (2005). Psicobiología del aprendizaje y la memoria: fundamentos y avances recientes. *Revista de Neurología*, 40(5), 289-297.

Ramón y Cajal, S. (1899). *Reproducción de dibujo de las células del cerebelo de una paloma*. Madrid: Instituto Cajal.

Samsonovich, A. V., Klimov, V. V., & Rybina, G. V. (2016). Biologically inspired cognitive architectures (BICA) for young scientists. Springer.



Accede a los ejercicios de autoevaluación a través del aula virtual

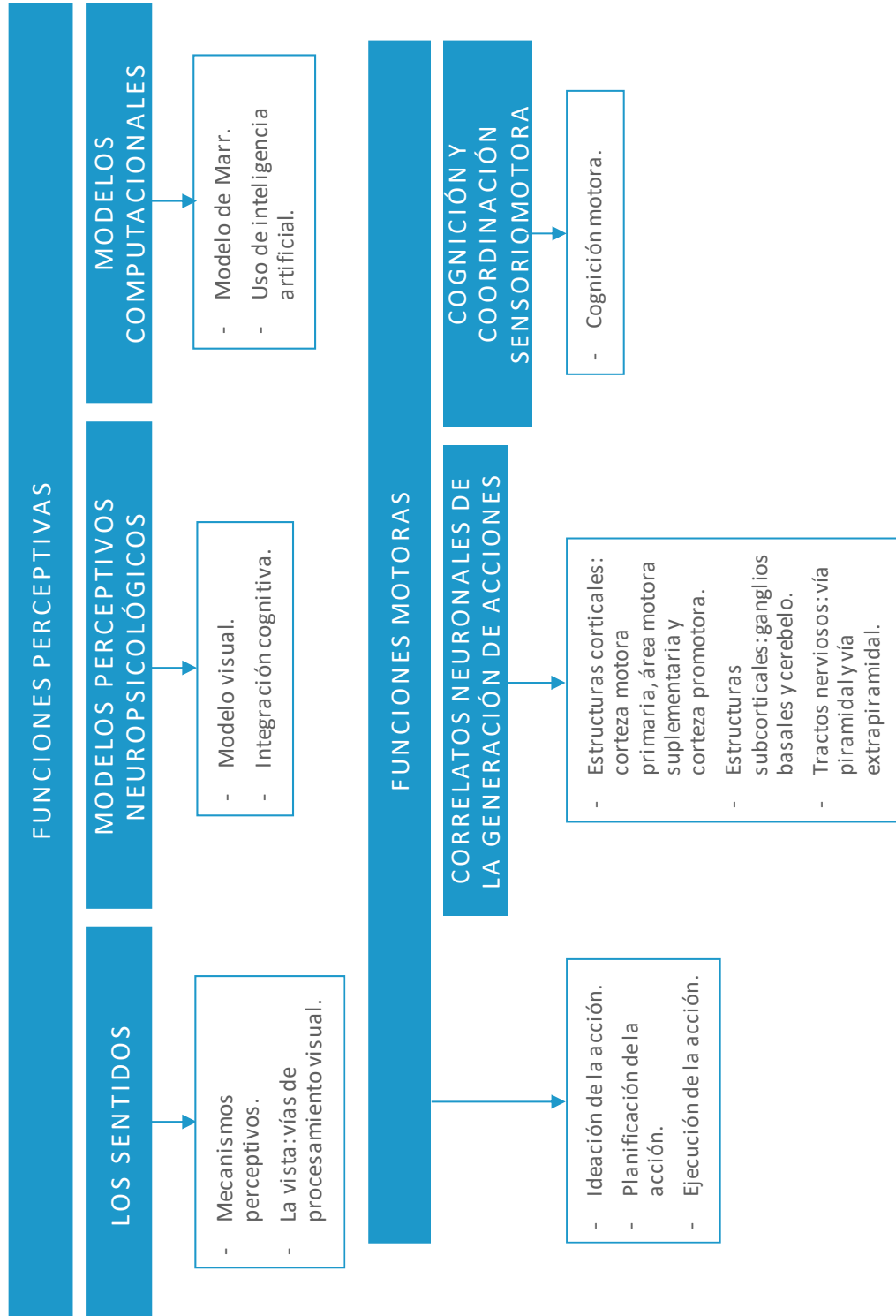


Accede al vídeo «Resumen del tema» a través del aula virtual

Inteligencia Artificial y Computación Cognitiva

Funciones perceptivas y motoras

Esquema. Tema 6



Ideas clave. Tema 6

6.1. Introducción y objetivos



Accede al vídeo «Introducción. Funciones perceptivas y motoras» a través del aula virtual

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

En este tema iniciamos el estudio de los procesos mentales que soporta el sistema nervioso. Como primer aspecto importante de la cognición se analizan los mecanismos para incorporar información del entorno en los procesos mentales: los mecanismos perceptivos. También nos centraremos en la producción de la conducta en los humanos, estudiando de forma superflua los complejos sistemas de coordinación sensoriomotora existentes en el sistema nervioso. Los principales objetivos de este tema son:

- ▶ Realizar un análisis pormenorizado del sistema perceptivo visual, ya que es el mejor estudiado y más predominante en los humanos.
- ▶ Comprender la relación existente entre los sistemas perceptivos y los sistemas motores del sistema nervioso.
- ▶ Conocer los modelos computacionales de referencia del procesamiento visual.
- ▶ Analizar el proceso de generación de la conducta motora en los humanos.
- ▶ Comprender la relación existente entre los sistemas perceptivos y los sistemas motores del sistema nervioso.
- ▶ Conocer los modelos de representación de acción y de coordinación entre sentidos y sistema locomotor.

En este tema pondremos foco en el sistema de percepción visual humano, siendo este un enfoque paradigmático e ilustrativo de la percepción humana. Así como en la representación de las acciones, ya que este aspecto enlaza el procesamiento cognitivo con la implementación de la conducta. Así mismo, el análisis de los modelos del funcionamiento motor en humanos servirá de inspiración para la construcción de modelos computacionales orientados al control de sistemas autónomos.

6.2. Mecanismos perceptivos



Accede al vídeo «¿Qué son los sentidos y cómo funcionan?» a través del aula virtual

Los **mecanismos perceptivos** del cuerpo humano permiten obtener información relevante para la supervivencia **tanto del entorno exterior como del interior del cuerpo** (sentidos exteroceptivos e interoceptivos respectivamente). Como sabemos, el cerebro no se limita a recibir e interpretar los estímulos de forma pasiva, sino que **transforma y elabora la información recibida** para generar una representación útil y adaptativa del mundo. Los humanos contamos con diferentes sistemas senso-perceptivos que se ocupan de la función perceptiva, y aunque cada uno se ocupa de una modalidad sensorial diferente (gusto, tacto, dolor, temperatura, vista, oído, equilibrio, etc.), todos tienen una estructura funcional similar.

En el sistema nervioso humano los sentidos son independientes a nivel funcional y anatómico. Es decir, tienen sus propias estructuras (sentidos propiamente dichos) y una **forma análoga de transmitir la información** de la modalidad sensorial correspondiente al cerebro (las vías aferentes sensitivas).

Los sistemas perceptivos exteroceptivos captan la información del exterior a través de receptores sensoriales. Estos receptores o sensores están diseñados para **transducir** una forma específica de energía en impulsos eléctricos. Por ejemplo, los conos y los bastones de la retina captan las variaciones de energía electromagnética

correspondiente al espectro visible y las traducen en patrones de activación en forma de impulsos eléctricos. El proceso de **transducción sensorial** transforma la medida de una magnitud física en potenciales de acción.

Este patrón de activación en forma de potenciales de acción se transfiere a células de segundo orden que se encargan de transmitir la información hacia el encéfalo. El **tálamo es el centro neuronal donde se recibe la información proveniente de todos los sentidos** (a excepción del olfato, que cuenta con una vía más directa de acceso a la corteza olfatoria). Después de un procesamiento inicial en diferentes núcleos talámicos, la información se propaga hacia la corteza primaria, y finalmente llega a la corteza de asociación. Los núcleos subcorticales se asocian con el procesamiento inconsciente de la información sensorial, mientras que las áreas corticales somatosensoriales se asocian con la percepción consciente.

En general, el procesamiento de la información sensorial sigue un esquema jerárquico y paralelo, que se correlaciona con los procesos que dan lugar a la percepción a partir de la sensación.

La información que proporcionan los sentidos es de vital importancia para el organismo. En el caso de los humanos, **la visión ocupa un puesto central entre todos los sentidos**. Prueba de ello es que el procesamiento visual está presente en una gran cantidad de espacio en el cerebro. Además de las principales vías de procesamiento comentadas anteriormente, la visión humana implica procesos muy complejos de coordinación motora, tanto del cuello como de los propios ojos, etc. Pero para poder generar los procesos motores que nos permiten, por ejemplo, manipular objetos es necesario que previamente reconozcamos con éxito los objetos que nos rodean.

Desde el punto de vista cognitivo, los objetos que percibimos visualmente han de tener una **representación** de sus estructuras en el cerebro, de forma que también podamos tener una representación mental de los mismos. Gracias a estas representaciones mentales podemos comparar los objetos que ya conocemos con los

objetos que vemos en un momento dado y ser capaces, entre otras cosas, de reconocerlos.

Los sentidos, como el de la vista, **se integran con el resto de las funciones cognitivas** para poder generar conductas complejas, como las de perseguir una presa o pelar una manzana. La percepción visual implica, entre otras muchas cosas, dónde mirar, qué hacer y qué buscar. Por lo tanto, la conducta final se produce como resultado de la interacción de diversas funciones cognitivas e incluso afectivas. Aunque aquí tratamos el tema de la visión desde un punto de vista aislado, es conveniente recordar que los procesos perceptivos operan en constante integración con procesos de memoria, atención, emoción, aprendizaje, etc.

Por ejemplo, es impensable concebir el sistema visual humano sin un sistema de atención. Decidir dónde se fija la mirada, qué buscar en el campo visual, identificar objetos, etc. son procesos que ocurren gracias a la **integración de los mecanismos perceptivos y los mecanismos atencionales**. Los ojos tienen un mecanismo de escáner que permite mover estos en determinadas direcciones. Dependiendo de la información que se obtiene en cada instante se tomarán decisiones (a menudo inconscientes) sobre dónde ha de posarse la mirada en los momentos posteriores.

6.3. Modelos computacionales de la visión



Accede al vídeo «Funcionamiento de la visión» a través del aula virtual



Accede al vídeo «Modelos cognitivos y computacionales de la visión» a través del aula virtual

La construcción de modelos computacionales de la visión parte de la base de que se puede considerar la percepción visual como un procesamiento de la información, gracias al cual podemos realizar operaciones como el reconocimiento de objetos.

Para reconocer un objeto se requiere tener una representación tridimensional del mismo. Marr (1982) propuso un modelo basado en tres etapas para construir dicha representación.

El modelo de Marr está basado en un enfoque modular.

- Una representación inicial (el *bosquejo inicial* o *primal sketch*) referida a cambios de intensidad, brillo, sombras, contornos, etc.



Figura 1. Imagen original (izquierda) y detección de bordes (derecha)

Fuente: <https://es.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/51124/versions/1/screenshot.jpg>

- Una representación centrada en el observador (el **esquema en $2\frac{1}{2}D$**), donde se aprecian las partes del objeto que son visibles desde el punto de vista del observador.

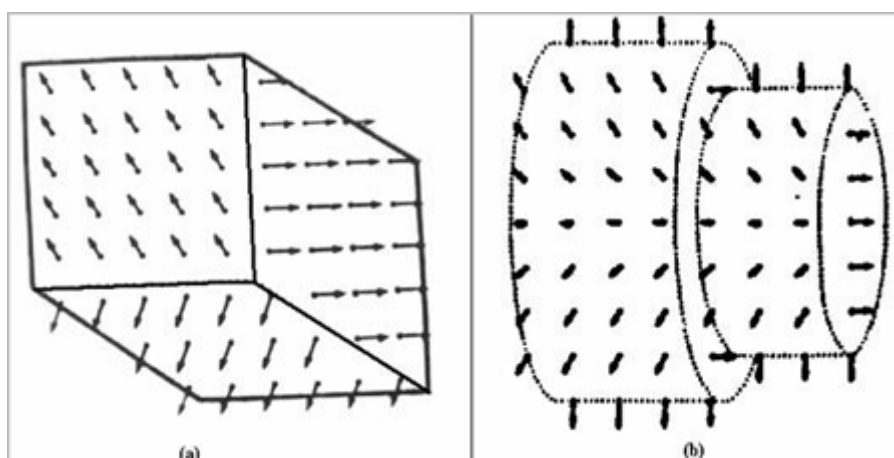


Figura 2. Esquema en $2\frac{1}{2}D$ de un cubo (a) y de dos cilindros acoplados (b)

Fuente: <http://www.ub.edu/pa1/node/93>

- Una representación centrada en el objeto (el **modelo 3D**), donde, independientemente de la posición del observador, se obtiene la forma real del objeto y su relación con otros objetos.

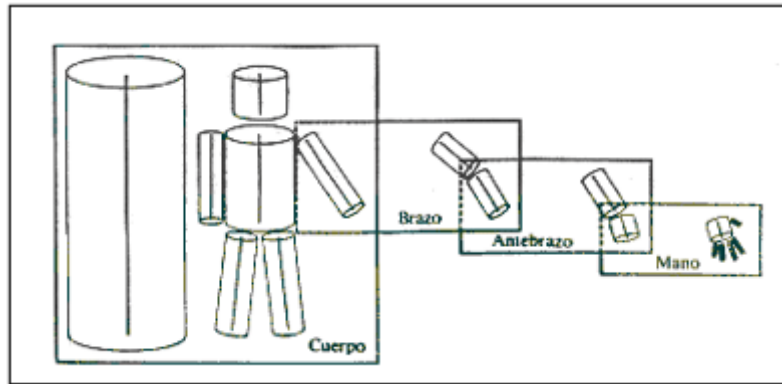


Figura 3. Representación 3D de una figura humana.
Fuente: Marr y Nishihara (1978).

Ellis y Young (1992) proponen un modelo similar al de Marr:

- Primer nivel: representación inicial del objeto en sí mismo.
- Segundo nivel: representación centrada en el observador.
- Tercer nivel: acceso del observador a las representaciones centradas en el objeto.
- Cuarto nivel: reconocimiento de las propiedades y atributos del objeto.

Según estos autores, contamos con una unidad de reconocimiento de cada objeto conocido, que puede acceder a una memoria semántica cuando se produce el reconocimiento. Hoy en día, se aboga por una representación mucho más distribuida basada en las características de los objetos, prescindiendo de la necesidad de contar con un módulo específico para cada objeto (lo cual sería bastante ineficiente y limitaría el número de objetos que se podrían detectar).

Una de las contribuciones más interesantes de estas teorías es la caracterización computacional que hacen del problema de la percepción visual. Por ejemplo, Marr (1982) hace una distinción de los distintos niveles en los que se puede describir la tarea de percepción visual:

- ▶ Nivel de teoría computacional: definición del problema de procesamiento de la información, cuya solución es la meta de la computación. Caracterización de las propiedades abstractas de la computación. Descubrimiento de las propiedades del mundo visible que restringen el problema computacional.
- ▶ Nivel algorítmico: estudio de los algoritmos que se pueden usar para realizar las computaciones deseadas.
- ▶ Nivel de implementación: realización física del algoritmo para un *hardware* especificado. Arquitectura de la máquina que ejecuta el algoritmo.

Esta clasificación es interesante desde el punto de vista de la inteligencia artificial, pues cualquier sistema de visión artificial necesariamente ha de tener en cuenta todos estos factores. De hecho, es importante hacer la distinción de estos niveles de descripción porque **permite desacoplar el problema del algoritmo que lo pretende solucionar**, y en última instancia de la implementación que lo hace funcionar.

Lo más importante de estos modelos es que nos brindan una explicación plausible de cómo funciona la percepción visual en los humanos, de forma que podemos iniciar **procesos de bioinspiración basados en estos modelos**. Marr describe el proceso de percepción visual como una agrupación inicial de líneas, curvas, manchas, fragmentos, etc. (lo que desde el punto de vista de inteligencia artificial podemos denominar características de nivel bajo). El análisis de estos estímulos genera la creación de elementos primitivos en cada escala, esto permite crear un bosquejo (lo que desde el punto de vista de aprendizaje automático podemos denominar detección de patrones). Así, se continúa ascendiendo por capas de procesamiento que se basan en los resultados de capas anteriores (lo cual se asemeja bastante al procesamiento que se realiza actualmente en las técnicas de Deep Learning).

6.4. Organización del sistema motor



Accede al vídeo «Control e integración de la acción» a través del aula virtual



Accede al vídeo «Mecanismos neurofisiológicos de la acción motora» a través del aula virtual

El **sistema motor humano** se basa en los músculos, que a su vez controlan uno o varios efectores. Llamamos **efectores** a las partes del cuerpo que podemos mover. Los músculos se conectan con el sistema nervioso a través de las **motoneuronas alfa**. Estas parten de la espina dorsal y a través de la vía ventral llegan a las fibras musculares, que inervan mediante **uniones neuromusculares**. Las motoneuronas reciben entradas de diferentes fuentes, incluyendo **niveles de control inferior**, como las aferencias de fibras sensoriales musculares, y de **niveles de control alto**, provenientes del cerebro a través de las fibras descendentes.

El sistema motor **se organiza en una jerarquía** con procesamiento distribuido y **diferentes niveles de control**. Cada nivel de control se corresponde con una región específica de la corteza motora. La idea del movimiento se genera en los niveles superiores, que se encuentran en el área motora suplementaria y la corteza premotora. Más tarde, los niveles inferiores de control, que se sitúan en la médula espinal, convierten las órdenes motoras en contracciones específicas de los músculos. Por otro lado, la médula espinal es capaz de realizar movimientos sin intervención de niveles superiores, como es el caso de los **reflejos motores**.

La jerarquía de control se hace más patente en **los movimientos más complejos**. Los movimientos reflejos no requieren intervención de niveles superiores de procesamiento ni estructuras complejas, por lo que se desarrollan independientemente a nivel medular. Sin embargo, la mayoría de los movimientos complejos requieren la **coordinación de múltiples estructuras**, entre las que

destacan las áreas corticales (corteza premotora, área motora suplementaria y corteza motora), el cerebelo, los ganglios basales y la médula espinal.

Los movimientos complejos son compuestos, lo que significa que requieren una **secuencia específica de acciones motoras**. Estas secuencias no son la simple unión de movimientos independientes, sino que están coordinadas gracias a las representaciones jerárquicas que existen en todo el sistema motor. Todos los componentes de un movimiento complejo se integran de forma efectiva gracias a la coordinación entre las diferentes áreas involucradas.

Las acciones están encaminadas a la consecución de metas. Por lo tanto, toda acción puede considerarse como guiada o desencadenada por la necesidad de cumplir una meta. En este sentido podemos distinguir dos tipos de metas:

- ▶ Metas internas: motivación interna (del propio individuo), intención o voluntad.
- ▶ Metas externas: locomoción, manipulación de objetos o interacción con personas.

La verdadera diferencia entre metas internas y externas reside en los sucesos de ocurren antes de la selección de la acción. Las metas internas se generan por **estímulos internos**, mientras que las metas externas provienen de **estímulos externos**. La secuencia desde la estimulación hasta la ejecución de una acción sigue las siguientes fases:

1. Estímulo (externo o interno).
2. Selección de la acción.
3. Planificación.
4. Programación motora.
5. Ejecución de la acción.

Por otro lado, **cada meta** puede necesitar de la ejecución de **múltiples acciones** para conseguirse, y también puede ser que una misma acción sirva a más de una meta, con lo que **no existen una relación de uno a uno entre acciones y metas**. Además,

una misma acción puede realizarse de diferentes formas y un mismo movimiento puede contribuir a la consecución de varias metas.

La planificación del movimiento se realiza en el **área motora suplementaria (AMS)** y la **corteza premotora**. El AMS está relacionada con la preparación, selección y decisión de comenzar un movimiento. Por otro lado, la corteza premotora envía órdenes a la corteza motora primaria (CMP).

En la corteza motora es donde se produce la integración y la transformación de la entrada sensorial, así como la planificación y ejecución de la respuesta motora. La CMP o M1 es **el origen de la vía piramidal** que proyecta a la parte contralateral del cuerpo y conecta las motoneuronas con los músculos.

Inicialmente, el AMS genera un **plan abstracto** de movimiento, después la corteza premotora traduce ese plan en una **codificación concreta de determinadas acciones** motoras y finalmente la CMP se encarga de ejecutar las **órdenes musculares**.

6.5. Referencias bibliográficas

Ellis, A. W. y Young, A.W. (1992). *Neuropsicología cognitiva humana*. Barcelona: Masson.

Jacob, P. y Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in cognitive sciences*, 9(1), 21-25.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self* (No. 42). Oxford: Oxford University Press.



Accede a los ejercicios de autoevaluación a través del aula virtual

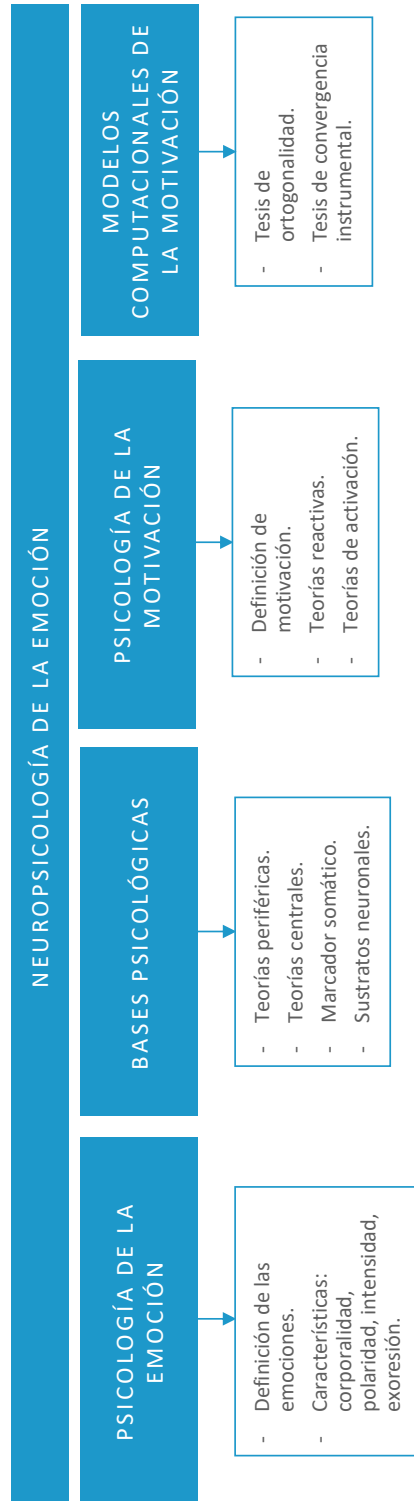


Accede al vídeo «Resumen del tema» a través del aula virtual

Inteligencia Artificial y Computación Cognitiva

Emoción y motivación

Esquema. Tema 7



Ideas clave. Tema 7

7.1. Introducción y objetivos



Accede al vídeo «Introducción. Emoción y motivación» a través del aula virtual

Para estudiar este tema lee las Ideas clave disponibles a continuación y:

Las páginas 9 a 21 del siguiente artículo: Vila, J. (2016). Neurociencia afectiva: Entre el corazón y el cerebro. Discurso de apertura, Universidad de Granada, curso académico 2016-2017.

Accede al artículo a través del aula virtual o de la siguiente dirección web:

http://digibug.ugr.es/bitstream/10481/42748/1/B-032-066_7.pdf

Barberá, E. (2002). Modelos explicativos en psicología de la motivación. *Revista electrónica de motivación y emoción*, 2(1).

Accede al artículo a través del aula virtual o de la siguiente dirección web:

<http://reme.uji.es/articulos/abarbe7630705102/texto.html>

En este tema nos centramos en la dimensión afectiva y emocional. Aunque la neurociencia cognitiva ha parecido centrarse tradicionalmente en los procesos cognitivos, en realidad la emoción siempre ha estado incluida en la agenda de investigación. De hecho, hay autores que proponen el uso del término «neurociencia afectiva» para dejar claro que su interés va más allá de los procesos cognitivos. Los principales objetivos de este tema son:

- ▶ Entender la diferencia entre los procesos psicológicos cognitivos y afectivos.
- ▶ Comprender los procesos básicos en psicología de la emoción.

- ▶ Analizar la representación computacional que se hace de la emoción, estableciendo las diferencias principales con los procesos biológicos.
- ▶ Comprender los procesos básicos en psicología de la motivación.
- ▶ Analizar la representación computacional que se hace de la motivación, estableciendo las diferencias principales con los procesos biológicos.

En este tema haremos hincapié en una caracterización pragmática de las emociones, sin entrar en detalle en la dimensión de los sentimientos (las representaciones conscientes de las emociones), puesto que la consciencia se tratará en detalle en otro tema. También, pondremos foco en el análisis de los procesos motivacionales. La construcción de modelos computacionales en este apartado está menos evolucionada y queda mucho trabajo de investigación por hacer, por lo que se repasarán algunos de los pocos trabajos existentes en este dominio.

7.2. Psicología de la emoción



Accede al vídeo «Las emociones y sus propiedades» a través del aula virtual



Accede al vídeo «Base neurofisiológica de las emociones» a través del aula virtual

Las emociones juegan un **papel vital** en los mamíferos, proporcionándonos un conocimiento inmediato y fundamental para la **supervivencia y la adaptación al medio**. En esencia, las emociones son una guía instantánea para la conducta, que sin pensar (sin entrar en procesos cognitivos de orden superior) nos permiten actuar en un instante determinado en función de la situación. Podríamos decir que las emociones también nos ayudan a saber qué es lo que tenemos que hacer ante una situación, pero ese conocimiento explícito ya implica un nivel de consciencia relativo al procesamiento cognitivo de la emoción.

Para no mezclar las respuestas emocionales básicas con otros procesos mentales más elaborados, es preciso **distinguir entre emoción y sentimiento**. Mientras que la emoción ocurre a nivel inconsciente y en escalas temporales muy cortas, los **sentimientos son percepciones conscientes** cuyo proceso lleva más tiempo. Simplemente podemos decir que los sentimientos son elaboraciones conscientes realizadas a partir de las emociones que siente el cuerpo. Por lo tanto, las emociones sirven en muchas ocasiones como guía para la acción, que no requiere de un procesamiento mental explícito (por ejemplo, no se requiere el uso del lenguaje para asustarse ante la presencia de un depredador).

Aunque hemos definido las emociones como un **mecanismo de adaptación al medio**, ciertamente en los humanos y en la cultura actual aparecen respuestas emocionales que parecen ser no adaptativas. Y de hecho muchas patologías mentales derivan de la existencia de **respuestas emocionales desadaptativas**. Se ha argumentado que esto es debido a que las emociones en los humanos han evolucionado para adaptarse a un entorno muy distinto al que tenemos hoy en día. Por lo general, consideraremos en este tema las emociones en su faceta adaptativa, dejando de lado todos los aspectos relacionados con la psicopatología de la emoción (como, por ejemplo, la depresión). En los humanos, el entorno social es una fuente principal de estimulación que provoca emociones. Como seres sociales que somos una parte importante de nuestra adaptación al medio implica las **relaciones con nuestros congéneres**.

Las emociones son eminentemente físicas. Una primera definición de emoción podría referirse a un **patrón específico de activación fisiológica**. Por lo tanto, la **dimensión corporal** cobra un aspecto de gran relevancia en cuanto a las emociones se refiere. Las emociones son en esencia estados internos del organismo que producen **tendencias específicas a la acción**. Simplificando, las emociones generan dos tipos básicos de conductas: de **acercamiento** y de **alejamiento**. Tendemos a alejarnos de aquellos objetos que nos provocan miedo y tendemos a acercarnos a aquellos objetos o sujetos por los que sentimos amor.

En el caso del miedo y del amor, siendo emociones básicas y muy relevantes, es fácil comprender cuál es la tendencia a la acción que provocan. No obstante, **una misma emoción puede dar lugar a diferentes conductas**. Es bien conocido que ante un miedo intenso el organismo puede responder huyendo, atacando o paralizándose (son las famosas respuestas de «**lucha o huida**»). En el caso de otras emociones es un poco menos directo entender la utilidad de la tendencia a la acción que provocan. Por ejemplo, la tristeza genera una conducta mucho más lenta, eliminando la mayoría de las actividades que realizaba el organismo habitualmente. Se piensa que la función de la **tristeza** es forzar una **ralentización del ritmo de actividades del organismo**, para favorecer que aparezca un proceso reflexivo profundo sobre un problema (o una pérdida) que necesita ser resuelto.

Entender las emociones requiere atender a su multidimensionalidad. Por un lado, tenemos diferentes emociones: amor, odio, ira, alegría, tristeza, etc. Cada una de ellas cuenta con un **patrón de activación fisiológico característico** (aunque a veces dos emociones distintas pueden tener en común respuestas fisiológicas, por ejemplo, tanto la ira como la alegría aumentan el ritmo cardíaco). Por eso, la interpretación consciente de las emociones también tiene aspectos que han de aprenderse (no son puramente innatos). Sobre todo cuando hablamos de emociones que no son básicas, es decir, **emociones secundarias** que surgen como combinación de las emociones básicas (envidia, orgullo, etc.).

A parte de la identificación de la propia emoción, existen otras dimensiones interesantes para clasificar las emociones. Un aspecto obvio, aunque a menudo mal entendido, es la **polaridad**. Se habla de emociones **positivas** y emociones **negativas** en función de si provocan estados placenteros o displacenteros. Por lo tanto, se dice que la ira, el miedo o el odio son emociones negativas y el amor y la alegría son emociones positivas. Esto no implica en absoluto que sean buenas o malas, en principio **todas las emociones son igualmente adaptativas y necesarias** para la supervivencia (de hecho, en intento de represión de las emociones negativas suele crear problemas psicológicos e incluso psicopatologías).

Otra dimensión importante acerca de las emociones es la intensidad. Podemos experimentar una emoción en diferentes niveles de intensidad, podemos estar ligeramente enfadados o muy enfadados. Por lo tanto, para definir un perfil emocional mínimo tendríamos que hablar de la identificación de la emoción, su polaridad y su intensidad.

En cuanto a la duración en el tiempo, **las emociones** se consideran con una **duración media**, del orden de minutos. Siendo las reacciones instantáneas **respuestas afectivas** y el estado emocional general que se desarrolla a lo largo de días el **estado de ánimo**.

En general, el estudio de las emociones requiere tener en cuenta al menos los siguientes aspectos:

- ▶ «Corporalidad»: patrones de activación fisiológica asociados.
- ▶ Sentimientos: percepción subjetiva de la emoción.
- ▶ Identificación de la emoción: emociones primarias y secundarias.
- ▶ Polaridad: positivas y negativas.
- ▶ Intensidad: nivel de intensidad de la emoción.
- ▶ Expresión: las emociones se expresan a través del rostro, la voz y el lenguaje no verbal.



Figura 1. Expresión de las emociones a través del rostro.

Fuente: <https://psicologiaymente.net/psicologia/emociones-basicas-cuatro-seis>

En cuanto a la identificación de las emociones, se han descrito una serie de **emociones primarias o básicas** y otro conjunto de **emociones secundarias** (estas listas de emociones varían ligeramente en función de los autores, pero existe un consenso bastante amplio acerca de las principales emociones básicas):

- ▶ Emociones básicas: tristeza, alegría, miedo, ira, sorpresa y asco.
- ▶ Emociones secundarias: culpa, vergüenza, envidia, humillación, desprecio, etc.

Existen multitud de clasificaciones diferentes (algunos autores sostienen que solo hay cuatro emociones primarias) y los autores argumentan sobre cómo aparecen las emociones secundarias, si son combinaciones de las primarias en forma de diadas o si tienen otros procesos de formación más complejos.

7.3. Modelos computacionales de la emoción

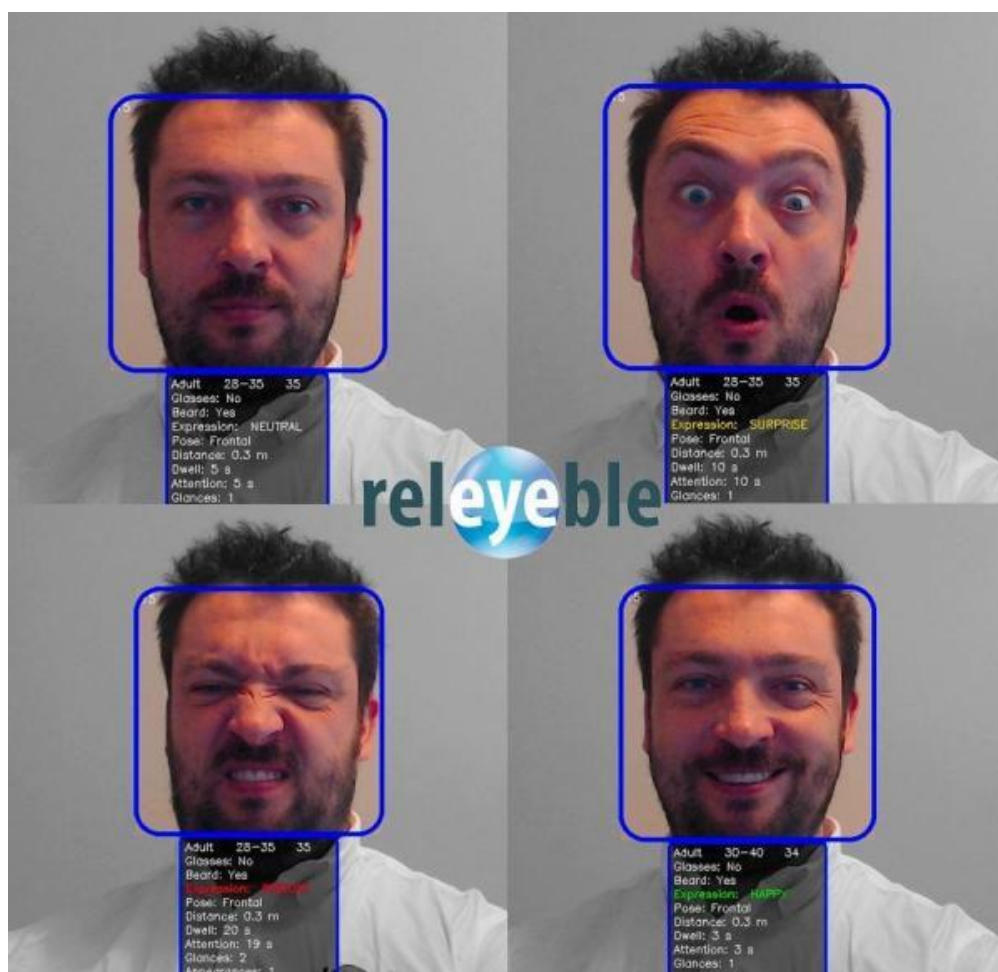
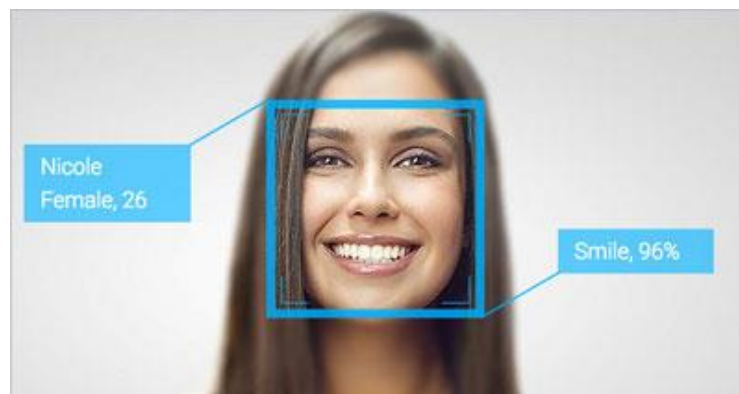


Accede al vídeo «Implementación computacional de la emoción» a través del aula virtual

En la actualidad contamos con **innumerables modelos computacionales** de la emoción. En general, todos se basan en los principales componentes estudiados sobre la emoción (polaridad, tendencia a la acción, etc.) **evitando la dimensión de la percepción subjetiva** y la consciencia de la emoción.

En los modelos computacionales se representa habitualmente **la emoción como una etiqueta** que toma valores positivos o negativos y se usa para guiar las respuestas del sistema. La etiqueta emocional se calcula como un indicador del **grado de consecución de los objetivos del sistema**. De esta forma, si el funcionamiento del sistema es óptimo y las metas se van alcanzando según lo previsto se activará una etiqueta de «alegría», mientras que, si las metas no se cumplen según el plan, aparecerá un indicador de «tristeza». Este tipo de simulaciones de las emociones puede **ser útil de cara a la interacción con humanos**, ya que puede dotar al sistema artificial de un modelo de la expresión de las emociones.

Por otro lado, también se cuenta con multitud de sistemas automáticos de **reconocimiento de la emoción en los humanos**. Para ello se usa primordialmente el reconocimiento de la expresión facial y se infiere el estado emocional de la persona. De nuevo, este tipo de sistemas automáticos sirven de base para mejorar la interacción entre humanos y máquinas.



Figuras 2 y 3. Ejemplos de sistemas automáticos de reconocimiento de la expresión facial.

Fuente: <https://nordicapis.com/20-emotion-recognition-apis-that-will-leave-you-impressed-and-concerned/>

7.4. Psicología de la motivación



Accede al vídeo «La motivación y sus propiedades» a través del aula virtual



Accede al vídeo «Teorías de la motivación» a través del aula virtual

La **motivación** es uno de los **procesos psicológicos básicos**, que, junto con la emoción, se suele asociar a la activación del organismo y la tendencia a la acción. En general, se asocia motivación con conducta, en el sentido de que la motivación es un proceso que dispone para acciones específicas. Es decir, **la motivación proporciona energía y dirección para la conducta**.

El concepto de energía se refiere a que en un momento determinado estamos **motivados para realizar ciertas conductas, pero no para hacer otras**. En otras palabras, nuestro organismo es capaz de encontrar energías para ir al bar con los amigos, pero no parece tener energía para ponerse a estudiar un tema de Neurociencia Cognitiva (o viceversa). En este ejemplo también está implícito el concepto de dirección, ya que **la conducta es dirigida por la motivación**, haciendo disponible la energía para unas tareas con más facilidad que para otras.

Los organismos cuentan con unos **motivos básicos** que son **innatos**: el hambre, la sed, el sueño, etc. Por ejemplo, si los niveles de glucosa en sangre son bajos nos veremos motivados a realizar conductas de ingesta (hambrientos). Además de los motivos básicos también se considera que existen unos **motivos secundarios** (más susceptibles de ser adquiridos por aprendizaje). Ejemplos de motivos secundarios son la motivación de logro, de poder o de afiliación.

Existen diversas teorías que tratan de explicar el funcionamiento de los procesos motivacionales en los humanos. Podemos distinguir entre **teorías reactivas** y **teorías de activación**.

Las teorías reactivas consideran al sujeto como sistema que **da respuestas a cambios en su ambiente**. Esto es similar a lo indicado para las emociones, que también se consideran como procesos que predisponen para la realización de ciertas conductas en base a los estímulos percibidos. En este contexto, se considera que las emociones son también agentes motivacionales. Lo importante acerca de las teorías reactivas es el concepto de **homeostasis**.

El concepto de homeostasis se refiere a la capacidad de **mantener un equilibrio dinámico de una magnitud** (por ejemplo, el nivel de glucosa en sangre) dentro de una variabilidad controlada. Cualquier cambio en los niveles de variables controladas que atenten contra el nivel de equilibrio provocará una **acción correctiva**. De esta forma, cuando hay un descenso del volumen del plasma sanguíneo (que implica una pérdida de agua y sodio) se produce la sed (sed hipovolémica). La motivación para beber será más grande cuanto más grande sea el desequilibrio (la deshidratación en este caso). Algo similar ocurre con la motivación para dormir, para comer, etc.

La teoría de **reducción del impulso** de Hull supone un modelo plausible para la motivación y concuerda muy bien con las observaciones realizadas con respecto a los motivos básicos (hambre, sueño, deseo sexual, sed, etc.). El organismo detecta una necesidad (desequilibrio) y aparece una motivación para recuperar la homeostasis. Sin embargo, estas explicaciones de corte tan fisiológico se quedan cortas para explicar los procesos asociados con motivaciones secundarias.

Así, además de los modelos basados en la homeostasis, han aparecido otras explicaciones de corte más cognitivo. Este es el caso de las **teorías E/V (expectativas/valencias)**. Estas teorías ofrecen un modelo racional del comportamiento, basado en la consecución intencional (consciente) de un objetivo. Las teorías E/V también son reactivas porque la motivación se entiende como reacción a los cambios en los estímulos (siendo las expectativas y valencias considerados como estímulos). En las teorías E/V la motivación se calcula mediante una multiplicación de los factores E y V. Por ejemplo, es mucho más probable que se dé una conducta si el valor que se le atribuye es muy alto y hay una gran expectativa

de conseguir la meta. Por el contrario, ante una baja expectativa y un bajo valor es muy poco probable que se inicie la conducta de consecución del objetivo.

Como se explica en el artículo de Barberá (2002), diversos autores hacen consideraciones distintas acerca de las expectativas y la valencia. Se pueden considerar las expectativas como de situación y resultado o como de resultado y consecuencia. Por otro lado, la valencia se puede considerar como comparación social, coste/beneficio, etc.

Las teorías de la activación consideran al sujeto como agente de su propia conducta. Por lo que la motivación se considera como un proceso espontáneo (no reactivo). La **conducta desplegada** desde este marco conceptual se considera **propositiva**, es decir, **dirigida por planes, metas y objetivos**.

En este punto conviene distinguir entre **motivación intrínseca** y **motivación extrínseca**. Mientras que la última se basa en la consecución de fines instrumentales que proceden del exterior (como, por ejemplo, trabajar por dinero), la motivación intrínseca se refiere a la activación, energía y dirección que generan conductas que buscan el mero placer de ser realizadas. Es decir, **las tareas motivadas intrínsecamente no se realizan en busca de una recompensa externa**, sino que son en sí mismas la recompensa. Un ejemplo típico de motivación intrínseca en niños es el juego y la curiosidad. En adultos, aunque menos frecuente, sería trabajar por placer (no por dinero). Desde el punto de vista del rendimiento óptimo en diversas tareas se busca la forma de combinar de forma efectiva la motivación intrínseca y la extrínseca (aceptar un trabajo más atractivo para la persona por menos sueldo que otro que no representa una motivación intrínseca).

También desde el punto de vista de las teorías de la activación, se han desarrollado las **teorías de la acción**. En este caso se consideran los **procesos volitivos complejos que aparecen entre la intención y la conducta**. En otras palabras, se considera que los modelos E/V son muy limitados para explicar el proceso motivacional. Aquí aparecen aspectos de más alto nivel como el **compromiso**, la **voluntad** y la **valoración de la meta** una vez conseguida.

En general, todos los modelos resuelven solo parte del complejo rompecabezas que es la motivación. El reto está ahora en **combinar adecuadamente los aspectos motivacionales, emocionales y cognitivos**. Aquí se hace patente de nuevo la dicotomía tan profunda que ha habido en el pasado entre cognición y los procesos de activación como la emoción y la motivación.

7.5. Modelos computacionales de la motivación



Accede al vídeo «Modelos computacionales de la motivación» a través del aula virtual

Mientras que existen multitud de modelos computacionales basados en otros procesos psicológicos, parece que **la motivación ha sido uno de los «grandes olvidados»**. Aunque la motivación tiene una relación estrecha con la emoción, el aprendizaje y la selección de acciones y, por lo tanto, puede estar representada hasta cierto punto en otros modelos, no se han realizado muchos trabajos abordando el tema central de la motivación desde un punto de vista computacional.

Un motivo puede ser que los motivos primarios (hambre, sed, deseo sexual, etc.) no aplican en realidad a los sistemas artificiales debido a las **grandes diferencias de sustrato físico con un organismo natural**. Sin embargo, podemos pensar que en el caso de sistemas artificiales físicos (robots) aparecen también motivos primarios como la disposición de energía eléctrica para el funcionamiento del sistema. Por otro lado, los **motivos secundarios**, como la motivación de logro, poder o afiliación, solo parecen tener sentido en **entornos sociales complejos**. Sin embargo, con la necesidad cada vez más acuciante de interacción natural entre personas y máquinas parece razonable investigar el modelado de estos motivos desde el punto de vista de la inteligencia artificial.

Algunos autores también argumentan que la motivación y la inteligencia están muy relacionados. En este sentido se establecen dos tipos de teorías (Bostrom, 2012):

- ▶ La tesis de ortogonalidad: que la inteligencia y las metas (motivaciones) son ejes ortogonales, sobre los cuales los agentes artificiales pueden situarse libremente.
- ▶ La tesis de convergencia instrumental: que implica que un agente con un nivel suficiente de inteligencia perseguirá las mismas metas intermedias por una mera cuestión instrumental. Es decir, por qué para conseguir los objetivos finales es necesario cubrir las necesidades básicas intermedias (por ejemplo, tener un suministro suficiente de energía).



Accede a los ejercicios de autoevaluación a través del aula virtual

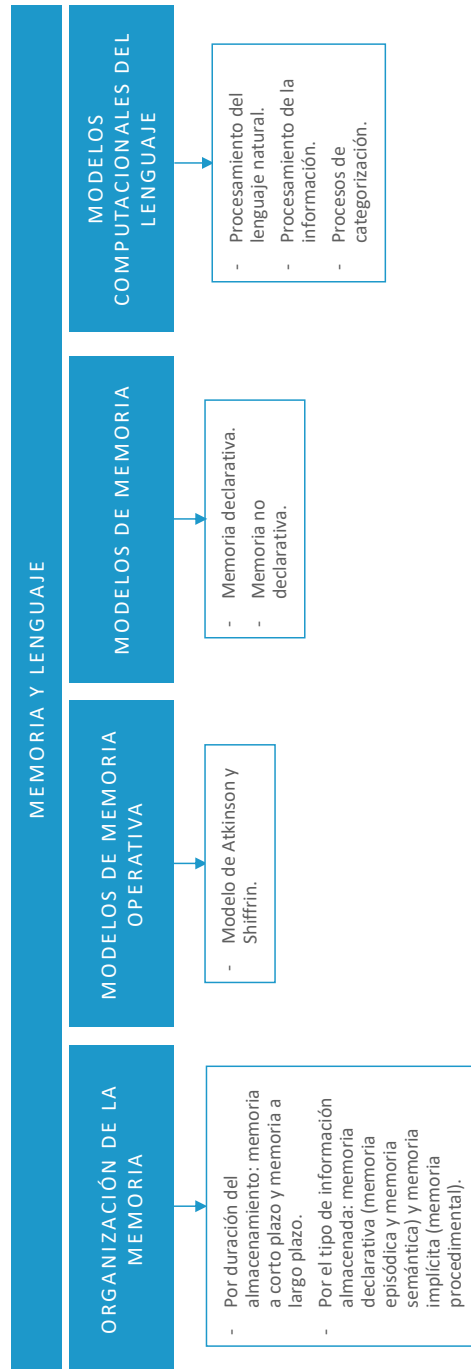


Accede al vídeo «Resumen del tema» a través del aula virtual

Inteligencia Artificial y Computación cognitiva

Memoria y lenguaje

Esquema. Tema 8



Ideas clave. Tema 8

8.1. Introducción y objetivos



Accede al vídeo «Introducción. Memoria y lenguaje» a través del aula virtual

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

En este tema nos centramos en el estudio de los sistemas de memoria. El uso de la memoria está implícito a menudo en la descripción de otras funciones cognitivas, pero merece especial atención hacer un repaso de las capacidades de la memoria humana y también de los principales modelos explicativos. Los principales objetivos de este tema son:

- ▶ Conocer los principales modelos explicativos de los procesos de memoria.
- ▶ Comprender la diferencia entre los diferentes sistemas de memoria identificados en los humanos.
- ▶ Analizar los modelos computacionales de memoria derivados de los estudios neuropsicológicos.

Dado que la memoria es un aspecto fundamental de los sistemas artificiales, este tema sirve también para comparar las características presentes en los mecanismos de almacenamiento y recuperación de la información que existen tanto en los organismos naturales como en los sistemas artificiales.

8.2. Organización de la memoria



Accede al vídeo «Qué es la memoria y qué tipos existen» a través del aula virtual

El primer aspecto importante acerca de la **memoria** es que **no se trata de un proceso unitario** o nuclear, sino que la memoria engloba en realidad múltiples sistemas que, aunque relacionados, se pueden distinguir claramente.

A mitad del siglo pasado se planteó la existencia de distintos almacenes de memoria en los humanos. La primera división más obvia se refiere al tiempo de almacenamiento de la información, y en base a ello podemos distinguir entre:

- ▶ **Memoria a corto plazo:** donde la información se retiene por espacio de unos segundos.
- ▶ **Memoria a largo plazo:** donde la información permanece almacenada por un tiempo indefinido.

Desde el punto de vista de las **bases neurológicas de la memoria**, se identificó la memoria a corto plazo con **circuitos reverberantes** que involucraban a la corteza frontal y la **memoria a largo plazo con cambios más permanentes** (neuroplasticidad) que sucedían en diversas partes del sistema nervioso central. Hoy en día se sabe que el **hipocampo** juega un papel esencial en la fijación de la memoria a largo plazo (gracias a los estudios de caso único y la amnesia anterógrada generada por lesiones en esa estructura del cerebro).

Además de la división que se establece en función de la duración del almacenamiento de la información, se crearon otras clasificaciones basadas en el tipo de contenidos almacenados. Desde este enfoque se distingue entre:

- ▶ **Memoria declarativa:** contenidos que pueden hacerse conscientes y explícitos y que, por lo tanto, pueden verbalizarse y ser usados en el razonamiento.

- **Memoria implícita:** contenidos inconscientes, que apenas pueden ser verbalizados y se refieren a la forma de realizar tareas.

En el ámbito de la memoria declarativa, se distingue a su vez entre dos tipos de sub-almacenes: la **memoria semántica** y la **memoria episódica**.

La **memoria semántica** se refiere a toda la información que almacenamos que se basa en **conocimientos atemporales**. Por ejemplo, que París es la capital de Francia. Es decir, la memoria semántica almacena **conocimiento sobre el mundo**, pero que no está directamente relacionado con el sujeto. Por otro lado, la **memoria episódica** almacena contenidos que están directamente relacionados con hechos o **sucesos relativos al propio sujeto**, y por lo tanto están **contextualizados temporal y espacialmente**. Por ejemplo, el recuerdo del primer beso.

Gracias al estudio de pacientes amnésicos y los avances en neuroimagen se sabe que cada uno de los sistemas de memoria descritos **tienen sustratos neurológicos diferentes**. Es decir, hay áreas del cerebro especializadas en cada uno de estos tipos de almacenes de memoria. La investigación actual se centra en comprender bien cómo funcionan estos diferentes tipos de memoria y cuál es la interrelación existente entre ellos y otras funciones cognitivas y afectivas.

8.3. Modelos de memoria operativa



Accede al vídeo «La memoria operativa y sus modelos» a través del aula virtual



Accede al vídeo «La memoria a largo plazo y sus tipos» a través del aula virtual

En las teorías más actuales ya no se usa el concepto de memoria a corto plazo, ya que se ha ido sustituyendo por los modelos de **memoria operativa**.

Una de las primeras descripciones de este sistema es la de Atkinson y Shiffrin, que distinguen entre los siguientes componentes:

- ▶ La memoria sensorial.
- ▶ La memoria a corto plazo.
 - Ejecutivo central.
 - Lazo fonológico.
 - Agenda visuoespacial.
 - Retén episódico.
- ▶ Memoria a largo plazo.
 - Memoria declarativa (semántica y episódica).
 - Memoria procedimental.

La memoria a largo plazo se corresponde con los aspectos previamente analizados y la **memoria sensorial** se refiere a los **almacenes temporales** de memoria que tienen los **diferentes sentidos** (memoria ecoica para el oído, memoria icónica para la vista, etc.).

Centrándonos en el modelo de memoria operativa, nos interesa estudiar los componentes asociados a la memoria a corto plazo. Como veremos, la **memoria operativa** está directamente relacionada con el **procesamiento consciente de la información, el lenguaje y el pensamiento**. Por este motivo, en la actualidad, se usa el término de memoria operativa (quedando relegado el uso de la expresión de memoria a corto plazo, que no refleja el funcionamiento real de este sistema de memoria).

En realidad, cuando hablamos de memoria operativa, nos estamos refiriendo a un **complejo sistema de coordinación y gestión de recursos limitados** en el que se pueden distinguir diferentes funcionalidades o módulos encargados de **almacenar temporalmente la información y procesarla**. En cierto modo, además de otras

funciones, la memoria operativa puede ser una puerta de entrada a la memoria a largo plazo.

En el famoso artículo de Miller (1956) se hacía referencia a la **limitada capacidad de almacenamiento de la memoria operativa**. Aunque los modelos descritos de memoria operativa son posteriores, se sigue manteniendo que la capacidad de almacenamiento de la memoria a corto plazo se sitúa en torno a los **siete elementos**. Además, como ya hemos mencionado anteriormente, la duración de ese almacenamiento es de unos pocos segundos (no siendo habitual que llegue al orden de minutos).

Veamos en detalle los **principales componentes de la memoria operativa**: el **ejecutivo central** se localiza en la **corteza prefrontal dorsolateral** y está relacionado con el **control de la atención y la selección de estrategias** para coordinar los otros módulos de la memoria operativa. También integra la información proveniente de diferentes modalidades sensoriales.

El **lazo fonológico** se localiza en la **corteza fronto-parieto-temporal izquierda** (en sujetos diestros) y está formado a su vez por dos subsistemas: el **almacén fonológico** y el **sistema de repetición articulada**. Su principal misión es la retención de contenidos basados en codificación fonológica.

La **agenda visuoespacial** se localiza en ciertos **circuitos frontoparietales del hemisferio derecho** y se encarga del procesamiento espacial y visual.

Por último, el **retén episódico** (o *buffer* episódico) se encarga de **integrar los contenidos** provenientes del lazo fonológico y la agenda visuoespacial, integrando la información verbal y espacial y **añadiendo la faceta episódica** al procesamiento. Aquí es donde se establecen las relaciones directas con la memoria a largo plazo.

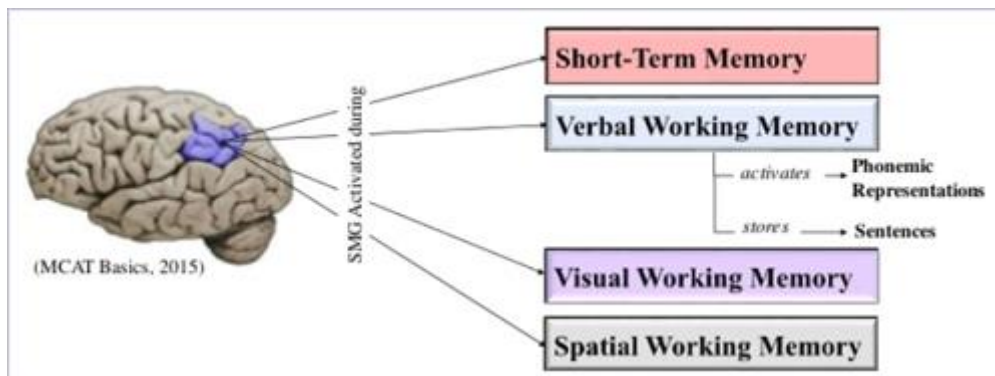


Figura 1. Áreas cerebrales involucradas en la memoria operativa.

Fuente: : <https://www.slideshare.net/OzellaBrundidge/neural-correlates-of-working-memory>

Para comprender mejor los procesos que corresponden al sistema de memoria operativa se pueden distinguir sus **principales funciones**: mantenimiento, manipulación, selección, control de inferencias y planificación de objetivos.

El **proceso de mantenimiento** consiste en conservar la información relevante durante la ausencia de cualquier otro estímulo distractor. En este proceso, el bucle fonológico mantendría la **información verbal** y la agenda visuoespacial mantendría la **información no verbal**. La actividad neuronal que correlaciona con estos procesos se sitúa en áreas de la corteza parietal y frontal. Además, se distingue entre dos grandes circuitos, el **fronto-parietal ventral** para la información verbal y el **fronto-parietal dorsal** para la información espacial.

El proceso de **selección** se basa en la codificación y elección de los estímulos. Esta selección se produce en función de diversos criterios y cualidades de los mismos. Se asocian con este proceso zonas del **área prefrontal dorsolateral** (en tareas de selección de elementos) y el **área prefrontal ventrolateral**, el **cíngulo anterior** y la **corteza parietal media** (en tareas de localizaciones espaciales).

Los procesos de **control de interferencias** son necesarios para conservar la atención estable, dando una respuesta adaptativa y manteniendo los contenidos en memoria operativa durante el tiempo suficiente para lograr los objetivos. Son en definitiva procesos de inhibición de estímulos no relevantes para la tarea en curso. Esta

inhibición y cambio atencional se asocia al **cíngulo anterior** y su interrelación con la **corteza prefrontal dorsolateral**.

La **manipulación** es el proceso más complejo, ya que requiere recuperar, mantener, seleccionar y supervisar la información. Al ser un proceso tan complejo es difícil estudiarlo de forma aislada, aunque hay consenso acerca del correlato neuronal que se situaría en la **corteza prefrontal dorsolateral**.

Por último, la **planificación de objetivos** es necesaria para establecer un equilibrio entre la estimulación externa, las representaciones internas y la generación de respuestas (conducta). Como veremos en el tema dedicado a las funciones ejecutivas, todas estas funciones se asocian a la **corteza frontal**.

Los sistemas de **memoria a largo plazo** se encargan de mantener indefinidamente la información relativa a conocimientos **procedimentales, semánticos y episódicos**. Tradicionalmente se mantiene la división en dos sistemas: la **memoria declarativa** y la **memoria no declarativa** (o procedimental).

La **memoria procedimental** está directamente relacionada con la adquisición implícita (inconsciente) de **habilidades motoras**, como los efectos de facilitación (*priming*), los hábitos motores o el aprendizaje por condicionamientos asociativos. Desde el punto de vista neurobiológico, esta memoria se asocia principalmente con los **ganglios basales**, ya que se ha observado que las lesiones en estos centros provocan trastornos como el Parkinson o la corea de Huntington. El **cerebelo** también tiene un papel muy relevante en la memoria procedimental, siendo también el locus de habilidades sensoriomotoras.

Centrándonos ahora en la **memoria declarativa**, se sabe que está correlacionada con la activación de un conjunto de centros neuronales de **áreas temporo-diencefálicas**. Es decir, las redes de memoria están distribuidas a lo largo de toda la corteza cerebral, aunque se han detectado zonas cuya lesión producen claros déficits en la memoria a largo plazo.

Se ha identificado un circuito de codificación de la información que conecta el **área medial del lóbulo temporal**, las **regiones diencefálicas** y el **núcleo basal de Maynert**. Las regiones temporo-mediales se encargan de recibir la información de todo el cerebro, que proviene de los diferentes sistemas perceptivos. Estas **áreas de asociación** combinan la información supramodal en elementos integrados de una experiencia (multimodal). En las zonas subcorticales, la **amígdala** se encarga de procesar los contenidos emocionales y las **estructuras diencefálicas** se ocupan de realizar una secuenciación temporal de los recuerdos. Al parecer, las **regiones prefrontales** también juegan un papel importante, ofreciendo información contextual y gestionando las estrategias para el correcto aprendizaje/memorización (aunque podríamos considerar esta regulación como un proceso metacognitivo).

El **modelo de asimetría hemisférica de codificación y recuperación** (HERA) propone que algunos tipos de memoria involucran a procesos de codificación del hemisferio izquierdo, mientras que los procesos de recuperación se sitúan en el hemisferio derecho (esencialmente en regiones ventrales prefrontales). Sin embargo, investigaciones llevadas a cabo más recientemente demuestran que la actividad en el lado izquierdo durante la codificación **es simplemente más intensa**, pero que **intervienen también las áreas prefrontales ventrolaterales derechas** cuando la carga de memoria es alta. Además, la corteza izquierda también se activa durante la recuperación de información.

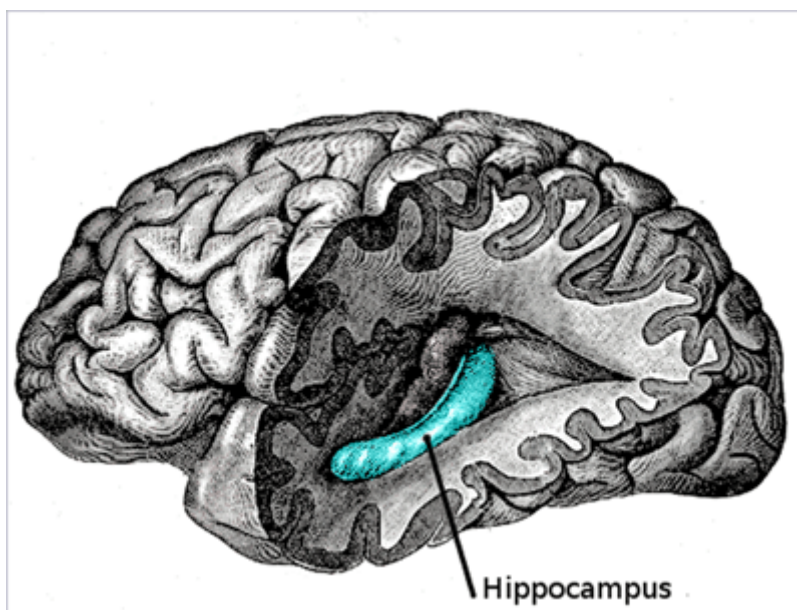


Figura 2: Localización del hipocampo en la parte medial del lóbulo temporal.

Fuente: <https://commons.wikimedia.org/w/index.php?curid=3907047>

El **hipocampo** es otra de las estructuras neuronales que tiene un papel importante en la formación de memorias a largo plazo, especialmente en la **fijación** de las mismas y en la **recuperación** de contenidos episódicos y autobiográficos. Al hipocampo se le identifica como parte esencial de un sistema más amplio, como el descrito anteriormente, que implementa el **sistema de memoria declarativa**. El hipocampo sería el encargado de **detectar sucesos, lugares y estímulos novedosos**.

Las lesiones en el hipocampo producen diferentes formas de **amnesia**, tanto **anterógrada** (no poder fijar nuevas memorias a partir de la lesión), como **retrograda** (no poder recordar sucesos anteriores a la lesión). Se piensa, por lo tanto, que el hipocampo se encarga de la consolidación de los recuerdos en la memoria a largo plazo. También se sabe que las lesiones hipocampales no afectan a la memoria procedimental (que como hemos visto está más relacionada con los ganglios basales y el cerebelo).

El funcionamiento detallado de la memoria semántica se estudiará en el tema dedicado a la categorización de los contenidos mentales y los modelos semánticos asociados.

8.4. Psicología del lenguaje



Accede al vídeo «El lenguaje y su estructura» a través del aula virtual

En un primer acercamiento parecería que el **lenguaje** es un proceso bien conocido por los humanos y fácil de definir, pero lo cierto es que el lenguaje es **uno de los fenómenos más complejos que se conocen** y entrar en el detalle excede con creces los objetivos de este tema.

El lenguaje se relaciona con la **comunicación verbal**. Sabemos que existen otras formas de comunicación que no se basan en el lenguaje (la comunicación no verbal), pero también sabemos que la riqueza del conocimiento que se puede transmitir gracias al lenguaje es mucho más grande (es cierto que, en la comunicación humana, la parte no verbal transmite mucha información, pero es una información relativa a las emociones, actitudes, etc., mientras que el lenguaje verbal se centra en descripciones detalladas que transmiten **una información mucho más precisa**). También es cierto que se puede considerar que hay lenguajes que no son verbales, como los que usan otros animales o las personas signantes. Sin embargo, a efectos del presente tema, utilizaremos solo el contexto del lenguaje humano verbal, dejando al alumno el estudio autónomo de otras formas de lenguaje que existen en el planeta.

Una definición simplificada de lenguaje podría ser la siguiente: se trata de un **sistema complejo de reglas y símbolos** que permiten comunicarnos. Los símbolos que usa el lenguaje son las **palabras**. Las palabras son **objetos que representan a otros objetos**. Es decir, no hace falta tener una mesa delante para poder hablar sobre las mesas, gracias al lenguaje sirve con utilizar la palabra «mesa» en español. Las **reglas** indican la forma en la que se pueden usar las palabras **para construir frases**. Es decir, lo que llamamos **gramática**.

Lo cierto es que **no hay un consenso global** entre lingüistas, psicólogos, filósofos, etc. acerca de lo que es el lenguaje. De hecho, siguen existiendo grandes debates acerca del lenguaje de otros animales, la relación entre pensamiento y lenguaje, etc. También se puede abordar el lenguaje desde **múltiples perspectivas**. Por ejemplo, desde el punto de vista del desarrollo podemos estudiar cómo los niños adquieren el lenguaje (y como lo niños que no están expuestos al lenguaje tristemente no lo adquieren).

Al hablar del lenguaje verbal humano, tenemos que tener en cuenta una serie de aspectos que son de vital importancia y son los que vamos a analizar también desde el punto de vista neurobiológico. Al tratar el problema de la comprensión del lenguaje verbal tenemos que hablar de **sonidos**, del **significado de las palabras**, y de la **gramática**. Además, podemos considerar la estructura del lenguaje a diferentes niveles, así distinguimos entre:

- ▶ Semántica: estudio del significado.
- ▶ Sintaxis: estudio del orden de las palabras en una frase.
- ▶ Morfología: estudio de las palabras y como estas se forman.
- ▶ Fonética: estudio de los sonidos del lenguaje.
- ▶ Fonología: estudio del uso de los sonidos en el lenguaje.

La palabra se define como la **unidad del lenguaje**. Es decir, la unidad que por sí sola puede constituir una expresión completa (y está separada por espacios en el lenguaje escrito). Las palabras están constituidas por **sonidos, letras y sílabas**. Por ejemplo, la palabra «palabra» tiene tres sílabas y cinco sonidos diferentes. Las palabras también se pueden analizar por los **morfemas** que contienen. Por ejemplo, la palabra «casita» está formada por una raíz «cas-», que indica el significado, y un sufijo «-ita», que proporciona la idea de tamaño.

Las palabras se crean a partir de elementos más sencillos llamados morfemas. La **morfología inflexiva** se refiere a cambios que no alteran el significado de la palabra, mientras que la **morfología derivativa** se ocupa de los que sí. Un ejemplo de cambio

inflexivo es «lápiz» que pasa a ser «lápices». Un ejemplo de cambio derivativo es «trabajo» que se convierte en «trabajador».

Se piensa que almacenamos las palabras en el cerebro en una especie de **diccionario mental** que contiene todas las palabras que sabemos usar y comprender. A este diccionario mental se le llama **lexicón**. El lexicón contiene toda la información que usamos sobre las palabras, es decir, sus sonidos, significado, apariencia escrita, función sintáctica, etc. Por tanto, el cerebro debe manejar la fonología, semántica, ortografía, etc. para ser competente en el uso del lenguaje. El lexicón de un adulto contiene en torno a 70 000 palabras (las estimaciones varían entre 15 000 y 150 000 palabras, aunque todos conocemos personas que hacen su vida normal usando menos de 1000 palabras, sobre todo los niños pequeños).

Gracias a este diccionario mental de palabras podemos realizar procesos de **reconocimiento**, al conocer una palabra se accede a toda la información relativa a ella (lo que significa, cómo se escribe, etc.). En este contexto hay muchos interrogantes abiertos: ¿cuál es la diferencia entre comprender un discurso y comprender una palabra? ¿Cómo se accede a las representaciones del lexicón? ¿Cómo sabemos si una palabra está almacenada o no?, etc.

Hay muchos **otros aspectos sobre el lenguaje** que merecen una atención especializada, pero que quedan fuera del alcance del presente tema. Por ejemplo, cómo ha evolucionado el lenguaje a lo largo de la historia de los humanos, cuáles son los mecanismos de adquisición del lenguaje, cuáles son las diferencias básicas entre diferentes idiomas, cuáles son los mejores métodos para estudiar el lenguaje, etc. A continuación, nos centraremos en las **estructuras cerebrales** que dan soporte a los procesos de **producción y comprensión del lenguaje** en los humanos.

8.5. Modelos computacionales del lenguaje



Accede al vídeo «Categorización y modelos computacionales del lenguaje» a través del aula virtual

Aunque existen multitud de modelos computacionales orientados a las técnicas de **procesamiento de lenguaje natural**, aquí nos centraremos en los modelos que tienen una clara filosofía de bioinspiración.

En el ámbito de la neurociencia cognitiva se considera al cerebro como un centro complejo de **procesamiento de la información** que es capaz de **dar sentido** a los datos provenientes de los sentidos. En estos procesos el lenguaje juega un papel esencial, ya que a menudo, y sobre todo los adultos, **pensamos con palabras**. Es decir, comprendemos el mundo narrándolo, aunque sea a nosotros mismos. Estos procesos de comprensión del mundo que rodea a la persona se realizan mediante diferentes funciones. Se dice que el cerebro humano está especialmente diseñado para **clasificar, categorizar y conceptualizar**. En otras palabras, pensamiento y lenguaje tienen la capacidad de **etiquetar los objetos y los sucesos** del mundo en base a una serie de conceptos o categorías aprendidos.

Desde el punto de vista puramente computacional, los **procesos de categorización** se pueden basar en **sistemas de clasificación**, para los que se definen un conjunto de clases previamente definidas, y tras una fase de aprendizaje, el sistema es capaz de categorizar nuevos objetos (que nunca ha visto previamente). En cierta medida, estos sistemas artificiales funcionan de forma similar a como lo hacen los humanos cuando están **aprendiendo nuevas categorías**.

Por ejemplo, un niño que nunca ha visto antes un gato, pero que sí conoce bien la categoría «perro», la primera vez que vea un gato dirá que ese animal es un perro. Es decir, lo categorizará como animal y como perro, porque cumple con las características conocidas de la clase perro: tener cuatro patas, pelo, dos orejas, etc. Sin embargo, si un adulto le corrige y le indica que se trata de un gato, después de unos cuantos episodios de resistencia, el niño **no tendrá más remedio que generar**

una nueva categoría para los gatos, en la que incluya algunas características distintivas, como por ejemplo hacer «miau» y no hacer «guau».

El lenguaje juega un papel en todos estos procesos, pues la «etiqueta», es decir, la **palabra**, es el **símbolo** que se usa para determinar la identificación de una categoría, para comunicársela a los demás y para comunicársela a uno mismo. La mayoría de los modelos computacionales de categorización no usan un lenguaje complejo, sino que simplemente se limitan a usar etiquetas o códigos para cada una de las categorías que han de ser clasificadas. Sin embargo, en los sistemas más complejos, que persigan imitar los procesos cognitivos de los humanos, es necesario incorporar algún tipo de lenguaje que permita establecer narrativas sobre el mundo. Estas narrativas requieren necesariamente de categorías para poder generar un discurso abstracto y no «perderse» en los detalles.

Los sistemas que no son capaces de categorizar y se quedan en un mero **análisis de las características de los objetos y eventos**, se dice que son sistemas que se **sobreajustan** (*overfitting*). Es decir, no son capaces de **abstraerse obteniendo categorías conceptuales**, para ellos cada nuevo objeto es totalmente una entidad nueva.

Sin embargo, los sistemas que logran categorizar con éxito son capaces de **generalizar** y, aunque detecten la presencia de un nuevo objeto, son capaces de saber qué tipo de objeto es. Por ejemplo, un humano no necesita haber visto todas las sillas del mundo para poder decidir si un nuevo mueble que ve es una silla o no. La ventaja de la **categorización** es que **permite razonar en un entorno lleno de ruido y ambigüedad**, y ante la presencia de objetos totalmente desconocidos, con el simple hecho de categorizarlos ya podemos obtener mucho conocimiento acerca de ellos. De hecho, cuando nos encontramos con una nueva silla que jamás hemos visto (incluso una de diseño sueco), somos capaces de inferir cómo se usa y qué tenemos que hacer para sentarnos en ella. Muchos sistemas artificiales adolecen de estas capacidades.



Accede a los ejercicios de autoevaluación a través del aula virtual

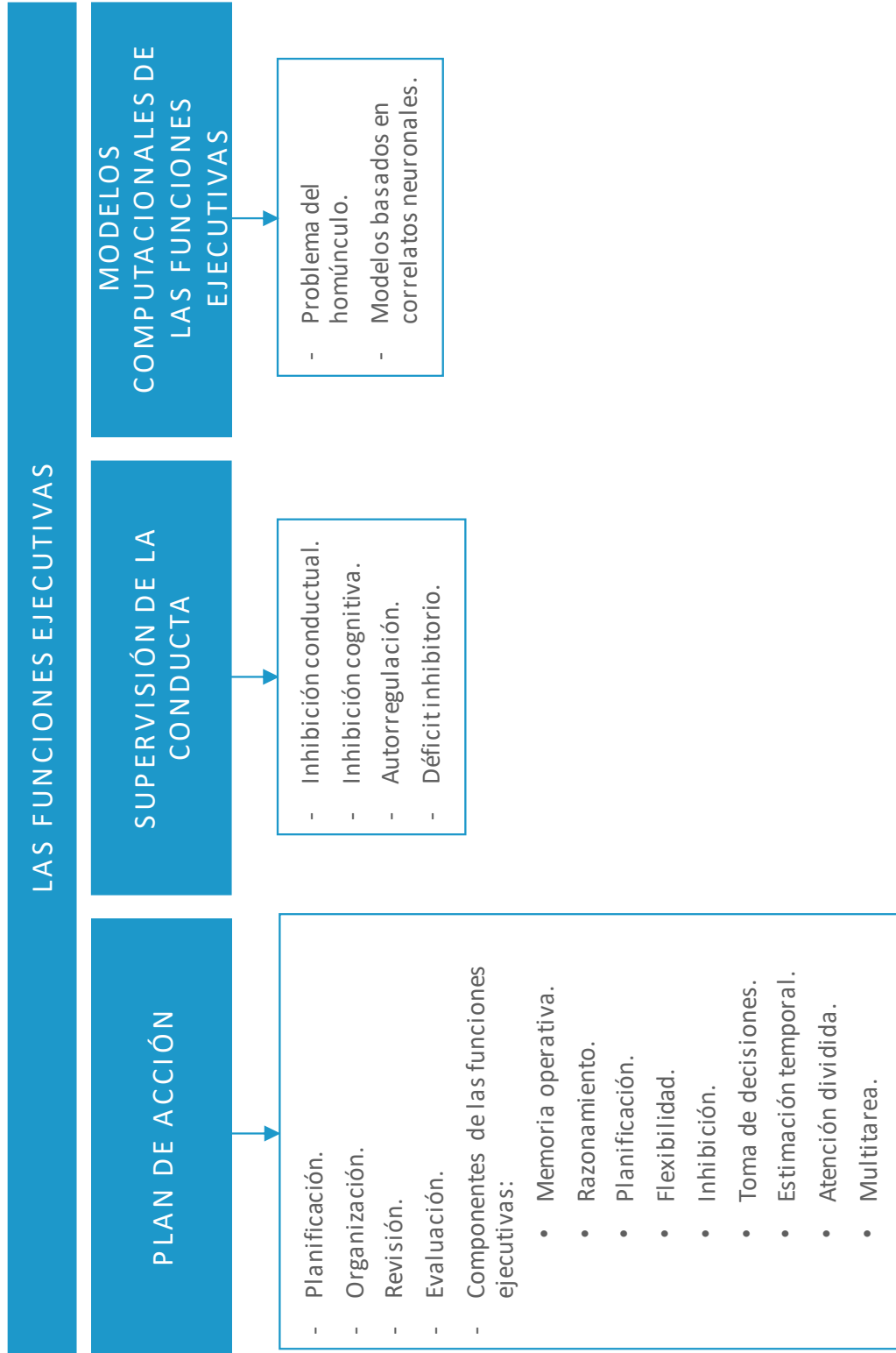


Accede al vídeo «Resumen del tema» a través del aula virtual

Inteligencia Artificial y Computación Cognitiva

Las funciones ejecutivas

Esquema. Tema 9



Ideas clave. Tema 9

9.1. Introducción y objetivos



Accede al vídeo «Introducción. Las funciones ejecutivas» a través del aula virtual

Para estudiar este tema lee las Ideas clave disponibles a continuación y el siguiente artículo: Tirapu-Ustárrroz, J., Muñoz-Céspedes, J. M. y Pelegrín-Valero, C. (2002). Funciones ejecutivas: necesidad de una integración conceptual. *Revista de neurología*, 34(7), 673-685.

Accede al artículo a través del aula virtual o de la siguiente dirección web:

<https://pdfs.semanticscholar.org/f6f1/b7ec7671afea34cb9cfae7804ee13c369368.pdf>

Como vemos, según vamos avanzando en el análisis de las funciones cognitivas es más clara la interrelación entre las mismas. En este caso, las llamadas funciones ejecutivas tiene que ver con el control de alto nivel de la conducta y del pensamiento. Concretamente, en este tema se persiguen los siguientes objetivos:

- ▶ Identificar los factores clave de las funciones ejecutivas en cuanto al despliegue de la conducta del organismo.
- ▶ Analizar la funcionalidad que implementan los modelos computacionales de las funciones ejecutivas.

Además de estudiar las funciones ejecutivas desde una perspectiva biológica se tienen en cuenta en este tema las posibles conclusiones que se puedan aplicar como bioinspiración para modelos computacionales y sistemas artificiales.

9.2. Componentes de las funciones ejecutivas



Accede al vídeo «Funciones ejecutivas. Concepto y principales funciones» a través del aula virtual

Las **funciones ejecutivas** son un conjunto de mecanismos que habitualmente se agrupan bajo el mismo concepto. Cuando hablamos de funciones ejecutivas estamos refiriéndonos a mecanismos mentales complejos relacionados con la **planificación, organización, revisión y evaluación del comportamiento**. Las funciones ejecutivas son, por lo tanto, necesarias para implementar con éxito las conductas orientadas a metas.

Según los autores se pueden identificar diferentes funciones como pertenecientes al grupo de funciones ejecutivas. En general, se consideran que los siguientes son componentes de las funciones ejecutivas:

- ▶ Memoria operativa, que como vimos se encarga del procesamiento de la información verbal, visuoespacial y la manipulación de símbolos durante cortos espacios de tiempo.
- ▶ Razonamiento o pensamiento consciente, que nos permite resolver problemas y establecer relaciones entre elementos de información (a su vez, podemos distinguir entre razonamiento lógico deductivo e inductivo).
- ▶ Planificación, que permite establecer los pasos y las secuencias necesarias para la consecución de metas, anticipando el resultado de las acciones.
- ▶ Flexibilidad, que hace posible que podamos incluir cambios en la planificación (manejando los cambios de contexto) para adaptarnos a las demandas del entorno.

- ▶ Inhibición, para no procesar la información no relevante para una tarea ni dejarse llevar por impulsos que van en contra de la consecución de una meta.
- ▶ Toma de decisiones, que nos permite elegir entre varias alternativas en función de nuestras metas, recursos, expectativas y predicciones.
- ▶ Estimación temporal, que nos ayuda a calcular la duración de determinados sucesos o actividades, para así poder predecir y planificar secuencias de acciones.
- ▶ Atención dividida, que nos permite realizar más de una tarea al mismo tiempo (cosa que solo es posible si las tareas requieren recursos de procesamiento diferentes).
- ▶ Multitarea, lo cual nos permite que podamos intercalar varias tareas en un mismo período de tiempo, retomando cada tarea en el punto en el que se interrumpió la última vez gracias a un cambio de contexto efectivo.

Como vemos, cuando hablamos de funciones ejecutivas estamos hablando de una serie de procesos, todos muy complejos, y que además han de funcionar **de forma conjunta**. Las funciones ejecutivas hacen que el **comportamiento humano sea tan rico, variado y complejo**, pues nos dan la capacidad de **perseguir varios objetivos al mismo tiempo**, estando cada uno de ellos circunscritos a escalas de tiempo diferentes. Además, las funciones ejecutivas hacen posible que podamos lidiar con la tremenda **complejidad, ambigüedad e incertidumbre del mundo real**, siendo capaces de detectar errores, evaluando nuestro propio rendimiento e inhibiendo las tendencias que aprendemos que no nos benefician.

Se puede decir también que las funciones ejecutivas cumplen un papel de **autorregulación** de nuestra actividad cognitiva y afectiva. Como señala Goldberg (2004), serían el director general que dirige el resto del cerebro. Las funciones ejecutivas, y su carácter metacognitivo, nos diferencia de otras especies, que reaccionan de forma más automática ante las demandas del entorno. Gracias a estas

funciones el ser humano puede **resolver problemas más complejos, realizar proyectos novedosos** y, en general, ser capaz de salir airoso de situaciones novedosas. Las funciones ejecutivas cumplen su rol sobre todo cuando las tareas que enfrentamos no se pueden realizar con éxito de forma automática.

Planificación de la acción

Desde el punto de vista de la **planificación**, el organismo necesita identificar y organizar una **secuencia de pasos y metas intermedias** con el fin de llegar al objetivo planteado. La planificación requiere que se plantee una meta inicialmente y determinar la mejor forma de conseguirla. En el desarrollo de los humanos se observa cómo los niños utilizan un habla exteriorizada referida a ellos mismos cuando están aprendiendo a guiarse en sus propias planificaciones. Más adelante en el desarrollo, ese lenguaje autoreferido se convierte en habla interior. En estos casos también se observa la interrelación entre lenguaje y funciones ejecutivas.

9.3. Supervisión de la conducta



Accede al vídeo «Planificación, supervisión y lenguaje interno» a través del aula virtual

Cuando hablamos de supervisión de la conducta, nos referimos a la coordinación, planificación, inhibición, detección de errores, etc. En general, la supervisión de la conducta requiere un nivel superior de control (o meta-control), por eso este concepto está relacionado con la metacognición.

También es importante destacar que, aunque hablamos principalmente de conducta, la inhibición también afecta a los procesos cognitivos. Es decir, podemos hablar de inhibición cognitiva e inhibición conductual. La **inhibición conductual**, como hemos visto anteriormente, se refiere a la capacidad de detener una respuesta automática en el momento preciso. Por otro lado, la **inhibición cognitiva** está asociada con las

diferentes funciones ejecutivas. Así, podemos hablar de inhibición de la atención o inhibición emocional.

La **inhibición de la atención** puede ser útil para evitar distracciones y que pensamientos, emociones u otros estímulos se interpongan en la respuesta necesaria para la tarea actual. La pausa inhibitoria permite que haya tiempo suficiente para evaluar la situación y que jueguen su papel los procesos de planificación y toma de decisiones. También el funcionamiento de la memoria de trabajo se beneficia de la inhibición atencional, ya que cambios de contexto continuo provocarían un mal funcionamiento de la misma y una conducta y pensamientos desorganizados.

Cuando se produce un **déficit inhibitorio**, como sucede en el TDAH, el sujeto tiene gran dificultad para resistirse a las gratificaciones inmediatas, provocando los típicos problemas de baja tolerancia a la frustración y aversión a la demora. Como se sabe, los niños con TDAH tienen por lo general malos resultados académicos y es que al fallar las funciones ejecutivas también resultan gravemente afectadas otras funciones como el aprendizaje.

Una mente compleja necesita autorregularse, autogestionarse y autoplanificarse. Las funciones ejecutivas son el grupo de mecanismos que hacen posible esto en los humanos, y uno de los aspectos que más nos diferencia de otros animales, junto con la complejidad del lenguaje humano. Todos estos mecanismos pueden además complementarse con una percepción de sí mismo, cerrando así un ciclo de metacognición.

Como decíamos, el **lenguaje también juega un papel** importante en la autorregulación. **El habla interior** (la producción verbal privada que aparece en la memoria operativa) se usa habitualmente para emitir mensajes internos que regulan nuestros propios actos. Esto es más patente en los niños pequeños, que antes de ser capaces de internalizar esta conversación privada verbalizan públicamente las autoinstrucciones que se dan. Por ejemplo, dicen «ahora voy a poner la pieza azul sobre la pieza blanca y luego voy a tirar la torre». Este lenguaje intrínseco se convierte

en un mecanismo adicional de autorregulación del comportamiento basado en autoinstrucciones.

9.4. Modelos computacionales de las funciones ejecutivas



Accede al vídeo «Base neurofisiológica y modelos computacionales de las funciones ejecutivas» a través del aula virtual

El estudio de las redes neuronales que dan lugar a las funciones ejecutivas en el cerebro permite construir modelos computacionales basados en hipótesis sobre el funcionamiento de las mismas.

La **corteza prefrontal** se considera desde hace tiempo como el centro ejecutivo que controla la selección de acciones y coordina las funciones cognitivas en general. Sin embargo, la base computacional de las funciones ejecutivas no se ha especificado claramente (volviendo al **problema del homúnculo** al considerar que la corteza frontal es como «un cerebro que controla el cerebro»).

En general, los modelos existentes siguen planteando muchas dudas: ¿cómo sabe la corteza prefrontal qué acciones o planes seleccionar? ¿Cómo influye la experiencia en la planificación? ¿Qué diferencia hay entre los procesos neuronales de la corteza prefrontal que dan lugar a estas propiedades y otros centros de procesamiento? El problema del homúnculo viene porque no somos capaces de responder a estas preguntas.

Ha habido varios intentos recientes de evitar el problema del homúnculo mediante la **formulación de modelos computacionales** precisos de los mecanismos neuronales subyacentes a las funciones ejecutivas (para una revisión detallada de estos modelos ver Hazy, Frank y O'Reilly, 2007).

Los modelos actuales se basan en el papel que juegan los **ganglios basales** y los **sistemas frontales** en el control motor y la selección de acciones. Los ganglios basales proporcionarían una modulación de la decisión de ejecutar o no una acción representada en la corteza frontal.

Otros enfoques abogan por la idea de que los **ganglios basales modulan las representaciones** de la memoria operativa facilitando así las funciones ejecutivas más abstractas. En el trabajo de Hazy, Frank y O'Reilly, (2007) se describe un sistema computacional que es capaz de tener un rendimiento similar a un humano en memoria operativa y **control ejecutivo a través de tareas de prueba y error**. El sistema implementa un mecanismo de aprendizaje por refuerzo que emula el sistema dopaminérgico mesencefálico (sistema de recompensa) y su activación a través de los ganglios basales y la amígdala.

9.5. Referencias bibliográficas

Goldberg, E. (2009). *The new executive brain: Frontal lobes in a complex world*. Oxford: Oxford University Press.

Hazy, T. E., Frank, M. J. y O'Reilly, R. C. (2007). Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1485), 1601-1613.



Accede a los ejercicios de autoevaluación a través del aula virtual



Accede al vídeo «Resumen del tema» a través del aula virtual