

La ética de los robots

Colección Teorema
Serie mayor

Mark Coeckelbergh

La ética de los robots

Traducción de Lucas Álvarez Canga

CÁTEDRA
TEOREMA

Título original de la obra:
Robot Ethics

1.ª edición, 2024

Reservados todos los derechos. El contenido de esta obra está protegido por la Ley, que establece penas de prisión y/o multas, además de las correspondientes indemnizaciones por daños y perjuicios, para quienes reprodujeren, plagiaren, distribuyeren o comunicaren públicamente, en todo o en parte, una obra literaria, artística o científica, o su transformación, interpretación o ejecución artística fijada en cualquier tipo de soporte o comunicada a través de cualquier medio, sin la preceptiva autorización.



© 2022 Massachusetts Institute of Technology, 2024

© De la traducción: Lucas Álvarez Canga, 2024

© Ediciones Cátedra (Grupo Anaya, S. A.), 2024

Valentín Beato, 21. 28037 Madrid

Depósito legal: M. 31.096-2023

ISBN: 978-84-376-4706-7

Printed in Spain

Índice

CAPÍTULO PRIMERO. Introducción: ¿de qué trata la ética de los robots?	9
¿Están los robots de camino o ya están aquí?	9
Objetivos de este libro: cuestiones éticas y reflexión filosófica	12
¿Qué son los robots? ¿Qué es la ética? Y ¿qué quiere decir «ética de los robots»?	13
Enfoque, estructura y alcance del libro	20
CAPÍTULO 2. Robots industriales, seguridad y el futuro del trabajo	29
Trabajando con robots: Karl Marx, industria 4.0 y algunas cuestiones éticas	30
La cuarta revolución industrial: el futuro de la economía, del trabajo y de la sociedad	37
El significado del trabajo, la equidad de la sociedad y el futuro del planeta	42
CAPÍTULO 3. Compañeros del hogar robóticos, privacidad y engaño ...	49
Compañeros y asistentes del hogar robóticos como robots sociales con IA	51
Privacidad, seguridad y protección	52
Engaño, el valor de las relaciones humanas reales, diferencia y sesgo de género y racial	56
La ética de los robots personales más allá de los humanos	67
CAPÍTULO 4. Robots asistenciales, pericia y la calidad de la atención sanitaria	71
Los robots en la atención sanitaria	72

Problemas éticos	74
¿Qué es un buen cuidado?	81
CAPÍTULO 5. Coches autónomos, agencial moral y responsabilidad	91
Algunas cuestiones con respecto a los robots autónomos	94
¿Máquinas morales? ¿Pueden los robots ser agentes morales?	94
La atribución de responsabilidad en el caso de los robots autónomos	99
La responsabilidad distribuida dentro de redes de responsabilidad	107
CAPÍTULO 6. Androides inquietantes, apariencia y paciencia moral	111
¿Es aceptable amar o torturar a las máquinas?	112
¿Debería importar éticamente la apariencia de los androides?	115
¿Cuál es la posición moral de los robots?	119
La cuestión con respecto a la posición moral directa	119
La cuestión con respecto a la posición moral indirecta	122
Pensar de otra manera y el enfoque relacional de la posición moral	124
CAPÍTULO 7. Drones asesinos, distancia y la existencia humana	131
El uso de robots para fines militares	133
Cuestiones éticas y otros problemas filosóficos planteados por los drones asesinos	136
La guerra justa y la responsabilidad	137
¿Las matanzas selectivas mediante drones son guerra o asesinato?	140
Matar a distancia	141
Justicia y virtud	145
Explicabilidad y responsabilidad	146
¿Deberíamos permitir que las máquinas maten a seres humanos?	147
CAPÍTULO 8. Espejos robóticos más allá de lo humano: la ética de los robots como una ética medioambiental	153
Espejos robóticos: todo sobre los humanos	154
Más allá de lo meramente humano	157
AGRADECIMIENTOS	171
GLOSARIO	173
BIBLIOGRAFÍA	179
ÍNDICE ANALÍTICO	193

CAPÍTULO PRIMERO

Introducción: ¿de qué trata la ética de los robots?

En la película de ciencia ficción estadounidense de 2004 *Yo, robot*, los robots humanoides sirven a la humanidad. Sin embargo, no todo va bien. Tras un accidente, un hombre es rescatado del coche que se está hundiendo por un robot, pero una niña de doce años no se salva. El robot calculó que el hombre tenía una probabilidad de supervivencia más alta: los humanos puede que hubieran hecho otra elección. En la película, los robots intentan más tarde tomar el control del poder que tenían los humanos. Los robots están controlados por una inteligencia artificial (IA), VIKI, que decide que restringir el comportamiento de los humanos y matar a algunos de ellos asegurará la supervivencia de la humanidad. La película ilustra el miedo a que los robots humanoides y la IA tomen el control del mundo. También apunta a hipotéticos dilemas éticos diferentes en el caso de que los robots y la IA llegasen a alcanzar inteligencia en un sentido general. Pero ¿es de esto de lo que trata o debería tratar la ética de los robots?

¿ESTÁN LOS ROBOTS DE CAMINO O YA ESTÁN AQUÍ?

Normalmente, cuando la gente piensa sobre los robots, la primera imagen que nos viene a la cabeza es la de un robot altamente inteligente con forma humana. A menudo esta imagen proviene de la ciencia fic-

ción, donde encontramos robots que se parecen y se comportan más o menos como los humanos. Muchas narraciones nos alertan de que los robots tomarán el poder. El miedo que estas nos producen consiste en que dejarán de ser nuestros sirvientes y pasarán a convertirnos en *sus* esclavos. El mismo término «robot» significa «trabajo forzoso» en checo y aparece en la obra de teatro *R.U.R.* de Karel Čapek, puesta en escena en Praga en 1921: justamente hace cien años. La obra es parte de una larga historia de narraciones sobre máquinas rebeldes similares a los humanos, desde el *Frankenstein* de Mary Shelley a películas como *2001: Una odisea del espacio*, *Terminator*, *Blade Runner* y *Yo, robot*. En el imaginario público, los robots son, a menudo, objeto de miedo y fascinación al mismo tiempo. Tenemos miedo de que tomen el poder, pero, al mismo tiempo, resulta emocionante pensar sobre la creación de un ser artificial que sea parecido a nosotros. Como parte de nuestra herencia romántica, los robots son proyecciones de nuestros sueños y pesadillas sobre crear un otro artificial¹.

En primer lugar, estos robots son ante todo terroríficos: son monstruos y causan asombro. Pero al comienzo del siglo XXI surge en Occidente una imagen diferente de los robots: el robot como compañero, amigo y, quizás, incluso pareja. La idea es ahora que los robots no deberían estar limitados a las fábricas industriales o a lugares remotos en el espacio. En el imaginario contemporáneo, están liberados de su sucio trabajo esclavo, y entran al hogar como compañeros sociables útiles, agradables y, a veces, sexis, con los que se puede hablar. En algunas películas, al final se rebelan (piénsese en *Ex Machina*, por ejemplo), pero, generalmente, se convierten en lo que los diseñadores de robots llaman «robots sociales». Están diseñados para mantener una interacción «natural» humano-robot: esto es, para interactuar de la forma en la que estamos acostumbrados con otros seres humanos o mascotas. Están diseñados para no dar miedo o parecer monstruosos, sino para ser bonitos, útiles, entretenidos, divertidos y seductores.

Esto nos conduce a la vida real. Los robots no están de camino: ya están aquí. Pero no son exactamente como los robots que nos encontramos en la ciencia ficción. No son como el monstruo de Frankenstein o Terminator. Son robots industriales y, a veces, «robots sociales». Sin embargo, estos últimos no son tan inteligentes como los humanos o sus

¹ Coeckelbergh, *New Romantic Cyborgs*.

Los robots no están
de camino: ya están aquí.
Pero no son exactamente
como los robots que nos
encontramos
en la ciencia ficción.

parientes de la ciencia ficción y, a menudo, no tienen forma humana. Incluso los robots sexuales no son tan inteligentes o tan conversacionalmente capaces como el robot representado en *Ex Machina*. A pesar de los avances recientes en IA, la mayoría de ellos no se parecen a los humanos en ningún aspecto. Dicho esto, los robots están aquí, y están aquí para quedarse. Son más inteligentes y tienen mayor capacidad de funcionar de forma autónoma que antes. Y tienen aplicaciones más orientadas al mundo real. Los robots no se usan solamente en la industria, sino también en la atención sanitaria, el transporte y la ayuda doméstica.

A menudo, esto hace que la vida humana sea más fácil. Sin embargo, también hay problemas. Algunos robots pueden ser ciertamente peligrosos, no porque puedan intentar matar o seducirte (a pesar de que los «drones asesinos» y los robots sexuales también están en el menú de la ética de los robots), sino, normalmente, por razones mucho más mundanas, como que puedan quitarte el trabajo, puedan engañarte y hacer que pienses que son una persona, y puedan causar accidentes cuando los usas de taxi. Miedos tales como estos no son ciencia ficción: pertenecen al futuro próximo. Dicho de manera más general: desde el impacto de las tecnologías nuclear, digital y de otro tipo en nuestras existencias y en el planeta, hay una conciencia y un reconocimiento cada vez más presente de que las tecnologías conllevan cambios fundamentales en nuestras vidas, sociedades y entorno, y que, por lo tanto, mejor es que pensemos más, y de forma más crítica, sobre su uso y su desarrollo. Hay aquí una cierta urgencia: debemos comprender y evaluar las tecnologías ahora, antes de que sea demasiado tarde, esto es, antes de que conlleven impactos que nadie quiera. Este argumento también se puede utilizar con respecto al desarrollo y uso de la robótica: consideremos las cuestiones éticas planteadas por los robots y su uso en la fase de desarrollo en lugar de después de los hechos. Diré ahora más sobre los objetivos y el alcance de este libro.

OBJETIVOS DE ESTE LIBRO: CUESTIONES ÉTICAS Y REFLEXIÓN FILOSÓFICA

En la medida en que las tecnologías robóticas y de automatización abandonen el reino de la ciencia ficción y penetren cada vez más en nuestras vidas cotidianas, es importante no solo ver los beneficios y oportu-

tunidades potenciales, sino también discutir las cuestiones éticas y sociales que plantean, ahora y en el futuro cercano. Considérese, por ejemplo, los robots industriales que son cada vez más inteligentes y trabajan con humanos en fábricas, los robots usados por usuarios vulnerables como es el caso de los niños, los coches autónomos que están siendo desarrollados por prácticamente todos los fabricantes de coches, los robots usados para cirugía en hospitales y los drones letales que se usan como armas de guerra. Este libro responde a estas cuestiones ofreciendo una concepción general de algunos problemas éticos y sociales clave junto con una gama de herramientas conceptuales de la ética de los robots para pensar sobre estos problemas. Muestra ejemplos de aplicaciones robóticas prácticas, identifica problemas éticos y sociales que plantean estas aplicaciones, y ofrece herramientas conceptuales para lidiar con estos problemas. La idea que está detrás de este ángulo práctico es que, de esta forma, la ética de los robots puede ayudar a hacer que el desarrollo y gobierno de la robótica procedan de una forma más éticamente responsable.

Además, el libro busca ofrecer una reflexión filosófica sobre qué son los robots y qué significa pensar sobre robótica para pensar sobre lo humano. Por ejemplo, nos hace reflexionar sobre cómo pensamos los seres humanos acerca del estatus moral. Esto proporciona una perspectiva más amplia que es muy necesaria en debates que normalmente están limitados a una preocupación ética más inmediata, y ayuda a explicar la profunda fascinación que ejercen sobre el público general los robots y las máquinas. Enseguida diré algo más sobre el significado filosófico de la ética de los robots. Pero haré primero una pregunta sencilla: ¿qué se quiere decir con ética de los robots?

¿QUÉ SON LOS ROBOTS? ¿QUÉ ES LA ÉTICA?
Y ¿QUÉ QUIERE DECIR «ÉTICA DE LOS ROBOTS»?

Si queremos discutir en este libro sobre la ética de los robots, ¿de qué queremos hablar? ¿Qué es «robot», y qué es «ética»? Esto ya nos conduce a problemas de definición, una preocupación típica de filósofos.

En primer lugar, ¿qué se quiere decir con el término «robot»? Ya he mencionado su significado etimológico, relacionado con sirvientes y esclavos. Sin embargo, cuando tratamos de habérnoslas con los robots en

la práctica, este es solamente un significado entre muchos. El significado del término es controvertido. Ni los expertos en robótica ni los filósofos se ponen de acuerdo sobre la definición de robot. Por ejemplo, el Instituto Internacional de Ingenieros Eléctricos y Electrónicos (Institute of Electrical and Electronics Engineers) define los robots como sigue: «un robot es una máquina autónoma capaz de detectar su entorno, llevando a cabo computaciones para tomar decisiones y realizando acciones en el mundo real»². Pero ¿hace esto que un termostato sea un robot? ¿Y un lavavajillas? ¿Y un piloto automático? ¿Deberían los robots ser capaces de moverse? ¿Dónde está la línea que los separa de otros tipos de *hardware*? ¿Es un coche autónomo un robot? ¿Cómo de material tiene que ser? Los robots tienen componentes de *hardware* (físicos) y de *software* (código). Si un agente artificial consistiera solo en *software*, se llamaría un «bot»: no se consideraría un robot. Patrick Lin y sus colegas han argumentado que, en contraste con los ordenadores o la IA no corporeizada, un robot «puede ejercer influencia sobre el mundo de forma directa»³. Aun así, ¿por qué es menos directa la influencia de un «bot»? ¿Y qué sucede si el *software* está conectado al *hardware*, sin tener la forma humana o animal que normalmente asociamos con «robot»? Finalmente, ¿cómo de autónomo e inteligente debería ser un «robot»? Por ejemplo, a veces los robots se combinan con IA, pero esto no tiene por qué ser necesariamente así. Dada la importancia cada vez mayor de la interacción entre los robots y los humanos, un término usado habitualmente es el de «interacción humano-robot»: el énfasis no está en el robot como un artefacto material, sino en la interacción entre humanos y robots. Hay definiciones amplias y estrechas de robots. En este libro, me centraré en aquellos robots que tienen *hardware*, y resaltaré a menudo problemas éticos planteados por robots que tienen un alto nivel de inteligencia, autonomía e interactividad, pero también incluiré otros tipos de robots.

Sin embargo, las definiciones técnicas no son suficientes. La controversia sobre el futuro de la robótica y las imágenes de la ciencia ficción apuntan tanto al contexto del robot, así como a las dimensiones humanas, sociales y culturales de la robótica. Lo que un robot «es» no puede ni debe ser reducido al artefacto material «robot», sino que debe

² Disponible en: <<https://robots.ieee.org/learn/>>.

³ Lin, Abney y Bekey, «Robots Ethics», 943.

Lo que un robot «es» no puede ni debe ser reducido al artefacto material «robot», sino que debe conectarse más bien con su uso y sus contextos sociales y culturales.

conectarse más bien con su uso, y sus contextos sociales y culturales. Lo que un robot «es» siempre está moldeado por el uso humano, (inter) acción, subjetividad y cultura. Por ejemplo, si ciertos humanos particulares usan un robot particular de una forma en la que lo tratan como una mascota, el significado de ese robot en ese contexto y situación no está descrito exhaustivamente llamándolo cosa o máquina. Las narrativas sobre robots mencionadas al comienzo de este capítulo no son irrelevantes a la hora de pensar sobre los robots: es mejor que les prestemos atención, ya que influyen no solo en la percepción pública de la robótica, sino también en su desarrollo. Los ingenieros y diseñadores no son inmunes a la ciencia ficción: también les influye *a ellos*. Algunos sueñan con construir una máquina con forma humana o, como usuarios del robot, proyectan todo tipo de significados antropomórficos (con forma humana) sobre el robot. Algunos evocan tales significados a propósito para hacer que su robot sea más «social» y aceptable. Usando de nuevo la imagen anterior, dar a los robots un nombre personal (el nombre de un humano o animal) no es una excepción: es una práctica común. Las definiciones técnicas de los robots, por lo tanto, son necesarias, pero no suficientes en la ética de los robots. Los robots no son simples máquinas: siempre son al mismo tiempo humanos, sociales y culturales. Su significado no puede reducirse a definiciones técnicas.

Los robots comparten este sentido multifacético con otras tecnologías. Así como la tecnología en general, el término «robot» se puede referir a una variedad de fenómenos. La tecnología no trata solo sobre artefactos materiales. También trata sobre el conocimiento necesario para desarrollar y usar estos artefactos, así como sobre la ciencia relacionada; normalmente de los programas de computación inmaterial; actividades como el uso, diseño y mantenimiento, y de los actores humanos y sociales involucrados en el desarrollo y uso de las tecnologías. Su despliegue tiene lugar dentro de un marco socioeconómico concreto (por ejemplo, el capitalismo) y en una cultura particular, su uso y desarrollo implican suposiciones sobre lo humano y la tecnología (por ejemplo, que los robots son simples herramientas, que son lógicos y racionales, etc.), y la tecnología trata (como ha mostrado Martin Heidegger en su famoso ensayo sobre tecnología) acerca de nuestra actitud hacia nuestro mundo y la concepción que tenemos de él⁴. Por lo tanto, los «ro-

⁴ Heidegger, *Question concerning Technology*.

bots» evocan todos estos significados y contextos de la tecnología. Son artefactos, pero su desarrollo y uso también requiere de tipos particulares de conocimiento y ciertas capacidades. Además, existe la ciencia de la robótica. El *hardware* del robot es material, pero los robots también tienen componentes inmateriales (código). Están vinculados con muchas actividades que realizan los seres humanos, como el diseño de los robots, así como el uso y la interacción con ellos. Su desarrollo involucra tanto a actores sociales como a las empresas y el Estado. Se desarrollan y se usan en sociedades capitalistas como los Estados Unidos, pero también en otras distintas. Pueden ser parte de la cultura occidental, o su desarrollo puede tener otras raíces culturales. Su desarrollo acepta normalmente un modelo de lo humano (por ejemplo, el humano como máquina) o de humanos particulares (por ejemplo, ancianos infantilizados). A menudo están conectados con una concepción particular del mundo, por ejemplo, la concepción según la cual la naturaleza está ahí para ser explotada, o que los consumidores son una especie de ganado que proporciona información para la economía de datos. Como muestran estos ejemplos, definir a los robots implica ya concepciones particulares, incluyendo concepciones *normativas*, y que estas ya están *haciendo* ética de los robots y «filosofía robótica»: todo gira alrededor de comprender y evaluar a los robots junto con (su relación con) los humanos. De forma más general, el lenguaje que usamos para hablar sobre robots (como desarrolladores, usuarios o, ciertamente, como filósofos) tampoco es neutral. Lo que un robot «es» depende de cómo hablemos *con* y *sobre* el robot⁵.

En segundo lugar, ¿qué es la «ética»? La ética es en sí misma un término controvertido que también puede tener muchos significados. Los filósofos tienden a estar de acuerdo en que tiene que ver con preguntas normativas, particularmente con *qué debemos* hacer y *cómo debemos* vivir. La ética se puede referir a los principios y valores morales, pero también a la rama de la filosofía académica denominada «ética», que discute los principios, teorías y conceptos éticos. Algunos filósofos (por ejemplo, los pragmatistas) ponen más énfasis en la práctica ética y en la experiencia moral que en principios éticos. Dependiendo de la teoría moral usada, la ética se puede centrar en las obligaciones morales, las consecuencias morales, el carácter moral u otros elementos. Además,

⁵ Coeckelbergh, «You, Robot».

la «ética» puede ser entendida como una cuestión sobre límites (para evitar el comportamiento inmoral), pero de forma más positiva, puede formularse como una cuestión sobre la vida buena. Puede tratar sobre lo que es bueno para los individuos, así como lo que es bueno para la sociedad. El término «ética» normalmente se refiere a la ética que concierne a los humanos, sin embargo, también se puede referir a éticas que conciernen a no humanos como animales o, ciertamente, robots. «Ética» puede significar ética dirigida *hacia* los humanos, animales, y así sucesivamente (son el objeto de la ética), o puede significar ética *de* o *para* los humanos y otras entidades (y son, en ese caso, el tema de la ética).

En tercer lugar, la expresión «ética de los robots» se puede referir a la ética de cómo los *humanos* deberíamos usar, interactuar y desarrollar robots de una forma que conduzca al bien para los humanos u otras entidades, tales como animales o, quizás, incluso robots (los humanos son los sujetos éticos; los robots son los medios para alcanzar el fin de la ética humana), o se puede referir a una ética para *robots*. Aquí se usa a menudo el término «ética de las máquinas»; nos referimos en este caso a la ética que puedan tener los robots: los robots son vistos como (potenciales) sujetos éticos. Es importante no confundir estos significados diferentes. La ética de los robots *no* trata necesariamente sobre «dotar a los robots de una ética» o de que «los robots tengan una ética». Por ejemplo, Peter Asaro distingue entre tres significados de ética de los robots: «los sistemas éticos incorporados en los robots, la ética de los que diseñan y usan los robots, y la ética de cómo debemos tratar a los robots»⁶. El primer significado tiene que ver con la ética de los robots en tanto sujetos éticos, como cuando se dice que los coches autónomos deberían tener una ética incorporada. El segundo y tercer significados tienen que ver con la ética de los humanos en tanto sujetos éticos, como cuando se discute si los que diseñan robots deberían ser corresponsables de su uso (por ejemplo, para propósitos comerciales o en una guerra) o se afirma que los robots no deberían ser «torturados».

Hablando de forma descriptiva, la ética tiende, al menos parcialmente, a ser diferente entre culturas y sociedades e, incluso, dentro de las propias sociedades. La gente puede tener diferentes opiniones éticas sobre cuestiones particulares. No todo el mundo está de acuerdo con

⁶ Asaro, «What Should We Want from a Robot Ethic?», 9.

cómo gobernar o dirigir el comportamiento de la propia vida, y lo que es mejor para la sociedad. La ética en este sentido descriptivo ha variado históricamente. Por ejemplo, muchos de nosotros tendemos a pensar que los animales tienen algunos derechos y, de forma más general, un estatus moral más elevado que las cosas. Esta variación cultural e histórica es también relevante para la ética de los robots: las actitudes éticas hacia los robots y las creencias sobre ellos tienden a diferir entre culturas, y pueden variar en el tiempo. Japón se pone a menudo como ejemplo de cómo varían las actitudes hacia los robots. Frecuentemente se dice que la gente en Japón tiene una actitud más positiva hacia los robots debido a su específica historia cultural y popular (con robots representados como ayudantes de los humanos) y la continua influencia de concepciones del mundo en las que los no humanos pueden tener un estatus espiritual⁷. Algunas personas sostienen que en el futuro otorgaremos o deberíamos otorgar derechos a los robots.

Nótese que en este libro usaré el término «ética» y «moralidad» de forma indistinta. Sin embargo, quienquiera que reflexione sobre ética de los robots es libre de distinguir entre estos términos si se muestra que puede ganarse algo con tal distinción. Nótese también que, en términos académicos, el campo de la ética de los robots puede ser visto como una rama de la ética (aplicada) y de la filosofía práctica. Pero se puede relacionar también con el campo de la filosofía de la tecnología. Es, por lo tanto, vista como una de las tecnologías a las que se les puede aplicar conceptos y teorías de la filosofía de la tecnología.

En este libro usaré «ética» y «ética de los robots» en muchos sentidos diferentes. En cada capítulo o argumento haré explícito el tipo de «ética» de la que se está hablando. Por ejemplo, el libro incluirá discusiones sobre si los robots pueden ser agentes morales (sujetos éticos) y pacientes morales (objetos éticos). No solo se tratará de la ética como preocupación por hacer lo correcto, sino también de la ética de la vida y de la sociedad buenas. Estas elecciones y mi tratamiento de estas cuestiones reflejan de forma inevitable mi propia concepción de cómo hacer ética. Por ejemplo, en contraste con muchos libros de texto sobre ética, lo predeterminado no es centrarse en la ética del individuo. El capítulo siguiente se hace preguntas concernientes al impacto sobre la sociedad en lugar de solamente sobre los individuos. Además, en contraste con muchos libros de

⁷ Coeckelbergh, «Robotic Appearances».

ética aplicada, su proyecto no está centrado en la aplicación de una teoría moral normativa. El lector no encontrará capítulos con nombres como «deontología», «consecuencialismo» o «ética de la virtud». En su lugar, el libro toca diversas cuestiones éticas, como la responsabilidad, la dignidad humana y las cuestiones concernientes al estatus moral. Estas discusiones implicarán otros conceptos y teorías, incluyendo, por ejemplo, una teoría sobre la responsabilidad, así como una teoría moral normativa, como el consecuencialismo (capítulo 5) y la ética de la virtud (capítulo 6). Pero el *punto de partida* no es la teoría como tal: son las cuestiones éticas que se plantean desde el uso y la práctica tecnológicos.

No todos los expertos en ética de los robots estarán de acuerdo con este enfoque ni con las definiciones que se han proporcionado. Mi presentación de lo que queremos decir con robots, ética y ética de los robots puede (como todas las definiciones de conceptos filosóficos) ser discutida. Otros pueden poner mayor énfasis sobre la ética individual o las teorías morales. Por ejemplo, se puede encontrar otro tratamiento en la obra de Keith Abney, quien responde a la cuestión de lo que es la ética de los robots discutiendo no solo qué son la moralidad y la ética, sino también qué queremos decir cuando hablamos de derechos y obligaciones morales, cuáles son las principales teorías morales contemporáneas y cómo prestan apoyo la ética de los robots, y qué es una persona⁸. Diré algo más sobre el enfoque de mi libro.

ENFOQUE, ESTRUCTURA Y ALCANCE DEL LIBRO

Mientras que este libro se centra en introducir conceptos y teoría filosófica en el campo de la ética de los robots, también busca hacerlo de una forma que muestre su relevancia para las cuestiones del mundo real. Para esto, liga ciertas aplicaciones robóticas específicas con discusiones filosóficas. Por ejemplo, conecta los coches autónomos con las cuestiones que conciernen a la responsabilidad moral. Así, introduce al lector al primer y más conocido objetivo, así como a una definición de la ética de los robots: contribuir (mediante conceptos y deliberación filosóficos) a la comprensión y al tratamiento de *cuestiones éticas con*

⁸ Abney, «Robotics, Ethical Theory, and Metaethics».

robots en el mundo real. Este es un objetivo importante que puede, y tiene que, continuar siendo adicionalmente promovido mediante investigación inter- y transdisciplinar, implicando en ello diálogos con, por ejemplo, los profesionales de campos como la ingeniería, la política y el derecho. Cada vez más gente, dentro y fuera de la academia, se ve atraída por este emocionante proyecto.

Este énfasis sobre las cuestiones éticas del mundo real con robots actuales y de un futuro próximo tiene como consecuencia el que se ponga menos atención a temas como la superinteligencia (esto es, máquinas con una inteligencia hipotética que excede por mucho la humana), o máquinas que tengan inteligencia artificial general (la inteligencia hipotética de una máquina que puede entender y aprender cualesquiera tareas intelectuales que pueda acometer un humano). En mi opinión, los debates sobre estos temas nos desvían a menudo del tratamiento de los problemas del mundo real y del futuro cercano en el ámbito de la ética de los robots. Dicho esto, reconozco sin reservas que tales discusiones pueden contribuir a el (los) objetivo(s) filosófico(s) de la ética de los robots (ver más abajo). Por lo tanto, en el último capítulo, introduzco el tema de la superinteligencia en el contexto de los enfoques transhumanistas de la ética de los robots.

Además, mi énfasis sobre la ética de los robots en el mundo real, más bien que sobre los robots de ciencia ficción, no implica que crea que la ciencia ficción sea completamente irrelevante para la ética de los robots. Según ya he sugerido, es relevante estudiar cómo se imaginan, y han sido imaginados, los robots en la ciencia ficción, cómo esta imaginación y estas narrativas influyen sobre cómo los robots se perciben y se usan hoy en día, y cuáles son las implicaciones normativas. He contribuido a este tipo de trabajo con mi libro *New Romantic Cyborgs*. La ciencia ficción también ofrece experimentos mentales interesantes para los filósofos. Uso algunos de ellos en este libro, especialmente en mis introducciones a los distintos capítulos, cada uno de los cuales se refieren con frecuencia a películas de ciencia ficción. Y algunas veces pueden extraerse lecciones éticas y políticas de la ciencia ficción. Por ejemplo, como ha mostrado Eileen Hunt Botting, *Frankenstein* puede ser un recurso para pensar sobre la responsabilidad y los derechos⁹. En el contexto de nuestro tema, esto podría significar que la historia nos ayude a

⁹ Botting, *Mary Shelley*.

hacernos preguntas sobre la responsabilidad que conlleva la creación de robots. ¿Es aceptable que la gente que crea y vende robots «abandone» sus creaciones, y deje los desafíos éticos que plantean en manos de los usuarios y su sociedad, o deberían asumir su responsabilidad por tales creaciones? Para muchos expertos contemporáneos de la ética de los robots, esta es una pregunta retórica: creen que los diseñadores y desarrolladores de la tecnología, entre otros, deberían ser responsables de la tecnología (ver también el sumario que presento de la investigación y la innovación responsables en el capítulo 4). Por el contrario, para aquellos que están involucrados en robótica y que aún no se han tomado en serio la ética de los robots, o para aquellas personas procedentes de las humanidades que aún no han conectado con la tecnología contemporánea, reflexionar sobre la responsabilidad de los ingenieros bajo la luz de *Frankenstein* puede ofrecer un buen punto de partida para la reflexión y, así, contribuir al principal objetivo de la ética de los robots: tratar los problemas éticos de la robótica en el mundo real.

Sin embargo, en mi opinión, la ética de los robots debería tener un segundo (pero no secundario) objetivo, que recibe menos atención, pero que quizás nos ofrezca una explicación más profunda de por qué el público más general está tan interesado en este campo, a saber: contribuir a la *filosofía*, y no solo a la reflexión sobre la tecnología, sino, por ejemplo, también a la reflexión sobre lo humano. Reflexionar sobre los robots no es solamente pensar sobre los robots, sino que también toca muchas otras cuestiones que tienen una relevancia filosófica más amplia. Es por esto por lo que a lo largo del libro la ética de los robots está conectada con cuestiones de otros subcampos de la filosofía. Por ejemplo, muestra que la investigación sobre la ética de los robots conduce a cuestiones ontológicas (ver la sección previa sobre lo que es un robot), epistemológicas (por ejemplo, qué es la pericia, y qué tipo de conocimiento tenemos sobre el estatus ontológico y moral de otras entidades) y filosófico-antropológicas (qué son los humanos y qué significa ser humano). El libro muestra que preguntarse sobre los robots es también siempre preguntarse sobre los humanos junto con su moralidad, sus prácticas y sus instituciones: esto se muestra en cada capítulo, y finaliza con una reflexión sobre las relaciones entre los robots y los humanos. Por supuesto, reflexionar sobre los robots nos ayuda a comprender y evaluar mejor la (una particular) tecnología. Como tal, es parte de la filosofía de la tecnología y de la ética de la tecnología. Pero también trata sobre los humanos. Un tema común que recorre este libro es que los robots funcionan como herramientas para compren-

Los robots son espejos
que nos muestran
las partes a menudo
bellas, pero también más
oscuras, de los humanos,
así como de sus
reflexiones y actos
morales.

ernos mejor a nosotros mismos: herramientas usadas por los científicos para comprobar y mejorar sus modelos sobre los seres humanos y otros seres naturales, pero, además, herramientas usadas por los *filósofos* para reflexionar sobre lo que es ser humano. De ahí que el título inicial de este libro fuera *Robotic Mirrors* (Espejos robóticos): usamos los robots como espejos para reflexionar, entre otros objetivos, sobre lo humano y la ética. Los robots son espejos que nos muestran las partes a menudo bellas, pero también más oscuras, de los humanos, así como de sus reflexiones y actos morales.

La ética de los robots es, entonces, parte de la ética (aplicada) y de la filosofía de la tecnología, pero también puede enmarcarse como parte de una «filosofía robótica» más amplia que tenga dimensiones tanto prácticas como teóricas, e incluya temas tales como la epistemología, la metafísica y la filosofía política y, al final y en su máxima expresión, es filosofía *tout court*, como cuando nos hace reflexionar sobre la naturaleza del mundo o nos ofrece formas de reflexionar sobre lo que significa ser humano. Esto, en cambio, implica que no existe una forma «neutral» de hacer o presentar la ética de los robots: cómo se haga y se presente depende siempre del marco conceptual y del enfoque filosófico de cada uno. A lo largo del libro, indicaré a veces tales direcciones y trasfondos.

Finalmente, las cuestiones éticas y los problemas filosóficos discutidos en este libro son solo una selección de ellos y, a menudo, son tratados en más detalle en otras partes o están esperando que haya más investigación en este campo. Por ejemplo, a pesar de que me refiero a contextos y culturas no occidentales en los capítulos 3-4 y 7-8 (por ejemplo, los robots en Japón), se puede decir más sobre la ética de los robots a la luz de las diferencias y desafíos culturales que plantea una ética global de los robots.

Y la cuestión de los problemas medioambientales planteados por los robots, que se han mencionado anteriormente, figuran en varios capítulos, pero se necesita trabajar más en esta área. También he decidido limitar el alcance de este libro a la ética de los robots de una forma que incluya referencia a la IA y *algunas* de las cuestiones éticas que plantea, pero no ofrecer una exploración extensa y completa; se pueden encontrar libros completos que traten este tema en otras partes¹⁰. Además,

¹⁰ Boddington, *Towards a Code of Ethics*; Coeckelbergh, *Ética de la Inteligencia Artificial*; Gunkel, *An Introduction*; Liao, *Ethics*; Bartneck et al., *An Introduction*.

aunque la ley y la regulación son una forma de responder a las cuestiones éticas que aquí se plantean, no describo por completo los aspectos legales de la robótica ni introduzco discusiones sobre la reglamentación; el lector encontrará otros puntos de vista interesantes al respecto en otras partes¹¹. El libro tampoco pretende ser una introducción a la ética, o a cómo funcionan los robots. Especialmente para colegas que se ocupen de este campo, es bueno tener en cuenta que este libro es una *introducción*. Doy referencias para los lectores que quieran ir más allá en cualquiera de las cuestiones, y se están publicando cada vez más trabajos de forma continua. Lo mismo se aplica a los profesores que usen esta publicación como libro de texto y quieran ampliarlo con el material aquí proporcionado. A pesar de que el libro se lee como un trabajo independiente, recomiendo usarlo en combinación con parte de la bibliografía citada más importante. Por ejemplo, cada capítulo puede combinarse con uno o más artículos clave o capítulos de libro que figuran en la bibliografía correspondiente. También hay colecciones excelentes de artículos y actas de congresos importantes, tales como los contenidos en los simposios sobre *Robophilosophy* (robofilosofía)¹². Se invita a los profesores a añadir clases o sesiones extra basadas en su propio trabajo o en intereses específicos en este campo. El tema también se presta perfectamente al uso adicional de medios de comunicación distintos del texto, tales como películas, (imágenes de) obras de arte, ¡y (por qué no) robots!

Permítaseme hacer una breve descripción de los capítulos y temas que se abordan.

El capítulo 2 plantea problemas sobre los robots industriales y de servicio que se vuelven más inteligentes y relevan a los humanos de ciertas tareas. ¿Cuáles son las consecuencias para nuestras economías? ¿Cuáles son las implicaciones para los seres humanos? ¿Conducirá esto a un desempleo masivo y a nuevas formas de explotación? ¿Cuál es el significado del trabajo? ¿Es aceptable que los consumidores tengan que tratar cada vez más con máquinas en lugar de humanos? ¿La interacción humano-a-humano será solo algo de lo que puedan disfrutar los ricos? ¿Cómo transformarán nuestras sociedades la robótica y la automatización?

¹¹ Ver, por ejemplo, Hildebrandt, *Smart Technologies*; Turner, *Robot Rules*; Fosch-Villaronga, *Robots*.

¹² Lin, Bekey y Abney, *Robot Ethics*; Lin, Abney y Jenkins, *Robot Ethics 2.0*; Coeckelbergh *et al.*, *Envisioning Robots*.

¿T IENEN agencia moral los robots? ¿Pueden ser considerados responsables de sus acciones? ¿Nos reemplazarán en el trabajo? ¿Les debemos algo los seres humanos? Estas son algunas de las cuestiones de carácter ético y moral a las que tenemos que enfrentarnos en un contexto en el que el uso de robots y otros dispositivos inteligentes cada vez más sofisticados se ha vuelto habitual. En este libro, planteado como una guía ante este escenario inédito, Mark Coeckelbergh aborda directamente estas preguntas, para lo cual tiene en cuenta una gran variedad de dispositivos —desde el robot de compañía hasta el dron de combate—, sosteniendo en todo momento una premisa esencial: los problemas de la robótica no atañen solo a las máquinas, sino a la esencia misma de los seres humanos.



colección TEOREMA
serie mayor

0112139

ISBN 978-84-376-4706-7



9 788437 647067