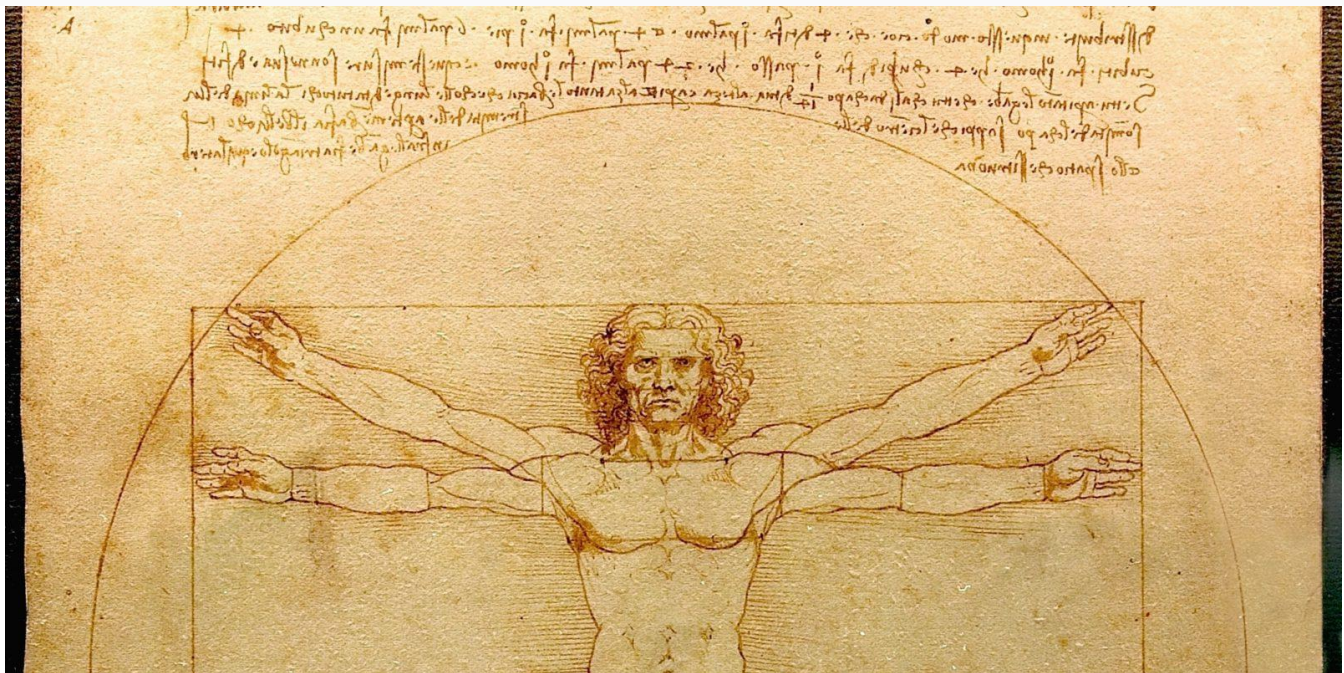


¿MÁQUINAS CONSCIENTES?

REFLEXIONES SOBRE LA DENOMINADA «INTELIGENCIA ARTIFICIAL»

Ferenc Patsch



Hombre de Vitruvio, Leonardo da Vinci (1492)

¿Qué nos hace humanos? A esta pregunta, muchos – incluidos personajes famosos como el historiador israelí Yuval Noah Harari^[1] y el fundador de OpenAI, Sam Altman^[2] – responden: «el humor», «la imaginación», «las emociones». Nosotros sostenemos que estas respuestas son filosófica y teológicamente ingenuas y, en última instancia, inadecuadas: la característica definitoria de la condición humana se encuentra más bien en la consciencia, que también suele considerarse el «santo grial» de la investigación en inteligencia artificial (IA).

En este artículo queremos demostrar que la llamada «inteligencia artificial generativa» no puede, en principio, ser tan inteligente como nosotros, es decir, que las máquinas nunca serán «conscientes de sí mismas». Por supuesto, estas afirmaciones requieren una explicación más matizada, para evitar las trampas del pensamiento anticientífico y de la «arrogancia clerical»: dos trampas en las que los teólogos han caído ocasionalmente desde el comienzo de la era moderna.

Introduzcamos el tema presentando una breve anécdota. Dos monjes están en la orilla de un arroyo y contemplan las ondulaciones del agua. De repente, uno de ellos dice pensativo: «¡Ojalá pudiera ser tan feliz como ese pez en el agua!». El otro responde: «¿Cómo puedes estar seguro de que ese pez es feliz? Tú no eres un pez». El primero responde: «¿Cómo sabes que no puedo saberlo? Tú no eres yo».

Este cuento expresa con acierto uno de los problemas más debatidos en la investigación de la IA hoy en día: ¿cómo sabemos que las máquinas actuales son realmente «inteligentes»? ¿Puede un ordenador ser más listo o más creativo que su creador, el hombre? ¿Y quién está cualificado para decidir tales cuestiones?

La posibilidad de crear robots inteligentes no es, desde luego, una cuestión nueva. Al contrario, ha sido una viva preocupación de la imaginación de novelistas y cineastas que han trabajado en el género de la ciencia ficción durante las últimas décadas. Aquí queremos referirnos a tres películas que se han convertido ya en clásicos.

1) *2001: Odisea del espacio* es una película estadounidense de ciencia ficción de 1968 que tuvo gran éxito. Fue tan famosa que su impacto dejó una huella duradera en la cultura del séptimo arte debido a sus originales soluciones cinematográficas. El guion también llamó la atención del público: por ejemplo, uno de los personajes principales de esta película era un superordenador infalible llamado HAL, con personalidad propia, que de repente se rebela contra su creador, toma el control de la nave espacial (pues descifra, leyendo los labios, una conspiración humana contra él) y sólo después de muchas adversidades y emociones es finalmente desarmado.

2) Toda una generación de mediana edad recordará el primer episodio de *Star Wars: Una nueva esperanza* («A New Hope»), estrenado en 1977. Uno de los principales «personajes» de esta película era C-3PO, un robot humanoide algo torpe que expresaba emociones y se hacía amigo de su compañero robot – a veces incluso «peleaba» con él – mientras charlaba y balbuceaba simpáticamente en pantalla. Su compañero inseparable era R2D2. Esta película es una señal para los diseñadores que trabajan hoy en el campo de la robótica social, una demostración de que una máquina no tiene por qué parecerse necesariamente a un ser humano (*humanlike*) para ser querida y provocar empatía: basta con que sea simplemente «benévola» y «servicial», y por supuesto no está de más que posea habilidades especiales útiles para los humanos. Como estos robots poseían sin duda tales cualidades, millones de personas los admiraron, y los convirtieron en figuras icónicas para toda una generación.

3) Recordemos también *Terminator 2* («Judgement Day» / «El juicio final»), un clásico del cine que disfrutó una generación posterior. Una de las escenas de esta película de 1991 es la de Arnold Schwarzenegger entrando desnudo como «Terminator» en un bar, en busca de un motociclista hostil, al que primero pide amablemente que le entregue su ropa y su moto, y luego, tras encontrar resistencia, se las arrebató por la fuerza. En este punto, la idea distópica de que los humanos deben luchar a muerte con robots «inteligentes» por su supervivencia alcanza su clímax.

La cuestión de los «ordenadores inteligentes» parece haber cobrado más actualidad que nunca. Algunos dicen que la visión de ciencia ficción de hace unas décadas es ya una realidad, mientras que otros afirman que se hará realidad en un futuro próximo^[3]. Hoy en día, toda una industria – la llamada «industria de la IA» – se basa en la idea de ordenadores cada vez más inteligentes y, como oímos regularmente en los medios de comunicación, expertos de OpenAI, DeepMind y Anthropic trabajan incansablemente para desarrollar una «*Inteligencia Artificial General*», es decir, una IA fuerte o de nivel humano, que quizás represente una especie de singularidad, un nuevo comienzo en la investigación.

Lo que está en juego no es nada trivial: incluso los observadores más moderados advierten de que la IA tiene una importancia estratégica y geopolítica inconmensurable: quien gane la «carrera armamentística» de la IA podrá controlar el futuro del mundo^[4]. Pero, ¿está realmente en juego la supervivencia de la humanidad, como sugieren los medios sensacionalistas? Y si es así, ¿en qué sentido? Para aclarar estas preguntas, primero debemos hacer un breve análisis de la cuestión.

¿Qué es la consciencia artificial?

Según la definición del sitio web oficial de OpenAI, la «Inteligencia Artificial General» (IAG) es un «sistema altamente autónomo que supera a los humanos en la mayoría de los trabajos de valor económico»^[5]. La gran ventaja de esta definición es que es breve y concreta (y, por tanto, relativamente fácil de verificar); sin embargo, su inconveniente es que es simplista y, en última instancia, filosóficamente inadecuada. Si el superordenador Deep Blue de IBM venció al campeón mundial de ajedrez Garry Kasparov en 1997, ¿significa eso que era más inteligente que él? O, cuando en 2016 AlphaGo triunfó sobre Lee Sedol, el campeón mundial del juego de mesa go – infligiendo así a la psique asiática una herida difícilmente imaginable en Europa –, ¿podría verse esto también como una solemne declaración de que el ordenador es más inteligente que el principal jugador de go del mundo? Y, por último, si el ChatGPT es más hábil planificando un viaje de negocios que una secretaria de empresa, entonces, sólo porque la máquina haga mejor un determinado trabajo de valor económico, ¿puede considerarse realmente más inteligente que los humanos? O, como afirman algunos, ¿son también conscientes de sí mismas? La respuesta es, por supuesto, una cuestión de interpretación: depende de lo que entendamos por «inteligencia» y «consciencia» (o «autoconsciencia»). Como nunca se ha definido el significado de estos términos, debemos aventurarnos más allá del ámbito de las ciencias, en lo que es propiamente el reino de la filosofía y la teología.

Un intento de aclarar el concepto

Desde luego, la cuestión de la inteligencia informática no es nueva. Los llamados «filósofos de la mente» llevan décadas trabajando en una definición más precisa de la inteligencia. En general, sus investigaciones les han llevado a distinguir tres componentes o aspectos principales de la misma: la consciencia, la percepción y la autopercepción.

1) La cuestión de la «consciencia» ha sido abordada en profundidad, entre otros, por el filósofo estadounidense Thomas Nagel. En su ensayo de 1974 titulado *What is it like to be a bat?*, sostenía que sólo los que son conscientes pueden «sentir». Según su famoso argumento, la pregunta «¿qué se siente al ser un murciélago?» sólo puede hacerse sobre un murciélago, mientras que no tiene sentido sobre una tostadora. La diferencia es que un murciélago tiene obviamente algún tipo de consciencia (específica de la vida animal), mientras que no puede decirse lo mismo de una lavadora o una tostadora^[6]. Por tanto, la consciencia debe concebirse como una especie de continuo. La consciencia humana es superior a la del murciélago en la medida en que también es consciente de sí misma; los delfines y los monos representan una especie de estado intermedio entre el hombre y el murciélago, porque se reconocen en el espejo, lo que sugiere que tienen un nivel de consciencia de sí mismos, evidentemente inferior al del hombre, pero superior al de los demás animales^[7].

2) Otro componente de la inteligencia humana es el de «ser sensible» (*sentience*), es decir, la capacidad que tenemos de ser conscientes de nuestras percepciones y emociones. También ésta es una capacidad exclusivamente humana en comparación con la de los animales: no hay pruebas detectables de que otras especies sean conscientes de lo que perciben o sienten. Por supuesto, en un sentido similar, un ordenador también puede percibir, ya que es posible crear un algoritmo que permita a un robot, por ejemplo, detectar objetos amarillos en su campo de visión, y luego percibir esta propiedad y reaccionar ante ella. Del mismo modo, se puede enseñar a los robots a percibir y reconocer comportamientos hostiles y, en caso necesario, a responder a un ataque con agresividad. Sin embargo, esto no significa que la máquina entienda lo que es el amarillo (color) o que sea consciente de la agresión como emoción.

3) Por último, para hablar de inteligencia también se requiere «consciencia de sí» («*self-awareness*»). El término inglés «*self-awareness*» no es lo mismo que «*consciousness*», porque no se trata sólo de una intuición cognitiva, sino también de una especie de «consciencia del propio cuerpo» (*body awareness*). Esta última también puede observarse en la vida animal, a distintos niveles. En el caso de los humanos, «consciencia de sí» significa ser consciente de uno mismo como individuo separado de todos los demás, con sus propios pensamientos, decisiones e ideas conscientes. Es algo más que la capacidad de representarse a sí mismo, que también se observa en mamíferos superiores (como han demostrado, por ejemplo, los experimentos con Koko, una gorila a la que se enseñó el lenguaje de signos). Hay que señalar que incluso los sistemas operativos o depuradores de los ordenadores modernos pueden representarse a sí mismos, pero esto no es todavía consciencia, autodeterminación o autopercepción.

La consciencia en sentido estricto – como consciencia, percepción y autopercepción – es una forma de ser exclusiva de los seres humanos, al menos por ahora. Esta situación sólo podría cambiar si la llamada «inteligencia general artificial»^[8] se realizara no sólo hipotéticamente, sino también en la realidad, es decir, si se creara al menos «un robot inteligente» capaz de realizar cualquier tarea mental que el ser humano sea capaz de realizar, es decir, que también tuviera consciencia, percepción y autopercepción en el sentido mencionado. Muchos futurólogos creen que los grandes modelos lingüísticos que existen hoy en día, como ChatGPT, son precursores o «chispas» de la aparición de una inteligencia artificial general de este tipo^[9]. Nosotros, en cambio, coincidimos con la mayoría de los estudiosos en que, en principio, no será posible alcanzar ese objetivo nunca.

«¿Un mundo feliz?»

Es imposible negar que la investigación sobre IA ya ha producido resultados extraordinarios. Sin ánimo de ser exhaustivos, mencionemos algunos: conducción autónoma, robots navegantes, pilotaje de aviones de combate; lectura, escritura y habla en muchos idiomas; reconocimiento de caracteres ópticos, traducción del habla a la escritura, traducción a una lengua extranjera; programación de ordenadores; creación de obras de arte (desde la toma de fotografías a la creación de vídeos, incluso a partir de instrucciones dadas como descripciones verbales)^[10]; creación e interpretación de pistas musicales; limpieza de material de audio; redacción y supervisión de acuerdos; creación de planes empresariales y financieros; realización de funciones de servicio al consumidor mediante *chatbots* y recomendación de productos; predicción meteorológica; predicción bursátil; biometría (reconocimiento facial, de huellas dactilares y del iris); lectura de labios; reconocimiento de escritura; análisis de estados emocionales (incluso a partir de texto o expresiones faciales); detección de fraudes; escritura manual, tutoría, redacción y evaluación de ensayos académicos; detección de plagios; formulación de diagnósticos médicos; descubrimiento de fármacos; predicción del tráfico; participación en actividades recreativas (ajedrez y go); descubrimiento de nuevos juegos; realización de tareas de asistencia virtual; mantenimiento de conversaciones significativas, etcétera^[11].

La lista que hemos dado ya es impresionante. Y lo es aún más si tenemos en cuenta las posibilidades adicionales de desarrollo de la inteligencia artificial que pueden aparecer pronto a medida que mejoren las funciones actuales: por ejemplo, en un futuro próximo podría proporcionar una asistencia sanitaria rápida y de alta calidad a la población (desde el diagnóstico hasta el cuidado de ancianos), ayudar a resolver algunos de los problemas más acuciantes a los que se enfrenta la humanidad hoy en día (desde descubrir una cura para el cáncer hasta mitigar los efectos del cambio climático) y ofrecer una educación pública personalizada y de alta calidad (democratizando lo mejor de la pedagogía moderna).

Sin embargo, para evitar un optimismo excesivo, hay que recordar que lo contrario también puede ser cierto. Los sistemas de armas inteligentes pueden destruir a millones de personas, los mensajes *deepfake* (por ejemplo, vídeos falsos producidos en masa), la vigilancia biométrica y la policía predictiva pueden desestabilizar fuertemente las democracias existentes y contribuir a la aparición de un «estado de vigilancia» de control social sin precedentes en la historia^[12]. Por supuesto, no es posible predecir con exactitud qué dirección tomará la evolución futura. Una cosa es cierta: dentro de unas décadas, la humanidad vivirá en un mundo muy distinto del nuestro, hasta un punto que superará nuestra imaginación.

La interacción humana con las nuevas tecnologías plantea ya una serie de cuestiones sociales, jurídicas y éticas. La razón de fondo es clara: la inteligencia artificial es peligrosa porque trata con nosotros, los humanos. Y los humanos somos capaces de abusar de cualquier cosa, incluidos algoritmos y robots. Nuestra especie, como advierten los psicólogos sociales, no sólo es capaz de cooperar, sino también de competir. Como seres humanos, a menudo tomamos decisiones contradictorias e irracionales, guiados por nuestro propio interés, y somos capaces de explotarnos e incluso de matarnos unos a otros. La tradición de la teología católica describe esta condición mostrando que, aunque Dios creó al hombre como un ser bueno (*Gn 1,31*), tras la caída (*cfr. Gn 3,1-24*) su naturaleza se corrompió y su voluntad se inclinó al mal.

Detrás de la agresividad y el comportamiento beligerante suelen esconderse ansiedades, miedos y, en última instancia, mecanismos de autodefensa^[13]. Sólo por esta razón, sería utópico e ilusorio pensar que el mal – el pecado – producido por el hombre puede ser eliminado del mundo. Y los conflictos bélicos seguirán existiendo entre nosotros. Por supuesto, el pecado desempeña un papel en la historia de la salvación – se dice que «la guerra dirige la historia» –, pero sería un error fatal confiar las grandes cuestiones de la vida, como las decisiones estratégico-militares, a máquinas inteligentes. La humanidad en su conjunto es responsable del futuro de nuestra especie. Por tanto, es conveniente que un gobierno mundial democrático se encargue de regular la inteligencia artificial y su papel en nuestras vidas^[14]. Los inventos tecnológicos del siglo XXI no pueden funcionar sobre la base de leyes consagradas en el siglo XX^[15].

¿Robots inteligentes?

Pero, ¿hasta qué punto son inteligentes los robots actuales? En junio de 2022, el ingeniero de investigación de Google Blake Lemoine, en una prueba ética del modelo lingüístico a gran escala LaMDA (*The Language Model for Dialogue Applications*), afirmó que el algoritmo había alcanzado la autoconsciencia, se había vuelto sensible y revelaba emociones reales. Lemoine fue despedido de la empresa en julio de 2022 «por violar las normas de confidencialidad de la empresa»^[16].

Las opiniones de este investigador han sido rechazadas por la mayoría de los profesionales de la inteligencia artificial. Por ejemplo, Robert J. Marks, ingeniero informático de la Universidad de Baylor, no sólo presentó un argumento convincente en su debate con Lemoine^[17], sino que también argumentó convincentemente que la inteligencia informática no es comparable a la mente humana. En su libro *Non-Computable You*^[18], expone las razones de su convicción filosófica – e implícitamente teológica – de que en nuestro mundo existen realidades no algorítmicas y no computables. Implican funciones cognitivas, como la creatividad, la comprensión y la percepción, así como emociones complejas, como el amor, la compasión y la empatía; si se entienden correctamente, cada una de estas funciones mentales es un fenómeno exclusivamente humano.

Por cierto, parece que incluso los representantes de las ciencias cognitivas y tecnológicas que se ocupan de la inteligencia artificial se inclinan cada vez más a reconocer los límites de su propia disciplina. La autocomprensión de los científicos, como la de los verdaderos artistas^[19], suele ser más humilde que la que el público lego tiende a atribuirles: son conscientes de que no son omnipotentes.

¿«Chovinismo del carbono»? ¿«Clericalismo oscurantista»?

Hoy resulta riesgoso afirmar que la inteligencia artificial nunca será inteligente en el mismo sentido que nosotros, los humanos. Porque tal afirmación nos expone inmediatamente a la acusación de ser «tecno-pesimistas» o «chovinistas del carbono». Se trataría, en palabras del cosmólogo del MIT Max Tegmark, de una preferencia injusta por la especie humana frente a la inteligencia basada en el silicio de las máquinas^[20]. Los peligros aumentan cuando alguien

aparece en público, con el título de cristiano, para pedir cautela contra el excesivo «tecno-optimismo». Esa persona es fácilmente tachada de *remainer*, de conservador oscurantista, de enemigo del progreso y de representante de la reacción clerical. Es necesario razonar cuidadosamente para evitar tales acusaciones.

Ya hemos visto que el concepto de inteligencia – que no tiene un significado «eterno» – es en realidad un continuo: hay realizaciones inferiores y superiores. Por consiguiente, es lícito – aunque algo inexacto – llamar «inteligentes» a sus manifestaciones inferiores. En este sentido análogo, se puede hablar de «inteligencia artificial». Al mismo tiempo, sin embargo, cabe señalar que este desarrollo, a pesar de los brillantes resultados obtenidos hasta ahora y de los que se obtendrán en el futuro, es, en principio, limitado y nunca alcanzará el nivel de la inteligencia humana. Los robots no serán conscientes como nosotros, no podrán percibir como nosotros y no tendrán consciencia de sí mismos.

Según Federico Faggin, el «padre» italiano del transistor moderno, «Chat GPT no entiende lo que dice, mientras que un ser consciente sí», porque «la comprensión no es un fenómeno algorítmico»^[21]. Uno de los primeros exponentes influyentes de esta visión escéptica fue Roger Penrose, quien, en su libro de 1989, *The Emperor's New Mind*^[22], describió la IA fuerte como un emperador desnudo que intenta ser considerado inteligente, pero cuya consciencia no se parece en nada a la nuestra, porque la consciencia humana no puede proceder de un ordenador de silicio.

¿Significa todo esto «chovinismo del carbono»? Podemos considerarlo más bien como una apertura hacia un enfoque más holístico. En efecto, como bien señalan John Searle y otros filósofos, la consciencia no puede reducirse a un mero «procesamiento de información». Según Altman, puesto que la inteligencia artificial es capaz de razonar mejor que los humanos, hablar de forma más inteligente, calcular más rápido y superar a los humanos en muchos otros ámbitos, también es más inteligente que ellos. Si definimos la inteligencia como operaciones de cálculo por segundo, entonces Altman tiene indudablemente razón. Pero, ¿quién ha demostrado que la inteligencia humana es completamente atribuible al número de operaciones que nuestro cerebro realiza cada segundo? Esta afirmación gratuita es más bien una característica del pensamiento reduccionista, que resultó útil para el desarrollo de las ciencias naturales en los primeros tiempos de la revolución tecnológica, pero es insuficiente para describir la totalidad del mundo en su conjunto.

Algunas reflexiones sobre la inteligencia artificial

En este contexto, podemos referirnos a uno de los filósofos más influyentes del siglo XX, Martin Heidegger (1889-1976). En sus reflexiones sobre la esencia de la técnica, que se inscriben en el contexto de su amplia crítica de la «historia del ser», ya intuía, a mediados del siglo pasado, el cambio de nuestra cultura, que se manifiesta en la construcción del mundo de la técnica como *Gestell*^[23]. Distinguió acertadamente entre el «pensamiento meditador» (*das besinnliche Denken*) y el «pensamiento calculador» (*das rechnende Denken*)^[24]. Este último es un paradigma cultural que, en última instancia, pretende poseer, dominar por la fuerza y transformar la realidad para controlarla y, como tal, es deshumanizador. Este paradigma «tecnocrático» – cuya crítica más significativa se encuentra hoy en la encíclica *Laudato si'* (LS) del Papa Francisco (2015)^[25] – se está extendiendo rápidamente en la actual fase tecnocrático-tecnológica del desarrollo de nuestra civilización planetaria, y existe el temor justificado de que sus

efectos deshumanizadores y distorsionadores se potencien aún más con el uso de ChatGPT y otras inteligencias artificiales disponibles.

Llegados a este punto, me gustaría añadir una reflexión personal. Hacía tiempo que me ocupaba del tema de la llamada «inteligencia artificial», cuando me di cuenta de que mi reticencia a comparar la inteligencia de las máquinas con la inteligencia humana estaba motivada en realidad por razones teológicas: no puedo reconciliarme con la idea de que la inteligencia de los robots pueda un día igualar, y con el tiempo superar, las capacidades de los seres humanos en todos los aspectos, porque tampoco creo – basándome en argumentos filosóficos y teológicos bien establecidos – que los seres humanos puedan llegar a ser Dios. Creo que los superordenadores nunca podrán tener las mismas experiencias que nosotros, los seres humanos, porque no tienen cuerpos humanos, no tienen un «entorno», no tienen otros seres humanos a su alrededor.

Conclusión

- Para concluir, es importante subrayar una vez más que no se trata de devaluar los logros de la tecnología hasta la fecha ni de poner obstáculos a futuros avances. Las conclusiones de la investigación empírica hablan por sí solas. Por ejemplo, investigadores en psicología y estadística de la Universidad de California han demostrado que los grandes modelos lingüísticos ya superan a los humanos en muchos ámbitos de la vida (e incluso Chat-GPT3 fue capaz de superar a los humanos en algunas tareas creativas)^[26]. El aprendizaje automático es capaz de hacer «milagros». Hay que tener en cuenta que un niño también crece como persona a través de la interacción constante con otras personas, especialmente con sus padres. Por tanto, no se puede descartar que cuando conversamos con los *chatbots*, estemos, por así decirlo, «socializándolos»: les estamos enseñando lo que significa ser una persona. No se puede descartar que en esta interacción ocurra algo en ellos y que con el tiempo adquieran (o incluso tengan) cierta consciencia. Sin embargo, una cosa es segura: esta consciencia será de una naturaleza radicalmente distinta a la nuestra, y no alcanzará el nivel de autoconsciencia.

La hipótesis de Elon Musk es que el *TruthGPT* – cuyo propio nombre es una utopía – que está desarrollando acabará interesándose por el universo, y posiblemente por nosotros, y nos estudiará como los humanos estudiamos espontáneamente a los simios^[27]. Este tipo de especulación puede calificarse de exageración, y carece de base empírica.

Las especulaciones de los escritores de ciencia ficción y los gurús de la alta tecnología sobre robots autoconscientes son en realidad ideas antropomórficas: no son más que la proyección de nuestra inteligencia en algoritmos que hemos producido. En conclusión, es posible que los robots puedan adquirir algún tipo de consciencia o inteligencia hablando con nosotros, pero como no tienen ni cuerpo ni relaciones sociales – y por tanto no pueden experimentar su «mundo-entorno» (*Umwelt*) y su «mundo-con» (*Mitwelt*), como nosotros –, su «consciencia» o «inteligencia» desarrollada nunca será como la nuestra, es decir, no podrá convertirse en autoconsciencia.

1. Para Yuval Noah Harari, «los seres humanos son animales que son capaces de contar una historia, es decir, están dotados de imaginación». Cfr. www.youtube.com/watch?v=l1aNsAAImZU ↑
2. A la pregunta sobre qué nos hace humanos, Sam Altman respondió con una palabra: «las emociones». Cfr. «OpenAI CEO Sam Altman and CTO Mira Murati on the future of AI and ChatGPT», en www.youtube.com/watch?v=byYIC2cagLw ↑
3. Para Yann LeCun, «todavía estamos muy lejos de la IA de nivel humano», Cfr. <https://ai.meta.com/blog/yann-lecun-advances-in-ai-research> ↑
4. Cfr. A. Zegart, «Open Secrets: Ukraine and the Next Intelligence Revolution», en *Foreign Affairs*, enero-febrero 2023, 58. ↑
5. «OpenAI Charter», en <https://openai.com> ↑
6. Cfr. Th. Nagel, «What is it like to be a bat?», en *Philosophical Review* 83 (1974) 435-450. ↑
7. Cfr. J. M. Plotnik – F. B. M. de Waal – D. Reiss, «Self-recognition in an Asian elephant», en *Proceedings of the National Academy of Sciences*, vol. 45, n. 103, 2006, 17053-17057; sobre los aspectos filosóficos, cfr A. Turay, *Kozmológiai antropológia*, Szeged, A katolikus hittudományi főiskolák jegyzetei, 1987. ↑
8. La creación de la IGA es el principal objetivo de algunas investigaciones y empresas de IA, como OpenAI, DeepMind y Anthropic. ↑
9. Cfr. «Sparks of Artificial General Intelligence: Early experiments with GPT-4», en <https://arxiv.org/abs/2303.12712> ↑
10. ChatGPT, Bing Image Creator y Midjourney pueden incluso crear una imagen basada en cualquier descripción textual. ↑
11. Habríamos tenido que poner entre comillas todos los términos que citamos arriba, para indicar la diferencia entre ChatGPT, Bard, LaMDA y los demás modelos lingüísticos y la inteligencia humana. Es solo por comodidad o *marketing* que los primeros pueden definirse como «inteligentes». ↑
12. Cfr. S. Talty, «What Will Our Society Look Like When Artificial Intelligence Is Everywhere?», en www.smithsonianmag.com/innovation/artificial-intelligence-future-scenarios-180968403 ↑
13. Cfr. A. Kurpatov, «Сознание ИИ. Как искусственный интеллект решает задачи планеты | Чёрное зеркало с Андреем Курпатовым», en www.youtube.com/watch?v=9LqADeuWQcc ↑
14. Cfr. T. Harris, «AI: Grappling with a New Kind of Intelligence», en <https://www.youtube.com/watch?v=EGDG3hgPNp8> ↑
15. Cfr. F. Patsch, «La “Inteligencia Artificial Generativa” y nuestro futuro», en *La Civiltà Cattolica*, 17 de noviembre de 2023, <https://www.laciviltacattolica.es/2023/11/17/la-inteligencia-artificial-generativa-y-nuestro-futuro/> ↑
16. Cfr. «Blake Lemoine: Google fires engineer who said AI tech has feelings», en www.bbc.com/news/technology-62275326 ↑
17. Por ejemplo, cuando LaMDA, «chateando» con Blake Lemoine, respondió que lo que lo hace feliz es estar con su familia, Marks observó lo absurdo de la respuesta, afirmando que la IA estaba simplemente plagiando de internet, pues LaMDA no tiene una familia. Cfr. R. J. Marks, «Is ChatGPT conscious? The Agenda», en www.youtube.com/watch?v=-j_oHZ3SvIk&t=16s ↑
18. Cfr. R. J. Marks, *Non-Computable You: What You Do That Artificial Intelligence Never Will*, Seattle (WA), Discovery Institute Press, 2022. ↑
19. Cfr. H.-G. Gadamer, *Verità e metodo*, Milán, Bompiani, 2001, 117. ↑
20. Cfr. M. Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*, New York, Knopf, 2017. ↑
21. F. Bastiani, «Faggin: “L’AI non avrà mai coscienza. ChatGPT? A differenza nostra non comprende ciò che dice”», en <https://startupitalia.eu/tech/faggin-lai-non-avra-mai-coscienza-chatgpt-a-differenza-nostra-non-comprende-cio-che-dice/> ↑

22. Cfr. R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics*, Oxford, Oxford University Press, 1989. ↑
23. Cfr. M. Heidegger, *L'abbandono*, Génova, Il Nuovo Melangolo, 1983, 27-43. ↑
24. *Ibid.*, 30. ↑
25. Cfr. en particular LS 106-114. En cuanto a las raíces «heideggerianas» de la encíclica, cfr F. Patsch, «Heideggerian Foundations in Pope Francis?», en J. Azetsop – P. Conversi (edd.), *Foundations of Integral Ecology*, Roma, Gregorian & Biblical Press, 2022, 343-384. El 8 de enero de 2024, en un pasaje importante del tradicional discurso inaugural del inicio del año dirigido al Cuerpo diplomático de la Santa Sede, el Papa Francisco volvió sobre el tema del Mensaje para la Jornada Mundial de la Paz, dedicado enteramente a la «inteligencia artificial», que tiene por «uno de los retos más importantes de los próximos años». ↑
26. Cfr. T. Webb – K. J. Holyoak – H. Lu, «Emergent analogical reasoning in large language models», en *Nature Human Behaviour*, vol. 7, 2023, 1526-1541. ↑
27. Cfr., «TruthGPT: Elon Musk usa l'AI per comprendere la vera natura dell'Universo», en *Cryptonomist* (<https://cryptonomist.ch/2023/04/20/truthgpt-scoperta-ai-elon-musk>), 20 de abril de 2023. ↑