

“Ética y robótica”

Adriano Fabris, Sergio Bartolommei, Edoardo Datteri¹

1. Cuestiones previas sobre el concepto de ética.

Para responder adecuadamente a la cuestión referida en el título, es necesario aclarar antes de nada los diferentes significados que el término “ética” puede asumir dentro del debate moderno. Surgen de este término, si se mira con atención, diferentes significados según el nivel elegido y la perspectiva adoptada. Intentamos, seguidamente realizar una sintética (y necesariamente esquemática) enumeración de los aspectos y de las principales cuestiones en juego. Aquí tenemos las diferentes perspectivas a partir de las cuales es posible profundizar el concepto de ética:

1. *Ética como institución.* Desde un punto de vista institucional (como universo social de prácticas y de enunciados distintos de otros) la ética es una estructura normativa caracterizada por la *prescribibilidad*: un conjunto de comandos, prohibiciones, mandatos, etc. que empujan a las personas a actuar en un cierto modo o a probar determinados sentimientos. La finalidad de la ética como institución social normativa es realizar vínculos a los deseos de las personas para garantizar el nivel de cooperación social necesario a la supervivencia, el bienestar o la autorrealización de los miembros de la sociedad (los fines cambian con la variación de las concepciones del valor y de la buena vida sostenidas por las éticas normativas singulares: ver punto 7.2).

2. *Ética y derecho.* Esta finalidad general de la ética se acerca a la otra gran institución social normativa, el derecho. De esta última se diferencia, al menos por los siguientes motivos: el derecho es válido para los ciudadanos de un determinado territorio, prevé una pena física para el transgresor, en el derecho solo el juez está autorizado a aplicar la ley y emitir la sentencia; la norma jurídica es respetada por los ciudadanos por miedo a

¹.- Traducido del italiano por Daniel Buraschi Bresadola

las sanciones. Al contrario, la moral vale (o pretende valer) para cualquier ciudad, es seguida por una convicción personal, la sanción consiste o en el remordimiento o en la desaprobación social, más que en la coacción física o en la sanción pecuniaria. En la moral todos – y no solo los expertos de la moral - pueden aprobar o desaprobar.

3. *Ética y moral*. Son ambos términos actuales, a diferencia del pasado, son utilizados como sinónimos (ethos y mores designan en efecto un mismo objeto: el hábito y la costumbre), aunque sí que el uso del *primer* término puede prevaler cuando se aluda a aspectos *formales* o *teóricos* de la disciplina (se habla de teorías éticas y de métodos de la ética), mientras que el *segundo* se utiliza con particular referencia a los *aspectos de contenidos* (normas y principios morales sustantivos). Se puede también decir que mientras “moral” designa (más propiamente) la costumbre, el hábito, la ética, es todo lo que concierne el *estudio de la moralidad*. Este último ámbito de investigación parece ser competencia del filósofo *de la moral*, una figura que hay que diferenciar netamente del *moralista*, que se ocupa de la formulación de juicios de valor con fines pedagógicos y edificantes.

4. *Los términos opuestos a moral*. Una expresión como “lo que estás haciendo es moral” puede ser interpretada de dos formas según se este *aprobando* lo que se ha hecho (y en este caso “moral” es opuesto a “inmoral”), o se esté sencillamente afirmando que lo que se está haciendo *pertenece al ámbito* de la moralidad (en este caso moral es opuesto a “no – moral” o “extra – moral”). Mientras para aclarar las razones (plausibles o no) de un juicio de “inmoralidad” será suficiente explicitar las razones que sostienen el sistema de creencias de quien lo pronuncia, no parece posible individualizar en modo exacto un criterio unívoco para delimitar la moral de la no–moral. De hecho, cada gran familia de la ética normativa ha propuesto criterios diferentes, y probablemente toda la historia de la ética podría ser reconstruida a partir de estos criterios.

5. *Niveles de complejidad reflexiva interna*. La ética (o moral) puede diferenciarse en base al grado de reflexividad que caracteriza algunos de sus actos. Es posible distinguir al menos dos dimensiones de la ética: una ética del sentido común y una ética crítica. La

primera es el conjunto de las opiniones recibidas, interiorizadas en la infancia antes de ser sometidas a una evaluación crítica y racional. La ética del sentido común guía nuestra vida moral cotidiana y se puede paragonar a nuestra lengua madre. La ética crítica, al contrario, surge en momentos particularmente problemáticos conectados a la evolución de las costumbres colectivas o a ciertos eventos existenciales que sacuden la vida de las personas individuales. Son los momentos, como aquellos determinados por el desarrollo de las tecnologías biomédicas y de la necesidad de volver a reflexionar sobre los valores dados por obvios – en los cuales las certidumbres del sentido común vacilan y el juicio tiene que suspenderse. La ética crítica busca las “razones de los argumentos” – nunca analizados antes – que sostienen las opiniones recibidas para ver si tales razones puedan todavía ser sostenidas. A su vez, las razones pueden subdividirse en dos grandes categorías: empíricas (por ejemplo argumentos de la evolución de la ciencia) y normativas (distinguiendo entre razones deontológicas y razones teológicas o consecuenciales, ver punto 7.2). En el caso en el cual las opiniones recibidas no superen el test de la justificación racional las mismas deberán ser corregidas, perfeccionadas o abandonadas a la luz de las mejores razones alternativas disponibles.

6. *Universalidad del juicio moral*: sobre la cuestión de la justificación de los juicios morales podemos ir más allá y afirmar que no parece apropiado limitar la justificación de un juicio moral al solo autointerés o se tendrá que demostrar que aquello que es hecho por autointerés se concilia con principios más generales y completos. La conducta moral debe resultar justificable -como sostienen los filósofos pertenecientes a tradiciones morales diversas y alternativas- *desde un punto de vista universal*. En esta expresión está condensado tanto el sentido de la “regola aurea” (“no hacer a otros lo que no te gustaría que te hicieran a ti”, o “ama al prójimo como a ti mismo”), aquel que invita a “ponerse en lugar del otro”: la justificación de un principio ético no puede ser satisfecha en términos de intereses parciales y sectoriales (egoístas, étnicos, raciales, sexuales, de especie, etc...). Desde un punto de vista moral es irrelevante que el beneficiario de una cierta acción sea yo, tú, nuestra raza (o etnia, o especie), y desventurados los otros y quienes a nuestro grupo no pertenecen. Lo que cuenta en la moral es propiamente la posibilidad de *trascender intereses de grupos y facciones*

particulares para ganar un punto de vista general -el punto de vista del “observador imparcial” del cual han hablado, mencionándolo, teorías morales diversas y, al menos en el plano normativo, también entre ellos incompatibles.

7. *Articulación conceptual interna de la ética contemporánea*: se suele articular conceptualmente la ética contemporánea en cuatro ámbitos generales, que para algunos son muy diferentes, y para otros son superpuestas: ética descriptiva, meta-ética, ética normativa, ética aplicada.

7.1 *Por ética descriptiva (ED)* se entiende, en general, la ética en cuanto relevaciones de las doctrinas morales empíricamente difusas o estudio de los factores y de los movimientos (empíricos) a los cuales la ética puede ser reconducida. La locución es susceptible de dos interpretaciones distintas. En una primera interpretación la E.D. es en relación a la reconocimiento de los sistemas de creencias morales divididos de hecho de las personas y en base a las cuales las personas conducen la propia vida; en este primer significado la ética debería *limitarse a estudiar la moralidad en sus formas históricas concretas* y en su funcionamiento social normal. En cuanto al segundo significado de la E.D. designa un verdadero “programa de investigación” en la filosofía moral (de los cuales Hobbes y Espinoza son los principales exponentes), caracterizado por la intención de reducir la ética a los modos de proceder de la ciencia, como es el caso de las matemáticas, la geometría o las ciencias naturales, en este segundo significado, la locución E.D. envía a una *bien precisa forma de ética*, inspirada en el objetivo de restar validez a las pretensiones de autonomía de la ética reconduciéndola a la metodología en uso en otros órdenes del discurso para poner fin, en virtud de una ética *more geométrico* demostrada, a las disputas interminables que parecen ir en contra de su concreta distinción.

7.2 *La ética normativa* es el complejo de las teorías éticas sustantivas que se contienen en el campo en los debates morales concretos (personalismo ontológico-metafísico, teoría de la ley natural, personalismo kantiano, utilitarismo y neo-utilitarismo, neo-contractualismo, teoría de los derechos, ética de la virtud, etc.). No existe un solo modo, o un modo unánimemente compartido, de clasificar las teorías éticas. Uno de los

criterios más difusamente seguidos es el utilizado en 1.874 por Henry Sidgwick y consiste en distinguir entre éticas *teológicas o consecuencialistas* (denominadas también *éticas del bien*) de una parte y de otra éticas *deontológicas* (o *éticas del deber*). Las primeras individualizan una cierta idea de “bien” (cuyo contenido variará en base a la teoría de valor compartida), la noción central de la ética y, sobretudo, tienden a valorar como buena o mala las acciones en base a los efectos o consecuencias que las acciones mismas producen, independientemente de los principios o de las intenciones que sigue la acción o en las cuales está inspirada; y viceversa, las éticas deontológicas tienden a fijar previamente una norma, ley o propiedad de la acción que establece aquello que (en modo condicional o *prima facie*) es justo o injusto hacer, y las acciones son juzgadas en modo secundario en relación a la cualidad intrínseca de la acción misma, o de la conformidad o no de la acción a la norma prefijada, independientemente de las consecuencias producidas o previsibles. Expresión emblemática de las éticas del primer tipo es el utilitarismo para el cual la bondad de las consecuencias (o “estado final de las cosas”), es el único factor moralmente relevante; representativa del segundo grupo es, en cambio, la ética kantiana para la cual aquello que cuenta es la conformidad de la acción, un cierto tipo de deber, independientemente de sus consecuencias.

8. *La ética aplicada*: a partir de la segunda mitad de los años 60 del pasado siglo se empezó a manifestar una profunda insatisfacción en relación al modo tradicional de conducir el trabajo filosófico en relación a la moral. Surge en particular una mayor sensibilidad por los problemas morales de orden concreto (problemas ecológicos o conexos a la paz o a la guerra o a la relación entre el norte y el sur, problemas ligados al desarrollo de la biomedicina o de las telecomunicaciones, problemas del mundo de los negocios, etc...). Nace así el desafío de la “ética aplicada”, la cual solicita la reflexión filosófica de redimensionar la importancia de las investigaciones lógico-lingüísticas o las investigaciones de los principios normativos abstractos y generales, para ocuparse también o sobre todo de los dilemas concretos y de casos específicos. *La bioética, la ética ambiental, la ética animalista, la ética de los negocios, la ética de la comunicación, la tecno- y la robo-ética, la ética de la biocultura* son todas ramas de la ética aplicada. Esto significa que son ámbitos *especiales* de la reflexión ética *normativa*, susceptibles a su vez de ser interpretados, o como *aplicaciones* de teorías

sistemáticas más comprensivas, o como *ámbitos autónomos* de la investigación moral. Desde el punto de vista meta-ético, un aspecto que engloba algunas de estas nuevas éticas (o estos nuevos sectores de la reflexión moral tradicional) es la tentativa de extender la consideración moral más allá de los límites convencionales en los cuales hasta ahora estaba circunscrita: las relaciones de los seres humanos entre ellos y de cada uno consigo mismo. Por “confines de la consideración moral” se entienda un conjunto de tutelas instituidas en garantía del trato de quien este incluido dentro del “área protegida” de la “ciudad moral”, las nuevas éticas se preguntan respecto a la posibilidad de incluir en la ciudad también, reales entes no-humanos, sean éstos (al menos en parte) no producidos por los seres humanos (ética animalista o ética ambiental), o producidos por los hombres (ética de la biocultura, robo-ética). En este ámbito una de las direcciones más interesantes de la reflexión moral concierne la reflexión si, una vez individuados estos nuevos sujetos posibles de consideración moral, a éstos deban ser atribuidos auténticos *derechos* (que reclaman una serie de deberes directos por parte de los seres humanos hacia los “recién llegados”), o si es suficiente alargar la clase de los deberes tradicionales de los derechos humanos hacia sí mismos en modo de que ciertas entidades no humanas se conviertan al menos *indirectamente* objeto del cuidado y el respeto de las personas tradicionalmente interesadas. Otros sectores de la ética aplicada (entre los cuales encontramos la tecno-ética) están en cambio dirigidos a reflexionar sobre los problemas morales que surgen derivados de las nuevas posibilidades (biológicas o no biológicas) de la modificación de la corporeidad y de la mente humana, ya sea con fines terapéuticos o no (mejoradores o de potenciamiento). En este punto debemos preguntarnos sobre la licitud o no de las modificaciones que pueden producirse en relación a la creación de *prótesis artificiales* con funciones asimilables a las ya tradicionales funciones humana, también en relación a la construcción de *nuevas modalidades o configuraciones de los seres humanos*, ya sea *incrementando* facultades y funciones en relación a las existentes en su materia física, psíquica o biológica, o creando nuevas y nunca experimentadas funciones, más allá de los límites actuales (biológicos y no biológicos) de la “naturaleza humana”.

2. La responsabilidad en el ámbito de las relaciones ser humano-robot.

En el contexto de este amplio debate el grupo de investigación de la Universidad de Pisa, Departamento de Filosofía, ha llevado a cabo sus propias investigaciones siguiendo tres principales direcciones, coligadas a tres cuestiones de fondo: 1. Un análisis en profundidad de los conceptos principales que conciernen a las investigación tecno-ética; 2. Un reconocimiento de algunas de las consecuencias que se deducen sobre un plano moral con la aplicación de las nuevas tecnologías; 3. Una específica discusión sobre los problemas morales que tienen que ver con las relaciones que se instauran entre biotecnología y bioética.

En este punto tenemos intención sobre todo de hablar del primer orden de cuestiones. En particular, queremos discutir el concepto de “responsabilidad”, aplicado no solo a la relación entre hombre y máquina, considerado en general, sino también y más específicamente, a la relación entre ser humano y robot. En este último contexto, de hecho, pueden surgir problemas notorios concernientes a la responsabilidad. Podemos preguntarnos, de hecho: quién es propiamente responsable de una eventual mala acción realizada por un robot?. ¿El robot o su constructor?. ¿El constructor o el programador?.

Surgen en este punto, sin lugar a dudas, claras similitudes con lo debatido, y que actualmente se debate, en el ámbito de la investigación teológica.

Es suficiente pensar en el libro que Robert Wiener ha dedicado a la figura del Golem², o al título de un libro reciente de estudios religiosos: *Is God a Programmer?*³ Hoy, en definitiva, los programadores somos nosotros. Hoy nosotros somos los creadores, y los robots son nuestros Golem. Pero, ¿somos realmente como Dios?, y más aún, ¿cuál es la relación que podemos tener con nuestras criaturas? (o, mejor dicho, con nuestras creaciones), ¿qué responsabilidad tenemos en estos casos?, y ¿qué significa, en estos casos, tiene la responsabilidad?.

² N.Wiener, *God, Golem, Inc.*, The MIT Press, Cambridge (Mass.) 1964.

³ G.L.Simons, *Is God a Programmer? Religion in the Computer Age*, Harvester Press, Sussez, 1988.

Podemos empezar intentando contesta a esta última pregunta. La noción de “responsabilidad” tiene, al menos, dos significados.

1. El primero está relacionado con la posibilidad de considerar un agente como la *causa* de una acción (y de sus efectos). Los agentes son responsables de determinadas situaciones o estados de cosas en la medida en la cual han contribuido a provocar tales situaciones o estados de cosas.

2. Desde otra perspectiva, al contrario, la responsabilidad tiene que tener en cuenta las *intenciones* y las *motivaciones* que llevan a un agente a realizar ciertas acciones. Según este significado, entonces, un agente es responsable de una situación o de un estado de cosas en la medida en la cual su conducta ha sido regulada en base a aquellas razones o aquellos principios que guían las acciones que ha llevado a cabo.

Las concepciones *conceptualistas o teleológicas de la ética* (por ejemplo el *utilitarismo*) se basan en el primero significado de “responsabilidad”. Las acciones son evaluadas en base a un plan molar (considerándolas justas o equivocadas) basándose en sus consecuencias y sus resultados. La *ética deontológica* (como por ejemplo la *ética kantiana*) se basa, al contrario, en el segundo significado. La evaluación moral de las acciones se cumple considerando la conformidad de las acciones a ciertas leyes y a ciertos principios predefinidos, y teniendo en cuenta las intenciones del agente.

Hay que observar que muchos debates de hoy en día sobre el concepto de “responsabilidad moral” ponen el acento sobre *el número y sobre la identidad de quién respecto a los cuales yo puedo sentirme responsable*. Hasta hace no mucho tiempo estos sujetos podían ser solamente nuestros similares, nosotros mismos, eventualmente Dios. Hoy emergen nuevos sujetos, según los criterios que se utilizan para identificar los que son dignos de respeto. La ética aplicada contemporánea se ha comprometido específicamente a definir estos “otros” dignos de respeto (y esto puede incluir las generaciones futuras, los animales no humanos, el ambiente natural, los robots y los sistemas de inteligencia artificial)⁴.

⁴ Cfr. El fascículo 2004/1 de “teoría” dedicado al tema “Responsabilidad en las éticas aplicadas”.

3. **Ética de la responsabilidad: las cuestiones de fondo.**

Existen, resumiendo, *dos tipos de responsabilidad*. Hay, de un lado, la responsabilidad del sujeto que empieza un proceso, tanto que lo haga por su propia iniciativa, como que sea introducido en un contexto de determinaciones causales; hay, del otro lado, la responsabilidad de quien libremente asume algo que depende de él, ya sea que lo reconozca como vinculante para su hacer, o se considere a si mismo como responsable de una relación con este mismo principio de su hacer. En el primer caso, se habla propiamente, en italiano, de “responder de”, en el segundo de “responder a”.

¿En qué manera es posible pensar la robo-ética como una ética de la responsabilidad?. ¿En qué modo la ética de la responsabilidad puede ser el *background* de los diferentes sentidos de “ética” que hemos analizado anteriormente?. En que manera, entonces, tiene que ser entendida la responsabilidad en la relación hombre – máquina y, más específicamente, en la relación entre hombre y robot?. A nuestra manera de ver esto puede ser realizado siguiendo estos pasos:

1) Hace falta antes de todo evitar de pensar que las máquinas sean simples instrumentos, medios que pueden ser utilizados bien o mal. En este caso la ética sería *primera e independiente* respecto al uso de las máquinas. Las cosas no son tan sencillas. La utilización de las máquinas, de hecho, transforma el hombre. Existe una verdadera mutación antropológica en este uso, aunque si la máquina no es incorporada en el hombre mismo. Es suficiente pensar a lo que conlleva el uso del teléfono celular.

2) Esto sin embargo no significa que somos exentos de nuestra responsabilidad. a) No somos desresponsabilizados respecto al contexto tecnológico comprensivo o en el cual estamos inmersos (el mundo de la técnica, del progreso científico- tecnológico, de las telecomunicaciones: que se mueve según leyes ya autónomas, sobre las cuales el hombre no tiene capacidad de incidir). De hecho, podemos incidir al menos localmente, en nuestro campo de acción, sobre los efectos de esta tecnología global; por ejemplo utilizando o no determinados instrumentos. b) Además no somos de – responsabilizados ni siquiera respecto a las consecuencias de las actividades de máquinas que utilizamos,

pero que no hemos producido o programados nosotros. Es fácil pensar que haya una precisa responsabilidad por parte del constructor o del programador: y esta responsabilidad es una responsabilidad *directa*. Pero hay también una responsabilidad *indirecta*: se trata de la responsabilidad que es propia de quién acepta de utilizar, aunque simplemente, las máquinas que no ha construido ni programado. La utilización simple tiene sus reglas, que el sujeto lo acepta y del cual es responsable (como cuando aceptamos de suscribir las condiciones de un software descargado gratuitamente). Es necesario, entonces, distinguir la responsabilidad de la construcción, la de programación y la del uso. c) Por su parte, en el caso de máquinas dotadas de autorregulación, en el caso de máquina que pueden cumplir lo que, en analogía con los comportamientos humanos, podemos llamar “elecciones” – pensamos por ejemplo a los *learning robots*, aunque en estos casos hay una responsabilidad *indirecta*. Aclaremos, sin embargo, que cosa ha cambiado desde un punto de vista ético con la construcción de *learning robots*. Hemos pasados, en la relación hombre-robot, del modelo “*master-slave*”, al modelo “*companion-companion*”. Esto ha reducido ciertamente la capacidad de control del hombre y, en consecuencia, su responsabilidad *directa*. Pero no ha eliminado su responsabilidad en una forma *indirecta*. Existen, también en este caso, dos tipos de responsabilidades: 1.- La responsabilidad indirecta del constructor y del programador (que tienen una responsabilidad *directa* respecto a las condiciones de funcionamiento de la máquina capaz de autoregularse y, en parte, del ambiente en el cual el robot aprende, pero indirecta respecto a sus “elecciones”. Se trata de hecho de responsabilidades de estos sujetos humanos, en la medida en la cual las “elecciones” del robot capaz de aprender se desarrollan siempre en el ámbito determinado del software y del hardware que han sido inicialmente predispuestos por el proyectista y el programador). 2. Existe después otra acepción, la responsabilidad, siempre indirecta, del utilizador (que acepta, nuevamente, las reglas de la utilización o de un *learning robot* y juega con estas reglas de juego).

3) Existen diferentes sentidos del término “responsabilidad”, pero el hombre, como sujeto moral, no puede librarse totalmente de esta responsabilidad. Justamente porque es un sujeto moral: o sea, resulta responsable.

4) Este discurso, si se quiere, tiene también una interesante aplicación teológica. Si Dios es pensado, como pasa en muchas religiones, especialmente de tipo monoteísta, como creador (y la teología filosófica lo ha pensado como constructor, relojero, mecánico, programador, etc.), entonces Dios *no puede no ser responsable*, al menos de forma *indirecta*, del funcionamiento de su producto. La capacidad de autodeterminación de este producto lo puede salvar de hecho de la responsabilidad *directa* respecto a las consecuencias, pero no de las *indirectas*. Hay que ver si de todos modos esta concepción no sea demasiado antropocéntrica, si no aplica a lo divino categorías humanas, demasiado humanas. Preguntándonos: ¿Dios está sujeto a la moral propia en el sentido que viene pensado por el hombre?, o tenemos que pensar en una específica moral robótica, que puede eventualmente ser elaborada autónomamente, por pruebas y errores, por ejemplo por un *learning robot*?. Pero también de esta moral tendrían responsabilidad indirecta los constructores y los programadores de los robots.

5) Una última cuestión, más filosófica si cabe habría que realizar: ¿Por qué hay que hablar todavía de responsabilidad?, ¿porqué tenemos el sentimiento de deber asumir la responsabilidad en respecto a algo?, ¿tiene algún sentido hacerlo?. Para responder a esta cuestión no es suficiente hacer referencia a la responsabilidad, directa o indirecta respecto a las consecuencias (lo que hemos hecho hasta ahora; lo que hemos llamado el “ser responsable de”). Hay que hacer intervenir, además, la referencia a determinados principios y necesitamos motivar su asunción. Hay que pasar de *responder de* algo a la conciencia que hay que *responder a* los principios que cada uno considera puntos de orientación y condición de las elecciones de su propia vida. Cada uno de nosotros: también el investigador que construye y programa un robot.

En conclusión: no se trata de ser optimistas, ni pesimistas en relación a los progresos de las tecnologías y de sus aplicaciones. Se trata, solo, repito, de ser responsables. O sea, de ser conscientes que el hombre es un sujeto ético, en todos los significados que esta expresión puede asumir y que hemos analizado anteriormente.