

Filosofía política e inteligencia artificial. Primeras noticias. Comentario a Mark Coeckelbergh, Filosofía política de la inteligencia artificial

(2023) Cátedra
Madrid, 224 pp.

Manuel Arias Maldonado
Universidad de Málaga
ORCID ID 0000-0003-1448-4379
marias@uma.es

Cita recomendada:

Arias Maldonado, M. (2024). Filosofía política e inteligencia artificial. Primeras noticias. Comentario a Mark Coeckelbergh, Filosofía Política de la Inteligencia Artificial. *Eunomía. Revista en Cultura de la Legalidad*, 26, pp. 434-439
DOI: <https://doi.org/10.20318/eunomia.2024.8520>

Recibido / received: 20/02/2024
Aceptado / accepted: 07/03/2024

Aunque el desarrollo de lo que quizá impropiaemente ha venido en llamarse Inteligencia Artificial (en adelante, IA) no es ninguna novedad, la irrupción pública del ChatGPT parece haber despertado a las sociedades contemporáneas a una realidad de contornos aún desconocidos y quizá por eso percibidos como amenazantes. Tampoco esta sensación de alarma tiene nada de original; la disrupción tecnológica provoca forzosamente una sacudida en las sociedades que la experimentan. De un lado, se cuentan los temores que las nuevas tecnologías suscitan en el plano cultural y psicológico, donde a menudo se tiene la sensación de que la «autenticidad» de lo «humano» se ve comprometida cuando la máquina gana terreno en la organización económica o las relaciones sociales: la escritora Virginia Woolf llegó a decir que el carácter humano había cambiado con la invención del cinematógrafo. Del otro, descartada ya la premisa decimonónica según la cual el progreso tecnológico produce beneficios universales, hay colectivos que se preocupan legítimamente por el daño que una innovación particular puede infligirles; los luditas solo fueron la vanguardia de los perjudicados por ese cambio tecnológico constante que define a la modernidad capitalista.

En los últimos años, la inquietud se ha centrado principalmente en las tecnologías relacionadas con la producción y transmisión de conocimiento e información que se benefician de la conexión en red facilitada por Internet, así como por sus aplicaciones en las distintas esferas sociales. Se teme por el futuro de la democracia a la vista del impacto de las redes sociales en la conversación pública; se



alega que el poder político y las propias empresas tienen ahora más capacidad que nunca para vigilar y controlar la vida privada de los ciudadanos; se teme por el futuro del empleo en una sociedad automatizada. Y, desde luego, no faltan quienes juzgan que la robotización y la digitalización van de la mano a la hora de «deshumanizar» unas sociedades que cada vez se alejan más del propósito emancipador introducido en la historia por el pensamiento ilustrado, dificultando la realización del ideal que representa el sujeto autónomo capaz de decidir cómo quiere vivir sin sufrir interferencias por parte del poder público o sujetarse al dominio de los demás.

Así ha sucedido también, como no podía ser menos, con la IA. Que una máquina pueda ayudarnos a traducir un texto a otro idioma, crear imágenes a nuestra medida o responder a cualquier pregunta imaginable ha sido considerado como el anticipo de un cambio civilizatorio. ¿De qué no será capaz esa «inteligencia artificial» que se perfecciona a sí misma y cuánto tardará en volverse contra sus creadores? ¿Quién ha decidido introducir a esta criatura en el mundo y por qué no se pregunta a los ciudadanos si quieren vivir con ella o preferirían no hacerlo? ¿Cómo podemos asegurarnos de que sus efectos serán benéficos, si ni siquiera sabemos todavía cuáles llegarán a ser? ¿Y qué tienen que decir sobre ello los teóricos de las disciplinas académicas que se dedican a analizar fenómenos sociales? Porque algo tendrán que decir, si es que queremos comprender mejor en qué consisten y cuáles son sus implicaciones en los distintos órdenes, incluyendo ese orden normativo en el que se plantean preguntas acerca de lo deseable e indeseable.

Hay que agradecer al belga Mark Coeckelbergh, filósofo de la tecnología de larga trayectoria radicado en la Universidad de Viena, que se haya preguntado qué diría una «filosofía política de la IA»; aunque sería más exacto decir que su libro es una suerte de cartografía de la IA a la luz de la tradición filosófico-política. Su método consiste en identificar aquellos temas clásicos de esta última cuya formulación teórica y aplicación práctica puede verse influida por el desarrollo de la IA, si no lo está siendo ya, a fin de discutir a continuación cómo debe responderse a ese impacto en el plano normativo. De ahí que el libro –después de un capítulo introductorio– esté organizado alrededor de los grandes problemas de la filosofía política: la libertad; la igualdad y la justicia; la democracia y el totalitarismo; el poder y la vigilancia. A ellos se añaden el medio ambiente y los animales no humanos; una conclusión relativa a la cualidad política de la tecnología cierra el trabajo. El resultado es una guía de indudable utilidad para la reflexión filosófico-política sobre la IA; el autor ha desbrozado el sendero tras penetrar hacha en mano en el bosque y, si bien el libro no destaca por la originalidad de sus tesis ni ofrece planteamientos propios acerca de lo que *deba* hacerse con la IA, sienta las bases para que esa reflexión pueda llevarse a cabo. Digamos que en este notable libro es todo lo que está, pero no está todo lo que es.

A juicio del autor, es necesario traer a primer plano las implicaciones *políticas* de la IA, mucho menos atendidas que las consideraciones *éticas* de las que él mismo se ha ocupado con fruición. Pero no estoy seguro de que tenga razón cuando sostiene que tanto la IA como la robótica suelen ser contempladas como un asunto técnico raramente vinculado a la política y cuyos efectos indeseados permanecen fuera de foco; que los filósofos de la tecnología cultivan la ética más que la política, en cambio, está fuera de duda. En cuanto a los teóricos políticos, no cabe duda de que algunas tecnologías han sido objeto de mayor atención que otras y de que la producción de novedades tecnológicas se considera a menudo una circunstancia sobrevenida con la que deben aprender a lidiar las sociedades humanas. No obstante, basta asomarse a la literatura producida por los teóricos políticos medioambientales para dar con una animada discusión acerca del papel de la tecnología en la conformación de esas mismas sociedades.

En cualquier caso, es indudable que irrupción de la IA exige una reflexión suplementaria en la que el papel de filósofos y teóricos políticos está llamado a ser protagonista; al menos, debería serlo. Y ello porque, como señala Coecklbergh desde el principio, la IA es «política» de cabo a rabo; no es una tecnología «neutral» cuyos efectos dependan del uso que los seres humanos decidan darle. Ninguna lo es: todas vienen definidas por eso que han venido a llamarse «*affordances*» o «posibilidades», cuya materialización produce un impacto variable —en forma e intensidad— sobre la sociedad y la subjetividad. Dicho esto, una escoba tiene menos potencia disruptora que la IA, pese ser ambas innovaciones tecnológicas en sentido propio. De hecho, si algo caracteriza a la IA es una *polivalencia* que dificulta realizar juicios tajantes acerca de su moralidad: no es lo mismo preguntar al ChatGPT por la receta para fabricar un cóctel Molotov que hacerlo por la vida de Santa Teresa de Jesús. De manera análoga, sería incongruente lamentar que el proceso de robotización disminuya de manera drástica el número de trabajadores industriales obligados a hacer el tipo de tareas repetitivas que denunciaba con humor Charles Chaplin en *Tiempos modernos*; incluso si con ello se pone fin a una subcultura comunitaria en la que han germinado sentimientos de solidaridad colectiva y formas de identidad vinculadas al trabajo manual.

Resulta desconcertante, en ese sentido, que la breve introducción del libro no contenga una descripción de la IA que permita al lector comprender el tipo de tecnología a la que el título ya se refiere. Hay que entender que Coecklbergh se ocupa —así se trasluce de los ejemplos que emplea— de las distintas aplicaciones conocidas de la IA, entre ellas la IA generativa (que se alimenta de los datos que se le proporcionan para hacer su propio aprendizaje imitativo); su reflexión se limita así a los confines de la llamada «IA débil», que es la que conocemos y que se diferencia de una futurible «IA general» cuyas capacidades excederían las de los seres humanos que las diseñan. Asistentes virtuales, vehículos autónomos, traductores automáticos, sistemas de recomendación y chatbots, instrumentos de reconocimiento facial e *tutti quanti* conforman el presente cotidiano de la IA; lo que pueda venir o cómo puedan desarrollarse estas aplicaciones, todavía no lo sabemos. En cuanto a la robótica, podemos encuadrar dentro de la misma aquellos artefactos capaces de realizar las tareas para las que han sido programados. Salta a la vista que la *fusión* de la IA y la robótica ofrece posibilidades inéditas.

La vocación didáctica e introductoria del libro lleva a Coecklbergh a comenzar cada sección con un desbrozamiento conceptual destinado a aclarar el sentido de los términos que emplea. Así sucede con la libertad, noción demasiado amplia para resultar operativa. Para el autor, es la libertad negativa —libre disposición personal de una esfera de no interferencia— la que está en riesgo con el despliegue de la IA que se dedica a la vigilancia: «La cámara o el robot están siempre ahí». Coecklbergh tira de Foucault y de la proverbial imagen del *Panopticon*: la monitorización a distancia sería ya suficiente para «disciplinar» a los individuos sin necesidad de coerción física directa. La vigilancia digital está a menudo escondida, advierte el autor, pudiendo incluso ser horizontal: «los usuarios de las redes sociales se vigilan unos a otros con la mediación de los algoritmos».

Sin embargo, la figura del *Panopticon* no puede aplicarse tan alegremente al tipo de «observación» que llevan a cabo esos algoritmos que rastrean las interacciones digitales en busca de patrones o términos que delaten un tipo particular de actividad. Nadie nos está mirando; son nuestras huellas las que se convierten *ex post* en indicios detectables por el algoritmo. Y lo mismo vale para la presunta vigilancia horizontal entre usuarios, atenuada de entrada por el hecho de que usamos nuestros teléfonos de manera voluntaria e incluso disfrutamos interactuando en las redes sociales. Afirmar de manera rotunda que se está interfiriendo con nuestra

libertad negativa en estos casos parece requerir matices por lo general ausentes en la filosofía política que lidia con esta nueva realidad.

En todo caso, el filósofo belga es el primero en admitir que la libertad negativa no es el único valor a considerar; la justicia o la seguridad reclaman también su lugar y eso genera conflictos entre bienes para cuya resolución carecemos de criterios incontestables. Algo parecido sucede con la libertad positiva, que él entiende como la posibilidad de decidir de manera autónoma cómo queremos vivir. Coeckelbergh recurre a la idea del *nudging* popularizada por Cass Sunstein y Richard Thaler, que como es sabido consiste en «empujar» al individuo hacia un uso más responsable de su autonomía sin por ello mermar su libertad para decidir. Para Coeckelbergh, es legítimo que el Estado obligue a las tabaquerías a incluir en las cajetillas a la venta una advertencia explícita sobre los riesgos que fumar tiene para la salud, pues uno podría comprar tabaco y conservar su libertad de decisión; en cambio, el *nudging* que se realiza mediante IA constituiría –tanto si lo hacen las empresas como si lo hace el poder público– una amenaza para la libertad positiva. Bien, pero ¿por qué? Leemos que «operando sobre la psicología subconsciente de las personas, [la IA] las manipula sin respetarlas como personas racionales». Por desgracia, Coeckelbergh no proporciona ningún ejemplo de esta manipulación y, en consecuencia, resulta difícil estar de acuerdo con él: no se ve por qué esta clase de influencia sería más eficaz o cualitativamente diferente de la que llevan a cabo muchos otros agentes persuasivos: de los medios de comunicación a las empresas que hacen publicidad, pasando por el gobierno que vende sus logros y el amigo que quiere convencernos de que invertamos dinero en su negocio. Se echa de menos la referencia a análisis empíricos capaces de probar –en la medida en que eso sea posible– que la IA nos empuja a hacer cosas que otras instancias no logran forzarnos a hacer. Este problema metodológico reaparece cuando el autor se ocupa de la automatización del trabajo y sostiene que el problema está en que una sociedad injusta no resuelve nada reemplazando a los trabajadores industriales por robots.

La tendencia a la hipérbole se aprecia también allí donde Coeckelbergh habla del empleo que las grandes empresas de telecomunicaciones hacen de nuestros datos. A su juicio, compartido por autoras como Carissa Véliz o Shoshana Zuboff, el capitalismo de la vigilancia se basa en el trabajo que los usuarios de las redes sociales hacen para esas empresas: somos nosotros quienes producimos una mercancía (los datos) que se vende a los clientes (anunciantes). Sentencia Coeckelbergh: «Esto es una forma de explotación». Pero no está nada claro que lo sea, porque el usuario ha aceptado un contrato y disfruta de los servicios que les proporcionan tales empresas a cambio de regalarles sus datos; es una transacción que, sea cual sea el juicio moral que nos merezca la actividad subsiguiente, no parece atormentarnos. Hablar de «explotación laboral» es, en consecuencia, inadecuado; necesitamos un vocabulario distinto y adaptado a los fenómenos emergentes de los que este libro trata se ocupa. En ese sentido, aplicar el viejo vocabulario marxista a la relación que mantenemos con las empresas tecnológicas no parece ayudarnos a comprenderla: igual que un trabajador industrial allá por 1834 hubiera dejado sin dudar la fábrica para retirarse con una pensión, pocos de nosotros renunciaríamos a seguir usando nuestro teléfono. En una sección posterior, Coeckelbergh llegará a invocar a Hannah Arendt para sugerir la posibilidad de un «totalitarismo digital»; sus contornos, empero, no son dibujados con la precisión exigible a tan fenomenal aseveración.

No tengo aquí espacio para discutir todos los temas que aborda el autor; deteniéndome en su tratamiento de la libertad, he querido dar una impresión de las fortalezas y debilidades del libro. Coeckelbergh es exhaustivo a la hora de presentar los distintos enfoques que, dentro de la filosofía política, pueden aplicarse a un problema clásico de la disciplina –como la justicia o la democracia o el poder– y

explora la manera en que la IA puede afectarle. Su posición personal no es clara; suele limitarse a presentar posiciones contrapuestas y a formular preguntas sin llegar a responderlas. Es comprensible: aquí se transita un territorio virgen. Por ejemplo, ¿quién sabría decir si la IA fortalecerá o debilitará la democracia? Dado que Coeckelbergh tampoco lo sabe, empieza por señalar que la IA tiene un marcado carácter tecnocrático y nos recuerda que la tecnocracia posee serios problemas de déficit democrático; eso conduce a la pregunta sobre cuál debe entonces ser el papel de los expertos y la tecnología en una democracia. De ahí transita a una distinción entre formas representativas y formas deliberativas de democracia, a las que añade la democracia agonista; todas ellas son presentadas con detalle y con el auxilio de la literatura pertinente. Surgen nuevas preguntas: ¿pueden la ciencia y la tecnología ayudarnos a superar la ignorancia de la mayoría? Si la democracia ha de ser más deliberativa y participativa, ¿cuál es el rol de la IA en ella? Y volvemos al principio: ¿nos lleva la IA hacia la tecnocracia, o una democracia de tintes habermasianos se hace viable de su mano? El autor no da respuesta. Para colmo, avala la discutible idea de las «cámaras de eco» creadas por las redes sociales, saca a colación a Cambridge Analytica como causa mayor del *Brexit* y llega a la conclusión de que «la IA no es solo una herramienta para la política, sino que cambia la política misma». Se nos advierte también de que si la IA es empleada para la toma de decisiones sin la transparencia debida, nos encontraremos con un problema de rendición de cuentas; y se concluye que las políticas públicas relativas a la IA «tendrán que equilibrar valores políticos diferentes que potencialmente entran en conflicto entre sí». Es difícil mostrarse en desacuerdo con afirmaciones tan genéricas.

Algo parecido sucede cuando Coeckelbergh vuelve su atención sobre los problemas medioambientales y su relación con la IA. De un lado, nos dice, esta tecnología podría ayudarnos a analizar datos climáticos y a conservar ecosistemas, pero también es posible que contribuya a la explotación industrial de los animales y, bueno, no deja de ser cierto que la computación a gran escala consume enormes cantidades de energía... A estas alturas, se ha constatado ya que la IA y sus aplicaciones se caracterizan por una *ambivalencia* inerradicable. Pese a la insistencia del autor en la idea de que la tecnología en sentido genérico nunca es neutral, porque cada tecnología hace posible unos usos particulares que definen su carácter, hay que preguntarse si una tecnología tan diversa y versátil como la IA no acredita esa neutralidad en mayor medida que otras; si, como se deduce de su propio trabajo, lo determinante no será que definamos cuáles serán los usos legítimos de la IA por contraste con aquellos otros que nos parecen indeseables. Asunto distinto, que el libro no aborda, es que eso sea factible; aunque la tecnología es siempre política, parece difícil someter la innovación a control democrático.

Así es como llegamos a la breve conclusión con la que se cierra el trabajo, donde el filósofo belga recuerda su propósito: mostrar cómo los conceptos procedentes de la filosofía política y la teoría social pueden ayudarnos a enmarcar y comprender los desafíos normativos que trae consigo la IA. Pese a que el libro tiene vocación introductoria, ofrece una base sólida sobre la que construir un marco normativo que haga posible pensar sobre los aspectos políticos de la IA; así lo piensa su autor. No seré yo quien diga que no hace tal cosa: este libro cumple con su cometido de manera admirable, especialmente si consideramos las dificultades que comporta elaborar una filosofía política de nuevo cuño para un fenómeno emergente. Pero eso es lo que no encontrará el lector en estas páginas: aquí no hay conceptos originales ni tesis novedosas acerca de las implicaciones políticas de la IA, sino un esforzado trabajo de aplicación conceptual —por momentos lastrado por una excesiva complacencia hacia las tesis más apocalípticas sobre la naturaleza y el destino del capitalismo liberal— que nos permite ver más claro el horizonte que se dibuja delante

de nosotros y nos recuerda la importancia de formular las preguntas correctas, rehusando sin embargo a responderlas.

