

eISSN: 1989-3612

DOI: <https://doi.org/10.14201/art2024.31973>

RESEÑA / BOOK REVIEW

Coeckelbergh, Mark (2024). La ética de los robots. Madrid: Cátedra. Traducción de Lucas Álvarez Canga

Virginia YOLDI LÓPEZ

Profesora-tutora de filosofía en UNED Pamplona

Marc Coeckelbergh, autor de referencia en la filosofía de la tecnología, presenta en *La ética de los robots* una amplia panorámica de las cuestiones éticas que suscita el uso de los robots, entendidos estos como máquinas capaces de realizar numerosas tareas de modo autónomo o semiautónomo.

Partiendo de la constatación de que los robots no son ciencia ficción (“los robots no están en camino: ya están aquí” p.11), nos muestra la necesidad de pensar en profundidad sobre ellos. El libro, que se dice introductorio, tiene los objetivos explícitos de responder a cuestiones éticas relacionadas con los robots, ofrecer herramientas conceptuales para pensar sobre esos problemas y mostrar que pensar sobre robótica nos ayuda a pensar lo humano.

El texto se divide en ocho capítulos, cada uno de ellos dedicado a una tipología de robots, salvo el primero, que es una introducción, y el último, de temática antropológica. Es notoria la voluntad de claridad en la exposición sin pérdida de rigor filosófico. Para ello, el autor, utiliza recursos como la subdivisión en apartados cortos, abundantes resúmenes y recapitulaciones, frases destacadas, ejemplos tomados de la ficción (*Westworld*, *Frankenstein*) y referencia a noticias o anécdotas conocidas (como

atropellos con coches autónomos). Todo ello completado con un índice analítico y un glosario, además de una nutrida bibliografía.

Cada capítulo describe los usos e impactos de los distintos tipos de robots en la sociedad y plantea preguntas éticas al respecto. Buscando las respuestas, el autor se adentra en la reflexión filosófica, cuya profundidad varía de unos capítulos a otros. Coeckelbergh dialoga con otros autores a propósito de la demarcación de los problemas y de las posiciones éticas posibles. Por ejemplo, cuando habla de robots asistenciales, se adentra en la definición del buen cuidado y el papel que los robots pueden tener en él, explicando los puntos de vista de pensadores como Linda y Robert Sparrow, cautelosos con el uso de la máquina en estos entornos porque puede provocar deshumanización, y la suya propia proponiendo un decálogo o ideal normativo para el buen cuidado.

El análisis ético de las modalidades de robots a las que se refiere (industriales, del hogar, asistenciales, vehículos autónomos, androides y drones asesinos) va abriéndose a campos de pensamiento más amplios en cada caso. Así el segundo capítulo nos muestra cómo la reflexión ética sobre los robots en la industria nos lleva inevitablemente a plantearnos cómo organizamos el trabajo en las sociedades capitalistas actuales o qué impacto en el medioambiente tienen esas máquinas. Del mismo modo, los robots asistenciales nos hacen repensar los sistemas sanitarios, los drones asesinos reflexionar sobre la guerra en general o los androides sobre las relaciones humanas.

El autor aborda los temas con enfoque multifactorial e integrador, ejemplo de esto es el tratamiento de la cuestión del engaño. Tanto en la utilización de los robots del hogar como de los asistenciales para sanidad, de los androides o, incluso, de los drones asesinos, se pregunta hasta qué punto hay un déficit ético debido a que se engaña al usuario. Engaño conectado a la posible infantilización, cosificación, humillación o trato indigno de las personas. Coeckelbergh cita a pensadores que consideran inaceptable la faceta engañosa de la máquina y exigen retirarla o darle otros usos (como Noel y Amanda Sharkey) y explica también lo que piensan autores que no consideran que haya engaño, o que lo minimizan o que cuestionan el engaño recurriendo al concepto de ilusión co-creada por el propio usuario, que participa en ella conscientemente. Posibilidades todas ellas éticamente relevantes.

Es central en esta obra el problema de la responsabilidad. Cómo distribuir responsabilidades cuando quien actúa es una máquina, hasta qué punto los humanos son responsables y qué humanos de la cadena de producción y uso lo son. Así, explorando el caso de los vehículos autónomos, surge la reflexión sobre si podemos responsabilizar a los robots

de las consecuencias que provocan sus actuaciones. Se hace eco de quienes, como John Danaher, proponen un conductismo ético en el que no sería necesaria la conciencia del agente para atribuirle responsabilidades éticas¹. Coelkebergh, por su parte, mantiene la necesidad de centrar la responsabilidad en los humanos, al menos, por ahora: "bajo las actuales circunstancias, sería altamente irresponsable no hacer responsables a los humanos de las acciones y consecuencias de los artefactos... que llamamos robots" (p. 110).

La velocidad con la que se avanza en estos campos hará que tengamos que enfrentar problemas nuevos, siendo consciente de ello, el autor nos invita a ir pensando nuestro presente y ser cautelosos. Así, la cuestión de atribuir derechos a los robots ligada a la posibilidad de que tengan conciencia, con ser interesante, no es central para él, dado que, por el momento, no estamos en esa realidad.

Los robots nos sirven de espejo para pensarnos como humanos. Es una idea que recorre el libro y que se desarrolla en profundidad en el último capítulo donde insiste en la honda conexión entre ética y antropología y explora algunas de las distintas concepciones de lo humano presentes en la filosofía actual atravesada por la tecnologización. Estas posiciones serían el humanismo, el transhumanismo, el posthumanismo y una ecología integradora y abarcativa. La exposición sintética del transhumanismo y, sobre todo, del posthumanismo con su defensa de la hibridación y la ruptura del binarismo categorial resultan brillantes, así como su toma en consideración de una ética que saca al humano del centro y coloca al medioambiente global como prioridad para actuar éticamente.

El recorrido del autor tiene gran cantidad de desviaciones y recovecos a explorar. Toma en serio casi cualquier derivada o posibilidad, transmitiéndonos que, en un mundo tan cambiante, cualquier cosa podría pasar, y la filosofía no puede ignorarlo. En ocasiones, Coelkebergh señala senderos sin continuarlos, plantea una gran cantidad de preguntas invitando a la reflexión dejando muchas de ellas sin intento explícito de respuesta. Por eso el lector agradecerá y disfrutará los planteamientos explorados con detenimiento, las argumentaciones completas y las respuestas más o menos convincentes, como cuando explica, en el capítulo 6, su propuesta de la posición moral indirecta de los andróides basada en un enfoque ético relacional.

1. En el libro *Tecnofilosofía* de Anibal Astobiza se explora un conductismo relacional de este tipo proyectado hacia un futuro más robotizado.

Este libro es también, a su modo, un compendio de teorías éticas, que van apareciendo al hilo de los distintos problemas. El autor las explica con sencillez y muestra las posibilidades de su aplicación a casos planteados por la robótica. Por ejemplo, cuando trata de robots androides y cómo enfocar el maltrato a estos, dice: “si alguien maltrata a un robot, esto puede conducir al maltrato a humanos (un argumento consecuencialista) y/o a que humanos no ejerzan sus obligaciones hacia otros humanos (un argumento deontológico) y /o a un carácter moral malo y vicioso (un argumento de la ética de la virtud)” (p. 122).

El hecho de que los robots estén dotados de Inteligencia artificial y tengan cierta agencia hace que sean recurrentes en el libro temas como la seguridad, la privacidad, la vigilancia, la justicia y la desigualdad. Además de la explicabilidad, la opacidad y la responsabilidad. Esto lo pone en relación con otras obras recientes que tratan similares problemas en las tecnologías digitales. Kate Crawford en *Atlas de IA* o James Bridle en *La nueva edad oscura*, Shoshana Zuboff en *El capitalismo de la vigilancia*, Carissa Véliz en *Privacidad es poder*, Remedios Zafra en *El bucle invisible* o los autores de *Robótica, ética y política* editado por Norbert Billbeny se han sumergido en estas cuestiones, que se conectan con la necesidad de regular, educar y debatir a fondo sobre máquinas autónomas, como insiste Coeckelbergh. Dado que la normativa y el control sobre estas tecnologías son aún escasos y están en proceso de implementación, todo ejercicio de reflexión informada es interesante.

En muchos de esos autores, las perspectivas que se nos presentan ante las tecnologías más sofisticadas son de advertencia, con tono globalmente pesimista. Sin embargo, Coeckelbergh, aunque no olvida inconvenientes y peligros, aborda estos temas sin poner por delante los miedos, sino adentrándose con rigor, con mente abierta a la búsqueda de soluciones. No se adhiere a la utopía ni al solucionismo tecnológico, tampoco al catastrofismo. Confía en la inteligencia del lector y en su capacidad de pensar por sí mismo cuando se le proporcionan los datos y las herramientas filosóficas adecuadas. Esto concuerda con su énfasis en que el pensamiento ético ha de extenderse a todos los implicados en la fabricación y uso de los robots; todos deben sentirse concernidos y todos deben ser escuchados. Por ello propone también integrar la ética en el plan de estudios de las materias técnicas relevantes.

Como señala el autor: “Hay que comprender y evaluar las tecnologías ahora, antes de que sea demasiado tarde y conlleven impactos que nadie quiera” (p 12). *La ética de los robots* es un buen modelo de cómo hacerlo.

REFERENCIAS BIBLIOGRÁFICAS

- Astobiza, Anibal (2023). *Tecnofilosofía*. Madrid: Plaza y Valdés.
- Bilbeny, Norbert (ed.) (2023). *Robótica, ética y política*. Barcelona: Icaria.
- Bridle, James (2020). *La nueva edad oscura*. Barcelona: Debate.
- Crawford, Kate (2023). *Atlas de IA*. Madrid: NED.
- Véliz, Carissa (2021). *Privacidad es poder*. Barcelona: Debate.
- Zafra, Remedios (2022). *El bucle invisible*. Madrid: Nobel.
- Zuboff, Shoshana (2020). *La era del capitalismo de la vigilancia*. Barcelona: Paidós.