

Desconfianza estética y conflictos ético-políticos ante el reto de la inteligencia artificial confiable: una propuesta ilustrada con casos cinematográficos recientes (2019-2023)*

Marcos Jiménez González
Universidad de Zaragoza
m.jimenez@unizar.es

Antonio Luis Terrones Rodríguez
Universidad Loyola Andalucía
alterrones@uloyola.es



© de los autores

Fecha de recepción: 25/3/2024
Fecha de aceptación: 16/6/2024
Fecha de publicación: 25/7/2024

Resumen

En 2019, el Grupo de Expertos sobre Inteligencia Artificial, impulsado por la Comisión Europea (CE), publicó *Directrices éticas para una inteligencia artificial fiable*. Este documento, junto con otros, pretende cimentar las bases europeas para una inteligencia artificial (IA) confiable. Es decir, una serie de condiciones indispensables, principios éticos y estrategias que señalan el camino para asegurar un proyecto europeo humanocéntrico, alineado con las necesidades y las expectativas de los seres humanos. En el presente artículo se detecta que esta propuesta se encuentra ante dificultades relacionadas con la imagen esencialmente negativa de la IA, que surge del ámbito de la literatura (complejo de Frankenstein), influye en el imaginario colectivo y se representa actualmente en las artes audiovisuales. De acuerdo con ello, se analizan algunos casos recientes con el objetivo de establecer un diagnóstico y ofrecer posibles soluciones, cuestión que dificulta el camino ético y político que se pretende emprender en el espacio europeo y requiere de una educación estética profunda, de la que actualmente hay una carencia sistémica.

Palabras clave: cine; complejo de Frankenstein; inteligencia artificial; Unión Europea; arte

* El presente estudio se integra en los resultados del grupo de investigación Estética y Filosofía de la Imagen [DGA, H22-17D] y en el proyecto de investigación *Los sótanos de la desinformación: De usuarios a terroristas en la sociedad digital* (DESTERRA), financiado por el Ministerio de Ciencia e Innovación, ref. TED2021-130322B-I00. Asimismo, la publicación es parte del proyecto *Ética cordial y democracia inclusiva en una sociedad tecnolozada* (ETICORDIAL), PID2022-139000OB-C22, financiado por MCIU/AEI/10.13039/501100011033/FEDER, UE.

Abstract. *Aesthetic distrust and ethico-political conflicts in the face of the challenge of trustworthy Artificial Intelligence: A proposal illustrated by recent films (2019-2023)*

In 2019 the Expert Group on Artificial Intelligence, an initiative of the European Commission, published Ethical Guidelines for Trustworthy Artificial Intelligence. This document, along with others, aims to lay the foundations in Europe for trustworthy Artificial Intelligence (AI). A set of prerequisites, ethical principles and strategies point the way to a human-centric European project that aligns with the needs and expectations of human beings. This article argues that this proposal faces difficulties due to the essentially negative image of AI, which has its roots in literature (the Frankenstein Complex), influences the collective imagination, and is currently found in the audiovisual arts. Thus some recent film cases are analysed, with the aim of establishing a diagnosis and offering possible solutions. A question that hinders the ethical and political path that is intended in the European space and requires an extensive aesthetic education, which is currently systemically lacking.

Keywords: cinema; Frankenstein Complex; artificial intelligence; European Union; art

Sumario

- | | |
|--|---|
| 1. Introducción | 4. El problema de Frankenstein en el cine: casos de estudio |
| 2. El deseo de una inteligencia artificial confiable en el espacio europeo | 5. Conclusión |
| 3. Imagen y representación de la inteligencia artificial | Referencias bibliográficas |

1. Introducción

A pesar de los importantes beneficios que ha suministrado en numerosos ámbitos y al dominio del pensamiento solucionista en las sociedades tecnolizadas (Morozov, 2016), es cada vez más evidente que la inteligencia artificial (IA)¹ no constituye una «solución mágica», pues conlleva importantes riesgos, debido a sus efectos controvertidos y ambivalentes. Una muestra del reconocimien-

1. El origen de este término se sitúa en los años cincuenta, específicamente durante la celebración del Darmouth Summer Research Project on Artificial Intelligence (McCorduck, 1991). Existen numerosas definiciones de IA originadas en el seno de diversas perspectivas científicas, no obstante, con el propósito de esclarecer conceptualmente esta tecnología altamente compleja, se ofrece la siguiente definición: «[...] consideramos como IA cualquier tecnología de software con al menos una de las siguientes capacidades: percepción, que incluye audio, visual, textual y táctil (p. ej., reconocimiento facial), toma de decisiones (p. ej., sistemas de diagnóstico médico), predicción (p. ej., pronóstico del tiempo), extracción automática de conocimiento y reconocimiento de patrones a partir de datos (p. ej., descubrimiento de circulación de noticias falsas en las redes sociales), comunicación interactiva (p. ej., robots sociales o bots de chat) y razonamiento lógico (p. ej., desarrollo de teorías a partir de premisas). Este punto de vista abarca una gran variedad de subcampos, incluido el aprendizaje automático» (Vinueza et al., 2020: 1-2).

to de la complejidad de esta tecnología disruptiva a partir de una perspectiva de riesgo es la reciente Ley de IA de la Unión Europea (Comisión Europea, 2021), el primer marco jurídico aprobado por la alianza de países sobre este sistema sociotécnico. Además, en respuesta a la creciente conciencia científica de los desafíos inducidos por la IA, se han expresado diversos llamados para la integración de la ética. De este modo, se han propuesto una IA responsable (Université de Montréal, 2017; Dignum, 2019), una IA ética (Floridi et al., 2018), una IA beneficiosa (Future of Life Institute, 2017), una IA alineada con los objetivos del desarrollo sostenible (Vinuesa et al., 2020) o incluso una IA sostenible (Van Wynsberghe, 2021).

Desde que, en 2018, el Grupo de Expertos de Alto Nivel sobre IA (AI HLEG, siglas en inglés de High-Level Expert Group on Artificial Intelligence), impulsado por la CE, subrayó la importancia de articular estrategias para promover un ecosistema de IA fundamentado en la confianza, se han invertido importantes esfuerzos en diversos ámbitos, como el industrial, el científico, el educativo o la Administración pública. Si bien la búsqueda de mecanismos técnicos es fundamental para asegurar la propuesta europea, es conveniente atender a otros aspectos muy importantes en el terreno político. El carácter normativo del proyecto humanocéntrico europeo en torno a la inteligencia artificial confiable requiere evaluar las estrategias desplegadas para el logro de los objetivos planteados por los expertos.

En general, la comunicación visual y, en particular, las imágenes transmitidas sobre la IA en el cine son fundamentales para el establecimiento de un ecosistema ético asentado en la confianza. La imagen transmitida a la opinión pública convierte a esta tecnología disruptiva en un fenómeno social y cultural (Romele, 2022). A este respecto, consideramos que existe una brecha entre la concepción ética de la IA y sus modos de representación en imágenes, especialmente en el ámbito cinematográfico. Además, apreciamos la cultura, en particular la cinematográfica, como una manifestación esencial para reflexionar sobre las implicaciones de la automatización tecnológica e invitar a la ciudadanía a cuestionar los imaginarios y el lenguaje público común (Pasquale, 2024: 312-313). En un mundo en el que el relato que equipara el progreso con la tecnología está profundamente arraigado en el imaginario colectivo de Occidente e integrado a través de la tecnocracia en las grandes ideologías políticas (Nowotny, 2022: 108), ciertamente, es conveniente pensar hasta qué punto las producciones cinematográficas sobre IA pueden contribuir a generar un ecosistema basado en la confianza a través de una mirada crítica.

Por tanto, el objetivo principal de este trabajo consiste en examinar la aceptabilidad de las producciones cinematográficas sobre IA mayormente consumidas en territorio europeo, pues, aunque el cine constituye una actividad insuficiente para el cultivo de la confianza, desempeña un papel fundamental. Para alcanzar este propósito, nuestra investigación se estructura de la siguiente manera: en primer lugar, destacamos el marco desiderativo asumido en el espacio europeo a partir de la IA confiable, examinando los entre-

sijos de la confianza en el terreno tecnológico; en segundo lugar, subrayamos la importancia de la imagen y la representación, indicando que la idea mental sobre la IA se ha conformado a partir de un imaginario ficticio (cinematográfico o audiovisual) y de premisas epistemológicas erróneas; en tercer lugar, analizamos un conjunto de producciones cinematográficas fundamentadas en el complejo de Frankenstein, puesto que, a partir de diferentes tramas narrativas, promueven el conflicto entre el humano y la máquina en el imaginario colectivo.

2. El deseo de una inteligencia artificial confiable en el espacio europeo

En diciembre de 2018 el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial expuso un primer borrador del documento *Directrices éticas para una inteligencia artificial fiable*, iniciativa que fue definitivamente presentada en abril de 2019. En este trabajo se aboga por un enfoque humanocéntrico, articulado esencialmente a partir del concepto *inteligencia artificial confiable*, con el que se insiste en la necesidad de alinear el desarrollo de esta tecnología con las necesidades y los intereses del ser humano (Bryson y Theodorou, 2019). Para este propósito los expertos concretan un conjunto de condiciones indispensables que posibilitan la confianza y señalan una serie de principios éticos sustentados en los derechos fundamentales de la UE: respeto a la autonomía humana, prevención del daño, equidad y explicabilidad. En la misma senda, más tarde, el grupo de expertos entregó un documento de orientaciones bajo el título *Recomendaciones de política e inversión para una IA fiable*, planteando estrategias específicas para la articulación de políticas y regulaciones. En particular, apuntan a ámbitos europeos, enfatizando la competencia, la diversidad y la inclusión, la educación, el medio ambiente, la infraestructura, la inversión y los derechos laborales. En tercer lugar, fue desarrollada una lista de evaluación para suministrar una herramienta operativa de las directrices, facilitando a los usuarios una experiencia beneficiosa mediante una adaptación práctica de los requisitos claves para la confianza y los principios éticos. Finalmente, también es posible conocer el proyecto europeo basado en la confianza a través del *Libro blanco sobre inteligencia artificial: Un enfoque europeo orientado a la excelencia y la confianza* (Comisión Europea, 2020).

A partir de estas iniciativas europeas la confianza ha ocupado una parte importante de la literatura centrada en los intelectos sintéticos (Floridi, 2021; Liang et al., 2022; Li et al., 2023; Terrones Rodríguez y Rocha Bernardi, 2023). El grueso de esta producción ha servido para evidenciar que nos encontramos ante un concepto complejo como resultado de los diversos espacios en los que es integrada, ya que existen diferencias notables entre un diagnóstico clínico, una evaluación crediticia, la conducción autónoma de un vehículo o la predicción de la reincidencia de delitos en el sistema penitenciario. Al margen de las evidentes diferencias, es importante comprender los entresijos de un concepto que constituye un referente normativo para el ecosistema científico y tecnológico de los Estados miembros.

Joseph C. Pitt (2010) ha dedicado sus esfuerzos a analizar las condiciones de posibilidad del surgimiento de la confianza. Según Pitt, la confianza surge a partir de una promesa articulada en el seno de la comunicación lingüística. En su exposición afirma que una promesa entraña conocimiento en el contexto de las relaciones sociales, de tal forma que desplaza a los animales de estas relaciones y sitúa la confianza como una cuestión esencialmente humana y, por ende, interpersonal (2010: 448). En lo respectivo a la tecnología, el filósofo norteamericano asegura que no es posible confiar en ella, en vista de que es exclusivamente un medio utilizado por los humanos. Por consiguiente, Pitt considera que la confianza hunde sus raíces en la comunicación y en el conocimiento, insistiendo en que es un asunto que atañe a las personas.

S. D. Noam Cook (2010) también se ha ocupado de analizar la confianza, aunque en un sentido diferente a Pitt. En su caso, ha situado la reflexión en el ámbito de la experiencia humana presente en las relaciones que implican expectativas mutuas (2010: 455). El cumplimiento de las expectativas que las personas tienen entre sí es uno de los pilares indispensables que sostienen las relaciones. Por tal razón, este cumplimiento es una condición *sine qua non* para el florecimiento de la confianza. Mientras que Pitt dirige su atención al componente lingüístico, Cook se sitúa en la estela de la praxis, apreciando que la promesa es un elemento resultante de la confianza y no al contrario. Conforme a esta secuencia, Cook articula un enfoque de la confianza situado en la práctica y en la experiencia humana. Pese a las diferencias, tanto Pitt como Cook aseguran que la tecnología se encuentra integrada en un entramado de relaciones condicionadas por diversas circunstancias contextuales, imposibilitando que pueda ser depositaria de confianza.

El reconocimiento de la tecnología como un asunto humano implica asumir que la IA constituye un sistema sociotécnico, producto de un conglomerado de materiales, prácticas, ideología, agencia, economía, instituciones, etc. (Sartori y Theodorou, 2022) que dan forma a una unidad sistemática sostenida en diversas relaciones. En este sentido, y recogiendo el testigo de Pitt y Cook, no es adecuado afirmar que un determinado producto tecnológico es, en sí mismo, depositario de confianza. De cualquier forma, en virtud de la posición que ocupa la tecnología en una red de relaciones, lo razonable es proyectar la confianza hacia el sistema en su conjunto. En consecuencia, la comprensión de la confianza precisa asumir el peso de las relaciones y las cualidades de los sistemas sociotécnicos:

La confianza en el desarrollo, despliegue y utilización de sistemas de IA no concierne únicamente a las propiedades inherentes a esta tecnología, sino también a las cualidades de los sistemas sociotécnicos en los que se aplica la IA. De forma análoga a las cuestiones relativas a la (pérdida de) confianza en la aviación, la energía nuclear o la seguridad alimentaria, no son simplemente los componentes del sistema de IA los que pueden generar (o no) confianza, sino el sistema en su contexto global. Por lo tanto, los esfuerzos dirigidos a garantizar la fiabilidad de la IA no solo atañen a la confianza que suscita el propio sistema de IA, sino que requieren un enfoque integral y sistémico que

abarque la fiabilidad de todos los agentes y procesos que forman parte del contexto sociotécnico en el que se enmarca el sistema a lo largo de todo su ciclo de vida. (Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, 2019a: 6)

Por tanto, ¿qué vínculo mantienen la tecnología y la confianza? Philip J. Nickel et al. (2010) han investigado la dificultad de la confianza y su proyección en el espacio tecnológico, señalando que, si bien es un elemento muy presente en la tecnología, se manifiesta en un sentido derivado. El enfoque sistémico apunta a la red de promesas en relación con el funcionamiento y los resultados de la IA, motivando que la confianza no sea depositada en los artefactos tecnológicos, sino en los agentes humanos que tejen las relaciones y proyectan los discursos que, posteriormente, son traducidos en imaginarios, como se examinará a continuación, tomando como muestra un conjunto de producciones cinematográficas que objetivan la dominación de la máquina sobre el ser humano. Por tanto, es conveniente analizar la aceptabilidad de algunas iniciativas centradas en el potencial de la IA, como son los filmes ampliamente consumidos por el público europeo, con el interés de identificar qué elementos pueden contribuir al proyecto humanocéntrico europeo o dificultarlo.

3. Imagen y representación de la inteligencia artificial

Después de este recorrido, se entiende que la confianza debe ser concebida como un aspecto moral presente en el entramado relacional de los sistemas sociotécnicos². La relación desconcertante mantenida en la actualidad con la IA surge, en gran medida, del prejuicio ontológico-relacional que trata al ser humano y a la tecnología como elementos ajenos entre sí, situación que suele fomentarse en notas de prensa, noticias y reportajes. Un ejemplo muy representativo es el reciente documental *Lo desconocido: Robots asesinos* (Sweet, 2023), en el que se analiza el desarrollo de la IA, su aplicación militar y la influencia del cine en la configuración de su imagen pública. Por lo tanto, un punto de partida fiable para conocer las contradicciones de las relaciones con la IA se encuentra en sus diversas representaciones, no solo las relativas a su avance y desarrollo, sino también las representadas en los imaginarios creados en la ficción.

Bajo este prisma, es importante esclarecer si el tipo de imágenes que han sido transmitidas de la IA contribuyen notablemente al alcance de los objetivos del proyecto europeo. En este sentido, la comunicación científica y, en particular, la comunicación visual sobre la IA son fundamentales de cara a configurar un ecosistema ético preocupado por la confianza. Según Alberto Romele (2022), la ética aplicada a la IA no ha mostrado suficiente interés en la comunicación. El autor considera que la comunicación sobre la innovación

2. En *Meditación de la técnica* (2015), José Ortega y Gasset sostuvo que una de las características principales del ser humano es su condición técnica. Actualmente se presenta otro escenario en el que este asunto debe ser entendido de acuerdo con los avances tecnológicos.

en IA contribuye a configurar expectativas o imaginarios. En virtud de esta idea, es conveniente pensar cómo es comunicada la IA, es decir, qué imagen es transmitida a la opinión pública, pues esta tecnología, al margen de ser considerada un sistema sociotécnico, se ha convertido en un fenómeno social y cultural (Romele, 2022: 2-3). Las conclusiones de Romele son translúcidas al respecto: hay una brecha entre la concepción ética de la IA y sus modos de representación en imágenes:

While the ethics of science communication generally promotes the practice of virtues like modesty, humility, sincerity, transparency, openness, honesty, and generosity, stock images and other popular visual representations of AI are arrogant, pompous, and overconfident. [Mientras que la ética de la comunicación científica suele promover la práctica de virtudes como la modestia, la humildad, la sinceridad, la transparencia, la franqueza, la honestidad y la generosidad, las imágenes de archivo y otras representaciones visuales populares de la IA son arrogantes, pomposas y demasiado seguras de sí mismas.] (2022: 16)

Estas palabras introducen una de las premisas fundamentales para nuestra investigación; a saber, que existe una escisión entre la actitud que presenta la sociedad ante el incremento y el desarrollo de la IA en numerosas esferas de la vida y la forma que adopta su contemplación y concepción estética. Dicha escisión es la que subyace en los argumentos de Romele, concretamente cuando recurre a *El reparto de lo sensible*, de Rancière (2014), como marco teórico básico. Si se entiende que la comunicación en imágenes es efectivamente distinta a la comunicación oral o escrita, tal y como sostienen los padres del giro icónico (Boehm, 1994) o del giro pictorial (Mitchell, 1994), se podrá afirmar que los «puntos ciegos» a los que se refiere el autor surgen al conjugarse dos modos de valoración: la conceptual y la visual. Este fenómeno no ocurre solo con las representaciones visuales de la IA en imágenes de archivo, sino que, desde hace años, el mundo de la ficción, sobre todo de la mano del cine y de las series, ha creado una imagen de ella esencialmente negativa. Esta imagen entorpece la convivencia del ser humano con esta tecnología, dificultando el surgimiento de la confianza y ocasionando que las distintas representaciones de la IA, tanto en el cine como en la publicidad o en la red, puedan resultar problemáticas para el proyecto humanocéntrico por el que apuesta la UE. A este respecto, las representaciones de la IA en el cine no difieren mucho del análisis realizado por Romele (2022), cuyos resultados se centran más en las redes sociales, pero inciden, de igual modo, en nuestra investigación, ya que existe una amplia diferencia entre el modo de imaginar la IA y la organización que debe adoptar este sistema sociotécnico en el marco desiderativo de la UE. En consecuencia, el ecosistema fundamentado en la confianza puede encontrarse con la dificultad de un discurso comunicativo y estético desfavorable para su propósito ético. Por lo tanto, si el planteamiento europeo solicita un conjunto de condiciones indispensables para la confianza, el marco de la ficción que engloba películas sobre IA consumidas por el público europeo fomenta imaginarios controvertidos y requiere una comprensión crítica.

En un reciente artículo, Saffari et al. (2021) realizan un estudio detallado de la representación de los robots en el cine y de su posible influencia en la relación establecida con los robots existentes en la vida real. Además de enumerar las características de los robots de la ficción de acuerdo con su apariencia, interacción y comportamiento ético-social, dan cuenta de una serie de conexiones y disonancias entre el marco de la ficción y el de la realidad, llegando a conclusiones interesantes y contradictorias en estos términos. Una de ellas señala que mientras que en la vida real los robots y la IA no aspiran a una equiparación en materia de derechos y libertades similares a los de los seres humanos, debido a su función esencialmente instrumental, los robots ficticios sí presentan esta ambición. Para los investigadores, este fenómeno responde a un interés exclusivamente narrativo (Saffari et al., 2021: 6). Según esta consideración, la rebelión del robot ficticio contra el ser humano se encuentra motivada por un argumento basado en la fuerza estética, que, aunque funciona narrativamente, muestra una situación muy alejada de la realidad actual de la IA. Con todo ello, surge un interrogante: ¿en qué medida la imagen del robot autónomo rebelde adoptada por la ciencia ficción afecta negativamente al imaginario colectivo sobre la IA, dificultando el establecimiento de una concepción asentada en la confianza, esencial para el proyecto humanocéntrico europeo?

El principal problema consiste en que la imagen mental consolidada sobre los intelectos sintéticos se ha configurado a partir de un imaginario ficticio (cinematográfico o audiovisual) y de premisas epistemológicas erróneas, por ejemplo, la de presuponer que la IA posee un estatus cognitivo similar a la inteligencia de las personas (Larson, 2022). El problema de la semejanza con las capacidades o con la apariencia humana siempre ha sido sustancial, desde que se debatía la creación de máquinas antropomórficas³, hasta la llegada de otros factores, como la equiparación y la superación de la IA a la humana. La introducción de la IA en diversos espacios de la vida humana suscita numerosos desafíos. Aquí analizaremos una de las causas de la disociación entre la IA y su representación, profundizando en los imaginarios estéticos de los que partimos a la hora de concebirla.

4. El problema de Frankenstein en el cine: casos de estudio

La imagen amenazante presentada mayoritariamente en el cine sobre la IA se encuentra fundamentada en el imaginario de la creación de vida artificial y se muestra en un conjunto importante de películas donde se representan máqui-

3. A principios del milenio el debate no era el mismo que en la actualidad, pues versaba sobre los conflictos morales que generaba una IA que presentara una apariencia similar a la humana, cuando existen notorias diferencias ontológicas. Jose Luis Molinuevo (2004), basándose en las declaraciones de Peter Weibel (Post y Mulder, 2000), mencionó que gran parte del rechazo existente hacia «los nuevos seres artificiales» residía, en gran medida, en que eran «indiscernibles del ser humano» (p. 113). Una idea similar a la sostenida por Saffari et al. (2021: 5).

nas (Balló y Pérez, 2010: 277). Desde las más conocidas y antiguas, como *Metrópolis* (Lang, 1927), *2001: Una odisea del espacio* (Kubrick, 1968) o *Blade Runner* (Scott, 1982), hasta las más recientes, como *Ex machina* (Garland, 2014) o *Megan* (Johnstone, 2022), estos filmes muestran un contexto de oposición conflictiva entre un ser creado frente a su creador, beben de obras literarias como *Frankenstein, o el moderno Prometeo* (Shelley, 2011) y establecen una analogía con el mito clásico en lo referente a un espíritu de libertad y rebeldía⁴. Entre las múltiples interpretaciones y metáforas existentes (Hernández de la Fuente, 2015: 197), es importante destacar la adaptación moderna, llevada a cabo por Mary Shelley en *Frankenstein* (2011), pues constituye la primera obra en la que se muestra el imaginario de la creación de vida artificial o segunda creación, envuelto en los conflictos morales generados por el mito prometeico⁵.

4. El término que también ha sido utilizado para designar a este imaginario es igualmente denominado *segunda creación*, debido al establecimiento de una analogía entre la creación divina o titánica de la humanidad —de acuerdo con algunas versiones del mito, Prometeo fue quien creó a los hombres a partir del barro (García Gual, 2022: 206)— y la creación humana de la máquina, que suscita el planteamiento de interrogantes morales en torno a los límites de las personas y el alcance de su libertad. El mito de Prometeo es uno de los más ricos en cuanto a la representación de la condición humana, su dimensión técnica e intelectual (Jaran, 2021: 16) y, sobre todo, de la aspiración a equipararse a su creador, en términos de poder. Generalmente, la idea aceptada sobre Prometeo consiste en que se situó junto a los hombres entregándoles el fuego y potenciando en ellos todas las capacidades representantes del progreso técnico. Mediante este ejercicio de dotación, el ser humano puede asemejarse más a los dioses que a los animales, y ocupar una posición intermedia e híbrida entre dos espacios.
5. La similitud existente entre el mito de Prometeo y el doctor Frankenstein se origina a partir del fenómeno de la creación de vida, mediante la suplantación de funciones que tradicionalmente corresponden a Dios, como es posible apreciar en la película *El doctor Frankenstein* (Whale, 1931). Respecto a esta cuestión, Diercyck Barrena asegura lo siguiente: «El nuevo Prometeo de Mary Shelley [...] es un simple mortal que supera los límites permitidos recurriendo únicamente al conocimiento racional [...] El científico será, por lo tanto, la nueva figura que, a través del conocimiento de la tecnología, consigue infundir vida. Se unifican así en la novela los dos significados del fuego (conocimiento tecnológico y vida) en una misma figura» (2023: 9). A través de la unión de restos de muchos cadáveres en un solo cuerpo, Víctor Frankenstein consigue dotar de vida a un monstruo por el que debe asumir responsabilidad, en vista de unas consecuencias morales abrumadoras. En primer lugar, el doctor Frankenstein se siente culpable por la creación impropia de un ser vivo; posteriormente, considera que debe tutelar al monstruo, estableciendo, de ese modo, un vínculo de responsabilidad afectiva no deseada con la criatura; por último, experimenta una gran culpabilidad por haber creado un ser solitario e inadaptado socialmente, así como por causar una rebelión contra su padre y contra la humanidad, similar a la del titán Prometeo contra Zeus. Mientras que Prometeo se posicionó del lado de los humanos, la criatura se situó con los monstruos, con los diferentes. La explicación detallada de la metáfora del mito de Prometeo responde a la necesidad de comprender los elementos esenciales que configuran el relato y la representación de los robots autónomos en el medio audiovisual, pues entendemos que los creadores de las inteligencias artificiales ocupan una posición similar a los científicos o los dioses, al igual que los robots y las máquinas en el caso del monstruo.

En la década de 1940, en el origen de su *serie de los robots*⁶, Isaac Asimov detectó la excesiva presencia de este fenómeno en la narrativa y lo denominó «complejo de Frankenstein» —*Frankenstein complex*—, en referencia a la preocupación existente en el ser humano a ser superado por la tecnología y los avances técnicos que son fruto de su creación. Este complejo consiste en la paranoia generada por el terror a la rebelión de los robots. Asimov estableció las tres leyes de la robótica con el propósito de reducir el influjo de este complejo en el espacio narrativo, a pesar de que ni él mismo pudo evitarlo en sus relatos, tal y como sostuvo ya en los años ochenta Beauchamp (1980)⁷. Además de estar presente en las historias del propio Asimov, que intentó evitarlo, este complejo va más allá de la narrativa, de la literatura o del audiovisual, llegando a permear el dominio sociopolítico y a instalar una determinada concepción en el imaginario colectivo, en uno de los múltiples ejemplos en los que la ficción se entremezcla con la realidad.

En *El robot en el cine: De la obediencia a la rebelión* (Jiménez González, 2019) se analizaron las relaciones entre el mito de Prometeo y la representación de los robots, androides y ginoides en el cine. En este texto se hace hincapié en que la faceta rebelde de las inteligencias artificiales, como HAL 9000, de *2001: Una odisea del espacio*, o Ray, el Nexus 6 de *Blade Runner*, presenta una gran fuerza estética y potencia el escenario tantas veces representado de la rebelión de las máquinas, dado que el paso natural después de la rebelión individual tiende hacia una rebelión colectiva, tal y como ocurre en *Terminator* (Cameron, 1984), *Matrix* (Wachowsky y Wachowsky, 1999) o *Yo, robot* (Proyas, 2004). Además, en la misma obra se anticipa un interrogante que se desarrolla en el presente trabajo, a saber, si la producción audiovisual actual potenciaría la misma premisa de oposición entre el humano y la máquina, inspirada en el complejo de Frankenstein (Jiménez González, 2019: 79). Para despejar esta cuestión, a continuación se analizarán algunos de los filmes más recientes que cuentan con el protagonismo de la IA.

El componente narrativo de oposición entre el ser humano y las máquinas se observa en filmes que siguen la línea de rebelión de una IA al estilo de HAL

6. Se denomina así al conjunto de relatos que Asimov escribió desde la década de 1940 hasta la mitad de la de 1980. Entre otros, destaca *Yo, robot* (2009), donde se ubican *El círculo vicioso* o *Razón*.
7. Aunque Asimov defendía que en sus relatos y novelas los robots no eran «pseudohombres» que se rebelaban contra su creador, este autor defiende lo contrario: «Asimov in *I, Robot* and several of his other robot stories actually reinforces the Frankenstein complex—by offering scenarios of man's fate at the hands of his technological creations more frightening.» [«Asimov, en *Yo, robot* y en otras de sus historias de robots, refuerza en realidad el complejo de Frankenstein, ofreciendo escenarios del destino del hombre a manos de sus creaciones tecnológicas más aterradoras.»] (Beauchamp, 1980: 85). En las conclusiones llega a decir que «if my reading of Asimov's robot stories is correct, he has not avoided the implications of the Frankenstein complex, but has, in fact, provided additional fictional evidence to justify it.» [«Si mi lectura de las historias de robots de Asimov es correcta, no ha evitado las implicaciones del complejo de Frankenstein, sino que, de hecho, ha proporcionado pruebas ficticias adicionales para justificarlo.»] (Beauchamp, 1980: 93).

9000 en 2001: *Una odisea del espacio*, donde una tecnología domótica representada en una nave espacial pretende destruir a los navegantes. Bajo esta premisa, durante los últimos años han sido estrenados algunos filmes cuyo centro narrativo consiste en la rebelión de una casa inteligente —*smarthouse*— contra los inquilinos, como se puede apreciar en *Margaux* (Miller, 2022) o *Bigbug* (Jeunet, 2022), ambas estrenadas en Netflix y producidas en Estados Unidos y Francia, respectivamente. A pesar de la similitud narrativa existente entre estas dos producciones, se perciben una serie de diferencias en lo que respecta a los conflictos entre el humano y la máquina. *Margaux* ha sido concebida como un *slasher* clásico, en el que unos amigos habitan una casa rentada durante un periodo de descanso. La diferencia entre esta y otras películas con una trama similar, como *Viernes 13* (Cunningham, 1980) o *Scream* (Craven, 1996), radica en la naturaleza de la amenaza, pues en el caso de *Margaux* se representa a través de sistemas domóticos. Esta situación motiva en el espectador el establecimiento de una representación negativa y la identificación de los medios tecnológicos con una figura amenazante, equiparable a la de un asesino en serie en las películas de terror clásicas. En cambio, la relación entre los seres humanos y la domótica adquiere una connotación distinta en *Bigbug*, dado que el enfrentamiento surge debido a un exceso de seguridad por parte de la IA. A causa de una situación de emergencia, los inquilinos no pueden evacuar su domicilio, debido a que las salidas han experimentado un cierre automático. Pese a las notables diferencias entre *Margaux* y *Bigbug*, hemos identificado el influjo de un elemento esencial en torno a la convivencia conflictiva entre los humanos y la tecnología que conduce al enfrentamiento⁸.

También hemos reconocido otros productos cinematográficos situados en la frontera entre los androides y la domótica, como pone de manifiesto *Muñeco diabólico* (en inglés *Child's Play*) (Klevberg, 2019) —recreación de la versión de 1988—, donde el muñeco Buddi es, además de un juguete para niños, una herramienta de domótica diseñada con IA que ofrece soporte para la realización de las tareas básicas del hogar, estableciendo, de ese modo, una evidente diferencia con la versión de la década de 1980. El filme de 2019 mantiene incuestionables diferencias respecto a la original (Tom Holland, 1988), en la que Chucky encarna a un muñeco poseído por el demonio. En esta misma senda se encuentra *Megan*, un filme con elementos propios de la década de 1980, como es el caso de los muñecos diabólicos, pese a que son introducidos aspectos novedosos de carácter tecnológico, en particular, bajo la influencia de la IA. Megan es una ginoide diseñada por Gemma, especialista en robótica, para acompañar a su sobrina Cady, una menor que sufre el prematuro fallecimiento de la madre. La historia de Megan y Cady posee elementos semejantes

8. Una diferencia de este filme respecto al anterior es que no categoriza a los seres humanos como objetos positivos ni a las máquinas como negativos, sino que ambos pueden pertenecer a los dos extremos, respectivamente, por lo que la trama presenta una complejidad mayor. No obstante, el contexto se sitúa en un futuro de carácter totalitario y dominado por una sociedad en la que las máquinas controlan el mundo, razón por la que el complejo de Frankenstein está muy presente.

a los representados al inicio de *Muñeco diabólico*, pues coincide en la maldad progresiva que adquiere el muñeco. No obstante, a diferencia de Chucky, Megan, al igual que Buddi en la recreación de 2019, cuenta con IA en su diseño, de manera que la desconfianza generada tras su desobediencia es posible asociarla de un modo más directo al complejo de Frankenstein⁹.

Estos recursos cinematográficos constituyen una muestra para señalar las diversas formas adoptadas por las producciones y permiten subrayar el elemento esencial que permanece, en este caso, del complejo de Frankenstein. Si durante la segunda mitad del siglo xx y los primeros años del xxi este complejo fue expresado mediante la figura de andróides y gínoides que reivindicaban un estatus político y moral equiparable al de sus creadores humanos, en la actualidad son planteadas, además, nuevas formas de representación¹⁰, integrando a la IA como un sistema sociotécnico superior al ser humano en gran cantidad de procesos cognitivos. A este respecto, es importante destacar que, si bien se utilizan mecanismos narrativos muy atractivos para reflejar el despliegue de los sistemas tecnológicos, como sostiene Saffari et al. (2021), se motivan una serie de concepciones sobre las tecnologías disruptivas, en general, y la IA, en particular, que precisan un análisis más profundo. Entre estas concepciones es importante identificar las que manifiestan preocupación por una posible rebelión de las máquinas, pese a que este problema se encuentra muy alejado de los desafíos reales que las sociedades actuales afrontan respecto a la IA.

En otras películas el problema no radica en la asunción del complejo de Frankenstein como hilo conductor, sino en que el escenario de la rebelión de las máquinas constituye el punto de partida, es decir, se instala en el imaginario colectivo como una evidencia a la que el ser humano está destinado con motivo del avance de la IA. En la línea de *Terminator* o *Matrix*, referencias ambientadas en un futuro distópico en el que los humanos llevan años en guerra contra las máquinas, algunas producciones recientes optan por asumir incondicionalmente que el actual desarrollo de la IA conducirá a un escenario similar al de los filmes mencionados. En la película de las hermanas Wachowski, las máquinas, en cualquiera de sus formas, simbolizan el elemento negativo de la representación, mientras que los humanos supervivientes —los habitantes de Zion— representan el elemento positivo. De este modo, se genera una narrativa binaria que opone al ser humano y a la máquina, sin dar oportunidad a la convivencia entre unos y otros, pues el argumento de la saga está inspirado en el enfrentamiento entre ambas partes. Bajo esta ambientación en un futuro

9. En *Los Mitchell contra las máquinas* (Rianda y Rowe, 2021) hay una escena en la que los electrodomésticos cobran vida y atacan a los seres humanos. A pesar de ser una película de animación y de comedia, esta escena está construida con elementos tenebrosos y culmina con la rebelión de varios Furbby, unos muñecos populares durante las décadas de 1980 y 1990 que presentan las mismas connotaciones que Chucky, Buddi o Megan.

10. La representación de andróides sigue también presente, aunque ya no sea la única manifestación. Es posible observar este elemento en la película *Alpha Test* (Mirtes, 2020) o en *Wife Like* (Bird, 2022), donde las figuras que potencian el complejo de Frankenstein son andróides y gínoides, herederos de filmes e historias clásicas, como *Blade Runner* o *Yo, robot*.

en el que los seres humanos están en guerra con las máquinas se encuentran *Jung_E* (Sang-ho, 2023) y *The Creator* (Edwards, 2023). A pesar de estar situadas en este contexto¹¹, destacan ciertos avances en lo relativo a la convivencia con la IA en relación con otras producciones de finales de la década de 1990 y principios de la de 2000, como la trilogía de *Matrix*.

En el mismo escenario de crisis climática, en el filme *Jung_E* los seres humanos fundan colonias espaciales para sobrevivir y asegurar una convivencia pacífica con las máquinas. No obstante, esta película recurre al complejo de Frankenstein para fundamentar la situación de guerra, dado que las máquinas deciden rebelarse y formar repúblicas independientes en tres de las colonias. La diferencia básica con otras producciones que parten de esta premisa radica en que la narrativa no juzga esta situación. Por el contrario, es aceptada con el propósito de potenciar la relación entre madre e hija (las protagonistas) e insistir en los conflictos morales que despierta la emulación del cerebro y la identidad de un ser humano en un soporte mecánico. Así pues, en este caso particular, la rebelión de las máquinas configura el contexto, sin la necesidad de determinar el epicentro narrativo, que pivota sobre los conflictos identitarios ocasionados por el volcado de la personalidad en un cuerpo robótico.

The Creator presenta una mayor complejidad, ya que refleja la posibilidad de una sociedad en la que la IA está integrada en los modos de vida humanos, quedando la oposición entre humanos y máquinas reducida a ciertos sectores de la población. La historia expone un mundo dividido en dos bloques, encabezados por Estados Unidos y China, en los que el primero ha prohibido la investigación y el uso de la IA, debido a la explosión por accidente de una bomba nuclear en Los Ángeles, y el segundo opta por el avance y la convivencia con ella. El escenario, entonces, no es de conflicto entre el ser humano y la IA, sino entre los seres humanos que conviven con ella y aquellos que no. El problema sería, pues, humanocéntrico, dejando más abierta la posibilidad a la convivencia, como de hecho hace en la película el bloque oriental, que demuestra ser más tolerante y estar dispuesto a encontrar un acuerdo para la paz. En este sentido *The Creator* abre un camino de reflexión poco explorado en la narrativa de ciencia ficción hasta el presente: el cuestionamiento de que los problemas generados por la IA son responsabilidad humana, en última instancia, tal y como se advierte en el propio filme: «La bomba nuclear de Los Ángeles fue un error de código, ¿lo sabías? Un error humano» (Edwards, 2023: 71'41").

Una vez examinadas diversas producciones cinematográficas y analizados los conflictos existentes, hemos formulado el siguiente diagnóstico. A pesar de

11. Cabe mencionar que en los últimos años muchas producciones han elegido este tema como elemento contextual o como escenario en el que desarrollar una historia que versa sobre otro tema muy distinto. Es el caso de *Madre / androïde* (Tomlin, 2021), película ambientada en una sociedad postapocalíptica en la que los humanos y la IA están enfrentados, cuyo tema principal es otro muy distinto: la maternidad y las relaciones sentimentales. Algo parecido sucede con *Los Mitchell contra las máquinas* (Rianda y Rowe, 2021), donde la guerra entre humanos e IA establece el contexto para abordar el tema de la unión y la convivencia familiar.

la relevancia de la propuesta política y científica europea de carácter desiderativo en relación con un ecosistema confiable de IA fundamentado en una perspectiva humanocéntrica, la educación sentimental y la potenciación de imaginarios catastróficos en el marco de la ficción ocasiona que la relación con esta tecnología disruptiva cobre su sentido en prejuicios surgidos de la literatura y del cine. Aunque el tema de la rebelión de las máquinas es propio de la narrativa de ciencia ficción y de otros subgéneros, como el *cyberpunk* (Masís González y Castro López, 2021: 133), es preciso cuestionar su aceptabilidad desde un punto de vista ético. Asumir sin reservas la rebelión de las máquinas en este marco implicaría presuponer que cualquier realidad existente en el imaginario ficticio puede formar parte de la cotidianidad, cuando queda lejos de ser así. Y aunque la temática de la rebelión de las máquinas constituye un recurso atractivo para el público y estimula el trabajo científico (Alfonseca et al., 2021) y periodístico (Zentner, 2023), es necesario subrayar que han surgido voces discordantes que consideran que este conflicto debe permanecer en el ámbito de la ficción. Una de estas voces ha sido expresada por el científico Ramón López de Mántaras, quien sostiene lo siguiente:

[...] queda muy bien en la ciencia ficción, pero las máquinas no van a despertarse por sí mismas un día y decir, «venga, vamos a dominar el mundo», porque no tienen intencionalidad ni conciencia. Por mucho que aprendan a hacer cosas extraordinarias, por mucho que las programemos para ello, no podrán hacer eso. Es cierto que a veces incluso un programador se sorprende ante lo que pueden llegar a aprender, pero eso no significa que puedan aprender a ser malas o a tener intenciones de rebelarse contra las personas. Estas ideas hay que dejarlas para la ciencia ficción. (López de Mántaras, 2017)

5. Conclusión

El análisis de las producciones cinematográficas anteriormente señaladas nos permite insistir en la necesidad de incrementar la sabiduría práctica¹² durante la interpretación de las tramas inspiradas en el complejo de Frankenstein. La

12. La sabiduría práctica que proponemos hunde sus raíces en el pensamiento aristotélico y consiste, según Jesús Conill, en «aquel “tacto” por el que el juicio moral en situación, animado por la convicción, es capaz de responder a los conflictos que suscita una moral de la obligación, con el objetivo de lograr una vida buena con y para los otros en instituciones justas» (2006: 230). Recogiendo el testigo de Paul Ricoeur, Conill destaca dos elementos importantes de la sabiduría práctica: el peso de la experiencia y la conflictividad existente por las exigencias de aplicación de la norma moral universal. En este sentido, examinar de un modo crítico la aceptabilidad de las producciones cinematográficas actuales en un horizonte de IA confiable no constituye una insinuación de la censura, sino un llamado para el compromiso ético que toda actividad con un impacto social debe asumir. Hans-Georg Gadamer también destaca la impronta aristotélica en este concepto: «es precisamente también su ‘filosofía práctica’, la obra maestra de ese genio de lo razonable, la que abre nuevos caminos. Aristóteles separó conscientemente la filosofía práctica de la teórica, ofreciendo con ello perspectivas teóricas a las que hoy deberíamos prestar atención renovada, hoy, quiere decir, en la situación crítica de una cultura de la humanidad seducida hacia una aplicación desmedida de nuestro saber científico y técnico» (1998: 192).

representación de la IA que estimulan estos filmes puede dificultar el establecimiento de confianza y, por tanto, inducir una extraña sensación de inseguridad en la ciudadanía respecto a un sistema sociotécnico cada vez más relevante en sus vidas. La realización del proyecto humanocéntrico europeo expresado en la IA confiable requiere ejercitar el juicio crítico para desentrañar los posibles discursos políticos implícitos en las películas. En este sentido, es importante cuestionar la aceptabilidad del hilo argumentativo centrado en el conflicto entre el humano y la máquina en aras de un ecosistema europeo de IA fundamentado en la confianza. Asumir sin reservas que los humanos son dominados por las máquinas no constituye una respuesta adecuada y, por ende, una muestra de sabiduría. Al contrario, supone una estrategia desalentadora para discernir la dimensión política inherente a la tecnología, la puesta en valor de la experiencia humana y el desarrollo de la IA alineado con las necesidades de la humanidad, es decir, para mantener un nivel razonable de responsabilidad sobre la tecnología.

Aunque todo sistema sociotécnico conlleva efectos ambivalentes y controvertidos, especialmente la IA, es conveniente cuestionar la objetivación del complejo de Frankenstein, aun cuando las estrategias cinematográficas empleadas son muy atractivas. El consumo cada vez más extendido de filmes que representan historias conflictivas de la relación entre el humano y la máquina precisa una interpretación crítica para distinguir realidad y ficción. Ciertamente, no se trata de articular mecanismos únicamente desde la racionalidad técnica, sino de invertir esfuerzos en otros campos, como el del cine, para reflexionar sobre la complejidad del fenómeno de la IA.

El problema viene dado cuando desde el cine se muestra una imagen reduccionista en la que la oposición entre el humano y la máquina se agota en la problemática del complejo de Frankenstein, recreada de manera reiterada en la narrativa. A este respecto conviene mencionar una posible solución que, lejos de sugerir que la aceptabilidad del contenido dependa de si este está en sintonía o no con aquello que es deseable, proponga una educación profunda en la sensibilidad artística, pues, tal y como hemos mencionado ya, aquellos elementos que funcionan narrativamente en el arte no tienen por qué ser los que funcionen moral y políticamente en la sociedad. El rechazo de hacer concordar el contenido ficticio con lo políticamente deseable resulta necesario, debido a que una solución de este estilo requiere adecuar el juicio estético al juicio moral, conduciendo inevitablemente a la censura. El motivo de su rechazo es que no se puede pretender que los elementos de la ficción reflejen única y exclusivamente aquellas cosas que «deben ser», sino que es posible, y ahí reside la riqueza del arte, imaginar mundos alternativos y realidades ficticias que no se ajusten a la cotidianidad de los hechos ni a las reglas concretas de una sociedad¹³. Por eso, y en ello se basan algunos autores que se han citado

13. Estas afirmaciones coinciden con la tesis defendida recientemente por Ortiz Villeta (2024), en la que el saber diferenciar entre lo ficticio y lo real resulta fundamental. La autora afirma que «Si medimos constantemente a la ficción según el grado de aproximación a la realidad

en el texto, se debe entender que gran parte del temor hacia el progreso tecnológico, así como la concepción simplista de la IA, proviene de la cantidad de ocasiones en las que tanto el complejo de Frankenstein como otras representaciones negativas han dominado la narrativa de los productos culturales que mayoritariamente consumimos. Si se tiene en cuenta que este complejo es ficticio, es decir, que sirve de manera extraordinaria para ofrecer el conflicto narrativo necesario en una novela o en una película, se entenderá que hace una buena función en el marco de la ficción literaria y audiovisual. Lo que se tiene que evitar es pensar que ese motivo extraordinario de la narrativa es el paradigma del que se debe partir a la hora de afrontar los retos actuales que tenemos como sociedad respecto a la IA.

Referencias bibliográficas

- ALFONSECA, Manuel; CEBRIÁN, Manuel; FERNÁNDEZ ANTA, Antonio; COVIELLO, Lorenzo; ABELIUK, Andrés; RAHWAN, Iyad (2021). «Superintelligence cannot be Contained: Lessons from Computability Theory». *Journal of Artificial Intelligence Research*, 70, 65-76.
- ASIMOV, Isaac (2009). *Yo, robot*. Barcelona: Editora y Distribuidora Hispano Americana.
- BALLÓ, Jordi y PÉREZ, Xabier (2010). *La semilla inmortal*. Barcelona: Anagrama.
- BEAUCHAMP, Gorman (1980). «The Frankenstein Complex and Asimov's Robots». *Mosaic: A Journal for the Interdisciplinary Study of Literature*, 13 (3/4), 83-94. Recuperado de <<http://www.jstor.org/stable/24780264>>
- BIRD, James (Director) (2022). *Wifelike* [Película]. Estados Unidos: SP Media Group / Wonder Street.
- BOEHM, Gottfried (ed.) (1994). *Was ist ein Bild?* Múnich: Fink.
- BRYSON, Joanna J. y THEODOROU, Andreas (2019). «How society can maintain human-centric artificial intelligence». En: TOIVONEN-NORO, Marja Irene y SAARI, Eveliina (eds.). *Human-Centered Digitalization and Services*. Singapur: Springer.
<https://doi.org/10.1007/978-981-13-7725-9_16>
- CAMERON, James (Director) (1984). *Terminator* [Película]. Estados Unidos: Hemdale, Pacific Western, Euro Film Funding.
- COMISIÓN EUROPEA (2020). *Libro blanco sobre la inteligencia artificial: Un enfoque europeo orientado a la excelencia y la confianza*. Recuperado de <<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065>>
- (2021). *Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act, AIA) and Amending Certain Union Legislative Acts*. Recu-

estamos a) propiciando la confusión entre ficción y realidad y b) traicionando y matando la ficción. Porque su esencia es, justamente, aunque se utilicen elementos de ella, no ser realidad. Nos permite contar la complejidad del mundo y del ser humano y eso requiere subtexto, imaginación y libertad creativa» (p. 68).

- perado de <<https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AI-Act-FullText.pdf>>
- CONILL, Jesús (2006). *Ética hermenéutica*. Madrid: Tecnos.
- COOK, S. D. Noam (2010). «Making the Technological Trustworthy». *Knowledge, Technology & Policy*, 23, 455-459.
<<https://doi.org/10.1007/s12130-010-9126-4>>
- CRAVEN, Wes (Director) (1996). *Scream* [Película]. Dimension Films / Woods Entertainment.
- CUNNINGHAM, Sean S. (Director) (1980). *Viernes 13* [Película]. Paramount Pictures / Georgetown Productions Inc. / Sean S. Cunningham Films.
- DIERYCK BARRENA, Manuel (2023). «Inmersión en el problema del mito de Frankenstein». *Análisis. Revista de Investigación Filosófica*, 10 (1), 3-21.
<https://doi.org/10.26754/ojs_arif/arif.202316251>
- DIGNUM, Virginia (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham: Springer.
- EDWARDS, Gareth (Director) (2023). *The Creator* [Película]. Estados Unidos: 20th Century Studios.
- EUBANKS, Virginia (2007). «Popular technology: Exploring inequality in the information economy». *Science and Public Policy*, 34 (2), 127-138.
<<https://doi.org/10.3152/030234207X193592>>
- FLORIDI, Luciano (2021). «Establishing the Rules for Building Trustworthy AI». En: FLORIDI, Luciano (eds.). *Ethics, Governance, and Policies in Artificial Intelligence*. Cham: Springer.
<https://doi.org/10.1007/978-3-030-81907-1_4>
- FLORIDI, Luciano; COWLS, Josh; BELTRAMETTI, Monica; CHATILA, Raja; CHAZERAND, Patricie; DIGNUM, Virginia; LUETGE, Cristoph; MADELIN, Robert; PAGALLO, Ugo; ROSSI, Francesca; SCHAFFER, Burkhard; VALCKE, Peggy y VAYENA, Effy (2018). «AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations». *Minds and Machines*, 28 (4), 689-707.
<<https://doi.org/10.1007/s11023-018-9482-5>>
- FUTURE OF LIFE INSTITUTE (2017). *Asilomar AI Principles*. Recuperado de <<https://futureoflife.org/open-letter/ai-principles/>>
- GADAMER, Hans-Georg (1998). *El giro hermenéutico*. Madrid: Cátedra.
- GARCÍA GUAL, Carlos (2022). *Prometeo: El mito del dios rebelde y filántropo*. Madrid: Turner.
- GARLAND, Alex (Director) (2014). *Ex Machina* [Película]. Reino Unido: Universal Pictures International (UPI), Film4, DNA Films.
- GRUPO DE EXPERTOS DE ALTO NIVEL SOBRE INTELIGENCIA ARTIFICIAL (2019a). *Directrices éticas para una IA fiable*. Recuperado de <<https://data.europa.eu/doi/10.2759/14078>>
- (2019b). *Recomendaciones de política e inversión para una IA fiable*. Recuperado de <<https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>>
- HERNÁNDEZ DE LA FUENTE, David (2015). *Mitología clásica*. Madrid: Alianza.

- JARAN, François (2021). *El animal artificial*. Madrid: Guillermo Escolar Editor.
- JEUNET, Jean-Pierre (Director) (2022). *Bigbug* [Película]. Francia: Eskwad / Gaumont.
- JIMÉNEZ GONZÁLEZ, Marcos (2019). «El robot en el cine: De la obediencia a la rebelión. Análisis de los imaginarios estéticos predominantes en el cine de ciencia ficción». En: URRACO SOLANILLA, Mariano y MARTÍNEZ MESA, Francisco José (eds.). *De esclavos y robots y esclavas: Paisajes transmediáticos*. Madrid: Libros de la Catarata.
- JOHNSTONE, Gerard (Director) (2022). *Megan* [Película]. Estados Unidos / Nueva Zelanda: Atomic Monster / Blumhouse Productions.
- KLEVBORG, Lars (Director) (2019). *Muñeco diabólico* [Película]. Canadá / Estados Unidos: KatzSmith Productions.
- KUBRICK, Stanley (Director) (1968). *2001: Una odisea del espacio* [Película]. Reino Unido y Estados Unidos: Metro-Goldwyn-Mayer (MGM) / Stanley Kubrick Productions.
- LANG, Fritz (Director) (1927). *Metrópolis* [Película]. Alemania: Universum Film (UFA).
- LARSON, Erick J. (2022). *El mito de la inteligencia artificial: Por qué las máquinas no pueden pensar como nosotros lo hacemos*. Barcelona: Shackleton Books.
- LI, Bo; QI, Peng; LIU, Bo; DI, Shuai; LIU, Jingen; PEI, Jiquan; YI, Jinfeng y ZHOU, Bowen (2023). «Trustworthy AI: From Principles to Practices». *ACM Computing Surveys*, 55 (9), 1-46.
<<https://doi.org/10.1145/3555803>>
- LIANG, Weixin; ADEBE TADESSE, Girmaw; HO, Daniel; FEI-FEI, L.; ZAHARIA, Matei; ZHANG, Ce y ZOU, James (2022). «Advances, challenges and opportunities in creating data for trustworthy AI». *Nature Machine Intelligence*, 4, 669-677.
- LÓPEZ SÁENZ, M.^a Carmen (2016). «La aplicación gadameriana de la phrónesis a la praxis». *Contrastes: Revista Internacional de Filosofía*, 6, 79-98.
<<https://doi.org/10.24310/Contrastescontrastes.v0i0.1624>>
- «Lo último que me preocuparía es una supuesta rebelión de las máquinas». Entrevista a Ramón López de Mántaras. CSIC (20 de diciembre de 2017). Recuperado de <<https://www.csic.es/es/ciencia-y-sociedad/iniciativas-de-divulgacion/historico-de-iniciativas-de-divulgacion/lo-ultimo-que-me-preocuparia-es-una-supuesta-rebelion-de-las-maquinas>>
- MANOVICH, Lev y ARIELLI, Emanuele (2021). *Artificial Aesthetics: A Critical Guide to AI, Media and Design*. Manovich. Recuperado de <<http://manovich.net/index.php/projects/artificial-aesthetics>>
- MASÍS GONZÁLEZ, Tadeo y CASTRO LÓPEZ, Roberto (2021). «Distopías y la rebelión de las máquinas: Sobre los tópicos del Cyberpunk». *Pensamiento Actual*, 21 (36), 131-138.
<<https://doi.org/10.15517/pa.v21i36.47019>>
- MCCORDUCK, Pamela (1991). *Máquinas que piensan*. Barcelona: Tecnos.
- MILLER, Steven C. (Director) (2022). *Margaux* [Película]. Estados Unidos: Lighthouse Pictures.

- MIRTES, Aaron (Director) (2020). *Alpha Test* [Película]. Estados Unidos: Exit 10 Films.
- MITCHELL, William John Thomas (1994). *Picture Theory*. Chicago: University of Chicago Press.
- MOLINUEVO, José Luis (2004). *Humanismo y nuevas tecnologías*. Madrid: Alianza.
- MOROZOV, Evgeny (2016). *La locura del solucionismo tecnológico*. Buenos Aires: Katz Editores.
- NICKEL, Philip J.; FRANSSSEN, Maarten y KROES, Peter (2010). «Can We Make Sense of the Notion of Trustworthy Technology?». *Knowledge, Technology & Policy*, 23, 429-444.
- NOWOTNY, Helga (2022). *La fe en la inteligencia artificial: Los algoritmos predictivos y el futuro de la humanidad*. Barcelona: Galaxia Gutenberg.
- ORTEGA Y GASSET, José (2015). *Meditación de la técnica: Ensimismamiento y alteración*. Madrid: Biblioteca Nueva.
- ORTIZ VILLETÁ, Áurea (2024). *El arte de inventar la realidad*. Valencia: Barlin Libros.
- PASQUALE, Frank (2024). *Las nuevas leyes de la robótica: Defender la experiencia humana en la era de la IA*. Barcelona: Galaxia Gutenberg.
- PITT, Joseph C. (2010). «It's Not about Technology». *Knowledge, Technology & Policy*, 23, 445-454.
<<https://doi.org/10.1007/s12130-010-9125-5>>
- POST, Maaikje y MULDER, Arjen (2000). *Book for the Electronic Arts*. Ámsterdam: Art Data.
- PROYAS, Alex (Director) (2004). *Yo, Robot* [Película]. Estados Unidos, Alemania: Twentieth Century Fox, Mediamstream Vierte Film GmbH & Co. Vermarktungs KG, Davis Entertainment.
- RANCIERE, Jacques (2014). *El reparto de lo sensible. Estética y política*. Buenos Aires: Prometeo libros.
- RIANDA, Michael y ROWE, Jeff (Directores) (2021). *Los Mitchell contra las máquinas* [Película]. Estados Unidos / Hong Kong / Canadá: Sony Pictures Animation / Columbia Pictures.
- ROMELE, Alberto (2022). «Images of Artificial Intelligence: a Blind Spot in AI Ethics». *Philosophy & Technology*, 35, 4.
<<https://doi.org/10.1007/s13347-022-00498-3>>
- SAFFARI, Ehsan; HOSSEINI, Seyed Ramezan; TAHERI, Alireza y MEGHDARI, Ali (2021). «Does cinema form the future of robotics?»: A survey on fictional robots in sci-fi movies». *SN Applied Sciences*, 3 (655), 1-13.
<<https://doi.org/10.1007/s42452-021-04653-x>>
- SANG-HO, Yeon (Director) (2023). *Jung_E* [Película]. Corea del Sur: Climax Studio.
- SARTORI, Laura y THEODOROU, Andreas (2022). «A sociotechnical perspective for the future of AI: Narratives, inequalities, and a human control». *Ethics and Information Technology*, 24 (4), 1-11.
<<https://doi.org/10.1007/s10676-022-09624-3>>
- SCOTT, Ridley (Director) (1982). *Blade Runner* [película]. Estados Unidos: Warner Bros. Pictures.

- SHELLEY, Mary W. (2011). *Frankenstein, o el moderno Prometeo*. Madrid: Alianza Editorial.
- SKAUG SÆTRA, Henrik (2022). *AI for the sustainable development goals*. Boca Raton: CRC Press.
- SWEET, Jesse (dir.) (2023). *Lo desconocido: Los robots asesinos* [Documental]. Estados Unidos: Netflix, Story Syndicate.
- TERRONES RODRÍGUEZ, Antonio Luis y ROCHA BERNARDI, Mariana (2023). «El valor de la ética aplicada en los estudios de ingeniería en un horizonte de inteligencia artificial confiable». *Sophia: Colección de Filosofía de la Educación*, 36, 221-245.
<<https://doi.org/10.17163/soph.n36.2024.07>>
- TOMLIN, Mattson (Director) (2021). *Madre / Androide* [Película]. Estados Unidos: Miramax.
- UNIVERSITÉ DE MONTRÉAL (2017). *Montreal Declaration for a Responsible Development of AI*. Recuperado de <<https://montrealdeclaration-responsibleai.com/the-declaration/>>
- VALLEJO CAMPOS, Álvaro (2004). «El concepto aristotélico de *prhónesis* y la hermenéutica de Gadamer». En: ZÚÑIGA GARCÍA, José Francisco; SÁEZ RUEDA, Luis; PÉREZ TAPIAS, José Antonio; NICOLÁS MARÍN, Antonio y ACERO FERNÁNDEZ, Juan José (eds.). *El legado de Gadamer*. Granada: Editorial Universidad de Granada.
- VAN WYNSBERGHE, Aimee (2021). «Sustainable AI: AI for sustainability and the sustainability of AI». *AI Ethics*, 1, 213-218.
<<https://doi.org/10.1007/s43681-021-00043-6>>
- VINUESA, Ricardo; AZIZPOUR, Hossein; LEITE, Iolanda; BALAAM, Madeline; DIGNUM, Virginia; DOMISCH, Sami; FELLÄNDER, Anna; LANGHANS, Simone Daniela; TEGMARK, Max y FUSO NERINI, Francesco (2020). «The role of artificial intelligence in achieving the Sustainable Development Goals». *Nature Communications*, 11, 1-10.
- WACHOWSKI, Lana y WACHOWSKI, Lilly (Directoras) (1999). *Matrix* [Película]. Estados Unidos: Warner Bros. / Village Roadshow Pictures / Groucho Film Partnership.
- WHALE, James (Director) (1931). *El doctor Frankenstein* [Película]. Estados Unidos: Universal Pictures.
- ZENTNER, Gonzalo (2023). «¿Es posible una “rebelión de las máquinas”?»: Qué respondió la Inteligencia Artificial». *El Litoral* (2 de abril). Recuperado de <https://www.ellitoral.com/internet-y-tecnologia/video-viral-posible-rebelion-maquinas-inteligencia-artificial-chatgpt-respuestas-dudas-certezas-terminator-skynet-james-cameron-control-total_0_qqH7wUWwMP.html>

Marcos Jiménez González es graduado y doctor en Filosofía. Se especializó en la rama de Estética y Teoría de las Artes, centrándose en la Estética del cine, ámbito en el que lleva a cabo su labor docente e investigadora. En 2018 publicó su primer libro, *Romanticismo, técnica y poder en la Arquitectura de Albert Speer* (editorial Ápeiron), y en 2022 el segundo, *Fritz Lang y el expresionismo* (Shangrila). Ha sido profesor asociado en la Universidad Rey Juan Carlos (URJC) y colaborador en la Universidad a Distancia de Madrid (UDIMA), así como investigador posdoctoral Margarita Salas en la Universidad de Salamanca (USAL), contrato con el que realizó estancias de investigación en el Instituto de Filosofía (IFS) del CSIC y en la facultad de Filosofía de la Universidad Complutense de Madrid (UCM). Actualmente es profesor en la Universidad de Zaragoza (UNIZAR), en la Unidad Predepartamental de Filosofía (área de Estética y Teoría de las Artes).

Marcos Jiménez González has a degree and a PhD in Philosophy, and specialised in the Aesthetics and Theory of the Arts, focusing on the aesthetics of cinema, a field in which he carries out his teaching and research work. In 2018 he published his first book, *Romanticismo, técnica y poder en la Arquitectura de Albert Speer* (editorial Ápeiron), followed by *Fritz Lang y el expresionismo* (Shangrila) in 2022. He has been associate professor at the Universidad Rey Juan Carlos (URJC) and collaborator at the Universidad a Distancia de Madrid (UDIMA), as well as Margarita Salas postdoctoral researcher at the Universidad de Salamanca (USAL), a contract under which he carried out research visits at the Instituto de Filosofía (IFS) of the CSIC and at the Faculty of Philosophy of the Complutense University of Madrid (UCM). He is currently a lecturer at the University of Zaragoza (UNIZAR), in the Predepartmental Unit of Philosophy (Aesthetics and Theory of the Arts).

Antonio Luis Terrones Rodríguez es profesor e investigador en la Universidad Loyola Andalucía. Licenciado en Filosofía por la Universidad de Murcia, máster en Ética y Democracia por la Universitat de València y doctor en Filosofía por la Universitat de València. Sus principales líneas de investigación se centran en la ética aplicada a la inteligencia artificial y en la filosofía de la tecnología. Ha impartido docencia en la Pontificia Universidad del Ecuador, en la Universidad de La Habana y en la Universidad de Costa Rica, entre otras instituciones académicas. Entre sus publicaciones destacan «Ética para la inteligencia artificial sostenible» (*Arbor*, 198 (806): a683) y «El valor de la ética aplicada en los estudios de ingeniería en un horizonte de inteligencia artificial confiable» (*Sophia, Colección de Filosofía de la Educación*, 36, 221-245).

Antonio Luis Terrones Rodríguez is professor and researcher at the Loyola Andalusia University. He has a degree in Philosophy from the University of Murcia, a Master's in Ethics and Democracy from the University of Valencia, and a PhD from the University of Valencia. His main lines of research focus on ethics applied to artificial intelligence and the philosophy of technology. He has taught at the Pontifical University of Ecuador, at the University of Havana and at the University of Costa Rica, among other academic institutions. His publications include "Ethics for Sustainable Artificial Intelligence" (*Arbor*, 198 (806): a683) and "The Value of Applied Ethics in Engineering Studies in a Horizon of Reliable Artificial Intelligence" (*Sophia, Colección de Filosofía de la Educación*, 36, 221-245).
