

Prolegómenos a una ética para la robótica social*

Prolegomena to an Ethics for Social Robotics

Júlia Pareto Boada

Institut de Robòtica i Informàtica Industrial (IRI),
CSIC-UPC
jpareto@iri.upc.edu

ISSN 1989-7022

RESUMEN: La robótica social presenta un elevado potencial disruptivo, al expandir el ámbito de aplicación de la tecnología inteligente a contextos prácticos de naturaleza relacional. Por su capacidad de interactuar con las personas "intersubjetivamente", los robots sociales pueden asumir nuevos roles en nuestras actividades cotidianas, multiplicando las implicaciones éticas de la robótica inteligente. En este artículo ofrecemos algunas consideraciones preliminares para la reflexión ética sobre la robótica social, para clarificar cómo orientar acertadamente el pensar crítico-normativo en esta ardua tarea. Defendemos la 'perspectiva del ser' y sus categorías vinculadas de 'teleología' e 'interés' como aquellas desde las que articular la reflexión. Argumentamos que atender primariamente al 'ser' de los robots, antes que a su 'hacer', permite acercarnos correctamente al fenómeno nuclear de toda preocupación ética sobre la robótica social (la externalización de nuestra agencia en estas entidades), evitando los reduccionismos en la mirada ética a que éste puede conducir.

PALABRAS CLAVE: agencia, ética, inteligencia artificial, interacción humano-robot, robótica social

ABSTRACT: Social robotics has a high disruptive potential, for it expands the field of application of intelligent technology to practical contexts of a relational nature. Due to their capacity to "intersubjectively" interact with people, social robots can take over new roles in our daily activities, multiplying the ethical implications of intelligent robotics. In this paper, we offer some preliminary considerations for the ethical reflection on social robotics, so that to clarify how to correctly orient the critical-normative thinking in this arduous task. We defend the 'being perspective' and its linked categories of 'teleology' and 'interest' as the ones from which to articulate the reflection. We argue that attending primarily to the robots' 'being', before their 'doing', allows us to correctly approach the core phenomenon of all ethical concerns on social robotics (the outsourcing of our agency in these entities), avoiding the reductionisms in the ethical gaze to which the latter may lead.

KEYWORDS: agency, artificial intelligence, ethics, human-robot interaction, social robotics

1. Ante la fuerza disruptiva de la robótica social

El despliegue de sistemas de inteligencia artificial en el ámbito de la robótica ha multiplicado, en las últimas décadas, el potencial disruptivo de este campo tecnocientífico. La "encarnación" de inteligencia artificial en dispositivos robóticos da lugar a "agentes artificiales" capaces de llevar a cabo, con cierto grado de autonomía tecnológica (Funk y Coeckelbergh, 2020), acciones¹ con objetivos específicos en entornos reales con independencia de control externo. Podemos encomendarles no sólo tareas automatizables, sino conductas que requieren de un previo procesamiento cognitivo de información sobre el entorno y de elecciones a partir de ella, siempre respecto a un fin predeterminado.

Con ello la robótica inteligente altera radicalmente el abanico de tareas y roles que pueden asumir los robots, extendiendo sus posibles ámbitos de aplicación a contextos prácticos que hasta ahora quedaban reservados a la agencia humana. En este sentido hablamos de disrupción, pues la expansión del potencial robótico a tareas que exigen una interacción y adaptabilidad al entorno nos sitúa ante una posible transformación de las actividades humanas, tanto a nivel práctico-estructural como conceptual. A su

* Este artículo ha sido posible gracias a la financiación recibida por parte del Ministerio de Ciencia e Innovación en el marco de las Ayudas para contratos predoctorales para la formación de doctores - FPI (PRE2018-084286), y por la Agencia Estatal de Investigación a través del Sello de Excelencia María de Maeztu del Institut de Robòtica i Informàtica Industrial (IRI), CSIC-UPC (MDM-2016-0656-18-2).

Quiero expresar mi gratitud a la Dra. Begoña Román, quien, con su generosidad intelectual, su entusiasmo profesional y su incansable predisposición al diálogo, ha contribuido inestimablemente a la publicación de este artículo.



Received: 01/12/2020
Accepted: 08/01/2021

vez, se ensancha el conjunto de implicaciones éticas de la robótica inteligente y, por tanto, el alcance de la reflexión normativa necesaria para que la introducción de sus artefactos en nuestra existencia se halle en consonancia con los valores morales y derechos humanos. A esto responde, precisamente, la aparición de nuevas disciplinas como la 'Roboética' (Operto y Veruggio, 2008), la 'Ética para Máquinas' (*machine ethics*) y la todavía más incipiente 'Robo-filosofía' (Seibt, 2017).

Uno de los grandes focos de disrupción de la robótica inteligente se encuentra en los "robots sociales", destinados a servir en nuestra cotidianidad de forma radicalmente nueva, a saber, como entidades capaces de entrar en nuestra dinámica humana de interacción social, no sólo física sino también semánticamente (Rodogno, 2016). Esto les confiere un cierto estatuto de agentes sociales que parece situarlos en una ontología resbaladiza, por cuanto presentan características de 'casi-sujetos'. Así, los robots sociales, al asumir un rol en prácticas socio-culturales, irrumpen en nuestra forma de concebir y organizar la vida en comunidad. A esta clase de sistemas inteligentes autónomos pertenecerían, por ejemplo, los robots personales en el ámbito doméstico para la asistencia en situaciones de vulnerabilidad, los robots para los cuidados en el sector sociosanitario, los robots asistenciales para contextos terapéuticos de rehabilitación cognitiva, o los robots de apoyo a la educación.

La fuerza disruptiva de estas inteligencias artificiales "socialmente situadas" ubica la rama de la robótica social bajo un fuerte escrutinio ético. Esta actividad tecnocientífica y sus productos desatan crecientes preocupaciones sobre un sinfín de cuestiones relativas a la dignidad humana, la autonomía personal, la libertad, la privacidad, la responsabilidad, la devaluación de prácticas humanas como los cuidados, la degeneración de nuestras facultades morales, la justicia distributiva, etc. Así, en la actualidad, la explosión de debates y reflexiones sobre las múltiples implicaciones de la robótica social a diferentes niveles (individual, interpersonal, socio-estructural, profesional, etc.) responde al reto colectivo de orientar el desarrollo, implementación y uso de estas aplicaciones de la inteligencia artificial (Nørskov et al., 2020; Coeckelbergh et al., 2018; Seibt et al., 2016).

El objetivo del presente artículo es contribuir a esta deliberación crítico-normativa mediante una serie de consideraciones preliminares a la reflexión ética sobre la robótica social. De ese modo, no nos ocuparemos de discutir un problema concreto de la robótica social, sino que, en un estadio previo, señalaremos algunos rasgos fundamentales de la perspectiva desde la cual dirigir la mirada ética. Se pretende, en definitiva, establecer ciertas bases que ayuden a clarificar la compleja cuestión de cómo pensar la robótica social desde el punto de vista de la ética, asentando algunos fundamentos para poder desvelar los retos relevantes que plantea a nivel ético-normativo. Y es que, como advierte Žižek (2013), para encontrar respuestas debemos formular antes la pregunta correcta.

Con vistas a este propósito, se empezará abordando, en el segundo apartado, el controvertido concepto de "robot social", para acotar, a su vez, la rama de actividad tecnocientífica en la que nos centraremos. En el tercer apartado expondremos la complejidad que supone reflexionar en torno a la robótica social desde una perspectiva ética. En el cuarto apartado, se identificará el fenómeno nuclear de las preocupaciones éticas sobre la robótica social (que designamos como 'externalización de la agencia'), y se advertirá del posible reduccionismo en la mirada ética a que éste puede conducir. De este modo, se revelará lo que consideramos una aproximación desenfocada, articulada primariamente en el 'hacer' del robot. En el quinto

apartado se argumentará a favor de una perspectiva ética centrada prioritariamente en la cuestión del 'ser'. Se desvelarán dos categorías (teleología e interés) clave para la reflexión, que parecen haber pasado inadvertidas en el marco de discusión actual sobre la robótica inteligente.

2. Sobre la noción de robot social

Como evidencian Sarrica et al. (2020) en su revisión sistemática de literatura científica y popular, no existe un consenso en la definición de "robot social", si bien hay algunas conceptualizaciones de referencia. Su recopilación de definiciones nos permite concluir que estas difieren según los rasgos en que se ponga el énfasis (función, objetivo, aspecto, capacidades del robot, etc.).

Para los propósitos de este artículo, consideramos pertinente caracterizar los robots sociales como aquellos concebidos para realizar tareas en el marco de prácticas que involucran cierta interacción social, como es el caso de los contextos de cuidados, educación o compañía. Su particularidad consiste en ser "agentes artificiales" que proveen asistencia *por medio de* "interactuar socialmente" con los humanos (aunque no necesariamente en exclusiva con éstos).

En el ámbito tecnocientífico que nos ocupa, dicha interacción se entiende en un sentido muy concreto: que el robot se interrelaciona con nosotros de "forma interpersonal" (Breazeal et al., 2008); lo que a veces lleva incluso a calificar los robots sociales como aquellos destinados a relacionarse con las personas a nivel emocional (Campa y Campa, 2016). Con ello básicamente se pretende distinguir el carácter específico de este tipo de bidireccionalidad comunicativa de otras interacciones que se dan entre humanos y robots sin necesidad (o, mejor dicho, sin intención predeterminada) de "relación" psico-emocional.

El rasgo más distintivo de los robots sociales es, pues, su capacidad de interactuar "humanamente", siguiendo patrones básicos de lo que se consideran interacciones intersubjetivas significativas. Ahora bien, es fundamental no perder de vista el marco 'funcionalidad-finalidad' donde se ubica esta capacidad de interacción, pues su ejercicio siempre se orienta a un objetivo concreto. En este sentido, por hacer hincapié en el esquema de finalidades que entra en juego en un robot social, resulta clarificadora la definición de Sheridan (2020), que señala que el objetivo del robot es establecer con la persona una interacción afectiva o de otro modo útil –esto es, siempre provechosa con vistas a un fin.

Esta puntualización permite situarse más rápidamente en las ambigüedades de la noción "robot social". Para empezar, se entiende mejor el sentido en que se usa comúnmente el término "social" en relación a este tipo de inteligencias corporeizadas. A veces, el contraste entre los robots sociales y los industriales puede confundir, conduciendo incluso a una equiparación inexacta entre robot social y robot de servicios. Fundamentalmente, 'social' se dice de un robot por como este artefacto desarrolla sus tareas, es decir, interactuando. Cuál sea la finalidad última de esta interacción es ya otra cuestión, que tiene que ver con su campo de aplicación o el perfil del usuario principal. Con ello se dilucida también la razón de ciertas categorizaciones clásicas en la literatura sobre robótica social. Dado que la interacción humano-robot tiene como finalidad servir a algún objetivo dentro de un determinado contexto re-

lacional, los robots sociales se pueden clasificar en distintos grupos de acuerdo con criterios relativos a esta finalidad.

Uno de ellos sería, por ejemplo, el ámbito de las prácticas en las que asiste el robot (educativo, sociosanitario, de cuidados, asistencial, terapéutico, doméstico, etc.). A esto responde la gran diversidad terminológica (a menudo poco rigurosa) para categorizar robots sociales en distintos (sub)grupos: “robots tutores o profesores”, “robots sanitarios”, “robots de cuidados” (*care robots*), “robots asistenciales”, “robots personales”, “robots de compañía”, etc.

Otro criterio detectable es lo que podríamos entender como el carácter de la finalidad última de la interacción del robot con el humano. Parece haber cierto empeño en diferenciar los robots sociales en función de si la razón última de su interacción es externa o interna a la misma —es decir, si la interacción es parte necesaria pero no suficiente de la funcionalidad del robot (esto es, la interacción no agota la tarea del robot) o si la interacción y la funcionalidad se equiparan (la interacción es la tarea).

A esta distinción respondería, por ejemplo, la taxonomía de Feil-Seifer y Matarić (2005), en la que el concepto de “robots socialmente asistenciales” (*socially assistive robots* o *SAR*) —más oportunamente redefinidos como “robots sociales asistenciales” en (Payr, 2015)— se introduce para singularizar aquellos robots cuya interacción social es un medio para conseguir desarrollar una tarea de apoyo en un ámbito determinado. Éste sería el caso de un robot que asistiera en el campo de la salud mental ayudando a un usuario a llevar a cabo un ejercicio cognitivo a través de una comunicación verbal y gestual. Un ejemplo de ello es el sistema cognitivo robótico para asistir a pacientes² con demencia en un juego de estimulación cognitiva, desarrollado por Andriella et al. (2020) en el Institut de Robòtica i Informàtica Industrial (CSIC-UPC), y en colaboración con la Fundació ACE. Gracias a su capacidad de percepción, aprendizaje y reacción al comportamiento del usuario, el robot interactúa de forma adaptativa, moldeando autónomamente en cada momento, según el historial de sus acciones y el estado del ejercicio, el nivel y tipo de asistencia a ofrecerle (animando, recomendando o desvelando la solución).

La intención de Feil-Seifer y Matarić (2005) es contraponer esta clase de robots sociales a otros cuya interacción social es la finalidad última, no habiendo así una tarea posterior con respecto a la cual esta es un medio. Esto no impide discernir los motivos por los que se busca esta interacción, lo que da lugar a (sub)clasificarlos en robots de compañía, entretenimiento, apoyo emocional en diferentes ámbitos, etc. Cabe señalar que a los robots cuya funcionalidad es la interacción por la interacción, Feil-Seifer y Matarić (2005) los identifican con los que Fong et al. (2003) habían tipificado de “robots socialmente interactivos” (*socially interactive robots* o *SIR*), si bien no parece que originalmente esta categoría excluyera los robots cuya función interactiva tuviera como objetivo una tarea externa a la misma. Por el contrario, esta terminología vendría a reforzar la idea de que el rasgo distintivo de los robots sociales es, precisamente, interactuar como *manera de* desarrollar una tarea, independientemente de si ésta se agota o no en la interacción.

Una distinción excesivamente perfilada del carácter de la finalidad última de la interacción tiene implicaciones éticas. Estimamos importante señalar que, en relación a ciertas prácticas de tipo relacional a las que se pretende asistir tecnológicamente, centrar el desarrollo robótico en una visión desmesuradamente fijada en tareas puede suponer un empobrecimiento

de la actividad. Por ello, cuando la interacción social se reduce a una forma de llevar cierta labor a buen puerto, limitando la interacción a un objetivo específico y discerniendo la tarea del acto social, se puede reforzar una forma perjudicial de fragmentar actividades humanas de dimensión holística. La robótica social se insiere siempre, en última instancia, en el contexto de cuidados, entendido en un sentido comprensivo —esto es, no como relativo a un paradigma clínico/médico-rehabilitador, sino a uno social-hermenéutico cultural—. En este sentido, el cuidado no significa paliar déficits de salud, sino que se vincula a una noción integral de salud en que entran en juego la prevención y el bienestar psicoemocional, entre otros; es decir, se trata de una actividad orientada al sustento del “mundo” de las personas (Tronto, 1993). Esto problematiza distinguir demasiado estrictamente entre las finalidades externas e internas de las funcionalidades interactivas para las que se diseña un robot en ámbitos prácticos de esta índole, pues en las tareas pertenecientes a este tipo de prácticas (asistencia física, cognitiva o emocional), ‘asistir’ y ‘cuidar’ se interrelacionan, y la finalidad de la interacción difícilmente se puede delimitar con exactitud.

3. ¿Cómo orientar el pensar ético en torno a la robótica social?

En la creciente literatura académica sobre las implicaciones éticas de la robótica social constatamos un desorden abrumador. Existe una gran diversidad de criterios y perspectivas —no siempre explicitados y todavía menos frecuentemente justificados— desde los que se identifican y se toma posicionamiento con respecto a las problemáticas (Vandemeulebroucke et al., 2018). Esto da lugar a preocupaciones de índole muy distinta³. Sin duda, este embrollo es sintomático de la complejidad que entraña la robótica social como objeto de análisis crítico-reflexivo, es decir, a la hora de ser abordada desde la ética.

3.1. La ética como actividad reflexiva

En general, se da por supuesta la comprensión de lo que significa aproximarse a la robótica social desde una perspectiva ética. Rara vez encontramos una explicación del término “ética” en la literatura crítico-normativa sobre robótica social, y mucho menos una explicitación del sentido de la relación entre ambas actividades.

El desafío reflexivo al que nos emplazan los nuevos horizontes abiertos por la robótica inteligente requiere distinguir la ‘ética’ de la ‘moral’. Ambas tienen que ver con el ejercicio de orientación de la acción humana. Ahora bien, aunque son etimológicamente equivalentes y se usan popularmente de forma indistinta, hay consenso acerca de la necesidad de entender estos dos términos como connotativamente distintos (Ricoeur, 2008). El motivo es poder capturar dos dimensiones de nuestra vida práctica que necesitamos diferenciar: las acciones y las razones. La ética se entiende como una reflexión crítico-racional sobre los fundamentos de la moral, esto es, sobre la validez de las normas y valores de nuestros hábitos y costumbres. De ahí la conocida contraposición entre la moral como “vivida” y la ética como “pensada” (López Aranguren, 1994). La ética se ocupa de la legitimidad, de deliberar acerca de la necesidad de mantener, abandonar o (re)generar las distintas morales —que, en definitiva, son prácticas relativas a determinados contextos socioculturales e históricos y, por tanto, revisables—. La moral se entiende como adhesión a valores, reglas, a partir de los cuales las acciones se estiman como ‘buenas’ o ‘malas’. Por esto la ética se define como una metamoral,

al ser una “reflexión de segundo grado sobre las normas” (Ricoeur, 2008, 48). De este modo, y siguiendo a (Cortina, 2007), podemos diferenciar la moral y la ética por su forma inmediata o mediata de orientar la esfera de las acciones humanas respectivamente. Pues mientras la moral responde a la pregunta “¿qué debo hacer?” —indicando una acción u omisión—, la ética contesta la cuestión de “¿por qué debo [hacerlo o no]?” (Cortina, 2007, 62) —ofreciendo, por tanto, argumentos.

Es a la ética así entendida, y no a la moral, a la que corresponde afrontar el reto de orientar el despliegue de la robótica social. Ante este nuevo escenario, lo que necesitamos es poder fundamentar las acciones a emprender en relación al desarrollo, implementación y uso de estos sistemas inteligentes autónomos en base a unos valores crítica y explícitamente analizados y revisados, aprovechando el potencial disruptivo de la innovación para poner entre paréntesis el ‘automatismo moral’ que suele regir nuestra cotidianidad. Esto es una tarea colectiva que requiere de interdisciplinariedad, y no sólo se opone a las posiciones extremas de rechazo o acogida acrítica de los avances tecnológicos, sino que contrasta con ellas por tomar posición activa en la (re)configuración de la realidad social.

Cabe mencionar que, desde un punto de vista ético, hay distintas perspectivas desde las que se puede fundamentar el porqué de una acción, según el carácter de las razones desde las que se examina su corrección o incorrección. Principalmente, la reflexión se puede articular a partir de principios, consecuencias o virtudes, lo que explica la tradicional clasificación entre éticas deontológicas, teleológicas y de la virtud (Camps, 2017). En el panorama actual de deliberación ética sobre la robótica social hay una considerable heterogeneidad a nivel de perspectivas. La detección y exposición de los problemas se suele argumentar desde uno de estos tres criterios o ángulos de valoración. Es importante recordar que una actividad ética debería ser inclusiva, fundamentando la acción no solo en principios o consecuencias, sino integrando armónicamente a ambos en la consideración (Weber, 2012), y siempre en vista del tipo de sociedades que queremos. Todo ello sin desdeñar la atención al contexto, cuya relevancia ha sido bien advertida por la ética del cuidado (Busquets Surribas, 2019): pues no sólo hay valores inherentes a este que deberán introducirse en la ponderación (como cuando hablamos de contextos profesionales), sino que también las consecuencias de una misma acción pueden variar según las circunstancias y sus afectados.

3.2. Las complejidades de una deliberación ética

Hay ciertos elementos que hacen notablemente complicado reflexionar sobre la robótica social desde una perspectiva ética y dirigir el pensamiento en este terreno de forma apropiada y fructífera.

En primer lugar, existe una considerable diversidad de esferas de acción implicadas por la actividad de la robótica social y respecto a las cuales, por tanto, enfocar el pensar. Dado que la ética es una actividad de reflexión centrada en la argumentación del porqué de cursos de acción, esto supone una primera dificultad de orientación, pues, con vistas a este fin, hay una multiplicidad de ámbitos de acción a barajar: (1) la robótica social como actividad tecnocientífica con diversos sujetos implicados, pluralidad de intereses y valores (Echeverría, 2003); (2) las decisiones concretas de los ingenieros en cada una de las fases constitutivas de la actividad (diseño, desarrollo y experimentación de los productos); (3) las acciones relativas a la im-

plementación de los sistemas inteligentes resultantes (adopción institucional, integración de los artefactos en ámbitos profesionales); (4) las acciones de los usuarios; etc. Esto comporta, a su vez, una variedad de sujetos a quienes conciernen las reflexiones y determinaciones que se derivan de la reflexión ético-normativa. Este intrincado entramado de dimensiones es al que quiere dar respuesta la 'Roboética' en su sentido primario (Veruggio y Abney, 2014), esto es, como disciplina centrada en el actuar humano en las diferentes fases de despliegue de la robótica y sus productos y, por tanto, ocupada con los fundamentos de acción de sujetos pertenecientes a campos de actividad muy diversos.

Parece que a ello se añadiría, además, otro ámbito para la consideración ética: el hacer de los robots sociales. Dado su estatuto de sistemas inteligentes con autonomía para (inter)actuar con el entorno para el logro de fines —esto es, para percibir, procesar información y obrar en consecuencia, adaptándose y aprendiendo en el curso de interacción—, estos robots plantean cuestiones relativas a su comportamiento y capacidad de sopesar los cursos del quehacer en contextos humanos que requieren de cierta "sensibilidad para consideraciones morales" (Wallach y Allen, 2009, 34). Esta es la línea de reflexión de la 'Ética para máquinas', la segunda de las ramas de la 'Roboética'. Volveremos a ello más adelante.

En segundo lugar, la naturaleza tecnológica de la robótica social dificulta el abordaje ético de la misma, al precisar de un 'pensar ético por esferas (conjugadas)', que debe atravesar varias capas de análisis. El carácter instrumental de la tecnología requiere contextualizar la reflexión ética de acuerdo con el campo de aplicación de la robótica. Dado que la finalidad de la robótica social es, en última instancia, externa a la misma —consistente en servir a los fines del contexto práctico en que sus productos se introducen como medios de apoyo—, la reflexión ética no puede desvincularse de la esfera de actividad humana concreta para la que se desarrollan los artefactos. Esto requiere filtrar el pensar ético a la luz de un determinado campo de aplicación de la robótica social (asistencial, educativo, sociosanitario, terapéutico, etc.), e interpretar los retos éticos que comporta en el marco de valores y finalidades propios de estos contextos —sin olvidar el universo más general de principios y valores de la ética cívica donde se legitiman.

Por consiguiente, una reflexión ética en torno a la robótica social debería proceder, al menos en parte, como una ética aplicada, en el sentido de enfocarse a cada uno de los diferentes sectores para los que se desarrolla esta actividad tecnocientífica. Esto es importante, pues en realidad no es riguroso hablar de una 'ética aplicada a la robótica social', si por esto se entiende que el terreno de reflexión es exclusivamente relativo a esta actividad y/o sus artefactos sin enmarcarla en las coordenadas de principios, valores y finalidades particulares de cada contexto práctico para el que se conciben y se desarrollan. Por eso preferimos hablar de ética *para* la robótica social. Se trataría de un ejercicio de ética aplicada como hermenéutica crítica de las actividades humanas (Cortina, 1996). No ayuda a resolver problemas, que es a lo que se dedica la ética aplicada (Román Maestre, 2016), plantear las problemáticas de la robótica social sin "situarlas" en una determinada actividad.

En tercer lugar, en el seno de los sistemas inteligentes de la robótica social se da una cierta convergencia entre dos estatutos ontológicos hasta ahora claramente diferenciados, a saber: sujeto y objeto. Esto distorsiona con facilidad el enfoque de la reflexión ética acerca de esta actividad tecnocientífica. Los robots sociales, como entidades con capacidad de hacer como resultado de un proceso cognoscitivo previo, plantean retos éticos no solo en tanto que ob-

jetos —ahí surgen cuestiones sobre privacidad, impacto laboral, desigualdad social, robotización de los cuidados, etc.—, sino también en tanto que “sujetos agentes” que desarrollan una tarea dentro del contexto de actividades humanas cotidianas, y lo hacen mediante una interacción pretendidamente “intersubjetiva”. Es por esta capacidad de agencia que la ‘ética para máquinas’ se orienta a garantizar el alineamiento de los sistemas inteligentes autónomos con los valores humanos. Por el momento, adelantaremos que la piedra de toque de una aproximación ética rigurosa a la robótica social consiste en una determinada concepción de esta “agencia” artificial de sus artefactos.

4. Autonomía tecnológica y externalización de la agencia: ¿de qué hablamos?

La creciente autonomía tecnológica de los robots inteligentes abre la posibilidad de transferirles parte de nuestra agencia para la realización de tareas concretas. Es lo que denominaremos ‘externalización de la agencia humana’, por cuanto tiene que ver con *encomendar* a entidades tecnológicas determinados quehaceres que requieren de facultades cognitivas e interactivas con el entorno.

Es primordialmente en virtud de este fenómeno que la robótica inteligente, así como su rama de la robótica social, urge a ser pensada desde la ética. El hecho de externalizar parte de nuestra agencia, delegándola o asignándola a los robots, nos sitúa ante un desdibujamiento de los pilares de nuestra vida práctica. Se plantean interrogantes sobre qué roles confiarles a los robots, para qué y las transformaciones que ello supone, tanto en un contexto de prácticas determinado, como en el plano social e individual.

Efectivamente, si examinamos las diferentes preocupaciones éticas planteadas por la robótica social en los debates actuales descubrimos que todas ellas pivotan, ya sea más explícita o implícitamente, sobre el fenómeno de la externalización de la agencia. Piénsese, por ejemplo, en las recurrentes controversias sobre la vulneración de la dignidad humana, como la objetificación de los usuarios de robots sociales asistenciales o el engaño de una relación “intersubjetiva” con robots de compañía (Coeckelbergh, 2012; Noori et al., 2019; Sharkey y Sharkey, 2012; Sparrow y Sparrow, 2006). En la misma línea están las discusiones sobre los límites de la interferencia de un robot social con la autonomía o la privacidad personal (Feil-Seifer y Mataric, 2011), y los debates acerca de la responsabilidad por posibles daños inesperados. O repárese en las inquietudes sobre la denigración de facultades morales humanas que conllevaría tanto la sustitución del profesional por robots en prácticas de cuidado (O’Brolcháin, 2019 Vallor, 2015), como la interacción humano-robot (Cappuccio et al., 2020); o la corrupción de prácticas de cuidados a raíz de la robotización de tareas en contextos hermenéuticos. Todas estas problemáticas vienen desencadenadas por el hecho de que nuestra capacidad de realizar tareas específicas sea asumida por esta tipología de agentes artificiales.

Ahora bien, resulta crucial advertir los reduccionismos en la mirada ética a que puede conducir el fenómeno de externalización de la agencia humana. Éste, sin un previo análisis crítico de la perspectiva bajo la cual entenderlo debidamente, parece impelernos a estructurar la aproximación ética a la robótica social de forma desacertada, centrada en el ‘hacer’ del robot.

En efecto, si los robots sociales son concebidos como entidades a quienes encargar (parte de) nuestras funciones en prácticas cotidianas de naturaleza relacional, fácilmente puede darse

una fijación en cuestiones relativas a su dimensión “agente”. Dada su capacidad de asumir un rol encomendado, deviene esencial determinar cómo estos robots deberán comportarse e (inter)actuar para que puedan entrar legítimamente en nuestro espacio social y hacerse cargo de susodichas tareas de una determinada manera (apropiada a la nuestra, no a la suya).

Según la premisa de que depende necesariamente de su comportamiento que un robot sea o no beneficioso para la humanidad (Veruggio y Abney, 2014), se ha dirigido la atención a cuestiones propias de la ‘ética para máquinas’, orientada a dotar (en caso de ser posible) a los robots de competencia moral. Se aduce (Allen et al., 2006) que esta es el requisito para asegurar que el impacto de su quehacer en entornos inherentemente éticos —no estructurados y, por tanto, complejos e impredecibles— sea positivo, en el sentido de estar alineado con los principios y valores humanos (Cave et al., 2019). A pesar de existir sólidas refutaciones a la idea de que implementar capacidad de razonamiento ético a sistemas inteligentes sea garantía de resultados sociales positivos (Brundage, 2014), se destinan numerosos esfuerzos a ello (Sharkey, 2020), para que estos puedan sopesar adecuadamente los posibles cursos de su quehacer de acuerdo con ciertos principios, cuya concreción y translación computacional es otro motivo de debate.

Sin duda, las cuestiones relativas al ‘hacer’ del robot son relevantes para la orientación normativa de la robótica social. No obstante, centrar la reflexión *primariamente* en esta dimensión supone pasar por alto el foco de atención verdaderamente principal y, por tanto, un reduccionismo de la mirada ética. En ello, juega un papel determinante el marco desde el que nos acerquemos al fenómeno de la externalización de la agencia.

Se pueden identificar una serie de aspectos implícitos y constitutivos del modelo de aproximación ética desenfocada a la robótica social, en el que las problemáticas se piensan, se definen y se juzgan prioritariamente a la luz de la cuestión por el ‘hacer’. Los introducimos someramente.

El más relevante es una concepción de los robots como agentes artificiales que se insieren en nuestra cotidianidad en calidad de sujetos o “casi-otros”. Esta perspectiva es la que Coeckelbergh (2011) considera como perteneciente a una “ontología individualista”, bajo la que los robots sociales son vistos como “individuos” —es decir, como un foco bien demarcado y ‘cerrado’ de agencia—. Desde esta óptica, el fenómeno de externalización de la agencia casi naturalmente induce a una aproximación ética deficiente, que parte de una idea de los robots como “sujetos” diferenciados de los humanos, que entran en nuestros contextos prácticos como “alteridades” que tienen su propia agencia (aunque delimitada, por heterónoma y hetero-referente).

De ahí que los objetivos de la ‘ética para máquinas’ se presenten como apremiantes y se articule la reflexión ética sobre la robótica social primeramente en torno a cuestiones relativas al estatuto moral de los robots, no sólo como sujetos agentes, sino también pacientes (Coeckelbergh, 2020; Gunkel, 2018).

A la par, todo esto se vincula a una fijación ética en los impactos que los robots tienen por su conducta, y que, además, se limitan al ámbito individual de la vida humana. Pues cabe señalar una forma diádica de entender la relación humano-robot, que articula la reflexión ética sobre este binomio. Se centra entonces la atención en los impactos que el robot social puede tener

sobre la persona con quien interactúa en el desarrollo de una tarea, a riesgo de descuidar algo fundamental: que dicho artefacto se introduce en un entramado de por sí relacional, esto es, en una red de interrelaciones entre distintos actores que configuran la totalidad de la actividad (Vallès-Peris et al., 2018).

5. Una aproximación ética a la robótica social desde el 'ser': teleología e interés

Aquí defendemos que la cuestión fundamental a cuya luz acercarnos a la robótica social desde una perspectiva ética no es la del 'hacer', sino la del 'ser'. Antes que 'hacer correctamente', un robot social debe 'ser correcto', lo cual tiene que ver con la idea de legitimidad. Preguntarse por el ser implica tener en cuenta la coherencia entre la razón de ser del agente artificial y las finalidades a las que sirve a través de sus funcionalidades; es decir, significa pensar en términos de adecuación entre acciones, valores y objetivos, entrando todos ellos en la ecuación como objetos de la actividad crítico-reflexiva.

La relevancia de esta dimensión no ha pasado totalmente desapercibida, pues hay autores que reflexionan sobre la pertinencia de otros tipos de tecnología partiendo de aspectos como, por ejemplo, los problemas que pretenden solventar (Baumer y Silberman, 2011) o los valores respaldados por su diseño (Millar, 2015). Sin embargo, consideramos imperativo explicitar la centralidad de esta cuestión, ya que es aquella que debería moldear primariamente la perspectiva desde la que reflexionar éticamente en torno a la robótica social, por lo que respecta a todas las fases de su despliegue.

Desde el prisma del 'ser' se explicitan dos categorías fundamentales para una aproximación ética pertinente a la robótica social: teleología e interés. Ambas promueven una mirada adecuada de los robots sociales como artefactos que, como productos fabricados por (y para) la actividad humana (Johnson G., 2011), nacen en un marco de finalidades concretas que responden a intereses determinados —que están siempre vinculados a un contexto político, sociocultural e histórico específico y que también deben ser tomados en consideración en la reflexión ética.

Con ello se hace manifiesto que la agencia de estos artefactos inteligentes responde a un propósito último, que le viene impuesto, asignado y designado externamente. Esto tiene una doble connotación: por un lado, el objetivo del robot viene fijado por los humanos y, por otro, éste se vincula a una(s) finalidad(es) relativas al contexto práctico concreto donde ejerce las tareas. Así, deberíamos precisar que los robots sociales son *externalizaciones de una agencia interesada* y, por tanto, son agentes de naturaleza subrogada, que plantean interrogantes no solo sobre su ejecución de cursos de acción, sino prioritariamente sobre su razón de ser, es decir, de las finalidades e intereses de los que estos artefactos emergen, representan y a los que sirven.

De ese modo, pensar en términos de teleología e interés en materia de robótica social supone volver la atención ética hacia donde principalmente pertenece, esto es, la acción humana. Se trata de dos conceptos que deberían empezar a reivindicarse abiertamente en el marco de las reflexiones actuales sobre la actividad de este campo tecnocientífico.

Principalmente, se pueden alegar dos razones capitales a favor de la perspectiva del 'ser' como fundamento de una aproximación ética a la robótica social, que exponemos a continuación.

1) *La coherencia con la naturaleza tecnológica de la robótica social*

En primer lugar, poner el foco en la cuestión del 'ser' resulta coherente con la naturaleza tecnológica de la robótica social. Como expusimos antes, la finalidad última de esta actividad tecnocientífica siempre se vincula con los fines particulares del campo práctico donde se destinan los artefactos. En consecuencia, para la reflexión ética se hace necesario tener en cuenta las finalidades, valores y principios específicos a la esfera de actividad para la que los robots sociales se conciben como herramientas. Solo desde esta contextualización se puede deliberar críticamente sobre la pertinencia de sus funcionalidades, esto es, su quehacer.

Para esclarecerlo, tomemos la particularidad de los robots sociales, a saber, su capacidad para interactuar con las personas. Esta es su principal funcionalidad, objeto que requiere de examen ético. Ahora bien, dado que, por el carácter instrumental de la tecnología, la interacción como funcionalidad siempre se presentará como un medio para un determinado fin, una reflexión ética consecuente deberá tomar en consideración el objetivo predeterminado de la interacción. Pues la problemática de la interacción humano-robot diferirá según su finalidad sea, por ejemplo, facilitar la ejercitación cognitiva del usuario en el ámbito terapéutico u ofrecer compañía en un contexto de asistencia sociosanitaria.

La perspectiva del ser conduce a ello de forma natural, pues parte de un interrogarse sobre el "sentido" del robot, atendiendo a la armonía entre sus funciones específicas y los objetivos de estas, dentro de un marco más amplio de finalidades, valores y hermenéuticas abiertas. Así, se garantiza una aproximación a la robótica social y sus artefactos centrada en su carácter *instrumental* respecto a algo concreto.

Esto significa atender al '*cómo*' que el artefacto representa con relación a unos objetivos. Con ello no nos referimos simplemente a evaluar si la robótica cumple *de facto*, a través de las funcionalidades desplegada en sus productos, las labores que pretendemos delegar. Es decir, no se trata de adoptar un paradigma restringido en que las finalidades de una actividad sean simplificadas a tareas. Para clarificarlo, aterramos un poco la cuestión: debería examinarse minuciosamente, por ejemplo, qué implica desarrollar artefactos robóticos para asistir en el comer. Dado que la finalidad es dar de comer, ¿pensaremos en términos de instrumentos que faciliten la ingesta, traduciendo el acto de ayudar a comer a garantizar el nutrirse? ¿Optaremos, entonces, por cubrir las necesidades de los pacientes que no pueden valerse por sí mismos a la hora de comer mediante el despliegue de un brazo robótico capaz de interacción física con el entorno, que le habilite para llevar a cabo su función sin peligro de dañar corporalmente a las personas? ¿O esto significaría dejar fuera la dimensión social del comer? ¿Desarrollamos mejor una aplicación robótica social, capaz de interacción "interpersonal", o entendemos que serán los humanos quienes complementarán la asistencia asegurando a la actividad el carácter relacional, que es interpersonal e impredecible? En caso que optemos por desarrollar artefactos sociales, ¿su interacción irá destinada, únicamente, a cumplir con la tarea de alimentar, o debería pensarse también desde la dimensión de los cuidados, del "estar al lado de", hacer compañía, y dar conversación o tacto? Tener en cuenta el estatuto de 'instrumento' de la robótica significa, fundamentalmente, tomar en consideración la forma

en que sus artefactos ejecutan unas funciones —u, otramente dicho, la representación de valores y concepciones de la actividad que éstos personifican—. Pues debe haber consonancia entre el ‘qué’ se hace y el ‘cómo’ se lleva a cabo, en referencia a unos valores propios de la actividad.

En esta línea de pensamiento, ha surgido una interesante propuesta de aproximación normativa a la robótica para el ámbito de los cuidados, basada en el marco teórico de la “naturaleza de las actividades” (Santoni de Sio y van Wynsberghe, 2016). Se la conoce como “*nature-of-activities approach*”⁴, dirigida, en este caso, a delimitar qué funciones delegar a los robots para los cuidados y cuáles no y cómo diseñarlos para que estas respondan a valores constitutivos del ejercicio de atención sanitaria. El mérito de la propuesta estriba en tomar en consideración la distinción entre actividades “orientadas a la práctica” y “orientadas a objetivos” y partir de ella a la hora de reflexionar sobre qué porciones (tareas) constituyentes de la actividad del cuidado se podrían asignar al robot sin que, con ello, se dañe la dimensión que concerniría a los bienes internos de la misma (MacIntyre, 2019).

Resulta fundamental puntualizar que reflexionar éticamente desde una visión instrumental de la tecnología no implica desatender su carácter “mediador” (Verbeek, 2006). Es decir, examinar la coherencia de las funciones con las finalidades, en consonancia con valores definitorios de la actividad, no agota el objeto de reflexión ética. Pues los artefactos tecnológicos no son instrumentos neutrales, sino que condicionan decisivamente la forma en que se ejecutan las prácticas y, por tanto, su entramado de valores, por cuanto invitan a unas e inhiben otras posibilidades de “percepción y acción, experiencia y existencia” (Verbeek, 2006, 364). Este aspecto es crucial para la reflexión ética, porque anticipar qué tipo de mediación tecnológica queremos puede contribuir a desarrollar un tipo de robots con funcionalidades o características determinadas en lugar de otras, y todo ello puede aumentar o mermar la riqueza de los “entornos de funcionamiento” en que se introducen los artefactos (Toboso et al., 2020).

2) Una mirada ética extensa más allá y acá de los impactos

En segundo lugar, adoptar la perspectiva del ‘ser’ posibilita una aproximación ética comprensiva a la fuerza transformativa de esta actividad para la vida humana. Esto se explica por dos motivos.

Por un lado, porque al partir de unas coordenadas conceptuales que revelan a los robots como productos socioculturales —cuya razón de ser responde a un determinado sistema de valores y forma de entender y estructurar la vida en comunidad, esto es, a intereses humanos—, amplía la reflexión ética más allá de un mero ejercicio de “impactología”. Es decir, evita reducirla exclusivamente a una evaluación crítica de los (predecibles) impactos de los robots.

En efecto, reducir a los impactos la evaluación ética de la tecnología entraña complejidades bien conocidas, como la imprevisibilidad de sus consecuencias (Jonas, 2015) o la cuestión sobre la perspectiva a partir de la cual éstas deberían estimarse (desde qué valores o principios, y en relación a quién, exactamente). Pero todavía más trascendente resulta el riesgo de caer en un “conservadurismo moral”, al convertir la reflexión ética en una evaluación moral. Ciertamente, atender solo a los impactos producidos por la introducción robótica en nuestras prácticas podría precipitarnos a atrincherar nuestra aproximación en el marco de

los valores actuales, sin atender la necesidad constante de repensarlos, tal como implica el ejercicio de la ética.

Bajo la perspectiva del ‘ser’, los robots sociales no son simplemente agentes artificiales que pueden asumir tareas esencialmente humanas, sino que representan una encarnación de cierta cosmovisión e intereses. Por ello, la reflexión no se limita a abordarlos como entidades “quienes” producen impactos en su quehacer, los cuales deben ser positivos —responsabilidad que, como dijimos, desde una posición ética desenfocada se suele interpretar como primariamente exigiendo desarrollar competencia moral en estos artefactos—. Contrariamente, se focaliza la atención antes que nada en la acción humana, tomando como objeto del análisis normativo el entramado de intereses, finalidades y valores a los que responden los productos de la actividad robótica. Esto supone, pues, no solo deliberar sobre ‘medios’ en un marco de ‘fines’ ya dados, sino abrir la consideración sobre ‘qué medios’ para ‘qué fines’.

Por otro lado, la perspectiva del ‘ser’ incluye en la reflexión ética una consideración a distintas esferas de la vida humana, más allá de la individual. Al centrar la atención primordialmente en los humanos —por ser los sujetos que deciden el ‘para qué’ del robot y que transforman el panorama de sus prácticas al incluir en ellas estas tecnologías—, la perspectiva del ‘ser’ no limita el ámbito de reflexión a la vida de los individuos concretos con quienes estos artefactos interactúan en el inmediato contexto de su uso, sino que también lo extiende a otras esferas, como la interpersonal, sectorial o institucional. De este modo, promueve una reflexión sobre el conjunto de nuestra forma de estructurar la vida humana, las prácticas y arquitecturas subyacentes a su lógica, esto es, una aproximación ética exhaustiva y, por tanto, madura. El interrogante está, en primer lugar, en qué tipo de prácticas (y valores) queremos fomentar y cuáles no —cuestión cuya importancia acertadamente enfatizan Sandel (2015) y MacIntyre (2001)—, y qué parte de nuestra agencia queremos delegar o no y porqué. Se trata, en definitiva, de un preguntarse por el tipo de sociedades y vida que queremos construir.

6. Conclusiones

El potencial robótico para asumir nuevos roles en prácticas humanas de dimensión relacional —es decir, en la esfera de las interacciones intersubjetivas— sitúa la robótica social como una de las grandes fuerzas disruptivas actuales de nuestra vida práctica. Esto urge a aproximarnos a dicha actividad tecnocientífica desde la disciplina de la ética. Sin embargo, para ello es importante establecer previamente los fundamentos de un pensamiento crítico-reflexivo bien orientado hacia este campo.

El presente artículo ha pretendido contribuir en esta línea, clarificando la perspectiva desde la que dirigir la mirada ética a la robótica social, que será aquella centrada primariamente en la cuestión del ‘ser’. Así nos ubicamos dentro de las coordenadas conceptuales trazadas por las categorías de teleología e interés, que son clave para acercarnos correctamente al fenómeno nuclear de preocupación ética, relativo a la posible externalización de nuestra agencia en los robots sociales. Al volver el foco de atención ética a la acción humana, la ‘perspectiva del ser’ permite emprender una reflexión crítica verdaderamente capaz de orientar el potencial de (re)configuración de la vida humana que tiene la actividad tecnocientífica.

Bibliografía

- Allen, C., Wallach, W., y Smit, I. (2006). Why machine ethics? *IEEE Intelligent Systems*, 21(4), 12–17. <https://doi.org/10.1109/MIS.2006.83>
- Andriella, A., Torras, C., y Alenyà, G. (2020). Cognitive System Framework for Brain-Training Exercise Based on Human-Robot Interaction. *Cognitive Computation*. <https://doi.org/10.1007/s12559-019-09696-2>
- Arendt, H. (2005). *La condición humana*. Paidós.
- Baumer, E. P. S., y Silberman, M. S. (2011). When the implication is not to design (technology). *Conference on Human Factors in Computing Systems - Proceedings*, 2271–2274. <https://doi.org/10.1145/1978942.1979275>
- Breazeal, C., Takahashi, A., y Kobayashi, T. (2008). Social Robots that Interact with People. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1349–1369). Springer. https://doi.org/https://doi.org/10.1007/978-3-540-30301-5_59
- Brundage, M. (2014). Limitations and risks of machine ethics. *Journal of Experimental and Theoretical Artificial Intelligence*, 26(3), 355–372. <https://doi.org/10.1080/0952813X.2014.895108>
- Busquets Surribas, M. (2019). Descubriendo la importancia ética del cuidado. *Folia Humanística*, 12.
- Campa, R., y Campa, R. (2016). The rise of social robots : a review of the recent literature. *Journal of Evolution and Technology*, 26(1).
- Camps, V. (2017). *Breve historia de la ética*. RBA Libros.
- Cappuccio, M. L., Peeters, A., y McDonald, W. (2020). Sympathy for Dolores: Moral Consideration for Robots Based on Virtue and Recognition. *Philosophy and Technology*, 33(1), 9–31. <https://doi.org/10.1007/s13347-019-0341-y>
- Cave, S., Nyrup, R., Vold, K., y Weller, A. (2019). Motivations and Risks of Machine Ethics. *Proceedings of the IEEE*, 107(3), 562–574. <https://doi.org/10.1109/JPROC.2018.2865996>
- Coeckelbergh, M. (2011). Is ethics of robotics about robots? Philosophy of robotics beyond realism and individualism. *Law, Innovation and Technology*, 3(2), 241–250. <https://doi.org/10.5235/175799611798204950>
- Coeckelbergh, M. (2012). Are emotional robots deceptive? *IEEE Transactions on Affective Computing*, 3(4), 388–393. <https://doi.org/10.1109/T-AFFC.2011.29>
- Coeckelbergh, M. (2020). Should We Treat Teddy Bear 2.0 as a Kantian Dog? Four Arguments for the Indirect Moral Standing of Personal Social Robots , with Implications for Thinking About Animals and Humans. *Minds and Machines*, 0123456789. <https://doi.org/10.1007/s11023-020-09554-3>
- Coeckelbergh, M., Loh, J., Funk, M., Seibt, J., y Nørskov, M. (Eds.). (2018). Envisioning Robots in Society - Power, Politics, and Public Space. In *Proceedings of Robophilosophy 2018 /TRANSOR 2018* (Vol. 311). IOS Press. <https://doi.org/10.1017/CBO9781107415324.004>
- Cortina, A. (1996). El estatuto de la ética aplicada. Hermenéutica crítica de las actividades humanas. *Isegoría*, 13, 119–134.
- Cortina, A. (2007). *Ética mínima. Introducción a la filosofía práctica*. Tecnos.
- Echeverría, J. (2003). *La revolución tecnocientífica*. Fondo de Cultura Económica de España.
- Feil-Seifer, D., y Matarić, M. J. (2011). Socially assistive robotics: Ethical issues related to technology. *IEEE Robotics and Automation Magazine*, 18(1), 24–31. <https://doi.org/10.1109/MRA.2010.940150>
- Feil-Seifer, D., y Matarić, M. J. (2005). Defining Socially Assistive Robotics. *9th International Conference on Rehabilitation Robotics*, 465–468.
- Fong, T., Nourbakhsh, I., y Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3–4), 143–166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)

- Funk, M., y Coeckelbergh, M. (2020). (Technical) Autonomy as Concept in Robot Ethics. In *Biosystems and Biobotics* (Vol. 25). Springer. https://doi.org/10.1007/978-3-030-24074-5_12
- Gunkel, D. J. (2018). The other question: can and should robots have rights? *Ethics and Information Technology*, 20, 87–99. <https://doi.org/10.1007/s10676-017-9442-4>
- Johnson G., D. (2011). Computer Systems: Moral Entities but Not Moral Agents. In M. Anderson & S. L. Anderson (Eds.), *Machine Ethics* (pp. 168–183). Cambridge University Press. [https://doi.org/https://doi.org/10.1017/CBO9780511978036](https://doi.org/10.1017/CBO9780511978036)
- Jonas, H. (2015). *El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica*. Herder.
- López Aranguren, J. L. (1994). *Ética. Obras completas, II*. Trotta.
- MacIntyre, A. (2001). *Animales racionales y dependientes. Por qué los seres humanos necesitamos las virtudes*. Paidós.
- MacIntyre, A. (2019). *Tras la virtud*. Austral.
- Millar, J. (2015). Technology as Moral Proxy: Autonomy and Paternalism by Design. *IEEE Technology and Society Magazine*, 34(2), 47–55. <https://doi.org/10.1109/MTS.2015.2425612>
- Noori, F. M., Uddin, Z., y Torresen, J. (2019). Robot-Care for the Older People: Ethically Justified or Not? 2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 43–47. <https://doi.org/10.1109/DEVLRN.2019.8850706>
- Nørskov, M., Seibt, J., y Santiago Quick, O. (Eds.). (2020). Culturally Sustainable Social Robotics. In *Proceedings of Robophilosophy 2020*. IOS Press.
- O’Brolcháin, F. (2019). Robots and people with dementia: Unintended consequences and moral hazard. *Nursing Ethics*, 26(4), 962–972. <https://doi.org/10.1177/0969733017742960>
- Operto, F., y Veruggio, G. (2008). Roboethics: Social and Ethical Implications of Robotics. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1499–1524). Springer. <https://doi.org/10.1007/978-3-540-30301-5>
- Payr, S. M. (2015). Towards Human-Robot Interaction Ethics. In R. Trappl (Ed.), *A Construction Manual for Robots’ Ethical Systems. Cognitive Technologies*. Springer. <https://doi.org/10.1007/978-3-319-21548-8>
- Ricoeur, P. (2008). *Lo justo 2. Estudios, lecturas y ejercicios de ética aplicada*. Trotta.
- Rodogno, R. (2016). Ethics and social robotics. *Ethics and Information Technology*, 18(4), 241–242. <https://doi.org/10.1007/s10676-016-9412-2>
- Román Maestre, B. (2016). *Ética de los servicios sociales*. Herder.
- Sandel, M. (2015). *Contra la perfección. La ética en la era de la ingeniería genética* (2a). Marbot Ediciones.
- Santoni de Sio, F., y van Wynsberghe, A. (2016). When Should We Use Care Robots? The Nature-of-Activities Approach. *Science and Engineering Ethics*, 22(6), 1745–1760. <https://doi.org/10.1007/s11948-015-9715-4>
- Sarrica, M., Brondi, S., y Fortunati, L. (2020). How many facets does a “social robot” have? A review of scientific and popular definitions online. *Information Technology and People*, 33(1), 1–21. <https://doi.org/10.1108/ITP-04-2018-0203>
- Seibt, J. (2017). Robophilosophy. In R. Braidotti & M. Hlavajova (Eds.), *Posthuman Glossary* (pp. 390–393). Bloomsbury Academic.
- Seibt, J., Nørskov, M., y Schack Andersen, S. (Eds.). (2016). What Social Robots Can and Should Do. In *Proceedings of Robophilosophy 2016 / TRANSOR 2016*. IOS Press.
- Sharkey, A. (2020). Can we program or train robots to be good? *Ethics and Information Technology*, 22, 283–295. <https://doi.org/10.1007/s10676-017-9425-5>

- Sharkey, A., y Sharkey, N. (2012). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology*, 14(1), 27–40. <https://doi.org/10.1007/s10676-010-9234-6>
- Sheridan, T. B. (2020). A review of recent research in social robotics. *Current Opinion in Psychology*, 36, 7–12. <https://doi.org/10.1016/j.copsyc.2020.01.003>
- Sparrow, R., y Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, 16(2), 141–161. <https://doi.org/10.1007/s11023-006-9030-6>
- Toboso, M., Morte, R., Monasterio, A., Ausín, T., Aparicio, M., y López, D. (2020). Robotics as an Instrument for Social Mediation. *Biosystems and Biorobotics*, 25, 51–58. https://doi.org/10.1007/978-3-030-24074-5_11
- Tronto, J. (1993). *Moral Boundaries. A Political Argument for an Ethic of Care*. Routledge.
- Vallès-Peris, N., Angulo, C., y Domènech, M. (2018). Children's imaginaries of human-robot interaction in health-care. *International Journal of Environmental Research and Public Health*, 15(5). <https://doi.org/10.3390/ijerph15050970>
- Vallor, S. (2015). Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. *Philosophy and Technology*, 28(1), 107–124. <https://doi.org/10.1007/s13347-014-0156-9>
- Vandemeulebroucke, T., Dierckx de Casterlé, B., y Gastmans, C. (2018). The use of care robots in aged care: A systematic review of argument-based ethics literature. *Archives of Gerontology and Geriatrics*, 74(August 2017), 15–25. <https://doi.org/10.1016/j.archger.2017.08.014>
- Verbeek, P.-P. (2006). Materializing Morality. *Science, Technology, & Human Values*, 31(3), 361–380.
- Veruggio, G., y Abney, K. (2014). Roboethics: The Applied Ethics for a New Science. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot Ethics. The Ethical and Social Implications of Robotics*. MIT Press.
- Wallach, W., y Allen, C. (2009). *Moral Machines. Teaching Robots Right from Wrong*. Oxford University Press.
- Weber, M. (2012). *El político y el científico*. Alianza Editorial.
- Žižek, S. (2013). *The Purpose of Philosophy is to Ask the Right Questions*. Big Think. <https://bigthink.com/videos/the-purpose-of-philosophy-is-to-ask-the-right-questions>

Notas

- 1 El concepto de 'acción' se ha adoptado comúnmente para hablar de las tareas o funciones de los robots inteligentes, de lo que estos *hacen*. Inspirándonos en la diferenciación señalada por Arendt (2005) entre 'labor', 'trabajo' y 'acción', sería oportuna en la ética para la robótica social la distinción entre la capacidad de obrar (acción) y la capacidad de hacer (labor y trabajo), para diferenciar entre dos tipos de agencia: aquella autónoma y moral, caracterizada por la espontaneidad, la imprevisibilidad y la pluralidad y, por tanto, exclusivamente humana, y aquella agencia no moral, a la que pertenecerían los actos heterónomos de entidades fabricadas, como los robots. Esto conllevaría diferenciar connotativamente entre los términos 'acción/acto/obrar' y 'quehacer/tarea/conducta/comportamiento'. Referirnos con propiedad a la dimensión agente del robot implicaría usar siempre las nociones pertenecientes al segundo grupo, por lo que no hablaríamos de 'actos' o 'acciones' de los robots, sino de 'quehaceres' o 'conductas'. En este trabajo trataremos de mantenernos fieles, a partir de ahora, a esta distinción sustancial.
- 2 Adviértase que este sistema cognitivo robótico se concibe principalmente como una herramienta de apoyo a los terapeutas, por cuanto les ayuda a realizar esta tarea concreta en el marco de sus sesiones con los pacientes. Son estos profesionales quienes se encargan de preprogramar el robot para su función con cada usuario específico, a quien luego el sistema se va adaptando durante el transcurso de la tarea.
- 3 Así lo muestran los resultados del trabajo de revisión bibliográfica (en curso) sobre las preocupaciones éticas planteadas en relación a la robótica social asistencial, llevado a cabo en coautoría con las Dras. Begoña Román y Carme Torras. En la fecha de publicación del presente artículo, susodicho trabajo se encuentra en proceso de redacción para su edición.

El proceso de revisión se ha centrado en la literatura de la base de datos Scopus comprendida entre los años 2006 y 2020. Para la búsqueda, se han usado cuatro entradas terminológicas: “ethics” AND “assistive robot*” OR “care robot*” OR “social robot*” OR “human-robot interaction”. Con ello se han identificado una gran diversidad de cuestiones éticas asociadas a este campo de la robótica inteligente. Para cada entrada, las problemáticas se han compilado y categorizado en tablas que facilitan su visualización, en base a las cuales se ofrece un análisis cuantitativo-descriptivo del estado de la cuestión, complementado con un análisis cualitativo-crítico posterior. Estas tablas se incluirán en este próximo artículo previsto.

- 4 Podría discutirse la idoneidad del término “naturaleza” para denominar esta aproximación, por cuanto puede llevar a pensar que la propuesta de dichos autores se erige alrededor de una concepción de las actividades humanas como algo estático, inmodificable, con valores inmutables. Esto sería equivocado, pues de hecho se preocupan de explicitar una legítima diversidad interpretativa en torno a lo que significa exactamente una actividad, los valores asociados a ella. Por ello, sería una opción renombrar este enfoque como “aproximación a la dinámica de las actividades”, haciendo hincapié en la idea de la actividad como un proceder lanzado a una constante regeneración y, por tanto, axiológicamente revisable.