

Título: El rol del ser humano dentro de la ciencia en la edad de la inteligencia artificial.

Title: The role of human beings in science in the age of artificial intelligence.

Autor: Eduardo C. Garrido Merchán

Institución: Universidad Pontificia de Comillas.

Resumen

El avance de la computación ha hecho posible experimentar con modelos de inteligencia artificial de billones de parámetros, como Dalle-2 o GPT-3, con complejas arquitecturas y capacidad tan alta que representan lenguajes, videos, sistemas físicos o emular el discurso de cualquier persona si se tienen los suficientes datos de ella. Estos sistemas son capaces de solucionar algunos problemas con mayor eficacia que la de un ser humano, como por ejemplo el problema del plegamiento de proteínas, el comercio de alta frecuencia, la elaboración de nuevas teorías matemáticas y físicas, escribir códigos de programación para solucionar problemas genéricos o jugar mejor que otros humanos, exhibiendo nuevas estrategias, al Go o al ajedrez. En este contexto, es posible preguntarse sobre la hipótesis de que si la inteligencia artificial continúa su imparable avance, el ser humano se verá relegado a un papel secundario en la ciencia. En este artículo, se proporciona un argumento que se apoya en que el ser humano cuenta con información única que no está disponible para las máquinas, que desde un punto de vista fenomenológico percibimos como el qualia de los fenómenos. En concreto, se estudia esta percepción como una información que proporciona un valor diferencial para solucionar problemas científicas, filosóficos y artísticos irresolubles para las máquinas.

Abstract

The advancement of computing has made it possible to experiment with artificial intelligence models with billions of parameters in complex architectures with such a high capacity that they can represent languages, videos, physical systems or emulate the speech of any person if they have enough data about them. These systems are capable of solving problems with greater efficiency than a human being, such as the problem of protein folding, high frequency trading, elaborating new physical and mathematical theories, writing programming codes to solve generic problems or playing better than other humans exhibiting new strategies to games like Go or chess. In this context, it is possible to reason about the hypothesis that if artificial intelligence continues its unstoppable advance, the human being could be relegated to the background in science. In this article, a basic argument is provided on how human beings have unique information that is not available to machines, which from a phenomenological point of view we perceive from the qualia of the phenomena we perceive, and how this information provides a value differential to solve scientific, philosophical and artistic problems insoluble for machines.

Palabras clave: Inteligencia artificial, ciencia, fenomenología.

Keywords: Artificial intelligence, science, phenomenology.

Breve contenido de la propuesta

La psicología popular actual contempla a la inteligencia artificial [8] como un paradigma que será capaz de, en palabras de César Alierta, ser capaz de resolver cualquier problema. No solo el hardware ha permitido implementar modelos de capacidad cada vez más elevado, sino que ha posibilitado la búsqueda de arquitecturas más eficientes y capaces de generalizar para resolver problemas de dominio general. Si a esta situación le añadimos la exponencial evolución de los datos adquiridos o compartidos en internet, que se le conoce como el paradigma del Big Data [11], y la

potencial llegada del metaverso [14], los modelos basados en inteligencia artificial podrían razonar mediante el uso de esta información en formas que los seres humanos, dadas nuestras limitaciones biológicas, no podemos, especialmente cuando estuviéramos en realidades simuladas por un ordenador donde todos los problemas a resolver fueran estrictamente computacionales.

Algunos ejemplos de problemas que en la actualidad son resueltos exclusivamente por algoritmos y que los seres humanos no podemos solucionar con la misma eficiencia son el problema del plegamiento de proteínas, razonamiento automático de teoremas matemáticos [12] y generaciones de nuevas teorías físicas, rendimiento en juegos de estrategia como el Go o algunos videojuegos, codificación automática de todo tipo de problemas [13], regresión y clasificación de ingentes cantidades de datos estructurados y no estructurados como texto, vídeo y audio o high frequency trading. Ante este novedoso paradigma que comienza, la edad de la inteligencia artificial, el que el ser humano, dadas sus limitaciones biológicas y asumiendo que las tesis transhumanistas [7] que abogan por una posible carga de conciencia en un ordenador son falsas [5] o, en su escenario más optimista abismalmente complejas, según queda argumentado en numerosas fuentes, nos queda reflexionar de forma urgente sobre la hipótesis de que el rol que el ser humano en la ciencia continúe siendo significativo si un meta-modelo de inteligencia artificial general es capaz de elaborar un modelo general del universo que nos rodea o el metaverso en el que nos relacionaremos [14]. Este trabajo desarrolla una argumentación basada en nuestra naturaleza humanista que aporta argumentos teóricos a favor de dicha tesis. Considérese que el ser humano es consciente de sí mismo, cualidad de la que carecen algoritmos implementados en ordenadores, tradicionales o cuánticos, cuya lógica es emulable por una máquina de Turing universal.

Este artículo comienza describiendo el tipo de problemas resolubles por una potencial inteligencia artificial general, continúa desde un punto de vista fenomenológico exponiendo el carácter único de la consciencia fenomenológica del ser humano y la información única que a través del qualia y nuestra percepción consciente somos capaces de incorporar en nuestra toma de decisiones para plantear preguntas cuya solución no es computacional como la del sentido o la de la belleza. Por último, se cierra la presentación argumentando que, debido a este procesado de información no computacional el ser humano tiene un rol diferencial en la ciencia, la filosofía y el arte que es independiente de la complejidad de la inteligencia computacional que tenga la inteligencia artificial general de mayor capacidad que creemos.

Las redes neuronales profundas [9] y el meta-aprendizaje [15] son los principales modelos que han sufrido un crecimiento exponencial en los últimos años debido al auge de la computación y el hardware. En concreto, las redes neuronales artificiales masivas de billones de parámetros que discriminan que información es relevante para solucionar un problema mediante mecanismos de atención llamadas transformadores son capaces gracias a algoritmos de aprendizaje automático de elaborar modelos del lenguaje, visión y predicción de eventos futuros en un entorno codificando la información que les rodea en un modelo interno gracias a redes codificadoras y pudiendo emular discursos de cualquier persona, imágenes o acciones gracias a redes internas generativas desde su espacio de conceptos latente. Si un meta-modelo de transformadores [16] es capaz de activar distintos transformadores en base al problema encontrado en el mundo que nos rodea, entonces, se puede potencialmente culminar en una inteligencia artificial general [10] capaz de resolver cualquier problema computacional del subespacio de problemas resolubles por una máquina universal de Turing.

Es crítico en este punto enfatizar que, por complejo que sea el modelo de inteligencia artificial representado, el espacio de problemas que resuelve es el de los problemas computacionales resolubles por una máquina universal de Turing ya que estas redes están implementadas en

ordenadores tradicionales cuyos procesadores siguen arquitecturas de Von Neumann o en ordenadores cuánticos que resuelven los mismos problemas que los tradicionales con la única diferencia de que lo hacen exponencialmente más rápido. En cualquier caso, quedan excluidos de esta clase de problemas a resolver aquellos problemas existentes en el universo que no son computacionales.

Como seres humanos, cualquier medición aproximada, y como en el caso del coeficiente de Stern sesgada culturalmente, de nuestra inteligencia computacional nos situaría en una posición mucho más inferior a la de un sistema de inteligencia artificial como el descrito anteriormente. Sin embargo, gracias a poseer consciencia fenoménica, los seres humanos somos capaces de percibir el *qualia* del universo que nos rodea, una información subjetiva, privada y relativa a cada observador particular. Esta propiedad de nuestro ser, la existencia de sí, el estar ahí en términos del *Dasein* de Heidegger [4] es algo de lo que carecen las redes neuronales implementadas en hardware que emulan máquinas de Turing universales de acuerdo a las tesis planteadas por Roger Penrose [1,2], entre otros argumentos en contra de la inteligencia artificial fuerte, del funcionalismo y de la realizabilidad múltiple [6].

Estas máquinas se situarían en la habitación de Searle, encontrando correlaciones pero sin comprensión ya que la comprensión requiere de la percepción del *qualia* únicamente expresada en la consciencia fenoménica, se situarían en el problema del conocimiento de Mary al carecer de la percepción de la "rojez del rojo", incapaces de sentir que un amanecer muy bonito no es solo mejor que un amanecer bonito, según Bergson cuando se refiere a las intensidades de la percepción que nos rodea. La fenomenología [3] sería una ciencia inescrutable para la más avanzada de las inteligencias artificiales. Se puede emular artificialmente que alguien muy enamorado está más enamorado que alguien enamorado, pero en ningún momento se puede emular la percepción del enamoramiento puesto que requiere de la existencia del ser. En consecuencia, al carecer de existencia, una máquina no se puede preguntar, también por carecer de libre albedrío al estar sometida a la rigidez de la secuencialidad de sus operaciones en una CPU, GPU o TPU o en un conjunto de ellas, por el fin de la existencia, o por su sentido.

En ciencia, el rol del ser humano es clave, ya que la meta que le mueve es salvaguardar la existencia del prójimo y sentir que lo que está investigando mediante la herramienta de la ciencia contribuirá a una existencia mejor para la sociedad en su conjunto. En cambio, una inteligencia artificial al proponer nuevos teoremas matemáticos o leyes físicas no puede sentir las implicaciones de los descubrimientos efectuados, por más compleja que sea su descubrimiento y exclusivo de su artificio al poseer una inteligencia computacional más elevada, en términos reduccionistas, que el ser humano. En consecuencia, se argumenta que el propósito final de la ciencia, que es comprender la verdad del universo que nos rodea y mejorar nuestro bienestar como comunidad, es únicamente alcanzable por el ser humano ya que el ser humano es consciente de su propia existencia y la máquina carece del ser ahí, en términos del propio Heidegger.

Referencias

- [1] Penrose, R. (1991). The emperor's new mind. *RSA Journal*, 139(5420), 506-514.
- [2] Penrose, R. (1994). *Shadows of the Mind* (Vol. 4). Oxford: Oxford University Press.
- [3] Giorgi, A., & Giorgi, B. (2003). *Phenomenology*. Sage Publications, Inc.
- [4] Heidegger, M. (2010). *Being and time*. Suny Press.

- [5] Garrido Merchán, E. C. (2021). Transhumanismo y Consciencia Fenoménica. Congreso Razón Abierta.
- [6] Chalmers, D. J. (2002). Philosophy of mind: Classical and contemporary readings.
- [7] Bostrom, N. (2001). What is transhumanism. Nick Bostrom.
- [8] Russell, S., & Norvig, P. (2002). Artificial intelligence: a modern approach.
- [9] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- [10] Goertzel, B. (2007). Artificial general intelligence (Vol. 2). C. Pennachin (Ed.). New York: Springer.
- [11] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448.
- [12] Blanchette, J. C., Böhme, S., & Paulson, L. C. (2013). Extending Sledgehammer with SMT solvers. *Journal of automated reasoning*, 51(1), 109-128.
- [13] Competitive programming with AlphaCode, DeepMind Research. The AlphaCode team.
- [14] Dionisio, J. D. N., III, W. G. B., & Gilbert, R. (2013). 3D virtual worlds and the metaverse: Current status and future possibilities. *ACM Computing Surveys (CSUR)*, 45(3), 1-38.
- [15] Vanschoren, J. (2019). Meta-learning. In *Automated Machine Learning* (pp. 35-61). Springer, Cham.
- [16] Lin, T., Wang, Y., Liu, X., & Qiu, X. (2021). A survey of transformers. *arXiv preprint arXiv:2106.04554*.