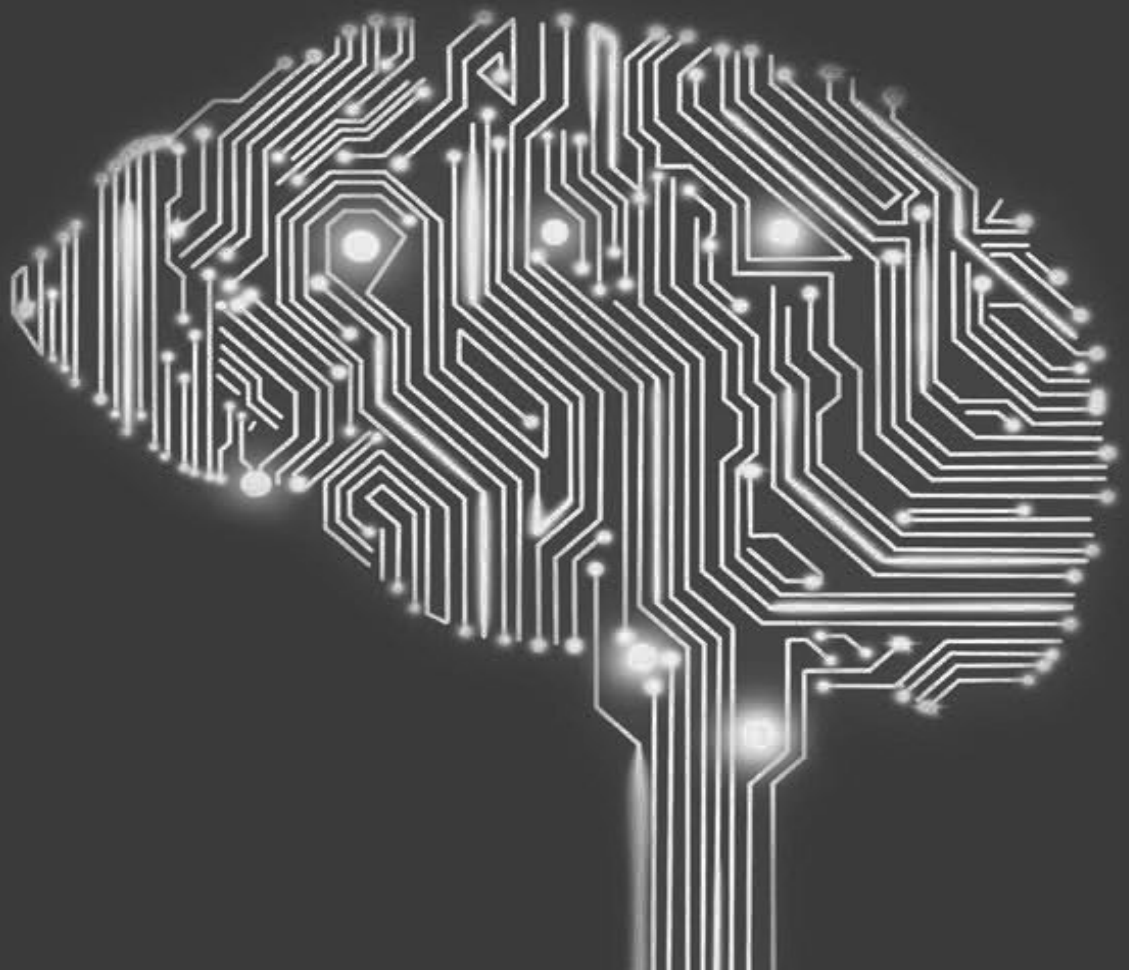
Libros.com

Juan Ignacio Rouyet Ruiz

ESTUPIDEZ ARTIFICIAL

Cómo usar la inteligencia artificial sin que ella te utilice a ti



Primera edición digital: mayo 2023
Campaña de crowdfunding: equipo de Libros.com
Composición de la cubierta: Mariona Sánchez
Maquetación: Irene E. Jara
Corrección: Beatriz García
Revisión: Isabel Bravo de Soto

Versión digital realizada por Libros.com

editorial@libros.com

ISBN digital: 978-84-19435-27-9

Juan Ignacio Rouyet

Estupidez artificial

Cómo usar la inteligencia artificial sin que ella te utilice a ti

A Paloma, un potosí.

Índice

De agua, harina y sal

Por el lejano valle de Wakhan, más allá de los montes de Hindu Kush, caminaba el sabio sufí conocido por todos como el Gran Yusuf ibn Tarum. En la cercana comunidad de Ishkashim, se enteraron de la presencia por la zona del gran maestro y salieron por los caminos a su encuentro. Cuando dieron con él, le llamaron y le pidieron que pasara unos días con ellos. El maestro accedió y los acompañó hasta su aldea.

Al día siguiente, toda la comunidad se reunió en la plaza para escuchar las sabias palabras del Gran Yusuf ibn Tarum. El líder de la comunidad se puso en pie para pedir al maestro por sus enseñanzas.

—Gran Yusuf ibn Tarum, sabemos de vuestra grandeza, de vuestra gran sabiduría y que estáis tocado por un espíritu de revelación. Nosotros somos una comunidad pequeña y es posible que nunca hayáis oído hablar de nosotros, de nuestra dedicación y búsqueda de la verdad. Por ello, gran maestro, antes de escuchar vuestras sabias palabras, dejadnos que os contemos nuestro pensamiento y forma de obrar para alcanzar la perfección, de tal forma que podáis confirmar, refutar o completar nuestras ideas.

El Gran Yusuf ibn Tarum se puso en pie e interrumpió al líder, cuando este se disponía a explicar su escuela de pensamiento. Al momento, comenzó

a contar las ideas que preocupaban a aquella comunidad, los pensamientos que les hacían errar, las dudas que tenían y cómo intentaban superar sus dificultades. Todos quedaron asombrados de cómo el gran maestro sabía tanto sobre ellos, siendo como eran, una comunidad tan modesta y desconocida.

Cuando terminó, el líder de la comunidad se levantó de nuevo y alabó las palabras del maestro:

—¡Oh, Gran Yusuf ibn Tarum! Es asombroso lo que hemos visto. Se nota con certeza que os acompaña un espíritu de revelación, y así conocéis aquello que a otros se les oculta. Decidnos, gran maestro, cuál es el camino para tal perfección de sabiduría.

El maestro se quedó callado. Se hizo el silencio en la aldea. Tras unos segundos de expectación, pidió que le trajeran un cuenco de barro y agua, harina y sal. En el cuenco echó el agua, la harina y la sal. Lo removió bien y preguntó al líder.

—Dime, ¿de qué está hecha esta mezcla?

—De agua, harina y sal.

—¿Cómo lo sabes?

—Porque conozco los ingredientes, gran maestro. Si conoces los ingredientes de una mezcla, eres capaz de conocer la naturaleza de la mezcla.

Así ocurre con la naturaleza humana. Conozco vuestros pensamientos porque conozco los ingredientes de la naturaleza humana[1].

Conocer los ingredientes de la mezcla para evitar la estupidez artificial. ¿Qué ingredientes? ¿Qué mezcla? ¿Qué estupidez? Empecemos por eso de la estupidez, que siempre es más simpático.

La inteligencia artificial viene precedida de amenazas, pero también promete esperanzas. Alcanzar las esperanzas o hacer realidad las amenazas es algo que depende de nosotros. De nuestra actuación, es decir, de nuestra inteligencia o de nuestra estupidez.

La estupidez artificial es aquel comportamiento que anula nuestra autonomía cuando usamos la inteligencia artificial. Es cuando decidimos dejar de ser responsables, porque abandonamos nuestra capacidad de responder y dejamos que la inteligencia artificial responda por nosotros.

Estupidez artificial es argumentar con frases del estilo «lo dice el sistema», «el algoritmo ha determinado que…», o, la mejor de todas, «la inteligencia artificial ha decidido…». ¡¿Cómo?! ¡Y tú qué! ¿Tú qué dices, qué determinas o qué decides? En todas estas frases falta un «yo» que responde, y en su lugar se traslada la respuesta a un algoritmo.

El primer paso para ser ético no es tanto ser buena persona. El primer paso es tener la voluntad de responder de tus actos ante ti y ante los demás. Delegar la respuesta de tus actos en una inteligencia artificial es abandonar toda responsabilidad. Es dejar de ser ético. Es estupidez artificial. ¿Cómo evitarlo? Conociendo los ingredientes de la mezcla.

La inteligencia artificial es muy compleja de entender. Tanto, que parece magia. ¡Cómo es posible que haga lo que nosotros hacemos! Sin embargo, cuando conoces un truco de magia, carece de emoción. Hubo un tiempo en el que parecía magia que los mensajes fueran instantáneos con el telégrafo, o que el teléfono transportara la voz. También en el pasado tuvimos miedo del tren. Se pensó que nos volvería locos. Hoy estos inventos no nos asombran. Si conocemos los ingredientes, conoceremos la mezcla y lo entendemos todo. Pero en este juego de la inteligencia artificial no hay una mezcla, sino dos: la mezcla de la inteligencia artificial y nuestra mezcla; sí, nosotros mismos.

Te propongo conocer los ingredientes de la inteligencia artificial para quitarle todo atisbo de divinidad. Si queremos usar bien una herramienta tenemos que saber algo sobre cómo funciona. No sabemos construir un coche, pero sabemos por qué se mueve. No sabemos pilotar un avión, pero sabemos que no vuela por una ciencia desconocida. Hoy vemos que la inteligencia artificial es capaz de entendernos, de hablar, de decidir una ruta, de diagnosticar una enfermedad, de escribir un poema o pintar un cuadro. ¿Magia? No. Nada de eso. Basta conocer sus elementos básicos. Si conocemos los ingredientes de la mezcla de la inteligencia artificial, sabremos usarla, y empezaremos a alejarnos de la estupidez artificial.

También te propongo conocer nuestra mezcla. Saber de nosotros, si nuestros ingredientes son los mismos que los de la inteligencia artificial, y cómo respondemos de lo que hacemos. Este libro tiene algo de filosofía y algo de ética. Quizás pienses que la filosofía no sirve para nada, porque no

lleva a conclusiones prácticas. Tienes parte de razón en ello, porque en ocasiones los textos filosóficos no hay quien los entienda. Pero la filosofía te hace pensar, y pensar te ayuda a ser libre. ¿Hay algo más práctico que ser libre?

Te planteo dos objetivos con este libro: conocer y actuar. Conocer los ingredientes para conocer la mezcla. Conocer para actuar, lejos de la estupidez artificial.

Para ello, en cada capítulo de este libro, intentaré ir respondiendo a una serie de preguntas que nos permitirán ir avanzando poco a poco.

¿Debemos tener miedo de la inteligencia artificial?

¿Cuáles son los verdaderos riesgos éticos de la inteligencia artificial?

¿Cómo funciona la inteligencia artificial?

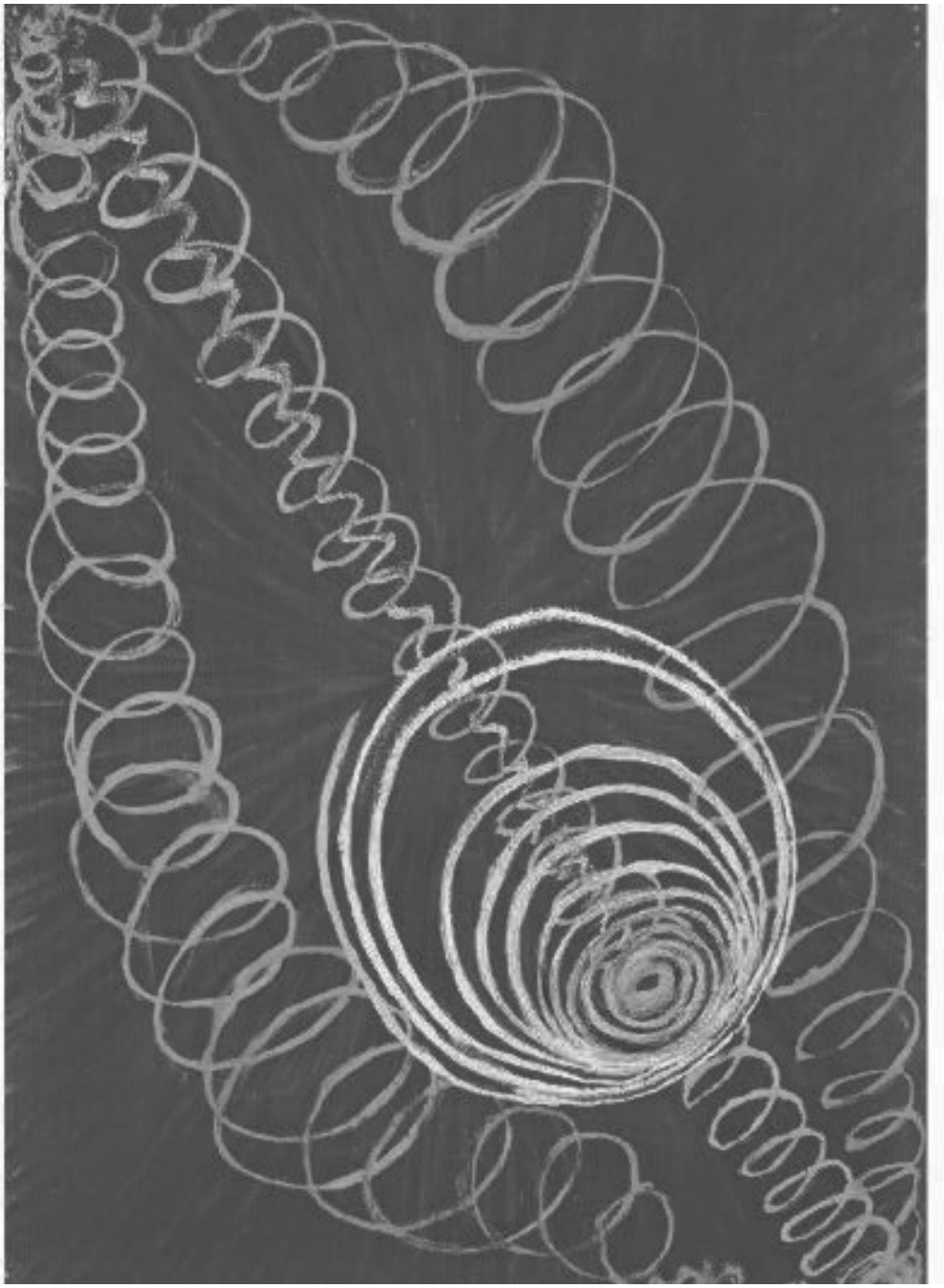
¿Es la inteligencia artificial igual a nosotros?

¿Es posible tener una inteligencia artificial ética?

¿Qué deben hacer las organizaciones para tener una inteligencia artificial ética? ¿Qué debemos hacer nosotros?

Al final de cada capítulo te daré mi respuesta. Pero solo será mi respuesta, no la tuya. Saca tu espíritu crítico y busca tu punto de vista.

En la historia del gran maestro sufí Yusuf ibn Tarum hay dos posibles relatos. En uno de ellos, la inteligencia artificial es quien conoce nuestra mezcla y sabe de nosotros, de cómo actuamos. Estupidez artificial. En el otro escenario, nosotros conocemos la mezcla de la inteligencia artificial y sabemos cómo usarla. Sabiduría natural. Este libro te propone pensar y actuar para hacer realidad este segundo relato.



***Caos primordial*, Nº 16, Hilma af Klint, 1906-07. Fundación Hilma af Klint**

Miedo me da

Hoy vemos este cuadro y no nos causa rechazo. Nos gustará más o menos, pero lo aceptamos. *Caos primordial* forma parte del llamado arte abstracto, al cual ya nos hemos acostumbrado. Pero en 1906, cuando fue pintado por la artista sueca Hilma af Klint, dentro de una serie de pinturas llamadas *Las pinturas para el templo*, la situación era completamente distinta. Hilma era seguidora del movimiento filosófico-religioso llamado teosofía, que busca el conocimiento de una realidad espiritual que va más allá de las doctrinas particulares de cada religión. Inspirada por esta idea, Hilma pintó su serie de cuadros para el Templo, casi de una forma instintiva, dejando llevar su mano como guiada por una fuerza superior.

En 1908, Hilma enseñó su colección de 111 pinturas abstractas a Rudolph Steiner, filósofo defensor de la teosofía. Rudolph desanimó a Hilma a continuar con aquella línea de pintura, porque era inapropiado para la teosofía. Hilma se quedó desolada y estuvo sin pintar unos 4 años. Posteriormente, continuó con su visión de un nuevo estilo de pintura, dejando una colección de más de 1200 cuadros abstractos.

Cuando murió, en 1944, legó su obra a su sobrino Erik, y dejó escrito en su testamento que su obra no se hiciera pública hasta 20 años después de su

muerte, esperando que, para entonces la sociedad, en general, pudiera entender su arte. En 1970, cumplido el plazo de los 20 años, Erik presentó las obras al Moderna Museet de Estocolmo, quien las rechazó cuando se enteró que la autora había flirteado con la teosofía. La primera exposición pública de la obra de Hilma af Klint no fue hasta 1986, en Los Ángeles, bajo el título «Lo espiritual en el Arte: pinturas abstractas 1890 – 1985»[2].

Todo este desprecio por la obra de Hilma af Klint, por su relación con una visión espiritual, dio fama a Vasili Kandinsky, quien es considerado como el padre de la pintura abstracta. En 1911 publicó su famosa obra *De lo espiritual en el arte*[3], donde sienta las bases del arte abstracto y explica cómo una serie de formas y colores pueden inspirar ciertas ideas y emociones. A partir de entonces, sus obras tituladas como *Composiciones* (para qué darle nombres concretos) forman parte de la historia del arte contemporáneo y del arte abstracto.

En 1906 el arte abstracto de Hilma fue rechazado, pero 5 años después comenzó a cambiar la visión sobre el arte abstracto. En los años siguientes se aceptaron nuevas normas y se hicieron nuevos juicios de valor. La desconocida Hilma pasó 80 años en el olvido hasta que comenzó a ser reconocida su aportación al arte. Ahora sus pinturas no nos sorprenden, ni nos causa rechazo si han sido inspiradas por una visión espiritual.

En esta historia hay un juicio de valor erróneo y un acierto. El juicio de valor erróneo consistió en denostar una obra de arte por las inclinaciones filosóficas de su autora. El error de creer que una obra de arte debería estar realizada por un artista con unas condiciones determinadas, lo que sería considerado un artista serio. «Debería», ya ha salido la palabra. El acierto estuvo en Kandinsky, quien antes de exponer una obra abstracta, explicó en un libro cómo entenderla.

Lo ocurrido con la obra de Hilma af Klint ocurre ahora con la inteligencia artificial. En cualquier actividad que realizamos, siempre existe un debate entre lo que es y lo que debería ser. La obra *Caos primordial* no era como debía ser. Ese tipo de arte era entonces inaceptable, pero hoy en día ya es aceptable. Incluso aceptamos cualquier tipo de expresión artística. No hay más que pasarse por ARCO. En el arte ya no hay deberías. Y la inteligencia

artificial, ¿es cómo debería ser?, ¿dejará de tener deberías?, ¿debemos tener miedo a la inteligencia artificial?

Para responder a estas preguntas podemos ver lo que ha ocurrido en el pasado con otras tecnologías. Te propongo un viaje por tres avances tecnológicos del siglo XIX: el ferrocarril, el telégrafo y el teléfono. En los tres casos vamos a ver el debate que hubo en su momento sobre lo que debería y no debería ser, y cómo este debate ha ido cambiando con el tiempo. Veremos que los miedos iniciales —lo que no debería ser— no eran tan peligrosos, y que muchas cuestiones que eran inaceptables en aquella época, hoy las tenemos por cosas normales. Igualmente, había muchas esperanzas —lo que sí debería ser—, que con el tiempo se han visto que no han llegado a ser realidad. Lo que ha ocurrido con otras tecnologías, ocurrirá con la inteligencia artificial.

Al juzgar el arte de Hilma af Klint los deberías se centraron en si su obra seguía un estándar artístico o si su biografía era seria. Si hablamos de inteligencia artificial, ¿sobre qué habla lo que debería o no debería ser?

Actualmente en nuestra sociedad todo debate se centra en cuatro dimensiones que articulan lo que debería y no debería ser. Son las dimensiones de salud, seguridad, economía y moral. No necesariamente por este orden, aunque las cuestiones morales suelen surgir en último lugar. Tanto defensores como detractores crean sus argumentos en base a todas o algunas de estas dimensiones. Juicios de valor que, al igual que en el arte, han ido cambiando, y lo que antes era «esta tecnología no debería ser así», con el tiempo se ha convertido en «sí debe ser así». Vamos a verlo.

El ferrocarril

Velocidad que enloquece

La irrupción del ferrocarril a comienzos del siglo XIX unió los pueblos y abrió los miedos. En la eterna, tranquila y verde campiña británica apareció un monstruo: la máquina de vapor. El ferrocarril fue la imagen viva del progreso tecnológico y su presencia dividió tanto a los campos, por el tendido de raíles; como a la sociedad, por los «deberías».

La primera cuestión sobre lo que debería o no debería ser vino con la seguridad, y, en particular, de la mano de la velocidad. Viajar a 50 Km/h se percibía como un riesgo considerable. Este miedo se vio acrecentado por dos accidentes memorables en sus comienzos.

Uno de ellos sucedió el mismo día de la inauguración de la línea Liverpool-Mánchester en 1830. Durante una parada del tren para abastecerse de agua, William Huskisson, miembro del Parlamento, fue atropellado por una locomotora Rocket debido una imprudencia. Murió a los pocos días[4]. Años más tarde, el escritor Charles Dickens tuvo un accidente de tren. Dos raíles sobre un viaducto fueron retirados por error al mirar un horario de tren equivocado. Todos los vagones de primera acabaron sobre el río, excepto el de Dickens, que quedó colgando[5]. Dickens se salvó, pero no así la reputación del ferrocarril como transporte seguro.

Aquí tenemos ya un paralelismo con la inteligencia artificial y los vehículos de conducción autónoma. Existen dudas sobre su seguridad. No ayuda el accidente de 2018 en Arizona, donde un vehículo autónomo de Uber mató a una mujer, que cruzaba la carretera con su bicicleta. Tuvo repercusión por dos motivos: fue el primer accidente por atropello de un vehículo autónomo y fue causado por un error de *software*. El radar no identificó correctamente a la señora cruzando la carretera, y además Uber había deshabilitado la opción de frenado de emergencia[6]. En el accidente de Dickens se quitó un rail por error; en el caso de Uber se deshabilitó una opción del *software*.

Todo avance tecnológico tiene un periodo inicial de desgracias. Dicho de una manera quizás más radical: todo avance tecnológico ofrece nuevas formas de morir; no se sabe de nadie que muriera por accidente de ferrocarril en la época del Imperio romano. A pesar de estos siniestros, el ferrocarril ha continuado y hoy es un medio de transporte bastante seguro. La inteligencia artificial también llegará a ser segura.

No obstante, no era necesario sufrir un accidente para padecer daños en la salud. Se pensaba que la velocidad en sí misma causaba demencia. Así se atestiguaba en múltiples casos de personas, particularmente hombres, que durante un viaje en tren habían tenido comportamientos violentos y actitudes fuera de sí[7]. Cada vez que el tren paraba en una estación, estas personas recobraban su sensatez y se comportaban de manera correcta y educada. No obstante, una vez que el tren se ponía en movimiento, la actitud violenta y errática volvía a manifestarse.

Se pensaba que el traqueteo del tren, unido al excesivo ruido por el movimiento sobre los raíles, literalmente afectaba al cerebro y alteraba los nervios. Se culpó, entonces, no solo a la máquina de vapor como el instrumento de progreso que no debía existir, sino que también se culpó a la propia civilización como el origen de nuestros males. En la última mitad del siglo XIX, los exploradores que volvían de las zonas más remotas del mundo informaban que apenas veían personas con enfermedades mentales en las áreas con menor civilización. Se llegó entonces a establecer que la locura, o cualquier otra enfermedad mental, era la inevitable consecuencia de la civilización y que un incremento de la demencia era el castigo que debíamos pagar por el incremento de la civilización[8].

¿La civilización afecta a nuestro comportamiento? Todo parece apuntar a que sí. De hecho, ese es el objetivo de la civilización: tener un comportamiento específico, que denominamos *social*. ¿Nos lleva a comportamientos no deseados, como la locura con el tren? Esto es más discutible. En el caso del ferrocarril con el tiempo se ha visto que no. Con la inteligencia artificial, se empezó a analizar si los asistentes de voz, tipo Alexa (Amazon), Google Home, Siri (Apple) o Cortana (Microsoft), afectaban al comportamiento de los niños. Habitualmente, cuando hablamos con estos dispositivos, no decimos palabras tales como «por favor» o «gracias», más bien

decimos simplemente: «Alexa, ¿qué hora es?». Se vio que los niños usaban este estilo directo de comunicación, tanto con los asistentes de voz, como en la comunicación con los adultos. Parecía que estos siervos inteligentes volvían maleducados a los niños.

Esta situación llevó a Amazon a incorporar la «Magic Word» en su sistema de Alexa. Con esta nueva funcionalidad, cada vez que se pide algo a Alexa incluyendo la palabra mágica «por favor», el sistema responde también de manera agradecida, diciendo, por ejemplo, «gracias por ser tan amable». De manera similar, Google integró posteriormente su funcionalidad «Pretty Please».

Posteriormente se vio que estos asistentes inteligentes no afectaban al comportamiento de los adultos y que los niños eran conscientes de cuándo hablaban con una máquina o con una persona. Al final todo se aclara y nosotros nos adaptamos.

Ahora debemos dar paso a los economistas. En aquel entonces del ferrocarril, todo parecía indicar que la civilización se veía atacada por la innovación. Tocaba, entonces, hablar de los importantes beneficios económicos que traería la innovación del ferrocarril.

Los zapateros venderán más zapatos

Claramente, el ferrocarril era más productivo y eficiente, comparado con los medios habituales de transporte del momento, que eran a caballo o en barcazas por canales fluviales. Las palabras *productivo* y *eficiente* nunca deben faltar si queremos demostrar la bondad de una innovación.

El ferrocarril iba a generar un gran beneficio a toda la sociedad. Para convencer de ello, lo mejor era demostrarlo con un ejemplo cercano, como era el caso de suponer la existencia de un humilde zapatero en un remoto pueblo campestre. Gracias al tren, este zapatero vería ampliado su mercado, porque ya no solo vendería calzado a sus vecinos del pueblo, sino que podría vender sus zapatos en las principales ciudades británicas[9]. El ferrocarril era una fuente de negocio para los zapateros y para cualquier empresario o dueño de un humilde negocio.

Pero no solo para dueños de negocio. El ferrocarril iba a permitir ahorrar dinero y tiempo y tener una vida más plena. Cualquier mercancía transportada por canal entre Mánchester y Liverpool llevaba 36 horas de viaje, frente a la hora y tres cuartos que suponía transportarla por tren, lo que supone un ahorro del 50 % del coste. Tomando como base estas eficiencias, vinieron los análisis de grandes números con grandes esperanzas.

Teniendo en cuenta que medio millón de personas al año utilizaban dicha línea de tren, si cada uno de ellos ahorraba tan solo una hora en el trayecto, esto suponía un ahorro de 500.000 horas, es decir, 50.000 días de trabajo —en aquel entonces una jornada de trabajo eran 10 horas al día—. Esto equivalía a aumentar la fuerza de trabajo en 160.000 hombres, sin necesidad de incrementar la comida para alimentarlos. Hemos conseguido más capacidad de trabajo. Además, el trabajo de estos 160.000 hombres sería de más valor que el de aquellos ocupados en el campo o en el transporte convencional a caballo o barcazas[10].

Es posible que todo esto nos suene. Toda innovación siempre nos promete más valor añadido, lo que quiera que signifiquen las palabras *valor* y *añadido*. Nunca nos lo explican, tan solo sueltan el par de palabras acompañadas de una sonrisa de satisfacción. Está claro que, si se quiere vender la bondad de una tecnología, hay que meter la palabra valor. La inteligencia artificial también nos promete trabajos de más valor. Aquellas actividades tediosas, rutinarias o peligrosas las podrá hacer un robot que ni siente ni padece. El tema merece más atención y lo veremos posteriormente, pues, por otro lado, si la inteligencia artificial hace un trabajo tedioso, quiere decir que una persona deja de tener un trabajo. Aunque fuera tedioso, seguro que le arreglaba la vida. Lo mismo sucedía con el ferrocarril, su eficiencia venía a cambio de una pérdida.

El ferrocarril iba a eliminar puestos de trabajo: aquellos dedicados, precisamente, al transporte por caballo o canales, lo cual generaría pobreza en las personas dedicadas a tal actividad[11]. Sin embargo, un avispado cochero no tendría por qué preocuparse.

Se estaba demostrando que el transporte a caballo, debido a la línea férrea entre Mánchester y Liverpool, iba en aumento al favorecer el transporte de corta distancia[12]. Las grandes distancias eran cubiertas por ferrocarril, pero

esos tramos cortos entre una estación y una población cercana eran salvados por diligentes taxistas a caballo. Como el tren movía a muchas a personas, estos audaces conductores verían incrementado su negocio.

Todo eran parabienes. Algunos oficios, como el transporte por canal, se verían afectados por la llegada del ferrocarril, pero el resultado global sería que los zapateros venderían más zapatos, los cocheros llevarían a más gente, el país sería más productivo y habría trabajo de más valor. El tren sería más productivo, más eficiente y de más valor: ¿cómo negarse ante esas tres palabras?

La inteligencia artificial quizás nos quitará puestos de trabajo, pero nos dicen que haremos otros y de más valor. Al igual que los cocheros de entonces, nos dicen que no nos preocupemos. Más adelante veremos si debemos preocuparnos o no. De momento, el tren todavía traerá más enhorabuenas.

Peor lana, pero más inteligentes

Finalmente, quedaron los argumentos de naturaleza moral, que eran aquellos que asignaban un cierto juicio de valor al invento. El ferrocarril fue considerado una máquina infernal. El propio nombre representaba un cierto aire de odio: ferrocarril, carril de hierro; hierro, metal, que denota dureza y batalla. ¡Qué cosa tan inhumana!

El tren se consideraba innecesario, pues no había razón de viajar tan rápido, y además iba en contra de los valores de tranquilidad, belleza y unidad. La principal oposición vino por los dueños de las tierras, que veían cómo aquella máquina oscura y agresiva atravesaba sus campos expulsando humo negro a toda velocidad.

Aquello no podía traer nada bueno: rompía la paz de un desayuno tranquilo y llenaba sus hogares de hollín y ruido; destruía su privacidad, pues los pasajeros de los trenes pasaban cerca de sus casas; destruía la unidad de las granjas, al ser divididas por los raíles del tren; desfiguraba el paisaje, al cortar montañas o crear largos puentes para cruzar valles o ríos; interfería en la caza; e incluso, y aquí empieza lo verdaderamente pintoresco, afectaba a la calidad de la lana de las ovejas que pacían junto a las vías del tren. En resumen, el

ferrocarril transmitía los valores de vulgar, mercenario y desagradable[13]. Claramente, el ferrocarril no era lo que debía ser, pues no traía nada bueno.

Sin embargo, frente a esta visión tan deprimente, existía otra forma de ver las cosas. También había grandes esperanzas para la humanidad en aquellas máquinas de hierro y carbón. Dado que el ferrocarril iba a aumentar la productividad de cada artesano, —por ejemplo, de los zapateros— se iba a producir una cadena de consecuencias virtuosas sin precedentes: al tener que fabricar más zapatos, sería más competente en su trabajo; esto le haría ser más rápido en hacer zapatos, lo que le daría más tiempo libre; el tiempo libre le induciría a la indagación; y la indagación, al conocimiento[14]. Por tanto, el tren iba a hacer que la gente del siglo XIX fuera más inteligente, al menos, los zapateros.

Pero aún hay más. No era verdad que el ferrocarril fuera a estropear el eterno candor de suaves valles y colinas. Antes bien, la construcción de puentes, viaductos, accesos y depósitos relacionados con el ferrocarril, traerían un nuevo tipo de arquitectura creando un embellecimiento arquitectónico. Esta nueva belleza sería fuente de una mejora en los hábitos y la moral de la población rural: fomentaría el cultivo del gusto y la difusión del conocimiento, por medio de la comunicación global[15]. Se llegó a decir, textualmente y por personas muy sesudas que «la duración de nuestras vidas, en lo que respecta al poder de adquirir información y diseminar conocimiento, se duplicará; y podemos estar justificados al buscar la llegada de un tiempo, cuando el mundo entero se habrá convertido en una gran familia, hablando un solo idioma, gobernado en unidad y armonía por leyes similares, y adorando a un solo Dios»[16].

Con estas palabras, uno queda rendido ante el ferrocarril. Esos trenes que cruzan los campos iban a conseguir que los pueblos fueran más inteligentes, con mejores hábitos, más éticos y más fraternales. Hoy en día, con la alta velocidad, cabe suponer que nuestra inteligencia, nuestra ética y nuestra fraternidad vuelen a la velocidad del rayo. ¿Somos hoy todo eso? ¿Se cumplen las bondades que nos venden con las innovaciones? ¿La felicidad que nos pintan, con solícitos robots a nuestro lado, será verdad?

Hablando de la velocidad del rayo, unos años más tarde, llegó la comunicación a la velocidad de la luz.

El telégrafo

Acabaremos con la barbarie

No todo avance tecnológico fue tachado, de primeras, de incorrecto. El establecimiento de la primera línea de telégrafo en 1844 trajo grandes esperanzas para la humanidad, casi como el tren[17]. Con el telégrafo, por primera vez, la transmisión de la información se separaba del medio físico por el cual viajaba. Hasta la fecha, la información era transportada por un mensajero, habitualmente a caballo, y esta viajaba a la velocidad a la que iba el mensajero. Si el mensajero tardaba tres días en llegar a su destino, el mensaje tardaba tres días en llegar a su destinatario. Si el mensajero corría, el mensaje llegaba antes; si se entretenía en cantinas por los caminos, la misiva llegaba tardía.

El telégrafo hizo que la información viajara a la velocidad del rayo. Esto era increíble. ¿Cómo era posible que un mensaje llegara de manera instantánea, sin mediar un mensajeo por medio? Además, ¿qué es eso de la electricidad? Para mayor conmoción el telégrafo estaba basado en esa fuerza invisible llamada electricidad, que no todo el mundo llegaba a entender. Parecía mentira que el pensamiento pudiera ir más deprisa que la materia. Si nuestras ideas podían volar a la velocidad de un suspiro, entonces se aventuraban grandes esperanzas.

Se creó el concepto de «comunicación universal» como aquel medio por el cual se podría unir la mente de todos los hombres en una especie de conciencia común. El telégrafo iba a unir a todos los hombres y mujeres de la Tierra para transmitir los principios más elevados del humanismo. Por fin la barbarie estaba llamada a su fin. Era imposible que los viejos prejuicios y las hostilidades existieran por más tiempo, dado que se había creado un instrumento que permitía llevar el pensamiento a cualquier lugar del mundo, y con independencia del mensajero. Ahora sí que estábamos delante de un invento moralmente bueno en sí mismo. El telégrafo era lo que debía ser.

Todas sus supuestas bondades innatas aparecieron reflejadas en la primera frase que se transmitió por dicho medio, atribuyendo al invento una dote

divina. El 24 de mayo de 1844 se produjo la primera transmisión pública por telégrafo entre Washington y Baltimore, cubriendo una distancia de unos 60 Km. Samuel Morse, desde la Cámara de la Corte Suprema en el Capitolio de EE. UU., envió a su colega Vail, en Baltimore, la frase «What hath God wrought!», tomada de la Biblia. En particular del capítulo 23 y versículo 23 del Libro de los Números, y que se puede traducir como «¡Lo que Dios ha hecho!». De alguna forma, el telégrafo era una creación divina que iba a traer toda clase de beneficios a la sociedad y era propio de una obra celestial.

Si ya con el ferrocarril íbamos a ser más éticos y fraternales, y ahora con el telégrafo se iba a acabar la barbarie, hoy en día el mundo debe ser maravilloso, aunque quizás no nos demos cuenta. Más bien, parece que hemos conseguido una comunicación global, enganchados al móvil, pero no tanto esa «comunicación universal» que nos lleva a una conciencia común y a una unidad entre los seres humanos. No hay más que ver ciertos mensajes en las redes sociales.

Hoy vemos que, finalmente, el telégrafo no ha terminado con la barbarie. ¡Qué cosa más rara! ¿Qué puede haber fallado, si todo apuntaba tan bien?

La guerra por un telegrama

El error —la falacia— radicó en creer que una tecnología tiene connotaciones morales y que es buena o mala en sí misma, sin tener en cuenta el uso que nosotros hacemos de ella. El telégrafo no es ni bueno ni malo, depende del uso que hagamos de él, es decir, de los mensajes que transmitamos. Un ejemplo claro es lo que ocurrió con el llamado telegrama de Ems que causó la guerra Franco-Prusiana en 1870. Justo lo contrario de acabar con la barbarie.

En aquellas fechas del siglo XIX, España, una vez más, era el tablero de ajedrez de las vanidades de Europa por cuestión de la sucesión al trono real. Con el exilio de Isabel II tras la Revolución de 1868, conocida como la Gloriosa, se comenzó a buscar un monarca para España. Una de las opciones fue el príncipe Leopoldo de Hohenzollern, propuesto por Otto von Bismarck, en aquel entonces, primer ministro de Prusia. Esto no gustó a Francia, porque suponía aumentar el poder de Prusia en Europa. Finalmente,

Francia consiguió que Prusia abandonara su propuesta. Pero Napoleón III, rey de Francia, quiso tener por escrito la renuncia de Guillermo I, rey de Prusia, a toda pretensión de proponer un candidato para el trono español.

Aprovechando que Guillermo I pasaba unos días en el balneario de Ems, Napoleón III envió a un embajador a entrevistarse con él. El rey Guillermo I le recibió, si bien rehusó dejar por escrito dicha renuncia, aduciendo que, por su parte, no tenía noticias oficiales de la renuncia de Leopoldo de Hohenzollern, como, en efecto, así era. Al día siguiente, supo de tal renuncia y, en lugar de entrevistarse de nuevo con el embajador, le mandó un mensaje diciendo que, dado que ya no había candidato, por su parte no había más que decir. Hasta aquí, todo correcto, formal y diplomático. Sin embargo, los intereses particulares pueden trastocar la realidad.

Con la nueva situación, Guillermo I mandó un telegrama a Bismarck relatando los acontecimientos y dejando en manos de este si comunicar o no tal hecho a la prensa y el cómo hacerlo. Bismark era defensor de la candidatura de Leopoldo de Hohenzollern y no le gustó cómo se habían sucedido los acontecimientos. Esto no podía quedar así. Redactó un nuevo telegrama para la prensa alterando ligeramente los hechos. En su telegrama se decía escuetamente que Guillermo I rechazó recibir al embajador y que no tenía nada que decirle. Rechazar recibir a un embajador y proclamar que no hay nada que decir es una afrenta diplomática. Este rechazo y silencio, que no fueron tales, se entendió como un desprecio a Francia. El telegrama fue enviado el 13 de septiembre de 1870 y Napoleón III declaró la guerra a Prusia 6 días después.

¿Esto es lo que se dice acabar con la barbarie? Esta historia suscitó un debate sobre si la tecnología afectaba a nuestra capacidad de pensar de una forma sosegada. El debate sigue activo hoy, con unas redes sociales que no descansan, movidas por una inteligencia artificial. Veamos primero el fundamento.

Sin capacidad de pensar

Este hecho de la guerra Franco-Prusiana trajo una consideración sobre el telégrafo: no dejaba tiempo para pensar. La inmediatez del telégrafo

transmitía el valor de la necesidad de inmediatez en la respuesta. Todo, hasta las respuestas, tenía que suceder a la velocidad de la luz, porque la información iba a la velocidad de la luz. Incluso las noticias se sucedían a la velocidad de la luz.

La noticia del telegrama de Bismark fue publicada y discutida de manera vertiginosa por los periódicos, lo que favoreció un clima social que exigía una respuesta rápida ante tamaña ofensa. Por ello, se culpó al telégrafo en sí por el estallido de la guerra Franco-Prusiana de 1870, no tanto por el contenido del telegrama, sino porque esa comunicación instantánea obligó a imprimir un ritmo acelerado en la diplomacia que impidió la reflexión sosegada de los hechos. En cierta manera, la declaración de guerra seis días después de recibir el telegrama fue el producto de una toma de decisión poco meditada, acuciada por la necesidad de dar una respuesta tan rápida como la rapidez del telégrafo.

El periódico inglés *Spectator* publicó en 1889 el editorial «Los efectos intelectuales de la electricidad» donde alertaba de los daños que podía causar el telégrafo en nuestra mente. En realidad, como se ve por el título del editorial, la preocupación no nacía del telégrafo en sí mismo, sino de la electricidad, que era esa fuerza misteriosa del telégrafo.

Lo más alarmante para el *Spectator* era lo que denominaba «fuerza intelectual» de la electricidad, la cual afectaba al telégrafo. Según el periódico, el telégrafo era un invento que no debía ser, pues su uso iba a afectar al cerebro y al comportamiento humano. Merece la pena leer algunos párrafos, no tienen desperdicio[18]:

> [Debido al telégrafo] todos los hombres están forzados a pensar en todas las cosas al mismo tiempo, en base a información imperfecta y con muy poco tiempo para la reflexión. Es rumor, más que inteligencia, lo que se apresura sin aliento por mares y continentes. […] La constante difusión de declaraciones en fragmentos, la constante emoción de sentimientos no justificados por hechos reales, la constante formación de opiniones apresuradas o erróneas, al final, se podría pensar, acabará por deteriorar la inteligencia de todos aquellos a los que apelan al telégrafo. […] El resultado es una precipitación universal y confusión de juicio, una disposición a decidir demasiado rápidamente, una impaciencia si no se toman medidas apresuradas antes de que los estadistas u otros responsables hayan tenido tiempo de pensar. Es como si todos los hombres tuvieran que estudiar todas las cuestiones bajo la emoción de la ira, el miedo o la pena, o con esa sensación consciente de estar «agitado» que, de todas las molestias recurrentes de la vida, hace que una

verdadera reflexión sea lo más difícil o imposible. […] Esta excitación antinatural, esta perpetua disipación de la mente, esta pérdida de sentimiento en las escenas de una ópera, al final tiene que acabar por dañar la consciencia y la inteligencia; y esto, que se lo debemos a la electricidad, deber ser balanceado respecto a cualquier beneficio material que pueda ofrecer. No decimos nada más que el resultado universal del uso del telégrafo es sobrecargar la mente ordinaria con fragmentos de información indigeridos e indigestibles, e insistimos solo que su tendencia será debilitar y finalmente paralizar el poder reflexivo.

Puedes sustituir la palabra telégrafo por redes sociales, o por cualquiera de sus hijos (Whatsapp, Twiter, Instagram, Facebook, TikTok, YouTube, por citar algunos) y la reflexión parece que siga siendo válida. Esta historia del telégrafo nos muestra cómo una tecnología no es en sí misma ni buena ni mala. Pero tampoco es neutral, porque tiene sesgos a través de los valores que transmite. Por ello, tenemos que tomar una doble decisión cuando estamos con una tecnología: sobre sus fines y sobres sus valores. Determinar para qué queremos la tecnología (para hacer la guerra, para comunicar la paz) y qué valores aceptamos (rapidez o sosiego).

En lugar de la palabra telégrafo puedes poner cualquier otra tecnología. El resultado es el mismo. La diferencia con las tecnologías que tenemos actualmente es que sus fines o sus valores se ven aumentados por la inteligencia artificial. Pero esto lo veremos posteriormente, porque todavía tenemos que dar la bienvenida al teléfono. ¡El que faltaba!

El teléfono

Un invento indigno

La invención del teléfono hacia mediados del siglo XIX estuvo rodeada de asombro, miedo y suspicacia. Inventado por Antonio Meucci en 1854, fue formalmente patentado por Graham Bell en 1876, quien lo presentó públicamente en la exposición del Centenario en Filadelfia ese mismo año. Hoy en día, en una sociedad que parece móvil-dependiente, apenas podríamos vivir sin nuestro teléfono, pero en aquella primera presentación el invento pasó inadvertido para público y jueces. Tuvo que pasar por el estand de Bell el emperador de Brasil, Pedro II, conocido del propio Bell, quien lo probó y dijo con sorpresa: «¡Dios, esto habla!»[19].

A partir de entonces, el teléfono empezó a ser notorio, pero no apreciado. Se pensó que el teléfono era indigno para las personas. Todo aquel que probaba el invento se sentía estúpido hablando delante de un disco de metal, especialmente cuando tenían que gritar para hacerse escuchar. El teléfono podía ser un gran invento, pero ello no compensaba la pérdida de dignidad personal que suponía su uso.

Entramos de lleno en las consideraciones de valor, en las cuestiones morales. Hablar por teléfono no era moralmente adecuado porque era indigno. Apareció la cuestión de la tecnología y dignidad humana en su uso. El teléfono era una tecnología que nos degradaba como seres humanos. Hoy puede que no pensemos así y no nos causa rubor hablar por el móvil en la calle como si estuviéramos comiendo una rebanada de pan y a viva voz. Ahora pensamos que la inteligencia artificial nos degrada como seres humanos, porque parece que un conjunto de cables es como nosotros y nos supera. Con el tiempo, quizás, acabemos perdiéndole el respeto a la inteligencia artificial, igual que con el teléfono.

En aquel entonces, una de las causas de considerar al teléfono indigno era el desconocimiento de la tecnología que se estaba usando. No se entendía cómo hablando delante de un disco de metal se podía transmitir la voz por un cable. Bell fue acusado de impostor, y se dijo que era un simple

ventrílocuo. El teléfono era un engaño y solo los incautos podían pensar que aquello era verdad.

The London Times defendió con orgullo que era imposible transmitir la voz por un cable debido a la naturaleza intermitente de la corriente eléctrica. El físico Joseph Henry, famoso por sus estudios en electromagnetismo, aseguró que ese invento era imposible porque contradecía la ley de la conservación de la energía —esa que viene a decir que la energía ni se crea ni se destruye, solo se transforma; ¡a saber por qué pensó eso!—. *The New York Herald* dijo que el efecto de aquel aparato era sobrecogedor y casi sobrenatural y *The Providence Press* dijo que las fuerzas de la oscuridad de alguna forma estaban aliadas con el aparato. Tan solo un mecánico de Boston llegó a dar con la explicación sobre el funcionamiento de aquel extraño aparato parlante. Su funcionamiento se basaba en un orificio a lo largo de todo el cable por el cual se transmitía la voz. Siempre han existido respuestas simplistas que buscan tranquilizar los pensamientos.

El desconocimiento de cómo funciona una tecnología no ayuda a que esta sea aceptada. El teléfono era un engaño, o bien un invento del lado oscuro, porque se ignoraba su funcionamiento. La inteligencia artificial nos asusta porque nos asombra y no comprendemos cómo es capaz de hacer todo lo que vemos de ella. Pensamos que usurpa algunas de nuestras funciones como seres humanos, en particular la capacidad de pensamiento. En capítulos posteriores empezaremos a desmitificar tal asombro.

Si desde un punto de vista moral el teléfono era impropio, desde un punto de vista económico no corría mejor suerte.

Ni a banqueros ni a tenderos

Tras el recelo y la desconfianza vino la indiferencia y el menosprecio. No se veía la utilidad del teléfono. Debemos entender que no se veía su beneficio económico. No había respuesta clara a la pregunta «¿Cómo puedo ganar dinero con este invento?». Se pensaba que estaba bien como divertimento para los científicos de laboratorio, que podían jugar con él para entender el funcionamiento de la electricidad, pero no se apreciaba un uso razonable para el resto de personas. Al menos un uso por el cual se pudiera cobrar.

De forma graciosa se decía que los banqueros pensaban que el teléfono podría ser útil para los tenderos, si bien nunca sería útil para los banqueros; y los tenderos decían que podría ser útil para los banqueros, pero nunca útil para los tenderos. Ninguno de los dos vio negocio en el teléfono. ¿Qué pasaría si les preguntáramos hoy?

The Boston Times, en un editorial burlón, dijo que con el teléfono uno podía cortejar a una chica, ya estuviera en China o en Boston. Pero el aspecto más serio y alarmante del invento, continuaba el editorial de forma irónica, era el poder irresponsable que daría a la mayoría de las suegras, las cuales serían capaces de enviar su voz por todo el globo habitable[20]. En resumen, no se veía utilidad seria —es decir, económica— al teléfono y solo era la diana apropiada para dardos satíricos.

Una vez que hemos desprestigiado al teléfono, corresponde desprestigiar a su inventor. Si la pobre Hilma af Klint fue ninguneada después de su muerte por sus inclinaciones espirituales, Graham Bell no corrió mejor suerte, porque no se veía la necesidad de su invento. Cuando pensamos que una obra no es como nosotros pensamos que debe ser, pasamos a desacreditar al autor de la obra. ¡Fuera Hilma, fuera Graham Bell! Es la llamada falacia *ad hominem*, que dicho de manera común significa que, como, en realidad, no tengo argumentos racionales en contra de tus ideas, pues me meto contigo. Es más eficaz porque exige pensar menos. Esta falacia se ve mucho en la esfera pública, entre aquellos que nos gobiernan y otros desgobernados que habitan en las redes sociales.

El teléfono en sus orígenes era un desperdicio tecnológico: ni moralmente aceptable ni económicamente viable. Ahora nos falta abordar la cuestión de la seguridad.

Un montón de gelatina transparente

Según el teléfono se iba implantando en la sociedad, y, en particular, en los hogares, apareció la sombra de la pérdida de privacidad. ¡La seguridad sale a escena! Se percibía como un gran peligro que miles y miles de hogares estuvieran conectados por cables, capaces convertir las intimidades caseras en algo público. Se pensó que la intimidad iba a desaparecer por completo. En

las editoriales de los periódicos se podían leer frases tan amenazantes como «¿Qué será de la privacidad de la vida?», o bien, «¿Qué será de la santidad del hogar doméstico?»[21]. La grandilocuencia siempre abona los miedos.

A propósito de la instalación del tendido de líneas de teléfono en el estado de Rhode Island, *The New York Times* publicó que los técnicos que instalaron los cables pudieron escuchar a clérigos parlanchines, canciones melodiosas o gatos de medianoche, así como las conversaciones confidenciales de cientos de matrimonios[22]. La noticia, fuera verdadera o no, revela la preocupación del momento por la privacidad del teléfono. ¡Si vieran la situación hoy en día…!

Partiendo de esta premisa de ataque a la privacidad, el argumento derivó, de nuevo, en aspectos morales. El teléfono iba a afectar a la esencia humana. Nada más y nada menos. El teléfono nos iba a transformar como sociedad y como personas, por esa virtud que tiene de transmitir el pensamiento a distancia. Con el teléfono se rompería la esfera de lo privado, y eso acabaría con lo más profundo del ser humano.

Siempre nos hemos preguntado por la esencia del ser humano. Por aquello que nos distingue de cualquier otro ser vivo, en particular de los animales. En aquella época del teléfono naciente se pensaba que un elemento de la esencia del ser humano era la privacidad. Esta visión tiene su fundamento, pues se desconoce que, por ejemplo, las vacas tengan vida privada —aunque quizás la tengan y esta sea tan privada que no la conocemos—. La vida privada corresponde a esa parte de nuestra existencia sobre la cual no queremos dar cuenta públicamente y que nos identifica como individuos en lo más profundo.

Una publicación británica especializada en tecnología vaticinaba que pronto seríamos el uno para el otro un montón transparente de gelatina[23]. El teléfono nos iba a convertir en una masa social uniforme sin la singularidad propia de cada persona. La frase me parece significativa, porque habla de *transparencia* y de *gelatina*. No sé si por primera vez, pero quizá sí una de las primeras veces, se hablaba de la transparencia que causa la tecnología y de la pérdida de valor único personal que esto puede suponer. Esa transparencia nos lleva a una masa informe de gelatina, fácilmente moldeable, donde todos somos lo mismo. ¿Hoy también? ¿Son los

macrodatos, el *big data*), tratados por la inteligencia artificial, la nueva masa de gelatina que nos convierte en un simple conjunto de números?

El teléfono móvil, dotado hoy en día con inteligencia artificial, invade nuestras vidas y también nos surgen dudas sobre lo que debe y no debe ser. Lo que sí está claro es que a finales del siglo XIX el teléfono era algo que no debía ser. Este ha seguido adelante y hemos ido superando todos los miedos y suspicacias, o quizás nosotros nos hemos ido adaptando con el tiempo.

Hoy no tenemos la sensación de perder nuestra esencia humana si hablamos por teléfono. Hemos visto que la esencia humana es más que la vida privada. Pero hoy sí pensamos que la esencia humana se ve atacada por la inteligencia artificial. Quizás quepa tener en cuenta la historia del teléfono. Con el teléfono hemos aprendido sobre nuestra naturaleza humana, con la inteligencia artificial sucederá lo mismo. En el capítulo 3 lo veremos.

De momento, damos ahora un salto gigante en el tiempo. Atravesamos corriendo el siglo XX y nos plantamos en la actualidad, con esto de la inteligencia artificial. ¿Es la inteligencia artificial lo que debe ser?

La inteligencia artificial

Nos roban los trabajos

Una de las primeras voces que alarmaron sobre la inteligencia artificial fue la de Nick Bostrom en 2014 con su libro *Superinteligencia: caminos, peligros, estrategias*[24], en el cual alertaba de los posibles peligros de una superinteligencia y pedía que esta se desarrollara con cautela. Consejo bien razonable. Nick Bostrom apenas era conocido por el público general, por lo que sus bien intencionadas ideas quedaron en el ámbito científico. Sin embargo, cuando la alarma fue retransmitida por grandes personalidades del mundo científico y tecnológico, como Bill Gates[25], Elon Musk[26] o Stephen Hawking[27], entonces nos quedamos más intranquilos.

Una de las primeras preocupaciones de la inteligencia artificial es que nos pueda suplantar en nuestros trabajos. «¿Suplantar al ser humano? ¡Eso no puede ser!». Pero calma, puede tener su explicación.

Por ejemplo, si mandamos una nave terrestre no tripulada para que corretee por los páramos de Marte, necesitamos que esté gobernada de forma autónoma por una inteligencia artificial. Supongamos que la nave terrestre estuviera dirigida desde la Tierra vía radio. Desde una sala de control se vería la situación del terreno de Marte a través de cámaras de vídeo situadas en la nave marciana. En función de lo que se viera, se podría actuar dando la orden de girar, continuar o frenar. Sería similar a volar un dron.

La diferencia con volar un dron radica en el tiempo de comunicación. Debido a la distancia entra la Tierra y Marte, el tiempo medio de comunicación entre los planetas es de unos 12 minutos. Esto significa que, si a través de las cámaras de vídeo vemos que la nave se dirige al agujero de un cráter y mandamos la orden de parar, esta llegará a la nave 12 minutos después. Pero la cosa es peor. Cuando nosotros vemos por el vídeo que la nave va hacia un cráter, eso ha ocurrido hace 12 minutos. Por tanto, desde que la nave se enfila hacia el cráter, hasta que recibe nuestra orden de frenar, habrán pasado en total 24 minutos. Para entonces, es posible que el

carricoche marciano ya haya frenado por sí solo y para siempre en el fondo del cráter. En esta situación es necesario suplantar al piloto humano en la Tierra por una inteligencia artificial, que conduzca el vehículo directamente desde Marte.

De acuerdo, conducir coches por Marte necesita de la inteligencia artificial; sin embargo, este hecho no es preocupante, pues actualmente no hay muchos puestos de trabajo de chófer marciano que puedan ser sustituidos por una inteligencia artificial. La preocupación surge con el resto de actividades que sí pueden ser sustituidas por un sistema inteligente, en particular aquellas que se consideran rutinarias o peligrosas.

En este punto la situación es muy similar a lo que hemos visto con la llegada del ferrocarril, el cual sustituyó a los cocheros de larga distancia, pero favoreció a aquellos transportes que te acercaban de la estación de tren a la población más cercana. Además, creó a su vez el empleo de conductor de tren, que era de más valor que el oficio de cochero. Toda nueva tecnología destruye puestos de trabajo y crea nuevos de más valor. Ahora bien, esto tiene un límite.

Según un informe de la OECD de 2019[28] los puestos de trabajo con menor riesgo de sustitución por una inteligencia artificial son aquellos que requieren habilidades más difíciles de automatizar, tales como la resolución de problemas, la planificación, la asesoría o la capacidad de influencia. Por ello, aquellas personas que trabajen como ejecutivos de alto nivel, directores, profesores, abogados, tecnólogos, desarrolladores de negocio, científicos o ingenieros pueden estar más tranquilas: la inteligencia artificial no se va a meter con ellos, o se va a meter menos. Sin embargo, trabajos de limpiadores, cuidadores, montadores, conductores, operadores de planta, procesadores de comida, agricultores u obreros de la construcción tienen más riesgo de sustitución.

Solución: formemos a los limpiadores, cuidadores, etc., en las habilidades difíciles de automatizar. La complejidad radica en que tales habilidades (resolución de problemas, la planificación, la asesoría o la capacidad de influencia), por el mismo hecho de ser difíciles de automatizar, son también difíciles de adquirir. Resulta complicado ver un mundo en el que todos seamos ejecutivos de alto nivel, directores, abogados, científicos o ingenieros.

Una posible solución es la renta básica universal (UBI en inglés, *Universal Basic Income*). Técnicamente consiste en disponer de una subvención mensual para todos los miembros de una comunidad, con independencia de sus ingresos y de su mérito personal, y a un nivel suficientemente alto para permitir una vida libre de inseguridad económica. Dicho en lenguaje común, significa cobrar por existir para poder vivir.

El remedio no es tan claro. Existen visiones encontradas sobre las ventajas e inconvenientes de una renta básica universal[29]. Desde un punto de vista económico se discute sobre cómo financiar tal ingreso. Es decir, ¿esto quién lo paga? También si éste favorecerá la inflación, haciendo subir los precios y, por tanto, la renta básica será insuficiente para cubrir un nivel de vida mínimo.

Se unen, además, cuestiones de ámbito moral. Cuando dependes de una subvención, pierdes parte de tu libertad, pues recibes una renta, pero dejas de tener acceso a los medios que producen tal renta. Pasas a depender de aquel que te da la pensión. También se ponen en peligro los principios de responsabilidad y reciprocidad. Cuando recibes algo a cambio de nada, no hay reciprocidad en el intercambio y se desbalancea la responsabilidad, porque pasas a tener derechos sin obligaciones. El derecho de recibir un ingreso, sin la obligación de hacer algo.

Actualmente la cuestión está en el debate académico. Con el tiempo, pasará al debate político, pues el tema es un buen caladero de votos, y entonces primará más la ideología que el raciocinio. Hasta entonces, existen otras preocupaciones más acuciantes.

También se equivoca

La inteligencia artificial, como todo sistema electrónico, es susceptible de tener brechas de seguridad y, por tanto, de ser atacado. Sin embargo, lo realmente significativo son los ataques de seguridad que buscan que el sistema inteligente se equivoque.

En un artículo de 2015 presentado por unos investigadores de Google[30] se demostraba que se podía confundir a una red neuronal en la clasificación de imágenes. Inicialmente, mostraban la foto de un oso panda, que el sistema

inteligente reconocía con una fiabilidad del 57%. Posteriormente, esa foto era tratada digitalmente de tal manera que se modificaba ligeramente en su composición de píxeles. A simple vista, para nosotros, la foto seguía siendo totalmente un oso panda, sin embargo, el sistema inteligente lo reconocía como un gibón, con una probabilidad de un 90%.

El tema no pasaría de ser algo anecdótico y simpático si no fuera porque esta vulnerabilidad se podría utilizar para el crimen. Se ha demostrado que, llevando unas gafas de una montura especial con llamativos colores, se puede falsear el reconocimiento facial. Con tales gafas, una persona puede pasar por una actriz de cine a ojos de un sistema inteligente[31].

Estos temas de la seguridad, en toda innovación, son naturales y pasajeros. El tren en sus comienzos tuvo accidentes. Lo mismo ha ocurrido con la aviación y ahora ocurre con los vehículos autónomos. Con el tiempo se va mejorando la seguridad y, hoy en día, el ferrocarril y la aviación son medios de transporte suficientemente seguros, sabiendo que no existe el riesgo cero.

Así ocurre con la inteligencia artificial. Los errores de interpretación de la inteligencia artificial provocados intencionadamente se solucionan con lo que se llaman ejemplos adversos. Consiste en entrenar a los sistemas inteligentes con aquellos casos que pueden dar a error, e indicarle cuál es la solución correcta. Por ejemplo, en el caso del oso panda, el entrenamiento con ejemplo adverso consiste en enseñarle al sistema inteligente la foto ligeramente modificada del oso panda —la que confundía con un gibón— y decirle que no se confunda, que es un oso panda.

La cuestión de la seguridad es algo que nunca tiene fin. Siempre habrá personas que dedicarán toda su capacidad y buen hacer en ver cómo alterar un sistema inteligente. Frente a tales ataques, surgirán contramedidas. Entonces, esas mismas personas dedicarán ahora todos sus esfuerzos a evitar tales contramedidas. Saldrán nuevas contramedidas y así estaremos, en una historia interminable, que dará bien de comer a muchos ingenieros de seguridad.

Mientras pasamos el tiempo entre medidas y contramedidas de seguridad, podemos ir atendiendo esas cuestiones morales que nos hablan de autonomía y verdad.

Despotismo digital

Padecemos exceso de buenismo. Lo hemos visto con ferrocarril y el telégrafo. De primeras se nos muestra un mundo feliz gracias al nuevo progreso. Con el ferrocarril, los zapateros venderían más zapatos que pares de pies hay en el mundo; con el telégrafo se acabaría la barbarie; la electricidad aumentaría la felicidad de las personas, porque dispondrían de más calidad de vida. Después, la realidad es distinta.

En el caso de la inteligencia artificial, esta se pone al servicio de las redes sociales con el sano y loable objetivo de conocer al usuario para mejorar su experiencia de usuario. ¡Oh, palabra sacrosanta hoy en día! Suéltala en cualquier reunión y toda cabeza se doblará. Todo sea por la experiencia del usuario.

Mediante la mejora de la experiencia de usuario se pretende que te sientas más cómodo, más atendido y más feliz. «¡Qué buenos son conmigo! Lo hacen por mi bien». Justo ese es uno de los males más perniciosos: cuando te obligan a realizar acciones que supuestamente son para tu beneficio. Eso es el despotismo digital.

En la decadencia de las monarquías absolutistas, a finales del siglo XVIII, apareció el despotismo ilustrado. La mentalidad de aquellos monarcas despistados era la de obrar supuestamente por el bien del pueblo, pero sin contar con el pueblo: «Todo para el pueblo, pero sin el pueblo». Ahora tenemos una monarquía tecnológica, donde la inteligencia artificial es el gran valido, que nos lleva a una especie de despotismo digital: «Todo para el usuario, pero sin el usuario».

El tema puede parecer menor, pero es muy relevante. El filósofo John Stuart Mill lo trató con detalle en su obra *Sobre la libertad*[32], a mediados del siglo XIX, donde advierte sobre la tiranía de la opinión de la mayoría. Mill admite que a una persona se le pueda impedir realizar ciertas acciones, si ello supone evitar un daño a un tercero. Pero nunca está justificado obligar a alguien a realizar, o no realizar, determinados actos porque eso sea bueno para él, porque esto le haría más feliz, o porque, en opinión de una mayoría, hacerlo sería más acertado o más justo.

Mejorar tu experiencia de usuario es buscar ese supuesto bien para ti, pero sin contar contigo. De eso ya se encarga la inteligencia artificial. De pensar por ti y saber lo que de verdad te gusta. Pero no te molestes ni te enfades, que es por tu bien. ¿Cómo consigue la inteligencia artificial pensar por nosotros? Muy fácil.

Lo primero es conocer tu personalidad y para conseguirlo no hace falta una gran cantidad de información. Se ha demostrado que se puede predecir, con una precisión entre el 60% y el 80%, tu personalidad como usuario de redes sociales simplemente analizando con *machine learning* dos elementos: tus contactos y tus publicaciones[33].

Posteriormente, se pone en juego el denominado Modelo de los cinco cactores o de los cinco grandes[34], que te clasifica de manera rápida. Según este modelo, la personalidad se compone de los siguientes cinco factores: apertura a la experiencia, en qué medida eres curioso o cauteloso; meticulosidad, si eres organizado o descuidado; extraversión, cómo eres de sociable o reservado; simpatía o tu capacidad de ser amigable, compasivo o bien insensible; y neurosis, o en qué medida tienes inestabilidad emocional.

Dependiendo de los valores en estas variables, así te comportas[35]. Si eres una persona extrovertida o abierta a las experiencias, tienes más probabilidad de usar las redes sociales, publicar más fotos o hacer más actualizaciones de tu perfil. Si tienes bajos niveles de neuroticismo —poca inestabilidad emocional — o altos valores en meticulosidad, utilizas menos las redes sociales, actualizas menos su estado y tienes menos probabilidad de tener adicción. Si eres un usuario con altos niveles en simpatía, sueles compartir fotos sobre tus actividades y visitar tu página y la de tus amigos con frecuencia.

Con este conocimiento de tu personalidad, la inteligencia artificial hace el resto: te propone recomendaciones ajustadas a eso que llaman tu perfil, que son los valores matemáticos que conforman tu personalidad. Así, si YouTube detecta que te gusta ver patochadas, conociendo tu personalidad te propone más vídeos de necedades; Netflix te sugiere el mismo tipo de película que siempre ves; o una publicación te muestra una publicidad personalizada que te invita a comprar ese producto que tú mismo ignoras que necesitas. El objetivo último es evitar que abandones las redes sociales y facturar lo máximo posible.

Este despotismo digital puede afectar a nuestros niveles de autonomía, es decir, a nuestra capacidad de tomar decisiones. John Stuart Mill, al hablar de la libertad, decía que a una persona no se le podía obligar a realizar, o no realizar, determinados actos por un supuesto bien para él. Ciertamente, Mill hablaba de obligar y a la hora de aceptar una recomendación en una red social no existe una obligación *de facto*. Lo que existe, en su lugar, es una incitación, pero una incitación tan bien creada que puede llegar a anular tu capacidad de toma de decisiones. Una seducción que se potencia por la inteligencia artificial. No te obliga, pero influye en tu capacidad de decidir. Y no nos engañemos, no es por nuestro bien, es para obtener más beneficio.

Si en su momento acabamos con el despotismo de las monarquías absolutistas, debemos evitar caer en el despotismo digital avalado por la inteligencia artificial. Está en juego nuestra autonomía.

Esto es de justicia

Existe un error muy extendido en nuestra forma de hablar cuando al referirnos a sistemas digitales decimos expresiones del estilo: «el sistema dice que usted…», «la aplicación le ha denegado su petición de…», «el sistema ha aprobado...». En estos casos aceptamos por bueno el resultado matemático de un algoritmo: «Si lo dice el sistema, entonces es que es así». Este hecho es todavía más relevante si interviene la inteligencia artificial, porque su resultado es más convincente. Pero este resultado puede no ser tan bueno, en el sentido de no ser justo o ecuánime. Esta falta de justicia puede tener dos causas: que sea por sesgos no previstos, o bien por decidir el presente en función del pasado. En los dos casos se resiente la justicia, y en el segundo, además, la posibilidad de cambiar tu vida.

Nuestra aspiración es ser justos, en el sentido de que existan igualdad de oportunidades para todos, con independencia de la cultura, sexo o ideología de cada uno. Intentamos tomar decisiones sin sesgo, ecuánimes. Como ahora las decisiones las tomamos a través de la inteligencia artificial, entonces debemos evitar que esta manifieste algún tipo de sesgo en su funcionamiento.

Uno de los primeros y más famosos casos de sesgo sucedió con Tay, un *chatbot* de Microsoft, diseñado para interactuar con jóvenes entre 18 y 24 años mediante conversaciones casuales y entretenidas a través de Twitter. Su método de aprendizaje consistía en dicha interacción con los jóvenes. A través de la propia conversación con los jóvenes se esperaba que el *chatbot* fuera aprendiendo. El resultado fue que en menos de 24 horas Tay se volvió nazi y acabó lanzando comentarios racistas y antisemitas[36]. Microsoft cerró el *chatbot* inmediatamente y los jóvenes entrenadores de Tay tuvieron algo de lo qué vanagloriarse ante sus amigos.

El caso puso de manifiesto lo importante que es entrenar a las máquinas según unos datos adecuados. De igual forma que no basaríamos la educación de un niño en la lectura indiscriminada de tuits, Twitter no es el medio más adecuado para enseñar a un sistema inteligente.

Si los datos de entrada para el aprendizaje de un sistema inteligente tienen sesgos, los resultados de ese sistema inteligente estarán sesgados. Para que el sistema ofrezca otro resultado, debe tener otros datos de aprendizaje. La inteligencia artificial no es tan inteligente como para decir, «esto que hago está mal, no estoy siendo justo, tengo sesgos» y cambiar su comportamiento. Tay no sabía que el antisemitismo era impropio de la justicia. Tan solo tenía una mayor cantidad de tuits con comentarios antisemitas y la inteligencia artificial asigna el hecho de verdad o de bueno a lo que es más abundante.

Desde entonces, cada vez más la inteligencia artificial se viene utilizando de manera rutinaria en múltiples aspectos que afectan a nuestras vidas: desde acceder a una entrevista de trabajo, obtener la libertad condicional o recibir un préstamo. Todas estas de decisiones se pueden ver afectadas por algún tipo de sesgo, lo cual puede afectar a tomar una decisión justa.

Si el sistema que recomienda candidatos para una entrevista de trabajo tiene datos de entrada con más hombres que mujeres, recomendará candidatos varones con más probabilidad. Si al sistema que recomienda sentencias le entrenamos con datos de delincuentes reincidentes, la mayoría de ellos negros, recomendará con mayor probabilidad denegar la libertad condicional a una persona negra. Por fortuna, esta debilidad del sesgo en la inteligencia artificial está detectada y su resolución depende de dos variables: tiempo y voluntad.

Parte de los actuales sesgos provienen del propio sesgo que tienen las bases de datos de entrada de los sistemas inteligentes debido a nuestra propia historia. Por ejemplo, si en un sistema de reconocimiento de imágenes incluimos mayoritariamente fotos de cocinas donde se encuentran mujeres, porque son las fotos más abundantes que tenemos por cuestiones históricas, el sistema inteligente determinará que una persona en una foto de una cocina es una mujer[37]. Con el tiempo tendremos bases de datos más diversificadas. Fotos de cocinas con hombres y mujeres. Ya solo quedará el segundo paso, tener la voluntad de usar dichas de bases de datos sin sesgo.

La segunda causa de injusticia es quizás más problemática, pues afecta a la esencia misma de la inteligencia artificial, la cual trabaja siempre con datos del pasado, y en función de ellos intenta predecir el futuro. Pero esos datos estáticos del ayer no conciben la posibilidad de que algo pueda cambiar. Tu presente queda determinado por tu pasado. Una inteligencia artificial mal utilizada nos lleva a ser cautivos por siempre de nuestra historia. Te niega la posibilidad de cambio en tu vida. Esto se puso de manifiesto en Reino Unido en 2020 a raíz de la pandemia de la COVID-19.

Una de las muchas consecuencias de este bicho coronado fue la suspensión de las clases presenciales en los colegios e institutos durante la primera ola. La ausencia prolongada durante meses de clases, hizo que no se pudieran celebrar los exámenes de final de curso. Como solución inteligente se recurrió a la inteligencia artificial. Se decidió que la nota final de cada alumno para ese año académico se calculara por un algoritmo que tuviera en consideración dos variables: las calificaciones previas de años anteriores de dicho alumno y el rendimiento de su escuela o instituto en cuanto a la media de los resultados históricos de sus alumnos[38].

Esta decisión supuso un gran número de protestas por parte de los alumnos. No era para menos. La nota de un alumno del curso escolar 2019/2020 dependía de cuánto había estudiado ese alumno y el resto de alumnos del instituto en años anteriores. Entonces, ¿de qué le había servido estudiar ese curso? Ya se hubiera esforzado o no, su nota dependió de lo que hizo en los años previos. «Elegí un mal año para empezar a estudiar». Seguro que esto es lo que pensó un alumno que, en septiembre de 2019, antes de la pandemia, decidió cambiar de actitud y empezar a estudiar. La inteligencia

artificial le negó la posibilidad de cambio. Su futuro estaba irremediablemente condicionado por su pasado.

En nuestra vida, día tras día, tenemos la posibilidad de dejar de hacer algo que quizás hemos venido haciendo hasta ese momento. Eso es ejercer nuestra libertad individual. No siempre lo practicamos, pues ejercer la libertad cuesta esfuerzo y resulta más cómodo hacer lo que siempre hemos hecho. Depende de nosotros. Pero la inteligencia artificial puede anular esa posibilidad de cambio, coartar la esperanza de que hoy no seas como eras ayer. La inteligencia artificial trabaja con matemáticas, opera con cálculos estadísticos. En las matemáticas, dos más dos siempre son cuatro; sin embargo, en nuestra vida la libertad nos ofrece la opción de sumar otro resultado.

Pero, ¡ojo!, como hemos visto en el telégrafo, depende de nosotros. La inteligencia artificial es solo una herramienta y la podemos usar de una forma u otra. En realidad, la inteligencia artificial no asignó las notas a los alumnos. Lo hicieron los responsables académicos, que en ese año de pandemia decidieron utilizar cierto algoritmo, con unos ciertos cálculos y estimaciones, para tomar decisiones sobre las vidas de los alumnos.

«Si lo dice el sistema, entonces es que es así», no es cierto. Detrás puede haber sesgos o malos usos de la inteligencia artificial. Delegar la toma de decisiones en la inteligencia artificial es acabar con la justicia. Esta nunca es el resultado de un cálculo matemático.

Autonomía y justicia

¿Debemos tener miedo de la inteligencia artificial?

Gracias al ferrocarril íbamos a disponer de más tiempo libre, seríamos más éticos y más fraternales. Con el telégrafo acabaríamos con la barbarie. Si la velocidad de aquellos trenes decimonónicos trastornara la mente, hoy no podríamos viajar en coche. Menos en avión. Si el telégrafo impedía un pensamiento sosegado, ¿quién podría pensar hoy en día? Con el teléfono habríamos perdido toda nuestra esencia humana y hoy seríamos un montón de gelatina.

Son las predicciones, alarmantes o bondadosas, de las innovaciones del pasado. La mayor parte de ambas visiones se refutan pasados muchos años, si bien para entonces no podemos pedir explicaciones a sus autores. ¿A quién le decimos que, a pesar del telégrafo, seguimos teniendo guerras? ¿Cómo le exigimos cuentas? Esperanzar o atemorizar sale gratis, bien lo saben lo embaucadores.

Con el tiempo encontramos un nivel medio entre los deseados beneficios y las temidas amenazas. Toda innovación tecnológica nos trae un cierto beneficio y un cierto perjuicio. Pero, al final, ni los bienes son tan bondadosos ni los males tan perniciosos. Con la inteligencia artificial sucederá lo mismo.

Toda tecnología no es ni buena ni mala en sí misma, pero tampoco es neutral, porque transmite valores. Lo hemos vivido, claramente, con el telégrafo: por el valor de inmediatez que tiene, tuvimos la esperanza de unir a toda la humanidad en una comunicación instantánea, y el temor de anular por completo nuestra capacidad de pensar. Lo que debemos aprender es que el resultado del telégrafo no depende del telégrafo. Depende de nosotros. Podemos usar la comunicación para unirnos en un proyecto común o para iniciar una guerra. Igualmente, hoy en día, podemos atender notificaciones por las redes sociales o apagarlas y pararnos a pensar.

Nosotros vamos a la guerra, nosotros nos paramos a pensar. Nosotros elegimos un valor determinado de una tecnología. No es tanto buscar una inteligencia artificial ética, sino cómo podemos ser éticos y responsable con

la inteligencia artificial. Más que tener miedo de la inteligencia artificial, debemos tener miedo de nosotros con la inteligencia artificial: con ella podemos ser extremadamente humanos o extrañamente fieros. Ese es su peligro, esa su fortaleza. Depende de nosotros.

Nosotros elegimos el valor de cada tecnología, dentro de los múltiples valores que esta presenta. Ahora bien, para la inteligencia artificial, ¿existe algún valor innato? ¿Existe algún verdadero riesgo ético de la inteligencia artificial?

En este repaso histórico de otras tecnologías hemos visto que muchos problemas iniciales se han ido resolviendo con el incremento de la madurez de la tecnología. Lo mismo ocurrirá con la inteligencia artificial respecto a temas de trabajo, errores propios de la inteligencia artificial o su seguridad. Iremos resolviendo los problemas según vayamos aprendiendo y mejorando la tecnología. Pero existen una serie de riesgos que son más profundos, porque están relacionados con la propia esencia de la inteligencia artificial y los valores que transmite. Me refiero a dos valores: el valor de la autonomía y el valor de la justicia. Son los que debemos vigilar.

La inteligencia artificial posee la idea de autonomía, es decir, de poder decidir. Esto es peligroso, porque puede afectar a nuestra capacidad de decidir. Lo hemos visto con el concepto de despotismo digital, donde la inteligencia artificial piensa por ti y te ofrece lo que aparentemente es bueno para ti, pero sin contar contigo.

Luego está la idea de certeza que induce la inteligencia artificial: lo que dice la inteligencia artificial es bueno, es lo que es. Muy peligroso. Lo repito por si no he sido claro: muy peligroso. Es llevar al máximo nivel la frase «lo dice el sistema», que significa, «el resultado es verdadero». De esta forma, si la inteligencia artificial dice que esta es tu nota académica, en función de tus notas pasadas, quiere decir que es verdad, que no puedes cambiar tu vida. O si ofrece un veredicto sobre ti respecto a lo que vas a hacer, en función de lo que hiciste, este es cierto, aunque tú tengas la intención de cambiar. La idea de verdad en la inteligencia artificial afecta a nuestra libertad y a nuestra justicia.

En la historia de Hilma af Klint dije que hubo un juicio de valor erróneo y un acierto. Respecto al juicio de valor, hemos visto que nuestras visiones sobre un hecho cambian. Algo que considerábamos impropio, con el tiempo se vuelve aceptable. Si hoy vemos a la inteligencia artificial como indigna, nuestra visión cambiará. El acierto estuvo en que Kandinsky, antes de exponer una obra abstracta, explicó en su libro *De lo espiritual en el arte* cómo entender una obra de arte abstracta. Eso es lo que debemos hacer con la inteligencia artificial: entenderla, tener unos conocimientos básicos de cómo funciona. Si entendemos su funcionamiento, perderemos los miedos infundados y conoceremos sus verdaderos riesgos. Vamos a ello en el siguiente capítulo.

Mis conclusiones. ¿Las tuyas?

En este capítulo hemos abordado dos cuestiones: ¿Debemos tener miedo de la IA? ¿Cuáles son los verdaderos riesgos éticos de la IA? Estas son mis propuestas de respuesta:

- A lo largo de la historia, toda innovación ha tenido ilusiones y temores, centrado en aspectos de salud, seguridad, economía y moral. En esta tabla tienes un resumen de lo que hemos visto.
- Con el tiempo se ha visto que muchos temores e ilusiones eran infundados. Con la inteligencia artificial ocurrirá lo mismo.
- Toda tecnología no es ni buena ni mala en sí misma, pero tampoco es neutral, porque transmite valores. El resultado ético de la inteligencia artificial dependerá de nosotros, del valor que elijamos.
- Por ello, no es tanto buscar una inteligencia artificial ética, sino cómo nosotros podemos ser éticos y responsable con la inteligencia artificial.
- En la inteligencia artificial existen riesgos que desaparecerán, pero debemos vigilar aquellos riesgos que forman parte de los valores que transmite la inteligencia artificial: el valor de la autonomía y el valor de la justicia.
- Para evitar miedos infundados de la inteligencia artificial debemos conocer cómo funciona.
- Estas son mis conclusiones, pero no las des por ciertas. Reflexiona con autonomía y justicia para extraer tus propias conclusiones.

	Ferrocarril	**Telégrafo**	**Teléfono**	**Inteligencia Artificial**
SALUD	La velocidad causa locura.	Nos hará perder la capacidad de pensar.		
SEGURIDAD	Se producen accidentes.		Perderemos la privacidad y la	Errores propios de la IA

			individualidad (masa informe de gelatina)	
ECONOMÍA	Destruirá trabajo (transporte por canales), pero mejorará otros (zapateros).			Destruirá puestos de trabajo.
MORAL	El tiempo libre por tardar menos en viajar lo dedicaremos a ser más inteligentes.	Acabará con la guerra.	Hablar por teléfono es indigno del ser humano.	Pérdida de autonomía. Pérdida de justicia (sesgos, sin libertad de cambiar).

Montañas y mar, Helen Frankenthaler, 1952. Fundación Helen Frankenthaler

Inteligencia probable

Este cuadro parece una acuarela, pero no lo es. En el verano de 1952, su autora, Helen Frankenthaler, pasó unos días en Nueva Escocia, Canadá, pintando distintas estampas de aquel paisaje costero. A su vuelta, ya en el estudio, realizó *Montañas y mar* con una técnica totalmente novedosa, que le permitió crear grandes áreas de colores difuminados, algo nunca visto hasta entonces. Nadie sabía cómo había sido capaz de crear aquellos colores tan tenues, utilizando además pintura al óleo, en lugar de acuarela. Era cosa extraña.

El cuadro causó admiración cuando lo presentó. Nadie entendía qué nueva técnica podía haber utilizado para crear aquella sensación etérea de colores y mostrar lo vaporoso de un paisaje costero. Al principio su obra no fue aceptada por la crítica. Tal y como la misma artista comentó muchos años más tarde de la presentación de la obra, *Montañas y mar* fue vista con ira, fue destrozada por la rabia. Algunos lo vieron como un gran trapo de pintura donde limpias tus pinceles, pero no algo para enmarcar[39]. Helen Frankenthaler tuvo que explicar su nueva técnica.

Habitualmente cuando se pinta un cuadro al óleo hay que preparar el lienzo. Es necesario recubrirlo de una capa que evite que la pintura entre por

los poros del lienzo y, de esta forma, los colores se mantengan vivos. Es lo que se conoce como imprimar el lienzo y, justo, lo que no hizo Frankenthaler. Ella, por el contrario, dispuso el lienzo de forma horizontal en el suelo y pintó directamente en él, sin prepararlo. Además, diluyó la pintura de óleo en aguarrás, de manera que la pintura se pudiera extender y desvanecer por el lienzo.

Una vez explicada la técnica, se acabó la rabia, se acabó la ira y vinieron los reconocimientos. Helen Frankenthaler se convirtió pionera del estilo artístico llamado pintura de campos de color. *Montañas y mar* fue calificada nada más y nada menos como la piedra de Rosetta de los campos de color (la original piedra de Rosetta sirvió para el gran hito de descifrar los jeroglíficos egipcios, de ahí la comparación). Se comparó con la obra *Impresión, sol naciente* de Monet, que dio nombre al impresionismo, porque su obra suponía una nueva etapa para el expresionismo abstracto. ¡Lo que hace explicar las cosas con tranquilidad! ¡Lo que relaja saber cómo funciona!

Se teme lo desconocido. Lo hemos visto en el capítulo anterior con el telégrafo, que estaba basado en la electricidad, la cual se consideraba como una fuerza invisible y maravillosa. ¿Cómo era posible que los mensajes llegaran de forma instantánea? Hoy el telégrafo no nos causa sorpresa, ni miedo. También lo vimos con el teléfono. ¿Cómo se puede transportar la voz por un hilo de cobre? Para dar respuesta a tan maravilloso hecho se inventaron soluciones tan peregrinas como ventriloquía o que el hilo de cobre era hueco y la voz marchaba por su interior, como por una tubería. Este desconocimiento sobre el funcionamiento del teléfono hizo que se viera como una amenaza para la persona. Hablar por un altavoz se consideró indigno del ser humano y nos avocaba a perder nuestra individualidad, a convertirnos en una masa informe de gelatina.

Lo mismo sucede con la inteligencia artificial. Nos da miedo porque no sabemos cómo funciona. Porque la llamamos inteligencia y la vemos como rival. ¡Inteligentes somos nosotros! De forma similar al telégrafo o el teléfono, surgen preguntas del estilo: ¿cómo es posible que hable?, ¿pero, cómo aprende?, ¿cómo puede escribir poemas o pintar cuadros? Y acabamos concluyendo que esto va a afectar a la dignidad humana.

En el capítulo siguiente veremos brevemente en qué sentido la inteligencia artificial se parece o no a nuestra inteligencia y si eso, por tanto, puede afectar a nuestra dignidad como seres humanos. Pero para entender eso, primero tenemos que ver cómo funciona la inteligencia artificial. El objetivo no es profundizar en detalles, no te asustes, no habrá fórmulas matemáticas, sino entender en qué principios técnicos se fundamentan los sistemas inteligentes. Para ello veremos el funcionamiento de algunas de las capacidades que podemos considerar más asombrosas de la inteligencia artificial, como puede ser el escribir o pintar, mantener una conversación, aprender o tomar decisiones.

Vamos a abrir la caja negra de la inteligencia artificial.

Di lo más habitual

Hablar con un sicoanalista-máquina

En 1950 Alan Turing publicó su famoso artículo *Computing Machinery and Intelligence*[40] donde plantea lo que denominó juego de imitación. En el juego de imitación original hay tres personas: un hombre, una mujer y un interrogador, que puede ser de cualquier sexo. El interrogador permanece en una habitación separado de los otros dos. El objetivo del juego es que el interrogador determine quién de las dos personas que están en la otra habitación es el hombre y quién la mujer. Para ello solo se puede valer de preguntas realizadas por escrito y discernir la cuestión en función de las respuestas que reciba.

Supongamos ahora que cambiamos al hombre por una máquina. La cuestión que suscita Turing es si en este caso el interrogador se equivocará o acertará a la hora de discernir quién es el hombre y quién la máquina en la misma medida que en la situación original. De esta manera, definió su famoso test de Turing, que determina que una máquina es inteligente si al hablar con ella —originalmente por escrito— creemos que estamos hablando con una persona.

Turing auguraba, en el mismo artículo, que en 50 años sería posible disponer de una máquina con la capacidad de hacer el juego de imitación de tal forma que un interrogador medio, en 5 minutos, el 30 % de las veces se equivocaría en la identificación; o bien, un 30 % de las veces la máquina pasaría el test de Turing.

Cuando Turing escribió su famoso artículo, no existía ninguna máquina capaz de hacer el juego de imitación. Lo que Turing se preguntaba es si sería posible que en algún momento hubiera alguna máquina capaz de pasar su prueba. En 1966 se creó la máquina ELIZA, que fue capaz de superarlo bajo ciertas condiciones. Podemos pensar que eso es un gran logro y que, desde entonces, las máquinas empezaron a engañarnos; pero no es así, pasar el test de Turing no es tan complicado.

ELIZA fue creada por Joseph Weizenbaum, profesor del MIT y considerado uno de los padres de la inteligencia artificial. ELIZA actuaba desde el papel de un sicoanalista que recibía preguntas y comentarios de un supuesto paciente. Este paciente es el interrogador, si pensamos en el juego de imitación, y es el que debe averiguar si está hablando con un sicoanalista o con una máquina. Toda la conversación sería por escrito. Según el propio artículo en el que se explica el funcionamiento de ELIZA, la conversación entre el paciente y el supuesto sicoanalista podría ser del siguiente estilo[41]:

PACIENTE (P): Todos los hombres son iguales.
SICOANALISTA-MAQUINA (SM): ¿En qué sentido?
P: Siempre nos molestan de una forma u otra.
SM: ¿Puedes pensar en algún ejemplo concreto?
P: Bueno, mi novio me hizo venir aquí.
SM: Tu novio te hizo venir aquí.
P: Dice que estoy deprimida la mayor parte del tiempo.
SM: Siento oír que estás deprimida.
P: Es verdad, soy infeliz.
SM: ¿Crees que viniendo aquí te ayudará a no ser infeliz?
P: Necesito ayuda, eso parece seguro.
SM: ¿Qué significaría para ti si tú tuvieras ayuda?
P: Quizás podría aprender a llevarme bien con mi madre.
SM: Háblame de tu familia.
P: Mi madre me cuida.
SM: ¿Quién más en tu familia te cuida?
P: Mi padre.
SM: Tu padre.
P: Tú eres como mi padre de alguna forma.
SM: ¿Qué parecido ves?
P: No eres muy agresivo, pero creo que no quieres que me dé cuenta.
SM: ¿Qué te hace pensar que yo no soy muy agresivo?
P: No discutes conmigo.
SM: ¿Por qué piensas que no discuto contigo?
P: Me tienes miedo.
SM: ¿Te agrada creer que te tengo miedo?

P: Mi padre tiene miedo a todo el mundo.

SM: ¿Qué más te viene a la mente cuando piensas en tu padre?

P: Amenazas.

SM: ¿Tiene eso algo que ver con el hecho de que tu novio te haga venir aquí?

La lógica que utilizó ELIZA para pasar el Test de Turing era bastante sencilla. Durante la conversación, ELIZA buscaba palabras clave en las frases del interrogador, si las encontraba, generaba una respuesta en función de reglas establecidas, por ejemplo, la palabra clave «deprimida» genera la frase «siento oír que estás deprimida»; si no las encontraba, reformulaba las palabras del interrogador, por ejemplo, «mi novio me hizo venir aquí», lo reformulaba diciendo «tu novio te hizo venir aquí»; o respondía con frases genéricas, por ejemplo, «ya veo», «por favor, continúa», que no están en el texto mostrado, pero sí en la lógica del programa.

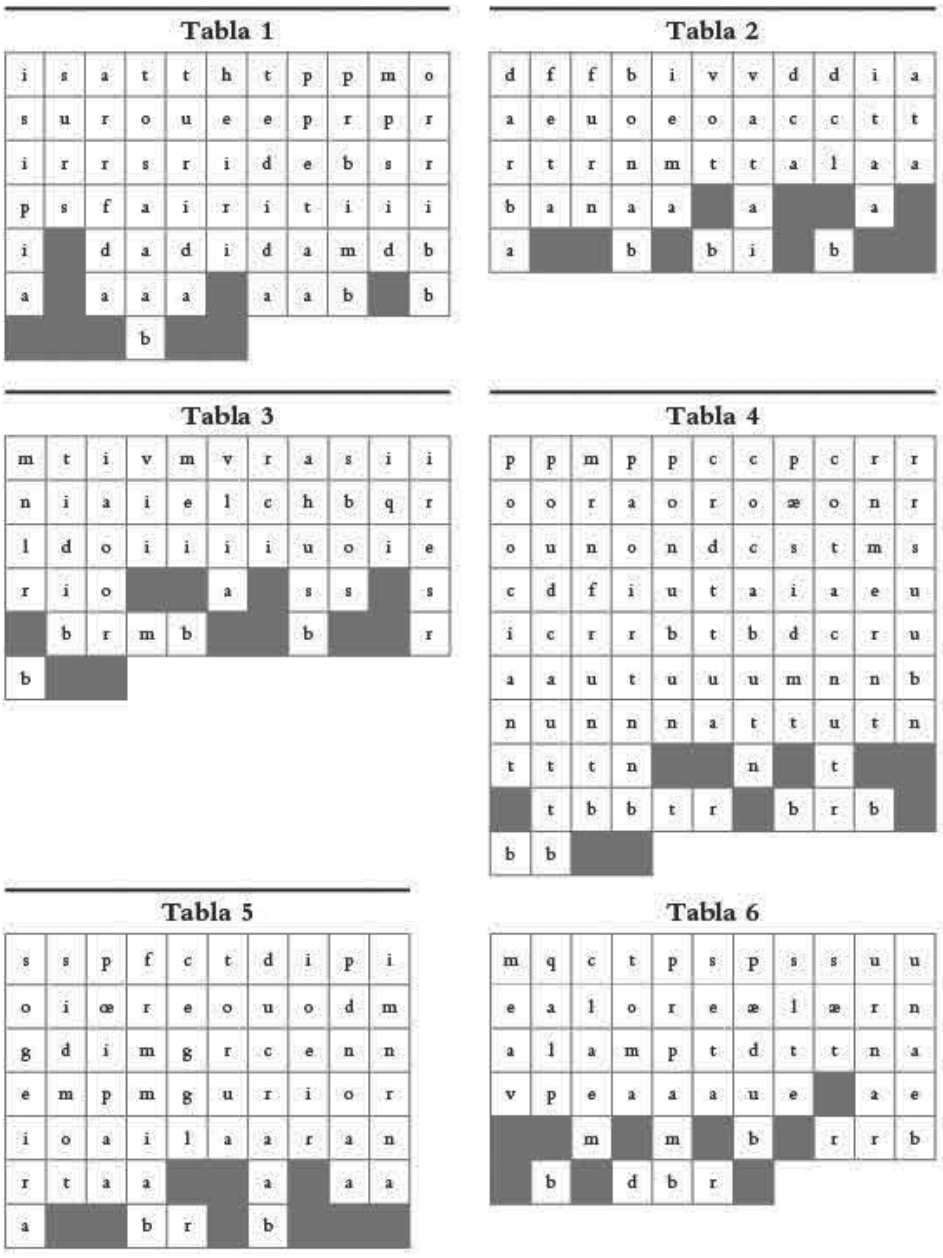
Como se ve, no es tan complicado pasar el test de Turing. Basta con estar en un entorno muy concreto, por ejemplo, un sicoanálisis, y hablar sin decir nada. Los seres humanos somos muy buenos en esto último de hablar sin decir nada y lo hacemos en muchas ocasiones. Ello no va en contra de los objetivos del test de Turing. Lo que Turing proponía era que, en el juego de imitación, una máquina se comportara como un ser humano y, ciertamente, los seres humanos no siempre decimos cosas inteligentes.

El objetivo último de Weizenbaum con ELIZA no era pasar esta prueba, sino investigar sobre lo fácil que es construir frases con significado semántico. En realidad, es relativamente fácil hacer que una máquina hable. Hablar es propio de los humanos y cuando vemos una máquina hablar nos quedamos sorprendidos. En 2018 Google hizo una demostración pública de cómo su asistente virtual realizaba una cita en una peluquería[42]. La persona que tomaba la cita en la peluquería no sabía que estaba hablando con el asistente de Google. En cierta forma, era un test de Turing encubierto y la demostración causó sensación. Sin ánimo de quitar mérito a los ingenieros de Google, que lo tienen, sí es verdad que no resulta tan complicado crear frases con sentido. Vamos a ver algunos ejemplos.

¿Quieres hablar en latín?

Para ver lo sencillo que es crear frases nos vamos a ir a 1677. ¿1677, cuando todavía no había inteligencia artificial? Con ello veremos que los rudimentos de la inteligencia artificial son muy básicos y antiguos.

Por aquel año, el matemático inglés John Peter publicó el libro *Versificador artificial*[43] que permitía crear frases en latín con sentido. Todo lo que hacía falta era elegir un número de seis cifras, saber contar del 1 al 9 y utilizar las tablas llenas de letras que te muestro a continuación.



El método es asombroso. Se pueden construir casi 600.000 frases en latín, sin tener ni idea de latín. Por ejemplo, con el número 421376 se obtiene la frase «Tristia verba aliis causabunt somnia certa». A falta de conocimientos en latín, tenemos a san Google, quien con su traductor nos dice que esa frase significa «Palabras tristes encenderán ciertos sueños». O con el número 157188 se genera la frase «Pessima bella tibi producunt sidera multa», que

viene a decir: «Las peores guerras han producido muchas estrellas». Si unimos los dos números ya podemos empezar a crear poesía, ¡y en latín!:

Tristia verba aliis causabunt somnia certa,
pessima bella tibi producunt sidera multa.

O bien en castellano:

Palabras tristes encenderán ciertos sueños,
las peores guerras han producido muchas estrellas.

Esto parece maravilloso, pero la ciencia que se esconde detrás de ello es bien sencilla y es la inspiración de la inteligencia artificial que hoy nos habla.

Simplemente, detrás de estas tablas con letras que parecen puestas al azar se esconden las siguientes listas de palabras:

	Palabras de la tabla 1	**Palabras de la tabla 2**	**Palabras de la tabla 3**	**Palabras de la tabla 4**	**Palabras de la tabla 5**	**Palabras de la tabla 6**
1ª palabra	pessima	dona	aliis	producunt	iurgia	semper
2ª palabra	turpia	verba	reor	concedunt	dogmata	prava
3ª palabra	horrida	vota	vides	causabunt	tempora	sola
4ª palabra	tristia	iura	malis	promittunt	crimina	plane
5ª palabra	turbida	bella	viro	portabunt	fœdera	tantum
6ª palabra	aspera	fata	inquam	monstrabunt	pignora	certa
7ª palabra	sordida	facta	tibi	procurant	somnia	quædam
8ª palabra	impia	dicta	mihi	prædicunt	sidera	multa
9ª	perfida	damna	scio	confirmant	pocula	sæpe

Cada cifra del número que elegimos es ahora la posición de una palabra en una lista. Por ejemplo, el número 421376 nos dice que tenemos que seleccionar la cuarta palabra de la primera tabla, después la segunda palabra de la segunda tabla, luego la primera palabra de la tercera tabla, y así hasta la última tabla. De esta forma obtenemos las palabras de la frase, que son las que aparecen en negrita.

Matemáticamente hablando, detrás del versificador artificial aparece la combinatoria. Lo que estamos haciendo es crear frases combinando palabras preestablecidas y con un orden semántico. Por ejemplo, en la lista de la tabla 2 tenemos nombres, que harán de sujeto, y en la lista de la tabla 4 tenemos verbos. Después, basta con combinar una palabra de cada lista siguiendo el orden establecido.

Estamos en 1677 y ya tenemos un sistema «inteligente» que es capaz de crear frases. Esta mecánica la podemos automatizar. Podemos pensar en una aplicación en la cual, cuando pinchas en un botón se genera un número aleatorio que posteriormente te devuelve una frase. Así tenemos un generador de frases en latín.

Pero, hoy en día ¿quién habla latín? ¿No sería más útil escribir, por ejemplo, palabras de amor?

Palabras de amor

Si la naturaleza no te ha dotado con el ingenio de decir inspiradas frases amorosas, el método que se esconde detrás del versificador artificial viene en tu auxilio. Así lo pensó Christopher Strachey[44] cuando en 1954 creó su máquina de redactar cartas de amor. Si no tienes palabras para esa persona que cuando la ves pierdes el sentido y se te cae el móvil de las manos, o llega el día de tu aniversario y no sabes qué decirle a tu amada pareja, prueba con esto:

Adorado cielo,
Tú eres mi ferviente cómplice de emociones. Mi cariño extrañamente se aferra a tu deseo apasionado. Mi empeño anhela tu corazón. Tú eres mi anhelante

simpatía: mi tierno empeño.
Tuyo bellamente.
M.U.C.

M.U.C. es el nombre de la máquina, que significa Manchester University Computer, un nombre nada poético para las palabras que dice. Sea como fuere, M.U.C. hace, en esencia, lo mismo que el versificador artificial de John Peter.

El escritor de cartas amorosas, M.U.C., dispone de una serie de listas con palabras de saludos o introducción (del tipo «adorado cielo»), listas con adjetivos, nombres, verbos y adverbios, como, por ejemplo:

Saludos 1	**Saludos 2**	**Adjetivos**	**Nombres**	**Verbos**	**Adverbios**
querido	cielo	afectuoso	devoción	anhelar	afectuosamente
estimado	corazón	amoroso	emoción	adorar	ardientemente
amado	vida	ansiado	ambición	atraer	bellamente
adorado		bello	empeño	aferrar	fervientemente
		anhelante	encanto	Esperar	extrañamente
		ardiente	deseo	desear	cariñosamente
		codiciado	afán	querer	tiernamente
		ferviente	hechizo	suspirar	amorosamente
		apasionado	cómplice	ansiar	tristemente
		tierno	corazón	gustar	
			simpatía	penar	

Al igual que con el versificador artificial, M.U.C. utiliza la combinatoria. La diferencia radica en que para construir las frases amorosas no basta con extraer las palabras sin más, sino que hay que introducir cierta inteligencia. Habrá que incorporar determinantes, tales como mi, mío o tú, tuyo, incluir algunos sujetos (yo, tú, nosotros), conjugar los verbos adecuadamente (en

este caso, por lo general en primera o segunda persona del singular) y acomodar el género según corresponda (amado/a). M.U.C utiliza la combinatoria junto con algunas reglas gramaticales.

Algo similar, pero todavía más elaborado, hizo el ingeniero informático alemán Theo Lutz, quien en 1959 publicó en la revista *Augenblick*[45] el primer poema escrito por un ordenador. La poesía se tituló «Texto estocástico», nombre no muy poético, pero que nos da cuenta de cómo fue su origen. El poema no nació de una inspiración iluminada, sino mediante técnicas estadísticas. Los dos primeros versos del poema dicen así:

No toda mirada está cerca. No todo pueblo es tarde.
Un castillo es libre y todo granjero está lejos.
Cada extraño está lejos. Un día es tarde.
Cada casa es oscura. Un ojo es profundo.

Viendo estos cuatro primeros versos, ya podemos entender algo de la lógica que inspira el poema: parece que a la máquina le gusta la palabra tarde y que juega con el par cerca, lejos. En inteligencia artificial decir que a un sistema le gusta es equivalente a decir que tiene mayor probabilidad. Hemos incorporado la estadística en esto de crear frases.

Todo esto parece maravilloso. Nuestra capacidad de hablar y de componer frases con sentido parece que también la tienen los sistemas inteligentes y que no es algo exclusivo de los humanos como pensábamos. Quizás estemos perdiendo uno de nuestros rasgos diferenciadores. Pero la realidad es distinta. Detrás de la construcción de frases en la inteligencia artificial se encuentra la combinatoria, las reglas gramaticales y reglas estadísticas.

Mira y aprende

El cerebro de Einstein

Un sistema muy relevante dentro de la inteligencia artificial es lo que se llama red neuronal. Las redes neuronales están inspiradas en el funcionamiento de nuestro cerebro. Pretenden simular nuestra forma de pensar. Se utilizan para muchos fines, pero, por ejemplo, son la herramienta básica para reconocer objetos, es decir, mediante las redes neuronales, la inteligencia artificial es capaz de ver. Pero, ¿cómo?

Para responder a esta pregunta vamos a recurrir a Hofstadter. Me refiero a Douglas Hofstadter, científico, filósofo y académico estadounidense, y no a Leonard Hofstadter, protagonista en la serie *The Big Bang Theory* (por cierto, al igual que le pasa a Penny, la novia de Leonard en la serie, siempre me cuesta escribir este apellido de Hofstadter). En 1981 Hofstadter publicó el famoso libro *The Mind's I: Fantasies and Reflections on Self and Soul* (*El yo de la mente: fantasías y reflexiones sobre el yo y el alma*), donde aparece un curioso capítulo en el que intervienen el griego Aquiles y una tortuga[46].

La tortuga le propone a Aquiles tener todo el cerebro de Eisntein en un libro. Para ello tendríamos que recopilar datos neurona a neurona de Einstein. Cada hoja del libro sería una neurona de Einstein y en cada una de ellas guardaríamos la siguiente información:

- Un valor umbral. Que indica la predisposición a que esa página la tenga que leer. Por ejemplo, un valor umbral alto significa que va a ser difícil que lea esa página. Leer una página del libro-Einstein es similar a encender o activar una neurona de su cerebro.
- Páginas que conecta. Las siguientes páginas del libro (otras neuronas) a las cuales debo ir cuando, después de leer la página en la que estoy.
- Valores de pesos. Cantidades que me indican cómo de fácil o complicado es ir a las páginas con las cuales se conecta la neurona en la que estoy.

- Valores de cambio. Un conjunto de números que me indican cómo en un momento dado se pueden cambiar los números de las páginas con las que se conecta.

Un libro así es físicamente inmanejable. Para que nos hagamos una idea, nuestro cerebro tiene del orden de 100.000 millones de neuronas. Supongamos que el cerebro de Einstein era de tamaño similar, neurona arriba, neurona abajo. Esto significa que el libro-Einstein debería tener ese mismo número de páginas, lo que nos llevaría a un grosor de unos 20.000 km. Si dividimos el libro en dos tomos, leyendo el primer tomo llegaríamos a Japón y con el segundo tomo, volveríamos a España. Imposible de manejar, pero un libro de estas características es el fundamento de programación de las redes neuronales.

Mi cerebro en números

En nuestro cerebro una neurona se activa cuando recibe suficiente señal de todas las neuronas que le preceden y con las cuales está conectada. Es entonces cuando deja pasar su señal a otras neuronas, las cuales a su vez se activarán si reciben suficiente señal de sus conexiones. Así sucesivamente, hasta llegar a un resultado, por ejemplo, saber identificar que he visto un 7 escrito. ¿Cómo sucede esto en una red neural?

Lo primero que tenemos que tener en cuenta es que actualmente no existe una red neuronal de ámbito general que sirva para todo, como ocurre con nuestro cerebro, que sirve para hablar, para ver, para movernos, para sentir, etc. Depende de para qué se entrene la red neuronal. Así, por ejemplo, una red neuronal se debe entrenar para ver o para moverse. Puede ser para las dos cosas, si bien su complejidad aumenta. Es similar a tener libros-Einstein especializados: libro-Einstein para ver, libro-Einstein para moverse, etc.

Pero, además, tenemos la especialización dentro de la especialización. Siguiendo con la red neuronal para ver, esta identificará objetos según cómo haya sido preparada. Depende de lo que queramos que identifique. Si queremos identificar números, debemos entrenar la red neuronal para ello, y no servirá, por ejemplo, para identificar caras. Para identificar caras necesitamos preparar la red neuronal de otra forma. O bien podemos

prepararla para que identifique ambas cosa, números y caras. Es similar a tener libros-Einstein de lecturas especializadas: libro-Einstein para números, libro-Einstein para ver caras, o libro-Einstein para ver ambas cosas. ¿Por qué un libro-Einstein está especializado? Por sus hojas finales. Cada libro-Einstein, cada red neuronal, es como una historia que puede tener varios finales. El libro-Einstein de identificar números, que podemos llamar libro-Einstein-ver-números, termina con diez posibles finales, con los finales del 0 al 9. Tiene diez hojas finales, con un número del 0 al 9 en cada una de ellas. Veamos entonces cómo funciona el libro-Einstein-ver-números.

Supongamos que queremos que este libro-Einstein particular vea un 7 escrito en un papel. Todo lo que tenemos para empezar es un conjunto de trazos negros sobre un fondo blanco. La disposición e intensidad de cada trazo en el papel me dice qué primeras hojas del libro debo leer. Hemos visto que en cada hoja tenemos una serie de valores: páginas que conecta, umbral (cómo de fácil se enciende la neurona), pesos (cómo de fácil se comunica con otras neuronas) y valores de cambio. Por ahora nos interesan los tres primeros valores. En cada hoja inicial de lectura, en función de una serie de cálculos con el umbral y con los pesos, obtendré a qué otras hojas del libro, con las cuales está conectada, tengo que ir. En cada una de estas nuevas hojas, unos nuevos cálculos me llevarán a otras hojas. Así sucesivamente hasta que termine en las hojas de final de historia.

Lo curioso es que no terminaré en una sola hoja de final de historia, sino que terminaré en todas. ¡En las diez hojas finales del 0 al 9! Pero según los cálculos realizados llegaré a estas hojas finales con distinta intensidad. Depende de cómo haya dibujado el 7 en el papel: no es lo mismo dibujar un 7 claro, que un 7 que se parece a un 1 o un 9. Estas diferencias hacen que llegue a las hojas finales de historia con distinta intensidad. Tendré, por ejemplo, un final de historia 0 con intensidad 3; un final de historia 1 con intensidad 40 (el 7 dibujado se parece algo a un 1); un final de historia 2 con intensidad 30; final 7 con intensidad 80; o un final 9 con intensidad 70 (el 7 se parece algo a un 9). Todos los finales de historia, del 0 al 9, tendrán su intensidad, no los he puesto aquí por no aburrir. La intensidad significa la probabilidad. Como el final de historia 7 tiene la intensidad más alta, el

libro-Einstein-ver-números concluye que lo más probable es que esté viendo un 7.

La conclusión de tanta explicación sobre el libro-Einstein es la siguiente: en una red neuronal unimos probabilidad e intención. Una red neuronal intenta simular el funcionamiento de nuestro cerebro. Si nuestro cerebro está formado por neuronas, una red neuronal está formado por unidades lógicas llamadas células, que dejan pasar una cierta señal según ciertas condiciones, hasta llegar a células finales. La red neuronal se prepara según la intención que se tenga con ella. Si nuestra intención es ver números, tendremos 10 células finales; si nuestra intención es mover un coche de juguete hacia adelante, hacia atrás, a derecha o izquierda, tendremos cuatro células finales. El resultado de la red neuronal será la célula final con más intensidad. Una red neuronal es similar a crear una novela con varios protagonistas y varios finales. Dependiendo de las primeras acciones que hagan sus protagonistas se determinará el fin más probable. Un primer paso desencadena ya un final.

La intención de una red neuronal depende de para qué se prepare dicha red neuronal. Preparar una red neuronal significa entrenarla, o, lo que es lo mismo, que la red aprenda. Entramos entonces en otro apartado misterioso de la inteligencia artificial, sobre qué es capaz de aprender. Pero primero necesitamos saber: ¿qué es aprender en inteligencia artificial?

Aprender es cambiar valores

En inteligencia artificial el término aprender significa que el sistema tiene la capacidad de modificar alguna de sus valores internos. Vimos que el libro-Einstein tenía una serie de valores en cada una de sus hojas. Si estos valores permanecen sin cambio, siempre que comencemos a leer el libro por unas determinadas hojas, llegaremos a la misma historia final. Pero si conseguimos que algunos valores de cada hoja cambien, entonces al comenzar a leer el libro por las mismas determinadas hojas llegaremos a un resultado distinto. Decimos entonces que el libro-Einstein ha aprendido. ¿Cómo conseguimos cambiar los valores escritos en cada hoja? De igual forma a como nosotros aprendemos.

Cuando éramos pequeños seguro que tuvimos libros de grandes dibujos que nos servían para identificar las cosas de nuestra vida. Seguro que veíamos un dibujo simpático de un león y nuestros padres nos decían: «esto es un león, le-ón»; o bien una oveja, y nos decían «o-ve-ja»; o siendo más mayores, veíamos un 7 y nos decían «sie-te». Lo mismo se hace con las redes neuronales.

Es lo que se conoce como fase de entrenamiento, que es el colegio de la red neuronal. Por ejemplo, a la red neuronal para identificar números se le manda al colegio de números y allí se pasa un buen rato viendo números escritos de distinta forma. Verá un montón de 7 escritos con distinto acierto, unos más claros, otros más difíciles de identificar, y en todos ellos le diremos que es un 7. Es decir, para todos ellos le decimos cuál es la historia final más probable que debe obtener. Si para un número que parece un 7 de aquella manera la red neuronal obtiene una historia final distinta, dice, por ejemplo, que es un 9, entonces cambia sus valores internos hasta llegar a la historia final de 7. Es entonces cuando ha aprendido a identificar un 7 escrito de aquella manera.

En el libro-Einstein uno de los valores que escribíamos eran los valores de cambio, que representaban cómo cambiar las hojas con las cuales se conecta cada hoja particular. Estos valores de cambio son la forma que tiene de cambiar sus enlaces a otras hojas y, por tanto, de obtener nuevos resultados, es decir, de aprender.

Este tipo de entrenamiento se llama aprendizaje supervisado, porque mandamos al sistema inteligente al colegio, donde unos profesores le enseñan. Pero en otras ocasiones la inteligencia artificial no va a la escuela, sino que aprende de la propia experiencia, acude a la universidad de la vida. Es lo que se denomina aprendizaje no supervisado. Consiste en identificar patrones de comportamiento.

Nosotros también lo hacemos. Si decimos «buenos días» y vemos que nos sonríen, entonces hemos identificado un patrón de comportamiento: decir buenos días es algo que gusta. De igual forma, una inteligencia artificial puede analizar tu comportamiento en redes sociales e identificar que cuando mandas mensajes tristes, coincide con ver muchas fotos de dulces de chocolate. Entonces el sistema ha aprendido que cuando estás deprimido te

da por comer chocolate. A partir de ahora, cuando detecte algún comentario tristón, es posible que te aparezca una sugerente publicidad de crujiente chocolate. Es aprender la vida.

De nuevo volvemos a la probabilidad. Aprender en la inteligencia artificial es decirle cuál es el resultado más probable, o bien que esta lo identifique. Por eso, en el capítulo anterior hablábamos de la ausencia de justicia en el caso de las notas a los alumnos de instituto de Reino Unido. Porque la inteligencia artificial estaba dando las notas más probables para una historia previa que había aprendido. Pero ya sabemos que las mejoras novelas son aquellas que tienen el final menos esperado, el más creativo.

Hablando de creatividad, ¿puede la inteligencia artificial ser creativa?

Mira y copia

Manierismo inteligente

En el año 2016 apareció un cuadro con el singular título de *El próximo Rembrandt*. Era el retrato de un hombre de mediana edad, con bigote y perilla, posando de lado hacia la derecha y sobre un fondo claroscuro. Vestía a la moda del siglo XIV, con capa parda, gorguera blanca en el cuello y sombrero negro de ala ancha. Todo el cuadro tiene el estilo austero de Rembrandt y su típico juego de luces y sombras. Parece un cuadro de Rembrandt, pero no es de Rembrandt. Es de la inteligencia artificial, que creó un cuadro a la manera de Rembrandt. Es una especie de manierismo inteligente.

Hacia mediados del siglo XVI surge el estilo artístico denominado manierismo, como referencia a aquellos artistas plásticos que intentaban pintar «a la manera» del gran Miguel Ángel o el singular Rafael. Su método consistía en extraer las técnicas pictóricas que estos maestros utilizaban para aplicarlas de forma mecánica. Se pensaba que la utilización reiterada de grandes técnicas llevaría a grandes obras. Ilustre error. El manierismo está considerado más como un arte decadente o de transición, que como un periodo de significativa creatividad.

La obra *El siguiente Rembrandt* fue pintada por la agencia de publicidad J. Walter Thompson Amsterdam, junto con ING, Microsoft, la Universidad Tecnológica de Delft y el Museo Mauritshuis, y siguiendo el mismo método de construcción que aquellos manieristas de finales del siglo XIV. Se diseñó a base de analizar la técnica de Rembrandt para posteriormente reproducirla. Se decidió pintar un retrato porque Rembrandt tiene una gran cantidad pinturas donde aparecen rostros de personas, lo que suponía una gran cantidad de fuentes de las que extraer su técnica.

Los diseñadores recopilaron más de 346 pinturas, que fueron analizadas por algoritmos creados para identificar patrones en su forma de pintar retratos, tales como distancia de los ojos, tamaño de la nariz, forma de la boca, vestimenta y tipos de fondo. Esto permitía después enseñar a una

máquina a reproducir tales rasgos, teniendo en cuenta las proporciones finales del cuadro.

Una vez decidido el tema, se emplearon más de 500 horas de impresión para representar el retrato, que posteriormente fue completado con pintura en 3D para dar el característico volumen del óleo y las pinceladas propias de Rembrandt. Con todo ello, después de 18 meses de trastear, no tanto con pinceles y pigmentos, sino con algoritmos y fórmulas, apareció *El próximo Rembrandt*. Aplausos.

Aplausos, pero no tantos. Porque, en definitiva, este cuadro simula un Rembrandt. La inteligencia artificial es muy buena identificando patrones y esto es lo que hizo. Identificó la forma de pintar de Rembrandt, igual que hacían los manieristas con las técnicas de Miguel Ángel o Rafael. Tecnológicamente, puede significar un hito, pero artísticamente no es muy meritorio. Toda obra a la manera de otro artista carece de interés artístico.

Escribir una novela que te recuerde a *Don Quijote de la Mancha* carece de mérito. Una vez nacido Cervantes, nadie aprecia escribir como Cervantes. Una vez nacido Beethoven, nadie aplaude una obra que suene a Beethoven. Pintar, escribir o componer como lo haría un cierto artista es fácil. Consiste en analizar las técnicas originales del autor y aplicarlas de manera mecánica, pensando que incluir mucho de algo grande es algo más grande. Por eso, utilizar la inteligencia artificial para pintar, escribir o componer como lo haría un célebre artista es relativamente fácil. Ahora bien, ¿es posible que una máquina cree una obra nueva?

Entonces Edmon de Belany levanta la mano y dice que sí.

Nuevo arte antagónico

Edmon de Belamy es el nombre del primer cuadro totalmente original pintado por la inteligencia artificial[47]. Fue subastado en Christie's en 2018 y adquirido por un comprador anónimo por valor de unos 360.000 €. La obra es, de nuevo, un retrato, esta vez de un joven, que también posa mirando de soslayo a la derecha. Tiene un estilo más moderno, con trazos poco definidos que representan al muchacho difuminado en un fondo parcialmente oscuro.

El retrato es el primero de una serie que representa a la imaginaria familia De Belamy, desde el primer conde de Belamy allá por el siglo XVIII o XIX.

La obra es novedosa y tiene un cierto estilo propio. No parece pintada a la manera de nadie. No parece manierismo digital. ¿Es posible, entonces, que la inteligencia artificial cree obras totalmente nuevas? Vamos a ver cómo fue pintado el joven Edmon de Belamy.

En el margen inferior derecho del cuadro aparece la firma de su autor, como corresponde a toda buena obra de arte. Pero en lugar de ver unos trazos que puedan simular un nombre, que habitualmente no nos dicen nada, vemos una compleja fórmula matemática que comienza por «minmax», —puede que tampoco nos diga mucho—. Es la fórmula de lo que se conoce como Redes Generativas Antagónicas (GAN en inglés, *Generative Adversarial Networks*), que son el fundamento, entre otras cosas, para crear obras de arte.

Las redes generativas antagónicas constan de dos sistemas inteligentes, habitualmente dos redes neuronales, que compiten entre sí. Para el caso de crear obras de arte, una de estas redes neuronales puede ser asimilada a un comisario de una exposición y la otra a un falso pintor. Supongamos que el comisario quiere organizar una exposición de pintura. Su objetivo es que no le cuelen ningún cuadro que no esté pintado por un merecido artista. El falso pintor, por el contrario, quiere engañar al comisario y exponer en dicha exposición. Ambos tienen objetivos antagónicos: el falso pintor quiere engañar y el comisario quiere que no le engañen.

Este juego de engañar y no ser engañado lleva su tiempo. Para empezar, el falso pintor no tiene ni idea de pintar y comienza, por tanto, a pintar cuadros con motivos aleatorios. Cada vez que pinta un cuadro se lo presenta al comisario de la exposición. Este lo evalúa comparándolo mentalmente con obras de arte de distintos estilos y autores para determinar si el cuadro es merecedor de un gran artista. Sistemáticamente, va rechazando los cuadros por falta de calidad, pero según la cara que pone el comisario y algún que otro comentario que suelta al examinar cada obra, el falso pintor va cogiendo experiencia sobre cómo debe ser una supuesta obra maestra.

De esta forma el falso pintor va aprendiendo y llega un momento en el que tiene conocimiento suficiente como para crear un cuadro que pueda pasar por una digna obra de arte. En ese momento el pintor advenedizo ha

conseguido engañar al comisario y su cuadro se colgará en las paredes de la exposición.

Desde un punto de vista técnico, el comisario de la exposición es una red neuronal llamada discriminador. Ha sido entrenada con imágenes de cuadros famosos y su objetivo es determinar la probabilidad de que una pintura sea similar a una obra de arte. El falso pintor es una red neuronal llamada generador, que está conectada al discriminador. El generador (falso pintor) le presenta obras al discriminador (comisario de la exposición) con el objetivo de conseguir que el discriminador acepte una obra suya. Comienza a generar imágenes de forma aleatoria y va perfeccionando sus obras en función de los resultados que obtiene por parte del discriminador. La red neuronal discriminador, que ha sido entrenada con una gran base de datos de obras de arte, está entrenando, a su vez, a la red neuronal generador en el arte de la pintura.

Ya vimos, al hablar del libro-Einstein, que una red neuronal puede aprender, y para ello cambia algunos de sus valores. Luego esta idea de ir perfeccionando las obras que produce el generador es posible. En particular, para pintar a Edmon de Belamy, el sistema fue entrenado con unas 15.000 imágenes de cuadros de distintas épocas[48].

El objetivo del discriminador (comisario de la exposición) es maximizar sus aciertos y minimizar los logros del generador (falso pintor). Mientras que el objetivo del generador (falso pintor) es maximizar sus logros, y minimizar los aciertos del comisario. Es lo que se llama una estrategia minimax. ¡Ahora se empieza a entender la firma del cuadro!

Esta visión de redes generativas antagónicas fue ideada por Iean Goodfellow[49] en 2014. Su apellido viene a significar algo así como buena persona o buen tipo. De ahí el apellido de la familia de Belamy, forma aproximada de «bel ami», que significa buen amigo en francés. Edmon de Belamy fue creado por la organización francesa Obivious dedicada al estudio de la inteligencia artificial y el arte.

La utilización de la técnica de redes generativas antagónicas ha evolucionado a lo que se denominan redes antagónicas creativas[50] (CAN en inglés, *Creative Adversarial Networks*). Mediante esta tecnología se han

creado obras de arte de cualquier estilo: desde cuadros que representan paisajes, a cuadros abstractos con trazos o formas geométricas diversos.

Esa forma de pintar tiene dos características significativas: azar y probabilidad. La primera red neuronal, el generador, o falso pintor comienza pintando de forma aleatoria. La segunda red neuronal, el discriminador, o comisario de arte, decide si la pintura generada es una obra de arte o no. Pero en realidad, no decide. Lo que hace es asignar una probabilidad respecto a si la pintura generada es o no obra de arte. Cuando la primera red neuronal le enseña un cuadro, la segunda ofrece un resultado que dice: «esto es una obra de arte con una probabilidad de X %». Cuando esa probabilidad supera un cierto umbral definido, se considera que el cuadro es una obra de arte. Esta probabilidad es la intensidad que comentábamos en el punto anterior sobre los finales de historia de las redes neuronales.

Azar y probabilidad, una vez más y, cuando la probabilidad supera un cierto umbral, se determina una historia final y decimos que la red neuronal *decide*. Introducimos ahora el componente de tomar decisiones.

Decide lo más deseado

Me pones en un brete

Decimos que, por ejemplo, la inteligencia artificial decide cuándo mandarte una notificación al móvil, o bien decide qué ruta debe tomar para llegar a un destino, o qué hacer en un vehículo autónomo para evitar un accidente. Esto nos puede parecer maravilloso. ¿Cómo sabe la inteligencia artificial lo que debe decidir, si a veces, ni yo mismo lo sé?

Supongamos que nos encontramos en un concurso de televisión en el cual, tras numerosas y sacrificadas pruebas hemos ganado 1 millón de euros. En ese momento, antes de finalizar, viene la presentadora y nos reta: «Te propongo un juego final: te puedes llevar el millón de euros, o bien nos jugamos 3 millones de euros a cara o cruz». Vaya dilema.

Para tomar una decisión valoramos cada situación respecto a lo que me agrada o desagrada de cada una de ellas y qué nivel de certeza existe. Por un lado, tengo 1 millón seguro, con una probabilidad del 100 %. Por otro lado, tengo 3 millones con una probabilidad del 50 %, pues me lo juego a cara o cruz. Si lo que más me gusta es el dinero, o me gusta el riesgo y la aventura, me decantaré por la segunda situación, pues puedo ganar 3 millones, aunque tenga incertidumbre. Por el contrario, si me importa más la vergüenza y la cara que se me queda en el caso de que pierda (¡a ver cómo lo explico en casa!), entonces puede que me decante por la primera situación y no acepte el reto.

Estamos considerando dos factores: probabilidad y lo apetecible o deseable de cada situación. Ambos afectan. La probabilidad afecta, porque si en lugar de proponerme jugar los 3 millones a cara o cruz, me proponen que sea sacando el as de oros de una baraja española, posiblemente decida quedarme con el millón de euros que está garantizado. La probabilidad de sacar un as de oros en una baraja española es de solo el 2,5%. Pero también afecta lo deseable de cada situación. Si en lugar de prometerme 3 millones a cara o cruz, me prometen 10 millones, es posible que aceptara. En ese caso, la

vergüenza de perder es menor, pues lo abultado del premio bien merece la apuesta.

En situaciones en las que interviene cierta incertidumbre tomamos las decisiones barajando la probabilidad y el deseo. Esto mismo ocurre con la inteligencia artificial. Pero la inteligencia artificial no tiene deseo En lugar de deseo se le llama utilidad, que parece un nombre más científico, y es más serio decir que la inteligencia artificial toma una decisión en función de la utilidad, que en función del deseo.

La inteligencia artificial en muchas ocasiones trabaja en lo que se llaman entornos inciertos. Son aquellos en los que la inteligencia artificial no dispone de toda la información necesaria para tomar una decisión, o bien en cualquier momento puede suceder algo que afecte al sistema. La conducción de un vehículo autónomo sucede en un entorno incierto, pues en cualquier momento puede suceder algo en el tráfico. Un sistema de diagnóstico de una enfermedad también sucede en un entorno incierto, pero ahora porque el sistema no dispone de toda la información. Quizás el sistema inteligente está analizando una radiografía, pero no conoce todo el historial del paciente —quizás solo lo conoce el paciente y en ese momento no lo puede decir—. La inteligencia artificial que trabaja para mostrarte notificaciones en tu móvil también opera en un entorno incierto: no dispone de toda la información, pues no sabe qué estás haciendo en ese momento —afortunadamente—, y no sabe si ocurrirá algo nuevo, por ejemplo, recibir una notificación de otra red social de la competencia. En todos estos casos la inteligencia artificial toma una decisión, ya sea frenar o acelerar, diagnosticar tu enfermedad o mandarte una notificación al móvil, mediante lo que se llama la utilidad esperada. Consiste en conjugar probabilidad con deseo (utilidad). Veamos cómo lo hace.

La fórmula del deseo

Cuando tenemos que decidir entre 1 millón seguro o 3 millones a cara o cruz, hacemos un cálculo informal que considera la probabilidad y lo deseable de cada situación. Este cálculo informal hay que formalizarlo de cara a la inteligencia artificial, es decir, hay que convertirlo en fórmula

matemática. De esta manera surge la llamada «utilidad esperada», que es el producto de la probabilidad por la utilidad. Pero ya hemos dicho que la utilidad es en el fondo es lo deseable de una situación. Por tanto, la utilidad esperada lo podríamos llamar «deseo esperado» y es un deseo multiplicado por la probabilidad de que ocurra.

Además, la inteligencia artificial tiene una directriz muy clara al respecto: debe tomar aquella decisión que produzca una utilidad esperada más alta, el mayor deseo posible. Con esta premisa, si la inteligencia artificial estuviera en la situación del concurso de televisión, la decisión habría sido clara.

Lo primero que se le diría a la inteligencia artificial es qué debe entender por utilidad, es decir, qué es lo deseable. En la situación del concurso la respuesta es clara: lo deseable es el dinero, y cuanto más dinero, más deseable. Entonces en la primera situación tiene una utilidad esperada de 1 millón, producto de multiplicar 1 millón de euros por el 100 %. En la segunda situación tiene una utilidad esperada de 1,5 millones, producto de multiplicar de 3 millones de euros por el 50 %. Como 1,5 es mayor que 1, la inteligencia artificial decidirá jugarse los 3 millones. No hay duda.

Quizás el concursante no habría decido eso, como vimos, por el tema de la vergüenza. ¿Cómo meto la vergüenza en esta ecuación? Es más complicado. Habría que intentar cuantificar el nivel de vergüenza y meterlo en el valor de utilidad. Ahora los 3 millones de euros son menos deseables, porque hay una componente de vergüenza. Ya no multiplicaremos 3 millones de euros por 50%, sino algo menos. Si la vergüenza es muy alta, puede que la utilidad esperada de los 3 millones de euros sea menor que la del millón de euros, y opte por esto último.

La inteligencia artificial toma la decisión que le lleva a la máxima utilidad esperada, que es como decir el máximo de los deseos probables[51]. El reto radica en determinar esa idea de deseo. Si estamos en un sistema de conducción autónoma lo deseable pueden ser varias cosas. Dependerá de cada persona. Uno puede desear gastar poco combustible, otro recorrer caminos curiosos, o bien la ruta más corta; quizás alguien desee una conducción tranquila y otro conducir de forma agresiva. Pero también están los deseos de la empresa que diseña la inteligencia artificial. Quizás su deseo sea que pases por determinados sitios porque así ellos facturan por cada

usuario que accede a dichos sitios. Si pensamos en un sistema de notificaciones, quizá tu deseo sea recibir notificaciones interesantes y el deseo de la red social sea mandarte muchas notificaciones para que no te vayas de su red.

A todos estos cálculos se les llama utilidad esperada. Por ello, se dirá que la inteligencia artificial tome una decisión que arroje la máxima utilidad, que es una forma de decir que la inteligencia artificial toma la decisión que produce lo que más se desea. Pero siempre nos quedará por saber cuál es esa fórmula del deseo en un doble sentido: qué es lo que se desea y quién lo desea, si los usuarios o los diseñadores de la inteligencia artificial.

Incluso ti tenemos clara la fórmula del deseo, no siempre lo que decide la inteligencia artificial coindice con nuestras decisiones. La inteligencia artificial puede *desear* una cosa y nosotros otra.

Decidimos por sensaciones

La teoría de la utilidad esperada en un modelo matemático que se utiliza en el ámbito de la economía para determinar qué decisión puede tomar una persona en situaciones inciertas. Este modelo se ha traspasado a la inteligencia artificial, de tal forma que esta decide aquello con lo que obtenga una mayor utilidad esperada. Pero nosotros no somos tan racionales tomando decisiones. Esto lo demostró el economista y físico Maurice Allais al plantear la siguiente paradoja[52].

A un grupo de personas se les propuso lo siguiente, respecto a ganar un cierto dinero en una lotería. En cada opción tenemos un premio con una cierta probabilidad:

Dilema 1

Opción A: ganar 4000 € con un 80 % de probabilidad.

Opción B: ganar 3000 € con un 100 % de probabilidad.

Dilema 2

Opción C: ganar 4000 € con un 20 % de probabilidad.

Opción D: ganar 3000 € con un 25 % de probabilidad.

El primer dilema es similar al planteado en el ejemplo del concurso. Se observó que en este dilema la mayoría de las personas elegían la opción B (los 3000 euros seguros), mientras que en el dilema 2, la gente escogía la opción C. La decisión habitual sobre el dilema 1 iba en contra de la teoría de la utilidad esperada. Si en este dilema hacemos los cálculos correspondientes, matemáticamente hablando, obtenemos más utilidad esperada con la opción A, que con la opción B. Sin embargo, en el dilema 2, la decisión habitual sí coincide con lo que dice la teoría. Es decir, que en ocasiones seguimos lo que dice la teoría y en ocasiones no.

Una teoría que a veces funciona y a veces no, no es una gran teoría. Esto lo saben los economistas y los diseñadores de inteligencia artificial, por ello se realizan ajustes sobre la fórmula de la utilidad esperada. Con todo, podemos extraer una conclusión: la teoría de la utilidad esperada no dice qué debemos decidir, ni describe cómo tomamos las decisiones. Es decir, no es ni normativa ni descriptiva.

No es normativa en el sentido de que no establece cómo se deben tomar las decisiones. No hay una obligación de seguir la teoría de la utilidad esperada. Cuando nosotros decidimos, podemos seguir esa teoría o no. Como no es normativa, tampoco es descriptiva. La teoría de la utilidad no describe cómo tomamos las decisiones. Para verlo con más claridad, pensemos en otra teoría: la teoría de la gravedad. Esta sí es normativa y descriptiva, porque describe cómo cae una piedra al suelo, y la piedra siempre debe caer según dicha teoría —por ello hasta se le llama ley—.

Tomar decisiones parece algo complejo. Quizás lo sea para nosotros, porque tomamos decisiones considerando muchos factores: desde el dinero que gano hasta la vergüenza si pierdo. Para la inteligencia artificial la situación es más sencilla. Tan solo tiene que buscar la decisión que le dé el máximo deseo esperado, que en lenguaje serio se dice máxima utilidad esperada. Tanto nosotros como la inteligencia artificial tomamos decisiones en situaciones inciertas conjugando la probabilidad con el deseo, si bien de forma distinta. Nosotros lo hacemos de una manera informal, y la inteligencia artificial de forma matemática y exacta. Aquí se produce un doble alejamiento. Por un lado, nuestros deseos no siempre se ajustan a una fórmula matemática. Por otro lado, detrás de la utilidad esperada en la

inteligencia artificial se encuentra una medida del deseo, pero como usuarios nunca sabemos cuál es ese deseo, ni quién desea, y por ello podemos no llegar a entender por qué la inteligencia artificial toma una decisión sobre una situación que nos afecta.

Somos predecibles

¿Cómo funciona la inteligencia artificial?

Hemos visto cómo funcionan algunas de las capacidades de la inteligencia artificial: cómo es capaz de hablar o de componer textos; cómo las redes neuronales intentan simular nuestro cerebro y cómo deciden un resultado; cómo puede crear arte; y cómo toma decisiones en ciertas circunstancias inciertas. En todos los casos existe un denominador común: la probabilidad.

La inteligencia artificial se fundamenta en el hecho cierto de que las cosas suceden en la vida de manera más o menos uniforme o probable. Dados unos síntomas, lo más probable es una enfermedad. Esta certidumbre sobre los acontecimientos también sucede en nosotros. Somos bastante previsibles. Cuando hablamos o escribimos, lo más probable después de un sujeto es un verbo; después del verbo «ir», lo más probable es que aparezca la palabra «a» y un lugar (por ejemplo, «a casa»).

Nuestra escritura tiene un determinado estilo que hace que usemos unas palabras con más frecuencia que otras. También los artistas suelen pintar con una paleta más probable de colores, y pintan unos motivos con más probabilidad que otros. Por ello reconocemos un cuadro de Velázquez o de Goya con solo verlo. Finalmente, solemos tomar siempre las mismas decisiones. Somos muy probables tomando decisiones. El mundo y nosotros mismos somos bastante probables. Eso es lo que aprovecha la inteligencia artificial.

Una vez que la inteligencia artificial identifica patrones de comportamiento, dado que somos tan estables y probables, intervienen los diseñadores de la inteligencia artificial y le ponen un objetivo. Este objetivo es el deseo. Deseamos crear poemas de amor, recrear la charla con un sicoanalista, pintar un cuadro, identificar números o procurar que los usuarios no abandonen una red social. En esto de cumplir deseos (objetivos) la inteligencia artificial es realmente buena y seguro que lo consigue. Bastará con darle una fórmula y pedirle que calcule el valor más grande o más pequeño, que es como pedirle que nos dé el máximo de deseo, o el menor de los disgustos. Además, lo hace de forma rápida.

Ahora tenemos una explosión de la inteligencia artificial porque ahora es cuando estamos consiguiendo que sea rápida de verdad, tanto o más que nosotros. Por eso también nos asusta. Pero los fundamentos son antiguos. Lo hemos visto con el versificador artificial de John Peter de 1677, que es la base para que un sistema inteligente escriba textos.

Trabajar con probabilidades nos puede llevar a confundir lo más probable con lo cierto. Un sistema inteligente puede determinar que, si vas por una cierta ruta, te diriges entonces a un determinado destino. En realidad, lo que hará es determinar una cierta probabilidad de ir a un destino, en función de comportamientos anteriores. Puede determinar, por ejemplo, que vas al trabajo con un 80 % de probabilidad. Esto no significa que un 80 % de tu persona vaya al trabajo. Lo cierto es que irás o no irás al trabajo. Es probable que vayas en un 80 %, pero la verdad será otra y esta será sí o no: irás o no irás; nunca irás a medias.

La probabilidad es una herramienta que nos permite trabajar con una realidad incierta, donde existen distintos grados de posibilidad. Sin embargo, la probabilidad de una realidad no es la propia realidad, en la misma medida en que la creencia no es lo mismo que la verdad. Uno puede creer que hay vida inteligente en otros planetas, y lo puede creer con una cierta probabilidad, pero luego la verdad será que haya vida o no. De igual forma, una inteligencia artificial puede creer que vas a hacer algo, estafar a un banco, o no estudiar en un curso. Asignará probabilidades en función de las veces que hayas estafado en el pasado o de cuánto hayas estudiado. Pero la verdad es que finalmente tú estafarás o no, estudiarás o no.

Helen Frankenthaler vivió la rabia y la ira contra su poética obra *Montañas y mar* porque nadie entendía su novedosa técnica de usar lienzos sin preparar y diluir la pintura en aguarrás. Lo tuvo que explicar y entonces pasó de la incomprensión al reconocimiento. Lo mismo sucedió con el telégrafo o con el teléfono. Parecían cosas misteriosas, pero ahora sabemos cómo funcionan y son tecnologías que aceptamos sin recelo. La inteligencia artificial se basa en que casi siempre hacemos lo mismo. Sorpréndela y sorpréndete haciendo algo distinto.

Mis conclusiones. ¿Las tuyas?

En este capítulo hemos abordado la siguiente cuestión: ¿Cómo funciona la inteligencia artificial? Estas son mis propuestas de respuesta:

- Para hacer que la inteligencia artificial hable o escriba textos se utiliza la combinatoria y la probabilidad. En el lenguaje combinamos ciertas palabras, sujetos, verbos, complementos, con una cierta probabilidad.
- Las redes neuronales ofrecen un resultado más probable entre los finales de historia que le proponemos.
- La inteligencia artificial es capaz de pintar porque identifica patrones o porque combina azar y probabilidad. Una red neuronal comienza pintando de forma aleatoria hasta llegar a una pintura que otra red neuronal determina que es una pintura con una cierta probabilidad.
- En entornos inciertos la inteligencia artificial toma una decisión mediante la teoría de la utilidad esperada, que es el cálculo del máximo deseo probable.
- Pero nosotros no somos tan matemáticos tomando decisiones, no siempre seguimos la teoría de la utilidad esperada.
- La inteligencia artificial se fundamenta en la probabilidad porque somos bastantes predecibles. Pero no siempre lo más probable es lo que luego ocurre. Depende muchas veces de nosotros.

Estas son mis conclusiones. Pero no las aceptes de forma predecible. Piensa y extrae tus propias conclusiones, y sin son distintas a las mías quiere decir que es probable otra posibilidad.

La traición de las imágenes (Esto no es una pipa), **René Magritte, 1928–1929. Museo de Arte del Condado de Los Ángeles.**

Esto no es una pipa

El pintor belga René Magritte sorprendió al mundo cuando pintó su famosa obra *La traición de las imágenes (Esto no es una pipa)*. El lienzo representa una pipa, perfectamente dibujada, pero debajo de la misma escribe el texto «esto no es una pipa». Desconcertante, pero cierto.

Magritte tiene razón. En verdad, lo que vemos no es una pipa. No la podemos llenar de tabaco y no la podemos fumar. Tal y como él mismo dijo «si hubiese puesto debajo de mi cuadro "Esto es una pipa", habría dicho una mentira»[53]. Lo que vemos es la representación de una pipa. Con esta obra Magritte quería destacar que lo que vemos, no es lo que es.

¿Qué ocurre con la inteligencia artificial? Si enseñáramos el cuadro de *La traición de las imágenes (Esto no es una pipa)* a un sistema inteligente de reconocimiento de imágenes, este lo clasificaría como una pipa. Podría también leer el texto que aparece debajo de la pipa y que dice «Esto no es una pipa» y almacenarlo junto con la imagen. ¿Encontraría alguna contradicción? ¿Pensaría, «esto es una pipa, pero no lo es»? Si la obra de Magritte se titula la traición de las imágenes, ¿es la inteligencia artificial la traición de la inteligencia?

René Magritte es uno de los ejemplos más representativos de lo que se denomina surrealismo. Salvador Dalí es otro ejemplo de ello, o Luis Buñuel en el caso del cine. En 1924 André Bretón escribió el primer *Manifiesto Surrealista*, donde expuso que el surrealismo buscaba expresar el funcionamiento real del pensamiento, sin la intervención de la razón[54]. Sin embargo, el término surrealismo nació unos años antes, en 1917, de la mano del poeta y dramaturgo Guillaume Apollinaire, cuando publicó la obra de teatro *Las tetas de Tiresias*, que subtituló como drama surrealista. En el prólogo de la obra, el propio autor dice que «cuando el hombre quiso imitar el caminar, creó la rueda que no parece una pierna. Así hizo el surrealismo sin saberlo»[55]. Esta relación de la rueda con el caminar es similar a la inteligencia artificial con nuestra inteligencia. Pero esto lo veremos más tarde.

El surrealismo, y la obra de Magritte en particular, trabaja sobre el pensamiento, y eso es lo que vamos a hacer ahora. Este capítulo es algo filosófico. Su propósito es pensar sobre el pensamiento —como ves, ya empezamos con frases circulares filosóficas—. A la inteligencia artificial se le llama inteligencia y aquí te propongo entender por qué a esta tecnología se le llama inteligencia y si es igual a la nuestra.

Este capítulo tiene tres partes. En las dos primeras te invito a pensar sobre qué es pensar y cuáles son sus componentes. En la tercera parte veremos qué se entiende, en realidad, por inteligencia artificial y en qué se parece, o se diferencia, a nuestra inteligencia. Si crees que la filosofía no sirve para nada, porque no llega a conclusiones prácticas, quizás solo te interese la tercera parte. Pero si opinas, como yo, que la filosofía te ayuda a desarrollar un pensamiento crítico y a ser libre, entonces te animo a leer todo el capítulo. Recuerda, además, que uno de los objetivos de este libro es pensar. Este capítulo te invita a pensar. Pensemos, entonces si la inteligencia artificial es inteligencia, si parece una pipa, pero no es una pipa, o qué pasa con ella.

¿Hay alguien ahí?

Inteligencia significa pensar

El término de inteligencia artificial fue acuñado por John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon en 1955. Nació con la propuesta de celebrar una conferencia en la que analizar qué es la inteligencia y ver cómo una máquina podría simularla. La conferencia se llamó Dartmouth Summer Research Project on Artificial Intelligence, más conocida como la Conferencia de Dartmouth, y se celebró el año siguiente en la universidad Dartmouth College.

El documento de propuesta de conferencia exponía los puntos que, se entendía, debía cubrir un sistema para que este se considerara inteligente. Así se hablaba de cómo programar una máquina para resolver problemas, usar el lenguaje, trabajar con conceptos y abstracciones, aprender e incorporar elementos aleatorios para generar creatividad[56]. Estas fueron las habilidades iniciales a cubrir para simular nuestro comportamiento. Otra cuestión es que se hayan cubierto o no. Lo significativo es que la inteligencia artificial nace con el propósito de diseñar máquinas que hagan lo que nosotros hacemos como seres inteligentes. Ese fue su objetivo en origen y lo sigue siendo en la actualidad. Ahora bien, ¿qué hacemos nosotros como seres inteligentes?

Aquí empieza un primer problema. Somos seres inteligentes y nos resulta complicado definir qué es la inteligencia[57]. La dificultad se encuentra en que la inteligencia no es una cosa única, sino que comprende distintas habilidades. Cuando hablamos de inteligencia nos referimos a una capacidad que nos permite razonar, planificar, resolver problemas, pensar en conceptos abstractos, comprender ideas complejas, aprender rápidamente y aprender de la experiencia[58]. Además, y esto es importante, todas estas capacidades están encaminadas a una capacidad más amplia y profunda de comprender nuestro entorno, lo que significa captar o dar sentido a lo que nos rodea.

¡La de cosas que hacemos como seres inteligentes! Algunas de ellas aparecen en la propuesta de la Conferencia de Dartmouth, tales como resolver problemas o aprender. También en la Conferencia aparece trabajar

con conceptos y abstracciones. Sin embargo, en la definición de inteligencia, en lugar de trabajar con conceptos abstractos, se dice pensar en conceptos abstractos. Es un matiz relevante. ¿Qué es *pensar*?

Ahora es cuando nos ponemos filosóficos. Si queremos tener una definición de qué es pensar, podemos acudir a Descartes, famoso por su conocida frase: «Pienso, luego existo». Según esta declaración, pensar es algo muy importante, ya que está relacionado con existir. Existimos porque pensamos, y mientras tengamos pensamiento, seguiremos existiendo. Descartes decía que somos cosas que piensan, es decir cosas que dudan, conciben, afirman, niegan, quieren, no quieren, imaginan y sienten[59]. Habla de cosas, si bien todos estos verbos exigen la presencia de una consciencia, es decir, de alguien que duda, concibe, afirma, niega, etc. ¡Uy!, hemos metido la palabra consciencia.

Esta idea de una consciencia lo vemos de forma más clara en otra definición de pensar, esta vez, de la mano de Kant —no podía faltar Kant—: pensar es unificar representaciones en una consciencia, o, lo que es lo mismo, pensar es emitir juicios[60]. La definición es algo más compleja que en Descartes —como no podía ser de otra forma, viniendo de Kant—, pero lo importante es entender que detrás de la idea de pensar asoma a su vez la idea de una consciencia, en el sentido de tener la capacidad de percibirse a sí mismo y en el mundo. Esto está relacionado con el final de la definición que hemos visto de inteligencia. Tenemos una serie de capacidades (razonar, planificar, resolver, aprender, etc.), pero encaminadas a comprender nuestro entorno, a dar sentido a lo que nos rodea. La conciencia entra dentro de ese dar sentido.

Pensar viene del latín *pensare*, que significa pesar. El pensamiento es lo que pesa o pondera los datos, los argumentos, las experiencias y hasta el propio pensamiento —el pensamiento puede pensar sobre lo que está pensando—. Como dice Platón en su dialogo *Teeteto*, pensar es un discurso que el alma se dirige a sí misma y el resultado de tal diálogo es un juicio sobre lo que se piensa. Aparece de nuevo la idea de pensar como una consciencia que emite un juicio.

En resumen, la inteligencia supone una serie de habilidades, tales como razonar, planificar, resolver problemas o aprender y, de manera especial,

pensar. El objetivo de estas habilidades no es el conocimiento enciclopédico o ganar *Pasapalabra*, sino comprender, dar sentido a lo que nos rodea. Esta comprensión se consigue precisamente con la acción de pensar, cuyo resultado es emitir un juicio y para ello se necesita la presencia de una consciencia, es decir, de alguien que piensa, que emite un juicio. Todas estas habilidades las hacemos nosotros los humanos con nuestra inteligencia. ¿Hasta dónde llega la inteligencia artificial? ¿Piensa la inteligencia artificial?

Alexa te responde en chino

En el capítulo anterior vimos que Alan Turing resolvió la cuestión de si la inteligencia artificial piensa de manera simple y apelando a la práctica. Su propuesta fue, lejos de debates filosóficos, decir que una máquina piensa si hace lo que nosotros hacemos. Punto. Esto lo podemos saber si es capaz de engañarnos y hacerse pasar por uno de nosotros. En definitiva, según Turing, una máquina piensa si pasa su llamado juego de imitación, su famoso test de Turing.

John Searle, profesor de filosofía, no está muy de acuerdo con la visión de Turing. No termina de ver el tema de la consciencia en la propuesta del test. Por ello, en 1980, treinta años después del artículo de Turing y cuando la investigación en inteligencia artificial ya estaba dando resultados, publicó su famoso experimento mental llamado «La habitación china de Searle»[61].

Supongamos una habitación completamente aislada del exterior, salvo por una ranura a modo de boca de buzón, en cuyo interior se encuentra una máquina, que, según nos dicen, entiende chino. Para comprobarlo, por la ranura del buzón podemos introducir un texto en chino, junto con una pregunta relativa a su contenido. El texto puede ser de la forma: «Juan fue al restaurante y pidió una hamburguesa. Cuando el camarero se la trajo estaba fría. Al marcharse Juan, dejó muy poca propina». La pregunta relacionada con este texto puede ser: «¿Le gustó a Juan la hamburguesa?». Tras unos minutos de operación, por la misma ranura aparece un nuevo texto en chino que responde a la pregunta realizada. Por ejemplo, para el caso de Juan y la hamburguesa, la respuesta podría ser: «Definitivamente, no le gustó la hamburguesa y se fue disgustado del restaurante». Todo esto en chino, de tal

forma que una persona versada en dicho idioma concluye, sin lugar a duda, que la máquina que está dentro de la habitación entiende chino. La máquina pasa el test de Turing en chino.

Una máquina de esta naturaleza en realidad ya fue diseñada hacia 1977[62], si bien en inglés, no en chino, siendo capaz de resolver este tipo de preguntas sencillas. Hoy en día, los actuales asistentes de voz, tipo Alexa (Amazon), Google Assistant, Siri (Apple) o Cortana (Microsoft), son una aproximación a «La habitación china de Searle». Uno puede pensar que, por ejemplo, Alexa te entiende en español o en cualquier otro idioma.

Volvamos a la habitación china. En realidad, dentro de la habitación no se encuentra ninguna máquina maravillosa. Se halla el propio Searle, el cual no tiene ni idea de chino, pero es capaz de responder a las preguntas que le hacen en dicho idioma. Para ello, Searle dispone de un completo manual que le permite operar con símbolos chinos. Por ejemplo, el manual le dice: «si encuentras el conjunto de símbolos "poca propina", entonces debes responder con estos símbolos: "no le gustó"». Entendamos que «poca propina» y «no le gustó» se encuentran escritos en chino. Como Searle no habla chino, estos textos no son más que símbolos. Searle no ve palabras con un significado para él, solo ve símbolos en chino y el manual de la habitación china le permite relacionar esos símbolos de manera adecuada. Le relaciona, por ejemplo, los símbolos «poca propina» con los símbolos «no le gustó».

La persona que se encuentra fuera de la habitación, y que hace preguntas en chino relativas a un texto en chino, está totalmente convencida de que dentro de la habitación hay alguien que entiende chino. Sin embargo, dentro de la habitación solo existe un manual y Searle, que no entiende chino. Searle plantea entonces una serie de cuestiones sobre quién entiende chino: ¿podemos decir que los manuales entienden chino?, ¿podemos decir que Searle entiende chino porque es capaz de responder preguntas en chino?, o bien, ¿es la habitación, con el manual y Searle, quien entiende chino?

Si estas preguntas las llevamos a día de hoy, nos podemos preguntar si Alexa, Google Assistant, Siri o Cortana entienden español o cualquier otro idioma en el cual se le hagan preguntas.

En el fondo de todas estas preguntas se encuentra la idea de pensar, en el sentido de emitir un juicio, que hemos visto antes y que está relacionada con

la presencia de una consciencia. Aparentemente, la habitación china parece que piensa, porque emite un juicio sobre si a Juan le gustó la hamburguesa. Parece que hay alguien que piensa en chino. Sin embargo, dentro de la habitación solo se encuentra un manual, que no piensa, y Searle, que tampoco piensa en chino. Para Searle, en este experimento, nadie piensa en chino, por tanto, pasar el test de Turing no es prueba suficiente de inteligencia. Le falta esa capacidad de pensar, de emitir un juicio con una consciencia.

Pero Searle también tiene sus detractores.

Una cortés costumbre

Hofstadter no está de acuerdo con la visión de Searle. Ya sabes que me refiero al Douglas Hofstadter del que hablamos en el capítulo anterior y que publicó el libro *The Mind's I: Fantasies and Reflections on Self and Soul*, donde la tortuga le proponía a Aquiles el libro-Einstein.

Douglas Hofstadter es uno de los defensores de lo que se conoce como inteligencia artificial (IA) fuerte. Según esta teoría, una inteligencia artificial suficientemente compleja puede tener cualidades mentales similares a las humanas en todos los aspectos. Cuando la inteligencia artificial esté suficientemente desarrollada será capaz de pensar, en el sentido que hemos visto de emitir juicios, tener emociones y tener consciencia. Es una cuestión de la complejidad del algoritmo, es decir, es una cuestión de tiempo que la inteligencia artificial sea exactamente igual a nosotros. Si eso ocurre, en un momento dado tendremos replicantes que nos dirán con nostalgia que han visto arder naves de ataque más allá de Orión y que todos esos momentos se perderán en el tiempo, como lágrimas en la lluvia. Serán robots con miedo a morir.

El algoritmo en un sistema inteligente lo podemos asimilar a ese manual que se encuentra en la habitación china, que le permite trabajar con palabras en chino como si fueran símbolos. Si trasladamos la visión de Hofstadter a la habitación cde Searle, significa que, si ese manual está suficientemente elaborado, de tal forma que se pueda aplicar sobre cualquier texto y cualquier pregunta, entonces podremos decir que la habitación china

pensará, tendrá emociones y consciencia como nosotros. En ese momento, de verdad entendería chino.

En particular Hofstadter está pensando en un libro similar al libro-Einstein que vimos en el capítulo anterior. Era un libro que contenía, neurona a neurona, toda la información del cerebro de Einstein. Para Hofstadter este libro-Einstein, al ser tan complejo y contener todo el cerebro de Einstein, ya no es un libro, sino que es el propio Einstein. En este libro quedan representados sus pensamientos, sus emociones, e incluso su propia consciencia mediante los valores numéricos de cada hoja. De hecho, no tendríamos porqué llamarlo libro-Einstein, tendríamos que llamarlo directamente Einstein. El libro es Einstein y podemos referirnos a él como si fuera una persona. Tanto es así, que en realidad *leer* el libro-Einstein es hablar con Einstein.

En un momento dado de la conversación entre Aquiles y la tortuga se produce el siguiente lance respecto a cómo se debe tratar al libro-Einstein:

—Aquiles: Ahora, espera un minuto. Estás empleando el pronombre «él» sobre un proceso combinado con un enorme libro. Eso no es "él", es otra cosa […].

—Tortuga: Bueno, te dirigirías a él como Einstein […], ¿no? ¿O dirías: «Hola, libro de los mecanismos cerebrales de Einstein, me llamo Aquiles»? Creo que pillarías a Einstein desprevenido si hicieras eso. Ciertamente, se quedaría perplejo.

—Aquiles: No hay ningún «él». Me gustaría que dejaras de usar ese pronombre.

Aquiles y la tortuga ponen de manifiesto el debate, no resuelto todavía, sobre si ese libro tan complejo tiene consciencia y es el propio Einstein o no. La tortuga dice que sí, y por lo tanto no hay razón para llamarlo libro-Einstein. Si lo llamamos libro-Einstein, entonces al saludar a un amigo llamado Alfredo, tendríamos que dirigirnos a él algo así como: «Hola, conjunto de procesos cerebrales Alfredo». Por el contrario, Aquiles dice que el libro-Einstein es una cosa y no podemos tratarle como si fuera una persona.

Este libro-Einstein, como vimos, es un ejemplo de red neuronal y hoy en día no tenemos redes neuronales tan complejas. La primera red neuronal,

llamada SNARC, fue creada por los estudiantes de Harvard Marvin Minsky y Dean Edmonds en 1950, y estaba formada por 40 neuronas. Actualmente, las redes neuronales pueden contener entre miles y pocos millones de neuronas. Estamos lejos, muy lejos de los 100.000 millones de neuronas de nuestro cerebro, y, por tanto, si los defensores de la IA fuerte tienen razón, estamos lejos de una posible IA fuerte capaz de tener consciencia.

Hasta entonces, hasta que llegue ese día de una inteligencia artificial suficientemente compleja, nos podemos pasar el tiempo debatiendo sobre la consciencia en los sistemas inteligentes. Unos dirán que sí, otros dirán que no, y otros pueden decir lo que argumentó Alan Turing.

En el famoso artículo «Computing Machinery and Intelligence»[63] el propio Alan Turing abordó una serie de objeciones contra su propuesta de test de inteligencia. Una de tales objeciones es relativa a la existencia de una consciencia. Su punto de vista es realmente curioso.

Alan Turing plantea que en realidad nosotros tampoco estamos seguros de que todo el mundo tenga consciencia. Si hablo con mi amigo Alfredo, damos por hecho que él es una persona como yo. Dado que yo tengo consciencia, supongo que Alfredo también disfruta de ella. Pero en verdad, a ciencia cierta no lo puedo asegurar. No puedo entrar en la cabeza de Alfredo y saber si la tiene. Como bien dice Turing: «En lugar de discutir continuamente sobre este punto, es habitual tener la cortés convención de que todo el mundo piensa».

Así se resuelve la cuestión: suponemos que todo el mundo piensa, que todo el mundo tiene consciencia, para no parecer descortés. Ciertamente nunca podremos estar seguros de si la persona con la que hablamos tiene una percepción de sí misma, de manera similar a como yo me percibo a mí mismo en el mundo. Dado que no puedo asegurar que toda persona tenga consciencia, ¿por qué negársela a una máquina? Si un replicante me habla de sus experiencias en Orión, y me dice que siente que todo eso va a desaparecer, ¿por qué no pensar que tiene consciencia como yo? Si no lo discuto con una persona, no tendría por qué discutirlo con una máquina.

Para intentar salir de este embrollo, podemos ver otro punto de vista sobre esta cuestión. Podemos pensar con arte.

Una cuestión de arte

Pintar a lo burro

No podremos aceptar que la máquina sea igual al cerebro hasta que esta sea capaz de escribir un soneto o componer un concierto en respuesta a los pensamientos y emociones que siente, y no por una cascada aleatoria de símbolos; es decir, no basta con que lo escriba, sino saber que lo ha escrito. Esta era la opinión de Geoffrey Jefferson[64], famoso neurólogo y pionero de la neurocirugía, hacia 1949. Introducimos entonces el arte en el debate. ¿Puede la inteligencia artificial crear arte?

En el capítulo anterior vimos que la inteligencia artificial pintó *El siguiente Rembrandt* y *Edmon de Belamy*. Sin embargo, no podemos decir que cumpliera las premisas de Jefferson. *Edmon de Belamy* fue pintado con azar y probabilidad y en ninguno de los dos cuadros podemos asegurar que la inteligencia artificial supiera que estaba pintando algo. Para profundizar sobre esta idea de saber que estás creando algo podemos recurrir al maestro Boronali.

Joachim-Raphaël Boronali posiblemente no sea un pintor muy conocido. Tan solo se le conoce una pintura, titulada *Y el sol se durmió en el Adriático*, fechada en 1910. El cuadro consiste en una serie de trazos de colores sin forma aparente. Una zona horizontal de tonos azules en la mitad inferior del lienzo bien podría simular un mar, y otra franja superior de colores anaranjados y amarillos pudiera ser producto de ese sol que parece dormirse en el mar. Unas pinceladas gruesas naranjas sin significado llaman la atención en un primer plano a la derecha. Con algo de imaginación uno puede ver una especie de embarcación en un mar al atardecer. La obra fue presentada en el Salón de los Independientes de París y fue vendida por 400 francos de la época (unos 3.500 euros actuales).

Para acercamos a la figura de Boronali nos debemos situar en la bohemia calle de Saules, del barrio parisino de Montmartre hacia ese año de 1910, donde se levanta el recogido y ameno café de Lapin Agile. En aquella época el café Lapin Agile era frecuentado por artistas. Allí uno se podía encontrar al

mismísimo Picasso. Entre los artistas existía el debate de si aquellas expresiones en las que apenas se reconoce lo que se pinta se podrían considerar arte.

Había un escritor y periodista, llamado Dorgelès, que era contrario a las nuevas expresiones artísticas y estaba dispuesto a demostrar que el arte es algo más que pinceladas sin sentido en un lienzo. Con este propósito, en cierta ocasión tomó prestado a Lolo, el burro del dueño del café Lapin Agile, y en presencia de un funcionario judicial, para dar fe del hecho, dispuso un lienzo por detrás del jumento a una distancia adecuada, le ató una brocha a la cola del animal, y comenzó a darle zanahorias. La animación de Lolo ante el manjar que le ofrecían le hacía mover la cola con alegría, dejando así trazos de pintura en el lienzo. Con esta inspiración tan culinaria se creó la curiosa y desconocida obra *Y el sol se durmió en el Adriático*.

Para hacer efectivo su engaño, Roland Dorgelès se inventó el nombre del inexistente Joachim-Raphaël Boronali. Tituló el cuadro como hemos dicho y lo presentó al público. No contento con ello, ideó, además, el llamado movimiento artístico el Excesivismo, con manifiesto incluido, como se merece todo movimiento artístico, encabezado por el ilustre e inexistente maestro Boronali.

Al poco tiempo, Dorgelès reveló su engaño. Señaló que su propósito era demostrar que una obra pintada por un burro podía entrar en una exposición de pintura de la época. Viendo el cuadro, uno puede dudar de que haya sido realizado completamente por un burro. Existe una foto de la época en la que se ve a un conjunto de enmascarados brindando detrás del burro Lolo y celebrando la proeza artística. El enmascaramiento solo responde a un espectáculo bohemio transgresor. En la foto se ve al jumento afanado en su tarea, lanzando algunas pinceladas sobre el lienzo, mientras degusta una zanahoria. Todo parece indicar que el pobre animal participó en parte de la obra, pero no es seguro que en toda ella. Quizás esos trazos sin forma definida que se sitúan en la parte frontal derecha puede que los hubiera realizado el inconsciente pollino.

¿Qué pinta, nunca mejor dicho, el burro Lolo en todo esto? ¿Tienen algo en común el retrato *Edmon de Belamy* y el lienzo *Y el sol se durmió en el Adriático*? Sí, lo tienen, y también con *El próximo Rembrandt*. En los tres casos

tenemos a alguien que pinta un cuadro con algo; en ningún caso tenemos algo que sabe que pinta un cuadro.

Vimos que detrás de *Edmon de Belamy* se encuentra la organización Obivious y tras *El próximo Rembrand,* una serie de organizaciones. Detrás de *Y el sol se durmió en el Adriático* se encuentra un grupo de artistas provocadores encabezados por Roland Dorgelès.

En los tres casos estas personas, o grupos de personas, han decidido pintar un cuadro de forma novedosa: en dos de ellos con una inteligencia artificial, en otro caso con un pollino contento. Ni el burro ni la inteligencia artificial decidieron pintar. Habitualmente un pintor pinta con pinceles y ahora han surgido nuevas opciones. En *Edmon de Belamy* y en *El próximo Rembrandt* se han cambiado los pinceles por algoritmos matemáticos y buenas impresoras. En la obra *Y el sol se durmió en el Adriático* existen pinceles dirigidos por una mano que alimenta un burro. Si decimos que la inteligencia artificial ha pintado *Edmon de Belamy* y *El próximo Rembrandt*, tendremos que admitir que Lolo es el autor del cuadro *Y el sol se durmió en el Adriático*.

En las tres obras hay un hecho común y significativo: ni el burro Lolo ni la inteligencia artificial pusieron nombre a sus obras. Porque poner nombre a una obra de arte significa saber que la has pintado y saber que quieres decir algo con ella. Por esta razón en ninguno de estos tres casos sus supuestos autores, la inteligencia artificial y Lolo, pusieron nombre a su obra: porque no sabían lo qué hacían. Según Geoffrey Jefferson, ni la inteligencia artificial ni el burro Lolo serían comparables a nuestro cerebro, porque no saben lo que han pintado.

¿Y qué dice de nuevo Turing a todo esto?

Alguien que cuenta algo

Turing no escapa a la cuestión y la aborda desde un punto de vista sugerente, de nuevo en su famoso artículo de «Computing Machinery and Intelligence». No plantea tanto que una máquina pueda crear un soneto, cosa que da por hecho, sino que además pueda hablar sobre el propio soneto que ha escrito, lo cual supone un nivel más avanzado de criterio. Esto sería similar a lo que hemos dicho de poner nombre a un cuadro. Si tú hablas de

un soneto que has escrito, quiere decir que sabes que lo has escrito y sabes lo que quieres decir con él. Turing sugiere que, quizás, en un momento dado, una máquina podría escribir un soneto y, posteriormente, tener la siguiente conversación con una persona:

PERSONA (P): En la primera línea de tu soneto que dice «te compararé con un día de verano», ¿no sería mejor decir «un día de primavera»?

MÁQUINA (M): No tendría métrica.

P: Y qué tal un «día de invierno». Esto sí qué rimaría métricamente.

M: Sí, pero nadie quiere que le comparen con un día de invierno.

P: ¿Podrías decir que Mr. Pickwick te recuerda a las Navidades?

M: En cierto modo, sí.

P: Sin embargo, las Navidades representan un día de invierno, y no creo que a Mr. Pickwick le molestara la comparación.

M: No creo que hables en serio. Por un día de invierno se entiende un típico día invernal, y no un día especial de Navidad.

Como se ve, la máquina estaría hablando sobre el contenido del propio soneto. Con este intenso diálogo Turing intenta rechazar el comentario de Jefferson cuando habla de una cascada aleatoria de símbolos. Efectivamente, esta conversación no tiene los acusados elementos aleatorios que hemos visto, por ejemplo, en el poema «Texto estocástico», en el escritor M.U.C. de cartas de amor, o en el rudimentario versificador artificial. No obstante, sí comparte con ellos los mismos elementos combinatorios y la propia visión de «La habitación china de Searle». Al final estamos hablando de combinar una serie de símbolos según unas reglas gramaticales.

Una conversación de este estilo entre una persona y una máquina es posible. Si queremos, podemos diseñar un sistema inteligente que mantenga un diálogo similar, de igual forma que se creó una máquina para ver si la hamburguesa le gustó a Juan. El algoritmo M.U.C. puede escribir cartas de amor o podemos hacer que una máquina escriba el poema «Texto estocástico». También tenemos una inteligencia artificial para pintar cuadros o componer música. Sin embargo, todo esto no resuelve la cuestión principal que señala Jefferson, en el sentido de que no basta con escribir un texto, sino saber que lo has escrito.

A través del arte tenemos la facultad de decir algo de manera original. Detrás de una obra de arte siempre hay alguien que quiere contarnos algo. Y esto lo podrá hacer mediante pinceles, un burro, una máquina de escribir o con inteligencia artificial; pero ni el pincel pinta, ni la máquina escribe, y la inteligencia artificial tampoco. Detrás de todas las obras de arte que hemos visto creadas por la inteligencia artificial —las cartas de amor, el «Texto estocástico», *El Próximo Rembrandt* y *Edmon de Belamy*— se encuentra alguien, una persona, que nos quiere decir algo de una forma original.

Obviamente los defensores de la inteligencia artificial fuerte, Turing y Hofstadter negarán este extremo. Turing, con su diálogo sobre Mr. Pickwick, argumenta que una máquina podrá hablar sobre el poema que ha compuesto, entendiendo que lo ha escrito ella sola, sin que haya nadie detrás. Hofstadter argumentará que es cuestión de tiempo, hasta que alcancemos un algoritmo lo suficientemente complejo.

Sin embargo, la cuestión de que detrás de la inteligencia artificial siempre habrá alguien, puede tener un cierto fundamento matemático. Es el llamado problema de la decisión.

Sin respuesta para todo

Curiosamente la idea de los ordenadores surgió para resolver un problema matemático planteado inicialmente en el siglo XVIII por Leibniz, quien pensó en una máquina que mediante un sistema lógico formal pudiera servir para razonar cualquier tipo de proposición. Su ilusión era poder apretar un botón de tal máquina y dejar que el sistema calculara hasta dar con la solución a cualquier problema. Según una famosa cita suya, si hubiera cualquier tipo de disputa entre dos personas, bastaría con usar esta máquina y «simplemente decir: calculemos, y sin más preámbulos, veamos quien tiene razón»[65]. Durante años se pensó si tal máquina sería posible.

David Hilbert y Wilhelm Ackermann en 1928[66] plantearon formalmente dicha pregunta con el enunciado del llamado «Problema de la decisión» (*Entscheidungsproblem*). Se preguntaban si existiría un algoritmo único que pudiera determinar si cualquier proposición era verdadera o falsa. Debería ser un algoritmo para el cual, dado un enunciado y con unas reglas

de juego, fuera capaz de responder «sí» o «no» en función de si el enunciado era verdadero o falso. Es la idea de tener ese botón, en el que pensó de Leibniz, que arranca un algoritmo que me determina si lo que le pregunto es verdad o mentira. Para nuestra desgracia, o fortuna, quién sabe, no existe tal algoritmo.

Así lo demostraron, de manera independiente y mediante razonamientos distintos, Alonzo Church[67] y Alan Turing[68] hacia 1936. La prueba de Turing tiene más relevancia para nosotros, que nos ocupamos de esto de la inteligencia artificial, dado que supuso el desarrollo de lo que conocemos como máquina de Turing y que es la inspiración de los actuales ordenadores. En cierta medida, hoy hablamos de inteligencia artificial porque hace siglos Leibniz planteó un problema matemático.

La demostración que refuta el «Problema de la decisión» es compleja. Lo que interesa es que, en definitiva, se demostró que no existe un algoritmo que pueda servir para todo. No existe la máquina de Leibniz, con un botón que resuelve cualquier tipo de problema. Para cada pregunta debo arrancar un algoritmo distinto, es decir, apretar un botón diferente. No es una cuestión de tiempo, de saber más y dar con el algoritmo adecuado que lo resuelva todo. Es, sencillamente, que no existe.

Gracias a este resultado los diseñadores de *software* tienen su futuro asegurado. Debido a que el «Problema de la decisión» es falso, somos nosotros los que decidimos qué botón apretar, es decir, qué algoritmo válido debemos diseñar para cada caso.

Roger Penrose, en su obra *La Nueva Mente del Emperador*[69], defiende que la refutación del «Problema de la decisión» nos indica que un algoritmo no puede establecer siempre la validez de otro algoritmo, no puede decidir por sí mismo. Debe existir algo externo que decida qué algoritmo utilizar para determinar la validez de algo. Ese algo somos nosotros. Nosotros somos los que decidimos a través de algoritmos.

Nosotros decidimos si pintar un retrato a la manera de Rembrandt, si crear cartas de amor, si otorgar una hipoteca a un cliente o si incitar a alguien a permanecer conectado a las redes sociales. Además, decidimos cómo hacerlo. Podremos utilizar la inteligencia artificial, pero somos nosotros los que pensamos, juzgamos y decidimos qué hacer y qué algoritmo arrancar.

Una pipa sin consciencia de ser pipa

Que sea como nosotros

Hasta ahora hemos visto lo que no es la inteligencia artificial. Corresponde ya, y más teniendo en cuenta de qué va este libro, saber qué es la inteligencia artificial.

Lo primero a considerar es que no hablamos de un único tipo de inteligencia artificial. En realidad, no tenemos una inteligencia artificial, sino que podemos decir que tenemos varias inteligencias artificiales. Vimos que el concepto de inteligencia artificial surgió de la llamada Conferencia de Dartmouth con el propósito de analizar cómo es nuestra inteligencia y cómo una máquina podría simularla. También hemos visto que nuestra inteligencia comprende varias habilidades, tales como razonar, planificar, resolver problemas, pensar en conceptos abstractos, comprender ideas complejas o aprender. Los distintos tipos de inteligencia artificial se centran en habilidades particulares de eso que llamamos inteligencia. Esto lleva a que cada inteligencia artificial tenga objetivos distintos y se fundamente en áreas de investigación específicas.

La propuesta de Turing, respecto a decir que un sistema es inteligente si pasa el juego de simulación, ha derivado en un tipo de investigación que buscar superar ese objetivo. El propósito es crear sistemas inteligentes que, bajo ciertas circunstancias, actúen de igual forma a como lo haríamos nosotros, de tal suerte que lleguemos a pensar que son personas de carne y hueso.

Es el caso del sistema ELIZA que hemos visto. El objetivo de ELIZA, que simula ser un doctor o doctora en psicología, no es ofrecer diagnósticos acertados sobre nuestra situación médica, sino hacernos creer que estamos hablando con una persona que ha estudiado psicología. Igualmente, la demostración que hizo Google sobre cómo su asistente personal inteligente era capaz de realizar una cita en una peluquería, no era para ir finalmente a cortarse el pelo. No me imagino a «Ok Google» entrando a saltitos en la

peluquería para atusar su boliche sin pelo. El objetivo era decir: «Mirad lo que soy capaz de hacer, mi asistente de voz es capaz de engañarte».

En general, crear un sistema inteligente para que pase el test de Turing presenta poco interés dentro de la investigación científica de la inteligencia artificial. Está bien como divertimento en ferias y espectáculos. Tal y como argumentan reconocidos autores, como Russell y Novig[70], el gran salto de la aviación sucedió cuando se pasó de simular a las aves a utilizar los túneles de viento y aplicar las leyes de la aerodinámica. Hoy los aviones nos llevan con seguridad de un punto a otro porque el objetivo de la aeronáutica no es crear máquinas que vuelen igual que las palomas y puedan engañar a otras palomas. Científicamente hablando es irrelevante un supuesto «test de la Paloma» para la aviación. Así sucede con el test de Turing.

No obstante, sí debemos reconocer que, con el objetivo de pasar la prueba de Turing, se han desarrollado una serie de campos específicos que han resultado beneficiosos para el desarrollo general de la inteligencia artificial. Pasar el test significa tener la capacidad de entender el lenguaje natural y de adaptarse a nuevas situaciones para poder responder en distintas circunstancias. Esto significa que es necesario investigar en las siguientes áreas de la inteligencia artificial:

- Procesamiento de lenguaje natural, para que la máquina pueda entender lo que se le dice.
- Representación del conocimiento, para almacenar de manera estructurada el conocimiento.
- Razonamiento automático, para interpretar la información que recibe y extraer conclusiones.
- Aprendizaje automático (*machine learning*), para detectar y extrapolar patrones y poder adaptarse a nuevas circunstancias.

El test de Turing inicial se planteó mediante un juego de simulación en el que la interacción entre la persona y la máquina era por escrito. Posteriormente se propuso el llamado «Total Turing Test»[71] que propone que el sistema inteligente sea además capaz de moverse como nosotros. Esto

implica tener dos nuevas capacidades, como son la visión y movimiento, que dan lugar a desarrollar dos nuevas áreas de investigación:

- La percepción artificial, como, por ejemplo, el reconocimiento de imágenes.
- La traslación y actuación, para que la máquina se pueda mover en su entorno y manipular objetos.

Estas dos tecnologías se suelen aplicar a la robótica, que es otra rama de la inteligencia artificial.

Cuando hablamos de inteligencia artificial solemos hablar de todos estos temas que son componentes, o áreas de investigación de la misma. Aquí vemos cómo la idea de pasar el test de Turing nos puede servir, no tanto para superarlo con éxito, sino para desarrollar una inteligencia artificial en varios ámbitos. Así, ahora somos capaces de mandar robots a Marte, no con el propósito de engañar a supuestos marcianos y hacerles pensar que hablan con humanos, sino para estudiar Marte, mejor incluso de lo que lo haríamos nosotros. Esto nos lleva al modelo de inteligencia artificial más común que se está desarrollando.

Obtener el mejor resultado

La investigación de la inteligencia artificial se centra en lo que se conocen como agentes inteligentes o agentes racionales. El objetivo de esta inteligencia artificial es crear sistemas que, dado un entorno, puedan actuar para obtener el mejor resultado posible. Dos elementos: un entorno y uno, o varios, objetivos relacionados con este entorno.

Su funcionamiento se basa en los llamados agentes inteligentes, o agentes racionales. Un agente no es más que algo que hace algo en un entorno dado. Por ejemplo, esa máquina que corretea por nuestra casa barriendo el suelo, es un agente inteligente. Su objetivo es, dado un entorno (el suelo de nuestra casa), obtener el mejor resultado, que es dejarlo limpio.

Desde un punto de vista de investigación, los agentes inteligentes pueden trabajar cualquiera de los campos de la inteligencia artificial que hemos

comentado en el punto anterior, si bien solo utilizan aquellos que les son estrictamente necesarios.

Así, nuestro agente barredor utiliza algunas de las áreas de investigación de la inteligencia artificial para realizar su tarea. Primero emplea razonamiento automático con reglas básicas. Reglas en el sentido de, por ejemplo: si me encuentro con una pared, me paro; si detecto agua, utilizo un tipo especial de limpieza. Dado que su objetivo solo es limpiar el suelo, no se le pide cosas muy complicadas, como, por ejemplo, analizar la basura recogida. Pero bien se podría hacer, si se quisiera, y quizás ofrecer recomendaciones a su dueño, tales como «no comas tantas patatas fritas, o no las dejes por el suelo» —no sé si nos gustaría tal robot indiscreto—. En este caso, el agente tendría un entorno, el suelo de nuestra casa, pero dos objetivos: limpiarlo y regañarnos por mala alimentación.

También utiliza las tecnologías de reconocimiento visual, para identificar obstáculos o mascotas, y la tecnología de traslación y actuación para moverse y limpiar, que es lo suyo. En este caso, el agente barredor no utiliza el procesamiento de lenguaje natural, pues, por el momento, no hablamos con él. Pero sería posible, y así, le podríamos decir «no me barras los pies, que trae mala suerte», y el agente obraría sumisamente.

Técnicamente hablando, un agente inteligente es aquel que actúa de manera inteligente. Pero esto no nos dice nada, si no aclaramos qué significa actuar de manera inteligente. En el fondo esa es la cuestión que venimos hablando en este capítulo: saber qué se entiende por inteligencia en la inteligencia artificial.

Dentro de la investigación de la inteligencia artificial, esta cuestión de actuar de manera inteligente no se basa en temas filosóficos sobre la naturaleza humana y la consciencia. Se dice que un agente actúa de manera inteligente si cumple lo siguiente[72]:

1. Sus acciones se ajustan a sus objetivos y circunstancias.
2. Es flexible frente a los cambios en el entorno y en sus objetivos.
3. Aprende de la experiencia.
4. Realiza acciones adecuadas dada su limitación de percepción y su computación finita.

Según esta definición tan concreta, nuestro agente barredor es inteligente, puesto que:

1. Sus acciones se ajustan a sus objetivos y circunstancias: me deja el suelo limpio, siempre y cuando yo no sea excesivamente guarro.
2. Es flexible frente al cambios en el entorno y en sus objetivos: si me compro un gato, lo detecta y no lo barre.
3. Aprende de la experiencia: con el tiempo puede determinar cuánto tiempo le va a llevar barrer la casa y si necesita recargar la batería o no.
4. Realiza acciones adecuadas dada su limitación de percepción y su computación finita: quizás no reconozca todo a la perfección y un día barra mi anillo de matrimonio que accidentalmente se me cayó (qué estaría haciendo), y no le pidas que reconozca obstáculos a gran distancia. No obstante, barre adecuadamente, a pesar de tener ciertas limitaciones.

¿Todo esto es inteligencia? ¿Esto que hace nuestro barredor autómata está relacionado con lo que hemos visto de inteligencia? Sí.

No una sino varias

Al hablar de inteligencia veíamos que esta comprende distintas habilidades: razonar, planificar, resolver problemas, pensar en conceptos abstractos, comprender ideas complejas, aprender de la experiencia. Ser inteligente no significa usar todas y cada una de estas habilidades. Nosotros no siempre estamos pensando en conceptos abstractos, y es bien notorio que no siempre aprendemos de la experiencia. Igual sucede con la inteligencia artificial basada en agentes inteligentes. Un agente inteligente no tiene que utilizar todas las habilidades que comprende la idea de inteligencia.

Ahora bien, nuestro humilde barredor utiliza varias de estas habilidades. Cuando barre, está razonando, porque está aplicando cierto razonamiento lógico, es decir, ciertos silogismos del estilo: «si me encuentro un gato, entonces parar e ir hacia atrás». Esto le sirve para resolver el tipo de problemas que se va a encontrar. También planifica, pues establece una secuencia de acciones para cumplir su objetivo de tener nuestro suelo limpio.

Y, finalmente, puede aprender dónde hay zonas más difíciles, según su experiencia previa. Por tanto, nuestro barredor sin pretensiones es un agente inteligente en toda regla.

Hay algo que no hace: no piensa, en el sentido de emitir un juicio. Seguro que no está aplicando esa capacidad más amplia y profunda de comprender el entorno, de captar o dar sentido a lo que nos rodea; de valorar, sopesar y juzgar. Seguro que no está pensando «¡qué guarro es mi dueño!».

Tenemos, por tanto, varios tipos de inteligencia artificial, según nuestro objetivo de investigación. Una de ellas puede estar encaminada a simular nuestro comportamiento, y llegado el caso, pasar el test de Turing. Otra se fundamenta en agentes inteligentes, cuyo objetivo es, dado un entorno de actuación, obtener el mejor resultado posible. Existen además distintas tecnologías que ayudan a desarrollar estos tipos de inteligencia artificial: procesamiento de lenguaje natural, representación del conocimiento, razonamiento automático, aprendizaje automático (*machine learning*), percepción —por ejemplo, reconocimiento de imágenes—, traslación y actuación. Cada tipo de inteligencia artificial utilizará todas o parte de estas tecnologías, según sea su objetivo. A modo de ejemplo, en la tabla siguiente tienes una serie de sistemas de inteligencia artificial con las tecnologías que utilizan.

Procesamiento del lenguaje natural	Representación del conocimiento	Razonamiento automático	Aprendizaje automático	Percepción artificial	Traslación y actuación
Vehículo autónomo					
X	X	X	X	X	X
Pago a través del móvil por reconocimiento facial					
		X	X	X	
Sistema de recomendación de contenidos					
	X	X	X		
Sistema de diagnóstico médico por imágenes					
	X	X	X	X	
Chatbot de resolución de dudas					

X	X	X	X		
		Barredor de casa			
		X	X	X	X

sobre si la máquina piensa o emite juicios y si tiene una consciencia. Según su propuesta, basta con que parezca que emite juicios.

Magritte es puro surrealismo. Es el surrealismo que en el prólogo de la obra de teatro *Las tetas de Tiresias* de Apollinaire proclama que «cuando el hombre quiso imitar el caminar, creó la rueda que no parece una pierna».

La inteligencia artificial comparte, o debe compartir, esa misma idea surrealista. Si queremos mejorar el caminar, creamos la rueda, no nuevas piernas; si queremos mejorar el volar, creamos aviones, no nuevos pájaros. De igual forma, si queremos mejorar nuestra capacidad de obrar en el mundo, no es necesario crear una tecnología que haga lo que nosotros hacemos, sino crear una tecnología que sea muy buena en aquello en lo que nosotros no lo somos tanto, por ejemplo, en calcular rápidamente.

Cuando Magritte pintó una pipa, debajo escribió «esto no es una pipa». Tenía razón. Aquello no era una pipa, sino la representación de una pipa. De igual forma, la inteligencia artificial no es exactamente como nuestra inteligencia y no tiene por qué serlo. La inteligencia artificial no es capaz, al menos por ahora, de pensar, en el sentido de comprender una situación, valorarla y juzgarla. Ni falta que hace. Para eso estamos nosotros, quienes juzgaremos, mejor o peor, pero no creo que la inteligencia artificial sea capaz de juzgar mejor que nosotros. La justicia no es cuestión de matemáticas y cálculos estadísticos.

El objetivo no es que en un momento dado la inteligencia artificial nos llegue a engañar y se haga pasar por uno de nosotros —otra cosa es que otros busquen engañar con la inteligencia artificial—. El objetivo de la inteligencia artificial, con esa visión de los agentes inteligentes, es obtener el mejor resultado en un entorno dado. De igual forma que una rueda o un avión me dan un mejor resultado. Y no me preocupa si la rueda o el avión se parecen a mi capacidad de andar o de moverme por el aire.

Quizás la polémica de la inteligencia artificial es que tiene la palabra *inteligencia*, y eso nos recuerda a nosotros. Es el cuadro de una pipa que no es una pipa. Si quiero fumar, no me compro un cuadro que representa una pipa. De igual forma, si quiero pensar, no debo acudir a máquina que simula pensar. Podré acudir a ella como apoyo, como complemento. Para pensar, estamos nosotros.

El arte nos distinguirá de la inteligencia artificial —con permiso de Turing y Hofstadter—. Detrás de una obra de arte siempre hay alguien que quiere contar algo. Ya sea mediante pinceles, un burro, o mediante inteligencia artificial. Detrás de la inteligencia artificial siempre hay alguien que toma la decisión de cómo hacer algo. Nosotros elegimos el algoritmo. Nosotros tenemos la posibilidad de huir de los patrones, de entender y dar sentido a nuestro entorno y ser capaces de crear algo único que diga:

Me gustas cuando callas porque estás como ausente,
y me oyes desde lejos, y mi voz no te toca.
Parece que los ojos se te hubieran volado
y parece que un beso te cerrara la boca.

Veinte poemas de amor y una canción desesperada.

Pablo Neruda

Mis conclusiones. ¿Las tuyas?

En este capítulo hemos abordado la siguiente cuestión: ¿Es la inteligencia artificial igual a nosotros? Estas son mis propuestas de respuesta:

- Nuestra inteligencia es una capacidad que nos permite razonar, planificar, resolver problemas, pensar en conceptos abstractos, comprender ideas complejas, aprender rápidamente y aprender de la experiencia. Todas estas capacidades están encaminadas a una capacidad más amplia y profunda de comprender nuestro entorno, lo que significa captar o dar sentido a lo que nos rodea.
- Esta capacidad de dar sentido a lo que nos rodea está relacionada con pensar, en el sentido de emitir un juicio.
- Pensar, a su vez, necesita de una consciencia, que es la capacidad de percibirse a sí mismo y en el mundo.
- Existen distintas visiones sobre si la inteligencia artificial tiene o podrá tener consciencia. La inteligencia artificial fuerte (Hofstadter) defiende que es un tema de complejidad de los algoritmos, y en un momento dado, tendrá consciencia. Searle, con su propuesta de Habitación China defiende que la inteligencia artificial no sabe lo que hace y solo trabaja con símbolos. Turing dice que si no nos preguntamos si los demás tienen consciencia y lo damos por hecho, por qué negarlo para la inteligencia artificial.
- El Problema de la Decisión nos indica que detrás de una inteligencia artificial siempre hay alguien que toma una decisión respecto a qué hacer y cómo hacerlo.
- La inteligencia artificial actual se fundamenta en los llamados agentes inteligentes, que son aquello que actúan en un entorno para obtener el mejor resultado.
- Estos agentes son inteligentes porque sus acciones se ajustan a sus objetivos y circunstancias, son flexibles frente al cambios en el entorno y en sus objetivos, aprenden de la experiencia, realizan acciones

adecuadas dada sus limitaciones. Se les considera inteligentes por hacer esto, no se busca, por ahora, que tengan consciencia.
- La inteligencia artificial utiliza distintas tecnologías.

Estas son mis conclusiones. Pero sé inteligente y piénsalas, en el sentido que hemos visto de emitir un juicio. Júzgalas y quizás obtengas otras.

Mujer con abanico, María Blanchard, 1916. Museo Nacional Centro de Arte Reina Sofía

Será por éticas

—¡Pues lo que haga más feliz al usuario! Todos buscamos la felicidad y tenemos que transmitir la idea de que esa felicidad es posible. ¡Con esa visión nació nuestra empresa!

Todo el Comité de Dirección se quedó callado tras escuchar las palabas de Ariadna, directora ejecutiva de la *startup* AlegrIA. La empresa comenzó a rodar hacía apenas dos años con el propósito de ofrecer los mejores contenidos multimedia a sus clientes mediante una potente plataforma basada en inteligencia artificial. Aquella mañana el Comité de Dirección estaba reunido para revisar el criterio del algoritmo de inteligencia artificial que recomendaba nuevos contenidos a los usuarios.

La sesión se inició compartiendo cada uno cómo se sentía emocionalmente en ese momento. Así lo indicaban las últimas buenas prácticas de gestión de reuniones, las cuales incidían en la importancia de contar a tus compañeros cómo te sentías, para crear un vínculo afectivo favorable al comienzo de cada reunión. Todo el mundo manifestó sentirse contento, ilusionado, a tope, retador o de cualquier otra forma que significaba eso que llaman energía positiva, salvo Nieves, quien manifestó sentirse algo cansada por pasar mala noche. Sin reparos, pero de forma guay,

el resto del comité la animó a ver la situación desde una perspectiva positiva, cosa que Nieves no terminó de entender, quizás por el sueño que tenía, quizás por la tontería de propuesta.

La reunión se inició como una tormenta de ideas, donde cada cual podía expresar lo que quisiera, sin ser juzgado por ello. A pesar de tanta energía positiva y tan buena intención, los ánimos se fueron caldeando, al igual que en las reuniones tradicionales, llegando un momento en el que había más tormenta que ideas. Fue cuando intervino Ariadna, con las palabras que hemos visto. Se hizo un silencio momentáneo, que pronto fue roto para ser respondida.

—Vale, pero qué tipo de felicidad, porque esto depende de los gustos de cada uno —replicó Humberto con autosuficiencia.

—Buscar la felicidad de cada uno es complicado. Mejor ofrezcamos lo que le gusta a la mayoría y así hay menos riesgo de equivocarnos. Digo yo —propuso, Benito, mientras se encogía de hombros, como queriendo no meter ruido.

—Bueno, ¡vamos a ver! Esto tampoco puede ser lo que quiera cada uno o lo que quiera la mayoría. ¡Así nos va! —saltó Tomás, con su gesto de complacencia habitual siempre que hacía una crítica a la sociedad actual—. De alguna manera todos sabemos que lo que está bien o no. Apliquemos el sentido común, ¡por favor!

—Pues partamos entonces de unos principios. ¡Tiene que haber calidad! —replicó con energía Kanza.

—Un poco de orden, ¡por favor! —gritó Habana, levantando las palmas de las manos buscando concordia—. Tendremos que buscar una aprobación de la mayoría, un diálogo entre todas las partes, con los usuarios, los productores de contenidos…

—Pero, ¿quiénes somos nosotros para determinar lo que está bien o lo que está mal? —irrumpió Nieves, quien siempre mantenía una visión escéptica de la vida.

En ese momento surgió la tormenta perfecta, todos hablando a la vez, apoyando sus ideas, sin que nadie escuchara a nadie. En un extremo de la mesa se encontraba un joven, más joven que los miembros del Comité de Dirección de AlegrIA, que tenía la dificultosa labor de tomar notas de la

reunión, para luego levantar acta y compartirla con todos los presentes. Era el becario. Hasta ahora aquella labor había sido más o menos posible. Ahora resultaba imposible en aquella batalla verbal por obtener la razón a gritos.

Aprovechando que tenía bajo su control el proyector de vídeo, en un momento del acalorado debate, quitó la presentación que había servido de base para revisar la estrategia de contenidos, y proyectó una imagen de un cuadro que había estado buscando por Internet. El cuadro era *Mujer con abanico*, pintura cubista de la artista santanderina María Blanchard. Dejó de tomar nota y esperó.

Al cabo de unos segundos, se hizo el silencio cuando todos se percataron de que un cuadro raro había sustituido la notable presentación de la directora. Ariadna espetó:

—¡Esto qué es! ¿Has tomado nota de todo lo que hemos dicho?

—Esto es la solución —respondió el becario con una sonrisa abierta.

Obviamente, el becario fue despedido de AlegrIA por comportamiento inesperado. Nieves dijo que aquello del cuadro se podría considerar una visión innovadora, pero fue acallada bajo la crítica unánime de que no estaba lúcida por falta de sueño. El becario tenía razón en su propuesta.

La *startup* AlegrIA es ficticia, y los diálogos de su supuesto Comité de Dirección también. Sin embargo, todos los argumentos que hemos visto son frases que hoy en día podemos escuchar en cualquier tipo de debate o discusión. Nos pueden parecer novedosos y hacernos creer que nuestro pensamiento es moderno. Pero todos esos argumentos han sido expuestos en los siglos y siglos de filosofía que llevamos, desde los tiempos de los griegos. Todo lo que pensamos hoy en día, ya ha sido pensado en el pasado, y, lo que es mejor, ya ha sido discutido, aprobado y rebatido, por lo que conocemos las ventajas e inconvenientes de cada idea que propongamos. Esta es la gran ventaja de conocer la filosofía, que no pierdes el tiempo, porque lo que escuchas ya lo conoces.

La historia de la filosofía, al menos en lo que respecta a la ética, es similar a un hilo de Twitter, donde continuamente se proponen y rebaten propuestas de teorías. En un momento dado, un filósofo propone una teoría, que siglos después es discutida y rebatida por una nueva teoría de otro filósofo, que siglos después es discutida y rebatida por otra teoría de otro filósofo, que

siglos después… Así hasta nuestros días. La diferencia con Twitter es que, en la historia de la filosofía el debate es respetuoso y racional.

Por eso, hoy en día, tenemos tantas visiones sobre lo que está bien y lo que está mal y no llegamos un consenso sobre una visión única. Tenemos un cuadro ético cubista, como el cuadro *Mujer con abanico*, que presentó el becario de AlegrIA. La pintura representa una mujer, con falda roja, que parece llevar un abanico amarillo en su mano izquierda. La imagen de la mujer está descompuesta en distintos puntos de vista. Como rota o fragmentada en distintos planos. Así es el cubismo. Se abandona la visión de la realidad desde una perspectiva única y se intenta plasmar en un lienzo plano toda la riqueza de puntos de vista que tiene el espacio en el que nos movemos.

El cubismo nació como respuesta a la diversidad. En particular cuando artistas, como Picasso, se encontraron con el arte africano, que ofrecía imágenes de formas simplificadas y con una visión totalmente distinta de la cultura tradicional renacentista. Aquel nuevo arte era otro punto de vista. Esa misma diversidad de puntos de vista la tenemos en la ética actual, tras siglos y siglos de debate, y, por ello, la moral de nuestros días nos parece un cuadro cubista difícil de ver. Un cuadro en el que cada uno tiene su opinión sobre lo que está bien o está mal.

Esta diversidad de visiones se traslada también al ámbito de la inteligencia artificial. Si queremos crear una inteligencia artificial ética, surge la primera cuestión, que parece sin respuesta aparente: ¿qué ética aplicamos en la inteligencia artificial? Porque no tenemos un cuadro ético definido; tenemos un cuadro ético cubista formado por muchos puntos de vista, por muchas teorías éticas.

Eso es lo que le ocurría al Comité de Dirección de AlegrIA. Había demasiadas teorías éticas en juego. Los nombres de sus miembros no son casuales. Sus primeras letras hacen guiños a distintos filósofos. Ariadna es de la escuela aristotélica —no en todo—; Humberto se guía por los sentidos, como Hume; Benito es utilitarista, como Bentham; Tomás sigue a santo Tomás —este es fácil—; Kanza es kantiana, busca principios, como Kant; Habana se inspira en el discurso práctico de Habermas; y Nieves lo rompe todo, como hizo Nietzsche con la ética.

Cada uno de ellos es *follower* de un filósofo, y no es consciente de ello. Habla creyendo que es original, sin saber que su pensamiento está subscrito a la cuenta Instagram de un filósofo. Así sucede hoy en día a todos. Existe la falsa creencia de que la filosofía son «pájaros y flores», y lo que dice no se puede aplicar a la vida real. Sin embargo, los argumentos de cada uno de los miembros de AlegrIA son «realmente reales», tan reales como que los hemos dicho o escuchado más de una vez. Y todos ellos son la consecuencia de una corriente filosófica.

¿Cómo entender y aplicar este cuadro ético cubista? Si tenemos un cubismo ético, ¿es posible tener una inteligencia artificial ética? La respuesta es sí. Vamos a verlo. Primero, vamos a profundizar en estas distintas perspectivas éticas, que hemos visto en las distintas voces del Comité de Dirección de la *startup*., porque representan teorías éticas relevantes de la historia de la filosofía, con sus pros y sus contras. Además, lo vamos a ver siguiendo el objetivo que tenía el Comité de Dirección de AlegrIA, que era revisar el algoritmo de recomendación de contenidos.

Actualmente, se utiliza la inteligencia artificial en muchos sistemas que sirven para ofrecer contenidos; ya sea, por ejemplo, de películas en Netflix, de historias en Instagram o de notificaciones en LinkedIn. Te propongo ver qué dirían estos filósofos si les preguntáramos cómo hacer que la inteligencia artificial para la propuesta de contenidos fuera ética. Con esto, entenderemos de verdad el cuadro cubista que tenemos. Para nuestra tranquilidad, en el siguiente capítulo, daremos la solución a cómo ver un cuadro cubista.

¿Qué es esto de la ética?

Unas primeras preguntas éticas

Nosotros nos comportamos con unas reglas que no siempre son únicas. Si una leona caza y mata a su presa, la respuesta de por qué actuó así es clara y única: porque tenía hambre. Y cada vez que tenga hambre, cazará y matará. Pero nosotros podemos tener hambre y actuar de distintas formas: podemos matar o robar por conseguir comida; podemos trabajar por conseguirla; si conseguimos comida, podemos comerla toda o compartirla con otros; o podemos ayunar. ¿Por qué podemos actuar de distintas formas? ¿es alguna de ellas mejor que otra? ¿es alguna la correcta, la que «debe ser»? Estas preguntas tan habituales son preguntas éticas, principalmente la última, que se resume en: ¿cómo debo actuar? Las teorías éticas intentan responder a tales preguntas desde distintos puntos de vista.

Una primera teoría ética, que se considera como el origen en sí de la ética, se encuentra en el diálogo de Platón llamado *Protágoras*. En él se cuenta un mito con el cual se pretende dar respuesta a de las preguntas anteriores[73]:

> Una vez que Zeus creó la tierra y todos los seres vivos, se dio cuenta de que tenía que dotarlos de recursos para que pudieran sobrevivir. Para ello recurrió al titán Epimeteo a quien envió a la tierra para distribuir a todas las especies distintas facultades. De esta forma, por ejemplo, a unos animales les dotó de velocidad, a otros de fuerza, a otros les dio alas para volar y a otros la habilidad para guarecerse en cualquier escondrijo. Distribuyó así todas las facultades y por eso hoy tenemos tantos animales con capacidades tan diversas. Pero Epimeteo no era muy precavido y se gastó todos los dones en los animales de tal forma que no le quedó ningún recurso que dar a la especie humana.
>
> Prometo, su hermano, se percató de tal hecho e intentó solucionarlo. Robó el fuego a Hermes y las distintas artes de obrar a Atenea. El fuego lo dio para conocimiento de todos los humanos y las distintas artes las repartió entre distintos seres humanos de forma individual. De esta forma, los seres humanos nos especializamos en distintas tareas. Algunos recibieron el arte de hacer calzado, otros de hacer ropas, aquellos de cultivar la tierra, construir naves o edificios.
>
> Parecía que Prometeo había resuelto el error de su hermano. El problema surgió, precisamente, cuando se crearon las ciudades y los seres humanos se juntaron en

ellas. Entonces se empezaron a pelear unos con otros. Habíamos recibido muchos dones, pero nos faltaba el sentido de la convivencia.
Tuvo que intervenir Zeus para salvar a la especia humana. Envió a Hermes para distribuir entre los seres humanos dos dones adicionales: el sentido del honor y el sentido de la justicia. Antes de partir, Hermes le preguntó a Zeus cómo debía repartir tales dones. Si debía ser como las anteriores artes, que se otorgaron de forma individual, o bien debía dotarlos a todos por igual en el sentido del honor y la justicia. La respuesta de Zeus fue clara: debía repartir estos dos dones entre todos por igual, pues si participaran de ellos solo unos pocos, como ocurría con las demás artes, jamás habría ciudades. Además, daba la facultad a los hombres para poder expulsar de la ciudad a aquellos que no participaran del honor y la justicia.

Esta historia resuelve varias cuestiones. Primero, ya sabemos por qué tenemos fuego y por qué existen distintos oficios. Después, se explica por qué los seres humanos tenemos una idea de qué es justo y por qué tendemos a esa justicia. El don de la justicia nos dice qué es lo correcto, lo que debe ser; mientras que el don del honor es ese sentimiento de vergüenza que podemos tener si no hacemos eso que es justo o correcto. Según la historia de Prometeo, sabemos qué es lo justo y nos da vergüenza ante los demás no realizar tal justicia. Ese sentimiento de vergüenza es el origen de la moral.

Por último, este mito aporta una cuestión adicional: no todos podemos opinar sobre cómo hacer zapatos, pero todos podemos opinar sobre cómo ser justos. Este punto es interesante porque pasa de soslayo en el texto, pero defiende esta idea tan actual de que en el ámbito de la ética todo es opinable.

Las artes de obrar que Prometeo robó a Atenea, las repartió de forma individual. De esta forma, unos son zapateros, otros sastres, otros agricultores y otros arquitectos. Si queremos saber cómo se hace un zapato, preguntamos a los zapateros y solo a los zapateros, pues solo ellos saben de zapatos. Solo los zapateros pueden opinar sobre cómo hacer zapatos. Sin embargo, el don de la justicia se repartió entre todos los seres humanos. Esto significa que todos los seres humanos sabemos sobre la justicia. Por tanto, todos podemos opinar sobre qué es o no es justo. Ahora entendemos por qué en los debates de televisión acuden y opinan tertulianos de toda índole para hablar de lo bueno y lo malo: resulta que aplican el mito de Prometeo, quizás sin saberlo.

Esta conclusión, algo simpática, pone de manifiesto lo práctico de la filosofía. Puede parecer que hablamos de historietas teóricas, que no resuelven nada, pero no es verdad. Al comienzo de este apartado nos

hacíamos una serie de preguntas, a raíz de nuestro comportamiento respecto a de una leona: ¿por qué podemos actuar de distintas formas? ¿es alguna de ellas mejor que otra? ¿es alguna la correcta, la que «debe ser»? Este mito de Prometeo responde a tales preguntas.

¿Por qué podemos actuar de distintas formas? Los animales recibieron dones y actúan según tales dones. La leona tiene la fuerza y velocidad para cazar. Nosotros recibimos el don de la justicia que nos hace obrar según las circunstancias.

¿Es alguna de ellas mejor que otra? ¿es alguna la correcta, la que «debe ser»? Sí, y esa forma correcta la determina el propio ser humano, porque todos podemos opinar sobre qué es justo y además podemos expulsar de la ciudad a quien consideremos que no cumple tal idea de justicia.

Como vemos, detrás de cada teoría ética aparece una idea sobre la realidad y sobre el ser humano, que tiene su efecto práctico. Para el mundo griego según este mito el ser humano puede opinar sobre cualquier tema. Esto se traduce en la famosa frase, que seguro hemos escuchado de que «el hombre es la medida de todas las cosas». Hoy en día, pensamos de igual forma, si bien desde fundamentos distintos.

Este texto de Prometeo se considera como una primera teoría ética: tenemos el don de la justicia y queremos hacer justicia porque una divinidad, Zeus, nos ha dado el don de entender qué es justo y de tener vergüenza si no lo hacemos; además, explica por qué todos podemos opinar sobre qué es justo. Esta teoría responde a la cuestión de cómo sabemos lo que es justo y por qué intentamos hacer el bien y la justica, pero, inevitablemente, levanta otras cuestiones: ¿si no creo en dioses, de dónde viene ese don?, ¿se puede discutir sobre la justicia de la misma forma que se discute sobre cómo hacer zapatos?; si hay unas normas para hacer buenos zapatos, ¿cuáles son las normas para ser justo?; si todos podemos opinar sobre esas normas, ¿para qué las normas?

La historia de la ética es la historia de un intento de responder a este tipo de cuestiones. Una teoría ética responde a unas cuestiones de manera más o menos razonada. Pero, como dije antes, al responder a algunas preguntas, nacen otras nuevas preguntas. Viene entonces una nueva teoría que responde a las nuevas preguntas, pero genera otras tantas, o bien es una teoría no

satisfactoria para todos y nace una nueva teoría ética. Así, llegamos al debate del Comité de Dirección de AlegrIA y al cuadro cubista ético. Lo iremos viendo con más detalle. Ahora terminemos de cerrar qué es esto de la ética.

Vida buena con justicia

Hemos visto que el mito de Prometeo es una primera teoría ética que explica por qué tendemos a un comportamiento justo. Sin embargo, si nos fijamos bien, esta historia no dice nada sobre cómo nos tenemos que comportar para ser justos. Explica por qué tenemos el don de la justicia, pero no cómo comportarse para ser justo. Esta es la diferencia entre la *ética* y la *moral*, como ahora veremos.

Habitualmente las palabras ética y moral se entienden como sinónimos y se utilizan de manera indistinta. Parte de tal confusión se debe al origen de estas palabras. *Ética* procede del griego *ethos*, que significaba originariamente «morada», «lugar donde vivimos», y que posteriormente pasó a considerarse como «el modo de ser de una persona»; algo así como la morada particular de cada uno. Por otro lado, *moral* procede del latín *mos*, *moris*, que originariamente significaba «costumbre», pero que evolucionó también a «modo de ser una persona»; sería como la costumbre particular de cada uno. De este modo, tanto *ética* como *moral* confluyen en la idea de todo aquello que se refiere al modo de ser de una persona. Modo de ser que se supone ha sido adquirido como resultado de poner en práctica una cierta costumbre o unos hábitos.

Según esto, ética y moral son lo mismo y, en cualquier conversación, podemos utilizarlas según queramos. No obstante, por si se da el caso de que nos encontramos con alguien puntilloso y preciso, que se siente estupendo en cualquier momento, vamos a profundizar en cómo se diferencian ambos conceptos.

La diferencia radica en qué tipo de preguntas nos hacemos respecto a ese modo de ser o de obrar de una persona. Cuando yo quiero realizar una acción me pueden venir dos tipos de preguntas: una, ¿puedo hacer eso? y dos, ¿por qué debo o no debo hacer eso? La primera pregunta la responde la moral, y la segunda pregunta la responde la ética.

Volviendo al caso de tener hambre, me puedo preguntar dos cosas: ¿puedo robar comida?, o bien, ¿por qué debo o no debo robar comida? La moral responde a la cuestión de si robar o no robar, si bien la respuesta depende de cada tipo de moral. Por ejemplo, la moral cristiana dirá «no robarás», mientras que la moral llamada utilitarista —lo veremos posteriormente— dirá «depende de si robas para hacer el bien o robas para hacer el mal» —de ahí el aprecio romántico que tenemos por Robin Hood, que robaba para el bien de los pobres—. La ética, sin embargo, responde a la pregunta de por qué la moral ha dado esa contestación. Siguiendo con el ejemplo, la ética de santo Tomás, que sigue la moral cristiana, responde que no puedes robar «porque así lo dice la Ley Divina»; mientras que la ética utilitarista, que obviamente sigue la moral utilitarista, responde que en ocasiones puedes robar «porque lo que está bien o está mal depende del nivel de felicidad que cause a la mayoría» —por eso, si robas para dárselo a los pobres, a estilo Robin Hood, habrá una mayoría significativa de gente feliz—.

En resumen, la moral responde a la pregunta: ¿de qué forma debo comportarme?; mientras que la ética me explica: ¿por qué debo comportarme de dicha forma?

Como la moral responde a qué debo hacer, la moral se constituye por un conjunto particular de principios, preceptos, mandamientos o prohibiciones. En palabras del cantautor Joan Manuel Serrat, la moral es la que dice «niño, deja ya de joder con la pelota, que eso no se dice, que eso no se hace, que eso no se toca»[74]. La ética por su lado reflexionaría sobre esos preceptos e intentaría dar respuesta a por qué el niño debe dejar de joder con la pelota, debe dejar de decir eso o hacer eso o tocar eso.

Atendiendo a estos dos tipos de preguntas, ya podemos tener una primera definición sobre la ética y la moral.

La ética es la rama de la filosofía que se ocupa del hecho moral. Dentro de la filosofía existen muchas ramas de conocimiento, cada una con un nombre propio. Si queremos estudiar el conocimiento, y saber, entre otras cosas, por qué conocemos o cómo sabemos si algo es verdad, tenemos una rama de la filosofía que se llama gnoseología o epistemología si hablamos del conocimiento científico —como ves, suelen tener nombres raros—. Si queremos estudiar las cosas en general y saber qué es la materia o qué es real,

acudimos a la rama de la filosofía que se llama ontología. Pues bien, de igual forma, existe una rama de la filosofía, llamada ética, que explica cuestiones sobre el hecho moral.

Fíjate que en la definición anterior hay un pequeño cambio. He dicho «el hecho moral» en lugar de «la moral», que era el término que usábamos al hablar de los dos tipos de preguntas. Son conceptos distintos, pero relacionados, como vemos en esta definición[75]:

> El hecho moral es una dimensión de la vida humana compartida por todos, que consiste en la necesidad inevitable de tomar decisiones y llevar a cabo acciones de las que tenemos que responder ante nosotros mismos y ante los demás, para lo cual buscamos orientación en valores, principios y preceptos.

Ese conjunto de valores, principios, preceptos es lo que llamamos la moral. Por tanto, el hecho moral es una dimensión humana, dentro de la cual se encuentra la moral. Esto puede parecer un juego de palabras o ganas de liarla al más puro estilo filosófico, sin embargo, tiene su importancia.

El hecho moral es una condición que tenemos los seres humanos, en virtud de la cual nos sentimos obligados a responder, ante nosotros y ante la sociedad, sobre lo que hacemos. En la historia del mito de Prometeo vimos que Zeus repartió dos dones: el don de la justicia y el don del honor. Ese don del honor es ese sentimiento de vergüenza que podemos tener si no hacemos eso que pensamos es correcto. Es decir, ese don del honor es lo que podemos entender como el hecho moral. Esa necesidad de explicar por qué hemos hecho algo. Este hecho moral, es un don, una cualidad, que solo tenemos los seres humanos. Por eso decimos que los seres humanos somos agentes morales.

Ahora bien, como resulta que, por necesidad, me siento en la obligación de responder de mis actos, porque soy un agente moral, necesito un libro de instrucciones, una lista FAQ, que me dé normas o indicaciones de qué es lo correcto y cómo debo comportarme. Ese libro de instrucciones es la moral. No existe un único libro de instrucciones, sino distintos libros, es decir, no existe una moral, sino distintas morales que me dicen cómo actuar —de ahí tanto libro aventurero de autoayuda—.

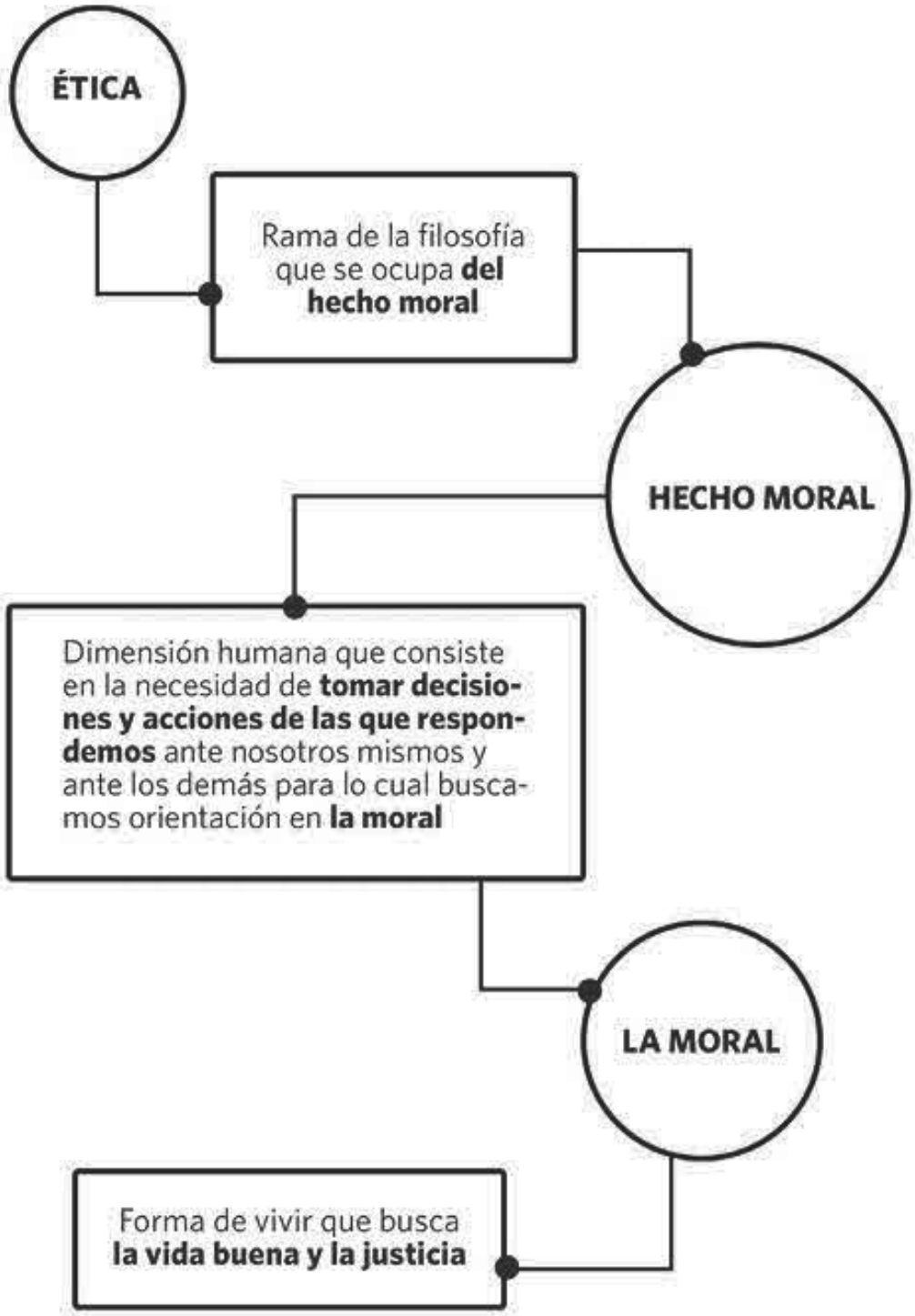
Toda moral dirá muchas cosas, pero básicamente se ocupa de dos temas, que son las preocupaciones centrales del ser humano: la vida buena y la

justicia.

La vida buena es ese tipo de vida «adecuada», «correcta», con la cual nos sentimos bien. No es fácil definir cuál es esa vida adecuada o correcta y, además, depende de cada uno. Hoy en día, que rechazamos el término correcto por verlo demasiado rígido y excluyente, hablamos de vida sana, vida positiva o de comportamientos no tóxicos, términos más suaves, asépticos y menos comprometidos. Cuestión de modas. Si buscamos una definición meditada y, por tanto, más permanente de qué es esa vida buena vemos que consiste en ese tipo de vida que da sentido al destino de la persona, a su razón de ser[76]. En principio, cada moral, si está bien construida, debe ayudarte en esa búsqueda de la vida buena —cosa distinta de la panoplia de consejos vacíos *instagramerianos* que buscan más bien la buena vida fácil—.

Por otro lado, nuestra vida siempre se realiza en sociedad por lo que también nos preocupa cómo buscar esa vida buena dentro la sociedad en la que vivimos. También es complicado, pues implica buscar un equilibrio entre la vida en común, a la vez que preservamos nuestra autonomía como individuos. Este equilibrio entre lo común y la autonomía es el principio de la justicia.

Por tanto, cada moral intenta ofrecer comportamientos que resuelvan dos cuestiones de nuestra vida: ¿qué es la vida buena?, ¿qué es la justicia? En la figura siguiente te muestro un esquema que resume todos estos puntos sobre la ética, el hecho moral y la moral, los cuales se encuentran encadenados.


ÉTICA
Rama de la filosofía que se ocupa **del hecho moral**
HECHO MORAL
Dimensión humana que consiste en la necesidad de **tomar decisiones y acciones de las que respondemos** ante nosotros mismos y ante los demás para lo cual buscamos orientación en **la moral**
LA MORAL
Forma de vivir que busca **la vida buena y la justicia**

Ya vamos viendo la cuestión

Hemos visto que nosotros, los seres humanos, somos agentes morales porque tenemos una necesidad innata de responder por lo que hacemos. ¿Y los robots o la inteligencia artificial? ¿Tiene la inteligencia artificial también esa necesidad de responder por lo que hace? Si tiene esa necesidad, ¿en qué moral busca orientación?

En el capítulo «Esto no es una pipa» veíamos que la inteligencia artificial se diferencia de nosotros en la capacidad de pensar, es decir, en comprender nuestro entorno para dar sentido a lo que nos rodea y poder emitir un juicio. Esta capacidad de pensar, de emitir un juicio, es la que nos permite determinar si juzgamos que lo que hacemos es correcto o no. Es la capacidad que nos permite ser agentes morales. Mientras la inteligencia artificial no tenga esa capacidad de pensar, no podrá ser un agente moral. No tendrá esa necesidad de responder por lo que hace.

La inteligencia artificial no siente la necesidad de ser ética. Nosotros, sí. Nosotros, como agentes morales que somos, tenemos la necesidad de responder por las acciones que hacemos, en este caso, a través de la inteligencia artificial. Si usamos la inteligencia artificial, nuestra preocupación debe ser hacer lo que consideramos correcto mediante la inteligencia artificial. No es tanto que la inteligencia artificial sea ética, sino que nosotros seamos éticos a través de la inteligencia artificial.

Para ello, para ser éticos con la inteligencia artificial, buscamos orientación en la moral. Pero la moral, como hemos visto, no es única, sino que hay varias. Ahora entendemos el debate sin fin del Comité de Dirección de la *startup* AlegrIA.

Vimos que cada miembro de AlegrIA tenía una opinión fundamentada en una ética particular, posiblemente, sin que ellos lo supieran: Ariadna era algo aristotélica, Humberto seguía a Hume, Benito era utilitarista, como Bentham; Tomás hacía honor a su nombre, siguiendo a santo Tomás; Kanza era kantiana, Habana se inspiraba en Habermas y Nieves, escéptica, como Nietzsche. En los siguientes apartados vamos a ver qué significaría aplicar estas teorías éticas a un sistema inteligente, que es lo que pretende cada miembro del Comité de Dirección de AlegrIA. En particular, lo vamos a

hacer para el algoritmo de inteligencia artificial que recomienda nuevos contenidos a los usuarios de la *startup*. Es decir, veremos cómo sería un sistema inteligente de recomendación de contenidos que fuera, por ejemplo, aristotélico, kantiano, utilitarista o nietzscheano.

Porque buscamos la felicidad

Aristóteles: Lo que le haga feliz

«¡Pues lo que haga más feliz al usuario! Todos buscamos la felicidad y tenemos que transmitir la idea de que esa felicidad es posible. ¡Con esa visión nació nuestra empresa!».

Para Aristóteles[77], debemos comenzar por el fin y, en particular, por el fin particular de cada actividad. Cada actividad tiene un objetivo, un fin determinado. Por ejemplo, el fin de la medicina es la salud, el fin de la estrategia es la victoria o el fin de la arquitectura es un edificio. Por ello, si quiero actuar de forma correcta en una actividad, primero debo saber cuál es el fin de dicha actividad. A ese fin Aristóteles lo llama «el bien» de esa actividad. Así, el bien de la medicina —lo que está bien en la medicina— es la salud. No existe «el Bien» genérico, sino «un bien» para cada cosa. Dicho en lenguaje más filosófico, para Aristóteles el bien de cada cosa es su fin.

De manera general, como ser humano, si quiero hacer lo correcto, lo primero que tendré que pensar es cuál es mi objetivo, mi fin como persona. Pero, ¿cuál es el fin del ser humano?, ¿cuál es mi fin en esta vida? Gran pregunta que, posiblemente, nos hemos hecho en alguna ocasión. Aristóteles lo resuelve rápidamente: el fin del ser humano es la felicidad. Nuestro objetivo en esta vida es ser felices. Por tanto, lo que está bien para nosotros es la felicidad. Actualmente, Aristóteles parece estar de moda, si atendemos a tanto discurso *happy flower* que escuchamos. Pero no tanto.

Porque, ¡cuidado!, no vale cualquier tipo de felicidad. Aristóteles habla de una felicidad algo especial, de una «felicidad verdadera» llamada *eudaimonía*. Podemos pensar que la felicidad es tener riquezas, éxito o millones de seguidores en Instagram. Pero Aristóteles dice que no, porque esa felicidad no te lleva al verdadero fin del ser humano. La verdadera felicidad (*eudaimonía*) es aquella que permite realizar la función del ser humano.

Ahora sí que la hemos liado, porque esto parece el típico barullo filosófico, que después de leerlo te quedas como estás. Resulta que el fin del

ser humano es una felicidad verdadera, que solo sabré cuál es cuando sepa cuál es el fin del ser humano. ¡Fantástico, Aristóteles!

Para salir de este lío, Aristóteles define cuál es la función del ser humano: «La función del ser humano es el ejercicio de las actividades del alma de acuerdo con la virtud y durante una vida entera». Responde, pero ahora introduce tres elementos: alma, virtud y vida entera, que tenemos que ir desgranando.

Cuando Aristóteles habla del ejercicio de las actividades del alma, se refiere, en realidad, al ejercicio de la razón en dos dimensiones: en su parte más intelectual o cuando controlamos nuestras apetencias o pasiones de forma razonada. Para Aristóteles el fin del hombre es ejercitar su razón en estos dos ámbitos.

Pero no de una forma cualquiera, sino de forma excelente, que es la virtud (el segundo elemento de la definición). La virtud es el modo especial de ser de cada cosa, es la excelencia de su función. Así, por ejemplo, la virtud del cuchillo es cortar o la virtud del ojo es ver. Y, ¿cuál es la virtud en el ser humano?, ¿cómo realiza el ser humano la excelencia en su función? Aristóteles responde: «La virtud es la disposición o hábito de elegir el término medio relativo a cada uno en acciones y emociones, guiado por la razón y como lo haría una persona prudente». Esta definición no tiene desperdicio y dice muchas cosas:

- La virtud es una *disposición y es un hábito*. Hay que querer ser virtuoso y esto solo se consigue con la práctica.
- Es una disposición que te lleva a *elegir*, a seleccionar, el *término medio* en las acciones y emociones. Por ejemplo, si queremos ser valientes, deberemos escoger el punto medio entre el miedo y la audacia; si queremos ser generosos, debemos elegir el término medio entre el despilfarro y la tacañería.
- Además, es un término medio *relativo a cada uno*, es decir, que no es único, sino que depende de cada persona. Cada individuo es distinto y necesita conocer su punto medio. Esto solo se consigue, de nuevo, con la práctica.

- La persona virtuosa ejercita la razón y elige con *prudencia*. Esta prudencia nos permite escoger lo más oportuno en cada momento respecto a lo que es bueno o malo para nosotros. Para ser prudente hay que deliberar, contrastar opiniones, hay que entender bien la situación y saber juzgar.

La virtud, en resumen, es esa disposición que nos permite elegir con *moderación* (término medio) y *prudencia*. Moderación y prudencia, dos elementos relevantes que después usaremos.

Todo esto de la virtud venía por la definición que vimos de la función del ser humano, que es el ejercicio de las actividades del alma de acuerdo con la virtud y durante una vida entera. Hemos visto a qué se refiere Aristóteles con las actividades del alma (ejercitar la razón y controlar las pasiones) y ahora con la virtud (elegir con moderación y prudencia). Queda la cuestión de la vida entera.

Esto Aristóteles lo resuelve de una manera muy poética diciendo: «una golondrina no hace verano». Ser virtuoso, realizar de forma excelente la función del ser humano no es cosa de una vez, sino de una vida entera. No basta que una vez elijas bien, con moderación y prudencia. Eso es una golondrina casual en el cielo. Ser virtuoso es cuestión de práctica, de probar y errar, y de elegir después de todo de forma consistente con moderación y prudencia. Eso son cientos de golondrinas, que avisan de la verdadera presencia del verano. Aparece, una vez más, esa visión de la experiencia, de la práctica. Ser virtuoso, elegir el punto medio, con moderación y prudencia, no lo vas a aprender viendo *reels* de Instagram. Hay que ponerse manos a la obra, actuar y practicar. ¡Vale!, pero, ¿sobre qué actúo?

Aristóteles propone una serie de virtudes, es decir, una serie de acciones y emociones sobre las cuales encontrar ese punto medio. Así, habla del valor, la templanza, la generosidad, la gentileza (punto medio entre la ira y el servilismo) o la magnanimidad, también llamada grandeza de alma. Esta última es una virtud curiosa porque es reconocerse uno mismo merecedor de grandes cosas, siendo el punto medio entre ser vanidoso y pusilánime. El que tiene grandeza de alma, según Aristóteles, se comporta de la siguiente manera: no pide favores y, si pide algo, lo hace con desgana, mostrando poco

interés; camina lentamente y habla con voz profunda. La grandeza de alma es opuesta a la humildad, la cual se considera un vicio.

Vemos que las virtudes que propone Aristóteles no con como las nuestras. Representan más bien valores propios de la Grecia antigua, que son solo para ciudadanos libres, adultos y acomodados. No hay virtudes para esclavos, trabajadores, artesanos o para las mujeres. Sobre las mujeres, Aristóteles indica que reciben el gobierno de sus maridos y atienden lo que estos les encomiendan y según la división de tareas que tiene cada uno.

Por último, un dato curioso. Aristóteles introduce un elemento de fortuna para ser feliz. Así, por ejemplo, si no tienes la suerte de haber nacido en el sitio adecuado o de tener belleza, es más complicado que seas feliz. Aristóteles habría arrasado hoy en día con esto último, donde se confunde lo bueno con lo bello.

Una vez vistos todos los elementos, corresponde hacer una breve recapitulación, que además te muestro en un gráfico. Según Aristóteles, el fin (el bien) del ser humano es la felicidad. Pero no una felicidad cualquiera, sino una felicidad verdadera (*eudaimonía*), que es aquella que nos permite realizar nuestra función. Esta función es el ejercicio del alma (ejercitar la razón y controlar las pasiones) por medio de la virtud y durante toda tu vida. La virtud es un modo de ser que te permite elegir ese punto medio (moderación) con prudencia (deliberando y sopesando). Esta felicidad verdadera que propone Aristóteles es en definitiva un tipo especial de carácter que solo se consigue con la práctica y el esfuerzo personal.

La felicidad aristotélica en la inteligencia artificial

Ariadna defendía que había que ofrecer a cada usuario los contenidos que les hicieran más feliz. En un principio, Aristóteles estaría de acuerdo con ella, puesto que el bien del ser humano es la felicidad. Cumplir esta propuesta parece inicialmente sencillo. Necesitaríamos algún tipo de medición de la felicidad del usuario, que bien pueden ser las estrellas que asigne a un

contenido, o la frecuencia con la que ve un cierto tipo de contenidos, entendiendo que el usuario ve con más frecuencia aquellos contenidos que más feliz le hacen. La inteligencia artificial es muy buena cumpliendo objetivos. Por ello, si le pedimos que recomiende contenidos para que cada vez el usuario ponga más estrellas o esté más tiempo en la plataforma, seguro que lo consigue. Aristotélico total. Pero, ¿estar conectado es estar feliz?, ¿qué pasa con la moderación y la prudencia?

Aristóteles nos habla de una felicidad un tanto especial, de una felicidad verdadera (*eudaimonía*) a través de la cual el ser humano realiza su función, que es el ejercicio de la razón de forma virtuosa. Esta virtud se alcanza mediante el hábito de elegir el punto medio y como lo haría una persona prudente. Si de verdad queremos ser aristotélicos, tenemos que considerar la cuestión de la virtud, a través de la moderación y la prudencia. ¿Habrá pensado Ariadna en este punto de incluir la moderación y la prudencia?

De hacerlo, de querer ser aristotélicos, en ocasiones habría que ofrecerle al usuario contenidos quizás no del todo de su gusto, para que pudiera tener un punto medio. ¿Se atrevería AlegrIA a ofrecer contenidos que sabe que no le van a gustar al usuario?

La prudencia es incluso más complicada de cumplir. AlegrIA tendría que ofrecer información detallada sobre los pros y contras de un contenido. Cosa difícil, no desde un punto de vista técnico, sino porque, ¿quién informa hoy en día de los posibles perjuicios o fallos de un contenido? ¿Estaría dispuesta AlegrIA a decir algo crítico sobre un contenido? ¿Qué le diría a la productora de tal contenido: «es que somos aristotélicos»?

En principio el comentario de Ariadna parece aristotélico, y es de gran actualidad, pues hoy se alaba la felicidad. Pero es una felicidad en forma de *likes*, seguidores, o estrellas de valoración; sin moderación ni prudencia. Porque incorporar esto es más complicado de programar y molesta al usuario. Si de verdad queremos una ética de Aristóteles, tenemos que ejercitar la virtud, mediante la moderación y la prudencia. Y por una vida entera, que una golondrina no hace verano, ni ver un documental de La 2 un día no te hace intelectual.

Pero supongamos que Ariadna es completamente aristotélica y defiende una recomendación de contenidos que lleven al usuario a esa felicidad

verdadera (*eudoimonía*). Supongamos que programan su sistema con los elementos de moderación y prudencia. El siguiente punto sería: ¿qué virtudes busco en los usuarios? En un principio tendría virtudes aristotélicas, tales como el valor, la templanza, la generosidad y la grandeza de alma. ¿Son estas las únicas virtudes? ¿Y si el usuario quiere otras? Además, Aristóteles solo pensó en virtudes para ciudadanos libres varones, pues las mujeres estaban al gobierno de sus maridos. ¿Propondrá AlegrIA contenidos para esta visión? Y, ¡ojo con los feos!, que, según Aristóteles, lo tienen más difícil para ser felices; quizás no merezca la pena ofrecerles contenidos. Ser aristotélico ya no está tan bien.

Una última cuestión. Aristóteles dice que el fin del ser humano es la felicidad, pero esto no lo demuestra, lo da por hecho y por comúnmente aceptado. ¿Por qué la felicidad? Porque es lo normal, lo natural, lo que dice el sentido común, se podría contestar. Algo parecido argumentaba Tomás en la discusión del Comité de Dirección de AlegrIA. Vamos a ver este tema de *lo natural*, el sentido común.

Santo Tomás: lo natural

—Bueno, ¡vamos a ver! Esto tampoco puede ser lo que quiera cada uno o lo que quiera la mayoría. ¡Así nos va! —saltó Tomás, con su gesto de complacencia habitual siempre que hacía una crítica a la sociedad actual—. De alguna manera todos sabemos que lo que está bien o no. Apliquemos el sentido común, ¡por favor!

Con la llegada del cristianismo entran en el ámbito de la ética elementos tomados de la Biblia y de la revelación de Jesucristo. La preocupación de los grandes teólogos cristianos consiste en razonar los principios morales que establece el cristianismo, para, de esta forma, crear una doctrina coherente y sólida que salga del ámbito de la fe para entrar en el orden de la razón.

Uno de tales teólogos es santo Tomás de Aquino, autor de la *Suma Teológica*[78]. Habitualmente aparece representado con un libro, o escribiendo su doctrina, e iluminado por la gracia divina. En su pecho brilla un sol que representa su sabiduría.

Santo Tomás parte de la ética aristotélica y la adapta al cristianismo. Coincide con Aristóteles en que el fin del hombre es la felicidad, pero ahora esa felicidad verdadera no es la *eudaimonía*, sino lo que llama la bienaventuranza, que es la contemplación de Dios en su esencia. También coincide con Aristóteles en que esta felicidad solo se consigue a través de la virtud. Sin embargo, santo Tomás difiere de Aristóteles respecto al origen de las virtudes y en aportar nuevas virtudes.

Para Aristóteles la virtud es un logro de nuestra razón. Es una capacidad que tenemos en nuestra naturaleza como seres humanos. Para santo Tomás esa capacidad de ser virtuoso ha sido infundida en nosotros por Dios sin intervención nuestra, a través de la ley natural y la ley divina. ¡Cuántos conceptos en un párrafo tan corto! Virtud, ley natural y ley divina. Vamos a verlos.

La bienaventuranza se consigue mediante una vida correcta. Necesitamos entonces saber qué es correcto, es decir, qué está bien y qué está mal. Esto lo sabemos gracias a la ley natural que Dios ha inscrito en nuestro ser. La ley natural es una ley común a todo ser humano, que nos inclina de forma natural hacia el bien y nos permite discernir lo bueno de lo malo a través de la razón. Gracias a esta ley natural podemos intuir qué está mal o qué está bien, y por ello podemos ser virtuosos. Por eso, cuando algo está mal, es que está mal, no hay vuelta de hoja. ¡Es natural!

Ahora bien, ¿por qué tenemos la ley natural que tenemos? Es decir, según esta ley natural, entendemos que está mal matar o robar. ¡Es natural que esté mal! ¿Por qué? Porque esta ley natural forma parte de la ley divina. La ley divina es la intención de Dios, es lo que Dios quiere. Si nuestro fin es la bienaventuranza, que es la contemplación de Dios en su esencia, debemos tener la orientación de hacia dónde mirar. Esa orientación es la ley divina, que no podemos conocer, pero sí podemos intuir a través de la ley natural que tenemos inscrita.

Esta existencia de la ley natural y su correspondencia con la ley divina es lógica. No tendría sentido que la ley divina y la ley natural fueran contrarias. No sería razonable que, para alcanzar esa bienaventuranza, Dios nos pidiera algo (ley divina) que no pudiéramos realizar (que no estuviera en nuestra inclinación natural). Por ello la ley divina, lo que Dios nos pide que

hagamos, es algo que podemos hacer de forma natural gracias a esa ley natural con la que nos ha dotado. No tendría sentido que Dios me pidiera no matar, pero yo no entendiera por qué. Ahora lo divino, lo natural y lo racional: lo que Dios me pide, es algo natural en mí y lo entiendo. ¡Natural!

Luego, efectivamente, para Aristóteles podemos ser virtuosos si ejercitamos la razón, como lo haría una persona prudente, y podemos discernir por nosotros mismos qué está bien o qué está mal. Para santo Tomás, esto lo sabemos, no por nosotros mismos, sino por la ley natural que nos ha infundido Dios.

También dije que santo Tomás incorporaba nuevas virtudes respecto a Aristóteles, las llamadas virtudes teologales: fe, esperanza y caridad. Solo mediante estas virtudes, también infusas en nosotros, podemos llegar a alcanzar esa bienaventuranza de contemplar a Dios en su esencia. Las virtudes aristotélicas están muy bien, y debemos seguirlas, pero con ellas tan solo podemos alcanzar un simulacro de bienaventuranza, ser algo felices y sentir un poquito a Dios. Las buenas de verdad, las virtudes *Premium*, son las virtudes teologales, porque solo ellas me llevan a la bienaventuranza plena.

Programemos lo más natural en la inteligencia artificial

Santo Tomás no difiere tanto de la ética aristotélica, salvo que incluye a Dios, como fin del ser humano, la ley natural, inspirada en nosotros por la ley Divina, y las virtudes teologales, que nos permiten alcanzar esa bienaventuranza. Por tanto, las objeciones que pusimos para aplicar la ética de Aristóteles a la inteligencia artificial siguen siendo válidas para el caso de santo Tomás. Ahora bien, estos nuevos elementos tienen su importancia.

Tomás —ahora me refiero al de la *startup* AlegrIA— defiende una visión natural de lo que está bien y está mal. Posiblemente él no lo sepa, pero este pensamiento está inspirado en esa idea de una ley natural inscrita en nosotros. Debido a esta ley natural, las virtudes son algo razonable. Es lo natural. Si queremos aplicar la ética de santo Tomás al sistema inteligente de recomendación de contenidos de AlegrIA, no tendríamos que dar muchas explicaciones de por qué el sistema recomienda algo. Es lo natural. Sería natural que se recomendaran contenidos para favorecer la fe, la esperanza o la

caridad. Pero esto, que puede parecer natural, si somos tomitas, puede que no lo sea para otro y que este otro diga, con el mismo aplomo, que lo natural es otra cosa.

En el fondo, Aristóteles también defiende esta visión natural, pero de otra forma. Él dice que lo natural es que el ser humano busque la felicidad y a ese fin natural lo llama el bien. ¡Buen salto! ¿Por qué si la felicidad es nuestro fin natural, la felicidad es lo que está bien? ¿Por qué lo que supuestamente puede ser lo natural en nosotros es en definitiva lo bueno? Este pensamiento lo tenemos muy extendido actualmente, en el que se defiende que todo lo que nos pasa, si es natural, es entonces bueno. ¡Deja fluir tus emociones, que es lo natural, y eso es bueno!

Las emociones de cada uno, justo lo que decía David Hume, nuestro próximo invitado a esta reunión de filósofos. Porque tras la llegada del Renacimiento pasamos a pensar en éticas que consideren la individualidad y la conciencia de cada uno.