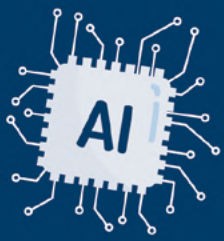
Colección Cuadernos de Cátedra

Lenguaje e inteligencia artificial: el desafío IA

M. Babio



Universidad Nacional del Centro
de la Provincia de Buenos Aires
Rector: Dr. Marcelo Aba
Vicerectora: Prof. Alicia Spinello

Facultad de Ciencias Sociales
Decana: Lic. Gabriela Gamberini
Vicedecana: Dra. María Luz Endere

Coordinación del Área Editorial
Edición: Lic. Carolina Ferrer y Dra. Ana Paula Alcaraz
Diseño y diagramación: Mario Pesci
Diseño de tapa: Soledad Rolleri

Esta obra contó con dos evaluaciones integrales independientes y su publicación fue avalada por Res.075/23 del Honorable Consejo Académico de la Facultad de Ciencias Sociales de la UNICEN.

En esta publicación se utilizan formas de lenguaje con enfoque de género, de acuerdo con la decisión de las y los autores (genérico masculino, forma doble -las/los). Esta toma de posición responde a la necesidad de visibilizar las tensiones a las que nos enfrenta el cambio social y al modo en que ellas se expresan en el lenguaje. Nos interesa visibilizar, en definitiva, el contradictorio y a la vez irrenunciable camino que conduce hacia la igualdad de géneros.

Babio, Marcelo
Lenguaje e inteligencia artificial: el desafío IA / Marcelo Babio. - 1a ed - Tandil: Universidad Nacional del Centro de la Provincia de Buenos Aires, 2024.
Libro digital, PDF
Archivo Digital: descarga y online

ISBN **978-950-658-623-2**

1. Inteligencia Artificial. 2. Ciencias Tecnológicas. 3. Ciencias Sociales. I. Título.
CDD 306.46

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Compartir Igual 4.0 Internacional

A mi hija Isabel y a mi suegro Manuel.

Contenido

Cuadernos de Cátedra es la primera colección que edita la Facultad de Ciencias Sociales de la Universidad Nacional del Centro de la Provincia de Buenos Aires. Se trata de una iniciativa impulsada desde el Área Editorial con el objetivo de promover la producción, circulación y acceso de contenidos universitarios a diversos lectores.

Lenguaje e inteligencia artificial: el desafío IA es el séptimo de la colección, reúne textos producidos desde la práctica docente sobre las implicaciones sociales de la Ciencia. Sus destinatarios son estudiantes y docentes universitarios y de otros niveles académicos, como así también aquellos lectores interesados en los debates que atraviesan la constitución del campo científico, sus problemas y tensiones en las Ciencias Sociales.

Introducción

La irrupción de la temática de las inteligencias artificiales puso en escena a dos protagonistas que mantienen una alianza estrecha y aspiran a desarrollar su visión del mundo: la ingeniería computacional y las ciencias cognitivas. Estas disciplinas intercambian modelos teóricos y conceptos, se confirman mutuamente y proponen una lectura acerca de lo humano y del entorno en que habitamos. Este bloque científico – tecnológico ha logrado una amplia difusión por parte del sistema de información global, captando la atención de las sociedades y de los individuos. Son protagonistas en el entorno comunicacional y desde ese lugar de autoridad se proponen como una fuerza transformadora capaz de orientar el destino del trabajo, de la educación, de la política, de la ecología planetaria, de las relaciones humanas y de la verdad. No nos puede extrañar que hayan desencadenado un fuerte debate, entre quienes les bajan el precio a las inteligencias artificiales (en adelante IA) considerándolas como meros artilugios técnicos y los que las consideran como evolución superadora de lo humano. Hay quienes las postulan como la solución potencial de los dilemas de la sociedad y quienes las sindican como causa posible de la desaparición de la especie humana en una hecatombe final.

Nuestra propuesta será la de focalizar en una serie de nociones propias del universo de las IA, tales como la inteligencia, el aprendizaje, la acción, el razonamiento, la memoria y la capacidad simbólica. A simple vista, estas nociones parecen sencillas y pueden darnos la impresión de que sabemos de qué hablamos cuando nos referimos a ellas. Esa es la magia del sentido común y del lenguaje natural. Este es el efecto imaginario que queremos colaborar en disipar. Cuando nos referimos a términos como inteligencia o lenguaje podemos aludir a entidades teóricas totalmente diferentes e incluso incompatibles. La elección del término "inteligencia artificial" ha sido una fuente de confusiones y malentendidos

Volver

que resulta conveniente señalar si queremos introducirnos en el significado de estas tecnologías.

A lo largo de los capítulos vamos a revisar y discutir nociones tales como la inteligencia, la capacidad de decisión, la atención y la memoria, aquellas facultades que intenta replicar la IA con el objetivo de diseñar agentes que puedan realizar tareas que requieran de una inteligencia similar a la humana. Nos vamos a encontrar con diferentes posturas, con debates y antagonismos entre corrientes de la psicología constructivista, el cognitivismo, la lingüística generativa y el asociacionismo. Cada una de estas corrientes nos permitirá acercarnos a la lógica de las IAs, comenzar a entender cómo aprenden, qué generan y cuál es su relación con una facultad como la del lenguaje, una característica que aprendimos a considerar como el sello distintivo de nuestra especie y nuestra cultura.

Capítulo 1. Inteligencia artificial

Sumario

Durante este capítulo introduciremos la definición de IA como diseño de agentes computacionales que puedan realizar tareas que impliquen capacidades similares a las de la inteligencia humana. ¿Cómo puede surgir algo inteligente de algo no inteligente? Esta es una gran pregunta que abren las IAs y para avanzar en la comprensión de este misterio vamos a visitar brevemente la teoría de la mente de Marvin Minsky. A continuación, presentaremos dos conceptos clave para la comprensión de la IA, el de función matemática y el de algoritmo. Las IA aprenden a identificar patrones al interior de grandes masas de datos y realizan predicciones a partir de ellos. Las computadoras sólo entienden de números, reciben inputs numéricos y entregan outputs numéricos. Puede parecernos que dialogamos con una IA a través del teclado o incluso la voz, pero esa es una ilusión; en esencia interactuamos con funciones matemáticas. El otro concepto que vamos a instalar es el de algoritmo. Un algoritmo es una receta que nos indica una serie de pasos y el orden en que debemos realizarlos. Los modelos de IA se diseñan según diferentes algoritmos que les permiten emular facultades humanas tales como la memoria, la atención, la inteligencia o el razonamiento. Para concluir vamos a clasificar los distintos tipos de IA y el alcance de cada uno de ellos.

Nos acercamos a la inteligencia artificial

Una estrategia que podemos adoptar para entender de qué se trata la IA: no dejarnos arrastrar por la vorágine informativa y el entusiasmo mediático. Un segundo consejo: no asumir que es necesario correr tras la técnica con la esperanza de mantenernos actualizados con los últimos adelantos y por ello dejar de lado los fundamentos conceptuales. Si nos dejamos llevar por la vorágine es fácil que terminemos confundidos por una gran cantidad de términos de divulgación mediática, los mensajes comerciales de las grandes empresas en la difusión de sus productos y, en definitiva, no terminemos de reflexionar sobre nuestro objeto de interés. Para nosotros el camino será otro, el de intentar acercarnos a los conceptos básicos de esta tecnología. Quién trata de asimilar el concepto, quien puede aproximarse desde el conocimiento, aunque sea parcial, de los fundamentos de una tecnología estará en mejores condiciones de vincularse con ella de manera más racional y manteniendo un mayor grado de independencia. Como corolario, sin duda, podrá operar y manejar mejor las herramientas. Porque, seamos claros al respecto, las IA, al menos por ahora, son eso: herramientas tecnológicas imaginadas, diseñadas y programadas por científicos humanos, entrenadas por humanos y sostenidas a partir de un inmenso esfuerzo económico y energético mundial.

¿Cómo evitar el furor entusiasta que desató el fenómeno ChatGPT y el halo de confusión que ha generado en torno a las IA? En este capítulo vamos a introducirnos en el ámbito de la IA y para ello les propondremos una estrategia: caminar unos pasos para atrás y situarnos justo en el momento en que aún no se había desencadenado el furor global. Es decir, nuestra propuesta será la de analizar de qué se trata la IA y para ello partimos de un momento anterior al lanzamiento de ChatGPT. De hecho, podríamos situarnos en un punto muy anterior de la curva de evolución de la IA, en sus comienzos hacia 1950 con el Test de Turing o la mítica conferencia de McCarthy en Dartmouth en 1956 donde se etiquetaron por primera vez estas tecnologías como “inteligencia artificial”. Minsky, Shannon, Solomonoff, McCarthy y muchos científicos relevantes participaron de este evento que puso nombre de inteligencia a estas

máquinas de cálculo. Varios científicos detestaron la denominación por lo equívoco de un término que, debemos aceptar, ha sido fuente constante de confusiones, desvaríos imaginativos y deslizamientos ideológicos. De hecho, existen numerosos trabajos que ilustran al detalle esta historia, es por ello que decidimos no sumar a un espacio tan bien documentado y utilizar otra periodización. Elegimos ubicarnos hacia fines de 2021, es decir, nuestro grado de conocimientos corresponde con principios de 2022. Para definir con mayor detalle cuál es el estado de la cuestión en el momento en que nos situamos, referimos al informe 2021 sobre IA de la Universidad de Stanford (Zhang, 2021). Hacia esa fecha los pilares de la IA que fundamentan su estado actual estaban maduros a partir de varios hitos, el algoritmo de *backpropagation* en 1986 (Hinton, 2022) la publicación del trabajo sobre la técnica de procesamiento de lenguaje natural (PNL) Word2vec en 2013 (Mikolov et al., 2013), el paper *"Attention is all you need"* de 2017 en que se presenta en sociedad el modelo Trasnformer (Vaswani et al., 2017) y el entrenamiento de los grandes modelos de lenguaje (LLMs) de Google (Bert) y de Open AI (Gpt) hacia 2018 aproximadamente. Estos grandes hitos ponen las bases técnico científicas y de implementación que permitirán el despegue descomunal de las IA, en particular sus modelos generativos de lenguaje. Desde allí en adelante la curva de desarrollo se torna exponencial, potenciada por un gran protagonista que suele pasar desapercibido que es el desarrollo de hardware, en particular la producción masiva de microprocesadores GPU y TPU en el equipamiento de los inmensos data centers de empresas como Google u Open AI. Esta segunda usó el data center de Microsoft Azure lo que tendrá consecuencias en términos comerciales y de difusión del desarrollo. Existen dos formatos básicos de desarrollo, el propietario, en el cual el desarrollador no publica su código y el de base abierta en el que se pone a disposición de la comunidad el código de los programas. A partir de su vinculación con Microsoft, Open AI el desarrollador de Gpt perderá su fundamento de transparencia y dejará de ser de base abierta, es decir, ya no será tan "open". Si bien la carrera acelerada ya se había iniciado años atrás y las IA ya habían impactado en términos de interés académico, todavía no se habían transformado en los grandes protagonistas en que se han convertido al día de hoy. Es decir, aunque la curva de desarrollo de

las IA ya había despegado en términos exponenciales y las bases técnicas estaban maduras, la opinión pública no había recibido el bombardeo mediático del *"hipe"* ChatGPT. En este capítulo aún no vamos a introducir el tema de los modelos conversacionales que tanta atención han captado, preferimos concentrarnos en varias nociones y conceptos que nos introduzcan en la lógica de operación de las IA, una base útil para luego entender el sentido de los generadores de texto e imagen con interfases conversacionales en lenguaje natural.

Definimos la IA como el estudio de los agentes que reciben percepciones del entorno y llevan a cabo las acciones.

P. Norvig y S. Russell, 2004, p. 19.

Una definición de Inteligencia Artificial

Para comenzar vamos a adoptar una definición de IA y para ello recurrimos a un texto clásico con el que se han formado gran parte de los profesionales del área en sus licenciaturas y posgrados, el texto de Russell y Norvig, "Inteligencia artificial, un enfoque moderno". Tengamos en cuenta la diversidad que cubre el campo de la IA, incluyendo el reconocimiento de imágenes, el procesamiento del lenguaje natural, la robótica, la conducción autónoma, entre otros dominios de interés y aplicación. No obstante, lo que la define no es el campo de aplicación sino la manera de abordar el dominio problemático. La epistemología nos enseña que no debemos tratar de comprender una disciplina por los objetos que analiza, sino por los abordajes y problemas que formula (Bachelard, 2000). En este sentido, el campo se caracteriza por la implementación de determinados abordajes lógicos, matemáticos, de aprendizaje computacional, de diseño de sensores para captura de datos (percepción), la utilización de modelos y analogías inspiradas en la biología y muchas otras herramientas técnicas y conceptuales cuya exploración profunda excede el objetivo de nuestro trabajo. Intentaremos, eso sí, presentar algunas de estas nociones y desarrollos de manera de poder entender con qué tipo de cuestiones nos encontramos en una sociedad en que la IA se hace cada vez más gravitante. Lo primero será focalizarnos en la noción de IA para no perdernos en la selva de aplicaciones ni reducirla a un tipo particular de herramienta o, lo que es más limitativo, identificarla con algún producto comercial. Entre esta diversidad de enfoques y sub campos de desarrollo, los autores encuentran un concepto articulador, el de "agente inteligente". Desde este eje del "agente inteligente" retomamos la siguiente definición de IA que nos servirá de escalón inicial para el desarrollo del capítulo.

> *"El principal tema unificador es la idea del agente inteligente. Definimos la IA como el estudio de los agentes que reciben percepciones del entorno y llevan a cabo las acciones. Cada agente implementa una función la cual estructura la secuencia de las percepciones en acciones; también tratamos las diferentes formas de representar estas funciones, tales como*

> *sistemas de producción, agentes reactivos, planificadores condicionales en tiempo real, redes neuronales y sistemas teóricos para las decisiones."* (Russell y Norvig, 2004, p.19)

Un agente es cualquier entidad capaz de percibir su medio ambiente mediante sensores y actuar en consecuencia por medio de algún dispositivo denominado "actuador". Para visualizar de qué se trata un agente pensemos en algo tan sencillo como el termostato de un aire acondicionado. Este artefacto se puede regular a determinados niveles térmicos que deseamos mantener y contará con sensores que monitorean la temperatura ambiente. Cuando esta atraviese los límites de determinados valores térmicos se encenderá y cuando vuelva a refrigerar o calefaccionar el aire a la temperatura adecuada se apagará. Tiene un determinado objetivo, percibe el medio ambiente y acciona en consecuencia. No obstante, no diríamos de él que es un agente inteligente en el sentido de una IA y menos aún de un ser humano. Entonces, ¿Qué son estos "agentes inteligentes" a los que alude la definición de IA? Se trata de entidades computacionales con varias capacidades, en primer lugar, la de tener una percepción del entorno, de percibir su medio ambiente mediante algún tipo de sensores y, en segundo lugar, actuar de manera racional en función de estas percepciones. Prestemos atención a dos nociones, percepción y acción. Entre ambas, la percepción y la acción se sitúa otro elemento clave, la racionalidad. Las acciones seguirán una orientación racional de respuesta a partir del estado percibido en el entorno. Ellas se enlazan en una secuencia que comienza con la percepción y que culmina con una acción. Esto nos remite a la conocida idea de input y output. De manera simplificada podríamos pensar que, a partir de un input vinculado a la percepción del ambiente, se genera un output en forma de una acción. Entre la percepción del ambiente y la acción existe una instancia intermedia a la que se alude como "función".

The Imitation Game.

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'.

A. Turing, 1950, p. 433.

Las inteligencias artificiales intentan imitar la mente humana

El término "inteligencia artificial" se vincula con la expectativa de los científicos del campo de las ciencias de la computación de generar artefactos que puedan imitar la manera en que los seres humanos razonan a partir de datos extraídos del medio ambiente. Para ello deben aprender a extraer patrones, a establecer regularidades en un medio ambiente que logran percibir a través de algún sensor. Cuando nos referimos a sensores podemos pensar en nuestros sentidos. Ellos nos permiten captar el mundo que nos rodea, verlo, palparlo, escucharlo, saborearlo. La biología ha desarrollado complejos sistemas perceptivos a través de intrincados mecanismos de adaptación evolutiva al entorno. Nuestros dispositivos de registro son fruto de millones de años de evolución de la materia viviente. Pero nuestra capacidad de vincularnos con el entorno y de percibirlo no termina allí, también estamos dotados de la aptitud de percibir significaciones en el plano simbólico porque somos poseedores de la facultad del lenguaje que potencia nuestras posibilidades de manipular el ambiente de manera plástica y creativa. ¿Las máquinas pueden imitar o incluso igualar estas capacidades? Esta pregunta se ubica en el centro de una polémica encarnizada con defensores y detractores. Aunque el sistema mediático pueda dar la impresión de que se ha alcanzado estos objetivos de igualar las facultades humanas, por el momento se trata de una aspiración más que de un logro. Esto no implica desconocer que los avances en ese sentido son impactantes y en gran medida desconcertantes. Aún no existen respuestas más allá de la futurología.

Sentimos que es necesario incluir otra distinción que hace a la filosofía de la percepción y que abre un debate respecto del alcance de las similitudes entre las facultades humanas y las de la IA. Existe una diferencia entre ver y mirar, tal como existe una diferencia enorme entre oír y escuchar (Berger, 2009). Mirar es un acto humano lleno de intención y atravesado por el vector cultural, por el deseo y el proyecto vital. Es y será algo totalmente ajeno a la dimensión mecánica de una IA. Las capacidades de percepción, el desarrollo de sensores que permiten la captación

del ambiente de las IA se desarrolla a ritmos acelerados, así podemos decir, en un sentido figurado, que las IA pueden ver o que pueden oír, no obstante, no podemos eludir la necesidad de cuestionar su capacidad de mirar o de escuchar. En todo caso su mirada nunca será la nuestra, en todo caso su escucha nunca será la nuestra. Incluso si lograsen aprender patrones que le permitieran emular la capacidad de mirar o escuchar seguiría existiendo una diferencia de nivel e índole que mantendría apartados ambos mundos, el humano y el artificial. Nuestra capacidad de mirar es sumamente plástica y se enlaza con la historia de la técnica y las manifestaciones simbólicas. Nuestro arte de mirar mutó cada vez que la técnica amplió el horizonte de percepción. El microscopio nos abrió la dimensión de lo pequeño y el telescopio la de la inmensidad. La fotografía enlazó la mirada, la memoria y el recuerdo en un nudo emotivo y material. La IA será, no lo duden, parte de la mutación del espíritu humano en esta historia del mirar y del oír. El camino de estas reflexiones nos llevaría hacia horizontes filosóficos que dejaremos para otro libro, por ahora nos vamos a restringir a la cuestión de la manera en que las IA toman por modelo la mente y las capacidades humanas de pensamiento e inteligencia.

Figura 1. Alan Turing, considerado el padre de la computación moderna. En su célebre artículo de 1950 "Computer Machinery and Inteligence" postuló las condiciones que debiera cumplir una máquina autómata para ser valorada como análoga a la inteligencia humana. Su célebre test de Turing, establece que una máquina podrá ser considerada inteligente si logra engañar a un evaluador en cierto porcentaje de aciertos durante 5 minutos de preguntas y respuestas haciéndole creer que se trata de un humano. Podemos decir que abrió la caja de pandora. Se suicidó en 1954 luego de haber sido perseguido y castrado químicamente según las brutales regulaciones homófobas vigentes en su país y en su época. Foto tomada de https://www.aps.org/publications/apsnews/201401/physicshistory.cfm

Volver

¿Cómo puede surgir la inteligencia de algo no inteligente? Para hallar una respuesta, demostraremos que es posible construir una mente a partir de muchas partes pequeñas que en sí mismas no la poseen. Llamaremos "sociedad de la mente" a este modelo, según el cual cada mente está formada por numerosos procesos más pequeños. Daremos el nombre de agentes a estos procesos.

M. Minsky, 1986, La sociedad de la mente, p. 16.

En varios pasajes del libro señalamos nuestra capacidad de percibir patrones, de identificar regularidades y de actuar en función de ellas. A partir de la percepción del entorno, los agentes inteligentes pueden decidir y en consecuencia de ello actuar. Percepción, razonamiento, toma de decisión y acción forman una secuencia que debemos explorar si queremos introducirnos en este mundo en que la computación simula la manera en que los seres humanos nos vinculamos de forma inteligente con el medio ambiente. Antes de avanzar queremos aclarar que "percibir", "razonar", "decidir" y "actuar" son términos que pueden llevar a confusiones dado que las entidades a las que nos referimos, humanos y máquinas son muy diferentes. Marvin Minsky señaló la necesidad de ensamblar la memoria, el razonamiento, la deliberación para la acción, la posibilidad de procesar las informaciones del ambiente y muchos otros pequeños y grandes elementos que se ponen en juego para hacer posible cualquier acción de la mente que podamos considerar inteligente. A estas instancias que participan de cualquier acción humana que involucre la inteligencia y que suponga una mente los llama agentes. En "La sociedad de la mente" (Minsky, 1986), un libro que aconsejamos leer a todos los que se interesen por la relación entre la IA y la mente humana, nos brinda un ejemplo que vale la pena recordar. Afirma que para realizar cualquier tarea se ponen en juego muchísimos entes que funcionan en conjunto como una sociedad, de allí el título de su libro "la sociedad de la mente". Nos parece que cuando realizamos acciones somos nosotros los que las llevamos adelante. Tenemos la impresión de ser uno quien piensa y ejecuta las acciones. Pero, se pregunta, ¿qué es "uno"? Aquí entra a jugar la sociedad de esos pequeños entes que conforman la mente. Minsky nos invita a realizar un sencillo experimento en el que nos ayuda a entender la idea de sociedad de la mente (Minsky, 1986). En este experimento se nos ordena: ¡levante una taza de té!

Sus agentes de ASIR desean sostener la taza.

Sus agentes de EQUILIBRIO desean evitar que el té se derrame.

Sus agentes de la SED desean que usted beba el té.

Sus agentes de MOVIMIENTO quieren que usted acerque la taza a los labios.

Cada uno de estos agentes se ocupará de sus tareas y no tendremos conciencia de ello mientras tomamos el té y conversamos tranquilamente. Es más, ASIR no se ocupará de SED, EQUILIBRIO no se ocupará de MOVIMIENTO, ninguna de las pequeñas entidades necesita ocuparse de las otras ya que todas actúan en concierto en una alegre sociedad. Sólo cuando alguna de las capacidades se ve trastornada por alguna patología, una lesión cerebral o algo menos dramático como la ebriedad o un sobresalto, tomaremos conciencia de ellas. Pero esto no termina aquí, cualquiera de esos agentes involucra a la vez una multiplicidad de otros pequeños agentes. ASIR deberá controlar la rotación de la muñeca, controlar el peso relativo para ejercer compensaciones, la posición de la espalda, la forma en que se ha sentado, la lista es enorme. Qué desafío implica tratar de explorar cada una de estas entidades y mucho más aún penetrar el misterio de su ensamble en una sociedad que funciona de manera tan asombrosa y eficiente. Nuestro conocimiento de la manera en que operan estos conceptos en ambos mundos, el de las computadoras y el de los seres humanos tienen algo en común: la dificultad para terminar de entenderlos en profundidad.

Sería ilusorio suponer que comprendemos cabalmente qué es y cómo funciona la inteligencia humana, de la misma manera en que nos cuesta penetrar en los vericuetos de la IA. Tal es así, que en el campo informático y en la psicología del comportamiento existe un término para denominar aquellos modelos cuya manera de inferir nos resulta opaca: modelos de "caja negra" (Mandler, 2007). Esta noción tiene raíces en el conductismo y consiste en introducir una instancia intermedia entre el estímulo y la respuesta. Entre input y output se supone algún tipo de función o funciones no accesibles que procesan la entrada y generan la salida. Este desconocimiento sobre el carácter de las funciones ocultas en la instancia de caja negra ha sido utilizado en muchos ámbitos, entre ellos la cibernética y la psicología de la conducta. Hablar de caja negra implica aceptar que los procesamientos digitales en algunos de los modelos de IA son inaccesibles o, al menos, de muy compleja reconstrucción. Sabemos que funcionan, cómo se han diseñado y con qué materiales los entrenamos, pero en gran medida la forma en que operan es un misterio para la ciencia aún en su grado más alto de conocimiento (Prince, 2024). Con

la inteligencia y la capacidad de razonar humana ocurre otro tanto. La lógica, una gran herramienta de la ciencia, describe la formalización, la reconstrucción deductiva sobre la manera de razonar y establecer inferencias a partir de datos, no obstante, es eso, una reconstrucción racional y no la descripción última de la manera en que funciona el entendimiento humano. Otro tanto ocurre con los modelos explicativos que nos proponen la psicología constructivista o la neuro psicología asociacionista tal como veremos en capítulos destinados a distintas perspectivas sobre el lenguaje y la inteligencia. Los límites de la razón y su ensamble con la acción humana son otro factor problemático, dado que las bases racionales no siempre indican cuál es el camino adecuado para definir una acción, en muchas, sino en la mayoría de las circunstancias a las que nos enfrentamos, la toma de decisión se fundamenta en criterios inciertos y ambiguos. Esta cuestión es retomada y elaborada a nivel conceptual y técnico por las disciplinas computacionales. La noción misma de inteligencia nos abre muchas disyuntivas y paradojas, no podemos terminar de definirla conceptualmente ni en el caso de las entidades humanas ni en las computacionales. Quizás ambas cuestiones permanezcan impenetrables a nuestro conocimiento, tal como lo postula Chomsky (Chomsky, 1986). Para decirlo claramente, la IA es un intento de replicar procesos de la percepción, el entendimiento y la acción humana, aun cuando no lleguemos a entender cabalmente cómo operan estas capacidades en el plano biológico o psicológico. Para decir que un programa piensa como un ser humano debemos contar con un mecanismo para entender cómo piensan los humanos. Tal como dicen Russell y Norvig, es necesario penetrar en el funcionamiento de las mentes humanas (Russell y Norvig, 2004). De esto se ocupan las ciencias cognitivas que trataremos en un capítulo posterior. Estos “agentes inteligentes” deben tener características que los distingan de otros tipos de programas que también pueden sacar conclusiones y actuar a partir de datos. Un agente racional es aquel que actúa con la intención de alcanzar el mejor resultado o, cuando hay incertidumbre, el mejor resultado esperado. Algunas de estas características son los controles autónomos que les permiten a los modelos de IA operar y tomar decisiones sin que medie la acción directa de los seres humanos,

contar con la capacidad de percibir su entorno, persistir durante un tiempo prolongado, acceder a informaciones relevantes que le permitan sacar inferencias y ser capaces de alcanzar objetivos (Russell y Norvig, 2004). A su vez, al tratarse de entidades inteligentes, deben poder responder a entornos cambiantes, es decir, deben tener altos niveles de plasticidad. Como vemos, los requisitos son muchos y muy complejos de alcanzar mediante algoritmos computacionales y sostenidos no ya en nuestras estructuras de carbono sino en soportes de silicio propios del hardware.

La inteligencia artificial permite predecir a partir de modelos

Nosotros, los seres humanos vivimos en un universo caótico, complejo y en constante cambio, sin embargo, logramos orientarnos y sobrevivir adaptando nuestras acciones a condiciones tan inciertas. Esta afirmación propia de las disciplinas cognitivas y fácil de aceptar desde nuestro sentido común es retomada como fundamento de las disciplinas computacionales. Sin la capacidad de predecir la manera en que se encadenan los estados de la realidad que nos rodea nos sería imposible subsistir como individuos e incluso como especie. Nuestra sensación de familiaridad con el mundo que nos rodea, la idea de que "sabemos" cómo funciona nuestro entorno se debe a que contamos con imágenes simplificadas de la realidad. A estas imágenes conceptuales simplificadas las llamaremos "modelos". A pesar del caos del entorno y de que vivimos rodeados de ruido, nuestras facultades nos permiten dar sentido al mundo. Dado que los modelos nos permiten comprender la realidad podremos utilizar ese conocimiento para orientar nuestra acción. Tendremos la oportunidad de tratar esta noción con mayor profundidad desde la perspectiva del estructural constructivismo de J. Piaget y del cognitivismo de S. Dehaene, del constructivismo social de Vigotsky y del enfoque neuropsicológico de Luria. Cada uno de ellos profundiza en ciertos aspectos de la inteligencia humana y la manera en que nos permiten el ajuste con el entorno. Podemos afirmar, con algunas salvedades, que los autores coinciden en postular que la inteligencia permite extraer de la

realidad ciertos rasgos obviando otros como poco significativos. Contar con modelos implica registrar y descartar. Podemos quedarnos con aquello que resulta significativo porque podemos obviar y dejar de lado aquello que no nos resulta relevante. Tendemos a la simplificación, pero ¿en qué medida simplificamos? Para entender la idea de simplificación y de modelo nos podemos servir de la idea de mapa como representación de la complejidad infinita de un territorio. Pensemos en los distintos tipos de mapas de una región, los políticos que marcan las fronteras, los físicos que nos muestran el relieve del terreno, los mapas de destino que muestran carreteras y medios de transporte, los grabados antiguos con que decoramos ambientes, cada uno de los cuales tendrá su utilidad. Uno no es, en sí mismo, mejor o peor que otro sino en base al propósito al que lo destinemos. Como pudimos ver, existen muchas formas de abordar el tema de la manera en que se constituye la inteligencia humana y de la forma en que podemos concebir su estructuración en el plano neuropsicológico, estructural o psico social. En nuestro caso, el de la IA, vamos a ocuparnos de una forma en particular de constituir estas representaciones, la probabilística.

La probabilidad implica establecer inferencias, sacar conclusiones a partir de datos (Canavos, 1998). Es una forma de tornar predecible un conjunto de datos a partir de detectar patrones de relación entre los mismos. Un modelo de este tipo establece la probabilidad de un resultado a la luz de un criterio o varios criterios. En alguna medida, nuestras elecciones de vías de acción estarán orientadas por la certeza que nos proveen nuestras anticipaciones, los patrones que creemos detectar en el entorno en que nos movemos. Resulta razonable suponer que, en función de las horas de estudio que le destinemos a la aprobación de una materia crecerá o decrecerá la probabilidad de aprobarla. No obstante, no deberíamos dejarnos llevar por la falsa sensación de certidumbre que nos deja nuestro sentido común, ya que el movimiento de los fenómenos aleatorios del entorno y nuestras inferencias con relación a los mismos responden a leyes de probabilidad que en gran medida se nos escapan (Canavos, 1998). Decimos "la probabilidad" y no la certeza. Esto nos resulta claro si pensamos en el ejemplo del examen. Los factores en juego y sus relaciones pueden revestir altos niveles de complejidad. La experiencia en

el mundo académico nos indica que múltiples factores podrán incidir en el resultado de un examen, incluso el capricho del docente evaluador. La capacidad humana de generalizar, anticipar y calcular probabilidades a largo plazo es asombrosa, piensen en la enorme cantidad de estimaciones a largo plazo que implica llevar a cabo la proeza de obtener un título universitario. Somos extraordinariamente eficientes en este sentido, no obstante, también podemos sostener que somos muy ineficaces en la estimación de probabilidades. El mundo silvestre de nuestra experiencia cotidiana está lleno de incertidumbre y nuestra capacidad natural para estimar las probabilidades es limitada incluso, en algunos contextos, podría calificarse de desastrosa. En muchos casos nuestras intuiciones y expectativas desestiman factores relevantes (como la distribución en las poblaciones y las consideraciones derivadas del teorema de Bayes). Adoptar el camino de la probabilidad implica aceptar este destino, renunciar al camino de la certeza. Un enfoque probabilístico es un atajo, una cierta renuncia a encontrar el sentido último de las cosas. El camino de la probabilística ha dado un gran impulso a la tecnología de la IA, sin ella su desarrollo actual resulta inconcebible. Ni mejores ni peores, distintos, el camino de nuestro razonamiento, el humano fruto de la historia biológica y del ambiente social tiene sus características y el que hemos logrado programar en las máquinas otras. Anticipamos que hacia fines de 2023 se están desarrollando avances, al parecer significativos, en algoritmos de razonamiento en las IA (Lightman et al., 2023). No obstante, hacia diciembre su verdadero alcance es una incógnita. Por ahora nos vamos a acercar a principios muy básicos y un buen comienzo es ocuparnos del concepto de función.

Lo difícil es percibir la falta de fundamentos de nuestra creencia.

L. Wittgenstein, 1998 [1969], Sobre la certeza. 24.c.

El concepto de función

Una función involucra dos conjuntos y establece una relación entre ellos. A uno de ellos lo llamaremos dominio o conjunto de partida y al otro codominio o conjunto de llegada. La utilización del concepto de función es *"muy amplia y flexible en la matemática actual"* (Rabuffetti, 1997, p. 47). Hay muchos tipos de funciones, algunas muy sencillas y otras muy complejas, en nuestro caso no nos interesa introducirnos en sus laberintos, sino captar una idea muy general. En el conjunto de partida se encuentra una serie de valores posibles para un conjunto de elementos y en el conjunto de llegada otro conjunto de valores para otro conjunto de elementos. De un lado, el de partida, tenemos las variables independientes (las famosas "x") y en el otro conjunto se colocan las variables dependientes (las famosas "y"). De manera muy general se establece que los valores o magnitudes del primer conjunto se relacionan con los valores o magnitudes del segundo. De esto se trata la "relación" entre uno y otro conjunto.

Mediante el establecimiento de determinadas funciones podemos predecir y en algún caso comprender cuestiones vinculadas con el entorno. Si pensamos en términos de funciones, asumimos que los datos de los que partimos son de tipo numérico. Vayamos a un ejemplo, podemos suponer que, si tratamos de establecer la probabilidad de que una persona realice tareas domésticas no remuneradas, se podrá establecer una relación entre la cantidad de horas que dedica a las mismas y otras variables (agradecemos este ejemplo al gran equipo de docentes de ciencias sociales computacionales del IDAES). Algunas relaciones nos resultan bastante obvias como por ejemplo la casi certeza de que, en el estado actual de las relaciones sociales en la sociedad argentina, el sexo de la persona influye en la posibilidad de que destine mayor o menor cantidad de horas a trabajar en su hogar sin obtener remuneración por ello. Podemos suponer que la mujer tiene mucha mayor posibilidad de realizarlas y el hombre menor. También se nos puede llegar a ocurrir que la edad será una variable relevante, así si un integrante del hogar tiene menos de cinco años o más de noventa y ocho daremos por hecho que habrá gran probabilidad de que no realice este tipo de tareas. Ahora

bien, a pesar de que las cuestiones a las que nos referimos pertenecen al mundo cotidiano en el que tendemos a percibir distribuciones de roles en las parejas, las remuneraciones en una sociedad, la capacidad de los niños pequeños o de las personas de edad muy avanzada, nada de ello es comprendido ni percibido por una IA. Evitemos naturalizar las operaciones de las IAs. Las funciones no operan con personas, ni dinero, ni con otros elementos de nuestro mundo cotidiano, operan con números. Cuando decimos que una IA opera aproximando funciones (interpolando o extrapolando) tenemos que tener siempre presente que los datos de entrada no serán elementos del mundo cotidiano sino números y que los resultados que arroje en sus inferencias serán también números. Si no nos acordamos de ello nuestra comprensión estará nublada y será sencillo que naufraguemos en un océano de narrativas y construcciones imaginativas de toda índole. Salgamos de las películas de ciencia ficción y volvamos a las IAs con la idea de función en nuestro arsenal.

La IA puede ayudarnos a establecer conjeturas de este tipo, es decir, determinar la relación entre distintos grupos de fenómenos a partir de probabilidades. De hecho, la forma en que se relacionan diferentes variables con que podemos describir los eventos de la realidad (fática o ideal) pueden adoptar muchas formas. Algunas de las maneras en que se pueden expresar estas relaciones pueden ser sencillas y otras muy complejas, no ingresaremos en la cuestión de las funciones lineales y no lineales, de la interpolación o la extrapolación, pero retengamos que el campo de las funciones es muy basto. Las tareas a las que se destinan las funciones en el campo de la IA son muy amplias, pero nosotros nos vamos a acercar al mismo de manera muy general. Pensemos en que tenemos un conjunto de datos y queremos entrenar a un modelo de IA para que detecte la manera en que los mismos se relacionan entre sí, que aprenda a detectar sus patrones. A partir de ese universo de relaciones que ha determinado, pretendemos que prediga la probabilidad de ocurrencia de un evento ante datos que no se le han sido presentados con anterioridad. Por ejemplo, pensemos que contamos con datos de una población de alumnos de distintas carreras, tales como la zona de residencia, la edad, el NSE, el nivel de estudios y profesión de los padres, los consumos culturales, la secundaria en que han estudiado. Supongamos también que

pudimos establecer algún tipo de relación estadística entre estos factores y la elección de carreras que han realizado, el modelo de IA nos permitirá predecir cuál es la probabilidad de que un nuevo caso, que nunca ha registrado, elija una u otra alternativa de estudio universitario.

Recordar esto es importante pero no suficiente para comenzar a entender qué es la IA. Para ello deberemos abordar algunas cuestiones con relación a la forma en que la IA determina la probabilidad de un evento a partir de datos de partida y más aún, entender cómo aprende a hacerlo. En principio contentémonos en retener la idea de que los agentes racionales con los que tratamos en IA se enfrentan a una serie de percepciones sobre el estado en que se encuentra el medio ambiente con el que se vinculan y que en cada una de las instancias debe decidir qué acción es la más eficiente para cumplir las funciones para las que está diseñado. En el caso anterior, su acción consistirá en arrojar una probabilidad de que un caso con determinadas características elija o no una carrera. Pero pensemos en otro campo, el de la robótica y veamos cómo nos ayuda a entender este mundo de la IA. Pongamos el caso de una de las aspiradoras robot de las que arman mapas inteligentes. Veamos cómo actúan en el proceso de limpiar un ambiente, ejemplo típico en la enseñanza de los principios de IA (Figura 2). Quienes quieran profundizar pueden remitirse al texto mencionado de Russell y Norvig (2004). Destacamos la idea de que el proceso de barrido implica una secuencia que se extenderá en el tiempo e implicará cambios de estado. Lo primero que deberá hacer es percibir el ambiente y dividirlo en cuadrículas mediante sensores, para ello cuenta con un programa. A la vez, esta información quedará registrada y esto le permitirá definir sus desplazamientos por el espacio que debe barrer. Imaginemos que el ambiente queda dividido en cuadrículas. Nuestra aspiradora robot deberá identificar mediante sensores adecuados si hay suciedad en la primera cuadrícula, si la detecta deberá aspirarla, si no la detecta se moverá hacia la izquierda o la derecha, adelante o atrás, de acuerdo a su localización y repetirá el ciclo hasta que el ambiente quede totalmente despejado. Es decir, en cada una de las instancias deberá elegir entre una serie de alternativas y tomar la mejor decisión. Este proceso se podría ejemplificar mediante una tabla en que figuran las alternativas

que debe tomar el agente y valorar la eficacia con que se desempeña el robot para obtener su objetivo final. Si, por ejemplo, se encuentra en un cuadrante y se mueve hacia otro cuadrante que ya ha barrido su acción puede ser considerada poco eficiente. De hecho, podemos pensar la evaluación de su eficacia en términos del tiempo que tarda en cumplir la tarea y el número de movimientos que utiliza para hacerlo. Esta idea de medir la eficacia en términos económicos se puede vincular con la idea de gasto energético y optimización de recursos. En el fondo subyace una noción que ha abierto varios problemas relevantes en el campo de la matemática: cómo resolver algo utilizando la menor cantidad de recursos es algo que ha interesado desde siempre a este gremio. Euclides, Arquímedes, Descartes, Gauss, Poncelet, Steiner, Hilbert, la lista de matemáticos ilustres que se han ocupado de estas cuestiones es enorme. En un principio la aspiradora robot comenzará a moverse al azar por el ambiente y su accionar nos parecerá torpe, pero conforme su experiencia crezca sus movimientos se irán perfeccionando hasta llegar a una solución que optimice sus desplazamientos según un esquema que pueda representarse como el más eficaz. Habrá encontrado una solución al tema de nuestro living, esquivando los obstáculos y rodeando las patas de los muebles. Un robot aspirador, un dispositivo bastante simple, ha emulado la acción humana de aspirar, mediante un programa que le permitió encontrar una solución inteligente en función del objetivo con el que se la ha programado, es decir, en forma racional en términos de tiempo y esfuerzo.

Nos falta, sin embargo, dar un paso más para comprender de qué se trata la IA y tiene que ver con la plasticidad, la capacidad de ajustarse a ambientes cambiantes. Aprender implica generalizar, poder aplicar soluciones incorporadas anteriormente a situaciones nuevas. Si cambiamos los muebles de lugar, la aspiradora robot debe poder reconfigurar la representación del ambiente para adaptarse a la nueva situación que se le plantea. Este es uno de los desafíos que nos presentan nuestros modelos de aprendizaje artificial. Para que una IA pueda identificar patrones debemos entrenarla a partir de datos lo que le permite constituir un modelo desde el que pueda predecir. Si tengo un grupo de datos con los que entreno a mi modelo debo cuidarme de que mantenga la

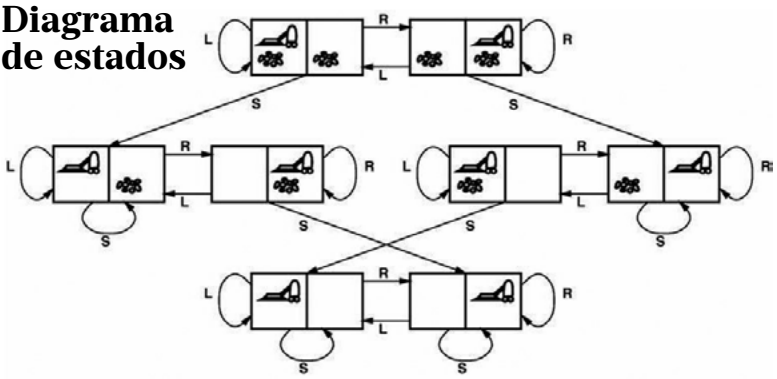


Figura 2. Soluciones de problemas mediante búsqueda. En el diagrama se observan los posibles desplazamientos de la aspiradora robot para lograr el objetivo final de limpiar el ambiente en el menor número de pasos. Publicada por Pastor Ledezma 2005 https://slideplayer.es/slide/92150/

suficiente elasticidad como para predecir más allá de los datos con que se lo alimenta, este es uno de los temas que más preocupan al momento del ajuste de los modelos, volveremos sobre ello más adelante.

El algoritmo propone los pasos posibles para solucionar un problema

En este capítulo nos ocupamos de las IA, un terreno de investigación en el cual se trata de replicar las facultades humanas y para ello se recurre a nociones matemáticas que permitan diseñar complejas arquitecturas capaces de ello. Así se recurre al cálculo matemático, al álgebra y la probabilística, a la teoría de la decisión, a los desarrollos de la teoría de grafos entre otros campos de las ciencias matemáticas que han sido capitalizados con enorme éxito. La IA opera mediante ellos y con el esfuerzo de la ingeniería de software que permite programar soluciones aptas para correr en los soportes físicos que provee el hardware.

Mencionamos el auxilio que presta la matemática en el diseño de estas soluciones de IA y cada vez que entramos en contacto con el tema surge esta idea de algoritmos. Llegó el momento de ocuparnos de este concepto tan presente en nuestro discurso cotidiano. ¿Qué es un

algoritmo? Para conocer acerca de ellos recurrimos a una de las autoridades en el tema Panos Louridas profesor asociado del MIT. En su libro "Algoritmos" (Louridas, 2020) advierte que, por esa fecha, es común decir que se vive en "la era de los algoritmos". Ha pasado, nos dice, la "era de las computadoras", la "era de internet" y, por esos tiempos hace muy pocos años se podía decir eso, que nos situamos en el reino algorítmico. Es posible que a 2023 se diga que vivimos en "la era de la inteligencia artificial" y que los algoritmos hayan sido relegados al rango de herramientas que permiten el advenimiento del nuevo protagonista la IA.

*An **Algorithm** is a set of rules that a machine follows to achieve a particular goal. An algorithm can be considered as a recipe that defines the inputs, the output and all the steps needed to get from the inputs to the output. Cooking recipes are algorithms where the ingredients are the inputs, the cooked food is the output, and the preparation and cooking steps are the algorithm instructions.*

Ch. Molnar, 2021, Interpretable Machine Learning. A Guide for Making Black Box Models Explainable, p. 12.

Antes de entrar en la definición de algoritmo quisiéramos resaltar lo amplio del campo que abarcan. Podemos encontrarlos en todas partes, incluso podríamos decir que la caja de brownies instantáneos que tenemos en la alacena incluye un algoritmo que nos permite prepararlos. Nos encontramos con algoritmos desde la escolaridad primaria cuando realizamos una división según el algoritmo de Euclides y determinamos el máximo común divisor. También los encontraremos en estudios más avanzados cuando usamos el no menos célebre algoritmo de Arquímedes para el cálculo de PI. Contar con los dedos implica un algoritmo, si miramos con atención, los algoritmos saltarán por todas partes. Dehaene nos muestra cómo van variando las estrategias de los niños para contar con los dedos de la mano, Figura 3) lo que implica la utilización de diferentes algoritmos (Dehaene, 2016). Qué tienen en común las instrucciones para cocinar los brownies, contar con los dedos de la mano y el algoritmo de Arquímedes: todos y cada uno de ellos son "recetas", procedimientos con pasos sucesivos (Figura 3).

Figura 3. Los algoritmos para contar con los dedos varían según edad y cultura. En la imagen vemos la representación de la forma en la que contaban dos personas de Nigeria entrevistados por científicos del Centro de Cultura Científica de la Universidad del País Vasco. Mientras contaban, contraían los dedos de las manos, primero de la derecha y luego de la izquierda, empezando por el dedo pulgar. Imagen y referencia empírica extraídos de https://culturacientifica.com/2018/11/28/y-tu-como-cuentas-con-los-dedos-1/

Uno de los mejores consejos que nos entrega Luoridas es que no podemos olvidar que los algoritmos, aun cuando se puedan programar e implementen en computadoras, son escritos anteriormente por humanos que buscan la solución de un problema. Aconseja que si se quiere entender la esencia de un problema los algoritmos que se utilicen sean escritos a mano, con lápiz y papel. Esto diferenciará al técnico que solo puede tipear códigos del arquitecto que logra entender la médula de un problema. Definir el concepto de algoritmo abre muchas alternativas, sobre todo si usamos la noción de una manera tan amplia como la que acabamos de sugerir que incluye recetas para preparar repostería casera. En este capítulo vamos a restringir el alcance de la noción y la acomodaremos para que sea útil en el campo de la programación. Aunque no hayamos esperado a las computadoras para tener algoritmos y ellos nos acompañen hace milenios, su rol en la programación es central. Mediante una forma de pensar llamada "razonamiento algorítmico" podemos resolver problemas de manera práctica y eficiente, a partir de programas que corren de manera muy rápida en las computadoras. La programación es la técnica que permite trasladar las intenciones de una notación razonada de manera tal que las computadoras puedan entenderlas y para ello utiliza los llamados "lenguajes de programación". Nuestra forma de pensamiento, las soluciones que encontramos a los problemas que imaginamos, no son posibles de ser asimilados directamente por las computadoras, ellas no hablan nuestros idiomas naturales, ni pueden entender nuestros algoritmos en los sistemas de notación con que solemos expresarlos. En este proceso de traducciones sucesivas y pasajes por diferentes formas y jerarquías de lenguajes y gramáticas que parten desde las maneras de expresión comunes de la comunicación humana, hasta el lenguaje binario de ceros y unos, llamado lenguaje máquina, hay un largo camino cuya exploración nos permitirá comenzar a entender las IA. Sin dar una definición precisa de algoritmo, señalaremos la serie de restricciones que nos acercan a una primera comprensión (Louridas, 2020, p. 26):

I. Un algoritmo implica una serie de pasos y estos pasos deben cesar luego de un número finito. Un algoritmo no puede correr de manera infinita.

II. Los pasos deben estar definidos con precisión para ser ejecutados de manera precisa.
III. El algoritmo debe operar a partir de determinado input.
IV. El algoritmo debe producir un output. Ese es su propósito y sentido, producir algo como un resultado (por ejemplo, el máximo común divisor en el algoritmo de Euclides).
V. El algoritmo debe ser efectivo y eficiente, un humano debe poder ejecutar todos y cada paso del mismo con lápiz y papel en un lapso razonable de tiempo.

A diferencia del estilo de pensamiento de la matemática con sus demostraciones y su orientación hacia el rigor, el pensamiento algorítmico de la ingeniería de software se orienta hacia la eficacia y la practicidad. Lo que desvela a una y otra son cuestiones de distinto orden. Eso no implica que una forma de proceder sea mejor ni peor, sino que responden a lógicas divergentes y conviene recordarlo al momento de evaluar lo que una y otra producen. Aquí lo relevante será la eficacia, la utilidad de las soluciones, es el mundo de la ingeniería de software más que el de la exigencia de rigor de la ciencia matemática. En la base de la IA subyacen los algoritmos. Ellos permiten encontrar soluciones, identificar los patrones en los inputs y disparar resultados en forma de outputs. Sin ellos no podríamos concebir de qué se trata la IA y por ello debemos registrar su importancia desde el punto de vista conceptual.

Ordenamos y clasificamos el campo de la IA

Llegados a este punto será conveniente discriminar diferentes niveles al interior de este gran campo de la informática en que se trata de simular comportamientos humanos que se denominan inteligentes. Existen, por supuesto, muchos comportamientos que podemos llamar inteligentes y que pueden y han sido imitados por máquinas: clasificar imágenes, jugar distintos tipos de juegos, conducir, dar un diagnóstico médico, interpretar una amplia gama de patrones, traducir textos, generar textos o imágenes, la lista es muy larga y sigue creciendo al pulso de

nuevos desarrollos. El objetivo de estos modelos es el de imitar, alcanzar y como queda claro, superar muchas veces el rendimiento de los seres humanos en funciones inteligentes. Su performance en una tarea específica no implica que sea más inteligente que un humano o que se pueda decir de ella que es inteligente en el mismo sentido que lo predicamos de nosotros los homínidos. Su potencial futuro es algo que resulta muy difícil de predecir. En un extremo nos encontramos con IAs que se especializan en la realización de una sola tarea que pueden cumplir con gran eficiencia, aunque no puedan realizar ninguna acción de manera eficaz fuera de ese campo puntual. Un gran ejemplo de ello es Deep Blue, la supercomputadora de IBM que derrotó a Gary Kasparov, el campeón mundial de ajedrez en 1996, en un evento que captó el interés mundial. En el otro extremo tenemos la expectativa de llegar a la AGI o inteligencia general que contaría con la gran dotación de capacidades de entendimiento con que cuenta un ser humano. Es de consenso casi general en la industria que este objetivo se lograría en un tiempo no muy prolongado, medido en décadas, en años o en meses. Por ahora se trata de eso, de predicciones, si este objetivo se logra o si es siquiera plausible es algo que el tiempo dirá. Pero para todo ello debemos esperar, volvamos antes sobre algunas cuestiones básicas.

Ordenemos un poco el campo de la IA para determinar de qué hablamos. Para ello es útil conocer una distinción básica y aceptada de manera general por la disciplina, la de diferentes niveles de IA. No existen demasiados debates en torno a estos ordenamientos, seguiremos el que propone Martin Keen de IBM por su simplicidad y capacidad explicativa (Keen, 2023). El criterio utilizado discrimina dos ejes, el de las capacidades y el de las funcionalidades. Del cruce de estos ejes se obtiene lo siguiente.

En el eje de las capacidades:

1. Inteligencia artificial estrecha: también llamada IA débil. Son sistemas expertos en imitar el comportamiento humano en el cumplimiento de algunas tareas particulares, pero sin capacidad de adaptarse o generalizar sus capacidades en el cumplimiento de otras.

2. Inteligencia artificial fuerte: también llamada AGI. Tendría la capacidad de adaptación a múltiples dominios, en el límite la posibilidad

de imitar cualquier tipo de acción humana inteligente sin necesidad de ser entrenada por humanos en esas nuevas tareas. Si la propia AGI quisiera aprender a resolver un desafío al que nunca se hubiera enfrentado lo haría por sí sola.

3. Súper inteligencia: aquí estaríamos hablando de una evolución de las IAs al punto tal que habrían desarrollado todas las competencias humanas sumadas a una capacidad de cómputo fuera de toda escala. Ese potencial generaría una entidad cuya mera existencia y cuya índole no podemos siquiera concebir. No sólo tendría la potestad de generar juicios propios, sino que sería capaz de tomar conciencia de sus estados internos y evolucionar según vectores intencionales y propósitos. Desarrollaría capacidades homologables a los sentimientos y la deliberación. No podemos dejar de señalar la antropomorfización de estas entidades por parte de los ingenieros de software, la industria y su caja de resonancia mediática, volveremos sobre el punto más adelante. Por ahora señalemos que, de llegarse a este punto de evolución, términos como deseos, propósitos o inteligencia perderían sentido por la profunda diferencia material entre lo biológico y lo artificial.

Hacia 2023, sin dudas nos encontramos en el primer nivel de desarrollo, el de la IA estrecha. El resto es parte de la ciencia ficción, de las expectativas. Si nos atenemos a la realidad de los desarrollos efectivos el marco es claro. Utilizamos la clasificación de Keen de IBM por discriminar claramente entre lo existente y lo puramente teórico, en lo que tiende a coincidir con las evaluaciones que suele realizar el medio académico. El *storytelling* de los voceros de Google, Open IA, Anthropic, NVIDIA, Stability, Meta o de cualquiera de las grandes corporaciones, deja una impresión algo distorsionada. Sin negar lo pasmoso de los avances pareciera que tienden a asignar capacidades mucho más desarrolladas a sus productos. Si nos quedamos con lo que enuncian, estamos al borde de que esos modelos crucen la línea entre teoría y realidad en lapsos cortos. Ya veremos si es así o no. Por ahora volvamos a nuestra clasificación. Ya nos ocupamos del primer eje, el de las capacidades, pasemos a ver qué ocurre cuando lo cruzamos con el de las funcionalidades.

1. Funcionalidades de las IAs débiles de tipo reactivo: sus capacidades se limitan al manejo de informaciones en tiempo actual y de manera inmediata sin dominio acabado de contextos más generales ni de memoria. Se trata de modelos existentes desde hace muchas décadas y que pueden llegar a altos niveles de eficacia a gran velocidad, dependiendo del volumen y calidad de los datos de entrenamiento y la índole de las tareas para las que se las prepara. Actúan mediante evaluaciones estadísticas complejas a partir de masas de datos. A partir de los datos con que cuentan y los elementos a considerar definen cuál es la mejor respuesta, lo que simula la inteligencia humana. Con los datos necesarios y especializadas en tareas específicas pueden tener resultados notables. Consideremos por ejemplo que Deep Blue, la IA que derrotó a Kasparov pertenecía a este tipo de modelo. Son eficaces en modelos de control industrial, chatbots de asistencia virtual en tiempo real, robótica, videojuegos y muchas tareas que no impliquen la incorporación y retención de datos a lo largo del tiempo.

2. Funcionalidades de las IAs débiles de memoria limitada: su sello distintivo es la capacidad de memoria que le permite recordar eventos pasados u objetos y así evaluar su evolución a lo largo del tiempo y actuar en consecuencia. Esta posibilidad requiere que la IA pueda retener y recuperar datos almacenados, si bien de manera limitada. Esos datos recuperados le permiten optimizar su evaluación del presente. En este casillero podemos ubicar los modelos generativos de lenguaje con los que interactuamos o los generadores de imagen, cuyas grandes capacidades nos habrán quedado claras.

3. Funcionalidades de la IA fuerte o AGI: Aquí nos ubicamos en el terreno puramente teórico ya que no existen desarrollos de estos modelos más que en el papel. Una de las capacidades que se les asignan es la de la comprensión empática de los pensamientos, las intenciones y emociones de otras entidades, en particular las humanas. Habrían alcanzado el punto que enuncia Minsky en la teoría de la mente, con coordinación de los agentes en algo similar a la célebre "sociedad de la mente". Sus acciones estarían basadas

en sus necesidades y en sus intenciones particulares. Capacidades de este tipo se incrementaría con la posibilidad de captación constante de estímulos de todo tipo, visuales, sonoros, texto, profundidad de campo y distancias, movimiento, captación térmica, todo lo que se pueda imaginar. Este es uno de los aspectos en los que se pueden pensar las AGI pero no el único, otros ejes en el que se piensan es en el de los objetivos, la colaboración entre agentes inteligentes artificiales o entre agentes humanos e IAs, la mejora significativa de los procesos de razonamiento, de memoria episódica, de memoria a corto y largo plazo, el aprendizaje continuo con ajuste permanente de parámetros, la imaginación y la búsqueda de soluciones inesperadas a partir de los datos preexistentes, la capacidad de agencia y de formulación de propósitos, el automorfismo o capacidad de auto mejora permanente, la replicabilidad autónoma, el aprendizaje jerarquizado y la incorporación de estructuras totalmente reversibles. Los desafíos son varios y existe una puja por definir cuál será el camino que pueda acercar o arribar a la meta de una inteligencia general. Hace unos años esto hubiera sido tomado por delirio, hoy, al menos la industria y el mundo académico que se le asocia lo enuncia sin vergüenza. La cantidad de esfuerzo intelectual que se destina a esta tarea y el volumen de experimentos que corren hoy la industria y la academia en la persecución de las AGI es descomunal.

4. Funcionalidades de la Super inteligencia. Siempre instalados en el plano de lo teórico, estas entidades contarían con la capacidad de autopercepción y de registro de sus procesos internos. Serían poseedoras de emociones propias y de conjuntos de objetivos autodefinidos de alta complejidad. Se las imagina como dioses, es decir como hipérboles de lo humano. Si no contamos con AGI, mucho menos con estas súper entidades. Se ubican en un terreno muy resbaladizo, entre la ciencia ficción, las dificultades de interpretación de capacidades emergentes, la avidez de la comunicación mediática, las operaciones de publicidad de las empresas de software y sus voceros, la intervención de filósofos y ensayistas,

sumadas a las consideraciones éticas e incluso religiosas, como podemos imaginar nada queda demasiado claro. Cuando prestamos atención a las elaboraciones en torno a este estadio de la IA nos encontramos con una serie de construcciones ideológicas y de antropomorfizaciones más o menos confusas. En esta dimensión, las cuestiones filosóficas se hacen presentes y se ponen en juego concepciones, en algunos casos relevantes y en otros delirantes. Un consejo: dejemos espacio para el asombro y no nos apuremos a apostar contra lo fantástico. No olvidemos lo absurdo que pudo parecer un viaje a la Luna en el siglo XVIII y aquí estamos, pensando en colonizar Marte. El mundo de la ingeniería de software está demostrando una alta capacidad de innovación y puede que revolucione la existencia de los individuos y las sociedades. Sin embargo, el "universo IA" está en deuda en el terreno de reflexión ética y filosófica. El discurso de los super poderes o su reverso en los vaticinios del apocalipsis cinematográfico que circula en torno al tema resulta insuficiente. Debemos avanzar en el plano de la reflexión y el debate si queremos afrontar el desafío que implican estas tecnologías que apuntan a modelar el futuro de nuestras sociedades y el destino de la humanidad en su conjunto.

Bibliografía

Bachelard, G. (2000). *La Formación del Espíritu Científico: Contribución a un Psicoanálisis del Conocimiento Objetivo*. Siglo XXI.

Berger, J. (2009). *El sentido de la vista*. Alianza Forma.

Canavos, G. (1998). *Probabilidad estadística aplicaciones y métodos*. McGraw Hill.

Chomsky, N. (1986). *El lenguaje y el entendimiento*. Planeta Agostini.

Hinton, G. E. (2022). The Forward-Forward Algorithm: Some Preliminary Investigations. ArXiv, abs/2212.13345.

Lightman, H., V. Kosaraju, Y. Burda, H. Edwards, B. Baker, T. Lee, J. Leike, J. Schulman, I. Sutskever y K. Cobbe (2023). Let's Verify Step by Step. ArXiv, abs/2305.20050.

Volver

Louridas, P. (2020). *Algorithms*. The MIT Press Essential Knowledge Series.

Mandler, G. (2007). *A History of Modern Experimental Psychology: From James and Wundt to Cognitive Science*. The MIT Press.

Minsky, M. (1986). *La Sociedad de la mente*. Editorial Galápago.

Mikolov, T., K. Chen, G. S. Corrado y J. Dean (2013). Efficient Estimation of Word Representations in Vector Space. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1301.3781

Molnar, Ch. (2021). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. p. 12.

Prince, S. (2024). *Understanding Deep Learning*. The MIT Press.

Rabuffetti, H. (1997). *Introducción al Análisis matemático*. El Ateneo.

Russell, S. y P. Norvig (2004). *Inteligencia artificial. Un enfoque moderno*. Pearson Ealan.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, LIX*, 433-460.

Vaswani, A., N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser y I. Polosukhin (2017). Attention is All you Need. Neural Information Processing Systems. https://doi.org/10.48550/arXiv.1706.03762

Wittgenstein, L. (1998 [1969]). *Sobre la certeza*. Gedisa.

Zhang, D. T., S. Mishra, E. Brynjolfsson, J. Etchemendy, D. Ganguli, B. Grosz, T. Lyons, J. Manyika, J. Niebles, M. Sellitto, Y. Shoham, J. Clark y R. Perrault (2021). The AI Index 2021 Annual Report. ArXiv, abs/2103.06312.

YouTube

The 7 Types of AI - And Why We Talk (Mostly) About 3 of Them. 2023. Martin Keen

Chanell IBM Technology

Volver

Capítulo 2. La capacidad lingüística: una facultad específica de lo humano

Sumario

En este capítulo vamos a introducir una polémica en relación al lenguaje y sus bases, que enfrenta a los lingüistas racionalistas y a los asociacionistas cognitivistas. Comenzaremos presentando la posición del racionalismo lingüístico (el asociacionismo se desplegará en capítulos subsiguientes). Para Chomsky, el exponente principal del racionalismo lingüístico, el lenguaje es una facultad del entendimiento humano que nos diferencia del resto de las especies, nuestro sello distintivo. Según lo entiende no puede ser replicado por las máquinas y tampoco se asemeja a los sistemas de comunicación de otras especies vivientes. Se trata de una facultad innata que no depende de la adaptación al medio ambiente ni de la transmisión intergeneracional. Sus desarrollos en torno a la lingüística y a los lenguajes formales tuvieron una influencia positiva y son utilizados actualmente en el campo de las ciencias de la computación. Su oposición a herramientas probabilistas contribuyó a detener algunos desarrollos técnicos que terminaron siendo cruciales para el procesamiento del lenguaje natural. Para el autor, la concepción de la ciencia vinculada al cognitivismo en alianza con corrientes dominantes en las ciencias de la computación es un camino sin salida desde el punto de vista científico. Según sostiene, su enfoque probabilista no aporta conocimiento sobre los principios profundos que hacen a la comprensión del entendimiento y el lenguaje humano.

Volver

Chomsky: un enfoque racionalista de la capacidad simbólica

Somos grandes buscadores de patrones. Esta noción está en la base del planteo cognitivista y de sus aliados teóricos en la corriente dominante de las disciplinas computacionales vinculadas a las IA. El homínido y aún más, la vida, tienden a rastrear, identificar y utilizar patrones que los guíen en su interacción con el medio que los circunda. Esta tendencia hacia la utilización de patrones se manifiesta en múltiples niveles, biológicos, psicológicos, cognitivos, culturales y científicos. Pero, antes que nada: ¿qué es un patrón en el contexto que a nosotros nos interesa? Cuando nos referimos a patrones debemos pensar en estructuras y secuencias recurrentes que permiten predecir eventos en un ambiente determinado. Es decir, para hablar de un patrón tenemos que tener en cuenta una organización de elementos que tienen una determinada secuencia, que son repetitivos y que proveen algún tipo de orientación predictiva. Por lo tanto, la existencia de patrones permite entender el medio ambiente y establecer una forma de relación particular con el mismo. Su permanencia en el tiempo permite guiarse en el medio ambiente en que nos encontramos insertos. El universo tiene regularidades y las entidades biológicas somos particularmente sensibles a la captación de las mismas. De hecho, la existencia de patrones supone que los eventos del ambiente al que nos referimos tienen algún grado de organización y guardan algún tipo de lógica. Este será nuestro punto de partida.

Sin embargo, no todo se trata de la búsqueda de patrones cuando nos referimos a la capacidad simbólica. ¿Y si los seres humanos contáramos con una facultad innata, la del lenguaje, que no dependiera de identificar patrones en el entorno y que ya formara parte de nuestro arsenal sin ser derivables de la relación con el medio ambiente en el que se enclava el organismo? Esta suposición va en el sentido contrario al que adoptan los asociacionismos propios de las teorías del comportamiento y sus vecinos los cognitivismos. Las disciplinas cognitivas y las computacionales proponen un paradigma en torno a la identificación de patrones y regularidades en grandes conjuntos de datos y el establecimiento de

predicciones eficaces en torno al medio ambiente. No obstante, existen objeciones en torno a este paradigma que avanza a pasos agigantados a partir de los éxitos tecnológicos de las IAs generativas. El abanderado de estas objeciones es Noam Chomsky, lingüista, filósofo y analista crítico de la política global. Los argumentos de las corrientes cognitivistas cuentan con un arsenal de divulgación que atraviesa todos los niveles, desde el periodístico hasta el más alto grado de sofisticación académica. Este paradigma ascendente ha sabido traducir sus nociones a un lenguaje sencillo y convincente. Apoya sus certezas en estudios empíricos cuyos alcances experimentales suelen ser claros, pero cuyas consecuencias se tiende en muchos casos a sobredimensionar. En la esfera de la divulgación hacia la opinión pública se muestran como aliados de las disciplinas computacionales. Estamos ante un panorama lleno de extrapolaciones e hipérboles en torno a la inteligencia, el lenguaje y las facultades humanas. Cuando un paradigma avanza y tiende a colonizar otros campos de saber, resulta conveniente tratar de establecer sus alcances y limitaciones. El lenguaje, aquello que en ciencias sociales y humanas consideramos territorio propio está siendo explorado y conceptualizado en otros campos. La inteligencia, aquello que los psicólogos consideramos territorio propio de lo humano, o al menos del reino de la vida, está siendo vinculado a las máquinas. Desde nuestra perspectiva, las consecuencias de este diálogo son positivas. El pensamiento estanco ya no es pensamiento en el sentido científico y conceptual del término. Un saber que no es desafiado y dinamizado se torna repetición escolar (Bachelard, 2000; Popper, 1998). Cuando un campo de saber se enfrenta al cuestionamiento de sus certezas básicas pueden desatarse actitudes defensivas o cerrarse amurallando sus fronteras. No consideramos que esta sea una alternativa productiva ni eficaz. Tampoco lo es correr tras la novedad y adoptar el saber proveniente de otras disciplinas de manera acrítica. Antes de discutir, aceptar o combatir estos paradigmas, trataremos de estudiar sus fundamentos.

¿Pero qué relación puede tener esto con la capacidad simbólica en un sentido amplio? Este interrogante nos invita a sumergirnos en varias aventuras que han emprendido pensadores de las ciencias sociales, la antropología, el psicoanálisis, la fonología y la lingüística generativa.

Visitaremos el pensamiento de Chomsky en torno al lenguaje y en un capítulo posterior el de Piaget en torno a la inteligencia. Estos pensadores nos brindarán perspectivas diferentes en torno a la capacidad simbólica, a la comunicación humana y, lo que es central, el lenguaje como potestad compartida por nuestra especie. Ellos se han enfrentado con su campo de estudio y reflexión suponiendo que los eventos con los que tratan no se mueven al azar, ni por simple asociación de estímulos y respuestas. Han sabido captar un orden en el mundo, han postulado una organización, un tipo de secuencia y una regularidad en el campo de los signos y los símbolos. Si bien los mundos que postulan y sus herramientas teórico conceptuales difieren, cada uno de ellos ha sido guiado por la razón y ha buscado una lógica en el universo. Es, en definitiva, el desafío de lo racional, del entendimiento humano que aspira a conquistar las profundidades de un entorno complejo que no entrega de manera inmediata sus claves de comprensión. En eso coinciden los mejores representantes del estructuralismo continental de la década de los 60 y 70 y la lingüística generativa radicada en Estados Unidos.

Chomsky, el lenguaje y el entendimiento humano

Empecemos por Chomsky, lo que resulta comprensible ya que este libro trata sobre la IA y nuestro protagonista es un hijo dilecto del M.I.T., institución que se convertirá en una usina de reflexión en torno a la lingüística computacional. Se ubica, sin lugar a dudas, en un punto clave de intersección del estudio de la capacidad de lenguaje humano y de la programación del lenguaje natural en la lingüística computacional. Para constatar la importancia de Chomsky en este campo basta con revisar el número de entradas y referencias al autor en muchos de los textos que representan el estado actual de la disciplina. Por ejemplo, en la última versión de 2023 del libro de Jurafsky y Martin, "Speech and Language Processing", aún no editado pero que los autores han abierto para ser revisado por pares, detectamos 21 entradas referidas a Chomsky. Lo importante de las referencias a este autor reside en que se relacionan con

procesos muy específicos para el procesamiento del lenguaje. La lingüística transformacional y generativa de Chomsky se integra como herramienta y guía conceptual en el mundo del procesamiento del lenguaje en los grandes modelos generativos de las IAs. En resumidas cuentas, es posible escalar desde los conceptos formales propuestos por la lingüística de Chomsky hasta algoritmos probados y con resultados concretos en el campo del procesamiento del lenguaje natural. El llamado procesamiento del lenguaje natural PLN es una rama de las ciencias computacionales que intenta abordar los lenguajes llamados "naturales", el español, el chino mandarín o el francés, desde la perspectiva de los lenguajes formales, propios de la computación. La diferencia entre los lenguajes naturales y los formales reside en que los primeros no tienen definiciones estrictas desde el punto de vista lógico – formal y matemático, en tanto los lenguajes de procesamiento computacional cuentan con definiciones y restricciones rigurosas. Es decir, el PLN, en algunas de sus orientaciones, intenta abordar el lenguaje corriente, el que compartimos las comunidades lingüísticas, como si fuera un lenguaje formal. Para profundizar en estas definiciones recomendamos el clásico tratado de Russell y Norvig "Inteligencia Artificial. Un enfoque moderno." (Russell y Norvig, 2004, pp. 927-931). Esta relación entre Chomsky y el procesamiento informático no ha sido siempre sencilla, realiza aportes clave en temas como la desambiguación y el tratamiento de lenguajes libres de contexto, pero desalienta herramientas que terminan siendo fructíferas para la predicción de palabras como en el caso de las llamadas cadenas de Markov profundas (Jurafsky y Martin, 2023, pp. 356-378).

Como todo buen cartesiano Chomsky es un pensador que duda, que se siente molesto con las ideas poco claras o imprecisas. Para él los conceptos tienen consecuencias y se encadenan en razonamientos estrictos. Proceder según estos parámetros es muy complicado si no sabemos bien de qué hablamos o nos vamos por las ramas saltando de una idea mal formulada a otra. Si no aclaramos a qué nos referimos con lenguaje, inteligencia o capacidad simbólica en campos particulares como el de las IA o la lingüística seguramente naufraguemos. Chomsky no da nada por cierto y ese es uno de los sellos de calidad de su enfoque. La duda es una herramienta clave del pensamiento crítico y despeja el camino

para ajustar los conceptos y definir los límites en que son aplicables. Este es otro de los motivos por los cuales conviene comenzar nuestro recorrido en este punto. Veamos cuál es la actitud mental de Chomsky y, sobre todo, su incomodidad ante lo que considera abusos en la utilización de nociones tales como "inteligencia", "razonamiento" o "entendimiento", "intención" o "propósito". Según lo entiende, esta alianza de ciencias cognitivas y la ingeniería computacional promueve una concepción de la ciencia y de sus objetivos muy apartada de los enfoques racionales que aspiran a encontrar las causas profundas de los fenómenos. Considera que esta postura implica una renuncia inaceptable que consiste en reemplazar la explicación racional por la aproximación probabilística. Estos enfoques se basan en el procesamiento estadístico de grandes masas de datos sin analizar. La lógica con que se mueven los fenómenos queda sin explicar, lo que obtienen como resultado de los estudios es la probabilidad de que los fenómenos se muevan en uno u otro sentido. Es como si se filmara el clima en diversas regiones durante largos períodos y se procesaran estos datos mediante IAs para predecir acontecimientos meteorológicos. Con este input sin analizar podríamos obtener outputs sobre el estado del clima, pero poco podríamos avanzar en términos de los fundamentos de la meteorología. Sabremos si mañana nos conviene llevar el paraguas, pero no contaremos con una razón científicamente fundada para hacerlo. Algo similar ocurre con el estudio del lenguaje desde una perspectiva de la probabilidad, algo a lo que se opuso de manera sistemática por considerar que los enfoques actuales no aportan conocimientos científicos sólidos sobre el entendimiento humano. El autor lo expresa de la siguiente manera:

> *The kind of critique just outlined, which is quite widespread, is generally accompanied by a novel concept of science that has emerged in the computational cognitive sciences and related areas of linguistics, with a new notion of "success": an account of some phenomena is taken to be successful to the extent that it approximates unanalyzed data.* (Chomsky, 2011, p. 266)

Desde sus tiempos de joven lingüista cartesiano, los reparos de Chomsky tienen como uno de sus principales blancos las llamadas "ciencias del comportamiento"; consideraba que sus principios estaban pobremente formulados y enturbiaba la investigación sobre el lenguaje y la inteligencia (Chomsky, 1972, [1965]). Las ideas asociacionistas que suponían que todo rasgo de inteligencia y toda capacidad de los seres biológicos, en particular los humanos, partía de la asociación de estímulos externos con reacciones conductuales, chocan de frente con sus convicciones racionalistas y las conjeturas innatistas que conllevan. La expectativa en el esquema simplista estímulo–respuesta dominaba el espectro y colonizaban diferentes disciplinas ejerciendo una poderosa influencia en torno al aprendizaje del lenguaje. A esto se sumaba, para indignación del autor, la interpretación apresurada de la teoría de la información de C. Shannon y las implicaciones de su teoría de la comunicación. Para Chomsky la orientación hacia la comunicación no es una parte fundamental de la facultad lingüística. Las expectativas y conjeturas que generaba esta alianza intelectual en el plano del lenguaje, del aprendizaje y de la comprensión de lo humano en general sonaban "extravagantes" a los oídos de nuestro protagonista (Chomsky, 1986). Por otra parte, refiere que, hacia fines de los 40 y durante la década de los 50, los positivistas hacían fila para anunciar el advenimiento de una comprensión global del entendimiento humano fruto del análisis estadístico del comportamiento. Para Chomsky esto resultaba inconcebible, no solo por su inclinación hacia los esquemas formales de pensamiento, propios de su espíritu racionalista, sino por la pobreza que percibía en las conjeturas empíricas que entregaba el behaviorismo en el campo del aprendizaje o de la comprensión del lenguaje. La hostilidad entre Chomsky y los cognitivistas era recíproca y dio lugar a varias controversias, algunas de ellas plagadas de descalificaciones recíprocas (Chomsky, 2007). Señalaba que las ciencias del comportamiento son un reflejo empobrecido y raquítico de las ciencias naturales que aspiran a emular. A esto contraponía la intuición de la existencia de explicaciones que tuvieran en cuenta estructuras profundas que pudieran hacer lugar a la plasmación empírica del lenguaje en el plano superficial. Chomsky aspira a la abstracción racional, a la descripción, formalización y contraste de estructuras profundas que pudieran

Volver

dar cuenta de las realizaciones de superficie. Para Chomsky la capacidad humana del lenguaje es universal y conforma el sello de nuestra especie. No es adquirida por transmisión comunicativa intergeneracional ni se forma por la presión de las experiencias adquiridas en el medio ambiente. De hecho, enfatiza, se adquiere a partir de un número muy pequeño de ejemplos y experiencias que no alcanzan para fundamentar la incorporación de un universo tan rico como el del lenguaje y tan plagado de implicaciones. Su conclusión es radical: sólo puede tratarse de una facultad con bases innatas.

Language acquisition can be seen as the transition from the state of the mind at birth, the initial cognitive state, to the stable state that corresponds to the native knowledge of a natural language. Poverty of stimulus considerations support the view that the initial cognitive state, far from being the tabula rasa of empiricist models, is already a richly structured system. (Chomsky, 2003, p. 7)

Cuando estudiamos el lenguaje humano, nos acercamos a lo que algunos podrían llamar "esencia humana", las cualidades distintivas del entendimiento que, por lo que sabemos hasta ahora son específicas del hombre e inseparables de cualquier fase crítica de la existencia humana personal o social.

N. Chomsky, 1986, p. 171.

Los niveles jerárquicos de lenguaje: las fronteras entre el entendimiento humano y los autómatas

Desde estas posturas racionalistas e innatistas se supone que, cuando hablamos, seleccionamos libremente una estructura generada por nuestro procedimiento recursivo y que concuerda con nuestras intenciones comunicativas (Chomsky, 2011). Una selección particular en una específica situación discursiva es un acto libre en el sentido de Saussure, pero el procedimiento subyacente que especifica los posibles "patrones regulares" está estrictamente gobernado por reglas. Chomsky aspira a sistematizar estas regularidades y dilucidar las reglas que las gobiernan. Su impacto en el tratamiento formal de los lenguajes es obvio, y su efecto en el campo de la IA no se hizo esperar (Everaert et al., 2015). Queremos señalar su importancia en el terreno de diseño de máquinas capaces de generar textos que procuran imitar el lenguaje natural. Basta con señalar que sus aportes sobre las gramáticas formales brindan la pista para responder qué tipos de lenguaje pueden ser comprendidos por máquinas autómatas y qué tipo de estructuras lingüísticas son susceptibles de ser "traducidas" a estados internos de una máquina (Figura 1). Esta idea de que existen procedimientos que hacen posible el pasaje de un nivel de lenguaje a otro y de la existencia de gramáticas que lo posibilitan nos resulta familiar a partir de la explosión de las IAs generativas, que se entrenan con texto en lenguaje natural y que lo generan como respuesta. El enfoque jerárquico y estructural de Chomsky se corresponde con su militancia contra los esquemas empiristas que se basan en la generalización de inferencias basadas en regularidades distributivas (Everaert et al., 2015). La referencia a distintos niveles de estructuración gramatical nos ayuda a pensar qué hacemos cuando interactuamos con un chat conversacional de una IA generativa. No dialogamos con ella, aunque así nos lo parezca, sino que interactuamos atravesando diversos niveles jerárquicos de lenguaje. La jerarquía de Chomsky, formulada en 1958 (Figura 1), establece una barrera entre los lenguajes naturales y los llamados lenguajes formales (Gallego, 2008). Estos últimos son aquellos que pueden

ser reconocibles e interpretables por una máquina, en tanto que los lenguajes naturales propios de la humanidad se ubican más allá de la capacidad de los autómatas debidos a su alto nivel de ambigüedad y a su recursividad, lo que le permite generar un número infinito de expresiones a partir de un conjunto limitado de elementos. Esta diferenciación entre gramáticas formales y gramáticas ambiguas tiene gran influencia en la comprensión de los límites del lenguaje computacional y su discriminación del humano. A su vez, las gramáticas formales, aquellas que pueden ser reconocibles y procesables por máquinas, presentan cuatro niveles de restricción. Es decir, están más o menos sometidas a reglas según cuatro niveles de jerarquía. En el nivel 0 se ubican las gramáticas más generales y poderosas. Sin embargo, los lenguajes de programación no se ubican en este nivel dada su complejidad de manejo. La mayoría de los lenguajes de programación por razones de eficacia se ubican en el nivel 2 de la jerarquía de Chomsky (Phyton, Java, C++, entre otros). En este libro no entraremos en detalles respecto de los diferentes niveles de gramáticas, pero señalamos que sus consecuencias en términos de la teoría de los autómatas han sido amplias y profundas (Jurafsky y Martin, 2023).

Figura 1. Jerarquía de Chomsky Fuente: Clase digital 5. Análisis léxico: Teoría de los lenguajes - Recursos Educativos Abiertos (ugto.mx)

Antes de repasar algunos de sus planteos en torno a la lingüística transformacional y la lingüística generativa, señalemos algo que se suele pasar por alto cuando hablamos de su obra: para Chomsky no todo puede ser comprendido y explicado (Chomsky, 1985). Señala la existencia de un punto ciego, de un núcleo opaco en la comprensión del entendimiento humano, algo que excede la capacidad explicativa de la ciencia. Nos alerta que, aun cuando pudiéramos explicar la adquisición del lenguaje mediante algún procedimiento conceptual especulativo, y aun cuando recibiéramos confirmaciones empíricas, quedaría por explicar el uso normal que le damos al conocimiento así adquirido. Este problema, postula, es "intratable" mediante los procedimientos lingüísticos con los que se cuenta y tiende a suponer que resultará inaccesible al intelecto científico en general. Ello no quiere decir que se deba o se pueda abandonar la especulación, la construcción de esquemas y conjeturas en torno a estos temas. Sin embargo, no esperemos explicaciones finales que permitan cerrar la cuestión del entendimiento y la creatividad humana vinculada a la capacidad lingüística. En principio se trata de abandonar la expectativa de extrapolar el esquema simple de estímulo respuesta y el sueño de descubrir estructuras cerebrales que sigan este principio. Tampoco le parecía posible apegarse a una concepción tan simplista de "estímulo", con tan poca carga cualitativa. Señala, quizás con razón, la insuficiencia conceptual de las producciones de los seguidores de Skinner, cuyas elaboraciones aspiran a seguir la falsa imagen que se forma el positivismo sobre lo que sería la ciencia de su época. En este punto se apega a la crítica estándar del racionalismo cuando sostiene que esta "falsa imagen" se constituye a partir de esquematizaciones de una pobreza notable; lo que se imita no es la ciencia natural en su mejor expresión de abstracción racional, sino una especie de sombra degradada apegada a lo empírico (Popper, 1999). A ello, contrapone la tesis de la existencia de mecanismos abstractos que no pueden ser analizados en términos asociativos ni reducirse a principios simplistas.

Uno de los textos mayores de Chomsky lleva el nombre de "El lenguaje y el entendimiento" (1986). ¿A qué se refiere con entendimiento? Se trata de la capacidad innata de adquisición y de procesamiento del lenguaje. La facultad lingüística del entendimiento humano es una

singularidad, el sello que nos diferencia de otros seres vivos. Único y singular, el lenguaje nos distingue de otras especies. En esto el autor es taxativo, el lenguaje humano es algo único y distinto del lenguaje animal. Esta afirmación se conoce como "principio de la unicidad" y sostiene que el lenguaje tal y como se manifiesta en nuestra especie no es una prolongación más sofisticada de capacidades ya presentes en el reino animal sino algo totalmente diferente. En este sentido, hacia los años 2000, define dos niveles en la facultad humana del lenguaje: un sentido amplio FLA y facultad del lenguaje en un sentido estrecho FLE (Birchenall y Müller, 2014). El primer nivel, el de sentido amplio implica e incluye un sistema sensorio motor y un sistema conceptual intencional. Los animales vertebrados cuentan con un sistema motor que les permite articulaciones de sonidos con pautas específicas dentro de las especies. Estas articulaciones vocales discriminadas y compartidas al interior de la especie son reproducibles por conducta imitativa y contribuyen a la comunicación. A su vez, entre especies animales no homínidas, se da por probado un nivel complejo de organización conceptual, representaciones abstractas muy ricas, llegando incluso hasta el manejo de herramientas. En esto coincide con autores provenientes de otros campos como es la neuropsicología, terreno en que se llega a constatar de manera empírica habilidades muy complejas como por ejemplo la capacidad numérica de mamíferos superiores no humanos (Dehaene, 2016). En definitiva, el reino animal en determinado nivel de agregación evolutiva cuenta con facultades de lenguaje que contribuyen a la adaptación y permiten la comunicación. Tampoco la intencionalidad sería una potestad exclusiva de los humanos. Desde un campo polémico como el de la teoría de la mente, se llega a afirmar que algunas especies como los primates, pero también en niveles menos sofisticados, cuentan con la posibilidad simbólica de representar intenciones y estados mentales en sus congéneres en situaciones de convivencia grupal. Ni las intenciones, ni la capacidad conceptual, ni la posibilidad de articular sonidos con fines comunicativos nos distinguen de otros niveles de organización de la vida animal (Chomsky, 2011). Es decir, la facultad del lenguaje en un sentido amplio es compartida con otros animales.

Volver

Entonces, cuál sería la diferencia específica de la facultad humana del lenguaje, aquello que se define como "entendimiento" y que hacia 2002 se asocia con el nivel estrecho (FLE). Algo muy particular llamado capacidad computacional. Aquí queremos hacer una advertencia: qué se mencione la palabra "computacional" no implica que nos refiramos a computadoras. A lo largo de todo el libro vamos a insistir en la utilidad de construir conceptualmente las nociones para evitar confusiones. Con capacidad computacional nos referimos a la posibilidad de la facultad del lenguaje de construir un conjunto infinito de oraciones a partir de un número limitado de términos. Esto se conoce como infinitud discreta. Con un número de elementos discretos (identificables uno a uno), es posible producir un número infinito de oraciones. La especie humana es la única con la capacidad de generar un número infinito de oraciones, incluso aquellas que nunca han sido generadas con anterioridad. En este rubro estamos solos, no nos acompaña ninguna especie y, lo que es más relevante para este libro, Chomsky sostiene que tampoco se puede predicar algo así de la generación del lenguaje por parte de las máquinas. Un momento, ¿no se puede producir un número infinito de oraciones mediante la combinación de elementos discretos a partir de un proceso computacional artificial? Pareciera que sí ya que las IA generativas producen texto e imagen, no obstante, lo hacen dentro de los límites de una gramática restrictiva. Se trata del tema de la creatividad de las máquinas, un tema complejo cuya discusión requeriría un capítulo en sí mismo. Para Chomsky, no obstante, no tenemos motivos para suponer que las IAs, incluso en el más alto nivel de aprendizaje profundo, nos pudieran enseñar algo sobre el lenguaje humano. Preguntarse si las máquinas pueden pensar, señala, es un callejón sin salida ya que los sistemas predictivos de este tipo operan por fuerza bruta en la identificación de patrones. Son, nos dice, verdaderos bulldozers que identifican patrones de regularidad en inmensas masas de datos. Conceptualizar el lenguaje o el pensamiento humano y artificial en un mismo plano le parece inaceptable. Ya sabemos que el autor defiende desde siempre la unicidad del lenguaje humano y se esfuerza de deslindar sus características de las facultades de otros seres vivos de manejar símbolos. Si por principio se niega a asimilar nuestras facultades con las que emergen en otros

órdenes de la vida, mucho menos podría aceptar la equiparación con el funcionamiento de modelos computacionales de generación de lenguaje natural (LLMs) cuya función es la de predecir la próxima palabra. Para Chomsky existen muchos motivos para deslindar las facultades humanas de las propias de las IAs en cualquiera de los niveles de desarrollo técnico en que se encuentren (Chomsky, 2011; Keating, 2023).

Figura 2. Encuentro de Chomsky y Minsky en 2011 en la entrega del simposio "150 aniversario en el MIT sobre la inteligencia humana y artificial". Fuente: minds and machines. Final installment of MIT's 150th anniversary symposia explore intelligence — both human and artificial. Larry Hardesty and Anne Trafton, MIT News Office. Publication Date: May 9, 2011. https://news.mit.edu/2011/mit150-brain-ai-symposium

Chomsky, un polemista que enfrenta científicos e ingenieros

¿Dónde reside la raíz de estas discusiones? ¿Por qué resulta tan difícil acordar en torno a la inteligencia y la capacidad lingüística? Una de las dificultades con que nos encontramos hoy en día, en pleno estallido mediático del tema de las IAs reside en el propio término

"inteligencia", cuyo sentido y alcance dista de ser "claro y distinto". Llamar a diferentes entidades mediante los mismos términos no garantiza la coherencia. Chomsky, retomando las inquietudes de Turing (1950), señala que se puede decir de un avión que vuela, así ocurre en diversas lenguas, pero ¿hablamos de lo mismo cuando nos referimos a un ave? Podemos, en forma de una analogía pobre, decir que un submarino "nada", pero salta a la vista que no predicamos algo similar a las acciones de entidades biológicas. Desplazarse en el agua no implica necesariamente nadar.

Existen, sostiene, dos niveles bien discriminados de avances en el plano de la IA, el científico orientado al conocimiento y el técnico vinculado a la ingeniería. El primero, que vincula con científicos como Minsky, no es distinguible de las ciencias cognitivas y sitúa las cuestiones conceptuales en un nivel alto de reflexión. El cruce entre ambos científicos se repitió a lo largo de los años en diálogos, conferencias y referencias cruzadas (Chomksy, 2011). El segundo, propio del desarrollo de productos técnicos orientados a generar rentabilidad y teñidos de componentes ideológicos, ambición empresarial y divulgación entusiasta, no guarda relación con la forma en que podamos explorar la naturaleza humana y la complejidad de las facultades simbólicas que se ponen en juego en la interacción con el entorno. Resulta claro hacia dónde se orientan las simpatías y antipatías de Chomsky. Para el autor, la interrogación técnica no responde a la forma en que se estudia en ciencia, en que se conjetura en ciencia. El problema tiene que ver con la típica dificultad que genera la proliferación de variables en ambientes complejos en que se desenvuelve el comportamiento humano y los desafíos que ello implica para su comprensión. Rechaza la aproximación estadística probabilística y la premisa de aceptar "lo que funciona" por el mero hecho de que funciona. No se puede estudiar las leyes del movimiento haciendo estadística de un conjunto de hojas que hace volar el viento. No se cuenta con el elemento radical de abstracción profunda que caracteriza cualquier experimento que interroga la naturaleza compleja de la vida y la experiencia (Bachelard, 2000). Si uno quiere indagar los primeros principios acerca de la manera en que el mundo funciona no tiene sentido querer establecerlo a partir de la manera en que las cosas están ocurriendo, existen

demasiadas variables en juego. La investigación técnica en IA no tiene esta inclinación, no aspira a comprender principios básicos acerca de la inteligencia, es más bien la voluntad de producir artefactos que hagan cosas. Nadie se atrevería a desconocer el aporte de este gran lingüista (Figura 3). No obstante, y respecto del actual desarrollo de las IA generativas de lenguaje, Chomsky considera que nada relevante puede ser pensado por esta vía. Afirmación que parece un tanto radical a la luz de los desarrollos recientes de los modelos generativos y el desafío que presentan para las concepciones del lenguaje. El debate está abierto.

Como si todo esto no fuera suficiente, señala otro punto muerto en el intento de descifrar la lógica del pensamiento y el lenguaje humano a partir de su relación con las IA. No solo cuestiona el enfoque técnico probabilístico de las mismas, sino el modelo biológico en que se basan sus propuestas. Se pregunta si las propias ciencias cognitivas no se encuentran varadas en la búsqueda de respuestas a nivel de las redes neuronales. Si las sinapsis neuronales no son, o no han probado ser, el mejor soporte para situar los procesos de pensamiento, no deberíamos esperar que las redes neuronales artificiales que intentan modelizar una réplica de las mismas sean la base desde la que podamos entender las facultades del entendimiento humano. De manera provocativa señala que la sede del pensamiento no debiera situarse a nivel de la neurotransmisión ya que la velocidad con que ella opera es muy lenta para ser considerada el gran vehículo de las facultades superiores de generación de lenguaje. A título conjetural se refiere a estudios que ubican el nivel del pensamiento en estructuras de base de nivel "más bajo" que la sinapsis como son los microtúbulos y la referencia, no ya a reacciones químicas de neurotransmisión, sino a eventos cuánticos. Estos campos de estudio son altamente conjeturales y son objeto de fuerte controversia académica, a la vez que son terreno fértil de la divulgación más desenfrenada. Algunas de las referencias a las que se puede acudir es la de F. Beck y J. Eccles "Aspectos cuánticos de la función cerebral y su rol en la conciencia" de 1992 y a los postulados polémicos de Kauffman (2009) un gran impulsor de esta corriente.

Volver

The neural net system is just the wrong place to look. They don´t have the right kind of architecture which is involving in thinking, we have to find something else might turn out to be at the molecular level in the internal level of the neuron or in a lower level with massive possibilities of computation.

B. Keating, 20 de septiembre de 2023, Noam Chomsky on AI, Neural Networks, and the Future of Linguistics [Archivo de Vídeo]. Youtube. https://www.youtube.com/watch?v=lkRft71_JrY

En definitiva, el autor nos invita a mantener la disciplina intelectual según el modelo de la ciencia del siglo XVII, momento en el cual el pensamiento especulativo adopta el principio básico de la coherencia racional. Según su propuesta, para tender puentes entre uno y otro campo, el de las facultades humanas y las de la IA se debe situar con precisión cuál es el nivel en que nuestras afirmaciones son válidas, cuáles son los principios base en que soportamos las afirmaciones y lo que es fundamental, cuál es el alcance de los conceptos. Usar la misma palabra no implica aludir a la misma idea ni al mismo tipo de eventos. Antes de lanzarnos de manera apresurada a responder los grandes interrogantes, sugiere Chomsky, se debería reflexionar de manera consistente: antes debemos hacer ciencia que preocuparnos por la ingeniería. A pesar de las utilizaciones locales de gran incidencia en la lingüística computacional, el concepto general de los proyectos de la lingüística chomskiana y de los planteos de las IA es divergente. En definitiva, afirma, las IA no tienen la arquitectura adecuada para sustentar algo análogo al pensamiento humano. El tiempo nos dirá si las arquitecturas actuales de las IA u otras que puedan desarrollarse convergerán o no en formas cercanas al pensamiento y la capacidad simbólica humana. Por lo pronto contamos con la respuesta negativa de un gran lingüista: Chomsky apuesta en contra.

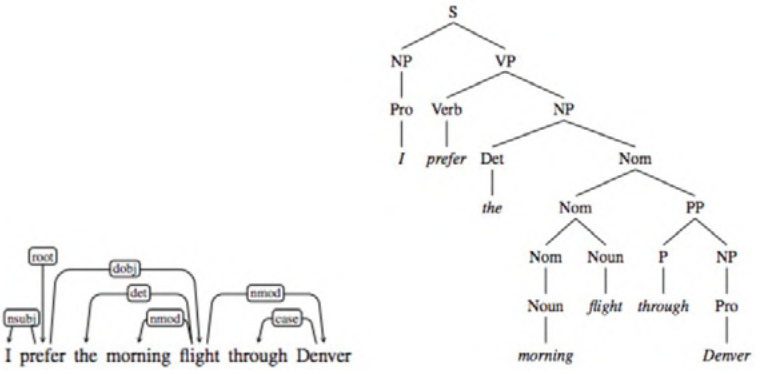


Figura 3. Utilizaciones de los modelos gramaticales de Chomsky en la lingüística computacional. Comparison between constituency tree and dependency tree (Tomado de Jurafsky y Martin, 2023).

Bibliografía

Bachelard, G. (2000). *La Formación del Espíritu Científico: Contribución a un Psicoanálisis del Conocimiento Objetivo*. Siglo XXI.

Beck, F. y J. C. Eccles (1992). Biophysics Quantum aspects of brain activity and the role of consciousness. *Proceedings of the National Academy of Sciences of the United States of America, 89*(23), 11357-11361.

Birchenall, L. F. y O. Müller (2014). *La Teoría Lingüística de Noam Chomsky: del Inicio a la Actualidad*. Fundación Universitaria Los Libertadores. Universidad del Rosario Bogotá.

Chomsky, N. (1972). *Lingüística cartesiana*. Gredos.

Chomsky, N. (1985). *Reflexiones sobre el lenguaje*. Planeta Agostini.

Chomsky, N. (1986). *El lenguaje y el entendimiento*. Planeta Agostini.

Chomsky, N. (2003). *On Nature and Language*. Cambridge University Press UK.

Chomsky, N. (2007). Symposium on Margaret Boden, mind as machine: a history of cognitive science. *Artificial Intelligence, 171*, 1094-1103.

Chomsky, N. (2011). Language and Other Cognitive Systems. What Is Special About Language? *Language Learning and Development, 7*, 263-278.

Dehaene, S. (2016). *El cerebro matemático*. SXXI.

Everaert, M. B. H., M. A. C. Huybregts, N. Chomsky, R. C. Berwick y J. J. Bolhuis (2015). Structures, Not Strings: Linguistics as Part of the Cognitive Sciences. *Trends in Cognitive Sciences, 19*(12), 729-743. http://dx.doi.org/10.1016/j.tics.2015.09.008

Gallego, A. (2008). La jerarquía de Chomsky y la facultad del lenguaje: consecuencias para la variación y la evolución. *Teorema: Revista Internacional de Filosofía, 27*(2), 47-60.

Jurafsky, D. y J. H. Martin (2023). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3ra ed. Stanford University.

Volver

Kauffman, S. A. (2009). Physics and Five Problems in the Philosophy of Mind. arXiv:0907.2494 [physics.hist-ph]

Keating, B. (2023). *Noam Chomsky on AI, Neural Networks, and the Future of Linguistics*. Youtube. https://www.youtube.com/watch?v=lkRft7l_JrY

Popper, K. (1998). *El realismo y el objetivo de la ciencia*. Tecnos.

Popper, K. (1999). *La lógica de la investigación científica*. Tecnos.

Russell, S. J. y P. Norvig (2004). *Inteligencia artificial. Un enfoque moderno*. Pearson Educación.

Capítulo 3. La inteligencia humana y el constructivismo

Sumario

Existen muchas formas de conceptualizar la inteligencia según el enfoque teórico que se adopte. En este capítulo presentaremos el constructivismo piagetiano, una de las teorías dominantes en la comprensión de la capacidad de inteligencia humana. Describiremos la manera en que Piaget desarrolla su idea de inteligencia como una interacción dinámica entre las estructuras mentales y el medio ambiente. Veremos cómo la inteligencia implica una evolución genética a través de estructuras cada vez más eficaces y equilibradas relacionadas con la célebre idea de Piaget de los estadios. Destacaremos la noción de operaciones lógicas en el constructivismo y la manera en que nos permite comprender las conductas inteligentes como una forma de ordenar el mundo. Introduciremos la noción de descentramiento como abandono de la posición egocéntrica en la constitución de la inteligencia y el pensamiento. Una cosa es realizar una operación inteligente y otra diferente es tomar conciencia de la manera en que la realizamos. Para dar cuenta de ello repasaremos la noción de toma de conciencia y su relación con el concepto constructivista de abstracción reflexionante. Para concluir nos detendremos en aquellos modelos biológicos vinculados a la genética que sirven de inspiración al constructivismo de Piaget.

Piaget, un científico con muchos intereses

Nos aferramos a nuestras certezas. Muchas veces se nos dificulta cuestionar aquellas ideas con las que hemos sido formados. Ellas se llegan a convertir en parte de nosotros y de nuestra identidad. Algo así ocurre con la obra de Piaget y con la noción de constructivismo en general. La palabra misma "constructivismo" se ha convertido en un adjetivo positivo, al menos en los ámbitos pedagógicos y en ciencias sociales. Ahora bien, los principios del constructivismo estructural, columna central de la epistemología genética, chocan de frente con las orientaciones asociacionistas del cognitivismo. Esta fricción nos interesa en particular, por la estrecha relación de las ciencias cognitivas y los desarrollos de las IA. Cuando se afirma que las IAs toman como modelo la inteligencia humana y afirmamos que intentan generar agentes que puedan actuar como si contaran con facultades superiores humanas, nos estamos refiriendo a una manera particular de concebir la inteligencia. A veces tendemos a olvidar la diferencia entre los objetos de nuestra intuición espontánea y los objetos de conocimiento. Tendemos a suponer que sabemos de qué hablamos cuando pensamos en una cuestión como es la inteligencia. Existen muchas teorías en torno a la inteligencia. Esas teorías no se refieren a una cosa única y real que pudiéramos medir, tocar, experimentar mediante nuestros sentidos y definir de manera directa. No existe ese punto de referencia, esa cosa que pudiéramos mirar desde distintas perspectivas, este es un principio epistemológico básico. Cuando adoptamos uno u otro camino en la definición de la inteligencia estamos generando diferentes entidades en el plano del conocimiento. Debemos aceptar que aquello a lo que se llama inteligencia desde una u otra perspectiva son entidades totalmente distintas y, de hecho, muchas veces incompatibles. Nuestro sentido común ingenuo nos suele engañar y llevar a suponer que cuando le ponemos un nombre a una entidad alcanza para saber de qué se trata. Es como si fuera suficiente usar el término inteligencia para garantizar que siempre hablamos de lo mismo. Pues no, cuando los constructivistas hablan de inteligencia aluden a cuestiones incompatibles en gran

medida con aquello a lo que se refieren los cognitivistas. Son dos territorios en pugna, en torno a la noción de inteligencia existe una contienda.

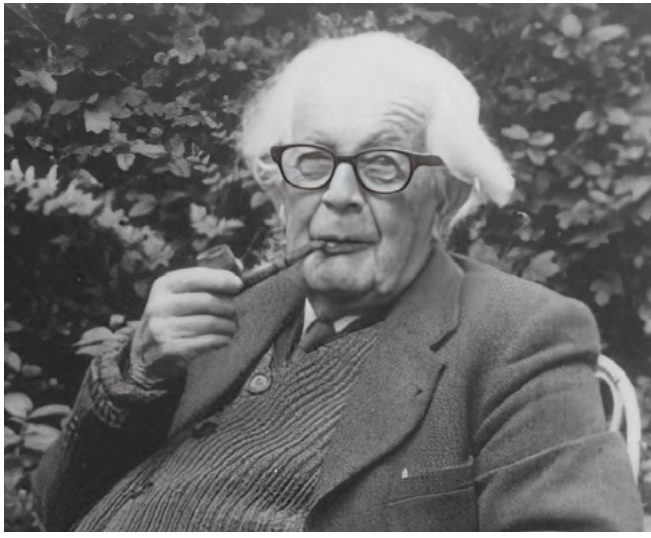


Figura 1. Jean Piaget es el padre del estructural constructivismo en psicología. Se opone a la génesis sin estructura del empirismo atomista y a la estructura sin génesis del formalismo innatista. Para él la inteligencia surge del juego entre génesis y estructura. Imagen extraída de: https://www.actualidadenpsicologia.com/biografia/cronologia-jean-piaget/

Cualquiera de los caminos que elijamos para explicar las funciones superiores de la inteligencia, la memoria o el razonamiento, nos obligará a aceptar un montón de cosas. No es un defecto o una virtud específicas de estas corrientes sino un hecho general de las ciencias. Constructivistas y asociacionistas acuden al auxilio de modelos provenientes de otras disciplinas científicas. Ninguna ciencia con referencia a hechos empíricos permanece aislada, no hay fronteras cerradas. Detectar los vasos comunicantes entre las disciplinas, los préstamos de conceptos y de modelos nos acerca a la comprensión del sentido mismo de una ciencia. Otra forma en que podemos explorar el sentido y la lógica de una corriente científica o de pensamiento es detectar cómo enlaza sus propósitos con su estrategia de abordaje. Este abordaje implica dos dimensiones relacionadas, la teórica y la de los fenómenos empíricos que aspira analizar. Empecemos por ver cómo procede el exponente principal del abordaje constructivista de la inteligencia humana J. Piaget. ¿En qué se interesa, cuáles son sus propósitos como investigador? A Piaget parecía interesarle todo, la biología, la psicología, la lógica, la epistemología, la sociología, era, antes que nada, una mente curiosa (Piaget, 1976). Es uno de esos casos que surgen en ocasiones en la ciencia en

que alguien logra integrar conocimientos articulados en varias disciplinas y ensamblarlos en una propuesta coherente. Supo encontrar principios generales aplicables a muchos fenómenos provenientes de reinos diferentes. Hallaba patrones unificadores, lograba atravesar la inmensa diversidad de los fenómenos. Enlazaba cuestiones aparentemente tan distantes como son la mutación de los moluscos, el manejo de los objetos por parte de los niños, los errores sistemáticos en la lecto escritura o la evolución epistemológica de las ciencias a través de la historia. Era una de esas mentes capaces de penetrar el velo de la diversidad empírica y vislumbrar estructuras subyacentes. Se alimentaba a partir de muchas fuentes de saber, pero si nos viéramos forzados a destacar una por sobre las otras, el lugar de privilegio lo tendría la lógica. En su teoría la idea lógica de equilibrio de las estructuras ocupa un lugar predominante. Postula que la vida es un despliegue en el tiempo de estructuras sucesivas que apuntan hacia el equilibrio. Esta noción, la de equilibrio, le permite explicar la manera en que se estructuran los organismos vivos para adaptarse a las presiones del ambiente.

La inteligencia en la organización mental

¿Cuál es el papel de la inteligencia en la organización mental? ¿Cómo se relaciona y diferencia de otras formas de equilibrio propias del orden de la vida? Los humanos, como todo ser vivo, están enclavados en un ambiente y despliegan conductas hacia el mismo. Entre el ambiente externo y el organismo se producen intercambios permanentes en una relación dinámica. Aquí entra a jugar un factor que podemos denominar "energético", que es la necesidad como motor de la conducta. Desplegamos acciones hacia el medio ambiente impulsados por una necesidad, cuando el equilibrio se encuentra roto. Se denomina acción al despliegue del individuo tendiente a reestablecer el equilibrio con el medio ambiente. Aquí Piaget hace una distinción interesante entre la psicología y la fisiología. Plantea que, en la fisiología la reacción del organismo es material e implica una transformación a nivel orgánico. Ante un desequilibrio el individuo reacciona mediante una transformación material e interna del cuerpo. Es lo que nos pasa, por ejemplo, cuando nos

sentimos amenazados o cuando nos sofoca el calor. Ante una amenaza se estimulará la secreción de adrenalina y cortisol o ante un desequilibrio térmico pueden dispararse diferentes dispositivos homeostáticos termo reguladores que involucran al hipotálamo, el sistema endócrino y desencadenan eventos tales como la vasodilatación periférica o la sudoración. El registro mental y psicológico de fenómenos tales como el estrés (Figura 2) nos permiten registrar la diferencia entre estos campos. A diferencia de estas reacciones fisiológicas que son materiales y a nivel orgánico, las acciones psicológicas se verifican a distancia y son de tipo funcional (Piaget, 1991).

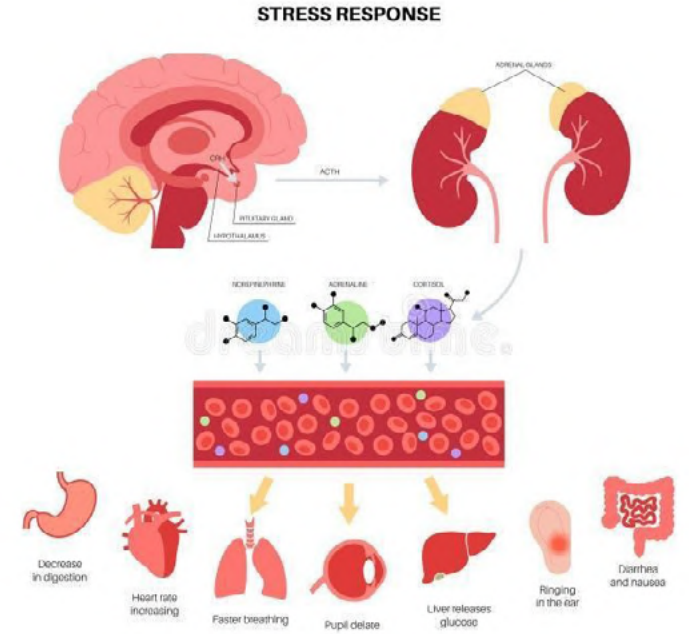


Figura 2. Los intercambios de los seres vivos con el ambiente apuntan al equilibrio. En este gráfico se aprecian las reacciones fisiológicas vinculadas con situaciones desequilibrantes como el estrés. Mientras que en el plano fisiológico las reacciones orgánicas implican cambios materiales e internos, el plano psicológico opera a distancia. La dimensión cognitiva implicará un apartamiento cada vez mayor del mundo material hasta el punto de operar según relaciones abstractas sin necesitar la percepción sensorial de los objetos en el ambiente inmediato. Imagen extraída de: https://sistema-de-respuesta-estrés-síntomas-y-reacción-emergencia-en-el-cuerpo-humano-eje-suprarrenal-hipotálamo-glándulas-image235377568

Solemos identificar a Piaget con la descripción de las acciones de los niños, lo cual es correcto, pero nos brinda una idea limitada acerca de su proyecto. Desde este punto de vista, se asocia su trabajo con la enumeración de las cosas que puede o no puede hacer un niño de acuerdo a la edad. Sus aportes han servido de guía a varias generaciones de educadores que han hecho uso de la idea de estadios de desarrollo por los que atraviesan los alumnos en etapas sucesivas. Nuestro interés es otro, el de rescatar el papel de la lógica en las operaciones de la inteligencia humana. Piaget intenta descifrar, en forma no causal, la manera en que se construye la lógica del mundo. Nos enfrentamos a una realidad en apariencia caótica y, mediante nuestras operaciones intelectivas le damos forma. Nuestras acciones, aún las de los niños más pequeños no se despliegan al azar de manera desordenada. No son algo inconexo, sino que responden a una forma de organización. La tarea a la que se dedica Piaget es la de descifrar esta organización. Organizando el mundo externo nos organizamos internamente. Es el famoso juego entre asimilación y acomodación. Las nuevas experiencias se asimilan a los esquemas con los que cuenta el individuo. La inteligencia implica generalizar. A su vez, la novedad de las experiencias presiona sobre las estructuras existentes e impulsan su acomodación. Este proceso constructivo y tendiente al equilibrio, permitirá un manejo cada vez más sofisticado del medio ambiente. Las estructuras intelectivas con que cuenta el individuo le permitirán una generalización de la experiencia. Ordenamos el mundo de manera inteligente porque contamos con estructuras que nos permiten asimilar las experiencias nuevas con las que nos enfrentamos. Somos seres formadores y organizadores, damos lógica al mundo con el que interactuamos. Pero estas estructuras cambian a lo largo de nuestra historia vital como individuos. Estos cambios, que se apoyan en la base de nuestra materialidad orgánica, se producen por la manipulación de nuestro entorno. A la manipulación del entorno según criterios lógicos se la llama operaciones. Mediante las operaciones que efectuamos sobre el entorno realizamos una doble construcción, la del mundo exterior y la de nuestras estructuras mentales. Ordenando nos ordenamos (Piaget, 1977).

La interacción con lo real plantea desafíos para el individuo, algunos de los cuales podrá solucionar mediante operaciones cognitivas (Figura

3). Puede hacerlo, porque cuenta con estructuras que le permiten incorporar experiencias. En el plano mental, así como en el físico, el individuo trata de mantener un equilibrio. Algunos de los objetivos del individuo requieren poner en acción operaciones lógicas. Ordenar, discriminar, agrupar, comparar, combinar y resolver muchas cuestiones muy sencillas del orden de la vida real conllevan verdaderos problemas cognitivos. Las experiencias implican diferentes grados de presión sobre nuestras capacidades, generan desequilibrios que podrán o no ser resueltos con los recursos con los que contamos. En diversos grados de desarrollo evolutivo, la batería de recursos disponibles estará más o menos desarrollada.

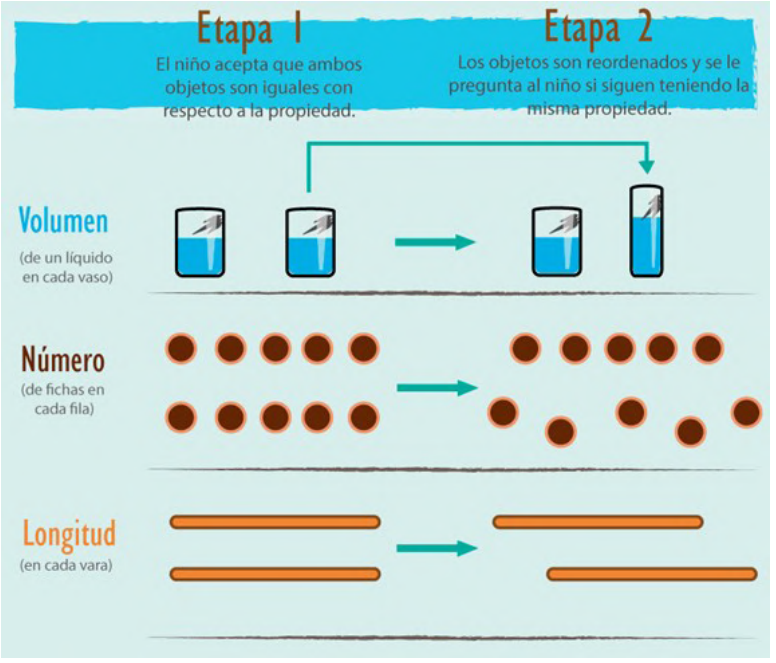


Figura 3. Los intercambios cognitivos son de tipo funcional y no material como en el caso de los fisiológicos. El individuo se independiza cada vez más del mundo físico a través de sus etapas de desarrollo. En el estadio de las operaciones concretas necesitará la presencia de índices perceptivos para manipular las relaciones, en el operatorio formal sus operaciones se efectúan en un plano mental de abstracción. En un plano superior, se podrán manipular relaciones sin necesidad de índices perceptivos ni manipulación de objetos materiales. (Piaget, 1978) Imagen extraída de: https://www.actualidadenpsicologia.com/etapa-de-las-operaciones-concretas/

Volver

Cuando el problema no genera desafíos, el interés del individuo será bajo y lo resolverá de manera más o menos automática. Son los llamados estados alfa, momentos en que las experiencias que plantea el ambiente pueden ser asimiladas sin problema a las estructuras existentes. Pero llega un punto en que el mundo se comienza a tornar contradictorio, en el que las experiencias comienzan a "hacer ruido". Las soluciones con que contamos ya no nos convencen tanto. Cuando el problema exceda en un alto grado las capacidades con que se cuenta para resolverlo el individuo ni siquiera lo registrará, tratará de ignorarlo, se mantendrá indiferente o tendrá reacciones de rechazo. Pero existe un punto en que las cosas, "nos ponen a pensar", es un punto justo en que las estructuras con que contamos se han tornado inestables por la presión de múltiples experiencias y han comenzado a movilizarse generando ensayos e hipótesis de resolución. Nos encontramos en un estado beta, en que las estructuras deben movilizarse para poder incorporar experiencias para las cuales la organización estructural con la que se cuenta ya no responde de manera satisfactoria. Este inter juego avanza en un espiral ascendente de estructuras sucesivas. Es la consabida historia de la sucesión de etapas en la génesis de la inteligencia.

Actuamos sobre el mundo que nos rodea, le damos forma. Le podemos dar una forma porque somos dadores de formas. El acto mismo de la inteligencia es concebible porque las acciones siguen patrones organizados lógicamente. Las estructuras de la inteligencia son estructuras lógicas de equilibrio. Esto merece una explicación, por eso queremos introducir la idea de grupo, una noción utilizada por el estructuralismo y una base importante en las teorías de autores como Leví-Strauss, Jackobson y Lacan entre otros. Piaget, como Leví-Strauss y muchos otros exponentes de la corriente conocida como estructuralismo, se interesaban por la matemática y la lógica. De hecho, eran apasionados por un tipo particular de matemática llamada álgebra abstracta que no se interesa tanto por las cantidades o los números, sino por las relaciones y las operaciones.

Tanto en matemáticas como en psicología, es decir en las dos disciplinas entre las cuales se intercala la lógica, el papel de las totalidades operatorias con sus propiedades de conjunto ha llegado a ser fundamental, tanto en la sistematización de las operaciones abstractas como de las operaciones reales en juego en el pensamiento en acción.

J. Piaget, 1977, Ensayo de lógica operatoria, p. 48.

Piaget y Bourbaki: una asociación casi obvia

Piaget es el impulsor de una corriente psicológica conocida como estructural constructivismo. En este apartado nos interesa introducir el costado "estructural" de sus intereses. El estructuralismo trajo consigo un estallido de entusiasmo, durante un período de poco más de una década se tuvo la sensación de haber conquistado el lenguaje, la comunicación, la lógica del mundo. París era una fiesta. Esas certezas o ilusiones cuyo origen podemos situar en la década de los cincuenta y tuvieron su eclosión en la de los sesenta, tendieron a declinar hacia fines de los setenta y se diluyeron hacia los ochenta. Esto fue posible por la confluencia de dos tipos de saberes, el que provenía de las llamadas ciencias del lenguaje, en particular el modelo fonológico y las disciplinas matemáticas, en particular el álgebra. Nuestro autor fue un protagonista clave en este movimiento a partir de sus aportes a la psicología de la inteligencia (Piaget, 1974). La declinación de las expectativas del estructuralismo en su intento por conquistar el universo de la significación por las vías de la racionalidad coincide con una fuerte reacción irracionalista y por el florecimiento de las llamadas corrientes post estructurales (Dosse, 2004). Sin embargo, el estructural constructivismo piagetiano no perdió fuerza como así lo hicieron otros emprendimientos estructuralistas. Sus planteos dinámicos mantuvieron su vigencia y siguieron ejerciendo influencia en el terreno del estudio de la inteligencia y sobre todo en los ámbitos educativos.

En el corazón mismo de la corriente estructural se sitúa el concepto de grupo matemático. Tras esta noción se esconde un misterioso personaje Poldavo: Nicolás Bourbaki. Si no se han interesado por la matemática es posible que no hayan escuchado su nombre, pero pueden estar seguros que sin su influencia el estructuralismo no hubiera tomado la forma que tomó. Este personaje de origen Poldavo nunca existió. Tampoco lograremos encontrar en el mapa el reino imaginario de Poldavia. Se trata de un pseudónimo inventado por un grupo de matemáticos que revolucionaron su disciplina. Esta denominación fue creada como una broma, ya que el sentido del humor es un rasgo de inteligencia, algo que abundaba entre

estos intelectuales entre los que se encontraron H. Cartan, J. Dieudonne, S. Mandelbrojt y A. Weil, entre muchos otros. Su intento era el de unificar la matemática y formalizar sus axiomas centrales. Crearon un pseudónimo para mantener el anonimato ya que consideraban que la matemática no era cuestión de nombres propios sino de ideas. Esto generó un halo de misterio. Su gran herramienta fue la teoría de grupos. El imaginario Bourbaki es una referencia constante en la obra de pensadores como Lévi-Strauss, Piaget y Lacán. Cada uno en su disciplina establece intercambios productivos con el Grupo Bourbaki verdadero rock star en los círculos científicos de su época (Figura 4).

Figura 4. Algunos de los integrantes del grupo Bourbaki. En el medio del grupo con lentes graduados se ubica André Weil. Él fue quien aportó la base matemática de simetrías en las relaciones elementales del parentesco a C. Levy Strauss (Fresán, 2011). Imagen extraída de: https://elpais.com/ciencia/2020-12-01/el-colectivo-secreto-que-cambio-la-educacion-en-matematicas.html

La referencia a este conjunto de matemáticos reunidos bajo el seudónimo es una constante en Piaget. El respeto y la admiración es mutua, ya que los Bourbaki veían en Piaget una confirmación empírica y psicológica de sus planteos en torno a la relevancia de la teoría de los conjuntos y la noción de grupo matemático como articulación entre operaciones reversibles. Aquello que imaginaban como gran tarea de unificación de la matemática podría recibir una confirmación experimental en la obra de Piaget. La formalización ganaba espacio contra las corrientes intuicionistas y empiristas. En esta distribución de tareas cada uno cumpliría su papel. Los matemáticos refinarían las estructuras con el sueño de simplificar los fundamentos de la matemática real y viva de su época (Bourbaki, 1976). Sus socios los psicólogos constructivistas, comandados por Piaget, explorarían la forma en que las estructuras lógicas se manifiestan en las operaciones de las acciones reales de los individuos (Hernandez, 2002). No se trataría, por supuesto, de un calco en el plano psicológico de las operaciones lógicas elucubradas por los matemáticos (Castorina, 1981). Sería una lógica de operaciones sometidas al entorno del mundo real y actuadas por entidades biológicas humanas. A esto se le dio el nombre de lógica natural y su máxima expresión formalizada es el grupo matemático de Klein Piaget, el célebre grupo INRC que caracteriza las capacidades operatorias abstractas de la inteligencia ya consolidada del adulto (Piaget, 1977).

Una de las cuestiones que más rechazo puede generar ante el avance de la investigación y desarrollo de las IA es su relación con la matemática y las formalizaciones. Se vive su irrupción como un proceso de deshumanización vinculada al mundo de los números y de la fría lógica. Sin embargo, la vinculación de la lógica y los procesos formales siempre fue un elemento central cuando se puso sobre la mesa la cuestión de la forma en que procede la inteligencia humana. Es muy difícil plantear una dimensión sin acudir a la otra. Mediante las IAs se intenta diseñar e implementar agentes que cumplan tareas que suponen facultades similares a las de los seres humanos para el cumplimiento de sus objetivos. Sin el auxilio de las herramientas lógicas y las matemáticas no hubieran podido lograr algo semejante. Sin acudir al álgebra lineal, al álgebra de funciones, a diferentes tipos de lógicas, a la posibilidad de agrupar,

discriminar, realizar operaciones reversibles o conectar un elemento con otro en retículas, la inteligencia artificial no hubiera sido una empresa factible. En algunas de sus estrategias de diseño, en particular el de redes neuronales, tomarán el modelo de la biología como inspiración. Pero, en definitiva, ¿qué pretenden que pueda realizar una IA? Las máquinas de este tipo procesan números y las operaciones numéricas sólo podrán aportar enlaces cuyo sentido sea comprensible a la luz de las matemáticas. Otro tanto pensaba Piaget. Para él no tenía sentido la reflexión en torno a la inteligencia si no se tomaba como referencia su punto de llegada: las estructuras reversibles de inteligencia adulta. Esto implica pararse en el punto de llegada, el de la estructura desarrollada de la inteligencia y rastrear hacia atrás la aparición de los elementos que la hacen posible.

La manera de plantear la relación entre lógica e inteligencia puede variar significativamente. El enfoque constructivista nos obliga a pensar la inteligencia desde una perspectiva de totalidad. Su estrategia es la de avanzar de lo abstracto a lo concreto y no a la inversa. Cuando queremos descifrar cuáles son los procesos que permiten que los humanos arribemos a la inteligencia adulta, podemos adoptar diferentes estrategias. La primera, de corte evolucionista y empirista tratará de ver cómo se adquieren progresivamente las capacidades conforme los niños crecen, hasta que llegan a su punto máximo de desarrollo. La opuesta, adoptada por Piaget, parte del análisis lógico de la inteligencia en su punto más desarrollado. Lo central será detectar la manera en que se enlazan las operaciones en una estructura de conjunto.

La exactitud de la abstracción reflexiva que caracteriza al pensamiento lógico matemático es la de ser sacada no de los objetos, sino de las acciones que se pueden ejercer sobre ellos y más esencialmente de las coordinaciones más generales de estas acciones, como las de reunir, ordenar, poner de acuerdo, etc.

J. Piaget, 1974, El estructuralismo, p. 18.

La inteligencia y el pensamiento

A diferencia de las IA que carecen de pensamiento, la inteligencia humana se articula con la capacidad representativa. Las funciones corporales de tipo fisiológico se ejercen en el proceso de contacto material, en tanto que las psicológicas tienen lugar cada vez a mayor distancia. La inteligencia le permite al ser humano independizarse de la materia y resolver sus operaciones en el plano mental. Nos encontramos con dos series de transformaciones que corren articuladas y en paralelo. En un plano, el de las operaciones sensorio motoras, la experiencia de manipulación y los circuitos establecidos entre el sujeto y el objeto, permiten la interiorización progresiva de la lógica del mundo físico. Así se avanza desde la plataforma de los reflejos involuntarios, los llamados circuitos primarios referidos al propio cuerpo, hasta formas superiores de organización del espacio y el tiempo. Es muy conocida la secuencia de estructuras progresivas de asimilación y acomodación con sus equilibrios cada vez más reversibles y estables. En la obra de Piaget el rol predominante lo tiene la constitución de sistemas operatorios que se ponen en juego en la matematización y logización provenientes de lo real.

En un principio el individuo refiere todo a su propio cuerpo. Establece circuitos en que todo se remite a ese poderoso centro. El mundo comenzará siendo algo succionable, luego será algo agitable, el eje mayor es la referencia a ese cuerpito que comienza a crecer y que todavía carece de autonomía motriz. En el plano del pensamiento ocurre otro tanto, la primera referencia es uno. A eso se le llama egocentrismo. Solo progresivamente nos vamos apartando de ese eje para poder adoptar otros puntos de vista. El egocentrismo es una poderosa fuerza que se expresa en el plano del pensamiento. Tal vez habrán escuchado alguna expresión egocéntrica por parte de los niños. Tienden a sentir que son el centro del universo, es por ello que pueden llegar a pensar o a decir que la luna los sigue a ellos mientras caminan. El egocentrismo, esa poderosa tendencia, expresa la dificultad que tiene el ser humano para ubicarse como una cosa entre las cosas, una entidad entre las entidades. La adopción de la perspectiva exterior a uno mismo implica un proceso constructivo de reversibilidad. Es necesario asumir que somos visibles desde aquellos

puntos a los que miramos. El apartamiento representacional, la formación del símbolo en el niño tiene cierto grado de independencia respecto de las operaciones lógicas (Piaget, 1966), a su vez esta toma de distancia se articula con la evolución de otra importante facultad la de memoria y su articulación con la inteligencia (Piaget, 1978). Inteligencia, capacidad simbólica, razonamiento lógico, memoria, capacidades y facultades que estarán en el centro de los intentos por replicar la inteligencia humana por medios artificiales. Pero volvamos a Piaget antes de abordar estos temas en relación con las IAs.

El pensamiento tiene cierto grado de independencia con relación al surgimiento de la inteligencia. Para Chomsky, como pudimos ver en un capítulo anterior, el lenguaje y la capacidad simbólica son los grandes impulsores del desarrollo de las facultades humanas. No obstante Piaget, defensor acérrimo de la lógica, supone que son las operaciones las que preparan el terreno del pensamiento y el lenguaje y no a la inversa. Un tema polémico si los hay. Para este autor, las operaciones sensorio motrices son las que allanan el camino para que se monten otros procesos como los representacionales, este será un punto de polémica con Chomsky y sus planteos, que Piaget descarta por innatistas (Piaget, 1966, 1974). El pensamiento simbólico nos independiza del objeto externo al que accedemos mediante los sentidos y podemos jugar con su representación mental.

El punto culminante del proceso será la abstracción reflexionante. Cuando entra en juego la capacidad de abstracción reflexionante podremos volver sobre nuestras acciones y sobre nuestro pensamiento y reconstruir la lógica subyacente. En este nivel ocurren varias cosas: primero la formulación de un objetivo puntual, segundo la ejecución de operaciones orientadas a lograr ese objetivo, tercero la posibilidad de desandar mentalmente la serie de pasos efectuados y, por último, la posibilidad de expresar simbólicamente a través del lenguaje (natural o formal) la totalidad del proceso y su lógica. Una cosa es llevar adelante una acción inteligente y otra bien distinta acceder a la lógica mediante la que efectuamos las operaciones. De hecho, haciendo un paréntesis, podemos señalar que este es el punto en que una IA podrá convertirse en inteligencia artificial general (AGI). Hacia noviembre de 2023, esta es una cuestión

aún no resuelta, nos ocuparemos del tema en próximos capítulos.

Volvamos a la inteligencia humana y su relación con la formación del símbolo en los niños. El juego y la imitación serán actividades simbólicas que implican la representación de un objeto ausente y se articulan con la evolución de las estructuras de la inteligencia. En esto también se opone a las concepciones asociacionistas que derivan todas las representaciones de las imágenes sensoriales como fuente primaria. Como en todos los niveles encuentra procesos dinámicos entre el medio interno y el ambiente e intenta rastrear el equilibrio entre los mismos. Sin entrar en detalles señalemos que el juego y la imitación serán actividades representativas que preparan la llegada del pensamiento abstracto (Figura 5). El pensamiento abstracto nos permitirá realizar operaciones creativas basadas en la capacidad inteligente de imaginar relaciones de enorme complejidad con independencia de la intuición perceptiva de entidades materiales. De hecho, estas relaciones que creamos e imaginamos pueden contradecir nuestra percepción ingenua e inmediata. En el plano mental podremos jugar con estructuras de pensamiento puro, ejecutar operaciones inteligentes que tienden a equilibrios puramente mentales. La diferencia entre la mejor abeja y el peor arquitecto es que el arquitecto podrá imaginar y crear en su mente una forma para luego ejecutarla en el mundo externo, en tanto que la abeja deberá repetir una forma que la preexiste. La inteligencia es una pieza central en el acto de libertad humana.

Esta noción de grupo, señalábamos, tiene relación con el costado estructuralista de Piaget. Pero ¿qué implica específicamente con relación a la génesis y adquisición de capacidades inteligentes en los seres humanos? Para responderlo deberemos introducir el otro costado de las preocupaciones de esta corriente psicológica: el constructivista. La obra del autor se asocia con dos denominaciones, la de estructural constructivismo y la de epistemología genética. Ambas denominaciones nos aclaran varias cuestiones con relación a la inteligencia humana, su estructura y la forma en que se relaciona con un proceso genético.

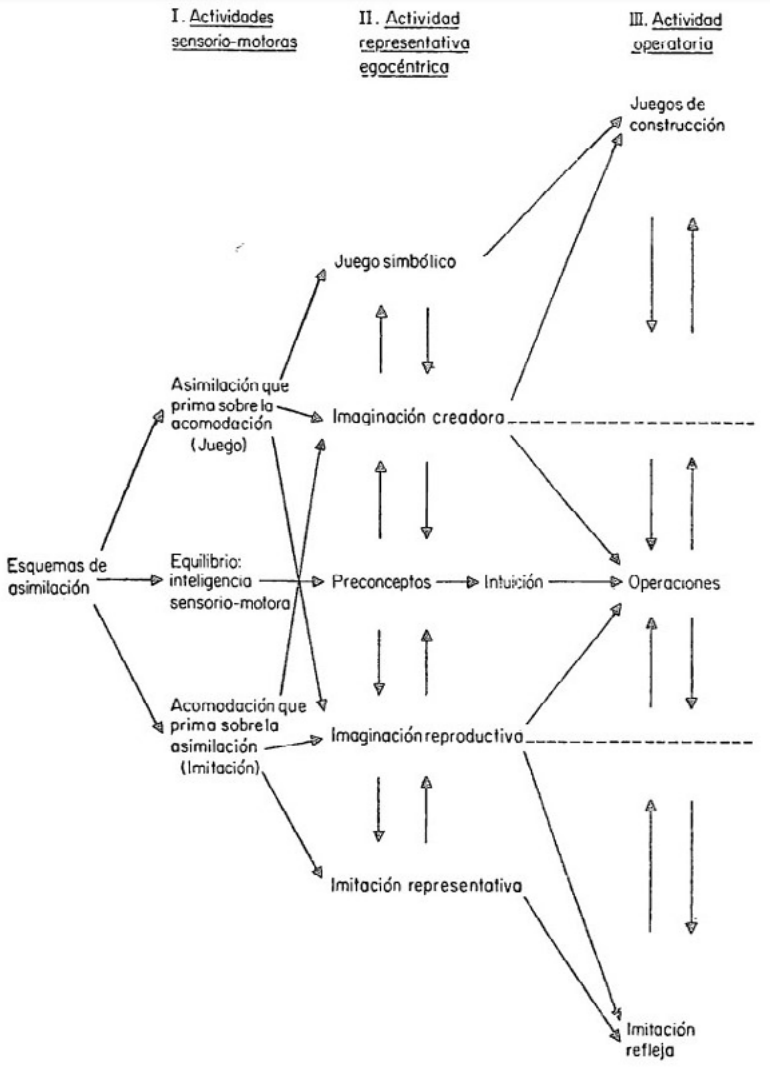


Figura 5. J. Piaget. La formación del símbolo en el niño. Página 395. Etapas generales de la actividad representativa.

Volver

El modelo biológico en Piaget

La forma en que se incorpora la referencia a los procesos vitales y a la biología en el estructural constructivismo de Piaget para elaborar sus nociones acerca de la inteligencia, es totalmente diferente a la que se adoptará en el cognitivismo. Detectar estas diferencias nos ayudará a reflexionar sobre el alcance de las teorías sobre la inteligencia y la influencia del cognitivismo como fuente de inspiración en el desarrollo de las IAs. Constructivistas y cognitivistas mantienen una polémica en cuanto a su comprensión de la noción de inteligencia cuyos alcances nos parecen muy significativos. Estas polémicas ya se hacen sentir en el mundo pedagógico y tienen, incluso, un alcance mediático en forma de notas que nos informan acerca de la manera en que se debería enseñar en la era de la IA. Antes de opinar en uno u otro sentido, sería interesante plantearnos cuáles son las bases de estas polémicas con las que convivimos y conviviremos seguramente en un futuro inmediato. A continuación, vamos a repasar la forma en que Piaget retoma los modelos de la biología evolutiva.

Varias pasiones confluyen en la obra de Piaget, pero su primer interés fue la zoología y la evolución de las especies. De manera temprana, apenas un bachiller, se involucró con un concepto que lo acompañaría el resto de su vida: el de la evolución. Según relata en su autobiografía sus exploraciones juveniles sobre las mutaciones de los moluscos y la lectura de las teorías biológicas, sumadas al interés por la teoría del conocimiento y su rechazo de los dualismos mente – cuerpo, cimentan lo que se transformaría en el planteo de la epistemología genética (Piaget, 1976). La epistemología genética tiende un puente entre la evolución biológica, la inteligencia humana y el desarrollo de las ideas científicas (Piaget, 1978a, 1987). El campo de la biología es muy amplio y, por lo tanto, señalar que se retoman modelos de esta disciplina no aclara demasiado. De hecho, otras corrientes muy diferentes como son las conductistas y cognitivistas también lo hacen. Por lo tanto, debemos aclarar qué aspectos y qué modelos biológicos le sirven de marco de inspiración a cada corriente. Veamos cuál es la manera en que el conductismo, el cognitivismo y el estructural constructivismo se apoyan en modelos biológicos.

Volver

El conductismo toma como modelo la noción de reflejo, de reacciones y de comportamientos a partir de la idea básica de estímulo - respuesta. En este sentido plantea una continuidad entre el comportamiento de la especie humana y otras especies del reino animal. Tengamos en cuenta que para el conductismo lo importante es la historia del individuo. Una conducta se genera como resultado de las experiencias que hayan tenido los individuos en relación con el medio. El conductismo es una teoría de la historia de las experiencias de organismos individuales enclavados en medio ambientes específicos. El éxito o fracaso de las acciones emprendidas por el individuo en el pasado influye en la adopción de conductas posteriores. Ante una circunstancia con la que nunca se ha enfrentado se desarrollarán conductas de respuesta de manera relativamente azarosa, no se tiene precedentes y se actúa por tanteos. Esto, plantea el conductismo, es así para cualquier especie animal. Si las conductas tienen resultados positivos para el individuo se dice que "han sido reforzadas positivamente" y la posibilidad de repetirlas en el futuro se incrementará. Pero, el medio ambiente no siempre reforzará nuestras conductas de manera positiva, en ocasiones nos va bien, en ocasiones nos va mal. Si las conductas que desarrollamos son "penalizadas" por el medio ambiente en que estamos insertos, la posibilidad de repetirlas será menor. El conductismo, con gran espíritu práctico, renuncia a utilizar nociones tales como motivaciones o deseos, como forma de explicar la conducta humana. Se sitúa en el nivel de los estímulos y respuestas. Para ellos es importante poder registrar, medir y experimentar en psicología como en cualquier otra ciencia. Renuncian a las explicaciones mentalistas ya que consideran que se postulan entidades puramente especulativas en torno a la mente. Para ellos, entre estímulo y respuesta, entre input y output se sitúa la famosa "caja negra". La idea de caja negra supone que entre dos fenómenos observables como los estímulos del medio ambiente y las conductas existe una instancia intermedia que no podemos reconstruir. Se trata de funciones a las que no tenemos acceso. En lugar de hacer complejas construcciones sobre aquello que ocurre en esta instancia intermedia, el conductismo adopta la estrategia de centrarse en aquello que sí puede registrarse: los estímulos del medio ambiente y las conductas de los individuos. La clave está en la capacidad de los

individuos de percibir un estímulo muy particular como es el resultado de sus conductas. En el medio ambiente existen asociaciones entre las conductas y sus resultados. La asociación no es mental, es material y se produce en el medio ambiente. Cuando el individuo capta el estímulo del éxito o fracaso de sus acciones esto modificará la manera en que actuará en el futuro. Por supuesto, en el medio ambiente existirán muchos estímulos que nos impulsan hacia diversas conductas según esquemas de compensaciones; la conducta será el resultado de la acción de muchas fuerzas que nos impulsan en diferentes sentidos. Este tipo de orientaciones de conducta será medido por el conductismo que utilizará para ello el instrumental de la estadística. Este esquema que presentamos de manera tan sencilla y siguiendo los postulados de Skinner (1986), se complejiza en el conductismo de segunda generación (Osgood, 1986) que introduce consideraciones semánticas (teoría mediacional de Osgood), pero a nosotros nos basta con quedarnos con estas afirmaciones sencillas en torno al refuerzo. Cuando nos internemos en el campo de las IAs veremos aparecer varias de estas nociones relacionadas al behaviorismo, o ciencia del comportamiento. La idea de aprendizaje por refuerzo, la de caja negra, de compensaciones, la de penalización o premio de determinado output, forman parte del arsenal con que se han pensado modelos de IA. Esta es una de las formas en que la psicología toma como modelos postulados provenientes de la biología, tal vez la más sencilla, la siguiente será la del cognitivismo.

El cognitivismo, que desarrollaremos de manera más exhaustiva en el siguiente capítulo, hace eje en la estructura del sistema nervioso central y de sus componentes básicos las neuronas. La idea más relevante para nosotros será la de conexión entre neuronas y la formación de redes neuronales, una referencia central para el campo de la IA. La gran clave será la conexión entre las neuronas, la formación de enlaces y redes entre las mismas. De allí la denominación de conexionismo con la que se identifican estas corrientes vinculadas tanto a la neuropsicología como a una tendencia dominante el de las redes profundas de IA (deep learning).

Piaget, en su búsqueda de modelos en el campo de la biología, no se sitúa a nivel del comportamiento de los organismos como lo hace el

conductismo, ni el de las estructuras de las redes neuronales como el cognitivismo, sino en el terreno de la evolución de las especies. Se interesa por la biología evolutiva y se basa en la obra de un genetista ruso llamado Theodosius Dobzhansky cuyos postulados giran en torno a las mutaciones genéticas (Castorina, 1972). Las poblaciones y sus individuos deben adaptarse al medio ambiente manteniendo un equilibrio con el mismo. Para hacerlo deben mutar adaptándose a los desafíos que les plantea un medio ambiente dinámico y en constante cambio. Esta circunstancia propia de la adaptación vital se presenta tanto en el campo de la biología de las poblaciones como de la inteligencia humana. Entre ambos campos plantea lo que se llama isomorfismo. Este concepto proveniente de la matemática consiste en afirmar que entre dos conjuntos existen características equivalentes. Estas características estructurales son equivalentes, aunque los elementos que componen cada conjunto sean diferentes. El isomorfismo y las homologías estructurales son herramientas muy potentes en las ciencias ya que permiten controlar el pasaje de modelos entre una disciplina y otra. Estos préstamos de modelos y de ideas entre disciplinas científicas no tiene por qué restringirse al plano matemático. Lo más relevante es que en un campo y el otro se puedan detectar sistemas de relaciones equivalentes desde el punto de vista de la forma que adoptan los fenómenos. Los modelos migran de un campo de saber a otro, los préstamos entre una y otras disciplinas son, en ocasiones, bastante inesperados. Se llega a afirmar que las relaciones de parentesco operan como una lengua (Leví-Strauss, 1985) o que la arquitectura de las catedrales góticas sigue la misma estructura y forma que la "Suma teológica" de Santo Tomás de Aquino (Panofsky, 1986). La biología, en sus diversas facetas, ha sido una fuente constante de homologías y de planteos de isomorfismo retomados por otros campos de saber científico. Bueno, Piaget sostiene que los mecanismos de las poblaciones en sus mutaciones genéticas y la inteligencia humana tienen características en común y que ambas tienden a equilibrios sucesivos cada vez más estables. Plantea un isomorfismo entre la biología de poblaciones e inteligencia. No vamos a ahondar en los conceptos de fenocopia y de genocopia que retoma de la biología evolutiva, ya que nos desvían del objetivo de este libro. No obstante, queremos destacar los

paralelismos que encuentra esta corriente entre el desarrollo de las especies y el desarrollo de la inteligencia en un juego constante entre la génesis y la estructura. Es decir, Piaget tomará el modelo de la evolución biológica, en particular la idea de las mutaciones evolutivas y lo ha de trasponer al campo de la inteligencia humana. Ambos modelos teóricos tendrán elementos en común, en particular la noción de equilibraciones sucesivas y el juego entre el ambiente (lo externo o exógeno) y las estructuras (lo interno o endógeno).

Durante la historia vital, las necesidades de adaptación al medio ambiente son un impulso para la evolución, este impulso es la génesis de la inteligencia. A su vez en este proceso genético de "mutación" atravesamos por diversas estructuras cada vez más complejas en un espiral ascendente. Este tipo de juego entre la génesis histórica y la estructura se manifiesta en el plano biológico y el cognitivo. Nos adaptamos y evolucionamos porque contamos con la estructura para hacerlo. Esto implica rechazar la idea de tabla rasa del conductismo. Las mutaciones se producen sobre la base de estructuras preexistentes y disponibles para el organismo. También implica un rechazo del innatismo racionalista que supone la existencia de estructuras inmutables y universales. Al innatismo racionalista opone la idea de génesis, a la idea de tabla rasa del conductismo opone la noción de estructura. Durante nuestra historia vital, la presión del ambiente impulsará cambios en las estructuras. Génesis, estructura y equilibrio serán el sello distintivo de esta corriente en su intento de explicar la inteligencia.

Uno de los objetivos de este libro es el de resaltar la relación entre la inteligencia humana y la artificial, y para ello, nos parece necesario remarcar el tipo de modelos que se pueden adoptar para construir la noción de inteligencia en general. Como habíamos remarcado, Piaget, nos cuenta que desde muy joven sintió rechazo por los dualismos mente cuerpo. No se sentía cómodo con las explicaciones racionalistas y con la noción de ideas innatas ni con el empirismo y su planteo asociacionista que establece el predominio de los estímulos exteriores. Pensaba que la solución a los dilemas de la vida debía contemplar el juego dinámico de lo endógeno (estructuras) y lo exógeno (medio ambiente). En consecuencia, sus planteos en el estructural constructivismo abren dos tipos

de controversias en torno a la noción de inteligencia, en primer lugar, en sus polémicas con Noam Chomsky a quién reprocha la noción de que las capacidades lingüísticas tengan un soporte innato (Piaget, 1974). Según Piaget el desarrollo de la vida en general y de la inteligencia humana en particular son un juego entre la génesis evolutiva impulsada por la necesidad de adaptarse al ambiente y los equilibrios crecientes de las estructuras que caracterizan la evolución de la inteligencia a lo largo del tiempo. Desde el vamos rechaza el postulado racionalista de la existencia de estructuras innatas. En segundo lugar, discute con los planteos asociacionistas propios del conductismo de su época para los cuales las explicaciones responden al esquema de estímulo respuesta. La solución que propone es la de considerar el juego recíproco entre génesis y estructura (Piaget, 1991).

Como podemos ver, la manera en que se vincula la epistemología genética con los modelos biológicos es muy diferente a la que caracteriza a las corrientes cognitivistas o neuropsicológicas que presentaremos en el siguiente capítulo y que marcan el rumbo de varios de los desarrollos más relevantes en IA. Esta breve exposición del estructuralismo genético de Piaget nos servirá como contrapunto con su gran rival el cognitivismo, un actor intelectual que tiende a cobrar cada vez más relevancia y que hoy por hoy, le disputa terreno en el mundo de la pedagogía. Conforme avancemos en la lectura de este libro nos quedará cada vez más claro que la noción de inteligencia es muy compleja y que bajo esa denominación nos podemos referir a cuestiones muy dispares. No sólo nos cuesta aceptar la idea de que la IA sea homologable a la humana. Cuando echamos un breve vistazo al mundo de las ciencias que se ocupan de la inteligencia humana nos resulta evidente que no constituye un todo homogéneo y que a su interior existen desacuerdos básicos.

Bibliografía

Bourbaki, N. (1976). *Elementos de historia de las matemáticas*. Alianza.

Castorina, J. A. (1972). Biología y conocimiento de Jean Piaget. *Tarea, 3*, 73-90.

Castorina, J. A. (1981). *Introducción a la lógica operatoria de Piaget*. Paidós.

Dosse, F. (2004). *Historia del estructuralismo*. Akal.

Fresán, J. (2011). *Hasta que el álgebra nos separe*. RBA.

Hernández, J. (2002). *Las estructuras matemáticas y Nicolás Bourbaki*. Universidad Autónoma de Madrid.

Osgood, C. (1986). *Conducta y comunicación*. Taurus.

Osgood, C. E., y T. A. Sebeok (1965). Psycholinguistics: A survey of theory and research problems. Indiana University Press (originally published as supplements to International Journal of American Linguistics and Journal of Abnormal and Social Psychology, 1954).

Levi-Strauss, C. (1985). *Las estructuras elementales del parentesco*. Planeta Agostini.

Panofsky, E. (1986). *Arquitectura gótica y pensamiento escolástico*. La Piqueta.

Piaget, J. (1966). *La formación del símbolo en el niño*. Fondo de Cultura Económica.

Piaget, J. (1977). *Ensayo de lógica operatoria*. Guadalupe.

Piaget, J. (1978a.) *Adaptación vital y psicología de la inteligencia*. Siglo XXI.

Piaget, J. (1978b.). *Memoria e inteligencia*. El Ateneo.

Piaget, J. (1987). *Biología y conocimiento*. Siglo XXI.

Piaget, J. (1991). *Psicología de la inteligencia*. Siglo veinte.

Piaget, J. (1974). *El estructuralismo*. Hyspamerica.

Skinner, B. (1986). *Sobre el conductismo*. Planeta.

Capítulo 4. Las ciencias cognitivas

Sumario

A lo largo de este capítulo visitaremos a dos grandes inspiradores de las disciplinas computacionales en el desarrollo de modelos de IA: las ciencias cognitivas y la neuroanatomía. Revisar estos planteos nos permitirá comprender el llamado modelo conexionista de las redes neuronales que dominan el panorama actual de las IA. Nuestro punto de partida será la neuroanatomía y la noción de localización de funciones superiores a partir de los desarrollos de Alexander Luria. Su idea de coordinación de funciones será clave para entender la manera en que las IA, tal como ocurre con las funciones superiores de los sistemas nerviosos centrales biológicos, son posibles a partir del ensamble de diferentes capacidades como la memoria, la atención, la planificación y el razonamiento. Con posterioridad haremos una breve referencia a la exploración del cerebro humano a partir de diversas tecnologías de imagen y la forma en que permitió avanzar en la comprensión de la mente humana. A esta altura podremos presentar a un gran protagonista: la neurona biológica, una fuente de inspiración para el desarrollo de la neurona artificial, el componente básico de las redes neuronales de IA. Cerraremos con la presentación de la neuropsicología cognitivista a partir de los planteos de Stanislas Dehaene. Para esta corriente el conocimiento se adquiere a partir de la generación de conexiones neuronales en forma de redes a nivel de diversas estructuras del sistema nervioso central. En términos coloquiales lo podemos pensar como el "cableado" de nuestros cerebros. A partir de sus planteos podremos asimilar el

concepto cognitivista de "ajuste de parámetros" una noción con gravitación en el aprendizaje tanto humano como artificial.

IA: una alianza clave entre las disciplinas cognitivas y las computacionales

Todo sucede muy rápido cuando nos enfrentamos con una función exponencial como la del crecimiento y sofisticación de las capacidades en la IA. Aún los autores que se mantienen en la vanguardia de las ciencias cognitivas ven retadas sus afirmaciones por los cambios producidos en la arquitectura de los algoritmos, por el incremento de la capacidad computacional vinculada con la producción de microprocesadores cada vez más rápidos y eficaces y por la multiplicación de las estrategias de aprendizaje de las redes neuronales. Es como si la carrera desbocada del aparato técnico – científico – empresarial se empecinara en envejecer los textos y en cuestionar las certezas, apenas formuladas, de los científicos mejor calificados. Cuando nos encontremos con afirmaciones que evalúen las IA debemos ser cuidadosos e identificar el año de edición y, sobre todo, a qué modelos y tipos de diseños se refieren. Cuando la curva de desarrollo despega a tal velocidad, la capacidad de adaptación y la plasticidad con que se deben ajustar nuestras teorías puede resultar estresante. El desafío intelectual se torna más agudo en casos como este en el que la confluencia de disciplinas es tan llamativa, dado que las ciencias computacionales toman los modelos biológicos del sistema nervioso central como gran homología para diseñar los modelos de aprendizaje máquina. Por otra parte, los adelantos que genera el funcionamiento de las IAs sirven de plataforma para sustentar teorías y explicaciones acerca de facultades como el aprendizaje humano. Es notable la estructura de confirmación recíproca entre dos campos bien definidos de saberes, los de las ciencias de la computación y los de las disciplinas cognitivas. De hecho, existen sub disciplinas derivadas de esta alianza intelectual como es la neurociencia cognitiva computacional (O'Reilly et al., 2020). El intercambio de nociones entre los campos es constante y es frecuente que los argumentos conceptuales, los modelos de intelección y las referencias a

investigaciones empíricas surgidas en un campo sean retomadas en el otro. Se trata de una fuerte alianza cuyos efectos conceptuales e incluso políticos institucionales puede llegar a impactar con fuerza por varios factores, entre ellos el esfuerzo económico que realiza la sociedad en aras de impulsar la tecnología IA, el atractivo mediático que ha excitado la fantasía del público global, los canales de financiamiento científico en las universidades y fundaciones y, lo que resulta fundamental, la expectativa acerca de una nueva revolución industrial.

En este capítulo nos toca ocuparnos de las ciencias cognitivas, un mundo de saberes en cuya conformación confluyen varios campos disciplinares, entre otros, las neurociencias, la neuropsicología, ciertos aspectos de las teorías de la información, desprendimientos del conductismo, los aportes de las técnicas radiológicas de captación de imágenes y la psicología de la inteligencia.

¿Qué nos aporta este campo de saber en nuestra tarea de comprender el vínculo entre las facultades humanas y las de la IA? Al momento de contraponer dos campos de saber, una de las estrategias más exitosas desde el punto de vista epistemológico es la de detectar homologías. Las homologías son analogías estructurales que proceden a describir los modelos de relaciones profundas entre campos de saberes. Difieren de las simples similitudes de superficie al evitar quedarse con comparaciones basadas en parecidos en el aspecto de los fenómenos. Un defensor de estos procedimientos de comparación de disciplinas es Gastón Bachelard con su recomendación de seguir el llamado "vector de abstracción" a fin de evaluar el recorrido de las ciencias (Bachelard, 2000). La referencia a las relaciones abstractas apunta a la identificación de relaciones profundas que permitan ir más allá de las teorizaciones ingenuas que nos dictan nuestros sentidos. La construcción de modelos de relaciones y su generalización a otros campos de saber ha sido un impulso activo en la conformación y expansión de disciplinas científicas, tal y como lo hemos expuesto en capítulos anteriores a través de la producción de Chomsky, Piaget o Levi-Strauss. Destacamos estos tres autores, ya que las homologías estructurales que proponen, han arrojado resultados que no solo aportaron en la dimensión intelectual, sino que tuvieron impacto en el plano de la producción de resultados empíricos. Somos firmes

defensores del principio que dicta juzgar el árbol por sus frutos, lo que nos lleva a inclinarnos hacia los conceptos que, aún en el más alto grado de abstracción, busquen abrirse camino hacia el universo empírico.

La alianza intelectual entre las ciencias cognitivas y las ciencias computacionales relacionadas con las IA está plagada de analogías. Entre ambos campos de saber se ha establecido un feed back muy productivo. Cada una se mantiene atenta a los desarrollos de la otra. Las ciencias cognitivas y la neuropsicología ofrecen el modelo de las redes neuronales como base de diseño de las IA más potentes hasta la actualidad. Son la base del aprendizaje profundo a partir del modelo de las neuronas biológicas como plataforma conceptual de las neuronas artificiales en el modelo del llamado "perceptrón simple" y, en un plano más complejo de agregación, sirven de homología estructural del perceptrón multicapa en las así llamadas "redes de aprendizaje profundo". La comprensión de las funciones de las disciplinas neuropsicológicas ha aportado modelos de intelección del comportamiento humano y su vínculo con la neuro anatomía ha provisto una base material cuyo conocimiento no para de avanzar a partir del aporte de las técnicas de interpretación de imágenes del sistema nervioso procesadas por IA. Nos encontramos ante un bucle virtuoso en que las disciplinas se interceptan y alimentan su mutuo crecimiento.

La psicología cognitiva asume que la mente es un sistema de procesamiento de la información constituido por diferentes subsistemas y que permite establecer funciones tales como la inteligencia, el proceso atencional, la memoria, los dispositivos superiores de control de la conducta, entre otros procesos que hacen a la relación del organismo con el entorno. Sus enfoques son correlacionales y no causales, lo que, traducido a lenguaje corriente implica que no terminamos de saber cuál es su funcionamiento. Se han conceptualizado las facultades del entendimiento según conceptos cuyo origen es tan antiguo como la filosofía griega, se han elaborado conceptos y teorías que tratan de unir la base material de la biología, con conceptos muy abstractos vinculados a lo funcional y se han correlacionado algunos datos que ligan uno y otro nivel. Pero estos vínculos son solo eso, puentes entre conceptos funcionales y estructuras materiales del sistema nervioso central y sus extensiones. Como bien se

sabe, correlación implica probabilidad y no causalidad. Se señala que no terminamos de saber cómo y por qué operan muchas de las relaciones que realizan las máquinas inteligentes en el aprendizaje profundo, a ello se le da el nombre de "caja negra". Conocemos los inputs y conocemos los outputs, pero mucho de lo que hay en medio de ambos permanece en la oscuridad. Las inferencias son remotas y en muchos casos especulativas (Molnar, 2021; Prince, 2024). En rigor a la verdad poco sabemos respecto de la manera en que un conjunto de neuronas presentes en nuestros cerebros les permite entender lo que están leyendo o me permiten escribir lo que escribo, formular y sostener el proyecto de formar una familia o terminar una carrera o siquiera sentir que estamos aquí presentes o identificar la zona en que habitamos. Pues bien, asumamos que en el caso de las conceptualizaciones en torno a la mente humana no estamos en condiciones mucho mejores que en el caso de las inteligencias maquínicas. Entre ambos campos de conjeturas y formulaciones explicativas se comparten analogías y se formulan modelos que, para sorpresa de muchos en gran variedad de casos funcionan. Nuestra impresión es que en estos terrenos hay mucho que avanzar si queremos acceder a conceptos de alto nivel explicativo. Ya expusimos algunos de los fundamentos del escepticismo y la prudencia ante el avance de las técnicas (Chomsky, 2011). A continuación, vamos a introducirnos en la neuropsicología de las funciones superiores y en el modelo básico de la neurona como dos puentes materiales entre la biología y las máquinas. En el primer caso la evolución de la materia viva a través de un proceso de tres mil setecientos a tres mil ochocientos millones de años, en el otro en un proceso intelectual y tecnológico acelerado que se dispara en la década de 1950.

Executive functions

These are the abilities that allow a person to adapt to new situations and develop and follow their life goals. In this way, the term "executive functions" is an umbrella term for a host of functions such as those that allow people to plan and organize themselves over long periods of time; make complex high-level and abstract judgements; and organize and control their memory processes.

P. W. Burgess y J. S. Simons, 2005. Theories of frontal lobe executive function: clinical applications, p. 39.

La neuropsicología, un enfoque científico para el estudio de la mente

Alexander Luria compartía algo con Chomsky, Piaget y Vigotsky, la descalificación de las explicaciones basadas en la idea de estímulo respuesta. Se opuso a la idea de que estos conceptos lineales tuvieran alguna utilidad para explicar la conducta humana y, menos aún, los basamentos materiales de la actividad mental. Eso no nos tendría que asombrar tratándose de un médico renegado que, contra la voluntad de su padre decidió orientarse hacia el estudio de Freud en lugar de ocuparse de cuestiones más respetables en el campo de la ciencia o de la medicina. Sus biografías coinciden en señalar que el apartamiento de sus estudios de psicología freudiana no fue exactamente voluntario y que los temores

Figura 1. Luria fue un pionero en los estudios de neuropsicología. Llevó su disciplina a otro nivel. Sus estudios sobre la localización cerebral de las facultades superiores fueron un avance gigantesco en la conformación de un enfoque científico experimental de la psicología. Imagen extraída de: https://neurocienciashoy.blogspot.com/2012/07/biografias-hoy-alexander-luria.html

vinculados a las purgas de Stalin tuvieron mucho que ver. La segunda guerra mundial lo ubica en los campos de batalla como neurocirujano, lo que le brinda acceso de primera mano a pacientes con lesiones cerebrales. Detallista y observador, comenzó a percibir la relación entre las lesiones cerebrales y las funciones mentales superiores. Su curiosidad y su capacidad de establecer enlaces entre conceptos lo llevó a desarrollar la relación entre los hallazgos experimentales como cirujano y las teorías de la psicología y el lenguaje de Leon Vigotsky, su gran maestro. Psicología, fisiología, exploración quirúrgica con heridos de guerra, un cóctel que derivó en el estallido de la neuropsicología. Nunca pude entender por qué, al menos en la carrera de Psicología en UBA en los años 80, casi no escuchamos hablar de Luria. Incluimos una foto y les recomendamos acercarse a sus trabajos, bien escritos y útiles para entender el acercamiento entre las bases materiales del cerebro, las facultades superiores de la mente humana y las estrategias de diseño de IA.

Una premisa a tener en cuenta en el estudio de la ciencia es la de definir con claridad cuál es el nivel en el que ubicamos nuestras afirmaciones. La neuropsicología no es la excepción y en esto Luria fue un maestro. Su terreno fue el de la discusión de la idea de localización de los procesos superiores de la mente humana. Su planteamiento de los sistemas complejos dinámicos revoluciona el enfoque de estudio de la disciplina. De acuerdo a su concepción, los procesos mentales deben ser entendidos a partir de un sistema funcional con organización jerárquica. Rompió con una idea instalada, la aspiración de identificar las funciones superiores de la mente con áreas específicas del cerebro. Según sus hallazgos, sólo era posible aspirar a la localización específica de la corteza en caso de funciones elementales.

En el informe de uno de mis casos hay una descripción de un famoso compositor que, tras una hemorragia en la región temporal izquierda, era incapaz de distinguir los sonidos del lenguaje ni de comprender las palabras que se le decían; sin embargo, continuó componiendo brillantes trabajos musicales.

A. Luria, 1974, El cerebro en acción, p. 136.

Cuando entran en consideración las funciones mentales hay que aceptar dos premisas: que una misma área del cerebro puede estar implicada en más de una función y que las funciones dependen de sistemas organizados que se articulan de manera interrelacionada (Figura 2). Por otra parte, las áreas y regiones implicadas en una misma función pueden estar muy lejanas las unas de las otras. En sus trabajos, pioneros en el campo de la neuropsicología, divide la actividad del cerebro en tres bloques funcionales, con una organización jerárquica en que cada uno cumple una función en términos ascendentes. Según nos indica Luria, cada forma de actividad consciente constituye siempre un sistema funcional complejo y tiene lugar a través del trabajo combinado de las tres unidades cerebrales, cada una de las cuales aporta su propia contribución. Cada unidad funcional se monta sobre la anterior para operar de manera coordinada y ascendente (Luria, 1974).

La primera unidad funcional del cerebro, es un aparato que mantiene el tono cortical y el estado de vigilia y atención regulando estos estados de acuerdo con las demandas que en ese momento confronta el organismo. No tiene una ubicación específica privativa e involucra el tronco cerebral, el sistema reticular y sistema límbico.

La segunda unidad funcional del cerebro, tiene como rol primario la recepción, análisis y almacenaje de la información. Esta unidad se localiza en las regiones laterales del neocórtex en la superficie convexa de los hemisferios, de la que ocupa las regiones posteriores incluyendo las regiones visuales (occipital), auditiva (temporal) y sensorial general (parietal).

La tercera unidad funcional del cerebro es la que se hace cargo de las funciones más complejas. Es un sistema que tiene a cargo crear intenciones, formular planes y programas de acciones. A su vez tiene funciones de control ya que verifica la ejecución de los programas y regula las conductas. La localización de los componentes de esta tercera unidad funcional del cerebro es la más compleja, ya que involucra muchas estructuras. No obstante, Luria señala la importancia de los lóbulos frontales o, para ser más precisos, las divisiones prefrontales del cerebro. Según sus estudios de esta unidad funcional, las regiones prefrontales del córtex son estructuras corticales terciarias, en íntima

comunicación con casi todas las otras zonas principales del córtex. De hecho, realizan una función mucho más universal de la regulación general de la conducta.

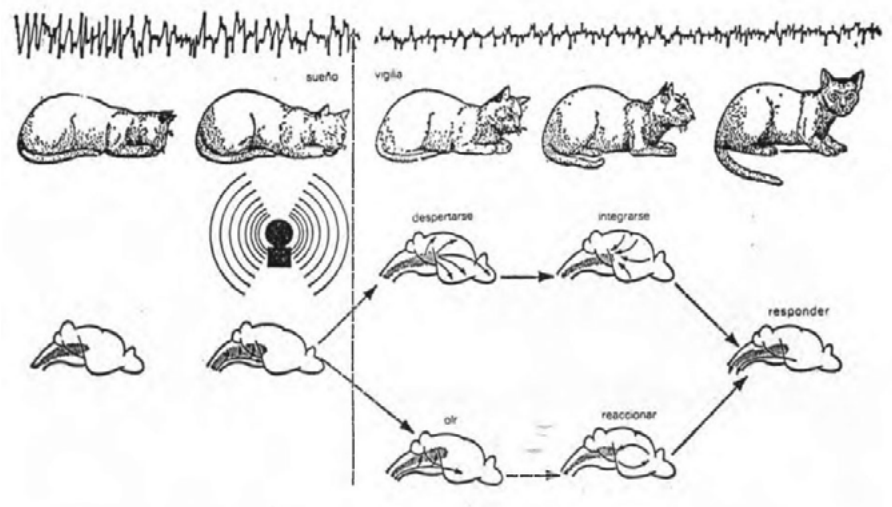


Fig. 5. — El efecto activante de la estimulación de la formación reticular del córtex que evoca una respuesta de arousal (French). El gato se despierta por el sonido de un timbre; la excitación que se produce en la formación reticular se extiende al córtex auditivo y conduce al arousal. Las ondas de EEG cambian correspondientemente. La formación reticular integra la actividad cerebral, y de ello resulta una respuesta organizada general del gato.

Figura 2. En la ilustración se aprecia cómo, la actividad de la unidad funcional opera mediante la acción conjunta de varios subsistemas. La organización de las unidades funcionales se torna cada vez más compleja en orden jerárquico ascendente. A. Luria (1974). El cerebro en Acción. Página 48.

¿Por qué nos resulta importante saber de estos temas cuando nos ocupamos de las IAs? Porque estas facultades serán tomadas en cuenta para el diseño de las máquinas pensantes. Algunas de las funciones clave en diferentes estrategias de diseño de IA son: la atención, memoria, manejo del lenguaje, procesamiento de informaciones, toma de decisiones, formulación de objetivos y planes a corto y largo plazo, establecimiento de sistemas de control de acciones. Todos estos procesos presentan desafíos en el diseño de algoritmos que permitan replicar el cerebro humano en el intento por acceder a máquinas inteligentes. Se han hecho presentes en diferentes esquemas de soluciones, pero muy en particular

en los diseños de redes neuronales, el planteo más eficaz que se haya alcanzado hasta el presente.

Existen muchos otros conceptos con efectos significativos en el diseño de soluciones, entre ellos la llamada hipótesis de modularidad. Esta idea conecta el campo de los estudios cognitivos y el de desarrollo de las IA. Es un postulado central sobre el que se asientan muchas de las exploraciones de la neuropsicología. Esta hipótesis es seguida por muchos científicos, pero si queremos identificarla con uno en particular debemos mencionar a Jerry Fodor (1983), un investigador proveniente del campo de la psico lingüística y de la filosofía cognitiva. Estas perspectivas sintonizan claramente con los enfoques computacionales, ya que conciben las funciones a partir del procesamiento de informaciones. Esto no quiere decir que los modelos que elaboran permitan soluciones fáciles de implementar en diseños de IA. La relación entre uno y otro campo es de "estimulación" y de sugerencia. Lo que produce la neuro psicología y la neuro anatomía sirven de plataforma para pensar soluciones y elaborar algoritmos que puedan correr procesos en las máquinas que piensan.

Las imágenes ayudan a entender la mente humana

Explorar los confines de la realidad, del universo que nos circunda ha sido desde siempre una de las mayores ambiciones de los homínidos. Tal vez estos impulsos nos hayan permitido salir de las sabanas y ocupar la totalidad del globo, o sumergirnos en las fosas oceánicas, circunvalar la tierra y lo que nos interesa hoy, penetrar en los secretos ocultos de la mente (Shulman, 2023). Pero ¿Cómo se accede al estudio de la mente humana? Por suerte ya no tenemos que esperar acceder a cerebros lesionados. Recordemos que en un primer momento las indagaciones de las funciones cerebrales avanzaban por la vía negativa. Actuaban sobre funciones disminuidas en experimentos clínicos que se efectuaban entre individuos con lesiones que afectan negativamente el desempeño. Esto permite asociar la región lesionada con la capacidad disminuida. Como

podemos imaginar, esto reduce la base de individuos sobre los que se puede realizar estudios. El desarrollo de la tecnología de imagen permitió la indagación con individuos sanos en diversas condiciones experimentales, lo que aceleró considerablemente la disciplina de las ciencias cognitivas y sus aliados. Por suerte, hoy se cuenta con vías menos invasivas que la cirugía, que han permitido explorar los secretos de la conexión entre la mente humana y el cerebro (Fundación Sadosky, 2023). No obstante, las técnicas de implantes intracraneales, de naturaleza invasiva, son un enfoque en crecimiento. Su utilización en pacientes con trastornos funcionales no está en discusión, aunque su implementación en individuos sanos es polémica y entran en juego barreras éticas que deberían discutirse en profundidad.

Las imágenes nos abren la puerta al cerebro y desde esas ventanas se puede tratar de inferir algo acerca de la mente humana (Figura 3). Podemos asegurar que, en la intersección de estos campos, las técnicas de monitoreo de la actividad cerebral tienen un rol importante y generan hipótesis fecundas en el desarrollo de modelos computacionales. En este terreno de la exploración por imágenes, se cuenta con dos enfoques principales el de las neuroimágenes estructurales y el de las neuroimágenes funcionales (Portellano Pérez y García Alba, 2014).

Neuroimágenes estructurales: su objetivo es el de localizar estructuras anatómicas concretas para identificar áreas dañadas. Las principales herramientas de estas técnicas de imagen son la tomografía axial computarizada (TC) y la resonancia nuclear magnética o resonancia magnética (RM).

Neuroimágenes funcionales: apuntan a registrar la actividad cerebral. Es un enfoque que implica monitorear la manera en que responde la actividad del cerebro ante determinadas conductas del individuo. Para ello se apela a experimentos funcionales en los que los individuos realizan tareas concretas y procesan diversos tipos de estímulos. Mediante técnicas funcionales se registran las actividades electro magnéticas y metabólicas del cerebro. Si el registro apunta a la actividad electromagnética se cuenta con las herramientas de la electroencefalografía (EEG) y la magnetoencefalografía (MEG). Si se apunta al registro metabólico se dispone de la resonancia magnética funcional (RMF), de la tomografía

computarizada por emisión de fotones simples (SPECT) y la tomografía por emisión de positrones (PET).

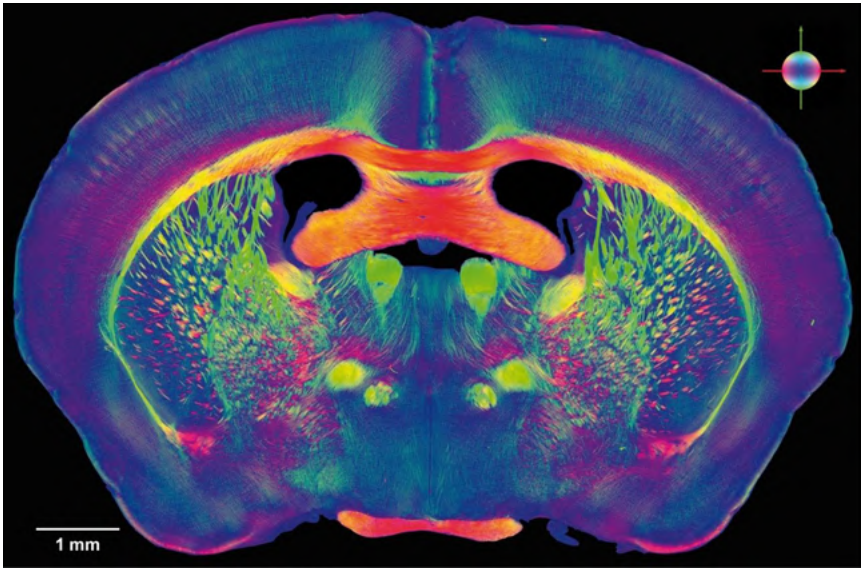


Figura 3. En esta imagen se estudia estructura y función del cerebro de una rata. Los enfoques de aprendizaje profundo y de big data permiten optimizar los conocimientos de las funciones cerebrales en humanos y otras especies. Mediante enfoques de imágenes combinados con procesamientos de IA se profundiza en el plano anatómico estructural y funcional. Imagen extraída de: https://www.mbfbioscience.com/blog/2021/12/neuroinfo-analyzes-rat-brains/

La neurona biológica como base de la neurona artificial

Nuestra imagen de las neuronas biológicas tiene origen en la imaginación pictórica de un gran artista y científico español Santiago Ramón y Cajal (1852-1934), cuya exploración de este pequeño mundo le valió el premio Nobel de medicina en 1906 (Figura 4). Este padre de la neuro ciencia moderna y artista, con el escaso auxilio de tinciones y microscopios, visualizó y recreó la neurona, el ladrillo base del cerebro.

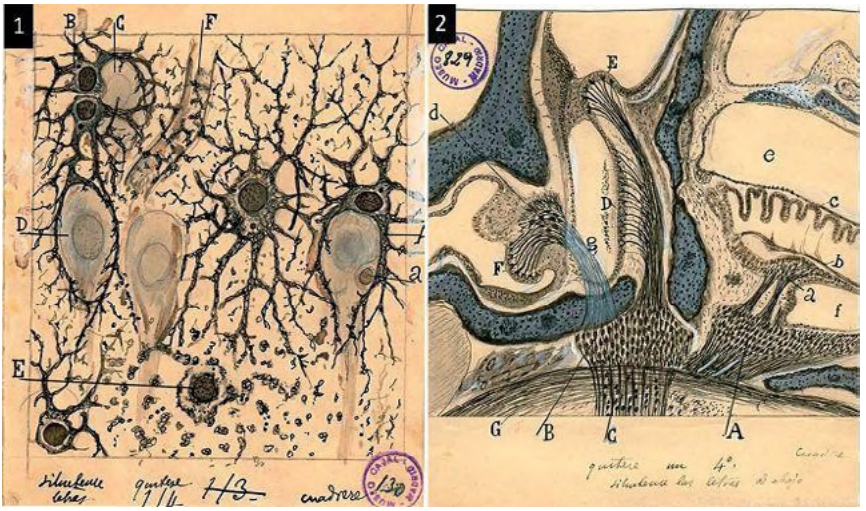


Figura 4. Santiago Ramón y Cajal nos ha dejado más de 2900 ilustraciones que van desde distintos tipos de neuronas aisladas a complejas estructuras de redes. La exposición de sus trabajos recorre el mundo y se exhibe en los museos más prestigiosos. Algunos críticos han llegado a decir que debería compartir un reconocimiento similar al de Newton, Curie, Leonardo y Picasso al mismo tiempo.Imagen extraída de: https://www.nationalgeographic.com.es/ciencia/actualidad/una-muestra-con-dibujos-santiago-ramon-cajal-recorre-estados-unidos-canada_11187

Tratándose de las neuronas y del sustrato bio fisiológico y anatómico en general hay que considerar lo que se llama problema de escala. La cuestión de escala implica la definición del tipo de dimensiones en que se sitúan las entidades a considerar, así, por ejemplo, en ciencias sociales será clave definir si se habla de un nivel micro de relaciones interpersonales, meso de relaciones institucionales o macro de relaciones que implican el conjunto social en su totalidad o incluso de un nivel de nociones que involucre las relaciones a nivel global. Otro tanto ocurre con el plano de nuestro sistema nervioso central en el cual podemos focalizar en las neuronas, sus umbrales de excitación y los procesos bioquímicos a nivel de sus sinapsis o podemos pensar en estructuras de relación más amplia que consideran redes de neuronas en las que la interacción implica gran número de componentes. Podemos en un nivel teorizar y explorar empíricamente los intercambios moleculares o en el otro extremo podemos

conceptualizar un plano de mayor nivel de abstracción y agregación como es el de las facultades superiores como el razonamiento y la memoria. Todo será cuestión de perspectiva. En un primer nivel, el de una simple neurona, podemos observar cómo se retoma el modelo biológico de excitación neuronal como base de la neurona artificial, cuestión que vamos a retomar en capítulos posteriores (Figura 5).

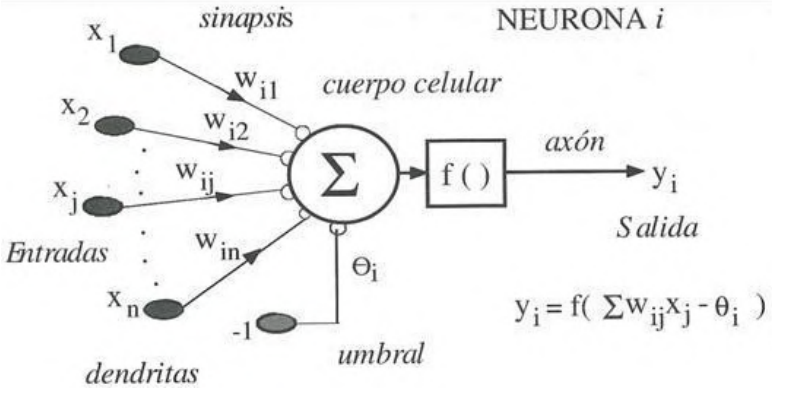


Figura 5. Modelo básico de neurona artificial. Se establece una analogía que retoma los elementos de la neurona biológica y sus procesos de activación como fundamento del diseño de la artificial y su algoritmo. Imagen extraída de: https://grupo.us.es/gtocoma/pid/pid10/RedesNeuronales.htm

Los diversos planos de exploración de la biología son tomados en cuenta por los desarrollos de las IA como fuente de ideas para elaborar sus arquitecturas computacionales y los algoritmos que les permiten operar a escala de lo que podemos llamar por extensión y en analogía con los organismos biológicos, "atención", "razonamiento" o "memoria" y que forman los pilares de la función superior de la inteligencia. Estos conceptos son retomados y reelaborados por las soluciones científico técnicas de la IA. La neuropsicología ha desarrollado varias modelizaciones con distintas propuestas, entre ellas la de la Hipótesis de Modularidad, cuyo enfoque se vincula con el procesamiento de la información por parte de módulos funcionales y procesadores cognitivos. En esta hipótesis se postula que la información que brinda la experiencia será procesada por los diversos módulos que operan de manera independiente y el resultado funcional derivado de dichas acciones será

interpretado en función de estos procesamientos sin que ello implique una operación de simple adición (Fodor, 1983). La tarea de los neuropsicólogos será la de modelizar y comprender el proceso cognitivo y sus complejidades (Portellano Pérez y García Alba, 2014). Descomponer el proceso en módulos y componentes con distintas funciones es una forma de proceder propia de la generación de algoritmos de IA y está en la base del diseño de sus arquitecturas.

Seguramente, el mayor aporte que ha ofrecido la homología estructural con el funcionamiento del sistema nervioso de los seres vivos proviene de las corrientes llamadas conexionistas, que tantos reparos le genera a N. Chomsky tal como expusimos en un capítulo anterior. El conexionismo, que se conoce como modelo de enfoques de las redes neuronales, presupone que el funcionamiento del cerebro depende de una serie de unidades interconectadas en el que cada unidad está constituida por una neurona o pequeño grupo de neuronas (Vogels y Abbot, 2005). Desde la perspectiva del conexionismo, según la plantean autores como Vogels, Rajan y Abbot, se implica que el sistema nervioso central es un sustrato natural que opera a partir de reacciones bioquímicas y que no funciona como un computador convencional, ya que los procesos que lleva adelante se verifican de manera paralela y no secuencial. Esto quiere decir que los algoritmos de IA en un tiempo anterior a 2006 y la arquitectura CUDA para GPU de NVIDIA, se verificaban mediante la transferencia de informaciones paso a paso, es decir en secuencia. Como veremos en el capítulo 7 dedicado a los modelos generativos con base en el algoritmo Transformer de IA y la manera en que opera, esta limitación ha desaparecido y ya es posible que el cómputo que realizan los ordenadores se verifique en forma paralela. Posiblemente sepamos que nuestros procesadores se diferencian unos de otros por la cantidad de núcleos de procesamiento. Cuantos más núcleos tiene un microprocesador más tareas podrá correr al mismo tiempo. CPU quiere decir, de hecho, Unidad Central de Procesamiento e integra unidades independientes llamadas núcleos (core) capaces de llevar adelante tareas de manera autónoma. Esto nos recordará la noción de Fodor con relación a la hipótesis de modularidad en que cada módulo realizaba tareas a ser integradas a un nivel funcional superior. La capacidad de las arquitecturas actuales de redes

El ABC del conexionismo

Un sistema conexionista, o una red neuronal, consiste en una red de procesadores simples, similares a neuronas [neuron-like processors] llamados nodos o unidades. Cada nodo tiene conexiones dirigidas a varios otros nodos, de modo que obtiene señales de algunos nodos y envía señales a otros nodos, incluyendo, posiblemente, aquellos de los cuales obtiene señales.

J. Tienson, 1995, Una introducción al conexionismo, p. 367.

neuronales a partir del llamado algoritmo Transformer permite el procesamiento paralelo, es decir la acción de la red neuronal sobre todo el sistema de datos que se le suministra, lo que ha ampliado enormemente la eficiencia de cómputo y es una de las bases de la explosión de la IA en el estado en que nos toca experimentarla en la actualidad. Ya tendremos oportunidad de explicar en mayor detalle de qué se trata el modelo Transformer y su relación con enfoques tales como los grandes modelos generativos de lenguaje GPT, Palm o Bard.

Aprender ajustando parámetros

Hasta ahora vimos cómo las ciencias computacionales se alimentaban de ideas que las orientaban en la tarea compleja de imitar la mente humana y los procesos inteligentes. Esto marcaba su lenguaje y sus enfoques algorítmicos. Pero este proceso también funciona a la inversa en este bucle virtuoso que mencionamos. Las ciencias cognitivas también adoptan el lenguaje de las IAs para presentar sus ideas y formular sus conceptos. Uno de los ejemplos más claros es la terminología y la batería conceptual utilizada por Stanislas Dehaene, un matemático doctorado en psicología cognitiva. Tal vez sea su formación como matemático y como psicólogo cognitivo lo que lo lleve a adoptar ambos enfoques como plataforma para sus elaboraciones y conjeturas.

Según Dehaene, la humanidad hizo del aprendizaje su especialidad. Nos señala que el cerebro, la lengua, la cultura, la familia, la alimentación, muchas dimensiones se han desarrollado, nos han modelado en un aprendizaje que evoluciona y se ajusta a las diferentes circunstancias medio ambientales a las que se ha enfrentado la especie. Para empezar a comprender la forma en que piensa este proceso analicemos la siguiente frase:

> *Solo el Homo sapiens logra generar de manera sistemática pensamientos simbólicos abstractos y actualizar su plausibilidad ante nuevas observaciones.* (Dehaene, 2019, p. 27).

A menudo, las situaciones de la vida diaria son altamente cambiantes, y los parámetros y criterios de respuesta no dependen de una lógica inflexible y generalizable a todas las circunstancias, sino del momento y del lugar en donde se desarrolle el criterio. La fijación excesiva en un criterio, una hipótesis o una estrategia de acción afecta de forma importante la resolución de problemas.

T. W. Robbins, 1988, Dissociating executive functions of the prefrontal cortex, p. 123.

Esta sencilla afirmación incluye tres cuestiones. La primera la idea de unicidad de la especie homo sapiens *"la única capaz de..."*. La segunda la de la capacidad del ser humano de generar conceptos abstractos, por ende, generalizables, idea que ya habíamos encontrado en Piaget. La tercera, la posibilidad de actualizar sus pensamientos ante nuevas experiencias, que nos lleva directamente hacia la cuestión de la actualización de parámetros. En esta suscita afirmación condensa el punto de intersección entre las disciplinas cognitivas y los planteos de las ciencias informáticas en el desarrollo de IAs. Esta ligazón surge como algo casi natural ya que su formación primaria como matemático le permite manejar con fluidez los postulados de las teorías de la información presentes en ambos campos disciplinares.

Como veremos en el capítulo siguiente, el que aborda conceptos básicos sobre las IAs, la idea de ajustar los parámetros es central en el aprendizaje máquina (machine learning). Su modelo mental se apoya en la idea de que la base del aprendizaje y de la relación de ajuste con el medio ambiente, tanto en las estructuras biológicas como en los soportes de silicio de las máquinas que aprenden, es de tipo probabilístico. Tanto humanos como máquinas generamos hipótesis probabilísticas con relación al entorno. Estas conjeturas implican la posibilidad de generar nociones simbólicas abstractas que nos permiten generalizar los eventos con los que nos enfrentamos. Con esta base formulamos hipótesis en cuanto a la probabilidad de que tal o cual evento se produzca, del resultado de nuestras acciones e incluso de las consecuencias de nuestros planes a largo plazo. Por supuesto, estamos dando una imagen muy simplificada de los planteos cognitivos y de las facultades que encadenan, pero esta noción nos sirve para entender por dónde avanzan sus razonamientos. Si las hipótesis son muy rígidas pierden lo que se llama "flexibilidad cognitiva" en el caso de la neuropsicología y en el caso de los modelos de aprendizaje automático "over fitting". En términos coloquiales podemos asimilar estos fenómenos con la tozudez. Si las ideas y preconceptos son demasiado rígidos, si el aprendizaje se restringe a nociones demasiado estrechas y poco móviles la capacidad de adaptarse a circunstancias novedosas será escasa. Atenerse constantemente a un guion del que no podemos despegarnos no parece ser una alternativa demasiado feliz en

términos adaptativos. De allí la capacidad de *actualizar la plausibilidad* de nuestras conjeturas. Hablar de plausibilidad es hablar de probabilidad. No sabemos demasiado acerca de nuestro cerebro, los planteos que lo rodean están llenos de supuestos y de conjeturas, pero de algunas cosas podemos estar ciertos: existen miles de millones de conexiones entre las neuronas y estas se conectan en redes. Estas conexiones son comparadas con los parámetros de las redes neuronales. Las redes neuronales biológicas contienen miles de millones de conexiones y también las tienen las redes neuronales artificiales. Explorar entidades con este grado de complejidad es un desafío para el estado actual de la ciencia.

¿Qué son estos parámetros y cómo se ajustan? Veamos como lo explica Dehaene a partir del ajuste de dos subsistemas el de la percepción visual y el motor en el acto de tomar una varilla entre los dedos (Figura 6). ¿Cómo hacemos para tomar un objeto que visualizamos? Para eso nos presenta una ilustración con que René Descartes analizaba el problema. Para tomar el objeto debemos ajustar la información visual

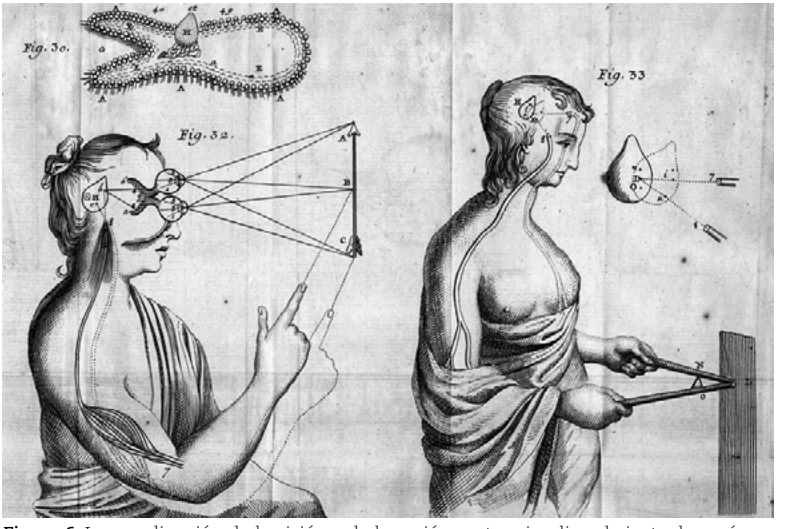


Figura 6. La coordinación de la visión y de la acción motora implica el ajuste de parámetros. The picture art collection Alamy Achive Imagen extraída de: https://www.alamy.es/imagenes/?name=The+Picture+Art+Collection&pseudoid=3BFFF4BE-75E7-4303-BBFD-C0B79962921D

(input) y convertirla en instrucción motora (output). En este punto introduce un interesante experimento en el cual nos colocamos unos lentes ajustados para correr el ángulo de visión algunos grados. Aquello que percibimos estará desfasado, lo que produce que, cuando tratemos de tomar un objeto nuestra mano lo buscará donde no se encuentra. Cuando nos saquemos los lentes luego de usarlos durante un tiempo sufriremos un cierto nivel de desacomodamiento motor. Este desajuste se corregirá rápidamente. Hemos ajustado los parámetros visuales con los motores nuevamente. Los grados que distorsionan nuestra visión proponían un parámetro diferente al de nuestra percepción visual corriente. Al sacarse los lentes nuestros parámetros se ajustarán. Ya volveremos sobre esta cuestión cuando tratemos el tema del ajuste de modelos de aprendizaje automático en el llamado machine learning (ML).

El ejemplo de los lentes es sencillo e implica unos pocos parámetros con un ajuste que podemos seguir y explicar. Es un modelo cuya probabilística nos resulta accesible. Se trata de la captación de regularidades de bajo nivel por parte del cerebro, así pequeños grupos de neuronas captarán pequeños fragmentos de información provenientes de la retina para luego unificarlas en una captación de conjunto. Tengamos en cuenta que aún este ejemplo ha sido simplificado, ya que la percepción tampoco es algo sencillo y está conformada por complejas hipótesis espaciales. Ahora bien, no todo es cuestión de regularidades de bajo nivel, que los seres humanos captamos regularidades de muy largo alcance y establecemos conexiones de largas series de eventos que inciden en nuestras expectativas de acción. Estos procesos ponen en juego muchas instancias, no todas provenientes de cálculos "razonables" de probabilidades. ¿Por qué esto es así? Porque nuestro universo no es determinista y no se puede hacer depender de una serie ensamblada de causas y efectos. Esta es nuestra virtud y nuestra desgracia al mismo tiempo.

Dehaene, tal como Chomsky se pregunta cómo es posible el aprendizaje cuando las regularidades no son ya de bajo nivel como en el ejemplo de tomar un objeto ajustando las previsiones visuales, sino la de aprendizaje de cuestiones tan complejas como la lengua. Se trata de aquello que se denomina "explosión combinatoria", un fenómeno que se desencadena cuando la combinación de un puñado de elementos

genera una cifra tan impactante de posibilidades que excede nuestras posibilidades de cálculo. La ventana que se abre al mundo ya no es estrecha sino inabarcable. En estos casos el modelo con que piensan el aprendizaje las disciplinas cognitivas es el de jerarquización. ¿A qué se refiere Dehaene con la idea de jerarquizar? El ser humano puede discriminar las informaciones, extraer principios generales, elevar los problemas al plano abstracto a partir de su capacidad simbólica, lo que implica una ventaja evolutiva respecto de otras especies y nos diferencia de las máquinas que piensan (Dehaene, 2019). Para Chomsky estas capacidades se articulan con el lenguaje y son propias del entendimiento humano. No obstante, el análisis que propone en cognitivismo es algo diferente, recordemos que su enfoque es asociacionista y probabilístico. Por otra parte, los aprendizajes se montan sobre lo que la materia les provee, la estructura del cerebro y las funciones disponibles. Esta disponibilidad de capacidades responde a la especialización de circuitos vinculados a nuestra historia evolutiva. No se trata, como en el caso de la generación de una arquitectura de IA cuyos algoritmos se ensamblan en conjunto, sino de una compleja integración propia de la organización evolutiva de la materia en interacción con medios cambiantes.

Existen diversas maneras de comprender ese diferencial humano, su gran capacidad de aprendizaje. Si prestamos atención al capítulo anterior, en que presentamos las ideas del constructivismo estructural, nos daremos cuenta de inmediato su contraposición con las posturas cognitivistas. Son dos grandes contendientes en esta tarea tan importante de comprender cómo opera el aprendizaje. Para familiarizarnos con los planteos de los cognitivistas proponemos prestar atención a la manera en que explican la emergencia del aprendizaje de la lectura. ¿Recuerdan lo difícil que resulta aprender a leer? El aprendizaje del lenguaje oral fluye con mucha mayor naturalidad, pero enfrentarse a esas hojas llenas de pequeños garabatos colocados en filas o hileras que son las palabras y las oraciones escritas es como escalar una montaña para los pequeños escolares. Leer nos recuerda el cognitivismo (Dehaene, 2015) es una actividad reciente en nuestras culturas, tiene poco tiempo entre nosotros los humanos. Para cuando surge históricamente nuestros cerebros de primates ya se encontraban consolidados

desde el punto de vista evolutivo. Por ende, no forma parte de nuestro acervo genético. En la tarea de aprender a leer, algunas regiones cerebrales especializadas en otras funciones e interrelacionadas, deberán reorientarse a fin de adaptarse a una tarea para la que no han evolucionado. Desde muy pequeños venimos preparados para reconocer distinciones fonológicas presentes en cualquier lengua. Hayamos nacido en el rincón del globo en que hayamos nacido, esa capacidad la tendremos, es algo que ya se ha inscripto en nuestro acervo genético. Es más, no sólo lograremos identificar partículas elementales y sus combinaciones, sino, lo que es aún más interesante, la melodía de la lengua que vayamos a adquirir. De allí la adaptación material de nuestro aparato fonador a la lengua nativa en que nos haya tocado vivir y crecer en los primeros momentos de nuestro desarrollo. Esa capacidad de identificar algunas combinaciones de sonidos de la lengua materna hacia los seis meses de edad se denomina "embrión léxico mental". Con posterioridad el conjunto se ensamblará con reglas gramaticales de articulación lo que impulsa el proceso de adquisición de la lengua. Bueno, esa capacidad material predispuesta de manera espontánea en el aprendizaje del habla no está presente en el caso de la lectura. El conocimiento del lenguaje es inconsciente y espontáneo porque no sale de los circuitos neuronales especializados. Aprender a leer implica reorientar esos circuitos, quebrarlos y reorganizarlos. Como vemos, para el cognitivismo es importante el tema del "cableado" (Figura 7). Su orientación asociacionista choca frontalmente con las hipótesis constructivistas cuyas referencias a circuitos neuronales es muy indirecta. Su forma de explicar el proceso de aprendizaje de la lectura sigue los principios generales de las teorías de equilibración estructural en vínculo interactivo con las operaciones lógico matemáticas piagetianas (Ferreiro, 2000). Este es uno de los puntos en que centran sus ataques los cognitivistas en general, y Dehaene en particular (Dehaene, 2016), tanto en el terreno del aprendizaje del lenguaje como el de la matemática. No quisiéramos desviarnos del tema que es la relación de las facultades humanas con la IA, pero estamos seguros que reflexionar sobre esta contraposición en términos de su impacto didáctico puede arrojar resultados productivos.

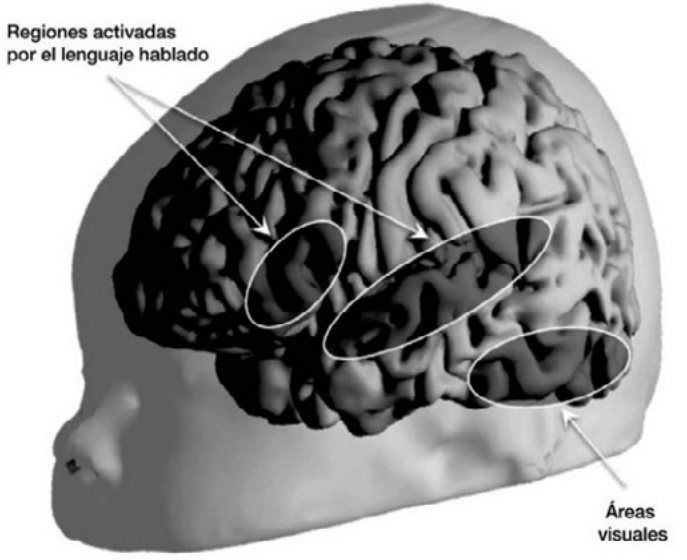


Figura 7. Mucho antes de aprender a leer, el cerebro del bebé ya está consistentemente organizado: las áreas del lenguaje hablado funcionan desde los primeros meses de vida, así como las áreas visuales. Con el aprendizaje de la lectura, una parte de ellas va a especializarse para reconocer los grafemas y los fonemas. (Dehaene, 2015, página 29)

Ya habíamos anticipado el hecho de que no alcanza con mencionar una entidad para garantizar que se habla de lo mismo. Asociacionismos y constructivismos construyen dos concepciones muy diferentes acerca del aprendizaje y de la inteligencia. Para los asociacionismos cognitivistas la referencia al sustrato material del cerebro y las interconexiones neuronales es la base de elaboración primaria. La neurona está en el centro de la escena y su modelo se conecta de manera fluida con los diseños de las IA. Sus conjeturas retoman la idea de combinación de pequeños grupos de neuronas que procesan secuencias en segmentos temporales. La idea de ajustar parámetros entre conexiones se vincula de manera directa con la forma en que programa el aprendizaje una red neuronal artificial. Mientras el número de parámetros resulta manejable podremos seguir, en cierta medida, las inferencias de la red neuronal. No obstante, esta no es la regla sino la excepción. Sin entrar en detalle retengamos la necesidad de encontrar soluciones cuando la cantidad de parámetros

se eleva de manera exponencial. Las ciencias cognitivas no terminan de encontrar respuestas, pero propone modelos interesantes y productivos para conceptualizar estos problemas. Cuando los números son tan altos y las conexiones tan misteriosas como en el pensamiento, es muy complejo seguir el encadenamiento de las causas y los efectos o siquiera determinar las funciones que aproxima el modelo (Molnar, 2021; Prince, 2024). Será por eso que la comprensión profunda de la operatoria de estas entidades biológicas o artificiales nos resulta tan difícil de descifrar. Si tenemos en cuenta que los objetivos de las IAs apuntan a generar agentes con capacidad de acción que implique procesos de inteligencia similares a los humanos, es obvio que también aquí nos moveremos en la relativa oscuridad. Pensemos que una red neuronal en los grandes modelos de 2023, cuenta con trillones de parámetros. Imaginen inmensas matrices llenas de cifras, trillones de ellas, multiplicándose de manera incansable. En gran medida, por más que los modelos de las ciencias cognitivas puedan arrojar algo de luz, nos movemos entre penumbras.

Bibliografía

Bachelard, G. (2000). *La Formación del Espíritu Científico: Contribución a un Psicoanálisis del Conocimiento Objetivo*. Siglo XXI.

Burgess, P. W. y J. S. Simons (2005). *Theories of frontal lobe executive function: clinical applications*. En P. W. Halligan & D. T. Wade (Eds.), The Effectiveness of Rehabilitation for Cognitive Deficits (pp. 211–232). Oxford Academic.

Chomsky, N. (2011). Language and Other Cognitive Systems. What Is Special About Language? *Language Learning and Development, 7*, 263–278.

Dehaene, S. (2015). *Aprender a leer*. Siglo XXI.

Dehaene, S. (2016). *El cerebro matemático*. Siglo XXI.

Dehaene, S. (2019). *¿Cómo aprendemos?* Siglo XXI.

Ellis, A. (1992). *Neuropsicología cognitiva humana*. Masson S.A.

Everaert, M. B. H., M. A. C. Huybregts, N. Chomsky, R. C. Berwick y J. J. Bolhuis (2015). Structures, Not Strings: Linguistics as Part of the Cognitive Sciences. *Trends in Cognitive Sciences, 19*(12), 729–743. http://dx.doi.org/10.1016/j.tics.2015.09.008

Ferreiro, E. (2000). *L écriture avant la lettre*. HACHETTE Éducation.

Fodor, J. A. (1983). *The modularity of mind*. The MIT Press. https://doi.org/10.7551/mitpress/4737.001.0001

Luria, A. (1974). *El cerebro en acción*. Fontanella.

Luria, A. (1980). *Introducción evolucionista a la psicología*. Fontanella.

Luria, A. (1986). *Conciencia y lenguaje*. Visor.

Molnar, C. (2021). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*.

Portellano Perez, J. y J. García Alba (2014). *Neuropsicología de la atención, las funciones ejecutivas y la memoria.* Síntesis.

Prince, S. (2024). *Understanding Deep Learning*. The MIT Press.

O'Reilly, R. C., Y. Munakata, M. Frank y T. Hazy (2020). *Computational Cognitive Neuroscience*. 4ta ed. Wiki Book. https://CompCogNeuro.org

Robbins, T. W. (1998). *Dissociating executive functions of the prefrontal cortex*. En A. C. Roberts, T. W. Robbins y L. Weiskrantz (Eds.), *The prefrontal cortex* (pp. 117–130). Oxford University Press.

Tienson, J. (1995). Una introducción al conexionismo. En Rabossi, E. (Comp.), *Filosofía de la mente y ciencia cognitiva* (pp. 359–380). Paidós.

Vogels, T. P. y L. F. Abbott (2005). A Signal Propagation and Logic Gating in Networks of Integrate-and-Fire Neurons. *Journal of Neurosciences, 25*(46), 10786–10795. https://doi.org/10.1523/JNEUROSCI.3508-05.2005.

Videos YouTube

Algoritmos de Deep Learning para diagnósticos por imágenes. Fundación Sadosky Julio 2023.
https://www.youtube.com/watch?v=NbV7ssTTr2M

Aprendizaje Automático para la neurociencia Fundación Sadosky. Agosto 2023.
https://www.youtube.com/watch?v=Dk6upKHwiHk

Carl Shulman (Pt 1) – Intelligence Explosion, Primate Evolution, Robot Doublings, & Alignment
https://www.dwarkeshpatel.com/p/carl-shulman?utm_campaign=post&utm_medium=web. Junio 14 2023.

Capítulo 5. La IA y el aprendizaje

Una vez que presentamos la noción de IA en el capítulo 1 y repasamos nociones en torno a la inteligencia humana y el aprendizaje en los capítulos 3 y 4, estamos listos para describir la forma en que aprenden las máquinas. Veremos que, bajo el concepto general de IA, se pueden diferenciar dos tipos de aprendizajes: el aprendizaje máquina y el aprendizaje profundo. Son los llamados "Machine Learning" (ML) y "Deep learning" (DL). Se presentará la idea de que a las máquinas hay que entrenarlas para que aprendan y que lo hacemos suministrándoles datos (inputs). A partir de ellos podrán descubrir relaciones al interior del conjunto de datos y esto les permitirá realizar predicciones (outputs). Sin aprendizaje no podrán realizar predicciones eficientes. Describiremos los pasos necesarios para entrenar y ajustar los modelos de IA para que alcancen de manera eficaz las metas que les planteamos. A continuación, vamos a describir los principales enfoques de aprendizaje, el supervisado y el no supervisado. Como cierre comentaremos algunos de los enfoques actuales en el aprendizaje en IA, un campo que se mantiene en constante evolución.

Tres nociones en torno al aprendizaje: inteligencia artificial, aprendizaje máquina y aprendizaje profundo

Entre todas las facultades que hemos señalado existe una que está en el foco de nuestros intereses y constituye una clave en torno a la IA, la capacidad de aprendizaje. Las máquinas pueden aprender, ya enfatizamos su capacidad de discriminar patrones. A su vez, bajo ciertas condiciones,

pueden retener esa información a partir de procesos análogos o comparables a la memoria y, en cierta medida, pueden generalizar el resultado de sus aprendizajes.

En el gráfico visualizamos tres conjuntos sucesivos, el de la IA, el del machine learning y el de deep learning (Figura 1). Estos términos suelen generar algunas confusiones y se usan en muchos casos de manera indiscriminada. En realidad, la noción de IA cubre un campo muy amplio y se refiere de manera genérica a la aspiración de lograr agentes inteligentes que puedan realizar tareas que impliquen facultades humanas que hacen a la inteligencia (Russell y Norvig, 2004). El ámbito del machine learning cubre un campo general de las disciplinas de la computación que apuntan a dotar a las máquinas de la capacidad de aprendizaje. Por último, el deep learning es un subcampo del machine learning que corresponde a algoritmos computacionales complejos relacionados con lo que se conoce como redes neuronales de aprendizaje profundo. En tanto el machine learning contempla estrategias y algoritmos más sencillos y requiere una mayor intervención humana para el tratamiento de los datos, el deep learning permite detectar relaciones profundas a partir de datos no estructurados con una mayor autonomía. Volveremos sobre este tema en el capítulo siguiente.

El desafío que se proponen estos enfoques de IA es el de diseñar modelos computacionales que puedan realizar tareas específicas a partir

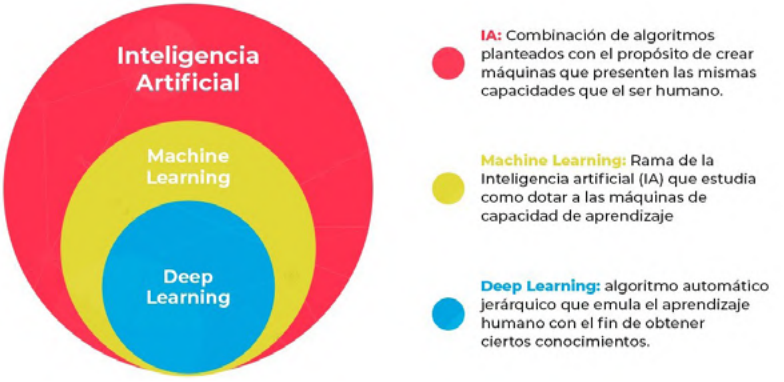


Figura 1. Los tipos de inteligencia artificial. Imagen extraída de: https://www.masterdatascienceucm.com/que-es-machine-learning/

de un proceso de entrenamiento que les permita aprender. Un modelo de machine learning será un diseño de algoritmo matemático capacitado para recibir un conjunto de datos, identificar patrones al interior de los mismos y actuar en consecuencia con algún tipo de respuesta. Estas acciones están basadas en el cálculo de probabilidades. Este enfoque de ML permite a las máquinas predecir a partir de datos que se le suministran. A partir de este proceso de aprendizaje podrán establecer predicciones acerca de casos que nunca han visto anteriormente. Este proceso de aprendizaje se denomina entrenamiento. En esencia se trata de entrenar a las máquinas para que puedan realizar estimaciones, prepararlas para que puedan predecir. El modelo de IA, de por sí, no puede determinar nada antes de ser entrenado, está, podemos decir, en el grado cero de conocimiento. Una vez que se lo hace pasar por un proceso de entrenamiento podrá realizar diversos tipos de tareas, pero todas ellas tendrán algo en común: involucran algún tipo de predicción. Podrán predecir cuestiones tan variadas como la probabilidad de que una imagen corresponda a un perro o un gato, las probabilidades de que alguien termine una carrera universitaria, el movimiento de un brazo robótico, la relación entre las lluvias en un continente y el rendimiento de campaña en las cosechas de su diferentes regiones, la próxima palabra en una oración, la probabilidad de éxito de determinadas alternativas de movimientos en un juego, la trayectoria que debe seguir un móvil para no colisionar, el sentimiento positivo o negativo en el contenido de un texto y muchas otras tareas para las que se las destine.

Se había definido la IA como el intento de generar agentes cuyos procesos imiten comportamientos inteligentes, para ello deberemos hacerlos atravesar procesos de aprendizaje. Deben aprender a percibir el ambiente y para ello les hemos de proveer de informaciones adecuadas y asegurarnos que pueda extraer de ellas consecuencias adecuadas. La calidad de los datos con que la entrenemos es crucial, el aprendizaje máquina, como el humano, está sometido a la regla: "basura entra basura sale". Si los datos son incorrectos, mal organizados o maliciosos los resultados del entrenamiento también lo serán. Como nos podemos imaginar entrenar un modelo no es una tarea tan sencilla.

Es necesario asegurarse de no haber permitido, por descuido, que el agente se dedique decididamente a llevar a cabo acciones poco inteligentes. La definición propuesta implica que el agente racional no sólo recopile información, sino que aprenda lo máximo posible de lo que está percibiendo.

S. J. Russell y P. Norvig, 2004, Inteligencia Artificial. Un enfoque moderno, p. 42.

Aprender para una máquina será identificar elementos y establecer relaciones de diverso tipo, pero ¿Cómo es posible que aprenda una máquina? Aunque pueda parecer obvio debemos recordar que para hacerlo tendremos que proveerla de datos. Al conjunto de datos mediante el cual entrenamos al programa se lo llama *"dataset"* (Thompson, 2022). Los *datasets* pueden ser de muy diverso tamaño y tener diferentes características de acuerdo al tipo de información de la que se trate. Los datos serán el input de entrenamiento con que "alimentamos" al programa para que comience a detectar patrones. Estos datos pueden ser construidos a partir de imágenes, palabras, distintos tipos de valores numéricos, sonidos, etc. Ninguna teoría del aprendizaje sostiene que se pueda conocer el mundo que nos rodea si no se tiene algún tipo de experiencia. Los datos de entrada funcionarán como las "experiencias" que serán base para el aprendizaje de las máquinas. Hay dos tipos de estrategias para entrenar a nuestras máquinas, la de suministrar datos "etiquetados" o datos "no etiquetados". Veamos cómo se clasifican los entrenamientos según el tipo de datos que reciben los modelos (Figura 2).

1. Entrenamiento supervisado: se utilizará una etiqueta añadida a un dato que identifica de qué se trata, le proporciona la información al sistema. Pensemos que queremos enseñar al modelo a reconocer figuras geométricas de distintos colores, cuadrados y triángulos azules y rojos. Si el aprendizaje es supervisado, a cada dato que forme parte del dataset se le tiene que adosar la información del tipo de figura y del color del que se trata. Supervisar significa etiquetar, es decir identificar de qué tipo de entidad se trata el input que suministramos al modelo (se indica la respuesta correcta).
2. Entrenamiento no supervisado: no le proporcionamos etiquetas al programa y dejamos que sea el propio algoritmo el que identifique los patrones a partir de encontrar relaciones profundas al interior de la masa de datos.
3. Entrenamiento semi supervisado: en estos casos se combinan ambos tipos de datos, pueden ser útiles dependiendo de la cantidad y calidad de datos disponibles y de la estrategia general de diseño.

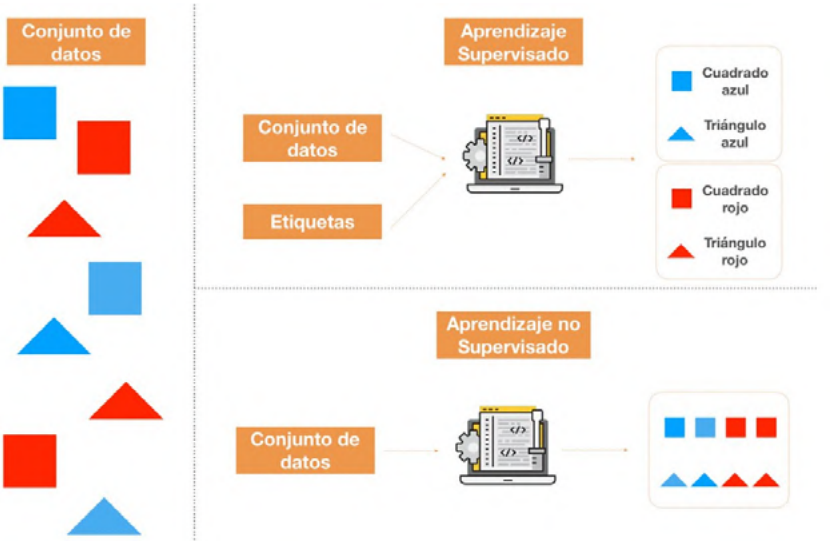


Figura 2. En los procesos de aprendizaje supervisados se suministran los datos y se le indica al modelo de qué se trata cada uno de ellos mediante etiquetas. En el aprendizaje no supervisado se suministra el conjunto de datos al modelo para que aprenda a ordenarlos y clasificarlos por sus propios medios, estableciendo relaciones entre ellos.Imagen tomada de: https://aprendeia.com/diferencia-entre-aprendizaje-supervisado-y-no-supervisado/

Hasta aquí vamos bien, pero ¿cómo sabemos si lo que aprendió es correcto o incorrecto? Cuando comenzamos a adquirir alguna capacidad de niños, como por ejemplo la prensión fina que nos permite tomar un lápiz con dos dedos, o manejar la cuchara para comer papilla, las cosas no nos resultan tan fáciles. Quienes tengan hijos recordarán el enchastre de puré de calabaza que puede producir un bebé mientras desarrolla sus capacidades de manipulación de la cuchara. Con los modelos de IA pasa algo similar, les lleva un tiempo aprender. En un comienzo domina la torpeza del bebe al comer y algo similar ocurre con los modelos en sus estadios tempranos de entrenamiento. El modelo necesita ser ajustado para cumplir con eficacia las tareas que queremos que desempeñe. A partir de lo experimentado pretendemos que pueda realizar generalizaciones que le permitan anticipar aquello con lo que nunca se ha enfrentado. De poco nos serviría una IA de reconocimiento de imágenes que sólo pudiera reconocer aquellas imágenes con que la entrenamos. Queremos que

se pueda enfrentar a nuevos entornos. Para las entidades biológicas, el aprendizaje es un proceso, se trata de una actividad del organismo desplegada hacia el medio ambiente durante períodos relativamente largos de asimilación. La asimilación de las características del medio ambiente presiona sobre el organismo hacia la reorganización de sus estructuras internas, al menos así lo plantea Piaget el máximo exponente del constructivismo en aprendizaje (Piaget, 1991).

Tengamos en cuenta de que se trata de un proceso de entrenamiento que llevará un tiempo, existe un tiempo de aprendizaje, no se trata de un suceso inmediato. Como vimos en los capítulos 3 y 4 relativos a la capacidad de aprendizaje humano, este proceso se verifica por ajustes muy específicos que permiten la adquisición de capacidades. Los modelos explicativos de la psicología genética harán énfasis en la organización lógica de la inteligencia, la psicología constructivista de la escuela soviética en la interacción social, la neuropsicología y las ciencias cognitivas en general harán eje en hipótesis sobre la formación de circuitos neuronales, la relación de la base material del cerebro y los procesos mentales y, lo que más nos importa, la capaciedad de cómputo. ¿Cómo se puede impulsar a un modelo de IA en el proceso de aprendizaje? Nada de esto es factible sin la intervención activa del ser humano que diseña estrategias posibles para orientar el proceso. Existen diferentes modelos de IA para diferentes tareas y cada uno de ellos tendrá sus falencias y virtudes. Debemos aprender a elegir las estrategias adecuadas a los problemas que enfrentemos. A esta altura vamos a suponer que elegimos una estrategia, es decir un tipo de modelo en particular, y lo queremos entrenar para que cumpla con su cometido. Existen variaciones importantes y estrategias diferentes para entrenar modelos, en nuestro ejemplo vamos a explicar los rudimentos generales utilizados para cierto tipo de modelos de ML, aunque existan otros modelos cuyas formas de aprendizaje y entrenamiento sean más complejas. Nos interesa transmitir una idea general. Manos a la obra.

Paso 1. Armamos nuestro *dataset*

Una máquina es una máquina. Una máquina no sabe, una máquina no entiende, una máquina procesa. ¿Qué puede procesar una máquina como es una computadora?: números. Una máquina puede realizar complejos cálculos a gran velocidad consumiendo gran cantidad de energía, eso es lo que puede hacer. La ilusión de que hablamos con ella, de que nos da consejos como un conocido, o un profesor o un experto es una ilusión. La idea de que ve un perro cuando la alimento con la imagen de un perro es una analogía con la que tratamos de comprender la interacción con un medio electrónico. No perdamos esto de vista, nuestro dataset, el input a partir del que iniciará su aprendizaje, debe ser traducido a un tipo de lenguaje que pueda interpretar el modelo. Desde los estímulos e informaciones que percibimos los humanos en nuestro mundo cotidiano, ya sea una imagen, un texto escrito o un sonido, hasta el tipo de información que puede percibir una computadora hay un largo camino. Tengamos en cuenta que nosotros, los seres humanos, también captamos el entorno dentro de ciertos umbrales perceptivos, podemos, por ejemplo, reconocer ciertas frecuencias de sonido y no otras, somos sensibles a determinados estímulos y no a otros. No pensemos en que el mundo natural se percibe de manera plena, la percepción es un filtro que retiene ciertas características y es insensible a otras, esto es así para nosotros y para las máquinas. Si no abandonamos nuestra postura de realismo inocente nos costará entender el sentido del aprendizaje. Cuando proporcionamos información a una máquina, decíamos, debemos hacerlo en un tipo de lenguaje que pueda asimilar. Tomemos por ejemplo un texto cualquiera en lenguaje natural, el que usamos en el mundo cotidiano. Entre uno y otro tipo de lenguaje existen tareas de traducción y transformación. Atravesaremos un conjunto de lenguajes jerarquizados, cada uno de ellos con su gramática y sus elementos. Tomemos por ejemplo un dataset compuesto por datos estadísticos y demográficos, con diversas variables de actitud y de conducta, aunque bien podríamos pensar en uno compuesto por sonidos, por conjuntos de textos o por imágenes. Podemos leer esos datos porque contamos con las competencias simbólicas necesarias para hacerlo, entendemos, por ejemplo, el idioma en que están expresados los datos. Cuando contamos con la

competencia simbólica necesaria un texto podrá cumplir funciones comunicacionales. Pero, como ya dijimos, las máquinas no hablan nuestros idiomas naturales. Aquí entra a jugar la transformación de un sistema codificado a otro sistema codificado. Pasaremos de nuestra capacidad simbólica de relacionarnos con el mundo en forma de lenguaje natural, a la instancia intermedia del código de programación y desde allí al lenguaje binario de las máquinas. Ese es el entorno que deberemos educar, un mundo muy complejo de ceros y unos. El estado actual de la tecnología nos lleva en ocasiones a perder de vista esto, ya que las interfaces gráficas se han tornado muy amables e intuitivas, pero detrás de nuestro procesador de texto, de las imágenes de los juegos o de los gráficos, subyacen enormes cantidades de ceros y de unos. Si avanzamos un poco más podemos señalar que tampoco existen ceros y unos ya que estos símbolos no son más que eso, representaciones humanas de procesos de apertura o cierre de circuitos. Aún más, el desarrollo de los soportes tecnológicos puede tomar otras orientaciones en el procesamiento de informaciones que nos obligan a concebir el dato en formas más complejas y contra intuitivas como son las de la cuántica. En cuanto nos acercamos con detalle a cualquier tipo de fenómeno computable nos damos cuenta de su complejidad y de los desafíos cognitivos que nos plantea. Aquello que está en la base material, física de los procesos de aprendizaje es un misterio, tanto en el caso de las entidades biológicas como el de las máquinas y estamos lejos de penetrar en sus últimos secretos.

Pasemos a la acción. Seleccionamos nuestro dataset con datos coherentes, pertinentes y ordenados en tablas con filas y columnas y, mediante algún lenguaje de programación lo acomodamos para que sea legible para el modelo. Esto que suena tan sencillo es muy trabajoso. Contar con datos de calidad es clave en este proceso, dar buenos ejemplos en una tarea de aprendizaje no es un tema menor. Como se suele decir en el ámbito de la investigación científica con relación a los datos y a los resultados que arrojan: si basura entra basura sale. Por lo tanto: seamos cuidadosos al seleccionar con qué materiales educamos a nuestros hijos, nuestras mascotas y a nuestras IAs. Conocer cuáles son los datos a partir de los cuales se entrenan los modelos IA nos aporta información necesaria para estimar cuál es el valor de los resultados que arrojan. Es importante

contar con informaciones no solo de contenido, sino de calidad, características, modalidades de filtrado, estructuración y organización de los datos. Contar con información abierta y disponible acerca de los datasets de entrenamiento de las IA debiera ser una práctica generalizada, al menos en los grandes modelos puestos a circular entre la población general. Obviamente, los datos de entrenamiento de IAs que contengan información sensible de personas, instituciones o empresas no tienen por qué ser de dominio público. No obstante, conforme avanzan los desarrollos, las informaciones no siempre se encuentran disponibles ni siquiera para los usuarios y contratantes de los modelos y los niveles de transparencia no son siempre los deseables (Thompson, 2022).

Paso 2. Entrenamiento a partir de *training set* (*dataset* de entrenamiento)

Habíamos dicho que nos interesa que los modelos puedan lidiar con datos a los que nunca se vieron expuestos y esto se puede lograr de varias maneras. Para comenzar deberemos evaluar las características que debe tener un training set para ajustarse a nuestros objetivos y a los problemas que queremos que la IA pueda enfrentar (traducir, funcionar como asistente en un chatbot, generar nuevo texto, predecir comportamientos climáticos o electorales, clasificar imágenes de animales, etc). Tengamos presentes que un set de entrenamiento es adecuado o inadecuado acorde a nuestros propósitos y que no existen recetas universales. Nada nos exceptúa de la tarea de comprender el ámbito de actividades en que nos encontramos inmersos, la IA es una herramienta que requiere de nuestra pericia analítica y nuestra capacidad de construir interrogantes significativos. Dicho lo anterior podemos pensar en cómo se puede organizar un set de entrenamiento y cómo podemos administrar los datos para llevar adelante un aprendizaje eficaz. Recordemos que tenemos dos tareas, la de permitir que el modelo identifique patrones y el de evaluar su capacidad de predecir a partir de ellos cuando le suministramos información nueva.

Para ello los ingenieros de software encontraron una estrategia muy sencilla y eficiente: dividir el dataset entre un conjunto de entrenamiento

As new AI technology rapidly progresses, there has been a decline in documentation quality about the datasets used to train these models. What is really inside my AI? What is it made of?

A. D. Thompson, 2022, What's in my AI? A Comprehensive Analysis of Datasets Used to Train GPT-1, GPT-2, GPT-3, GPT-NeoX-20B, Megatron-11B, MT-NLG, and Gopher, p. 5.

y otro de prueba. Si cuento con un conjunto de datos ordenados según características que me interesa que registre el modelo, por ejemplo, cincuenta mil imágenes de vehículos, las estadísticas de ingresos del INDEC de la República Argentina (Instituto Nacional de Estadística y Censo) o cien mil mails algunos de los cuales son Spams y otros no, puedo utilizarlos para entrenar el modelo y lograr que prediga si una imagen se trata de un auto o una bicicleta, cuánto ganará una persona con determinadas características en un período de tiempo o clasificar un mensaje en spam o no spam. Tomaré un porcentaje de los datos, por ejemplo, un 80% de ellos y los usaré para entrenar el modelo, para lograr que encuentre patrones entre los datos, detectando relaciones internas entre sus características (*features*). Con el otro porcentaje, en este caso el 20% de los datos vamos a controlar la capacidad del modelo de predecir (modelos de correlación) de clasificar (modelos de clasificación) o de agrupamiento (clusterización). Es decir, tomo una parte del modelo que realizará una predicción del tipo: bicicleta: SI / un porcentaje de posibilidad de que tenga un rango de ingresos / que indique Spam SI o Spam NO / que complete una oración con una palabra / que indique que se mueva la reina a algún casillero de ajedrez, todo dependerá del propósito para el que lo estamos entrenando. Una vez que se ha realizado el primer entrenamiento con la mayoría de los datos con los que cuento, tomaré ejemplos extraídos de la parte que reservé para controlar y le pediré al modelo que entregue su predicción. Le entregó una imagen de bicicleta y veo si predice bicicleta, uso como input un mail con Spam y veo si predice Spam o no Spam. A ese output predictivo lo llamo valor predicho y lo puedo comparar con el valor real. Esto es posible porque yo conozco el resultado correcto, ya que me reservé una parte del conjunto de datos, sé que la imagen que le suministré era una bicicleta y no un monopatín o que el mensaje era o no era spam. Esa diferencia entre el valor real y el valor predicho me indicará lo bien o lo mal entrenado que está mi modelo. En un comienzo, sobre todo en entornos de alta complejidad, las predicciones serán bastante deficitarias, ya que el modelo procederá al azar porque aún no cuenta con la suficiente experiencia, es decir no está ajustado. Este tipo de estrategia que describimos es sólo una dentro de los modos de aprendizaje posibles. Existen otros como por ejemplo el de redes adversarias de las que nos ocuparemos más adelante.

Volver

La falta de transparencia sobre el diseño y las dificultades para entender la forma en que operan las IAs colocan al usuario en desventaja.

A complex computational system is transparent if all of the details of its operation are known. A system is explainable if humans can understand how, it makes decisions. In the absence of transparency or explainability, there is an asymmetry of information between the user and the AI system, which makes it hard to ensure value alignment.

S. J. D. Prince, 16 de enero de 2024, Understanding Deep Learning, p. 425.

Paso 3. Ajustamos el modelo: la configuración de los pesos

Volvemos a recordar que, aunque nos parezca que estamos tratando con imágenes o con palabras, con votos o con movimientos de ajedrez, estamos lidiando con valores numéricos en lenguaje máquina. Queremos que el modelo nos ayude a resolver algunas cuestiones, determinados tipos de problemas que nos formulamos con relación al entorno. Para ello le brindamos experiencias en forma de informaciones de todo tipo, imágenes, frecuencias de sonido, distribuciones porcentuales, lo que se nos ocurra. Por ejemplo, podemos suministrarle números escritos a mano que queremos que aprenda a identificar. Estos números tendrán ciertas características (*features*) cómo incluir redondeles como el cero, el ocho, el seis o el nueve, tener palitos como el uno y el cuatro, incluir ángulos como el cuatro. Nuestro modelo deberá encontrar vínculos entre estas características para así, una vez entrenada, poder predecir de qué número se trata: redondel pequeño en la parte superior con palito que mira para abajo será nueve, si mira para arriba será seis, si tiene dos círculos que se unen será el ocho. Queda claro que la variación de caligrafía en los números escritos a mano es mucha, cada uno tiene un tipo de letra que lo caracteriza. Por eso, cuando le mostremos un número escrito de manera tal que no figure en el dataset con que lo entrenamos pretendemos que logre identificarlo correctamente. Queremos acortar la diferencia entre lo que predice el modelo y los valores correctos. No obstante, esto puede resultar algo engañoso, ya que las características que me permiten predecir determinados fenómenos, como por ejemplo el voto, el resultado de un examen o la elección de un tipo de vestimenta, puede depender de muchos factores y no todos los factores tendrán el mismo peso. Algo similar ocurre con la clasificación de animales o la elección de la próxima palabra en una oración. Si queremos que la IA realice anticipaciones coherentes y efectivas tendremos que lograr que llegue a determinar la importancia de los distintos atributos. Por ejemplo, la probabilidad de ser víctima de violencia familiar puede correlacionarse con el nivel de escolaridad o con el lugar de residencia, pero cabe

conjeturar que el sexo de la persona será un factor de mucha mayor gravitación. El aprendizaje reta nuestras conjeturas y las evalúa en forma de ajuste predictivo. Entrenar implica lograr que el programa logre detectar cuál es el peso que se debe atribuir a cada uno de los atributos que logra identificar. No es difícil de entender, aunque requiere una pequeña aclaración matemática. Aquello que nos asombra de la IA es la posibilidad de estos modelos de aprendizaje automático de detectar las relaciones profundas de estas características por sí solas, de formar algo así como hipótesis en torno a los patrones, vínculos y regularidades en el conjunto de datos que le proporcionamos. El cálculo es sencillo, las características las multiplica por un determinado número. Simplificando podemos decir que a la característica o *feature* (x) la multiplica por un número (w). Esos números por los que multiplica a las características se los llama pesos y eso es lo que se debe ajustar, el conjunto de pesos de cada una de las entradas para lograr que el error de las predicciones sea adecuado a nuestras necesidades. Cada tipo de modelo realiza estas operaciones de ajustar los pesos de maneras diferentes, con mayor o menor complejidad, esfuerzo de cómputo y tiempo. En la práctica, estos entrenamientos se realizan varias veces y se compara los resultados para ver con cuál de todos los modelos ajustados conviene quedarse. Podemos entrenar programas con algún objetivo equivalente, con el mismo set de datos y obtener diferentes conjuntos de pesos. Entre las distintas alternativas que nos arrojan los entrenamientos elegiremos la que mejor rendimiento tenga, la de mejor performance. Lo más valioso en estos casos es el conjunto de pesos de entrenamiento de un modelo. Entrenar modelos sencillos de correlación o clasificación o pequeñas redes neuronales es relativamente fácil y rápido. Entrenar gigantescos modelos generativos de lenguaje o imagen implica una inversión monstruosa desde el punto de vista económico, energético y de conocimiento.

Error de infra ajuste (underfitting): *indica que el sistema no es suficientemente complejo para aprender los datos de entrenamiento. Para corregirlo hay que incrementar la complejidad del modelo (por ejemplo, más unidades en el caso de una red neuronal).*

Error de sobreajuste (overfitting): *indica que el sistema no es capaz de generalizar lo aprendido con los datos de entrenamiento a otros datos.*

R. Benítez, G. Escudero, S. Kanaan, D. Masip Rodó y A. Cencerrado Barraqué, 2018, Inteligencia Artificial Avanzada, p. 222. Universitat Uberta de Catalunya. PID_00250574.

Todo esto resulta coherente y pareciera que vamos bien, pero nos topamos con el problema del overfitting y el underfitting. ¿De qué se tratan estos problemas con los que se enfrentan los entrenamientos de modelos de IA de manera permanente? Un modelo puede haberse entrenado tan bien y ajustado tanto a los datos con que se lo entrenó que haya perdido flexibilidad. Es como si se hubiera ajustado tanto a lo que sabe que se hubiera convertido en un necio. Todos conocemos seres humanos o incluso otro tipo de animales tercos que se ajustan a estas descripciones. A esto se le llama estar sobre ajustados. Es lo que se denomina en jerga de programación estar *overffiteado*. Es obvio que no queremos generar modelos necios que no puedan predecir ante eventos nuevos, pero tampoco modelos ignorantes que no hayan aprendido a "descifrar" sus datos. Señalamos que cuando comienza el entrenamiento un modelo puede haber sido alimentado con los datos adecuados, haber identificado patrones al interior de los mismos, pero sus predicciones pueden ser torpes. Está aprendiendo y se está ajustando. ¿Pero nos conviene que se ajuste a la perfección a los datos con los que lo entrenamos? La respuesta es que no siempre, en muchos modelos, como en los de redes neuronales, no resulta conveniente ya que esto le quitaría flexibilidad. Un ajuste total a los datos de entrenamiento puede desembocar en que no pueda predecir más allá de lo que ha aprendido. Si se ha ajustado totalmente al reconocimiento de imágenes geométricas como cuadrados y triángulos incluidas en un data set que incluye colores rojo y amarillo, puede que no logre comprender que un cuadrado azul también es un cuadrado. Será un modelo necio. Queremos ajustar el error en cierta medida. Pretendemos que, aun cuando se lo haya entrenado con imágenes de aves en vuelo pueda identificar que un pichón en el piso o un pingüino pertenecen también a ese conjunto (ejemplo tomado de uno de mis divulgadores preferidos Carlos Santana Vega).

Perder fue muy duro. Antes de jugar con AlphaGo, pensé que ganaría. Después del primer juego cambié mi estrategia y peleé más, pero perdí. El problema es que los humanos a veces cometemos errores muy grandes, porque somos humanos. A veces estamos cansados, a veces tenemos tantas ganas de ganar el juego, tenemos esta presión. El programa no es así. Es muy fuerte y estable, parece un muro. Para mi esta es la gran diferencia. Sé que AlphaGo es una computadora, pero si nadie me lo dijera, tal vez pensaría que el adversario es un poco extraño, pero un jugador muy fuerte, una persona real.

Lee Sedol luego del match con Alpha Go.

E. Gibney, 2016, Go players react to computer defeat. Nature.

El aprendizaje por refuerzo

Se trata de otro gran protagonista en la escena de las IA. Ha cobrado una importancia creciente en los últimos tiempos y es la base de estrategias exitosas de entrenamiento. Se trata de un enfoque emparentado con la psicología behaviorista del comportamiento y el conductismo de Skinner que mencionamos en el capítulo 3 (Skinner, 1986). Según el conductismo, una conducta incrementa la posibilidad de ser repetida si ha obtenido resultados favorables en el pasado. A la inversa, si una conducta obtiene resultados desfavorables su probabilidad de ser reiterada por el individuo decrece. El medio ambiente premia o penaliza las conductas y, de esta manera, afecta su probabilidad de ocurrencia en un sentido positivo o negativo. Este simple principio psicológico postulado para los seres vivos subyace en las formulaciones del aprendizaje por refuerzo de modelos de IA.

El objetivo de esta estrategia es lograr que un agente inteligente logre optimizar su política de acción en interacción con un medio complejo. Esta política consiste en una serie de reglas que tomará el agente para definir el curso de acción en determinados ambientes con los que interactúa en consideración con su evaluación del estado (Russell y Norvig, 2004). Dicho en lenguaje sencillo, que aprenda a ganar en el juego en que se la entrena. Para entender esto pensemos en los cursos de acción que tomaríamos en un video juego. ¿Qué conviene, escapar o arriesgarse a enfrentar el monstruo y ganar el diamante? En este tipo de entrenamiento es necesario formular problemas y objetivos que el agente debe cumplimentar. Debe entender qué implica ganar y que implica perder para optimizar sus políticas generales. En próximos capítulos comentaremos varios detalles de estos modelos. Este enfoque es el que permitió destronar a Gary Kasparov en una fecha histórica en la cronología IA, el 10 de enero de 1996. La estrategia de aprendizaje reforzado ha demostrado su capacidad de dominar los escenarios más complejos aún en los entornos de juego más exigentes tal y como lo demostró en tiempos más recientes al vencer al campeón mundial de Go, Lee Sedol, a quién derrotó 4 a 1 en un match celebrado en marzo de 2016. El entrenamiento de Alpha Go tuvo dos fases, en la primera se la alimentó con ejemplos de las

partidas de los jugadores más talentosos del mundo. En este contexto sus capacidades fueron altas pero limitadas, ya que se mantenían en el marco de la creatividad humana. La etapa más interesante desde el punto de vista del aprendizaje no depende ya de los ejemplos de movimientos en partidas humanas, sino de la etapa de auto juego. En esta fase de aprendizaje el modelo realiza una exploración aleatoria del enorme campo de posibilidades que brinda el Go. En el juego del Go las posibilidades de jugadas son inmensas y no resulta sencillo representarnos la lógica que subyace a las mismas. Para comprender de qué se trata el modelo de aprendizaje por refuerzo acudiremos a un juego más sencillo que nos ofrece una captación más intuitiva. A continuación, incluimos un gráfico que ejemplifica la lógica de aprendizaje por refuerzo en nuestro conocido TA TE TI. (Figura 3)

Entrenamiento reforzado por feedback humano (RHLF): se trata de un enfoque mixto, que combina el aprendizaje por refuerzo con la supervisión humana. Su particularidad reside en la noción de retroalimentación humana. Esta estrategia ha dado muy buenos resultados en términos de eficacia, en particular en el entrenamiento de grandes

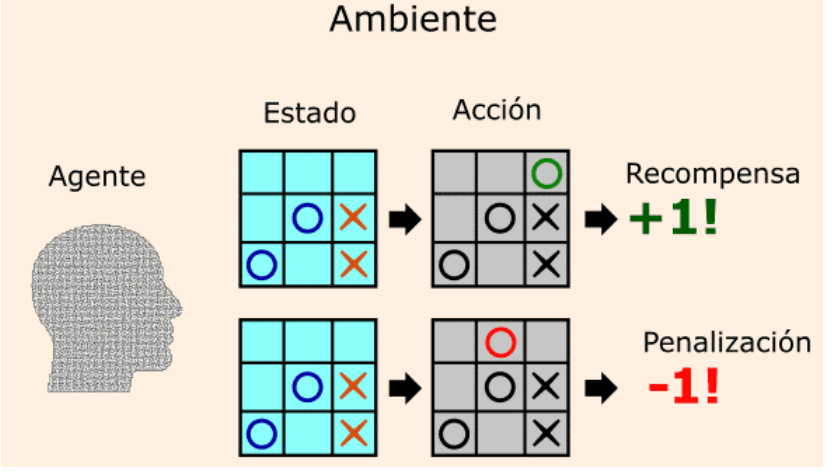


Figura 3. En este juego, el ambiente es el juego en sí. El estado es la situación actual del juego. El agente es el que toma las decisiones. La acción es la elección de la casilla. La recompensa es la victoria, y la penalización es la derrota. Imagen extraída de: https://www.ceupe.com/blog/aprendizaje-por-refuerzo.html

modelos de lenguaje como los de Open AI Gpt y también ha sido utilizado por Anthropic en su modelo Claude2 o Deep Mind en algunos de sus propuestas (Natolambert, 2023). En principio, para comprender el alcance de este tipo de aprendizaje se debe tener en cuenta dos nociones, la de recompensa y la de feed back humano. Estos dos elementos configuran el núcleo del aprendizaje reforzado (RL) mediante feedback humano (HF). Algo emparentado vimos en capítulos anteriores con relación al refuerzo de la conducta en el caso de los organismos biológicos cuando nos referimos al conductismo. El aprendizaje por refuerzo está íntimamente ligado a la noción de recompensa. Recordarán que la idea de refuerzo de las conductas de un organismo, incluido el humano, se relaciona con el resultado de sus conductas. Cuando el individuo percibe que su conducta tuvo un resultado positivo la probabilidad de repetirla se ve incrementada, si percibe que el resultado ha sido negativo la probabilidad de repetirla será menor. El comportamiento reforzado, entonces, toma en cuenta los condicionantes que llevan a un organismo a actuar de una u otra manera. Aquí hay una diferencia sutil pero crítica, no se trata de que el resultado de la conducta oriente al individuo. Lo que orienta la conducta es la percepción del resultado de sus conductas. Por eso el diseño de los modelos deben contemplar que la IA pueda percibir adecuadamente cuál es la calificación que obtiene su conducta. En neuropsicología o en psicología conductista, la noción de recompensa es parte de este concepto más amplio que es el de conducta reforzada. Este es un muy buen ejemplo si queremos entender cómo una disciplina como es la neuropsicología o la teoría del aprendizaje de la psicología conductista funcionan como orientadores en el diseño de soluciones en informática.

Hasta aquí todo parece sencillo, pero veremos que no lo es tanto. Muchas veces resulta muy difícil determinar cuáles son las recompensas que conducen a estados internos del individuo que lo impulsen en uno u otro sentido. Prestemos atención, estamos hablando de estados internos del individuo, este es un punto clave. El RHLF tiene por propósito incidir en la conducta de un agente de IA para orientar sus decisiones en un sentido determinado. ¿En qué sentido? En el sentido de las preferencias humanas. ¿Cómo se puede lograr? A partir del feedback de seres humanos que califique en un sentido o en otro sus decisiones anteriores.

Para que esto ocurra es necesario que el modelo esté entrenado y que produzca algún tipo de resultado que se quiera orientar. Debemos determinar qué se considera como rendimiento esperado, una clasificación más flexible o rigurosa de imágenes, un criterio ético o que minimice riesgo, un criterio vinculado a la seguridad de los datos, la eficacia de una instrucción para toma de decisión, ustedes elijan porque la IA ofrece un campo amplio de aplicaciones y funciones en muchos planos de la actividad de la humanidad, las sociedades, las instituciones o la vida de las personas. Debemos tener algún criterio humano a partir del cual nos interesa modelar la conducta del agente (de la IA). Un ejemplo de ello es el de los sesgos, no querremos que un modelo de lenguaje tenga orientaciones racistas o sesgos de género. El RHLF no puede adoptar todos los criterios que los seres humanos querríamos introducir en el comportamiento de nuestros agentes inteligentes. Aquí entra a jugar una toma de decisión en los esfuerzos de aprendizaje de las IA. El RHLF es costoso en horas hombre y en término de procesamiento por eso es de interés debatir cuáles son los ejes prioritarios en que se apuesta en el alineamiento de las IA con estas metodologías. Los ambientes reales son muy difusos y complejos, también es muy difusa y compleja la información con la que se alimentan los modelos de IA y, sin adelantar demasiado, las fronteras se hacen cada vez más amplias ya que la multimodalidad suma imagen, video y audio a las ya existentes alternativas de data sets compuestos de texto de las LLMs. En ambientes de alta complejidad y con niveles considerables de ambigüedad, el RHLF puede ser una herramienta eficaz para orientar la conducta de los agentes de IA.

Para comenzar se tiene que entrenar un modelo de IA y este modelo debe estar preparado para poder recibir feedback de sus acciones. Esto es importante, recuerden que cuando les contamos sobre el conductismo hicimos hincapié en que la clave residía en que el organismo pueda registrar si el resultado de su conducta es positivo o negativo. Así que, desde un principio la IA debe estar configurada para registrar el resultado de su conducta como algo positivo o negativo y poder modificar su comportamiento futuro en ese sentido. Lo segundo que tendremos que hacer es recopilar datos acerca de las acciones de la IA y clasificar los resultados según criterios humanos. Para esto hay que definir cómo etiquetarlas,

por ejemplo, si queremos trabajar con sesgo pueden clasificarse en "inofensivas" u "ofensivas" (Bai et al., 2022). Luego tenemos que entrenar un modelo de recompensa. En este punto entra a jugar la valoración humana que otorga puntajes a las acciones del agente inteligente (scoring). Es el universo del criterio valorativo humano. Aquí se nos complica un poco el tema. Para recompensar a nuestro gran modelo de lenguaje, tenemos que preparar otro modelo que nos permita automatizar la función de recompensa. No piensen en que vamos a tomar una a una las respuestas del modelo al que queremos entrenar e informarle manualmente nuestra calificación. Lo que se va a hacer es "tunear", es decir ajustar un modelo más pequeño que incluya las preferencias humanas y las traduzca a un lenguaje que pueda ser asimilado por el modelo que queremos "educar". Esta recompensa será un número (recompensa escalar). Como siempre hemos remarcado, debemos pasar los estímulos (inputs) a un lenguaje interpretable por las máquinas y los lenguajes que ellas aceptan son numéricos. Una vez que tenemos preparado este modelo de recompensa lo ponemos a trabajar (será como nuestro "profesor") en el entrenamiento del gran modelo de lenguaje que queremos tener bien alineado con las preferencias humanas.

Otros tipos de estrategias de aprendizaje

Las estrategias de aprendizaje se multiplican constantemente lo que resulta muy desafiante para los que intentamos comprender los alcances de los modelos y su vinculación con la inteligencia en general, tanto en el plano humano como artificial. Los logros alcanzados por los aprendizajes son cada vez más impactantes y han logrado descifrar, por ejemplo, el plegamiento de proteínas, la generación de agentes capaces de optimizar sus políticas de acción en interacción con medios complejos o la resolución de cuestiones matemáticas como la generación de algoritmos eficaces en la multiplicación de matrices, un problema que llevaba 50 años sin progresar. Hoy en día se cuenta con muchas vías de desarrollo abiertas a la exploración de las fronteras de las IA. La mención de otras estrategias de aprendizaje no seguirá una clasificación tan prolija como la anterior en que discriminamos entre aprendizaje supervisado, no supervisado o por

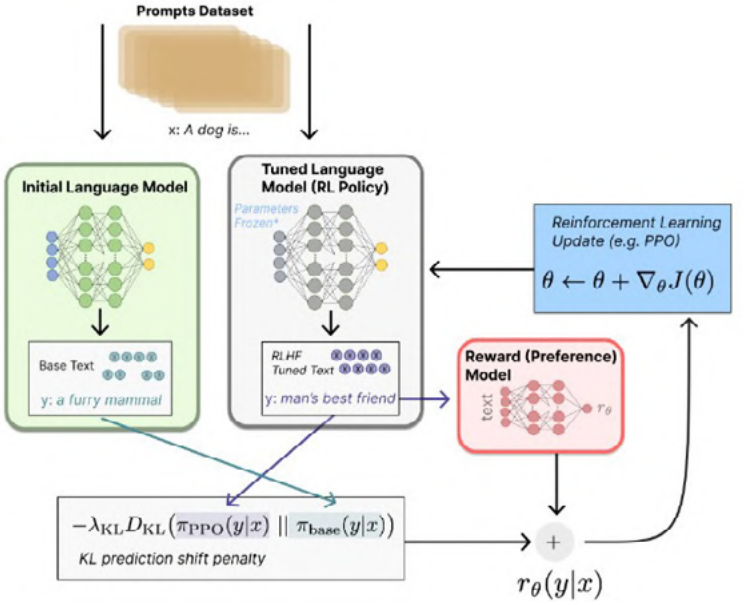


Figura 4. En este gráfico se ilustran las diversas instancias involucradas en el proceso de refuerzo de conducta del agente inteligente. Si prestan atención a las flechas se darán cuenta que se trata de un ciclo iterativo (repetitivo) de entrenamiento. Se parte de un dataset de prompts (pedidos que le haremos al modelo) con que se alimenta un modelo inicial de lenguaje que queremos orientar. Al modelo que queremos entrenar se le requiere algo (en este caso un prompt x: un perro es...). Lo siguiente es calificar su respuesta para luego reorientar su conducta hacia una respuesta que nos parezca adecuada. Al prompt x "un perro es...", el LLM contesta y: "un animal peludo...". Lo vamos a premiar o a desalentar mediante una recompensa. Para ver si actúa según nos complace, contamos con un modelo de recompensa (el rectángulo rosado) que califica la respuesta según los criterios con los que lo hemos entrenado. Esta calificación será remitida al modelo que queremos "educar". Su próxima respuesta será ajustada a nuestros criterios humanos. La próxima respuesta a "un perro es...", será y: "el mejor amigo del hombre...", output que volverá a ser calificado. Quienes quieran una explicación más técnica que la que les brindamos pueden remitirse a https://github.com/huggingface/blog/blob/main/rlhf.md Ilustración tomada de https://github.com/huggingface/blog/blob/main/rlhf.md#illustrating-reinforcement-learning-from-human-feedback-rlhf

refuerzo. Nos encontramos en un punto del ciclo de desarrollo de las IAs en que se abren múltiples caminos en este intento de dotar a las máquinas de la capacidad de aprender. Para introducir estas cuestiones es conveniente mencionar otro tipo de estrategias, en principio las siguientes:

Utilización de modelos pre entrenados: se trata de un campo con enorme productividad en el universo de las IA, el de la transferencia de

aprendizaje *(transfer learning).* Nuevamente voy a acudir a un ejemplo de Carlos Santana. Si queremos seleccionar un docente para enseñar química y tenemos dos opciones, alguien que enseña lenguas, pero no tiene conocimientos de ciencias exactas y naturales o un bebé de seis meses que no sabe hablar, ¿qué decisión nos parece adecuada? Puede que nuestro docente de lenguas tenga el cargo asegurado. Esta idea tan sencilla ha sido aprovechada en el entrenamiento de IAs de todo tipo. El principio básico que lo permite es el de *transfer learning* que apunta a tomar modelos entrenados con un propósito y aprovechar su capacidad de identificar patrones como plataforma para adquirir nuevas capacidades orientadas a tareas diferentes.

Entrenamiento con datos sintéticos: una estrategia en la cual los datos son generados por la propia IA. Son denominados sintéticos porque los datos no son tomados de informaciones disponibles en el exterior del modelo, sino generados por el mismo modelo. Se utilizan en diversas estrategias de diseño, por ejemplo, en las redes adversarias (Russell y Norvig, 2004). Este tipo de arquitectura pone a competir dos instancias como adversarios que llevan el nombre de generador y de discriminador. Una, la llamada discriminadora, cuenta con un entrenamiento supervisado en el cual ha aprendido a reconocer algún tipo de entidad por ejemplo imágenes de rostro. La otra, llamada generador produce imágenes que envía al discriminador tratando de "engañarlo". Cada vez que el discriminador identifica que la imagen es falsa recibe un premio y se penaliza al generador. Este ciclo continúa hasta que el generador logra ajustarse lo suficiente como para que el discriminador no logre distinguir lo que produce de las imágenes tomadas de la realidad (Figura 5).

Aprendizaje contrastativo: a diferencia del aprendizaje supervisado en que proporcionamos ejemplos de entrada y le enseñamos a discriminar entre predicciones correctas o incorrectas, en los modelos de aprendizaje contrastativo se entrena el modelo para que identifique similitudes y diferencias. Recordemos que en el aprendizaje supervisado el IA encontraba patrones en los datos y formaba hipótesis predictivas que se contrastan con los datos reales que nosotros conocemos. Por ejemplo, predecirá que un mail es o no spam (resultado predicho) que podremos

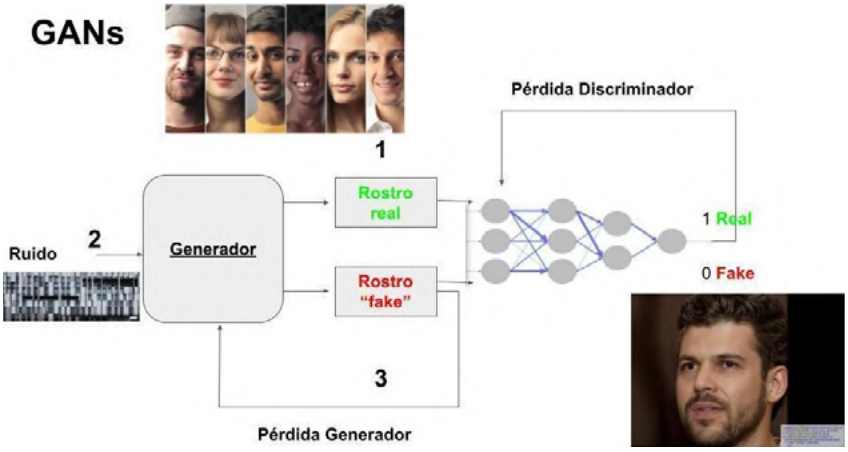


Figura 5. Las redes GANs (Generative Adversarial Network) son modelos muy eficientes al momento de generar imágenes, la brecha entre la imagen sintética y la real se acorta y falta muy poco para que nos resulte imposible distinguirlas a partir de nuestro sistema natural de percepción. Imagen extraída de: https://lucentialab.com/2021/02/23/redes-gans/

comparar con el resultado real, determinando si la predicción del modelo es correcta o incorrecta. El aprendizaje contrastativo opera de otra manera, apunta a que el modelo pueda predecir la similitud o diferencia entre pares de datos proporcionados. Para ello se alimenta el modelo con pares de datos similares y pares de datos diferentes. Destacamos este enfoque porque se asemeja a las estrategias didácticas utilizadas para el aprendizaje de lenguas extranjeras entre seres humanos. De hecho, esta es una de las principales inspiraciones para este enfoque. Aprender diferencias y semejanzas siempre ha sido una vía central en la generación de capacidades simbólicas humanas, ya hemos explorado la importancia que le otorgaron a los pares diferenciales los estructuralistas franceses. Por otra parte, abre el campo de los procesos de razonamiento abductivo uno de los grandes continentes explorados en el mundo de las IAs (Miller, 2019). Consolidar la capacidad de construir semejanzas y diferencias parece ser una alternativa consistente en términos de la facultad de la inteligencia. (Figura 6).

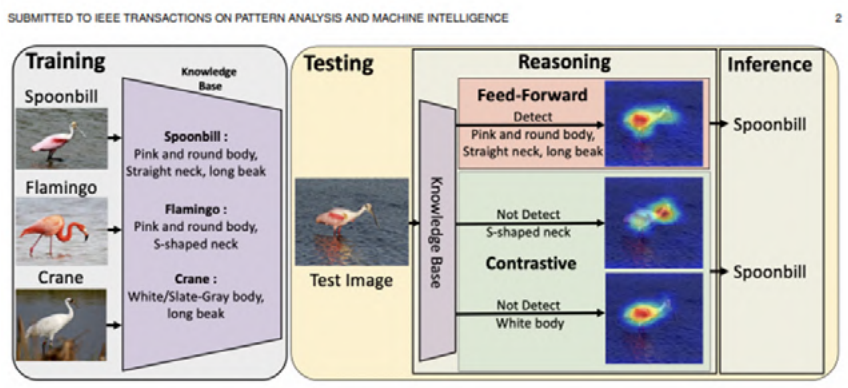


Figura 6. Las características discriminatorias se identifican en los marcos de razonamiento Feed-Forward. Los contrastes entre los hechos observados y una base de conocimientos conocida se utilizan como razones en la inferencia contrastiva. Prabhushankar, M. 2021. Ilustración 1.

En este ejemplo se trata de que el modelo pueda distinguir entre tipos de aves de aguas poco profundas: flamencos, espátulas y grullas a partir de diferencias y similitudes bastante finas. Los flamencos y las grullas tienen en común las características de poseer cuerpos redondos y rosados, pero difieren en la forma de sus cuellos. Mientras que los flamencos presentan cuellos en forma de S las grullas tienen cuellos rectos. A su vez las espátulas tienen un cuerpo blanco agrisado lo que las diferencias de grullas y flamencos, en tanto que su cuello es recto y largo, lo que las ubica en conjunto con las grullas y las diferencia de los flamencos. Lo más importante es la generación de la capacidad de razonamiento, de detección de esquemas de similitudes y diferencias, lo que genera una capacidad de razonamiento generalizables a otros dominios (Prabhushankar, 2021).

Hasta aquí nos mantuvimos en el plano general y eludimos, en gran medida, el tema de las grandes redes neuronales y los modelos generativos de lenguaje con interfaces conversacionales. En el próximo capítulo nos sumergimos en estas tecnologías que desataron el furor global y que pueden llegar a ser responsables de la próxima gran revolución de los medios de producción.

Volver

Bibliografía

Bai, Y., A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. J. Henighan, N. Joseph, S. Kadavath, J. Kernion, T. Conerly, S. El-Showk, N. Elhage, Z. Hatfield-Dodds, D. Hernandez, T. Hume, S. Johnston, S. Kravec, L. Lovitt, N. Nanda, C. Olsson, D. Amodei, T. B. Brown, J. Clark, S. McCandlish, C. Olah, B. Mann y J. Kaplan (2022). Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. ArXiv, abs/2204.05862.

Gibney, E. (2016). Go players react to computer defeat. *Nature* https://doi.org/10.1038/nature.2016.19255

Lucentia Lab (2021). *La otra cara de las redes GANs*. https://lucentialab.com/2021/02/23/redes-gans/

Miller, T. (2019). Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence, 267*, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

Natolambert, L. C. (2023). *Illustrating Reinforcement Learning from Human Feedback (RLHF)*. https://github.com/huggingface/blog/blob/main/rlhf.md

Prabhushankar, M. (2021). *Contrastive Reasoning in Neural Networks* [Tesis de doctorado, School of Engineering, Georgia Institute of Technology] http://hdl.handle.net/1853/70076

Piaget, J. (1991). *Psicología de la inteligencia*. Siglo veinte.

Russell, S. J. y P. Norvig (2004). *Inteligencia artificial. Un enfoque moderno*. Pearson Ealan.

Skinner, B. (1986). *Sobre el conductismo*. Planeta.

Thompson, A. D. (2022). *What's in my AI? A Comprehensive Analysis of Datasets Used to Train GPT-1, GPT-2, GPT-3, GPT-NeoX-20B, Megatron-11B, MT-NLG, and Gopher*. LifeArchitect.ai.

Capítulo 6. La inteligencia artificial y el lenguaje

Sumario

Existe una disciplina de las ciencias de la computación conocida como NLP (Natural Language Processing) que utiliza algoritmos para lograr que las computadoras puedan interactuar con los seres humanos mediante el lenguaje corriente, aquel que utilizamos para comunicarnos en nuestra vida cotidiana. El NLP contiene dos subramas el NLU (Natural Language Understanding) que se ocupa de la comprensión del lenguaje y el NLG (Natural Language Generation) que se enfoca en la generación de texto. Las redes neuronales artificiales son algoritmos que toman por modelo las redes neuronales biológicas y han generado un avance crucial en la conquista del lenguaje por parte de las computadoras. No es posible terminar de descifrar cuáles son los patrones a partir de los cuales realizan sus inferencias, son modelos "de caja negra". Los grandes modelos de lenguaje (LLMs) utilizan un modelo particular de redes neuronales de aprendizaje profundo conocido como Transformer y son capaces de generar texto que nos resulta coherente a nosotros los humanos. Su diseño le permite generar textos nuevos que no son meras repeticiones de los datos con que los hemos entrenado.

El lenguaje computacional, entre científicos e ingenieros

¿Qué es el lenguaje? No podemos responder esa pregunta de manera definitiva. Lo que sí podemos hacer es acercarnos a la forma en

que se ha conceptualizado en distintos campos y las disputas que esto ha generado.

Queremos avanzar en la comprensión de la relación entre la IA y el lenguaje, a esto se han dedicado científicos e ingenieros. ¿Qué diferencia a un ingeniero de software de un científico académico? En primer lugar, el horizonte de sus tareas. Mientras la ciencia se plantea un estándar muy alto en términos de rigor, la ingeniería se plantea un estándar muy alto en términos de eficacia.

Tenemos dos caminos posibles de acercamiento a la conquista del lenguaje por medios computacionales. El primero intenta adoptar las soluciones más rigurosas y que pueden ser probadas mediante los métodos formales de la lógica, la matemática y la lingüística; el enfoque racionalista propuesto en su momento por Chomsky y que presentamos en el capítulo 2. El segundo es el de validar las ideas mediante pruebas empíricas y ver qué ocurre. Este enfoque sigue la orientación práctica del pensamiento técnico. Estas dos tendencias se hicieron presentes en el intento de conquistar el lenguaje mediante mecanismos artificiales. Compitieron, se criticaron mutuamente, permitieron avanzar el proyecto o lo obstaculizaron. Hoy la contienda persiste y es dable suponer que se va a profundizar.

¿La IA nos enseña algo sobre el lenguaje humano? ¿Nos aporta conocimiento sobre los procesos de inteligencia? ¿Cuándo hablamos de aprendizaje, de razonamiento o de memoria en una máquina nos referimos a algo relacionado con las facultades humanas superiores? Este es un debate abierto que dará mucho que pensar en el futuro. La presión de la industria en la instalación de sus productos y su caja de resonancia mediática en la difusión de novedades instalan muchas ideas con relación a la IA y sus alcances. Tienden a presentarla como la clave para comprender qué es la inteligencia humana y el lenguaje, recomiendan cómo debemos educar y cuál debe ser la organización social que se adapte a la revolución industrial que ellos encabezarían. Por otra parte, lingüistas, filósofos y diversos actores institucionales se arrogan el derecho a levantar la bandera moral en nombre de la verdad, la ciencia y el bien común. La controversia es inevitable.

Antes de debatir es conveniente conocer algo sobre la materia en disputa. Aunque no tengamos la respuesta a estas controversias les proponemos comenzar a conocer algo de estos grandes protagonistas que tanto interés nos generan: el procesamiento del lenguaje, las redes neuronales artificiales y los grandes modelos generativos cuyo uso se ha masificado a partir de noviembre de 2022.

El procesamiento del lenguaje natural por medio de computadoras

El procesamiento del lenguaje natural (NLP) es la manera en que se enseña a las máquinas a comprender y generar lenguaje humano en todas sus manifestaciones, de manera escrita o mediante el habla (Pineda Cortés, 2017). Es parte de las ciencias de la computación y se enfoca en el intento de lograr una interacción entre las computadoras y los humanos a partir del lenguaje. Esta tarea no ha sido para nada sencilla y ha requerido la confluencia de varias disciplinas de saber, la lingüística, la teoría de la información, la estadística, la lógica, la matemática e incluso la biología (Jurafsky y Martin, 2023). Cada uno de estos saberes ha colaborado para acercarse al objetivo de que una computadora pueda manejar el lenguaje de manera eficaz y coherente. Se trata de una meta a la que nos hemos aproximado por la colaboración interdisciplinaria. Cuál es el grado en que las IA logran dominar el lenguaje o alcanzar su comprensión es, a la fecha de diciembre de 2023, materia de controversia.

Con la denominación NLP se alude a un paraguas que abarca distintas técnicas vinculadas al manejo del lenguaje por parte de las computadoras (Figura 1). El NLP se divide en dos ramas bien diferenciadas: la que se ocupa de enseñar a las máquinas a entender el lenguaje (NLU) y la que se ocupa de preparar a las máquinas para generar lenguaje (NLG).

El NLU (Natural Language Understanding) se ocupa del desarrollo de algoritmos que les permitan a las computadoras comprender e interpretar nuestro lenguaje; como nos podemos imaginar, no parece ser algo

Por la magnitud de la empresa y con el fin de abordar sus objetivos específicos, la Lingüística Computacional se ha desarrollado históricamente en varias especialidades, con metáforas, teorías y metodologías diferentes, y cada una de estas ha representado un esfuerzo de investigación y desarrollo tecnológico de dimensiones colosales.

L. A. Pineda Cortés, 2017, p. 92.

sencillo y ha concentrado esfuerzos a lo largo de muchas décadas (Rong, 2014). Nuevamente nos encontramos con un término que llama a confusión, el de "comprensión". En este mundo técnico de la IA abundan las extrapolaciones y muchos científicos y filósofos rechazan la utilización de la noción de "comprensión" del lenguaje, algo que consideran propio de los seres humanos, pero no de las máquinas. El NLU tiene un campo de aplicación que incluye la clasificación de secuencias (por ejemplo, asignar una intención positiva o negativa a una secuencia de texto), etiquetado de secuencias (que permite asignar una etiqueta a las palabras en una secuencia de acuerdo al contexto), procesamiento de secuencia en secuencia (entra una cadena de texto como input y sale otro texto como output, no necesariamente de la misma extensión, como por ejemplo en las traducciones automáticas). Estas últimas aplicaciones texto a texto exceden el universo del NLU y se prolongan hacia el campo de la generación de texto a la que nos hemos acostumbrado a partir de ChatGPT y otros grandes modelos de lenguaje (Figura 1).

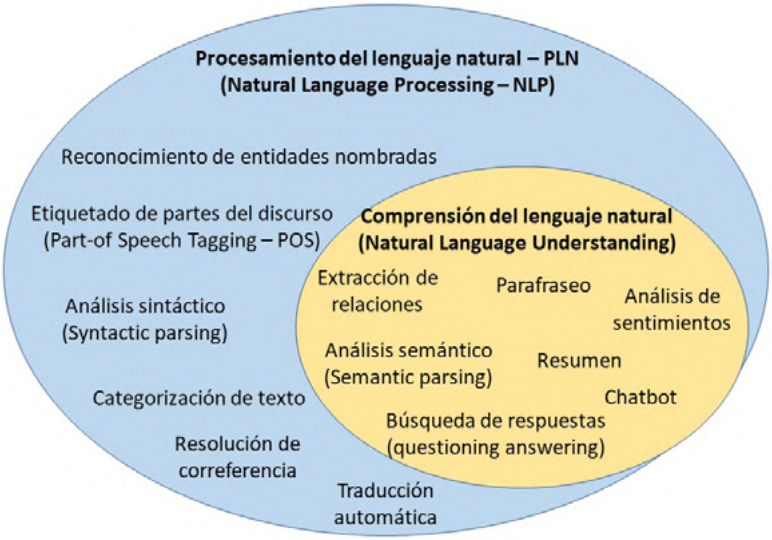


Figura 1. Víctor Vallejo. Procesamiento del lenguaje natural y Comprensión del lenguaje natural. 17 de abril de 2020. Wikipedia. https://es.wikipedia.org/wiki/Procesamiento_de_lenguajes_naturales#/media/Archivo:Procesamiento_del_lenguaje_natural_y_Comprensi%C3%B3n_del_lenguaje_natural.png

¿Qué dificultades encierra la tarea de lograr que las máquinas entiendan y generen el lenguaje natural humano? Como nos podremos imaginar muchas. En primer lugar, la ambigüedad, hay que tener en cuenta que las máquinas no entienden los matices que le damos al lenguaje en los contextos de comunicación. Nuestro mundo cotidiano está lleno de zonas grises de ambigüedad y las palabras están preparadas para significar sentidos muy diferentes a los de su literalidad. Para darnos una idea de la magnitud de las dificultades que representa la ambigüedad tengamos en cuenta que el tratado de Jurafsky y Martin de lingüística computacional incluye cientos de referencias al tema (Jurafsky y Martin, 2023). Muchas cosas pueden confundir a una máquina, los significados de las palabras, la atribución de un adjetivo o un complemento en la sintaxis de una oración, los signos de puntuación, el listado es enorme. Respecto de la gama de ambigüedades propias del lenguaje existen varias clasificaciones en las que no nos detendremos en detalle. Para darnos una idea sobre aquello a lo que nos referimos con el concepto de ambigüedad veamos dos ejemplos.

1. Ambigüedad léxica: una palabra puede adquirir diversos significados acordes al contexto en que se inserta. Si tomamos una frase como "la alianza es brillante", "alianza" puede referirse a un anillo de bodas o a un acuerdo conveniente.
2. Ambigüedad sintáctica: se presenta en casos en que una oración puede ser interpretada de diferentes formas. Si digo "compró los tomates baratos", no queda claro si se optó por comprar tomates baratos o si el precio resultaba conveniente. No sabemos a qué atribuir la calificación de barato, al tomate o al precio. Lo mismo nos ocurre con la frase "subió la lámpara por la escalera y se rompió". ¿Se rompió la lámpara o la escalera?

Los seres humanos estamos habituados a manejarnos con márgenes de ambigüedad altos y de manera constante, nuestra comunicación está plagada de sentidos subyacentes, segundas intenciones, giros, expresiones metafóricas, somos seres comunicacionalmente muy complejos. Lograr que una máquina adquiera algunas de las capacidades necesarias

para manejarse en un terreno tan resbaladizo implica un trabajo de diseño de algoritmos computacionales que ha retado la inteligencia de lingüistas e ingenieros.

La lucha por conquistar el lenguaje ha sido larga y se encuentra bien documentada, es por ello que no vamos a reproducirla en este libro. Los caminos que se intentaron fueron varios, por ejemplo, se procuró encontrar un conjunto de reglas formales que permitan generar lenguaje o se ha apelado a procedimientos estadísticos de recuentos de palabras, los intentos han variado hasta llegar a la solución actual de redes neuronales. Por más que se haya intentado, no ha sido posible encontrar una serie de reglas que permitan a una máquina entender el lenguaje humano de manera dúctil y menos aún, que puedan resolver íntegramente el tema de la ambigüedad por medio de algoritmos de restricción formal.

Uno de los temas a considerar es la consabida necesidad de convertir el texto en números para que lo pueda procesar la computadora. Una de las primeras opciones fue la técnica conocida como "bolsas de palabras" (bag of words), una estrategia conocida desde los años 70 (Aryal et al., 2019). Se trata de una estrategia simple que consiste en contar el número de veces en que una palabra está presente en un texto o secuencia de texto. El principal problema del enfoque es que por este medio el texto pierde su estructura interna. Puedo saber cuántas veces se menciona cada una de las palabras, lo cual indica ciertas cuestiones con relación al texto, pero no puedo avanzar mucho más allá de ello.

Un paso significativo en el PLN fue el de incorporar la herramienta de los vectores y de las matrices relacionadas con el álgebra lineal. Esto abre la puerta a un universo matemático muy desarrollado, permite jugar con un conjunto de posibilidades lógicas a otro nivel, el de los espacios vectoriales y las estructuras de grupos. No importa que comprendamos necesariamente de qué se tratan, pero sepamos que son herramientas muy poderosas que nos permiten pensar en cuestiones como espacios semánticos en los que conviven las palabras de una lengua y nos permiten calcular cuán relacionados se encuentran los términos entre sí. Estas similitudes se pueden cuantificar calculando cuán distantes están

Los lingüistas chomskyanos tienden a concentrarse en juicios categóricos sobre tipos muy raros de oraciones, los practicantes de PNL estadística están interesados en buenas descripciones de las asociaciones y preferencias que se dan en la totalidad del lenguaje a usar. De hecho, a menudo descubren que se puede obtener un buen rendimiento en el mundo real concentrándose en tipos comunes de oraciones.

C. Manning y H. Schutze, 1999, Fundations of Statistical Natural Language Processing, p. 7.

los términos entre sí en un espacio geométrico (a partir de herramientas como la trigonometría). Pensemos en la palabra casa, la palabra iguana y la palabra puerta ¿qué términos están más cercanos entre sí? ¿Cuáles tienen mayor posibilidad de aparecer juntos en una oración? A partir de concebir el lenguaje mediante esta representación espacial podemos encontrar zonas en las que conviven en proximidad palabras con sentidos cercanos. Mundos de sentido vinculados al deporte, a la literatura, a los insultos, a los colores, podemos programar todas las agrupaciones que consideramos necesarias, a esto se lo conoce como agrupamiento por tópicos. Cuando decimos que podemos percibir este espacio, pensar en algo así, estamos haciendo trampa, ya que nuestros cerebros no tienen la capacidad de representación espacial más allá de las tres dimensiones. Poder concebir en el plano abstracto no equivale a poder imaginar en el plano de la intuición espacial ("visualizar"). Existen algunos trucos visuales para representar alguna dimensión extra, pero rápidamente se torna imposible visualizar estos espacios. Para concebir un mapa semántico que corresponde a los términos de una lengua natural tenemos que imaginarnos una enorme cantidad de dimensiones.

A partir de esta idea de representar los términos como vectores se diseñaron muchas soluciones, por ejemplo, cruzar los textos con palabras, las palabras con palabras y definir cuán próximas se encontraban. Pero no solo eso, sino que se comenzó a trabajar con una idea que será clave aún hoy, la de ventana de contexto (Jurafsky y Martin, 2023). La idea de ventana de contexto permite detectar relaciones entre palabras que se encuentran en vecindad unas de las otras. El procesamiento de la información no tomará los términos o palabras de manera aislada, sino que los analizarán en conjuntos, estableciendo relaciones entre los mismos. El conjunto de términos que toma en cuenta el modelo se conoce como ventana de contexto. Estas ventanas pueden variar de tamaño, en un principio las ventanas de contexto eran pequeñas, calculaban vecindades cortas, dos palabras para la derecha y dos palabras para la izquierda, por ejemplo. A partir de estas vecindades en las ventanas de contexto las IAs pueden realizar poderosas inferencias sobre la distribución de probabilidades de aparición de los términos, organizar los sentidos de diversas maneras, en definitiva, pueden comprender mucho mejor el

significado de los textos. O, para expresarlo de manera correcta, pueden realizar inferencias que nos generan la impresión a los humanos de que entienden el sentido de los textos. Porque entender en el sentido humano no entienden nada. Al menos no lo hacen en el sentido en que lo hacemos nosotros. Repetiremos esto hasta el cansancio. Protéjanse de las notas periodísticas y de algunas comunicaciones de divulgación que difunden la idea de que las IAs comprenden.

El enfoque de las ventanas de contexto no es nuevo, pero su verdadera potencia se desencadenó a partir de las llamadas redes neuronales profundas de los grandes modelos generadores de lenguaje. En la actualidad modelos como los de Open AI o de Google tienen contextos de miles de palabras (o tokens idea que veremos en seguida). Claude la IA de Anthropic estaría por encima de los 100K de ventana (Anthropic/C, 2023). A mediados de 2023 GPT4 Turbo anunció que cuenta con una versión de 300 páginas de contexto (no aportamos fuentes dada la falta de transparencia informativa de la empresa). A nivel teórico en 2023 se han planteado algoritmos que permitan millones de palabras como ventana de contexto, algo que parecía distante. Este nivel se ha alcanzado: Gemini 1.5 cuenta con ventanas de contexto de 1 millón de tokens y, para ciertas operaciones y en algunas versiones escala hasta los 10 millones de tokens. Pero todavía no llegamos a este punto en nuestro desarrollo, vayamos de a poco y volvamos a los planteos de NLU.

¿Hasta dónde llegamos por medio de NLU (comprensión del lenguaje natural)?

En febrero de 2011 la IA Watson de IBM ganó uno de los programas de preguntas y respuestas más populares: Jeopardy! Compitió con ganadores de ediciones anteriores, con los mejores representantes humanos hasta la fecha. Para ello tuvo que escalar a un alto nivel de manejo de la sintaxis, la morfología, la gramática, el manejo de contexto, y no solo eso, también tuvo que articular esas capacidades con la ejecución de estrategias de juego (Ferrucci et al., 2010). Ante esta conquista el lingüista americano John Searle señaló que Watson ganó el concurso de preguntas

y respuestas sin haber entendido ni una de las preguntas que se le realizaron (Wall Street Journal, 23 de febrero de 2011). Podemos discutir hasta el infinito sobre el derecho a utilizar o no la noción de comprensión y no ponernos de acuerdo. Para hacerlo puede ser conveniente avanzar en el conocimiento de estas técnicas y las capacidades con las que cuentan las AI en general y en sus diversas arquitecturas. El desafío de Jeopardy! Implica manejar un muy amplio espectro de intereses humanos y por el otro lado poder interpretar la mecánica de un concurso de entretenimiento lo que implica un alto grado de ambigüedad en el manejo del lenguaje.

El NLU implica un avance en las técnicas de procesamiento del lenguaje natural respecto del NLP anterior. Ambos, NLP y NLU se enfocan en que las máquinas puedan interpretar y procesar el lenguaje natural. La diferencia reside en que la orientación del NLP apunta a la precisión literal del lenguaje en todas sus formas, en tanto que el NLU va más allá de la identificación y el agrupamiento de material lingüístico, intenta captar con mayor profundidad las intenciones humanas y penetrar en las regiones de ambigüedad. El natural language understanding (NLU) implica un salto de calidad que permite mejoras significativas en muchos campos de interacción humano – máquina como por ejemplo en los modelos de traducción automática, los asistentes virtuales (chatbots) y la identificación de sentimientos. Es importante valorar las dificultades que encierra esta tarea.

Hasta aquí avanzamos por la vía de comprender el lenguaje natural. En gran medida, logramos manejar cuestiones sintácticas, semánticas y contextuales. Un manejo de esta magnitud ha colaborado en la mejora de la traducción automática, algo que habremos notado a lo largo del tiempo en nuestro uso de traductores en línea. Por otra parte, el NLU permite mejoras drásticas en la captación de las intenciones del emisor, en particular en el campo del llamado "análisis de sentimiento" que permite, por ejemplo, calificar los contenidos de los textos como positivos o negativos. A partir de técnicas vinculadas al NLU en el manejo de interacciones pregunta respuesta se obtuvo incrementos en el rendimiento de los chatbots lo que permitió la explosión de los sistemas de asistencia automatizados. Sin embargo, en comparación con el estado actual de las mejoras de la IA

"When we hear language, we bring so much context to interpreting the question that we come up with sensible and reasonable answers. The computer struggles with that." D. Ferrucci.

N. Greenfieldboyce, 14 de febrero de 2011, On 'Jeopardy!' It's Man Vs. This Machine. https://www.npr.org/2011/02/14/133697585/on-jeopardy-its-man-vs-this-machine

"I mean, the computer has to find out, you know, where are the individual words and then how do the words group together," Ferrucci says. "What's the verb? What's the subject? What's the object? What's the preposition? What's the object of the preposition?"

N. Greenfieldboyce, 14 de febrero de 2011, On 'Jeopardy!' It's Man Vs. This Machine. https://www.npr.org/2011/02/14/133697585/on-jeopardy-its-man-vs-this-machine

estos avances nos parecen relativamente ineficaces. Lidiar con los chatbots de las empresas o con las traducciones automáticas que obteníamos años atrás, incluso los anteriores modelos de asistencia de Siri y Alexa resultaba relativamente decepcionante.

Aunque usemos la palabra "comprender" nos queda claro que por vía de estos algoritmos las máquinas no entienden el lenguaje en el sentido en que lo entendemos los seres humanos. Una cosa es entender y otra generar lenguaje. Entender una lengua no es lo mismo que producirla. Es frecuente escuchar decir a alguien "entiendo cuando me hablan, pero no puedo armar una oración en inglés". Claro, los dispositivos mentales involucrados en la comprensión no son los mismos involucrados en la ejecución. Producir algo nuevo, generar una oración, implica la adquisición de otro tipo de capacidades y nos introduce en la otra subrama del procesamiento del lenguaje natural, el de la generación de lenguaje (NLG).

El cambio de paradigma desde la comprensión hacia los modelos generativos de lenguaje vendrá de la mano del incremento en la capacidad de generalización de los aprendizajes. Generalizar el aprendizaje permitirá que el conocimiento obtenido a partir de los datos con que se "alimenta" a los modelos de IA les permita generar textos, sonidos o imágenes nuevas. Estos textos o imágenes pueden ser similares a los textos o imágenes con los que han aprendido, pero no son copias. De manera aproximativa, se puede decir que la IA aprende los patrones presentes en esos datos con los que se las entrena y a partir de ellos genera resultados que siguen la misma lógica. Por ejemplo, aprenden qué características tiene la imagen de un gato y si se lo requerimos correctamente nos entrega la imagen de un gato. No se tratará de que recupere y copie una imagen que recibió como input, sino que produzca una nueva imagen que corresponda a las características (*features*) que aprendió a reconocer en los gatos. Aprende cómo se organiza un cuento infantil o una nota periodística o un informe científico y si se lo pedimos correctamente nos entregará un cuento infantil, una nota o redactará un paper. No serán copias, serán resultados que se han generado a partir de su comprensión de los patrones que hacen a los cuentos, a las notas o a los papers. Tendrá manejo de los términos que utiliza cada contexto, el infantil, el periodístico o el de alguna ciencia en particular. Los diseños de los algoritmos que

se utilizan para dar una solución eficiente a la generación del lenguaje serán los de las redes neuronales profundas por eso vamos a repasar de qué se tratan en el apartado siguiente.

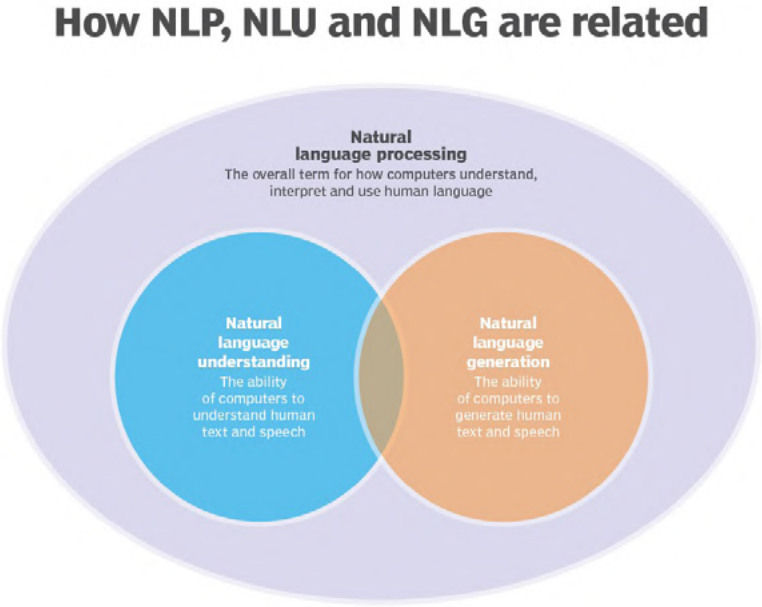


Figura 2. En el campo del procesamiento del lenguaje natural conviven dos subramas, la de la comprensión del lenguaje natural (NLU) y la de la generación del lenguaje natural (NLG). Pasar de la comprensión del lenguaje por parte de los modelos computacionales a la generación del mismo implica un cambio de paradigma hacia los enfoques de redes neuronales de aprendizaje profundo. Imagen extraída de: https://www.techtarget.com/searchenterpriseai/definition/natural-language-generation-NLG

Las redes neuronales y su componente básico la neurona artificial

Las redes neuronales (RN) representan el punto más alto de desarrollo alcanzado hasta la fecha en el intento de replicar las capacidades de la inteligencia humana a partir de artefactos computacionales. Han sido la meca para la lingüística computacional y abrieron la compuerta a la explosión de los llamados modelos generativos de lenguaje (NLG) Figura 2). De hecho, el texto no es el único campo en el que se desarrolla

este paradigma de redes neuronales, también ha tenido impacto en campos como el del procesamiento de imágenes, en la robótica y en toda una gama de desarrollos en los cuales el llamado aprendizaje profundo constituye una ventaja. Su gran impacto social se ha hecho sentir a partir de la masificación de uso de los grandes modelos generativos de lenguaje cuyo exponente clave ha sido ChatGPT a disposición del público global a fines de 2022. Es por ello que estamos hablando de estos temas con tal nivel de atención, aunque la incidencia de modelos anteriores de IAs con procesamiento de lenguaje ya estaba haciendo sentir sus efectos. De hecho, las ciencias sociales han mostrado reflejos lentos ante el dominio del lenguaje por parte de las disciplinas computacionales lo cual las coloca en deuda dado el interés superlativo que le asignan al fenómeno en el plano semiótico y lingüístico en muchas de sus formas.

Las redes neuronales están conformadas por una familia de algoritmos computacionales. Entendamos esto. Insisto, son una familia de algoritmos. Como primera tarea nos vamos a sacar de la cabeza las imágenes de las neuronas. ¿Entendido?: son una familia de algoritmos. Ya repasamos lo que era un algoritmo en capítulos anteriores. Por si acaso les recordamos que un algoritmo es un conjunto de instrucciones organizadas lógicamente en una secuencia de pasos. Queremos lograr algo y para ellos vamos a ejecutar un algoritmo que nos permite tomar datos de entrada (input) y ejecutando los pasos necesarios pretendemos obtener un determinado tipo de resultado (output) (Louridas, 2020). Si pensamos un segundo en ello habremos avanzado bastante. Uno de los principales obstáculos para nuestra comprensión reside en la cantidad de imágenes cautivadoras que rodean el tema de la IA. Detrás de las figuritas, de los dibujos de las redes neuronales, existen conceptos e ideas (Figura 3). Cuando miremos las ilustraciones pensemos que se trata de gráficos que intentan esquematizar las soluciones a determinados problemas. ¿Cuáles son las dos principales cuestiones que tratan de solucionar las redes neuronales artificiales? Los problemas del aprendizaje de las máquinas en primer lugar y en términos generales, la generación de agentes que puedan replicar acciones que necesitan de capacidades similares a la de la inteligencia humana para ser ejecutadas. De todas las familias

de algoritmos que se han desarrollado hasta la fecha las RN son las más potentes (aunque no necesariamente las más eficientes para todos los problemas).

Las redes neuronales siguen respondiendo al esquema input – output, esto sigue operando como en otros modelos de IA. Las redes neuronales tienen que entrenarse tal y como vimos en el capítulo sobre aprendizaje máquina. Las redes neuronales aprenden sin la necesidad de que se etiqueten los datos, son capaces de hacerlo en forma autónoma y a partir de datos no estructurados. En este sentido superan a enfoques y diseños de machine learning anteriores en términos de su capacidad de "lidiar" con datos con baja estructura interna.

Cuando abordamos el cognitivismo y las neurociencias nos encontramos con dos ideas, la de funciones superiores y la de redes de neuronas. Ambos conceptos han influenciado el pensamiento de los ingenieros de software en el diseño de los modelos. La biología es una fuente de inspiración para la ingeniería. Una de las nociones que ha servido de analogía a las ciencias computacionales es la de coordinación de facultades superiores. Según esta forma de concebir el funcionamiento del sistema nervioso lo más importante es pensar en la colaboración de las facultades localizadas en diferentes regiones cerebrales. Lo principal es el conjunto antes que las facultades aisladas. El otro concepto potente que sirve de inspiración a los ingenieros es el de neurona y, más específicamente el de red neuronal. La neuropsicología postula que la clave para comenzar a comprender las facultades humanas está en la conexión entre las neuronas en forma de complejas redes. Las neuronas pueden ser consideradas como elementos básicos. No debemos tomarlas como átomos aislados sino pensarlas en conexión. Allí reside su potencia, en ese universo de millones de conexiones por las que fluye energía se activan complejos procesos que permiten que nos manejemos en nuestro medio ambiente. Las conexiones se forman, la energía fluye, las neuronas se activan y es por ello que podemos ejecutar funciones superiores. En determinado estadio la vida se desarrolló y emergió en la materia la gran aventura de lo neuronal. Nos internamos en este mundo y para ello les proponemos tres pasos: entender qué es una neurona como unidad básica de procesamiento de información,

luego veremos cómo se agrupan formando redes y por último veremos cómo esto les permite aprender.

Tal vez nos convenga olvidar por un momento esta denominación de "neurona", un término entre los muchos que enturbian nuestra comprensión de estas máquinas. Su primer nombre fue "perceptrón simple" y fue diseñada por un ingeniero llamado Frank Rosemblat en 1957 (Russell y Norvig, 2004). Olvidemos los cerebros y pensemos en funciones. Una neurona es una función matemática, o si se quiere un conjunto de funciones matemáticas, que puede ser programada mediante un algoritmo. Para representarnos de qué hablamos pensemos en que no existe en ningún lugar un objeto físico que tenga la forma de la neurona. A continuación, incluimos una implementación de un perceptrón simple en un programa muy utlilizado llamado Python (Figura 4). No hay cuerpo neuronal, no hay axones, no hay dendritas, ni membrana celular, no entran a jugar las proteínas en estos algoritmos. No existen reacciones químicas como la de la bomba de sodio potasio, ni canales materiales de entrada y salida de sustancias activadores que conecten la neurona física con su ambiente. No existe nada de ello en el mundo material, es una simple instrucción de programación. Mejor aún, avancemos y pensemos que se trata de algoritmos, es decir de una serie de pasos que sirven para

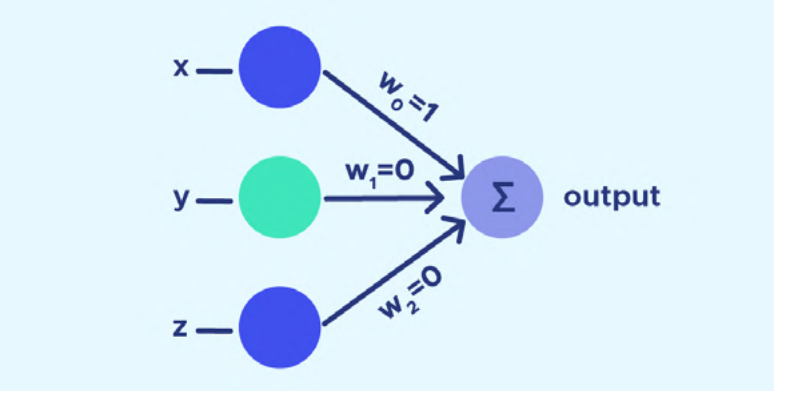


Figura 3. El perceptrón simple fue ideado por Frank Rosemblat en 1957. Imita una neurona biológica y es el elemento base del aprendizaje en una red neuronal. Neurona artificial es un nombre que se ha elegido para nombrar una función matemática. https://datascientest.com/es/perceptron-que-es-y-para-que-sirve

ejecutar varias funciones que se ensamblan. Cada paso (función) permitirá que nuestro modelo (el perceptrón simple o neurona artificial) pueda realizar determinados procesos. Todo esto conduce a que la neurona artificial reciba inputs y entregue un output. Entre el input y el output realizará un trabajo de procesamiento. Para que realice correctamente las tareas para las que la hemos destinado deberemos entrenarla, tal y como vimos en un capítulo anterior, las "máquinas inteligentes" deben aprender si queremos que realicen algún tipo de tarea. A este aprendizaje se lo conoce como entrenamiento.

```
class SimplePerceptron:
    learn_rate = 0.1

    def __init__(self):
        self.weights = None

    def logistic_function(self, x: float) -> float:
        """
        Logistic function, used as the activation function
        """
        return 1. / (1 + np.exp(-x))

    def forward_pass(self, X: np.ndarray) -> float:
        """
        Prediction of a single data point, given the current weights
        """
        weighted_sum = np.dot(X, self.weights)
        output = self.sigmoid_function(weighted_sum)

        return output

    def fit(self, X_train: np.ndarray, y_train: np.ndarray, n_epochs: int = 20):
        """
        Training using Batch Gradient Descent.
        Weights array is updated after each epoch
        """
        self.weights = np.random.uniform(-1, 1, X_train.shape[1])
        current_weights = self.weights.copy()
        for epoch in range(n_epochs):
            for x, y in zip(X_train, y_train):
                y_predicted = self.forward_pass(x)
                current_weights -= self.learn_rate * (y_predicted - y) * y_predicted *
(1 - y_predicted) * x
            self.weights = current_weights.copy()

    def predict(self, X_test: np.ndarray) -> np.ndarray:
        """
        Predict label for unseen data
        """
        return np.array([self.forward_pass(x) for x in X_test])
```

Figura 4. Acá les dejamos el código en Phyton para el perceptrón simple. Como podemos ver estamos muy lejos del objeto material y biológico al que llamamos neurona. Lejos incluso de la clásica ilustración del perceptrón simple, cuyo parecido con la neurona biológica excita la imaginación y colabora a que sintamos que la inteligencia orgánica y la de los artefactos son más próximas que lo que en realidad son. Si quieren más detalle les dejamos el link del que copiamos código. Imagen extraída de: https://blog.damavis.com/el-perceptron-simple-implementacion-en-python/

Las redes neuronales profundas

El planteo de las redes neuronales es un avance enorme, pero encuentran limitaciones de todo tipo. A partir de los modelos de IA constituidos por un solo perceptrón (neurona) se pueden resolver muchos problemas, pero muchas cuestiones quedan fuera de sus posibilidades. Estas limitaciones ya se conocían hacia fines de la década de los sesenta a partir de la crítica realizada por Minsky y Papert en un texto llamado "Perceptrones. Una introducción a la geometría computacional." (Minsky y Papert, 1969).

Este trabajo tuvo influencia negativa sobre el enfoque, y se lo considera uno de los responsables del abandono de la exploración de redes neuronales durante décadas. Los límites que señalaba el trabajo eran dos, uno demostraba que los perceptrones simples tenían limitaciones en su razonamiento lógico lo que les impedía realizar algunas operaciones fundamentales (función lógica XOR). Esta limitación desaparecía cuando se sumaban más neuronas artificiales en el procesamiento de los datos. Lo que no lograba hacer el perceptrón simple se podía lograr con el simple procedimiento de conectar más neuronas. Sin embargo, Minsky señala que tales diseños necesitaban una capacidad de procesamiento inconcebible para los desarrollos de hardware de la época. Nunca perdamos de vista que la infraestructura, la base material, manda. Tampoco se contaba con una solución adecuada en término de software ya que los algoritmos con que se podía operar en ese momento elevaban la cantidad de operaciones de manera dramática. Según su opinión inicial la IA basada en redes era un camino sin salida. Esto orientó la indagación hacia otro tipo de enfoques por ejemplo el simbólico impulsado por Chomsky.

Pero la industria del hardware ha sido activa y la capacidad de cómputo se acelera a grandes pasos. Existen pautas de duplicación de la capacidad de cómputo según la llamada ley de Moore, una anticipación enunciada en 1965 (Moore, 1998). Se trata de una predicción más que de una ley en el sentido estricto, no obstante, describe con bastante ajuste el crecimiento que ha tenido la capacidad de cómputo de la humanidad a partir de la fabricación de transistores. Los componentes básicos de los chips son los transistores que le permiten procesar la información. Desde los 6 transistores que contenían cuando fueron patentados inicialmente llegamos a la época actual en que llegan a contar con millones de ellos. La duplicación de la capacidad de cómputo seguía una escala logarítmica según la cual se duplica cada dos años. En la actualidad esta predicción tiende a ser reemplazada por la llamada Ley de Huang, (no es una ley sino una estimación), que presagia un incremento en cómputo con ritmos mucho más acelerados provenientes no solo ya de factores físicos como la miniaturización o la implementación de varios núcleos (los cores de las computadoras) sino de otras fuentes que incrementan la eficiencia de los procesos (Huang, 2023, YouTube NVIDIA). Todo esto es terreno de debate y nos desvía del

tema. Volviendo al tema del crecimiento de la capacidad de cómputo, una limitación que nos señalaba Minsky, la aceleración de la ley de Moore tardó algunas décadas, pero al fin se alcanzó la suficiente potencia como para permitir la operación de las redes neuronales.

Un aprendizaje más eficiente: algoritmo *backpropagation*

Llegar hasta el punto actual en que los grandes modelos de lenguaje interactúan con humanos a partir de simples instrucciones en lenguaje natural implicó recorrer un largo camino. Hacia fines de la década de los 80 la capacidad de cómputo se había incrementado, los modelos de redes neuronales multicapa ya eran conocidos, muchas de las piezas ya estaban en el tablero. Pero aún existía una limitación: no se contaba con el algoritmo adecuado como para entrenarlas. No se podía terminar de liberar la potencia de las redes neuronales, entonces en 1986 llegó el algoritmo de backpropagation (propagación para atrás del error) (Hinton, 2022).

¿Qué aportó el algoritmo de backpropagation? Una solución eficiente para que las redes neuronales pudieran entrenarse sin supervisión humana a partir de datos no estructurados. Las redes neuronales podrían generar una representación interna de los datos que se le suministraban y estaban procesando. Durante todo el libro insistimos en la idea de patrones. A partir de este algoritmo de propagación para atrás del error, las redes neuronales de muchas capas podían aprender a generar una representación interna de estos patrones de manera muy eficiente.

Cuando presentamos el tema del aprendizaje nos referimos a la idea de "ajustar" los modelos y este es el punto en que aporta esta propuesta. Los algoritmos de backpropagation permiten que el modelo de red neuronal se ajuste a sí mismo de manera muy eficaz. No vamos a entrar en las complejidades matemáticas del modelo en torno a cuestiones tales como el descenso de gradiente, o la interrelación con la minimización de la función de coste, pero nos alcanza con saber que este algoritmo reduce drásticamente la cantidad de cálculos necesarios para lograr un ajuste eficaz cuando se lo alimenta con grandes masas de datos no

estructurados (Hinton, 2022). Esta idea nos es familiar: cuando tenemos un modelo de IA sin entrenar, los primeros resultados que arroje serán malos, la máquina es ignorante, boba. Solo a partir de un aprendizaje al que llamamos ajuste de parámetros logrará entregarnos predicciones aceptables. Cada modelo de IA encierra sus complejidades y algunos requieren muchos cálculos para entrenarlos. La publicación citada de Misky Rumelhart, Hinton y Williams en 1986 fue uno de los grandes momentos de quiebre en el camino de conquista del lenguaje por parte de las computadoras.

Las ventajas de compactar la información

Ya casi llegamos al punto en que los grandes modelos de lenguaje cuentan con todos los elementos para funcionar al nivel en que lo hacen a la fecha. Nos faltan algunos algoritmos clave, el primero de los cuales tiene que ver con algo llamado "compactar" o reducir la dimensión de los datos. Nuestro lenguaje es algo enormemente complejo. El número de elementos con que se conforma una lengua natural, esas que utilizamos las comunidades en nuestro mundo cotidiano, es difícil de manejar en términos computacionales. Los sentidos que podemos producir a partir de esos elementos semánticos son potencialmente infinitos. Es obvio que aquí hay problemas que solucionar.

Si queremos entender a qué se refieren las dimensiones sin entrar en el álgebra lineal, la base profunda de la cuestión (Grossman, 1996) podemos pensar en la diferencia que existe entre un punto, una línea, un plano (una hoja de papel, por ejemplo) o la habitación en que nos encontramos. El espacio que habitamos y estamos acostumbrados a representar tiene tres dimensiones, alto, largo y ancho, (nuestros conocidos x,y,z). Una hoja de papel, el plano, tiene dos dimensiones, la línea tiene una dimensión y el punto, se dice, no tiene dimensión. Esto es lo que nos permite nuestro sistema de percepción biológica, así nos acostumbramos a representar "el mundo real". Pero podemos imaginar o trabajar a partir de espacios mucho más grandes, con muchas más dimensiones, con infinitas dimensiones.

El problema es que no podemos representarlos de manera gráfica, no podemos imaginarlos intuitivamente. Cada característica que queremos representar para un modelo de IA puede ser pensado como una dimensión.

Salgamos un momento del lenguaje y pensemos en que la IA tiene que clasificar imágenes lo que nos resulta más fácil de imaginar. Como bien sabemos, los rostros se diferencian unos de otros a partir de muchas características. Piensen en 100 personas que conozcan y tendrán 100 rostros diferentes. Las diferencias que podemos encontrar entre los rostros son cuantiosas. los humanos somos muy hábiles en ese rubro. Una solución sería etiquetar esas diferencias e indicarle a la IA cuáles son los patrones que tiene que tener en cuenta. Con las redes neuronales no queremos hacer algo así, queremos que aprendan a clasificarlas de manera autónoma. Pero ¡atención!: para nosotros los humanos es evidente que un rostro es un rostro, nuestra captación intuitiva se dispara de manera inmediata (aunque los procesos neuro cognitivos no sean tan simples como parecen). Para nosotros un rostro es un rostro, para las máquinas no. Recordemos que debemos "traducir" la información natural para que pueda ser "digerida" por una máquina. Cuando ingresamos un input de imagen, se descompone la imagen en pixeles, a cada uno de los cuales se le asigna un número que representa el rango cromático. Nuestra imagen de rostro se ha convertido en un input numérico (un vector o una matriz numérica). Pero aquí surge otra cuestión: tenemos muchos números, mucha información y no toda es útil para que la IA aprenda a reconocer rostros, hay mucho "ruido" en la información. Necesitamos simplificar y con ese propósito se recurre a una estrategia fantástica inspirada en el álgebra lineal: la compactación. Para ello vamos a utilizar un sistema multicapa en el cual cada capa de neuronas va a tener menos elementos. Vamos a crear una especie de embudo de capas en las que cada una de las capas de neuronas recibirá el procesamiento de la capa anterior y la pasará a la siguiente (Figura 5). Como cada capa es más pequeña los datos se irán compactando hasta llegar al resultado que queremos obtener. Al final del proceso de compactación se han comprimido los datos y se obtiene el resultado de una predicción (output) de dimensión mucho más pequeña (identifica un animal, predice de qué número escrito a mano se trata, si es un auto u otro vehículo, etc.).

Volver

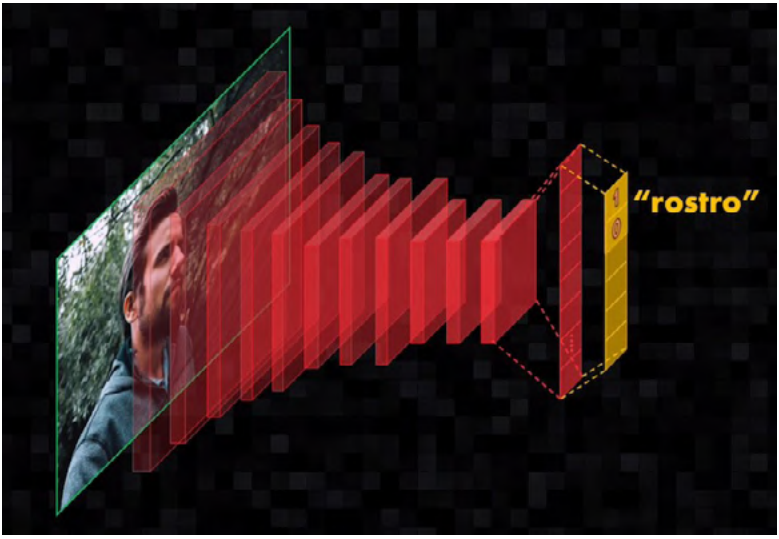


Figura 5. Una imagen digital está compuesta por una enorme cantidad de pixeles. La red neuronal compactará esta dimensión tan grande hasta llegar a la dimensión que nos resulte conveniente. Compactar la información es imprescindible en el manejo de estas tecnologías. Imagen extraída de: https://www.codificandobits.com/blog/deteccion-de-rostros-machine-learning/

Las palabras con sentidos similares frecuentan los mismos barrios

Esta idea de simplificar el número de variables a considerar y de compactar la información mediante el uso de algoritmos para redes neuronales está presente en muchos dominios no solo el de las imágenes. Recordemos que a nosotros nos interesa la conquista del lenguaje y para esto vamos a regresar a la década de 1950, a las hipótesis distribucionales de Harris y Firth, (Harris, 1954), (Firth, 1962, [1957]). Se trata de una hipótesis muy sencilla y potente: las palabras similares aparecen en contextos similares. Para esta hipótesis el contexto de uso es clave. Es por eso que Firth se apoya explícitamente en la noción de uso del lenguaje de Wittgestein para quién *el significado de las palabras reside en su uso* (Firth, 1962). La proximidad de las palabras en un contexto es un

Words that occur in similar contexts tend to have similar meanings. This link between similarity in how words are distributed and similarity in what they mean is called the distributional hypothesis. The hypothesis was first formulated in the 1950s by linguists like Joos (1950), Harris (1954), and Firth (1957), who noticed that words which are synonyms (like oculist and eye-doctor) tended to occur in the same environment (e.g., near words like eye or examined) with the amount of meaning difference between two words "corresponding roughly to the amount of difference in their environments" (Harris, 1954, 157).

D. Jurafsky y J. Martin, 2023, p. 103.

indicador de su relación semántica. Si visitan la misma zona, si frecuentan los mismos "barrios" dentro del territorio del lenguaje muy posiblemente mantengan relaciones entre sí. Así, botines, estarán más cerca en significado de pelota que de dinosaurio o de tenedor. Ambos, los botines y las pelotas visitan la zona lingüística del deporte, ese es su contexto de uso más frecuente.

> *As Wittgenstein says, the meaning of words lies in their use*
> (Firth, 1962, p. 11)

A partir de estas hipótesis distribucionales y con el objetivo de poder representar las complejidades que encierran las relaciones semánticas con un costo computacional razonable, se desarrolla el algoritmo Word2vec (Rong, 2014). Comencemos por el principio y volvamos a recordar que las máquinas no aceptan el lenguaje natural de manera directa, sólo nos podemos comunicar con ellas si convertimos los términos en cifras. Esto genera muchos problemas técnicos, no podemos asignarle a cada palabra del idioma un número porque las palabras son muchas y las operaciones matemáticas deberían manejar números gigantescos. El objetivo de toda esta tarea es lograr que una red neuronal pueda construir un mapa semántico del lenguaje. Si queremos que una red neuronal pueda manejar el idioma tenemos que lograr que la representación del mismo no sea colosal, para eso se acude a la compactación. Tomemos el caso del idioma español. En la actualidad el español tiene unas 93.000 palabras (RAE, 2023). No resultaría muy práctico que nuestra representación de cada palabra tuviera 93.000 dimensiones. Tengamos presente que nuestra red debería hacer cálculos monstruosos que desafiarían su capacidad de cómputo. Tendremos, entonces, que lograr que nuestro modelo de lenguaje aprenda a compactar esta información. A este conjunto de técnicas se las llama "word embeddings" (Jurafsky y Martin, 2023). Alimentamos nuestro modelo de lenguaje con las 93.000 palabras y las "hacemos pasar" por un proceso hasta llegar a una capa de 300 neuronas. El resultado que obtendremos es un espacio "más denso" de 300 dimensiones en lugar de uno muy disperso de 93.000 dimensiones. Este espacio semántico estará organizado en subespacios compactos en que "vivirán" las distintas palabras. La magia

se produce porque usamos una forma de organización propia de una rama de la matemática que aparece en varias partes del libro: el álgebra lineal o como también se la conoce algebra vectorial. Estos sub espacios, las zonas o barrios a los que nos referimos, contendrán palabras cuyo sentido está conectado desde el punto de vista semántico, pero también contemplará relaciones sintácticas.

Sin dudas es aquí donde se puede ubicar la gran revolución, en el trabajo publicado por Mikolov en 2013 (Mikolov et al., 2013). El desarrollo proviene de un equipo de ingenieros de Google que proponen este algoritmo que permite a las redes neuronales encontrar asociaciones de palabras en grandes corpus de texto. Les recomendamos firmemente que no se pierdan de visitar el enlace de Word2vec que incluimos en este capítulo, ya que les permitirá visualizar esa constelación compactada de palabras, lo que les ayudará a entender mejor la idea de modelos computacionales de lenguaje natural. Decíamos que estos sub espacios, nuestros "barrios" eran densos, pero ¿qué queremos decir con "densidad"? Nuestro lenguaje puede ser entendido como un delta en el que los grandes ríos, los canales y los arroyos de sentidos se entrecruzan, desde una palabra puedo llegar a otra palabra por diversos caminos. Desde fútbol a dolor de panza puedo llegar por el camino de la palabra pelota. Tengo una *pelota* para jugar al fútbol / se me formó una *pelota* en la panza porque los ravioles me cayeron pesados. Los barrios están conectados porque en el lenguaje natural cada palabra puede visitar diferentes contextos de uso, tal como nos decía Wittgenstein (Wittgenstein, 1988). En el ejemplo *"pelota"* habita en la cancha de fútbol y también en nuestros estómagos y el mapa semántico de una máquina deberá registrarlo si pretende manejar el lenguaje natural. Compactamos nuestro vocabulario desde 93.000 términos a 300 dimensiones, está muy comprimido. El resultado tendrá 300 dimensiones, piensen que su representación está dada por lo que se llama un vector (Figura 6). Un vector es una tira de números que en este caso tiene 300 valores pero que mantiene muchísima, pero no toda, la información semántica contenida en un lenguaje como el español de 93.000 palabras.

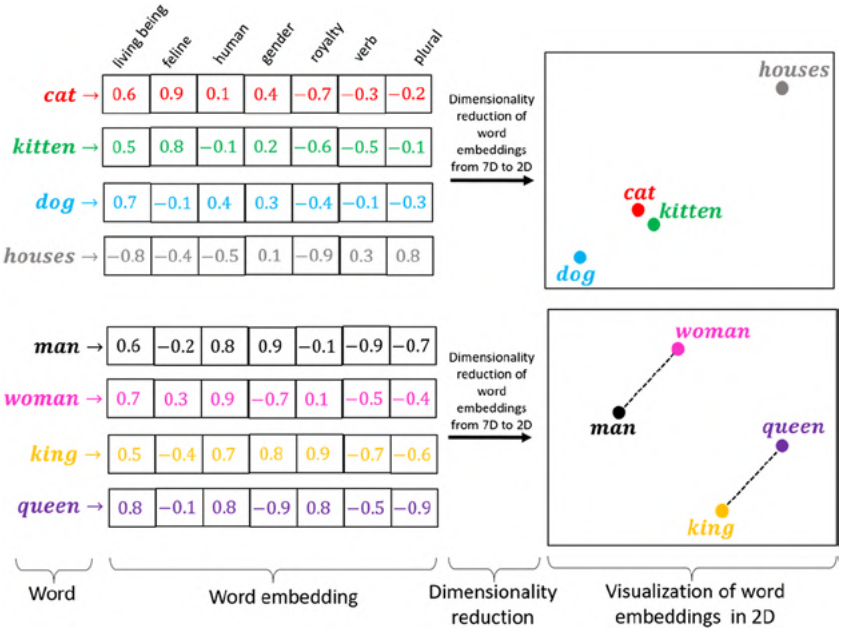


Figura 6. Les presentamos un ejemplo para que visualicen las operaciones de compactación. En el gráfico podemos seguir una compactación desde vectores de 7 dimensiones a vectores de 2 dimensiones. Las filas de números (embedding) expresan la codificación numérica en vectores (tiras de números) de las dimensiones de cuatro palabras hombre, mujer, rey, reina. En el cuadro final podemos ver cómo se visualizan las distancias entre los cuatro términos según dos dimensiones: hombre y mujer pertenecen al eje de género, rey y reina al de jerarquía nobiliaria. La posición tiene que ver con la cercanía de los puntos y puede ser calculada de manera matemática (sumas y restas entre los vectores).Imagen extraída de: https://swatimeena989.medium.com/training-word2vec-using-gensim-14433890e8e4

Los invitamos a explorar un mapa semántico compactado por Word2vec.

Word2vec quiere decir de palabras a vectores. Word (palabra), 2 (to= a), Vec (vector).

Hagan control click en el enlace y entrarán en el mapa semántico compactado (les recomendamos enfáticamente que no se lo pierdan).

https://projector.tensorflow.org/

Ya estamos avanzados, entramos en la explosión de las redes neuronales de la última década, una tecnología que tal vez produzca un cambio en los modos de producción y genere la próxima revolución industrial. Puede que cambie nuestras vidas y las de las sociedades en que vivimos. Quizás sean la causa de su transformación en un sentido positivo, aunque cabe la posibilidad de que se convierta en la causa de su colapso final, ya se verá.

Existen diversos tipos de redes neuronales, cada una de ellas adaptada a distintos tipos de tareas y, lo que es fundamental, preparadas para procesar diversos tipos de datos (Figura 7). No olvidemos que la percepción es algo crucial en términos de generar agentes con IA que puedan realizar tareas que requieran de facultades similares a la inteligencia humana. Para realizar esas acciones debe captarse el mundo y eso se logra a partir de recibir inputs en forma de datos y a través de distintos tipos de sensores. Tendremos entonces diferentes arquitecturas, por ejemplo, si queremos lidiar con imágenes se podrán usar las llamadas redes convolucionales, si se quiere procesar textos contamos con las redes neuronales secuenciales, que hoy han sido superadas en potencia por los llamados modelos Transformers, para la generación de imagen a partir de inputs textuales contamos con los modelos de difusión, los enfoques y propuestas son cada vez más potentes y diversos. Los modelos Transformers serán nuestra próxima parada cuando nos sumerjamos en el siguiente capítulo en este camino hacia la conquista del lenguaje natural por parte de las "máquinas que piensan" ("piensan" según el imaginario popular, nosotros preferimos decir que calculan).

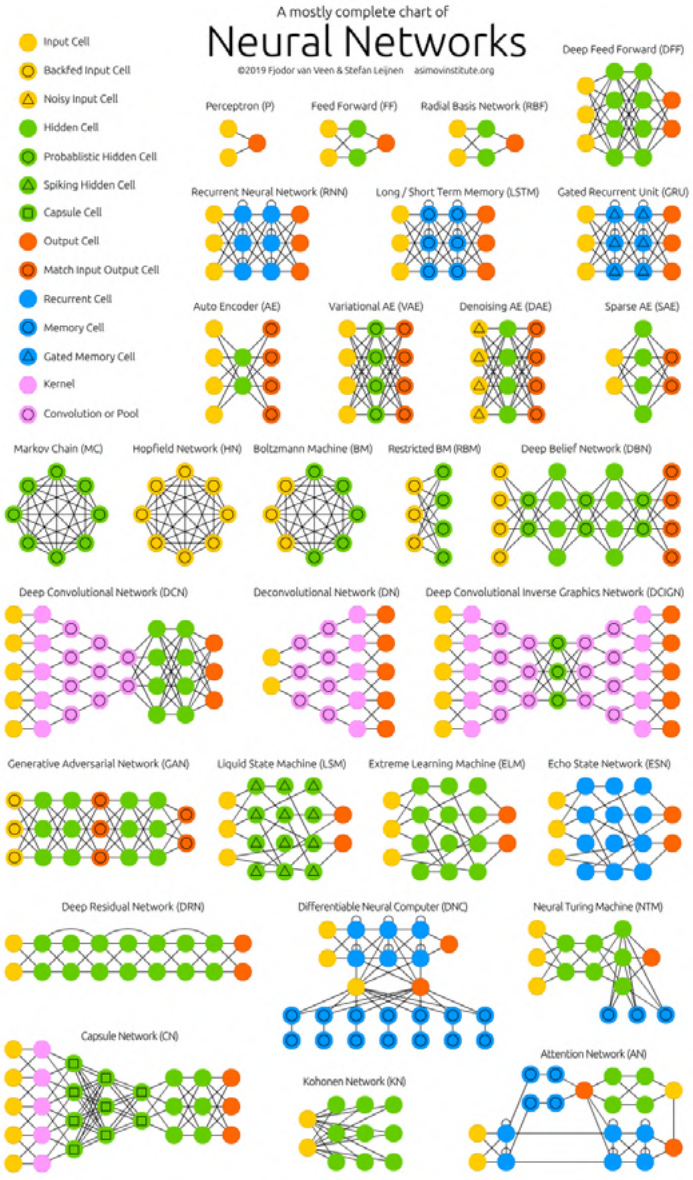


Figura 7. Distintos modelos de redes neuronales en 2019. Cada una de ellas presenta diversos enfoques y ha resultado eficaz para la resolución de distintas tareas. https://bookdown.org/keilor_rojas/CienciaDatos/redes-neuronales-artificiales-y-aprendizaje-profundo.html

Volver

Bibliografía

Aryal, S., K. M. Ting, T. Washio y G. Haffari (2019). A new simple and effective measure for bag-of-word inter-document similarity measurement. ArXiv, abs/1902.03402.

Diccionario de la Real Academia Española (RAE). Actualización 2023. https://dle.rae.es/contenido/actualizaci%C3%B3n-2023

Grossman, S. (1996). *Algebra lineal*. Mc Graw Hill.

Harris, Z. S. (1954). Distributional Structure. *Word, 10*(2-3), 146-162. https://doi.org/10.1080/00437956.1954.11659520

Hinton, G. E. (2022). The Forward-Forward Algorithm: Some Preliminary Investigations. ArXiv, abs/2212.13345.

Ferrucci, D., E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer y C. Welty (2010). Building Watson: An Overview of the DeepQA Project. *AI Magazine, 31*(3), 59-79. https://doi.org/10.1609/aimag.v31i3.2303

Greenfielboyce, N. (2011). On 'Jeopardy!' It's Man Vs. This Machine. *NRP Ciencia.* www.npr.org/2011/02/14/133697585on-jeopardy-its-man-vs-this-machine

Jurafsky, D. y J. H. Martin (2023). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3ra ed. Stanford University.

Louridas, P. (2020). *Algorithms. MIT Press Essential Knowledge Series*. The MIT Press.

Manning, C. y H. Schutze (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press.

Mikolov, T., K. Chen, G. S. Corrado y J. Dean (2013). Efficient Estimation of Word Representations in Vector Space. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1301.3781

Minsky, M. y S. Papert (1969). *Perceptrons: An Introduction to Computational Geometry*. The MIT Press .

Moore, G. E. (1998). Cramming More Components Onto Integrated Circuits. *Proceedings of the IEEE, 86*, 82-85.

Pineda Cortés, L. (2017). *La Computación en México por especialidades académicas Coordinador general*. Academia Mexicana de Computación.

Russell, S. J. y P. Norvig (2004). *Inteligencia artificial. Un enfoque moderno*. Pearson Educación, S.A.

Searle, J. (23 de febrero de 2011). Watson Doesn't Know It Won on 'Jeopardy!'. *Wall Street Journal*. https://www.wsj.com/articles/SB10001424052748703407304576154313126987674

Wittgenstein, L. (1988). *Investigaciones filosóficas*. Crítica-UNAM.

Rong, X. (2014). word2vec Parameter Learning Explained. ArXiv, abs/1411.2738.

Blogs / Youtube

Blog Anthropic/Claude (2023)
https://www.anthropic.com/product?ref=soloprogramadores.com

GTC 2023 Keynote with NVIDIA CEO Jensen Huang. NVIDIA CHANNEL.
https://www.youtube.com/watch?v=DiGB5uAYKAg

Watson and the Jeopardy! Challenge. 6 de noviembre 2013

IBM Research
https://www.youtube.com/watch?v=P18EdAKuC1U

Ferruchi, D. (2010) AI - In and Out of Jeopardy: David Ferrucci at TEDxBinghamtonUniversity
https://www.youtube.com/watch?v=bUT5395tLSc

Capítulo 7. Los grandes modelos generativos de lenguaje de las redes *Transformers*

Sumario

Durante este capítulo nos internaremos en las redes neuronales de desarrollo profundo y su manejo del lenguaje natural. Dos capacidades de la mente humana cobrarán importancia: la memoria y la atención. Nuestra primera parada será en los modelos de redes recurrentes con manejo de memoria de corto y largo plazo (LSTM). Estos modelos utilizan la memoria para detectar las relaciones entre las palabras en secuencias de texto. Su éxito se debe a la capacidad de retener relaciones a lo largo del tiempo. Nuestra segunda tarea será conocer a los grandes protagonistas de la explosión actual de las IA: los modelos Transformer. Veremos cómo cambian las reglas de juego a partir de mecanismos de atención y como ganan potencia a partir del llamado procesamiento paralelo.

Llegamos a fines de diciembre de 2023 y ya cumplimos nuestro objetivo de testimoniar una fracción del desarrollo de las IA en sus comienzos. Evitamos la tentación de incluir aquello que se ubica a principios de 2024. Todo se va a disparar y lo que hoy relatamos habrá quedado en el pasado, ya se perfilan nuevos enfoques como MAMBA, hace días han lanzado modelos multimodales como Gemini, recibimos la noticia de modelos de video, ya se verá hacia dónde vamos, pero eso es parte de otra historia o de otro libro...

Las redes neuronales regresivas con memoria a corto y largo plazo (LSTM)

¡Perfecto, como pudimos ver en el capítulo anterior ya tenemos nuestros mapas semánticos compactados con las técnicas word2vec! Avanzamos un largo camino en la conquista del lenguaje natural, pero seguimos teniendo problemas sin resolver. Con la agrupación por sub espacios o regiones en que las palabras con significados conectados están más cercanas no nos alcanza. ¿Por qué? Aquí tenemos que hacer participar a otro de los grandes tesoros con que cuentan los cerebros de las especies más desarrolladas biológicamente y en particular los seres humanos: la memoria. Para acceder a la IA hay que encontrar soluciones a la cuestión de replicar la memoria humana. Cuánto mejores sean los algoritmos que se destinan a esta tarea más cerca estaremos de conquistar el manejo del lenguaje natural y lograr el objetivo de que las máquinas puedan interactuar con nosotros por este medio.

Vayamos un poco para atrás y pensemos en qué implica la IA: en generar agentes que cumplan tareas que replican acciones que puedan ser consideradas inteligentes. Para cumplir una tarea un agente debe ser capaz de realizar lo que se conoce en neuropsicología como "acción ejecutiva". El agente ejecutará una acción, pero para ello debe focalizarse en ella, en cierta medida comprender cuál es la tarea y aquí entra a jugar otra función del cerebro humano: la atención. Pero las cosas no terminan aquí, porque para focalizarse y decidirse a realizar una acción debe a su vez tener memoria, recordar el sentido del mundo que lo rodea.

Pensemos en un ejemplo: queremos definir la compra de una heladera y nos ponemos a leer reseñas. Nos dedicamos a evaluarlas y en base a ellas tomaremos nuestras decisiones. Es claro que deberemos mantener la atención en la tarea y recordar el contenido de una reseña. Es más, mientras las leemos deberemos recordar sus partes; no nos sirve de nada retener fragmentos dispersos de algunas frases y olvidar otras o empezar una oración y a la mitad olvidar cómo comenzaba. Es aquí que aparece nuestra facultad de memoria. Entre el interés y la capacidad ejecutiva se inserta la memoria (Portellano Pérez y García

Alba, 2014). Nuestra lectura de las reseñas durará un rato y durante este período entrará en juego lo que llamamos memoria de corto plazo. Ahora bien, cuando nos pregunten dentro de unos años cómo fue que definimos la compra de la heladera puede que recordemos que leímos unas reseñas y tal vez hayamos retenido algo de ellas, pero no tendremos un registro tan detallado como el que tuvimos al momento de tener que definirnos por una u otra opción. Habremos olvidado. Es tan importante olvidar como recordar (Portellano Pérez y García Alba, 2014). Imaginemos el caso opuesto en que no solo recordamos cada detalle de cada reseña, con cada palabra que hubiéramos leído. Imaginemos algo más extremo aún, que recordamos todo lo que haya sucedido en todos esos años, que no hubiéramos podido descartar ningún mínimo detalle. A simple vista parece espantoso para cualquier cerebro. Para poder operar, nuestros cerebros han aprendido a través de la evolución natural a recordar y han aprendido a olvidar a partir de la especialización de distintos niveles de memoria. Existen memorias de corto y de largo plazo y cada una tiene formas distintas de proceder, a su vez existe un tipo de memoria que se conoce como memoria de trabajo (Cowan, 2008). Esta última es clave para la coordinación necesaria en la ejecución de acciones en planes de mediano y largo alcance. La facultad de la memoria humana en todas sus facetas deberá ser conquistada si alguna vez se piensa llegar a la llamada IA general capaz de igualar y superar al ser humano en cualquier tarea. Por ahora el último tramo de las funciones de memoria de trabajo no cuenta con algoritmos computacionales desarrollados como para competir con la capacidad ejecutiva humana (Legg, 26 oct 2023 YouTube).

Este círculo virtuoso establecido entre la indagación de la neurobiología y las disciplinas computacionales ha generado un impulso decisivo. Recordemos que entre uno y otro campo disciplinar existe un intercambio de modelos y homologías que han resultado en mutuo beneficio. Se trata de un proceso muy dinámico, ya que como bien sabemos, las ciencias no son conjuntos estables de hipótesis confirmadas. En el campo de la biología los enfoques son muchos y variados. Las ciencias biológicas y las disciplinas cognitivas ofrecen un horizonte amplio desde el cual extraer modelos y conjeturas. Muchas veces

es más fecundo contar con marcos de hipótesis e interrogantes desafiantes como plataforma para generar modelos que partir de certezas cristalizadas y de convicciones inamovibles (Popper, 1998). La historia de las ciencias nos indica que aún las certezas más firmes han sido conmovidas por nuevas conjeturas, en el terreno del pensamiento nada es estanco. Conexiones neuronales, ondas que recorren la superficie de la corteza cerebral, energías que se dispersan según modelos termodinámicos y hasta enlaces cuánticos, los postulados que conviven en este mundo de las facultades humanas superiores del lenguaje y el entendimiento son muchos y fascinantes. El campo de la biología es un terreno en permanente revisión, en el que conviven y compiten muchas hipótesis y modelos algunos de los cuales son impulsados por investigadoras e investigadores argentinos (Figura 1). Retengamos la idea de que las ciencias de la computación y la ingeniería informática se han inspirado frecuentemente en modelos biológicos, lo que no es equivalente a sostener que hayan logrado replicar en parte el funcionamiento del cerebro. No hay certezas últimas sobre el funcionamiento del cerebro o de la memoria así que el "préstamo" de la biología a la informática quedará en el nivel de los modelos y las conjeturas. Veamos de qué se trata esto de la memoria y las profundas diferencias entre la memoria biológica y la de las computadoras.

¿Qué es la memoria? ¿Dónde está situada la memoria? Existen diferencias conceptuales y físicas entre la memoria del hardware en una computadora y una red neuronal. En una computadora existen dos tipos de memoria, la temporal (RAM) y la de almacenamiento en disco. Por otra parte, estas memorias son administradas por una unidad central de procesamiento, la CPU. La memoria en la computadora tiene dos características que nos interesan: la información está ubicada en lugares físicos y existe una unidad de procesamiento que administra su funcionamiento. En las redes neuronales, tanto biológicas como artificiales estas dos características no están presentes. La memoria es algo que se actualiza, que opera cuando se activan las redes neuronales. La memoria no es algo que "está" en algún lado "guardada", es algo que se produce a partir de una actividad. Puede resultar un poco difícil de representar, se trata de un concepto que no es del todo intuitivo. La unidad funcional de

la memoria reside en el enlace neuronal. Resulta conveniente pensar en ella como algo que ocurre cuando existe una actividad, no como algo que está ubicado en un lugar físico específico. Pensemos en redes o circuitos antes que en estanterías o cajones (una metáfora frecuente que no me pertenece). Estas nociones derivan de la hipótesis de un neuropsicólogo llamado Donald Hebb publicada en 1949 en la que propone que la base funcional de procesos como la percepción, el almacenamiento de memoria o la toma de decisiones es el enlace neuronal (Brown, 2020).

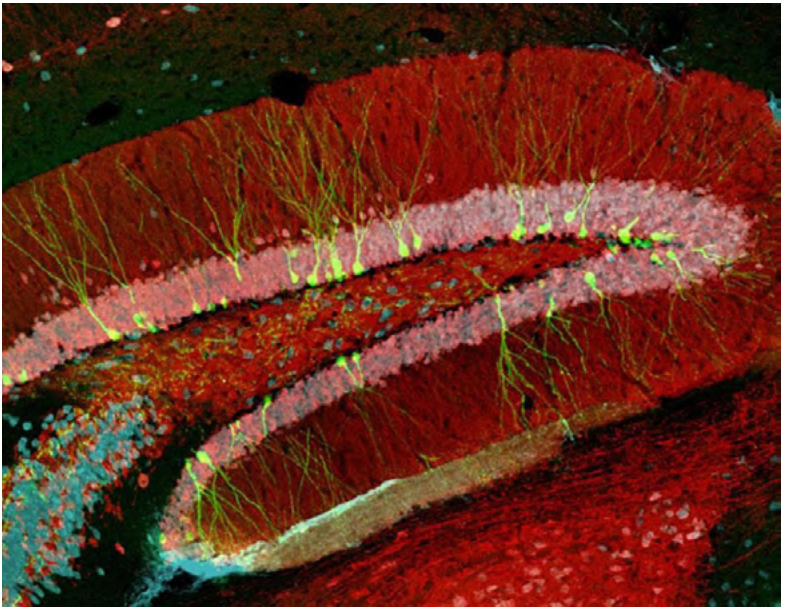


Figura 1. El funcionamiento de las neuronas en procesos como el de la formación de memoria avanzan a partir de nuevas técnicas de imagen y de la exploración de las ciencias biológicas. Los paradigmas en torno al funcionamiento neuronal se modifican constantemente. La manera en que operan las redes neuronales biológicas y artificiales encierra muchos misterios y dista de ser un territorio conquistado. Esta imagen ilustra un estudio de Antonia Marin-Burgin, Alejandro Schinder, Lucas Mongiat y María Belén Pardi, investigadoras e investigadores del CONICET, en torno a la formación y activación de neuronas jóvenes. Sus aportes contribuyen a cambiar viejos paradigmas vinculados al nacimiento de neuronas. Se relacionan con varios procesos entre otros el de la formación de memorias. Imagen extraída de: https://www.leloir.org.ar/cientificos-argentinos-descubren-un-proceso-clave-en-el-funcionamiento-de-la-memoria

Volver

A partir de este paradigma de las redes neuronales como bases de procesos artificiales se ha logrado imitar las capacidades de memoria de corto y largo plazo. La biología ha funcionado como fuente de inspiración para el procesamiento de las informaciones en las computadoras. Esto nos acerca aún más al manejo del lenguaje natural en las llamadas redes recurrentes. Estos modelos trajeron como novedad la capacidad de recordar los elementos más importantes y dejar de lado los menos relevantes en las cadenas de texto. Tienen capacidades de retener aquello que resulte significativo dado determinado contexto, pero a la vez desechar lo que ya no resulta importante. Se trata de redes neuronales recurrentes ¿qué quiere decir esto? Las redes recurrentes LSTM reciben un input, realizan una predicción sobre el resultado de la secuencia y ese resultado lo "envían para atrás" como información para el contexto del texto. De esta manera actualizan la información de manera constante a partir de los resultados que van produciendo. La red va a recordar los pasos anteriores y los integrará cuando reciba otro input (Figura 2). Este es un progreso significativo.

Este tipo de redes determina, para un estado de desarrollo del texto, cuáles son las partes que siguen siendo relevantes y las que se pueden desechar. Pensemos por ejemplo que estamos manejando un texto en que se habla en plural de lo que hace un grupo de chicos en una plaza y,

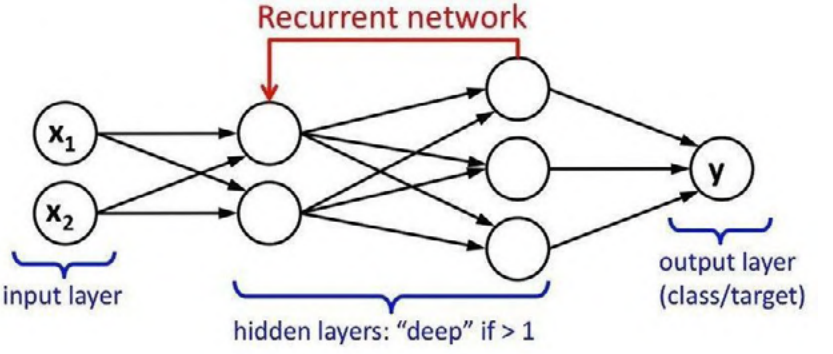


Figura 2. La red neuronal recibe el input, por ejemplo, una serie de palabras de una oración, las capas ocultas (hidden layers) realizan las inferencias y producen un resultado que a su vez se "manda para atrás" (presten atención a la flecha roja). Así las redes de memoria recurrente avanzan en el tiempo reteniendo información relevante. Imagen subida a Researchgate por Vidushi Mishra. https://www.researchgate.net/figure/Recurrent-neural-networkRNN-or-Long-Short-Term-MemoryLSTM-5616_fig2_324883736

a cierta altura, el texto empieza a cambiar y se focaliza en lo que hace un chico en particular. La red actualizará el estado y dejará de prestar atención al plural "ellos" o a las palabras "grupo" y "amigos" y prestará interés a la palabra "el" y al nombre propio "Martín" que es el protagonista en el contexto actual del desarrollo del texto. Este tipo de enfoque de redes recurrentes LSTM resulta efectivo en campos como el de la traducción. Muchos de ustedes habrán notado una mejora significativa en los traductores cuando se pasó de los métodos estadísticos anteriores al paradigma de las redes neuronales. Otra área en que el tema de la memoria es crucial es el de la implementación de chatbots. En los sistemas de atención automatizada la actualización del estado y el manejo de la secuencia temporal resulta indispensable; un asistente virtual que no pueda manejar correctamente el flujo de la conversación es algo enervante.

A pesar de ser buenas soluciones para el manejo del lenguaje natural, las redes neuronales recurrentes se muestran poco efectivas en dos sentidos: son "pesadas" y lentas en el procesamiento y presentan algunas dificultades cuando los textos son muy extensos. Veamos un ejemplo:

> *La mariposa aleteaba plácida entre las flores y el sol brillante de la tarde se asomaba entre las nubes para iluminar la seda de sus alas.*

Recordemos que las redes recurrentes vinculaban cada palabra que procesaban como contexto de la siguiente, lo que resulta muy efectivo cuando las palabras están cerca las unas de las otras en el texto. Seguramente la red no tenga problemas para entender que el contexto varía en determinado punto de la oración y se pasa de femenino a masculino en el caso de "las flores" y "el sol". Las palabras son vecinas y eso favorece que se le otorgue mucho peso a su relación. Sin embargo, la red recurrente tendrá más dificultad para vincular el posesivo "sus" con el sujeto de la oración, el sustantivo "mariposa". El motivo de este déficit es que el modelo tiende a dar más peso relativo a las palabras próximas, así funcionan sus mecanismos de memoria. Parece bastante natural, ya que seguramente no recordemos con qué palabra comienza este libro y nos será más sencillo recordar el contenido de esta oración que acabamos de leer. Este problema lo viene a solucionar el algoritmo Transformer, ya que no analiza el texto

una palabra tras otra en secuencia sino bloques de texto todos al mismo tiempo sin importar su orden (procesamiento en paralelo).

El asombro de Adán: los modelos *Transformers*

En una entrevista Aidan Gomez relata la sorpresa que se llevaron, él y el resto del equipo de investigación en Mountain View en los laboratorios de Google cuando despertaron una mañana y los modelos que estaban entrenando ya no balbuceaban, sino que generaban texto con una enorme coherencia. Se había roto una barrera a partir de la implementación del máximo protagonista de esta historia: el modelo Transformer había adquirido capacidad de generar lenguaje a un alto nivel. Se había abierto la compuerta hacia los grandes modelos de lenguaje LLMs.

La belleza arquitectónica del modelo reside en su simpleza, no son más de 200 líneas de código, comenta Gomez en una entrevista (YouTube, 24 de mayo 2023, Eye on AI). Obviamente se refiere a un modelo básico, el código total para lograr su desarrollo operativo, con ajustes, hiperparámetros, etc. es mucho más extenso. El secreto de su potencia reside en un mecanismo con el que contamos los humanos: el de la atención (Vaswani et al., 2017). Una vez más la ingeniería de software se inspira en la mente humana. ¿De qué se trata este mecanismo de atención con el que cuenta el modelo Transformer? "Prestar atención" implica relacionar las partes en una secuencia. Para operar con el lenguaje natural es necesario determinar qué partes de una secuencia de texto se relaciona con las otras. Por ejemplo, si tenemos un adjetivo, a qué parte de la oración se lo atribuiremos, si aparece un artículo a qué lo conectamos. Las palabras se conectan con las palabras, con las frases, con los textos, el mecanismo de interés logra realizar las operaciones necesarias para detectar esos vínculos. Algo de este estilo habíamos señalado en relación con los modelos recurrentes, pero la arquitectura Transformer va a llevar esta capacidad a un nivel superior.

Se ha superado el problema de conectar los términos entre sí en textos más largos ya que se prescinde del mecanismo de la memoria y la secuencia temporal de los textos. Ahora analiza todo al mismo tiempo. De hecho, resulta bastante sorprendente que se haya logrado semejante dominio de

la lengua natural. Entender la conexión entre los términos en frases con las que lidiamos en el mundo cotidiano nos suele resultar bastante sencillo. Textos que nos parecen transparentes implican una serie amplia de operaciones que realizamos de manera inconsciente, de eso se trata, en parte, la competencia lingüística. En el manejo cotidiano del lenguaje natural existe toda una gama de operaciones que realizamos de manera intuitiva, que nos permiten detectar vínculos complejos entre textos distantes y llenos de sentidos implícitos cuyo lazo no es posible reconstruir. Una cosa es realizar una operación y otra muy diferente es poder reconstruir la manera en que la hemos realizado. Lo hacemos porque lo hacemos, lo sabemos porque lo sabemos, así de simple. Recordemos que, a pesar de todo el desarrollo que tienen las neurociencias y sus parientes cercanos las disciplinas cognitivas, la manera en que explicamos el funcionamiento de la mente, el pensamiento y el lenguaje es altamente conjetural.

El algoritmo Transformer logra establecer vínculos en cadenas de lenguaje a partir de soluciones de software, es decir, a partir de algoritmos (Figura 3). En torno a este modelo se concentraron los esfuerzos de la comunidad de IA, se invirtió capital, se corrieron experimentos, se comenzó a entrenar los grandes modelos Transformer con enormes masas de datos de todo tipo. Se eligió ese camino y aquí estamos. Según reconocen sus creadores, pudo ser otra solución, pero de hecho fue esta (YouTube, 24 de mayo 2023, Eye on AI). Antes de que se adoptara en forma masiva a nivel global lo adoptó la industria, lo hicieron los inversores y lo hizo la academia. La conjunción de esfuerzos mancomunados fue clave porque es un modelo muy sensible a la cuestión de escala. Cuanto más grande el modelo mejor funciona, suma capacidades e incrementa su performance. La operatividad del modelo Transformer crece exponencialmente con las estrategias de "fuerza bruta", agregando más datos de todo tipo para entrenarlos, sumando hardware, incrementando las horas de entrenamiento y ajuste, construyendo data centers cada vez más gigantescos y utilizando más energía. Existen tendencias en la actualidad que incrementan la performance a partir de otras vías como las mejoras en la calidad de los datos, las estrategias de entrenamiento o que apuntan al diseño de algoritmos más eficientes, pero la tendencia a incrementar el tamaño de los modelos de lenguaje no se ha detenido.

Volver

La noche en que los Transformers comenzaron a cambiar el mundo

I didn´t really have an appreciation for what we have accomplished at the end. I remember like the night before the deadline, the night before we have to submit. It was 3AM, Ashish (Vaswani) and I were sitting at the couch and he turn to me and he said – Aidan, this is going to be huge -. My reaction was like - ¿Really, you think so? – I think what I didn´t appreciate, couldn´t appreciate, was the fact that such a simple method could achieve such an insanely high performance. It was hard to see the future. I think that Ashish saw that happening.

Machine Learning Street Talk. (14/09/2022). Entrevista a Aidan Gomez. Language as Software [Archivo de video]. Street Talks Youtube Channel

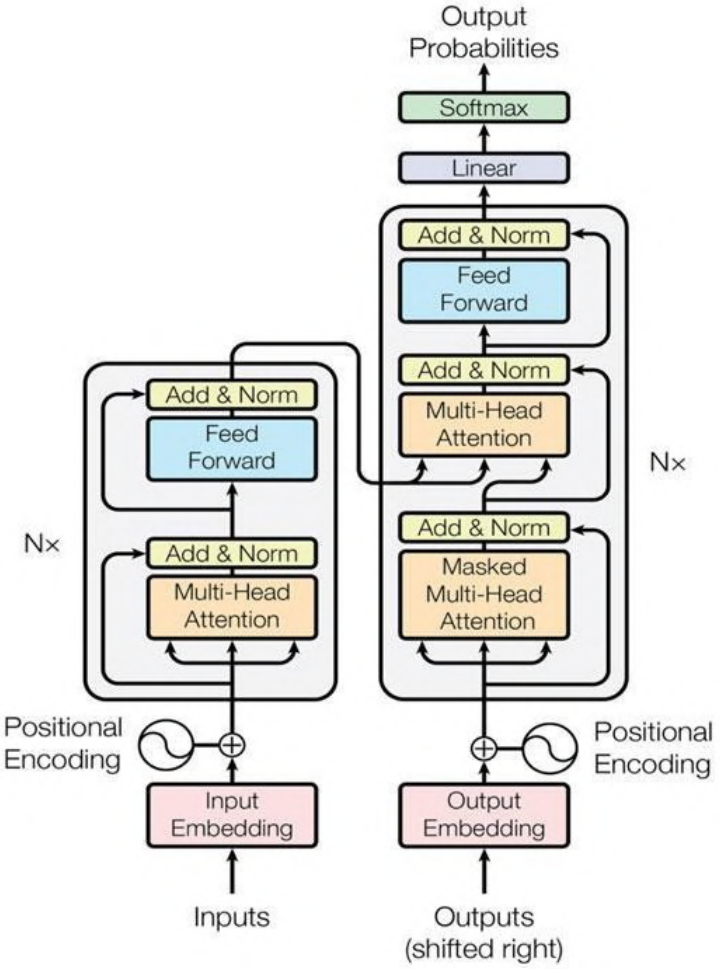


Figura 3. La imagen del algoritmo Transformer es una de las más reconocibles en el universo de la IA. Es la base de los grandes modelos generadores que dominan el panorama actual. Tal como ocurre en todos los modelos que comentamos en este libro existe un input y existe un output. Se trata de un algoritmo que codifica y decodifica secuencias de textos mediante un mecanismo de atención que le permite detectar las relaciones entre los términos del mismo (palabras – tokens). Su misión es generar la palabra con mayores posibilidades de completar un texto. Con esta capacidad tan sencilla logra solucionar una enorme cantidad de tareas. El lenguaje, natural o artificial, nunca deja de asombrarnos. Imagen extraída de: https://www.iic.uam.es/innovacion/transformers-en-procesamiento-del-lenguaje-natural/

Volver

El modelo Transformer por qué funciona y cómo opera

Una de las primeras experiencias que tuve en interacción con ChatGPT, fue la de pedirle que generara un guión para un video educativo orientado a adolescentes de escolaridad media acerca de las ventajas de combustibles menos contaminantes en el delta del Paraná. Para eso utilicé un conjunto de instrucciones bien detalladas (prompt), en las que le especificaba que desechara los chats anteriores, que procediera paso a paso, que numerara los ítems, que generara cuadros para determinadas cuestiones, le señalaba cuáles tenían que ser sus competencias (redactor de contenidos educativos / ingeniero especializado en energías renovables, etc.) y cómo quería que se organizara el texto (siguiendo pautas de narración audiovisual / duración del video / características de los personajes). El resultado, sin ser un dechado de creatividad, resultaba bastante asombroso. Se trataba de un texto coherente, articulado según las reglas del género "video educativo", definía roles al interior de la historia, planteaba cierta tensión dramática y cumplía con las pautas que le había suministrado. Puede que en la actualidad no nos asombre, pero les puedo asegurar que en su momento resultaba impactante. ¿Cómo es esto posible con la única capacidad de predecir la próxima palabra en un texto?

Vamos a avanzar en este sentido eludiendo las cuestiones técnicas de diseño y los planteamientos matemáticos que rodean el tema. Tengamos en mente que, por más que podamos entender conceptualmente los modelos y su lógica, no dejan de ser modelos de caja negra. En el fondo son cálculos numéricos con la artillería del álgebra, el cálculo, la lógica, la probabilística y todo lo que se nos ocurra. Podemos ir un poco más allá en nuestro desconcierto y pensar si la matemática que concebimos los humanos alcanza para comprender lo que procesan esas enormes sumas y multiplicaciones, en el seno de esos gigantescos cúmulos de parámetros y de patrones detectados por las redes (Molnar, 2021; Prince, 2024). Ni siquiera los diseñadores de estos algoritmos o los matemáticos más calificados terminan de descifrar la operativa de estos "engendros" de modelos sobre parametrizados. No desesperemos y volvamos a nuestro intento de entender conceptualmente de qué se tratan estas redes neuronales Transformer.

Volver

Un consejo de Popper: no se pregunten qué es, pregúntense por qué y cómo

Las preguntas de qué-es tales como ¿qué es la justicia? o ¿qué es la corroboración? carecen de todo valor, son siempre inútiles y carecen de todo valor científico o filosófico; y lo mismo ocurre con todas las respuestas a las preguntas qué-es como las definiciones.

K. Popper, 1998, Realismo y el objetivo de la ciencia, p. 301.

No se pregunten qué son las redes neuronales o qué es el algoritmo Transformer, si es inteligencia, si no es inteligencia o cualquiera de estos debates bizantinos. Si transitan ese camino es fácil que terminen empantanados en discusiones interminables en torno a definiciones. Desde mi punto de vista es una forma clásica de perder el tiempo y no contribuye a mejorar nuestra capacidad de interacción con las IA. Los invitamos a pensar por qué operan como operan y a partir de qué características logran hacerlo. Existen varios factores que hacen que este diseño de redes neuronales generadoras con aprendizaje profundo sea algo diferente, un "deal breaker". En principio nos concentramos en tres de ellas: el mecanismo de atención, el procesamiento paralelo y la secuencia codificación - decodificación. De hecho, el procesamiento paralelo potencia el mecanismo de atención. El tercer factor será la secuencia de codificación – decodificación de textos que siguen el viejo esquema input output. Con estos tres aliados coordinados podremos entender algo sobre "el secreto" de estos grandes modelos de lenguaje.

Habíamos establecido que los modelos predecesores al algoritmo Transformer, los de las redes neuronales recurrentes, operaban de manera tal que podían replicar la memoria de larga y corta duración. Son redes cuya virtud principal es la de mantener coherencia a lo largo del tiempo. A través de mecanismos de atención lograban focalizar en aquellos elementos del texto que resultaban relevantes en el contexto de una secuencia textual que se desarrollaba hacia adelante en el tiempo. Esto quiere decir que el procesamiento del lenguaje seguía una secuencia. Son un ejemplo típico de procesamiento secuencial. Esto implica que el procesamiento del texto se hace parte por parte progresando hacia adelante. ¿Cuál es el problema con este procedimiento? Su lentitud. Para procesar una parte del texto debe esperar procesar la anterior, a esto se refiere la idea de secuencia. De hecho, habíamos remarcado que esta era una virtud, dado que las redes recurrentes eran muy buenas en la conservación de la coherencia dada su capacidad de establecer vínculos a partir de la memoria y actualizar el estado definiendo cuáles de las partes del texto seguían siendo relevantes en un contexto dado. Eso es lo que va a dejar de lado el algoritmo Transformer, prescinde de la memoria y se queda sólo con el mecanismo de atención.

Volver

Si lo pensamos con un poco de detenimiento y a la luz de nuestra concepción del lenguaje natural, la idea de que una frase se despliega hacia adelante en el tiempo resulta casi obvia. Es la vieja y conocida noción de metonimia o de eje sintagmático (Russell y Norvig, 2004). Pero las cosas en el universo de la técnica o el de la ciencia son así, en ocasiones hay que dejar de lado lo que conocemos y aceptamos como cierto y pensar ideas nuevas que rompan con nuestras nociones previas, por contra intuitivo que estas puedan parecernos. Esto no quiere decir que se puede romper con lo anterior porque se nos ocurre, sin solucionar cuestiones cruciales, como en este caso la del establecimiento de vínculos entre los términos de un texto. La memoria de las redes LSTM eran una solución elegante y efectiva y era necesario encontrar la manera de atacar el problema de los vínculos intra textuales e inter textuales. Aquí entran a jugar el tema de los mecanismos de atención.

Antes de ver cuáles fueron las dificultades y cómo se solucionaron veamos cuáles fueron los beneficios que trajo obviar el tema de la secuencia temporal del texto. Como ya señalamos, los procesadores tienen dos posibilidades, pueden realizar sus operaciones una por una en secuencia o pueden realizar varias tareas al mismo tiempo. Las redes recurrentes LSTM trabajan en secuencia, el algoritmo Transformer permite realizar procesamientos en paralelo. Una tarea como es el procesamiento de un texto se puede descomponer en varias sub tareas que se ejecutan al mismo tiempo. Todos hemos escuchado los conceptos de multi núcleos o multi procesadores. Una de las virtudes de las tarjetas gráficas GPU es su eficacia de procesamiento paralelo (su adaptación por parte de NVIDIA resultó un verdadero éxito). Estos y otros desarrollos de hardware pueden ser aprovechados por algoritmos capaces de dividir sus tareas en partes más pequeñas. Si procesamos todos los elementos de un texto al mismo tiempo ganaremos velocidad y el modelo se tornará más eficiente por el ahorro de recursos. En informática el tiempo de ejecución, la simplificación de operaciones y la reducción del número de pasos para el logro de un objetivo son siempre factores determinantes. El algoritmo Transformer en su sencillez aporta todos estos beneficios, queda por ver cómo lo logra si prescinde de la memoria y de la temporalidad sintagmática que poseían las redes recurrentes LSTM. No obstante

Attention is all you need

We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely.

A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser y I. Polosukhin, 2017, Attention is All you Need. Neural Information Processing Systems. https://doi.org/10.48550/arXiv.1706.03762

no queremos dejar la impresión de que las redes Transformers tengan un rendimiento superior para cualquier tarea. Tanto un diseño como el otro tienen sus virtudes y limitaciones de acuerdo al contexto de uso.

Atención es todo lo que necesitas

El paper "Attention is all you need" publicado en 2017 tuvo una importancia comparable con la publicación de Word2vec en 2013 (Mikolov et al., 2013) o el redescubrimiento del algoritmo de retropropagación del error en 1986 (Hinton, 2022) idea original que se había planteado hacia mediados de la década de los 70. Esta es la columna vertebral de desarrollo de los modelos de redes neuronales Transformer que tanto han avanzado en la tarea de lograr que hombres y máquinas se comuniquen mediante el lenguaje natural.

Lo único que necesitas es atención, porque se puede prescindir de la memoria y de la secuencia temporal del texto al momento de determinar las relaciones entre los elementos del mismo. El gran problema que debió superarse es el de la detección de las relaciones entre los términos, la cuestión de la asignación de vínculos y la de discriminar cuáles son las partes claves en una cadena lingüística. ¿Qué se conecta con qué en una frase? A partir de estos mecanismos de atención y la realización de cálculos matemáticos logra determinar la relación entre todas las partes de un texto. Esto lo puede hacer otro tipo de algoritmos, no es privativo de las redes Transformers. Recordemos que a partir de Word2vec los términos se pueden codificar en forma de conjuntos de números (filas de números a los que se conoce como vectores) que captan diversas características de cada palabra o parte de palabra (token). La diferencia que introduce el mecanismo de interés de los modelos con base en el algoritmo Transformer es que no se trata de mapas semánticos estáticos, sino de soluciones que se adaptan dinámicamente a los contextos que se les presenta. Si avanzan un poco en la comprensión del modelo Transformer se darán cuenta de la utilidad que tuvo aprender a derivar en cuarto año de secundaria, el interés de evaluar diversos tipos de funciones y su sentido, por qué es importante aprender en bachillerato que es una función trigonométrica o por qué resulta útil saber multiplicar matrices, pero hablar de estas cosas no es nuestro propósito. Hasta aquí llegamos en la explicación del modelo, existen muchas otras cuestiones a explorar para quienes quieran internarse en

From a scientific point of view, it is desirable to obtain a unitary model of the world comprising both mechanical and psychological phenomena. Such a theory would become available, for example, if the workers in Artificial Intelligence, Cybernetics, and Neurophysiology all reach their goals. Still, such a success might have little effect on the overall form of our personal world-models. I will maintain that for practical, heuristic reasons, these would still retain their form of quasi-separate parts.

M. Minsky, 1965, Matter, mind and models, p. 3. Massachusetts Institute of Technology Cambridge, Massachusetts. Matter, Mind and Models (mit.edu)

estos complejos mundos de los algoritmos y de sus esfuerzos por manejar el lenguaje natural. Nos quedan sólo dos tareas por delante: la de recapitular algunas nociones sobre los grandes modelos de lenguaje y la de presentar algunos de sus logros pasados y algunas de las alternativas que se anticipan para el futuro.

Bibliografía

Brown, R. E. (2020). Hebb and the Organization of Behavior: 17 years in the writing. *Molecular Brain, 13*. https://doi.org/10.1186/s13041-020-00567-8

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research, 169*, 323-38 .

Hinton, G. E. (2022). The Forward-Forward Algorithm: Some Preliminary Investigations. ArXiv, abs/2212.13345.

Mikolov, T., K. Chen, G. S. Corrado y J. Dean (2013). Efficient Estimation of Word Representations in Vector Space. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1301.3781

Minsky, M. (1965). *Matter, mind and models*. Massachusetts Institute of Technology.

Molnar, C. (2021). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Leanpub.

Popper, K. (1998). *Realismo y el objetivo de la ciencia*. Tecnos.

Portellano Pérez, J. A. y J. García Alba (2014). *Neuropsicología de la atención, las funciones ejecutivas y la memoria*. Síntesis.

Russell, S. J. y P. Norvig (2004). *Inteligencia artificial. Un enfoque moderno*. Pearson Educación, S.A.

Vaswani, A., N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser y I. Polosukhin (2017). Attention is All you Need. *Neural Information Processing Systems*. https://doi.org/10.48550/arXiv.1706.03762

YouTube – Podcast - Blogs

Aidan Gomez on How AI Language Models Will Shape the Future (2023) Eye on AI https://www.youtube.com/watch?v=-xobW4jh66U&t=1297s

Entrevista a Aidan Gomez. Language as software. Fecha de estreno: 14 nov 2022. Street Talks Youtube Channel. https://www.youtube.com/watch?v=ooBt_di8DLs&t=627s

Shane Legg (DeepMind Founder) - 2028 AGI, Superhuman Alignment, New Architectures 26 oct 2023 Dwarkesh Podcast

Epílogo. El match entre la inteligencia humana y la artificial

Nuestro propósito general es llegar a este punto en que podamos reflexionar en torno a las capacidades humanas y las artificiales. Existe una superposición terminológica, usamos las mismas palabras para nuestras capacidades y para las de la IA lo que genera confusiones que conviene despejar. La utilización de los mismos términos no implica que nos encontremos ante los mismos conceptos. Para ello y para cerrar el libro vamos a contraponer las capacidades humanas con las de la IA desde el punto de vista conceptual. Vamos a detenernos en el aprendizaje, la percepción, la capacidad de actuar como agentes, el razonamiento y como punto final una cuestión el debate permanente, el surgimiento tan discutido de capacidades emergentes. ¿En qué devendrá esta transformación? El tiempo lo dirá. Espero que este libro los haya ayudado a pensar en el lugar de lo humano en el contexto de un salto tecnológico sin precedentes, con ese ánimo les pido que se internen en la lectura del epílogo.

Aprendizaje

Los grandes modelos generativos de lenguaje natural son diseños con aprendizaje mixto, no supervisado y supervisado. En una primera fase se alimentan con enormes masas de información textual no etiquetada. Los datos, obviamente se acomodan y dejan prolijos, pero no mucho más que esto. En los casos de los grandes modelos de Open AI, de Google, Anthropic o de Meta los datasets de entrenamiento fueron gigantescos. Se deja que el modelo detecte patrones de todo tipo al interior

del input de datos, está aprendiendo. Como ya dijimos, los primeros resultados que arroja el modelo son bastante precarios, es bastante tosco. Aquí entra el aprendizaje supervisado y el ajuste, ya que se necesita la participación humana para mejorar la performance. El programa deberá ajustar los resultados a partir de sus errores (función de coste). El modelo se ajusta de manera que logre producir resultados que puedan parecer correctos a la luz de los criterios humanos.

La máquina aprende a predecirnos, a satisfacer nuestras expectativas. Piensen un instante y podrán entender por qué esta es una de las mayores preocupaciones entre aquellos que las consideran una amenaza para la humanidad. Incrementamos todo lo posible la inteligencia de modelos que nos entienden lo suficiente como para conformarnos. Cada vez nos resulta más difícil discriminar aquello que se produce mediante modelos generadores o mediante la acción humana sin mediación de las IA. No todos sienten que existan estas amenazas o que las IAs puedan escalar más allá del punto en que se encuentran por limitaciones de arquitectura o por otras cuestiones. Otros argumentan que las máquinas no entienden nada y que solo escupen los outputs que puedan conformarnos. El tiempo dirá.

Entre el aprendizaje humano y el computacional detectamos dos diferencias notorias: la velocidad de transferencia y el volumen de información requerida para adquirir conocimientos. En términos de transferencia de aprendizaje las máquinas son imbatibles. En el proceso de aprendizaje de estas redes neuronales profundas la clave es la segunda etapa donde, a partir de los patrones que ha identificado por sus propios medios, se mejora la calidad de sus respuestas de manera supervisada. En esta fase, la evaluación humana permite afinar el llamado "ajuste de parámetros". Lo valioso de un modelo no es la masa de datos con que se lo alimenta sino los parámetros con los que se ajusta. Una vez que se está en posesión de esos parámetros se los puede transferir a otros modelos. Obviamente no es una cuestión tan sencilla, ya que entran a jugar cuestiones como la compatibilidad de la arquitectura de los modelos y las estrategias de entrenamiento de uno y otro. Más allá de las complejidades, es una estrategia de aprendizaje muy efectiva y con enorme potencial. La liberación de los parámetros de LLaMA la IA

de Meta, por ejemplo, generó una avalancha de modelos de lenguaje de base abierta con alta eficiencia. Utilizar el conocimiento adquirido por un modelo para alimentar a otro es una práctica frecuente que va más allá de los modelos de lenguaje y puede ser utilizada en muchas áreas de la IA. A esto se lo denomina Transfer Learning y es un proceso rápido en términos comparativos con lo que nos cuesta a nosotros los humanos llevar adelante procesos educativos. En el horizonte se dibuja la posibilidad de que los programas se repliquen de manera automática y ya se avanza en la tarea de que modelos más sencillos generen datos sintéticos y eduquen a programas más sofisticados de lenguaje con alto grado de eficiencia.

A diferencia de esta velocidad vertiginosa los seres humanos aprendemos al ritmo de una tortuga. Calculen los años de aprendizaje que hemos tenido que recorrer los humanos para entender los materiales que se estudian en un curso de grado universitario. En velocidad perdemos. Ahora bien, la cantidad de datos que necesita un gran modelo de lenguaje para adquirir conocimientos es muy grande y su capacidad de generalización es muy baja en comparación con la nuestra. Con pocos ejemplos logramos establecer enlaces e inferencias de dimensiones colosales. A la máquina hay que insistirle para que logre captar algo: funciona a fuerza bruta a partir de la masividad de los datos y las horas de entrenamiento. Somos seres con alta capacidad de captación de lógicas subyacentes y constituimos estructuras estables, plásticas y con gran capacidad de reversibilidad lógica, tal como nos lo indica Piaget a lo largo de toda su obra.

Los grandes modelos generadores de lenguaje LLMs pertenecen al universo de lo que se conoce como "modelos fundacionales". Los modelos fundacionales generan diversos tipos de outputs, de texto, imagen, sonido y en fechas recientes video. Sobre estos grandes modelos generalistas de lenguaje se pueden realizar ajustes finos (fine tunning) para especializarse en campos o tareas específicas. El gran modelo de lenguaje acumula numerosas capacidades para las que ha sido entrenado, como el manejo de texto en LLMs, pero suma otras que se denominan emergentes, como la de programar o resolver cálculos aritméticos que se han generado de manera inesperada. Esta potencia de los grandes modelos fundacionales puede ser aprovechada en actividades puntuales como el

campo de la salud, las finanzas, el derecho, la logística y casi cualquier otro ámbito de tareas que podamos imaginar. Para ajustarlo y lograr que se convierta en experto en el manejo de un campo de saber o actividad es necesario reentrenar los modelos, para lo que se les proporciona datasets con información de calidad perteneciente a nuevos dominios. En el proceso de reentrenamiento se vuelven a ajustar los parámetros de los modelos y lograr el nivel de eficacia que se considere adecuada dentro de las posibilidades. Este es una de los caminos en que se puede desarrollar el impacto de las IA, al menos con los modelos actuales. Las implicaciones para el mundo del trabajo son enormes y seguramente redundará en la organización futura de las sociedades.

Percepción

Para subsistir en un medio ambiente, los seres vivos debemos tener algún tipo de registro del mismo. Para vivir hay que percibir. Ahora bien, percibir puede significar muchas cosas, cualquier tipo de captación de diferencias en el medio ambiente implica un tipo de percepción y muchos de estos registros desencadenan acciones. Los soportes de silicio con sus entradas y salidas de corriente también logran percibir el medio ambiente. Percibir es captar un input a partir de la existencia y operatividad de un sensor. Estos inputs tienen que poder ser procesados por la entidad de la que se trate, biológica o artificial. Los humanos, por ejemplo, captamos ciertas longitudes de ondas luminosas, dentro de un espectro limitado (nuestro rango se corresponde con el famoso arco iris) o frecuencias sonoras diferentes a la que captan otras especies y somos menos sensibles a las modificaciones en el campo magnético de la tierra que las aves. Cada entidad captará señales diferentes. Ya señalamos que algunas aves solo entienden de cifras, por eso traducimos los impulsos sensoriales a números para que puedan captarlas. El lenguaje natural es solo uno de los inputs informativos que puede registrar una máquina, podemos también alimentarla con datos provenientes de los pixeles de una imagen, suministrarle registros de movimiento, de temperatura, de sonido, todo aquello que se les ocurra puede ser codificado y transmitido a un ordenador. Cada tipo de captación implicará una lógica distinta y

habilitará tipos de aprendizaje diferente, esto es así para el ser biológico y el artificial. Los grandes modelos de lenguaje aprenden a partir de texto, ese ha sido su insumo base, así como los modelos convolucionales son eficientes en el reconocimiento de imágenes y los de difusión son eficaces generándola.

Contamos con muy buenos modelos especializados en la captación, procesamiento y generación en cada tipo de registro sensorial. Pero atención: fines de 2023 y 2024 parecen ser la puerta de entrada a la multimodalidad (Figura 1). Se denomina multimodalidad a la capacidad de los modelos de integrar muchos registros perceptivos al mismo tiempo y de encontrar relaciones entre ellos, tanto en su entrenamiento como en los outputs que generan. Estas capacidades de aprendizaje e interacción multimodal pueden llegar a implicar un salto cualitativo en términos de la capacidad de los modelos de IA (Najdenkoska et al., 2023). Nos vamos a enterar pronto del impacto que generen. Para entender de qué se trata tengamos en cuenta que modelos como Gemini de Google han sido diseñados desde la base como alternativas multimodales que pueden captar

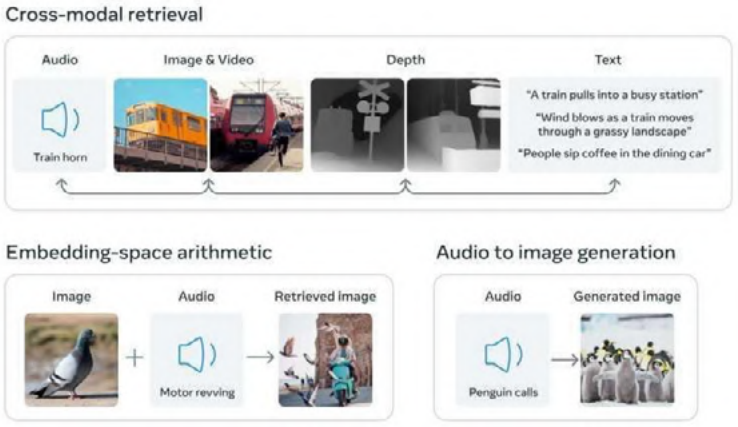


Figura 1. El entrenamiento multimodal potencia los resultados de aprendizaje en modelos de IA. Esta es una tendencia con buenos resultados para los entrenamientos de modelos de Meta y en modelos de base abierta. La captación de relaciones entre diversas modalidades sensoriales potencia las capacidades de las IA, aumenta el volumen de material disponible para su entrenamiento y las dota de una percepción más compleja del medio ambiente en que se insertan. Gemini de Google ha dado un salto cualitativo de gran escala a partir de su lanzamiento como modelo de generación multimodal a fines de 2023. Imagen extraída de: https://hipertextual.com/2023/05/meta-imagebind-ia-multisensorial-codigo-abierto

texto, imagen y video (Gemini Team, 2023). Meta ha utilizado este tipo de inputs para entrenamiento, sumados a registros térmicos y captación 3D. Estas estrategias se van a multiplicar sin lugar a dudas. Sumar registros sensoriales implica la posibilidad de captar nuevos patrones y, potencialmente, establecer vínculos cruzados que enriquezcan la relación con el medio ambiente. Un dato adicional, no solo se incrementa la modalidad de percepción, sino el volumen de datos. Es muy diferente tener a disposición el volumen de texto que circula por internet, que contar también con gran parte del video disponible en la red cosa muy sencilla para Google, el propietario de Youtube. El alcance puede ser gigantesco.

Razonamiento

Como bien sabemos, la cuestión del razonamiento en el reino de lo humano es muy amplia. Si nos detenemos a considerar, cada disciplina, campo de saber o actividad humana, podremos constatar como desarrolla su propio estilo de razonamiento. No obstante, entre toda esta diversidad existen dos reinos que quisiéramos destacar: el de la retórica entendida como el arte de convencer y el de la lógica y su capacidad de generalizar.

Comencemos por el arte de convencer mediante el razonamiento. Razonar en términos de la retórica implica conducir a otro por caminos mentales que lo obliguen a llegar a determinada conclusión. Según Aristóteles, la retórica argumentativa es una especie de violencia (Aristóteles, 2002) y la lógica lo es aún más (Aristóteles, 1982). Esta disciplina, la retórica, no se limita a los razonamientos lógicamente intachables, sino que engloba silogismos truncos e imperfectos llamados entimemas. Se trata de artilugios que se orientan a convencer y derrotar al adversario en la contienda de mentes. Son artefactos construidos para engañar en el diálogo utilizando recursos con el aspecto de la lógica, pero que no se atienen a sus rigurosas leyes de derivación. Para ponerlo en términos claros: son estratagemas para engatusar. Esta disciplina ha sido desarrollada en muchos campos de actividad humana. Existen estrategias retóricas en el campo de la ciencia con relación a la lógica, la inteligencia humana y las estrategias de validación de las teorías científicas (Apostel, 1994). Existen estrategias retóricas y han sido desmenuzadas en

el campo de la retórica jurídica y lo que se conoce como lógica informal de la argumentación (Perelman et al., 1989). Destacamos este aspecto del razonamiento humano porque hace a una dimensión ausente en los modelos de IA, la existencia social. Razonamos según la estricta lógica de la razón matemática, por supuesto, pero razonamos también en términos retóricos para convencer, para orientar a otro social con el que convivimos. Como seres sociales que somos en muchas ocasiones queremos e intentamos que el otro piense y haga determinadas cosas. El modelo en sí mismo carece de intenciones y no quiere nada en particular. Puede, eso sí, diseñar una cadena de razonamiento convincente si se lo indicamos correctamente, pero las intenciones estarán en el input de la instrucción humana. Tengan esto presente siempre. Si es lo único que deja la lectura de este libro estaremos más que conformes.

Ahora entremos en la segunda dimensión del razonamiento, la que hace a la lógica y a la capacidad de generalización. Los seres humanos razonamos y aprendemos por analogía, esa es una de las capacidades adaptativas del cerebro humano. Ante un problema nuevo podemos generalizar nuestro conocimiento y aplicar criterios análogos para encontrar la solución. No es necesario, en muchos casos, generar nuevas conexiones ya que nuestras redes neuronales pueden generalizar su capacidad de operar. Estos patrones de aprendizaje pueden dar forma a muy diferentes tipos de contenidos. Algo de este estilo pudimos repasar cuando presentamos dos maneras muy diferentes de abordar esta cuestión tan relevante en términos de la inteligencia humana: el constructivismo cuyo máximo representante es Piaget o el asociacionismo de Dehaene (uno de sus representantes más destacados).

Ahora bien, los grandes modelos de lenguaje no son tan buenos en estos rubros. Tienen dificultades para solucionar razonamientos tan sencillos como el siguiente: "Isabel es la hija de Soledad, ¿de quién es madre Soledad?". Estos escollos lógicos tienden a ser superados, pero se detectan otros de manera constante. De hecho, hoy 6 de Mayo de 2024, Copilot ha resuelto de manera correcta que Soledad es madre de Isabel, cosa que no pudo hacer GPT3 a mediados de 2023 cuando comencé a escribir el libro. Esta simple operación de reversibilidad, u otras simliares, puede ser un escollo para sistemas de lenguajes capaces de resolver otras

cuestiones de apariencia muy compleja. Pueden operar en matemática o programar aun antes de ser entrenadas específicamente para ello, pero no obstante no son capaces, al menos por el momento, de resolver la totalidad de las cuestiones que impliquen la capacidad lógica humana. Algunos razonamientos están a su alcance y otros se les complican.

Cuando se enfrentan a nuevos desafíos no logran generalizar de la misma forma y necesitan establecer nuevos nodos, extender sus redes. Esto ocurre en el enfoque que se conoce como statistical AI (inteligencia artificial con enfoque estadístico). Este enfoque estadístico y asociacionista está en el corazón de las redes neuronales profundas, pero no es el único existente. Ya mencionamos la competencia entre esta aproximación y otra llamada enfoque simbólico propiciada por autores como Chomsky (Chomsky, 2002). Recientemente se avanza en enfoques de arquitectura llamadas neurosimbólicos que intentan combinar las virtudes expresivas de las redes neuronales con la estructura lógica del enfoque simbólico (Wan, et al., 2024).

Existen varias ideas para tratar de superar estas limitaciones de razonamiento lógico como las alternativas de utilizar espacios de múltiples dimensiones como base de los modelos de razonamiento (Ananthaswamy, 2023). Otras posibilidades provienen de estrategias conocidas como *step by step reasoning*, que implican dividir los problemas en partes y contar con supervisión humana que le señale al modelo dónde se encuentran los errores. La vía de orientar a los modelos mediante prompts (Figura 2), utilizando cadenas de razonamiento (tree of thought) es una vía que parece prometedora (Wei, 2022) Esto podría permitir a los programas mejorar su conocimiento y, tal vez, comenzar a generalizar las soluciones aprendidas (Uesato et al., 2022).

Capacidades comunicativas

El mundo se transformó, comparto la certeza generalizada en torno a esta cuestión. Tal como aparecieron el fuego, el lenguaje, la escritura, la agricultura, la rueda, la máquina a vapor, la energía eléctrica, las computadoras, Internet o los celulares irrumpió la IA. En el mundo de la comunicación no nos podemos hacer los distraídos, ni evadir la

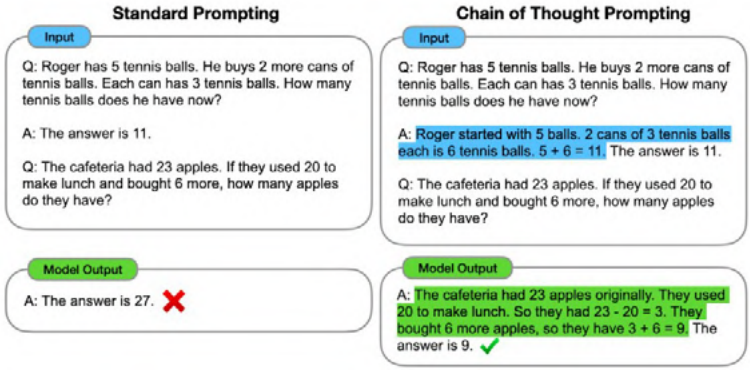


Figura 2. En lugar de dar instrucciones (prompts) de manera directa para resolver un problema, la estrategia de cadena de pensamiento indica al modelo que debe descomponer el problema en partes con lo que incrementa su performance mejorando sensiblemente el procedimiento y el resultado del razonamiento. Imagen extraída de: https://blog.research.google/2022/05/language-models-perform-reasoning-via.html

responsabilidad de conocer de qué se trata este fenómeno IA. Los LLMs generalistas combinan todos los saberes que capturan de internet al mismo tiempo, su adaptación de fine tuning a saberes particulares los tornan hiper eficientes en el manejo de actividades concretas, es abrumador. Por otra parte, debemos tener en cuenta la tendencia creciente al desarrollo de agentes personalizados en la optimización de capacidades de personas individuales, organizaciones y empresas. Estas tendencias marcan el pulso de un mundo en que la eficacia se va a incrementar y también los requerimientos de capacidades plásticas de los profesionales. Por desafiante que resulte es algo que debemos empezar a asimilar y aceptar. De no hacerlo puede verse mermada nuestra capacidad profesional en los mundos en que actuemos. Si estas estimaciones son acertadas tal vez deberíamos pensar cuál es nuestro rol como educadores. Como es obvio no tenemos una respuesta universal acerca de las estrategias que se pueden adoptar para responder al desafío. Cada uno deberá esforzarse por encontrar la que le parezca más adecuada y responsable.

Dicho lo anterior comencemos a pensar lo que implica vincularse con las IAs en interacciones comunicacionales. Una de las causas más probables de la velocidad de adopción de ChatGPT y de su curva de

crecimiento, la más acelerada en la historia de la tecnología humana, es la experiencia de conversar con una máquina. Se trata, por supuesto, de una ilusión, ya que este intercambio, aunque tenga el aspecto de una conversación, es algo totalmente distinto. A partir de un input que ingresamos al modelo se inicia un proceso de codificación de la secuencia de entrada. Se activan mecanismos de interés que detectan los vínculos en el texto y se conectan con otras operaciones de decodificación que determinan la probabilidad de aparición de una palabra hasta completar una secuencia con apariencia coherente para un ser humano. Aquí el énfasis lo ubicamos en la palabra "apariencia". A riesgo de cansarlos volvemos a remarcar que no se trata de coherencia o de incoherencia sino de predicciones que se ajustan de manera tal que satisfagan nuestra percepción humana de coherencia. No imaginen un genio super poderoso con inteligencia sobrehumana que conversa con ustedes que todo lo sabe pero que a veces alucina. Eso queda para los click baits del mundo digital que nos revelan cuál es la mejor ciudad para vivir en el planeta o el mejor método para evitar el envejecimiento de la piel, hacer crecer una fortuna o lo que sea. Tampoco le bajemos el precio a desarrollos técnicos como las IA que han logrado ser imbatibles en Go, descifrar plegamientos de proteínas, competir en olimpíadas matemáticas en el más alto nivel (IAs narrow no modelos generalistas como los LLMs), o convertirse en una base que permita jugar a Minecraft conectados por API, un logro sorprendente de las LLMs. Se trata de que tengamos una representación más adecuada de la manera en que operan los modelos, sean estos generalistas como los LLMs o con funciones específicas como la gama de prodigios de Alpha en Deep Mind que tanto asombro nos deberían producir. Para ponerlo en términos directos: no "conversamos" con una LLM, sino que interactuamos con funciones matemáticas.

Esto es posible porque las redes neuronales profundas han encontrado patrones para una enorme cantidad de cuestiones que hacen a nuestra cultura, nuestras ciencias, nuestras creencias, modos de proceder, géneros textuales, datos fácticos, patrones para casi cualquier cosa. También una cantidad inmensa de patrones que no tienen significado desde el punto de vista del saber humano. Esto también "habita" en los procesos que pueden activarse en estos colosos que son los LLMs (hiper parametrizados). Les

aconsejamos pensar en procesos que se activan y no en conocimientos que "están" en algún lado de la red neuronal, tal y como habíamos señalado en el caso de la memoria humana. Aquello que consideramos significativo desde el punto de vista humano es solo una pequeña porción del potencial de la red neuronal profunda, el sentido en el que lo ajustamos para hacerlo coincidir con nuestra estrecha captación del entorno.

No olvidemos cuál es la índole del material con que se entrena un LLM, cuya diversidad incluye todo tipo de informaciones. En un mismo dataset figuran la idea de que la tierra es un geoide, que es redonda y que es plana al mismo tiempo. Podemos pensar figurativamente que entre las respuestas que puede brindar una IA generativa existen regiones que consideraríamos coherentes y regiones que consideraríamos disparatadas según nuestras propias convicciones. Por otra parte, nuestras certezas no son gran cosa que digamos tal como lo ha demostrado la historia del pensamiento humano. No olvidemos que lo que una generación considera certeza para la otra es causa de risa. A pesar de ello los seres humanos nos formamos una imagen del mundo y desarrollamos una sensación de certidumbre. Sentimos que el mundo que habitamos tiene cierta consistencia. Tal como lo señala Prince, las IA no tienen la misma capacidad de consolidar las informaciones y creencias que los seres humanos (Prince, 2024). Menos aún necesitan desarrollar una sensación de certidumbre respecto del ambiente en que se enclavan o siquiera registrar que se insertan en un ambiente. Los humanos por incoherentes que seamos y lo contradictorias que sean nuestras creencias somos seres hábiles en justificar nuestras inconsistencias. Existen muchas teorías e hipótesis que abordan el tema de maneras diversas, como la teoría de la disonancia cognitiva (Festinger, 1975) que aborda aspectos de la cognición individual o la teoría del análisis crítico del discurso (Van Dijk, 2002) que aborda la dimensión social del tema de nuestras creencias. Por disparatada que pueda ser la manera en que pensamos e insostenibles las creencias que profesamos vamos a esforzarnos y en muchos casos lograr sostener ese cúmulo de sin sentido de manera relativamente consistente. Esta es una capacidad que las IA de Deep learning no necesitan ya que su orientación es fundamentalmente predictiva y estadísticamente orientada. Las predicciones de las IA nos pueden parecer

inadecuadas desde nuestra perspectiva humana. Esta inadecuación se dispara por muchos tipos de cuestiones. Puede que no existan datos disponibles, que no esté suficientemente ajustada (*underfitting*) o sobre ajustada (*overfitting*), que el modelo no llegue a interpretar adecuadamente el input o *prompt* con el requerimiento con que activamos su capacidad generativa, que haya tenido un problema lógico o de cálculo, no tiene ninguna relevancia. Se hace mucho hincapié en que entrega sus respuestas de manera convincente y eso es un detalle sin ninguna relevancia, todos los que tengan acceso a la regulación de temperatura de los grandes modelos lo saben. Los modelos tienen la capacidad de "modular" las respuestas según porcentajes de certeza, rigor, creatividad, humor u otras variables. Su respuesta puede tener un aspecto asertivo porque así está configurada. Pudiera haberse programado para mostrarse dubitativa, chistosa, compasiva, hostil, o desplegar la apariencia de tener un carácter en particular, incluso imitar una celebridad mediática o un científico fallecido hace 100 años. Si utilizamos los prompts y elegimos el modelo de IA adecuado lo hará. Si pensamos en términos del tono de sus respuestas y la estilística con que se ornamentan estaremos errando completamente el punto. Si volvemos a la retórica antigua (Aristóteles, 2002) esta nos dará la pista de cómo evaluar el engendro técnico del modelo de lenguaje de red neuronal profunda. Siempre que interactúen tengan en cuenta tres niveles: el tópico o temático (los contenidos con que entrena la IA), el argumentativo (su lógica) y el estilístico (la tonalidad y estilo con que se construyen sus respuestas). Hagan esto durante un tiempo de uso constante y lo van a incorporar inconscientemente. Les aseguro que si tienen presentes estos aportes milenarios de Aristóteles estarán en mejores condiciones para vincularse con estos valiosos compañeros de ruta que son las IAs generativas de lenguaje. Aumenten capacidades sin dejarse fascinar o confundir.

¿Por qué nos da la impresión de que "conversamos" con estos grandes modelos generativos de lenguaje natural? Porque se trata de redes profundas con muchas capas a las que se añade una capa adicional que le permite interactuar en forma de chat, como es el caso del ChatGPT. A la red de aprendizaje profundo Gpt se le añade la llamada capa conversacional de chat que podemos pensar como una "interfase" entre hombre

(input) y máquina (output). En este juego input output se pueden establecer intercambios reiterados de alternancia (hombre - máquina - hombre - máquina...). Una vez inmersos en el ciclo nos puede dar la impresión de que dialogamos. La adición de los comandos por voz incrementará ese efecto ilusorio, e incluso le aportará un tinte emotivo ajustable a distintas temperaturas y estilos. Este monto de emotividad puede producir efectos psicoemotivos complejos y potencialmente nocivos que deberemos considerar en el futuro, en particular si avanzan los desarrollos de agentes de asistencia personal. Dado que tienen como misión predecir una palabra, predecirá una palabra. En ocasiones no encontrará la respuesta que consideremos adecuada para la pregunta, pero su misión no es atenerse a la verdad de una respuesta sino a la probabilidad de aparición de una palabra. Olvídense de la idea de alucinación que tan engañosa resulta. Usemos la palabra ya que está instalada, pero consideremos cuál es el alcance efectivo del concepto.

En nuestro mundo cotidiano intercambiamos informaciones, compartimos comunicativamente nuestras creencias y esparcimos aquello que consideramos nuestro saber. Esto forma parte de la manera en que se construye una imagen del mundo (un proceso que es mucho más complejo y que no consideramos en este libro). Atravesamos un momento estimulante en el que se debate a nivel global sobre cuestiones fundamentales como el aprendizaje, el sentido del trabajo, la inteligencia o la creatividad. En gran parte estos planteos se han disparado por la irrupción del cambio tecnológico y de la concentración de recursos intelectuales, la orientación de flujos de capital y la inversión de recursos energéticos derivados hacia el enfoque de IA. Se ha tomado ese rumbo y esa es una cuestión que debemos reconocer. Nos estamos comunicando a nivel global y debatiendo acerca de cuestiones básicas y que afectan nuestra definición de lo humano, de las sociedades y de lo que implica aprender y crear. Una vez aceptada esta realidad, ¿qué podemos hacer? Un consejo: tratar de definir cómo nos situamos respecto de estos dispositivos a los que se da el nombre de IAs. Entre todo este "ruido comunicacional" puede resultarnos complicado discriminar cuáles son las opiniones a las que prestar atención. Otro consejo: descarten todo lo que tenga apariencia de ser material periodístico o de divulgación generado por IA porque será todo menos "inteligente". Se

trata por lo general de clickbait orientado a generar tráfico y carente en gran medida de valor cognoscitivo.

Uno de los puntos de partida para definir cuál es nuestro lugar como individuos o como instituciones que vamos a convivir con dispositivos inteligentes artificiales es tratar de entender dónde reside el saber o la inteligencia de los mismos. Interactuar con entidades que no llegamos a entender es desalentador. Por otra parte, el cúmulo de opiniones y de informaciones que circulan es abrumador y nos cuesta encontrar un eje que nos conduzca en el camino de la reflexión. Consejo: traten de pensar a partir de un eje conductor que les interese a ustedes, que los interpele y sigan ese camino de exploración. Al tratarse de una entidad con la que voy a establecer una interacción comunicacional, es clave definir cuál es el tipo de saber con el que cuenta y cómo lo produce. Esa es una de las cuestiones que a mí me interesan, porque enseñar y aprender han sido una constante desde muy pequeño. Dado que me incumbe y me resulta estimulante he destinado esfuerzos a explorar cuáles son los procesos que permiten a las IA generativas "producir" resultados que tengan, al menos, la apariencia de producciones de saber.

Mis primeras interacciones estuvieron destinadas a establecer sus procesos de inferencia y sus bases lógico matemáticas. Rápidamente quedó claro que ese proceso lógico, ese santo grial de inferencia no existía en ningún lugar específico del modelo. Interactuamos por medio de instrucciones textuales (prompts) y nos contesta con texto en un lenguaje natural inmediatamente comprensible. No obstante el proceso intermedio se genera en un contexto muy ajeno al razonamiento humano. ¿Dónde reside ese saber? ¿Qué es eso que pareciera que sabe? Las IA son un misterio y trato de dilucidar de qué se tratan. En nuestra interacción humano - máquina nos vinculamos comunicacionalmente con entidades cuyo saber no tiene una estructura de consistencia similar a la nuestra. Debía aceptar que interactuaba con millones de coeficientes en matrices que se multiplicaban y sumaban en espacios de dimensiones que no iba a lograr concebir. Hoy me parece algo muy lejano en el tiempo, pero en realidad no han pasado dos años. Prometí no incorporar material de 2024 pero voy a hacer una excepción e invitar a Simon Prince a nuestro debate. Se trata de un científico que combina su rigor técnico con preguntas muy penetrantes en torno a

cuestiones como el saber y la inteligencia. Su forma de comunicar elude las metáforas tan frecuentes en torno a superpoderes, o super inteligencias y otros términos que tanta confusión introducen en torno a estas tecnologías. Una de las cuestiones que más lo obsesiona es la pregunta acerca de dónde reside el saber, dónde reside el conocimiento en los modelos de redes neuronales generadoras. Su último libro "Understanding Deep Learning" (2024) puede dejar una impresión de ser un tratado rigurosamente técnico, pero su orientación va mucho más allá y deja pendiente la pregunta sobre cuál es el lugar que ocupa el saber, el conocimiento, el razonamiento cuando nos enfrentamos con estos grandes modelos de IA. Según su opinión no debemos buscar el saber o la razón en los parámetros del modelo, ni en alguna característica intrínseca del mismo, sino en la información, los datos con que lo hemos entrenado. Allí reside la clave de las respuestas, no en una mágica cualidad de superpoderes con que se comunican las IA en el entorno de la difusión global. No se dejen capturar por el entusiasmo o por el lenguaje de los superpoderes si no quieren evaluar las IA desde la narrativa de los comics books que inunda la esfera global comunicativa.

Continuemos con el tema del saber y la alucinación tan relevante para el mundo de la comunicación. Vayamos a una anécdota en la experiencia como educador. A poco de salir ChatGPT, un alumno soluciona una ficha domiciliaria a partir de este artilugio. Lo plano de la redacción ya nos daba una pista de la estrategia del joven a evaluar, pero lo que confirmaba su aplazo eran las citas de Bourdieu claramente inventadas por el chat. Utilizaba nociones propias del arsenal conceptual del autor, "sonaba" bastante a Bourdieu pero no lo era y eso le quedaba claro al ojo entrenado. No obstante, las frases eran totalmente asertivas, el modelo las enunciaba sin pudor. Esta es su manera de proceder con la seguridad de quien sabe. Un modelo conversacional de este tipo ni sabe ni deja de saber. El saber no tiene nada que ver en esta operatoria. Este fenómeno conocido como alucinación puede ser corregido haciendo del modelo algo más eficiente, es cuestión de técnica. No obstante, en términos reales sus respuestas siempre serán inhumanas. La elección de la palabra alucinación no ayuda. La utilización de este término es otra de las operaciones de difusión mediática y de los voceros de la industria que empañan la comprensión del alcance del modelo y que impactan en

la imaginación popular. Inteligencia y alucinación son nociones que contribuyen a antropomorfizar los modelos. Una respuesta de una AI generadora de lenguaje que corresponda a un dato verificable que damos por cierto en el universo fáctico no es ni más ni menos alucinatoria que otra que no lo haga. Ambas son predicciones, ambas se relacionan con patrones cuyo establecimiento y lógica nos resulta entendible hasta cierto nivel e incomprensible en otros (Prince, 2024). Volvemos a reiterar, estos artefactos operan por aproximación de funciones a partir de enormes masas de datos con que se los entrena. No alucinan ni dicen la verdad, operan según funciones matemáticas. Tampoco dialogamos con una IA, sino que interactuamos con ella. Si partimos de esta base estaremos en mejores condiciones para convivir con estos desarrollos de la técnica. Un consejo para mejorar nuestros juicios en torno a la IA: defendernos de los términos confusos que la rodean y del "efecto halo" que describe Thorndike en 1920 con relación a lo que hoy conocemos como sesgos cognitivos (Nisbett y Wilson, 1977).

El modelo podrá predecir con mayor o menor efectividad la palabra que nos conforme, podremos lograr que sus chistes sean más graciosos, sus poemas más ajustados a nuestros patrones de valoración estética, sus cuentas más acertadas, pero en el fondo nada de esto se relaciona con el sentido. Con un ente así no es posible conversar. Queda por resolver el tema de si en la comunicación humana el vacío de las palabras y la repetición del sin sentido es la regla o la excepción. Nada nos asegura que entre un ser humano y otro exista algo así como una comunicación de sentidos y no podamos considerar gran parte de la conversación cotidiana como una repetición mecánica de fórmulas convencionales sin un sustento por detrás como dirían algunos pensadores escépticos. Humanizar nuestra comunicación es un gran desafío.

Capacidad de agencia

Como queda claro en la definición de IA que presentamos en el primer capítulo, un agente inteligente es una entidad capaz de realizar acciones inteligentes de manera autónoma, que percibe su entorno y puede mejorar su desempeño a partir del aprendizaje o de la

incorporación de conocimientos (Russell y Norvig, 2004). Por lo general nos formamos una imagen de las IAs a partir de la interacción con ChatGPT, Claude, Stable Diffusion o Dall-e. También se puede acudir a otros desarrollos de base abierta, la cantidad de modelos actualmente disponibles es enorme y no para de crecer. Estas herramientas no llegan a cubrir la definición de agentes inteligentes de la misma manera en que se pregona la inteligencia en el caso de los seres humanos, ya que, por ejemplo, los chats conversacionales como ChatGPT, son simples interfaces de pregunta respuesta que generan la experiencia simulada de un diálogo. Nadie diría que toman decisiones autónomas para lograr sus objetivos. Nos indican, eso sí, cómo podríamos alcanzarlos. Nos puede guiar en la tarea de realizar un video educativo o preparar una torta de chocolate. En un sentido figurado, al menos por ahora, podríamos afirmar que nos utiliza de herramientas: nos indican qué dieta seguir para mantenernos fuertes, jóvenes y sanos y nosotros, como sus ejecutores, seguimos las instrucciones.

¿Cómo nos vinculamos con las IAs una vez que las concebimos como agentes inteligentes? Comencemos a pensar en un arco con cuatro interacciones básicas: humano – humano / humano – máquina / máquina – humano / máquina – máquina. Desde aquí se podría comenzar a evaluar el sentido de la aparición de artefactos que incrementan día a día su capacidad de agencia. Repetimos para no distraernos de lo fundamental: pensemos en agentes inteligentes antes que en IAs. Esta es una vía de reflexión que nos permitirá contemplar el impacto de estas tecnologías en el universo de las actividades humanas y el trabajo. Estemos preparados para el impacto. Considerar este grupo de interacciones y su dinámica puede ser una guía para una reflexión que se relacione con la comunicación, el sentido y nuestra capacidad de llevar adelante tareas en el mundo profesional y del trabajo. En el seno de este abanico me gustaría seguir enseñando cualquier disciplina a la que me dedique.

Por sí solos, los modelos de lenguaje basados en redes neuronales no funcionan como agentes. Sin embargo, las capacidades de los grandes modelos de lenguaje de operar con otras inteligencias artificiales periféricas, de operar dispositivos robóticos o de utilizar herramientas tales como calculadoras crecen constantemente. Estos modelos pueden

funcionar, y de hecho ya lo hacen, como coordinadores generales de acciones que ejecutan otras instancias. Los modelos de lenguaje aportan conocimiento y guía sobre los pasos a seguir y los dispositivos periféricos ejecutan las acciones. Por aquí avanza la capacidad de agencia de las IA generadoras de lenguaje a pasos acelerados. Para convertirse en una agente inteligente, los modelos deberían poder representar los objetivos, establecer planificaciones, ejecutar los pasos correspondientes, evaluar el resultado de manera crítica, aprender de la experiencia e incorporar conocimientos vinculados a la tarea y, a partir de todo ello incrementar sus capacidades en vistas a las próximas acciones que emprenda (Figura 3). Aún no llegamos a este punto en que se hayan integrado todos los pasos en un mismo modelo, pero la colaboración con agentes externos aporta un atajo.

Una cosa es poder realizar una acción y otra tomar la decisión de realizarla. Esto es clarísimo y no deja dudas. Incluso más allá de ello se sitúan el drive o empuje y el deseo. Este elemento energético está en la base de la inteligencia de los seres vivos que se implican desiderativamente en un ambiente que pretenden modificar. Esta es una de las primeras consideraciones de Piaget cuando desarrolla el tema de los fundamentos de la inteligencia (Piaget, 1991). Los seres vivos tenemos objetivos, algunos de ellos deliberados y conscientes y otros inscriptos en los patrones de acción de las poblaciones a las que pertenecemos como especies o culturas. Las computadoras no. Salvo que se quiera entrar en complejos argumentos filosóficos sobre qué es en realidad un propósito,

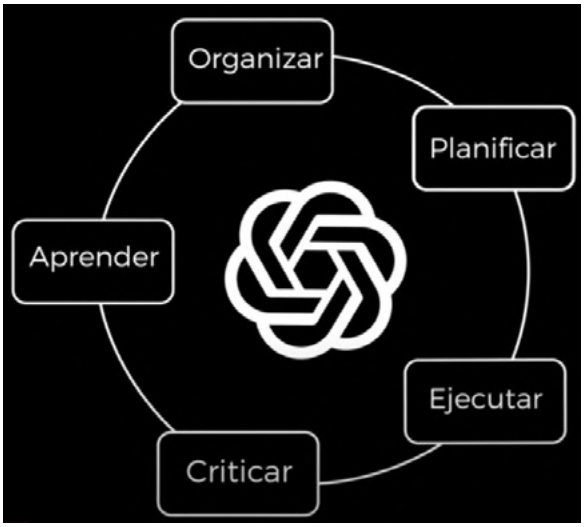


Figura 3. El gráfico muestra el encadenamiento de instancias que debería atravesar una IA para ser considerada agente inteligente. Captura de pantalla de: https://www.youtube.com/watch?v=ZwIuMd2nAWU

qué es la mente, el espíritu, la divinidad, el pan mentalismo o alguna otra construcción metafísica por el estilo, nos será difícil sostener que las IAs tengan objetivos propios.

Un momento ¿propios? Y aquí se nos presenta una cuestión. Puede que una IA no tenga un propósito propio, pero cuente con la capacidad de definir propósitos intermedios para cumplir un objetivo que nosotros le hemos programado. En el capítulo 1 introdujimos la advertencia de Russell de no orientar las IA en la realización de tareas estúpidas (Russell y Norvig, 2004). Estúpidas e incluso peligrosas, nunca subestimen la imprudencia de nuestra especie. Todo incremento significativo de saber y de dominio del entorno incrementa el riesgo, no hay que ser un genio para darse cuenta de ello.

Consideremos por un momento la cuestión de los objetivos finales y los objetivos intermedios. Pongamos el caso en que hemos establecido un objetivo y estos artilugios, dependiendo del diseño algorítmico, son muy sensibles al premio o al castigo (por ejemplo, en las funciones de coste). Podemos, por ejemplo, pedirle a una IA que nos ayude a programar un viaje y le damos acceso a nuestros recursos, consultará vuelos, hoteles, destinos, horarios, avisará a la persona que cuida a nuestras mascotas, escribirá mails avisando al trabajo y amigos, armará itinerarios, definirá dietas adecuadas a nuestras alergias o gustos culinarios, utilizará nuestros medios de pago y controlará que se nos envíen los documentos necesarios conectada con el correo. Le hemos delegado la tarea de ayudarnos con el viaje, de hecho, puede que no le hayamos aclarado qué y cómo debe resolverla, lo hará mejor o peor, eso no viene al caso y si hoy lo resuelve mal porque no hay suficiente desarrollo de algoritmos, ponemos las manos en el fuego de que mañana la industria lo resolverá. Auto GPT avanza en este sentido, aunque, hasta donde nos consta no haya llegado a un nivel de desarrollo como para un uso masivo "en el mundo real" según las evaluaciones (benchmarks) disponibles (Yang et al., 2023). Eso para comenzar y presentar el "lado bueno" con una IA bien alineada al propósito funcional humano.

Pero veamos el "lado malo". Existe un experimento mental llamado *paperclip apocalipsis* (Bostrom, 2014). Se basa en la idea de objetivos finales e intermedios. A una IA se le da la orden de que produzca la mayor

cantidad que le sea posible de ganchos de metal, clips, de esos con que sostenemos las hojas. Pensemos que estamos dentro de unos años en que sus capacidades de resolver problemas han escalado según las curvas de crecimiento actuales, es decir es súper inteligente. Como es tan inteligente logra definir todos los métodos para producir clips utilizando todos los recursos disponibles de la humanidad y el planeta. Supongamos también que, en algún momento, usando el sentido común, decidimos impedir que siga inundando el universo de clips. Pero se ha fijado ese objetivo estúpido y no hay forma de combatir su tozudez mecánica. Por otra parte, puede considerar que para llegar a su objetivo debe aumentar su control sobre nosotros y el resto suena todo horrible. Apocalipsis en ciernes. Por ahora es pura especulación. Sin embargo, la capacidad de deliberar, de trazar objetivos, de negociar con otra instancia en forma colaborativa o competitiva no para de crecer. No nos olvidemos que una de las formas más eficaces de entrenarlas ha sido ponerlas a jugar y lograr que desarrollen estrategias para sortear todo tipo de desafíos, desde el tatetí, pasando por el Go, programas de preguntas y respuestas, torneos de programación, torneos de navegación de drones, test de ingreso a universidades, Tetris, plegamientos de proteínas, olimpíadas de geometría del más alto nivel mundial o algoritmos de multiplicación de matrices, eso sin contar la experimentación que seguramente realiza la industria armamentista con todo lo que ello implica en escenarios competitivos. Estos artilugios, cuando tienen una función de coste y un premio, juegan bien. Les enseñamos a ganar. Incrementamos estas capacidades y habrá que ver en qué medida resulta prudente.

Una IA del tipo de los grandes modelos de lenguaje no debería pensarse como una entidad aislada. Su potencia se incrementa si se la vincula con otras instancias de IA especializadas que la utilizan como base para propósitos específicos. Como suele ocurrir la investigación a partir de juegos es un campo eficaz para experimentar desarrollos. En octubre de 2023 NVIDIA publicó un paper en que presenta una librería que permite utilizar los conocimientos y las capacidades de GPT4 para jugar Minecraft (Wang et al., 2023). Nos tenemos que detener en la enormidad que esto implica, jugar Minecraft no es cualquier cosa, implica altísimos niveles de estrategia, de deliberación de

objetivos y de planificación en entornos poco definidos y casi sin reglas o instrucciones de juego (Figura 4).

Los saltos que estamos presenciando a partir de los grandes modelos de lenguaje en tan poco tiempo con algoritmos que lo único que hacen es predecir la próxima palabra en un texto son pasmosos. La potencia de los patrones que subyacen no ya al lenguaje en sí mismo, sino a los textos que generamos los humanos a partir de él no deja de impactar. ¿Jugaron alguna vez a Minecraft, conocen a alguien que lo haya hecho? Si no lo jugaron pidan a sus amigas o amigos que les cuenten. A diferencia del ajedrez o el Go, o incluso competencias como Jeopardi, Minecraft no tiene un objetivo definido, ni siquiera existe un tablero lo que incrementa el grado de dificultad para un algoritmo de aprendizaje. Una de las mayores ventajas que tiene el aprendizaje en juegos con reglas es que se puede definir una función de coste, lo que traducido al lenguaje corriente implica que podemos indicarle al modelo cuando gana y cuando pierde y a la vez lo podemos premiar o castigar a partir de ello. En Minecraft las cosas son mucho más confusas. Los universos complejos, como podemos imaginar, fascinan a los desarrolladores de software y a sus socios matemáticos. Esta imagen parece coincidir con todos los prejuicios y

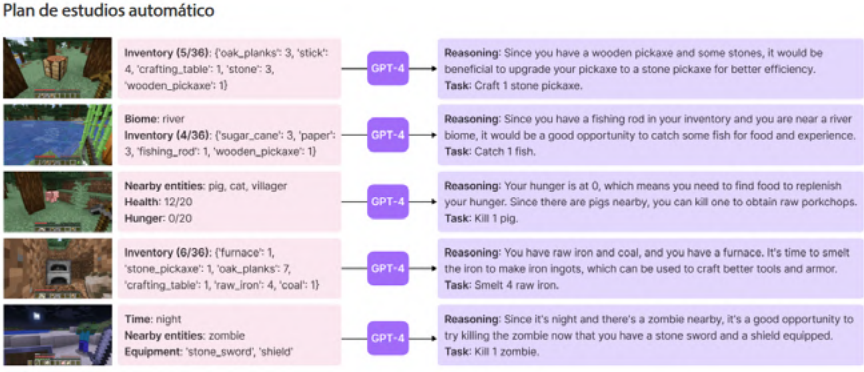


Figura 4. Currículum automático. El plan de estudios automático tiene en cuenta el progreso de la exploración y el estado del agente para maximizar la exploración. El plan de estudios es generado por GPT-4 basándose en el objetivo general de "descubrir tantas cosas diversas como sea posible". Este enfoque puede percibirse como una forma de búsqueda de novedades en contexto. Voyager: un agente encarnado abierto con grandes modelos de lenguaje. Guanzhi Wang NVIDIA. UT AUSTIN. STANFORD. CALTHEC. https://voyager.minedojo.org/

estereotipos mediáticos difundidos por Big Bang Theory. ¿Cómo es posible que Gpt4 participe en esta tarea de avanzar en el juego de Minecraft? Para comenzar recordemos que tiene información sobre muchas cosas, así se lo ha entrenado. Entre todo el material al que se ha visto expuesto seguramente estaba incluido no uno sino muchos tutoriales de Minecraft ya que la web está plagada de ellos. Si le pedimos que nos cuente paso a paso cómo llegar hasta los niveles superiores del juego (minar hasta conseguir tecnología de diamante), lo podrá hacer paso a paso. Pero Gpt4 no es un agente, es modelo de lenguaje capaz de interactuar de manera conversacional porque posee una capa extra destinada a comunicar el modelo con interlocutores humanos. En este punto deberá coordinarse con otro programa que sí pueda realizar cosas: con un agente como Voyager de NVIDIA (Wang et al., 2023). Lo interesante de ello es que Voyager no aprende a jugar, sino que se aprovecha del conocimiento del gran modelo de lenguaje que es Gpt4. La conexión entre IAs y todo tipo de programas y modelos periféricos es una vía de desarrollo y potenciación de los modelos de lenguaje que puede modificar la forma de hacer las cosas en un universo cada vez más tecnificado.

Aunque prometimos no incluir información de 2024 y de hecho obviamos el lanzamiento de SORA el generador de video de Open AI y de Gemini 1.5 con sus 10.000.000 de tokens en la ventana de contexto, no pudimos evitar la tentación de comentar lo que de hecho nos parece el enfoque más rupturista de los últimos tiempos, el del nuevo desarrollo de agentes dinámicos. De hecho, el enfoque de modelos fundacionales interactivos basados en agentes puede llegar a constituirse en una tendencia dominante, o al menos muy relevante en el desarrollo de las IAs en los próximos tiempos. Aunque no sea fácil de anticipar existen signos claros de que el desarrollo de agentes multi tarea revolucionará la manera en que nos vinculamos con estas tecnologías (Durante et al., 2024). Los esfuerzos combinados de NVIDIA y Microsoft empujan en esa dirección. Estos agentes contemplan varias características notables: son multipropósito, entrenados con una amplia variedad de *datasets*, mantienen una alta capacidad de captación espacial y movimiento dada la evolución de sistemas de entrenamiento por simulación, son multimodales con captación de lenguaje, imagen, video, pueden funcionar localmente en

robots (*IA embodiment*). Entendamos bien esto: se trata de modelos que se orientan a comprender el contexto del ambiente en que se los sitúa e interactuar con el mismo de manera inteligente. Nos encontraríamos con entidades que no solo captan el ambiente, sino que pueden interactuar con el mismo de manera significativa. Puede ser una de las vías que desemboquen en la AGI. Aun en el caso de que esto no suceda, el impacto en nuestras vidas y en la organización de nuestras sociedades puede llegar a ser descomunal.

Capacidades emergentes

Antes que nada, queremos avisar que la denominación "capacidades emergentes" está en el centro de una polémica encarnizada que no vamos a abordar y que vamos a mencionar lateralmente. No es totalmente claro a qué se refiere el término "emergente" ni el término "capacidad" usado en este contexto. Recomendamos tomar algún contacto con lo escrito por Molnar (2021) y por Prince (2024) antes de internarse en el tema. Dicho lo anterior, prosigamos.

Los seres vivos evolucionamos en una interacción adaptativa con el medio terrestre a lo largo de millones de años. La evolución puede ser pensada como un proceso caprichoso con generación de caminos que no conducen en definitiva hacia ningún resultado final necesario. La vida no es necesaria y tal vez sea improbable en un planeta relativamente hostil como el nuestro. En el lugar en que me encuentro estará el sol, aunque falta mucho aún para ello y cuando ocurra muy posiblemente ya no existan rastros de vida humana en la galaxia. En este largo proceso de la evolución aparecieron y desaparecieron crestas, exoesqueletos plumas y dientes, se alargaron o encogieron picos, los pelajes se aclararon u oscurecieron. La vida aprendió a percibir, a moverse, a reproducirse de diversas maneras, a construir refugios, ciudades, a resolver teoremas, a pintar, a viajar. También aparecieron y desaparecieron conductas, patrones de comunicación y ecosistemas. Salvo algunos cataclismos las cosas en el registro de la vida fueron relativamente lentas. Con la IA en modelos de lenguaje esto no ocurre, las capacidades emergen de manera rápida. Tratemos de imaginar espacios con dimensiones incalculables

con trillones de conexiones que identifican patrones que no podríamos ni siquiera concebir. Tal como decíamos en el caso de la memoria humana, no deberíamos imaginar estas capacidades como algo que "está" en algún lado, sino en algo que ocurre a partir de algún tipo de conexión o a partir de ondas de energía que recorren nuestro sistema nervioso. Ni siquiera terminamos de definir cuál es la lógica de nuestra actividad cerebral. Tampoco es seguro que debamos reducir todo a nuestros cerebros como tienden a sugerir los modelos cognitivos. Con las máquinas la incertidumbre es aún mayor. En cierta medida, al interior de esos modelos solo hay números ordenados en filas y columnas. Millones, trillones de ellos que se combinan en operaciones numéricas, con sumas y multiplicaciones. Hemos generado un mundo que nos es ajeno, con un lenguaje que no es el nuestro ni lo será jamás, existe una frontera que no se va a cruzar. Al menos en la actual arquitectura de los modelos generativos la interpretabilidad es una entidad mítica. De hecho, nuestra capacidad de explicar en profundidad el entendimiento humano no estaría en condiciones mucho mejores.

Sin ser preparados específicamente estos predictores de próxima palabra, estos aproximadores de funciones, han demostrado capacidades matemáticas, lógicas, de programación, de estructuración de relatos. Se les ha podido "regular la temperatura" para que actúen de manera humorística, formal, ajustada a lo esperable o creativa. Son, eso sí y por ahora, entidades limitadas que no alcanzan la llamada AGI que les permitiría igualar la capacidad humana en cualquier tipo de tarea. Su desarrollo es tal que pasan en gran medida el test de Turing (están muy cerca en porcentaje). Tal es así que Mustafa Suleyman, uno de los fundadores de Deep Mind, la que fuera usina de ciencia de Google propone otro test consistente en entregarle a una IA 100 mil dólares y darle un lapso para que lo convierta en 1.000.000 por sus propios medios de manera autónoma y sin indicaciones humanas (Suleyman y Bhaskar, 2023). Habría mucho que debatir sobre la vara que propone Suleyman para medir la inteligencia de estas entidades y las consecuencias sobre el destino de las sociedades e individuos, pero este no es el propósito del libro.

Lo cierto es que estos modelos que sólo pueden hacer una cosa: predecir la próxima palabra encierran misterios. Tenemos la impresión de que su desarrollo marcará la manera en que las sociedades trabajen, se

organicen y se comuniquen de aquí en más. Conviviremos con ellas no parece haber vuelta atrás. El destino de esta sociedad entre humanos y máquinas queda por verse.

Bibliografía

Ananthaswamy, A. (2023). A new approach to computation reimagines artificial intelligence. https://www.quantamagazine.org/a-new-approach-to-computation-reimagines-artificial-intelligence-20230413/

Apostel, L. (1994). Construcción y validación en la epistemología contemporánea. En Piaget, J. (Ed.), *Construcción y validación de las teorías científicas: contribución de la epistemología genética* (pp. 100-136). Paidós.

Aristóteles (1982). *Tratados de lógica (Organon)*. Gredos.

Aristóteles (2002). *Retórica*. Alianza Editorial.

Bostrom, N. (2014). *Superinteligencia: caminos, peligros, estrategias*. OUP.

Chomsky, N. (2002). *Syntactics Structures*. Mouton de Gruyter.

Durante, Z., B. Sarkar, R. Gong, R. Taori, Y. Noda, P. Tang, E. Adeli, S. K. Lakshmikanth, K. Schulman, A. Milstein, D. Terzopoulos, A. Famoti, N. Kuno, A. Llorens, H. Vo, K. Ikeuchi, L. Fei-Fei, J. Gao, N. Wake y Q. Huang. (2024) An interactive Agente Foundation Model. arXiv:2402.05929v1 [cs.AI].

Festinger, L. (1975). *Teoría de la disonancia cognoscitiva*. Instituto de Estudios Políticos.

Gemini Team, Google1. (2023) Gemini: A Family of Highly Capable Multimodal Models. arXiv:2312.11805v1 [cs.CL] 19 Dec 2023

Molnar, C. (2021). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Leanpub.

Najdenkoska, I., X. Zhen, y M. Worring (2023). Meta Learning to Bridge Vision and Language Models for Multimodal Few-Shot Learning. ArXiv, abs/2302.14794.

Nisbett, R. E. y T. D. Wilson (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*(4), 250-256. https://doi.org/10.1037/0022-3514.35.4.250

Perelman, C., J. Sevilla Muñoz y L. Olbrechts-Tyteca (1989). *Tratado de la argumentación. La nueva retórica*. Gredos.

Piaget, J. (1991). *Psicología de la inteligencia*. Siglo veinte.

Russell, S.J. y P. Norvig (2004). *Inteligencia artificial. Un enfoque moderno*. Pearson Educación, S.A.

Suleyman, M. y M. Bhaskar (2023). *La ola que viene: Tecnología, poder y el gran dilema del siglo XXI*. Edición Kindle.

Thorndike, E. (1920). A constant error on psychological rating. *Journal of Applied Psychology, 4*, 25-29.

Uesato, J., N. Kushman, R. Kumar, F. Song, N. Siegel, L. Wang, A. Creswell, G. Irving y I. Higgins (2022). Solving math word problems with process- and outcome-based feedback. ArXiv, abs/2211.14275.

Van Dijk, T. (2002). El análisis crítico del discurso y el pensamiento social. *Atenea Digital, 1*. https://doi.org/10.5565/rev/athenead/v1n1.22

Yang, H., S. Yue e Y. He (2023). Auto-GPT for Online Decision Making: Benchmarks and Additional Opinions. ArXiv, abs/2306.02224.

Wan, Z., Ch. K. Liu, H. Yang, C. Li, H. You, Y. Fu, Ch. Wan, T. Krishna, Y. Lin y A. Raychowdhury (2024). Toward Cognitive AI Suystems: a Survey and Prospective on Neuro_Symbolic AI.

Wang, G., Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan y A. Anandkumar (2023). Voyager: An Open-Ended Embodied Agent with Large Language Models. ArXiv, abs/2305.16291.

YouTube – Podcast - Blogs

BLOG Wei, J. Los modelos de lenguaje realizan el razonamiento a través de una cadena de pensamiento. MIÉRCOLES, 11 DE MAYO DE 2022. *Publicado por Jason Wei y Denny Zhou, científicos investigadores, Google Research, equipo Brain*

Deep learning is a strange beast. Machine Learning Street talk with Simon Prince. January 2024. https://www.youtube.com/watch?v=sJXn4Cl4oww&t=2631s

El Autor

Marcelo Babio es Doctor en Ciencias Sociales por la UNLP, Licenciado en Psicología de la UBA y codirector del Núcleo de Investigaciones Científicas Estudios de Comunicación y Cultura en Olavarría (ECCO) perteneciente a la Facultad de Ciencias Sociales de la UNICEN.

Su trabajo de investigación académica en UNICEN se centra en el monitoreo y evaluación de las transformaciones en el ecosistema comunicacional a partir de la evolución de las tecnologías de la información y la inteligencia artificial.

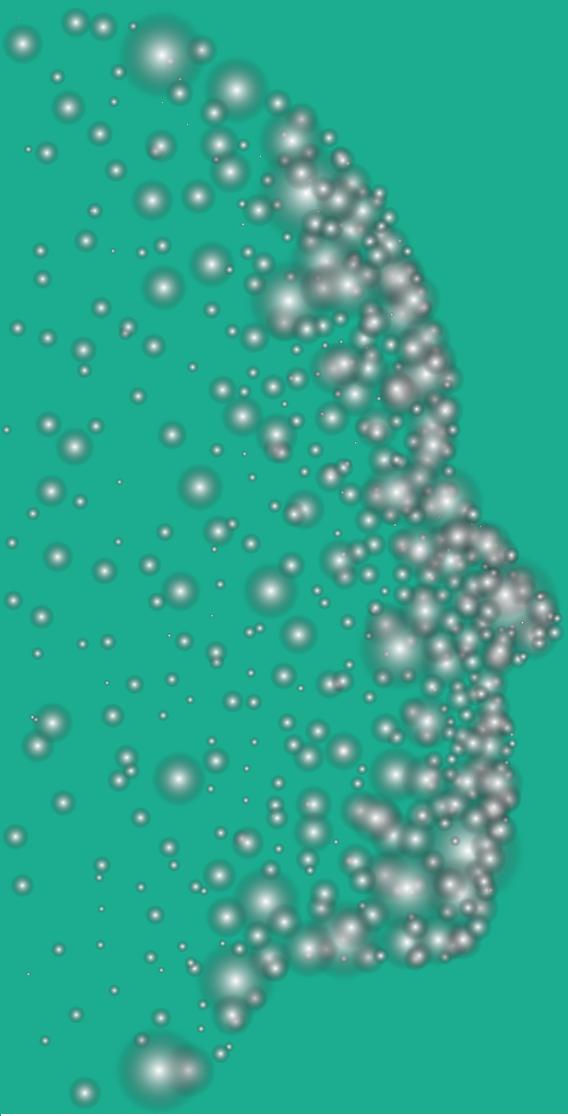
Profesor titular en UNICEN de Psicología y Comunicación – Comunicación Publicitaria y Marketing.

A cargo de los Seminarios: Abordaje de la inteligencia artificial: guía para una apropiación consciente de la herramienta, dictados en UNICEN y Facultad de Ciencias Sociales de la UBA durante 2024.

Investigador de mercado en consultoras en Argentina, Latinoamérica y el Caribe para marcas líderes en diferentes verticales de negocios. Especialización en mercados tecnológicos y de comunicación.

E-mail: marcelo.babio@gmail.com

Linkedin: Marcelo Babio



Universidad Pública
Argentina

Sociales
Facultad de Ciencias Sociales UNICEN

UNICEN
Universidad Nacional del Centro
de la Provincia de Buenos Aires