



¿Puede prestar cuidados la Inteligencia Artificial?

Vanessa Nurock¹

Recibido: 9 de enero de 2020/ Aceptado: 30 de mayo de 2020

Resumen. Este artículo tiene por objeto analizar la Inteligencia Artificial (IA) desde la perspectiva del género y más precisamente desde la ética del cuidado, para mostrar que los problemas planteados por la IA van mucho más allá de los que el género pone en evidencia y a los que con frecuencia se ven reducidos. La cuestión de si estos sesgos son incontestables y deben ser corregidos, me parece, sin embargo, que no es sólo coyuntural, como suele pensarse. Intento demostrar aquí que también es una cuestión estructural apoyándome en dos elementos complementarios: por una parte, afinando el análisis inspirado en la perspectiva de género; y, por otra parte, examinando los usos de la IA aplicados al cuidado. De este modo concluiré mostrando que no debemos preguntarnos solamente por el desarrollo objetivo de la IA sino también por nuestra relación subjetiva con la IA y las propiedades con la que la revestimos.

Palabras clave: Inteligencia Artificial; Ética; Género; Cuidado (Care); Sesgos

[en] Can AI care?

Abstract. This paper analyzes Artificial Intelligence through the lens of gender, and more particularly of the Ethics of Care, to show that the problems posed by Artificial Intelligence go much further than the gender biases to which they are often reduced. While these biases are not questionable and must be corrected, I argue that the issue here is not only conjunctural, as it is generally considered. I demonstrate that it is also structural by relying on two complementary elements: on the one hand, by refining the analysis of gender bias and, on the other hand, by looking at care robots including AI. I conclude by showing that it is not only the objective development of AI that needs to be questioned, but also our subjective relationship with AI and the properties we attribute to AI.

Keywords: Artificial Intelligence; Ethics; Gender; Care; Bias

Sumario. 1. Los sesgos: un problema importante para la IA. 2. La cuestión de los sesgos de género. 3. ¿La inteligencia artificial puede hacerse cargo del cuidado? 3.1. La IA de cuidados. 3.2. Las “máquinas morales”. 4. ¿Hacia una artificialización de la ética? 5. Referencias bibliográficas.

Cómo citar: Nurock, V. (2020). ¿Puede prestar cuidados la Inteligencia Artificial?, *Cuadernos de Relaciones Laborales*, 38 (2), 217-229.

La Inteligencia Artificial (IA) se despliega hoy en día en todos los campos de nuestra vida, desde el más técnico al más prosaico. Su extensión es tan amplia que afecta tanto a las tareas consideradas socialmente como las más “viriles”, aquellas que conciernen particularmente a lo militar – como lo demuestran, por ejemplo, las má-

¹ Profesora Asociada de Filosofía en el Departamento de Ciencia Política de la Université Paris 8-Saint Denis. Email: vanessa.nurock@univ-paris8.fr. Título original: « L'intelligence artificielle peut-elle faire du care ? ». Traducción: Carlos Prieto Rodríguez (UCM).

quinas de matar –, como a las tareas consideradas como más “femeninas”, particularmente aquellas que tienen que ver con el cuidado o con lo relacional, tal y como lo hacen ver los numerosos usos que se hace en los ámbitos de la medicina y de la paramedicina, del hogar, de la comunicación o, incluso, de la amistad, el amor o el sexo.

Podríamos, por lo tanto, suponer que la IA – que podría definirse groseramente como un conjunto de tecnologías relacionadas con las ciencias de la información y de la comunicación cuyo rasgo común es la automoción de las funciones intelectuales – no tiene género, ya que su uso iría desde aquello que consideramos prototípicamente más “femenino” – el cuidado de las personas vulnerables, por ejemplo – al más “masculino” – la IA aplicada a maquinaria de guerra, a drones mortíferos o a supersoldados, por ejemplo.

Podría parecer, por tanto, absurdo abordar la IA bajo el prisma del género. El género se caracteriza precisamente por comportamientos socialmente construidos (y con frecuencia heredados históricamente) que definen funciones binarias marcadas como femeninas o masculinas inscritas en unas relaciones asimétricas y sujetas a la dinámica del poder. Ahora bien, una de las razones por la que nos fiamos más de la IA es precisamente porque parece abstraerse de nuestros sesgos humanos demasiado humanos, lo cual quizás le permitiría, al menos a primera vista, abstraerse de los estereotipos de género. Se atribuye normalmente a la IA, en efecto, el carácter de neutralidad, o incluso de imparcialidad, lo cual constituye, según algunos, uno de los argumentos incuestionables para su utilización.

Este artículo tiene por objeto un análisis filosófico (apoyado en un análisis conceptual) de la IA entendida como un reto social, ético y político visto desde la perspectiva y de la cuestión del género. Se interesa por saber si la IA puede preocuparse del otro y de prestarle cuidados, cuestión que abordo formulando la pregunta genérica *Can AI care?*, ¿puede prestar cuidados la Inteligencia Artificial?

Intentaré mostrar que el género, y sobre todo la cuestión del cuidado (*care*), constituye un punto de vista particularmente interesante para captar algunos de los retos de la IA. Mi hipótesis es que la cuestión de los sesgos en general, y de los sesgos de género en particular, concernientes a la IA debe ser considerada no sólo como una cuestión coyuntural, según suele hacerse, sino también y sobre todo estructural, lo cual refuerza la hipótesis formulada por Joy Buolamwini² según la cual, junto a la mirada masculina (*male gaze*) y la mirada colonial (*colonial gaze*), puede sugerirse la existencia de una mirada codificada (*coded gaze*), que estructura el mundo tal y como es remodelado por la IA. Al considerar la cuestión de los sesgos de género como estructural antes que coyuntural, se puede plantear la cuestión de la atribución de tareas de cuidado a la IA y mostrar por qué la estructura de la IA no entra en una lógica de cuidado, tal y como ha sido caracterizada por la ética y la política del cuidado, sino todo lo contrario.

1. Los sesgos: un problema importante para la IA

Desde hace varios años, los trabajos sobre la cuestión de los sesgos relacionados con la IA, así como la expresión de sus estudios en la prensa, recordando ciertos escándalos, han cambiado con bastante radicalidad la mirada del gran público sobre la IA. El

² <https://medium.com/mit-media-lab/incoding-in-the-beginning-4e2a5c51a45d>

alcance y la extensión de estos sesgos tienen, en efecto, algo de inquietante: afectan a poblaciones minoritarias en la concepción interseccional de la IA (Buolamwini y Gebu, 2018): mujeres, personas de color, poblaciones jóvenes o personas de cierta edad; por ejemplo, en el software de reconocimiento facial, muy utilizados actualmente por la policía, particularmente en Estados Unidos, es mejor ser un hombre blanco joven que una mujer negra mayor cuando nos interesamos por la IA.

Estos sesgos afectan, cada vez más, a todos los sectores en los que interviene la IA: desde la asignación de recursos de la sanidad, de la justicia, a la educación o al empleo, y conciernen tanto a sectores que podrían parecer anecdóticos (y no lo son en modo alguno) como las máquinas relacionales (especialmente las asistentes personales) o los vehículos con sistemas de reconocimiento de voz incorporado. Este punto es esencial ya que, al insertarse en todos los aspectos de nuestra vida, incluso (y ¡sobre todo!) en aquellos en los que a priori no deberían estar, esos sesgos llegan a insertarse en nuestros modos de pensar (tanto natural como artificial).

Estos sesgos, con frecuencia categorizados entre el gran público como “el problema del hombre blanco³ (treintañero)” y calificados como “mirada codificada⁴” (*coded gaze*) por Joy Buolamwini, en referencia directa a la “mirada colonial” (*colonial gaze*) o la “mirada masculina” (*male gaze*) y que instalan, cada uno en su propio horizonte, una determinada visión del mundo y de sus relaciones de poder en función de un eje de dominación – patriarcal, colonial o algorítmico. Aquí la idea central es que codificar es tener el poder de imponer una cierta visión del mundo y ciertas relaciones de poder. Yo iría aún más lejos, y sugeriría un problema añadido, planteado por esta mirada codificada, es que pasa por el artefacto de una máquina que no es sólo una “máquina de ver”⁵, como puede ser el caso en la mirada masculina o colonial, sino también una máquina de prescribir. Esta dimensión prescriptiva es explícita en algunos casos, como en la asignación de recursos, en el de la salud o en el sistema judicial, por ejemplo. Durante mucho tiempo se ha pensado que estos sesgos no eran, después de todo, más que el espejo de la sociedad que hace la IA. Sin duda es el caso; pero la situación es probablemente más compleja, como tendremos ocasión de ver más abajo.

2. La cuestión de los sesgos de género

Entre esos sesgos, querría centrar mi análisis en este artículo en los sesgos de género. Para abordar la cuestión del género en relación con la IA, se puede recurrir a diferentes ángulos de aproximación. En primer lugar, es útil empezar por una mirada histórica para interrogarnos sobre la supuesta neutralidad de la IA, que podría – incluso debería – no tener género, dado que es generalmente presentada como objetiva.

³ Ver, por ejemplo, <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>

⁴ El término remite a visiones del mundo integradas, propagadas por aquellos que tienen el poder de codificar los sistemas (*embedded views that are propagated by those who have the power to code systems*). <https://medium.com/mit-media-lab/incoming-in-the-beginning-4e2a5c51a45d>

⁵ Es interesante observar que los argumentos técnicos utilizados para justificar los movimientos del código son a veces utilizados en referencia explícita junto con aquellos que habían sido utilizados para justificar que las fotografías no captan bien las pieles negras. Claramente la continuidad de estas miradas no es solo metafórica.

Uno de los proyectos que se encuentra en los orígenes de la IA es el proyecto ENIAC (*Electronic Numerical Integrator And Computer*), que comenzó en 1943 con el fin de poner en marcha dos años más tarde al primer ordenador Turing completo (es decir, capaz de resolver todos los problemas de cálculo), es particularmente revelador de la construcción histórica de los sesgos de género en la IA. En las fotografías que consagran el éxito del superordenador ENIAC, vemos ocho hombres blancos en la plenitud de su vida. Y sin embargo, tal y como lo ha hecho ver, entre otros, la socióloga Alison Adam (Adam, 1998), esa foto es particularmente reveladora de la historia del mundo digital en general, y de la de la IA en particular. En efecto, aunque se olvide frecuentemente, los equipos que han participado en el éxito de la IA eran, al principio, equipos mixtos. Estos equipos se han “masculinizado” a medida que se desarrollaba la IA.

Más exactamente, esa masculinización se ha operado a medida que la IA se convertía en un asunto importante, o reconocido como tal – lo cual hacía que no pudiera ser ya también un “asunto de mujeres”. Tal y como lo cuenta Betty Snyder, que participó en el proyecto ENIAC, se les “olvidaba” llamar a las mujeres cuando se hacían fotos oficiales. Del mismo modo, en los comunicados de prensa, se hacía normalmente referencia de manera vaga a un grupo de expertos citando sólo el nombre de algunos hombres⁶. La historia de la IA se ha visto así masculinizada en el sentido de que las mujeres han sido en gran parte borradas.

Dicho de otra manera, la concepción de la IA se halla dominada por una cierta visión patriarcal del mundo, que asigna a los hombres las tareas y los oficios más importantes y “objetivos”, aunque esto no se corresponda con la realidad de los hechos. Ahora bien, este fenómeno de la masculinización es un fenómeno histórico clásico de mayor alcance: cuanto más valor adquieren una posición, una tarea, o un oficio y cuanto más se les considera como “objetivos” más aparecen como un asunto de hombres. Por supuesto, no es que las mujeres se hayan hecho repentinamente incompetentes, lo que sucede es que la IA ha subido en el escalafón.

Y, sin embargo, tal y como lo denuncian hoy en día un número cada vez mayor de universitarios, y también de periodistas e incluso de empleados de los GAFAM (Google, Apple, Facebook, Amazon, Microsoft) esa objetividad, esa supuesta neutralidad de la IA, esa supuesta ausencia de género oculta en realidad, en el mundo occidental, un varón más bien blanco de unos treinta o cuarenta años⁷, tal y como lo he evocado un poco más arriba. Ahora bien, podemos subrayar que aquí toda la dificultad consiste en tener cuidado para no animar a las jóvenes a emprender carreras en las que se supervaloran las ciencias y las tecnologías, “duras”, en detrimento de las ciencias humanas y sociales – “blandas” y marcadas como más “femeninas” – lo

⁶ “In subsequent retellings, the women were skipped over repeatedly. In some historical images, the ENIAC Six are captioned as models, if pictured at all. “I wasn’t photogenic,” said Betty Snyder. “I wasn’t included on any of the pictures of the entire stupid thing.” When the army used a War Department publicity shot of the ENIAC for a recruitment ad, they cropped out the three women in the picture entirely. The War Department’s own press releases about the ENIAC cited a vague, genderless “group of experts” responsible for the machine’s operation, and mention by name only John Mauchly, J. Presper Eckert, and Herman Goldstine.” *Broad Band*, New York Portfolio, 2018, p. 51.

⁷ Ver, por ejemplo, el artículo *AI’s white guy problem isn’t going away* <https://www.technologyreview.com/2019/04/17/136072/ais-white-guy-problem-isnt-going-away/>, que es una referencia explícita al de Crawford ya citado: <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>. Consultado el 28 agosto de 2019.

que no sería, en cierto modo, más que otra forma de virilización de la IA bajo la apariencia de una feminización. Es por esta razón por la que, considerando la visión del mundo subyacente a este discurso, se puede sugerir la idea de que las modalidades de una “refeminización” de la IA deben reflexionarse con atención para, paradójicamente, no “masculinizar” a las jóvenes que siguieran esta vía.

En segundo lugar, la cuestión del género de la IA tiene que ser abordada igualmente a través del análisis de la cuestión de los sesgos que yo abordaba de un modo más general un poco más arriba. En efecto, uno de los problemas que plantea este proceso es que los algoritmos están cargados de sesgos que se hallan ligados a esa pretendida “neutralización” de la IA, y que, tal y como se ha visto, podemos analizarlos más bien como una virilización de la IA, y viceversa.

Un ejemplo revelador es el de los sistemas de reconocimiento de voz, los cuales han sufrido – y en gran medida lo siguen haciendo – de sesgos, desde el momento mismo en que salen de un marco de referencia “standard”, “neutro”, es decir, masculino y sin acento. Así, algunos vehículos que tienen un sistema de reconocimiento de voz que permite efectuar tareas como la de hacer una llamada de teléfono, sin utilizar las manos – que siguen fijas al volante – no funcionan correctamente más que para los hombres y no para las mujeres. Más allá del mal funcionamiento del sistema, se plantea la cuestión de la visión del mundo que subyace a esta realidad y de la que es testigo la posición particularmente reveladora de algunos dirigentes de estas empresas. Así, Tom Schalk, vicepresidente de una de las empresas responsables de la concepción de estos sistemas de reconocimiento de voz, ATX⁸, llegó a proponer sesiones de formación destinadas a las mujeres, con el fin de enseñarles a hablar más fuerte hacia el micrófono, de manera que fuera más viril e imperativo. Podríamos aquí retomar el argumento clásico según el cual este fallo no es más que el “reflejo” de la población que ha concebido y desarrollado estos instrumentos. Pero a la vez este ejemplo demuestra que las cosas son más complejas. Tanto más interesantes cuanto que apuntan a dos elementos fundamentales de las relaciones entre género y tecnología: por un lado, la idea de que la voz que da órdenes es necesariamente masculina; por otro, la de que tenemos que adaptarnos a la tecnología (¡para sacar el mejor partido de ella!).

Finalmente, otro ejemplo ilustrativo⁹ es el de las asistentes personales virtuales adornadas con nombres (femeninos) como el de *Siri* (asistente digital de Apple), *Cortana* (la de Windows) o el de *Alexa* (la de Amazon) que pueden cuidar de vuestros más mínimos deseos, preocupaciones, cuestiones de la cotidianidad. Estas asistentes personales son una prueba más de los sesgos significativos de género. A primera vista, podrían parecer completamente acorporales y neutros, situados fuera de la binariedad masculino/femenino, ya que se trata de máquinas desencarnadas y por definición asexuadas. Sin embargo, tal y como lo demuestra todo un conjunto de trabajos, estas asistentes son casi inmediatamente consideradas como pertenecientes al género femenino, puesto que es el género apropiado para los comportamientos de asistencia – en una línea bastante próxima al problema (que acabo de evocar) encontrado en los sistemas de reconocimiento de voz.

⁸ <http://techland.time.com/2011/06/01/its-not-you-its-it-voice-recognition-doesnt-recognize-women/>. Consultado el 18 de agosto de 2019.

⁹ Retomo en parte un análisis desarrollado ya en (Nurock, 2019).

Más aún, tal y como se subraya sobre todo en un informe de la UNESCO¹⁰ estas asistentes personales refuerzan el sistema de dominación patriarcal. “Me pondría colorada si pudiera” (“*I’d blush if I could*”) responde Siri, sumisa, a un interlocutor que le lanza “Hola Siri, eres una golfa” (*Hey Siri, you are a bitch*). Como muestra Hilary Bergen (Bergen, 2016), estas asistentes personales han sido programadas para invitar al flirteo (por así decirlo), hasta en los casos en los que se convierten un acoso agresivo, lo que no deja de ser un problema. Más aún, sus reacciones no son propicias para desarmar este tipo de comportamiento, hasta pueden francamente tender a estimularlo. Hasta el punto de que, incluso si podemos preguntarnos si es necesario hablar con cortesía a un robot que, después de todo, no es una persona, el comportamiento opuesto plantea un problema muy distinto: al interactuar de manera tan violenta con un asistente personal, ¿no se corre el riesgo de asentar ciertas actitudes condenables, de “automatizarlas” en personas que interactúan con estas máquinas? De nuevo, en este caso la cuestión de la neutralidad no se plantea al tratarse de asistentes personales, precisamente por la dimensión de género que lleva inscrita. El modo como son concebidas estas asistentes personales (podríamos decir, estas asistentas personales) refuerza cierta concepción del trabajo doméstico como una característica femenina y a la vez la idea de su calificación como un trabajo de domesticidad a un mismo tiempo virtual e invisibilizado (Schiller y McMahon, 2019).

Retomando a Nora Ni Loideain y Rachel Adams¹¹, para completar este análisis de los (las) asistentes personales, podemos referirnos al mito de Eco, la ninfa del bosque. Esta oréade, incapaz de responder de una forma adecuada, no existe más que para reflejar el narcisismo masculino: ella palidece hasta el punto de parecer incorporal. La referencia al mito de Eco parece tanto más pertinente cuanto que esas asistentas personales artificiales nos remiten no sólo a nuestros propios sesgos – y son así el eco de nuestras sociedades – sino también porque ellas encarnan la figura, claramente marcada por el género, de la comadre: la mujer que difunde a los cuatro vientos todo lo que escucha. Ahora bien, es esto lo que hacen esas asistentes personales, que reenvían los datos y las conversaciones personales a las empresas que las comercializan.

Estas asistentas personales completan así, en varios sentidos, un neocapitalismo particularmente desacomplejado (Woods, 2018) al utilizar su invisibilización como un medio no sólo de vigilancia, sino también de capitalización a través de la utilización y la posible comercialización de los datos. Sin duda alguna, sobre este punto es igualmente útil señalar aquí que es un problema que se da también en las IA sexuales (Lugano y otros, 2017), aunque sea con un estilo algo diferente. El servicio sexual prestado a sus usuarios es doblemente rentable en términos económicos: por el precio de la adquisición y también por el valor de las informaciones proporcionadas, que son únicas por el grado de conocimiento íntimo de sus usuarios que ellas procuran. La relación íntima es así ante todo reconducida a una forma de vigilancia a la que nada escapa y aún menos las formas monetarizadas de lo que podría parecer proceder de las formas de relación más fuertes: la asistencia y el sexo – la mamá y la prostituta, o la mamá tratada como una puta/golfa en el caso de Siri – lo cual

¹⁰ Nos referimos al excelente informe publicado por la UNESCO en 2019: <https://en.unesco.org/ld-blush-if-i-could>. Consultado el 28 de agosto de 2019

¹¹ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3281807

demuestra irónicamente hasta qué punto este tipo de automatización no ayuda a sus usuarios a resolver en modo alguno su complejo de Edipo....

3. ¿La inteligencia artificial puede hacerse cargo del cuidado?

3.1. La IA de cuidado

Después de plantear la compleja conexión entre género e IA bajo la forma de una relación de doble sentido – los sesgos de género son sin duda un reflejo del mundo en el que evolucionan quienes conciben las máquinas de IA, pero son igualmente susceptibles de ser reforzados por estas máquinas –, querría ahora dar un paso más con el fin de explorar la relación compleja entre cuidado e IA a través de dos casos concretos: el de los artefactos de IA de cuidado y el de las máquinas llamadas morales. Entiendo aquí el término de cuidado en el sentido técnico que le han dado las éticas y las políticas del cuidado, el de una forma de ética y de política no imparcial sino relacional, altruista sin abnegación e integrando todos los aspectos de los diversos significados del cuidado (Tronto, 2013): preocuparse por (*care about*), ocuparse de (*care for*), la relación entre dispensador y receptor del cuidado (*care giver/receiver*), pero también el hecho de prestar atención a las estructuras democráticas (*care with*). Apoyaré mi análisis en dos ejemplos que me parecen paradigmáticos de las relaciones entre cuidado e IA: por un lado, el de los artefactos de IA de cuidado que se sitúan en la intersección de lo afectivo y de recibir cuidado, especialmente en la relación parental, y, por otro, el de las “máquinas morales”.

Los campos de la salud y del cuidado orientado a la salud han experimentado un vertiginoso crecimiento de aplicaciones en IA, que van desde niñeras a los artefactos de IA que permiten efectuar diagnósticos médicos. Querría focalizar mi análisis en un punto particular desarrollado por estas IA, en la intersección del campo de la salud y de la relación afectiva con los niños, es decir, los robots y acompañantes IA utilizados en el seguimiento de las afecciones crónicas infantiles. Podemos pensar, por ejemplo, en Meyko o en Robin, que son dos robots/IA muy diferentes.

Robin es un robot Nao dotado de IA, desarrollado en Gran Bretaña y destinado a niños afectados de una diabetes de tipo 1¹². Ha sido programado “como un nene” (*like a toddler*) afectado por una diabetes con el fin de permitir a los niños aplicar los conocimientos que pueden tener sobre la patología de la que se ven afectados, y, a la vez, permitirles identificarse con este robot-juguete que ha sido programado de forma “autónoma”, es decir, ser capaz de adaptarse a situaciones diversas. Meyko es un robot dotado de IA, desarrollado en Francia y destinado a la educación terapéutica y al seguimiento de niños de 3 a 12 años afectados de una enfermedad crónica. Según lo presenta la web de la empresa que lo ha concebido, está destinado a combatir el “cansancio” del niño originado por tratamientos de larga duración y a “mejorar su calidad de vida” (que los padres pueden verificar en línea en la página web de la empresa¹³).

Concretamente, estos dos casos de IA responden a un problema de fondo, que es el del seguimiento de los niños afectados por alguna enfermedad crónica y que son

¹² <http://www.emotion-modeling.info/robin>

¹³ <https://www.hellomeyko.com/fr/>

demasiado jóvenes para ser capaces de gestionar la situación ellos solos. No obstante, mientras que Robin se halla orientado completamente hacia el niño – y aquí se juega con la identificación y una función de juguete –, Meyko está pensado para permitir que los padres deleguen en él parte del seguimiento y también para asegurarse de que el niño ha tomado bien su tratamiento. Asume, por lo tanto, una función de cuidado a la vez en las interacciones – ya que el niño tiene la impresión de que Meyko se preocupa por él (*care about*) – y en el cuidado en sí mismo (*care for*), ya que Meyko es *care-giver* en este doble sentido. Aunque se pueda comprender la importancia de poder descargarse de una parte de la carga mental y concreta de los padres de los niños afectados, aquí el problema reside en la confusión de los dos sentidos del cuidado: no porque Meyko recuerde al niño que tiene que tomar sus medicamentos demuestra que se ocupa de él. La transitividad de una de las formas del cuidado a la otra en el orden artificial es aquí susceptible de plantear problemas, puesto que el tipo de IA se ha configurado con el fin de jugar teniendo en cuenta la circulación entre los dos aspectos del cuidado: el niño tomará el medicamento “para complacer” a Meyko, que no se preocupa por él, y corre el riesgo de confundir el programa de la IA con la solicitud de sus padres. Actuando de esta manera se hace como si esos dos sentidos del cuidado fueran idénticos, mientras que no lo son, lo cual podría causar confusión en el niño a poco que su único recuerdo para tomar el tratamiento se deba a Meyko. Afortunadamente no es el caso y estos dispositivos son concebidos hoy en día sólo como refuerzos, por más que el caso de Meyko se encuentre un poco en el límite. Sin embargo, si hubieran sido desarrollados hasta el punto de reemplazar la solicitud paternal o si se hallaran presentes tempranamente en el desarrollo del niño para suplir, o incluso reemplazar, la solicitud parental, entonces el problema se plantearía en toda su dimensión ya que un uso semejante sería no sólo susceptible de modificar las relaciones (entre hijos y padres y a la vez entre niños y máquinas), operando al mismo tiempo una confusión entre los dos aspectos distintos del cuidado que pueden estar entrelazados (yo me ocupo de ti porque me preocupo por ti) pero no necesariamente en la medida en que el “ocuparse de” (*take care/care for*) puede ser parcialmente automatizado (incluso si la calidad del cuidado pueda ser cuestionada), lo que no es el caso de la solicitud (*care about*).

3.2. Las “máquinas morales”

El segundo ejemplo sobre el que me gustaría reflexionar es el de las “máquinas morales”. Aunque la tentación de automatizar la moral es recurrente desde la época clásica en filosofía (Nurock, 2019), la idea de construir una “máquina moral”¹⁴ conoce hoy en día un renacido interés en el sector de las ciencias cognitivas y, más en particular, en el de la IA, sobre todo a través de dos aplicaciones de gran entidad como son los vehículos sin conductor y los drones.

Desde el comienzo de los años 2000 las ciencias cognitivas han intentado proponer modelos de cognición moral. Uno de los instrumentos de los que se ha servido para hacerlo ha sido el de la “trolleylogía” (de trolebús): el uso de dilemas morales inspirados en Philippa Foot (Foot, 1967) y popularizados por John Mikhail (Mikhail, 2011) en ciencias cognitivas y en teoría del derecho: pensemos en el famoso ejemplo del trolebús, que se puso muy pronto de moda, según el cual se debe elegir entre

¹⁴ <https://www.hellomeyko.com/fr/>

dejar morir a varias personas sobre las que se lanza un trolebús y sacrificar a una persona para salvar a más.

Usar estos dilemas para probar nuestro sentido moral e incluso para programar “máquinas morales” plantea un buen número de problemas. Me conformaré con abordar específicamente los problemas éticos. Estos son, a un mismo tiempo, teóricos y empíricos. Desde un punto de vista concreto, se puede señalar que el modo como son presentadas las situaciones tiene poco que ver con aquellas a las que debería hacer frente un conductor, que se encontrará con personas aplastadas en el caso de un vehículo sin conductor – cosa que no sucederá en el caso de un dron, dado que el hecho de que se disponga de una palanca nos sitúa más cerca de la realidad.

En el plano teórico, parece delicado, incluso francamente aberrante, utilizar dilemas morales para encontrar las respuestas «correctas». En efecto, por definición, un dilema pone en escena dos situaciones que son igualmente buenas o igualmente malas, pero que no se pueden resolver al mismo tiempo. El problema proviene de esta igualdad, pues no hay ni buena respuesta ni buena razón para elegir una antes que la otra, y de ahí proviene la dificultad creada por el dilema.

Si, antes de la trolleylogía los dilemas han sido planteados en psicología moral por el gran psicólogo de la segunda mitad del siglo XX Lawrence Kohlberg, es precisamente porque, como no existía la posibilidad de una buena respuesta, el uso de los dilemas permitía precisamente focalizarse en los argumentos utilizados por los sujetos más que en la respuesta que daban. El fin no era automatizar la moral sino más bien lo contrario: se trataba de reflexionar sobre las modalidades de desarrollo moral con el fin de permitir un refuerzo de la educación moral, lo cual es algo muy diferente.

Por otra parte, incluso aunque no fueran utilizados para encontrar la buena respuesta, el uso de los dilemas para analizar el desarrollo moral de los individuos ha sido muy criticado por Carol Gilligan (Gilligan, 1982), que era en sus orígenes una colaboradora de Kohlberg. Al haber trabajado en el marco kohlbergiano, Gilligan demuestra que el uso de los dilemas obligaba a los individuos a elegir lo inelégible y les encerraba en una alternativa insostenible. Operando de esta manera el marco no permitía tener en cuenta a las personas que trataban de salir del dilema y de encontrar una vía distinta, precisamente porque el dilema es un callejón sin salida. Es el caso paradigmático de una niña de 11 años, Amy, que rehúsa escoger una opción y sostiene, más bien, que hay que poner a dialogar a los diferentes protagonistas (en lugar de obligarles a ser antagonistas) y tener en cuenta la dimensión situada de sus puntos de vista.

Apoyándose en esas declaraciones que aparecían en sus estudios, pero a las que el marco kohlbergiano negaba el derecho de ciudadanía, Gilligan ha propuesto la idea de que la ética no podía ser considerada en referencia al bien y al mal, ni a lo justo o a lo injusto, sino que debe integrar también el cuidado. En una palabra, esta ética, que no compete necesariamente con perspectivas diferentes, insiste más en las relaciones que en la imparcialidad y en una forma de altruismo sin abnegación. Esta ética no es femenina – por más que se la prescriba con frecuencia a las mujeres – sino *feminista* en el sentido de que rechaza las estructuras y los valores del patriarcado y procede de una exigencia democrática.

El problema que pone sobre la mesa la perspectiva del género en la IA no es, por lo tanto, solo el problema coyuntural de los sesgos de género. Se trata también de un problema estructural: son las visiones del mundo subyacentes a la estructura del pro-

ceso de programación (incluso hipotética, como sucede en el caso de los vehículos sin conductor) las que deben cuestionarse. Si no se actualizan, si no son puestas en cuestión y si no se cambian, poco importa quién sea la persona efectivamente encargada de recoger los datos, ya que, de todos modos, esos datos se hallarán integrados en una estructura patriarcal.

Efectivamente, en el ejemplo de los vehículos autónomos, el uso de dilemas para intentar elaborar una máquina moral no solamente reproduce los binomios dominantes (incluso en el terreno de la ética) en nuestra sociedad, sino que corre también el riesgo de reforzarlas. Y ese es precisamente el riesgo que quiero explorar a continuación.

4. ¿Hacia una *artificialización* de la ética?

Creo que observar la IA con las gafas del cuidado permite señalar un peligro mayor, que he propuesto denominar la *artificialización* de la ética¹⁵ (Nurock, manuscrito inédito).

Utilizo el término *artificialización* en referencia al término ‘naturalización’, no tanto en el sentido que reviste en ciencias cognitivas (el de remitir a las capacidades biológicas, lo que, de entrada, no parecería deberse hacer), sino más bien en el sentido que le da Bourdieu (Bourdieu, 1998) al término, sobre todo en su libro sobre *La dominación masculina*. Bourdieu entiende por “naturalización” la incrustación de las relaciones de dominación (sobre todo de género) en nuestros hábitos de pensamiento, por su asimilación a un fenómeno natural, entendido en el doble sentido de normal y de biológico. Esta asimilación se opera gracias a un movimiento de esencialización acompañado de una deshistorización. Mi hipótesis es que la artificialización se apoya en la naturalización para adelantar un paso más en el anclaje en el código – incluso en el caso de que el *deep learning* pudiera parecer prevenir una forma de fijación – de los hábitos estructurales, sosteniendo la idea de que, si son artificializadas, entonces son imparciales y, por lo tanto, moralmente buenas. Existiría así la posibilidad de una fosilización estratificada en la que, de algún modo, la artificialización se añadiría a la naturalización.

El problema que se plantea al analizar la IA a través del prisma del género tiene un doble filo: por una parte, la IA refleja los sesgos de quienes dominan esta concepción e implica discriminaciones, que son el reflejo de los sesgos que caracterizan a nuestras sociedades. Pone frente a nosotros un espejo. Por otra, esos sesgos son susceptibles de verse no sólo reforzados sino también ratificados y hasta justificados por la IA. De modo que somos nosotros quienes nos transformamos para adaptarnos aún más a la imagen que la IA, nuestro espejo, nos reenvía de nosotros mismos – así, lo hemos visto, se ha podido sugerir a las mujeres que masculinicen su voz, para hacerlas, de algún modo, dignas de ser obedecidas, ¡todo ello con el fin de que sus órdenes sean tenidas en cuenta por las máquinas de reconocimiento de voz! El anverso y el reverso de la medalla son ciertamente distintos, pero, quiérase o no, participan de la misma dinámica.

¹⁵ He propuesto un primer análisis de este riesgo en (Nurock, 2019), que retomo aquí en grandes líneas, y que he desarrollado en un artículo titulado « The Artificialistic Fallacy » (de próxima aparición).

Aquí el problema viene al mismo tiempo de la colisión entre la pretensión de imparcialidad con la artificialización y del modo como consideramos la IA, es decir, de las propiedades que le atribuimos. Como atribuimos a la IA la característica de imparcialidad, relacionada con su pretendida dimensión de neutralidad, cuando, según hemos visto, oculta de hecho una virilización, le permitimos con ello artificializar propiedades sociales particularmente cargadas de desigualdad y cuestionables desde un punto de vista tanto ético como político. Este fenómeno no viene de la IA en cuanto tal sino, a un mismo tiempo, del hecho de que se convierte en un espejo de nuestras sociedades y de que hacemos como si este espejo no fuera descriptivo sino potencialmente normativo, y hasta prescriptivo, en el sentido de que puede decirnos imparcialmente lo que tenemos que hacer.

De este modo, uno de los riesgos mayores del uso de estos dilemas morales para programar esas “máquinas morales” es el de artificializar la ética grabando en el código una determinada concepción de la misma, que podríamos, consiguiendo, considerar como neutra, imparcial, y hasta incluso incontestable. Lo cual no es el caso, según acabamos de ver, particularmente a causa de las críticas que tienen como base la ética del cuidado. Corremos el riesgo de que este problema se replantee cuando, como sucede hoy en día, se intenta introducir la IA en diferentes campos sociales y políticos, sobre todo cuando se trata de la asignación de recursos (Eubanks, 2018).

En conclusión, se comprende que el anverso de la medalla, es decir, la reproducción de las estructuras de dominación, que traducen, a la vez, la desaparición de las mujeres (*a fortiori* a unas posiciones no subalternas) en el campo digital y la feminización de los artefactos de IA relegados a posiciones subalternas, no es más que la otra cara de la moneda de estructuras binarias, patriarcales, que sustentan actualmente, desde muchos puntos de vista, los desarrollos de la IA. El problema no es, en este sentido, que la IA no se encuentre suficientemente feminizada en el sentido de que no encontremos en su entorno suficientes mujeres – a pesar de que, obviamente, sea parte del problema. El problema, sobre todo, es que la IA reproduce, refuerza y conlleva el riesgo de autonomizar unas estructuras sociales de dominación. No basta con diversificar los *data* o incorporar más mujeres al mundo digital, es la infraestructura misma lo que hay que repensar de manera crítica y reconstruirla.

Pero plantear la cuestión del cuidado en relación con la IA implica también ir más allá y preguntarse por la relación que estamos dispuestos a delegar en la IA, incluso las relaciones que estamos dispuestos a construir con los artefactos de IA desplegados en el ámbito de lo íntimo – tanto más cuanto que algunas investigaciones intentan hoy promover la posibilidad de concebir artefactos de IA dotados de empatía, y hasta de concederles derechos (Gunkel, 2018). Preguntarse ¿puede la IA cuidar? (“*can AI care?*”) supone también subrayar hasta qué punto es importante plantearse la cuestión de saber cómo concebimos la IA, qué propiedades le atribuimos y por qué. Dicho de otra manera, lo que debe plantearse es la cuestión de las modalidades de nuestra relación con la IA, más que la de su pretendida neutralidad o imparcialidad (que, en realidad, según hemos visto, son falsas).

Entonces, ¿puede la Inteligencia Artificial prestar cuidados? No es ese su papel. Mientras que el papel que nos concierne es el de estar atentos (*careful*) cuando concebimos y utilizamos la IA y cuando establecemos algún tipo de relación con ella,

con el fin de clarificar las propiedades que le asignamos, las actividades que delegamos en ella, y la manera que tenemos de considerarla. No se trata de ser tecnóforos, puesto que la IA puede ser una verdadera oportunidad para desarrollar algunos campos y tomar el relevo de algunas operaciones fácilmente automatizables, pero se corre el riesgo de estrechar el campo de nuestra vida moral al llevar a cabo una forma de artificialización basada en una falsa automatización. Esa es la razón por la que debemos prestar atención cuando desarrollamos y promovemos unos aspectos de la IA y no otros. He ahí por qué una reflexión ética sobre la IA en todo su proceso, desde la concepción a sus desarrollos y a sus usos, es hoy en día más indispensable y urgente que nunca.

5. Referencias bibliográficas

- Adam, A. (1998). *Artificial Knowing: Gender and the Thinking Machine*. Londres y N.York: Routledge.
- Bergen, H. (2016). “‘I’d blush if i could’: Digital assistants, disembodied cyborgs and the problem of gender”. *Word and Text. A Journal of Literary Studies and Linguistics*. 6: 95-113..
- Bourdieu, P. (1998). *La Domination Masculine*. París: Seuil.
- Buolamwini, J. & Gebru, T. (2018). “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”. *Proceedings of Machine Learning Research*. 81: 1-15.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Nueva York: Picador, St Martin’s Press.
- Foot, P. (1967). “The Problem of Abortion and the Doctrine of Double Effect”. *Oxford Review*. 5: 204-217.
- Gilligan, C. (1982). *In a Different Voice*. Cambridge, M.A.: Harvard University Press.
- Greenspan, P. (2005). *Practical Guilt: Moral Dilemmas, Emotions, and Social Norms*. s.l.:Oxford University Press.
- Gunkel, D. (2018). *Robot Rights*. s.l.:MIT Press.
- Kohlberg, L. (1981). *Essays on Moral Development, Vol. I: The Philosophy of Moral Development*. San Francisco, CA: Harper & Row.
- Lugano, G., H. M. I. M. & Oveček, T. (2017). “From the Mind to the Cloud: Personal Data in the Age of the Internet of things”. En: Y. Zhou & M. H. Fischer, eds. *AI love you* (pp.111-131). Cham: Springer.
- Mikhail, J. (2011). *Elements of Moral Cognition: Rawls’ Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press.
- Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.
- Nurock, V. (2007). “L’enfance morale: développement moral et éducation morale”. *La Revue Philosophique de Louvain*. 105(1-2): 132-160. doi: 10.2143/RPL.105.1.2020260
- Nurock, V. (2011). *Sommes-nous naturellement moraux*. París: Presses Universitaires de France.
- Nurock, V. (2019). “Généalogie de la morale mécanisée”. Dans: *Robots et sociétés: enjeux éthiques et politiques* (pp.31-50). Québec: Les Presses de l’Université Laval.
- Nurock, V. (2019). “L’Intelligence Artificielle a-t-elle un genre?”. *Cités*. 80: 61-74. doi. org/10.3917/cite.080.0061

Nurock, V. Manuscrito inédito. *The Artificialistic Fallacy*. s.l.:s.n.

Schiller, A. & McMahon, J. (2019). "Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence". *New Political Science*. 41(2): 173-191. doi.org/10.1080/07393148.2019.1595288

Tronto, J. (2013). *Caring Democracy*. NYC: NYU Press.

Woods, H. S. (2018). "Asking more of Siri and Alexa: Feminine persona in service of surveillance capitalism". *Critical Studies in Media Communication*. 35(4): 334-349. doi.org/10.1080/15295036.2018.1488082