

# ALGORITMO Y EMPATÍA: EL IMPACTO DE LA INTELIGENCIA ARTIFICIAL EN LA SALUD MENTAL

Javier VIDAL MARCIEL

## Resumen

Inteligencia Artificial es un concepto empleado para referirse al esfuerzo teórico y aplicado por construir sistemas cognitivos en soportes inorgánicos. En este artículo, se exploran algunas de las consideraciones y limitaciones que emergen de adoptar una perspectiva psicológica al respecto. Se propone también una clasificación intuitiva, para poder navegar de forma más comprensiva un campo que se encuentra en un momento revolucionario y convulso. Posteriormente, se analizan algunas de las implicaciones que la irrupción de este tipo de tecnologías tendrá en el ámbito de la salud mental, especialmente en el espacio aplicado de la terapia psicológica. Finalmente, con el objetivo de ilustrar esta transformación, se relatan una serie de innovaciones y avances más concretos.

## Abstract

Artificial Intelligence is a concept used to address the theoretical effort applied in the process of building cognitive systems on an inorganic architecture. In this paper, a psychological perspective is used to explore the various limitations and barriers that emerge from this scientific endeavor. A preliminary classification is also proposed to navigate in a more intuitive way this hectic and revolutionary field. Subsequently, we analyze some of the implications related to the irruption of AI technologies in the mental health sector, especially in the psychological therapy setting. Finally, with the aim of exemplifying this transformation, some specific advances and innovations are discussed.

**Palabras clave:** Inteligencia artificial. Salud mental. Procesamiento del lenguaje natural.

**Keywords:** Artificial intelligence. Mental health. Natural language processing.

## Introducción

La irrupción de la Inteligencia Artificial (IA) está desplegándose con increíble velocidad y descontrol sobre la sociedad y los diferentes sectores profesionales. La abrumadora dificultad teórica de los avances que se están produciendo y una incesante transgresión de barreras antes inalcanzables, nos posiciona ante ella con incertidumbre e indefensión. Durante la segunda mitad del 2023, hemos sido testigos de cómo múltiples gobiernos e instituciones han comenzado a elaborar marcos legales y éticos para tratar de regular este cambio. Incluso dentro de los núcleos tecnológicos que lideran estos avances, algunas autoridades dudan de poder predecir o controlar las consecuencias de sus propias innovaciones.

## Conceptualización, alcances y límites de la IA

Cuando se menciona la IA en círculos no especializados, generalmente se hace aludiendo a la aparición de un nuevo invento o para introducir a otros al uso de una nueva herramienta. Es inusual que surja en estas conversaciones en forma de concepto, abstraído y desprendido de la última aplicación del mercado. Pese a ser una revolución de gran altura teórica y técnica, la IA se está inventando dentro de la empresa. Ello dificulta que la discusión tome altura y permita ver el panorama de manera comprensiva. La innovación no está ocurriendo en laboratorios universitarios ni en centros de investigación, tampoco muchas de las publicaciones más importantes están editadas en revistas académicas. Las principales conferencias internacionales (*IEEE Transactions on Pattern Analysis and Machine Intelligence*) y los repositorios no indexados (arXiv o Huggingface), ocupan los puestos más altos en el volumen de artículos publicados. Los autores más importantes están directamente contratados por los gigantes de la industria tecnológica, son directivos de Google, Amazon, Microsoft, OpenAI, IBM o Meta. De los cinco artículos sobre IA más citados en 2022 -también en los años previos- cuatro están firmados por

autores afiliados a empresas privadas. El que conserva el primer puesto (Varadi et al., 2022) lo realizaron científicos de Google DeepMind, el tercero más citado (Ramesh et al., 2022) es de OpenAI, el cuarto (Liu et al., 2022) es de Meta y el quinto (Chowdhery et al., 2022) también de Google. Por lo tanto, los avances surgen más rápido de lo que pueden ser sistematizados por la academia, y cualquier intento por conceptualizar qué es la IA parece siempre llegar tarde al último desarrollo. Es crucial, por lo tanto, comenzar cualquier discusión al respecto en un primer intento por alejarse y plantear qué es la IA.

En su aspecto más abstracto o filosófico, la IA se refiere a la pretensión humana de desarrollar sistemas capaces de ejecutar procesos inteligentes en soportes inorgánicos o mecánicos. En este sentido, se incorporan ya dos ideas clave.

En primer lugar, la construcción de un sistema inteligente (SI) artificial por parte de un SI natural -el ser humano- es una tarea *extensiva*. Así, estos sistemas de procesamiento de la información son complementos virtuales al cerebro orgánico, utilizados para sustituir la cognición natural en funciones que no se acoplan bien a sus constricciones naturales. Desde nuestros orígenes, inventamos tecnologías para resolver problemas que escapan a nuestro repertorio de estrategias evolutivas innatas. Por lo tanto, la IA podría entenderse inicialmente como un conjunto de herramientas que extienden nuestras capacidades cognitivas, fundamentalmente para resolver problemas que requieren del análisis exhaustivo de cantidades masivas de datos. Es decir, pertenece al mismo orden que el resto de inventos que la ciencia va generando.

La segunda idea reconoce la creación de la IA como un acto *reflexivo*. El tipo de organismo que se busca diseñar es a imagen y semejanza del creador. Desde los primeros desarrollos en este campo (McCulloch y Pitts, 1943; Turing, 1950), la pretensión ha perseguido la construcción de modelos de circuitos neuronales; una ambición que posteriormente ha evolucionado hacia el objetivo de construir una *inteligencia general*. Es decir, si algo tiene de disruptora esta tecnología en referencia a cualquier invento previo del ser humano, es el deseo consciente de reproducir nuestro atributo distintivo. Por ello, el proyecto de la IA ha estado siempre vinculado al cognitivo, psicológico, neurológico; y, en el fondo, ha sido

desde el comienzo un acto de autorretrato. Cualquier SI artificial lleva incorporado el conocimiento -también el desconocimiento- que existe de la mente humana. Esta segunda noción marca una clara rotura entre la ciencia de la IA y el resto de las ingenierías.

Son estos dos atributos de la IA en conjunción, el *extensivo* y el *reflexivo*, los que colocan este proyecto intelectual en una posición nueva.

Desde otra perspectiva, la idea de la IA enmarca un grupo de disciplinas agrupadas alrededor del objetivo de construir agentes inteligentes. Si se atiende al volumen de publicación académica, la IA aglutina tres áreas del conocimiento (Bircan y Salah, 2022; Lei y Liu, 2019):

1. Las matemáticas, especialmente la ciencia computacional y la estadística. Incluso, en un nivel más abstracto, la lógica.
2. La biología, principalmente en lo relativo a la estructura y funcionamiento de las neuronas (neurobiología, neurociencia, neurocomputación, etc.). También se pueden incluir las teorías de los sistemas, la cibernética y aquellas que estudian la complejidad orgánica.
3. La psicología, en sus especializaciones dirigidas a la ciencia cognitiva y el paradigma del procesamiento de la información.

Existen, además, otras disciplinas que tienen un peso teórico importante, pero se fragmentan entre esas tres ramas principales. Un ejemplo es el estudio del lenguaje en IA, que es tratado desde su dimensión gramatical por la filología, desde su praxis por las matemáticas y la lógica, desde su función cibernética por la biología y, en su faceta semántica y léxica, por la psicología cognitiva.

Estas tres áreas -matemáticas, neurociencia y psicología- no están en ningún caso igualmente representadas. No reciben el mismo peso en la definición de metas y direcciones teóricas. Ni siquiera en esta frontera del conocimiento que es la IA, están participando muchas otras materias a las que sí apela. Por un lado, su naturaleza *artificial* hace que las ciencias exactas asuman el papel protagonista, pues en esencia son sistemas mecánicos. Sin embargo, cuando se trata su otra faceta *inteligente*, ocurren dos errores.

El primer error ocurre cuando la psicología acude con sus versiones más reduccionistas y simplificadas de la cuestión, asumiendo que la inteligencia depende estrictamente de la eficiencia, autonomía o adaptabilidad a tareas analíticas, y omitiendo otras partes importantes de la vida mental. Sin embargo, sabemos que los procesos emocionales, el filtro atencional o la escisión consciente-inconsciente, son también fundamentales para poder entender la inteligencia. Este error no es del todo grave si se consideran las limitaciones matemáticas y físicas de los sistemas de IA. Se trata entonces, de una simplificación por necesidad. Por ejemplo, no existe actualmente una intuición de cómo podría implementarse la interacción entre la percepción y la memoria en un cerebro artificial. Tampoco contamos con una explicación consensuada al respecto desde la psicología.

El segundo error, más grave y deliberado, es asumir que esta idea parcial de inteligencia es en algún momento parte de un continuum evolutivo en el que humano e IA conviven. Sabiendo que se está buscando replicar una versión reducida del fenómeno inteligente, que ya es en sí una simplificación de la actividad mental, sugerir que IA y ser humano están en una misma línea cualitativa de desarrollo es poco acertado. Incluso se supera esta falsa proposición cuando se insinúa que llegará a superarnos en un supuesto camino compartido del poder cognitivo.

La distorsión de fondo ocurre por una confusión en torno a la distinción entre las categorías de *agente* y *herramienta*, o de la posibilidad de realizar una transición de un estado a otro. Los que proclaman que la IA -una herramienta- está en camino de convertirse en un ser con agencia -como el ser humano-, definen el criterio que discrimina ambas entidades alrededor de dos perspectivas.

La primera sería el tipo de actividad que aparenta tener. Si la herramienta exhibe comportamientos similares a los de una agencia, debe por tanto estar en camino de serla. Esta es la base que sostiene el conocido *Test de Turing* (Turing, 1950), en el que se establece que si en el diálogo con una máquina no puedo distinguirla de un humano, ésta es, por lo tanto, *inteligente*. Existe una célebre argumentación que justifica por qué esta idea es errónea: *La Habitación China* de Searle (1980).

La segunda perspectiva para el criterio adoptado es la eficiencia para completar objetivos en dichas tareas. Si además de manifestar el mismo comportamiento, una IA puede incluso superar a la agencia en una serie muy concreta de problemas -jugar al ajedrez, clasificar objetos o conducir un coche, etc.-, debe por lo tanto estar en camino de adelantarla evolutivamente.

Sin embargo, cuando ambos criterios se aplican para convertir la IA en algo cualitativamente simétrico o superior al ser humano, fallan en salvar una de las premisas básicas de cualquier agencia: estar dotada de *voluntad* propia. Como principio, la característica esencial de un sistema agente es no depender necesariamente de otro<sup>1</sup>. Por lo tanto, la imposibilidad de que la IA actual o en su potencial desarrollo se convirtiese en una agencia similar al humano está presente en tres niveles:

- No existe una IA independiente de la agencia o voluntad humana. Su aparente autonomía es la de seguir unas directrices para las que ha sido programada.
- Tampoco se sabría cómo diseñar o programar dimensiones básicas de la agencia tales como la voluntad o la identidad.
- Ni siquiera existe una articulación sólida, consistente y consensuada de qué son la agencia, la voluntad, la identidad o la consciencia; ni de cómo emergen en el ser humano.

Es decir, ni es agencia ni se conoce cómo podría llegar a serlo. Por lo tanto, si el ser humano es un sistema agente y la IA no reúne ni reunirá previsiblemente las condiciones para serlo, es y seguirá siendo su herramienta. Para no ser instrumental debería no solo cambiar cuantitativamente, sino que tendría que hacerlo cualitativamente. Puede que aquellas ideas que rechacen esta argumentación, antes que elevando la IA a agencia, estén reduciendo al ser humano a herramienta para lograr este falso paralelismo.

Otra conceptualización de la IA se hace desde la clasificación de las formas que puede tomar: las actividades en las que se implemente y **casos de uso**<sup>2</sup> que logra resolver. Hay tres ejes que pueden servir para proponer una **taxonomía intuitiva**<sup>3</sup>:

1. Según el **nivel de autonomía** que se le otorga a la IA para definir sus parámetros de acierto/error. Es decir, la libertad para elegir en qué dirección debe moverse para establecer que está aprendiendo:

1a. **Aprendizaje supervisado:** se proporcionan de forma directa y completa las reglas y variables que componen el fenómeno que se pretende modelizar. Por ejemplo, en una tarea de clasificación de imágenes de perros y gatos, se le proporciona una base de datos con millones de fotos catalogadas como uno u otro animal. De esta forma, el sistema aprende tomando como referencia una etiqueta inicial para determinar el grado de error en sus predicciones. Es supervisado porque aquel que lo programa está indicando cuándo una predicción (*la imagen X es un perro*) es falsa o cierta. La IA busca, por lo tanto, reducir al máximo el número de veces en los que sus aproximaciones contradicen las del supervisor.

1b. **Aprendizaje no supervisado:** se introducen un conjunto voluminoso de variables posibles y se permite al modelo seleccionar unas u otras como criterios determinantes, obteniendo así los rasgos con mayor peso explicativo. En el caso del ejemplo anterior, después de observar millones de imágenes de gatos y perros, la IA comenzará a utilizar la presencia de variables no explicitadas (*bigotes, hocico, orejas puntiagudas o pupila ovalada*) para distinguir un animal de otro. El programador humano no ha indicado que la presencia de pupilas ovaladas sea un rasgo distintivo de los gatos. La IA lo aprende porque, tras procesar millones de ejemplos, establece que un grupo de fotos tiene pupilas ovaladas y otro no. Incorpora así la noción de que éste es un criterio adecuado para diferenciar ambos animales. Al ser un método de aprendizaje parsimonioso, no directivo, necesita un mayor volumen de datos. Pueden ser increíblemente útiles pues detectan variables que el agente que diseña el modelo pudo obviar.

1c. **Aprendizaje por reforzamiento:** aunque técnicamente todos los modelos aprenden por acierto o error, en este paradigma de la IA se introduce esta noción de forma más realista. Se incorpora un *ambiente o ecosistema* a la tarea de aprendizaje. Aquí las *decisiones* que el modelo toma *elicitán*<sup>4</sup> consecuencias ambientales y el entrenamiento tiene un sentido acumulativo. Es cierto que esta tercera categoría es más reciente y queda por ver si constituye un paradigma en sí mismo o es simplemente una sub-variante de alguno de los dos anteriores.

2. Según el **tipo de información** o registro sensorial con el que trabajan. En realidad, todos los modelos de IA operan internamente con información numérica. Al igual que ocurre con el estudio cognitivo del ser humano, clasificar la información según el modo sensorial pierde sentido cuando llega al sistema nervioso y todo se transforma en impulsos electroquímicos. Hallamos, pues, codificadores de IA que logran transformar diversos tipos de datos digitales (palabras, sonidos, imágenes o señales) a vectores y matrices. Son, pues, sistemas que trabajan con representaciones numéricas de los diferentes modos sensoriales<sup>5</sup>. Existen cuatro posibles clasificaciones (más una categoría híbrida) con sus ejemplos correspondientes:

2a. **Información numérica**: la serie histórica de un activo financiero o las notas de un alumno a lo largo de toda su vida académica.

2b. **Información visual**: los vehículos autónomos, el reconocimiento de firmas o el reconocimiento facial en la cámara de un móvil.

2c. **Información auditiva**: los sistemas de reconocimiento e interacción por voz, así como los modelos de transcripción automática.

2d. **Lenguaje** (Procesamiento del Lenguaje Natural): los chatbots o cualquier modelo de transformadores (Vaswani et al., 2017) como GPT-4 (OpenAI, 2023a), BERT (Devlin, 2018) o Bard (Manyika y Hsiao, 2023).

2e. **Multimodal**: se trata del esfuerzo por integrar múltiples formatos sensoriales en un mismo proceso de aprendizaje. Uno de los más famosos es DALL-E 3 (OpenAI, 2023b), que representa aquellos sistemas que traducen textos en imágenes o viceversa.

### 3. Según la **dirección del procesamiento de la información**

3a. **Inductiva** (IA Codificadora): la gran mayoría de los modelos toma observaciones concretas (una imagen de un perro o gato) de forma masiva y simultánea para extraer posteriormente un aprendizaje en forma de criterios que envuelvan adecuadamente el fenómeno (distinción física entre un animal u otro).

3b. **Deductiva** (IA Generativa): recientemente, se han desarrollado procesos a la inversa. Utilizando codificadores extremadamente potentes y muy entrenados (un clasificador de imágenes de



perros y gatos con elevada precisión), se puede realizar una *deco-dificación* de estas reglas abstraídas para que el modelo represente un ejemplo concreto y nuevo a partir de lo aprendido (dibujar un perro incorporando las características de todas las imágenes correctamente clasificadas como tal). Pocas son las excepciones que exclusivamente son IA Generativa sin incluir también capacidades inductivas o codificadoras (Aggarwal et al., 2021). Por lo tanto, la mayoría de los modelos generativos son en realidad combinaciones de ambas direcciones de procesamiento de la información.

Siendo la historia de la IA compleja y diversa, se pueden destacar cinco ejes de transformación a partir de esta clasificación intuitiva:

1. **De lo supervisado a lo no supervisado:** se busca la emancipación de la intervención humana, la reducción de las instrucciones previas o de los parámetros de actuación preestablecidos. Los nuevos modelos tienen directrices más complejas, pero menos detalladas y concretas, capaces a la vez de hacer emerger un mayor rango de habilidades. Las nuevas IAs son autoprogramables.

2. **De lo unisensorial a lo multimodal:** la integración de la información a priori incompatible, transformando inputs visuales, sonoros o lingüísticos en un lenguaje numérico común. Se persigue la capacidad para intercambiar intuitivamente de un modus sensorial a otro, convirtiéndolos. Logran así obtener una representación más completa de los objetos sobre los que se opera.

3. **De la IA inductiva/codificadora a la deductiva/generativa (simultánea).** Otro de los importantes vectores de desarrollo es la combinación de sistemas analíticos y sintéticos, capaces de entender y simultáneamente producir información. Un ejemplo de IA exclusivamente codificadora, podrían ser aquellos sistemas que procesan los comentarios de clientes en una tienda online, asociando una valoración con el número de *estrellas* que el usuario ha otorgado. Este modelo podría destacar los servicios que aparecen con más frecuencia en las valoraciones negativas, para que el equipo de atención al cliente pueda enfocarse en ellos. Un salto de este sistema unidireccional a uno que incorpore lo generativo, podría ser incluir una respuesta automatizada y personalizada a cada uno de los comentarios de los consumidores. Consistiría en entrenar a la IA en responder con un *Gracias* a las valoraciones positivas o un *Lo sentimos* a las negativas. Todo esto, puede complejizarse mucho más.

**4. De la tarea concreta a la experiencia *copilot*.** Si en un principio cada sistema se desarrollaba para resolver un problema específico (jugar al ajedrez), las proyecciones futuras se encaminan a generar la *experiencia copilot*. En ella, la IA es una acompañante que sabe resolver diversos problemas que contengan temáticas diferentes y que demanden metodologías o técnicas diversas. ChatGPT quizás sea el mejor ejemplo, ya que es capaz de ofrecer una amplia gama de soluciones: desde producir imágenes hasta escribir código, resumir artículos hasta analizar una base de datos, o realizar búsquedas en internet hasta escribir poesía. No sólo se diversifica horizontalmente, sus tareas descienden a la vida diaria. Un compañero para la cotidianidad.

**5. De lo inaccesible a la experiencia de usuario.** En sus orígenes, los sistemas de IA se configuraban y aplicaban a través de lenguajes de programación y complicados procedimientos informáticos, lo cual motivó que durante décadas se mantuviesen encapsulados en ámbitos de investigación muy especializados. El florecimiento que ha experimentado en la última década, fruto de su vinculación con el sector productivo e industrial, ha creado la necesidad de potenciar la experiencia de usuario. Cualquier modelo que pretenda actualmente tener un impacto en la vida cotidiana, incluye interfaces más accesibles y modos de interacción más universales. Por ejemplo, los asistentes por voz como Siri (Apple) o Alexa (Amazon) se integran perfectamente en el manejo de un televisor inteligente.

En una última definición, más operativa y práctica, podríamos decir que la IA es un sistema de detección-producción de patrones, capaz de extraer información comparando cantidades masivas de observaciones concretas, que aprende con mayor o menor autonomía a generar abstracciones, criterios y reglas discriminatorias que permitan predecir nuevas iteraciones del tipo de fenómeno específico para el que son entrenadas.

## **IA y salud mental**

En la atención a la salud mental, desde sus diversas disciplinas y ámbitos, la IA promete numerosas aplicaciones. Cuando se contempla este fenómeno emergente, muchos de los profesionales que trabajan en contacto directo con el paciente o cliente inevitablemente lo cuestionan.

En primer lugar, subyace una emoción de miedo ante esta posibilidad. Un espíritu ludita, no exclusivo del área de salud mental, se apodera de cualquier trabajador ante la idea de introducir una segunda *inteligencia* en su desempeño profesional. Es importante este inciso porque es la primera vez en la historia de las introducciones tecnológicas, que una de éstas se considera *inteligente*. Aunque, como se ha discutido en el anterior apartado, dicha etiqueta está llena de objeciones y matices.

Pero en una segunda mirada, hay una legítima oposición a la introducción de la IA en los servicios de salud mental, cuando en ocasiones se sugiere que ésta ocurrirá en el acto terapéutico. Ciertamente, desde los comienzos de la IA, siempre ha habido aquellos que sugerían con esperanza o miedo que versiones muy avanzadas podrían convertirse en dignas sustitutas del mundo social, tanto en su dimensión cognitiva (informativa, operativa, dialéctica...) como emocional (reciprocidad afectiva, seguridad identitaria, regulación emocional o incluso compartir una intimidad). Es por ello factible que se sugiera la posibilidad de sistemas de IA cumpliendo la función terapéutica. De hecho, llevan existiendo desde hace años *bots*<sup>6</sup> que buscan satisfacer necesidades de socialización y también otros que pretenden entablar una relación terapéutica. *JoBot* es un ejemplo de programa de IA modelado a partir del trabajo de J. Diederich, un psicólogo clínico. En su página web describe: “JoBot realiza entrevistas diagnósticas para un amplio rango de problemas de salud mental, incluida la depresión y la ansiedad” (Diederich, 2023, párr. 02 About). *Wysa* (Sinha et al, 2022) es otro ejemplo de un agente conversacional que aplica tratamientos cognitivo-conductuales para diferentes poblaciones.

Evidentemente, la mera introducción de esta idea junto con la confianza ciega en los futuros potenciales de la IA que tinta actualmente de entusiasmo este campo, más pronto que tarde va a generar una profunda crisis en la disciplina psicológica. Obligará posiblemente a dividir el servicio terapéutico en dos funciones generales, una **informativa** y otra **relacional**.

La función informativa contendría todo aquel espectro de conocimientos objetivos que un profesional puede transmitirle a su cliente o paciente sobre el estado de una condición patológica o acerca de la posi-

ción que ocupa respecto a un ideal de salud mental. También los datos que le facilite sobre sus síntomas y signos, tratamientos farmacológicos o posibles recursos auxiliares con los que coordinar la intervención. Incluso podrían incluirse aquí aquellas técnicas terapéuticas que sean mecanismos estructurados y procedimientos estandarizados en respuesta a un cuadro determinado (por ejemplo, una relajación muscular progresiva o una meditación guiada). Aunque claramente, hay otras que escapan lo puramente informacional y entran el terreno de lo relacional *-la silla vacía-*, que pese a tener un *protocolo* o *guión*, no podrían aplicarse de manera estereotipada a cualquier paciente. En esas técnicas menos estructuradas, los objetivos terapéuticos concretos, las defensas del paciente y el círculo social específico imposibilitan la ejecución de fórmulas de carácter estandarizado. Son, por lo tanto, un universo de acciones terapéuticas que podrían considerarse entre las dos funciones básicas de la relación de ayuda.

Por otra parte, la función relacional incluye todas las dinámicas implícitas o explícitas entre el profesional y su cliente, que aun pudiendo ser tecnicizadas, emanan de habilidades sociales fundamentales en el ser humano. Esto incluye todas las operaciones que el psicólogo realiza desde el diálogo y que encuentran en el vínculo terapéutico su fundamento (la escucha empática, la confrontación o la transferencia y contratransferencia pueden ser algunos ejemplos).

Ambas representan gamas de las diferentes herramientas que posibilitan la transformación y el cambio en un proceso de esta naturaleza. Se utilizan de forma conjunta y dinámica, pues cada una atiende a las diferentes necesidades que exige cada formulación de caso. Algunas de las destacadas diferencias entre escuelas y corrientes de la psicología clínica emergen del peso que colocan sobre una u otra. Algunas consideran lo relacional la esencia de su terapia y lo informacional un añadido, o, al contrario, otras entienden que dotar al cliente de información es más prioritario que entablar el vínculo. Podría también considerarse una diferencia entre profesionales, más que de enfoques teórico-prácticos.

El poder de esta dicotomía de funciones como criterio que ordene los tipos de práctica psicológica puede haber sido hasta ahora más o menos evidente o incluso no considerarse como válido. Sin embargo, la

actual y futura entrada de la IA en el dominio terapéutico exige aplicarla y profundizar en ella.

Esta cuestión forma parte de una discusión mayor acerca del grado, la profundidad y la deseabilidad de la sustitución de seres humanos por sistemas de IA en el nivel del ejercicio profesional. Cualquier predicción al respecto debe leerse con mucha cautela debido a ciertos rasgos que caracterizan a este tema, entre los que resaltan dos.

El primer rasgo tiene un sustrato principalmente emocional. La profesión y el empleo son y continúan consolidándose como una de las áreas existenciales básicas del individuo moderno, alrededor de las cuales organiza gran parte de su vida social e identidad. Cuestionar la posibilidad, además con suficientes evidencias, de que pronto pueda desaparecer, nos coloca en una angustiada crisis.

El segundo rasgo es el nivel de complejidad teórica y altura técnica que caracteriza a la IA, dificultando que se entiendan bien los avances y sus consecuencias. Se está hablando, al fin y al cabo, de generar modelos matemáticos de la mente humana, sin ser conscientes del precario entendimiento que actualmente se tiene de ella.

Partiendo de esto -de que angustia e incompreensión caracterizan este tema-, están ya entre nosotros herramientas de IA sobre las que se pueden delegar muchas de las tareas que realizamos cotidianamente en el trabajo. Un criterio para poder predecir cuáles serán las primeras, es mantener presente cuáles son los dos núcleos que ahora mismo enmarcan las habilidades más fuertes y prometedoras de la IA -lo cual nos devuelve a la clasificación entre IA codificadora y generativa que se ha propuesto previamente-.

En primer lugar, se encuentra el procesamiento analítico y automatizado. Todo aquello que consiste en detectar factores explicativos, extraer criterios que ordenen la información, clasificar ejemplos dentro de una taxonomía determinada, automatizar operaciones concretas o diagnosticar entidades a partir de parámetros; por nombrar algunas competencias.

En segundo lugar, la elaboración de conclusiones cuya síntesis es explícita y, por lo tanto, *configurable*. Debido a que el análisis de la información es bastante abierto y objetivo, una vez *entiende* una parte de la realidad, la IA es capaz de generar réplicas de ella de forma determinista.

Por ejemplo, puedo pedirle a un conocido que me resuma un libro -es decir, que a partir de todo el conjunto de información que contiene priorice segmentos generales-. Yo no sé qué ha llevado a esta persona a considerar más importante una parte que otra para asegurar que entienda de qué trata la novela. Incluso, más allá, puede no saberlo ni él mismo. ¿Por qué ha acudido a su consciencia que un personaje representa mejor que otro el eje argumentativo de la historia? Quizás ha conectado más con una parte que con otra por cómo hace referencia a su vida personal. Tal vez estuviese más motivado a leer durante el comienzo y por lo tanto ha omitido más contenido de los últimos capítulos. Si, por el contrario, uno de los creadores de ChatGPT pide a su máquina que *genere* un resumen de esa misma narración, es perfectamente consciente de los criterios que se utilizan para rescatar una información sobre otra. No sólo conoce qué parámetros ha utilizado el modelo para codificar la novela, también sabe que ChatGPT está usando ciertas páginas de Internet para contextualizar su sinopsis -como la biografía del autor o la página de Wikipedia dedicada a dicho libro-. Por lo tanto, la IA no sólo será capaz de producir versiones nuevas del contenido con el que es entrenada, sino que también sus programadores podrán configurar los sesgos utilizados para ello.

Devolviendo esta discusión de la sustitución laboral al contexto de la salud mental, podría sugerirse la siguiente predicción: aquellos profesionales que basen su práctica clínica en su *función informacional* compiten de forma más directa con la IA. Por el contrario, aquellos que priorizan el vínculo -la *función relacional*- están más blindados a que sus tareas sean delegadas a la IA. Cuanto más alérgico -o menos cómodo- esté un profesional con lo ambiguo, contradictorio e incierto que caracteriza a lo subjetivo, más expuesto estará a que sus futuros pacientes recurran a la IA.

A la vez, esta situación plantea una crisis que confronta no sólo a los profesionales, sino al papel que se pretende que tenga la psicología aplicada en la sociedad. Cuanto más se defienda la psicología como una

disciplina cuya tarea es diagnosticar, diseñar trastornos a partir de cuadros sintomáticos y proponer tratamientos estructurados o desarrollar técnicas protocolarias, más expuesta se encontrará a que aquellos que la requieren acudan a modelos de IA en un futuro. Es más, cuánto más numérica o cuantitativa sea la medición del éxito en todas estas tareas, más susceptible es la psicología que conocemos ahora de experimentar futuras intrusiones. Se podría proponer por lo tanto que, cuanto más se mecanice la práctica terapéutica, más permeable se hace a las aplicaciones de la IA.

Por supuesto éstas son predicciones transportables a prácticamente todos los sectores profesionales. El caso de la educación es parecido, aunque en ese sector la entrada de la IA está siendo más acelerada. Cuanto más se comprenda la tarea pedagógica e instructiva como una transferencia de exclusivamente información objetiva o como una evaluación de parámetros numéricos, más fácilmente se prescindirá de tener un ser humano en el aula.

La valoración de la deseabilidad y conveniencia de la sustitución de actores humanos en favor de aplicaciones de la IA en el ámbito de la Salud Mental es una cuestión que deben resolver los diferentes campos implicados<sup>7</sup>. Si algo es claro es que esta cuestión apela directamente a las humanidades, a lo fenomenológico, a lo relacional y lo subjetivo. Estamos asistiendo a una implosión de lo cuantitativo. En su proyecto expansivo por traducir las ciencias humanas y sociales a lo numérico, inevitablemente está haciendo desaparecer al humano de las disciplinas matemáticas y exactas.

## **Un esbozo de proyección**

En un reciente Trabajo Fin de Máster (Vidal Marciel, 2023), se realiza un acercamiento hacia una gama de herramientas con un alentador potencial para la intervención en salud mental. Como se ha mencionado antes, una de las áreas de producción más importantes de la IA es el Procesamiento del Lenguaje Natural (PLN). De forma general, atañe al entendimiento y la producción de información lingüística en un formato igual al humano -natural-. La principal razón por la que esta rama es tan importante y prometedora, es por aquello mismo que hace del lenguaje la

base de nuestro pensamiento y comunicación: su capacidad para encapsular información y significado. Es decir, si una IA comprende una lengua, podría operar con cualquier hablante sobre cualquier tema del que se tenga conocimiento, siempre sabiendo que en este caso el verbo *comprender* se emplea ilustrativamente, en absoluto en el mismo sentido en el que se aplica a un ser humano. Así, para una IA de PLN, escribir poesía no es sustancialmente diferente a escribir un contrato de arrendamiento. Ambos procesos se basan en la recombinación y correcta disposición del texto acorde a una solicitud determinada.

Este conjunto de tecnologías son los *transformadores* que actualmente representan el paradigma más avanzado en lo que a *comprensión y producción* lingüística se refiere. La barrera más importante de todos los intentos desde la década de 1950 para construir una IA con la que poder dialogar, es implementar la propiedad semántica del lenguaje. Introducir la morfología y la sintaxis fue más sencillo, pues se trata de un compendio de reglas lógicas y formales, más aptas para su manipulación matemática.

Los *transformadores* suponen un avance al lograr una aproximación parcial a la semántica a través del contexto. ChatGPT o BERT, dos aplicaciones de este conjunto de modelos, no pueden asimilar el significado de la palabra *rey*, por ejemplo. Cuando procesan esta palabra no acude a su espacio de asociación todo un universo de conceptos relacionados, como ocurre en el ser humano. En nuestra mente, pensar en la palabra *rey* no sólo elicitó otras asociadas léxicamente como *reino* o *monarquía*. También desencadena que otros conceptos vinculados cultural o emocionalmente a *rey* suban a consciencia, aunque no se parezcan morfológicamente. Para un republicano, conceptos como *opresión* o *privilegio* podrán emerger. Para un monárquico, quizás otros como *institución* o *lealtad*. Esta es la difícil naturaleza del significado, que no es exacto ni opera bajo reglas determinísticas. Estos modelos de IA superan este problema creando un mapa, una topografía léxica de una lengua<sup>8</sup>.

Siendo entrenado con millones y millones de frases, GPT consolida poco a poco la asociación entre la palabra *rey* con la palabra *reina*, siguiendo el ejemplo. En todos los artículos de Wikipedia que haya procesado, ambas palabras aparecen cerca en muchas más ocasiones que *rey* y otras como *estanque* o *árbol*. De hecho, lo que GPT calcula es la proba-



bilidad de que *reina* aparezca en la misma frase que *rey*. A través de la puntuación atribuida a la probabilidad, comienza a crear un mapa del castellano, en el que cada palabra tiene una distancia determinada con todo el resto de palabras, unas coordenadas exactas. En este caso, *rey* estará cerca de *reina*, pero estará más lejos de *femenino* y más cerca de *masculino*. Eso le permite a la IA diferenciar que el *rey* es un varón y la *reina* una mujer, siendo a la vez de la misma familia léxica. Así, aunque GPT jamás sepa lo que es un monarca, puede situarlo en relación con otros conceptos. Es decir, entiende parcialmente su semántica gracias a generar un contexto léxico, una topografía del vocabulario. Esto es una simplificación enorme de lo que constituye un complejísimo proceso matemático, pero sirve para acercarse a cómo puede lograrse una comprensión lingüística más acertada en la IA.

Una de las tareas más ejecutadas en este ámbito y empleada en la investigación reseñada en el presente apartado, es la clasificación de textos. En concreto, se diseñó un problema de clasificación supervisada y binaria. Es decir, donde las alternativas a clasificar eran dos y cada ejemplo se procesó con una etiqueta externa. La pregunta de investigación formulada fue: ¿podría una IA distinguir el lenguaje en función del nivel de ansiedad del emisor?

Para ello se entrevistó a 11 mujeres jóvenes, con preguntas típicas de cualquier interacción terapéutica: *¿Cómo te sientes en tus relaciones sociales actualmente?* *¿Cómo ves el futuro?* Además, se aplicó un test psicométrico para medir el nivel de ansiedad de forma válida. El *State-Trait Anxiety Inventory* (STAI; Spielberger et al., 1970), es una prueba muy consolidada que reporta un excelente comportamiento estadístico. Mide la ansiedad en dos dimensiones: como *estado* y como *rasgo*. La primera hace referencia a un conjunto de síntomas, pensamientos y sensaciones puntuales y atribuidas a una situación específica. La segunda, entendida como un rasgo de personalidad, se interpreta como la permanencia de dicho estado a través del tiempo y los contextos, como una característica distintiva de la persona. Una vez obtenidas todas las puntuaciones, se dividieron las 11 mujeres en dos grupos a través de la mediana (dos mitades, una con alta y otra con baja ansiedad).

El ejercicio de la IA consistía en comparar las transcripciones de las entrevistas de un grupo con otro, aprendiendo a distinguirlas basándose -aprendizaje supervisado- en las puntuaciones del STAI (*alta* o *baja*). Finalmente, para evaluar el nivel de aprendizaje, la IA debía predecir si un texto pertenecía a un grupo u otro sin contar con la etiqueta del STAI.

Para poder aplicar estos modelos BERT al español, se utilizaron unas versiones publicadas por el Departamento de Ciencias de la Computación de la Universidad de Chile (Cañete et al., 2020). Estos modelos han sido entrenados en el *Spanish Unannotated Copora*, una base de datos con tres billones de palabras. Incluye, entre otras, toda la Wikipedia, los artículos del Parlamento Europeo y todas las publicaciones JCR. Gracias a que la IA había sido *pre-entrenada*, se pudo realizar esta tarea con una muestra tan reducida. El modelo incorpora todo el conocimiento anterior y su aprendizaje reside en ir ajustando (*fine-tuning*) sus parámetros a los nuevos textos (las transcripciones de entrevistas).

Al utilizar diferentes versiones, unas más voluminosas que otras, se obtuvo una precisión media alrededor del 60%. Esto es, en las predicciones del modelo, aquellas que coinciden con la etiqueta del STAI. Es importante rescatar algunas de las limitaciones de la investigación. Primera, la muestra era muy reducida y poco representativa. Segunda, la utilización del STAI como la medida de la ansiedad es un reduccionismo de lo que representa este fenómeno. Tercera, la división de los niveles de ansiedad en sólo dos grupos era necesario, por el tamaño muestral, pero es una visión dicotómica y poco realista de la ansiedad. Con todo ello, haber alcanzado una precisión del 60% (dada esta muestra, la inexistente inversión económica y la escasa experiencia del autor en computación), permite visualizar el enorme potencial de adopción que tiene esta tecnología.

Esta investigación representa una versión tentativa de lo que podría ser un sistema de diagnóstico automatizado. La tecnología en este caso ya está disponible para poder superar las limitaciones mencionadas e incrementar la precisión del modelo, sólo se precisa más muestra. ¿Qué aplicaciones podría tener un modelo con un suficiente número de entrevistas? Un modelo que pudiese, por ejemplo, clasificar a las personas en percentiles con una exactitud mayor al 95%. Para concretar aún más esta propuesta, ¿cómo se aplicaría al deterioro cognitivo en la tercera edad?

Desarrollar estos métodos, para poder determinar el nivel en el que una persona se sitúa desde diferentes constructos clínicos, tiene diferentes ventajas -entendiendo siempre que es simplemente una parte a integrar dentro del diagnóstico-. Como la IA funciona de forma automática y permanente, un sujeto podrá, por lo tanto, obtener sus puntuaciones sin necesitar la asistencia directa de un profesional o sin tener que completar un test psicométrico concreto.

Imaginemos que el modelo de IA con el que contamos ha sido entrenado en entrevistas con millones de sujetos y contamos con sus respuestas en una gran variedad de pruebas para detectar el deterioro cognitivo. Podría tomarse todo el contenido lingüístico que la persona va depositando de forma digital -chats, notas, redes sociales o comentarios en sitios web- para monitorizar cómo éste se compara con el lenguaje emitido por personas de su edad o con mayor deterioro cognitivo. Con un sistema suficientemente certero, contaríamos de forma instantánea con una puntuación para muchas variables -rigidez del pensamiento, fluidez verbal o memoria-. Además, esto ocurriría en directo y de manera continua, obteniendo un recorrido de estas dimensiones a lo largo de meses o años. Podría también implementarse un aviso para que, al superar ciertos niveles, este individuo acudiera a un servicio de salud mental. Al entrar en consulta, el profesional que le atiende podría conectarse a esta herramienta de IA y visualizar toda esta información. Es más, gracias a todos los casos que el sistema se encuentra analizando de forma simultánea, podría reconocer el impacto que cada tratamiento está teniendo en los diferentes pacientes de forma instantánea. El sistema también podría recomendarle al profesional una opción u otra. Pensemos en cómo todas estas oportunidades aligerarían los sistemas de sanidad pública, cómo liberarían a los profesionales de procedimientos formales para poder dedicarle tiempo de calidad al paciente. También sería considerable el impacto que este sistema preventivo podría tener en la detección precoz del deterioro cognitivo. Es más, dado este nivel de sistematización y detalle automatizado, probablemente el profesional pueda publicar permanentemente sus resultados. Si antes estaba forzado a dedicar su tiempo y el del paciente para completar pruebas y cuestionarios en caso de querer publicar, ahora estaría recogiendo información numérica sólo con mantener una conversación en su despacho.

¿Cómo modificarán estas tecnologías las teorías psicopatológicas? En la mayoría de las grandes clasificaciones -DSM-5 (American Psychological Association, 2013) o CIE-11 (World Health Organization, 2022)- los criterios diagnósticos se construyen a partir de la valoración de expertos y marcadores psicométricos. Si se contase con titánicos volúmenes de lenguaje codificado sobre cada uno de los trastornos, ¿cambiarían en algún aspecto las definiciones? Si se puede contrastar con exactitud matemática el lenguaje de un Trastorno Límite de la Personalidad con un Trastorno Histriónico de la Personalidad, ¿aparecerían criterios diagnósticos diferentes? Resulta imposible conocer una respuesta real a esta pregunta, porque hasta la fecha no ha sido posible realizar esta evaluación. Por supuesto, numerosas consideraciones éticas emanan de este planteamiento.

Volviendo al caso del deterioro cognitivo, ¿qué pasaría si introducimos otro tipo de información? Muchas de las valoraciones neuropsicológicas se realizan a través de dibujos, por ejemplo. Además de todo el mecanismo de lenguaje, el modelo de IA sería capaz de analizar -y mezclar con lo mencionado antes- pruebas como el *Test del Reloj* (Aprahamian et al., 2009) o la *Figura Compleja de Rey-Osterrieth* (Rey y Osterrieth, 1941). Quizás sea más complejo, pero estas propuestas ya están llegando y continuarán introduciéndose en la clínica general. Es por ello que la IA puede también revitalizar las pruebas proyectivas, dándoles la misma validez científica que al resto de herramientas sin perder su riqueza. Es algo que ya se está desarrollando con el *Test de Apercepción Temática* (Camati et al., 2021) o con el *Test de la Casa-Árbol-Persona* (Pan et al., 2022).

## Conclusiones

Sin lugar a duda, la IA va a ser aplicada a la salud mental y, previsiblemente, lo hará primero en aquellas actividades más mecánicas, formales y protocolizadas. Para poder protegerse de una probable intromisión y sustitución contraproducente, los profesionales del sector tendrán que dar mayor peso al vínculo y a lo cualitativo. Esta crisis puede conducir a dos alternativas. En una, la psicología realiza esta transformación hacia lo cualitativo y consolida su razón de ser en nuestra sociedad. Con-

vive en simbiosis con la IA, liberando tiempo para el paciente, la reflexión y la supervisión. En el otro panorama, continua su actual camino hacia la mecanización y entrega sus funciones a la IA.

*Javier VIDAL MARCIEL (Madrid, 1999) es Psicólogo por la Universidad Pontificia Comillas y Máster en Investigación en Psicología por la Universidad Internacional de La Rioja. Contacto: javier.vidal.psy@gmail.com*

## Referencias

Aggarwal, A., Mittal, M-, y Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1), 1-9. <https://doi.org/10.1016/j.jjime.2020.100004>.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>

Aprahamian, I., Martinelli, J. E., Liberalesso Neri, A., y Sanches Yassuda, M. (2009). The Clock Drawing Test. A review of its accuracy in screening for dementia. *Dementia & Neuropsychologia*, 3(2), 74-80. <https://www.redalyc.org/pdf/3395/339529013002.pdf>

Bircan, T., y Salah, A. A. A. (2022). A Bibliometric Analysis of the Use of Artificial Intelligence Technologies for Social Sciences. *Mathematics*, 10(23), <https://doi.org/10.3390/math10234398>.

Camati, R. S., Scaduto, A. A., y Enembreck, F. (2021). Using the Projective Thematic Apperception Test for Automatic Personality Recognition in Texts. *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Australia*, 78-85. doi: 10.1109/SMC52423.2021.9659294.

Cañete, J., Chaperon, G., Fuentes, R., Ho, J., Kang, H., y Pérez, J. (2020, 26 de abril). *Spanish Pre-Trained BERT Model and Evaluation Data* [Presentación de artículo]. Practical ML for Developing Countries Workshop, Addis Aababa, Etiopía.

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Yat, Y., Shazeer, N., Prabhakaran, V., ... Fiedel, N. (2022). PaLM: Scaling Language Modeling with Pathways. *arXiv*. <https://doi.org/10.48550/arXiv.2204.02311>.

Devlin, J., Chang, M., Lee, K., y Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*. <https://doi.org/10.48550/arXiv.1810.04805>.

Diederich, J. (2023). *JobBot<sup>TM</sup> : Artificial Intelligence for Psychological Services*. Extraído el 21 de noviembre de 2023, de <https://jobot.ai/>.

Lei, Y., y Liu, Z. (2019). The development of artificial intelligence: a bibliometrics analysis, 2007-2016. *IOP Conf. Series: Journal of Physics: Conference Series*, (1168), 1-6. DOI: 10.1088/1742-6596/1168/2/022027.

Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., y Xie, S. (2022). A Convent for the 2020s. *arXiv*. <https://doi.org/10.48550/arXiv.2201.03545>.

McCulloch, W, S., y Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5, 115-133. <https://doi.org/10.1007/BF02478259>.

Manyika, J., y Hsiao, S. (2023). An overview of Bard: an early experiment with generative AI. *Google*. <https://ai.google/static/documents/google-about-bard.pdf>.

OpenAI. (2023a). GPT-4 Technical Report. *arXiv*. <https://doi.org/10.48550/arXiv.2303.08774>.

OpenAI. (2023b). DALL-E 3 System Card. *OpenAI*. [https://cdn.openai.com/papers/DALL\\_E\\_3\\_System\\_Card.pdf](https://cdn.openai.com/papers/DALL_E_3_System_Card.pdf).

Pan, T., Zhao, X., Liu, B., y Liu, W. (2022). Automated Drawing Psychoanalysis via House-Tree-Person Test. *2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), China*, 1120-1125. doi: 10.1109/ICTAI56018.2022.00171.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., y Chen, M. (2022). Hierarchical Text: Conditional Image Generation with CLIP Latents. *arXiv*. <https://doi.org/10.48550/arXiv.2204.06125>.

Rey, A., y Osterrieth, P. A. (1941). *Rey-Osterrieth Complex Figure Copying Test*. APA PsycTests. <https://doi.org/10.1037/t07717-000>.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424. doi:10.1017/S0140525X00005756.

Sinha, C., Cheng, A. L., y Kadaba, M. (2022). Adherence and Engagement With a Cognitive Behavioral Therapy-Based Conversational Agent (Wysa for Chronic Pain) Among Adults With Chronic Pain: Survival Analysis. *JMR Formative Research*, 6(5). doi: 10.2196/37302.

Spielberger, C. D., Gorsuch, R. L., y Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

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

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydin, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, A.,... Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50, 439-444. <https://doi.org/10.1093/nar/gkab1061>.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., y Polosukhin, I. (2017). Attention Is All You Need. En U. Von Luxburg, I. Guyon, y S. Bengio (Eds). *NIPS'17: Proceedings of the 13st International Conference on Neural Information Processing Systems*. Curran Associates, Inc.

Vidal Marciel, J. (2023). *Procesamiento del Lenguaje Natural (PLN) y Salud Mental: Afinando un Modelo BERT para Evaluar Ansiedad*. Trabajo de Fin de Máster. Máster en Investigación en Psicología, Universidad Internacioal de La Rioja.

World Health Organization. (2022). ICD-11: International classification of diseases (11th revision). <https://icd.who.int/>.

## Notas

1. Podría argumentarse que el ser humano depende de otros para ejercitar su voluntad plenamente. Si bien esto puede ser cierto a largo plazo, una persona sí es capaz de emitir acciones sin precisar de directrices de su entorno social.
2. *Casos de uso* es una expresión tomada del inglés (*use case*), muy utilizada en el contexto de ingeniería y ciencias de la computación. Se utiliza para denominar una serie de problemas paradigmáticos o utilidades para los que se ha construido una herramienta específica. Debido al carácter tan aplicado de estos campos, muchas de las innovaciones no se formulan desde lo que son en sí mismas, sino en función de su uso instrumental.
3. Esta es una taxonomía funcional, no hace referencia a que la IA también se puede clasificar en cuanto a la estructura de sus arquitecturas (ANN, RNN, CNN, etc.)
4. Un anglicismo del verbo *to ellicit*, empleado en psicología conductual y sinónimo de suscitar o provocar. Tiene una connotación más exacta para explicar dinámicas entre agente y ambiente, ya que sugiere que las consecuencias del contexto escapan al control del sujeto una vez emitida la acción.
5. Normalmente, cuando se habla de visión computerizada (*computer vision*) -la tecnología detrás del reconocimiento facial, por ejemplo- pareciera que la IA es



capaz de ver. Realmente, lo que ocurre es una codificación de los píxeles en diferentes capas en forma de matrices (capas convolucionales).

6. *Bot*, tomado también del inglés, es a su vez una abreviación de robot. Por el uso que se le ha dado a este término, parece que robot se atribuye a autómatas mecánicos, mientras que bot a sus versiones digitales. Actualmente, se usa, por ejemplo, para referirse a los sistemas de IA en los asistentes virtuales.

7. Aunque lo más probable, es que en la mayoría de los casos se generará una relación más simbiótica que competitiva con la IA. Al fin y al cabo, la gran mayoría de los profesionales combinan la función informacional con la relacional en su práctica terapéutica. Posiblemente, facilite el poder dedicarle más tiempo a lo vincular.

8. Esta ilustración hace referencia a una primera fase en la codificación de texto, denominada *word embedding*. En castellano, la traducción directa resulta algo extraña, porque *embed* significa algo parecido a incrustar o enclavar. En el caso de los transformadores ocurren otros procesos importantes con estos *word embeddings* -palabras codificadas como vectores-, principalmente el uso de matrices para determinar cuánta atención (*attention score*) debe recibir una palabra. Esto se realiza pasando los *word embeddings* por varias capas atencionales (*attention layers*), que asignan puntuaciones según las palabras que rodeen a cada *token* y de la posición que éste ocupe en la frase. Este proceso se llama *self-attention* o propio-atención y es bastante nuevo.