

EL AMOR EN LOS TIEMPOS DE LA INTELIGENCIA
ARTIFICIAL. PARADOJAS Y REALIDADES
DE LOS IMAGINARIOS SOCIOTÉCNICOS
SOBRE LAS INTELIGENCIAS ARTIFICIALES
DE ACOMPAÑAMIENTO Y LA SUSTITUCIÓN
DE LAS PERSONAS

*Love in the Times of Artificial Intelligence. Paradoxes and Realities
of Sociotechnical Imaginaries Regarding Artificial Intelligence Devices
as Companions and their Ability to Replace People*

Enrique ÁLVAREZ VILLANUEVA
Universidad de Oviedo
ORCID: <https://orcid.org/0000-0002-0985-9079>

Enviado: 4 de abril de 2022
Aceptado: 6 de junio de 2022

RESUMEN

Hockenhuil y Leavitt Cohn introducen en su análisis de la escena tecnológica danesa el concepto de «aire caliente» para describir las medias verdades y exageraciones que las compañías utilizan para atraer la atención pública y la inversión. La ciencia ficción y los anuncios de las empresas del sector tecnológico describen con frecuencia a los robots del futuro como seres inteligentes y con capacidad de sentir y, por tanto, con la capacidad de sustituir a las personas como amantes y compañeros de vida. El presente artículo tendrá como objetivo analizar estas expectativas presentes en los imaginarios sociotécnicos de la inteligencia artificial con la intención de descubrir sus problemas y facilitar un juicio menos inflado de «aire caliente».

Palabras clave: inteligencia artificial; robots; expectativas excesivas; imaginarios; aire caliente; computación afectiva; *sexbots*.

ABSTRACT

In their analysis of the Danish technology scene, Hockenull and Leavitt Cohn introduce the concept of «hot air» to describe the half-truths and exaggerations companies use to attract public attention and investment. Science fiction and advertisements from companies in the technology sector frequently describe robots of the future as intelligent, sentient beings with the ability to replace people as lovers and life partners. This paper will aim to analyze the expectations that are present in the sociotechnical imaginaries of artificial intelligence with the intention of uncovering their problems and facilitating a judgment less inflated of «hot air».

Key words: artificial intelligence; robots; hype; imaginaries; hot air; affective computing; sexbots.

1. INTRODUCCIÓN: INTELIGENCIA ARTIFICIAL Y AIRE CALIENTE

La inteligencia artificial (en lo que sigue se indicará con las siglas IA) ha irrumpido de forma imparable en la vida de las personas, creando una gran expectación y nuevas formas de entender tareas comunes como desplazarse en coche o comprar artículos para el hogar por Internet. La IA, pues, ha generado imaginarios que han ido evolucionando a lo largo de la historia hasta el momento presente. Sheila Jasanoff define los imaginarios sociotécnicos como

visiones de futuros deseables colectivamente sostenidas, estabilizadas institucionalmente y realizadas públicamente; animadas por entendimientos compartidos de formas de vida y orden sociales alcanzables a través de, y apoyados por, avances en la ciencia y la tecnología. (Jasanoff 2015, 4)

Para Jasanoff, los imaginarios se co-producen entre la ciencia y la tecnología y la sociedad a través de diferentes enfoques del conocimiento, la tecnología, la tradición y la identidad, redefiniendo el pasado y marcando pautas hacia el futuro de lo que consideramos (o no) aceptable, dándoles así una importante dimensión normativa.

La dimensión colectiva del análisis que Jasanoff propone basado en imaginarios sociotécnicos contempla que estos imaginarios se dan en determinados grupos humanos que pueden ser «tan grandes como naciones-estado o tan pequeñas como una comunidad utópica disidente» (Jasanoff y Simmet, 2021).

Los imaginarios de la IA han ido desde la conveniencia de un mundo donde las máquinas facilitan la vida a las personas con sus impresionantes capacidades hasta las previsiones más ominosas que sostienen que la evolución de las máquinas conducirá al fin de la humanidad. La tendencia a los extremos que las visiones sobre la IA han tenido y siguen teniendo, favorecida por los maximalismos conceptuales que las figuras públicas del campo han sostenido desde los años cincuenta, ha sido recogida por las artes en general, que han canalizado esas posibilidades y amenazas en forma de historias atractivas que a su vez han gozado de gran influencia sobre los imaginarios sociotécnicos sobre la IA actuales.

Alan Turing, el padre de la inteligencia artificial, abrió su célebre artículo «Computer Machinery and Intelligence» con la pregunta «¿pueden pensar las máquinas?» antes de proponer su famosa prueba (Turing 1950, 433). Nueve años después, se presentaba en la conferencia de París el programa GPS (*General Problem Solver*), que no solamente era un programa de resolución de problemas computacionales, sino que tenía como trasfondo la pretensión tanto de construir máquinas inteligentes como de descubrir los secretos y el funcionamiento de la inteligencia humana para imitarla por medio de estas máquinas (López de Mantarás y Meseguer 2017, 24). La gran ambición de conseguir generar una inteligencia artificial general a imagen y semejanza de la humana con el fin de superarla ha permeado toda la historia de la IA, convirtiéndose en el imaginario por excelencia que ha logrado sobrevivir a numerosos escollos, pasando por las críticas bien fundamentadas de personalidades como Herbert Dreyfus desde los años 60 (Mejía Fernández 2020, 127), dos períodos conocidos como «inviernos IA» donde la inversión en esta tecnología cayó dramáticamente por las expectativas excesivas (*hype*) que no pudieron ser correspondidas con resultados, o el cambio de paradigma hacia la inclusión de las emociones en la conceptualización de la IA en los años 90 abanderado por Rosalind Picard (2000), siguiendo las nuevas tendencias en psicología.

Hockenhuil y Leavitt Cohn introducen en su análisis de la escena tecnológica danesa el concepto de «aire caliente» para describir las medias verdades y exageraciones que las compañías utilizan para atraer la atención pública y la inversión. Según ellos, este concepto es a la vez «vacío y productivo» (2021, 306). El aire caliente se vale de diversas técnicas para atraer a grupos sociales y proporcionar proyecciones a futuro que permitan visualizar la estimación de a dónde se dirige la situación actual de una tecnología, para así incentivar al público a compartir un imaginario y actuar conforme a sus pautas. Al mismo tiempo, esta dinámica del aire caliente permite interpretar, canalizar y dar operatividad a conceptos más abstractos del imaginario general y convertirlos

en cuestiones más reducidas y locales, destinadas en muchas ocasiones a promociones de productos o marcas concretas.

Por supuesto, el aire caliente es ampliamente utilizado en el campo de la IA para promocionar diferentes productos. Un ejemplo de esto es la intensa promoción del robot Sophia, creado por David Hanson para ser un robot de compañía en entornos sanitarios, de atención al cliente y educativos. Sophia fue objeto de una enorme campaña de marketing, apareciendo en numerosos programas de televisión para ser entrevistada (en España, por ejemplo, en *El Hormiguero* en 2019), siendo presentada como un robot *que piensa*. En 2017 se le concedió la ciudadanía de Arabia Saudí, convirtiéndose en el primer robot en recibir tal consideración. Por otro lado, algunas voces, como la de la investigadora Joanna Bryson, han denunciado a Sophia como un «engaño masivo» que se aprovecha del desconocimiento y la ingenuidad de la gente sobre la IA y la robótica (Bryson *apud* Devlin 2018, 58).

La polarización de opiniones y la ubicuidad de la IA han creado una cada vez mayor incertidumbre entre la población sobre el futuro de esta tecnología, con el 34 % de los participantes en una encuesta realizada por el Centro de gobernanza de la IA de la Universidad de Oxford afirmando que continuar con la investigación en IA traerá consecuencias negativas, y el 12 % viendo como posible la extinción humana (Buchholz 2019). Sin embargo, en otros países, como India, China o Japón, al menos el 70 % de las personas encuestadas dicen que la investigación en IA hará su vida mejor (Boyon 2022).

El presente artículo sostiene que el imaginario sobre las inteligencias artificiales de acompañamiento (las inteligencias artificiales asistenciales, pero especialmente las destinadas a satisfacer las necesidades sexuales y afectivas) las contempla no como entidades que cumplen una función específica, sino como sustitutos de las personas a medio plazo, y así se muestra en diferentes obras artísticas y culturales en general. El robot social se define en la literatura académica como «una entidad artificial inteligente (humanoide o robot) que interactúa de una manera interpersonal para lograr realizar metas predefinidas y relativas a un dominio en el contexto práctico en el que sirven» (Breazeal, Takanishi y Kobayashi 2008, 1349), pero usualmente es presentado, como Sophia, como una entidad capaz de tener pensamientos y voluntad propios. El artículo, por tanto, sostiene que la piedra angular sobre la que se construye esta supuesta sustitución que está por llegar es que estas inteligencias artificiales sean realmente inteligentes. Por lo tanto, el esfuerzo principal del mismo será plantear un análisis sobre la posibilidad de que tal cosa suceda y las implicaciones que esto tendría para reducir el aire caliente que los imaginarios mencionados crean. Finalmente, se hará también una reflexión acerca de las expectativas excesivas y el aire caliente como elementos

de los que, pese a sus inconvenientes para el debate justificado, no se puede prescindir.

2. LA CUESTIÓN DE LA INTELIGENCIA ARTIFICIAL

Como se ha expuesto más arriba, la concepción inicial de las personalidades pioneras de generar una inteligencia análoga a la humana trajo consigo un sesgo insalvable en el imaginario colectivo hacia el hecho de que en algún momento se conseguiría emular –y superar– tal inteligencia de forma artificial. A pesar de que el conductismo, coetáneo e inspirador de muchas de estas ambiciones, quedó ampliamente superado hace décadas por otras teorías psicológicas más comprensivas, el atractivo conceptual y comercial de imaginar una inteligencia análoga a la humana en una máquina goza todavía de gran popularidad. Antes de pararnos a pensar en cuestiones más intrincadas, como si un robot podría ser considerado una persona, recibir derechos o poseer agencia moral, debemos plantearnos brevemente esta cuestión de si podemos esperar que las máquinas sean *inteligentes*.

En el seminario de Dartmouth de 1955, tomado tradicionalmente como el evento fundacional del campo de la inteligencia artificial, McCarthy, Minsky, Rochester y Shannon definieron el problema de la inteligencia general como el de crear una máquina que actúe de maneras que puedan ser consideradas inteligentes si un humano se comportase así, entendiéndose que «todo aspecto del aprendizaje y cualquier otra característica de la inteligencia puede, en principio, ser descrita con la precisión suficiente como para ser emulada por una máquina» McCarthy *et al.* 2006, 12). Esta forma de entender la inteligencia dio lugar a la aproximación conocida como *Good Old-Fashioned AI* (GOFAI), un enfoque *top-down* consistente en el establecimiento de una serie de reglas formalizadas lógicamente e implementadas en la *mente* artificial y que luego serían aplicadas en los diferentes contextos en los que la entidad artificial se viera envuelta.

De una forma más matizada, Russell y Norvig (2019, 2) establecen cuatro definiciones por las cuales podría decirse que una máquina es inteligente: que piense *como lo hacen los humanos*, que actúe *como lo hacen los humanos* (que pase un test de Turing), que piense *racionalmente* o que *actúe racionalmente*. Estas dos últimas no satisfarían probablemente lo que las personas entendemos comúnmente por *inteligencia*, ya que, como veremos más abajo, la inteligencia es más que solamente racionalidad. Las dos primeras definiciones son más complejas, pero también más interesantes. Debido a que los humanos aprendemos por medio de la experiencia, los enfoques *bottom-up*, que extraen el

aprendizaje de la experiencia y que toman precisamente como inspiración el aprendizaje humano, han ganado gran relevancia en las últimas décadas, especialmente debido a la nueva capacidad de almacenar y procesar ingentes cantidades de datos gracias a la mejora de la tecnología computacional.

Al preguntarnos cómo seríamos capaces de alcanzar el punto en el que una creación artificial pueda mostrar inteligencia *como la humana*, Nick Bostrom (2014, 33) ofrece dos opciones: la primera consistiría en emular completamente el cerebro humano, copiando *in silico* la fisiología de este con la idea de que con un modelo físico y funcional idéntico la consciencia y el comportamiento inteligente deberían surgir espontáneamente. Esta concepción no está exenta de problemas, ya que el sistema nervioso humano no se limita al cerebro, y las estrechas comunicaciones del encéfalo y el resto del cuerpo en un contexto de interacción *encarnada* con el medio son clave para concebir el comportamiento inteligente (Burden y Savin-Baden 2019, 145). Si optamos por la generación artificial de inteligencia por medio de la computación, sin necesidad de recurrir a una fisicalidad análoga a la humana, encontramos el problema que llevan enfrentando las investigadoras en IA desde la concepción del cambio: es necesario definir la inteligencia y positivizarla para poder programarla y hacerla funcional.

No solo esto, sino que también encontramos muchos problemas relacionados con el concepto de inteligencia, como el denunciado por John Searle (2002) de la necesidad de contemplar la consciencia como requisito para establecer una inteligencia no dependiente del observador. Por ejemplo, los asistentes de voz de nuestros teléfonos tienen un comportamiento que nos parece inteligente porque pueden darnos la receta para cocinar una paella de forma inmediata si la solicitamos, pero difícilmente podemos considerar esto como una inteligencia *completa*, no dependiente del observador, como la humana.

Otro factor importante en la inteligencia humana es el de las emociones, toda vez que ya es aceptado universalmente por la comunidad de especialistas que existe una estrecha interconexión entre la corteza prefrontal (habitualmente considerada el *centro de control racional* del cerebro), las áreas perceptivas y los sistemas subcorticales relacionados con el procesamiento de las emociones (Rolls 2018). Como denuncia Erik J. Larson (2021), un concepto de inteligencia artificial general, con las mismas capacidades de un ser humano, tiene un grave problema de planteamiento, ya que hasta ahora el progreso en la IA se ha dado en cuestiones específicas –capacidad de cálculo, reconocimiento facial, análisis de datos, etc.–, pero no hay nada que nos haga pensar que la unión y perfeccionamiento de estas capacidades dará como

resultado una inteligencia como la que se busca, ya que tampoco existe, ni se sabe cómo podría existir, un algoritmo para tal resultado.

Otro aspecto a tener en cuenta es que, precisamente, esa interconexión de las áreas tradicionalmente asociadas con la racionalidad y la emocionalidad hace que los seres humanos mostremos sesgos y comportamientos impulsivos. Como sostiene Stuart Sutherland (2015, 168), «el estrés, las recompensas, los castigos y las emociones fuertes disminuyen la flexibilidad del pensamiento y provocan conductas irracionales». Esto ha de ser tenido en cuenta si se pretende desarrollar inteligencias artificiales creíbles, aunque los seres humanos estemos sesgados hacia la aceptación de que las conductas programadas en inteligencias artificiales son genuinamente inteligentes, incluso si estas conductas son notablemente diferentes a las humanas (Nass y Moon 2000, 94, Vallverdú *et al.* 2018). Marvin Minsky, uno de los pioneros presentes en Dartmouth que décadas más tarde comenzó a aceptar las críticas de autores como el ya citado Dreyfus y a añadir elementos más complejos de programar al concepto de inteligencia –sentido común, emocionalidad, autoconciencia...–, acaba por admitir que, debido a la complejidad de incorporar todos estos elementos a la programación de una mente artificial, los errores en el desempeño de las nuevas inteligencias creadas serán una constante y que, incluso siendo posible en un futuro, «no será un camino fácil» (Minsky 2006, 346).

Llamar al campo de conocimiento «*inteligencia* artificial» debido a que es capaz de generar productos que de algún modo pueden ser considerados inteligentes pasando de puntillas por los problemas conceptuales que por el momento siguen siendo insalvables es una suerte de sinécdoque que claramente perpetúa el sesgo que los imaginarios alientan sobre la capacidad real de las máquinas en desarrollo. Este sesgo se ve reforzado no solo por los múltiples ejemplos de robots *realmente* inteligentes en ciencia ficción, sino también por los anuncios llenos de aire caliente de muchos expertos del propio campo, como Ray Kurzweil (2006) o Elon Musk, entre muchos otros (Lizana 2022).

3. SUSTITUCIÓN DE LAS PERSONAS

Una de las áreas en las que el aire caliente se deja notar de manera más intensa es en la eventual sustitución de las personas por seres sintéticos en distintos ámbitos. La Universidad de Oxford sostiene que, en el futuro cercano, un 90 % de los trabajos existentes será realizado por inteligencias artificiales. El Nomura Research Institute, además, calcula que el 49 % de trabajos actuales desaparecerá en 15 años (Haque 2020, 207). A pesar de lo rotundo

de estos datos, es necesario entender que cualquier estudio que analice la sustitución de humanos por máquinas en la cadena productiva deberá tener en cuenta muchos otros factores que a menudo, por la conveniencia del impacto mediático que la referencia a los robots y a la inteligencia artificial tiene, son ignorados. En el corto plazo, no parece que los robots vayan a tener un impacto tan grande como otros factores como la deslocalización o la realidad geopolítica. En 2018, Elon Musk trató de automatizar por completo la cadena de montaje de vehículos Tesla en su fábrica de Silicon Valley. Tras el intento inicial, fue totalmente incapaz de satisfacer siquiera la mitad de la producción que se había establecido como objetivo, y terminó por llamar al intento «infierno manufacturero» y decir que «los humanos están infravalorados» en Twitter (Darling 2021, 13). En otros ámbitos, como el de la educación, los intentos de sustituir a personas por robots tampoco han terminado de convencer a los estudiantes, además de plantear formidables retos que han sido analizados en otro artículo (Álvarez Villanueva 2021).

De todos modos, los imaginarios alimentados por las grandes historias de ficción van, como veremos más abajo, orientados hacia la sustitución de las personas no solo en lo laboral, sino en ámbitos que requieren que las inteligencias artificiales sean capaces de emular por completo a los seres humanos.

Aunque el presente trabajo se centrará en las inteligencias artificiales destinadas al acompañamiento sentimental y a la satisfacción de las necesidades sexuales, conviene mencionar, aún de pasada, el otro gran polo de robots sociales dedicados a la asistencia de personas con trastornos o necesidades especiales. En el terreno de los cuidados asistenciales, algunos robots no humanoides como la foca robótica Paro están teniendo una excelente acogida y ofreciendo prometedores resultados en pacientes con demencia y autismo (Šabanovi *et al.* 2013, Hung *et al.* 2019, Libin y Libin 2004, 1795). Sin embargo, numerosas voces han puesto en tela de juicio la necesidad real de que estos robots muestren una genuina inteligencia o sociabilidad para cumplir bien con su trabajo (Fong, Nourbakhsh y Dautenhahn 2003, 160), o incluso si la sustitución de personas por robots es moralmente aceptable en casos de vulnerabilidad tales como el cuidado de ancianos por el peligro de condenar a estas personas a una suerte de «ostracismo emocional» al dejar de interactuar con otros seres humanos (Laitinen, Niemelä y Pirhonen 2019). Pareto Boada, Román Maestre y Torras Genís (2021) hacen una revisión pormenorizada sobre los principales debates éticos en torno a los robots de cuidados asistenciales. En cualquier caso, este tipo de análisis recibe más atención académica que mediática, y por tanto estos robots están menos presentes en los imaginarios y sufren menos la influencia del aire caliente que los destinados al acompañamiento sentimental y sexual, de modo que nos centraremos en estos últimos.

El imaginario sobre el amor con entidades artificiales se retrotrae, como mínimo, al mito de Pigmalión. En este relato, el escultor del mismo nombre, incapaz de encontrar a la mujer perfecta, decidió crear él mismo varias esculturas femeninas, enamorándose de una de ellas, Galatea, que finalmente fue traída a la vida por mediación de Afrodita. En nuestra época, podemos encontrar numerosos ejemplos de robots con morfología humana y dotados de inteligencia que resultan atractivos para las personas. Por ejemplo, María, la protagonista de la película *Metrópolis* (Lang 1927) tiene un doble robótico que comparte el atractivo de aquella, y que incluso llega a actuar como una bailarina exótica. En la película *AI* (Spielberg 2001) Gigolo Joe es un robot que está programado para emular el romance y tener sexo. Otros ejemplos similares pueden ser encontrados en las series *Westworld* (Nolan 2016) o *Humans* (Arnold *et al.* 2015) o en la película *Her* (Jonze 2013), entre muchas otras.

Jennifer Rhee hace un interesante análisis de la novela *Do Androids Dream of Electric Sheep?* (Dick 1968), obra que inspiró la película *Blade Runner* (Scott 1982) a pesar de las grandes diferencias entre ambas, y de su precuela *We Can Build You* (Dick 1972). Según Rhee, lo que distingue a las entidades artificiales de las personas es la capacidad de sentir emociones, aunque durante el desarrollo de las novelas la trama llevará a la conclusión de que los no humanos muestran comportamientos más empáticos hacia las personas que al contrario (Rhee 2018, 118). Esto tiene su reflejo en el imaginario sobre los robots, ya que lo que hace especiales a todos los robots que aparecen en la ciencia ficción como amigos o amantes de las personas es su capacidad de sentir y relacionarse, en oposición de otras máquinas como los frigoríficos o los lavavajillas, que no muestran ningún rasgo emocional. Aunque las personas todavía tiendan a *humanizar* de algún modo a las máquinas no programadas para tener un comportamiento empático (Jones 2017), tenderán a calificar de «robot» solamente a aquella máquina que imite de algún modo el comportamiento humano –excepto en el caso de que lleven un adjetivo relacional (de cocina, de limpieza...), algo que sin embargo no está presente en absoluto en la definición que da la International Organization for Standardization (ISO), que solo se basa en su operatividad y autonomía para realizar diferentes tareas sin supervisión (Devlin 2020, 44).

Respecto a la sustitución de las personas por robots fuera de la ficción, existe también una abundante literatura cuyo pionero fue David Levy con su obra *Love and Sex with Robots* (2008). Levy hace un concienzudo análisis sobre cómo y de quién nos enamoramos las personas, y por qué, debido a la flexibilidad en la personalidad y características físicas que los robots pueden tener, estos pueden ser mejores amantes y más atractivos para nosotros y nosotras que otros seres humanos. Levy incluso se plantea la cuestión de si

podrá ser creíble la reciprocidad sentimental de los robots hacia las personas. Apoyándose en las investigaciones de Reeves y Nass (2002), que demostraron que las personas interactuamos con ordenadores y otras máquinas dotándoles de cierta entidad emocional y dignidad a pesar de ser plenamente conscientes de que no sienten emoción alguna, Levy concluye que «si parece que le gustas a un robot, si se comporta en todas las formas como si le gustaras, entonces puedes asumir con seguridad que realmente le gustas, en parte porque no hay evidencia en sentido contrario» (Levy 2008, 147). La misma línea de argumentación siguen Meacham y Studley, que sostienen que frente a la perspectiva inferencialista analógica, que consistiría en aceptar que importamos al sujeto con el que interactuamos siempre y cuando sea capaz de experimentar realmente emociones, se puede aceptar una perspectiva enactivista, que enfatiza los patrones de interacción, ya que realmente todas las personas actúan con los otros basándose en sus gestos y comportamientos y no en representaciones internas ajenas a las que no les es posible acceder (Meacham y Studley 2017, 103). Algunos estudios apuntan a que, en el caso del sexo, la conducta de los participantes en el acto es uno de los principales factores a la hora de alentar el deseo en las personas, llegando al punto en el que las personas que tienen encuentros sexuales con robots ponen en suspenso la incredulidad que sería esperable en estos casos siempre que los códigos sexuales establecidos por la cultura se cumplan y los participantes se sientan seguros y cómodos (Szczuka, Hartmann y Krämer 2019).

En sentido contrario, hay una gran bibliografía académica que sostiene que, incluso si los robots sexuales (*sexbots*) son capaces de satisfacer los gustos y necesidades relativas al sexo de algunas personas, probablemente la idea de robots que sustituyan a las personas en relaciones afectivas más amplias es una labor mucho más difícil. Por ejemplo, John Sullins (2012, 408) observa que el amor es más que tener un comportamiento específico y que incluye una dimensión ética de autorrealización, compasión y enriquecimiento mutuo, y concluye que un robot difícilmente podría contribuir a tales fines. Mark Coeckelberg (2010, 7) comparte esta visión, sosteniendo que, aunque no es totalmente necesario que el robot sienta empatía para ser amado por una persona, esta sí es importante cuando es capaz de compartir sentimientos de vulnerabilidad que contribuyen a que su acompañante humano pueda ver reflejados este tipo de sentimientos y se sienta comprendido. Por otro lado, Nishio, Taura y Ishiguro (2012) han encontrado que es posible para los robots, usando expresiones faciales, afectar los sentimientos de las personas con las que interactúan, aunque admiten que la investigación está en un episodio temprano. A pesar de que los robots sexuales actuales, como Roxxy, Solana o Harmony, no tienen realmente una capacidad de expresión emocional

realista, existen comunidades de usuarios de muñecas sexuales y androides en varios países, como Davecat, un hombre de Detroit que comparte su vida con sus muñecas Elena y Sidore, considerando a esta última su mujer y afirmando que ha desarrollado vínculos emocionales con ella análogos a los que desarrollaría con una mujer humana (Danaher 2018, 9).

Nyholm y Frank (2018, 234) opinan que, si un robot está programado para amar a una persona y no tiene una elección real, no es posible tener la misma experiencia que se tendría con un amante que tenga posibilidad real de elegir. Por tanto, sería deseable, como se puede ver en la ficción, que los robots tuvieran esa capacidad. Los autores dejan abierta la posibilidad de construcción de una máquina con estas características, pero reseñan que la dificultad de fabricar un robot así sería inmensa, y que quizá no mereciera la pena embarcarse en tal proyecto.

A pesar de lo expuesto, las personas siguen estableciendo relaciones serias y contrayendo matrimonio con muñecas sintéticas (Bates 2013) y, aunque este fenómeno todavía sea de nicho, algunos expertos han demostrado que factores como la soledad, que normalmente se toman como condicionantes para que una persona opte por una compañía no humana, no juegan un papel determinante en la elección de seres sintéticos sobre personas para desarrollar relaciones afectivas (Szczyka y Krämer 2017). La idea de que los robots tengan una conciencia y una capacidad de elección real, además de capacidad emocional genuina, es la versión que más ha permeado en el imaginario sobre los robots de acompañamiento gracias a las historias de ficción que, como se ha expuesto más arriba, los presentan como personas artificiales que sustituyen a las de carne y hueso.

Es interesante observar que, según una encuesta hecha en 2013 a 1000 adultos, hay más personas que dirían que si una persona en una relación sentimental con otra persona tuviera relaciones sexuales con un robot estaría cometiendo una infidelidad (42 %) que personas que no lo verían así (32 %) (McArthur 2017, 31). Según la versión prevalente en el imaginario sobre los robots de acompañamiento, pues, los robots son vistos como personas sintéticas. Para tal fin, parece claro que necesitan poseer capacidad emocional e inteligencia, lo cual abre una batería de problemas que serán analizados a continuación.

4. EMOCIONES, MORALIDAD Y DERECHOS

Qué son las emociones, al igual que la inteligencia, es algo más o menos evidente en la cultura folk, pero al tratar de definir las con exactitud para que

las inteligencias artificiales sean capaces de implementarlas, se encuentran con numerosas dificultades.

La primera de estas dificultades es que no hay, ni remotamente, un consenso sobre qué son las emociones. De hecho, hay tantas teorías diferentes que numerosos académicos las clasifican en varios grupos según qué parámetros son los principales para cada teoría (Scarantino 2018, Sánchez-Escribano 2018, Strongman 2006, Pineda Oliva 2019). En este momento, por tanto, los robots desarrollados dentro del campo de la computación afectiva están programados con estructuras emocionales que emulan el comportamiento emocional con mayor o menor complejidad seleccionando la teoría de las emociones que mejor se adapte a la tarea de programación. Las teorías OCC de Ortony, Clore y Collins (1988) en conjunción con la teoría de las emociones básicas de Ekman (Ekman, Friesen y Ellsworth 1972) son las más populares, pero estas teorías no están en absoluto exentas de críticas (Hudlika 2014, 8, Feldman Barret 2018, 78). Si, como sostiene Scarantino (2018), hay un amplio consenso entre la comunidad científica de que las emociones tienen una base cognitiva, y que la conceptualización y reflexión humanas y la emoción están ligadas (Damasio 1994, 145), se sigue que la capacidad para dotar de consciencia e inteligencia a las máquinas es una tarea mucho más compleja de lo anticipado.

La cuestión de que las máquinas no vayan a estar dotadas de inteligencia y emocionalidad genuina enfría otro de los temores más latentes sobre la IA: el hecho de que quieran tomar el control y dominarnos. Al menos, *conscientemente*. Nick Bostrom (2017) y Max Tegmark (2017, 43) han apuntado que todavía existiría la posibilidad de que una IA destruyera el mundo o a la humanidad si le son asignados los fines equivocados. Una máquina que tuviera como fin hacer el máximo de sujetapapeles posibles tendería a optimizar la producción de estos, fabricando todos los que le fuera físicamente posible utilizando todos los medios a su alcance. Literalmente. De todas formas, parece más asequible programar una máquina para que tenga ciertas limitaciones en su operatividad que dotarla de inteligencia real que le permitiera *desear* activamente la destrucción de la humanidad o empresas similares. Algunas personalidades del mundo de la IA, como Oren Etzioni, CEO del Allen Institute for AI, o Bill Gates, afirman que preocuparse de este tipo de amenazas no tiene mucho sentido (Russell 2019, 152).

Otra cuestión es importante a la hora de pensar sobre la IA: los imaginarios sociotécnicos, como sostiene Jassanof, están embebidos en las culturas políticas de las diferentes naciones (Jassanof y Kim 2015b, 335). Al mismo tiempo, son «estabilizados por las instituciones» (Jassanof y Simmet 2021, 4), siendo estas instituciones las que articulan su valor normativo y ayudan a apuntalarlos socialmente. Veremos brevemente a continuación algunos ejemplos sobre

cómo las diferentes naciones toman partido social e institucional en el asunto de los robots que se presentan como sustitutos de las personas.

Como se ha comentado anteriormente, las obras de ciencia ficción de la cultura occidental retratan a menudo estas tecnologías como amenazas existenciales. Por otro lado, la sociedad japonesa, por ejemplo, tiene más facilidad para ver a los robots como compañeros y amigos, quizá mediada por la religión sintoísta, que considera una multiplicidad de entidades no humanas poseen un espíritu (*kami*) (Turner 2019, 50, Klærke-Olesen *et al.* 2018). Así, Japón es líder en patentes para la creación de robots asistenciales (Meacham y Studley 2017, 101). Además, las costumbres sociales japonesas y la mayor aceptación de robots e inteligencias artificiales y entidades virtuales en general crea una situación propicia para la asimilación de estas (Corner 2018, Appel, Marker y Mara, 2019). China, por su parte, sigue este precepto introduciendo profusamente robots y tecnologías de IA con la intención de convertirse en líder de la tecnología e ignorando por lo general los grandes temores del imaginario occidental (van Noort 2022), a pesar de que en su imaginario tradicional sí existen ciertos prejuicios contra las máquinas como corruptoras del espíritu de las personas (Kroes y Verbeek 2014, 2). En términos de robots de acompañamiento, China tiene cada vez una mayor apertura, debido en parte a los problemas demográficos que enfrenta, habiendo mucha mayor abundancia de hombres que de mujeres. Así, el negocio de los *sexbots* que no solo son usados para el sexo, sino para perseguir un vínculo emocional significativo, sigue en aumento (Fong 2017). Por el momento, debido a lo incipiente de la tecnología, no hay estrategias nacionales para regular la producción y consumo de *sexdolls* o robots de acompañamiento, excepto en algunos estados de Estados Unidos (Forbes 2020). También pueden incluirse aquí una serie de condenas en que tuvieron lugar en 2017 en Reino Unido por la importación de robots sexuales que imitaban la anatomía infantil, basados en una ley generalista y muy antigua de «importación de objetos obscenos» (Devlin 2018, 246). Algunos grupos, como el CASR (*Campaign Against Sex Robots*) llevan tratando de paralizar la producción de robots sexuales desde 2015, logrando algunos éxitos parciales, como cuando influyeron en oficiales del gobierno malasio para cancelar la 2nd *International Conference on Love and Sex With Robots* organizada por David Levy, aunque en Occidente sus éxitos han sido más bien modestos (Danaher, Earp y Sandberg 2018, 47).

5. PERSONAS ROBÓTICAS Y SU ENTIDAD MORAL

El debate sobre la capacidad de las inteligencias artificiales de ser agentes y pacientes morales es uno de los más interesantes que se generan al contemplar

la inteligencia en las máquinas, lo que llevaría, de ser esto posible, al debate de si sería o no conveniente asignarles derechos.

Uno de los casos más paradigmáticos de decisión moral que podrían enfrentar las máquinas –por ejemplo, los coches con conducción automática que se enfrenten a un accidente inevitable– es el que Philippa Foot diseñó en 1967 sobre el tranvía que se dirige inexorablemente al atropello de cinco personas, excepto si un sujeto que tiene la capacidad de intervenir opta por mover una manivela que acciona un cambio de vía que posibilitaría que el tranvía atropellase solo a una que está en una vía paralela (Foot 1967, 2). Resulta más o menos evidente que el modo de acción más conveniente en términos puramente utilitaristas sería sacrificar a una persona y salvar a cinco, ya que el daño total sería menor. Sin embargo, resulta interesante ver cómo en estudios elaborados al respecto con personas reales, frente al 82 % de alemanes que moverían la palanca para sacrificar a un individuo y salvar a los otros cinco, solo el 58 % de los chinos actuaría de tal modo (Rahwan, Skork y Hungerland 2020). No existe, pues, un consenso moral mínimo sobre el que programar las decisiones morales en las máquinas, ni siquiera en casos como estos.

Si introducimos algunas variantes del mismo experimento, como la posibilidad de detener el tren que se dispone a atropellar a un grupo de personas empujando a una persona corpulenta para que caiga a la vía y se sacrifique bloqueando el paso del tren, tenemos algunas variedades en los resultados, pero lo más interesante es que estudios de neurociencia han determinado que al enfrentarse a esta versión del dilema se activan áreas del cerebro relacionadas con la reacción emocional (D'Olimpio 2016), lo que está en consonancia con numerosos estudios sobre cómo el pensamiento racional no es el único implicado en las decisiones que tomamos (Damasio 2005, 84).

Si contemplamos la decisión racional como la válida moralmente, podemos concluir, siguiendo la opinión de Joseph Emile Nadeau, que solo los robots pueden ser éticos ya que solo ellos pueden ser puramente racionales (Nadeau *apud* Gunkel 2017, 64). Sin embargo, no está claro que la moralidad puramente racional sea universalizable a máquinas y a personas, ya que, como reseña Mar Cabezas –y en consonancia con lo reseñado sobre la neurociencia de las decisiones morales–:

las emociones [...] tales como el miedo, la tristeza y la ira pueden funcionar como marcadores universales de posibles daños morales para todos los miembros de la especie humana, más allá de las diferencias culturales, como alarmas que advierten de un posible daño. (Cabezas 2014, 265)

Como hemos visto antes con las reacciones en encuestas sobre el caso del tranvía, no todas las personas decidirían intervenir para salvar a unos y condenar a otros. De hecho, en caso de intervenir, probablemente la persona que ejecute la acción sería procesada por un delito contra la vida de las personas. Es curioso ver también como las predicciones asignan a los coches autónomos una capacidad futura de ocasionar muchos menos accidentes que los conductores humanos y, aun así, solo un 43 % del público norteamericano estaría dispuesto a darles una oportunidad (Darling 2021, 63).

La percepción de la agencia moral, como hemos visto, también puede estar sujeta a variaciones según la cultura. Mitcham cita la reflexión de Collins y Kusch sobre que los británicos y los azande, un pueblo del centro-norte africano, no tienen el mismo concepto de agencia moral, ya que estos últimos tienen una cultura animista donde ciertas fuerzas naturales tienen una intervención mucho más directa en las acciones de las personas que en Occidente (Collins y Kusch *apud* Mitcham 2014, 20) –aunque muchas personas occidentales también creen en seres sobrenaturales y magia, el sistema legal no les asigna una agencia–. De este modo, no es disparatado pensar que algunas culturas que no establecen un corte tan grande entre lo humano y lo no humano, como la japonesa, puedan asignar un grado diferente de agencia a los robots. De hecho, una estatua japonesa juzgada en el siglo XVIII por ser la única testigo de un crimen y no haber hecho nada por evitarlo sigue cumpliendo condena a día de hoy (Turner 2019, 50). A lo largo de la historia muchos animales han sido condenados en tribunales organizados a la humana, pero este tipo de prácticas han caído, por lo general, en desuso por lo inadecuado de las mismas (Darling 2021, 61).

Hanson (2014, 63) nos recuerda que los castigos y las gratificaciones no son exclusivas para los seres humanos; los animales también reciben comida y muestras de afecto cuando se comportan conforme a lo esperado por las personas, o castigos cuando no lo hacen. Es interesante notar que estos castigos y recompensas solo tienen sentido cuando el receptor es capaz de sentir placer o disgusto, y también que cualquier recompensa dada por las personas tiene como referente el interés de estas. Cualquier sistema ético que se implemente hacia los robots tendrá en cuenta primeramente las consideraciones morales de las personas; por tanto, será para ellos una moral heterónoma, lo cual solo es justo si consideramos a los robots creaciones humanas al servicio de las personas. Si los robots son inteligentes o sienten emociones genuinas, sería justo que ellos tuvieran una participación en la configuración de sus propios sistemas morales. Puestos a imaginar, si esto ocurriese probablemente no sería plato de gusto para las personas, ya que todo sistema moral está asociado a una cultura y sirve a intereses propios, ya sea del colectivo nacional

o del humano. Así, principios morales en las máquinas que fueran diferentes a los humanos probablemente no serían bien recibidos por nosotros.

En cualquier caso, los sesgos que los humanos tenemos hacia las máquinas descritos por Reeves y Nass hacen que tenga sentido plantearse que las personas tratemos a las máquinas como a otras personas de igual entidad que nosotros. En opinión de Anne Foerst, es solo cuestión de tiempo que esto se dé: «Creo que una máquina puede ser definitivamente una persona. [...] No serán humanos porque están hechos de diferente material, pero serán parte de nuestra comunidad, no hay ninguna duda en mi mente [sobre eso]» (Foerst *apud* Gunkel 2017, 172).

El debate de si los robots podrán ser considerados personas legales pone de relieve un claro sesgo, toda vez que existen animales no humanos a nuestro alrededor que poseen una inteligencia y una capacidad emocional genuina, y que incluso tienen una complejidad intelectual capaz de generar desórdenes mentales y rituales de corte mágico-religioso (Herreros Ubalde 2018, 288). Sin embargo, salvo en excepciones muy contadas y en general como sujetos de derechos pasivos, no se les concede una personalidad jurídica (Turner 2019: 140). Es entendible, de todos modos, que haya una mayor tendencia a personificar a los robots debido al citado aire caliente que los investigadores del campo promueven y que supone un importante acicate para los imaginarios sociotécnicos sobre la IA.

Algunos académicos salen al paso de este debate negando taxativamente que se deba entender que los robots son *como* las personas, lo que él describe como la «falacia androide» (Turner 2019, 189). Kate Darling (2021) sostiene una interesante tesis sobre cómo el modo más cabal de entender a las máquinas, incluso a las que dan la impresión de ser inteligentes, es de una forma análoga a cómo nos hemos relacionado con los animales, teniendo en cuenta que hemos tenido una gran historia tomándolos como compañeros y estrechos colaboradores en tareas que no éramos capaces de desarrollar por nosotros mismos. De todos modos, no hay que perder de vista que, por lo menos en el corto y medio plazo, las máquinas no sufrirán realmente, y por mucho que nos empeñemos en verlos como amigos y compañeros de vida, no debemos olvidar que dotarles de derechos podría generar más problemas que beneficios, teniendo en cuenta que los robots sociales, que son los que más probablemente veamos como personas, son producidos para realizar funciones para nosotros los humanos en un régimen de esclavitud funcional. David Levy sostenía que los robots podrán ser los mejores amantes para nosotros, pero ¿decidirán ellos en alguna cuestión?

6. CONCLUSIÓN: AIRE CALIENTE Y EXPECTATIVAS

Parece claro que, pese a que la capacidad computacional aumente exponencialmente en los próximos años, y a pesar del criterio de algunos especialistas (Kurzweil 2006), la capacidad emocional y la verdadera inteligencia para las máquinas requerirá de la clarificación conceptual y operativa de ambas categorías o la capacidad de hacer un modelo funcional del cerebro y probablemente de algunas otras áreas del cuerpo. Mientras tanto, no obstante, el aire caliente producido por las historias de ciencia ficción y las constantes evoluciones técnicas anunciadas con titulares capciosos en prensa siguen haciendo pensar al público que estas capacidades son ya una realidad o están cerca de serlo.

Los imaginarios sobre las IA de acompañamiento han sido impulsados por las obras académicas, como el trabajo de Levy atestigua, pero también por la ciencia ficción, algunos expertos del campo y los titulares de prensa cargados de aire caliente que animan a creer que las personas sintéticas podrían ser una realidad. Estos imaginarios, como Jasanoff explica, a su vez co-producen e impulsan la tecnología creada, mostrando robots cada vez más capaces de imitar los comportamientos humanos.

Es posible delimitar con argumentos los límites del aire caliente en campos como la conveniencia de la asignación de derechos de las máquinas o la capacidad de dotarlas de capacidades emocionales; lo que no es posible es limitar de este modo las expectativas, emociones y sueños de las personas que interactúan diariamente con inteligencias artificiales de una forma u otra. Es importante reflexionar sobre el hecho de que las leyes las hacen las personas que el público vota, y que si dejamos que la sociedad se deje llevar por sensaciones alarmistas quizá acabemos enfrentándonos a los debates equivocados que terminen en leyes exageradas o injustas mientras ignoramos otros, como el verdadero trabajo en la sombra de millones de personas que trabajan sufriendo una inmensa precariedad para que grandes compañías puedan presumir de los flamantes resultados de sus sistemas de IA (Crawford 2021).

Por tanto, es importante aprender a convivir con el aire caliente que haga subir el globo del bombo publicitario más y más hacia las alturas, pero quizá sea conveniente también viajar en este globo con una brújula y un mapa, para al menos tener una idea, aunque sea vaga, de a dónde nos dirigimos.

7. REFERENCIAS BIBLIOGRÁFICAS

ÁLVAREZ Villanueva, Enrique. «Pensar (en) el futuro: Cuestiones sobre la enseñanza de la Filosofía y la Inteligencia Artificial», *Paideia* 116 (2021): 59-70.

- APPEL, Markus, Caroline Marker y Martina Mara. «Otakuism and the Appeal of Sex Robots», *Frontiers in Psychology* 10 (2019). <https://doi.org/10.3389/fpsyg.2019.00569>.
- ARNOLD, Lewis, *et al.* directores. 2015. *Humans*, Kudos Film and Television. https://www.primevideo.com/dp/OSXL4LD3ESSWRXOSW0VUU6JN6Z?ref_=dvm_pds_tit_ES_lb_s_g_mk_w_smwT3y2ystadc&mrntrk=p_crid_585690138231_slid__pgrid_131344918062_pgeo_1005509_x__ptid_kwd1637760113485&gclsrc=aw.ds&&gclid=CjwKCAjwTlaVBhBkEiwAsr7cwZai3GyaqpwnqgMXig4m3rnM0AfkpY5paOC_A9dfRfYS4uY1YP1RoC4CoQAvD_BwE.
- BATES, Daniel. «Meet Davecat - The 40-year-old who has shunned «organic women» to marry a synthetic doll and keep another as a mistress». *Mail Online*. Recuperado el 22 de abril de 2022, de <https://www.dailymail.co.uk/news/article-2439522/Davecat-40-shunned-organic-women-marry-synthetic-doll.html>.
- BOSTROM, Nick. *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014. <https://doi.org/10.1016/j.futures.2015.07.009>.
- BOYON, Nicolas «Opinions About AI Vary Depending on Countries' Level of Economic Development». *Ipsos*. Recuperado el 22 de abril de 2022, de <https://www.ipsos.com/en/global-opinions-about-ai-january-2022>.
- BRAEZEAL, Linda, Atsuo Takanishi y Tetsunori Kobayashi. «Social Robots that Interact with People». En Bruno Siciliano y Oussama Khatib (Ed.), *Springer Handbook of Robotics*, 1349-1369. Berlín: Springer, 2018. <https://doi.org/10.1007/978-3-540-30301-5>.
- BUCHHOLTZ, Katharina. «Americans Fear the AI Apocalypse». *Statista*. Recuperado el 22 de abril de 2022, de <https://www.statista.com/chart/16623/attitudes-of-americans-towards-ai/>.
- BURDEN, David y Maggi Savin-Baden. *Virtual Humans. Today and Tomorrow*. Boca Ratón: CRC Press, 2019. <https://doi.org/10.1201/9781315151199-14>.
- CABEZAS, Mar. *Ética y emoción. El papel de las emociones en la justificación de nuestros juicios morales*. Madrid: Plaza y Janés, 2014.
- COECKELBERGH, Mark. «Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations». *Studies in Ethics, Law, and Technology* 4 (3) (2010) <https://doi.org/10.2202/1941-6008.1126>
- CORNER, Natalie «Sex doll the size of a CHILD being made in a Japanese factory reduces a TV presenter to tears – as the manufacturer admits customers decide the robots' age». *Mail Online*. Recuperado el 09 de junio de 2022, de <https://www.dailymail.co.uk/femail/article-5567457/The-dark-sex-robots-discovered-eye-opening-TV-documentary.html>.
- DANAHER, John. «Should We Be Thinking About Robot Sex?», En John Danaher y Neil McArthur (Ed.), *Robot Sex. Social and Ethical Implications*, 3-14. Cambridge: MIT Press, 2018.
- DANAHER, John, Brian Earp y Anders Sandberg. «Should We Campaign Against Sex Robots?», En John Danaher y Neil McArthur (Ed.), *Robot Sex. Social and Ethical Implications*, 47-72. Cambridge: MIT Press, 2018.

- DAMASIO, Antonio. *Descartes' Error. Emotion, Reason, and the Human Brain*. Nueva York: Avon Books, 1994.
- DAMASIO, Antonio. *En busca de Spinoza. Neurobiología de la emoción y los sentimientos*. Trad. Joandomènec Ros. Barcelona: Planeta, 2005.
- CRAWFORD, Kate. *Atlas of AI*. Cornwall: Yale University Press, 2021. <https://doi.org/10.1080/15295036.2021.1935758>.
- DARLING, Kate. *The New Breed. How to Think About Robots*. Londres: Allen Lane, 2021.
- DEVLIN, Kate. *Turned On. Science, Sex, and Robots*. Londres: Bloomsbury, 2018. <https://doi.org/10.5040/9781472950888>.
- DICK, Philip. *Do Androids Dream of Electric Sheep*. Auckland: SF Masterworks, 2010.
- DICK, Philip. *We Can Build You*. Nueva York: First Mariner Books, 1972.
- D'OLIMPIO, Laura «The Trolley Dilemma: Would you Kill a Person to Save Five?». *The Conversation*, Melbourne. Recuperado el 22 de abril de 2022, de <https://theconversation.com/the-trolley-dilemma-would-you-kill-one-person-to-save-five-57111>.
- EKMAN, Paul, Wallace V. Friesen y Phoebe Ellsworth. *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Nueva York: Pergamon Press Inc., 1972.
- FORBES. «Piden prohibir la venta de muñecas sexuales con aspecto infantil en EU». *Forbes México*. Recuperado el 09 de junio de 2022, de <https://www.forbes.com.mx/piden-prohibir-la-venta-de-munecas-sexuales-con-aspecto-infantil-en-eu/>.
- FONG, Mei. «Sex Dolls Are Replacing China's Missing Women». *Yahoo Finance*. Recuperado el 09 de junio de 2022, de <https://uk.finance.yahoo.com/news/sex-dolls-replacing-china-missing-200634845.html>.
- FONG, Terrence, Illah Nourbakhsh y Kerstin Dautenhahn. «A Survey of Socially Interactive Robots», *Robotics and Autonomous Systems* 42 (2003): 143-166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)
- FOOT, Philippa. «The Problem of Abortion and the Doctrine of the Double Effect». *Oxford Review* 5 (1967): 5-15. Recuperado de <https://philpapers.org/archive/FOOTPO-2.pdf>.
- GUNKEL, David J. *The Machine Question. Critical Perspectives on AI, Robots, and Ethics*. Massachusetts: The MIT Press, 2017. <https://doi.org/10.5840/ipq201353110>.
- HANSON, F. Allan. «Which Came First, the Doer or the Deed?». En Peter Kroes y Peter-Paul Verbeek (Ed.), *The Moral Status of Technical Artefacts*, 55-73. Nueva York: Springer, 2014.
- HAQUE, Enamul. *The Ultimate Guide to Artificial Intelligence*. Londres: Amazon Inc., 2020.
- HERREROS Ubalde, Pablo. *La inteligencia emocional de los animales*. Barcelona: Ediciones Destino, 2018.
- HOCKENHULL, Michael y Marisa Leavitt Cohn. «Hot Air and Corporate Sociotechnical Imagineries: Performing and Translating Digital Futures in the Da-

- nish Tech Scene», *New Media and Society* 23, 2 (2021): 302-321. <https://doi.org/10.1177/1461444820929319>.
- HUNG Lillian. *et al.* «The benefits of and barriers to using a social robot PARO in care settings: a scoping review». *BMC Geriatr* 19, 232 (2019): 1-10 <https://doi.org/10.1186/s12877-019-1244-6>.
- JASANOFF, Sheila. «Future Imperfect: Science, Technology, and the Imaginations of Modernity». En Sheila Jasanoff y Sang-Hyun Kim (Ed.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, 1-33. Chicago: University of Chicago Press, 2015a.
- JASANOFF, Sheila. «Imagined and Invented Worlds», En Sheila JASANOFF y Sang-Hyun KIM (Ed.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, 321-341. Chicago: University of Chicago Press, 2015b.
- JASANOFF, Sheila y Hilton. R. Simmet. «Renewing the future: Excluded Imaginaries in the global energy transition», *Energy Research & Social Science* 80, número de artículo 102205 (2021). <https://doi.org/10.1016/j.erss.2021.102205>.
- JONCE, Spike. director. 2014. *Her*, Vértigo Films. 2:00:00. https://www.primevideo.com/detail/0FWQIRWMNPJ5J2PZR5HFZIKWGK/ref=atv_sr_def_c_unk_1_1_1?sr=11&pageTypeIdSource=ASIN&pageTypeId=B08K3VS7DB&qid=1654838042
- KROES, Peter y Peter-Paul Verbeek. «Introduction: The Moral Status of Technical Artefacts». En Peter Kroes y Peter-Paul Verbeek (Ed.) *The Moral Status of Technical Artefacts*, 1-9. Nueva York: Springer, 2014.
- JONES, Raya A. «What makes a robot “social”?». *Social studies of science* 47, 4 (2017): 556-57. <https://doi.org/10.1177/0306312717704722>.
- KLÆRKE-OLSEN, Anne *et al.* *Imagining Life with AI Robots* (Tesis de maestría). Universidad de Aalborg, Dinamarca, 2018. Recuperado el 22 de abril de 2021, de https://projekter.aau.dk/projekter/files/280914176/Imagining_Life_with_AI_Robots_Final.pdf.
- KURZWEIL, Ray. *The Singularity is Near*, Londres: Duckworth Publishers, 2006.
- LANG, Fritz, director. 1927. *Metropolis*, UFA. 02:30:00. https://www.primevideo.com/detail/0HW43DHUHDKXC78Y5RHJYZSJR0/ref=atv_sr_def_c_unk_1_1_1?sr=11&pageTypeIdSource=ASIN&pageTypeId=B0B2226TW8&qid=1654837015
- LARSON, Erik. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*, Cambridge: Harvard University Press, 2021.
- LAITINEN, Arto, Marketta Niemelä y Jari Pirhonen. «Demands of Dignity in Robotic Care: Recognizing Vulnerability, Agency, and Subjectivity in Robot-based, Robot-assisted, and Teleoperated Elderly Care». *Techné: Research in Philosophy and Technology* 23, 3 (2019): 366-401. <https://doi.org/10.5840/techné20191127108>.
- LEVY, David. *Love and Sex with Robots*, Wiltshire: Cromwell Press Ltd, 2008.
- LIBIN, Alexander y Elena Libin. «Person-Robot Interactions from the Robopsychologist's Point of View: The Robotic Psychology and Rotherapy

- Approach». *Proceedings of the IEEE* 92, 11 (2004): 1789-1803. <https://doi.org/10.1109/JPROC.2004.835366>.
- LÓPEZ de Mantarás Badía, Ramón y Pedro Meseguer González. *¿Qué sabemos de? Inteligencia artificial*, Madrid: CSIC, 2017.
- LIZANA, José Alberto «Tras charlar con una Inteligencia artificial, ahora este ingeniero de Google piensa que tienen plena conciencia» *Genbeta*. Recuperado el 14 de junio de 2022 de <https://www.genbeta.com/actualidad/charlar-inteligencia-artificial-ahora-este-ingeniero-google-piensa-que-tienen-plena-conciencia>.
- MEJÍA Fernández, Ricardo. *Pensar la inteligencia artificial. El legado de Hubert Dreyfus a la fenomenología*. Madrid: McGraw-Hill, 2020.
- MCARTHUR, Neil. «The Case for Sex Robots», En John Danaher y Neil McArthur (Ed.) *Robot Sex. Social and Ethical Implications*, 31-47. Cambridge: MIT Press, 2018.
- MCARTHUR, John *et al.* «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955». En *AI Magazine* 27, 4 (2006): 12-14.
- MITCHAM, Carl «Agency in Humans and in Artifacts: A Contested Disclosure». En Peter Kroes y Peter-Paul Verbeek (Ed.). *The Moral Status of Technical Artefacts*, 11-29. Nueva York: Springer, 2014.
- MEACHAM, Darian y Matthew Studley. «Could a Robot Care? It's All in the Movement». En Patrick Lin, Ryan Jenkins y Keith Abney (Ed.). *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*, 97-112. Nueva York: Oxford University Press, 2017.
- MINSKY, Marvin. *The Emotion Machine. Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Nueva York: Simon & Schuster Paperbacks, 2006.
- NASS, Clifford y Youngme Moon. «Machines and Mindlessness: Social Responses to Computers», *Journal of Social Issues* 56, 1 (2000): 81-103. <https://doi.org/10.1111/0022-4537.00153>.
- NISHIO, Shuichi, Koichi Taura y Hiroshi Ishiguro. «Regulating Emotion by Facial Feedback from Teleoperated Android Robot», En Shuzhi Sam Ge *et al* (Ed.). *Social Robotics*, 388-397. Berlín: Springer, 2012.
- NOLAN, Jonathan, director. 2016. *Westworld*, Warner Bros. Blu-Ray.
- NYHOLM, Sven y Lily Eva Frank. «From Sex Robots to Love Robots: Is Mutual Love with a Robot Possible?». En John Danaher y Neil McArthur (Ed.). *Robot Sex. Social and Ethical Implications*, 219-244. Cambridge: MIT Press, 2018.
- ORTONY, Andrew, Gerald Clore y Allan Collins. *The Cognitive Structure of Emotions*. Nueva York: Cambridge University Press, 1988.
- PARETO Boada, Júlia, Begoña Román Maestre y Carme Torrá Genís. «The Ethical Issues of Social Assistive Robotics: A Critical Literature Review». *Technology in Society* 67, número de artículo 101726. (2021). <https://doi.org/10.1016/j.techsoc.2021.101726>.
- PICARD, Rosalind. *Affective Computing*. Massachusetts: MIT Press, 2000.
- PINEDA Oliva, David. *Sobre las emociones*. Madrid: Cátedra, 2019.

- RAHWAN, Iyad, Kerstin Skork y Elena Hungerland. «Sacrifice One Person in order to Save Five?» *Max-Planck-Gesellschaft*. Recuperado el 20 de abril de 2022 de <https://www.mpg.de/14386104/trolley-dilemma-interntional>.
- REEVES, Byron y Clifford Nass. *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. Stanford: CSLI, 2002.
- RHEE, Jennifer. *The Robotic Imaginary. The Human & The Price of Dehumanized Labor*. Minneapolis: University of Minnesota Press, 2018. <https://doi.org/10.5749/j.ctv62hh4x>.
- ROLLS, Edmund T. *The Brain, Emotion and Depression*, Nueva York: Oxford University Press, 2018.
- RUSSELL, Stuart. *Human Compatible. AI and the Problem of Control*. Londres: Penguin, 2019.
- RUSSELL, Stuart y Peter Norvig. *Artificial Intelligence: A Modern Approach*. Uttar Pradesh: Pearson, 2019.
- ŠABANOVIĆ, Selma, et al. «PARO Robot Affects Diverse Interaction Modalities in Group Sensory Therapy for Older Adults with Dementia». *Proceedings of IEEE International Conference on Rehabilitation Robotics*, número de artículo 6650427. (2013). <https://doi.org/10.1109/ICORR.2013.6650427>.
- SÁNCHEZ-ESCRIBANO, María Guadalupe. *Engineering Computational Emotion-A Reference Model for Emotion in Artificial Systems*. Nueva York: Springer, 2018.
- SCARANTINO, Andrea. «The Philosophy of Emotions». En Lisa Feldman Barret, Michael Lewis y Jeanette M. Haviland-Jones (Ed.) *Handbook of Emotions*, 3-48. Nueva York: Guilford Press, 2018.
- SCOTT, Ridley, director. 1982. *Blade Runner*, Warner Bros. 01:56:00 . https://www.primevideo.com/detail/0NUACWRNIVONA1ZKTBOBTRBS5M/ref=atv_sr_def_c_unkc_1_1_1?sr=11&pageTypeIdSource=ASIN&pageTypeId=B07S2P6YWH&qid=1654838629
- SEARLE, John. *The Rediscovery of the Mind*, Massachusetts: The MIT Press, 2002.
- SPIELBERG, Steven, director. 2001. *AI. Artificial Intelligence*, Warner Bros. 02:25:00. https://www.primevideo.com/detail/OR5MZEMTYAL7D37GHK2A0XNAEK/ref=atv_sr_def_c_unkc_1_1_1?sr=11&pageTypeIdSource=ASIN&pageTypeId=B07RF7QJ5Z&qid=1654837202.
- STRONGMAN, Kenneth T. *The Psychology of Emotion, From Everyday Life to Theory*, Norfolk: Wiley, 2006.
- SULLINS, John. «Robots, Love, and Sex: The Ethics of Building a Love Machine». *IEEE Transactions on Affective Computing* 3, 4 (2012): 398-409. <https://doi.org/10.1109/T-AFFC.2012.31>
- SUTHERLAND, Stuart. *Irracionalidad: El enemigo interior*. Madrid: Alianza, 2015.
- TURING, Alan. «Computing Machinery and Intelligence». *Mind* 59, 236 (1950): 433-460.
- SZCZUKA, Jessica y Nicole Krämer. «Not Only the Lonely –How Men Explicitly and Implicitly Evaluate the Attractiveness of Sex Robots in Comparison to the Attractiveness of Women, and Personal Characteristics Influencing This Evaluation»,

- Multimodal Technologies and Interaction* 1 (1), 3 (2017). <https://doi.org/10.3390/mti1010003>.
- TEGMARK, Max. *AI Compatible. Being Human in the Age of Artificial Intelligence*. Londres: Penguin, 2017.
- TURNER, Jacob. *Robot Rules. Regulating Artificial Intelligence*. Londres: Palgrave Macmillan, 2019.
- VALLVERDÚ, Jordi *et al.* «Fake Empathy and Human Robot Interaction (HRI): A Preliminary Study». *International Journal of Technology and Human Interaction* 12, 1 (2018): 44-59. <https://doi.org/10.4018/IJTHI.2018010103>.
- VAN Noort, Carolijn. «On the Use of Pride, Hope and Fear in China's International Artificial Intelligence Narratives on CGTN». *AI & Society* (2022): 1-13. <https://doi.org/10.1007/s00146-022-01393-3>.

