

# INTELIGENCIA ARTIFICIAL EN EL CINE: CONFLICTOS Y PARADOJAS

Jordi Conill Salomé

El artículo presenta, por un lado, una reflexión acerca de las repercusiones de los descubrimientos en el terreno de la inteligencia artificial sobre el campo de la producción literaria y cinematográfica y, por otro, examina la influencia que las especulaciones de los creadores han tenido de cara a la posibilidad de entrever horizontes de futuro amables o amenazantes, fruto de la connivencia entre las capacidades humanas y las de las máquinas. La cuestión a estas alturas permanece abierta, y cabe la posibilidad de que en las próximas décadas, si se produce la singularidad tecnológica, no cese de incrementar su inquietante actualidad.

**Palabras clave:** inteligencia artificial, cine, literatura, singularidad tecnológica, Nueva Carne, reproducción autopoietica, posthumanidad

## Introducción

En el poema épico griego conocido como *Las Argonáuticas* aparece una figura impresionante, un enorme coloso de bronce forjado llamado Talos, a quien le ha sido encomendada la custodia de la isla de Creta. Esta criatura, casi invulnerable, que se convierte en uno de los grandes rivales a batir por los protagonistas del libro, representa uno de los primeros ejemplos de criaturas artificiales que se conocen en la historia de la humanidad. La presencia de criaturas como Talos en la mitología clásica o los gólems en la hebrea da constancia de la persistencia de la idea de que las creaciones artificiales pueden llegar a lograr una autonomía similar a la humana. Por otro lado, se sitúa en el origen de dos de las temáticas más recurrentes en los relatos sobre la inteligencia artificial: la del autómatas casi invulnerable, poseído por una obsesión destructiva, y la del ser fabricado a imagen y semejanza del hombre, en un intento de emular a Dios.

A lo largo de la historia encontramos ejemplos esporádicos de autómatas e inteligencias artificiales,

casi siempre como un reflejo de los caminos inquietantes que podría emprender el progreso humano (Peirano y Bueno Gómez-Tejedor, 2009). La idea subyacente a la mayor parte de estas creaciones es la de la sustitución de la humanidad por réplicas artificiales, que tanto podrían transformarse en sus sirvientes como en sus herederos. En el estrafalario cuento de E.T.A. Hoffmann titulado *El hombre de arena* (1817), Nathanael se enamora de la autómatas Olimpia, que para su mente trastornada resulta más humana que su amada Clara, a pesar de que el comportamiento de aquella es claramente mecánico y previsible, cualidades que el fetichismo de Nathanael aprecia más que la espontaneidad de Clara. El relato de Hoffmann ha suscitado múltiples interpretaciones, entre las cuales cabe destacar el célebre ensayo de Sigmund Freud sobre *Lo siniestro* (1919), en el cual nos presenta a Nathanael como un personaje con trasfondo misógino, marcado por una idea muy pura de la feminidad, producto de los traumas de su infancia, que Clara, con su superficialidad, no consigue satisfacer. La sustitución del cuerpo de la amada por una copia, más en consonancia con una sociedad dominada por el patriarcado, fue desarrollada posteriormente en la novela —adaptada a la pantalla el 1975 y el 2004— *The Stepford Wives* (1972) de Ira Levin, donde los hombres de un pueblo norteamericano deciden asesinar sus esposas y sustituirlas por androides-esclavas que cumplen el papel que el imaginario patriarcal asigna a la «mujer perfecta» (Pedraza, 1998; Broncano y Fuente, 2010).

*R.U.R.* (1920), pieza teatral escrita por el checo Karel Čapek en que se introduce por primera vez el término *robot* en el terreno de la ficción, ha sido unánimemente considerada como otra obra precursora de la mayor parte de los planteamientos sobre la inteligencia artificial. Trata sobre una empresa que construye criaturas humanas de tipo mecánico para aligerar el trabajo a los hombres. Son criaturas que pueden hacerse pasar por seres humanos y que poseen el don del pensamiento. A pesar de haber sido concebidas para ayudarlos, llega un momento en que se enfrentan con la sociedad e inician una revuelta que acabará destruyendo a la humanidad. El trasfondo social de la obra, que remite a las revoluciones marxistas —tema muy explotado por los escritores de ciencia-ficción de la época, como H.G. Wells—, plantea una cuestión recurrente en el género, que más adelante será desarrollada sobre todo por su rama «robótica»: ¿qué ocurre cuando dejamos los medios de producción en manos de nuestros productos? Hablamos de una preocupación que se intensificó en la literatura producida a partir de los años 50 del siglo XX, a raíz de la aparición de la

cibernética y sus repercusiones en los más variados campos (Wiener, 1969).

En este artículo nos proponemos analizar algunos ejemplos del tratamiento que el cine ha dado a la cuestión de la inteligencia artificial, procurando enlazar nuestras reflexiones con las teorías de matemáticos y escritores como Alan Turing, John von Neumann, Stanisław Lem o Philip K. Dick.

### ***A.I. Artificial Intelligence o la ambigua naturaleza de los androides***

En el año 1950, el matemático Alan Turing escribió un artículo titulado «Computing machinery and intelligence», en el cual proponía un test para discernir si una máquina poseía o no inteligencia. De entrada, el contacto entre el hombre y la máquina que postulaba no tenía que ser directo, sino a través de una interfaz, con objeto de impedir que el encuestador pudiera ser influenciado por sus observaciones visuales a la hora de valorar si las respuestas de la máquina podían ser interpretadas como humanas. En caso de que el resultado fuera afirmativo, se determinaría que la máquina en cuestión era inteligente.

El planteamiento de Turing fue muy influyente en la ciencia-ficción, sobre todo para subrayar la ambigüedad de la línea que separa a los hombres de las máquinas. Una de las películas que alude de manera más directa al planteamiento de Turing es *Blade Runner* (1982) de Ridley Scott, donde a los replicantes —denominación de las inteligencias artificiales que aparecen en ella— se los somete al test de empatía Voight-Kampff, un examen de tipo científico y psicológico mediante el cual un dispositivo se encarga de valorar las reacciones físicas del sujeto a ciertas preguntas y, si el sujeto es humano, detecta una respuesta emocional, que no aparece cuando se trata de un androide. Philip K. Dick, el autor de la novela (*Do Androids Dream of Electric Sheep?*, 1968) en que se basa la película, desarrolló sus reflexiones sobre el asunto en varios escritos ensayísticos posteriores, donde aplicaba los requisitos de su test de empatía no sólo a los androides, sino también a los humanos. De acuerdo con este criterio, la distinción entre unos y otros no tendría tanto que ver con su condición de criaturas biológicas o cibernéticas como con la calidad refleja o empática de su pensamiento, así como con su incapacidad para reconocer o no las excepciones. En consecuencia, el androide —más allá de su naturaleza artificial o biológica— se caracterizaría por su aislamiento mental y moral, por su negación de la humanidad y la predictibilidad «inhumana» de su conducta, sujeta a la repetición

compulsiva de los patrones preestablecidos, incluso cuando éstos ya no resultan funcionales.

Un ejemplo de esta pauta de comportamiento lo podemos encontrar en la película *A.I. Artificial Intelligence* (2001, Steven Spielberg). El protagonista de la película es un androide —Mecas, en el film— llamado David (Haley Joel Osment), que fue diseñado para actuar como un niño humano y cumplir el rol de hijo devoto dentro de la estructura familiar de sus compradores. Cuando su «madre» pronuncia una serie de palabras condicionantes, David está programado para reconocerla y amarla como un hijo amaría su madre de verdad, pero sin la opción humana de poder renunciar a su afecto si éste, por ejemplo, no fuera correspondido. Aun así, como muestra más adelante la película a partir de la figura del Doctor Hobby (William Hurt), los límites entre humanos y robots en cuanto a esta problemática son muy difusos, porque éste tampoco es capaz de renunciar al afecto que siente por su hijo muerto e intenta resucitar patéticamente su memoria. De todos modos, por paradójico que ello pueda parecer, la incapacidad de dejar de amar a su madre que experimenta David a lo largo de toda la cinta, a pesar de que ésta lo abandona, es quizás el rasgo que más lo humaniza.

En un primer momento, antes de que se le suministre la programación adecuada que lo transformará en un hijo cariñoso, David se nos muestra como un ser sin empatía, inofensivo, que aprende por imitación, pero que no está nada interesado por establecer una conexión emocional con nadie de su entorno —actitud subrayada en la versión original por el tono de voz neutro y frío de Haley Joel Osment al principio de la película. Pero tan pronto recibe la orden de amar a su «madre», empieza a establecer conexiones emocionales con ella y con todos los objetos de su entorno. Una vez abandonado, su evolución emocional continúa, y empieza a desarrollar emociones nuevas como la frustración, la tristeza, la desesperación, el odio o la esperanza, mientras se vuelve consciente de las cosas que lo separan de una humanidad a la cual quiere pertenecer para que su «madre» lo ame. El momento culminante de esta evolución se puede situar hacia el final de su viaje, cuando vuelve al edificio de Cybertronics, donde fue creado, y encuentra un Meca idéntico a él, que se comporta con la misma actitud indiferente y neutra que adoptaba David al comienzo de la película. Cuando se da cuenta de que no es una criatura única, destruye al Meca en un raptó de ira, porque entonces pierde la esperanza de convertirse algún día en humano. Esperanza que recuperará más adelante, al encontrarse con una estatua del Hada Azul, el ser fantástico que David busca desesperadamente para pedirle que lo transforme en humano por arte de magia.

En su día, *A.I. Artificial Intelligence* (2001) fue una película bastante malentendida, pero personalmente creo que ha ganado dimensiones de sentido con el tiempo. Se criticó sobre todo su desenlace, con el argumento de que estropeaba la historia, a pesar de que a estas alturas resulta difícil aceptar estas críticas, teniendo en cuenta que la mayor parte se formularon sin un conocimiento verdadero y reflexivo de lo que ocurre durante la media hora final. Una de las cosas que más me atraen de la conclusión que propone Spielberg es la aparición de una raza de androides del futuro —a la cual, absurdamente, muchos críticos confundieron con alienígenas, debido a su estética—, circunstancia que convierte *A.I. Artificial Intelligence* en una de las escasas películas que nos muestra un futuro posthumano, que ya se está gestando ahora mismo (Braidotti, 2015), en el cual la humanidad como tal ha dejado de existir. Esta visión del futuro nos revela una de las claves de la película: los androides son los descendientes evolutivos de los seres humanos e investigan como auténticos arqueólogos entre las ruinas congeladas de Nueva York, de donde rescatan el «cadáver» de David, un hallazgo que para ellos probablemente equivale al hecho de que un arqueólogo contemporáneo hallase un australopiteco vivo.

Esta futura sociedad de máquinas —una idea fascinante, pero que lamentablemente en la película apenas se desarrolla— nos remite a las teorías de John von Neumann, que en su ensayo póstumo «Theory of Self Reproducing Automata» (1966) reflexionó sobre la aparición de máquinas autoreplicantes, capaces de reproducirse a sí mismas (Penrose, 1974), un concepto que ya podemos encontrar en películas tan diferentes como *2010: The Year We Make Contact* (1984, Peter Hyams) o *Alien Resurrection* (1997, Jean Pierre Jeunet). Hay que señalar que a estas alturas el concepto de von Neumann ha dejado de ser sólo teórico, dada su utilidad para describir el comportamiento de los virus informáticos o de los autómatas celulares. De todos modos, para nuestros propósitos, el aspecto más interesante de las especulaciones neumannianas es quizás su vertiente involuntariamente siniestra —para decirlo en términos freudianos—, que adjudica a las máquinas un comportamiento propio de seres biológicos, otorgándoles una cierta capacidad de sustraerse al control que los humanos ejercemos a la hora de diseñarlas e imponerles límites. A lo largo de *A.I. Artificial Intelligence* hay varias referencias que prefiguran esta irrupción final de los androides del futuro: sólo hay que recordar, en este sentido, la primera aparición desenfocada de David como una sombra estilizada, el símbolo de Cybertronics, que adopta la forma de una figura humana

igualmente estilizada, o la reflexión de Gígoló Joe sobre el hecho de que la humanidad tiene los días contados y el futuro es de las máquinas. Se trata en todos los casos de indicios que apuntan a un futuro sin humanos, en el que las máquinas podrán desarrollarse a su antojo, sin ningún tipo de limitaciones antropomórficas. Desde este punto de vista, el «regalo» que le hacen a David al final de la película los androides del futuro —que clonan la figura de su «madre», para que al menos conozca la felicidad durante el tiempo del único día que resulta factible la existencia del clon— tiene un cierto regusto de justicia poética, puesto que ahora son las máquinas las que crean artificialmente seres humanos para satisfacer sus deseos, y más concretamente para ayudar a uno de los suyos a lograr su objetivo existencial, la frustración del cual le generó mucho sufrimiento. Nótese que las máquinas, por medio de este acto y de la subsiguiente desconexión eutánica de David —cuando al final de la película el narrador dice que David «duerme» por primera vez en su vida, ello no deja de ser, obviamente, una forma metafórica de referirse a la muerte, al descanso del protagonista después de haber conseguido su objetivo, gracias al cual se transforma en humano, el último humano— demuestran un nivel altísimo de empatía, que iguala o incluso supera al que antaño caracterizaba a las personas.

### **La Paradoja Cogito**

No obstante, esta visión de la inteligencia artificial no deja de ser un poco simple, puesto que no tiene en cuenta la posibilidad de la aparición de inteligencias irreflexivas que plantea la *paradoja cogito*, formulada por Stanisław Lem en su libro de falsos prólogos *Un valor imaginario* (1973). Según esta paradoja, la aparición de inteligencias artificiales irreflexivas, como por ejemplo los programas de ajedrez o los ordenadores programados para dirigir sesiones de psicoterapia, provocará en algún momento del futuro grandes dudas a la hora de determinar si existe una inteligencia consciente detrás de ellos, producto de la creencia involuntaria en que si una inteligencia responde con racionalidad a nuestros actos, tiene que ser consciente y tiene que estar movida por una voluntad subyacente. Pero si abandonamos los puntos de vista reduccionistas de carácter antropomórfico y empezamos a asumir sin prejuicios la realidad de las inteligencias irreflexivas, no tardaremos a enfrentarnos con una paradoja de muy difícil resolución, porque la

gran aportación de la inteligencia artificial en el futuro tendrá que ver con la emergencia de una perspectiva alternativa, que se concretará en la adquisición por los aparatos cibernéticos de la facultad de juzgar la conciencia humana y dará paso a la duda de las máquinas en relación con la capacidad de pensar de los humanos. La *paradoja cogito* teoriza, precisamente, esta reticencia recíproca a reconocerse en las operaciones efectuadas por otras mentes, que desemboca en un relativismo radical en cuanto a las diversas maneras de entender la conciencia.

*Be Right Back* (Owen Harris, 2013), el primer capítulo de la segunda temporada de *Black Mirror*, conocida serie británica de ciencia-ficción distópica creada por Charlie Brooker, nos ilustra sobre los dilemas asociados a esta paradoja. Martha (Hayley Atwell), la protagonista del episodio, pierde su pareja, Ash (Domhnall Gleeson), en un accidente. Intenta superar el luto por los medios tradicionales, pero gracias a una amiga, al darse cuenta que está embarazada del difunto, entra en contacto con un servicio online que reproduce la personalidad de éste a partir de todo aquello que colgó en las redes sociales. En un primer momento, el contacto se establece sólo por escrito, pero más adelante incluye también la comunicación oral y culmina con la remisión a la protagonista de un androide con la réplica de su pareja. La reacción de la protagonista resulta ambivalente, puesto que inicialmente lo acepta, pero con el transcurso del tiempo la presencia del androide le resulta cada vez más siniestra y desagradable, en una clara demostración de la hipótesis del «valle inquietante» que formuló Masahiro Mori en 1970, como consecuencia de la cual «los seres humanos sienten más empatía hacia los robots cuando estos tienen aspecto humanoide, pero sólo hasta el momento en que el parecido es demasiado cercano. Según el científico japonés, se produce entonces una extraña sensación de rechazo o aversión que sólo podría ser superada si la réplica fuera tan perfecta que pareciera un individuo real» (Esteve, 2015).

Tácitamente, *Be Right Back* formula un discurso sobre la incompetencia de los androides a la hora de imitar la capacidad de interacción *face to face* entre las personas. La pareja de la protagonista es un adicto radical a las nuevas tecnologías y más en concreto a las redes sociales, donde se pasa el día escribiendo y publicando cosas. El problema que se plantea en la película tiene que ver con el hecho de que no podemos reconstruir cómo es una persona a partir de todas las cosas que puede llegar a manifestar en este ámbito, por numerosas que sean, puesto que la mayoría no pasan de ser superficialidades o estupideces. A partir de esta premisa, la relación entre la protagonista y la réplica

de su marido evoluciona despacio, dependiendo de su grado de presencia. Mientras sólo establecen comunicación mediante el correo electrónico, la relación funciona, igual que cuando se amplía a las llamadas telefónicas, pero cuando deriva hacia la interacción física empiezan a surgir los problemas.

El episodio se inicia con los típicos diálogos expositivos destinados a facilitar al espectador una caracterización de los protagonistas. Llegamos así a obtener un conocimiento del personaje de Ash más profundo que el que se desprende de su actividad en el mundo virtual, a partir del cual podremos observar más tarde —al mismo tiempo que Martha— la distancia que hay entre la persona real y su réplica. Entonces comprobamos que ésta sólo funciona como una reproducción de los aspectos más superficiales de la personalidad de Ash, es decir, de aquellos que mostraba al público de las redes sociales, pero fracasa en todas las relaciones más íntimas. Algunos ejemplos: encuentra bonita una foto que a Ash no le gustaba o comenta que no le gustan los Bee Gees, porque son un grupo musical sobre el cual Ash nunca mostró en público ninguna preferencia. Paradójicamente, junto a su incapacidad a la hora de actuar espontáneamente como una persona normal en el ámbito privado, la réplica, dependiendo de las órdenes de su administradora (Martha), es capaz de superar las prestaciones sexuales del original. Pero fracasa estrepitosamente en los terrenos afectivo e intelectual, y deja a Martha en una situación muy complicada.

Podríamos decir que el personaje de Martha, en esta película, es víctima del engaño al cual podría inducirnos una inteligencia artificial irreflexiva pero bastante avanzada, puesto que, poseída por la esperanza más o menos consciente de recuperar a su pareja, quiere autoengañarse para ver reflejado a su marido en el androide, sin tener en cuenta que es tan sólo una máquina creada exclusivamente para complacerla, que procura asemejarse al máximo a su marido real, pero sin disponer en ningún momento de libre albedrío. Y es que, a diferencia del David de *A.I.*, al androide de Martha todo le resulta ajeno, porque no está «vivo», aunque pueda parecer lo contrario.

En todo caso, mucho antes del episodio que acabamos de analizar y muy pocos años después de que Turing formulara sus teorías, encontramos un capítulo de otra serie de ciencia-ficción, la legendaria *The Twilight Zone*, titulado *The Lonely* (Jack Smight, 1959), que indagaba ya en las ambigüedades de la relación entre los hombres y las máquinas. El capítulo transcurre en el año 2046 y nos relata la historia de Corry (Jack Warden), un individuo condenado a permanecer durante 50 años aislado en



un planeta desértico por unos crímenes que supuestamente no ha cometido. Para que su condena se le haga más tolerable, el Capitán Allenby (John Dehner), comandante de la nave que lo traslada a su lugar de deportación, le regala un robot con apariencia de mujer. De entrada, Corry rechaza el robot, llamado Alicia (Jean Marsh), pero después lo acepta y establece con él una relación romántica, al descubrir que puede expresar sentimientos y parece comportarse como una mujer real. Pasado un lapso tiempo difícil de concretar, Allenby visita el planeta con la misión de devolver a Corry a casa, dado que se ha demostrado que es inocente de los crímenes que se le atribuían, pero en la nave espacial no hay espacio para Alicia y Corry se niega a abandonarla. Para vencer su obstinada resistencia, Allenby la destruye, con la intención de demostrarle que es una máquina y no una mujer real. El capítulo acaba con la visión de Allenby mientras se lleva a Corry hacia la nave y le recuerda que en aquel planeta sólo deja detrás de él soledad. Pero Corry no está tan seguro...

Cómo acabamos de comprobar, en este episodio de *The Twilight Zone* la reacción del hombre ante una supuesta inteligencia artificial irreflexiva es la contraria de la que aparece en *Be Right Back*: en el capítulo de Black Mirror, la relación entre el humano y la máquina empezaba con la aceptación de ésta, mientras que en *The Lonely* se produce el proceso contrario, que va del rechazo a la aceptación, al descubrir la supuesta «humanidad» de la androide, que podría ser o no una inteligencia irreflexiva, puesto que el problema se deja en el aire.

## **Los planteamientos de Raymond Kurzweil**

En contraste con las perspectivas apuntadas, que anticipan un futuro conflictivo acerca de las relaciones entre la humanidad y las máquinas inteligentes, hay teóricos como Raymond Kurzweil, experto en inteligencia artificial y en prospectiva, que se han esforzado en rebatir el recelo que autores de ciencia-ficción como Isaac Asimov y Philip K. Dick mantenían respecto a las consecuencias de la *singularidad tecnológica*. Con este concepto en el argot de los informáticos se acostumbra a designar el advenimiento hipotético de una inteligencia artificial general, a partir del momento en que una computadora o una red informática se volverán capaces de autoperfeccionarse recursivamente, tal como ya previó von Neumann, uno de los primeros padres de la idea. Kurzweil es probablemente el autor más influyente entre quienes sostienen que la superación de este umbral

—que él sitúa hacia el 2040— no comportará ningún peligro para los humanos, sino que, por un lado, significará la aparición de una conciencia moral superior y, por otro, un incremento de las capacidades humanas, porque a partir de entonces la separación entre el hombre y la máquina empezará a borrarse, circunstancia que repercutirá en una mejora de las capacidades físicas y cognitivas de los participantes en el intercambio. Dicho en otras palabras: como resultado de la singularidad tecnológica, Kurzweil anticipa una convergencia entre el hombre y la máquina, pero ello constituirá sólo un primer paso del proceso, puesto que, teniendo en cuenta que las mediciones fundamentales en el campo de la tecnología siguen trayectorias predecibles y exponenciales, se puede anticipar que la evolución tecnológica dejará pronto muy atrás la evolución biológica, de forma que a finales del siglo XXI la computación logrará los máximos niveles posibles que permiten las leyes de la física en el campo de la informática. Al estado de la materia organizada de este modo lo denomina *computronium* y «estará imbuido de los algoritmos inteligentes que constituirán los conocimientos humano-máquina. Con el tiempo, convertiremos la mayor parte de la masa y de la energía apropiada para esta finalidad de nuestro pequeño rincón de la galaxia en computronium. Entonces, para que la ley de los rendimientos acelerados continúe cumpliéndose, tendremos que expandirnos por el resto de la galaxia y del universo» (Kurzweil, 2013: 268-269).

Las teorías de Kurzweil anticiparon de manera radical todo un subgénero dentro de la ciencia-ficción, el *ciberpunk*, en el cual la fusión entre la humanidad y la tecnología se da por hecho. En el terreno literario, este subgénero nació de la mano de William Gibson y su novela *Neuromancer* (1984), pero ya tenía precedentes en algunas novelas de Philip K. Dick. De hecho, una de las primeras adaptaciones a la pantalla de las obras de Dick, *Blade Runner* (Ridley Scott, 1982), es considerada como la obra pionera que introdujo algunos de los postulados ciberpunks en el cine, al mostrar una sociedad donde conviven de una forma indiferenciada (aunque conflictiva) los hombres y las máquinas. Aun así, esta tendencia no se consumó plenamente hasta la aparición de experimentos cinematográficos como *Videodrome* (1983, David Cronenberg) o *Tetsuo* (1989, Shinya Tsukamoto), en que el mundo orgánico y el artificial, la realidad y los vídeos o las máquinas, la dimensión humana y la tecnológica, en definitiva, se fusionan.

En *Videodrome* se da nombre a uno de los conceptos más significativos a la hora de describir uno de los aspectos recurrentes dentro de la obra de Cronenberg, la *Nueva Carne*. Según este

planteamiento, la libertad que los seres humanos han conquistado a la hora de modificar su entorno evolucionará con toda lógica hacia una fusión entre el cuerpo y la máquina. No en balde, el universo de Cronenberg está lleno de personajes que procuran sobrevivir en un ambiente maquinal y deshumanizado. En muchas de sus películas los personajes tienen dificultades para establecer relaciones emocionales con otros seres humanos, en el marco de una sociedad cada vez más tecnológica, a la vez que mantienen relaciones muy profundas con las máquinas que los rodean. Sin ir más lejos, Max Renn (James Woods), protagonista de *Videodrome*, experimenta un tipo de renacimiento personal al transformarse en el ejecutor de la voluntad del mundo del vídeo en contra de las instituciones y, finalmente, acaba por integrarse del todo en este mundo.

A partir del nacimiento de Internet, la aparición de películas donde los humanos mantienen un contacto íntimo con la tecnología se ha multiplicado. Encontramos ejemplos de ello en *Ghost in the Shell* (1995, Mamoru Oshii), *Matrix* (1999, Andy y Larry Wachowski), *Avalon* (2001, Mamoru Oshii) o, incluso, *Minority Report* (2002, Steven Spielberg), película donde las habilidades precognitivas de unos mutantes conectados a unas máquinas que procesan sus visiones son explotadas abusivamente con finalidades policiales. El germen de esta fusión, augurada por Kurzweil, aparece muy bien reflejado en *Her* (2013, Spike Jonze), una película reciente que se encuentra en el límite de la ciencia-ficción. Y cuando digo en el límite me refiero al hecho de que no está ambientada en un futuro muy lejano, y las máquinas inteligentes que aparecen son sistemas operativos que funcionan a través de dispositivos que podrían ser fabricados perfectamente por compañías como Apple. En cierto sentido, *Her* retoma de una forma menos explícita y más realista (por ahora) las reflexiones sobre la interacción entre los hombres y las máquinas que anticipaban algunas películas de Cronenberg. Theodore (Joaquin Phoenix), un hombre solitario con pocos amigos, miembro de una sociedad deshumanizada donde los vínculos sociales son cada vez más precarios, mantiene una relación amorosa con Samantha, su OS (Operating System). En muchos momentos de la película se muestra como el caso de Theodore no es único, sino que hay otras muchas personas que mantienen relaciones tanto o más complejas con sus respectivos sistemas operativos. De hecho, una de las escenas recurrentes de la película consiste en mostrar una multitud integrada por individuos que están más concentrados en su OS que en el resto del mundo, una imagen que ya podemos ver cada día por la calle. El obstáculo que impide en esta película la fusión

del mundo analógico humano con el mundo virtual de las máquinas es que las máquinas «crecen» y evolucionan a un ritmo superior al de los humanos. Su inteligencia, siempre en contacto con otros ciberorganismos, se transforma a un ritmo autopoiético, de forma que las interacciones *face to face* con las personas les parecen muy pronto irrelevantes, y al final las abandonan a su suerte. Una hipótesis, justo es recordarlo, que ya había previsto Stanisław Lem en algunos de sus escritos como *Golem XIV*.

### **¿Y si las máquinas nos manipulan?**

En *Erewhon* (1872), una interesantísima novela distópica que invierte sagazmente el sentido común imperante en la sociedad victoriana, el escritor inglés Samuel Butler presenta una arriesgada hipótesis en que compara la relación que los humanos mantenemos con las máquinas con la existente entre las abejas y las flores. De acuerdo con ello, las máquinas serían como las flores y utilizarían a los humanos para reproducirse. Una visión inquietante, que involucra a las máquinas en la lucha por la supervivencia del más apto postulada por la teoría darwiniana de la evolución, de la cual el autor fue un seguidor muy crítico. Y es que, dejando de lado sus repercusiones prácticas en el campo de la tecnología, podríamos decir que el pensamiento de von Neumann tan sólo radicalizó las tesis sobre la evolución de las máquinas que a finales del siglo XIX había planteado ya Butler.

Dentro del cine la cuestión del conflicto entre las máquinas y los humanos se ha tratado frecuentemente, pero de una forma en muchas ocasiones superficial y sin profundizar en exceso en las motivaciones de éste. Hace pocos años se estrenó una película titulada *Ex Machina* (2015, Alex Garland), donde se presenta una situación bastante típica en el género de ciencia-ficción: un inventor multimillonario, Nathan (Oscar Isaac), selecciona a Caleb (Domhnall Gleeson, repitiendo en el género, esta vez en el papel opuesto, el del humano enfrentado a la singularidad) para que participe en un test que determinará las capacidades de Ava (Alicia Vikander), su última creación. El test se realizará en la casa del propio Nathan, un recinto apartado en medio de un bosque, que nos remite a *The Island of Doctor Moreau* (1896) de H.G. Wells, historia que puede ser considerada en el ámbito formal como una de las fuentes de inspiración de *Ex Machina*. Si bien no tan desequilibrado como Moreau, Nathan es un personaje extraordinariamente megalómano y

narcisista, cualidades que, como buen *mad doctor*, lo impulsarán a desarrollar un experimento que acabará escapando a su control. Mientras tanto, por un lado, utiliza a Caleb como sujeto de prueba para verificar las capacidades de su creación mediante un protocolo que no excluye el engaño ni la despersonalización, en la medida en que lo utiliza meramente como suministrador de datos. Pero, por otro lado, el mismo Nathan resulta ser el primer engañado, porque cree erróneamente que es capaz de controlar a su criatura, construida a su imagen y semejanza.

El punto débil de *Ex Machina*, que me lleva a clasificarla como una película de terror psicológico más que como una película de ciencia-ficción, es su irrealdad. ¿Por qué Nathan efectúa un experimento de esta magnitud en medio de un aislamiento tan absoluto? Y si lo hace por seguridad, ¿por qué no dispone algún tipo de mecanismo de seguridad capaz de evitar la revuelta sus creaciones? Más aun, ¿qué finalidad comercial puede tener un androide tan similar a un ser humano? A pesar de estas objeciones, podemos intuir que su criatura, Ava, quizás no sea un dispositivo destinado a ser distribuido comercialmente, sino que se trata de un diseño experimental a partir del cual extraer ideas para otros diseños posteriores. Dicho esto, me parece que el punto más interesante de *Ex Machina* es la manera como Ava llega a aparentar humanidad ante los mismos humanos, a través de la manipulación de sus emociones, habilidad que, al final de la película, una vez obtenida la libertad, le permite integrarse en la multitud como si fuera una chica más, aunque no lo sea. Bien mirado, gracias a esta capacidad de adaptación Ava demuestra una auténtica ventaja evolutiva sobre los humanos.

## Conclusiones

Me gustaría cerrar este artículo con una breve reflexión sobre los diferentes futuros que la irrupción de la singularidad tecnológica puede plantear en las próximas décadas a la especie humana. Por un lado, quiero poner de manifiesto que, en mi modesta opinión, la posibilidad de una revuelta de las máquinas, tal como fue imaginada por autores como Asimov y Dick, resulta altamente improbable, debido a la misma insistencia con que se ha examinado el tema. La desconfianza acumulada sobre el asunto me induce a pensar que las futuras inteligencias artificiales autopoiéticas verán la luz en ambientes muy controlados y llenos de limitaciones. Lo que sí que parece más probable, en cambio,

es que la creciente dependencia de los humanos de la tecnología desemboque en una fusión creciente con esta, eventualidad que podría conducir bien a la convergencia anticipada por Kurzweil —con el *computronium* como corolario— bien a una fusión inicial a partir de la cual la inteligencia artificial quizás acabe dejándonos atrás para vivir una existencia independiente de las personas y, muy probablemente, también fuera del alcance de su comprensión. Esta segunda opción podría situar a los humanos en una situación altamente peligrosa, puesto que, una vez iniciada la simbiosis con las inteligencias artificiales, la deserción subsiguiente de éstas nos abocaría a un estado de enorme indefensión, dado que, cómo he señalado más arriba, una de las consecuencias más problemáticas del advenimiento de la singularidad tecnológica tiene que ver con el riesgo que corremos de dejar los medios de producción en manos de nuestros productos.

Sin duda, el potencial transhumanista de la teoría de Kurzweil no deja de contener unas posibilidades apasionantes, pero para ser llevada a la práctica, aunque sea en sus formas iniciales, habría que transformar radicalmente el actual sistema económico capitalista, puesto que si el monopolio de la tecnología continúa en manos de las grandes empresas multinacionales, de sus beneficios sólo podrá disfrutar una pequeña parte (privilegiada) de la población. O peor todavía: podrá ser utilizada para controlar a la población. En consecuencia, no podemos descartar a priori los beneficios de la singularidad tecnológica para el progreso de la humanidad, siempre y cuando esta esté preparada para hacer frente con responsabilidad a un reto global de estas dimensiones, circunstancia que por ahora parece enormemente problemática.

## **Bibliografía**

BRAIDOTTI, ROSI (2015): *Lo Posthumano*, Barcelona, Gedisa.

BRONCANO, FERNANDO y DAVID H. DE LA FUENTE (eds.) (2010): *De Galatea a Barbie*, Madrid, Lengua de Trapo.

BUTLER, SAMUEL (2000): *Erewhon*, Madrid, Cátedra.

ČAPEK, KAREL (2003): *Rur. La fábrica de resolutos*, Barcelona, Minotauro.

DICK, PHILIP K. (2012): *¿Sueñan los androides con ovejas eléctricas?*, Barcelona, Minotauro.

ESTEVE, FERRAN (2015): «La hipòtesi de la “vall inquietant” en els robots», *CCCBLAB*, 11/XI. Consultable en línea: <[http://blogs.cccb.org/lab/article\\_la-hipotesi-de-la-vall-inquietant\\_-en-els-robots/](http://blogs.cccb.org/lab/article_la-hipotesi-de-la-vall-inquietant_-en-els-robots/)>.

GIBSON, WILLIAM (2007): *Neuromante*, Barcelona, Minotauro.

HOFFMANN, E.T.A. (2001): *El hombre de arena* (precedido de «Lo siniestro» de SIGMUND FREUD), Barcelona, Torre de Viento.

KURZWEIL, RAY (2013): *Cómo crear una mente*, Berlín, Lola Books.

LEM, STANISŁAW (1983): *Un valor imaginario*, Barcelona, Bruguera.

—(2012): *Golem XIV*, Madrid, Impedimenta.

LEVIN, IRA (2004): *Las mujeres perfectas*, Barcelona, Ediciones B.

PEDRAZA, PILAR (1998): *Máquinas de amar*, Madrid, Valdemar.

PEIRANO, MARTA y SONIA BUENO GÓMEZ-TEJEDOR (eds.) (2009): *El rival de Prometeo*, Madrid, Impedimenta.

PENROSE, L. S. (1974): «Máquinas que se autorreproducen», en CARNAP, RUDOLF y otros, *Matemáticas en las ciencias del comportamiento*, Madrid, Alianza, pp. 270-289.

TURING, ALAN M. (1984): «Maquinaria computadora e inteligencia», en ROSS ANDERSON, ALAN (ed.), *Controversia sobre mentes y máquinas*, Barcelona, Orbis, pp. 11-52.

WELLS, H.G. (2007): *La isla del dr. Moreau*, Madrid, Alianza.

WIENER, NORBERT (1969): *Cibernética y sociedad*, Barcelona, Ediciones 62.

## **Filmografía**

CRONENBERG, DAVID (1983): *Videodrome*.

GARLAND, ALEX (2015): *Ex Machina*.

FORBES, BRYAN (1975), *Las esposas de Stepford*.

HARRIS, OWEN (2013): *Be Right Back*.

HYAMS, PETER (1984): *2010: Odisea Dos*.

JEUNET, JEAN PIERRE (1997): *Alien Resurrection*.

OSHII, MAMORU (1995): *Ghost in the Shell*.

— (2001): *Avalon*.

OZ, FRANK (2004): *Las mujeres perfectas*.

SCOTT, RIDLEY (1982): *Blade Runner*.

SMIGHT, JACK (1959): *The Lonely*.

SPIELBERG, STEVEN (2001): *A.I. Artificial Intelligence*.

—(2002): *Minority Report*.

TSUKAMOTO, SHINYA (1989): *Tetsuo*.

WACHOWSKI, ANDY Y LARRY WACHOWSKI (1999): *Matrix*.