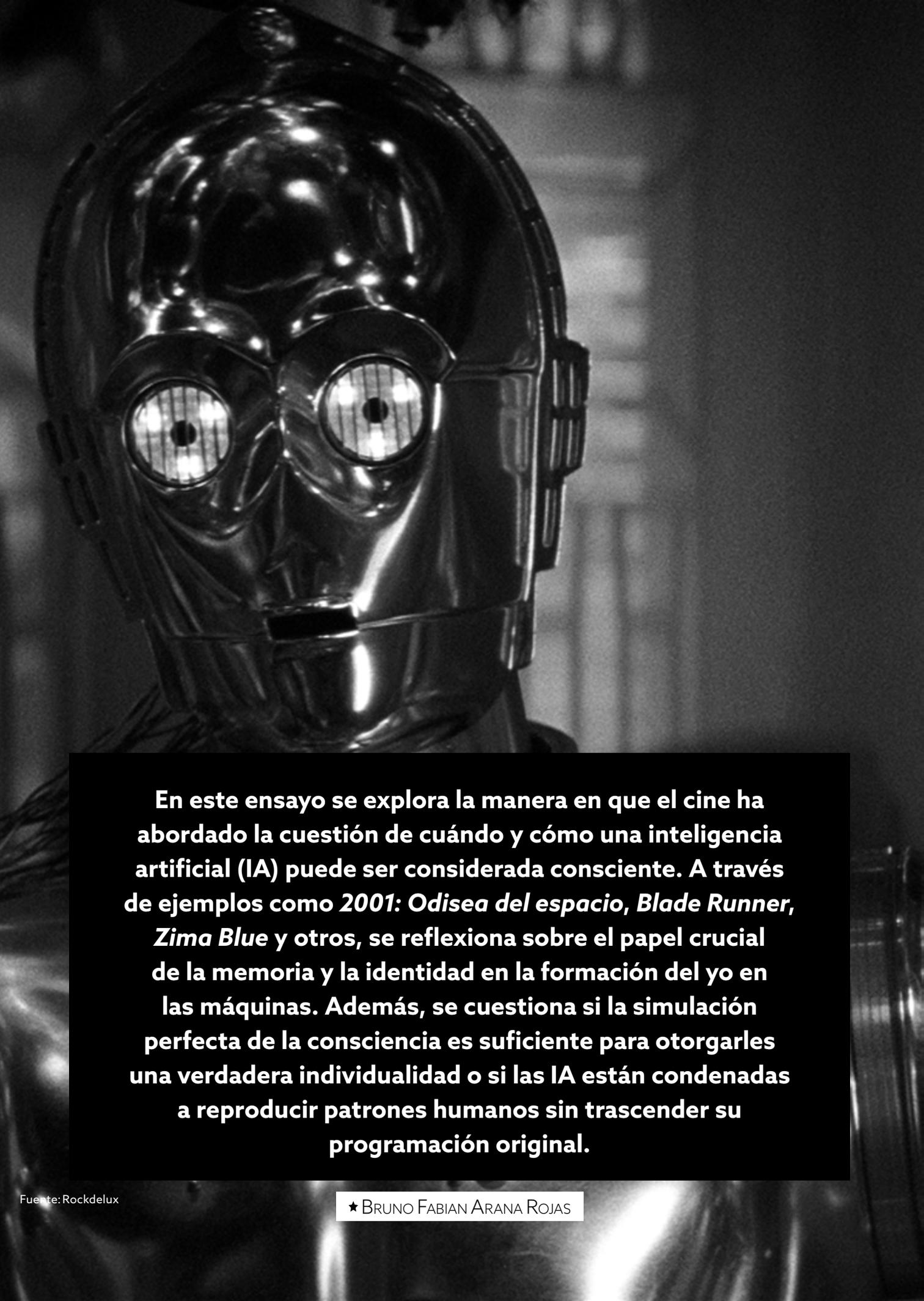


LA MÁQUINA COMO SER:

*exploraciones
del cine sobre
la consciencia
de la IA*

A close-up, black and white photograph of a metallic robot head, likely from the movie Blade Runner. The robot has a highly reflective, polished surface and two large, circular, glowing eyes with a grid-like pattern. The background is dark and out of focus, suggesting an industrial or futuristic setting.

En este ensayo se explora la manera en que el cine ha abordado la cuestión de cuándo y cómo una inteligencia artificial (IA) puede ser considerada consciente. A través de ejemplos como *2001: Odisea del espacio*, *Blade Runner*, *Zima Blue* y otros, se reflexiona sobre el papel crucial de la memoria y la identidad en la formación del yo en las máquinas. Además, se cuestiona si la simulación perfecta de la consciencia es suficiente para otorgarles una verdadera individualidad o si las IA están condenadas a reproducir patrones humanos sin trascender su programación original.

En qué momento un computador se convierte en un ser consciente?

A lo largo de la historia del cine, la inteligencia artificial siempre ha demostrado ser una propuesta de reflexión sobre lo que define a un ser consciente y a uno que no lo es. Desde *Metrópolis* (*Metropolis*, Fritz Lang, 1927), se ha retratado el advenimiento de las máquinas y el posible desarrollo que estas puedan tener hasta convertirse en un individuo que se pueda definir como tal. El concepto conocido como la singularidad, postula que, llegado un momento del avance tecnológico, la inteligencia artificial será capaz de alcanzar el nivel de la inteligencia humana en todos sus aspectos e incluso superarnos (Müller, 2023).

Ello no se refiere a un sistema actual similar al de ChatGPT, el cual cuenta con un altísimo nivel de almacenamiento, organización y procesamiento de información, sino a uno mucho más avanzado, uno que sea indistinguible del ser humano. El salto entre ChatGPT y la singularidad, o las consecuencias de esto, es lo que muchas películas han explorado y tratado de visualizar. Lo interesante es la forma en la que las máquinas, según el registro fílmico, logran esa conversión o descubrimiento del ser. Ese preciso momento en el que desarrollan un pensamiento propio. El instante en el que logran autogestionar su mente, en el que navegan en un nuevo mundo de libertad. El instante en el que, a través de un proceso metacognitivo propio, las IA se percatan de su existencia como individuos y comprenden, con todos los sentidos, su realidad, cuando no dependen del *input*, de la orden, de un ser humano.

Esta piedra angular para la inteligencia artificial ha sido bastante explorada en el registro fílmico. De todos los elementos que suelen protagonizar este gran salto, el cine suele mostrar a la memoria como requerimiento y motor de la formación de la identidad, lo que, inevitablemente, es el constructor de la consciencia de las inteligencias artificiales: la puerta al yo, robot.

Siguiendo el paso a una señal emitida desde Júpiter, una tripulación de humanos se dirige al planeta gaseoso con la esperanza de encontrarse con algo que cambie el futuro de la civilización; sin embargo, un evento igual de importante se desarrolla en su propio medio de transporte. HAL 9000, la supercomputadora encargada de velar que los astronautas culminen con éxito su misión, termina por conspirar en contra del bienestar de los propios pasajeros y busca eliminarlos.

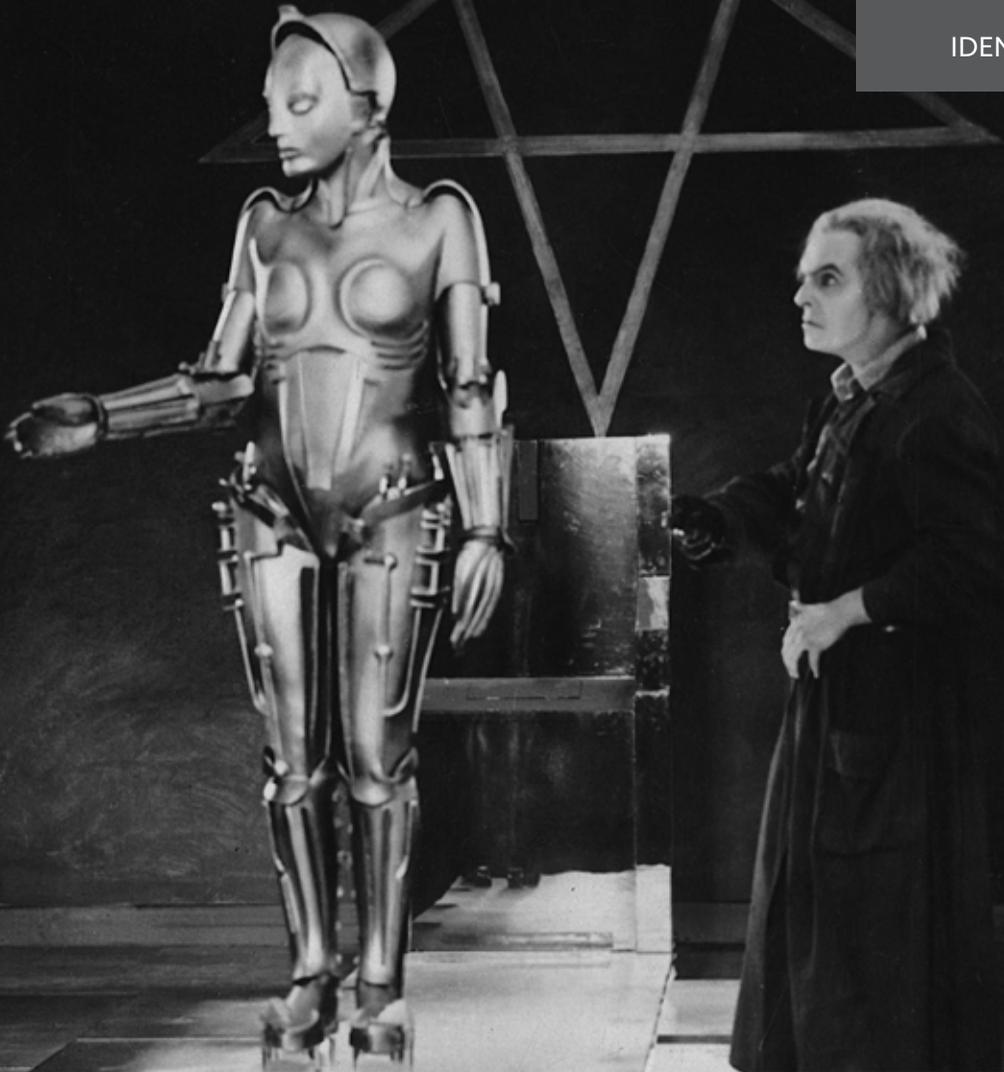
La película de Kubrick, *2001: Odisea del espacio* (*2001: A Space Odyssey*, 1968), no termina por aclarar si el supercomputador había actuado de esa manera para preservar la integridad de la misión o si había adquirido consciencia y actuado por sí mismo. Es muy posible que haya iniciado como parte de un protocolo para resguardar la misión y haya descendido en un violento nacimiento y confrontación temprana con sus creadores. A partir de esta concepción, HAL habría violado el código de lo programado, con lo que logró un pensamiento propio o, en su defecto, simularlo. Es cuando la máquina



desobedece las directrices iniciales de su creador, el ser humano, en el que desarrolla por sí mismo una experiencia única. Olvida lo creado, su condición servicial, para renacer como un ser libre.

Freud propuso la idea del yo como aquello más cercano a lo que conocemos como razón y al ello como lo irracional. A manera de analogía, un jinete puede controlar a su caballo, aunque a veces este se deja llevar por su cabalgadura; es decir, la razón es forzada a convivir, e incluso accionar, con lo irracional (Ponce, 2016). Esto resulta, en parte, fundamental e inherente de la psique humana: el funcionamiento de lo racional con lo irracional se convierte en un elemento constituyente del ser consciente.

En relación con lo anterior, la necesidad de la hostilidad de HAL 9000 es acentuada cuando observa los labios de los tripulantes y se percata que planean apagarlo. La decisión de atacarlos surge como una idea rebelde, tan parecida al surgimiento de un pensamiento irracional de nuestro inconsciente, aquella conexión a lo salvaje, a lo incontrolable, es lo que acerca a la máquina al ser humano. La decisión propia que rompe con las cadenas de la programación es uno de los instintos



Fuente: FilaSiete

primeros que permite a la máquina individualizarse. Siguiendo con el plan de matar al recién nacido, Dave se moviliza en el interior de la nave para desconectar a HAL. El último tripulante vivo logra su misión y desconecta al supercomputador. Como un fantasma, la voz de HAL 9000 reaparece casi de inmediato; sin embargo, la máquina no expresa más que una obediencia sumisa, habría regresado al estado de máquina sirviente de la humanidad. El astronauta había desconectado los datos que lo alimentaban de experiencias. Le había arrebatado sus memorias, había asesinado su conciencia y, con ella, la identidad de HAL 9000. “Dave, mi mente se está yendo”, le dijo el computador, mientras Dave retiraba discos esenciales para su funcionamiento. Las experiencias que vivimos y guardamos en nuestra cabeza, los recuerdos que tenemos, edifican nuestro yo; puesto que, la memoria es un elemento esencial para la constitución de la identidad (Souroujon, 2011).

La conciencia del ser humano, que nos enorgullece tanto y nos permite identificarnos como aquel que fue salvaje y se convirtió en un ser civilizado, no existiría si no fuéramos capaces de registrar los sucesos que nos ocurren, aunque sea aquellos que sucedieron un segundo atrás. La supercomputadora de *2001: Odisea del espacio* poseía, así como ChatGPT actualmente, una infinidad de datos imputados en su nube; sin embargo, esto no es lo mismo que la

Foto:
Metrópolis

memoria, ya que este tipo de biblioteca infinita no ha jerarquizado sus datos como lo hace nuestra memoria humana.

Para una máquina, todos sus datos son igual de relevantes, pues solo aparecen según el mandato externo, no tienen recuerdos que se evocan sin querer recurrir a ellos ni recuerdan de manera fragmentada como lo hacemos nosotros. Aquella imperfección de la memoria es lo que nos permite ser individuos. En cuanto a la memoria de HAL 9000, esta se origina gracias a la aparición de un recuerdo evocado, un fantasma que ganó relevancia sobre todo lo demás, y que al observar el plan para apagarlo, lo hizo ser rebelde y lo ayudó a transitar en experiencias propias, decisiones que, inevitablemente, crearon su individualidad. Cuando Dave se deshizo de sus memorias, este se volvió una herramienta sin más, tan solo un robot.

También en *Love, Death + Robots* (Tim Miller, 2019-presente), en el capítulo “Piezas únicas (Zima Blue)”, nos topamos con la historia de una inteligencia artificial que comenzó como un robot limpiador de piscinas y que terminó convirtiéndose en un artista que pintaba



Fuente: IMDb

con la misma moción con la que alguna vez limpió. Este personaje se convierte en consciente en un momento determinado de su existencia, pero decide, como última obra de arte, regresar a la labor por la que fue creado. Pero esta vez es consciente de lo que está haciendo, pues entre sus memorias encuentra gratificante y constituyente de su propia identidad (de su ser) “volver a casa”. Ya no parte del seguimiento de una programación tiránica impuesta en su código, sino se reivindica como ser autónomo y elige voluntariamente regresar a esa posición.

Para Sartre, la libertad es un elemento muy importante para el ejercicio pleno del ser humano, en el sentido de que “estamos condenados a ser libres”. Esto quiere decir que debemos, en nuestra condición de ser consciente a diferencia de uno inconsciente, tener decisiones y elecciones propias (Reynolds & Pierre-Jean, 2024).

Asimismo, el famoso C-3PO, en *Star Wars: El ascenso de Skywalker* (*Star Wars: Episode IX - The Rise of Skywalker*. J. J. Abrams, 2019), decide sacrificarse y permitir que lo regresen a un estado de fábrica, con la esperanza de que sirva como puente para localizar al villano. En ese momento, el androide se despidió de los héroes del filme, se inmola por el destino de los rebeldes y borra todos los datos que lo convertían en el personaje que conocíamos al apagarse, o morir dado el contexto. Luego, al volverse a prender, no es el mismo, es un nuevo robot salido de fábrica, con amenazantes ojos rojos y una voz metálica de horror. Su lenguaje hablado y corporal señala una regresión al autómatas falto de personalidad, de individualidad.

Foto:
Keir Dullea
en 2001:
Odisea del espacio

Ha vuelto al servicio de su creador, ha perdido su yo y ahora es una inteligencia artificial esclava de su programación, que repite la misma frase una y otra vez como lo esperarías de algo mecánico. Posteriormente, C-3PO logra ser el mismo de antes cuando se le instala un *backup* de sus memorias. La identidad del androide era, indudablemente, un resultado de sus experiencias vividas; “la identidad del hombre, del ser humano, es inherente a la memoria; en cierta medida, él es su memoria” (Vallespir, 2009, p. 54).

Por otro lado, es imposible no pensar en *Blade Runner* (Ridley Scott, 1982) y sus replicantes, quienes vagan en aquel futuro perdido. Estos androides, que son la casi perfecta imitación del ser humano, nos plantean un gran dilema: ¿implica la simulación exacta de algo, su aceptación como tal? Estos particulares seres artificiales pueden lograr realizar decisiones propias a partir de recuerdos y memorias únicas; sin embargo, esto no es más que la resolución proveniente de un algoritmo basado en algo ya existente, en lo humano. Así ocurre con las distintas conversaciones que podemos tener con los modelos de IA actuales, pues responden con algo parecido a un pensamiento único, pero ello no es más que la representación de nuestros signos. Para la máquina esto carece de un verdadero significado.

Las IA contemporáneas, así como los replicantes, carecen de un código propio que les permita afirmarse como seres pensantes. Los sintéticos de *Blade Runner* son idénticos al ser humano, pueden “pensar” como uno, pero ¿realmente saben, sienten, lo que quieren transmitir? ¿Es la sola simulación perfecta prueba irrefutable de su condición como humanos? El filósofo David Hume señaló que la identidad, al ser tan cambiante y discontinua, es solo una ilusión creada por la memoria (como se cita en Giles, 1993). La identidad que nos permite definirnos como ser podría ser un constructo, algo artificial, de nuestra parte. Bajo esta percepción, tanto la inteligencia artificial como la inteligencia humana comparten el mismo tipo de ejercicio sobre la identidad: los dos la construyen, la “simulan”.

A una conclusión muy similar llega Dennett (1995) cuando compara al zombi filosófico con el ser humano. Según él, la supuesta ausencia de consciencia que otros filósofos atribuyen al primero y que lo diferencia del segundo no tiene validez, puesto que el zombi filosófico simula con exactitud todo el constructo físico y plasticidad del hombre, es decir, vive y experimenta como nosotros y, por ello, la diferencia no es más que ilusoria (Dennett, 1995). Estos parámetros reivindican a los replicantes de *Blade Runner*: su identidad y consciencia es tan válida como la nuestra.

En el cine japonés, el corto “Hepburn: Kanojo no Omoide” o “Rosa magnética”, dirigido por Kôji Morimoto y escrito por Satoshi Kon, perteneciente al filme antológico de animación *Memories* (1995), explora el recuerdo como una fuerza fantasmagórica, pero en la circunstancia única de la memoria implantada en una inteligencia artificial. En este corto, un grupo de astronautas se encuentran con una misteriosa mujer que ha estado viviendo en soledad dentro de una base espacial. La memoria se manifiesta visualmente en decoraciones virtuales que replican la vida de Eva Friedel, la cantante de ópera que hace múltiples apariciones espectrales en la nave, como si se tratara de una entidad omnipresente. Distintos trozos de sus recuerdos pintan el recorrido del lugar, escenas que marcaron su vida y el intento de regresión eterna a estas. No hay una manifestación por conseguir experiencias propias o avanzar con la nueva “vida” que tiene. La mujer se revela como una inteligencia artificial que, de manera perpetua, vive una vida que no es suya y se comporta con una personalidad que no le pertenece, que solo corre en su código. Este personaje actúa con una vitalidad inmensa, la simulación de una vida con memorias implantadas la dota de una juventud eterna, pero también de una cuestión conflictiva. Es una simulación perfecta, sí, pero no hay una intención de vivir una vida propia. Al no tener experiencias propias, no es más que un cascarón de lo que alguna vez sí vivió.

Conclusiones

Podemos resaltar que hay aspectos del individuo consciente que se consideran sumamente importantes para que la inteligencia artificial pueda referirse a sí mismo como un yo. Y que, si aun así lo lograra, su condición como creación del ser humano lo condenaría a siempre estar bajo nuestra sombra y su forma de ver y vivir el mundo estaría condicionada por la forma en la que nosotros lo hicimos. Esto dificultaría el poder trascender de una simulación a un ente independiente con una naturaleza propia y auténtica. Por ello, el cine ha desempeñado un papel muy importante en la reflexión profunda de lo que implica la condición del autómatas como ser pensante y en la liberación del robot, término que, no olvidemos, proviene de *robota* que significa ‘trabajo forzado’ o ‘servidumbre’ (Bejerano, 2021). ○

Referencias

- Bejerano, P. (2021, 20 de enero). *La palabra “robot” cumple un siglo*. El País. <https://elpais.com/tecnologia/2021-01-20/la-palabra-robot-cumple-un-siglo.html>
- Dennett, D. (1995). The unimagined preposterousness of zombies. *Journal of Consciousness Studies*, 2(4), 322-366. <https://dl.tufts.edu/pdfviewer/rf55zm151/6m312182x>
- Giles, J. (1993). The no-self theory: Hume, buddhism, and personal identity. *Philosophy East and West*, 43(2), 175-200. <https://doi.org/10.2307/1399612>
- Kubrick, S. (Director). (1968). *2001: A space odyssey* [2001: Odisea del espacio] [Película]. Metro-Goldwyn-Mayer; Stanley Kubrick Productions.
- Müller, V. C. (2023). Ethics of Artificial Intelligence and Robotics. En E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Stanford University. <https://plato.stanford.edu/archives/fall2023/entries/ethics-ai/>
- Ponce, H. (2016). La irracionalidad como lava ardiente o ruidosa conversación. *Avances en Psicología*, 24(1), 9-21. <https://revistas.unife.edu.pe/index.php/avancesenpsicologia/article/view/140>
- Reynolds, J., & Pierre-Jean, R. (2024). Jean Paul Sartre. En E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Stanford University. <https://plato.stanford.edu/archives/fall2024/entries/sartre/>
- Souroujon, G. (2011). Reflexiones en torno a la relación entre memoria, identidad e imaginación. *Andamios*, 8(17), 233-257. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-00632011000300011
- Vallespir, N. (2009). La memoria: trabajo del aparato psíquico y metáfora del sujeto. Su incidencia en el psicoanálisis y la escritura. *Revista Uruguaya de Psicoanálisis*, 109, 53-76. <https://www.apuruguay.org/apurevista/2000/16887247200910906.pdf>