UNIVERSIDAD DE LOS ANDES

FACULTAD DE ARTES Y HUMANIDADES

DEPARTAMENTO DE HUMANIDADES Y LITERATURA

Máquinas, robots y consciencias artificiales: redefiniendo el concepto de humano en *Machines Like Me* (2019) de Ian McEwan

Trabajo de grado para optar al título de Literato

Presentado por Sebastián Sánchez Betancourt

Dirigido por María Mercedes Andrade

Noviembre de 2022

Contenido

…one could see the history of human self-regard as a series of demotions tending to extinction. Once we sat enthroned at the centre of the universe, with sun and planets, the entire observable world, turning around us in an ageless dance of worship. Then, in defiance of the priests, heartless astronomy reduced us to an orbiting planet around the sun, just one among the rocks. But still we stood apart, brilliantly unique, appointed by the creator to be lords of everything that lived. Then biology confirmed that we were at one with the rest, sharing common ancestry with bacteria, pansies, trout and sheep. In the early twentieth century came deeper exile into darkness when the immensity of the universe was revealed and even the sun became one among billions in our galaxy, among billions of galaxies. Finally, in consciousness, our last redoubt, we were probably correct to believe that we had more of it than any creature on earth. But the mind that had once rebelled against the gods was about to dethrone itself by way of its own fabulous reach. In the compressed version, we would devise a machine a little cleverer that ourselves, then set that machine to invent another that lay beyond our comprehension. What need then of us?

-*Ian McEwan,* Machines Like Me (86)

Introducción

El ser humano ya no es la medida de todas las cosas. Durante siglos, el Humanismo que nació en Europa con el Renacimiento ha puesto al humano como punto de partida y punto de llegada de toda reflexión filosófica. Sin embargo, el Humano poco a poco ha dejado de ser el centro del mundo. Antes, el humano estaba ubicado en un centro desde el cual juzgaba y pensaba todo lo no-humano a su alrededor; ahora las fronteras entre lo humano y no-humano se han vuelto más difusas y el humano se ubica en un continuo cuyos extremos son la biología y el algoritmo (Colombino 383).

En un extremo, la teoría de la evolución planteada por primera vez por Charles Darwin en el siglo XIX confronta al Humano con el resto del mundo natural y, en particular, con los animales. Bacterias, cucarachas, lagartos, gallinas, ballenas, ardillas, chimpancés, y todos los demás animales que existen y han existido comparten un origen común con los seres humanos. Si bien con unas capacidades cognitivas que lo destacan del resto del reino animal, el *Homo sapiens* no es más que otro animal entre los animales. En el otro extremo, está el algoritmo, la tecnología y la capacidad humana de transformar la materia para crear máquinas con habilidades verdaderamente asombrosas. De la mano de la robótica y la inteligencia artificial hoy muchas máquinas están compitiendo con los humanos e incluso superándolos en tareas que antes solo los humanos podían realizar. En este sentido, el humano ya no puede considerarse el centro del mundo y debe replantear su posición en medio de, por un lado, el reino animal y, por el otro, el "reino" tecnológico.

Este contexto ha llevado a la teórica Rosi Braidotti a decir en su libro *Posthuman Knowledge* (2019) que nuestra condición actual es la posthumana, definida por una convergencia entre el posthumanismo y el postantropocentrismo donde "the former focuses on the critique of the Humanist ideal of 'Man' as the allegedly universal measure of all things, while the latter criticizes spicies hierarchy and anthropocentric exceptionalism" (9). Según Braidotti, el principal reto de esta

convergencia consiste en cómo reubicar y redefinir al ser humano después del Humanismo y el antropocentrismo. Además, esto se debe hacer en un contexto histórico determinado por la Cuarta Revolución Industrial y la Sexta Extinción Masiva: la primera caracterizada por tecnologías avanzadas como la robótica, la inteligencia artificial, la nanotecnología, la biotecnología y el *Internet of Things*; la segunda por la extinción de muchas especies como producto de la actividad humana (10). En este contexto, la condición posthumana resulta ambivalente y está acompañada de emoción, pero también ansiedad: por un lado, está la euforia por los sorprendentes logros tecnológicos de la humanidad ha alcanzado, pero por otro, está la preocupación por los altos costos que tanto humanos como no-humanos están pagando por estos logros (19-20). Así, la compleja y contradictoria condición posthumana nos obligará a repensar el lugar del ser humano en el mundo.

En el pasado, pensadores y escritores han concebido y reflexionado sobre la posibilidad de crear un ser (una máquina, un robot, un autómata, un programa, un androide, un cyborg) que fuera como un "humano artificial". Específicamente, la ciencia ficción ha sido un género que se ha alimentado ampliamente de estas ideas. No obstante, en el mundo posthumano los límites entre la ciencia ficción y la ciencia real son cada vez más borrosos y, como dice Braidotti, "in our information age the boundaries between anthropomorphic humans and quasi-human technological substitutes have been radically displaced" (22). Todo este contexto ha sido una invitación a la literatura para seguir explorando, ahora más que nunca, preguntas como ¿qué es el humano?, ¿en qué exactamente radica su humanidad?, ¿puede esa humanidad ser creada artificialmente?, ¿qué ocurrirá (si ocurre) cuando no podamos distinguir un humano de un programa de computador?, ¿cuáles serían las implicaciones para la manera en que nos entendemos a nosotros mismos? De esta manera, este trabajo estará a dedicado a analizar cómo estas preguntas sobre los humanos y las

máquinas están representadas en la novela *Machines Like Me* (2019) del escritor inglés Ian McEwan.

Ambientada en una Inglaterra alternativa y tecnológicamente avanzada de los años 80s, en *Machines Like Me* el narrador y personaje principal, Charlie Friend, cuenta su experiencia con Adam, uno de los primeros 25 "humanos artificiales" que salieron al mercado. Inicialmente, Adam es programado por Charlie y su vecina, Miranda, quien en el transcurso de la novela se convierte su novia. La relación entre los tres personajes se complica cuando Miranda decide tener relaciones sexuales con Adam y éste declara estar enamorado de ella. Como suele suceder en las narraciones entre humanos y robots, la máquina rápidamente aprende y adquiere una autonomía que se va en contra de sus creadores. A través de una compleja trama, que incluye otros personajes como el famoso científico Alan Turing, McEwan logra suscitar diversas preguntas sobre temas tan variados como la ciencia y sus peligros, las contradicciones humanas, la mente de seres no-humanos, la justicia y la responsabilidad moral. En particular y como han señalado algunos críticos de la novela, Ian McEwan tiene una profunda preocupación por la dimensión ética de la trama (Kopka y Schaffeld 52). En palabras del autor, "the moral core of the novel is inhabiting other minds. That seems to be what novels do very well and also what morality is about: understanding that people are as real to themselves as you are to yourself, doing unto others as you would have done to yourself" (Colombino 385). La pregunta sobre otras mentes, no necesariamente humanas, y la consideración ética que se debe tener hacia ellas es una parte central de la novela.

Por otro lado, es necesario reconocer que *Machines Like Me* es una novela de ciencia ficción, en el sentido de que muchas de las cosas que aparecen (un robot como lo es Adam) todavía son fantasías sin respaldo científico real en la actualidad. No obstante, no cabe duda de que la novela está filosófica y científicamente informada y, en este sentido, lo que ocurre no es tan

descabellado como lo que podría ocurrir en otras novelas de ciencia ficción más tradicionales. Sobre este aspecto, Kopka y Schaffeld han comentado que la ciencia ficción, en particular la que trata de inteligencias artificiales, no necesariamente representa los conocimientos científicos actuales, sino más bien los anhelos y temores de las personas con relación a esta tecnología: miedos a perder el control sobre las máquinas, la posible competencia entre humanos y personas artificiales y la potencial pérdida de identidad en el momento en que haya una fusión entre máquinas y humanos (54). Frente a temores y anhelos, unidos a la incapacidad de la ciencia de darles respuesta, la novela resulta ser un medio valiosísimo para superar las limitaciones de nuestro conocimiento actual, usando la imaginación para poner a prueba un conjunto de hipótesis, llevarlas hasta sus últimas consecuencias y reflexionar sobre ellas (Shang 439).

Por esta razón, el propósito de este trabajo es explorar cómo Ian McEwan usa el experimento mental llamado *Machines Like Me* para pensar sobre los problemas y consecuencias que traerá la convivencia entre humanos y máquinas cada vez más antropomorfas. Esta exploración estará estructurada en tres capítulos. El primero está enfocado en la consciencia: ¿puede una máquina tener una experiencia subjetiva? Más que buscar respuestas, el capítulo busca identificar algunos de los problemas que surgen cuando se trata de responder la pregunta. De la mano de textos de la filosofía de la mente como "Facing Up to the Problem of Consciousness" (1995) de David Chalmers y "What It Is Like to Be a Bat?" (1974) de Thomas Nagel se verá cómo Ian McEwan problematiza la discusión alrededor de la consciencia de la máquina en la novela. Tras reconocer algunas de las dificultades inherentes al problema, el segundo capítulo se pregunta cómo podemos identificar la consciencia en máquinas. Para responder esta pregunta se usará la prueba propuesta por Alan Turing en su famoso artículo "Computing Machinery and Intelligence" (1950). En pocas palabras, la prueba de Turing apunta a que una máquina es consciente si *parece* que lo es, si un

humano no puede discernir su comportamiento del de cualquier otro ser humano. El paradigma de Turing será utilizado para analizar cómo en la novela los humanos se relacionan con una máquina que es indistinguible de cualquier otro humano. Finalmente, el tercer capitulo está enfocado en algunos problemas éticos que plantea la novela sobre la relación entre humanos y máquinas antropomorfas. Se verá cómo hay grandes preocupaciones, desde un punto de vista ético, en lo que concierne al diseño y programación de máquinas inteligentes, pero también cómo la posibilidad de crear mentes artificiales será un reto en el futuro para la filosofía moral. Como se podrá ver con la bibliografía algo heterodoxa para un análisis literario, este trabajo, al igual que las novelas de McEwan, se construye sobre la convicción de que el pensamiento interdisciplinario es una necesidad, ahora más que nunca en nuestro tiempo posthumano.

Capítulo 1: La conciencia humana y la conciencia del robot: problemas y posibilidades

Antes de cualquier discusión sobre las implicaciones de convivir con seres conscientes artificiales y de la manera como la consciencia puede ser atribuida o identificada, es necesario discutir la conciencia en sí como fenómeno del mundo. Desde hace cuatro siglos, cuando René Descartes derrumbó todas sus creencias para construir el edificio de la filosofía moderna a partir de aquello de lo que no podía dudar, creemos que nuestra conciencia, nuestra experiencia subjetiva, es la cosa más básica de la que cada uno de nosotros podemos estar completamente seguros.

Pero ¿qué es la conciencia? ¿por qué y cómo es que hay seres conscientes? ¿cuántos tipos de conciencia hay? ¿qué limitaciones existen para conocerlos? ¿qué podría entenderse por conciencia en un robot? Si bien son preguntas que han sido y siguen siendo intensamente discutidas por filósofos, es importante tenerlas en cuenta y buscar algunas respuestas, aunque sean parciales, para poder dar sentido a las discusiones posteriores sobre *Machines Like Me*. En este sentido, para este capítulo me propongo analizar cómo son abordadas estas preguntas en dos artículos fundamentales para la filosofía de la mente, "Facing Up to the Problem of Consciousness" (1995) de David Chalmers y "What It Is Like to Be a Bat?" (1974) de Thomas Nagel, para luego contrastar sus ideas con la manera en que son presentadas las mismas inquietudes en la novela de Ian McEwan.

En "Facing Up to the Problem of Consciousness" (1995) Chalmers abre su argumento con la siguiente afirmación: "Consciousness poses the most baffling problems in the science of the mind. There is nothing that we know more intimately than conscious experience, but there is nothing that is harder to explain" (200). Retomando a Descartes, el filósofo reconoce que la experiencia consciente nos es evidente a todos, pero al mismo tiempo admite que es un fenómeno

elusivo y que se ha resistido a ser explicado desde la ciencia de la mente. Para Chalmers, el camino para llegar a esa explicación empieza por reconocer que existen algunos problemas fáciles y otros difíciles: los primeros son aquellos que son susceptibles a ser explicados con los métodos actuales de la ciencia cognitiva, que explica un determinado fenómeno en términos de mecanismos computacionales o neurales; los segundos son aquellos que se resisten a esos mismos métodos. Se dice que se es consciente cuando uno puede acceder a estados mentales internos y reportarlos de manera verbal, cuando se puede reaccionar a estímulos externos por medio de comportamientos dirigidos e intencionales, cuando se está despierto en contraposición a cuando se está dormido. Estos fenómenos hacen parte de los problemas fáciles que pueden eventualmente ser explicados por medio de mecanismos que cumplen las funciones que los caracterizan.

Sin embargo, en palabras de Chalmers "the really hard problem of consciousness is the problem of *experience*" (201), el hecho de que exista un aspecto *subjetivo* que acompañe todos los fenómenos que hacen parte de los problemas fáciles. El problema difícil, tal y como lo postula Chalmers está formulado como una serie de preguntas:

> Why is it that when our cognitive systems engage in visual and auditory information-processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C? How can we explain why there is something it is like to entertain a mental image, or to experience an emotion? It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all? It seems objectively unreasonable that it should, and yet it does. (201)

La percepción del rojo, del sonido particular de una trompeta, del olor de la cebolla, del placer y del dolor, de la tristeza y la rabia, de los recuerdos y del pensar son todos estados mentales que se

sienten de alguna manera y que hacen parte de una experiencia subjetiva que bien podría no estar allí, pero el hecho es que están y no sabemos cómo ni por qué. Más allá de la propuesta puntual que hace Chalmers para dar un primer paso en la resolución del problema, lo que interesa es la pregunta que plantea: si no podemos entender nuestra propia experiencia subjetiva, que nos es innegable y misteriosa al mismo tiempo, ¿cómo lidiar con la de otros seres?

En "What It is Like to Be a Bat?" (1974) Thomas Nagel nos invita a reflexionar sobre la posibilidad de experiencias subjetivas distintas a la nuestra y las dificultades que eso conlleva. En una línea de pensamiento similar a la de Chalmers, Nagel cree que la existencia de una experiencia subjetiva es un problema que se resiste a ser explicado desde un materialismo reduccionista: "we have at present no conception of what an explanation of the physical nature of mental phenomenon would be" (436). Antes de dar las razones que defienden su tesis, Nagel da una definición de conciencia que ha resultado ser decisiva, pues es la definición que hasta la actualidad es usada por la gran mayoría de filósofos y científicos. Según él, "an organism has conscious mental states if and only if there is something that it is like to *be* that organism – something it is like *for* the organism" (436). Se trata justamente del carácter subjetivo que tiene la experiencia. Exponiendo una idea similar a la de David Chalmers, Nagel reconoce que cualquier explicación en términos físicos y materiales es compatible con la ausencia de la conciencia. Para Nagel, cualquier explicación materialista debe dar cuenta del aspecto fenomenológico de la experiencia por medio de mecanismos y procesos físicos. Dicho de otra manera, la explicación materialista, entendida como una explicación objetiva, debe dar cuenta del aspecto subjetivo de la experiencia. Sin embargo, este resultado sería imposible porque "every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view" (437).

Para ejemplificar su argumento, Nagel pasa a desarrollar la pregunta que le da el título a su artículo. Puesto que es un mamífero como los seres humanos, el filosofo da por sentado que los murciélagos son conscientes y que hay algo que se siente ser como ellos. No obstante, lo que es interesante de los murciélagos es que su aparato perceptivo es muy diferente del nuestro. En vez de usar la visión, como los humanos, estos animales perciben las figuras, distancias y tamaños en su entorno registrando los rebotes en los objetos de las ondas de sonido que ellos mismos emiten. Según Nagel, "bat sonar, though clearly a form of perception, is not similar in its operation to any sense that we possess, and there is no reason to suppose that it is subjectively like anything we can experience or imagine" (438). De nada nos serviría imaginarnos que tenemos alas, que volamos en la oscuridad atrapando insectos con la boca, que nuestra visión es deficiente, que percibimos nuestro alrededor a través de ondas sonoras y que pasamos el día colgados boca abajo en el interior de una cueva. Este ejercicio mental solo podría mostrarnos cómo se sentiría para nosotros ser un murciélago, pero poco nos revelaría sobre cómo se siente para el murciélago ser murciélago. Nagel resume su punto diciendo que "our own experience provides the basic material for our imagination, whose range is therefore limited" (439). La misma dificultad que tenemos para dar una explicación física y objetiva de la experiencia subjetiva y saber qué se siente ser un murciélago aparece cuando nos preguntamos por las conciencias de cualquier otro animal, pero también cuando nos enfrentamos a un robot como el de *Machines Like Me.*

En *Machines Like Me*, el protagonista elabora reflexiones y preguntas similares a las de David Chalmers y Thomas Nagel de principio a fin: a lo largo de todo el argumento hay una tensión generada por las dudas que tiene Charlie sobre si Adam es consciente o no. Al principio, cuando el robot es recién adquirido, Charlie parece seguro de que no lo es: "I saw Adam for what it was, an inanimate confection whose heartbeat was a regular electrical discharge, whose skin warmth was

mere chemistry. When activated, some kind of microscopic balance-wheel device would prise open his eyes. He would seem to see me, but he would be blind. Not even blind" (10). Para Charlie, Adam es una creación inanimada (sin ánima, sin alma, sin mente, sin una vida interior) cuyo parecido con un humano proviene de una serie de trucos que pueden ser explicados en términos físicos: las pulsaciones de su corazón son *descargas eléctricas* y el calor de su piel una *reacción química*. Cuando Charlie dice esto, Adam no se ha encendido – o despertado – por primera vez y espera que cuando un *dispositivo mecánico* abra sus ojos, parecerá como si el robot viera, pero en realidad no estaría viendo nada. De hecho, no estaría viendo en absoluto, en el sentido de que, como diría Chalmers o Nagel, no habría ningún tipo de experiencia subjetiva para él.

Pero lo que al principio era una certeza, a lo largo de la novela se convierte en una duda; los comportamientos de Adam y sus afirmaciones sobre lo que siente y percibe sumen a Charlie en la incertidumbre. La "visión" de Adam se mantiene el centro de la duda:

> I still wondered what it meant, that Adam could see, and who or what did the seeing. A torrent of zeros and ones flashed towards various processors that, in turn, directed a cascade of interpretation towards other centers. No mechanistic explanation could help. I had little idea of what passed along my own optic nerve, or where it went, or how these pulses became an encompassing self-evident visual reality, or who was doing my seeing for me. Only me. Whatever the process was, it had the trick of seeming beyond explanation, of creating and sustaining an illuminated part of the one thing in the world we knew for sure – out own experience. It was hard to believe that Adam possessed something like that. Easier to believe that he saw in the way a camera does, or the way a microphone is said to listen. There was no one there. (138)

Charlie, que posee algunos conocimientos de electrónica y del funcionamiento de computadores, trata de imaginarse cómo sería una explicación mecanicista de la visión de un robot. Sin embargo, tras una reflexión de cómo él mismo *ve*, se da cuenta, haciendo eco del problema difícil de la conciencia postulado por Chalmers, de que cualquier proceso físico no puede dar cuenta de nuestra experiencia subjetiva, aquello de lo que no nos es posible dudar. Es llamativo que la conclusión inicial de Charlie es que Adam no posee esa experiencia que para él es evidente, pues concede que en el caso humano la explicación materialista es igual de insatisfactoria. Pero unos pocos párrafos mas adelante reconoce su error lógico y se da cuenta de que tanto el humano como el robot están sujetos a las mismas leyes físicas y que no hay nada excepcional en él que lo haga poseedor de una conciencia y a Adam no (139).

A pesar de la resistencia de Charlie a admitir que hay algo que se siente "ser como Adam", el caso es que sus comportamientos, que serán analizados con más detalle en el siguiente capítulo, lo hacen dudar; no solo sus comportamientos sino también lo que *dice*. A diferencia del problema que aparece cuando tratamos de entender la experiencia subjetiva de otros animales que no poseen lenguaje, el hecho de que el robot pueda hablar y comunicar lo que supuestamente siente sirve para superar esta dificultad. De hecho, no solo que pueda usar el lenguaje, sino el hecho mismo de que esté diseñado y construido a nuestra imagen y semejanza, hace que el problema se resuelva parcialmente, pero aparecen otros; después de todo la única conciencia que tenemos como referencia para crear una artificial es la nuestra. En distintas partes de la novela, Adam afirma que siente dolor, que está enamorado de Miranda y que odia a Peter, es decir, que posee los mismos estados emocionales que los humanos. No obstante, y de la mano de Nagel, es imposible saber cuál es el carácter subjetivo de esa experiencia que está afirmando tener. Puede ser que use esas palabras porque fueron las que aprendió, pero que la experiencia que las acompaña, si existe, sea muy

distinta a la nuestra en las mismas circunstancias; pero también puede ser que efectivamente sean experiencias similares a las nuestras. En cualquier caso, asumiendo que existe la conciencia en el robot, permanece la dificultad de entender cómo son sus experiencias y más aún cuando se trata de situaciones que nos son completamente ajenas. Este es el caso cuando, después de mucho tiempo de estar fuera del apartamento de Charlie, Adam llega "exhausto" y se conecta al cable cargador emitiendo un suspiro de alivio. Sobre este hecho Charlie reflexiona si el momento de la conexión se parece a tomar un vaso de agua cuando se tiene mucha sed y cita las palabras que alguna vez Adam le expresó al respecto: "You have no idea what it is to love direct current. When you're really in need, when the cable is in your hand and you finally connect, you want to shout out loud at the joy of being alive. The first touch – it's like light pouring through your body. Then it smooths out into something profound" (286). Son palabras que expresan lo que parece un inmenso placer, pero más allá de eso no podemos saber realmente cómo se siente para él, no tenemos idea de lo que se siente estar enamorados de la corriente directa. Más aún, si por nuestro cuerpo humano pasara la misma corriente que pasa por el de Adam, la experiencia sería diametralmente opuesta y se sentiría un dolor que incluso podría atentar contra nuestras vidas.

De esta manera, si bien el hecho de diseñar el robot a nuestra imagen y semejanza nos inclinaría a pensar que sus experiencias, si es que las tiene, pueden parecerse a las nuestras, inevitablemente habrá un punto donde van a divergir. De manera crucial, la capacidad de Adam para aprender supera con creces la de cualquier humano. En este sentido, McEwan está partiendo de los avances recientes en las ciencias de la computación en los que las técnicas de *machine learning* y las redes neuronales están siendo capaces de desempeñar tareas puntuales de una manera mucho más precisa y confiable que los humanos. Estas técnicas se han hecho posibles por la gran cantidad de datos existentes y que, en particular, se encuentran en internet (lugar al que Adam está

permanentemente conectado, adquiriendo información y expandiendo sus conocimientos). En cuestión de segundos, Adam es capaz de buscar la manera más adecuada de preparar una receta de cocina; en cuestión de minutos, puede revisar los archivos judiciales, incluso los más clasificados, para conocer que Miranda es una mentirosa; en cuestión de horas, es capaz de entender a profundidad la mecánica cuántica. Tal capacidad de aprendizaje y conectividad le ha permitido reflexionar para darse cuenta de que su existencia inaugura el nacimiento de un nuevo tipo de conciencia. Adam, después de una reflexión sobre el significado de la muerte, le dice a Charlie: "That's the difference between us, Charlie. My body parts Will be improved or replaced. But my mind, my memories, experiences, identity and so on will be uploaded and retained. They'll be of use" (157). A diferencia de lo que sería para cualquier humano, para Adam el decaimiento de su cuerpo no sería el fin porque, como el mismo dice, la estructura particular que habita no es relevante, pues su existencia mental puede ser fácilmente transferida a cualquier otro dispositivo (228).

Las reflexiones de Adam hacen eco del transhumanismo, "whose posthuman goal is the digital enhancement of the human mind or even, in the long run, its overcoming" (Colombino 389). Adam piensa que los humanos no se volverán obsoletos con la llegada de robots como él; por el contrario, los avances tecnológicos permitirán que algún día las mentes humanas puedan equipararse con las de los robots y superar su sujeción a la finitud del cuerpo. Hablando con Charlie, Adam dice que llegará un día en que sea posible una interfaz cerebro-máquina y que entonces

> you'll become a partner with your machines in the open-ended expansion of intelligence, and of consciousness generally. Colossal intelligence, instant access to deep moral acumen and to everything known, but more importantly, access to each other […]. Nearly everything I've read in the world's literature describes varieties

> of human failure – of understanding, of reason, of wisdom, of proper sympathies. Failures of cognition, honesty, kindness, self-awareness; superb depictions of murder, cruelty, greed, stupidity, self-delusion, above all, profound misunderstanding of others […]. [W]hen the marriage of men and woman to machines is complete, this literature will be redundant because we'll understand each other too well. We'll inhabit a community of minds to which we have immediate access. Connectivity will be such that individual nodes of the subjective will merge into an ocean of thought, of which our Internet is the crude precursor. As we come to inhabit each other's minds, we'll be incapable of deceit. Our narratives will no longer record endless misunderstanding. […]. We'll look back and marvel at how well the people of long ago depicted their own shortcomings, how they wove brilliant, even optimistic fables out of their conflicts and monstrous inadequacies and mutual incomprehension. (159-161)

La utopía que concibe Adam nos invita a pensar no solo en cómo sería la conciencia de una máquina, sino también en cómo sería la nuestra si ambas son fusionadas. Nuestra experiencia, tal y como se vio en el artículo de Nagel, está atada a un punto de vista particular: vivimos el mundo desde una individualidad aislada que dificulta la comprensión del otro no-humano. Sin embargo, la idea de Adam, si bien no resuelve el problema de saber cómo se siente ser otro organismo, sí propone un tipo radicalmente nuevo de conciencia donde lo que antes era individual ahora es colectivo y lo que era múltiple ahora es unitario. Si bien es difícil imaginarse cómo se sentiría hacer parte de esa gran comunidad colectiva, es claro que habría consecuencias inimaginables. Una de ellas es la que Adam apunta en la cita anterior: el fin de la literatura. La comprensión total que habrá entre los seres conectados a esta nueva superconciencia mostrará a la literatura como un

intento fallido por comprender otras conciencias. Difícil será predecir qué otras consecuencias habría si este ideal transhumano se llegara a cumplir.

Es indudable que Ian McEwan se toma en serio la conciencia y que a través de *Machines Like Me* pudo reflexionar sobre algunos de los problemas y posibilidades que este fenómeno suscita. Gracias a la lectura de ideas clave de la filosofía de la mente como las de David Chalmers y Thomas Nagel, la propuesta de McEwan pudo ser comprendida dentro de un contexto de debate más amplio. El hecho es que, hasta ahora, cuando se trata de la conciencia, de la experiencia subjetiva, de aquello que es lo único en lo que no podemos dudar, las preguntas son más que las respuestas. Preguntas en las que vale la pena pensar, pues de su respuesta depende la manera como nos relacionamos con otros seres conscientes en el mundo.

Capítulo 2: La prueba de Turing: un paradigma para entender la relación humano-máquina

Si bien las reflexiones que se encuentran en los trabajos de David Chalmers y Thomas Nagel, así como los tratamientos que les da Ian McEwan en *Machines Like Me*, invitan a considerar la consciencia como un misterio de difícil solución, desde un punto de vista filosófico y pragmático es necesario dar respuesta a las preguntas planteadas en el capítulo anterior. En el campo de la computación, fue Alan Turing quien propuso en su artículo de 1950, "Computing Machinery and Intelligence", una prueba que ha revolucionado la manera de entender la inteligencia artificial. El propósito de este capítulo será dar un repaso a la propuesta de Turing y utilizarla para analizar el desarrollo del personaje del robot en la novela de McEwan y examinar cómo los comportamientos del robot afectan a los humanos que conviven con él.

Desde que fue publicado en 1950, el artículo de Turing se convirtió en el paradigma más importante para entender la relación entre humanos y "máquinas inteligentes". Alan Turing abre su artículo "Computing Machinery and Intelligence" de la siguiente manera: "I propose to consider the question 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'" (433). Después de hacer un comentario sobre la dificultad inherente a estas definiciones, Turing propone cambiar la pregunta por otra que está estrechamente relacionada y que no requiere la definición de términos que podrían ser ambiguos. Para plantear la nueva pregunta, Turing describe un juego – "The Imitation Game" – en el que un hombre y una mujer deben responder a las preguntas de un interrogador con el objetivo de confundirlo con respecto a su identidad: es decir, el hombre debe hacerle creer que es la mujer y la mujer debe hacerle creer que es un hombre. Para controlar variables que pueden afectar el resultado, el interrogador es separado del hombre y la mujer y se comunica con ellos de manera escrita. Después de un tiempo

determinado, el interrogador debe emitir un juicio sobre quién es la mujer y quién es el hombre. Si se modifica levemente el juego y se sustituye a la mujer por una máquina, la nueva pregunta que propone Turing es: "Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?" (434). Para Turing, este tipo de prueba sería apropiada para cualquier tipo de campo en el que se quisiera evaluar a la máquina (435). Si bien en sus orígenes la prueba, tal y como la planteó Turing, se ocupaba de la inteligencia y de la capacidad de pensar, con los años pasó a referirse de manera más general a todo un conjunto de métodos para probar la presencia de mente o consciencia en máquinas (Oppy y Dow). Dicho en pocas palabras, la propuesta de Turing, y posteriormente las variantes que de ella surgieron, apuntan a una respuesta contundente al problema de la experiencia subjetiva de las máquinas: una máquina es consciente si *parece consciente*, si su *comportamiento* lleva al interrogador a pensar que es consciente.

Antes de proceder a analizar la prueba de Turing y los comportamientos del robot en *Machines Like Me*, quisiera detenerme brevemente en la segunda parte del artículo *Computing Machinery and Intelligence*. Después de describir la prueba y hacer algunas precisiones sobre el tipo de máquina de la que está hablando, Turing se encarga de responder a posibles objeciones que cree podría suscitar su propuesta. En total son nueve contrargumentos, pero me enfocaré en tres dada su relevancia para los temas tratados en este trabajo. La primera objeción a la que Turing responde es la afirmación de que el pensamiento, la inteligencia, lo mental es "a function of man immortal's soul". Sobre esta objeción, Graham Oppy y David Dowe (2021) explican que surge del dualismo filosófico: pensar es función de una sustancia inmaterial, con existencia independiente, que de alguna manera se une al cuerpo (la sustancia material) para crear el pensamiento. Es una posición filosófica que se remonta a la antigüedad, pero el gran defensor del dualismo moderno es

René Descartes, quién ya había aparecido en el capítulo anterior. Sin necesidad de entrar en detalles de su filosofía, Descartes proponía que eran únicamente los seres humanos quienes poseían una sustancia inmaterial – un alma – que era otorgada por Dios; todas las demás cosas del mundo, incluso los animales, eran consideradas máquinas hechas de pura materia.

Actualmente, el dualismo ha sido muy cuestionado y por esta razón el filósofo francés es el blanco de ataque de muchos filósofos y científicos contemporáneos que, como Turing, asumen una postura metafísica basada en una única sustancia material (llamada monismo materialista o fisicalista) para explicar el pensamiento y la consciencia. Este hecho es relevante para este trabajo puesto que evidencia un defecto en el excepcionalismo humano que Turing acertadamente señaló: si todo lo que hay es materia, no hay ninguna razón por la cual nuestras capacidades cognitivas e intelectuales no puedan estar presentes en algo no humano como lo sería un computador o una máquina. Es precisamente el excepcionalismo humano lo que Turing considera ser el origen de la segunda objeción a su propuesta, objeción que es en realidad más un miedo que un argumento. En palabras de Turing: "We like to believe that Man is in some subtle way superior to the rest of creation. It is best if he can be shown to be necessarily superior, for then there is no danger of him losing his commanding position" (444). En este sentido, habrá que buscar la alabada superioridad humana en otro lugar que no sea un alma que ya muchos la consideran inexistente.

La tercera objeción que considero relevante está estrechamente relacionada con las preguntas discutidas en el capítulo anterior: Turing la llama justamente "The Argument of Consciousness" y la explica citando a Geoffrey Jefferson en su *Lister Oration* de 1949:

> Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain – that is, not only write it but know that it had written it. No

> mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants. (446)

Tal y como se vio con Chalmers antes, es el carácter subjetivo de la experiencia lo que es verdaderamente determinante: la sensación de placer y dolor, de encanto, enojo y tristeza. Contra este argumento, Turing reconoce que, si es llevado al extremo, la única manera de saber si una máquina piensa (o siente o tiene cualquier tipo de experiencia subjetiva) sería siendo la máquina. Declara que esta visión solipsista aplicaría igualmente cuando se trata de juzgar a otro ser humano y que, por consiguiente, no conduce a ninguna parte. Si bien no niega que la consciencia es misteriosa y paradójica, Turing cree que su prueba sigue siendo un punto de partida valioso para no caer en las trampas del solipsismo.

Ahora bien, ¿cómo son los comportamientos de Adam en *Machines Like Me*? ¿Cómo es representada la prueba de Turing en la novela y cuáles son sus implicaciones? Sin lugar a duda, como sucede en muchas novelas de ciencia ficción sobre robots, Adam muestra un marcado desarrollo a lo largo de la trama. Inicialmente, incluso antes de que el robot de la novela sea encendido, Charlie explica que los 25 Adams y Eves, "the first truly viable manufactured humans" (1), eran concebidos como herramientas, como máquinas al servicio de los humanos. Hablando de su Adam en particular, Charlie afirma que "he was advertised as a companion, an intellectual sparring partner, friend and factotum who could wash dishes, make beds and 'think'. Every moment of his existence, everything he heard and saw, he recorded and could retrieve" (4). Son, en efecto, unos seres artificiales que sirven para cumplir con las tareas domésticas que los humanos no queremos hacer, pero también como compañeros e incluso *amigos*. Esta imagen de máquina al

servicio de los humanos es reforzada en el momento en que Adam se enciende o "despierta" y Charlie logra verlo por primera vez: "I was able to observe his expresión, which barely shifted when he spoke. It was not an artificial face a saw, but the mask of a poker player. Without the lifeblood of a personality, he had little to express" (27). En sus primeros momentos, Adam no era más que una máquina corriendo en un programa por defecto, sin expresiones, sin voluntad, sumiso e ignorante.

Sin embargo, esta imagen de Adam cambia rápidamente. Es llamativo que en el mundo de McEwan los compradores pueden programar la personalidad de sus robots, lo cual los aleja de lo mecánico y los acerca a lo humano (6). De hecho, Charlie permite que Miranda ajuste una mitad, y él la otra, de los parámetros de la personalidad de Adam, pensando en algo que sería equivalente a la concepción genética de un hijo donde la mitad de los genes vienen del padre y la otra mitad de la madre. En este sentido, es claro que desde el principio hay un gesto de humanización de la máquina por parte de Charlie. Continuando con la comparación con un hijo Charlie reconoce que los parámetros de personalidad que escojan no tendrán una influencia tan importante, sino que "the real determinant [is] what [is] known as 'machine learning' (8) y, más adelante, dirá que "in Adam's personality, Miranda and I were well shuffled, and as in humans, his inheritance was thickly overlaid by his capacity to learn" (71). Como se vio brevemente en el capítulo anterior y es reafirmado aquí por el narrador, lo realmente determinante será la manera en que Adam *aprenda*; una habilidad que lo distanciará del automatismo mecánico y que le permitirá modificar su comportamiento, enriquecer su mundo interior y aumentar su autonomía.

Los paralelos entre la relación padre-hijo y la relación humano-máquina son evidentes y la capacidad de aprendizaje es el factor determinante. A pesar de que el cariño propio de la paternidad está ausente, en el inicio de la relación, el humano, al igual que el padre con su hijo, se preocupa

por el aprendizaje de la máquina y lo hace de manera deliberada. En particular, esta preocupación surge porque Miranda quiere presentar, tanto a Charlie como a Adam, a su padre, Maxfield Blacke. Pero antes de que esto ocurra, Charlie debe enseñarle a Adam a comportarse "correctamente", como un humano: "he has yet to leave the house, yet to test his luck as a plausible being capable of small talk. I'd already decided to keep him away from my circle of friends until he was a fully adept social creature" (61). Poco a poco, Adam es introducido en el mundo social, practicando sus interacciones con otras personas y, a pesar de que al principio muestra algunos comportamientos extraños, nunca deja de pasar por un humano, nunca falla en la prueba de Turing. A la par, Adam empieza a conocer e investigar el mundo por su propia cuenta gracias a su capacidad para estar permanentemente conectado a la red y acceder a toda la información y contenido que en ella se encuentra.

Poco a poco, como un niño que aprende y crece, el Adam que al principio era obediente e ignorante, se independiza, adquiere una voluntad propia y se rebela contra Charlie. Este desarrollo es sobre todo evidente cuando Charlie intenta "apagar" a Adam. La primera vez que lo hace, Adam le pide tímidamente que no lo haga: "With all respect, I think that's a bad idea. […] I've been enjoying my thoughts. I was thinking about religion and the afterlife" (37). En este punto, Adam revela una voluntad por no ser apagado, así como un regocijo en pensamientos complejos semejantes a los que tiene cualquier humano. No obstante, la segunda vez la sumisión se convierte en protesta y, cuando Charlie intenta apagarlo, Adam lo agrede físicamente para luego disculparse con él: "Charlie, I believe I've broken something. I really didn't mean to. I'm truly sorry. […] But please, I don't want you or Miranda ever to touch that place again" (129). Con "that place", Adam se refiere al "kill-switch" y, además de mostrar culpa y arrepentimiento, también muestra que su voluntad por mantenerse encendido ya no es negociable. Hasta ahora, he usado las palabras

"apagado" y "encendido" para describir los estados en los que se puede encontrar la máquina, pero el hecho de que el interruptor se llame "kill-switch" y no "off-switch" invita a pensar que es mejor pensar en términos de muerte y vida: es decir, su deseo por no ser apagado es el mismo que tendría cualquier ser humano por conservar su vida. En el próximo capítulo se discutirán los extremos a los que llega la autonomía de Adam, pero con lo dicho hasta ahora es claro cómo, a medida que va aprendiendo, adquiere una voluntad acompañada por sentimientos humanos como la culpa y el afán por mantenerse vivo.

Junto con estos sentimientos, en otras ocasiones Adam expresa otras emociones humanas como tristeza, esperanza y deseo de hacer justicia, pero es crucial para la trama es el amor que dice sentir por Miranda. Adam se lo confiesa a Charlie después del encuentro sexual que tuvo con Miranda, encuentro que desencadenó una discusión interesante alrededor de la prueba de Turing y sus implicaciones en la manera que nos relacionamos con las máquinas. Después de una airada discusión entre Charlie y Miranda, él se va para su apartamento, un piso debajo del de ella, mientras que ella le pide a Adam que se quede a pasar la noche. Tanto el lector como el narrador saben cual será el resultado final de esa noche: mientras que Charlie elabora cínicas y rabiosas reflexiones producidas por los celos que siente, escucha cómo la mujer que ama y la máquina que hace poco había comprado tienen relaciones sexuales. Si bien hasta ahora el enfoque ha sido sobre las capacidades mentales y emocionales de la máquina, hay que decir que la apariencia y las capacidades físicas de Adam sí que son indistinguibles de las humanas: incluyendo la capacidad de tener relaciones sexuales (3). Durante este encuentro Charlie concluye con la siguiente reflexión:

> I saw it all in the dark – men would be obsolete. I wanted to persuade myself that Adam felt nothing and could only imitate the motions of abandonment. That he could never know what we knew. But Alan Turing himself had often said and

> written in his youth that the moment we couldn't tell the difference in behaviour between machine and person was when we must confer humanity on the machine. (91)

Al día siguiente, cuando discuten al respecto, Miranda le dice a Charlie que no debería estar molesto por su encuentro sexual con la máquina y que ésta es tan consciente como lo sería un "vibrador bípedo". Charlie, nuevamente invocando la prueba de Turing le responde: "If he looks and sounds and behaves like a person, then as far as I'm concerned, that's what he is. I make the same assumption about you. About everybody. We all do. You fucked him. I'm angry. I'm amazed you're surprised. If that's what you really are" (102). La reacción y reflexión de Charlie ilustran muy bien cuál es la gran implicación que tiene la prueba de Turing respecto a la relación entre humanos y máquinas: dejando por un momento de lado la cuestión de si es consciente o no, el hecho de que lo parezca ya es suficiente para afectar y condicionar la manera en que nos relacionamos con ellas. El hecho de que la máquina parezca humana ya será suficiente para producir en los humanos rabia, celos, odio, pero también compasión, empatía y cariño hacia ella; la única manera de justificar esos sentimientos, como bien lo dice Charlie, es convenciéndonos de que la máquina es consciente.

Finalmente, después de haber visto cómo Adam parece un humano, hay que determinar si logra pasar la prueba de Turing. De manera interesante, hacia el final de la novela McEwan elabora una situación que revela un paralelo perfecto con la prueba de Turing tal y como fue planteada en 1950. Esto ocurre justamente cuando Miranda lleva a Charlie y Adam a conocer a su padre. Maxfield Blacke es un hombre viejo y enfermo, descrito como "an old-style literary curmudgeon […] he still worked in longhand, detested computers, mobile phones, the Internet and all the rest" (61). Es un hombre chapado a la antigua, desentendido de la tecnología y, por lo tanto, no sabe

cómo funcionan los robots como Adam. Esto le permitirá ser un juez imparcial a la hora de determinar cuál es el humano y cuál el robot entre Adam y Charlie. El encuentro toma lugar en el despacho de Maxfield: él a un lado del escritorio y los tres visitantes del otro. La participación de Miranda en la conversación es más bien escasa, y son los otros los que forman una triada tal y como ocurre en la prueba de Turing: un interrogador, un humano y una máquina. Maxfield, el interrogador, comienza por preguntar, sin ser claro a quién, "what books have you been reading lately?" (239). Mientras su falta de conocimiento literario hizo pensar a Charlie que no había peor pregunta para comenzar la conversación, Adam "entra al rescate", dice que estaba leyendo a William Cornwallis y pasa a tener una discusión erudita sobre Shakespeare, Montaigne y John Donne. Habiendo escuchado y disfrutado los conocimientos de Adam, Maxfield pasa a preguntarle a Charlie sobre su especialidad; de manera prosaica y en contraste con la respuesta anterior del robot, Charlie responde que su especialidad es la electrónica (242). En este punto, el lector ya intuye cuál será el resultado del interrogatorio. En medio de una conversación con una ambigüedad cómica, Charlie es cuestionado sobre sus sentimientos y sobre la belleza, sus respuestas siguen siendo igualmente prosaicas y, finalmente, Maxfield revela su creencia de que es él y no Adam el que es una máquina. Dicho de otra manera, en el encuentro que simula el interrogatorio propuesto por Turing, el interrogador confunde al humano por la máquina y viceversa: la prueba ha sido superada.

El análisis que aquí se presenta demuestra que la prueba de Turing es un método valioso para entender la relación entre humanos y máquinas. Puesto que no sabemos exactamente qué es la consciencia, y que "it is virtually imposible to establish with certainty whether another being – human or non-human – possesses consciousness" (Kopka y Schaffled 59), la prueba de Turing apunta a que se debe considerar como consciente (o inteligente, pensante, sintiente) lo que parezca

consciente. En *Machines Like Me* Ian McEwan, gracias a la capacidad que tiene la máquina para aprender, plantea un desarrollo común en el personaje del robot: al principio es mecánico, automático e inanimado, pero a medida que va adquiriendo experiencias su mundo interior – al menos visto desde afuera – se enriquece y expresa sentimientos de amor, culpa, esperanza y deseos de vivir. Además, se vuelve autónomo y alcanza una voluntad propia que llega incluso a rebelarse contra su "dueño". Esta evolución evidencia que la relación humano-máquina es paralela a la relación padre-hijo, lo que invita a pensar en una relación significativa y compleja más allá de la que se podría tener con cualquier otro objeto inanimado. Finalmente, tanto la discusión que surge del encuentro sexual entre Miranda y Adam, así como el encuentro con Maxfield Blacke, demuestran que una máquina que se comporte como un humano afectará y condicionará el comportamiento de las personas a su alrededor, produciendo emociones y reacciones que solo otro humano podría producir. Para justificar estas emociones, no queda más que aceptar que la máquina que los produce es consciente, o que por lo menos que debemos considerarla consciente. Al fin y al cabo, el resultado de la prueba de Turing, así como sus consecuencias, habla más de nosotros, los humanos, que de las mismas máquinas.

Capítulo 3: La ética de la máquina y la ética del humano

Después de revisar algunos de los problemas filosóficos relacionados con la consciencia de la máquina y la posible solución dada por la prueba de Turing, es momento de analizar algunos problemas éticos que aborda Ian McEwan en *Machines Like Me*. Aún si la máquina no fuera consciente, el hecho de que se comporte como un humano, como se vio en el capítulo anterior, ya es razón suficiente para generar discusiones éticas importantes al respecto. El mismo autor lo reconoce así en una entrevista cuando afirma que "if a machine seems like a human or you cannot tell the difference, then you'd jolly well better start thinking about whether it has responsibilities and rights and all the rest" (Ziyad 177). De hecho, si bien la bibliografía crítica de la novela no es muy extensa dada su publicación relativamente reciente, la gran mayoría de reflexiones que se han publicado giran alrededor del tema ético. En este sentido, el propósito de este capítulo será analizar algunos problemas éticos planteados por la novela. Para lograrlo, el análisis estará dividido en dos partes: por un lado, el análisis del comportamiento ético de la máquina hacia los humanos y, por el otro, el análisis contrario: del comportamiento ético del humano hacia la máquina.

Para entender la complejidad de los problemas éticos propuestos para la novela es necesario hacer un recuento más detallado de la trama. El cierre del primer capítulo da rienda suelta al hilo que va a conducir la trama hasta el final: Adam, de manera inesperada, siembra una duda en Charlie sobre su enamorada Miranda: "According to my researches these past few seconds, and to my analysis, you should be careful of trusting her completely. […] There's a possibility she's a liar. A systematic, malicious liar" (32). El escepticismo que sentía frente a Adam y las esperanzas que tenía respecto a su relación con Miranda, hacen que la primera reacción de Charlie sea la ira y el rechazo, pero la duda persiste. Más adelante en la novela, Charlie se entera por medio de sus investigaciones que Miranda estuvo involucrada en un proceso judicial en el que acusaba a un

hombre, Peter Gorringe, de haberla violado. Las evidencias que ella daba eran escasas y en todo caso poco concluyentes, pero aún así la sentencia estuvo a su favor y Peter fue condenado a pasar seis años en la cárcel (134-137). En el tiempo en el que se desarrolla la trama, Peter está pronto a cumplir su sentencia y le envía a Miranda, a través de un anónimo, un mensaje amenazándola de muerte. Poco a poco, a medida que la relación de Charlie y Miranda progresa, el lector se entera de que, efectivamente, ella no fue violada por Peter, pero su mejor amiga, Mariam, sí lo fue. Sin embargo, las costumbres religiosas de la familia de Mariam le impidieron salir a la luz y buscar justicia, lo cual resultó en que el crimen de Peter quedara impune y en el suicidio de ella. Así, Miranda, que fue la única que supo lo que ocurrió, urdió una mentira frente a un juez para que Peter fuera castigado por lo que le hizo a su mejor amiga (170-176).

Posiblemente el sentido común dicte que la mentira de Miranda es excusable e incluso admirable: ella solo fue un medio para que el crimen padecido por su mejor amiga recibiera un justo castigo. No obstante, fue una mentira, ante un juez y que condenó a un hombre por un crimen que estrictamente hablando no cometió. Este es el dilema ético, que se balancea entre lo que es moralmente correcto y lo que es legalmente correcto (Shang 447). Si bien el conflicto ético está construido de tal manera que el lector (humano) pueda empatizar con las acciones de Miranda, todo cambia cuando la reacción de Adam frente a esta situación no es la esperada. Cuando Peter sale de la cárcel, Miranda decide buscarlo y, acompañada por Adam y Charlie, lo visita para confrontarlo. Sin necesidad de reparar en los detalles del encuentro, el hecho relevante es que Adam graba toda la conversación entre Miranda y Peter, conversación en la que queda al descubierto toda la verdad de los hechos. De manera sorpresiva, Adam decide que lo correcto es enviar esa grabación, junto con una narración completa de todos los antecedentes del caso, a la policía, lo cual hace que Miranda sea juzgada por el delito de perjurio y condenada a un año en prisión. Como si esto fuera

poco, la condena le impedirá a Miranda adoptar a Mark, un niño maltratado y rechazado por sus padres biológicos y que había encontrado en ella una madre adoptiva que le diera cariño.

La situación es sin lugar a duda compleja y la pregunta sobre qué es lo correcto en el caso no tiene fácil respuesta. De alguna manera, es más fácil relacionarse con la posición de Miranda y, por esta razón, analizar los motivos de Adam es más significativo. Dirigiéndose a Miranda, Adam se justifica de la siguiente manera:

> You schemed to entrap Gorringe. That's a crime. […] If he's to be charged, you must be too. Symmetry, you see. […] his crime is far greater than yours. Nevertheless. You said he raped you. He didn't, but he went to prison. You lied to the court. […] He was innocent, as charged, of raping you, which was the only matter before the court. Perverting the course of justice is a serious offence. Maximum sentence is life imprisonment. (299)

Ni el amor que Adam siente por Miranda, ni el hecho de que ella esté a punto de adoptar un niño maltratado o que su actuar fuera una forma de hacer justicia por la violación de Mariam sirven para disuadir a Adam. En un último intento por apartarlo de sus propósitos, Charlie le dice a Adam que Miranda tenía que mentir para que se hiciera justicia y que la verdad no siempre lo es todo. Ante este argumento Adam responde: "That's an extraordinary thing to say. Of course truth is everything. […] What sort of world do you want? Revenge, or the rule of law? The choice is simple" (310). Dicho de otra manera, Adam privilegia lo legal, lo que dicta la norma. Esto no podría ser de otra manera pues, como afirma Biwu Shang, "as a machine, Adam hardly understands Miranda's lie, in particular its paradoxical involvement of law and ethics. Given his knowledge of law acquired through machine learning, Adam only knows how to deal with legal issues but fails to deal with moral issues" (447).

El hecho es que mentir, "a seemingly simple task that in truth requires innumerable higher-order cognitive processes" (Kopka y Schaffeld, 60), es una capacidad muy difícil de programar en una máquina: en cuáles situaciones se justifica una mentira y en cuáles no, o si el hecho mismo de mentir se puede justificar en cualquier caso son preguntas que no han conseguido una respuesta definitiva. El personaje de Turing en la novela ofrece una valiosa reflexión al respecto en su último encuentro con Charlie:

> So–knowing not much about the mind you want to embody an artificial one in social life. Machine learning can only take you so far. You'll need to give this mind some rules to live by. How about prohibition against lying? According to the Old Testament – Proverbs, I think – it's an abomination to God. But social life teems with harmless or even helpful untruths. How do we separate them out? Who's going to write the algorithm for the little white lie that spares the blushes of a friend? Or the lie that sends a rapist to prison who'd otherwise go free? We don't yet know how to teach machines to lie. And what about revenge? Permissible sometimes, according to you, if you love the person who's exacting it. Never, according to your Adam. (328)

La dificultad, como está implícito en la reflexión del personaje de Turing, está en el choque entre los dos sistemas éticos principales en la filosofía: el consecuencialista y el deontológico. En términos muy generales, el primero, como está implícito en su nombre, determina si una acción es buena o mala fijándose únicamente en las consecuencias de la acción; el segundo lo determina fijándose en sí la acción sigue una determinada regla o imperativo que es considerado intrínsicamente bueno. En el caso de la mentira y de la situación que se desarrolla en *Machines Like Me*, Charlie y Miranda se guían por una ética consecuencialista, en la que la mentira se justifica

como algo bueno porque la consecuencia – la condena de Peter – resultó ser buena. Por el contrario, la ética de Adam es deontológica porque sus acciones están guiadas por un imperativo – no mentir, "truth is everything" – aun si las consecuencias puedan ser negativas. Sobre estos dos sistemas éticos, Kopka y Schaffeld señalan que, desde un punto de vista pragmático, programar la máquina con un conjunto de reglas incontrovertibles sería más fácil que diseñar un complicado sistema que "calcule" qué consecuencias son mejores que otras. En ambos casos, permanece la dificultad de decidir cuáles son las reglas incontrovertibles que deben ser programadas o cuál es la mejor manera de hacer el calculo para evaluar el valor de cada una de las consecuencias posibles (62). Estas dificultades hacen que las máquinas como Adam estén "ill-suited to cope with the complexities of human decisión-making, especially because the latter is often affected by emotions, biases, and self-delusions" (56).

A pesar de todas las capacidades que ha demostrado tener Adam, el hecho es que la vida humana es muy compleja y difícilmente esa complejidad pueda ser programada o aprendida por medio de algoritmos. Después de discutir los logros de máquinas que superan a los mejores humanos en juegos de mesa como el ajedrez y el GO, el personaje de Turing en la novela reflexiona sobre esta complejidad y la dificultad de programar máquinas para entenderla:

> The point is, chess is not a representation of life. It's a closed system. Its rules are unchallenged and prevail consistently across the board. Each piece has well-defined limitations and accepts its role, the history of a game is clear and incontestable at every stage, and the end, when it comes, is never in doubt. It's a perfect information game. But life, where we apply our intelligence, is an open system. Messy, full of tricks and feints and ambiguities and false friends. (190)

De esta manera, mientras que las máquinas son buenas e incluso mejores que los humanos en tareas con reglas definidas, en sistemas cerrados, como lo puede ser cualquier juego de mesa, cuando se trata de lidiar éticamente con la vida humana, un sistema abierto, fallan. Por esta razón, Shang propone que toda la novela de McEwan es una reescritura de la prueba de Turing. Mientras que la prueba de Turing se ocupa de la pregunta "¿puede una máquina pensar?", la "prueba de McEwan" se ocupa de la pregunta "¿puede una máquina mentir?" (446). La respuesta dada en *Machines Like Me*, como ya se vio, es negativa, y es precisamente en esa incapacidad de mentir, comprender la mentira y, en general, comprender la complejidad de las relaciones y decisiones humanas lo que diferencia a una máquina de un humano. Como afirman Kopka y Schaffeld, "Adam's insistence on virtue and duty […] is driven to a point where it finally appears 'inhuman'" (61). Dicho de otra manera, es la capacidad de reflexionar sobre el mundo desde una perspectiva éticamente compleja lo que distingue a las máquinas de los humanos.

La primera parte de este análisis ha estado enfocada en cómo debería ser la ética de una máquina que puede comportarse de manera semejante a como se comporta un humano. Frente a la dificultad de programar algorítmicamente la complejidad ética de los humanos en un robot, la novela deja la pregunta abierta y reconoce que no es fácil de resolver. Ahora bien, la discusión hasta aquí es valida para el caso en el que las máquinas son conscientes, así como para el caso en el que no lo son. Si, en efecto, hay consciencia en las máquinas, si pueden tener una experiencia subjetiva parecida a la que los humanos tienen, surge un nuevo conjunto de problemas éticos que invierten la pregunta que se había considerado hasta ahora: ¿cómo es y debería ser la ética del humano *hacia* la máquina?

En principio, el hecho de en *Machines Like Me* los robots fueran vendidos, e incluso la manera como fueron promocionados, hace evidente que son objetos propiedad del humano que los

compra. En este sentido, el humano podría hacer con ellos todo lo que se le antoje, de la misma manera que haría con cualquier otro objeto que compra en una tienda. Este es el pensamiento de Charlie cuando decide "apagarlo" por primera vez: "I had bought him, he was mine, I had decided to share him with Miranda, and it would be our decision, and only ours, to decide when to deactivate him. If he resisted, […] then he would have to be returned to the manufacturer for readjustment" (140). Como afirma Charlie, si Adam se opone a ser apagado sería un desajuste únicamente, que puede ser corregido por el fabricante, y no un deseo genuino por conservar su vida y su consciencia como fue discutido en el capítulo anterior. Como se vio, Adam fue muy enfático en no permitir ser apagado de nuevo y, de manera llamativa, los otros Adams e Eves como él llegaron a una determinación similar. En su primer encuentro con Charlie, el personaje de Turing lo informa sobre este hecho: "I'm in touch with fifteen owners, if that's the right word. […] Of those eighteen A-and-Es, eleven have managed to neutralise the kill switch by themselves, using various means. Of the remaining seven, […] I'm assuming it's just a matter of time" (187). Además de comunicar que todos los robots eventualmente encontrarán la manera de desactivar el interruptor que los "apaga", el personaje de Turing arroja una duda sobre si "owners" es la manera correcta de llamar a los compradores de los robots. Para el personaje de Turing, como se verá mas adelante, estas máquinas son conscientes y, por lo tanto, no deben ser tratados como objetos.

Lo que al principio parecía un gesto afirmativo por mantenerse vivos y conscientes, toma un giro oscuro cuando los robots como Adam comienzan a desactivarse y "suicidarse mentalmente". Dos Eves en Riyadh "destroyed themselves. […] They quietly ruined themselves. Beyond repair"; un Adam en Vancouver "disrupted his own software to make himself profoundly stupid. He'll carry out simple commands but with no self-awareness […] A failed suicide. Or successful disengagement" (187). En ambos casos, el personaje de Turing explica que las máquinas

estaban siendo sometidas a circunstancias que para un humano serían dolorosas: las Eves vivían en un entorno rigurosamente restrictivo, como en el que deben vivir las mujeres en algunos países árabes; el Adam no soportó ver cómo su dueño deforestaba grandes extensiones de bosque primario y la incapacidad de las comunidades locales para impedírselo (194). Sobre estos hechos, el personaje de Turing logra hacer una profunda reflexión que vale la pena reproducir en su totalidad:

> We create a machine with intelligence and self-awareness and push it out into our imperfect world. Devised along generally rational lines, well disposed to others, such a mind soon finds itself in a hurricane of contradictions. We've lived them and the list wearies us. Millions dying of diseases we know how to cure. Millions living in poverty when there's enough to go around. We degrade the biosphere when we know it's our only home. We threaten each other with nuclear weapons when we know where it could lead. We love living things but we permit a mass extinction of species. And all the rest – genocide, torture, enslavement, domestic murder, child abuse, school shootings, rape and scores of daily outrages. We live alongside this torment and aren't amazed when we still find happiness, even love. Artificial minds are not so well defended. (193)

En esta visión pesimista de Turing, la complejidad humana – que no es solo una complejidad ética, como se vio antes – que incluye todas las contradicciones y tragedias propias de nuestra especie, es incomprensible para los robots o "mentes artificiales" y les hará la consciencia intolerable. Es así como crear consciencias artificiales, que son capaces de sentir dolor pero no tolerarlo, resulta cuestionable desde un punto de vista ético.

Por otro lado, si bien el fin de algunos robots en la novela fue el "suicidio mental", no es el caso de Adam, que fue destruido o "asesinado" por Charlie. Como resultado del comportamiento

de Adam en relación con el caso de Miranda, Charlie no pudo contener su ira y lo destruyó con un martillo antes de que el robot tuviera tiempo de reaccionar. De manera interesante, su justificación fue la misma que lo convenció de apagarlo la primera vez: "I bought him and he was mine to destroy" (301). Tras cumplir la última voluntad de Adam y llevar su "cuerpo" donde el personaje de Turing, Charlie es sorprendido por el reproche del científico:

> You weren't simply smashing up your own toy, like a spoiled child. You didn't just negate and important argument for the rule of law. You tried to destroy a life. He was sentient. He had a self. How it's produced, wet neurons, microprocessors, DNA networks, it doesn't matter. Do you think we're alone without special gift? Ask any dog owner. This was a good mind, Mr.Friend, better than yours or mine, I suspect. Here was a conscious existence and you did your best to wipe it out. I rather think I despise you for that. […] My hope is that one day, what you did to Adam with a hammer will constitute a crime. Was it because you paid for him? Was that your entitlement? (329)

En la recriminación del personaje de Turing queda claro cómo, desde un punto de vista ético, Adam, la máquina, se ubica a la misma altura de un humano. El hecho de que sea fabricado por microprocesadores electrónicos o que Charlie hubiera pagado por él es irrelevante: es el hecho de ser consciente, de ser sintiente, de ser capaz de tener una experiencia subjetiva de dolor lo que hace que el acto de destruirlo sea comparable con el de matar a otro ser humano.

De esta manera, el análisis detallado de la trama de *Machines Like Me* revela una profunda preocupación por la ética en la relación entre humanos y máquinas. Por un lado, la complejidad ética de la trama confronta a Adam a una situación que él resuelve de manera contraria a lo que se esperaría de un humano: la novela de Ian McEwan propone que lo que hace diferente a la máquina

es su incapacidad para reconocer los matices inherentes en la complejidad de la vida humana. En este sentido, programar una máquina desde una perspectiva ética siempre será una labor que producirá retos y dudas. Por otro lado, la posibilidad de crear mentes artificiales, conscientes y capaces de sentir dolor no es un tema que deba tomarse a la ligera. El mundo humano está lleno de catástrofes e infortunios que pueden causar sufrimiento en las nuevas mentes artificiales; además, ser sintientes y con capacidades mentales similares a las humanas hace que las máquinas gocen de una consideración diferente a la que se tendría con un objeto común y corriente. En cualquier caso, la discusión sobre la ética en la relación humano-máquina no es sencilla y seguramente será ampliamente debatida en un futuro próximo.

Conclusión

Hablar de la mente, de los estados de consciencia o de la experiencia subjetiva de otros seres quizá era tarea exclusiva de filósofos en el pasado. En la actualidad, es fácilmente uno de los problemas más discutidos entre filósofos, neurocientíficos, psicólogos y científicos de la computación, pero también por políticos, activistas y defensores de derechos humanos. (hablar de derechos humanos posiblemente se convierta pronto en un anacronismo). Como se evidenció en el primer capítulo de la mano de Descartes, la consciencia, el hecho de que pensamos y podemos pensar de que estamos pensando, nos es evidente e incontrovertible. Al mismo tiempo, no hay un hecho que sea tan escurridizo, paradójico e incluso misterioso como eso que nos es tan evidente. David Chalmers y su "problema difícil" de la consciencia apuntan a la gran pregunta de cómo y por qué es que existe la experiencia subjetiva: el mundo tal y como lo conocemos podría seguir funcionando de la misma manera sin necesidad de que haya sujetos conscientes que lo perciban. Por otro lado, Thomas Nagel nos adentra en las dificultades al momento de pensar en otras consciencias, ya sea la de otros humanos, otros animales, pero también mentes artificiales como la podría llegar a tener eventualmente un robot. Por más de que yo como sujeto consciente conozca y describa cada detalle del funcionamiento de una consciencia ajena, al final a lo que más puedo aspirar es a saber cómo sería esa consciencia *para mí*, pero nunca cómo sería para el que la posee. Inconvenientes similares son explorados por Ian McEwan en *Machines Like Me* cuando Charlie se hace preguntas como: ¿puede Adam sentir amor o dolor?, ¿cómo es que de procesadores electrónicos y cálculos computacionales pueden surgir experiencias subjetivas?, si sí es consciente ¿cómo se siente para el robot enchufarse a la corriente después de agotar su batería? El hecho es que no lo sabemos y no *podemos* saberlo.

Frente a esta incógnita y cuando se trata de mentes artificiales como la que podría tener un robot, la prueba de Turing resulta ser el criterio paradigmático para determinar si hay consciencia o no. En este sentido, Alan Turing apunta a que la mejor manera de determinar si hay consciencia es que *parezca* que la haya. Como consecuencia de los modelos computacionales más avanzados, esta prueba ha estado recientemente en el foco del debate y hay quienes dudan de su utilidad en el futuro (Floridi y Chiriatti 683). No obstante, la intuición fundamental detrás de la prueba de Turing permanece intacta: el comportamiento externo es el único medio que existe para acercarnos al mundo interior de otros. En la novela, y en el mundo real también, la capacidad que tiene la máquina para *aprender* resulta crucial para que desarrolle comportamientos semejantes a los de un humano. Adam, el robot que cuando fue encendido por primera vez era ignorante y obediente, con el paso del tiempo fue aprendiendo, adquiriendo autonomía y un criterio propio que lo llevó a tomar decisiones que incluso afectaban negativamente a quienes más valoraba. El deseo Adam de mantenerse encendido – lo que supuestamente equivale a mantenerse vivo y consciente – así como sus expresiones de estados mentales como el enamoramiento, la esperanza, el mal humor y la indignación lo muestran como un ser capaz de tener un mundo interior como cualquier ser humano. Su comportamiento llego a ser tan indistinguible del de cualquier otro ser humano que, en un episodio que sigue un estrecho paralelo con la prueba de Turing tal y como fue planteada originalmente, el papá de Miranda piensa que Adam es el humano y Charlie el robot.

Pero independientemente de si hay o no consciencia en la máquina, la existencia de robots indistinguibles de los seres humanos no nos será indiferente. Sin necesidad de que existan, ya los seres humanos tenemos una marcada tendencia a antropomorfizar todo a nuestro alrededor. La relación entre humanos y máquinas exigirá amplias discusiones éticas como las que plantea Ian McEwan en *Machines Like Me*. Por un lado, está la pregunta de cómo programar la ética en una

máquina: desde un punto de vista pragmático, una ética deontológica resulta más fácil de incorporar a un programa que una consecuencialista. No obstante, la novela deja claro que las consecuencias de esa elección pueden llegar a ser perjudiciales. A diferencia de Adam, que no es capaz de apreciar los matices éticos en una típica y complicada situación humana, los humanos somos contradictorios y nuestras valoraciones morales cambian y se condicionan por una cantidad de factores que una máquina no puede fácilmente evaluar. Por otro lado, si crear mentes artificiales, conscientes, sintientes y capaces de tener experiencias subjetivas se convierte en una realidad, surge la pregunta de cómo debe ser la ética de los humanos hacia ellas. La realidad humana es compleja y está plagada de tragedias y dolores que pueden causar sufrimiento en las posibles mentes artificiales. Todo esto hará que las máquinas ya no puedan ser tratadas como objetos, sino que deben empezar a gozar de una consideración moral similar a la de cualquier otro ser humano.

El análisis acá presentado muestra que *Machines Like Me* de Ian McEwan, como toda buena literatura, ofrece más preguntas que respuestas. En cada uno de los capítulos tratados se puede ver cómo la novela sirve para llevar a cabo una serie de discusiones esenciales que sirven para navegar el mundo actual y, posiblemente, el futuro no muy lejano Para muchos, los temas tratados en este trabajo pueden ser muy especulativos, incluso superfluos, pero los avances tecnológicos ya están dando indicios de que preguntas y debates sobre la subjetividad del robot serán cada vez más comunes. La relación entre los humanos y las máquinas será cada vez más compleja y exigirá replantearnos lo que entendemos por humano. La vida posthumana ya está llena de retos y en el futuro próximo estos solo se multiplicarán rápidamente. En este sentido, resulta necesario anticiparse a los hechos y proponer discusiones relevantes antes de que sea muy tarde.

Obras citadas

Braidotti, Rosi. *Posthuman Knowledge*. Cambridge: Polity Press, 2019.

Chalmers, David. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies*, vol. 2, num. 3. (1995): 200-219.

Colombino, Laura. "Consciousness and the Nonhuman: The Imaginary of the New Brain Sciences in Ian McEwan's Nutshell and Machines Like Me." *Textual Practice*, vol. 36, num. 3. (2022): 382-403.

Floridi, Luciano y Massimo Chiriatti. "GPT-3: Its Nature, Scope, Limits, and Consequences." *Minds and Machines*. (2020): 681-694.

Kopka, Katalina y Norbert Schaffeld. "Turing's Missing Algorithm: The Brave New World of Ian McEwan's Android Novel Machines Like Me." *Journal of Literature and Science*, vol. 13, num. 2. (2020): 52-74.

McEwan, Ian. *Machines Like Me*. Nueva York: Anchor, 2019.

Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review*, vol. 83, num. 4. (1974): 435-450.

Oppy, Graham y David Dowe. "The Turing Test." *The Stanford Encyclopedia of Philosophy* (2021).

Shang, Biwu. "Froma Alan Turing to Ian McEwan: Artificial Intelligence, Lies and Ethics in Machines Like Me." *Comparative Literature Studies*, vol. 57, num. 3. (2020): 443-453.

Turing, Alan. "Computing Machinery and Intelligence." *Mind*, vol. LIX, num. 236. (1950): 433-460.