LA INTELIGENCIA ARTIFICIAL: filosofía sistemática versus filosofía aplicada

23 de Noviembre 2023
Diego Rodríguez Gondonneau
Magíster en Filosofía
Profesor Asistente del Departamento de Filosofía de la Universidad de Concepción

Ante las dudas y temores suscitados por el proyecto científico de crear una Inteligencia Artificial, podríamos ensayar el método socrático. Si imitamos su reacción cada vez que algún cercano le pedía consejo, lo primero sería preguntar qué es la *inteligencia*. Luego deberíamos indagar si acaso es *posible* que haya inteligencia *artificial* (i.e. una inteligencia alojada no en un cuerpo o cerebro, sino en un aparato electrónico) y, sólo entonces, deberíamos preguntar si acaso es *bueno* que haya inteligencia artificial.

Sin embargo, debemos notar que las conversaciones del propio Sócrates suelen terminar en *aporía* y podría interpretarse que, en los escritos de Platón, se trata más de poner en entredicho la pregunta que de darle una respuesta, de algo así como "preguntarle a la pregunta": de hecho, según dichos escritos, los interlocutores de Sócrates le reprochan que "cambia el tema y termina siempre hablando de lo mismo".

Este *desfase* entre las "interrogantes del mundo" y el quehacer del filósofo no es sólo una peculiaridad de Sócrates y Platón: Nietzsche, acérrimo enemigo de Platón, pareciera aludir a dicho desfase cuando afirma "el que tiene el furor philosophicus en el cuerpo [...] se guardará sabiamente de leer los periódicos cada día [...]" (Nietzsche, Consideraciones Intempestivas, Schopenhauer Educador, *in* Obras, vol. I. Argentina: Aguilar; p. 142).

Si, a pesar de este desfase entre el quehacer del filósofo y las "interrogantes del mundo", insistimos en tomar a la filosofía como consejera sobre la Inteligencia Artificial, lo más razonable es evitar apoyarse exclusivamente en tal o cual filósofo, o bien en tal o cual escuela o época filosófica. Lo prudente es partir por echar un vistazo a la historia de la

filosofía, ya sea por nuestros propios medios o bien a través de algún filósofo que así lo haya hecho.

Además, puesto que se trata, entre otras cosas, de la posibilidad de construir una Inteligencia Artificial, lo más sensato sería indagar apoyándonos en filósofos menos alejados de este tipo de cuestiones y ver qué señalan sobre el panorama ofrecido por la historia de la filosofía respecto de la Inteligencia humana y su relación con el cuerpo humano. (Al pasar, podremos identificar algunos temas relevantes para la discusión cotidiana sobre la Inteligencia Artificial General.)

A.- Sólo sé que nada sé.

Blaise Pascal, filósofo reconocido por teoremas en secciones cónicas, por estudios en física sobre el vacío y, precisamente, inventor de la primera calculadora mecánica (s. XVII), respecto de la relación entre "espíritu" (sc. inteligencia) y "materia" (sc. cuerpo, cerebro o aparato electrónico), dice:

"Si somos simples y materiales, no podemos conocer [pues la materia no se puede conocer a sí misma], y si estamos compuestos de cuerpo y espíritu, no podemos conocer con nitidez ni las cosas simplemente espirituales ni tampoco las cosas simplemente corporales. Por esto es que todos los filósofos confunden las ideas de dichas cosas y hablan espiritualmente de las cosas corporales y corporalmente de las cosas espirituales, dicen que los cuerpos tienden hacia abajo… y hablando de los espíritus los consideran en un lugar y les atribuyen el movimiento de un lugar a otro. ¿Quién no creería, viéndonos describirlo todo como una mezcla de cuerpo y espíritu, que esa mezcla de cuerpo y espíritu es algo que sí comprendemos bien? Sin embargo, esa mezcla es lo que menos comprende el hombre y lo más prodigioso de la naturaleza (Ciudad de Dios, XXI.10 y Montaigne, Ensayos, II.12)". [traducción propia de: Pascal, Blaise. Pensées, *in* Oeuvres Complètes, Gallimard, 1999; §185, p. 613]

Más aún, el mismo Pascal (ibídem, §387,), mediante una alusión a Montaigne, consigna también el carácter oscuro de la pregunta por lo bueno para el hombre: "280 tipos del bien supremo en Montaigne".

Henri Bergson, filósofo de la primera mitad del siglo XX, cuyas obras presentan bibliografías que muestran el detallado estudio de la sicología y la neurología de su tiempo, también recalca la oscuridad de la relación entre “espíritu” y “materia”:

“Este libro afirma la realidad del espíritu y la realidad de la materia, e intenta aclarar la relación entre ambos a partir del ejemplo de la memoria. Este libro es netamente dualista. Sin embargo, considera el cuerpo y el espíritu de manera tal que espera reducir mucho, si no suprimir, las dificultades teóricas del dualismo. Realismo y el idealismo son dos tesis igualmente excesivas, es falso reducir la materia a la representación que tenemos de ella, también es falso considerarla algo que produciría en nosotros las representaciones, pero que sería de una naturaleza distinta a dichas representaciones. [traducción propia de Matière et Mémoire, in Bergson, Henri; Oeuvres; Presses Universitaires de France, 2001; p. 161.]

Si, por el contrario, decidiéramos abandonar la visión sinóptica de la historia de la filosofía, podríamos imaginar que alguno de los grandes sistemas de filosofía no acepta la posibilidad de un “espíritu” (i.e. inteligencia) fundado en un “cuerpo no-humano” (i.e. aparato electrónico) y, por ende, hace de la Inteligencia Artificial algo inverosímil. Sin embargo, esto también debería ser tomado con suma cautela:

1. Durante siglos, el vacío y el movimiento en el vacío eran una imposibilidad filosófica (e.g. para Aristóteles), sin embargo, la humanidad terminó aceptando su existencia.

2. La mismísima ciencia empírica a estado sometida no sólo a avances enormes, sino que también a profundas revoluciones, donde la ciencia anterior colapsa, surge algo radicalmente nuevo, y hasta pareciera que cambia “los respectivos científicos habitan mundos distintos” puesto que los presupuestos más elementales de sus teorías son tan distintos que dichas teorías llegan a ser lisa y llanamente “inconmensurables” (cf. Thomas S. Kuhn, Estructura de las Revoluciones Científicas).

B. La lección de Aristóteles ante la “Inteligencia Artificial”

A pesar de toda la incertidumbre ofrecida por la historia de la filosofía, todavía queda intentar sacar lecciones de la filosofía por otra vía: buscar si acaso el autor de alguno de los grandes sistemas de filosofía dió por hecho la existencia de algo así como la inteligencia artificial y, de ser así, ver qué actitudes “aconseja” su teoría o qué actitudes concretas tuvo ante ella dicho autor. Sobre todo, Josiah Ober recalca lo indispensable de la experiencia real y directa para el buen debate ético. Puesto que nosotros mismos todavía no estamos expuestos a la Inteligencia Artificial y puesto que es insuficiente el mero dejarse llevar por las fantasías de la ciencia ficción, no podemos dejar de aprovechar las lecciones ofrecidas por un gran filósofo que, en lo que a él mismo respecta, hubiera vivido en un mundo poblado de humanoides autónomos…

Pues bien, en la Atenas del siglo V y IV, abundaban los esclavos y, a diferencia de lo que ocurrió en el Sur de Estados Unidos, éstos se desempeñaban en las más diversas actividades, muchos eran de aspecto similar a los hombres libres y también había muchos que conseguían la libertad. En dichas circunstancias, Aristóteles (en los libros I y VII de la Política) menciona la existencia de esclavos por accidente y esclavos por naturaleza. Considera a los últimos como “instrumentos dotados de alma” e indispensables para el pleno desarrollo del hombre libre, dado que “la naturaleza no hace nada en vano”: es decir, la filosofía de Aristóteles reconoce y considera la existencia de algo así como “inteligencias no-humanas”, algo bastante cercano a la “inteligencia artificial” de nuestro tiempo.

Dichos “instrumentos dotados de alma”, considera Aristóteles, vivirían en un estado de subordinación que les sería provechoso, dada cierta superioridad intelectual de sus amos, así que la relación amo-esclavo sería cierta amistad. En efecto, los esclavos por naturaleza serían aquellas personas incapaces de anticiparse a lo necesario, carentes de

facultad deliberativa; capaces de *captar* la razón, pero no *poseedores* de ésta por sí mismos.

Por otra parte, Aristóteles recalca que no hay signos exteriores para distinguir acertadamente entre hombres libres esclavizados *de facto* y esclavos naturales: ambos tipos dominan el lenguaje, sienten placer y dolor, no se caracterizan por la raza. Así, lo primero que llama la atención es que el genio intelectual de Aristóteles no llegase a establecer un "test de Turing".

Además, pese a que afirma una y otra vez la necesidad y existencia de "esclavos por naturaleza", es decir personas con facultades disminuidas, su análisis conceptual de las facultades prácticas y teóricas del alma no permite un análisis completo del estatus concreto del "esclavo natural": en qué consistiría la inferioridad de un esclavo que, según el propio Aristóteles, debe ser capaz de dirigir a otros esclavos? ¿Qué ventaja tienen dichas "personas privadas de deliberación" que, a diferencia de los niños, no deben ser regidas por meras órdenes? Las categorías éticas y sicológicas de Aristóteles hacen difícil ver el camino entre el hombre pleno y el mero niño.

Aristóteles declara la utilidad de ofrecer la libertad como recompensa del buen trabajo a cualquier esclavo, lo que no calza con esa amistad que debería existir entre los esclavos por naturaleza y sus amos. Más aún, el testamento de Aristóteles decreta la libertad de muchos de sus esclavos, poniéndoles condiciones que nada tienen que ver con la superación de aquella precariedad intelectual que les atribuía.

Es decir, incluso para el propio Aristóteles, su obra filosófica no constituye una herramienta que permita resolver adecuadamente los dilemas concretos planteados por aquellas "herramientas dotadas de alma". Por el contrario, la discordancia entre la filosofía y la vida de Aristóteles en lo que a las "herramientas autónomas" se refiere, podría corroborar lo sugerido al comienzo, es decir la existencia de cierto *desfase* entre la Filosofía y "responder las interrogantes del mundo".

Eso sí, cabe señalar que numerosos autores de formación filosófica, pero dedicados a la "ética aplicada", han retomado la analogía de la esclavitud, sugiriendo o

desaconsejando tomar dicho fenómeno histórico como modelo para regular el diseño de Inteligencias Artificiales o el trato con ellas, me refiero a Joanna Bryson y a David Gunkel, cuyas principales ideas revisamos a continuación:

BRYSON aconseja

"Robot owners should not have obligations, but ensuring this is the responsibility of robot builders. Robot builders are ethically obliged – obliged to make robots that robot owners have no ethical obligations to. A robot's brain should be backed up continuously offsite by wireless network; its body should be mass produced and easily interchangeable. No one should ever need to hesitate an instant in deciding whether to save a human or a robot from a burning building. The robot should be utterly replaceable. Further, robot owners should know their robots do not suffer, and will never 'die' even if the rest of their owner's possessions are destroyed. The robot's brain state should be preserved off site. The robot can return to function exactly as before as soon as a new body can be acquired, though it may need some retraining if there is a new domicile to inhabit or slight variations between bodies. Robots then can be a relied on as no more than extensions of their owners. They should not be anthropoid if that can be helped, and their owners should have access to the robots' program-level interface as well as its more socially oriented one. This will help the owners form a more accurate, less human model for reasoning about their Companions […] Hopefully, […] a smaller proportion of everyone's time can be spent on mundane or repetitive tasks if they do not enjoy them. In that case, a larger proportion of time and resources can be spent on useful processes, including socializing with our colleagues, family, and neighbours." (Bryson, Joanna J., "Robots should be slaves", in Close Engagements with Artificial Companions, editado por Yorick Wilks, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2010; p. 73)

Así, la vía más fácil sería procurar mantener nuestros intentos circunscritos a crear Inteligencias relativamente limitadas, sin propósitos propios, sin placer ni dolor, para que en verdad sean como un órgano corporal separable; pero esto no agota los problemas…

GUNKEL advierte que “In The Phenomenology of Spirit, G. W. F. Hegel (1977, pp. 111–119)) famously demonstrated that slavery has negative consequences for the master who is, due to the very logic of the master/slave dialectic, incapable of achieving independence insofar as he is and remains beholden to the work performed by the slave. This philosophical insight has been borne out and verified by the historical evidence. As Alexis de Tocqueville (1899) reported about his travels through the southern United States, slavery was not just a problem for the slave, who obviously suffered under the burden of forced labor and dehumanizing racial prejudice; it also had deleterious effects on the master and his social institutions. “Servitude, which debases the slave, impoverishes the master” (de Tocqueville 1899, 361).” (Gunkel, David J., bosquejo de Person, Thing, Robot: a Moral and Legal Ontology for the 21st Century and Beyond, MIT Press, 2022; p. 13).

Es más, tomando ahora a pensadores menos técnicos, las cuestiones suscitadas por la Inteligencia Artificial no se limitan a las implicancias de tomarla como un esclavo.

Para Paul Mason (2020), son muchas las posibles consecuencias negativas de la eventual producción de computadores y máquinas dotados de Inteligencia Artificial General (AGI, en inglés): por ejemplo, reemplazo de trabajadores humanos, baja de los salarios; intromisión en nuestras vidas por parte de individuos, corporaciones privadas o estados, a través de nuestro abundante uso de aparatos electrónicos e Internet; pérdida de control sobre los ámbitos entregados al trabajo de la Inteligencia Artificial, sumisión a las órdenes o a los designios subrepticios de las inteligencias artificiales o de sus dueños.

En definitiva, el riesgo radicaría en una sumisión absoluta a la lógica del mercado y de las máquinas (p. 217), apoteosis del neoliberalismo, manipulación de nuestras conciencias, pérdida de autonomía para las personas (p. 200-208, i.a.), todo esto hecho plausible por la religión del siglo XX, i.e. el relativismo racional y moral suscitado por la convicción de que nuestras creencias metafísicas son totalmente contingentes, por creer que todo está totalmente determinado por ciertas leyes (Laplace): “from the complete

state of the universe at one moment of time, as described by the positions and velocities of all particles, it should be possible to predict all future states", *apud* Turing)" (p. 207), por considerar que nosotros estaríamos a merced de factores que escapan a nuestro control (biología, aleatoriedad, corrupción, los poderosos), por cierto, agnosticismo ocasionado por algunas interpretaciones de la Física Cuántica ("Interpretación de Copenhagen", que daría por obsoletas las nociones de causa y efecto, por imposible todo conocimiento prístino de la realidad) que, acompañadas de cierto "idealismo digital" que entiende al universo como un software o a la información como un Dios omnipresente, hace temer que podamos llegar a terminar presos al final de una caverna, cuyo fondo sería nuestro único horizonte y nuestro panorama estaría totalmente invadido por las sombras creadas para nosotros por "los dueños de la información". Todo "it" vendría de un "bit", todo "eso" sería producto de un oculto y ajeno "proceso": en suma, según Mason estaríamos sometidos al Genio Maligno imaginado por Descartes.

El mismo Mason propone un mandato global rigurosamente impuesto a los desarrolladores y controladores de la Inteligencia Artificial, sobre todo ahora que ésta (al menos en tareas como el juego de Go), ha demostrado ser capaz de su creatividad y, por lo tanto, "llevaría incorporada la capacidad de escapar al control humano". Algo similar es propuesto por Sam Altman, de Open AI, uno de los pioneros en el desarrollo de la Inteligencia Artificial (AI).

En suma, creemos que la Inteligencia Artificial suscita enormes desafíos e innumerables preguntas que, muy probablemente, pasarán a ser parte de nuestros inagotables debates en torno a lo justo y lo bueno.

BIBLIOGRAFÍA

- Aristóteles(2022). La Política. libros I y VII. Madrid. Gredos.

- Bergson, Henri. (2001)Matière et Mémoire; *in* Oeuvres; Presses Universitaires de France.

- Bryson, Joanna.(2010).“Robots should be slaves”, *in* Close Engagements with Artificial Companions, ed. Yorick Wilks, Oxford University.

- Fridman, Lex. (2023) (research scientist, Laboratory for Information and Decision Systems, MIT) entrevista a Sam Altman: https://www.youtube.com/watch?v=L_Guz73e6fw; descargado el 18 de octubre de 2023. Se puede revisar también en https://deeplearning.mit.edu/.

- Gunkel, David J.(2022) Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond. MIT Press.

- Mason, Paul. (2020). Por un futuro brillante. Madrid. Paidós.

- Nietzsche.(2012) Consideraciones Intempestivas, capítulo “Schopenhauer Educador”, in Obras, vol. I. Argentina: Aguilar.

- Ober, Josiah.(2023) Conferencia Inaugural Annual Lecture | Ethics in AI with Aristotle**;** descargado el 18 de octubre 2023 en https://www.youtube.com/watch?v=d1gTDGl7u1o&t=4332s .

- Pascal, Blaise.(1999) Pensées, *in* Oeuvres Complètes, Gallimard.

- Platón.(1981) La República, libro IX. Madrid.Gredos.