

INTELIGENCIA ARTIFICIAL Y ROBÓTICA SOCIAL

el poder de las apariencias

ARTIFICIAL INTELLIGENCE AND SOCIAL ROBOTICS

the power of appearances

<https://doi.org/10.26512/rfmc.v13i2.57815>

Darío Sandrone*

Universidad Nacional de Córdoba

<https://isep-cba.edu.ar/web/autores-98/>

<https://orcid.org/0000-0003-0201-6391>

dariosandrone@unc.edu.ar

* Profesor y Doctor en Filosofía por la Universidad Nacional de Córdoba

RESUMEN

En este artículo nos proponemos poner en relación dos campos de estudio científico y tecnológico, por un lado, el de la Inteligencia Artificial, por el otro, el de la Robótica social, con el propósito de abordar el estado actual del problema de la imitación tecnológica de lo humano. Desde sus comienzos, a mediados del siglo XX, el campo de la Inteligencia artificial se encontró atravesado por este problema, que generó debates en torno a si las máquinas inteligentes debían tomar como modelo a la inteligencia humana, o si solo debían aspirar a imitar sus productos. En paralelo, la Robótica, cuyo origen está marcado por un componente ficcional y especulativo, comenzó a consolidarse como una disciplina científica con desarrollos tecnológicos concretos que incorporan los productos de la Inteligencia Artificial. En consecuencia, la Robótica también se vio atravesada por la disyuntiva entre hacer robots al estilo humano o, por el contrario, generar seres técnicos que solo realicen las tareas humanas. El acelerado desarrollo de la robótica amplió la problemática de los aspectos intelectuales a la apariencia de los cuerpos. Queremos presentar aquí un recorrido histórico de esta articulación y brindar, a la vez, un panorama actual de esta problemática, sobre todo a partir de la consolidación en los últimos años de la Robótica Social.

Palabras clave: Inteligencia Artificial. Robótica social. Conocimiento. Apariencia. Cuerpo. Imitación.

Abstract: In this article we propose to relate the fields of scientific and technological study, on the one hand, to Artificial Intelligence, on the other, to Social Robotics, with the purpose of addressing the current state of the problem of technological imitation of the human. Since its beginning, in the middle of the twentieth century, the field of artificial intelligence has been plagued by this problem, which has generated debates around whether intelligent machines should take human intelligence as a model, or whether they should aspire to imitate their products. At the same time, Robotics, whose origin is marked by a fictional and speculative component, began to consolidate itself as a scientific discipline with concrete technological developments that incorporate Artificial Intelligence products. As a result, Robotics also found itself crossed by the divide between making robots in a human style or, on the contrary, generating technical beings who only perform human tasks. The accelerated development of Robotics expanded the problem of intellectual aspects to the appearance of bodies. We want to present here a historical tour of this articulation and provide, at the same time, a current panorama of this issue, especially based on the consolidation in recent years of Social Robotics.

Keywords: Artificial Intelligence. Social Robotics. Knowledge. Appearance. Body. Imitation..

IA: LAS DOS AGENDAS

El tópico acerca de la réplica, emulación y eventual sustitución de seres humanos a través de máquinas, a pesar de ser muy antigua, se ha centrado en las últimas décadas en los aspectos intelectuales. Este recorte se ha llevado a cabo sobre todo a partir de la emergencia de una máquina inédita en la historia humana, la computadora electrónica, que desde su aparición a mediados del siglo XX estimuló un área de investigación conocida como Inteligencia Artificial (IA)¹. Por un lado, se encuentran quienes afirman que la inteligencia artificial puede equiparar a la inteligencia humana, a la cual se tiene como modelo. Por otro lado, están quienes sostienen que la inteligencia artificial no necesariamente debe parecerse a la inteligencia humana, sino que puede adquirir rasgos propios y singulares. Luciano Floridi formula esta polarización en términos de las dos almas de la IA (Floridi, 2024, p. 80). Una es el *alma ingenieril*; la otra el *alma cognitiva*. Ambas se disputan “el predominio intelectual, el poder académico y los recursos financieros” (Floridi, 2024, p. 80). Según la caracterización de Floridi, el alma ingenieril tiende a ser reproductivista, esto es, a “«reproducir» mediante recursos no biológicos los *resultados exitosos* de nuestro comportamiento inteligente humano” (Floridi, 2024, p. 77). De esta definición debemos resaltar el término “resultados”, pues lo que le interesa a la línea reproductivista es lograr que los sistemas informáticos consigan los mismos productos que los humanos sin necesidad de que lo hagan al *estilo humano*. Al alma ingenieril no le interesa replicar la inteligencia humana, sino crear sistemas artificiales que garanticen sus logros o los superen. Por otro lado, el alma cognitiva de la IA intenta “«producir» el equivalente no biológico de nuestra inteligencia, es decir, la *fuentes* de dicho comportamiento” (Floridi, 2024, p. 77). Para este enfoque, la inteligencia humana es histórica y puntual, no es un arquetipo al que hay que aspirar. Es apenas lo que la evolución biológica ha podido conseguir hasta ahora. Si se puede dar con la fuente de la inteligencia, es posible producir una inteligencia más potente aún. En este sentido, Floridi caracteriza al alma ingenieril

¹ “Para que la inteligencia artificial pueda llegar a ser una realidad se necesitan dos cosas: inteligencia y un artefacto. El computador ha sido el artefacto elegido. El computador electrónico digital moderno...” (Russell y Norvig, 2004, p. 16).

reproductivista como aquella que quiere desarrollar “tecnologías inteligentes”, y al alma cognitiva productivista, en cambio, como aquella que quiere desarrollar “tecnologías verdaderamente inteligentes”. Aquí, el “verdaderamente” puede entenderse como “no solamente humano”.

Las dos almas de la IA han adoptado denominaciones diversas y algo incoherentes. A veces, se utilizan distinciones entre “IA débil” frente a “IA fuerte” o “IA a la vieja-usanza” (Good Old-Fashioned AI) frente a «nueva IA» (New o Nouvelle AI) para captar esta diferencia conceptual. Hoy en día se encuentran de moda expresiones como “inteligencia general artificial” (AGI) o “IA universal”, en lugar de “IA plena”. En el pasado, he preferido usar la distinción menos cargada de connotaciones entre “IA ligera” vs “IA fuerte” (Floridi, 2024, p. 80).

Floridi señala que existe una diferencia sustancial en términos de resultados concretos entre la primera y la segunda alma. El alma ingenieril “ha tenido un éxito asombroso, por encima de las expectativas más optimistas” (Floridi, 2024, p. 77), mientras que el alma cognitiva interesada en la producción de la inteligencia “sigue siendo ciencia ficción y ha supuesto una decepción lamentable” (Floridi, 2024, p. 77)^{II}. Asimismo, el éxito real de la IA contemporánea “no consiste en producir inteligencia humana sino en sustituirla” (Floridi, 2024, p. 80).

Por su parte, Russell y Norvig (2004), sostienen una distinción similar a la de Floridi, solo que en lugar de la búsqueda de tecnologías “verdaderamente inteligentes”, los autores utilizan el término “racional” como distintivo frente a la inteligencia humana. Según estos autores “un sistema es racional si hace «lo correcto», en función de su conocimiento” (Russell y Norvig, 2004, p. 2). Desde el punto de vista de estos autores, los humanos somos parcialmente racionales, pues existen determinados condicionamientos que nos limitan para hacer y pensar lo correcto. Muchas veces hacemos lo que nos conviene, lo que nos gusta, o lo que

^{II} Nos interesa particularmente la alusión a la ciencia ficción, pues como veremos más adelante, es una cuestión central.

deseamos. A veces hacemos lo que es bueno, o lo que es bello, aunque no sea correcto. En otras ocasiones razonamos sesgados por nuestra crianza, por la tradición y por las costumbres del lugar donde nacimos. Otras veces, simplemente cometemos errores de razonamiento. Esto no quiere decir que los humanos tomemos decisiones arbitrarias. Existe una explicación de las decisiones que tomamos, pero esa explicación no siempre se basa en un razonamiento correcto.

Conviene aclarar, que al distinguir entre comportamiento *humano* y *racional* no se está sugiriendo que los humanos son necesariamente “irracionales” en el sentido de “inestabilidad emocional” o “desequilibrio mental”. Basta con darnos cuenta de que no somos perfectos: no todos somos maestros de ajedrez, incluso aquellos que conocemos todas las reglas del ajedrez; y desafortunadamente, no todos obtenemos un sobresaliente en un examen (Russell y Norvig, 2004, p. 2).

Aceptando esta distinción: ¿Debemos diseñar una IA que piense como los humanos o que sea racional? Russell y Norvig señalan que existen investigadores y desarrolladores de los dos bandos, y que a lo largo de la historia han seguido cuatro enfoques: por un lado, “Sistemas que piensan como humanos” y “sistemas que piensan racionalmente”; por otro lado, “sistemas que actúan como humanos” y “sistemas que actúan racionalmente” (Russell y Norvig, 2004, p. 2). A los investigadores y desarrolladores que optan por la inteligencia y el comportamiento al estilo humano, les interesan las formas fácticas con las que piensan los humanos concretos; a los que persiguen sistemas racionales les interesa identificar y formalizar las “leyes del pensamiento”.

La distinción entre pensar y comportarse ya está, de alguna manera, en el artículo de Turing (1950). Precisamente, su test no tiene nada que ver con la forma de pensar, de hecho, afirma que la pregunta “¿Pueden pensar las máquinas?” le parece absurda y propone reemplazarla por la pregunta “¿Qué sucede cuando una máquina sustituye a A [humano] en el juego?” (Turing, 1950, p. 433). Lo importante no es “cómo piensa” la inteligencia artificial sino cómo se comporta, y si los agentes humanos pueden interactuar con esta de manera fluida, sin darse cuenta que

los enunciados provienen de una máquina. Se podría argumentar que para actuar como humano es necesario pensar y razonar como humano, aunque esto es más bien discutible. Justamente, el concepto de imitación propuesto por Turing nos libera de tener que defender ese principio: puedo imitar a un perro sin razonar como uno. Lo mismo estarían en condiciones de hacer las computadoras en relación con los humanos.

Por su parte, Erik Larson (2022) distingue dos corrientes en las que agrupa a los académicos y desarrolladores de la IA contemporáneos. Por un lado, se encuentran los que podríamos denominar “shakespeareanos”: son “mitólogos de la IA, [que] se entusiasman con la idea de que las máquinas posteriores a la singularidad vayan de disponer de conciencia, emociones, motivos y una vasta inteligencia. Con gran ironía mantienen viva la exploración filosófica trasladando temas shakesperianos a la informática” (Larson, 2022, p. 104). La principal figura de esta corriente es Raymond Kurzweil, un informático y escritor estadounidense que tuvo un paso como director de Ingeniería en Google. Kurzweil asegura que la IA superará el test de Turing en el año 2029, demostrando tener una mente (inteligencia, consciencia de sí mismo, riqueza emocional...) indistinguible de un ser humano. La alusión a Shakespeare proviene de que este tipo de planteos no solo estimulan la especulación filosófica, si no también la creación de dramas al estilo clásico, que pueden verse en películas como *Her* (2013), en la que un hombre se enamora de la inteligencia artificial en su computadora, o *Ex Machina* (2014), un culebrón de amor, sexo y muerte entre humanos y robots con inteligencia artificial. Esta corriente corresponde al alma cognitivista propuesta por Floridi.

Por otro lado, según la distinción de Larson, se encuentran los “russellianos”, en honor a Stuart Russell, a quien ya hemos mencionado. Esta corriente no cree que las máquinas puedan tener experiencias subjetivas, y pretenden “reducir la conversación de la superinteligencia a las ideas más respetables en lo matemático sobre una informática general que alcance «objetivos»” (Larson, 2022, p. 105). Como hemos visto más arriba, Russell sugiere que la racionalidad, antes que las emociones y pasiones humanas, es el valor más elevado. En ese sentido, posee una noción restringida de inteligencia humana, en la que su propósito es

solucionar problemas y conseguir objetivos. Una superinteligencia artificial no sería para ellos una máquina que sienta amor o deseo, si no una que pueda obtener logros y solucionar problemas a un nivel que los humanos no pueden hacerlo. Para Larson, ambas corrientes comparten un supuesto de base: en la interacción humano-máquina, se centran en la máquina.

Tanto los kurzweilianos como los russellianos proclaman una visión tecnocéntrica del mundo que a la vez simplifica la visión de la gente -en especial, con sus enfoques deflacionarios sobre la inteligencia como una forma de computación- y expande las visiones de la tecnología al promover un futurismo dedicado a la IA como ciencia y no como mito (...) Vemos posibilidades ilimitadas para las máquinas pero un horizonte restringido para nosotros (Larson, 2022, p. 105).

Por último, podemos incorporar la distinción que realizan en *The Nooscope Manifested* (2021), el filósofo italiano Matteo Pasquinelli, y el artista-activista digital, Vladan Joler. Para ellos, la inteligencia artificial tiene que ver con el proyecto racionalista, pero no en término de generar agentes racionales (como plantea Russell), sino un instrumento capaz de amplificar la racionalidad humana, como una suerte de telescopio de la razón. Gottfried Leibniz, que creía que la razón era más importante que la vista para conocer el mundo, trabajaba en un lenguaje universal que permitiera reducir la enorme variedad de ambiguas palabras y difusas ideas a unos pocos símbolos, de manera que el pensamiento se convirtiera en una actividad tan simple, mecánica y precisa como el cálculo matemático. Este instrumento, pensaba Leibniz, amplificaría las capacidades mentales que sirven para identificar patrones, tendencias y correlaciones entre los hechos del mundo. Para Pasquinelli y Joler, el instrumento imaginado por Leibniz se ha materializado en el siglo XXI, a través de la inteligencia artificial, el aprendizaje automático (machine learning) y la fenomenal extracción y acumulación de datos sobre personas y acontecimientos. Siguiendo con la analogía leibniziana con el microscopio, los autores comparan el flujo de información, que se obtiene con las *machines learning*, con un potente haz de luz, y los modelos estadísticos ejecutados por los algoritmos a escalas y velocidades

que la mente humana no podría, con una lente a través de la que podemos aprehender esa información. A través de este instrumento no se *ve* mejor el mundo, en cambio, se visualiza mentalmente con mayor claridad su dinámica. En definitiva, señalan los autores, es un *nooscopio* (del griego “noos”, que significa inteligencia). Al caracterizar la inteligencia artificial como un instrumento, los autores niegan que posea algún grado de existencia por fuera de los objetivos y las operaciones de la inteligencia humana. Rechazan el mito de la autonomía tecnológica que “insinúa las caricaturescas «mentes alienígenas» que se auto-reproducen en silencio” (Pasquinelli y Joler, 2021, p. 1). Para los autores, la operación ideológica de postular un agente inteligente artificial equivalente al humano cumple con el objetivo político de mistificar los procesos reales de alienación: “el crecimiento de la autonomía geopolítica de las compañías hi-tech y la invisibilización de la autonomía de los trabajadores a escala global” (Pasquinelli y Joler, 2021, p. 1). Para evitar esto, los autores proponen que hay que realizar la operación opuesta, secularizando la IA, negando que sea un proyecto que trate de construir “máquinas inteligentes” y postulando que lo que se intenta es construir un instrumento de conocimiento, susceptible de tener todas las fallas y distorsiones de cualquier instrumento.

De este primer recorrido nos interesa quedarnos con dos elementos. Por un lado, plantear la existencia de dos agendas académicas y tecnológicas al respecto de la relación maquina-humano, que en los tiempos que corren se han concentrado en la relación inteligencia artificial-inteligencia humana. Una de esas agendas plantea a las máquinas con un estatus ontológico propio, como un ser que *no puede* o *no debe* ser como el humano; la otra plantea la posibilidad real de que las máquinas puedan ser como los humanos. El otro aspecto que nos interesa señalar es que esta segunda agenda está asociada al mito, a la ciencia ficción o a la ideología política. Este punto es importante porque nos obliga a analizar qué rol cumple la ficción y el mito en estos debates. Más aún, qué rol cumple efectivamente en los desarrollos de inteligencia artificial. Nuestra intuición es que la ficción no debe entenderse por la vía negativa, como no-realidad, sino que constituye una heurística y una metodología de diseño positiva en tales desarrollos. Para comprender mejor este punto, es clave incorporar la figura del *robot* en este recorrido.

ROBÓTICA COGNITIVA, ROBÓTICA ECOLÓGICA Y ROBÓTICA DEL DESARROLLO

La palabra *robot* fue acuñada en la obra teatral *R. U. R. (Robots Universales Rossum)*, publicada en 1920 por el escritor checo Karel Čapek (2022). Se trata de la historia de una mega empresa que logra fama mundial al fabricar trabajadores artificiales, los “roboti”, exportados a todo el mundo desde la isla en donde la fábrica tiene sus instalaciones. Su fundador, el viejo Rossum, era un científico que había descubierto la manera de organizar la materia de manera de crear vida, y se propuso crear seres artificiales: “hacer todo como en el cuerpo humano, hasta la última glándula” (Čapek, 2022, p. 13). Sin embargo, el proceso era lento y no siempre salía bien. Luego de su muerte, tomó las riendas de la compañía el joven Rossum, su sobrino, que no era científico sino ingeniero. Su primera impresión fue la siguiente: “Esto no tiene sentido! Diez años para fabricar un hombre. Si no lo puedes producir más rápido que la naturaleza, no sirve para nada” (Čapek, 2022, p. 15). Se puso a estudiar el cuerpo humano y llegó a la conclusión de que era demasiado complicado. Un buen ingeniero podía simplificarlo, así que “comenzó a rediseñar la anatomía, para ver qué se podía omitir” (Čapek, 2022, p. 15). Su objetivo fue llevar a la empresa a otro nivel de productividad simplificando el producto, es decir, a los trabajadores: “Un ser humano es algo que, digamos, siente alegría, toca el violín, quiere salir a pasear y, en general, necesita hacer un montón de cosas realmente innecesarias” (Čapek, 2022, p. 15). Al menos innecesarias “cuando uno tiene que tejer o fabricar cosas”. El joven Rossum inventó el trabajador con menor número de necesidades. “Finalmente, se deshizo del hombre y creó el robot” (Čapek, 2022, p. 15).

El hecho de que la figura del robot surja de una obra teatral nos da la pauta del origen ficcional de esta figura. Por otro lado, en esta obra inaugural podemos encontrar las dos agendas actuales de la IA que mencionamos en la sección anterior. Por un lado, lo que Floridi llama el alma cognitiva y productiva, que Larson caracteriza como “shakespeareanos”, y que se referencian en el viejo Rossum, cuya aspiración era crear seres artificiales que sean réplicas del ser humano, con sus cuer-

pos y sus afectos. Por otro lado, el alma ingenieril y reproductivista, los “russellianos”, que se referencian en el joven Rossum, y cuyo principal propósito es que los robots reproduzcan exitosamente las tareas que realizan los humanos, prescindiendo de los rasgos humanos innecesarios, como sentir alegría o tocar el violín.

Más allá de este origen, en pocos años el robot ingresó al plano científico y al horizonte de los desarrollos tecnológicos como una posibilidad real. El mismo Turing especuló en reiteradas ocasiones con una computadora provista de un cuerpo similar al humano, con órganos sensores y efectores para interactuar con el entorno. Asumió, además, que esta era una condición para la cognición de las máquinas.

Una gran razón positiva para creer en la posibilidad de hacer maquinaria pensante es el hecho de que es posible hacer maquinaria para imitar cualquier pequeña parte de un hombre. Que el micrófono haga las veces de oído, y la cámara de televisión haga las veces de ojo, son lugares comunes. También se pueden producir robots controlados a distancia cuyas extremidades equilibran el cuerpo con la ayuda de servomecanismos” (Citado en Copeland, 2004, p. 420).

Turing creía que, incluso, se le podría incorporar “altavoces y ruedas” de manera que el robot pudiera “vagar por el campo” para aprender cosas por sí mismo (Citado en Copeland, 2004, p. 420). Como vemos, Turing, al igual que Čapek lo hizo en la ficción, postuló al robot como una réplica o imitación del ser humano, pero a diferencia de aquel no enfatizó las virtudes del robot para trabajar sino para conocer e interactuar inteligentemente con el mundo circundante. La concepción del robot como un ser técnico capaz de conocer y adquirir un comportamiento marcaría el objetivo principal de lo que posteriormente se denominó *robótica cognitiva* (Bisconti, 2001) o *robótica centrada en la cognición* (Dautenhahn, 2007, pp. 683-684). La orientación de este enfoque, que es también una línea de desarrollo al interior de la robótica contemporánea, enfatiza al robot como un sistema inteligente (en el sentido tradicional de la IA) y apunta al desarrollo de arquitecturas para el aprendizaje automático y la resolución de problemas (Dautenhahn,

2007, p. 682). Esto no necesariamente implica que el robot deba tener experiencias subjetivas similares a las humanas. Puede ser un complemento de la cognición humana que no necesariamente debe simularla, sino que puede, incluso, a través de otras operaciones, ampliarla. Esto supone tomar al robot como un instrumento al estilo que, como vimos, proponen los autores del *nooscopio*.

Por el contrario, si la cognición robótica toma como modelo a la humana y pretende replicarla, es inevitable que herede los problemas gnosológicos de la relación mente y cuerpo. El conocimiento humano se compone de elementos intelectuales, como el razonamiento lógico y los cálculos matemáticos, así como de elementos corporales, como la percepción a través de los sentidos y los procesos cerebrales. Así, la concepción del ser humano de la que se parte, sobre todo la manera en que se vincula la mente y el cuerpo, es fundamental para establecer el modelo al que aspira el diseñador de robots cognitivos. Desde que la robótica se estableció como disciplina ingenieril, en la segunda mitad del siglo XX, las relaciones entre la inteligencia artificial y el cuerpo artificial no han estado determinadas de manera definitiva y atravesaron varias etapas. En los inicios de la IA, en la década de 1950, el objetivo era “captar la estructura del pensar humano por medio de herramientas lógicas que representan la capacidad de razonamiento, y no por aquellas que busquen imitación del comportamiento del sistema físico que soporta al pensamiento (como el cerebro)” (Sandrone e Illich, 2024, p. 325). En el primer párrafo de las actas del coloquio de Dartmouth, escritas en agosto de 1955, donde nació el concepto de Inteligencia Artificial, se lee sobre el campo de estudio propuesto:

Ese estudio debe tomar como base la conjetura de que todos los aspectos del aprendizaje o de cualquier otro rasgo de la inteligencia se pueden describir por principio con la precisión necesaria como para que una máquina pueda imitarlos. Se intentará averiguar cómo lograr que las máquinas utilicen el lenguaje, establezcan abstracciones y conceptos, resuelvan tipos de problemas que ahora están reservados a los seres humanos y se perfeccionen a sí mismas. Creemos que se puede realizar un avance notable en uno o más de estos

problemas si un grupo de científicos cuidadosamente escogido trabaja de manera conjunta en ellos durante un verano”. (McCarthy; Minsky; Rochester; Shannon, 2006)

Notemos que, a diferencia de la cita de Turing, no se habla de interacción con el entorno, ni de sensores o efectores. Mucho menos de “vagar por el campo”. Se pone el foco en el funcionamiento de la mente como algo “descorporizado” y “simbólico”, que puede prescindir del cuerpo para pensar y que, una vez formalizado (la descripción precisa), puede insertarse en un soporte maquínico. La concepción cartesiana del *fantasma en la máquina* opera como modelo humano para diseñar robots cognitivos. En esta primera etapa, la idea de construir máquinas inteligentes está fuertemente inspirada y motivada por la búsqueda de los principios de la inteligencia humana, “el fantasma” (la *fente*, en términos de Larson). A su vez, se pone el foco especialmente en la capacidad de resolución de problemas. Los desarrollos en ese momento intentan replicar la facultad de planificación, de comprobar de teoremas, de ganar al ajedrez y cuestiones similares.

Posteriormente, a partir de 1980 comienzan a explorarse algunas formas de la IA basada en las habilidades sensorio motrices, en donde se enfatiza la naturaleza “corporizada” de la inteligencia humana (incluida la locomoción, manipulación de objetos y la interacción con el entorno). La inteligencia humana pierde especificidad y se diluye en la inteligencia de los seres biológicos en general: corporizada, situada, rodeada (Dautenhahn, 2007, p. 680). Al modificar la concepción de la inteligencia, y por lo tanto del ser humano, se modificó también el modelo de la robótica. Utilizar el lenguaje simbólico de la lógica de primer orden, pretendidamente representativa del razonamiento humano, comenzó a percibirse como insuficiente para diseñar la interacción inteligente de los robots en un entorno físico cambiante. La *Nouvelle IA* se alejó de la idea de que un robot es una “computadora” sobre ruedas (algo que, como vimos, planteaba Turing) y comenzó a pensar a la inteligencia artificial como inescindible del soporte material que interactúa con el entorno. El robot deja de pensarse como una máquina a la que hay que programar con altos volúmenes de lenguaje simbólico, y se la conc-

be como una entidad que puede generar sus propios comportamientos complejos (e inteligentes), a partir de la combinación de comportamientos simples en la interacción con el entorno. Por lo tanto, importa mucho más la información sensorial y los mecanismos para lograr el equilibrio ecológico que los esquemas simbólicos de razonamiento. Esto abre el camino a una *robótica ecológica*, no antropocéntrica, en la medida que el “organismo en el entorno”, antes que el “fantasma en la máquina”, se constituye como modelo a seguir para fabricar robots. Los animales se vuelven fuentes de inspiración: insectos, babosas, salamandras, se transforman en nuevos paradigmas de modelos cognitivos (Dautenhahn, 2007, p. 681). Se desarrollan “robots insectoides”, que no contenían modelos internos fijos del mundo, sino que los “construían” efímeramente a partir de un gran volumen de la información que recibían a través de sus sensores para eliminarlo a los pocos segundos, de manera que el “mapa” cambiaba a la medida que cambiaba el territorio.

En este tipo de robótica adquiere relevancia la noción de aprendizaje maquínico (*learning machines*), que también le debemos a Turing (1936, p. 454). La máquina de Turing es un esquema que permite fusionar el concepto de máquina con el de “criatura viva”, en la medida que describe procesos de adquisición de información del entorno y de sí misma, registra las mutaciones que se producen en función de esa información y conserva esos registros para tomarlos como nueva información. En este tipo de máquinas, las contingencias del entorno no configuran un obstáculo, ni llevarán necesariamente a un error en el funcionamiento, sino que serán un insumo para la transformación de la estructura y la forma de operar de la máquina. En base a ello, Turing propuso que, en lugar de crear una máquina que describa la manera en que la mente humana procesa información, “¿por qué no intentar producir una que simule la del niño? Si se sometiera a un curso de educación adecuado, se obtendría el cerebro de un adulto” (Turing, 1936, p. 456). En ese sentido, la máquina-niño es un esquema de mutación y selección (artificial) que no solo puede recrear mecánicamente el aprendizaje de la mente humana, sino de cualquier entidad, de cualquier organismo: “existe una clara relación por analogía entre este proceso y el de la evolución” (Turing, 1936, p. 456). En este tipo de máquina, lo here-

dato entra en una interacción dinámica con lo adquirido, provocando mutaciones en la estructura lógica como en su esquema de operaciones. Ya en R.U.R, mucho antes que Turing propusiera estas ideas, los robots aparecen como entidades que pueden aprender hasta convertirse en seres adultos: “como cualquier otra máquina, deben acostumbrarse a existir. Están creciendo por dentro o algo así. Hay que dejar un poco de espacio para el desarrollo natural (...) Como en una “escuela” humana. Aprenden a hablar, a escribir y algo de aritmética. Tienen una memoria increíble” (Čapek, 2022, p. 22). En la actualidad, y ya fuera de la ficción, la *robótica del desarrollo* es una subrama de la robótica, en la cual los robots no son programados con conocimientos lógicos sino que desarrollan habilidades cognitivas para interactuar con el entorno, e incluso habilidades sociales (Dautenhahn, 2007, pp. 682-683).

Aquí aparece un punto crítico. El aprendizaje humano no solo se nutre de las experiencias con el comportamiento del entorno físico y geográfico, sino también del entorno social. Moverse inteligentemente en el entorno físico es solo una parte del problema, el que aborda principalmente la robótica cognitiva. Tener habilidades sociales para moverse en una comunidad de humanos requiere aprender del comportamiento de otros en determinadas circunstancias. Podemos volver otra vez a Turing, quién intuyó que el aprendizaje posee un componente social. Hablando del robot, dice: “la criatura seguiría sin tener contacto con la comida, el sexo, el deporte y muchas otras cosas de interés para el ser humano” (Citado en Copeland, 2004, p. 420). Quizá aquí forcemos un poco a Turing, en su manera de pensar el problema de la interacción entre máquinas y humanos, pero sus agudas observaciones solían intuir el hecho de que la máquina inteligente debería satisfacer los problemas de la interfaz social, y eludir el rechazo humano, si quería lograr una inteligencia al *estilo humano*. En el siguiente pasaje Turing especula sobre la posibilidad de que un robot pudiera ir a la escuela.

No será posible aplicar exactamente el mismo proceso de enseñanza a la máquina que a un niño normal. No tendrá, por ejemplo, piernas, de modo que no se le podrá pedir que salga y llene el cubo de carbón. Posiblemente no tenga ojos. Pero por muy bien que estas deficiencias pudieran ser superadas por una ingeniería

inteligente, uno no podría enviar a la criatura a la escuela sin que los otros niños se burlaran excesivamente de ella. Sería necesario darle alguna instrucción. (Citado en Copeland, 2004, pp. 460-461).

En este apartado hemos planteado el origen ficcional y especulativo de la figura del robot, remarcado que esa misma especulación que nace de la ficción ingresó en el horizonte científico e ingenieril. Planteamos también que esto tiene que ver, de algún modo, con las ventajas cognitivas que ofrece el cuerpo a los seres humanos y otros vivientes para la interacción con el entorno geográfico y social. El robot, precisamente, no es solo una inteligencia artificial, sino también un cuerpo artificial. En consecuencia, la robótica se engarza en el proyecto del desarrollo de la inteligencia artificial, no como una aplicación de esta, sino como un campo de experimentación intrínseco. Por otra parte, el desarrollo de una IA corporizada, permite diseñar robots cognitivos más eficaces, que se vinculen con el entorno físico y geográfico y puedan generar conocimientos y formas autónomas de intervenir en este. Sin embargo, los humanos son seres singulares que habitan el entorno del robot, por lo que la manera de relacionarse con ellos es también singular. El hecho de hacer inteligencia artificial corporizada y perfeccionar los robots cognitivos no garantiza que estos puedan moverse con mayor inteligencia en entornos sociales y que mejoren la interacción humano-máquina (robótica social). Hay algo de esa interacción que requiere un tipo especial de investigación y desarrollo, y que apareció en el horizonte científico y tecnológico con fuerza hacia finales del siglo XX.

HRI Y ROBÓTICA SOCIAL: COMPORTAMIENTO Y APARIENCIA

Hasta ahora hemos sugerido que la inteligencia artificial no opera en abstracto, sino en algún tipo de entidad de la que se puede observar su comportamiento en algún entorno. Los teóricos de la IA como Russell y Norvig suelen referirse a estas entidades como *agentes*, a los que definen como “cualquier cosa capaz de percibir su medio ambiente con la

ayuda de sensores y actuar en ese medio utilizando actuadores” (Russell y Norvig, 2004, p. 37)^{III}. Lo que nos interesa aquí es pensar al robot como un agente artificial. De hecho, aunque no hay una definición unánime de qué es un robot, ni por parte de los expertos técnicos, ni por parte de los filósofos, la definición canónica de robot es muy similar a la de agente. El Instituto Internacional de Ingenieros Eléctricos y Electrónicos (Institute of Electrical and Electronics Engineers) define a los robots como “una máquina autónoma capaz de detectar su entorno, llevando a cabo computaciones para tomar decisiones y realizando acciones en el mundo real” (Citado Coeckelberg, 2024, p. 14). Esta definición tiene relevancia filosófica en la medida que implica la pregunta por lo real. Por ejemplo, Coeckelberg afirma llanamente que si un agente artificial está constituido solo de un software, “se llamaría un «bot»: no se consideraría un robot” (Coeckelberg, 2024, p. 14). Para ello se basa en las afirmaciones de Patrick Lin y sus colegas quienes “han argumentado que, en contraste con los ordenadores o la IA no corporeizada, un robot «puede ejercer influencia sobre el mundo de forma directa» (Lin, 2014, p. 943). Desde luego, para afirmar que un robot físico actúa sobre el mundo de manera directa, mientras que un *bot* no lo hace, es necesario tener una concepción de mundo y de acción específica. Como dijimos anteriormente, la robótica cognitiva ha encontrado algunos problemas cuando el “entorno real” que debe detectar, y en el que debe intervenir el robot, es el entorno social.

Para abordar ese problema específico, desde la década de 1990 comenzó a perfilarse una nueva área de investigación empírica que incluye a

III También Floridi habla de agencia como algo que tiene la capacidad de “1. recibir y utilizar datos del entorno, a través de sensores u otras formas de entrada de datos de entrada; 2. emprender acciones basadas en los datos de entrada, de forma autónoma, para alcanzar objetivos, mediante actuadores u otras formas de salida, y 3. mejorar su rendimiento aprendiendo de sus interacciones. Un agente similar puede ser artificial (por ejemplo, un *bot*), biológico (por ejemplo un perro), social (por ejemplo una empresa o un gobierno), o híbrido” (Floridi, 2023, p. 59). El punto 3, sin embargo, agrega la capacidad de aprender al agente. Larson, por su parte, también utiliza el término: “Adopto aquí una palabra habitual en la IA, «agentes» que significa cualquier cosa que actúe de manera independiente. Una persona puede ser un agente inteligente o cognitivo pero también puede serlo un sistema de IA o un alienígena” (Larson, 2022, p. 339)

la robótica pero que no se agota en esta. Se trata del área de Interacción Humano-Máquina (HRI, por su siglas en inglés). Dado que se trata de una disciplina que estudia la forma de vincularse de dos entidades, en principio, diferentes, humanos y robots, es un área de investigación interdisciplinaria donde convergen la robótica, la ingeniería y las ciencias de la computación, pero también la psicología, la lingüística, la etología, además de otras ciencias sociales como las investigaciones del comportamiento social, la comunicación y las ciencias cognitivas. El objetivo de la HRI es “llevar seres sintéticos al ámbito de la socialidad humana” (Bisconti, 2024, p. 1)^{IV}. Este objetivo puede dividirse en dos aspectos. Por un lado, el método de diseño del robot, en donde conviven dos enfoques: por un lado, el de *robot-centrado*; por otro, el de *humano-centrado*. Además, el otro aspecto, es el producto que se espera conseguir, en el que podemos distinguir al *robot asistente*, por un lado, y, por el otro, el *robot de compañía*.

Comencemos por el método de diseño. El primero que mencionamos es el enfoque de la HRI *robot-centrado* (Dautenhahn, 2007, p. 682). En esta línea, la interacción entre humanos y robots sirve como un insumo para “entrenar” al robot, ya sea para que se perfeccione en una tarea específica o para que se vuelva “más humano”, permitiéndole conectar a nivel emocional con los sujetos. En consecuencia, se diseñan o se persiguen el tipo de interacción social que satisfaga las “necesidades” que requiere el robot para aprender. Esas “necesidades” han sido identificadas por el diseñador del robot y modeladas en su arquitectura. Esta tendencia promueve la concepción del robot como una suerte de criatura viva, con amplios márgenes de autonomía, capaz de generar sus propios objetivos y motivaciones. En los proyectos más especulativos se asume que el robot puede ser susceptible de padecer impulsos y emociones. Para los robotistas que trabajan desde esta perspectiva, el norte de la investigación es desarrollar modelos de comportamiento, incluida la expresión de emociones, que sean regulados por la interacción social y el registro de los comportamientos humanos. Esta concepción es objetivista, en el sentido de que considera que el robot puede volverse una

IV Esta cita, al igual que el resto de las que se hagan de esta obra, es traducción propia.

criatura con un comportamiento propio, que no requiere la aprobación y aceptación de los humanos para lograrlo con éxito.

Por otro lado, un segundo enfoque metodológico de la *HRI es el enfoque humano-centrado* (Dautenhahn, 2007, p. 684). Esta línea se ocupa de cómo el robot puede cumplir su especificación de tareas de manera que sea *aceptable y saludable* para los humanos. En este enfoque, las percepciones y creencias humanas tienen relevancia, y se incluyen en el circuito de diseño (“humans in the loop”) (Dautenhahn, 2007, p. 684). Esto no solo abarca el sistema cognitivo humano, sino también las tradiciones, las culturas, las convenciones sociales. No solo importa como debería o podría ser el robot, sino como debería ser ante los ojos del humano. Se investiga cómo las personas reaccionan o interpretan su apariencia y su comportamiento, y lo que los humanos proyectan sobre el robot, con independencia de la arquitectura y las operaciones de este. Asimismo, el propósito de este enfoque es lograr un diseño equilibrado entre las operaciones que se desea que realice el robot en contextos sociales, y la *apariencia* que este *debe* tener para realizar esas operaciones, sin afectar de manera negativa ningún aspecto, ni físico, ni psíquico, ni social de la vida humana. Para ello, la percepción y psicología humana, así como los patrones de comportamientos grupales y las tendencias sociales, no son obstáculos a vencer, sino insumos para definir el diseño. Este enfoque es el que más se asemeja a aquella disciplina que proponía Simondon bajo el nombre de “psicología de la tecnicidad” (Simondon, 2017), centrada en cómo la psiquis humana recibe e interactúa con los objetos técnicos. Este enfoque de la HRI es también el que se interesa por el “valle inquietante” (Uncanny Valley) en los que los robots producen un rechazo psicológico por parte de las personas cuando exhiben rasgos muy similares a los humanos. Este sentimiento es difícil de comprender porque, como dijimos, la apariencia humana es clave para una aceptación de los robots, pero también, paradójicamente, es motivo de rechazo (Sandrone, 2023).

Nos interesa particularmente este enfoque. El propio Turing observa que la apariencia de la “máquina inteligente” es un obstáculo para la relación humano-máquina. Es por eso que la computadora y el humano deben estar ocultos tras una puerta en su célebre “juego de la imitaci-

ón”, de modo que el evaluador humano no los prejuzgue por su apariencia. Con este juego teórico, Turing anticipa que las dificultades de la integración social de las inteligencias artificiales iban a estar determinadas por la creación de una apariencia en la que los humanos no se pusieran a la defensiva cuando tuvieran que interactuar con estas, incluso cuando su comportamiento fuera indistinguible de un agente humano. Antes de su famoso artículo de 1950, en 1948, Turing ya afirmaba que “el grado en que consideramos que algo se comporta de forma inteligente viene determinado tanto por nuestro propio estado de ánimo y formación como por las propiedades del objeto en cuestión” (Citado en Copeland, 2004, p. 431).

Por otro lado, en términos del producto al que se aspira, el campo de la HRI hereda las dos almas de la IA que mencionamos al principio. Por un lado, el enfoque ingenieril y reproductivista plantea como horizonte *robots asistentes*, atravesados por una dimensión instrumental, en la cual la interacción social entre humanos y robots no es una finalidad en sí misma, sino un medio para entrenar a los robots en pos de que se perfeccionen en una tarea específica. Por ejemplo, robots que puedan interactuar socialmente en un bar con los clientes porque eso les permite aprender las mejores formas de atender las mesas; el objetivo de la interacción social es perfeccionar la tarea específica para reemplazar a los camareros humanos. Este objetivo no supone estrictamente que los robots *deben* tener todas las aptitudes sociales de los humanos, sino aquellas que sean necesarias para poder aprender la tarea específica a llevar adelante. Por otro lado, existe otra tendencia en la industria robótica que es la de elaborar lo que se conoce como *robots de compañía*, es decir máquinas “diseñadas y programadas específicamente para producir una interacción intersubjetiva que a menudo incluye una dimensión emocional con los humanos” (Bisconti, 2024, p. 1). La idea aquí es que los robots sean percibidos, no como instrumentos para una tarea específica, sino como otros seres con los cuales los humanos podemos pasar el tiempo, ser amigos, conversar y hasta enamorarnos.

Ahora bien, desarrollar un robot con capacidad de establecer vínculos personales y sociales con humanos requiere cierto tipo de inteligencia que no se agota en resolver problemas, comprobar teoremas o ganar

al ajedrez. La robótica cognitiva se muestra insuficiente y comienza a abrirse el camino hacia una *robótica social*. Esto reaviva el interés por estudiar la inteligencia humana y sus representaciones simbólicas, pero ya no solamente es su componente de inferencias lógicas, sino de la forma en que operan las representaciones sociales. Este tipo de inteligencia requiere habilidades como interpretar movimientos corporales y gestos faciales, predecir comportamientos de otros actores sociales, e incluso manipular sus expectativas y emociones. La premisa de la robótica social es que para hacer a los robots más inteligentes al estilo humano, es necesario hacer más sociales a los robots (Dautenhahn, 2007, p. 684). Esto implica asumir el método de diseño HRI humano-centrado. Si los humanos no perciben los comportamientos de los robots como agradables o aceptables, no querrán socializar con ellos.

La robótica social es “un campo multidisciplinario centrado en comprender, diseñar y evaluar sistemas robóticos para uso por o con humanos” (Bartneck, 2020). Abarca desde “el diseño de la apariencia y el comportamiento de los robots hasta la exploración de las implicaciones sociales del despliegue de robots” (Bisconti, 2024, p. 4). Requiere una combinación de diferentes técnicas derivadas de campos disímiles como el aprendizaje automático, la visión por computadora y el procesamiento de lenguaje natural, con el propósito de obtener un robot que pueda interactuar con los humanos de una forma personal, atractiva y eficaz, abriendo interesantes posibilidades para el futuro de HRI (Bisconti, 2024, p. 4).

En la actualidad, la Robótica Social es un campo que está creciendo día a día. Si pensamos en términos individuales, como propone el concepto de agente, y tomando el robot social^V como un tipo de agente específi-

^V Tal vez, no convenga hablar hoy de robot social, sino de robótica social distribuida en entornos sociovirtuales. Como afirma Piercosma Bisconti, no se trata de un tipo de robot, sino de un ecosistema que se expande confusamente en un “amplio espectro de máquinas diseñadas para interactuar con los humanos de una manera personal y emotiva, tan diverso como las tareas que realizan y los entornos que habitan” (Bisconti, 2024, p. 3). Parafraseando lo que Simondon dice de los autómatas (Simondon, 2007, p. 69), no hay una especie de robot social; solamente hay objetos técnicos que poseen una organización funcional en la que se consuman diferentes grados de robótica social.

co, es necesario que cumpla con algunos requisitos (Dautenhahn, 2007, p. 684). Por ejemplo, debe ser socialmente evocativo, es decir, debe tener la capacidad de evocar emociones y sentimientos en los humanos a partir de ciertos movimientos en su estructura material, capitalizando la tendencia de los humanos a antropomorfizar algunos rasgos de los objetos técnicos. Por otro lado, es importante que el robot sea socialmente situado, ya que se encuentra rodeado por un entorno social, el cual percibe y al que reacciona. Además, debe poseer la característica de ser sociable, es decir, tener impulsos y motivaciones propias que lo lleven a interactuar proactivamente con humanos (y no solo aprender de ellos). Para esto se requieren modelos profundos de cognición social que no excluyen las significaciones y las representaciones sociales.

Por ejemplo, un área de conocimiento que funciona como insumo de la robótica social es la proxémica, “una disciplina o rama de la semiótica dedicada al estudio de la organización del espacio en la comunicación humana” (Bisconti, 2024, p. 2). Cómo nos acercamos a las personas que tenemos más confianza, cómo saludamos a alguien a la distancia, cómo nos paramos cuando estamos orgullosos o cuando estamos avergonzados. Cómo usamos el espacio y que significa. Pero, también, cómo usamos el espacio cuando nos comunicamos. Por ejemplo, el Chat GPT podría poner en pantalla todas las palabras simultáneamente, y, sin embargo, coloca una a una, sucesivamente, para crear la ilusión de que “está escribiendo” y por lo tanto hay “alguien” que escribe.

Asimismo, es obvio que las significaciones estéticas son claves en la Robótica Social. Los estudios empíricos sostienen que “los seres humanos, cuando se enfrentan a un robot antropomórfico, tienden a comportarse en consecuencia” (Krämer, 2011). Además de sensores de visión y sonidos, los robots deben contar con ojos y orejas. Si quieren interactuar con humanos de manera más eficiente, necesitan un rostro. La comunicación a partir del rostro es distintiva de nuestra especie que, más que cualquier otra, tiene la capacidad de extraer información social y emocional de los gestos faciales. Un hito de la robótica en este sentido fue el famoso Kismet, diseñado por Cynthia Breazeal en el MIT, en la década de 1990. El robot Kismet era una cabeza robótica que tenía dos características fundamentales. Por un lado, un sistema de sensores de

visión y de audio, en combinación con un software de aprendizaje automático, que le permitía captar patrones en los rostros de las personas e identificar emociones en las expresiones faciales, como rabia, asco, miedo, alegría, interés, tristeza y sorpresa. A la vez, los sensores de audio le permitían identificar en la voz del hablante tipos diferentes de discurso afectivo: aprobación, prohibición, atención y neutralidad, entre otros. Asimismo, el Kismet contaba con una serie de rasgos mecánicos que le permitían simular expresiones faciales. Tenía orejas y pestañas que se podían mover para expresar alegría, sorpresa o tristeza. Estas reacciones eran fundamentales para que el robot realizara una comunicación con sujetos humanos y así aprender mejor de ellos, como una máquina-niño, incorporando poco a poco sus propias estrategias para realizar gestos y ganar más eficacia. El tipo de método de diseño empleado se acerca más al *HRI robot-centrado*, porque el foco está en cómo hacer que el robot mejore su interacción social, sin preocuparse en lo que los humanos sentían al interactuar con este.

En la actualidad, en entornos digitales, la cabeza Kismet encuentra descendencia en *avatars* y *softbots* que son representaciones gráficas que imitan la gestualidad humana en la pantalla de los dispositivos. Desde nuestro punto de vista, estas interfaces gráficas son también una deriva de la Robótica Social. Por tomar un caso reciente, en 2024 Ucrania presentó en sociedad a Victoria Shi, un avatar anunciado como la primera funcionaria pública digital del mundo, programada para cumplir funciones de portavoz en el Ministerio de Asuntos Exteriores de aquel país. A través de IA, el gobierno ucraniano creó la figura y la voz de este avatar, para usarlo como un “apoyo visual” en la comunicación del contenido de documentos oficiales. Victoria es, a fin de cuentas, una representación animada y antropomórfica de los documentos, una suerte de copia escaneada que tiene rostro humano y habla, o más bien, que es hablada por los textos que le ingresan. El aspecto de Victoria, por otra parte, no se creó de la nada. Se basó en la apariencia y las expresiones de una cantante e influencer humana muy popular en Ucrania, llamada Rosalie. Sin conocer los detalles, podemos aventurarnos a pensar que se entrenó un algoritmo de *machine learning* con fotos, videos y voces digitalizadas de Rosalie, y de miles de personas más, con el propósito de obtener como salida una figura digital de rostro humano, que

podiera simular expresiones y gestos sosegados, amables y sólidos, en un rostro familiar para los ucranianos, de manera que inspire en ellos sentimientos positivos como credibilidad, respeto e incluso admiración. A pesar de su origen en los circuitos de silicio, Victoria proyecta una sonrisa cálida, una mirada serena y un aire familiar, esculpido bit a bit en la pantalla, con la precisión que permiten las técnicas estadísticas del *machine learning*. Visto de esta forma desencantada, buena parte de las innovaciones digitales expresivas son el resultado de la recombinación y la comprensión de gestos humanos ancestrales. Este episodio, aunque impactante por la relevancia del ámbito en el que se aplicó, se suma a una serie de innovaciones en las que la IA contemporánea, lejos de explorar el camino de *La Singularidad* y la creación de *superinteligencias*, desarrolla automatismos que recrean y simulan comportamientos de los humanos para poder sustituir a estos por algo más dócil y menos costoso. Estos *avatars*, así como otros casos de robots físicos, se acercan más al diseño a partir de *HRI centrado en el humano*, porque no solo tienen como objetivo que el robot desarrolle una expresividad eficaz, sino que, además, se busca que genere empatía en los humanos. Lograr interfaces no humanas con expresiones humanas, que hagan sentir a gusto a los humanos, incluso más que los humanos mismos.

Este ejemplo nos muestra que existe un elemento distintivo que diferencia a cualquier agente informático, en general, de un tipo de agente específico, el *robot*. Este elemento es la apariencia, una forma de verse ante los ojos humanos o de escucharse (Alexa, de Amazon, y Siri, de Apple, por tomar dos casos de las marcas comerciales más populares, son voicebots -robot de voz- basados en inteligencia artificial que interactúan con humanos utilizando el lenguaje natural). Esta apariencia puede darse con un cuerpo artificial, si se trata de un robot físico, o como interfaz gráfica y sonora si se trata de un bot o un avatar que opera en entornos virtuales. En este sentido, la Robótica Social agrega elementos relevantes a los debates de la IA. Los robots, los softbots y los avatars tienen género, tienen peinados, tienen ropa. Cada una de esas decisiones de diseño implica una valoración cultural sobre el aspecto de los cuerpos humanos y sus accesorios. Como señala Piercosma Bisconti (2024), desde el punto de vista de una Filosofía de la tecnología, nos permite integrar de una forma más concreta los estudios y las es-

peculaciones sobre la IA en una teoría social acerca de cómo se generan “agentes sociales sintéticos” (Bisconti, 2024, p. 2). Desde la Robótica Social, el robot no es la máquina de trabajo de R.U.R ni el robot cognitivo de Turing, sino, además, un artefacto con el cuál comunicar ciertos valores sociales y culturales en relación a la inteligencia, pero también a la apariencia corporal y a través de la comunicación no verbal.

Cuando los diseñadores eligen un corte de cabello específico, tal vez largo y femenino, para una recepcionista robot, están realizando una acción prescriptiva no solo para el robot sino también para la sociedad misma. Esto va más allá del concepto de sesgo: el diseño del robot, su cuerpo, está escrito y moldeado por una competencia sociopolítica entre símbolos, narrativas y mecánicas de poder representacional. El robot es como una hoja en blanco donde las estructuras simbólicas y las narrativas contradictorias pueden chocar como en una guerra de poder ¶lejos de los cuerpos humanos, pero aún haciendo referencia a ellos. El asunto del robot como «lugar de producción simbólica alienada». (Bisconti, 2024, p. 11)

En la Robótica Social no solo importa la inteligencia humana como modelo a seguir para los desarrollos tecnológicos de agentes artificiales, también importa el poder de las apariencias en el aspecto visual y en el comportamiento. No solamente la manera en que se presentan estos agentes ante los ojos de los humanos importa, también la forma en la que hablan y escriben, la forma en que se disculpan si dicen algo equivocado, la forma en la que arquean las cejas o modulan el tono de voz si quieren mostrarse preocupados o contentos. El componente imitativo que Turing imaginó en un experimento de laboratorio, ocultando la apariencia de la máquina, se hace a cielo abierto en la sociedad misma a partir de la Robótica Social, que no busca la aceptación de los humanos por ocultar el aspecto de las máquinas, sino, justamente, se propone conseguir esa aceptación *precisamente* por el aspecto humano de estas. A partir de este desplazamiento, los aspectos éticos y políticos a tener en cuenta deben incorporar la cuestión de las apariencias humanas, lo que

supone un elemento inédito en los debates filosóficos sobre la automatización de la sociedad, que cuentan con una larga tradición.

ÉTICA DE LAS MÁQUINAS Y ROBOÉTICA: EL ARGUMENTO DEL ENGAÑO

Retomando el punto de inicio de este escrito, planteamos que existen dos agendas contemporáneas en relación a la IA. Por un lado, una agenda que tiene un propósito de generar agentes que reproduzcan las tareas humanas y sus productos, siendo inteligentes pero no al estilo humano, es decir, sin comprometerse con reproducir artificialmente las claves de la singularidad intelectual humana. La segunda agenda tiene como propósito, en cambio, replicar artificialmente la fuente de la singularidad humana. La robótica por su parte, según hemos planteado, se incorpora en espejo a esta doble agenda de la IA a través de la cuestión del cuerpo artificial. Por un lado, una agenda de la robótica que procura crear robots capaces de realizar tareas cognitivas o productivas que realizan los humanos, sin que la apariencia y el comportamiento del cuerpo humano sea un paradigma a seguir. Por otro lado, existe el proyecto de la Robótica Social que está comprometida con encontrar las claves de la comunicación humana basada en los cuerpos, incluido el rostro, para replicarla artificialmente. Conectando con lo que Floridi dice de la IA, podríamos decir que esta línea no solo busca crear robots corporizados, sino “verdaderamente corporizados”.

La primera de estas agendas tiene en mente principalmente la producción de robots asistentes; la segunda, incluye, además, la creación de robots de compañía, capaces de empatizar vitalmente con los humanos, más allá de las tareas puntuales que puedan realizar. Asumiendo esto, podemos preguntarnos ahora si estas dos agendas de la robótica deben enmarcarse en la misma matriz teórica para elaborar una teoría crítica, basada en aspectos éticos y políticos. Creemos que no, y nos interesa explorar y sugerir algunas diferencias, señalando una distinción entre las discusiones éticas en relación a las máquinas y la ética de los robots

(o la roboética) (Bisconti, 2024, p. 11), para poner el foco sobre esta última.

Por un lado, el proyecto de crear *robots asistentes*, que no repliquen al humano sino sus tareas y productos, es susceptible de ser analizado en términos de una ética de las máquinas tradicional. Estos robots pueden, sin ser como los humanos ni comportarse como los humanos, ser más rápidos, más fuertes, y más inteligentes. A pesar de ser entidades artificiales muy sofisticadas y asombrosas, los robots asistentes no dejan de ser máquinas con fines utilitarios que pueden tener consecuencias negativas para el individuo y la comunidad, como la dependencia, la pérdida de autonomía, o convertirse en un instrumento de opresión y explotación por parte de los poderes fácticos. Así entendido, el problema ético y político del robot asistente se inscribe en una larga tradición que se ha ocupado de la relación humano-máquina en esos mismos términos, sobre todo desde la revolución industrial (Sandrone, 2023a, 2023b). Marx hablaba de la maquinaria industrial como un “poderoso organismo” que se opone al obrero y se vuelve contra él (Marx, 2011, p. 219). Andrew Ure señaló que los humanos devendrían en “simples pastores de máquinas” (Ure, 1835, p. 20). Desde la ficción, Samuel Butler, en su célebre novela *Erewhon* de 1872, se preguntaba: “¿cuántos hombres, hoy día, viven en un estado de esclavitud con relación a las máquinas? ¿Cuántos pasan su vida entera, desde la cuna hasta la tumba, cuidando de ellas noche y día?” (Butler, 2012, p. 238). En la actualidad, estas críticas se renuevan explorando el asunto de los robots asistentes como una nueva forma de reemplazo de mano de obra humana por parte del capital, o como una tecnología de manipulación social. Dado que los robots asistentes parten de un enfoque centrado en el robots, a los que se concibe como criaturas vivas que pueden crecer y aprender, las críticas, generalmente, parten de un punto de vista humanista, en el cuál las sociedades humanas son algo distinto de la maquinaria robótica, y debe ser protegida de esta.

Por otra parte, existen algunas posturas críticas que más bien se insertan en un escenario posthumanista y que sostienen la tesis del continuum humano-máquina (Sandrone, 2023c). Desde este punto de vista, las máquinas han dejado de concebirse como meros instrumentos para

convertirse en parte constitutiva de la existencia humana. Al comienzo del siglo XX, Jacques Laffite consideraba que la vida social transcurría en una “mecanosfera”, donde la organización sociopolítica se daba en un entramado de humanos y máquinas (Sandrone, Vaccari, Lawler, 2022). En la década de 1950, Simondon hablaba de “acoplamiento humano-máquina” (Simondon, 2007, p. 137), y en la de 1960, Deleuze llamaba a este fenómeno “agenciamiento hombre-máquina” (Deleuze, 2014, p.236) La condición de posibilidad para que esto haya ocurrido es la aparición y expansión en el siglo XX de nuevas máquinas cibernéticas, no mecánicas, que pueden acoplarse de manera más íntima a las prácticas sociales humanas. Lo maquínico se escinde de lo mecánico y se funde con lo humano. El robot social, de alguna manera, es el *agente* artificial que mejor cristaliza esta transformación.

Una hipótesis que nos gustaría explorar es que una sociedad automatizada es muy diferente a una sociedad robotizada. En ambas, la singularidad humana, su forma específica de cognición y conducta está distribuida, expandida, replicada, complementada, por los sistemas artificiales: no existe lo humano sino es *con* lo artificial. Pero en la primera, los humanos son concebidos como insumos para la elaboración de sistemas artificiales autónomos; en la segunda, en cambio, los humanos interactúan socialmente con los sistemas artificiales. Por eso mismo, la ética orientada a los robots, a diferencia de la ética de las máquinas tradicional, no opera bajo la férrea distinción entre sociedad humana y maquinaria robótica, y pone en suspenso la dialéctica amo y esclavo entre humano y máquinas. En términos de Coeckelbergh, lo que interesa no es la opresión sino la “la co-interacción, el circuito de retroalimentación (feedback loop), que se desarrolla entre el objeto tecnológico y el sistema social e indagar cómo los dos elementos se co-construyen” (Coeckelbergh, 2010, p. 56). También este es el enfoque de Bisconti, quien plantea el concepto de *sociedades híbridas humano-robots* (Bisconti, 2024). Para comprender la sociedad en esos términos es necesario pensar al robot, no como una mera máquina asistente, sino como una máquina culturizada. Este sentido no admite la concepción de la esfera social y de la esfera tecnológica como conjuntos diferentes que interactúan, sino que obliga a pensar en clave de una misma esfera socio-técnica.

El objeto tecnológico da forma a la sociedad y, a su vez, es moldeado por ella. La roboética es el enfoque en el que se basa este trabajo, ya que considera las cuestiones éticas que plantean las máquinas sólo dentro del sistema social. Nuestro interés no se refiere a la máquina en sí, sino a la máquina como construcción cultural, como objeto-significante producido dentro de un determinado mundo social. Por lo tanto, la máquina se construye dos veces: una en el sentido literal de ser ensamblada y una segunda vez en la forma en que su valor está inextricablemente vinculado al contexto en el que se crea. En resumen, la roboética está mucho más interesada en el efecto social de la máquina que en la máquina (Bisconti, 2024, p. 9).

Sin embargo, siempre hay que resaltar el hecho de que una sociedad híbrida entre robots y humanos no es un acontecimiento espontáneo. En ese sentido, hay que tener muy presente que el proyecto de una robotización social no es aséptico y está plagado de intereses.

Por último, la roboética parece encontrar un foco de interés en los robots de compañía, en los cuales la interacción social no es un medio para entrenar al robot, sino el fin mismo. El interés en este tipo de robots radica en que se pretende que interactúen con la realidad emocional y psíquica de las personas. Esto amplía la discusión sobre la IA del plano cognitivo al plano de los afectos. A los humanos, además de conocer y actuar, nos gusta que nos acompañen, que nos valoren, que nos cuiden y que nos amen. Si los robots (o bots) nos hicieran sentir así podríamos incorporarlos a nuestra vida, junto a otros humanos. Existe entonces una conexión entre este producto de la robótica social y el alma productiva de la IA, pero no en clave cognitiva, sino en clave afectiva. Por otra parte, los robots de compañía no necesariamente deben ser agentes racionales, como señalaba Russell y Norvig, sino que es indispensable que piensen como humanos y se comporten como humanos, encuadrando la robótica social en esa tendencia de los desarrollos de la IA. Si esto es posible, es probable que entre humanos y robots existan muchos dramas “shakesperianos”, como señala Larson.

Todas estas discusiones son propias del campo *HRI humano centrado*, y, como mencionamos más arriba, se parece mucho a la disciplina que en la década de 1960 plantaba Gilbert Simondon bajo el nombre de “psicología de la tecnicidad” (Simondon, 2017, p. 45). Si bien en el *Modo de existencia de los objetos técnicos* (2007), Simondon sostiene que es posible establecer relaciones sociales con los nuevos individuos técnicos informacionales, pues “el hombre puede ser acoplado a la máquina de igual a igual como ser que participa en su regulación, y no únicamente como ser que la dirige o la utiliza” (Simondon, 2007, p.137), sin embargo, rechaza la figura del robot, a los que equipara con títeres, o marionetas, como una falsa máscara antropomórfica incrustada a los individuos técnicos. En la introducción califica al robot de “ser puramente mítico e imaginario ... producto de la imaginación y de la fabricación ficticia, del arte de la ilusión (Simondon, 2007, p. 11). Desde su punto de vista, el robot no es un individuo técnico autónomo, sino una representación ilusoria que se monta sobre su estructura operatoria que sí es real. La máquina es utilizada como escenario para una representación, o un lienzo sobre el que el artista dibuja fantasías, algo que, como vimos, es lo distintivo de la Robótica Social. La obra que se representa en la máquina es la simulación de lo humano, obra que, como toda obra artística, cobra existencia únicamente cuando existe un espectador que la aprecie. En *Filosofía y Cibernética* (2018), Simondon expresa esto del siguiente modo: “El robot es una imitación del hombre cuya finalidad funcional es imitar al hombre. El robot supone entonces un espectador para el cual existe la ilusión, un tercer término que es la conciencia espectadora más allá del imitador y de lo imitado” (Simondon, 2018, p. 47). Esta puesta en escena de lo humano sobre la máquina puede entenderse en un campo más amplio, que no solamente puede tener propósitos artísticos, sino prácticos y sociales, que en última instancia persiguen introducir la máquina a la cultura. Como afirma en *Psicología de la tecnicidad* (2017), un texto del 1960, los objetos técnicos novedosos tienen “la obligación de llevar un velo o un disfraz para penetrar en la ciudadela de la cultura” (Simondon, 2017, p. 45). Lo interesante del planteo de Simondon es que permite elaborar una teoría crítica de la robótica social porque, a diferencia de los robotistas sociales (y Turing), él ve un valor negativo en la imitación antropomórfica de las máquinas.

Esta valoración nos lleva a un problema filosófico que desde nuestro punto de vista es central, y que se conoce en la actualidad como *el problema del engaño o la objeción del engaño (the deception objection)*. Este problema está vinculado con la capacidad de la IA y la Robótica Social de reproducir la apariencia y el comportamiento humano, incluido la expresión de emociones y los gestos de afecto. La cuestión central es el grado de realidad que le atribuimos a esa simulación. Básicamente podríamos plantear el problema de la siguiente manera: los robots no pueden proporcionarnos amistad, amor o compañía, tan solo pueden simularla. Los robots no nos acompañan, operan *como* alguien que nos acompaña (Sparrow and Sparrow, 2006). Aquí el engaño no se entiende como una mentira, sino más bien como una ilusión o una ficción. Es cierto que se podría objetar que la ilusión y las falsas creencias son constitutivas del ámbito de las emociones y los sentimientos, incluso, y sobre todo, en los vínculos entre humanos. Pero eso no descalifica la objeción del engaño con relación a los robots, al contrario, la incorpora a una problemática de antropología filosófica, que tradicionalmente ha estado centrada en los seres humanos, pero que ahora se amplía a los debates sobre IA y Robótica Social. Al igual que lo hicimos más arriba, cuando esbozamos algunas definiciones de robots, nos volvemos a topar con el problema de la realidad: ¿qué es el mundo real, y qué significa influir sobre él? Justamente, el problema con la Robótica Social radica en que los robots interactúan con la realidad emocional y psíquica de las personas. No importa tanto lo que hacen sobre los humanos, sino lo que les hacen sentir.

En su libro, *Chamanes Y Robots* (2019), el antropólogo y sociólogo mexicano, Roger Bartra, plantea la cuestión en términos de lo que llama “efecto placebo robótico”: el artefacto es artificial pero el efecto sobre las psiquis de los seres humanos es real. Lo que hacen sobre los humanos se basa en lo que les hace sentir “realmente”. Los robots son artificios no reales que buscan proporcionar placer y disminuir el dolor, y, en muchos casos, lo consiguen. Asimismo, el problema de la ilusión no es algo propio de los robots. Nunca podemos saber si alguien nos quiere “realmente”, y, posiblemente, tampoco esa persona lo sabe “realmente”. El abanico de problemáticas que comienza a abrirse alrededor de los robots de compañía amplía la lista de disciplinas que intervie-

nen en la cuestión de la robótica social. Por ejemplo, podría incluirse un enfoque psicoanalítico a partir de Jentsch y Freud, que se ocuparon al respecto del sentimiento de lo siniestro (*unheimlich*) en los vínculos entre humanos y seres artificiales antropomórficos (Sandrone, 2023). Este es un debate que se prolonga, como hemos dicho más arriba, en el “valle inquietante” (Uncanny Valley), impulsado por el robotista japonés Masahiro Mori desde la década de 1970 (Mori, 2012).

Este punto también amplía la discusión de la IA a un problema de la estética o la Filosofía del Arte. Coeckelbergh, por ejemplo, plantea el problema del engaño en términos de un espectáculo

¿La creación de fantasías, que podría decirse es lo que hacen los desarrolladores de estos robots, es necesariamente mala? Por ejemplo, si se crea una fantasía en la que el robot es un ser vivo o un amigo, ¿es algo malo? Seguramente no tengamos un problema con lo ilusorio cuando vamos al teatro o a un espectáculo de magia. Estamos contentos de ser «engañados» durante el espectáculo (...) al mismo tiempo sabemos que lo que estamos viendo no es real...esto, aplicado a los robots personales, significaría que la gente suspendería la incredulidad de forma temporal durante la interacción con el robot para que, así, el robot pueda hacer su «magia» (Coeckelberg, 2024, p. 56).

La alusión a la magia también es interesante. Como ha afirmado Erik Davies, las tecnologías de la información y la comunicación son tecnologías del hechizo y la ilusión: “explotan por igual las leyes de la naturaleza como las de la percepción humana” (Davis, 2023, p. 44). Esto es, como hemos señalado anteriormente, el principio metodológico de la HRI centrada en el humano. A la vez, el arte de construir artificios que generen ilusión de humanidad no es algo nuevo, los fabricantes de marionetas y autómatas han buscado ese secreto desde hace siglos. Hay una destreza técnica notable en diseñar la mediación entre el mundo técnico y el mundo psicológico, entre los elementos mecánicos, geométricos, matematizables, calculables, y los elementos difusos, ambiguos, cualitativos de la mente humana. La expresividad es, sin duda, uno de

los aspectos humanos más difíciles de objetivar en una máquina. Diseñar un artefacto expresivo, es diseñar un artificio que parezca tener un yo, que logre la transición de “signos materiales a signos de alguien” (Gutierrez, 2021, p. 187). Es ponernos frente “un objeto moralmente visible por ser emocional, expresivamente significativo” (Gutierrez, 2021, p. 187).

Desde luego, esta problemática conlleva una discusión en el terreno de la ética y la política. En principio, porque el engaño como ilusión no habilita la forma del engaño como mentira y manipulación. Algo que Piercosma Bisconti llama la “falsa conciencia” del robotista (Bisconti, 2024, p. 11) En ese sentido, sería necesario regular el discursos de las corporaciones para que dejen en claro que se trata de un robot. Coeckelbergh señala que “se necesita un tipo de honestidad sobre lo que es realmente el robot y lo que puede proporcionar. Esto iría contra gran parte de la publicidad actual de los robots personales, que a menudo se venden como tu amigo»” (Coeckelbergh, 2022, p. 62). Una suerte de responsabilidad afectiva robótica. Por otra parte, aún la sociedad híbrida puede traer nuevos riesgos y segregaciones de humanos. Desde una Filosofía de la tecnología que aborde este asunto, más allá de las cuestiones conceptuales acerca de lo humano y lo artificial, la robótica social nos interpela sobre la pregunta por el poder: ¿Qué grupos de humanos se beneficiarán, y que grupos de humanos pueden resultar heridos?

CONSIDERACIONES FINALES

Hemos comenzado planteando las dos agendas de la IA contemporánea. En ambas, la cuestión de la imitación de lo humano está presente. En una de ellas, sin embargo, lo que se busca imitar es algo extrínseco: las tareas y productos que el humano realiza a través de su trabajo o su cognición. En la otra, lo que se busca es imitar al humano mismo, sus formas de razonar, de sentir y de hacer. Planteamos que estas discusiones adquieren otro espesor teórico cuando se las amplía incorporando discusiones en torno a la Robótica. La Robótica no es un campo

de aplicación de la IA, sino un banco de pruebas de muchas de las hipótesis sobre la inteligencia y sobre el vínculo entre mente y cuerpo. Planteamos luego, que la robótica además de problematizar la imitación tecnológica de lo humano, abre la problemática a la relación entre el humano y aquello que lo imita artificialmente. La interacción humano-robots es un ámbito de desarrollo científico y técnico, pero también un ámbito plagado de ficciones, especulaciones y problemáticas filosóficas que aportan elementos para abordar las dos agendas de la IA, además de algunos propios y singulares. La Robótica Social incorpora al debate sobre la posibilidad de hacer máquinas más humanas, una disputa acerca de las condiciones éticas y políticas en que debe darse la *ilusión* de que parezcan humanas.

Si tomamos en cuenta estos aspectos, debemos asumir que la Robótica Social es un proyecto científico y tecnológico que no se opone a las creencias infundadas de los seres humanos ni a sus convenciones sociales. Por el contrario, las asume como un insumo para los desarrollos tecnológicos en pos de generar productos que sean aceptados socialmente, no solo con fines utilitarios, sino en prácticas sociales más amplias en las cuales los robots aspiran a interactuar con los humanos como agentes sociales y socializados. En consecuencia, la Robótica Social no es solo un proyecto tecnológico, sino que debe entenderse como un proyecto psicosocial de alto impacto, que nos obliga a desarrollar un aparato crítico específico desde la Filosofía de la tecnología. Esta crítica debe tomar elementos éticos y psicológicos, así también como insumos provenientes de la teoría social y la teoría política. Esta matriz de pensamiento crítico, sin embargo, no se agota en la tradición de la crítica a la automatización de la sociedad que proviene del siglo XIX, sino que asume elementos inéditos que aún están en etapa de gestación.

REFERÊNCIAS

BARTNECK, Christoph; BELPAEME, Tony; EYSSEL, Friederike; KANDA, Takayuki; KEIJSERS, Merel; ŠABANOVIĆ, Selma. *Human-robot interaction: an introduction*. Cambridge: Cambridge University Press, 2020.

BARTRA, Roger. *Chamanes y robots: reflexiones sobre el efecto placebo y la conciencia artificial*. Barcelona: Anagrama, 2019.

BISCONTI, Piercosma. *Hybrid societies: living with social robots*. London; New York: Routledge, 2024.

BUTLER, Samuel. *Erewhon: o al otro lado de las montañas*. Madrid: Akal, 2012.

ČAPEK, Karel. *R.U.R.: drama colectivo en un prólogo cósmico y tres actos*. La Plata: Libros del Cosmonauta, 2022.

COECKELBERGH, Mark. *La ética de los robots*. Madrid: Cátedra, 2024.

COPELAND, B. J. (org.). *The essential Turing*. Oxford: Oxford University Press, 2004.

COPELAND, B. J. *Alan Turing: el pionero de la era de la información*. Madrid: Turner Noema, 2016.

DAUTENHAHN, Kerstin. Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 362, n. 1480, p. 679-704, 2007.

DAVIS, Erik. *Tecgnosis: mito, magia y misticismo en la era de la información*. Buenos Aires: Caja Negra, 2023.

DELEUZE, Gilles. *El poder. Curso sobre Foucault*, t. II. Buenos Aires: Cactus, 2014.

FLORIDI, Luciano. *Ética de la inteligencia artificial: principios, retos y oportunidades*. Barcelona: Herder, 2024.

GUTIÉRREZ AGUILAR, Ricardo. Más allá del *uncanny valley*: forma de lo humano, mimesis y expresión. *Bajo Palabra*, n. 28, p. 176-197, 2021.

LARSON, Erik J. *El mito de la inteligencia artificial: por qué las máquinas no pueden pensar como nosotros lo hacemos*. Bilbao: Shackleton Books, 2022.

LIN, Patrick; ABNEY, Keith; BEKEY, George A. *Robot ethics: the ethical and social implications of robotics*. Cambridge, MA: The MIT Press, 2019.

MARX, Karl. *Elementos fundamentales para la crítica de la economía política: borrador 1857–1858*, v. 2. Ciudad de México: Siglo XXI Editores, 2011.

McCARTHY, John; MINSKY, Marvin; ROCHESTER, Nathaniel; SHANNON, Claude. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence (agosto de 1955). *AI Magazine*, v. 27, n. 4, p. 12–14, 2006.

MORI, Masahiro. The uncanny valley. *IEEE Robotics & Automation Magazine*, v. 19, n. 2, p. 98–100, 2012.

PASQUINELLI, Matteo; JOLER, Vladan. El nooscopio: manifiesto. *La Fuga*, n. 25, 2021.

RUSSELL, Stuart; NORVIG, Peter. *Inteligencia artificial: un enfoque moderno*. 2. ed. Madrid: Pearson Prentice Hall, 2004.

SANDRONE, Darío; ILCIC, Andrés. Inteligencia artificial. In: KOZEL, A.; GRINBERG, S.; FARINETTI, M. M. (eds.). *Léxico crítico del futuro*. San Martín: UNSAM Edita, 2024. p. 324–327.

SANDRONE, Darío. Marx y Mumford: dos miradas sobre la maquinaria industrial y su génesis. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad*, v. 18, n. 54, 2023a.

SANDRONE, Darío. Cyborg y autómeta: dos modelos para pensar la inteligencia artificial. In: ROCCA, F.; BORISONIK, H. (comps.). *¿Un futuro automatizado? Perspectivas críticas y tecnodiversidades*. San Martín: UNSAM Edita, 2023b. p. 15–34.

SANDRONE, Darío. Bestiario de máquinas: la monstruosidad maquina en el siglo XX. *Revista Estudios Posthumanos*, v. 2, p. 115–148, 2023c.

SANDRONE, Darío; VACCARI, Andrés; LAWLER, Diego. The centrality of the machine in the thought of Jacques Lafitte. *Philosophy & Technology*, v. 35, art. 28, 2022.

SIMONDON, Gilbert. Psicología de la tecnicidad. In: SIMONDON, Gilbert. *Sobre la técnica*. Buenos Aires: Cactus, 2017. p. 35–130.

SIMONDON, Gilbert. Filosofía y cibernética. In: SIMONDON, Gilbert. *Sobre la filosofía*. Buenos Aires: Cactus, 2018. p. 37–68.

SIMONDON, Gilbert. *El modo de existencia de los objetos técnicos*. Buenos Aires: Prometeo, 2007.

SPARROW, Robert; SPARROW, Linda. In the hands of machines? The future of aged care. *Minds and Machines*, v. 16, n. 2, p. 141–161, 2006.

TURING, Alan. Computing machinery and intelligence. *Mind*, v. 59, n. 236, p. 433–460, 1950.

TURING, Alan. On computable numbers, with an application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society*, ser. 2, v. 42, p. 230–265, 1936–1937.

URE, Andrew. *The philosophy of manufactures*. London: Charles Knight, 1835.

Recebido em 06 de abril de 2025

Aprovado em 27 de agosto de 2025

Publicado em 21 de outubro de 2025