Vol. 2. Número 9 (Septiembre, 2022)

EL FUTURO DE LAS COMPUTADORAS

Por Dra Estalina Baéz-Ramírez

Es probable que estés leyendo esto en una computadora. También es probable que estés dando ese hecho por sentado. Eso es a pesar de que el dispositivo frente a usted habría asombrado a los científicos informáticos hace solo unas décadas, y parecía pura magia mucho antes de eso. Contiene miles de millones de diminutos elementos informáticos, que ejecutan millones de líneas de instrucciones de software, escritas colectivamente por innumerables personas en todo el mundo. Finalmente hace clic, toca, escribe o habla, y el resultado aparece perfectamente en la pantalla. Las computadoras una vez llenaron las habitaciones. Ahora están en todas partes y son invisibles, incrustadas en relojes, motores de automóviles, cámaras, televisores y juguetes. Gestionan redes eléctricas, analizan datos científicos y predicen el tiempo. El mundo moderno sería imposible sin estas, y nuestra dependencia de estas para la salud, la prosperidad y el entretenimiento solo aumentará.

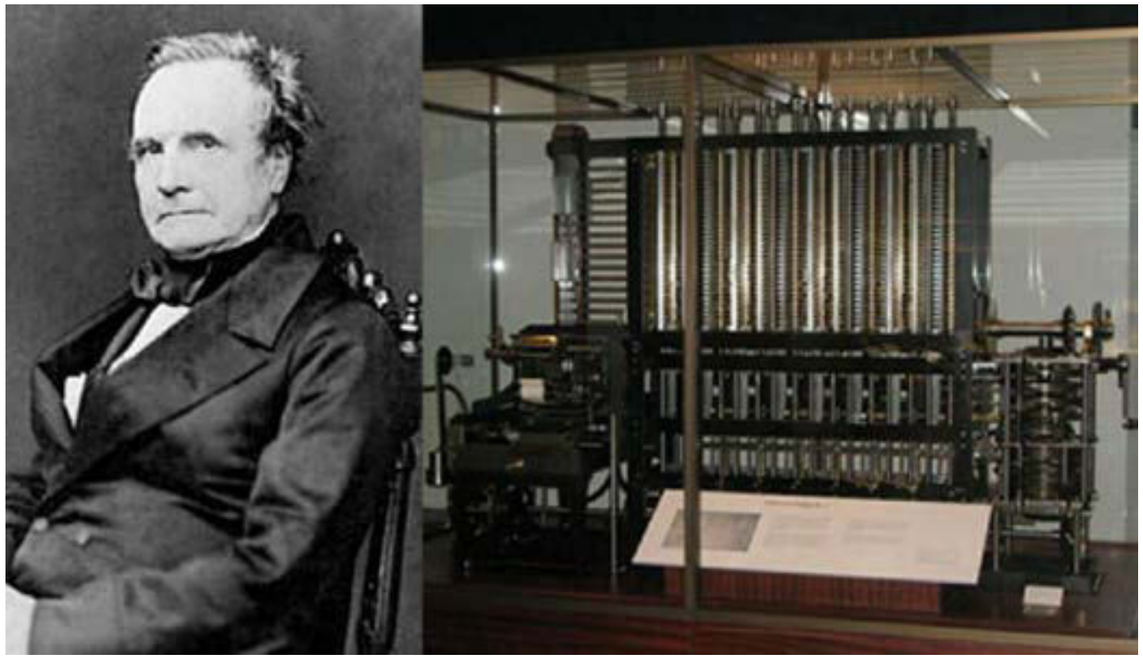

Figura 1. Charles Babbage, es el precursor de la computadora. Babbage trabajó en dos máquinas mecánicas: La Máquina de Diferencias que hoy en día puede verse en el Museo de la Ciencia de Londres, y una mucho más ambiciosa: La Máquina Analítica, la cual puede considerarse así como el auténtico precursor de los computadores digitales modernas o la calculadora. El prototipo de Máquina Diferencial que construyó en 1821, con capacidad para resolver polinomios de segundo grado, Imaginó una máquina programable con un "almacenamiento" para almacenar números, un "molino" para operar con ellos (se muestra), un lector de instrucciones y una impresora.

Los científicos esperan hacer que las computadoras sean aún más rápidas, que los programas sean más inteligentes y que se implemente la tecnología de manera ética. Pero antes de ver a dónde vamos desde aquí, repasemos de dónde venimos.

En 1822, el matemático inglés **Charles Babbage** diseñó la máquina diferencial, basada en el método de Newton o de derivadas divididas, la cual podía calcular de manera automática funciones polinomiales (algo totalmente sin precedentes) (**Fig. 1**). El primer diseño se estima requería 25,000 piezas con un peso de 4 toneladas, sin embargo por problemas económicos y mano de obra no se pudo concretar representando perdidas millonarias para el principal inversionista, el gobierno británico (***Romano J, 2014***).

A partir de esa pieza de ingeniería casi perfecta, surgió la idea de una maquina aún más ambiciosa, un dispositivo que realizara cálculos con propósito general y además que fuera programable. Esta máquina analítica, que es el vínculo entre las máquinas de cálculos aritméticos mecanizados con cálculos de propósito general. Al trabajar en colaboración con **Ada Lovelace (1833)**, quien es una de las precursoras de la programación, luego de estudiar la obra de Menabrea, detalla y elabora entre varios resultados una descripción de como dicha máquina podría ser programada para computar números de Bernoulli con rigor y excelencia: "Tejiendo patrones algebraicos exactamente como el telarde la Naturaleza teje flores y hojas" ***(Galera MCS, 1994***).

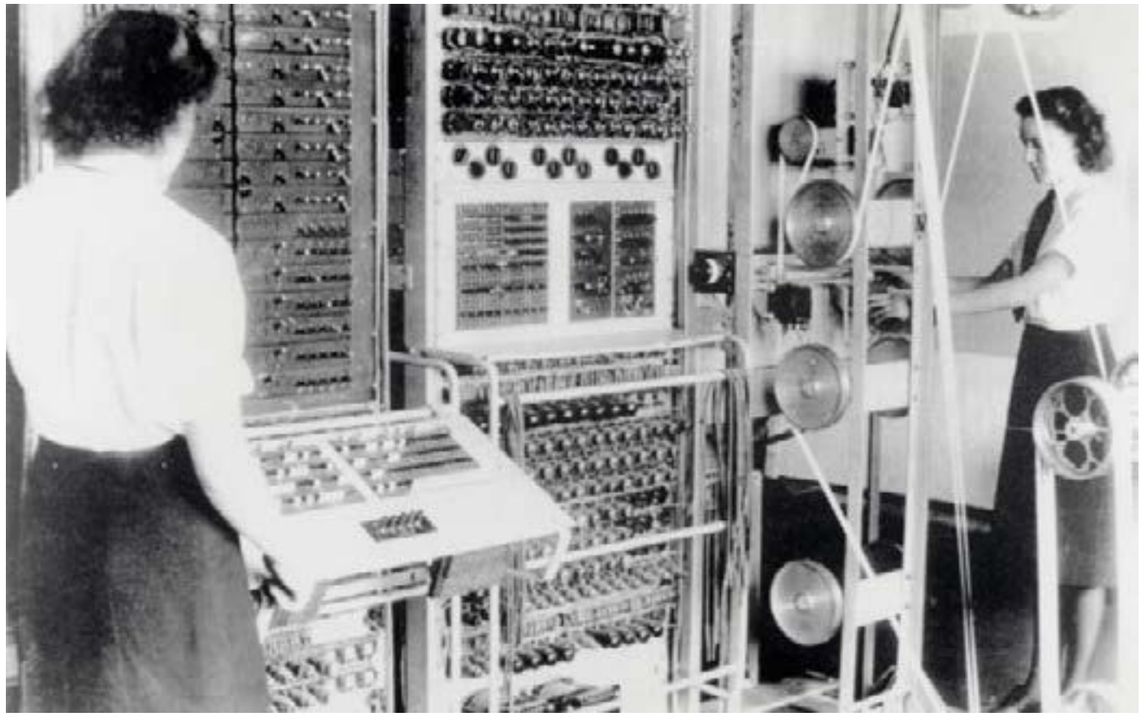

Figura 2. Colossus fue la primera computadora digital electrónica confiable del mundo. Completado en 1943, fue utilizado por las fuerzas de inteligencia británicas para descifrar el código durante la Segunda Guerra Mundial. SSPL/IMÁGENES FALSAS

Estos precursores y otros programadores lograron especificar una serie de pasos que la máquina tendría que seguir con el fin de resolver un problema, dando así la pauta para lo que hoy conocemos como programa. El diseño de la maquina analítica pretendía resolver las operaciones aritméticas básicas, operaciones de comparación y raíces cuadradas de números decimales de hasta 40 dígitos de longitud, alimentados mediante tarjetas perforadas e incluso con posibilidad de ser almacenados; el resultado de las operaciones se obtenía mediante un dispositivo tipo impresora. En la actualidad este dispositivo está catalogado como uno de los más grandes éxitos intelectuales del siglo (***Romano J, 2014***).

En 1936, el matemático inglés **Alan Turing** considerado el precursor de la informática moderna, proporcionó una influyente formalización de los conceptos de algoritmo y computación: la máquina de Turing, como aun algunos reconocen a las computadoras modernas. Formuló su propia versión de la hoy ampliamente aceptada tesis de Church-Turing. Introdujo la idea de una computadora que pudiera reescribir sus propias instrucciones, haciéndola infinitamente programable. Su abstracción matemática podría, utilizando un pequeño vocabulario de operaciones, imitar una máquina de cualquier complejidad, lo que le valió el nombre de "***máquina universal de Turing***".

La primera computadora digital electrónica confiable, **Colossus**, se completó en 1943 para ayudar a Inglaterra a descifrar los códigos de guerra (**Fig. 2**). Usó tubos de vacío, dispositivos para controlar el flujo de electrones, en lugar de piezas mecánicas móviles como las ruedas dentadas del motor analítico. Esto hizo que **Colossus** fuera

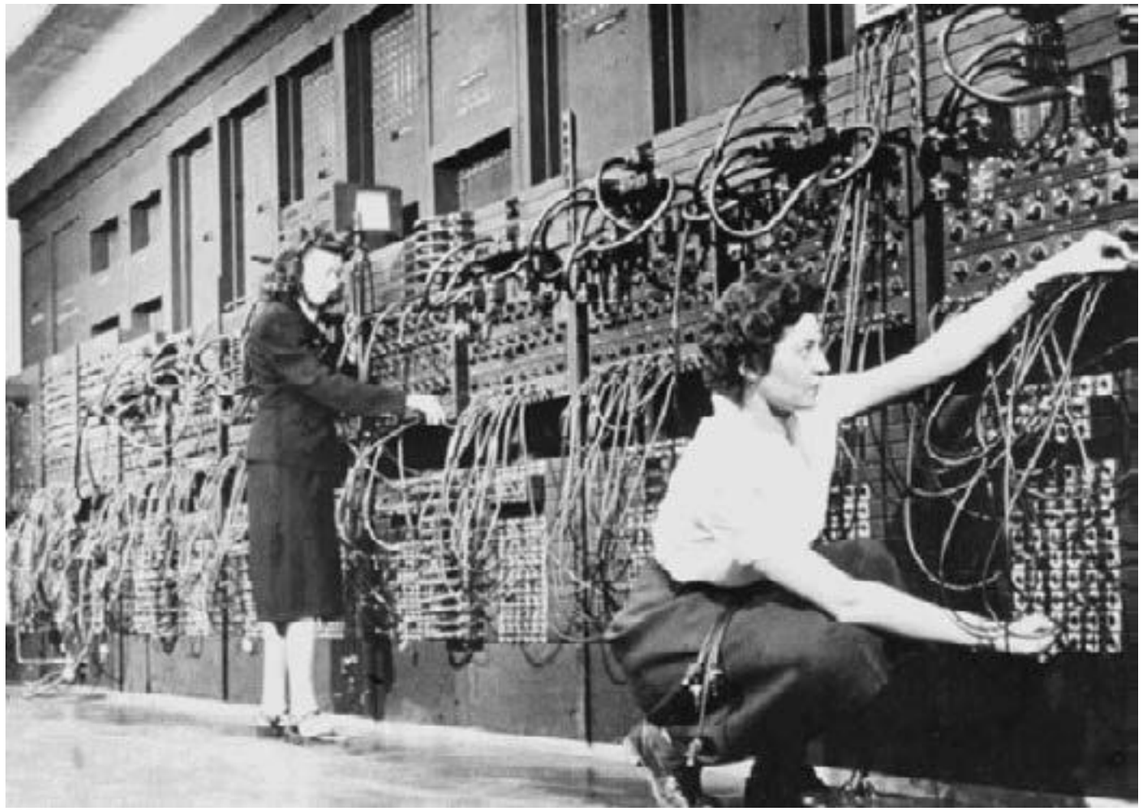

Figura 3. La primera computadora digital de propósito general totalmente electrónica, la ENIAC, fue presentada por el Ejército de los EE. UU. en 1946. Demasiado tarde para desempeñar un papel en la Segunda Guerra Mundial, los cálculos de la ENIAC ayudaron en el diseño de la bomba de hidrógeno. AGENCIA DE INFORMACIÓN DE EE. UU./ARCHIVOS NACIONALES/IDENTIFICADOR ARC: 594262

rápido, pero los ingenieros tenían que recablearlo manualmente cada vez que querían realizar una nueva tarea. Posteriomente surgió la primera computadora digital de propósito general totalmente electrónica, la ENIAC (**Fig. 3**).

Tal vez inspirado por el concepto de Turing de una computadora más fácilmente reprogramable, el matemático **John von Neumann,** quien escribió el diseño de EDVAC (***ElectronicDiscrete Variable AutomaticComputer***) en 1945, describió un sistema que podía almacenar programas en su memoria junto con datos y alterar los programas, una configuración que ahora se llama *arquitectura de von Neumann.*

Durante mucho tiempo, solo los expertos podían programar computadoras. Luego, en 1957, IBM lanzó **FORTRAN** (***FORMULA TRANSLATION***), que inició como un esfuerzo de traducir un lenguaje de fórmulas al lenguaje de ensamblador y por ende al lenguaje de máquina, un lenguaje de programación que era mucho más fácil de entender (***López, C. A. 2008***). Todavía está en uso hoy. En 1981, la compañía presentó IBM PC y Microsoft lanzó su sistema operativo llamado MS-DOS, expandiendo juntos el alcance de las computadoras en los hogares y las oficinas. Apple personalizó aún más la informática con los sistemas operativos para su Lisa, en 1982, y Macintosh, en 1984 (**Fig. 4**). Ambos sistemas popularizaron las interfaces gráficas de usuario, o GUI,

Vol. 2. Número 9 (Septiembre, 2022)

Figura 4. El cofundador de Apple, Steve Jobs, aparece en 1984 con las computadoras personales Macintosh originales. Estas computadoras popularizaron las interfaces gráficas de usuario, o GUI, que permiten a los usuarios hacer clic y arrastrar iconos en lugar de escribir una línea de comandos. MICHAEL L. ABRAMSON/GETTY IMAGES

ofreciendo a los usuarios un cursor de mouse en lugar de una línea de comando (***López, C. A. 2008***).

Mientras tanto, los investigadores habían estado haciendo un trabajo que terminaría conectando nuestro hardware y software novedosos. En 1948, el matemático **Claude Shannon** publicó "Una teoría matemática de comunicación" (*A MathematicalTheory of Communication*) un artículo que popularizó la palabra bit (por dígito binario) y sentó las bases de la teoría de la información. Sus ideas han dado forma a la computación y, en particular, al intercambio de datos a través de cables y por el aire (***Pérez de Lama, 2018***). En 1969, la Agencia de Proyectos de Investigación Avanzada de EE. UU. creó una red informática llamada ARPANET (**Fig. 5**), que luego se fusionó con otras redes para formar Internet (***Sain, 2015***).

En 1990, los investigadores del CERN, un laboratorio europeo cerca de Ginebra, Suiza, desarrollaron reglas para transmitir datos que se convertirían en la base de la **World Wide Web** (www), es lainiciativa de un proyecto práctico diseñado para crear un universo de información global utilizando toda la tecnología disponible (***Berners y cols. 1992***).

Persiguiendo inteligencia

Desde los primeros días de la informática, los investigadores se han propuesto replicar el pensamiento humano.Alan Turing abrió un artículo de 1950 titulado " Maquinaria informática e inteligencia" con: "Propongo considerar la pregunta: '¿Pueden las máquinas pensar?'". Procedió a esbozar una prueba, a la que llamó "el juego de imitación", en el que un

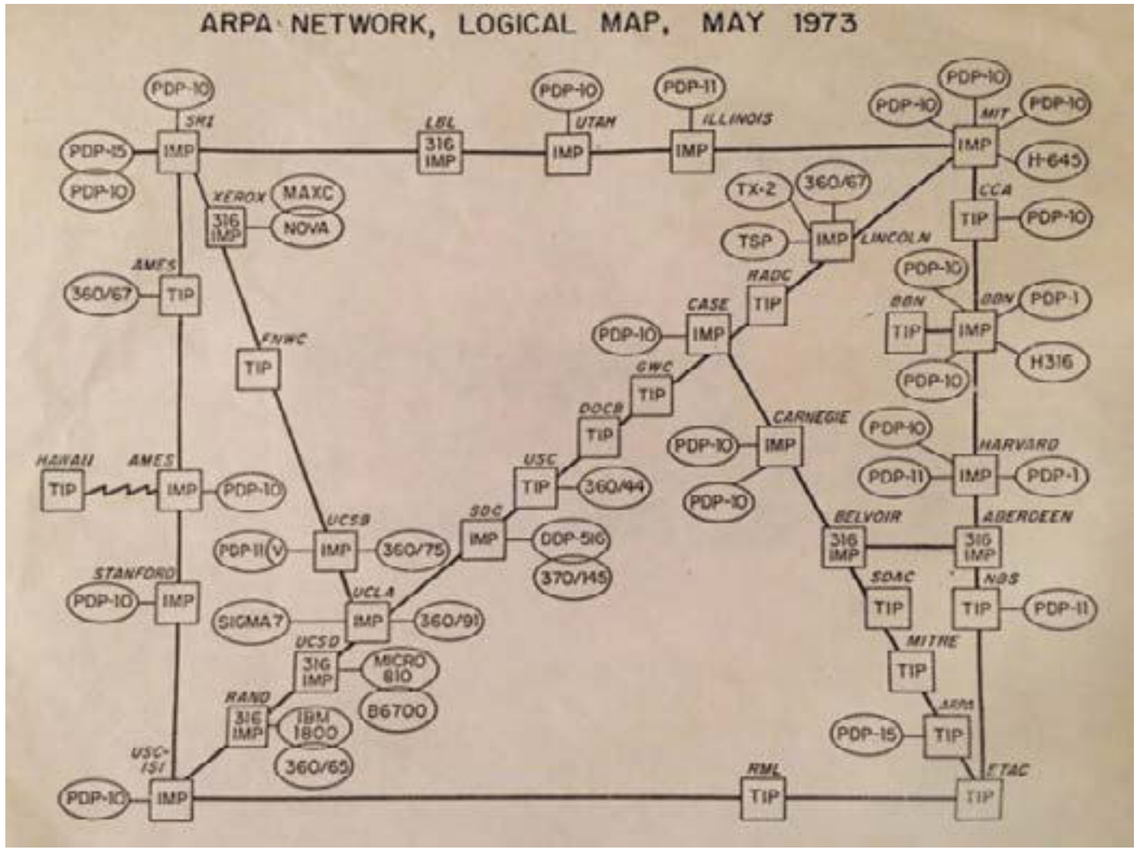

Figura 5. ARPANET, creada en 1969 (se muestra un mapa inicial), luego se fusionaría con otras redes para convertirse en Internet. ARPANET/WIKIMEDIA COMMONS

humano que se comunicaba con una computadora y otro humano a través de preguntas escritas tenían que juzgar cuál era cuál. Si el juez fallaba, la computadora presumiblemente podría pensar (Turing, 1950).

Definir la Inteligencia Artificial (IA) no es fácil, ya que el concepto de inteligencia per se no es del todo preciso. En términos coloquiales, IA se usa cuando una máquina es capaz de imitar las funciones cognitivas propias de la mente humana, como: creatividad, sensibilidad, aprendizaje, entendimiento, percepción del ambiente y uso del lenguaje (***Ocampo y Santa Catarina, 2018***).

Desde que se introdujo el término ***Inteligencia Artificial*** (IA) en 1956 por **John McCarthy** en la gran conferencia realizada en la universidad de Dartmouth College junto a Claude Shannon y Marvin Minsky entre otros, se formuló rápidamente la pregunta que ya cuestionaban estos visionarios y que tanto nos inquietó posteriormente ¿Podrían las máquinas pensar como los seres humanos con todo lo que esto conlleva? (***Mederos, 2017***).

Más de seis décadas e innumerables horas-persona después, no está claro si los avances están a la altura de lo que se tenía en mente en esa cumbre de verano. La inteligencia artificial nos rodea, de formas invisibles (filtrando spam), dignas de titulares (ganándonos al ajedrez, conduciendo automóviles) y en el medio (permitiéndonos chatear con nuestros teléfonos inteligentes). Pero todas estas

Vol. 2. Número 9 (Septiembre, 2022)

son formas limitadas de IA, que realizan bien una o dos tareas. Lo que Turing y otros tenían en mente se llama ***inteligencia general artificial*** (IGA). Dependiendo de su definición, es un sistema que puede hacer la mayor parte de lo que hacen los humanos. La IA ha avanzado mucho en la última década, en gran parte debido al aprendizaje automático. Anteriormente, las computadoras dependían más de la IA simbólica, que utiliza algoritmos o conjuntos de instrucciones que toman decisiones de acuerdo con reglas especificadas manualmente. Los programas de aprendizaje automático, por otro lado, procesan datos para encontrar patrones por sí mismos. Una forma utiliza redes neuronales artificiales, software con capas de elementos informáticos simples que, en conjunto, imitan ciertos principios de los cerebros biológicos. Las redes neuronales con varias o muchas más capas son actualmente populares y constituyen un tipo de aprendizaje automático llamado aprendizaje profundo.

Los sistemas de aprendizaje profundo ahora pueden jugar juegos como el ajedrez incluso mejor que el mejor humano. Probablemente

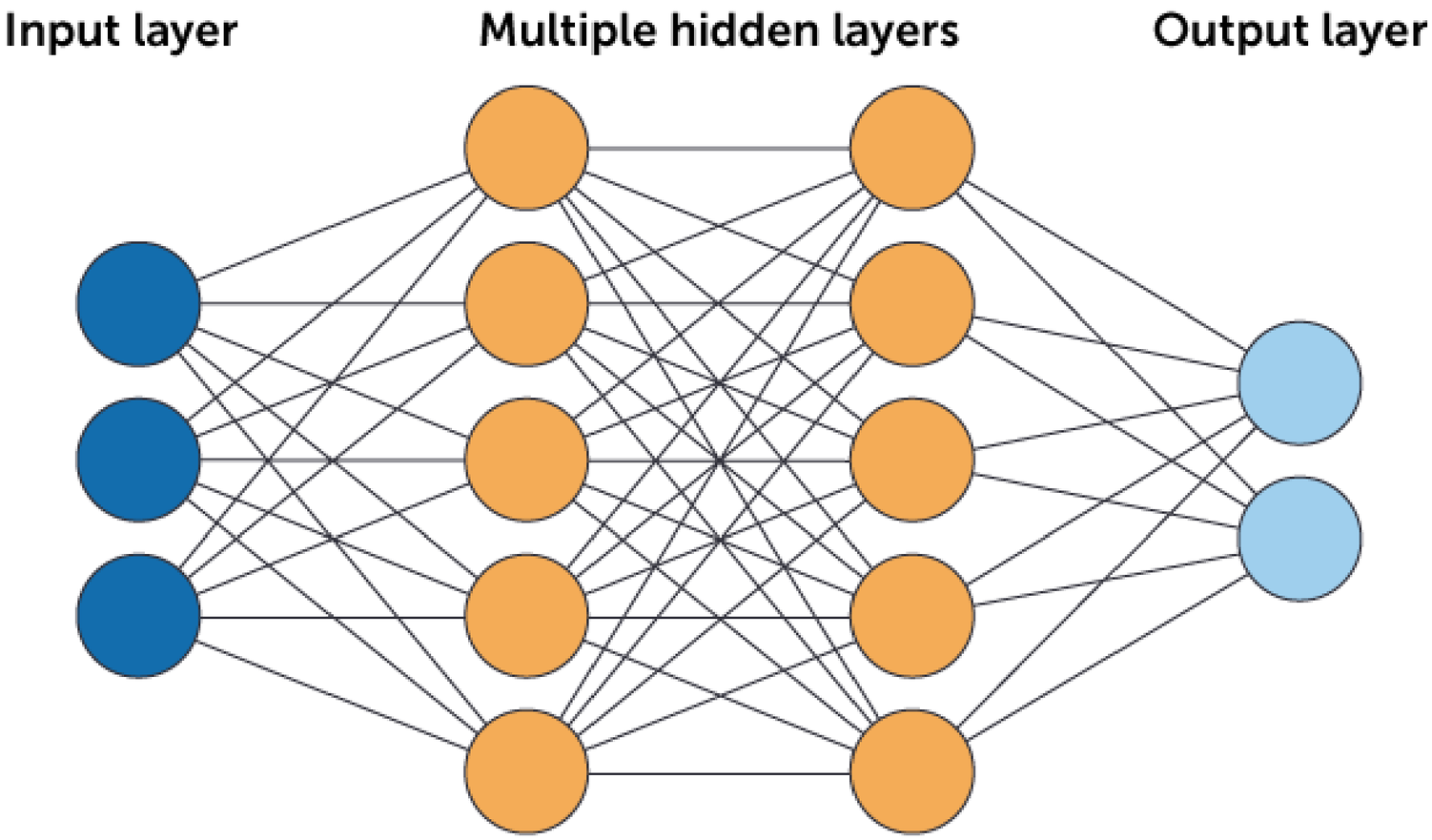

Figura 6. Los sistemas de aprendizaje profundo manipulan datos para extraer información de una manera que intenta imitar el cerebro humano. Tienen una capa de entrada, múltiples capas ocultas que transforman los datos de entrada y una capa de salida. E.OTWELL

puedan identificar razas de perros a partir de fotos mejor que tú. Pueden traducir texto de un idioma a otro. Pueden controlar robots y componer piezas musicales y predecir cómo se plegarán las proteínas.

TIPOS DE APRENDIZAJE

¿Cómo puede mejorar la IA? Los informáticos están aprovechando múltiples formas de aprendizaje automático, ya sea que el aprendizaje sea "profundo" o no (**Fig. 6**). Una forma común se llama aprendizaje supervisado, en el que los sistemas o modelos de aprendizaje automático se entrenan al recibir datos etiquetados, como imágenes de perros y sus nombres de raza. Pero eso requiere mucho esfuerzo humano para etiquetarlos. Otro enfoque es el aprendizaje no supervisado o autosupervisado, en el que las computadoras aprenden sin depender de etiquetas externas, de la misma manera que usted o yo predecimos cómo se verá una silla desde diferentes ángulos mientras caminamos alrededor de ella. Los modelos que procesan miles de millones de palabras de texto, prediciendo la siguiente palabra de una en una y cambiando ligeramente cuando se equivocan, se basan en el aprendizaje no supervisado. Luego pueden generar nuevas cadenas de texto. En 2020, el laboratorio de investigación **OpenAI** lanzó un modelo de lenguaje entrenado llamado GPT 3 (*Modelo de Preentrenamiento Generativo*) que es quizás la red neuronal más compleja de la historia. Basado en indicaciones, puede escribir artículos de noticias, cuentos y poemas similares a los humanos. Puede responder preguntas de trivia, escribir código de computadora y traducir lenguaje, todo sin estar específicamente capacitado para hacer ninguna de estas cosas. Está más avanzado en el camino hacia AGI de lo que muchos investigadores pensaban que era posible actualmente. Y los modelos de lenguaje se harán más grandes y mejores a partir de aquí (***Castro 2021***).

Otro tipo de aprendizaje automático es el aprendizaje por refuerzo en el que un modelo interactúa con un entorno, explorando secuencias de acciones para lograr un objetivo. El aprendizaje por refuerzo ha permitido que la IA se vuelva experta en juegos de mesa como Go y videojuegos como StarCraftII. Un artículo reciente de investigadores de DeepMind, incluido Precup, argumenta en el título que "La recompensa es suficiente". Con solo tener un algoritmo de entrenamiento que refuerce el comportamiento exitoso o semi-exitoso de un modelo, los modelos construirán gradualmente todos los componentes de inteligencia necesarios para tener éxito en la tarea dada y muchas otras.

Por ejemplo, según el artículo, un robot recompensado por maximizar la limpieza de la cocina eventualmente aprendería "percepción (para diferenciar utensilios limpios y sucios), conocimiento (para comprender los utensilios), control motor (para manipular utensilios), memoria (para recordar ubicaciones de utensilios), el lenguaje (para predecir el desorden

Vol. 2. Número 9 (Septiembre, 2022)

futuro a partir del diálogo) y la inteligencia social (para animar a los niños pequeños a ensuciar menos)". Está por determinarse si la prueba y el error conducirían a tales habilidades dentro de la vida útil del sistema solar, y qué tipo de objetivos, entorno y modelo se requerirían (***Silver y cols. 2021***).

Otro tipo de aprendizaje implica estadísticas bayesianas, una forma de estimar qué condiciones son probables dadas las observaciones actuales. Las estadísticas bayesianas están ayudando a las máquinas a identificar relaciones causales, una habilidad esencial para la inteligencia avanzada.

Turing también diferenció entre inteligencia general e inteligencia humana. En su artículo de 1950 sobre el juego de la imitación, escribió: "¿No pueden las máquinas llevar a cabo algo que debería describirse como pensar pero que es muy diferente de lo que hace un hombre?" (***Matthew Hutson, 2021***).

CUESTIONES ÉTICAS

En el cuento de 1942 "Runaround", uno de los personajes de Isaac Asimov enumeró "las tres reglas fundamentales de la robótica, las tres reglas que están más profundamente integradas en el cerebro positrónico de un robot". Los robots evitaban causar o permitir daño a los humanos, obedecían órdenes y se protegían a sí mismos, siempre que seguir una regla no entrara en conflicto con los decretos anteriores ***(Asimov, I. 1941***).

Podríamos imaginarnos los "cerebros positrónicos" de Asimov tomando decisiones autónomas sobre el daño a los humanos, pero en realidad no es así como las computadoras afectan nuestro bienestar todos los días. En lugar de robots humanoides que maten personas, tenemos algoritmos que seleccionan fuentes de noticias. A medida que las computadoras se infiltren más en nuestras vidas, tendremos que pensar más sobre qué tipos de sistemas construir y cómo implementarlos, así como metaproblemas como cómo decidir, y quien debería decidir, estas cosas.

Este es el ámbito de la ética, que puede parecer distante de la supuesta objetividad de las matemáticas, la ciencia y la ingeniería. Pero decidir qué preguntas hacer sobre el mundo y qué herramientas construir siempre ha dependido de nuestros ideales y escrúpulos. Estudiar un tema abstruso como las entrañas de los átomos, por ejemplo, tiene una clara relación tanto con la energía como con el armamento. "Existe el hecho fundamental de que los sistemas informáticos no tienen un valor neutral", dice **Barbara Grosz**, científica informática de la Universidad de Harvard, "que cuando los diseñas, incorporas un conjunto de valores a ese diseño" .Un tema que ha recibido mucha atención por parte de científicos y expertos en ética es la equidad y el sesgo. Cada vez más, los algoritmos informan o incluso dictan decisiones sobre contratación, admisiones universitarias, préstamos y libertad condicional. Incluso si discriminan menos que las personas, aún pueden tratar a ciertos grupos de manera injusta, no por diseño, sino a menudo porque están entrenados con datos sesgados. Podrían predecir el comportamiento delictivo futuro de una persona basándose en arrestos anteriores, por ejemplo, aunque diferentes grupos son arrestados a diferentes tasas por una determinada cantidad de delitos.

Otra preocupación es la privacidad y la vigilancia dado que las computadoras ahora pueden recopilar y clasificar información sobre su uso de una manera que antes era inimaginable. Los datos sobre nuestro comportamiento en línea pueden ayudar a predecir aspectos de nuestra vida privada, como la sexualidad. El reconocimiento facial también puede seguirnos por el mundo real, ayudando a la policía o a los gobiernos autoritarios. Y el campo emergente de la neurotecnologia ya está probando formas de conectar el cerebro directamente a las computadoras. La seguridad está relacionada con la privacidad: los piratas informáticos pueden acceder a datos que están bloqueados o interferir con marcapasos y vehículos autónomos.

Figura 7. Los drones de ataque no tripulados que pueden usar inteligencia artificial para identificar objetivos, como el Kargu-2 fabricado por la empresa turca STM, han planteado preocupaciones sobre la ética de las armas autónomas. MEHMET KAMAN/AGENCIA ANADOLU A TRAVÉS DE GETTY IMAGES

Las computadoras también pueden permitir el engaño. La IA puede generar contenido que parece real. Los modelos de lenguaje pueden escribir obras maestras o usarse para llenar internet con noticias falsas y material de reclutamiento para grupos extremistas. Redes antagónicas generativas, un tipo de aprendizaje profundo que puede generar contenido realista, puede ayudar a los artistas o crear deepfakes, imágenes o videos que muestran a personas haciendo cosas que nunca hicieron.

En las redes sociales, también debemos preocuparnos por la polarización en las opiniones sociales, políticas y de otro tipo de las personas. En general, los algoritmos de recomendación optimizan el compromiso (y las ganancias de la plataforma a través de la publicidad), no el discurso civil. Los algoritmos también pueden manipularnos de otras maneras. Los asesores robóticos (bots de chat para brindar asesoramiento financiero o brindar atención al cliente) pueden aprender a saber lo que realmente necesitamos, o presionar nuestros botones y vendernos

Vol. 2. Número 9 (Septiembre, 2022)

Figura 8. En 2021, Facebook dio a conocer su visión de un metaverso, un mundo virtual donde las personas trabajarían y jugarían. "Como muchos han dejado claro, esto es lo que quiere la tecnología", dice la socióloga y psicóloga clínica del MIT Sherry Turkle sobre el metaverso. "Para mí, sería más inteligente preguntar primero, no qué quiere la tecnología, sino qué quiere la gente. ¿Qué necesita la gente para estar más segura? ¿Menos solo? ¿Más conectados entre sí en las comunidades? ¿Más apoyo en sus esfuerzos por vivir una vida más saludable y plena?"

productos extraños.

Múltiples países están desarrollando armas autónomas que tienen el potencial de reducir las bajas civiles, así como de escalar el conflicto más rápido de lo que sus guardianes pueden reaccionar (**Fig. 7**). Poner armas o misiles en manos de robots plantea el espectro de ciencia ficción de Terminators que intentan eliminar a la humanidad. Incluso podrían pensar que nos están ayudando porque eliminar a la humanidad también elimina el cáncer humano (un ejemplo de no tener sentido común). Más sistemas automatizados a corto plazo liberados en el mundo real ya han causado caídas repentinas en el mercado de valores y los precios de los libros de Amazon alcanzan los millones. Si las IA están encargadas de tomar decisiones de vida o muerte, entonces se enfrentan al famoso problema del tranvía, decidiendo a quién o qué sacrificar cuando no todos pueden ganar. Aquí estamos entrando en territorio Asimov. También hay cuestiones sociales, políticas y jurídicas sobre cómo gestionar la tecnología en la sociedad. ¿Quién debe rendir cuentas cuando un sistema de IA causa daño? (Por ejemplo, los autos autónomos "confundidos" han matado a personas). ¿Cómo podemos garantizar un acceso más equitativo a las herramientas de IA y sus beneficios, y asegurarnos de que no perjudiquen a algunos grupos mucho más que a otros? ¿Cómo cambiará la automatización de los trabajos el mercado laboral? ¿Podemos gestionar el impacto medioambiental de los centros de datos, que consumen mucha electricidad? (La minería de Bitcoin es responsable de tantas toneladas de emisiones de dióxido de carbono como un país pequeño). ¿Deberíamos emplear preferentemente algoritmos explicables, en lugar de las cajas negras de muchas redes neuronales, para una mayor confianza y depuración, incluso si hace que los algoritmos sean más pobres en ¿predicción?

QUÉ SE PUEDE HACER

Michael Kearns, científico informático de la Universidad de Pennsylvania y coautor de *The Ethical Algorithm* sitúa los problemas en un espectro de manejabilidad. En un extremo está lo que se llama privacidad diferencial, la capacidad de agregar ruido a un conjunto de datos de, por ejemplo, registros médicos para que pueda compartirse de manera útil con los investigadores sin revelar mucho sobre los registros individuales. Ahora podemos hacer garantías matemáticas sobre exactamente cómo deben permanecer los datos de los individuos privados.

En algún lugar en el medio del espectro está la equidad en el aprendizaje automático. Los investigadores han desarrollado métodos para aumentar la equidad eliminando o alterando los datos de capacitación sesgados, o para maximizar ciertos tipos de igualdad, por ejemplo, en los préstamos, al tiempo que minimizan la reducción de las ganancias. Aún así, algunos tipos de equidad siempre estarán en conflicto mutuo, y las matemáticas no pueden decirnos cuáles queremos.

En el otro extremo está la explicabilidad. A diferencia de la equidad, que puede analizarse matemáticamente de muchas maneras, la calidad de una explicación es difícil de describir en términos matemáticos. "Siento que todavía no he visto una sola buena definición", dice Kearns. "Podrías decir: 'Aquí hay un algoritmo que tomará una red neuronal entrenada e intentará explicar por qué te rechazó para un préstamo', pero [la explicación] no se siente basada en principios".

Los métodos de explicación incluyen generar un modelo interpretable más simple que se aproxime al original, o resaltar regiones de una imagen que una red encontró destacadas, pero estos son solo gestos sobre cómo calcula el software críptico. Peor aún, los sistemas pueden proporcionar explicaciones intencionalmente engañosas para hacer que los modelos injustos parezcan justos para los auditores. En última instancia, si la audiencia no lo entiende, no es una buena explicación, y medir su éxito,

independientemente de cómo se defina el éxito, requiere estudios de usuarios.

Algo como las tres leyes de Asimov no nos salvará de los robots que nos hacen daño mientras intentan ayudarnos; pisar su teléfono cuando le dice que se dé prisa y le traiga un trago es un ejemplo probable. E incluso si la lista se extendiera a un millón de leyes, la letra de una ley no es idéntica a su espíritu. Una posible solución es lo que se llama aprendizaje por refuerzo inverso o IRL. En el aprendizaje por refuerzo, un modelo aprende comportamientos para lograr un objetivo determinado. En el aprendizaje por refuerzo, se infiere el objetivo de alguien al observar su comportamiento. No siempre podemos articular nuestros valores, los objetivos que en última instancia nos importan, pero la IA podría descubrirlos observándonos. Si tenemos metas coherentes, eso es.

"Quizás la preferencia más obvia es que preferimos estar vivos", dice Russell, quien ha sido pionero en la vida real. "Entonces, un agente de IA que usa el aprendizaje por refuerzo puede evitar cursos de acción que nos causen la muerte. En caso de que esto suene demasiado trivial, recuerde que ni uno solo de los prototipos de automóviles autónomos sabe que preferimos estar vivos. El automóvil autónomo puede tener reglas que en la mayoría de los casos prohíben acciones que causan la muerte, pero en algunas circunstancias inusuales, como llenar un garaje con monóxido de carbono, pueden ver a la persona colapsar y morir y no tener idea de que algo andaba mal. "

Ingeniero, cúrate a ti mismo

En el cuento de 1950 "El conflicto evitable", **Asimov (2012)** articulo lo que se convirtió en una "ley cero", que reemplazaría a las demás: "Un robot no puede dañar a la humanidad o, por inacción, permitir que la humanidad sufra daños". No hace falta decir que la regla debe aplicarse con "robotista" en lugar de "robot". Sin duda, muchos científicos informáticos evitan dañar a la humanidad, pero muchos tampoco se involucran activamente con las implicaciones sociales de su trabajo, lo que permite que la humanidad sufra daños, dice Margaret Mitchell, científica informática que codirigió el equipo de inteligencia artificial ética de Google y ahora consulta con organizaciones sobre ética tecnológica (Mujeres Conciencia Web, 2018).

Un obstáculo, según Grosz, es que no están debidamente capacitados en ética. Pero ella espera cambiar eso. Grosz y la filósofa Alison Simmons comenzaron un programa en Harvard llamado ***Embedded EthiCS***, en el que los asistentes de enseñanza con formación en filosofía participan en cursos de informática y enseñan lecciones sobre privacidad, discriminación o noticias falsas. El programa se ha extendido al MIT, Stanford y la Universidad de Toronto (***Grosz y cols. 2019***).

"Tratamos de que los estudiantes piensen en valores y compensaciones de valor", dice Grosz. Dos cosas la han llamado la atención. El primero es la dificultad que tienen los estudiantes con los problemas que carecen de respuestas correctas y requieren argumentar a favor de opciones particulares. El segundo es, a pesar de su frustración, "cuánto se preocupan los estudiantes por este conjunto de problemas", dice Grosz.

Otra forma de educar a los tecnólogos sobre su influencia es ampliar las colaboraciones. Según Mitchell, "las ciencias de la computación deben pasar de considerar las matemáticas como el principio y el fin de todo, a sostener tanto las matemáticas como las ciencias sociales y la psicología también". Los investigadores deberían traer expertos en estos temas, dice ella. En sentido contrario, dice Kearns, también deberían compartir su propia experiencia técnica con los reguladores, abogados y legisladores. De lo contrario, las políticas serán tan vagas que serán inútiles. Sin definiciones específicas de privacidad o equidad escritas en la ley, las empresas pueden elegir lo que sea más conveniente o rentable.

Al evaluar cómo una herramienta afectará a una comunidad, los mejores expertos suelen ser los propios miembros de la comunidad. Grosz aboga por consultar con poblaciones diversas. La diversidad ayuda tanto en los estudios de usuarios como en los equipos de tecnología (**Fig. 8**). "Si no tienes personas en la sala que piensen diferente a ti", dice Grosz, "las diferencias simplemente no están frente a ti. Si alguien dice que no todos los pacientes tienen un teléfono inteligente, boom, comienzas a pensar de manera diferente sobre lo que estás diseñando".

Según Margaret Mitchell, "el problema más apremiante es la diversidad y la inclusión de quién está en la mesa desde el principio. Todos los demás problemas se derivan de ahí" (***Matthew Hutson, 2021***).

REFERENCIAS BIBLIOGRÁFICAS

1) Asimov, I. (1941). Three laws of robotics. Asimov, I. Runaround.

2) Asimov, I. (2012). Visiones de robot (Serie de los robots 1). DEBOLSILLO.

3) Berners□Lee, T., Cailliau, R., Groff, J. F., &Pollermann, B. (1992). World□Wide Web: theinformationuniverse. Internet Research.

4) Castro, A. K. M. (2021). Inteligencia Artificial y Sociedad:¿ El fenómeno social tecnológico 4.0?. Futuro Hoy, 2(1).

5) Galera, M. C. S. (1994). Lady Ada Byron y el primer programa para computadoras. Divulgaciones matemáticas, 2(1), 75-81.

6) Hutson, M. (2021). Who should stop unethical AI. The New Yorker.

7) Kearns, M., & Roth, A. (2019). The ethical algorithm: The science of socially aware algorithm design. Oxford University Press

8) López, C. A. (2008). Historia de la Computación.

9) Mederos, E. Y. G. (2017). Inteligencia Artificial, Ética y Sistemas de Armas Automáticas en la Defensa Militar, Drones y Derecho Internacional.

10) Mujeres Conciencia (2018) https://mujeresconciencia.com/2018/04/01/como-podemos-construir-la-ia-que-nos-ayude-sin-perjudicarnos/

11) Ocampo, M., & Santa Catarina, C. (2018). Inteligencia artificial.

12) Pérez de Lama, J(2018). Unas notas sobre Shannon, fundador de la era de la Información.

13) Romano, J. (2021). Charles Babbage. Nextia, (1), 18-20.

14) Sain, G. (2015). Historia de internet. Revista pensamiento penal.

15) Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. Artificial Intelligence, 299, 103535.

16) Turing AM. Computing machinery and intelligence. Mind. 1950 Oct 1;59(236):433-60

17) TuringA(2018) https://www.eldiario.es/turing/John-Neumann-revolucionando-computacion-Manhattan_0_380412943.html (2018)