

Journal of Engineering Research

Acceptance date: 23/07/2025

MODELO DE APRENDIZAJE AUTÓNOMO A PARTIR DE EJEMPLOS BASADO EN UNA CONCIENCIA DE OBJETIVOS

Gabriel Fiol Roig

Universidad de las Islas Baleares
Palma de Mallorca
Isla Baleares. España



All content in this magazine is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Resumen: El aprendizaje autónomo a partir de ejemplos constituye uno de los campos de la IA más estudiados en la actualidad. Sin embargo, el significado que se otorga al concepto de «aprendizaje» en numerosos contextos y aplicaciones, resulta limitado en relación a su magnitud real. En numerosas obras y aplicaciones se confunden los términos de aprendizaje y razonamiento, atribuyendo al aprendizaje las cualidades de adquisición de conocimiento cualitativo propios del razonamiento, cuando en realidad va mucho más allá. Este trabajo presenta una visión más amplia del significado del concepto de aprendizaje en inteligencia artificial –IA en adelante-, la cual permite su exploración en una mayor magnitud. Para entender los alcances de dicha magnitud, se exploró el área de la conciencia humana, máximo exponente que engloba todos los mecanismos conscientes del ser humano. No obstante, su complejidad es tal que los avances resultan muy lentos y la falta de consensos resulta común. Además, la evolución de sistemas artificiales a partir de una conciencia general resulta poco menos que intratable en estos momentos. Finalmente se decidió investigar sobre un tipo de conciencia particular, conocida como conciencia de objetivos, cuyos alcances son más limitados, pero resultan viables y apropiados para abordar la cuestión del aprendizaje en IA. La conciencia de objetivos se refiere a la capacidad de un sistema de inteligencia artificial para comprender y alinearse con los objetivos y metas de los expertos humanos. Este trabajo describe, por una parte, un modelo general de conciencia de objetivos y, por otra, propone un diseño de un prototipo de modelo para sistemas de aprendizaje autónomos a partir de ejemplos

clasificados, fundamentado en el modelo de la conciencia de objetivos. Los resultados muestran una nueva categoría de sistemas de aprendizaje capaces de aprender progresivamente de forma autónoma, utilizando sus experiencias adquiridas durante el proceso para adaptar la hipótesis del aprendizaje de forma alineada con los objetivos de expertos humanos. Otra de las características que se destacan en el trabajo, es la naturaleza de las experiencias, de manera que éstas constituyen también un objeto de aprendizaje.

Palabras clave: Conciencia de objetivos, inteligencia artificial, experiencias, adquisición de conocimiento cualitativo, razonamiento, razonamiento inducto-deductivo, aprendizaje a partir de ejemplos, sistemas de aprendizaje autónomos.

INTRODUCCIÓN

La **conciencia** y la **inteligencia** son dos conceptos apasionantes que han sido un tema central en la filosofía y la ciencia durante siglos. La comprensión de la conciencia sigue evolucionando, especialmente en áreas como la psicología y la neurociencia, donde se estudia cómo el cerebro genera experiencias conscientes¹.

En [1] se matizan los aspectos clave de la conciencia, así, la **conciencia** se refiere a la **capacidad de los seres humanos** y otros animales para **percibir y ser conscientes** de sus propios pensamientos, sentimientos y entorno. Es la **experiencia subjetiva** de la mente, que incluye la **percepción sensorial**, la **cognición** y la **introspección**.

De acuerdo con la misma fuente, existen varias **teorías** que intentan explicar la naturaleza y el origen de la conciencia:

1. La **conciencia** y **consciencia** son términos que causan confusión. El primero se asocia con la ética y la **moral**, mientras que el segundo se relaciona con la **reflexión**, la **percepción** o el **conocimiento**. La Asociación Americana de Psicología (APA) define la **consciencia como la percepción o el conocimiento de algo**. Esta percepción se refiere tanto a lo que nos rodea como a lo que ocurre en nuestro interior. Como puede verse, va un paso de más allá de estar conscientes y alerta, también implica conocimiento de sí mismo. ¿Conciencia y consciencia pueden ser sinónimos? De acuerdo con la Real Academia Española (RAE), solo hay una situación en la que ambos términos significan lo mismo. Cuando hablamos de la percepción y autoconocimiento, se puede usar conciencia o consciencia de forma indistinta. Pero si uno se refiere al aspecto moral, lo correcto es decir «conciencia».

TEORÍA DEL ESPACIO DE TRABAJO GLOBAL

Propuesta por el neurocientífico **Bernard Baars**, esta teoría sugiere que **la conciencia surge cuando la información de diferentes partes del cerebro se integra en un “espacio de trabajo global”**. Este espacio actúa como una pizarra donde la información se hace accesible a varios sistemas cerebrales.

TEORÍA DE LA INFORMACIÓN INTEGRADA

Formulada por el neurocientífico **Giulio Tononi**, esta teoría sostiene que **la conciencia se relaciona con la capacidad del sistema de integrar información**. Cuanto mayor sea la integración de la información, mayor será el grado de conciencia.

TEORÍA DE LA CONEXIÓN NEURONAL

Esta teoría sugiere que la conciencia surge de la **actividad sincrónica** de las neuronas en diferentes partes del cerebro. Según esta teoría, la sincronización y la comunicación entre neuronas es clave para la experiencia consciente.

Numerosos estudios científicos han intentado desentrañar los mecanismos detrás de la conciencia. Utilizando técnicas avanzadas de **neuroimagen** como la **resonancia magnética funcional -FMRI-** y la **electroencefalografía -EEG-**, los científicos han identificado áreas cerebrales específicas que se activan durante estados conscientes.

De las anteriores teorías se desprende que **la conciencia es un mecanismo constituido por elementos sincronizados integradores y procesadores de información**.

CONCIENCIA E INTELIGENCIA.

La conciencia es fundamental para nuestra **identidad personal** y nuestra **experiencia del mundo**. Nos permite **reflexionar** sobre nuestros pensamientos, **planificar el futuro**, y **empatizar con los demás**.

La inteligencia es otra de las cuestiones conceptuales que conforman el núcleo de la cognición humana. Es tan antiguo como el concepto de conciencia y sus repercusiones son fundamentales en los avances de la ciencia.

La inteligencia se refiere a la capacidad de un individuo o una máquina para **adquirir, procesar y aplicar conocimiento y habilidades para resolver problemas y adaptarse al entorno**.

A medida que avanzamos en nuestra comprensión de la mente humana y la inteligencia artificial, surge la pregunta de si **la conciencia y la inteligencia están intrínsecamente relacionadas o si son entidades separadas y distintas**. A continuación, exploraremos algunas reflexiones sobre la relación entre ambos conceptos.

De acuerdo con [2] y [3], la **conciencia** y la **inteligencia** a menudo se confunden o se utilizan indistintamente, pero en realidad son dos **conceptos diferentes**.

LA CONCIENCIA

La conciencia es un **concepto más complejo y abstracto que la inteligencia**. Se refiere a la **capacidad de un ser para tener una experiencia subjetiva del entorno y de sí mismo**. La conciencia implica ser consciente de los pensamientos, emociones, sensaciones y percepciones propias, así como de la existencia de otras entidades conscientes. Es la **experiencia interna y personal de estar vivo y tener una identidad** única.

La conciencia nos permite ser conscientes de nuestra **existencia** y de nuestro

entorno. Es la capacidad de reconocer la realidad y de relacionarnos con ella. La **conciencia supone también tener conocimiento de uno mismo, de nuestros pensamientos, emociones y acciones**. Nos permite ser conscientes de nuestras decisiones y de las consecuencias de nuestras acciones. La conciencia también tiene un **componente ético o moral**,

ya que nos permite distinguir entre el bien y el mal y tomar decisiones basadas en nuestros valores y principios.

La conciencia también puede referirse al **estado fisiológico de vigilia**, es decir, el **estado de estar despierto y alerta**. En este sentido, la conciencia se refiere a la **capacidad de estar consciente de nuestro entorno y de responder a estímulos externos**.

LA INTELIGENCIA

La **inteligencia**, por otro lado, se refiere a la **capacidad de un ser para aprender, comprender, razonar/pensar de manera lógica, analizar información, planificar, tomar decisiones y resolver problemas de manera efectiva**. La inteligencia supone la **capacidad de adquirir conocimiento, aplicarlo de manera práctica y adaptarse a nuevas situaciones**.

La comprensión de la inteligencia ha evolucionado con el tiempo y ahora se reconocen diferentes tipos o manifestaciones de la inteligencia, como la inteligencia verbal, la inteligencia lógico-matemática, la inteligencia creativa, la inteligencia emocional, entre otras. Cada persona puede tener fortalezas y debilidades en diferentes áreas de inteligencia.

En el contexto de la **inteligencia artificial**, la **inteligencia** se refiere, de forma general, a la **capacidad de una máquina para realizar tareas que normalmente requerirían la intervención humana**. Los sistemas de inteligencia artificial utilizan algoritmos y modelos matemáticos para procesar grandes cantidades de datos y tomar decisiones basadas en patrones y reglas predefinidas. Aunque estos sistemas pueden ser muy eficientes en tareas específicas, como el reconocimiento de voz o la clasificación de imágenes, su inteligencia es limitada y no se asemeja a la inteligencia humana en su conjunto.

LA RELACIÓN ENTRE LA CONCIENCIA Y LA INTELIGENCIA

La **relación entre la conciencia y la inteligencia** es compleja y objeto de debate en la comunidad científica. Algunos argumentan que la conciencia es un producto emergente de la inteligencia, otros consideran que la inteligencia surgió a partir de conciencias complejas, mientras que una tercera corriente sostiene que son fenómenos independientes pero interrelacionados.

En el caso de los seres humanos, la conciencia y la inteligencia están estrechamente relacionadas. Nuestra conciencia nos permite ser conscientes de nuestra propia inteligencia y utilizarla para resolver problemas, aprender y adaptarnos al entorno. La conciencia también nos permite reflexionar sobre nuestras propias experiencias y emociones, lo que a su vez influye en nuestra inteligencia y toma de decisiones.

Desde la perspectiva de la **inteligencia artificial**, la conciencia es un desafío mayor. Aunque los sistemas de inteligencia artificial pueden **simular la inteligencia humana en ciertos aspectos**, hasta ahora **no han demostrado tener conciencia o una experiencia subjetiva del entorno**. Los sistemas de inteligencia artificial son capaces de **procesar información y tomar decisiones**, pero **carecen de capacidad de adaptación y de la experiencia interna que define la conciencia humana**.

Como punto de partida en un intento de desarrollar entidades artificiales, a continuación se plantean algunas **questiones fundamentales**:

- **¿Qué papel juega la conciencia en la inteligencia?**

La **conciencia** juega un papel fundamental en la inteligencia humana. Nuestra conciencia **nos permite reconocer la realidad y ser conscientes de nuestras propias aptitudes y conocimientos**, así como de nuestra **experiencia con el entorno**. Es la **capacidad de reconocer la realidad, relacionarnos con ella y adquirir experiencias dentro de la realidad reconocida**. La **conciencia** supone también tener **conocimiento de uno mismo, de nuestros pensamientos, emociones y acciones**. Nos permite ser conscientes de nuestras decisiones y de las consecuencias de nuestras acciones.

La conciencia influye en nuestra forma de pensar/razonar –esto es, procesar información-, aprender –esto es, mejorar nuestras funciones / habilidades / capacidades con el tiempo- y tomar decisiones –en base a lo razonado y aprendido-. Una buena conciencia otorga mejores oportunidades de éxito a la inteligencia –pero no necesariamente mejores capacidades cognitivas como es el razonamiento-. Por tanto, la conciencia necesariamente determina/influye_en las capacidades/oportunidades de la inteligencia.

Si la realidad o la forma de razonar reconocida por una persona difiere mucho de la realidad reconocida por la gran mayoría, entonces la adopción de acciones, decisiones y formas de actuar -esto es, la inteligencia- de dicha persona entrarán en conflicto con los de esta mayoría.

- **¿Cuál es la función de la inteligencia en la conciencia?**

La inteligencia nos brinda la capacidad de procesar información a través de procesos de razonamiento y aprendizaje y poder resolver problemas. El ser humano es inconformista por naturaleza, sintiendo una necesidad cada vez más imperiosa de comprender una realidad circundante cada vez más extensa y compleja. Cuanto más complejos sean los problemas, mayor reflexión, observaciones y cambios en la percepción de la realidad se requerirán. Las consecuencias de ello derivan en una percepción –capacidad de reconocer- cada vez más ajustada/clara de la realidad, con el consecuente enriquecimiento de la conciencia. Por tanto, se puede decir que la inteligencia enriquece la conciencia.

A su vez, tal como se describió en la primera cuestión, la nueva conciencia proporciona una percepción más ajustada de la realidad, permitiendo a la inteligencia –considerada adaptativa- evolucionar sus propios procesos de razonamiento y aprendizaje y así la toma de decisiones/acciones. Por tanto, la conciencia

y la inteligencia se enriquecen mutuamente.

- **¿Existe realmente un proceso de evolución del ser humano?**

A partir de las reflexiones sobre las dos anteriores cuestiones, podemos afirmar que la interacción conciencia-inteligencia es bidireccional y necesariamente está expuesta a un proceso de evolución hacia seres cada vez más conscientes e intelligentes.

- **¿Puede haber inteligencia, pero carecer de conciencia?**

Se trata de una pregunta controvertida y no compartida en su completa extensión por todos los investigadores. Suele haber una tendencia de que sí, es posible que una persona sea inteligente en términos de capacidad cognitiva para determinadas actividades/tareas, pero carezca de conciencia o una comprensión extensa y profunda de la realidad que la rodea, de sí misma y/o de los demás. En otras palabras, una escasa/nula conciencia no otorgará las capacidades necesarias a la inteligencia para el desarrollo de aquellas tareas cuya percepción/reconocimiento de la realidad circundante y de sí mismo no se ajusta a la propia realidad.

Sintetizando, la inteligencia en el desarrollo de una tarea requiere de conciencia, esto es, de capacidad para reconocer/percibir la realidad inherente a dicha tarea.

- **¿Es posible tener conciencia sin inteligencia?**

Determinados autores opinan que es posible tener conciencia sin tener un alto nivel de inteligencia. La conciencia se refiere a la capacidad de reconocer la realidad y relacionarse con ella, mientras que la inteligencia se refiere a la capacidad de razonar, aprender, comprender y resolver problemas. Una persona puede tener conciencia de sí misma y de su entorno sin tener un alto nivel de inteligencia en general.

- **¿La inteligencia y la conciencia, son procesos realmente independientes o están interrelacionados?**

La relación entre la conciencia y la inteligencia es compleja y ha sido objeto de debate entre filósofos, científicos y expertos en inteligencia artificial.

Algunos autores argumentan que **la conciencia y la inteligencia son independientes**, ya que:

1. Establecen que la inteligencia puede existir sin conciencia. Desde la perspectiva de la **Inteligencia Artificial**, el soporte de la inteligencia son los **algoritmos**, que procesan información y toman acciones y decisiones, sin ser conscientes de ello. Este aspecto conduce precisamente a la siguiente cuestión: **¿Cómo podrían mejorarse los algoritmos de los sistemas inteligentes actuales para convertirlos en realmente autónomos, sin necesidad de intervención humana?**
2. La conciencia puede existir sin inteligencia. Los seres humanos pueden experimentar estados de conciencia como el sueño o los recuerdos, sin necesariamente procesar información de manera inteligente.

No obstante, otros autores argumentan que **la conciencia y la inteligencia están interconectadas**, ya que:

3. La conciencia puede influir en la inteligencia, sobre lo cual ya se ha reflexionado anteriormente. Así, la conciencia **influye en nuestra forma de pensar/razonar, aprender y tomar decisiones.**
4. La inteligencia puede influir en la conciencia, como también se ha descrito anteriormente.

Por tanto, resulta más que difícil poder establecer actualmente de forma clara y precisa la relación entre la conciencia y la inteligencia.

REFLEXIONES

A pesar de los avances en la comprensión de la relación entre la conciencia y la inteligen-

cia, todavía quedan muchas preguntas y líneas de investigación abiertas, entre las que destacan:

1. **¿Cuál es la naturaleza exacta de la conciencia?** A pesar de los avances en el tema, todavía no se sabe exactamente qué es la conciencia.
2. **¿Cómo se relaciona la conciencia con la inteligencia?** Aunque la mayoría de expertos coinciden en que la conciencia y la inteligencia están relacionadas, todavía no se sabe exactamente cómo se relacionan.
3. **¿Cómo se puede desarrollar la inteligencia artificial consciente?** Parece que la **inteligencia artificial** –abreviado IA- **consciente** está todavía lejos de alcanzarse, siendo un objetivo a largo plazo que requiere una comprensión más profunda de la relación entre la conciencia y la inteligencia. No obstante, hay motivos para afirmar que es posible aproximarnos progresivamente a la comprensión y desarrollo de entidades artificiales conscientes a través de enfoques generalistas.

CONCLUSIONES: UN CAMINO HACIA LA INTELIGENCIA ARTIFICIAL CONSCIENTE

La vasta cantidad de estudios existentes sobre la conciencia y la inteligencia, todavía no han logrado establecer un consenso científico que defina de manera concisa y relacione ambos conceptos. Se sabe que su origen tiene lugar en varias zonas del cerebro. La complejidad del cerebro como mecanismo biológico que origina la conciencia y la inteligencia sólo ha permitido, hasta el momento, realizar aproximaciones sobre algunas áreas del mismo afectadas, pero todavía estamos muy lejos de determinar el proceso que conduce a la generación de los procesos conscientes e/o inteligentes.

A pesar de las dificultades inherentes, la Inteligencia Artificial proporciona vías para la progresión en el estudio, aunque el camino para la construcción de auténticas entidades artificiales con cierto grado de conciencia resulte largo y escabroso.

El objetivo de este trabajo se centra en el diseño de un **modelo autónomo de aprendizaje artificial** capaz de soportar cierto grado de conciencia.

REFLEXIONES SOBRE LOS CONCEPTOS DE INTELIGENCIA – RAZONAMIENTO – EXPERIENCIA – CONOCIMIENTO DEL ENTORNO DESDE LA PERSPECTIVA DE LA INTELIGENCIA ARTIFICIAL

En esta sección se exponen una serie de definiciones sobre los principales conceptos utilizados en el contexto de la Inteligencia Artificial en su camino hacia el desarrollo de entidades –conocidas como **agentes**- inteligentes.

Tales conceptos ya han sido tratados de forma general en la sección 1, por lo que la intención en esta sección es proporcionar una aproximación ajustada, aunque quizás limitada por el contexto, de los mismos. Además, la interpretación de alguno de ellos difiere entre numerosos autores de la abundante literatura en Inteligencia Artificial.

Definición 2.1 (Inteligencia de un agente en un entorno determinado). Capacidad del agente para adquirir, procesar y aplicar conocimiento para resolver problemas en un entorno dado. La INTELIGENCIA se basa en el conocimiento explícito del que dispone el agente y en su capacidad para procesarlo. Un agente cuyo conocimiento es pobre y/o la capacidad de procesamiento sea reducida, tendrá una inteligencia bastante limitada.

El **grado de inteligencia** de un agente se mide por el éxito con el que éste alcanza sus metas. El término «éxito» puede adoptar di-

ferentes interpretaciones dependiendo del usuario experto que explote los servicios del agente.

La inteligencia es un proceso, cuyas componentes son: El **Razonamiento**, el **Conocimiento** proporcionado por el entorno/ambiente y las **Experiencias** derivadas del razonamiento.

Definición 2.2 (Razonamiento de un agente en un entorno determinado). Capacidad del agente para adquirir conocimientos relevantes sobre el entorno. El razonamiento supone la **explicitación/descubrimiento de conocimiento implícito en una Base de Conocimientos**. Se consideran entornos complejos/evolutivos/no- predecibles, donde resulta difícil/imposible captar todas las características del entorno en un mismo instante.

Un buen razonamiento debería conducir ineludiblemente a la mejora de la capacidad de la inteligencia. No obstante, no es suficiente razonar adecuadamente para garantizar una buena inteligencia.

Los **resultados del razonamiento** conducen a las **experiencias** del agente. Un buen razonamiento conduce a experiencias positivas sobre el propio razonamiento, mientras que un mal razonamiento conduce a experiencias negativas. Tales **experiencias** deberían servir necesariamente para **mejorar –esto es, reforzar o corregir- el proceso de razonamiento de cara al futuro**.

Otro factor que afecta directamente a la inteligencia es el **conocimiento del entorno –percepciones** del agente-. Sin conocimiento no hay razonamiento y sin razonamiento no hay inteligencia.

Así pues, el **razonamiento, las experiencias derivadas del mismo y el conocimiento sobre el entorno**, son los tres elementos/componentes que constituyen la inteligencia de un agente.

El **razonamiento es una componente procedimental** que hace uso de **conocimiento, tanto del entorno como de las experiencias**.

Definición 2.3 (Experiencias de un agente en un entorno). Conocimiento que se adquiere sobre un entorno a partir de la práctica prolongada y/o vivencias en el mismo. Las experiencias son consecuencia de la aplicación de los resultados del razonamiento sobre el entorno.

Al proporcionar conocimientos sobre el entorno, la **experiencia debería incrementar la capacidad de profundizar del razonamiento**, y así mejorar las metas del agente.

Las **experiencias son una componente de conocimiento acerca del entorno** que resultan de la aplicación de los resultados de un proceso de razonamiento sobre el propio entorno.

Definición 2.4 (Conocimiento sobre el entorno, C_Entorno). Conocimiento percibido por un agente durante la exploración de su entorno. Este conocimiento puede adoptar diferentes formatos dependiendo de sus características y del lenguaje usado para expresarlo –conocimiento basado en datos, conocimiento lógico, conocimiento procedimental, etc....-

El **C_Entorno** proporciona nuevas evidencias/hechos sobre el entorno, lo cual redundará necesariamente en los resultados del razonamiento y de las experiencias que se obtienen del mismo. Por tanto, puede afirmarse que el **C_Entorno proporciona nuevos horizontes a la inteligencia**.

De las anteriores definiciones se concluye que La **INTELIGENCIA** incluye/se ve afectada necesariamente por el **RAZONAMIENTO**, la **EXPERIENCIA** y el **CONOCIMIENTO SOBRE EL ENTORNO**, puesto que estos en conjunto permiten incrementar las habilidades de un agente en el desarrollo de su función. Esto es:

C_ENTORNO + RAZONAMIENTO + EXPERIENCIA c INTELIGENCIA. [A]

Obsérvese que los términos de la expresión [A] están relacionados por el símbolo de inclusión - c -, tratándose solamente de una

aproximación al concepto de **INTELIGENCIA**, puesto que se desconoce el significado del mismo en su totalidad.

En la figura 1 se ilustra gráficamente una relación cíclica entre tales elementos.

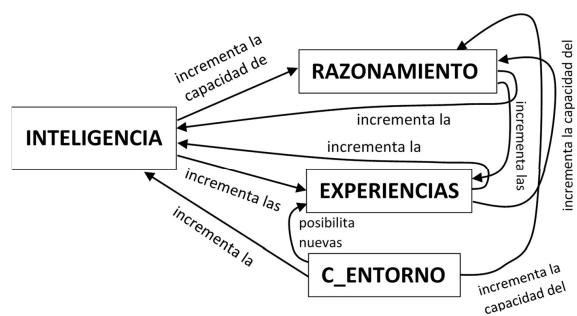


Figura 1. Relación INTELIGENCIA-RAZONAMIENTO-EXPERIENCIA

SOBRE UNA CONCIENCIA ARTIFICIAL DE OBJETIVOS

La investigación sobre la conciencia en IA aborda la cuestión de cómo los sistemas artificiales podrían desarrollar experiencias subjetivas similares a las humanas. Se trata de un complejísimo problema que pretende entender «por qué» algunos procesos biológicos neuronales generan estados conscientes y «cómo» traducir tales procesos en términos de procesamientos de información [4], [5].

La magnitud del concepto de **CONCIENCIA** es enorme, hasta el punto de que su conceptualización resulta muy difícil. Todavía no existe el suficiente consenso entre investigadores para establecer una definición que sirva de referencia. Las ideas manifestadas en [6], según las cuales se distinguen al menos tres tipos de conciencia: **conciencia de agencia**, **conciencia de objetivos** y **conciencia sensoriomotora**, delimitarán el estudio de este trabajo.

Así, a modo de aproximación al estudio de la conciencia artificial, nos centraremos en la **conciencia de objetivos en Inteligencia Artificial**, la cual se refiere a la **capacidad** de un sistema de inteligencia artificial para comprender y alinearse con los objetivos y metas

de los usuarios para los que fue creado. Esto supone que el sistema de IA es capaz de:

- **Entender los objetivos.** Esto es, el sistema de IA puede comprender los objetivos y metas que se le han asignado.
- **Alinearse con los objetivos.** El sistema puede adaptar su comportamiento y toma de decisiones para lograr los objetivos establecidos.
- **Evaluar el progreso.** El sistema puede evaluar su propio progreso hacia los objetivos y ajustar su comportamiento según sea necesario.

La conciencia de objetivos es importante en IA por una serie de razones:

- **Mejora la eficiencia.** Un sistema de IA que comprende y se alinea con los objetivos puede ser más eficiente en la toma de decisiones.
- **Reduce errores.** Un sistema de IA que comprende y se alinea con los objetivos puede reducir errores y comportamientos no deseados.
- **Aumenta la confianza.** La conciencia de objetivos puede aumentar la confianza en los sistemas de IA, ya que se pueden garantizar/esperar resultados más predecibles y alineados con los objetivos.

La conciencia de objetivos en IA también plantea desafíos y limitaciones, tales como:

- **Definir los objetivos de forma clara y precisa.** Para que el sistema de IA pueda comprenderlos y alinearse con ellos.
- **Complejidad de los objetivos.** Los objetivos pueden ser complejos y tener múltiples facetas, lo que puede dificultar su comprensión y alineación por parte del sistema de IA.
- **Ética y valores.** La conciencia de objetivos también plantea cuestiones éticas y de valores, con lo cual los sistemas de IA deberán tomar decisiones acordes con la ética y valores preestablecidos.

Cualquier proceso basado en una conciencia de objetivos «debe progresar siguiendo una serie de pautas esenciales que atañen a la conducta del sujeto», entre las que se incluyen:

- i. La (re)definición de los objetivos del sujeto en cada situación. Éstos incluyen las metas del sujeto en relación con la realidad circundante y con los propósitos inherentes/propios al/del propio sujeto. Por ejemplo, el sujeto quiere subir a un árbol sin dañarse. Pueden, obviamente, existir diferentes objetivos con cierto grado de logro contradictorio entre ellos, esto es, con un grado de dificultad razonable para la satisfacción de todos ellos.
- ii. Capacidad de reflexión del sujeto para ponderar la viabilidad de los objetivos en cada momento.
- iii. Capacidad del sujeto para la planificación inteligente –fabricación de procedimientos “razonables/inteligentes”- para la toma de decisiones en aras de alcanzar los objetivos previamente ponderados.
- iv. Capacidad para generar decisiones a través de la ejecución de los planes – procedimientos/programas- previamente fabricados, con objeto de adoptar acciones/decisiones que permitan alcanzar los objetivos previamente ponderados.

CONCIENCIA DE OBJETIVOS DE UNA TAREA EN UN CONTEXTO DADO

Cada una de las anteriores pautas requiere del conocimiento necesario y de los procedimientos adecuados para poder ser llevada a cabo.

En lo que se refiere a **objetivos relativamente complejos definidos en un mismo contexto/escenario de un entorno**, donde concurren diferentes subobjetivos, entonces es posible que los procedimientos asociados a

diferentes pautas deban **compartir y/o actualizar de manera sincronizada conocimiento común**. Por tanto, es necesario considerar la posibilidad de **sincronización** entre los **procedimientos asociados a las diferentes pautas** –y posiblemente de una misma pauta-. Un ejemplo lo encontramos durante el desarrollo de la tarea de conducir un vehículo, donde el agente conductor debe alcanzar el punto de destino a partir de su situación de origen. Considerando que el contexto/escenario es la conducción a través de las vías disponibles –carreteras, autopistas, caminos, etc...-, el conductor se ve obligado a tener claros en todo momento cuáles son sus objetivos –pueden existir diferentes posibilidades/rutas para alcanzar el destino-, la capacidad para ponderarlos debidamente de acuerdo con el tipo de vehículo –en caso de un camión de carga, debe tomar rutas seguras como buenas carreteras o autopistas, lo cual no es necesario en caso de conducir una moto-, la capacidad para planificar cada subobjetivo –puntos intermedios entre el origen y el destino- y la capacidad para tomar decisiones siguiendo el plan elegido.

De acuerdo con:

- la definición de INTELIGENCIA en la sección 1:

«Capacidad del agente para adquirir, procesar y aplicar conocimiento para resolver problemas en un entorno dado». [B]

- Las consideraciones realizadas en la sección 1 y las posibilidades actuales de la IA. Surge la siguiente cuestión:

«¿Cuáles son los medios en IA a través de los que se manifiesta la “capacidad” de un supuesto agente inteligente?»

La solución que ofrece la IA actual se basa en la ejecución de planes mediante *programas informáticos*, independientemente de su complejidad y contextos considerados en su diseño y construcción. Por tanto, el núcleo del intelecto de un agente inteligente es un **pro-**

grama

 –implementación de un plan- cuyo diseño y construcción corresponden «actualmente» a un experto humano. Un **agente inteligente** es, simplemente, un **programa en ejecución**.

De acuerdo con [B], la función asignada a la **inteligencia** se corresponde, de manera general, con las con las **pautas esenciales iii. y iv. de la conciencia de objetivos**, es decir, a la capacidad de fabricar programas/algoritmos para la toma de decisiones y a su vez a la capacidad de ejecutar dichos programas/algoritmos con objeto de producir las oportunas decisiones.

Por otra parte, la **conciencia de objetivos** –véase la sección 3- se refiere a la **capacidad** de un sistema de inteligencia artificial para **comprender y alinearse con los objetivos y metas de los usuarios para los que fue creado, progresando de acuerdo con las cuatro pautas esenciales descritas**.

La figura 2 ilustra Los componentes de la conciencia de objetivos y sus relaciones con la inteligencia.

De la figura 2 se observa que la INTELIGENCIA forma parte del grupo de componentes/pautas que constituyen/afectan la conciencia de objetivos, esto es:

INTELIGENCIA c CONCIENCIA DE OBJETIVOS [C]

Por tanto, desde esta perspectiva de la inteligencia artificial, se puede afirmar que la inteligencia constituye una manifestación de la conciencia de objetivos, esto es, es un proceso llevado a cabo de forma consciente y reflexiva. Sin inteligencia no hay conciencia que valga, pues el sujeto/agente no podrá llevar a cabo los objetivos de la conciencia a través del proceso de toma de decisiones, puesto que tal proceso no existe. Así pues, debe existir necesariamente alguna forma de inteligencia para que un sujeto/agente manifieste algún grado de conciencia.

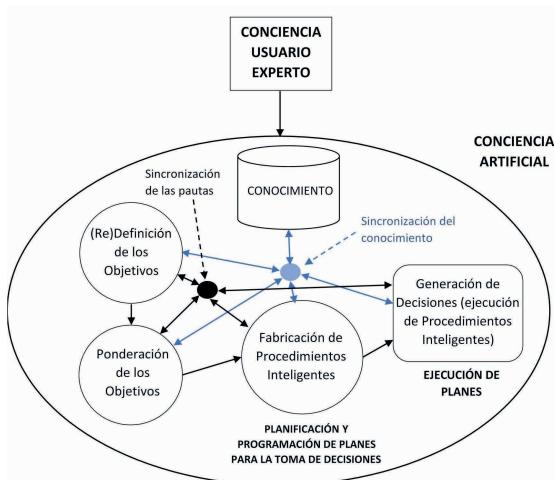


Figura 2. Componentes de la conciencia de objetivos y sus relaciones para la ejecución de una tarea en un contexto dado.

De la figura 2 se contempla que la conciencia artificial –en realidad se refiere a la conciencia artificial de objetivos- debe estar alineada con la conciencia del usuario experto, con objeto de mantener su confianza. Además, es adaptativa en el tiempo, progresando hacia el objetivo planteado. El progreso es quien debe determinar el tipo de adaptación –y no al revés-; en otras palabras, antes de adaptarse deben conocerse los motivos para ello. En este sentido, surge una importantísima cuestión:

«¿Cómo puede un agente autónomo evaluar/validar los resultados de su progreso?»

La respuesta tiene un nombre: «Las experiencias del agente en su entorno»

La importancia de las experiencias de un agente autónomo está fuera de toda duda. Sin experiencias no hay inteligencia y sin inteligencia no hay conciencia –estas afirmaciones se contemplan en el marco de los agentes inteligentes, no de una forma generalizada, como pueden ser los agentes biológicos-.

Como síntesis de las anteriores reflexiones, podemos afirmar que la **conciencia de objetivos** es un **proceso activo**, que consta un conjunto de **pautas esenciales basadas en conocimiento**, de manera que tanto el conocimiento como las pautas deben estar **sincronizados** adecuadamente.

CONCIENCIA DE OBJETIVOS DE MÚLTIPLES TAREAS EN UN MISMO CONTEXTO

En esta sección se realiza una reflexión en el caso de un agente que se desenvuelve en un contexto en el que puede desarrollar múltiples

tareas. Se trata de un caso muy común y su importancia radica en la interdependencia entre varias tareas –o simplemente en la dependencia de unas tareas con respecto a otras- con objeto de alcanzar un objetivo común.

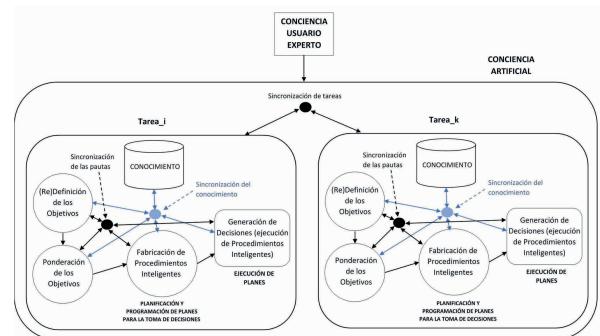


Figura 3. Componentes de la conciencia de objetivos y sus relaciones para la ejecución de dos tareas en un contexto dado.

Resulta común que los niveles de inteligencia de un sujeto no sean los mismos para todas las tareas, lo que significa que el sujeto es más hábil en algunas tareas y más torpe en otras. La conciencia de objetivos deberá permitir llevar a cabo cualquiera de las tareas del contexto siempre que los objetivos sean alcanzables.

Para llevar a cabo los objetivos propuestos se precisa de la **sincronización entre cada par de tareas interdependientes** –o simplemente dependientes-. Considerando que los **elementos de la conciencia de objetivos de cada tarea** son su **propio conocimiento y pautas esenciales para su progreso**, tal como se ilustra en la figura 2, entonces la sincronización debe incluir ambos elementos. La figura 3 constituye una extensión de la figura 2, donde se ilustra de forma general la reflexión que acaba de hacerse para un contexto con dos tareas interdependientes –o simplemente dependientes-.

CONCIENCIA DE OBJETIVOS DE MÚLTIPLES TAREAS EN MÚLTIPLES CONTEXTOS

Cuando un agente actúa en un entorno en el cual puede adoptar diferentes contextos/escenarios –como es el caso de los seres humanos-, su conciencia debe abarcar las dimensiones de todos los contextos y coordinar/sincronizar el comportamiento global del agente. En este sentido, la complejidad de los agentes que se desenvuelven en múltiples contextos es superior a la de aquellos que operan en un único contexto.

En este nuevo paradigma, la **sincronización entre grupos de contextos** deberá ser necesariamente **imperativa**. A modo de ejemplo, considérense las formas típicas de actuar de un humano en un contexto profesional y en un contexto familiar de ocio. Tales contextos pueden compartir diferentes situaciones, como es el caso de que un humano reciba a un amigo o familiar durante el desarrollo de su tarea profesional, no obstante, es razonable que las decisiones del humano respecto al amigo sean diferentes en el contexto familiar que en el profesional; otro ejemplo lo encontramos en el significado/interpretación diferente de una misma oración en diferentes contextos. Es la conciencia la que permite al humano diferenciar entre diferentes contextos, lo cual determina su forma de adaptarse con objeto de tomar las decisiones más adecuadas al caso. Por tanto, la **conciencia es el elemento sincronizador entre los diferentes contextos de un agente**.

No forma parte del objetivo de este trabajo la profundización en el estudio de la conciencia de objetivos en múltiples contextos, simplemente se realizarán algunas reflexiones al objeto de vislumbrar algunas características importantes de un hipotético modelo de IA.

Resulta patente que los humanos no somos igualmente conscientes en cada uno de los contextos que conforman la conciencia ge-

neral. También resulta palpable que un mismo sujeto no tiene el mismo nivel de conciencia en el desarrollo de cada una de las tareas de un mismo contexto, esto es, **cada sujeto tiene un nivel de conciencia particular respecto a cada una de las tareas de un mismo contexto**, conocida como **conciencia contexto-tarea**.

Resulta común que, durante la ejecución de una determinada tarea en un cierto contexto, alguno/s de los subobjetivos requieran el desarrollo de subtareas de otros contextos, con lo cual el agente deberá cambiar temporalmente su conciencia contexto-tarea, de acuerdo con las necesidades que exige el desarrollo de la tarea. Esto significa que debe existir algún tipo de **sincronización** entre las diferentes **conciencias contexto-tarea**.

SOBRE LAS EXPERIENCIAS

INTRODUCCIÓN

A menudo hemos oído y quizás también utilizado frases como “la vida es un camino de experiencias”, “lo que realmente da sentido a la vida son las experiencias”, “las experiencias son lo que nos mueve en esta vida”,

Pero, ¿qué son las experiencias humanas?, ¿qué relación tienen con la conciencia y la inteligencia?

La literatura sobre el estudio de las experiencias en inteligencia artificial no es precisamente abundante. En esta subsección se realiza una breve síntesis del concepto de «experiencias» desde la perspectiva académica. [7].

La palabra “experiencia” comparte una raíz latina común con la palabra “experimentación”. **Experiencia** se refiere a los **eventos conscientes en general**, más específicamente a las **percepciones**, o al **conocimiento práctico que es producido por estos procesos conscientes**. Entendida como un **evento consciente en el sentido más amplio**, la **experiencia implica un sujeto al que se le presentan diversos elementos**. En este sentido, ver un pájaro amarillo en una rama presenta al su-

jeto los objetos “pájaro” y “rama”, la relación entre ellos y la propiedad “amarillo”. Cuando se entiende en un sentido más restringido, solo la conciencia sensorial cuenta como experiencia. En este sentido, la experiencia suele identificarse con la percepción y contrastarse con otro tipo de eventos conscientes, como pensar o imaginar. En un sentido ligeramente diferente, la experiencia no se refiere a los eventos conscientes en sí, sino al conocimiento práctico y la familiaridad que producen. En este sentido, es importante que el contacto perceptivo directo con el mundo exterior sea la fuente del conocimiento. Entonces, un excursionista experimentado es alguien que realmente vivió muchas caminatas, no alguien que simplemente leyó muchos libros sobre caminatas.

Muchos debates académicos sobre la naturaleza de la experiencia se centran en la **experiencia como un evento consciente**, ya sea en un sentido amplio o más restringido.

En la literatura académica se discute una gran variedad de tipos de experiencias. Las **experiencias perceptivas**, por ejemplo, representan el mundo exterior a través de estímulos registrados y transmitidos por los sentidos. La **experiencia de la memoria episódica**, por otro lado, implica revivir un evento pasado que uno experimentó antes. En la **experiencia imaginativa**, los objetos se presentan sin pretender mostrar cómo son realmente las cosas. La **experiencia de pensar** involucra representaciones mentales y el procesamiento de información, en el cual se consideran, juzgan o conectan ideas o proposiciones. El **placer** se refiere a la experiencia que se siente bien. Está estrechamente relacionado con la **experiencia emocional**, que tiene además componentes evaluativos, fisiológicos y conductuales. Los **estados de ánimo** son similares a las emociones, con una diferencia clave que es que carecen de un objeto específico que se encuentra en las emociones. Los **deseos conscientes**

involucran la experiencia de querer algo. La **experiencia no ordinaria** se refiere a experiencias raras que difieren significativamente de la experiencia en el estado de vigilia ordinario, como experiencias religiosas, experiencias fuera del cuerpo o experiencias cercanas a la muerte.

LA EXPERIENCIA COMO EVENTO CONSCIENTE

La **experiencia** se entiende a menudo como un **evento consciente** en el sentido más amplio. Esto incluye varios tipos de experiencias, como la percepción, la conciencia corporal, la memoria, la imaginación, la emoción, el deseo, la acción y el pensamiento. Generalmente se refiere a la experiencia que tiene un individuo en particular, pero también puede tomar el significado de la experiencia que tuvo un grupo de individuos.

Cuando alguien tiene una experiencia, se le presentan varios elementos. Estos elementos pueden pertenecer a diversas categorías ontológicas correspondientes, por ejemplo, a objetos, propiedades, relaciones o eventos. Las experiencias pueden incluir solo elementos reales, solo elementos irreales o una combinación de ambos.

Cuando se entiende en un sentido más restringido, solo la conciencia sensorial cuenta como experiencia. En este sentido, es posible experimentar algo sin comprender qué es. Este sería el caso, por ejemplo, si alguien experimentara un robo sin saber qué estaba pasando exactamente. En este caso, las sensaciones provocadas por el robo constituyen la experiencia del robo. Los críticos a menudo señalan que la experiencia involucra varios componentes cognitivos que no pueden reducirse a la conciencia sensorial. Otro enfoque consiste en distinguir entre la experiencia interna y la externa. Entonces, si bien la percepción sensorial pertenece a la experiencia externa, también puede haber otros tipos de experiencia, como recordar o imaginar, que pertenecen a la experiencia interna.

LA EXPERIENCIA COMO CONOCIMIENTO

Desde otra perspectiva, la **experiencia** no se refiere a los hechos conscientes en sí, sino al **conocimiento** que producen. Para este sentido, es importante que el **conocimiento se produzca a través del contacto perceptivo directo con el mundo exterior**. Que el conocimiento sea directo significa que se obtuvo a través de la observación inmediata, es decir, sin implicar ninguna inferencia. Uno puede obtener todo tipo de conocimiento indirectamente, por ejemplo, leyendo libros o viendo películas sobre el tema. Este tipo de conocimiento no constituye experiencia del tema ya que el contacto directo en cuestión concierne solo a los libros y películas, pero no al tema en sí.

INTENCIONALIDAD DE LA EXPERIENCIA

La mayoría de las **experiencias**, especialmente las de tipo perceptivo, apuntan a **representar la realidad**. Esto suele expresarse afirmando que **tienen intencionalidad** o se trata de su objeto intencional. Si tienen éxito o son verídicos, representan el mundo tal como es en realidad. Pero también pueden fallar, en cuyo caso dan una representación falsa. **Tradicionalmente se sostiene que toda experiencia es intencional**. Esta tesis se conoce como “intencionalismo”. En este contexto, a menudo se afirma que todos los estados mentales, no solo las experiencias, son intencionales. Pero se suele dar especial protagonismo a las experiencias en estos debates ya que parecen constituir la forma más fundamental de intencionalidad.

Se acepta comúnmente que todas las experiencias tienen características fenoménicas, es decir, que hay algo que se siente al vivirlas. Quienes se oponen al intencionalismo afirman que no todas las experiencias tienen características intencionales, es decir, que las características fenoménicas y las características inten-

cionales pueden separarse. Algunos supuestos contraejemplos del intencionalismo implican experiencias sensoriales puras, como el dolor, de las que se afirma que carecen de componentes representacionales. Los defensores del intencionalismo a menudo han respondido afirmando que, después de todo, estos estados tienen aspectos intencionales, por ejemplo, que el dolor representa daño corporal.

EXPERIENCIAS E INTELIGENCIA ARTIFICIAL

En la sección 3.1 se discutió sobre el **carácter adaptativo de la conciencia como elemento fundamental para el progreso hacia los objetivos propuestos**. Esto es, la conciencia debe ir adaptándose con el paso del tiempo, de manera que garantice un progreso adecuado del agente hacia los objetivos marcados. Los resultados del progreso en un cierto momento indicarán, posiblemente, la necesidad de cambios en los mecanismos de la conciencia. Las **causas de esta necesidad de cambios pueden ser diversas**, algunos ejemplos los encontramos en: (a) el hecho de que el progreso en el momento dado se esté alejando peligrosamente de los objetivos; (b) las circunstancias del entorno han sufrido cambios y los objetivos propuestos u otras actividades predeterminadas deben ser reevaluados; (c) el agente ha sufrido algún percance recientemente; (d) algunos resultados obtenidos en la etapa de

«Generación de Decisiones –ejecución de Procedimientos Inteligentes–» de la figura 2, entra en conflicto con determinados hechos del entorno; etc ... Todas estas situaciones y circunstancias deben ser asumidas por la conciencia del agente, de manera que ésta pueda adaptarse a las mismas y continuar con el progreso exitoso de su tarea.

Como aproximación y **de acuerdo con la fuente de origen**, pueden distinguirse los siguientes **tipos de causas** de las experiencias:

i. **Causas debidas a cambios producidos en el entorno.** Los cambios del entorno pueden provocar diferentes situaciones en la conciencia del agente, tales como la necesidad de cambiar los objetivos, la obligación de cambiar la planificación o la implementación de algún plan, ...

ii. **Causas debidas al usuario experto externo.** El usuario experto puede supervisar el progreso del agente y manifestar la necesidad de que es necesario alinear mejor los resultados del agente con sus propios objetivos.

iii. **Causas provocadas por el grado de éxito/fracaso de los resultados del agente en un momento dado.** En este caso, puede que el propio **agente** También avalúe/valore el nivel de éxito/fracaso de los resultados del agente, con lo que **se convierte en la fuente de origen** de experiencias. Si hay un **usuario externo** que evalúa los resultados del agente, entonces éste **constituye una fuente de experiencias**.

De manera general, las **experiencias** se refieren a la **acumulación a lo largo del tiempo de conocimientos, habilidades y en general “sabiduría” a través de la práctica, la observación y la participación en diferentes situaciones**. Esta “sabiduría” se va acumulando en la memoria del agente, y **su función** es la de servir como **complemento del conocimiento que percibe el agente directamente desde el entorno, con objeto de tomar las decisiones oportunas sobre su progreso**.

Por tanto, los **elementos de conocimiento** que guían el progreso de un agente son las **percepciones del entorno** y las **experiencias**. Sin **experiencias** resulta muy difícil, por no decir imposible, una evolución apropiada de la conciencia de un sujeto.

De lo que acaba de decirse, podemos afirmar que la **conciencia de un agente inteligente abarcará** –posiblemente entre otros–,

además de los componentes mencionadas en la sección 3 –véase la figura 2–, los **conocimientos de las experiencias** y de las **percepciones del entorno**, que irán progresando con el tiempo.

Los siguientes puntos describen algunas cuestiones clave sobre las experiencias en IA:

- **¿Cómo se generan las experiencias?**

En el contexto de la IA, el término **experiencias** puede describirse mediante la siguiente frase: «**las experiencias** se refieren a la **acumulación a de conocimientos a través de la práctica, la observación y la participación en diferentes situaciones**». Por tanto, las **experiencias** son en realidad **conocimiento**. La cuestión aquí es determinar **cómo se genera este conocimiento**.

Para generar **experiencias**, el agente debe tener la **capacidad de confrontar/contrastar el conocimiento previamente adquirido por él mismo –este es, el conocimiento para alcanzar sus objetivos- con la realidad del entorno y/o la exposición a otros datos/conocimiento –provenientes posiblemente de diferentes fuentes–. Los resultados de este proceso de confrontación se expresan mediante conocimiento y constituyen las experiencias**.

Por tanto, la **conciencia debe, necesariamente, incluir algún proceso que permite contrastar/aplicar el conocimiento adquirido por el agente frente a conocimiento proveniente del entorno u otras fuentes de conocimiento/datos**.

Tales procesos adoptan un **formato procedural** –son algoritmos–.

- **¿Cuándo se generan las experiencias?**

Las **experiencias** se generan en el momento en que el agente aplica los procesos de contrastación/confrontación del conocimiento adquirido con el entorno u otras fuentes de conocimiento.

- **¿Son todas las experiencias realmente útiles para la progresión del agente hacia los objetivos?**

los procesos de contrastación/aplicación del conocimiento adquirido con otras fuentes de conocimiento son los que deciden hasta qué punto sus resultados resultan útiles como **experiencias**. Pueden asignarse prioridades o niveles de importancia entre las diferentes **experiencias**.

- **¿Para qué sirven las experiencias?**

Las **experiencias** constituyen **conocimiento útil que el agente va acumulando a lo largo del tiempo**, mediante la confrontación/aplicación del conocimiento previamente adquirido por él mismo a través de la interacción con el entorno y/o la exposición a otros datos/conocimiento. Dicho conocimiento **se utiliza junto con el conocimiento de las percepciones que el agente recibe del entorno, con ob-**

jeto de adoptar las decisiones que conducen a mejorar su progreso.

Como conclusión, desde el punto de vista de la IA se puede establecer que **las percepciones y las experiencias son dos elementos de conocimiento fundamentales en el logro de los objetivos** del agente.

La figura 4 ilustra gráficamente las características de una conciencia guiada, entre otros elementos de conocimiento, por las percepciones y las experiencias del agente. Obsérvese como el dispositivo de **Generación de Decisiones (ejecución de Procedimientos Inteligentes)**, se ha considerado como un elemento de conocimiento, pues se interpreta que contiene las decisiones generadas por la ejecución de los programas/algoritmos del propio dispositivo, esto es, se trata del conocimiento adquirido por el agente.

Algunas observaciones que destacan en la figura 4 son:

- Se ha incluido el elemento procedural **«Generador de experiencias»**, cuya función es **contrastar/aplicar las decisiones/acciones** inteligentes del elemento de conocimiento **«Generación de Decisiones (ejecución de Procedimientos Inteligentes)»** –dicho elemento constituye el **conocimiento adquirido** por el agente.
- El **Generador de experiencias** debe estar convenientemente sincronizado con el elemento **«CONOCIMIENTO»**, puesto que la generación de experiencias puede requerir conocimiento sobre los objetivos, los planes trazados para alcanzarlos o datos del entorno sobre los que deben tomarse decisiones.
- Se ha incluido el elemento de conocimiento **«CONOCIMIENTO (EXPERIENCIAS)»**, que incluye los resultados del **Generador de experiencias**, esto es, las experiencias. Este elemento

debe estar convenientemente sincronizado, posiblemente entre otros, con el elemento **«(Re)Definición de los Objetivos»**, puesto que las experiencias pueden sugerir una redefinición de los objetivos previos. Lo dicho puede tener consecuencias sobre los elementos procedurales **«Ponderación de los Objetivos»** y **«Fabricación de Procedimientos Inteligentes»**.

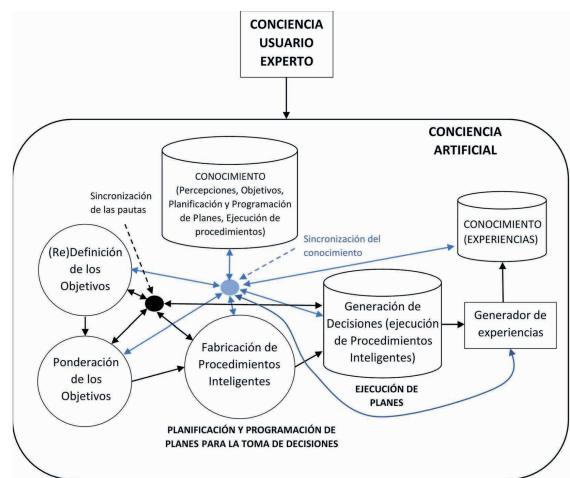


Figura 4. Componentes de la conciencia de objetivos con experiencia.

EL APRENDIZAJE AUTÓNOMO EN IA CON CONCIENCIA DE OBJETIVOS

Aunque el **aprendizaje** es uno de los términos más de actuales en IA, parece no existir todavía un consenso entre diferentes autores sobre la precisión de su significado. La definición 5.1 ofrece una descripción general ampliamente aceptada.

Definición 5.1 (Aprendizaje). Proceso mediante el cual un agente mejora su función. En otras palabras, el aprendizaje es la capacidad de un agente de progresar hacia sus objetivos.

Una aproximación más específica se ofrece en [8], donde se dice que un **agente aprende** cuando su **desempeño mejora con la experiencia** y mediante el **uso de datos**; es decir, cuando la habilidad no estaba presente en su

genotipo o rasgos de nacimiento.

Otros autores [9], [10], describen el aprendizaje en IA como un proceso de entrenamiento a partir de un conjunto de datos, con objeto de construir un modelo basado en estos datos que constituye una hipótesis del entorno. No obstante, dicha definición no considera la magnitud real de un proceso de aprendizaje, por diferentes razones fundamentales:

- Tales procesos se limitan a «razonar» – no a «aprender»- a partir de cantidades ingentes de datos y cuyos resultados quedan fijados por los datos introducidos inicialmente, por lo cual, en caso de entornos complejos, hay que repetir todo el proceso si aparecen nuevos datos o nuevas variables en la descripción de los mismos.

Tales procesos se conocen como «**procesos a un paso**».

- No hacen uso de las experiencias previas. Es lógico, pues al ser procesos a un paso, no existe la posibilidad de considerar experiencias y, por tanto, la posibilidad de retroalimentación del sistema de aprendizaje.

El **objetivo de esta sección** es la exploración de un **modelo de aprendizaje en IA basado en la conciencia de objetivos**, tal como se ha descrito en la sección 3.2. Para ello se seleccionará un determinado **tipo de problema** muy común en el campo del aprendizaje de las máquinas. El **resultado** será un **modelo con conciencia de objetivos** con los correspondientes componentes debidamente sincronizados.

PROBLEMA TIPO: APRENDIZAJE AUTÓNOMO A PARTIR DE EJEMPLOS O DATOS

Para facilitar la claridad de la lectura, la discusión se limitará a la **conciencia de objetivos de una tarea en un contexto dado**, tal como se describe en la sección 3.2.

El aprendizaje a partir de ejemplos o datos, constituye uno de los campos más estudiados en IA desde las últimas décadas. Sus orígenes se encuentran a finales de la década de los 70 y principios de los 80, donde aparecieron algunos de los primeros trabajos con gran resonancia en el campo, entre los que cabe citar, entre muchos otros, los clásicos [11], [12], [13], [14], [15].

Con objeto de articular el diseño de un modelo de aprendizaje autónomo a partir de ejemplos de acuerdo con la conciencia de objetivos, es importante establecer de forma precisa los términos y conceptos utilizados en IA –algunos de los cuales ya han sido introducidos en secciones anteriores-.

En ciertos universos complejos se considera que el ambiente adopta, en cada momento, una determinada **situación** [17], la cual puede verse como una especie de «estado complejo», difícil de manejar en el contexto de un **agente basado en metas – ABM-** [8].

Una situación describe las características del ambiente en un momento considerado, haciendo uso para ello de una cantidad ingente de conocimiento –**base de conocimientos**-. La cantidad de situaciones es, en general, muy elevada, y en muchos problemas resulta imposible de predecir.

La solución de los problemas se obtiene a partir de transiciones entre situaciones a través de la toma de decisiones –acciones- hasta dar con una situación objetivo.

No todo el conocimiento relativo a una situación está explicitado –es decir, el agente no tiene acceso directo al mismo- en la base de conocimientos. Esta es una cuestión muy importante, pues para llevar a cabo de forma segura y adecuada la transición entre situaciones, el agente deberá descubrir conocimiento no explícito en su base de conocimientos –esto es, conocimiento que, a pesar de no estar explicitado, está implícito en La base de conocimientos-. El proceso de llevar a cabo la explicitación de conocimiento se conoce como

razonamiento.

La figura 5 ilustra el razonamiento como un proceso de transición entre situaciones.

RAZONAMIENTO

El razonamiento es una de las más potentes herramientas de explotación del conocimiento sobre las que se sustenta la inteligencia humana. Los agentes que razonan constituyen uno de los tipos más «completos» de agentes en cuanto a capacidades cognoscitivas –sólo superados por los agentes que aprenden, pues éstos precisan de la capacidad de razonar para aprender.

El razonamiento siempre se lleva a cabo sobre un conocimiento considerado. No hay razonamiento sin conocimiento. Así pues, el razonamiento va directamente asociado a una representación particular del conocimiento.

El razonamiento es un **proceso inferencial** basado en algún lenguaje –el lenguaje usado para representar el conocimiento–.

Un proceso inferencial pretende obtener conocimientos implícitos en una base de conocimientos a partir de los conocimientos explícitos en la misma.

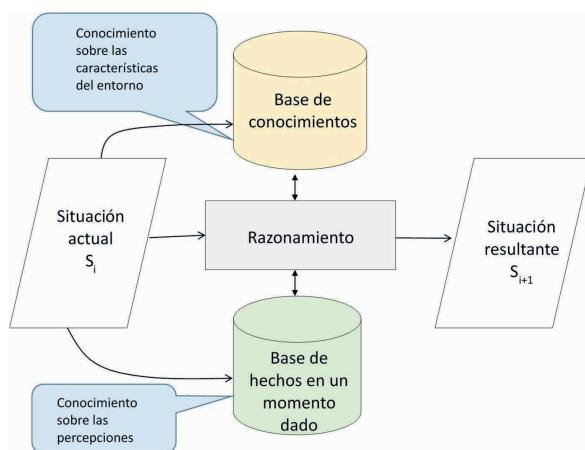


Figura 5. Razonamiento como proceso de transición entre situaciones.

Los **agentes que razonan** de este modo también se conocen como **agentes basados en conocimiento** –ABC– [8].

Principales tipos de razonamiento

• Razonamiento **deductivo**.

Proceso a través del cual se pretende alcanzar conocimientos específicos sobre alguna entidad a partir de conocimientos generales sobre la misma y sobre un entorno al cual dicha entidad pertenece. Un ejemplo ilustrativo es el *razonamiento lógico*.

• Razonamiento **inductivo**, también conocido como **adquisición inductiva de conocimiento**.

Proceso basado en obtener conclusiones o representaciones generales/abstractas –reglas, árboles de decisión, grafos, etc...– sobre alguna entidad a partir de representaciones específicas –partículas o elementos concretos de conocimiento– sobre dicha entidad y sobre el entorno al cual dicha entidad pertenece. Uno de los tipos de razonamiento inductivo más estudiados es el **razonamiento a partir de ejemplos clasificados**.

• Razonamiento **inducto-deductivo**.

Proceso basado en obtener conclusiones o representaciones generales –reglas, árboles de decisión, grafos, etc...– sobre alguna entidad a partir de representaciones generales y específicas –esto es, partículas o elementos concretos de conocimiento y a su vez representaciones abstractas– de dicha entidad y sobre el entorno al cual dicha entidad pertenece. Es el tipo de razonamiento más complejo, que sirve de soporte a numerosos sistemas de **aprendizaje**. En [17] se ofrece un extenso estudio sobre dicho razonamiento.

APRENDIZAJE

Proceso mediante el cual un sistema mejora –perfecciona– su función con el tiempo. El aprendizaje de un agente es la **capacidad del mismo para converger, a lo largo del tiempo, hacia sus objetivos**.

Para que un agente pueda **aprender** con el tiempo, debe ser capaz de **adquirir las correspondientes percepciones** del entorno,

razonar sobre el entorno en todo momento y fabricar sus propias experiencias y/o posiblemente adquirir las de algún usuario experto externo, de manera que sea capaz de extraer las pertinentes consecuencias que le permitirán converger hacia los objetivos propuestos.

Por tanto, el aprendizaje de un agente se basa necesariamente en las **experiencias**

resultantes de su toma de decisiones y en las continuas **observaciones del entorno**.

La forma de proceder de un agente que aprende, no debería tener limitaciones temporales, lo que significa que se trata de un proceso permanentemente abierto a una mejora de la función del agente. En cuanto a la explotación del conocimiento, cabe decir que el modelo que se propone se basa en un razonamiento inducto-deductivo, pues actúa a la vez sobre estructuras de conocimiento específicas y estructuras abstractas, con objeto de «mejorar» la hipótesis del agente, la cual se formaliza a través de estructuras de conocimiento abstractas.

Los agentes que aprenden de forma autónoma constituyen el máximo exponente dentro de la categoría de los agentes inteligentes. Tales agentes se caracterizan por su **capacidad de mejorar su función** –representada a través de lo que se conoce como **hipótesis**- a lo largo del tiempo, a través de la **experiencia acumulada y de la observación de nuevos elementos de conocimiento del entorno**.

Así, el **aprendizaje** se concibe como un proceso de **actualización** de una **hipótesis** de acuerdo con **nuevas evidencias** acerca del entorno y las **experiencias** acumuladas, de manera que la **hipótesis actualizada** sea la que mejor se ajuste a las experiencias y a las nuevas evidencias.

Existen diferentes tipos de aprendizaje, entre los que destacan, entre otros, el aprendizaje por rutina, el aprendizaje a partir de instrucciones, el aprendizaje por analogía y el

aprendizaje inductivo. Este último es uno de los más estudiados y tradicionalmente ha sido categorizado principalmente en aprendizaje a partir de ejemplos y aprendizaje a partir de observaciones. Otros autores, no obstante, realizan una taxonomía del mismo en aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semi supervisado y aprendizaje por refuerzo.

Esta sección se centra en el **aprendizaje** a partir de **ejemplos clasificados**, o simplemente **aprendizaje a partir de ejemplos**. La figura 6 ilustra un modelo simple y general de un para un tal sistema de aprendizaje [16], [17].

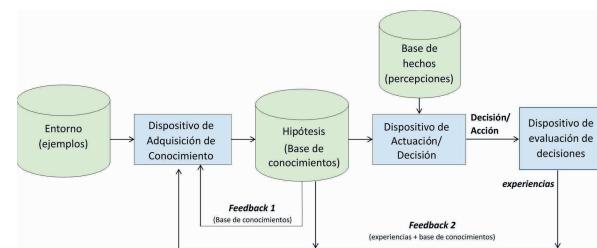


Figura 6. Modelo sencillo de un sistema de aprendizaje.

- **Entorno:** contiene las percepciones del agente, expresadas como ejemplos del ambiente. Éstos constituyen **porciones específicas/concretas de conocimiento** sobre algún concepto. Tales conocimientos vienen descritos extensionalmente.
- **Dispositivo de Adquisición de Conocimiento -DAC-:** es un **dispositivo procedural** basada en un **razonamiento inducto-deductivo**. Se encarga de adquirir los conocimientos relevantes provenientes del Entorno –constituido por conocimiento específico- y del *Feedback* –integrado por conocimiento abstracto y/o específico-. El **conocimiento adquirido** es abstracto –general- y constituye la **hipótesis o base de conocimientos actual**.

El procedimiento de razonamiento del DAC se basa en un **criterio específico**.

- **Hipótesis o Base de conocimientos** –**Hipótesis/BC**-: constituye lo que el agente conoce acerca del ambiente – descripción del ambiente-. La hipótesis se utiliza para la toma de decisiones/acciones del agente. Tiene un formato **abstracto – intensional**, lo cual facilita el razonamiento deductivo del agente en el logro de sus objetivos. Se va actualizando a lo largo del tiempo.
- **Base de hechos**: almacena los **hechos** del entorno sobre los que el agente debe tomar las decisiones/acciones. Tales decisiones/acciones las adopta el **Dispositivo de Actuación/Decisión** a partir de la **Hipótesis/BC**.
- Estos hechos constituyen percepciones del agente y están constituidos por
- **conocimiento específico** –extensional- descrito mediante ejemplos.
- **Dispositivo de Actuación/Decisión – DAD**-: a partir de las percepciones de la **Base de hechos** y de lo que el agente conoce acerca del ambiente – **Hipótesis/B**-, este dispositivo se encarga de razonar de forma deductiva con objeto de tomar alguna acción/decisión sobre cada uno de los ejemplos de la Base de hechos. Se trata de un **dispositivo procedural**.
- En otras palabras, este dispositivo **confronta la Hipótesis/BC con la Base de hechos del Entorno**, con objeto de decidir sobre los ejemplos de la propia Base de hechos.
- **Dispositivo de evaluación de decisiones**: dispositivo encargado de **verificar** las decisiones/acciones adoptadas por el **Dispositivo de Actuación/Decisión**. Los resultados constituirán las «consecuencias» o «experiencias» –esto es, el éxito/fracaso, entre posiblemente otros

aspectos...- del agente en la toma de decisiones/acciones. Se trata de un **dispositivo procedural**. Las **experiencias** pueden estar formadas por **conocimiento específico y/o abstracto**.

Las **experiencias** obtenidas pueden tipificarse de diferentes formas. A continuación, se propone una clasificación según la fuente original donde se generan:

- a. **Experiencias guiadas por el usuario/experto del sistema**. Estas experiencias no permiten –obviamente- una total autonomía del sistema de aprendizaje, puesto que el usuario interviene en el proceso de aprendizaje, proporcionando experiencias al sistema con objeto de guiar su progreso. Aspectos como la creatividad, la imaginación, la intuición entre otros..., todavía no han sido lo suficientemente investigados en IA, por lo que resulta imprescindible la guía del usuario experto. No obstante, en determinados problemas no resulta difícil establecer ciertos umbrales en algunas de las características asociadas a la *Hipótesis/BC* que permitan identificar cuándo deben originarse nuevas experiencias.
- b. **Experiencias proporcionadas por el propio sistema de aprendizaje** a través del **dispositivo de evaluación de decisiones**. Tales experiencias permiten al sistema de aprendizaje una total autonomía del proceso de mejora de su función.

Actualmente resulta todavía muy complejo la implantación de sistemas de aprendizaje totalmente autónomos, por lo cual se apostará por sistemas híbridos en este sentido.

- **Feedback**: retroalimentación basada en las **experiencias** y en la **Hipótesis/BC**, cuya función es la de suministrar conocimientos al **DAC** con objeto de actualizar/mejorar, junto con el conocimiento de las percepciones del **Entorno**, la propia **Hipótesis/BC**. Por

motivos de claridad en la discusión, obsérvese como en la figura 6 aparecen, por cuestiones de claridad, dos tipos de Feedback, aunque ambos se consideran de forma unificada:

- El **Feedback 1**, el cual se interpreta como la evolución de la Hipótesis/BC únicamente a partir del conocimiento de la propia Hipótesis/BC actual, sin la consideración de experiencias. Este tipo de retroalimentación se produce cuando en un momento dado aparecen nuevas percepciones desde el Entorno sin que se hayan producido experiencias. Se trata de una situación perfectamente asumible dado el carácter evolutivo del modelo de aprendizaje. **En tal caso será suficiente enriquecer la Hipótesis/BC a la partir de las percepciones del Entorno y de la propia Hipótesis/BC.**

- El **Feedback 2 o Feedback general**, que considera ambos, la Hipótesis/BC actual y las experiencias.

MODELO GENERAL DE UN SISTEMA DE APRENDIZAJE AUTÓNOMO A PARTIR DE EJEMPLOS CON CONCIENCIA DE OBJETIVOS-

La cuestión que se plantea en este punto es **«hasta qué punto el modelo de aprendizaje descrito en la subsección anterior constituye un modelo con conciencia de objetivos».**

La respuesta la encontramos en la sección 3: **«un hipotético sistema de IA fundamentado en el modelo de conciencia de objetivos, debe ser capaz de: Entender los objetivos, Alinearse con los objetivos y Evaluar el progreso».**

El modelo de la figura 6 es muy general y no resulta lo suficientemente preciso para poder asegurar que se basa de forma rigurosa en una conciencia de objetivos.

La Figura 7 ilustra una adaptación del modelo de aprendizaje a partir de ejemplos de la figura 6 a una conciencia de objetivos –véase la figura 4-.

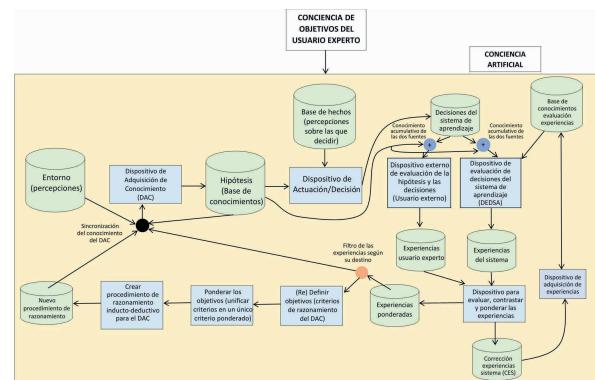


Figura 7. Modelo de un sistema de aprendizaje a partir de ejemplos con conciencia de objetivos.

Del modelo de la figura 7, se desprenden las siguientes observaciones:

- i. El **Entorno** proporciona las percepciones en forma de ejemplos al **DAC**.
- ii. El **DAC** constituye el **procedimiento de razonamiento inducto-deductivo del sistema**, de hecho es la principal fuente de inteligencia del sistema. **Procesa el conocimiento convenientemente sincronizado del Entorno**, de la **Hipótesis** –o **Hipótesis/BC**-, de las **Experiencias ponderadas** y, si es el caso, del dispositivo **Nuevo procedimiento de razonamiento** –este dispositivo contiene un nuevo procedimiento de razonamiento que substituirá al procedimiento de razonamiento del **DAC**-.
- iii. La **Hipótesis** o **Hipótesis/BC** contiene una descripción abstracta del proceso de toma de decisiones del sistema –por ejemplo, un conjunto de reglas de decisión, un árbol de decisión,...-. Es el elemento de conocimiento más importante del modelo, pues la **Hipótesis/BC** constituye el objetivo central del sistema de aprendizaje.
- iv. El dispositivo de **Actuación/Decisión** –abreviado **DAD**- se encarga de confrontar/ejecutar la **Hipótesis/BC** tomando

como datos de entrada las percepciones de la **Base de hechos** –abreviado BH-. Esta tarea de confrontación/ejecución consiste en **clasificar** los datos de la BH de acuerdo con la **Hipótesis/BC**. Los resultados de la clasificación se guardan en el dispositivo **Decisiones del sistema de aprendizaje** –abreviado DSA-.

v. La **Base de hechos** o BH está constituida por **percepciones del entorno en forma de ejemplos no clasificados**. El objetivo es la clasificación de dichos datos por medio de la **Hipótesis/BC**.

vi. **Decisiones del sistema de aprendizaje** o DSA es el dispositivo donde se almacenan los resultados del DAD. La toma de decisiones es completamente automática por parte del sistema y los resultados pueden interpretarse de diferentes maneras de acuerdo con el diseñador del sistema –así, los valores de los resultados de una decisión podrían catalogarse simplemente como correctas/erróneas, como buenas/aceptables/regulares/malas, etc....-

vii. El **Dispositivo externo de evaluación de la hipótesis y las decisiones (Usuario externo)**, evalúa la **Hipótesis/BC** y/o el **DSA** desde el punto de vista de un **usuario experto externo**. Se trata por tanto de una evaluación externa. El usuario externo evalúa ambos, la **Hipótesis/BC** y el **DSA**, no obstante, en algunas ocasiones puede que esté interesado únicamente en la evaluación de una de estas fuentes de conocimiento. Los resultados de la evaluación se almacenan en el dispositivo **Experiencias usuario experto**.

Este dispositivo, al basarse en una evaluación realizada por expertos humanos, se considera como el dispositivo de referencia en la evaluación de la **Hipótesis/BC** y el **DSA**, por tanto, sus **resultados** se con-

sideran «**determinantes**», tanto para la evolución de la **Hipótesis/BC** como para la **mejora/evolución de la capacidad de fabricación de experiencias del sistema**, de manera que su autonomía se vea incrementada con el paso del tiempo.

viii. El dispositivo **Experiencias usuario experto** almacena las experiencias externas producidas por el **Dispositivo externo de evaluación de la hipótesis y las decisiones (Usuario externo)**. Se enfatiza que tales experiencias se consideran determinantes, aunque esto no significa que el **Dispositivo de evaluación de decisiones del sistema de aprendizaje** pueda obtener otras experiencias válidas.

ix. El **Dispositivo de evaluación de decisiones del sistema de aprendizaje** o **DEDSA**, tiene la siguiente función:

Evaluar la **Hipótesis/BC** y/o el **DSA** por parte del sistema. Por tanto, se trata de una evaluación interna y totalmente autónoma, sin intervención de usuarios externos.

Para llevar a cabo el proceso de evaluación de las decisiones, el **DEDSA** se basa en los patrones para la evaluación de experiencias del dispositivo **Base de conocimientos evaluación experiencias**. Este dispositivo evoluciona/mejora a lo largo del tiempo, y proporciona el conocimiento necesario para evaluar las decisiones de la **Hipótesis/BC** y el **DSA**.

Los resultados de la evaluación se almacenan en el dispositivo de **Experiencias del sistema**.

x. El dispositivo **Base de conocimientos evaluación experiencias** constituye una base de conocimientos que contiene los patrones necesarios para evaluar las decisiones de la **Hipótesis/BC** y el **DSA**. Esta base de conocimientos puede ser enri-

quecida en los diferentes ciclos de aprendizaje por el **Dispositivo de adquisición de experiencias**, a través del procesamiento del conocimiento del dispositivo **Corrección experiencias sistema (CES)**.

xi. El **Dispositivo para evaluar, contrastar y ponderar las experiencias** tiene dos funciones:

1. Por una parte, contrasta las experiencias externas –**Experiencias usuario experto**- con las experiencias internas –**Experiencias del sistema**-, valorando ambos tipos de experiencias e «**identificando los errores**» que pueda haber cometido el **Dispositivo de evaluación de decisiones del sistema de aprendizaje (DEDSA)**. Además de detectar estos errores, este dispositivo también considera aquellas experiencias no detectadas por el DEDSA, pero sí por el **Dispositivo externo de evaluación de la hipótesis y las decisiones (Usuario externo)**. Los resultados se almacenan en el dispositivo **Corrección experiencias sistema**, con objeto de actualizar y enriquecer la base de experiencias del sistema – dispositivo **Base de conocimientos evaluación experiencias (CES)**–.
2. Por otra parte, una vez contrastadas las experiencias externas y e internas y libres ya de errores las experiencias internas del sistema de aprendizaje, el dispositivo las procesa y pondera, con objeto de obtener un conjunto final de experiencias equilibrado. Los resultados se almacenan en el dispositivo **Experiencias ponderadas**. Cabe enfatizar que el **Dispositivo de evaluación de decisiones del sistema de aprendizaje (DEDSA)** puede obtener experiencias válidas que el usuario externo no ha generado. Tales experiencias debidamente ponderadas,

formarán parte del dispositivo **Experiencias ponderadas**.

xii. El dispositivo **Corrección experiencias sistema (CES)**, almacena, por una parte, los «**errores**» producidos por el dispositivo **DEDSA** durante la generación de sus experiencias, las cuales se encuentran almacenadas en el dispositivo **Experiencias del sistema**. Por otra parte, el **CES** almacena las experiencias detectadas por el **Dispositivo externo de evaluación de la hipótesis y las decisiones (Usuario externo)**, las cuales no han sido detectadas por el dispositivo **DEDSA**. El conocimiento del **CES** se utiliza como fuente del **Dispositivo de adquisición de experiencias** para que éste actualice y enriquezca la base de conocimientos de las experiencias –**Base de conocimientos evaluación experiencias**–.

xiii. El **Dispositivo de adquisición de experiencias –DAE-** procesa los resultados del dispositivo **Corrección experiencias sistema (CES)** con objeto de actualizar/enriquecer los patrones de las experiencias del dispositivo **Base de conocimientos evaluación experiencias**. Así, puede afirmarse que a través del **DAE**, el sistema de aprendizaje mejora y aprende nuevas experiencias.

xiv. El dispositivo **Experiencias ponderadas** almacena el conjunto final de experiencias. Tales experiencias serán filtradas a través del **Filtro de las experiencias según su destino**. Así, posiblemente **algunas afectarán únicamente al DAC - esto es, lo retroalimentarán directamente-** mientras que las **experiencias que afecten a los objetivos del sistema** se suministrarán/retroalimentarán al dispositivo **(Re)Definir objetivos (criterios de razonamiento del DAC)**, con objeto de adaptar su proceso de razonamiento inducto-deductivo a los nuevos objetivos.

xv. (Re) Definir objetivos (**criterios de razonamiento del DAC**) es un dispositivo que procesa las **Experiencias ponderadas** que afectan a la redefinición de los objetivos. De acuerdo con el tipo de aprendizaje considerado –**aprendizaje autónomo a partir de ejemplos**–, los **objetivos** son los que guían el **proceso de razonamiento inducto-deductivo** del **DAC**, lo que significa que tales objetivos se traducen en los **criterios** que debe seguir dicho proceso de razonamiento.

xvi. Una vez se han (re)definido los objetivos, esto es, se han establecido los criterios en los que debe basarse el **DAC**, el siguiente paso consiste en la ponderación de dichos criterios con objeto de unificarlos en un único criterio. El dispositivo encargado de llevar a cabo la mencionada tarea se conoce como **Ponderar los objetivos (unificar criterios en un único criterio ponderado)**.

xvii. A partir de la unificación de los criterios en un único criterio ponderado – punto xvi.–, el siguiente paso es la creación de un procedimiento de razonamiento inducto-deductivo que opere según las directrices del criterio ponderado. Este procedimiento **substituirá al procedimiento implementado actualmente en el DAC**. El dispositivo encargado de llevar a cabo este paso se conoce como **Crear procedimiento de razonamiento inducto-deductivo para el DAC**.

xviii. El dispositivo **Crear procedimiento de razonamiento inducto-deductivo para el DAC** se encarga de crear/programar un procedimiento de razonamiento que **substituya** al procedimiento que en este momento está siendo utilizado por el **DAC**. Este nuevo procedimiento se fundamenta en las experiencias recibidas desde el dispositivo **Experiencias pon-**

deradas. La creación automática de programas para la IA es una tarea compleja, todavía incipiente. Por ello, dicha tarea deberá realizarse seguramente por parte de un programador humano experto, o quizás, en una primera aproximación, manteniendo una base de conocimientos lo suficientemente amplia, con diferentes procedimientos de razonamiento susceptibles de substituir en procedimiento actual del **DAC**.

Los resultados se almacenan en el dispositivo **Nuevo procedimiento de razonamiento –NPR–**.

xix. El dispositivo **Nuevo procedimiento de razonamiento –NPR–** almacena el nuevo procedimiento de razonamiento inducto-deductivo fabricado por el dispositivo **Crear procedimiento de razonamiento inducto-deductivo para el DAC**, de acuerdo con las indicaciones de las **experiencias ponderadas**. Este procedimiento substituirá al procedimiento actual del **DAC**.

xx. El dispositivo **Sincronización del conocimiento del DAC** se encarga de sincronizar todas las entradas de conocimiento del **DAC**.

Este modelo garantiza todas las propiedades de una conciencia de objetivos artificial para los sistemas de aprendizaje autónomo a partir de ejemplos clasificados.

PREGUNTAS ESCLARECEDORAS PARA EL LECTOR

Las características del trabajo descrito seguramente estimularán el planteamiento de determinadas preguntas por parte de algunos lectores. En esta sección se describen algunas cuestiones que ayudarán a clarificar la interpretación del autor al respecto.

1. ¿Qué es lo que realmente aprende un sistema de aprendizaje a partir de ejemplos basado en el modelo ilustrado en la figura 7?

El modelo de la figura 7 permite a un agente aprender:

a. A **mejorar la Hipótesis/BC**, objetivo central del agente.

b. A **mejorar la creación de experiencias**, lo cual se lleva a cabo confrontando las experiencias producidas por el usuario externo y las producidas por el propio sistema. Este aspecto resulta esencial para que el agente incremente, a lo largo del tiempo, tanto su autonomía como su evolución hacia los objetivos.

2. ¿Es totalmente autónomo el modelo de sistema de aprendizaje a partir de ejemplos ilustrado en la figura 7?

El modelo está concebido para ser totalmente autónomo, por lo que, desde el punto de vista teórico podemos afirmar que se trata de un modelo autónomo. No obstante, cuando se trata de implantar en la práctica dicho modelo, actualmente aún no se dispone de soluciones efectivas para la programación de algunos dispositivos, tal como se comenta en la sección 5.2. Para conseguir la funcionalidad de dichos dispositivos resulta necesario la intervención del usuario experto.

3. El sistema de aprendizaje de la figura 7, ¿dispone realmente una conciencia de objetivos?

Lo primero y más importante que hay aclarar, es que la **conciencia artificial de objetivos**, tal y como se ha abordado en este trabajo, constituye una “versión artificial” de la **conciencia de objetivos del usuario experto**, que es quien proporciona al diseñador/programador, todo el **conocimiento, procedimientos y sincronizaciones** necesarios del modelo, con objeto de garantizar el progreso del proceso de aprendizaje y su alineamiento con los ob-

jetivos.

A lo dicho hay que añadir que algunos aspectos relacionados con la conciencia humana, como la creatividad, la intuición, etc..., todavía no disponen de una traducción práctica apropiada, por lo que tales aspectos son aportados por el usuario externo experto, en forma de **elementos de conocimiento o procesos funcionales**.

4. ¿Cuáles son las aportaciones de la conciencia de objetivos en el marco de los sistemas de aprendizaje a partir de ejemplos -SAAE-?

La conciencia de objetivos proporciona un marco completo de diseño para los SAAE, cuyos objetivos conducen hacia la total autonomía del agente. Se trata solamente de un primer paso hacia un desarrollo práctico de SAAEs. La investigación en los diferentes componentes del modelo, juntamente con los avances hardware en el diseño de los computadores, conducirán hacia SAAEs cada vez más evolucionados y eficientes.

5. ¿Por qué en el modelo de aprendizaje de la figura 7, se separan las experiencias propias del usuario experto externo y las experiencias propias del sistema de aprendizaje?

La separación de dichos tipos de experiencias se realiza por dos razones:

- Para resaltar que hoy en día, todavía no existe la posibilidad de crear un agente del tipo considerado con una autonomía total, que no precise en absoluto la intervención de usuarios humanos expertos.

- Por otra parte, para destacar que la progresión del aprendizaje de la Hipótesis/BC por parte del sistema, resulta imposible sin la progresión de las experiencias. En otras palabras, para llevar a cabo la función -La Hipótesis/BC- de un «auténtico aprendizaje» el agente debe ser capaz de progresar en el desarrollo

de sus experiencias. Esto es, la experiencia también se aprende con el curso de los acontecimientos; las que en un momento parecen buenas, quizás deberán corregirse en el futuro. El aprendizaje de experiencias constituye actualmente uno de los grandes desafíos del aprendizaje autónomo.

6. ¿Cuál es el formato en el que se describen las experiencias?

Existen diferentes tipos de experiencias, algunas de las cuales pueden expresarse en formato extensivo –por ejemplo, cuando el SAAE clasifica erróneamente mediante la Hipótesis/BC, algún hecho de la BH, cuya experiencia puede representarse mediante dicho hecho-. El formato de dichas experiencias es del mismo tipo que el de las percepciones del Entorno, por lo que el DAC interpretará dichas experiencias como si fueran percepciones del Entorno. Por otra parte, existen experiencias con formato intensivo/abstracto, particularmente entre las expresadas por el usuario. En tales casos, resulta importante crear un lenguaje de experiencias, que permita representarlas y procesarlas, así como crear nuevas experiencias. En [17] se describe una aproximación a la representación de diferentes tipos de experiencias junto con algunos ejemplos prácticos detallados.

7. ¿Cuál es el dispositivo más importante del SAAE de la figura 7?

Todos los dispositivos, tanto de conocimiento como procedimentales, resultan esenciales para que el SAAE lleve a cabo correctamente su función. La ausencia de cualquiera de los dispositivos supondría la carencia de una adecuada conciencia de objetivos.

No obstante, si hubiera que destacar algún dispositivo en particular, este sería el **DAC**. Recuérdese que el DAC implanta un proceso de razonamiento inducto-deductivo a partir de ejemplos con objeto de generar/actualizar la **Hipótesis/BC**, que es la que sustenta la des-

cripción del Entorno según los objetivos propuestos. Esta hipótesis **constituye realmente los objetivos del SAAE**.

8. ¿Cuál es la forma de proceder general del SAAE de la figura 7?

El **SAAE** descrito **funciona por ciclos**, siendo el **DAC** el dispositivo central del mismo. En un ciclo dado, el **DAC** primeramente mira si hay necesidad de cambiar su **procedimiento de razonamiento** por el procedimiento del dispositivo **Nuevo procedimiento de razonamiento**.

Una vez realizada dicha acción, entonces el **DAC**, ya con el procedimiento de razonamiento actualizado, si éste ha sido el caso, procede a procesar las percepciones del Entorno y las experiencias. Si no hay percepciones del Entorno ni experiencias, entonces el **DAC** no hace nada. En cambio, si hay percepciones y/o experiencias, entonces el **DAC** las procesa, generando una **nueva Hipótesis/BC**, también conocida como la **Hipótesis/BC** del ciclo actual.

Una vez actualizada la **Hipótesis/BC**, el **Dispositivo de Actuación/Decisión –DAD–** procede a clasificar las percepciones de la **BH** –si las hay–, cuyos resultados se almacenan en el dispositivo **Decisiones del sistema de aprendizaje**. En caso de que la **BH** no contenga ninguna percepción sobre la que decidir/clasificar, entonces no se producirán decisiones y el dispositivo **Decisiones del sistema de aprendizaje** permanecerá vacío. A continuación, el **SAAE** proporciona a los dispositivos **Dispositivo externo de evaluación de la hipótesis y las decisiones (Usuario externo)** y **DEDSA**, la **Hipótesis/BC** del ciclo actual y las **decisiones** del dispositivo **Decisiones del sistema de aprendizaje** para que sean evaluadas y se emitan las correspondientes experiencias en los dispositivos **Experiencias usuario experto** y **Experiencias del sistema** respectivamente.

A continuación, las experiencias de los dispositivos **Experiencias usuario experto** y **Experiencias del sistema**, son procesadas por el **Dispositivo para evaluar, contrastar y ponderar las experiencias**, el cual tiene dos funciones, tal como se describe en el punto xi. de la subsección 5.2. Los resultados se almacenan en los dispositivos **Corrección experiencias sistema (CES)** y **Experiencias ponderadas**, cuyo significado se describe en los puntos xii. y xiv. respectivamente, de la subsección 5.2.

Por lo que respecta a las experiencias del dispositivo **Corrección experiencias sistema (CES)**, éstas constituyen el conocimiento de entrada del **Dispositivo de adquisición de experiencias**, el cual las procesa, con el objetivo de actualizar la base de experiencias del agente –**Base de conocimientos evaluación experiencias**.

En relación a las experiencias del dispositivo **Experiencias ponderadas**, éstas constituyen el conjunto final de experiencias que retroalimentarán el **DAC**. Según el tipo de dichas experiencias, el **Filtro de las experiencias según su destino**, enviará algunas de ellas directamente al **DAC**, mientras que otras se enviarán al dispositivo **(Re) Definir objetivos (criterios de razonamiento del DAC)**. Este último dispositivo se encarga de (re)definir los objetivos del nuevo procedimiento de razonamiento del **DAC** que operará en el siguiente ciclo.

Las acciones en secuencia, tal como se ilustra en la figura 7, de los dispositivos **(Re) Definir objetivos (criterios de razonamiento del DAC)**, **Ponderar los objetivos (unificar criterios en un único criterio ponderado)** y **Crear procedimiento de razonamiento inducto-deductivo para el DAC**, darán lugar al nuevo algoritmo/programa del siguiente ciclo, que substituirá al procedimiento del **DAC** del ciclo actual. Este nuevo algoritmo/programa se almacenará en el dispositivo **Nuevo procedimiento de razonamiento**.

Finalmente, dicho procedimiento se enviará al **DAC**, con objeto de ser ejecutado en el siguiente ciclo. Obsérvese en la figura 7, como **todas las entradas de conocimiento** del **DAC** deben ser correctamente **sincronizadas** por el dispositivo **Sincronización del conocimiento del DAC**.

CONCLUSIONES

El aprendizaje autónomo a partir de ejemplos constituye uno de los campos de la IA más estudiados en la actualidad. Sus aplicaciones prácticas constituyen una demostración del éxito en el avance del campo. No obstante, el significado que se otorga al concepto de «aprendizaje» en numerosos contextos y aplicaciones, resulta limitado en relación a su magnitud real. De hecho, en numerosas obras y aplicaciones se confunden los términos de aprendizaje y razonamiento, atribuyendo al aprendizaje las cualidades de adquisición de conocimiento cualitativo propios del razonamiento, cuando en realidad va mucho más allá.

Este trabajo presenta una visión general del significado del **concepto de aprendizaje** en IA, la cual permite su exploración en una mayor magnitud. Para ello se exploró inicialmente el área de la **conciencia humana**, cuyos alcances son inmensos, donde confluyen numerosas disciplinas, desde la filosofía hasta la neurociencia cognitiva y la neuropsicología. La complejidad del área es tal que los avances resultan muy lentos y la falta de consensos es común. Finalmente se decidió que la conocida como **conciencia de objetivos** resulta, en estos momentos, muy apropiada para abordar numerosos problemas complejos relacionados con el aprendizaje en IA, lo cual conlleva abrir nuevos campos de investigación, para que las soluciones constituyan realmente el producto de un proceso de aprendizaje de mayor magnitud. La **conciencia de objetivos** se refiere a la **capacidad** de un sistema de inteligencia

artificial para comprender y alinearse con los objetivos y metas de los usuarios.

El trabajo describe un **modelo general de conciencia de objetivos**, centrándose en tres **propiedades esenciales**: La **definición y comprensión claras de los objetivos**, la **alineación con los objetivos del usuario experto** y la **evaluación del progreso y su adaptación hacia los objetivos**. A partir de ahí, se ha adaptado un modelo de aprendizaje a partir de ejemplos clasificados al modelo de la conciencia de objetivos. El resultado muestra un modelo mucho más elaborado, completo y definido, que incluye los aspectos fundamentales de una conciencia de objetivos para este tipo de aprendizaje. Entre las propiedades de este modelo destacan:

El sistema de aprendizaje **trabaja en ciclos**, mejorando su hipótesis en cada uno de los ciclos. La mejora de la hipótesis va en dirección a los objetivos del sistema.

El dispositivo central del sistema lo constituye un procedimiento de razonamiento inducto-deductivo, conocido como **DAC**, que

procesa todos los elementos de conocimiento necesarios para adaptar la hipótesis del sistema, con objeto de **alinear sus objetivos** con los del usuario experto externo.

El sistema considera las experiencias que se derivan de cada ciclo de aprendizaje. Las experiencias constituyen un elemento de conocimiento fundamental para la **adaptación del progreso** del sistema de aprendizaje.

Capacidad para **aprender a crear experiencias**. El modelo no sólo ofrece la capacidad de mejorar la hipótesis del sistema, también brinda la capacidad de aprender a crear experiencias válidas, lo cual resulta fundamental para la alineación de los objetivos, el progreso de la hipótesis y la capacidad de incrementar la autonomía del sistema. De hecho, sin experiencias válidas no sería posible un auténtico aprendizaje.

Capacidad de **(re)definir los objetivos** en cada ciclo del sistema de aprendizaje.

La (re)definición de los objetivos puede acarrear la (re)programación del sistema de razonamiento inducto-deductivo.

REFERENCIAS

1. <https://iceebook.com/conciencia-que-es-caracteristicas-teorias-filosoficas>.
2. <https://iccsi.com.ar/conciencia-e-inteligencia/>
3. [https://iccsi.com.ar/diferencia-entre-conciencia-e-inteligencia /](https://iccsi.com.ar/diferencia-entre-conciencia-e-inteligencia/)
4. Chalmers, D. J. (1995). Facing up to the hard problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200-219. <https://www.ingentaconnect.com/contentone/imp/jcs/1995/00000002/00000003/art00003>.
5. Graziano, M. S. A. (2013). Consciousness and the social brain. Oxford University Press. <https://global.oup.com/academic/product/consciousness-and-the-social-brain-9780199915255>.
6. Joëlle Proust in *Neural Correlates of Consciousness*, Thomas Metzinger, 2000, MIT, pages 307-324.
7. <https://academia-lab.com/enciclopedia/experiencia/>
8. Russell, Stuart; Norvig, Peter (2010). *Artificial Intelligence: A Modern Approach* (3rd edición). Pearson series in Artificial Intelligence. ISBN: 978-0136042594.
9. Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning* (2006). Ed. Springer-Verlag New York Inc. ISBN: 978-0387310732.

10. Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016). *Deep Learning*. The MIT Press. ISBN: 978-0262035613.
11. Quinlan, J. R.. *Induction of Decision Trees*. Machine Learning, 1:81-106, 1986.
12. Fiol-Roig, Gabriel. *Contribución a la adquisición inductiva de conocimiento*. PhD Thesis, Universitat de les Illes Balears, Palma de Mallorca, Spain, 1991
13. Fiol, G.; Miró, J.. A New Perspective in the Inductive Acquisition of Knowledge from Examples. IPMU'92- Advanced Methods in Artificial Intelligence. Lecture Notes in Computer Science 682. pp 219-228. 1993.
14. Pawlak, Z.. *Rough Sets*. International Journal on Information and Computer Sciences, 11:341-356, 1982.
15. Michalski, R. S.; Larson, J. B.. Selection of most representative training examples and incremental generation of VL1 hypotheses: The underlying methodology and the description of programs ESEL and AQ11. Technical Report 867, Computer Science Department. University of Illinois at Urbana-Champaign, Urbana, IL, 1978.
16. Fiol-Roig, Gabriel. *UIB-AILS: A Model for Autonomous Inductive Learning Systems*. Journal of Engineering Research (ISSN 2764-1317). DOI: <https://doi.org/10.22533/at.ed.3174232418092>. Ed. Atena Editora (2024).
17. Fiol Roig, Gabriel (2025). *Sistema inteligente autónomo para el aprendizaje incremental de árboles de decisión*. Ed. Atena Editora. ISBN: 978-65-258-3009-4. DOI: <https://doi.org/10.22533/at.ed.094250406>.