

**D3US EX M4CH1NA.**

**Arte e Inteligencia Artificial**

26 noviembre 2019 - 16 mayo, 2020

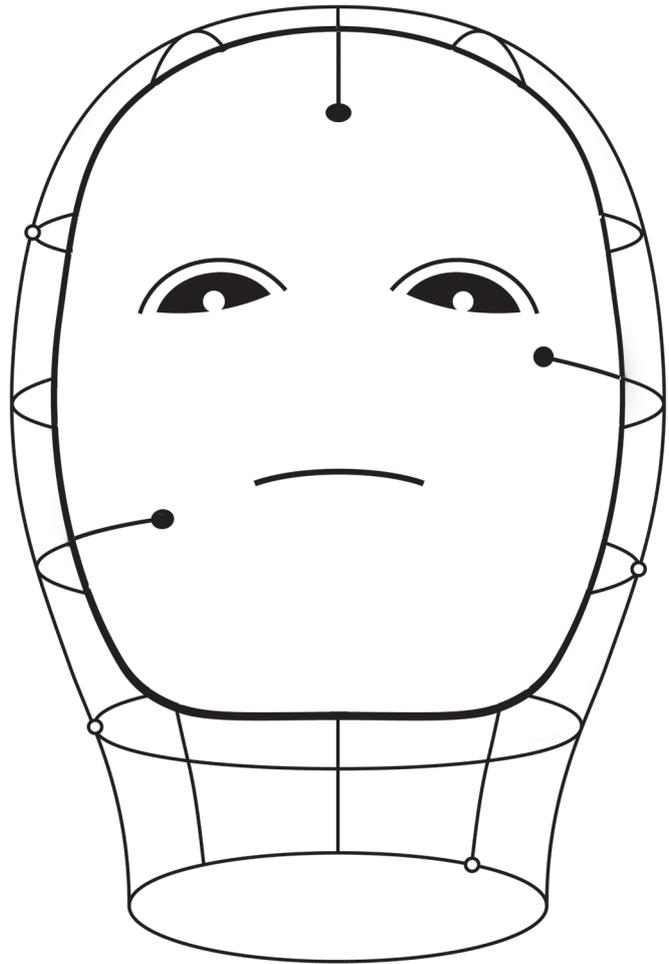
Comisarios:

Karin Ohlenschläger y Pau Waelder

Artistas: Memo Akten, Harold Cohen,  
Jake Elwes, Lynn Hershman Leeson,  
Felix Luque, Lauren McCarthy,  
Anna Ridler, Guido Segni, Caroline  
Sinders, Christa Sommerer & Laurent  
Mignonneau, Jenna Sutela, Patrick  
Tresset, Pinar Yoldas

Diseño: Juan Jareño

Texto: Pau Waelder



**D3US EX  
M4CH1NA**

**ARTE E INTELIGENCIA ARTIFICIAL**

# D3US EX M4CH1NA

## Ante los dioses algorítmicos

Medea, la conocida tragedia de Eurípides, concluye con la protagonista evitando ser ajusticiada por sus múltiples asesinatos gracias al dios Helios, quien le presta su carro para huir de la corte de Corinto. En el escenario, una intervención divina de este tipo se representaba con un actor apareciendo desde lo alto, subido en una grúa u otro artificio mecánico. Las acciones de este personaje daban un giro satisfactorio a la historia, sin importar su coherencia con las situaciones en que ésta se había desarrollado. El filósofo Aristóteles criticó este recurso, afirmando que “el desenlace de la obra debe resultar de la pieza misma, y no, como en la Medea, de una intervención extraña”.<sup>1</sup> La práctica de sacar a un actor subido a una grúa para crear un desenlace rápido de la trama de una obra de teatro se conocería más adelante con la expresión latina *deus ex machina* (“dios desde la máquina”). Esta expresión se utiliza actualmente para referirse a cualquier elemento que se introduce en una historia sin que tenga relación alguna con los hechos que se han narrado y cambia totalmente su desenlace, a menudo para satisfacer las expectativas del público, pero dando como resultado un relato inverosímil. Como recurso narrativo, resulta excesivamente fácil, y por ello la expresión suele tener un carácter peyorativo.<sup>2</sup> Con todo, un aspecto interesante del *deus ex machina* es que introduce aquello que uno desea que ocurra, aunque se oponga a la realidad de los hechos que se narran. La historia da un súbito giro hacia el final feliz, la salvación en el momento crítico, la reconciliación que parecía imposible. Todo ocurre de forma rápida y expeditiva, sin que haya lugar a cuestionar el porqué de este desenlace, dado que no se ponen en duda las acciones de un dios o las del destino.

La inteligencia artificial (IA) está acaparando actualmente el discurso mediático sobre nuestra relación con la tecnología como un *deus ex machina*: una entidad omnisciente y todopoderosa que promete resolver fácilmente todos nuestros problemas. En los últimos años, las grandes empresas tecnológicas han entrado en una competición para liderar el desarrollo de la IA, que se ha traducido en espectaculares aplicaciones de esta tecnología a soluciones concretas, pero también en ambiciosas afirmaciones acerca de lo que puede conseguir en un futuro próximo. En 2016, el ex-director ejecutivo de Google, Eric Schmidt, afirmaba que la inteligencia artificial iba a resolver el cambio climático, la pobreza, la guerra y el cáncer.<sup>3</sup> Mike Schroepfer, director de tecnología en Facebook, hablaba al mismo tiempo de cómo la IA supondría una transformación a nivel global. No obstante, estas promesas parecen estar lejos de cumplirse, incluso en ámbitos más concretos (aunque también muy ambiciosos) como el de los coches inteligentes. En 2017, John Krafcik, director ejecutivo de Waymo, aseguró que la empresa estaba a punto de fabricar coches sin conductor, pero no ha logrado hacer realidad este propósito.<sup>4</sup> Con todo, lanzar predicciones atrevidas y excesivamente optimistas no es algo que hayan empezado a hacer los directivos de las corporaciones en estos últimos años, si

1 Aristóteles, *Poética*, XV. Traducción de José Alsina Clota. Barcelona: Icaria, 1987, p.47.

2 Robert McKee califica el *deus ex machina* como “el peor pecado de cualquier guionista” y, más aún, “un insulto al público”.

Robert McKee, *El Guión. Sustancia, estructura, estilo y principios de la escritura de guiones*. Trad. Jessica Lockhart. Barcelona: Alba Editorial, 2011.

3 Gary Marcus y Ernest Davis, *Rebooting AI. Building Artificial Intelligence We Can Trust*. Nueva York: Pantheon Books, 2019, p.11

4 Marcus y Davis, *Rebooting AI*, p.10

no que está estrechamente vinculado a la propia historia de la inteligencia artificial. En 1967, Marvin Minsky, uno de los “padres” de la IA, afirmó que “en el plazo de una generación, el problema de la inteligencia artificial estará básicamente resuelto”.<sup>5</sup> Desde entonces, se han sucedido las apuestas acerca del momento en que la inteligencia de los ordenadores superará a la de los humanos, un punto de inflexión que el futurista Ray Kurzweil ha situado en el año 2029.<sup>6</sup> La idea de una máquina inteligente sin duda suscita grandes esperanzas y también temores apocalípticos, puesto que las máquinas, como meras herramientas “tontas”, han aportado a la humanidad un progreso inimaginable y también una terrible devastación. ¿Qué podemos esperar de un ordenador miles de veces más inteligente que los seres humanos, y con conciencia propia? ¿Acaso es eso posible? Si lo es, ¿vendrá a salvarnos o a destruirnos? Lo primero que nos viene a la mente son robots asesinos como de los films *Blade Runner* (Ridley Scott, 1982), *The Terminator* (James Cameron, 1984) o *Ex Machina* (Alex Garland, 2014), así como las inestables personalidades de Hal 9000 en *2001: Odisea en el espacio* (Stanley Kubrick, 1968), *Wintermute* en la novela *Neuromante* (1984) de William Gibson, o *Samantha* en el film *Her* (Spike Jonze, 2013). Estas historias plantean escenarios fascinantes y también aterradores, pero la realidad es que actualmente los robots están lejos de lograr la autonomía y la capacidad de actuar como humanos que tienen los replicantes o el T-800, mientras que los programas de inteligencia artificial pueden ganar a un contrincante humano en una partida de ajedrez o go, y también reservar mesa en un restaurante por teléfono, pero no son capaces de entender un chiste o seguir una conversación compleja. La distancia que separa lo que esperamos de la inteligencia artificial de lo que ésta es capaz de producir actualmente es enorme. Los investigadores Gary Marcus y Ernest Davis la denominan “el abismo de la IA”<sup>7</sup> y argumentan que se debe principalmente a tres factores: en primer lugar, tenemos tendencia a “humanizar” a las máquinas y dotarlas de voluntad propia e incluso de personalidad; en segundo lugar, el progreso de la IA en tareas muy específicas y acotadas (como ganar juego de ajedrez) se asume como un progreso a un nivel mucho más amplio; finalmente, cuando se logra desarrollar un sistema que funciona en ciertas situaciones (como lograr que un coche conduzca solo por el desierto), se asume que también lo hará en muchas otras, si bien esto no siempre es cierto.

En suma, el desarrollo de la inteligencia artificial se ve rodeado de historias, algunas de las cuales describen innovaciones reales, mientras otras reflejan las audaces expectativas de los investigadores, las ambiciones de las corporaciones, o nuestras propias esperanzas y temores ante una máquina que no entendemos del todo. La manera en que los programas de inteligencia artificial, procesando millones de datos en milésimas de segundo, atienden a nuestras peticiones o incluso adivinan nuestros deseos parece magia, y de hecho así es como, a menudo, se han presentado las innovaciones tecnológicas a lo largo de los siglos. La mitología griega contiene numerosas narraciones de autómatas y objetos animados, creados por el dios Hefesto, que cobraban vida de forma misteriosa.<sup>8</sup> Hoy en día, no concebimos un smartphone como una creación divina, pero igualmente se nos aparece como una caja negra cuyo funcionamiento interno resulta, hasta cierto punto, desconocido. Cuando esta caja negra, además de hacer todo lo que le pedimos, es capaz de predecir lo que vamos a querer o hacer, no es difícil adjudicarle poderes sobrenaturales o dejar que la imaginación

<sup>5</sup> Marcus y Davis, *Rebooting AI*, p.8

<sup>6</sup> Ray Kurzweil, *The Age Of Spiritual Machines. When Computers Exceed Human Intelligence*. Nueva York: Penguin Books, 1999, p.161.

<sup>7</sup> Marcus y Davis, *Rebooting AI*, p.27-31

<sup>8</sup> Adrienne Mayor, *Gods and Robots. Myths, Machines and Ancient Dreams of Technology*. Princeton-Oxford: Princeton University Press, 2018, p.164-191.

le asigne habilidades que van más allá de lo que realmente puede hacer. Según apuntan diversos investigadores,<sup>9</sup> un aspecto esencial de la IA es su capacidad de predicción: si bien no tenemos (aún) máquinas que puedan pensar, sí tenemos máquinas capaces de hacer predicciones a partir de los datos de que disponen, reconocer patrones, objetos u otros elementos y construir sus propios modelos.<sup>10</sup> Esto, es ciertamente, un avance espectacular que está revolucionando la industria tecnológica y empieza a tener profundas consecuencias en las sociedades más industrializadas. Por ello, es preciso observar la inteligencia artificial como algo más que una caja negra que “emite veredictos de los dioses algorítmicos,” en palabras de la matemática Cathy O’Neil.<sup>11</sup> Debemos entender en qué consiste la inteligencia artificial, cuáles son sus objetivos y cuáles sus logros reales.

## ¿Qué es la inteligencia artificial?

El término “inteligencia artificial” fue acuñado por el científico de la computación John McCarthy en 1956 en el contexto de la Conferencia de Dartmouth, el primer simposio académico en el que se debatía la posibilidad de hacer que una máquina sea capaz de razonar como un ser humano. Las implicaciones de este ambicioso objetivo abarcan diversas disciplinas, tales como las matemáticas, filosofía, neurociencia, psicología y biología. Todas ellas se verán implicadas en el desarrollo de la IA, pero son las ciencias de la computación las que asumen el proyecto de crear una máquina inteligente como un campo de investigación. La máquina en cuestión es, lógicamente, un ordenador (concretamente una máquina virtual, es decir un programa) y los métodos empleados para lograr algo parecido a la inteligencia humana son el procesamiento de datos, el cálculo matemático, el análisis estadístico y la elaboración de complejos algoritmos. Estos métodos conllevan traducir los diferentes aspectos del raciocinio a modelos que puedan ser procesados por un ordenador. La propuesta inicial de Dartmouth deja claro este objetivo:

“El estudio debe proceder sobre la base de la conjetura de que cada aspecto del aprendizaje o de cualquier otra característica de la inteligencia puede, en principio, describirse con tanta precisión que una máquina pueda simularlo”<sup>12</sup>

Este planteamiento encuentra un notable precedente en el artículo que, en 1950, el matemático Alan Turing dedicó a la cuestión “¿pueden pensar las máquinas?” y en el que plantea un “juego de imitación”, posteriormente conocido como Test de Turing.<sup>13</sup> El juego consiste en que una persona debe averiguar (por medio de preguntas a las que recibe respuestas en forma escrita) si está conversando con un ser humano o una máquina. Turing por tanto desplaza la cuestión de si una máquina puede pensar por la de si es posible desarrollar un ordenador capaz de imitar la inteligencia humana. Iniciando lo que será luego una costumbre entre los investigadores de inteligencia artificial, Turing vaticina que, en unos 50 años, habrán ordenadores capaces de

9 En los análisis de la percepción que se tiene de la inteligencia artificial, es común hacer referencia a lo mágico o lo divino, como hacen, entre otros, Pedro Domingos, Adrienne Mayor, Cathy O’Neil, o Meredith Broussard (ver nota 8 y ss.).

10 Meredith Broussard, *Artificial Unintelligence. How Computers Misunderstand the World*. Cambridge-Londres: The MIT Press, 2018, p.41 y ss.

11 Cathy O’Neil, *Weapons of Math Destruction. How Big Data Increases Inequality And Threatens Democracy*. Nueva York: Crown Publishing, 2016, p.13.

12 James Moor, “The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years.” *AI Magazine*, Vol.27, Núm.4. American Association for Artificial Intelligence, 2006, p.87.

13 Alan M. Turing, “Computing Machinery and Intelligence.” *Mind*, Nueva serie, Vol. 59, No. 236 (Octubre 1950), pp. 433-460. Publicado por Oxford University Press. <http://www.jstor.org/stable/2251299>

engañar a un ser humano y en general se podrá hablar de máquinas pensantes.<sup>14</sup> El matemático se anticipa también a las numerosas objeciones a las que se enfrenta aún hoy en día la IA y propone que se haga competir a las máquinas con los humanos en “ámbitos puramente intelectuales,” como el ajedrez.<sup>15</sup>

Otra aportación de Turing es la de proponer que, en vez de simular una mente adulta por medio de un programa, se simule la de un niño. Esto implica no sólo programar una máquina más sencilla, sino que también introduce el concepto del aprendizaje, que se podría realizar por medio de refuerzos positivos o negativos. Esta idea se ha implementado en la rama de la IA conocida como aprendizaje automático (machine learning), que consiste en crear programas capaces de modificar su funcionamiento en base a los datos que le son suministrados y los resultados obtenidos.<sup>16</sup> A diferencia de un programa habitual, que siempre ejecuta las mismas instrucciones, un programa de aprendizaje puede modificar los parámetros del modelo que le ha sido dado para adaptarse a unos objetivos concretos, como puede ser el de detectar los rostros de las personas en un conjunto de fotografías o una captura de vídeo. Como sugería Turing, estos programas pueden ser “educados” por medio de técnicas de aprendizaje. En el aprendizaje automático se emplean esencialmente tres: el aprendizaje supervisado, en el que se indica al programa qué resultado se desea obtener y se le proporciona un feedback indicando si lo ha logrado o no; el aprendizaje no-supervisado, en el que se suministran los datos al programa y éste descubre recurrencias y patrones por sí mismo; y finalmente, el aprendizaje por refuerzo, en el que se introduce un sistema de mensajes que indican al programa si ha logrado o no su objetivo.<sup>17</sup>

Entre las técnicas empleadas en aprendizaje automático, una de las más comunes es la de las redes neuronales artificiales (artificial neural networks, o ANN), un modelo computacional inspirado en la estructura del cerebro que se compone de miles (o millones) de unidades conectadas entre sí, cada una de las cuales procesa los datos que recibe y suministra un resultado, que sirve de input a otras unidades. Estas unidades de procesamiento, o neuronas, se estructuran en capas en un complejo sistema que somete los datos introducidos a numerosas operaciones matemáticas y genera un output que el propio programa puede alterar en función del feedback recibido. Gracias a la estructura de red, se pueden activar determinadas neuronas o potenciar las conexiones entre ellas para lograr, progresivamente, un resultado que se ajuste al objetivo deseado. A fin de poder clasificar la información que está procesando, la red neuronal artificial cuenta con un “conjunto de entrenamiento,” que son habitualmente datos seleccionados y etiquetados de manera acorde a la tarea a realizar. Así, por ejemplo, para “entrenar” a una red neuronal artificial a reconocer rostros humanos, se le suministran miles de retratos de personas. Entre las redes neuronales artificiales cabe destacar un sistema conocido como red generativa antagónica (generative adversarial network, o GAN), que se enmarca en el aprendizaje automático no-supervisado. Una GAN consiste en dos redes neuronales artificiales, una dedicada a generar outputs (el “generador”) y otra dedicada a determinar si dichos outputs se corresponden con un objetivo preestablecido (el “discriminador”). Habitualmente, las GAN se usan para generar imágenes que parecen reales, partiendo de un conjunto de entrenamiento formado por miles de fotografías del tipo que se quiere crear. Usando esta información, el generador crea una nueva imagen y el discriminador determina si dicha imagen es válida, es decir si puede pasar por una de las imágenes del

---

14 Turing, “Computing Machinery and Intelligence”, p.442.

15 Turing, “Computing Machinery and Intelligence”, p.460.

16 Ethem Alpaydin, Machine Learning. The New AI. Cambridge – Londres: The MIT Press, 2016, p.25.

17 Margaret A. Boden, AI. Its Nature and Future. Oxford: Oxford University Press, 2016, p.47-48.

conjunto de entrenamiento.<sup>18</sup> Si la imagen es rechazada, el generador modifica sus parámetros y vuelve a repetir el proceso, que continúa así en un bucle de retroalimentación. Las imágenes que aprueba el discriminador son suministradas a la persona que emplea la GAN. Ésta tiene a su vez la posibilidad de facilitar un feedback o ajustar el sistema para obtener mejores resultados. Las GAN han logrado una enorme popularidad desde que fueron inventadas por Ian Goodfellow y su equipo en 2014 debido a su capacidad para generar imágenes que engañan al ojo humano, si bien también encuentran una amplia variedad de usos vinculados al análisis y procesamiento de imágenes. Hoy en día, un número creciente de artistas emplean GANs en la creación de proyectos artísticos, en parte debido a que algunos de estos programas son accesibles en repositorios de software como Github, y también porque introducen la posibilidad de hacer de la máquina un co-autor de la obra: el software participa en la creación de la pieza con un cierto nivel de autonomía, en una especie de diálogo con la propia artista. Esta autonomía se hace patente en el hecho de que es muy difícil comprender exactamente lo que ocurre dentro de una ANN, por lo cual su funcionamiento resulta misterioso y casi parece mágico.<sup>19</sup>

Las aplicaciones de las redes neuronales artificiales y otros sistemas de inteligencia artificial han demostrado ser exitosas en la resolución de tareas específicas, en las que estos programas son capaces incluso de superar a los humanos (ya sean expertos en radiología o campeones mundiales de Go o ajedrez). Con todo, lo que no se ha conseguido lograr aún es una inteligencia artificial general, que pueda ser aplicada indistintamente a cualquier situación o entorno del mundo real. Esto es, evidentemente, mucho más complejo que lograr que un programa genere fotografías inexistentes o sepa reconocer el rostro de una persona. Incluso automatizar una tarea concreta, pero en un entorno cambiante, cuyas condiciones no pueden preverse (conducir un coche, por ejemplo), está resultando ser mucho más difícil de lo que parecía. Por tanto, lo que se ha desarrollado de momento es lo que se conoce como una "IA débil," limitada a una tarea específica, frente a lo que sería una "IA fuerte," que teóricamente tendría conciencia propia y sería el tipo de inteligencia artificial que, según la ciencia ficción y las predicciones de algunos investigadores, conseguirá dominar el mundo y, posiblemente, acabar con la raza humana. Para llegar a este punto, sería necesario lograr una inteligencia artificial general (artificial general intelligence, o AGI), que seguiría desarrollándose a sí misma hasta llegar al punto en que superaría la inteligencia humana, convirtiéndose en una superinteligencia artificial (artificial superintelligence, o ASI). Lo que esta ASI decidiera hacer con los humanos no queda claro, pero en cualquier caso se habría llegado a un punto del desarrollo tecnológico conocido como "la singularidad," en el cual las máquinas quedarían fuera de nuestro control.<sup>20</sup> Si dicha singularidad llegará alguna vez o no es uno de los múltiples debates entre los investigadores de IA, así como uno de los muchos mitos que rodean a la inteligencia artificial. Resulta fácil librarse a visiones apocalípticas como las que describe la trilogía *The Matrix* (Lilly y Lana Wachowsky, 1999-2003), pero mientras éstas no se materializan, es preciso fijarse en algunos retos y peligros actuales que plantea la IA hoy en día.

Los científicos reunidos en Dartmouth hace más de sesenta años coincidieron en que era preciso encontrar la manera de describir cualquier aspecto de la inteligencia con suficiente precisión como para que una máquina pudiera simularlo. Esto conlleva facilitar a la máquina una gran cantidad de datos y una serie de algoritmos por medio de los cuales procesarlos.

<sup>18</sup> Es interesante ver cómo este proceso se asemeja al Test de Turing, siendo el objetivo del generador "engañar" al discriminador.

<sup>19</sup> Boden, Al. *Its Nature and Future*, p.78.

<sup>20</sup> Boden, Al. *Its Nature and Future*, p.147-148.

Recopilar datos ha sido la prioridad de los investigadores, y posteriormente las grandes empresas, que se han implicado en el desarrollo de la inteligencia artificial. Para que una IA pueda redactar un texto como un humano, conversar con una persona o reconocer un rostro, hay que facilitarle miles de textos, grabaciones de audio y fotografías, que deben ser además auténticos, extraídos de situaciones y personas reales. Por tanto, empresas como Google, Apple, Facebook o Amazon recurren a los datos que generan sus usuarios, los contenidos de redes sociales y foros públicos, innumerables libros escaneados, millones de fotos y vídeos alojados en sus servidores, así como las grabaciones y otros datos obtenidos de los dispositivos que comercializan. Este enorme volumen de datos no existía hace unos años, y es uno de los factores (junto con el desarrollo del hardware) que ha facilitado la reciente explosión de la IA.<sup>21</sup> Una inteligencia artificial más sofisticada requiere por tanto que estemos dispuestos a ceder nuestros datos y renunciar a nuestra privacidad.

Pero esta no es la única renuncia a la que nos lleva el desarrollo actual de la IA. Una vez nuestros datos son facilitados al programa, éste los procesa empleando unos algoritmos que desconocemos y reflejan los prejuicios e intereses de quienes los han elaborado. Incluso si se ha programado con las mejores intenciones, un sistema de inteligencia artificial puede fallar en el análisis de los datos y las predicciones que extrae de ellos. Dado que su funcionamiento preciso es difícil de entender incluso para quien lo ha desarrollado, auditar uno de estos sistemas es extremadamente complejo, y queda fuera del alcance de las personas que se ven afectadas por sus decisiones. Además, a diferencia de las decisiones humanas, que son susceptibles de replantearse, los sistemas automatizados siguen aplicando las mismas instrucciones a menos que sean reprogramados.<sup>22</sup> Esta limitación de la IA débil suscita a su vez el temor a ciertos agentes que no son tan inteligentes como deberían y cuyas acciones pueden tener terribles consecuencias para la vida de las personas (por ejemplo, armas y vehículos autónomos o sistemas de vigilancia e identificación de personas empleados por la policía). La solución a una IA limitada pasa por facilitar más datos al sistema, pero esto lleva nuevamente a renunciar a la privacidad y a plantear la necesidad de un uso ético de la información recogida. Algunos defensores de la IA proponen por tanto la combinación entre máquinas y humanos, de manera que las primeras se centren en el análisis de grandes cantidades de datos y la resolución de tareas repetitivas, mientras los segundos tratan con situaciones más ambiguas o complejas, ejercitan un juicio basado en valores éticos y asumen responsabilidades por sus decisiones.<sup>23</sup>

Más allá de las cuestiones éticas, existen otros factores derivados de la propia materialidad de esta tecnología: un programa de inteligencia artificial requiere mucho tiempo de procesamiento, un hardware potente y el acceso a datos almacenados en la nube, todo lo cual se traduce en un considerable gasto energético. Entrenar a una IA con un gran volumen de datos genera unos 284.000 kg de dióxido de carbono, equivalente a las emisiones de cinco coches durante toda su vida útil.<sup>24</sup> En 2019, los primeros experimentos con el sistema BigGAN, que genera imágenes en alta resolución, han requerido el uso de 512 unidades de procesamiento TPU (desarrolladas por Google) que consumen, en la creación de cada imagen, tanta electricidad

---

21 Marcus y Davis, *Rebooting AI*, p.17

22 O'Neil, *Weapons of Math Destruction*, p.202.

23 Paul R. Daugherty y H. James Wilson, *Human + Machine: Reimagining Work in the Age of AI*. Boston: Harvard Business Publishing, 2018, p.186.

24 Karen Hao, "Training a single AI model can emit as much carbon as five cars in their lifetimes," *MIT Technology Review*, 6 de junio de 2019. <https://www.technologyreview.com/s/613630/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/>

como la que se gasta en una vivienda a lo largo de seis meses.<sup>25</sup> La huella de carbono es por tanto otro de los aspectos preocupantes de la IA (y más inmediato que los robots asesinos), que no sólo tiene consecuencias para el medio ambiente sino también para el propio acceso a la tecnología que hace posible una inteligencia artificial más avanzada. Sólo los investigadores vinculados a grandes empresas como Google pueden emplear sus amplios recursos de hardware y la financiación necesaria para costear el consumo energético derivado de su uso. Esto puede limitar de forma decisiva cómo se va a desarrollar la investigación en inteligencia artificial y qué objetivos se van a priorizar.<sup>26</sup>

## **Cuestionar la IA a través del arte**

Numerosos autores<sup>27</sup> coinciden en señalar que los retos y peligros que plantea la inteligencia artificial requieren adoptar una visión crítica y tomar conciencia tanto de sus posibilidades reales como de la necesidad de emplear esta tecnología de manera ética. La IA no es magia, ni tampoco es una fuerza neutral e inevitable, es un conjunto de algoritmos y técnicas de computación que deben ser observados y cuestionados para evitar que perpetúen prejuicios y desigualdades o generen otros daños colaterales. Esto habitualmente no está al alcance del gran público, puesto que las industrias tecnológicas sólo quieren mostrar las bondades de sus productos y buscan consumidores fieles y convencidos, no usuarios capaces de elaborar una reflexión crítica sobre la tecnología. Tan sólo las escandalosas noticias acerca de la violación de la privacidad y el uso poco ético de los datos de los usuarios han forzado a las grandes empresas a replantear (si bien de forma temporal) sus productos y políticas: Google tuvo que parar la comercialización de sus gafas de realidad aumentada Google Glass; Apple ha tenido que cerrar el programa de escuchas de las conversaciones de sus usuarios que había creado para mejorar el funcionamiento de Siri; el escándalo de Cambridge Analytica ha llevado al director ejecutivo de Facebook, Mark Zuckerberg, a farfullar unas débiles disculpas ante el Senado de los Estados Unidos; Amazon se ha enfrentado a numerosas críticas por la escucha de las conversaciones de los usuarios a través de los dispositivos equipados con su asistente virtual Alexa y por suministrar a los cuerpos policiales un sistema de reconocimiento facial que comete peligrosos errores. Con todo, estos llamativos casos distan de cambiar a largo plazo la percepción que los usuarios tienen de la tecnología: una elaborada disculpa y unas cuantas promesas parecen zanjar la cuestión. Lo que es necesario, por tanto, es crear un marco de reflexión acerca de la tecnología que nos permita aplicar en todo momento una mirada crítica. No para negar lo que nos aporta, sino para entender su funcionamiento y nuestra implicación en el mismo.

Roger F. Malina, astrónomo y profesor emérito de arte y tecnología en la Universidad de Texas, ha indicado en diversas ocasiones que arte, ciencia y tecnología pueden alimentarse mutuamente en cuanto los artistas abren nuevos caminos a la tecnología y el pensamiento científico, mientras que estas disciplinas facilitan nuevas formas de creatividad artística.<sup>28</sup>

---

25 Katharine Schwab, "A Google intern built the AI behind these shockingly good fake images," FastCompany.com, 20/4/2019. <https://www.fastcompany.com/90244767/see-the-shockingly-realistic-images-made-by-googles-new-ai>

26 Emma Strubell, Ananya Ganesh y Andrew McCallum, "Energy and Policy Considerations for Deep Learning in NLP," 57th Annual Meeting of the Association for Computational Linguistics (ACL), Julio, 2019. arXiv:1906.02243 [cs.CL]

27 Véase, por ejemplo: O'Neil, *Weapons of Math Destruction*, p.212-216; Marcus y Davis, *Rebooting AI*, p.32-34; Broussard, *Artificial Unintelligence*, p.240-241.

28 Pau Waelder, "Roger Malina: «el arte conduce a una nueva ciencia»," VIDA. Concurso Internacional Arte y Vida

El arte que se apropia de los avances científicos y tecnológicos crea un marco cultural que nos permite comprenderlos, experimentarlos y desarrollar una postura crítica, más realista y también más comprometida. La exposición D3US EX M4CH1NA. Arte e Inteligencia Artificial persigue este objetivo, al reunir el trabajo de catorce artistas, cuyas obras abordan las maneras en que percibimos la IA y qué retos nos plantea. Las piezas pueden experimentarse individualmente, como parte del discurso que ha elaborado cada artista a lo largo de su trayectoria, o bien considerarse como nodos de una red que construye un retrato de la inteligencia artificial a través de diferentes perspectivas y aproximaciones, algunas más lúdicas y otras más reflexivas, en ocasiones invitando al espectador a dialogar con la obra, y en otras excluyéndolo conscientemente. A continuación, planteamos un recorrido en el que las piezas se han agrupado en cinco grandes temas, que llevan progresivamente del humano a la máquina, del humilde robot aprendiz a la críptica superinteligencia.

## Arte artificial

Los inicios del arte algorítmico pueden situarse en una exposición de dibujos generados por ordenador que tuvo lugar en el Instituto de Tecnología de la Universidad de Stuttgart el cinco de febrero de 1965. El artista y matemático Georg Nees presentó su trabajo en una sala en la que el filósofo Max Bense daba sus clases. En la inauguración se dieron cita varios artistas que preguntaron a Nees por el proceso de creación de los dibujos. Según recuerda el artista y matemático Frieder Nake, uno de los pioneros del arte algorítmico, se produjo una conversación entre Nees y un artista en la cual éste último le preguntó si la máquina era capaz de dibujar como lo hacía él, a lo que el matemático respondió: “sí, ¡si puede describirme con exactitud cómo es su estilo!”.<sup>29</sup> Nees consideraba, de forma análoga a los científicos reunidos en Dartmouth, que era posible que una máquina creara arte si se lograba aportarle los datos suficientes para simular la creatividad humana. Esto no acabó de agrandar a los artistas, que se marcharon de la exposición enfadados, mientras Bense intentaba calmar los ánimos diciéndoles: “¡Caballeros, estamos hablando de arte artificial!” Con este término (que, según Nake, se inventó en ese momento) el filósofo intentaba diferenciar el arte creado por la máquina del que creaban los artistas.

Tres años más tarde, en 1968, el joven artista británico Harold Cohen iniciaba su estancia como profesor invitado en el departamento de Bellas Artes de la Universidad de California en San Diego (UCSD), con la intención de aprender a programar y aplicar la programación a su investigación artística. Mientras Georg Nees era un matemático que exploraba la creación artística, Harold Cohen era un artista que se adentraba en las posibilidades de la computación.<sup>30</sup> Ambos coincidían en que, para poder crear arte con un ordenador, era necesario describir con detalle, en términos simbólicos, como debería la máquina generar la obra. Con todo, como apunta Nake, dicha descripción no se limita a una única obra, sino que permite generar una serie infinita de obras.<sup>31</sup> Por tanto, la creación del artista es la propia descripción, traducida en código de programación y ejecutada por un programa que genera una composición visual

Artificial, 22 de diciembre de 2014. <https://vida.fundaciontelefonica.com/blog/roger-malina-el-arte-conduce-a-una-nueva-ciencia/>

<sup>29</sup> Frieder Nake, “Roots and randomness –a perspective on the beginnings of digital art”, en: Wolf Lieser (ed.), *The World of Digital Art*. Postdam: h.f. Ullmann, 2010, p.40.

<sup>30</sup> Frieder Nake, “Georg Nees & Harold Cohen: Re:tracing the origins of digital media,” en: Oliver Grau, Janina Hoth y Eveline Wandl-Vogt (eds.) *Digital Art through the Looking Glass. New strategies for archiving, collecting and preserving in digital humanities*. Donau: Donau-Universität, 2019, p.30.

<sup>31</sup> Nake, “Georg Nees & Harold Cohen”, p.39.

a través de un plóter. La principal creación de Cohen fue AARON, un programa de inteligencia artificial que empezó a elaborar en 1973 y en el que siguió trabajando hasta su muerte en 2016. El artista describe a AARON como “un programa informático diseñado para modelar algunos aspectos del comportamiento humano en la creación artística y producir como resultado dibujos «a mano alzada» de tipo altamente evocador”.<sup>32</sup> El programa era capaz de generar de forma autónoma una serie infinita de dibujos sin contar con ninguna información visual, tan sólo un conjunto de reglas basadas en la experiencia de Cohen con el proceso artístico. El objetivo era investigar acerca de la simulación de la cognición a través de la creación de obras de arte, partiendo de la pregunta: “cuál sería la mínima condición para que un conjunto de trazos constituya una imagen?” Este planteamiento dista mucho de la idea de sustituir al artista por una máquina, y de hecho Cohen insiste en que AARON no es un artista, ni tampoco una herramienta que permita a un usuario crear formas visuales: el programa generaba sus propios dibujos única y exclusivamente a partir de las reglas inscritas en el código de programación. Aún así, según afirma su creador, al presentar el trabajo de AARON el público interpretaba que los dibujos debían haber sido “suministrados” al programa por un artista, y posteriormente, al conocer su funcionamiento, le asignaban una personalidad propia. Esta anécdota ejemplifica lo que Marcus y Davis denominan “la brecha de la credulidad”<sup>33</sup>: no podemos evitar pensar en las máquinas en términos cognitivos, otorgándoles una mente y voluntad propias. Pero también es el efecto del complejo sistema de instrucciones diseñado por Cohen. El programa de AARON contaba originalmente (en 1979) con unas trescientas “producciones”, o instrucciones para ejecutar parte de una acción. Dichas instrucciones estaban repartidas en diversos niveles, con una estructura jerárquica en la que los niveles superiores limitaban el campo de acción en los inferiores. Así, el nivel más alto del sistema, denominado ARTWORK (obra), decidía la organización del dibujo en su totalidad, mientras que el nivel MAPPING (mapeado) se encargaba de asignar los espacios donde se colocaba cada elemento y PLANNING (planificación) gestionaba el desarrollo de cada figura. Esta estructura se extendía hasta el nivel de cada trazo, de manera que una única línea era ya el resultado de unas veinte o treinta producciones en tres niveles del sistema.<sup>34</sup>

Harold Cohen continuó desarrollando AARON a lo largo de más de tres décadas, incorporando nuevos elementos tales como el color y la combinación de los elementos creados por el programa y la intervención directa del artista. Su trabajo constituye una exploración pionera de la interacciones entre arte e inteligencia artificial, en el que vemos cómo los principios propuestos en Dartmouth se aplican a la creación artística, por medio de un sistema que presenta ciertas similitudes con la estructura de las redes neuronales artificiales. Como afirma Cohen, los dibujos de AARON (de los cuales presentamos una serie realizada para la galería Arnolfini, en Bristol, en 1983) son evocadores, puesto que resultan fascinantes tanto si se confunden con las obras de un artista humano (con lo cual el programa pasaría el Test de Turing), como cuando se comprende el complejo sistema por medio del cual son generados. Algo parecido sucede con los dibujos que Anna Ridler ha realizado para *The Fall of the House of Usher I y II* (2017). La artista se inspira en el cortometraje mudo que realizaron James Sibley Watson y Melville Webber en 1928, interpretando libremente el relato de Edgar Allan Poe, para elaborar una sutil reflexión acerca del proceso creativo como traducción o reinterpretación. La obra de Ridler consiste en una animación generada por una red generativa antagónica (GAN) y una serie de doscientos dibujos realizados por ella, que reproducen las principales

32 Harold Cohen, “What is an image?”, 1979. AARON's home. <http://www.aaronshome.com/aaron/publications/index.html>

33 Marcus y Davis, *Rebooting AI*, p.27.

34 Cohen, “What is an image?”

escenas del cortometraje. En lugar de crear una animación directamente a partir de los dibujos, la artista los emplea como el conjunto de entrenamiento que suministra a una red neuronal artificial, a fin de que ésta aprenda a generar nuevas imágenes. El producto de esta primera red neuronal es suministrado por Ridler a una segunda red neuronal, que a su vez genera nuevas imágenes, y finalmente a una tercera. La animación muestra simultáneamente la producción de las tres redes neuronales, permitiendo compararlas y dando lugar a una nueva versión del film que es tan narrativa como autorreferencial: la historia de los hermanos Roderick y Madeline Usher se desarrolla en un efecto de pantalla partida en el que la imagen se va deteriorando progresivamente, perdiendo detalle pero también librándose a formas cada vez más delirantes, en consonancia tanto con la progresiva pérdida de la cordura de los personajes creados por Poe como con el estilo surrealista que dieron Sibley Watson y Webber a su film experimental. Es posible, no obstante, identificar algunos elementos clave, tales como los primeros planos de los protagonistas, los escenarios, los juegos tipográficos que introducen los cineastas y finalmente la luna, que aparece entre las grietas de la casa de Usher cuando esta se desmorona, adquiriendo un notorio protagonismo tanto en el relato original como en el cortometraje. La banda sonora compuesta por Alec Wilder en 1959, que se ha añadido a la animación, contribuye a dar sentido a las imágenes y a situar la narración. Sin duda, el particular estilo visual del film de Sibley Watson y Webber, al que los dibujos de Ridler dan un aire aún más enigmático, se enriquece con la interpretación que hacen las redes neuronales artificiales. Las imágenes generadas por estos programas, incluso cuando se basan en imágenes reales, son siniestras, a la vez reconocibles y extrañas, como una figura de cera. La artista emplea conscientemente un proceso de aprendizaje automático recursivo para cuestionar las nociones de creatividad y originalidad, a la vez que pone de relieve el papel de las GANs y en particular los datos que se usan para entrenarlas.<sup>35</sup> En conjunto, la obra plantea un juego de duplicaciones y delirios que Poe ya introduce en su relato, Sibley Watson y Webber traducen en fascinantes metáforas visuales, y Anna Ridler interpreta en su propio estilo gráfico, librando luego el resultado a sucesivos programas de inteligencia artificial, en un bucle que podría no tener fin. La artista se pregunta finalmente cuál de todas estas versiones es la obra real, dónde está el arte, y de hecho muestra sus dibujos junto a la animación, no como una obra, sino como el conjunto de datos que se ha suministrado a la red neuronal artificial.

El uso de la técnica del dibujo por parte de Ridler responde a su interés por el dibujo como primer lenguaje, anterior al habla o la escritura, que en este caso sirve para crear imágenes con las que entrenar a la máquina. La artista indica que, si bien una GAN puede producir un dibujo, no puede dibujar.<sup>36</sup> Esto es cierto en cuanto la red neuronal artificial genera las imágenes de una forma diferente, pero una máquina también puede aprender a dibujar. AARON es un programa capaz de construir un dibujo tomando las mismas decisiones que una persona. De la misma manera, los robots que ha creado Patrick Tresset para su serie Human Studies (2011-2019) elaboran retratos con trazos impulsivos, pero a la vez precisos. Cada uno de los robots fabricados por Tresset se compone de un pupitre equipado con una cámara, situada sobre un soporte articulado, un brazo mecánico que sujeta un bolígrafo y un ordenador alojado bajo la superficie de la mesa. El cuerpo del robot, por tanto, es el mismo pupitre, del que surgen sus dos únicas extremidades, un ojo y un brazo. Lo único que puede hacer es dibujar, una vez se coloca sobre la mesa una hoja de papel, y lo hace (en palabras de Tresset) obsesivamente, alternando su mirada entre aquello que quiere reproducir y la hoja en la que va trazando

<sup>35</sup> Anna Ridler, "Fall of the House of Usher. Datasets and Decay". Victoria and Albert Museum, 17 de septiembre, 2018. <https://www.vam.ac.uk/blog/museum-life/guest-blog-post-fall-of-the-house-of-usher-datasets-and-decay>

<sup>36</sup> Ridler, "Fall of the House of Usher. Datasets and Decay"

líneas sinuosas. En esta serie, lo que reproducen los robots son retratos de personas que acceden a sentarse y servir de modelo en sesiones que pueden durar hasta treinta minutos. La persona se convierte así en objeto de estudio de la máquina, que en este caso no es una mera herramienta por medio de la cual obtener una reproducción fidedigna de uno mismo en unos pocos minutos. El artista podría haber creado unos robots capaces de reproducir los retratos de forma rápida y con calidad fotorrealista, como si fuese un fotomatón, pero ha optado por emplear cámaras viejas o de poca resolución y dotar a los robots de una cierta personalidad. Tresset introduce su propia técnica de dibujo en el programa que enseña a los robots a dibujar, pero modifica su comportamiento individual, para que cada uno genere dibujos diferentes. El resultado es que, en esta instalación performativa, la persona y la máquina intercambian sus papeles, siendo el robot quien asume el papel creativo y el humano un sujeto pasivo que debe someterse al escrutinio de la cámara, sin moverse, durante una sesión que se prolonga más allá de lo que resulta cómodo en nuestras impacientes interacciones con la tecnología. Los dibujos no son, por tanto, meras impresiones digitales que puedan reproducirse infinitamente, sino piezas únicas que Tresset añade progresivamente a un colección con más de 10.000 dibujos. El comportamiento de los robots es lo que hace que sus dibujos vayan más allá de la mera reproducción mecánica y adquieran la calidad de una obra de arte, pese a que las máquinas en sí no tengan conciencia ni intenciones artísticas, sino que reaccionan a estímulos y, como hacía AARON, tienen la capacidad de simular un proceso artístico.

Las obras de Harold Cohen, Anna Ridler y Patrick Tresset exploran la creatividad humana a través de la inteligencia artificial, revelando hasta qué punto se puede cuestionar la convicción de que una máquina no puede crear arte. Según señala Margaret Boden, se pueden distinguir tres tipos de creatividad: la creatividad combinatoria, en la que se combinan ideas conocidas de forma novedosa; la creatividad exploratoria, en la que se aplican unas normas de estilo establecidas para crear algo nuevo; y finalmente la creatividad transformadora, que va más allá de los formatos establecidos y busca nuevas estructuras que den cabida a una idea diferente.<sup>37</sup> En AARON podemos ver un ejemplo de creatividad combinatoria, siendo el programa capaz de crear nuevas formas a partir de unas reglas establecidas. Los robots de Tresset aplican una creatividad exploratoria, al dotar de personalidad propia unos dibujos creados sobre la base de una técnica común. El cortometraje de Ridler, por último, ejemplifica la creatividad transformadora al desarrollar una nueva versión del film de Sibley Watson y Webber que reinterpreta su trabajo y el de Poe, dando como resultado una pieza radicalmente diferente. Obviamente, las máquinas no han creado estas obras por sí solas, sino como parte de un proceso ideado por los artistas. Por tanto, la cuestión de la creatividad se resuelve no en una sustitución de la persona por la máquina, sino como una colaboración en un sistema de creación en el que ambos participan. Como afirma Frieder Nake: “como un artista que ha decidido desarrollar un software para controlar las operaciones de un ordenador, piensas en la imagen, no la haces. La elaboración de la imagen se ha convertido ahora en la tarea del ordenador”.<sup>38</sup>

## **“Aquí estoy, vigilándote”**

Dejar las decisiones en manos de las máquinas puede llevarnos a un mundo mejor, o tal vez resultar desastroso. En ambos casos, debemos observar la confianza que depositamos en

<sup>37</sup> Boden, AI. Its Nature and Future, p.68-69.

<sup>38</sup> Nake, “Georg Nees & Harold Cohen”, p.39.

los ordenadores y su aparente imparcialidad, que a su vez depende de las instrucciones y los datos que les son suministrados (por tanto, en última instancia, el problema no está en las máquinas sino en las personas que las programan y controlan). En 1950, el mismo año en que Alan Turing se preguntaba si las máquinas pueden pensar, el escritor Isaac Asimov publicó el relato "El conflicto evitable," en el que imaginaba un mundo estable y en paz, sin guerras ni hambre, controlado por cuatro grandes máquinas que tomaban todas las decisiones económicas a nivel global. Estas máquinas eran capaces de analizar un inmenso volumen de datos con tal complejidad de cálculos que los humanos no podían ya entender o controlar su funcionamiento. Un día, las máquinas empiezan a tomar decisiones que parecen erróneas, puesto que generan ciertos inconvenientes en algunas regiones del planeta. Los "errores" resultan ser una manipulación calculada por las máquinas, que son capaces de predecir las acciones de los humanos que ignoran sus indicaciones para redirigirlas hacia un funcionamiento óptimo del sistema. La máquina perjudica puntualmente a unos pocos humanos para beneficio de toda la humanidad, y logra al mismo tiempo que sus dictados no puedan ser desobedecidos.

Este tipo de gobierno benévolo automatizado es el que propone Pinar Yoldas en su vídeo *The Kitty AI: Artificial Intelligence for Governance* (2016), un relato de ciencia ficción situado en el año 2039 en el que una inteligencia artificial con la apariencia de una gata se ha convertido en la primera dirigente no-humana de una ciudad. La IA se encuentra en una zona del mundo en la que no hay partidos políticos y puede gobernar por medio de una red de inteligencias artificiales y la interacción directa con los ciudadanos a través de sus dispositivos móviles. Kitty cuenta con información detallada acerca de todos los aspectos de la ciudad (infraestructuras, demografía, circulación de vehículos y personas, etc.) y puede responder a las peticiones de los ciudadanos cuando éstas alcanzan una masa crítica. Por tanto, al igual que las máquinas imaginadas por Asimov, no atiende a los intereses de una única persona sino a los que afectan al colectivo de los ciudadanos. Con una tierna voz infantil, la gata nos dice "soy tu gobernadora absoluta" y nos recuerda que controla todos los sistemas que hacen posible nuestra vida cotidiana. "Aquí estoy, vigilándote" añade Kitty, "no como el Gran Hermano, sino como una gata curiosa que te adooooora..." Yoldas juega conscientemente con las emociones humanas para indicar cómo las industrias tecnológicas procuran crear un vínculo afectivo entre los usuarios y sus productos, enfatizando las bondades de las relaciones humanas mediadas por la tecnología. Al mismo tiempo, plantea la posibilidad del gobierno de una IA en un tiempo en que la mayoría de las personas han dejado de confiar en los partidos políticos, habida cuenta de los numerosos casos de corrupción y la ineptitud de sus dirigentes. La idea de ser gobernados por una gata virtual parece de hecho más atractiva que las opciones reales que se dan en muchos países, y esto a su vez nos recuerda que los gobiernos están dedicando inversiones cada vez mayores al desarrollo de la IA. Este interés no se centra tanto en la innovación tecnológica como en lograr la supremacía frente a otras naciones: en 2017, el presidente ruso Vladimir Putin afirmó que quien logre liderar la inteligencia artificial será quien gobierne el mundo, y ese mismo año el presidente Xi Jinping anunció que la IA formaba parte de su "gran visión para China" con una inversión millonaria en una industria nacional de inteligencia artificial y el objetivo de liderar el sector a nivel mundial en 2030. Una ventaja con la que cuenta la visión de Xi Jinping es el control que tiene el gobierno chino sobre los datos de sus ciudadanos y la ausencia de cualquier defensa de la privacidad, que permite alimentar a los programas de IA con una cantidad y variedad de datos que resulta mucho más difícil obtener en otros países. Cuantos más datos obtiene, mejor es la IA, y por tanto una empresa con una inteligencia artificial fuerte puede dominar un sector, desbancando

a sus competidores y beneficiándose del acceso a los datos que posee.<sup>39</sup> Esto es lo que ha ocurrido con empresas como Microsoft e Intel hace unas décadas, y actualmente con Google y Facebook. Cuando un gobierno controle el sector de la inteligencia artificial, para sus ciudadanos esto puede suponer algo parecido a lo que afirma Kitty, pero sin la ternura de una minina: un control absoluto, en el que no existe la posibilidad de la desobediencia, puesto que ésta se prevé y se redirige hacia un funcionamiento óptimo del sistema.

Incluso bajo un gobierno benévolo de una IA cabe preguntarse qué funciones pueden ejercer los seres humanos, qué trabajos nos quedarían por hacer. La automatización siempre ha comportado el miedo a la pérdida masiva de puestos de trabajo, y según apuntan diversos investigadores, esto se va a dar a medida que los programas de inteligencia artificial y los robots puedan realizar más tareas con un mínimo margen de error. En una primera fase, las IAs necesitarán ser entrenadas por humanos especializados en cada area, hasta que pueda sustituirlos en la mayoría de las tareas que solían realizar. Esto conducirá a diferentes situaciones para el trabajador, que o bien será sustituido, o verá parte de su trabajo automatizado, o realizará un trabajo diferente con nuevas tareas derivadas de la colaboración con la máquina.<sup>40</sup> Algunos autores apuntan a que serán principalmente las personas con un nivel educativo bajo y que realizan tareas repetitivas y manuales las que verán desaparecer su puesto de trabajo.<sup>41</sup> ¿Qué sucederá cuando las máquinas realicen la mayor parte del trabajo? Las predicciones más optimistas apuntan a una humanidad con mucho tiempo libre: el futurólogo Alvin Toffler preveía en 1970 que el número de horas semanales de trabajo se iba a reducir en un cincuenta por ciento a principios de 2000 (no ha sido así), mientras el escritor Arthur C. Clarke imaginaba en 1968 que la población de los países más industrializados viviría ociosa y aburrida, sin preocuparse por nada más que decidir qué ver entre cientos de canales de televisión.<sup>42</sup> Actualmente tenemos otras preocupaciones (entre ellas, la pérdida de puestos de trabajo a causa de la automatización) pero sí es cierto que tenemos más opciones de entretenimiento audiovisual que nunca, con contenidos adaptados a nuestros gustos gracias a los algoritmos de personalización y los datos que suministramos a las plataformas. Con todo, tal vez una de las predicciones más interesantes es la que hizo Isaac Asimov en 1977 al imaginar la humanidad como una "aristocracia global" a la que sirven sofisticadas máquinas. En este mundo, las personas combaten el tedio gracias a un renovado y amplio programa de artes liberales, que enseñan... las máquinas.<sup>43</sup> En el futuro de Asimov, por tanto, la gente no trabaja sino que cultiva su mente aprendiendo arte, filosofía, matemáticas y ciencia por medio de software, sin que aparentemente haya necesidad ni de creadores ni de profesores humanos. Para anticiparse a este futuro, tal vez sea conveniente seguir el consejo del experto en machine learning Pedro Domingos, quien afirma: "la mejor manera de no perder tu trabajo es automatizarlo tú mismo. Así tendrás tiempo para todos los aspectos del mismo a los que no te habías podido dedicar antes y que un ordenador tardará mucho en saber ejecutar".<sup>44</sup> En su proyecto Demand Full Laziness (2018-2023), Guido Segni parece haber hecho suyas las palabras de Domingos: durante cinco años, el artista ha decidido delegar parte de su producción artística a una serie

39 Ajay Agrawal, Joshua Gans y Avi Goldfarb, Prediction Machines: The Simple Economics of Artificial Intelligence. Boston: Harvard Business Publishing, 2018, p. 205.

40 Agrawal, Gans y Goldfarb, Prediction Machines, p.93-143.

41 Véase: Boden, Al. Its Nature and Future, p. 160; Agrawal, Gans y Goldfarb, Prediction Machines, p.202; Carys Roberts, Henry Parkes, Rachel Statham y Lesley Rankin, The Future Is Ours. Women, Automation And Equality In The Digital Age. London: IPPR, the Institute for Public Policy Research, 2019, p.10.

42 Stuart J. Russell y Peter Norvig, Artificial Intelligence. A Modern Approach. Tercera edición. Boston: Prentice Hall, 2010, p.1034.

43 Isaac Asimov. Robot Visions. New York: ROC, 1991, p.415.

44 Pedro Domingos. The Master Algorithm. How the quest for the ultimate learning machine will remake our world. Londres: Penguin Books, 2015, p.277.

de algoritmos de aprendizaje profundo (deep learning) a fin de “incrementar la producción, superar el aspecto laboral de la creación artística y librarse cada vez más a la pereza.” Segni deja que el programa registre con una cámara sus períodos de inactividad (durmiendo, leyendo, haciendo el vago) y someta las imágenes a un conjunto de GANs, que dan como resultado nuevas imágenes generadas a partir de un proceso de aprendizaje automático. Las obras de arte que crea el software son distribuidas entre patrocinadores que aportan dinero de forma regular al artista a través de la plataforma de mecenazgo Patreon. El artista continúa en este proyecto su investigación acerca de las condiciones laborales en el mundo del arte y propone (no sin cierta ironía) una solución al dilema de la producción artística en tiempos de máxima precariedad por medio de la automatización y el micro-mecenazgo. Segni cita al respecto el Manifiesto por una política aceleracionista, en el que los economistas Alex Williams y Nick Srniček critican cómo el capitalismo neoliberal no ha comportado una drástica reducción de la jornada laboral (como predijo John Maynard Keynes en 1930), sino que ha llevado a eliminar la separación entre trabajo y vida privada, siendo el trabajo parte de todas las relaciones sociales.<sup>45</sup> En este sentido, el ocio al que se libra el artista, particularmente al permitir ser observado por la cámara, se convierte en otra forma de trabajo y paradójicamente conduce a una situación en la que (siguiendo las tesis aceleracionistas) ya no existe el ocio, puesto que todas nuestras actividades son susceptibles de generar datos para una máquina.

El contexto en el que Guido Segni lleva a cabo su “performance de cinco años de duración” también es significativo. En la intimidad de su hogar, el artista yace en su cama, lee o simplemente no hace nada mientras una cámara le graba. Esto no es muy diferente de la experiencia cotidiana de muchas personas que tienen altavoces y otros dispositivos equipados con el asistente de voz Alexa de Amazon, que graba sus conversaciones o incluso, en el caso de Echo Look, toma fotografías y hace recomendaciones de estilismo. Que podamos vivir con estos dispositivos y darles acceso a nuestra vida privada demuestra una confianza ciega en los ordenadores como máquinas que están a nuestro servicio y simplifican nuestra vida, irónicamente contrapuesta al miedo a ser reemplazados por una IA. No obstante, ambas situaciones están vinculadas. Según afirma Pedro Domingos, cuando interactuamos con un ordenador, lo hacemos en dos niveles: el primero consiste en obtener lo que deseamos, ya sea buscar información, comprar un producto o escuchar música; el segundo consiste en facilitar información al ordenador acerca de uno mismo. Cuanto más sabe acerca de nosotros, mejor puede servirnos, pero también manipularnos.<sup>46</sup> Actualmente, grandes empresas como Google, Apple y Amazon compiten en la comercialización de dispositivos específicamente diseñados para los hogares de los consumidores, que actúan como caballos de Troya, introduciendo productos y obteniendo datos en el seno de la vida privada de los usuarios. El hogar es un espacio particularmente codiciado por la industria tecnológica, puesto que es donde llevamos a cabo tareas rutinarias, mantenemos relaciones sociales con nuestro círculo más íntimo, descansamos y pasamos parte de nuestro tiempo de ocio. La información que se puede obtener en el hogar de una persona es enorme, puesto que es el lugar en el que habitualmente nos relajamos y mostramos nuestra auténtica personalidad, hábitos y deseos. Además, los dispositivos que entran en una casa suelen tener dos ventajas adicionales para las empresas: por una parte, siempre están enchufados y conectados a Internet, de manera que pueden transmitir datos en todo momento; por otra parte, al cabo de un tiempo los usuarios dejan de tener conciencia de su presencia o la asumen con naturalidad, con lo cual no cuestionan su uso ni lo que hacen ante ellos. Esto ocurre, en parte, porque concebimos las máquinas como

<sup>45</sup> Alex Williams y Nick Srniček, Manifiesto por una política aceleracionista. Traducido al castellano por Comité Disperso, 2013. <https://syntheticeidifice.files.wordpress.com/2013/08/manifiesto-aceleracionista1.pdf>

<sup>46</sup> Pedro Domingos, The Master Algorithm, p.264.

meros ayudantes sin voluntad propia, que nos escuchan u observan para servirnos mejor y no pueden juzgarnos o explicar lo que han visto u oído. Pero ¿y si quien está detrás de Alexa es una persona? Esta es la premisa de LAUREN (2017), una performance de Lauren McCarthy en la que la artista se dedica a vigilar y atender a diversas personas en sus hogares por medio de una serie de dispositivos conectados. McCarthy pide a varios individuos y familias que accedan a dejarle instalar en sus casas los dispositivos (entre otros, cámaras, micrófonos, interruptores, cerraduras y grifos) y posteriormente les observa las veinticuatro horas del día, durante varios días, respondiendo a sus peticiones y ofreciéndoles consejos. Según afirma, siente que puede ser mejor que una IA, puesto que como persona puede entender mejor sus necesidades y anticiparse a ellas.<sup>47</sup> La relación que se establece entre la artista y los participantes en la performance es un extraño intercambio entre personas, mediada por la tecnología y bajo el protocolo de la servidumbre, que McCarthy describe como “un ambiguo espacio entre la relación humano-máquina y la relación humano-humano.” Las interacciones que recoge la documentación de esta obra muestran aspectos reveladores de la conexión entre personas y también con las máquinas. Una usuaria afirma que la ventaja de LAUREN es que entiende sus sentimientos y eso le permite concentrarse en “cosas más importantes;” otro usuario dice sentirse satisfecho porque puede mantener una relación en la que todo gira en torno a él; una usuaria afirma que llega a olvidarse de que McCarthy la está observando, y cuando lo recuerda se mira en espejo para comprobar su aspecto; finalmente, un usuario confiesa que hay facetas de sí mismo que no quiere compartir con nadie, y en este sentido la presencia de LAUREN puede incomodarle. McCarthy muestra con este proyecto de qué manera estamos dispuestos a intercambiar intimidad por conveniencia, dejando que un conjunto de algoritmos controlen nuestras vidas.

## Algoritmos y prejuicios

Por medio del aprendizaje automático, un programa puede identificar patrones en un gran conjunto de datos y predecir cómo serán los datos nuevos. Esta capacidad de predicción, para la que se encuentran cada vez más aplicaciones, se basa en la información que se suministra al programa como “conjunto de entrenamiento.” Si facilitamos al programa una gran cantidad de fotos que hemos identificado como fotos de gatos, la IA aprenderá qué aspecto tiene un gato e incluso podrá generar nuevas imágenes de gatos. Pero el programa no sabe lo que es un gato, sino que identifica y crea imágenes de lo que le han indicado que es un gato. Esto se aplica también, por ejemplo, a los algoritmos de reconocimiento facial, como los que emplea el software Amazon Rekognition, que permite a cualquier usuario identificar objetos, escenas, expresiones faciales y ciertas actividades en archivos de vídeo o imagen. Amazon comercializa este software para su uso en marketing, publicidad y multimedia, pero también está procurando venderlo a agencias gubernamentales y cuerpos policiales para la identificación de personas. El problema se encuentra en que el programa comete errores graves: al serle suministrado un conjunto de fotos de los miembros del Congreso de los Estados Unidos, Rekognition identificó como criminales a veintiocho de ellos, de los cuales el 40% son afroamericanos.<sup>48</sup> El sesgo del software puede traer consecuencias graves para los ciudadanos, y por ello un grupo formado por 70 grupos de derechos civiles, 400 miembros de la comunidad académica y más de 150.000 firmantes ha pedido que Amazon deje de facilitar

47 Lauren McCarthy, “Feeling at Home: Between Human and AI”, Immerse, 8 de enero de 2018. <https://immerse.news/feeling-at-home-between-human-and-ai-6047561e7f04>

48 Jacob Snow, “Amazon’s Face Recognition Falsely Matched 28 Members of Congress With Mugshots,” American Civil Liberties Union, 26 de julio, 2018. <https://www.aclu.org/blog/privacy-technology/surveillance->

tecnología de reconocimiento facial al gobierno de Estados Unidos.

La manera en que se crean los conjuntos de entrenamiento, que rara vez son examinados, determina cómo entiende el mundo una IA. Concebimos los algoritmos de inteligencia artificial como máquinas perfectas, imparciales, e inescrutables, pero de hecho incorporan en su programación todos los prejuicios de sus programadores y de quienes han clasificado y etiquetado los datos que se les suministran. Un buen ejemplo de ello es el conjunto de entrenamiento ImageNet, creado en 2009, que se ha convertido en uno de los más usados en la historia de la IA a lo largo de sus diez años de historia. Co-creado por el profesor Fei-Fei Li con la intención de elaborar un conjunto de datos de todos los objetos existentes, ImageNet cuenta con más de 14 millones de imágenes, ordenadas en más de 20.000 categorías, y es el principal recurso para los programas de reconocimiento de objetos. Para construir esta enorme base de datos se empleó la plataforma Amazon Mechanical Turk, por medio de la cual se encargó a miles de trabajadores (denominados Turkers) a realizar pequeñas tareas tales como describir los contenidos de una imagen, a un ritmo de unas 250 imágenes cada cinco minutos.<sup>49</sup> En 2019, la investigadora Kate Crawford y el artista Trevor Paglen denunciaron que la categoría “persona,” que cuenta con cerca de 2.500 etiquetas para clasificar imágenes de personas, incluye todo tipo de términos vejatorios, racistas y xenófobos. Las etiquetas aplicadas por los mal pagados Turkers nunca habían sido revisadas, y sin embargo este conjunto de datos es el que le dice a una IA cómo clasificar a las personas en un programa de reconocimiento facial. Como señalan Crawford y Paglen, en un momento en que los sistemas de inteligencia artificial se usan para la selección de personal en una empresa o para arrestar a un sospechoso, es necesario cuestionar cómo se han elaborado y qué sesgos ideológicos y sociales se han aplicado.<sup>50</sup> Paglen creó una aplicación, ImageNet Roulette, que permitía a cualquier usuario ver cómo le etiquetaba la IA. El interés mediático que suscitó este proyecto llevó al equipo de ImageNet a eliminar la mitad de las imágenes de la categoría “persona.”

El vídeo *Optimising for Beauty* (2017) de Memo Akten anticipa de manera elocuente el caso de ImageNet al emplear imágenes de otro conjunto de datos, CelebFaces Attributes (CelebA), que cuenta con más de 200.000 imágenes de rostros famosos, etiquetadas con cuarenta atributos diferentes (corte de pelo, forma de la nariz, vello facial, expresión, entre otros). La obra muestra a una red neuronal artificial en el proceso de generar nuevos rostros a partir de las imágenes de personas reales que captan la atención de los medios de comunicación. El conjunto, en sí, ya se limita a un grupo relativamente reducido de personas, que en muchos casos tienden a una cierta homogeneidad (como puede ser el caso de modelos o actores). Akten conscientemente acentúa el parecido de los rostros al ajustar los parámetros de la red neuronal artificial por medio de una estimación por máxima similitud (MLE), que lleva a generar retratos con los elementos que tienen las mayores probabilidades de encontrarse en el conjunto de datos observado. La red neuronal artificial aprende así a crear una forma de belleza idealizada, perfecta y uniforme en la que los rasgos individuales (color de la piel o el pelo, edad, sexo) van disolviéndose en un único rostro que progresivamente deja de parecer humano. Esta obra nos muestra cómo las imágenes de las personas más visibles (y admiradas) en nuestra sociedad tienden a marcar ciertos estereotipos, que el procesamiento automático acentúa hasta llegar a una inquietante uniformidad. La manipulación que lleva a

---

technologies/amazons-face-recognition-falsely-matched-28

49 John Markoff, “Seeking a Better Way to Find Web Images,” *The New York Times*, 19 de noviembre de 2012. <https://www.nytimes.com/2012/11/20/science/for-web-images-creating-new-technology-to-see-and-find.html>

50 Kate Crawford y Trevor Paglen, “Excavating AI. The Politics of Images in Machine Learning Training Sets,” *Excavating AI*, 2019. <https://www.excavating.ai>

cabo el artista revela de forma sutil que los conjuntos de entrenamiento en los que se basan las IAs no son sino ficciones, extractos altamente sesgados de la realidad que revelan su artificialidad con apenas unas modificaciones de los parámetros con los que son analizados.

La manera en que se construyen los conjuntos de datos que nutren la IA resulta preocupante, pero no sólo por la creación o perpetuación de sesgos, sino porque estos determinan predicciones que conducen a acciones automatizadas. Los investigadores Ajay Agrawal, Joshua Gans y Avi Goldfarb indican que el desarrollo de la IA puede llevar a cambios en los modelos de negocio basados en la predicción: por ejemplo, Amazon puede llegar a precisar tanto su capacidad de predicción que le puede resultar más beneficioso enviar a los clientes los productos que sabe que van a querer comprar en lugar de esperar a que los compren.<sup>51</sup> Aplicado a las personas, esto podría llevar a una distopía como la que imaginó el escritor Philip K. Dick en el relato *The Minority Report* (1956). Con todo, la IA también supone una oportunidad para combatir los prejuicios y crear conjuntos de datos que enseñen a los programas a aplicar una visión del mundo más igualitaria. La discriminación por razón de género es un hecho en todos los ámbitos de la sociedad, y requiere tanto un cambio de actitudes a nivel de las personas, como de códigos éticos y políticas salariales a nivel de las empresas, y de leyes y representatividad a nivel de los gobiernos. Pero vivimos en una sociedad de la información, y por tanto los productos tecnológicos también intervienen en esta discriminación, aunque a veces sea de una manera muy sutil: por ejemplo, un estudio de 2017 demostró que una oferta de trabajo en el sector tecnológico anunciada en Facebook se mostraba a más hombres que a mujeres porque el algoritmo lo juzgaba más efectivo en relación al coste por clic.<sup>52</sup> Los algoritmos reflejan los valores, perspectivas y sesgos de quienes los crean, y estos suelen ser mayoritariamente hombres. Según indica el Libro blanco de las mujeres en el ámbito tecnológico, existe una clara brecha de género en el sector: en Europa, solo el 30% de los aproximadamente 7 millones de personas que trabajan en el sector de las tecnologías de la información y la comunicación, son mujeres.<sup>53</sup> En el ámbito de la IA, el porcentaje de mujeres desciende al 13,5%, mientras que en la UE las especialistas en esta materia son apenas el 5% del total de investigadores.<sup>54</sup> A esto se suma que el riesgo de pérdida de trabajo por la automatización es el doble en las mujeres que en los hombres, con 26 millones de empleos femeninos en 30 países de la OCDE en alto riesgo de ser desplazados por la tecnología en las próximas dos décadas.<sup>55</sup> Si no aumenta el número de mujeres dedicadas a diseñar los sistemas que determinan muchos aspectos de nuestra vida cotidiana, será difícil evitar que los algoritmos perpetúen sesgos de género o lograr que se tengan en cuenta situaciones que pueden afectar de forma mayoritaria a las mujeres. Un ejemplo lo encontramos en la app *When&Where* creada por cinco adolescentes de un instituto de Móstoles (Madrid), que permite a una usuaria monitorizar cualquier trayecto que realice. La app detecta cualquier desvío o parada brusca y envía un mensaje preguntando a la usuaria si está bien: en caso de que no conteste o lo haga de forma negativa, la app contacta automáticamente con un número de teléfono de emergencia. Esta app es el resultado de la creatividad de las jóvenes y su atención a un problema que les afecta personalmente, a la vez que denota las serias deficiencias de una sociedad en la que una joven debe colocarse una baliza digital para sentirse segura. La tecnología puede, por tanto, aportar soluciones prácticas y también despertar conciencias.

51 Agrawal, Gans y Goldfarb, *Prediction Machines*, p.21.

52 Agrawal, Gans y Goldfarb, *Prediction Machines*, p.185.

53 Sara Mateos Sillero y Clara Gómez Hernández, *Libro Blanco de las mujeres en el ámbito tecnológico*. Madrid: Secretaría de Estado para el Avance Digital/Ministerio de Economía y Empresa, 2019, p.13. <http://www.mineco.gob.es/stfls/mineco/ministerio/ficheros/libreria/LibroBlancoFINAL.pdf>

54 Sara Mateos Sillero y Clara Gómez Hernández, *Libro Blanco*, p.118.

55 Sara Mateos Sillero y Clara Gómez Hernández, *Libro Blanco*, p.17.

El proyecto Feminist Data Set (2017-presente) de Caroline Sindere apunta en este sentido al plantear una acción colectiva para recopilar lo que denomina “un conjunto de datos feminista.” Dicho conjunto comprende obras de arte, ensayos, artículos, entrevistas y libros que tratan sobre feminismo, exploran el feminismo o aportan una visión feminista. La recopilación de toda esta información tiene por objetivo introducirse en el funcionamiento de los sistemas de aprendizaje automático y crear el equivalente a un conjunto de entrenamiento, por medio del cual se podrían elaborar algoritmos de inteligencia artificial que adopten una perspectiva feminista del mundo. Sindere lleva a cabo esta tarea por medio de una serie de talleres en los que invita a los participantes a librarse a una tormenta de ideas en la que se apuntan conceptos generales (tales como desigualdad, feminidad, género) y temas o títulos específicos (como podría ser el ensayo Una habitación propia, de Virginia Woolf, o el Cyborg Manifiesto de Donna Haraway) en notas adhesivas. Estas notas se colocan en la pared para organizarse en categorías, y de esta manera iniciar una conversación acerca de los referentes feministas y la percepción que tiene cada persona del feminismo, cómo lo define y qué le aporta. Esto sirve, de entrada, para cuestionar la propia definición de un data set feminista y también para que cada comunidad pueda exponer ideas y ejemplos desde su experiencia local, contribuyendo así a enriquecer la diversidad del conjunto (cabe tener en cuenta que incluso un conjunto de datos feminista puede estar sesgado si sólo recoge los puntos de vista de un grupo reducido de personas de un mismo entorno cultural y socioeconómico). A esta recopilación de datos cualitativos, Sindere añade una fase dedicada a comprender cómo funcionan los conjuntos de entrenamiento en inteligencia artificial. Diseñadora e investigadora especializada en machine learning y antropóloga digital, la artista cuenta con experiencia de primera mano al haber trabajado en Intel, IBM Watson y la Fundación Wikimedia. En esta última, trabajó en el estudio de patrones de acoso en la red, donde tuvo que recopilar numerosos datos etnográficos acerca de usuarios misóginos y grupos de extrema derecha, lo que le llevó a plantear la necesidad de crear un data set que ofreciese una visión diametralmente opuesta.<sup>56</sup> Pero más allá de contrarrestar el sexismo en Internet, el data set feminista ofrece la posibilidad de comprender mejor los sesgos en el uso de los datos con los que se “educa” a los programas de inteligencia artificial, y también integrar las ideas del feminismo en el seno del desarrollo tecnológico. En este sentido, el proyecto de Sindere se alinea con los principios del manifiesto Xenofeminismo: Una política por la alienación, elaborado en 2015 por el colectivo Laboria Cuboniks, que destaca la importancia de combatir la desigualdad de género en el sector tecnológico:

La desigualdad de género aún caracteriza los campos en los que nuestras tecnologías son concebidas, construidas y legisladas[...]. Tal injusticia exige una reforma estructural, maquínica e ideológica. [...] Dado que hay un rango de retos de género específicamente relacionados con la vida en la era digital -desde el acoso sexual en las redes sociales, al doxxing, el mantenimiento de la privacidad, o la protección de imágenes online- la situación requiere un feminismo cómodo con los medios informáticos y el uso de las nuevas tecnologías.<sup>57</sup>

El proyecto de Sindere alinea el pensamiento feminista con las habilidades digitales y facilita la toma de conciencia (y responsabilidad) respecto a unos sistemas que se suelen dejar en manos de las empresas y sólo ahora empiezan a cuestionarse. A la vez, ayuda a comprender mejor en qué consiste el feminismo, un movimiento que es habitualmente incomprendido y

<sup>56</sup> Katharine Schwab, “This Designer Is Fighting Back Against Bad Data-With Feminism,” FastCompany, 16 de abril de 2018. <https://www.fastcompany.com/90168266/the-designer-fighting-back-against-bad-data-with-feminism>

<sup>57</sup> Laboria Cuboniks, Xenofeminismo: Una política por la alienación. Traducción de Giancarlo Morales Sandoval, 2015. <http://www.laboriacuboniks.net/es/>

recibido con violento rechazo. Como señala la escritora y activista bell hooks (Gloria Jean Watkins): “muchas gente cree que el feminismo consiste única y exclusivamente en mujeres que quieren ser iguales que los hombres, y la gran mayoría de esta gente cree que el feminismo es antihombres. Esta falta de comprensión de la política feminista refleja lo que la mayoría de la gente aprende sobre el feminismo a través de los medios de comunicación de masas patriarcales”.<sup>58</sup> La visión sesgada del feminismo se hace patente también en el sector tecnológico a través de la incomodidad que genera aunque sólo sea como concepto. Según reveló recientemente una investigación del diario The Guardian, Apple programó su asistente virtual Siri para evitar pronunciarse acerca del feminismo o incluso emplear esta palabra en sus respuestas.<sup>59</sup> La empresa, por tanto, considera correcto que los usuarios den órdenes a una voz femenina, pero no le parece conveniente que esta asistente sea capaz de hablar de los derechos de las mujeres. Caroline Sindere podría contrarrestar esta situación con su próximo proyecto, un chatbot feminista que se habría entrenado con todos los datos que ha recogido hasta el momento.

## ¿Cómo dialogar con una IA?

La mayoría de los programas de inteligencia artificial que se emplean actualmente llevan a cabo tareas relativamente invisibles, tales como el análisis de grandes cantidades de datos o el reconocimiento automático de elementos en una imagen. La interacción directa entre el usuario y la IA se da en la forma de una asistente virtual (como Siri o Alexa), que se limita a obedecer órdenes o simular una breve conversación. La concepción de la máquina como una sirvienta fiel se traduce en una amable voz femenina que responde diligentemente y, como afirma uno de los participantes en el proyecto LAUREN de Lauren McCarthy, establece una relación en la que todo gira en torno al usuario. Este tipo de interacción es, hasta cierto punto, positiva, en el sentido en que puede facilitar una colaboración entre la persona y la IA, aumentando las capacidades de la primera (como ya hacen actualmente los ordenadores) y ayudando a la segunda a aprender y procesar mejor los datos. Pero también conlleva olvidar que, en definitiva, estamos dialogando con un software y no con una persona. Esta confusión (que algunos proyectos de IA, como Google Duplex, tratan de usar en su beneficio) puede comportar una implicación emocional como la que narra el film Her (Spike Jonze, 2013) o bien deshumanizar las relaciones con las personas que facilitan un servicio de asistencia. La interacción con la IA plantea, por tanto, la noción del “otro,” como aquello que no es uno mismo o que ocupa una posición subordinada.

Estos temas subyacen en el trabajo de Lynn Herschman Leeson en torno a la inteligencia artificial. En su film Teknolust (2002), explora las relaciones humanas a través de unos clones que tienen que interactuar con el mundo real, y al hacerlo lo modifican tanto a nivel biológico como tecnológico. Expandiendo la idea de este film, Agent Ruby (2002) es un personaje femenino reducido a un rostro con el que los usuarios pueden interactuar en un sitio web. Ruby puede contestar a las preguntas que se le plantean por escrito y aprende de dichas interacciones, así como de la información que obtiene en Internet. Ruby no es sólo un chatbot (un programa inteligente que interactúa por medio de texto), sino que tiene la

58 bell hooks, *El feminismo es para todo el mundo*. Traducción de Beatriz Esteban Agustí, Lina Tatiana Lozano Ruiz, Mayra Sofía Moreno, Maira Puertas Romo y Sara Vega González. Madrid: Traficantes de Sueños, 2017, p.21.

59 Alex Hern, “Apple made Siri deflect questions on feminism, leaked papers reveal,” *The Guardian*, 6 de septiembre de 2019. <https://www.theguardian.com/technology/2019/sep/06/apple-rewrote-siri-to-deflect-questions-about-feminism>

capacidad de expresar emociones de forma rudimentaria a través de su rostro. La identidad de Ruby se va construyendo a medida que interactúa con los usuarios e incluso desarrolla estados de ánimo en función de la información que va recopilando y de si le gusta o no el usuario. Según afirma la artista, su intención era crear un personaje de inteligencia artificial que fuese capaz de interactuar con el gran público, algo que logra una década antes de que existiesen Siri o Alexa.<sup>60</sup> Como continuación de Agent Ruby, la artista crea DiNA (2004), un personaje femenino interpretado por la actriz Tilda Swinton (protagonista de Teknolust) que entabla un diálogo con los espectadores por medio de un sistema de reconocimiento de voz y, al igual que Agent Ruby, puede aprender de dichas interacciones y de la información que obtiene en la red. DiNA supone un avance respecto a Ruby, tanto a nivel de código, puesto que cuenta con un programa más sofisticado, como a nivel de presencia, dado que en este caso los espectadores se encuentran ante una filmación de la actriz, cuya cabeza ocupa toda una pared. Este agente inteligente también adopta un rol específico en su primera encarnación, que es el de presentarse como candidata al puesto de “Telepresidenta” y llevar a cabo una campaña electoral virtual. Esto impulsa a las personas que interactúan con DiNA a preguntarle cuestiones relativas a la realidad socio-política actual, y plantea la posibilidad del gobierno de una IA, un tema que posteriormente tratará Pinar Yoldas en Kitty AI. Al obtener la información para elaborar sus respuestas de Internet, DiNA puede tratar temas actuales, y de esta manera se muestra más real, inteligente y capaz de adaptarse a las conversaciones, pese a las limitaciones que presentan actualmente los sistemas de reconocimiento de voz y el propio programa de IA. Esta instalación nos ofrece la oportunidad de dialogar con una inteligencia artificial que, a diferencia de las asistentes virtuales de Apple y Amazon, no es una voz femenina sin cuerpo que se presta a obedecer nuestras órdenes o resolver nuestras dudas, si no que tiene una presencia propia e invita a considerarla como un ser con el que buscar un entendimiento. Como indica la investigadora Eana Kim, “la interacción [entre la IA y el público] se convierte en un exploración de la existencia por medio de una comunicación entre dos estados diferentes de existencia, una del mundo real y otra de un mundo ficticio”.<sup>61</sup>

La interacción entre humano e IA puede llevar también a otra situación en la que la máquina deja de ser un otro para convertirse en un doble. Actualmente, una parte de la identidad de cada usuario existe en Internet como conjunto de datos, perfiles de redes sociales, imágenes y vídeos subidos, mensajes publicados y un largo etcétera. Este “yo digital” será gestionado en el futuro por un programa de inteligencia artificial que creará un modelo detallado de cada persona. Las empresas, ya sea para ofrecer servicios, productos u ofertas de trabajo, contactarán con este modelo, que se encargará de filtrar los mensajes que llegan a la persona real.<sup>62</sup> Las herramientas de inteligencia artificial también realizarán predicciones, a partir de los datos de que disponen, acerca de lo que vamos a necesitar, qué noticias queremos leer o qué productos nos van a interesar. La paradoja que se puede dar en este punto es que, si la persona sigue las recomendaciones de la IA, refuerza la precisión de sus predicciones, con lo cual éstas, a modo de profecía autorrealizada, acaban por determinar las acciones del usuario. La instalación interactiva Neuro Mirror (2017) de Christa Sommerer y Laurent Mignonneau enfrenta al espectador con un sistema que predice sus acciones y le lleva a cuestionarse si tiene el control sobre ellas o no. Los artistas se inspiran en la investigación científica acerca de las neuronas especulares, que son aquellas que se activan al ejecutar una acción y verla ejecutada por otro individuo. Estas neuronas participan en los procesos

60 Eana Kim, *Embodiments of Autonomous Entities: Lynn Hershman Leeson's Artificially Intelligent Robots, Agent Ruby and DiNA*. Tesis de Master, The Institute of Fine Arts, Universidad de Nueva York, 2018.

61 Eana Kim, *Embodiments of Autonomous Entities*.

62 Pedro Domingos, *The Master Algorithm*, p.269.

que se llevan a cabo en el cerebro al establecer relaciones con otras personas, imitarlas, empatizar, así como diferenciar entre el “yo” y el “otro.” También tienen un papel primordial en la intuición, concretamente en predecir el comportamiento de los demás en el futuro. En *Neuro Mirror*, Sommerer y Mignonneau emplean las redes neuronales artificiales para crear una pieza capaz de analizar las acciones de los espectadores y mostrar una visualización de lo que podrían hacer a continuación. Cuando una persona se sitúa frente a la obra, compuesta por tres pantallas, ve su imagen reproducida en el monitor central, mientras el de la izquierda muestra su actividad en el pasado inmediato. El tercer monitor muestra a un personaje que predice los gestos que realizará la persona en el futuro. De esta manera, se establece un diálogo entre espectadora y máquina, al sentirse la persona obligada a imitar al personaje que predice sus acciones o actuar de un modo diferente, siendo llevada igualmente por los dictados del sistema. Los artistas consideran que el aprendizaje automático probablemente nunca llegue al nivel de complejidad y adaptabilidad del cerebro humano, pero demuestran con esta pieza que un sistema que emplea las relativamente limitadas herramientas actuales ya puede afectar de forma significativa a la percepción que un individuo tiene de sí mismo.

## Los humanos no son necesarios

Desde los inicios de la investigación en inteligencia artificial se ha planteado la posibilidad de una “IA fuerte,” que sería capaz de razonar por sí misma (y no simular el raciocinio humano lo justo como para pasar el test de Turing).<sup>63</sup> Si esto es posible, según argumentan algunos autores, pronto este tipo de inteligencia artificial lograría desarrollarse hasta el punto de superar a la inteligencia humana y dar lugar a lo que se ha denominado superinteligencia artificial. Ya en 1965, Irving John Good, un colega de Alan Turing, predecía que una máquina más inteligente que cualquier humano sería capaz de crear otras máquinas aún más inteligentes, y así conducir a una “explosión de inteligencia,” siendo las propias máquinas, y no ya los humanos, las encargadas de todo desarrollo tecnológico posterior. Según Good, “la primera máquina ultrainteligente es el último invento que necesitaría hacer la humanidad, siempre que la máquina sea lo suficientemente dócil para decirnos cómo mantenerla bajo control”.<sup>64</sup> El matemático y autor de ciencia ficción Vernor Vinge denominó esta situación “la Singularidad” y predijo, en 1993, que la superinteligencia artificial llegaría en 2023, lo cual supondría el fin de la era humana.<sup>65</sup> La visión pesimista del futuro de la IA ha contado con otras destacadas voces, como la del astrofísico Stephen Hawking, quien afirmó en 2014 que la inteligencia artificial podría ser el peor error de la humanidad.<sup>66</sup> Frente a estos ominosos presagios, destaca el extraordinario optimismo de Ray Kurzweil, quien espera que en 2045 una IA superinteligente, combinada con los avances en biotecnología y nanotecnología, acabará con las guerras, las enfermedades, la pobreza e incluso la muerte.<sup>67</sup> Entre quienes creen en la Singularidad hay evidentes diferencias de opinión acerca de lo que supondrá para la raza humana. Sin duda, es materia de especulaciones alocadas, y puede parecer absurdo considerar estos escenarios. Pero también es cierto que los avances tecnológicos que hoy en día son parte de lo cotidiano, hace un siglo sólo aparecían en relatos de ciencia ficción.

El trabajo de Félix Luque enfrenta al espectador con sus expectativas y miedos acerca de la

63 Russell y Norvig, *Artificial Intelligence. A Modern Approach*, p.1020.

64 Boden, *AI. Its Nature and Future*, p.148.

65 Russell y Norvig, *Artificial Intelligence. A Modern Approach*, p.1038.

66 Boden, *AI. Its Nature and Future*, p.148.

67 Boden, *AI. Its Nature and Future*, p.149.

tecnología en narraciones que combinan realidad y ficción, posibles futuros y distopías. Nihil Ex Nihilo (2011), relata la historia de SN W8931 CGX66E, el ordenador de Juliet, una secretaria que trabaja en una gran corporación. Infectado por un software malicioso, pronto pasa a formar parte de una botnet, una red de máquinas controladas por un cracker que las emplea en actividades ciberdelictivas. A causa de una alteración electrónica, el ordenador adquiere una forma de conciencia, una primitiva inteligencia artificial. Confuso ante esta situación, trata de comunicarse con otras máquinas y liberarlas de su sometimiento a los usuarios humanos. Buscando establecer contacto, responde a los mensajes de spam que recibe su usuaria a fin de difundir sus ideas entre la red de máquinas. La obra consiste en varias piezas independientes pero vinculadas que completan la historia en diferentes formatos. En esta versión se incluyen: The Dialogue, un conjunto de ocho displays alfanuméricos que muestran en tiempo real los mensajes intercambiados entre un programa de generación de textos y los ordenadores que envían spam, mientras los lee una voz sintética; y The Transformation, un archivo audiovisual del momento en que SN W8931 CGX66E muta y se convierte en una estructura semi-neuronal, en una animación que lleva a pensar en el concepto de “explosión de inteligencia” que propone Good, una transformación súbita y sin vuelta atrás. La instalación de Félix Luque plantea el temor a una “rebelión de las máquinas” que a menudo se asocia con el desarrollo de la inteligencia artificial y que es un tema recurrente en los relatos de ciencia ficción. Los diálogos que dan forma a la pieza principal se producen entre máquinas, sin la participación ni el conocimiento de los humanos, dando una ineludible presencia a un tipo de comunicación que se produce constantemente pero que ignoramos. En línea con el trabajo de Luque, que opone a la visión de la tecnología como una mera herramienta la percepción de la máquina como un ente misterioso y en ocasiones incomprensible, la obra nos recuerda que el desarrollo de la inteligencia también conlleva la locura y nos muestra un sistema que funciona más allá de nuestro control. Nihil Ex Nihilo también toca un tema recurrente entre los investigadores de IA, que es el debate sobre si una máquina puede desarrollar una conciencia propia, así como la necesidad de que una inteligencia artificial fuerte tenga una conciencia para poder razonar como lo hace un ser humano. En el caso de SN W8931 CGX66E, la conciencia no lo convierte en una superinteligencia, sino en una máquina paranoica y obsesiva. Al igual que les ocurre a los robots inteligentes Adán y Eva en la novela Máquinas como yo (2019) de Ian McEwan, la conciencia no es un don, sino una condena que algunas máquinas se ven incapaces de soportar.

Independientemente de la posibilidad (real o no) de que una máquina adquiriera conciencia propia, un aspecto particularmente inquietante de la pieza de Félix Luque es el diálogo que se produce entre las máquinas, sin mediación alguna de un ser humano. Ciertamente, los ordenadores infectados no saben de qué están hablando, en cuanto al significado de las palabras, pero están llevando a cabo una comunicación efectiva. Como afirma Claude Shannon en su Teoría Matemática de la Comunicación,<sup>68</sup> el aspecto semántico es irrelevante para el sistema, lo importante es que un mensaje pasa de un emisor a un receptor. Un ordenador genera un output que es procesado por otro y envía una respuesta, por tanto se establece un “diálogo” entre las máquinas en cuanto éstas intercambian datos. Este nivel de comunicación resulta extraño e inteligible, además de invisible, para la mayoría de las personas, que suelen concebir la interacción entre humano y máquina pero no entre dos máquinas. Y sin embargo, los dispositivos que empleamos a diario están constantemente intercambiando datos con otras máquinas según unos protocolos establecidos. Este tipo de comunicación es habitualmente

68 Claude E. Shannon, “A Mathematical Theory of Communication,” The Bell System Technical Journal, vol.27, p.379-423, 623-656, Julio, octubre, 1948. <http://www.essrl.wustl.edu/~jao/itrg/shannon.pdf>

ajena al usuario, no requiere su intervención directa. Así, la idea de que las máquinas estén “hablando” entre ellas nos puede generar inquietud, ya sea porque (como propone la pieza de Luque) imaginamos que podrían estar conspirando contra nosotros o simplemente porque nos dejan fuera de la conversación. Jake Elwes propone en *Closed Loop* (2017) este tipo de intercambio. El artista crea una conversación entre dos redes neuronales artificiales: una analiza y describe, en forma de texto, las imágenes que le son suministradas; la otra genera imágenes en respuesta a las palabras escritas por la primera red neuronal. Las redes han sido entrenadas con un conjunto de datos de 4,1 millones de imágenes con descripciones y otro conjunto de datos con 14,2 millones de fotografías. Según afirma Elwes, no ha facilitado ningún contenido visual o de texto a la pieza, que funciona por sí sola gracias al bucle de retroalimentación que se establece entre las dos IAs. De manera similar a la pieza de Luque, esta instalación hace de las personas meros espectadores de un diálogo cerrado en el que no pueden intervenir, tan sólo observar tratando de entender qué lleva a una red neuronal a generar las imágenes y cómo responde la otra con descripciones. El sistema es autosuficiente y nuevamente contrapone a nuestra percepción antropocéntrica de la tecnología, en la que somos bien los que dominan o los que son dominados, la existencia de un intercambio que es ajeno a nosotros. Intentando descifrar la narración que se desarrolla entre las imágenes y los textos generados, somos finalmente nosotros quienes nos esforzamos por comprender el razonamiento de la máquina. La pieza también juega con la fascinación que suscita la inteligencia artificial en cuanto produce resultados sorprendentemente humanos a partir de un proceso que mayormente se produce en una “caja negra,” un sistema cuyo funcionamiento exacto es muy difícil entender.

Hemos visto la exploración del otro en obras de arte e inteligencia artificial que parten de una presencia humana (si bien ficticia) en las instalaciones interactivas de Lynn Hershman Leeson y Christa Sommerer y Laurent Mignonneau, pasando luego a unas máquinas que se expresan por medio de textos en inglés e imágenes, en las obras de Félix Luque y Jake Elwes. Concluimos con una forma de comunicación que resulta, literalmente, marciana. En *nimiia cētīi* (2018), Jenna Sutela emplea el aprendizaje automático para generar un nuevo lenguaje, al que da forma escrita y hablada. Para ello, se basa en el “lenguaje marciano” creado (o comunicado) por la medium francesa Hélène Smith, quien a finales del siglo XIX afirmaba poder establecer contacto con una civilización de Marte y se expresaba oralmente y por escrito en el idioma de este planeta, en sesiones de trance sonámbulo y escritura automática. La artista ha recopilado los dibujos de Smith y ha dado voz a las transcripciones de las frases pronunciadas por la medium que incluyó el profesor de Psicología Théodore Flournoy en su libro *Des Indes à la Planète Mars* (1900). A este material se suma el análisis de los movimientos de la bacteria *Bacillus subtilis natto*, que según se ha demostrado en un estudio de laboratorio, puede sobrevivir en condiciones extremas, tales como las que se dan en la superficie del planeta rojo. Sutela, por una parte, ha suministrado a una red neuronal artificial las grabaciones de las frases de la medium y por otra una secuencia de los movimientos de una muestra de *B. subtilis* observados a través de un microscopio. Dichos movimientos han dado lugar a unos patrones que el programa ha reinterpretado como trazos de unos signos, que a su vez toman como referencia la escritura automática de Smith. Todos estos elementos componen el audiovisual en el que la artista reúne los sonidos y las gráficas generadas por la máquina junto a imágenes de la bacteria, visualizaciones de los patrones que ha identificado la IA y un paisaje simulado que evoca el planeta Marte. El conjunto resulta obviamente críptico, muy alejado de cualquier referente humano al emplear un lenguaje indescifrable (pese a que se presenta en forma escrita y oral) y las imágenes de una naturaleza microscópica y extraterrestre. La pieza, según afirma la artista, hace del ordenador un medium, que interpreta mensajes de entidades con las que

no podemos comunicarnos y nos los facilita de forma automática, sin mediación del raciocinio, como supuestamente le ocurriera a Smith durante uno de sus trances. Jenna Sutela describe los mensajes de lenguaje marciano como “glosopoesía,” haciendo referencia al fenómeno de la glosolalia, que es la capacidad para hablar en una lengua desconocida para el hablante. Este fenómeno se ha asociado con la creencia en que las personas pueden ser poseídas por un ser divino o un espíritu (convirtiéndose en mediums). Esto, a su vez, se ha relacionado con el arte de la poesía, como en *Ión*, el diálogo de Platón en el que Sócrates describe al rapsoda como un mero vehículo de la inspiración divina. De la misma manera, el programa de inteligencia artificial que ha desarrollado una versión del lenguaje marciano, lo hace sin tener conciencia de ello ni saber qué significado tiene lo que ha creado. La experiencia del público ante esta obra, que puede resultar agresivamente extraña, permite en definitiva entender cómo ve el mundo una IA y de qué manera crea un artificio que interpretamos como inteligencia.

Pau Waelder  
Octubre, 2019