



Interacciones entre humanos y arquitecturas cognitivas

Alberto León Luengo
Toni Marí Marí
Jorge Martín Mínguez
Luis Moreno García

ÍNDICE

1. INTRODUCCIÓN

2. REVISIÓN DE LA LITERATURA RELEVANTE

2.1 DEFINICIÓN Y TIPOS DE ARQUITECTURAS COGNITIVAS

2.2 INTERACCIÓN HUMANO-ARQUITECTURA COGNITIVA

2.3 IMPACTO EN LA SOCIEDAD Y LA INDUSTRIA

3. ANÁLISIS DETALLADO DEL TEMA

3.1 MODELOS DE PROCESAMIENTO DE INFORMACIÓN

3.2 CASOS DE ESTUDIO

3.3 DESAFÍOS ACTUALES

4. CONCLUSIONES Y REFLEXIONES

5. BIBLIOGRAFÍA

[] 1. INTRODUCCIÓN

La convergencia entre la cognición humana y los sistemas artificiales ha dado lugar a un campo de estudio que a día de hoy se considera revolucionario en su ámbito: **Las interacciones entre humanos y arquitecturas cognitivas**.

Este tipo de arquitectura puede definirse como un modelo computacional que emula procesos mentales como el aprendizaje, la memoria y la toma de decisiones. Sin embargo, las arquitecturas cognitivas no sólo amplían el conocimiento que los seres humanos poseemos sobre la inteligencia humana, sino que también redefinen cómo éstos se relacionan con la tecnología.

Gracias al uso de herramientas como asistentes virtuales y robots sociales en este ámbito, se está ayudando a agilizar el desarrollo y la transformación de sectores como el educativo, el sanitario y el industrial, aunque ninguno de ellos queda exento de que, en un futuro, se le puedan presentar nuevos desafíos tanto de carácter técnico como ético.

En este trabajo, se abordará este tema desde cuatro perspectivas clave:

1. Revisión de la literatura relevante

- Se definirá qué son las arquitecturas cognitivas y su base teórica desde el punto de vista de la psicología y la Inteligencia Artificial.
- Se examinarán estudios pioneros que han marcado el desarrollo de este campo en la actualidad.

2. Análisis detallado de interacciones

- Se explicarán detalladamente algunos casos prácticos que se tienen en la actualidad, por ejemplo, tutores inteligentes en el sistema educativo, diagnósticos médicos asistidos por la Inteligencia Artificial y algunos de los usos más comunes en el ámbito de la robótica social.
- Se realizará una amplia comparación entre modelos simbólicos, en los que se encuentran diversos modelos como SOAR (el cual se explicará más en detalle), y los modelos conexionistas, en los que se encuentran las redes neuronales.

3. Desafíos actuales

- Se mencionarán algunas limitaciones técnicas que podrían presentarse mientras está teniendo lugar una interacción entre humanos y arquitecturas cognitivas, como por ejemplo, la falta de adaptabilidad de una arquitectura cognitiva cuando se encuentra en un entorno en el que no puede predecir comportamientos futuros.
- Se explicarán algunos de los dilemas éticos más importantes que se tienen en la actualidad, como pueden ser sesgos algorítmicos, dudas en el nivel de privacidad al que están limitadas estas arquitecturas y el nivel de autonomía que poseen en la toma de decisiones.

4. Conclusiones

- Y por último, se realizará una pequeña síntesis en los hallazgos logrados en los últimos años y se mencionarán algunas de las propuestas más relevantes que podrían ser utilizadas en investigaciones que puedan llegar a realizarse en un futuro próximo.

TEORÍA ➡ APLICACIONES ➡ PROBLEMAS ➡ SOLUCIONES FUTURAS

[] 2. REVISIÓN DE LA LITERATURA RELEVANTE

} 2.1 DEFINICIÓN Y TIPOS DE ARQUITECTURAS COGNITIVAS

En primer lugar, vamos a definir qué es una **arquitectura cognitiva**. Como definición general, una arquitectura cognitiva hace referencia tanto a una teoría sobre la estructura de la mente humana como a una instancia computacional de dicha teoría utilizada en los campos de la Inteligencia Artificial y la ciencia cognitiva computacional, donde los modelos más formalizados se pueden utilizar para perfeccionar aún más una teoría integral de la cognición o como un programa útil de Inteligencia Artificial.

Es importante mencionar que tanto la investigación sobre arquitecturas cognitivas como la instanciación de elementos software en arquitecturas cognitivas fue iniciada por **Allen Newell** en 1990.

Algunas de las arquitecturas cognitivas más exitosas y conocidas son **SOAR**, **ACT-R** y **CLARION**, las cuales se explicarán más en detalle a continuación.

La arquitectura **SOAR** (Security, Orchestration, Automation and Response), desarrollada por **John Laird** y **Allen Newell**, es una solución de software que permite a los equipos de seguridad integrar y coordinar herramientas de seguridad independientes, automatizar tareas repetitivas y agilizar los flujos de trabajo de respuesta a incidentes y amenazas.

Esta definición también podría aplicarse como definición de la letra **S (Security)** que compone el nombre de esta arquitectura. Sabiendo esto, el resto de letras son las que representan las 3 componentes principales, de las cuales se muestra a continuación una breve explicación de cada una de ellas:

O (Orchestration): Incluye la coordinación y la gestión de herramientas y procesos de seguridad para poder crear un flujo de trabajo integrado (recopilación y orden de datos operativos informáticos de seguridad de red).

A (Automatization): Incluye la automatización de tareas repetitivas y manuales, como pueden ser la gestión de alertas, investigaciones sobre incidentes y respuestas ante posibles amenazas.

R (Response): Posee la capacidad de responder de manera rápida y efectiva a cualquier incidente de seguridad.

El funcionamiento de una arquitectura SOAR puede dividirse en cuatro fases:

- **Fase 1 (Recopilación de Datos)**: Se recopilan datos de diversas fuentes, como pueden ser herramientas de seguridad o sistemas de información.
- **Fase 2 (Análisis y Detección)**: Se utilizan algoritmos y reglas para analizar los datos obtenidos anteriormente y detectar posibles amenazas y vulnerabilidades.
- **Fase 3 (Automatización de la Respuesta)**: Se ejecutan automáticamente acciones de respuesta predefinidas para abordar todos los incidentes detectados (análisis de antivirus o bloques de IPs).
- **Fase 4 (Orquestación de Flujos de Trabajo)**: Se coordinan las interacciones entre diferentes herramientas de seguridad para crear flujos de trabajo automatizados para la gestión de incidentes.

Sabiendo esto, se puede concluir que las principales ventajas de la arquitectura SOAR son la **mejora de la eficiencia** reduciendo la carga de trabajo manual y permitiendo que los equipos de seguridad se concentren en tareas más estratégicas, un **aumento de la velocidad de respuesta** ante posibles incidentes reduciendo así el tiempo de la misma y posibles daños, una **mejora de la seguridad** mediante la automatización de procesos y una respuesta rápida ante amenazas, la **reducción de errores** humanos mediante automatización y una mayor **optimización de los recursos**.

El segundo ejemplo es el modelo **ACT-R** (Adaptive Control of Thought-Rational), propuesto por **John Anderson**, que consiste en la búsqueda de la simulación y la comprensión de la cognición humana, es decir, cómo se organiza el conocimiento y se generan comportamientos inteligentes.

Este modelo está compuesto por los siguientes cuatro elementos:

- **Memoria declarativa:** Almacena conocimientos en forma de fragmentos, los cuales pueden verse como representaciones vectoriales de propiedades individuales.
- **Memoria procedimental:** Contiene las reglas de producción (instrucciones que llevarán a cabo las acciones).
- **Memoria de trabajo:** Espacio de memoria en el que se procesa la información en el que se recibe información procedente de la memoria declarativa y algunas reglas de producción.
- **Módulos:** Estructuras especializadas y relativamente independientes que procesan información específica como la percepción, el lenguaje o la atención.

También es importante mencionar las tres características clave de este modelo: La **simulación**, que permite a los investigadores probar sus hipótesis y comprender mejor la cognición humana; el **aprendizaje**, que permite a los agentes adaptarse y mejorar su desempeño a medida que van adquiriendo nueva información; y la **resolución de problemas** mediante la creación de sub-objetivos y aplicando el conocimiento obtenido en la memoria de trabajo.

En resumen, el modelo ACT-R es una herramienta poderosa para la investigación en ciencias cognitivas y la construcción de sistemas inteligentes, ya que permite modelar y simular la cognición humana y el aprendizaje.

Y por último, tenemos la arquitectura **CLARION** (Connectionist Learning with Adaptive Rule Induction ONLINE), creada para modelar la mente humana integrando tanto conocimiento implícito (inconsciente) como explícito (consciente), la cual se caracteriza por tener un aprendizaje bidireccional y una arquitectura modular.

Esta arquitectura está compuesta por cuatro componentes principales:

- **ACS (Action Subsystem):** Controla acciones usando reglas explícitas y redes neuronales implícitas.
- **NACS (Non-Action Subsystem):** Gestiona el conocimiento en general (memoria semántica y episódica).
- **MS (Motivational Subsystem):** Maneja impulsos y metas que guían el comportamiento.
- **MCS (MetaCognitive Subsystem):** Supervisa y regula los otros subsistemas.

Las principales **ventajas** de esta arquitectura comprenden la explicación de procesos tanto automáticos como conscientes, la adaptación a diferentes tareas y contextos, y su gran utilidad en aplicaciones de psicología cognitiva e Inteligencia Artificial.

En resumen, CLARION ofrece un modelo potente y flexible para estudiar la cognición humana, donde destaca su integración de distintos tipos de conocimiento y su aplicabilidad en múltiples áreas.

} 2.2 INTERACCIÓN HUMANO-ARQUITECTURA COGNITIVA

La forma en que los humanos interactúan con sistemas basados en arquitecturas cognitivas depende en gran medida del tipo de interfaz que se utilice. Estas interacciones varían notablemente según el medio, ya sea texto, robots con forma humana o robots con formas inspiradas en animales u objetos. A continuación, se analizará cómo se configuran estas relaciones en función del tipo de interfaz utilizada.

Entradas de texto:

Las interfaces conversacionales basadas en texto, como los chatbots, representan una de las formas más comunes y accesibles en la interacción humano-IA. Estos sistemas, alimentados por avances recientes en modelos de lenguaje natural, han pasado de ser simples programas de respuesta automática a agentes conversacionales complejos, capaces de mantener diálogos coherentes, adaptarse al tono del usuario e incluso detectar emociones a través del lenguaje.

Según el estudio *How Artificial Intelligence and Chatbots Have Changed Human-Robot Interaction*, este tipo de interacción ha demostrado ser particularmente útil para tareas de asistencia, atención al cliente y apoyo emocional, ya que elimina los posibles sesgos visuales que se generarían ante una forma física. La “desencarnación” del agente puede favorecer la proyección de rasgos humanos por parte del usuario, ya que el texto permite una mayor ambigüedad interpretativa. Además, al carecer de una presencia física, estos sistemas evitan el riesgo de generar rechazo asociado a una apariencia inadecuadamente humana, como ocurre con el valle inquietante.

Robots humanoides:

Los robots con forma humana ofrecen un canal de comunicación más rico en señales sociales, incluyendo lenguaje corporal, expresiones faciales, contacto visual y proximidad física. Esta complejidad multimodal permite establecer interacciones más intuitivas y naturales, lo que es especialmente valioso en entornos como la educación, la salud o el acompañamiento social.

En el estudio *Human-like interactions prompt people to take a robot's perspective* se observó que los usuarios eran más propensos a adoptar la perspectiva cognitiva de un robot cuando este mostraba conductas sociales sutiles como mirar directamente al interlocutor o asentir con la cabeza. Estos gestos, aparentemente simples, generaron una mayor empatía y colaboración, lo cual sugiere que las señales no verbales tienen un peso significativo en la calidad de la interacción.

Sin embargo, este tipo de robots también enfrenta un límite perceptual importante: si su apariencia o comportamiento se acerca demasiado al de un ser humano sin alcanzarlo completamente, puede provocar una sensación de inquietud, rechazo o incomodidad. Este fenómeno, conocido como *uncanny valley* o *valle inquietante* (Mori, 1970), plantea un reto serio para los diseñadores: cuanto más realista sea un robot, mayor será la expectativa del usuario sobre su inteligencia emocional y capacidades sociales. Cuando esas expectativas no se cumplen, el efecto puede ser contraproducente.

Robots no humanoides:

En respuesta a los problemas que plantea el valle inquietante, ha crecido el interés por diseñar robots con formas alternativas, como animales o criaturas estilizadas, que se alejan deliberadamente de la figura humana. Este enfoque busca conservar la expresividad emocional sin generar desconfianza o disonancia.

Un ejemplo paradigmático es *eBear*, un robot con forma de oso diseñado para expresar emociones mediante un sistema híbrido de expresión facial mecánica y visual. Según el estudio *eBear: An Expressive Bear-Like Robot*, este tipo de diseño permite mantener una rica interacción emocional sin desencadenar las reacciones negativas que pueda generar un robot casi humano. Los usuarios reaccionan positivamente ante la expresividad del robot, al percibirlo como “adorable” o “amistoso”, lo que facilita la formación de vínculos afectivos incluso cuando se trata de una máquina.

Estos robots no humanoides no solo son más aceptables en términos psicológicos, sino que también pueden ser percibidos como más accesibles y menos intimidantes. Esto los hace especialmente útiles en entornos educativos con niños, terapias con personas mayores, y situaciones donde se busca una interacción emocional segura pero no invasiva.

} 2.3 IMPACTO EN LA SOCIEDAD Y LA INDUSTRIA

La incorporación de arquitecturas cognitivas y sistemas de Inteligencia Artificial en distintas áreas de la sociedad está transformando tanto los procesos como las relaciones humanas con la tecnología. Estos avances no solo han optimizado tareas repetitivas, sino que han comenzado a intervenir en espacios donde la inteligencia humana parecía insustituible, como la medicina, la educación y la industria empresarial. A continuación, se exploran algunos de los casos más destacados

Medicina: Diagnósticos asistidos por IA

En el campo de la salud, la Inteligencia Artificial ha comenzado a desempeñar un papel crucial en el apoyo al diagnóstico clínico. Los sistemas cognitivos entrenados con grandes volúmenes de datos médicos pueden identificar patrones complejos y detectar anomalías que podrían pasar desapercibidas incluso para profesionales experimentados. Un ejemplo notable es el uso de redes neuronales profundas en la interpretación de imágenes médicas (radiografías, resonancias magnéticas, etc.), donde se ha demostrado que estos sistemas pueden alcanzar niveles de precisión comparables a los de los radiólogos humanos.

Sistemas como **DeepMind Health** o el algoritmo de detección temprana de cáncer de mama desarrollado por Google han mostrado resultados prometedores al reducir tasas de error y agilizar procesos diagnósticos. Además, plataformas como **IBM Watson for Oncology** han sido utilizadas para sugerir planes de tratamiento basados en literatura científica y casos clínicos similares, ofreciendo así una segunda opinión altamente informada. Aunque todavía es necesaria la supervisión médica, la Inteligencia Artificial se ha consolidado como un apoyo potente para mejorar la precisión y eficiencia del diagnóstico y tratamiento.

Educación: Plataformas adaptativas como Cognii

En el ámbito educativo, las arquitecturas cognitivas han dado lugar a sistemas de aprendizaje adaptativo que personalizan el contenido y el ritmo de enseñanza según las necesidades de cada estudiante. Estas plataformas no solo evalúan el rendimiento académico, sino que ajustan dinámicamente las tareas para mantener el nivel de desafío óptimo, mejorando la retención del conocimiento y la motivación del alumno.

Un ejemplo destacado es **Cognii**, una plataforma basada en Inteligencia Artificial que utiliza procesamiento de lenguaje natural para interactuar con estudiantes mediante respuestas abiertas. El sistema ofrece retroalimentación inmediata y fomenta el desarrollo del pensamiento crítico, especialmente en entornos donde el número de docentes por alumno es limitado. Este tipo de tecnología se está aplicando en niveles que van desde la educación básica hasta la formación corporativa, demostrando así su versatilidad.

Además de mejorar la eficiencia del aprendizaje, estas plataformas pueden reducir desigualdades educativas al ofrecer acceso a tutorías personalizadas a bajo costo, lo que representa un paso hacia una educación más inclusiva y accesible.

Empresa: Chatbots y asistentes virtuales (ej. IBM Watson)

En el sector empresarial, los asistentes virtuales y chatbots se han convertido en herramientas clave para mejorar la atención al cliente, automatizar procesos internos y reducir costos operativos. Estos sistemas, muchos de ellos basados en arquitecturas cognitivas como **IBM Watson**, pueden comprender preguntas en lenguaje natural, acceder a bases de datos complejas y ofrecer respuestas relevantes en tiempo real.

Las aplicaciones son múltiples: desde la atención al cliente en banca y seguros hasta soporte técnico, gestión de recursos humanos y análisis de datos. La disponibilidad 24/7, combinada con la capacidad de escalar a miles de interacciones simultáneas, ha hecho de estos sistemas un elemento central en la transformación digital de muchas organizaciones.

Más allá de los usos básicos, algunas empresas están comenzando a integrar chatbots en procesos de toma de decisiones, como análisis de sentimiento del consumidor o predicción de tendencias de mercado, utilizando capacidades de Machine Learning y minería de datos para detectar patrones que serían invisibles para los humanos.

[] 3. ANÁLISIS DETALLADO DEL TEMA

} 3.1 MODELOS DE PROCESAMIENTO DE INFORMACIÓN

En este apartado se irá realizando punto por punto una comparación entre **arquitecturas simbólicas** (SOAR) y **arquitecturas conexionistas** (Redes Neuronales), además de una tabla en la que se mostrará lo explicado a continuación de una forma más visual:

Representación del conocimiento:

Las arquitecturas simbólicas representan el conocimiento de forma explícita mediante símbolos, reglas y estructuras lógicas, mientras que las arquitecturas conexionistas lo hacen de forma implícita y distribuido en pesos de conexiones entre neuronas artificiales.

Aprendizaje:

Las arquitecturas simbólicas aprenden de forma manual basándose en la codificación de reglas proporcionadas por expertos, mientras que las arquitecturas conexionistas aprenden automáticamente a partir de grandes cantidades de datos ajustando sus parámetros internos.

Interpretabilidad:

Las arquitecturas simbólicas poseen una alta interpretabilidad por lo que les es fácil rastrear cómo se llegó a una conclusión siguiendo las reglas, mientras que las arquitecturas conexionistas poseen una baja interpretabilidad, donde el proceso de decisión suele ser opaco ("caja negra").

Adaptabilidad:

Las arquitecturas simbólicas poseen una baja adaptabilidad ya que para cambiar el conocimiento se requiere una intervención manual con ajustes explícitos, mientras que las arquitecturas conexionistas poseen una alta adaptabilidad y se ajustan dinámicamente a nuevos datos sin intervención humana directa.

Volumen de datos requerido:

Las arquitecturas simbólicas requieren pocos datos pero mucho conocimiento explícito previo, mientras que las arquitecturas conexionistas necesitan grandes volúmenes de datos de entrenamiento etiquetados.

Manejo de incertidumbre:

Las arquitecturas simbólicas tienen dificultades para manejar información ambigua, incierta o incompleta, mientras que las arquitecturas conexionistas son robustas frente a la incertidumbre gracias a su capacidad de generalizar patrones.

Escalabilidad:

Las arquitecturas simbólicas son difíciles de escalar a dominios muy complejos sin que el sistema se vuelva inmanejable, mientras que las arquitecturas conexionistas escalan mejor a tareas complejas aunque requieren recursos computacionales elevados.

} 3.2 CASOS DE ESTUDIO

Sistemas Tutores Inteligentes: Uso de ACT-R para personalizar el aprendizaje.

Los **Sistemas Tutores Inteligentes (ITS, Intelligent Tutoring System)** buscan simular la experiencia de un tutor humano personalizado, adaptándose en tiempo real al conocimiento, habilidades y errores de cada estudiante. Un enfoque destacado para construir estos sistemas es la utilización de **arquitecturas cognitivas**, como **ACT-R** (*Adaptive Control of Thought-Rational*).

ACT-R es una arquitectura cognitiva que modela el razonamiento y aprendizaje humanos mediante la representación explícita de reglas de producción y bloques de conocimiento. Dentro de un ITS, ACT-R permite **predecir el comportamiento del estudiante, diagnosticar errores y proporcionar retroalimentación adaptativa**. Los tutores diseñados sobre ACT-R ajustan la dificultad y la naturaleza de los ejercicios en función de las inferencias sobre el estado mental del estudiante.

Un ejemplo clásico es el **Cognitive Tutor de Carnegie Learning**, utilizado en matemáticas de secundaria. Este sistema emplea un modelo de estudiante basado en ACT-R que simula los procesos cognitivos de resolución de problemas. A medida que el estudiante interactúa con el tutor, el sistema rastrea su progreso en distintos "objetivos de habilidades" y ofrece ayuda específica en las áreas donde se detectan dificultades.

La ventaja de usar ACT-R en ITS es que no solo se enfoca en resultados correctos o incorrectos, sino que también intenta **explicar las estrategias cognitivas subyacentes** y, por tanto, guiar al estudiante a mejorar su forma de pensar, no solo sus respuestas.

Asistentes Virtuales: Procesamiento del lenguaje natural en Alexa o Siri.

Los **asistentes virtuales inteligentes**, como **Alexa** o **Siri**, representan uno de los desarrollos más populares de las arquitecturas cognitivas aplicadas a la interacción humano-máquina. Su capacidad para comprender y responder en lenguaje natural se basa en complejos sistemas de **Procesamiento del Lenguaje Natural (NLP)** y **Machine Learning**.

Estos asistentes utilizan un pipeline de procesamiento que generalmente sigue los siguientes pasos:

1. **Reconocimiento automático de voz (ASR):** Convierte la señal de audio en texto.
2. **Comprensión del lenguaje natural (NLU):** Analiza el texto para identificar la intención del usuario ("intents") y extraer entidades relevantes ("slots").
3. **Manejo del diálogo:** El asistente decide cómo responder o qué acción ejecutar, basado en la intención detectada.
4. **Síntesis de voz (TTS):** Genera una respuesta hablada en lenguaje natural.

Por ejemplo, en **Alexa**, se utilizan modelos basados en deep learning para mejorar la precisión en la detección de intenciones y en la desambiguación de frases. Además, estos sistemas integran técnicas de aprendizaje automático que permiten a los asistentes **aprender y adaptarse al estilo y preferencias del usuario** con el tiempo.

En el caso de **Siri**, Apple combina el procesamiento en la nube con procesamiento en el dispositivo para garantizar tanto **respuesta rápida** como **privacidad**. Siri utiliza redes neuronales recurrentes (RNNs) y modelos de lenguaje grandes que han sido entrenados para interpretar comandos, preguntas y acciones conversacionales en distintos contextos.

Estos asistentes representan un **avance crucial** en cómo las máquinas pueden interpretar, aprender y adaptarse a la comunicación humana, acercando la interacción a un nivel más natural e intuitivo.

} 3.3 DESAFÍOS ACTUALES

Ética: Sesgos algorítmicos y privacidad.

Uno de los principales desafíos éticos en la interacción entre humanos y arquitecturas cognitivas es el manejo de **sesgos algorítmicos** y la **protección de la privacidad**.

Los **sesgos algorítmicos** ocurren cuando los sistemas de Inteligencia Artificial reproducen, amplifican o incluso crean nuevos prejuicios injustos. Esto puede deberse a:

- I. Datos de entrenamiento que reflejan prejuicios históricos o sociales.
- II. Modelos de aprendizaje automático que generalizan de manera inapropiada.
- III. Errores de diseño en las arquitecturas cognitivas o en la selección de variables.

Por ejemplo, sistemas de contratación automática, algoritmos de predicción judicial y asistentes virtuales han mostrado comportamientos discriminatorios por raza, género o nivel socioeconómico. Esto plantea preocupaciones sobre **equidad, transparencia y responsabilidad** en la IA.

La **privacidad** también es un problema central. Muchas arquitecturas cognitivas, como los asistentes virtuales o los sistemas de recomendación, recolectan grandes volúmenes de datos personales (voz, ubicación, preferencias). El uso inadecuado o la filtración de esta información puede exponer a los usuarios a riesgos como la vigilancia masiva, la pérdida de anonimato o el abuso de datos personales.

Para abordar estos problemas, se proponen prácticas como:

- I. Implementar evaluaciones de impacto ético.
- II. Desarrollar algoritmos explicables y auditables.
- III. Diseñar arquitecturas que minimicen la recopilación de datos ("privacy by design").
- IV. Regular el uso de IA mediante normativas como el Reglamento General de Protección de Datos (GDPR) en Europa.

Capacidad de Generalización: Dificultad para emular la creatividad humana.

Aunque las arquitecturas cognitivas y los sistemas de IA han logrado avances notables en tareas específicas, todavía enfrentan serias limitaciones en cuanto a la **capacidad de generalización** y la **emulación de la creatividad humana**.

La **generalización** se refiere a la capacidad de un sistema para aplicar lo aprendido en un contexto a nuevas situaciones diferentes. Muchos modelos de IA actuales son **altamente especializados**: funcionan bien en tareas para las que han sido entrenados, pero fallan al enfrentarse a escenarios imprevistos o ambiguos.

La **creatividad humana**, por su parte, implica la generación de ideas originales y adaptativas que combinan información de maneras novedosas. Requiere:

- I. Pensamiento abstracto.
- II. Transferencia flexible de conocimientos.
- III. Comprensión profunda de contextos culturales y emocionales.

Los sistemas actuales de IA, aunque pueden producir arte, música o literatura basada en patrones estadísticos (por ejemplo, modelos como GPT-4 o DALL·E), no poseen una comprensión real de significado o propósito. Su "creatividad" es más una extrapolación sofisticada que un proceso genuinamente innovador.

Los principales obstáculos incluyen:

- I. Falta de modelos cognitivos que integren razonamiento lógico, emocional y social de forma natural.
- II. Dependencia de grandes cantidades de datos y entrenamiento supervisado.
- III. Limitada capacidad para establecer metas propias o redefinir problemas (aspecto fundamental de la creatividad humana).

El reto futuro es desarrollar **IA verdaderamente general (AGI)** que pueda aprender, adaptarse y crear de manera flexible y autónoma, integrando múltiples formas de inteligencia.

[] 4. CONCLUSIONES Y REFLEXIONES

Conclusiones generales

El análisis de las interacciones entre humanos y arquitecturas cognitivas ha permitido evidenciar el impacto profundo que estos sistemas tienen en diversas áreas de la sociedad. A medida que estas tecnologías avanzan, su integración en la educación, la salud y la industria plantea oportunidades significativas, pero también desafíos técnicos y éticos que deben abordarse con cautela.

Las arquitecturas cognitivas han demostrado ser herramientas poderosas para la optimización de procesos y la simulación del razonamiento humano. Sin embargo, su capacidad de adaptación y creatividad sigue siendo limitada en comparación con la cognición humana. Aunque modelos como ACT-R, SOAR y CLARION han logrado avances en la comprensión y procesamiento de información, aún queda un largo camino por recorrer para alcanzar sistemas con una verdadera Inteligencia Artificial general.

Reflexiones sobre la interacción humano-máquina

La forma en que los humanos interactúan con estos sistemas varía en función de la interfaz utilizada. Los chatbots y asistentes virtuales han facilitado la accesibilidad a la IA en tareas cotidianas, mientras que los robots humanoides han mejorado la comunicación en entornos educativos y médicos. Sin embargo, el fenómeno del *valle inquietante* sigue siendo un obstáculo en el diseño de robots, mostrando la importancia de considerar factores psicológicos en el desarrollo de estas tecnologías.

Por otro lado, los modelos no humanoides como eBear han demostrado que la interacción efectiva no depende exclusivamente de la similitud con la forma humana, sino de la capacidad del sistema para expresar emociones y generar confianza. Esta perspectiva abre nuevas líneas de investigación sobre cómo optimizar la comunicación entre humanos y arquitecturas cognitivas.

Desafíos y consideraciones futuras

Uno de los principales desafíos actuales radica en la ética y la seguridad de estos sistemas. La privacidad y la gestión de datos personales siguen siendo preocupaciones fundamentales en la implementación de arquitecturas cognitivas. La necesidad de transparencia en los modelos algorítmicos y la reducción de sesgos en la toma de decisiones son aspectos clave para garantizar el uso responsable de la IA.

Asimismo, el reto de la generalización y la creatividad en estos sistemas aún no ha sido superado. Mientras que los modelos actuales pueden generar contenido basado en patrones de datos, aún carecen de una comprensión profunda de significado y contexto. Las futuras investigaciones deberían enfocarse en la integración de enfoques neuro-simbólicos y en el desarrollo de arquitecturas más flexibles que permitan una mayor adaptabilidad a entornos dinámicos.

En resumen, las arquitecturas cognitivas han revolucionado la manera en que los humanos interactúan con la Inteligencia Artificial, pero su evolución futura dependerá de cómo se aborden sus limitaciones actuales. La combinación entre avances tecnológicos y consideraciones éticas será clave para garantizar un desarrollo que beneficie a la sociedad sin comprometer sus valores fundamentales.

[] 5. BIBLIOGRAFÍA

Bibliografía:

- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). *Cognitive Tutors: Lessons Learned*. The Journal of the Learning Sciences, 4(2), 167–207. https://doi.org/10.1207/s15327809jls0402_2
- Pavlik, P. I., Cen, H., & Koedinger, K. R. (2009). *Performance Factors Analysis – A New Alternative to Knowledge Tracing*. Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED).
- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press

Bibliografía:

- Hoy, M. B. (2018). *Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants*. Medical Reference Services Quarterly, 37(1), 81–88. <https://doi.org/10.1080/02763869.2018.1404391>
- Zhang, Y., Sun, S., Galley, M., Chen, Y. C., Brockett, C., Gao, X., ... & Dolan, B. (2020). *DIALOGPT: Large-Scale Generative Pre-training for Conversational Response Generation*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL).
- Amazon Alexa. (2023). *How Alexa Works*. Amazon Developer Documentation. <https://developer.amazon.com/en-US/alexa/alexa-skills-kit>

Bibliografía:

- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). *A Survey on Bias and Fairness in Machine Learning*. ACM Computing Surveys (CSUR), 54(6), 1-35. <https://doi.org/10.1145/3457607>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. Minds and Machines, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Goodman, B., & Flaxman, S. (2017). *European Union regulations on algorithmic decision-making and a "right to explanation"*. AI Magazine, 38(3), 50-57.

Bibliografía:

- Marcus, G., & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon Books.
- Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms* (2nd ed.). Routledge.
- Chollet, F. (2019). *On the Measure of Intelligence*. arXiv preprint arXiv:1911.01547. <https://arxiv.org/abs/1911.01547>

Otros recursos:

https://es.wikipedia.org/wiki/Arquitectura_cognitiva

<https://www.ibm.com/es-es/topics/security-orchestration-automation-response#:~:text=SOAR%20>

<https://www.purestorage.com/es/knowledge/what-is-soar.html#:~:text=SOAR:%20conclusi%C3%B3n&text=En%20%C3%BAltima%20instancia%2C%20permiten%20una,y%20minimizar%20las%20amenazas%20externas>

[https://psicologiaycienciacognitiva.wordpress.com/2013/03/23/la-teoria-act-r-de-john-anderson/#:~:text=Los%20sistemas%20de%20producci%C3%B3n%20depositan,memoria%20de%20trabajo%20\(EMT\).](https://psicologiaycienciacognitiva.wordpress.com/2013/03/23/la-teoria-act-r-de-john-anderson/#:~:text=Los%20sistemas%20de%20producci%C3%B3n%20depositan,memoria%20de%20trabajo%20(EMT).)

<https://thedecisionlab.com/es/reference-guide/neuroscience/adaptive-control-of-thought>

[https://es.wikipedia.org/wiki/CLARION_\(arquitectura_cognitiva\)](https://es.wikipedia.org/wiki/CLARION_(arquitectura_cognitiva))

<https://www.toolify.ai/es/ai-news-es/ia-simblica-vs-redes-neuronales-una-comparacin-2567626>

<https://www.datacamp.com/es/blog/what-is-symbolic-ai>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC10597719/>

<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1190620/full>

<https://arxiv.org/abs/1511.06491>