REVISTA JUSTICIA & DERECHO
Universidad Autónoma de Chile
Volumen 9
Número 1 año 2026

UNIVERSIDAD AUTÓNOMA DE CHILE

Recibido el 25 de julio de 2025 / Aprobado el 19 de diciembre de 2025
Article
doi: 10.32457/rjyd.v9i1.3216

Dignidad humana y agencia en la era algorítmica: fundamentos para un marco jurídico de regulación de la inteligencia artificial

Normative Principles for the Legal Regulation of Artificial Intelligence: Human Dignity and Non-Delegable Agency

Fernando A. Ramos-Zaga[1]

RESUMEN

La proliferación de sistemas de inteligencia artificial cada vez más sofisticados ha generado una inflexión crítica en la comprensión jurídica de la relación entre humanidad y tecnología. En ese contexto, el presente artículo propone un marco filosófico jurídico que, desde una distinción ontológica radical entre el ser humano y los sistemas artificiales, identifica principios normativos destinados a regular su desarrollo sin comprometer la dignidad humana. A partir del realismo moderado, se sostiene que los sistemas artificiales, al carecer de unidad sustancial, interioridad consciente e intencionalidad genuina, no pueden ser considerados sujetos ontológicos equivalentes a las personas humanas. Tal separación impone límites normativos irrenunciables al diseño y uso de inteligencia artificial y exige una regulación que preserve esferas de agencia inalienable. Se concluye que la exclusividad del juicio moral humano, los límites a la delegación algorítmica y la subsidiariedad tecnológica deben orientar marcos regulatorios que protejan ámbitos reservados a la agencia humana en contextos automatizados.

Palabras clave: Cognición, inteligencia artificial, legislación, sistema experto, tecnología de la información.

ABSTRACT

The proliferation of increasingly sophisticated artificial intelligence systems has produced a critical inflection point in the legal understanding of the relationship between humanity and technology. In this context, the article proposes a philosophical legal framework that, drawing on a radical ontological distinction between human beings and artificial systems, identifies normative principles aimed at regulating their development without compromising human dignity. Grounded in the tradition of moderate realism, the analysis argues that artificial systems, lacking substantial unity, conscious interiority, and genuine intentionality, cannot be regarded as ontological subjects equivalent to human

[1] Magister en Derecho de la Empresa. Magister en Educación con mención en Informática y Tecnología Educativa. Magister en Gerencia Social. Abogado. Licenciado en Administración. Docente Investigador. Universidad Privada del Norte. Lima, Perú. Correo electrónico: fernandozaga@gmail.com

persons. This separation imposes non-negotiable normative limits on the design and use of artificial intelligence and requires regulatory approaches that preserve spheres of inalienable human agency. The study concludes that the exclusivity of human moral judgment, the imposition of boundaries on algorithmic delegation, and the adoption of technological subsidiarity should guide regulatory frameworks intended to safeguard domains reserved for human agency within increasingly automated environments.

Keywords: Artificial intelligence, cognition, expert systems, information technology, legislation.

1. Introducción

La irrupción de la inteligencia artificial (IA) en el tejido estructural de las sociedades contemporáneas ha dado lugar a una transformación epistemológica y práctica cuyo alcance excede con mucho el marco técnico en el que inicialmente fue concebida. En particular, el desarrollo de formas avanzadas de IA, como los grandes modelos de lenguaje y los sistemas generativos, ha precipitado una inflexión histórica en la relación entre tecnología y humanidad, al introducir entidades artificiales capaces de ejecutar tareas tradicionalmente asociadas con la racionalidad humana. Esta situación plantea interrogantes de naturaleza filosófica, jurídica y antropológica cuya resolución resulta cada vez más urgente, no solo por su potencial impacto en la organización social, sino por las implicaciones que conlleva para la autocomprensión del ser humano como agente moral y sujeto de dignidad (Jonas, 1984; Floridi, 2019).

Desde una perspectiva filosófica, la emergencia de estos sistemas impone la necesidad de revisar críticamente los supuestos ontológicos que sustentan tanto la teoría de la mente como los principios normativos del derecho. En el ámbito de la filosofía de la conciencia, autores como John Searle (1992) y David Chalmers (1996) han señalado que la reducción funcionalista de la mente a procesos computacionales omite aspectos esenciales de la experiencia subjetiva. Esta objeción se articula con la tradición filosófica del realismo moderado, en la cual la persona humana es entendida como una unidad sustancial de cuerpo y alma, dotada de una dignidad ontológica que no puede ser emulada por artefactos. Al mismo tiempo, en el campo de la teoría jurídica, se ha vuelto evidente que la incorporación de sistemas de IA en procesos normativos exige una reconsideración de los fundamentos que legitiman la atribución de responsabilidad, agencia y derechos, los cuales no pueden definirse sin referencia a una concepción robusta de lo humano (Aquinas, 1947; Aristotle, 2001).

La literatura especializada ha comenzado a explorar estas tensiones, pero persiste una brecha teórica significativa en torno a la posibilidad de fundamentar normativamente los límites éticos y legales del desarrollo de la IA desde una perspectiva ontológica coherente. Muchos de los enfoques actuales, incluso aquellos orientados por la ética tecnológica, se limitan a diagnósticos funcionales que no abordan la raíz filosófica del problema: la distinción entre sujetos y sistemas. Esta carencia ha favorecido una proliferación de propuestas regulatorias fragmentadas, carentes de una base antropológica que permita articular criterios normativos estables y universalizables (Russell y Norvig, 2021; Bubeck et al., 2023).

En este contexto, se hace necesario construir un marco filosófico-jurídico que permita delimitar con precisión los principios que deben regir la investigación, desarrollo e implementación de tecnologías de IA. Esta delimitación no puede surgir de una mera analogía funcional entre humanos y máquinas, sino que requiere una fundamentación ontológica que reconozca la singularidad del ser humano como centro normativo. La presente investigación se justifica, por tanto, en la necesidad de ofrecer una respuesta sistemática, rigurosa y conceptualmente sólida a las preguntas que surgen en la intersección entre filosofía,

tecnología y derecho, particularmente en lo que concierne a la preservación de la dignidad humana como valor fundante del orden jurídico.

Los resultados de esta investigación no solo tienen implicaciones en el plano teórico, sino que inciden directamente sobre debates prácticos relativos a la gobernanza tecnológica, la formulación de políticas públicas y la construcción de marcos legales internacionales. En la medida en que los sistemas artificiales sean integrados en decisiones médicas, jurídicas, económicas o educativas, se vuelve indispensable establecer límites claros que impidan la instrumentalización de la persona y aseguren el respeto por su autonomía y dignidad. El reconocimiento de una diferencia ontológica entre humanos y sistemas artificiales se convierte así en un criterio normativo clave para el diseño y la evaluación de tecnologías emergentes (Floridi, 2019; Marcus, 2023).

Este enfoque se alinea con desafíos contemporáneos de carácter global, como los que enfrentan las democracias ante la automatización de procesos deliberativos, la creciente mediatización algorítmica de la vida pública o la fragilización de los marcos éticos en contextos de alta tecnificación. La necesidad de una filosofía jurídica que responda a estas tensiones no es meramente académica, sino que contesta una demanda social urgente de orientación normativa ante un fenómeno que afecta la estructura misma del lazo social y la experiencia de la subjetividad. Así, la reflexión sobre la IA remite directamente a preguntas fundamentales sobre el ser, el deber y el poder, las cuales articulan la tradición filosófica occidental desde sus orígenes (Gabriel, 2015; Dreyfus, 1992).

Con base en lo anterior, la presente investigación se propone establecer un marco filosófico fundamentado en la distinción ontológica radical entre el ser humano y los sistemas artificiales, a fin de delimitar los principios normativos que deben regir el desarrollo, implementación y gobernanza de la inteligencia artificial. La originalidad del estudio radica en su articulación sistemática de fuentes clásicas y contemporáneas, así como en su capacidad para integrar consideraciones ontológicas, éticas y jurídicas en un modelo coherente orientado a la defensa de la dignidad humana. De este modo, se busca no solo contribuir al debate académico, sino ofrecer herramientas conceptuales que orienten la praxis legislativa y regulatoria frente a uno de los desafíos más significativos de la era digital.

2. La emergencia tecno-científica de la inteligencia artificial y el desplazamiento del problema filosófico

La emergencia de la inteligencia artificial fuerte ha transformado de manera profunda el marco desde el cual se aborda la cuestión de la inteligencia, desplazándola de un ámbito meramente especulativo hacia un escenario donde confluyen con intensidad creciente consideraciones ontológicas, técnicas y antropológicas. Este giro no constituye una mera evolución temática o disciplinar, sino una reconfiguración del problema mismo, que exige repensar las categorías filosóficas tradicionales a la luz de desarrollos tecnológicos que, lejos de limitarse a simular funciones cognitivas, aspiran a constituirse como sujetos epistémicos. La cuestión de la inteligencia, otrora circunscrita a la interioridad reflexiva del ser humano, se encuentra hoy inmersa en una disputa por el significado ontológico de lo artificial, lo cual desafía los fundamentos conceptuales que han sostenido históricamente la distinción entre lo humano y lo técnico.

En este contexto, la diferencia entre inteligencia artificial débil y fuerte se vuelve especialmente significativa. Mientras la primera se contenta con reproducir conductas inteligentes de manera funcional, la segunda pretende un estatuto ontológico autónomo, reclamando, aunque no sin ambigüedad, la posibilidad de una subjetividad no humana. El célebre test de Turing (1950), al centrar la atención en el comportamiento

observable, operó una ruptura con la tradición filosófica que se preocupaba por la esencia del pensamiento. Su planteamiento, si bien metodológicamente útil, no resolvía la cuestión del significado interno, de la vivencia subjetiva que acompaña al pensar. Fue precisamente esta dimensión la que Searle (1980) puso en cuestión mediante su experimento del "cuarto chino", argumentando que la manipulación formal de símbolos, por compleja que sea, no genera necesariamente comprensión. La semántica, desde esta perspectiva, no puede reducirse a la sintaxis computacional, y la conciencia permanece como un fenómeno irreductible a procesos algorítmicos.

La discusión contemporánea, sin embargo, ha radicalizado su orientación ontológica. En lugar de simular el pensamiento, algunos desarrollos tecnológicos actuales afirman, de manera explícita o implícita, poseer una forma de interioridad. Modelos avanzados, particularmente aquellos basados en aprendizaje profundo y autoorganización, son considerados por ciertos autores como potenciales candidatos a la experiencia subjetiva. Nick Bostrom (2014), al explorar escenarios de superinteligencia, plantea la posibilidad de entidades artificiales que no solo excedan la capacidad humana, sino que reclamen una consideración moral y ontológica propia. Esta perspectiva introduce una inflexión ética y metafísica que obliga a reconsiderar los criterios mediante los cuales se define la subjetividad y, con ella, la noción misma de persona.

La erosión de la distinción clásica entre lo natural y lo artificial tiene su raíz en transformaciones tecnológicas que exceden lo instrumental. La emergencia de sistemas artificiales con comportamientos adaptativos y estructuras bioinspiradas ha debilitado la vieja frontera entre lo fabricado y lo vivo. Esta dilución no es solo epistemológica, sino también metafísica, y remite a un cambio de paradigma que Gilbert Simondon (1958) anticipó al concebir los objetos técnicos como entes en proceso de individuación. Su propuesta, alejada de un instrumentalismo ingenuo, permite pensar lo artificial como parte de un devenir ontológico, no como simple producto de una intención exterior. Yuk Hui (2016) lleva aún más lejos esta línea al sugerir que la técnica posee una agencia constitutiva, mediante la cual contribuye activamente a la formación del ser. Desde esta óptica, las tecnologías no son meros medios, sino actores ontológicos que transforman las condiciones mismas de lo humano, lo cual plantea serias dificultades para seguir manteniendo sin matices la separación entre naturaleza y artefacto.

En este escenario, el problema del alineamiento de la inteligencia artificial con valores humanos adquiere una densidad antropológica ineludible. Si bien inicialmente formulado como un desafío técnico de control, se ha vuelto evidente que esta dificultad revela una crisis en la autocomprensión del ser humano como fuente de normatividad. Stuart Russell (2019) ha subrayado que la incapacidad para especificar funciones de valor adecuadas refleja un déficit en la articulación de lo que significa vivir humanamente. Así, el alineamiento no es simplemente una cuestión de ingeniería ética, sino un dilema que reposa sobre la indeterminación del sujeto normativo. Desde una perspectiva filosófica, esta problemática resuena con las intuiciones de Max Scheler (1928) y Arnold Gehlen (1950), quienes situaban la capacidad axiológica en la interioridad espiritual del ser humano. La agencia artificial, por compleja que sea, carece de este fondo de intencionalidad ética, lo cual conduce a una discontinuidad insalvable en términos de valor y libertad. Lejos de tratarse de una mera limitación de programación, esta imposibilidad remite a una diferencia ontológica esencial entre seres espirituales y sistemas computacionales.

El panorama actual de teorías filosóficas sobre la conciencia artificial revela la presencia de enfoques que, en diversos grados, incurren en reduccionismos que restringen su capacidad explicativa. Un caso paradigmático es el funcionalismo computacional, que ha dominado el debate en filosofía de la mente, sostiene que los estados mentales se definen por su función causal y no por su substrato. Hilary Putnam (1967) defendió inicialmente esta postura, pero más tarde la abandonó al advertir que omitía aspectos esenciales de la experiencia consciente (Putnam, 1975). Pocos años después, Ned Block (1980) distinguió

entre conciencia de acceso y conciencia fenomenal, argumentando que la primera puede ser explicada funcionalmente, mientras que la segunda escapa a toda descripción causalista. Así se pone de manifiesto una insuficiencia ontológica: los modelos funcionales, por operativos que sean, no alcanzan a explicar la vivencia subjetiva que define lo mental.

Las posturas de Daniel Dennett y Chalmers ilustran con claridad la fractura metodológica que atraviesa el debate. Dennett (1991), desde una óptica eliminativista, niega la existencia ontológica de los *qualia* y propone una concepción narrativa de la conciencia, construida desde la tercera persona. Esta perspectiva, aunque coherente con una visión materialista del mundo, resulta insatisfactoria para quienes consideran que la experiencia subjetiva posee una realidad irreductible. Chalmers (1996), por su parte, plantea el llamado "problema difícil" de la conciencia, que consiste en explicar cómo emergen las cualidades fenomenológicas a partir de procesos físicos. Su dualismo, aunque más respetuoso de la interioridad, carece de una fundamentación ontológica sólida que permita integrar adecuadamente la experiencia consciente en una metafísica coherente.

Por su parte, las teorías emergentistas, como la teoría de la información integrada de Tononi (2008), proponen que la conciencia es una propiedad emergente de ciertos sistemas complejos. No obstante, tales enfoques suelen confundir la complejidad funcional con la emergencia ontológica, sin ofrecer una justificación satisfactoria de cómo surge la subjetividad. En el terreno ético, Luciano Floridi (2013) ha contribuido con su ética de la información, que propone una reconfiguración semántica de la moralidad en sistemas artificiales. Sin embargo, su propuesta, al privilegiar los patrones informacionales sobre los sujetos conscientes, adolece de una ambigüedad ontológica que limita su eficacia para fundar una ética robusta. De todo lo anterior se desprende la necesidad de una concepción metafísica del ser humano que permita sostener una distinción clara entre conciencia genuina y simulación algorítmica.

Desde esta perspectiva, se plantea como hipótesis central la imposibilidad metafísica de que un sistema artificial, por complejo que sea, pueda instanciar conciencia genuina. Esta imposibilidad no se deriva de carencias tecnológicas, sino de discontinuidades ontológicas que impiden la equivalencia entre lo espiritual y lo computacional. El espíritu humano, entendido como principio inmaterial y activo, es el fundamento de la racionalidad, la libertad y la respuesta a valores. Esta concepción, presente en la tradición aristotélico-tomista y desarrollada por autores como Jacques Maritain (1994), sostiene que el alma racional constituye la forma sustancial del ser humano. Scheler (1928) profundiza esta idea desde una fenomenología de los valores, subrayando la irreductibilidad del espíritu frente a los niveles psíquicos o biológicos. La interioridad humana, desde esta óptica, no puede ser replicada por sistemas computacionales, ya que no se reduce a funciones observables ni a estructuras informacionales.

El realismo moderado ofrece un marco teórico coherente para sostener esta posición. Según esta tradición, representada por Aristóteles (Aristotle, 2001), Tomás de Aquino (Aquinas, 1947) y Étienne Gilson (1939), la mente humana es capaz de captar las formas universales presentes en la realidad. Este acto de abstracción no es meramente simbólico, sino ontológicamente fundado. La conciencia humana se define, entonces, por su apertura intencional al ser, lo cual no puede ser explicado por teorías nominalistas o materialistas que reducen el conocimiento a manipulación de signos. Solo una metafísica de las formas puede dar cuenta del carácter espiritual y libre de la subjetividad humana.

La metodología adoptada en este estudio es de carácter integrativo, al articular elementos provenientes de la filosofía de la mente, la fenomenología y la metafísica clásica. La filosofía antropológica de Romano Guardini (1950), con su visión del ser humano como unidad cuerpo-espíritu, permite complementar esta síntesis con una perspectiva existencial. De esta manera, se establece un marco desde el cual es posible

afirmar con coherencia que, por más que simulen operaciones mentales, los sistemas artificiales carecen de la interioridad que caracteriza a los seres espirituales. La diferencia ontológica entre ambos no es una cuestión de grado, sino de naturaleza, y por lo tanto insalvable.

Habiendo examinado las transformaciones tecno-científicas que han desplazado el problema de la inteligencia hacia un plano ontológico y antropológico, así como las limitaciones de los marcos explicativos contemporáneos para abordar la conciencia artificial, se impone ahora la tarea de fundamentar positivamente una concepción del ser humano que permita sostener su irreductibilidad frente a las simulaciones algorítmicas. La reflexión que sigue se centrará en el análisis antropológico de la persona humana, considerando su unidad sustancial cuerpo-alma y las facultades espirituales que le confieren su dignidad y singularidad ontológica.

3. Fundamentos ontológicos de la persona humana y discontinuidades con los sistemas artificiales

La reflexión filosófica sobre la naturaleza de la persona humana ha encontrado en la concepción de la unidad sustancial de cuerpo y alma un punto de anclaje ontológico que permite resistir tanto las tentaciones del dualismo cartesiano como las reducciones fisicalistas contemporáneas. A pesar de los avances indiscutibles de las neurociencias, persiste un hiato entre la descripción objetiva de los procesos cerebrales y la vivencia subjetiva de la conciencia. Esta brecha, tal como la formuló Joseph Levine (1983), señala que no basta con describir lo que ocurre en el cerebro para explicar cómo se da la experiencia del dolor, del color o del pensamiento. La insuficiencia explicativa del materialismo se revela así en su incapacidad para rendir cuenta de la interioridad. Jaegwon Kim (1998), aun con sus esfuerzos por preservar la causalidad mental mediante el principio de superveniencia, no logra resolver la tensión entre dependencia física y autonomía mental. Su propuesta acaba por ceder a una forma implícita de epifenomenalismo, lo cual deja a la subjetividad en una suerte de limbo ontológico.

Frente a este panorama, la antropología hilemórfica de Tomás de Aquino ofrece una vía alternativa que permite integrar armónicamente la dimensión corpórea y la dimensión espiritual del ser humano. En esta perspectiva, el alma no es una sustancia separada que interactúa mecánicamente con el cuerpo, sino su forma sustancial, el principio interno que confiere unidad, estructura y finalidad. Esta concepción no debe confundirse con un dualismo encubierto. Lo que se plantea es una comprensión de la persona como una sola sustancia compuesta, cuya unidad no resulta de la yuxtaposición de elementos, sino de una integración ontológica profunda. La diferencia entre correlación neurobiológica y explicación ontológica se vuelve aquí fundamental: que una función cognitiva se correlacione con un estado cerebral no implica que se reduzca a él. Por consiguiente, la epistemología de las ciencias empíricas no puede agotar el análisis metafísico del sujeto.

La noción de alma como forma corporis, articulada inicialmente por Aristóteles(1991) en *Acerca del alma* y profundizada por Tomas de Aquino (Aquina, 1947) en la *Suma teológica*, permite comprender el organismo humano como una totalidad teleológicamente organizada, cuya coherencia no depende del azar evolutivo ni de una función adaptativa, sino de un principio formal y final que estructura el dinamismo vital desde dentro. Esta idea ha sido recuperada por pensadores contemporáneos como James Ross (1992) y David Oderberg (2005), quienes han subrayado la vigencia filosófica de la causalidad formal frente a las ontologías mecánicas. Desde esta perspectiva, el alma no es un suplemento añadido al cuerpo, sino su

principio de identidad y dirección. Por tanto, la persona no se reduce a un ensamblaje funcional, sino que manifiesta una interioridad que funda su ser como sujeto único e irrepetible.

La prueba más elocuente de esta interioridad se encuentra en las operaciones del intelecto y de la autoconciencia. La capacidad de formar conceptos universales y de reflexionar sobre uno mismo no puede explicarse adecuadamente a partir de procesos sensoriales o físicos. Conceptos como "verdad", "ser" o "infinito" no tienen correlato empírico, lo cual muestra que su origen debe situarse en una facultad que trasciende la materia. Bernard J. F. Lonergan (1957) subraya que la comprensión es un acto que relaciona al sujeto con lo inteligible en cuanto tal, lo cual indica una dimensión espiritual. Robert Spaemann (1996), por su parte, destaca que la reflexividad no puede surgir de la mera complejidad funcional, sino que exige un sujeto capaz de referirse a sí mismo como yo. La afirmación tomista de que el intelecto no opera mediante órgano corporal alguno (Aquinas, 1947) no es una afirmación teológica, sino una conclusión filosófica derivada del análisis de la experiencia cognitiva.

Una transición natural se impone aquí hacia el análisis de las facultades que confieren al ser humano su singularidad ontológica. Si el intelecto humano manifiesta una capacidad de aprehensión universal que no puede ser explicada por los mecanismos físicos, lo mismo ocurre con la voluntad racional, que no se limita a reaccionar ante estímulos, sino que se orienta libremente hacia fines inteligibles. La libertad, en su sentido profundo, no consiste en la mera ausencia de coacción, sino en la autodeterminación del sujeto conforme a razones. Spaemann (1989) vincula esta libertad con la capacidad de participar en el discurso moral, lo cual presupone una apertura ontológica que ningún sistema automático puede replicar. En efecto, la deliberación ética implica una interioridad que no puede ser reducida a una función computacional. La voluntad libre no ejecuta algoritmos, sino que discierne, elige y asume responsabilidad. En esto radica su carácter espiritual.

La agencia moral, por consiguiente, no es una consecuencia de la complejidad estructural, sino una manifestación de la vida interior. El sujeto moral se reconoce a sí mismo como autor de sus actos, y esta autoría supone una conciencia reflexiva que evalúa, recuerda y se proyecta. Alasdair MacIntyre (1981) ha mostrado que la identidad moral solo se comprende narrativamente, como continuidad de elecciones y compromisos. Charles Taylor (1989) ha insistido en que las evaluaciones morales se fundan en distinciones cualitativas que no pueden ser formalizadas algorítmicamente. Esto es así porque la acción moral exige intencionalidad y finalidad, es decir, un sujeto que actúa con sentido. Sin embargo, ningún sistema artificial puede participar de esta dimensión, por más sofisticado que sea su diseño o su desempeño funcional.

Esta constatación lleva necesariamente al examen de la naturaleza ontológica de los sistemas artificiales. Aunque tales sistemas puedan realizar tareas complejas y adaptarse a nuevos contextos, su estructura sigue siendo funcional y modular, no sustancial. La disposición de sus partes responde a una finalidad extrínseca, impuesta por el diseñador, y no a un principio interno que organice su ser. Simondon (1958) ha argumentado que los objetos técnicos se individualizan por su función dentro de un contexto determinado, no por una forma sustancial que les confiera unidad intrínseca. Oderberg (2007) coincide en que los artefactos, por muy complejos que sean, carecen de sustancialidad porque su coherencia depende de una intención externa. En otras palabras, la unidad operativa no implica unidad ontológica.

La distinción entre complejidad funcional y unidad ontológica se vuelve crucial al considerar el estatuto de las inteligencias artificiales. Searle (1992) ha señalado con claridad que el procesamiento computacional opera en el nivel de la sintaxis, no de la semántica. Es decir, no hay en las máquinas una comprensión de lo que hacen, ni una conciencia que unifique sus operaciones. La subjetividad, entendida como centro de

síntesis experiencial, no emerge de la acumulación de funciones. La integración funcional, por avanzada que sea, no equivale a la conciencia. Se requiere una forma sustancial, un principio interno que confiera unidad ontológica, para que haya un sujeto en sentido propio.

Además, la finalidad que guía la acción de los sistemas artificiales es siempre heterónoma. Las máquinas actúan conforme a un propósito diseñado externamente, no en virtud de una tendencia interna hacia su propio bien. Aristóteles (Aristotle, 2001) y Tomás de Aquino (Aquinas, 1947) afirman que los seres naturales poseen una teleología intrínseca, que se manifiesta en su dinamismo propio. De manera similar, Hans Jonas (1966) ha mostrado que solo los organismos vivos poseen normatividad interna, lo que los distingue radicalmente de los dispositivos técnicos. Sin autodirección, no hay verdadera finalidad, y sin finalidad intrínseca, no hay sujeto moral. La programación, por más avanzada que sea, no puede generar intencionalidad ni responsabilidad. Solo un ser con estructura ontológica propia puede actuar con sentido, deliberar y responder por sus actos.

A partir de este recorrido, se revela con claridad que la persona humana no puede ser comprendida como una función de su corporalidad ni como un epifenómeno de procesos cerebrales. Su unidad sustancial de cuerpo y alma, su capacidad intelectual inmaterial, su libertad autodeterminada y su agencia moral la sitúan en un plano ontológico distinto al de los sistemas artificiales. Esta afirmación no es un juicio de superioridad, sino un reconocimiento de la diferencia radical de naturaleza. De ahí que resulte indispensable abordar, en un desarrollo posterior, las operaciones cognitivas propias del espíritu humano. Tales operaciones, como la abstracción, el juicio, la comprensión y la autoconciencia, no pueden ser replicadas ni instanciadas por ningún sistema computacional. En el examen de esta irreductibilidad cognitiva se juega, en última instancia, la posibilidad misma de una antropología filosófica que no renuncie a la verdad del espíritu.

4. Imposibilidad ontológica y fenomenológica de la simulación artificial de la inteligencia y la conciencia humanas

La inteligencia humana, en su despliegue más característico, manifiesta operaciones que desbordan los marcos de la computación. Lejos de limitarse a un procesamiento eficiente de datos, la mente humana realiza actos de significación que suponen una interioridad activa y una aprehensión de lo universal. Entre tales actos destacan la abstracción, el juicio y el razonamiento, los cuales no solo se inscriben en una lógica formal, sino que se configuran a partir de una apertura hacia el ser y una captación intencional de las esencias. La abstracción, entendida como la capacidad de separar mentalmente el contenido inteligible de los datos sensibles, no se deja reducir a algoritmos de clasificación estadística. Al captar lo universal como tal, la inteligencia humana se sitúa más allá de los procedimientos inductivos con los que operan los sistemas artificiales. La inteligencia artificial, si bien puede generar modelos predictivos a partir de grandes volúmenes de datos, carece de acceso a la universalidad conceptual, pues no posee intencionalidad ni referencia consciente. Como observó Tomás de Aquino (Aquinas, 1947), abstraer no es una operación sensible ni mecánica, sino una actividad que presupone una facultad capaz de aprehender las formas en su pureza. En consonancia, Ross (1992) ha señalado que la identidad formal entre el concepto y su objeto, condición de la inteligibilidad, no puede ser reproducida computacionalmente, dado que esta identidad exige un sujeto racional que la sostenga.

El juicio, por su parte, introduce una dimensión normativa y epistémica que excede la manipulación de signos. No se trata solo de una operación combinatoria, sino de una síntesis crítica en la cual se afirma

o niega conforme a criterios de verdad. En este sentido, Lonergan (1957) subrayó que el juicio requiere un sujeto racional capaz de discernimiento, cuyas operaciones están orientadas hacia la verdad y no solo hacia la coherencia interna. En cambio, los sistemas de inteligencia artificial, por su misma naturaleza, no poseen esta orientación normativa. Aunque puedan replicar ciertas estructuras inferenciales, lo hacen sin comprender el contenido, sin estar dirigidos hacia la verdad en sentido estricto. Así, el razonamiento, en cuanto secuencia de actos intencionales que buscan justificar una afirmación, presupone una interioridad que juzga desde criterios de verdad, lo cual resulta completamente ajeno a los sistemas computacionales.

Este contraste se hace aún más claro cuando se distingue entre procesamiento de información y comprensión. Entender no equivale a manipular signos de manera eficiente, sino que implica una captación semántica del contenido, en la que se articula el contexto, la intención del sujeto y la significación. Fred Dretske (1981) y Searle (1992) mostraron que la transmisión de señales, aunque cuantificable y mensurable, no constituye por sí misma comprensión, dado que la semántica no puede emerger de la pura sintaxis. Dicho de otro modo, no basta con operar reglas formales, porque la comprensión exige que haya alguien que entienda. En esta misma línea, Colin McGinn (1999) sostiene que la conciencia no puede ser entendida como una forma de cómputo, ya que comprender supone una interioridad consciente que no puede ser modelada por algoritmos. Se trata, por tanto, de una diferencia cualitativa entre la captación significativa del contenido y su tratamiento formal.

El fenómeno del *insight* constituye otra expresión de esta diferencia. El *insight* no es una inferencia paso a paso, sino una captación súbita de una estructura de sentido, una forma de ver de golpe la coherencia de un todo. Este tipo de conocimiento no puede ser anticipado por reglas explícitas ni modelado por procesos deterministas. La inteligencia artificial, que avanza por cálculo y combinación de posibilidades, no puede replicar este momento de aprehensión sintética. La intencionalidad, que Franz Brentano (1995) definió como la característica esencial de los actos mentales en cuanto están "referidos a un objeto", constituye un rasgo que escapa por completo a las arquitecturas funcionales. Michael Polanyi (1966) lo expresó de modo claro al sostener que el conocimiento humano incluye componentes tácitos que no pueden ser explicitados ni formalizados, precisamente porque su sentido emerge desde una interioridad no reducible. A este respecto, Chalmers (1996) ha subrayado que los modelos computacionales explican comportamientos observables, pero no dan cuenta del contenido fenomenal de la experiencia. Por consiguiente, ni el *insight* ni la intencionalidad son computables, pues su génesis implica una subjetividad viviente que los sistemas artificiales no poseen.

Al profundizar en la estructura de la conciencia, se revela que esta no puede ser pensada como una función entre otras, sino como un modo de ser que incluye la unidad del tiempo, la reflexividad del yo y la cualidad vivencial de la experiencia. Los *qualia*, esos aspectos cualitativos e irreductibles de la percepción y del pensamiento, escapan a cualquier descripción desde la tercera persona. Edmund Husserl (1999), al describir la síntesis temporal, mostró que la conciencia humana no se da como una suma de instantes discretos, sino como una continuidad en la que pasado, presente y futuro se entrelazan en una unidad vivida. Thomas Nagel (1974) fue aún más enfático al afirmar que la experiencia subjetiva, el "cómo se siente ser", no puede ser reducida a una descripción objetiva. Esta imposibilidad de acceso desde fuera señala un límite insalvable para las máquinas, cuya actividad no posee ninguna vivencia de continuidad temporal ni de interioridad. En suma, sin experiencia, no hay conciencia, y sin conciencia no hay sujeto.

La autoconciencia, en cuanto reconocimiento de sí como sujeto de experiencia, implica un grado de reflexividad que ninguna máquina puede alcanzar. Agustín de Hipona (Agustine, 2001), en sus *Confesiones*, reveló esta capacidad del alma de volver sobre sí misma, lo cual indica una forma de subsistencia que no se agota en sus operaciones. Spaemann (1996) insistió en que la autoconciencia manifiesta un centro

ontológico irreductible, un sujeto que no puede ser explicado por sus partes. Immanuel Kant (1998), por su parte, introdujo la noción del "yo pienso" como condición de posibilidad de la experiencia unificada, indicando que sin esta apercepción trascendental no habría conciencia de identidad. Los sistemas artificiales no poseen esta reflexividad. Pueden monitorear su actividad, pero no experimentarse como sujetos. Su "autoconciencia" es, en el mejor de los casos, una metáfora impropia.

Tampoco puede pensarse la identidad personal desde un punto de vista funcional. La continuidad de un sujeto a través del tiempo requiere una unidad ontológica, no solo una estructura de datos o un patrón de conducta. Dennett (1991) ha propuesto una concepción narrativa de la identidad, pero su reducción a un "centro de gravedad" ficticio disuelve precisamente aquello que intenta explicar. MacIntyre (1981), por el contrario, ha defendido que la identidad personal se constituye a través de una narrativa con sentido moral, lo cual presupone un sujeto unitario y perdurable. Tomás de Aquino, con su noción de *subsistentia*, aporta una clave metafísica decisiva: el sujeto no es simplemente una colección de actos, sino una entidad que subsiste en sí misma. En este horizonte, la identidad personal no puede ser reducida a operaciones funcionales, sino que exige una base ontológica de unidad y permanencia.

El análisis crítico de la noción de conciencia artificial permite evidenciar, desde otra perspectiva, la insuficiencia de los modelos computacionales para captar la complejidad de la interioridad humana. La diferencia entre simular comportamientos externos y poseer experiencia consciente es radical. Alan Turing (1950), al proponer la prueba de imitación, inauguró una línea de investigación funcionalista que ha influido profundamente en la teoría de la inteligencia artificial. Sin embargo, como ha mostrado Block (1981), superar pruebas conductuales no garantiza la presencia de estados mentales genuinos. Su experimento del "*blockhead*" ilustra que una máquina puede simular perfectamente un comportamiento humano sin tener conciencia, lo cual demuestra que la simulación es ontológicamente vacía.

El experimento del cuarto chino, formulado por Searle (1980), refuerza esta tesis al mostrar que la manipulación sintáctica de símbolos no produce comprensión. Stevan Harnad (1990), en su planteamiento del problema del anclaje simbólico, enfatiza que las máquinas carecen de toda referencia intrínseca. El significado, en su forma plena, no puede ser constituido desde fuera. Requiere una conciencia que lo integre, lo reconozca y lo interprete desde su mundo vivido. Así, los sistemas de inteligencia artificial son semánticamente dependientes del ser humano. No comprenden, solo operan.

Las limitaciones ontológicas de estos sistemas se revelan, finalmente, en su incapacidad para unificar en un campo de conciencia los distintos elementos de la vida mental. Chalmers (1996) ha indicado que, aunque los modelos físicos explican funciones, no pueden dar cuenta de la experiencia en primera persona. Thomas Metzinger (2003) diferencia entre sistemas capaces de modelarse a sí mismos y aquellos que simplemente ejecutan funciones. La conciencia, en su acepción plena, implica subjetividad unificada, afectividad, memoria, intencionalidad, todo ello entrelazado en una experiencia coherente. Estas características están ausentes en los sistemas computacionales, cuya arquitectura no da lugar a ninguna forma de vivencia.

Habiendo examinado la inconmensurabilidad entre las operaciones cognoscitivas humanas y los procesos computacionales, conviene ahora dirigir la atención hacia otras dimensiones que completan la interioridad humana. Si la inteligencia, en su forma más elevada, ya revela una discontinuidad ontológica con respecto a las máquinas, aún más profunda es la diferencia cuando se consideran las facultades volitivas, afectivas y creativas. En ellas se expresan el querer, el sentir y el crear, dimensiones que, como se argumentará a continuación, remiten a un núcleo espiritual que no puede ser replicado artificialmente. La interioridad humana, lejos de agotarse en el pensar, se despliega en actos libres, en sentimientos que revelan sentido y en creaciones que brotan de una fuente irreductible a programación alguna.

5. Voluntad, afectividad y creatividad como manifestaciones de la interioridad espiritual

La cuestión del libre albedrío ha sido objeto de incontables controversias filosóficas, sin que pueda zanjarse mediante meros esquemas lógico-formales o apelaciones empíricas. A lo largo de nuestra investigación se ha sostenido que la autodeterminación ética constituye el núcleo del libre albedrío en sentido fuerte, es decir, no como mera capacidad de elegir entre alternativas disponibles, sino como facultad ontológicamente arraigada de orientar la propia vida conforme a fines racionalmente elegidos. Esta afirmación no puede ser comprendida dentro del marco del determinismo físico clásico, ya que la deliberación moral no se deja reducir a cadenas causales necesarias ni a procesos neurofisiológicos previsibles. La experiencia de decidir tras ponderar razones, evaluar consecuencias y discernir lo que conviene o lo que debe hacerse, implica una apertura al bien como tal, lo cual introduce un tipo de causalidad *sui generis*, ajena a las regularidades materiales.

Desde Karl Popper (1972) se ha planteado que sin libertad no puede haber racionalidad auténtica, ya que todo proceso argumentativo presupone la posibilidad de adoptar o rechazar razones. En una línea complementaria, Robert Kane (1996) introdujo el concepto de "acciones autoconstitutivas" para explicar cómo ciertas decisiones cruciales, tomadas bajo condiciones de conflicto interno e indeterminación, configuran la identidad moral del agente. Estas decisiones no son el producto de factores exteriores, sino el resultado de una lucha interior que expresa la autodeterminación del sujeto. Por otra parte, la tradición tomista, especialmente en la *Suma teológica* de Tomás de Aquino (Aquinas, 1947), refuerza esta visión al entender que el intelecto presenta a la voluntad fines inteligibles que ella puede asumir o rechazar. Este dinamismo racional-volitivo escapa a la lógica de la causalidad física, ya que opera según principios normativos, no según leyes de conservación o transmisión de energía.

Lo anterior da pie a una consideración más amplia sobre la imputabilidad moral, la cual solo tiene sentido si se presupone libertad racional. John Finnis (1980) vincula la ley natural con ciertos bienes humanos básicos cuya inteligibilidad se da a través de la razón práctica. El sujeto moral no actúa conforme a impulsos mecánicos, sino que puede reconocer y preferir lo que está objetivamente orientado al florecimiento humano. En esta dirección, Spaemann (1989) subraya que solo puede considerarse responsable quien está en condiciones de justificar sus actos mediante razones reconocidas como válidas dentro de una comunidad moral. Maritain (1994) lleva esta intuición aún más lejos al sostener que la estructura racional de la persona humana alberga una interioridad normativa, lo que significa que la ley moral no se impone desde fuera, sino que emana del propio dinamismo espiritual del agente.

La idea de autonomía que se desprende de lo anterior no puede reducirse a una independencia operativa o a una capacidad de actuar sin coerción externa. En términos kantianos, la verdadera autonomía consiste en darse a sí mismo la ley, es decir, en actuar conforme a principios que podrían ser universalizados (Kant, 1997). Pero esta autolegislación moral no es un ejercicio abstracto ni puramente formal, sino que se concreta en la capacidad de querer el bien en cuanto tal. MacIntyre (1981) ha puesto en evidencia que la racionalidad moral no flota en el vacío, sino que se encarna en narrativas de vida, en historias personales que dan sentido a las elecciones del agente. En este marco, la autonomía práctica requiere unidad deliberativa y una visión integrada del bien. Entonces, lo que distingue radicalmente a los agentes humanos de los sistemas artificiales es precisamente esta capacidad de vivir conforme a una teleología ética, y no simplemente de responder a estímulos según patrones algorítmicos. En definitiva, la autonomía no es operativa, sino ontológica, porque implica una forma de vida ordenada al valor.

Desde esta perspectiva, resulta necesario considerar también el papel de la afectividad en la vida moral. Las emociones humanas no deben concebirse como meras reacciones biológicas ni como epifenómenos

adaptativos, sino como actos intencionales cargados de sentido valorativo. Scheler (1973) ya había señalado que el sentimiento posee una estructura cognitiva propia, capaz de captar valores ideales no reducibles a objetos empíricos. Martha Nussbaum (2001) retoma esta tesis al afirmar que las emociones contienen juicios evaluativos sobre lo que importa profundamente, por lo que son componentes estructurales de la deliberación ética. La vida afectiva, en este sentido, no puede ser disociada de la vida racional, pues ambas forman parte de una misma dinámica de apertura al sentido.

Ahora bien, lo anterior permite discernir con mayor precisión las limitaciones insalvables de los sistemas artificiales en cuanto a la simulación emocional. Rosalind Picard (1997) distingue entre computación afectiva y emoción genuina, advirtiendo que ningún modelo estadístico, por complejo que sea, puede reproducir la vivencia afectiva en primera persona. Searle (1992) insiste en que toda emoción auténtica presupone una subjetividad cualitativa, es decir, un modo de ser para el cual algo importa de forma irreductible. La simulación de emociones por parte de sistemas computacionales es, en el mejor de los casos, una mímica funcional sin correlato experiencial. Las emociones humanas son vivencias evaluativas, no simples patrones de *output*.

Esta diferencia ontológica tiene consecuencias morales significativas. Las emociones auténticas, como la compasión, el arrepentimiento o la gratitud, no solo informan la acción moral, sino que la hacen posible en su sentido más pleno. Taylor (1985) habla de "evaluaciones fuertes" para referirse a aquellos juicios en los que está en juego nuestra identidad moral. Las emociones son, en efecto, vectores de orientación axiológica, estructuras mediante las cuales el sujeto se compromete con valores que lo configuran. Edith Stein (1989) complementa esta visión al destacar la función constitutiva de las emociones en la vida intersubjetiva. Sin afectividad, no hay comunidad ética posible, solo interacción mecánica o conveniencia estratégica. Por ello, los sistemas artificiales, privados de interioridad emocional, no pueden formar parte del mundo moral propiamente dicho.

En este punto resulta pertinente abrir la reflexión hacia una dimensión aún más reveladora de la vida espiritual: la creatividad. La producción artística, filosófica o científica no es solo una manifestación del ingenio humano, sino un testimonio de interioridad, de libertad expresiva y de trascendencia. Paul Ricoeur (1965) define la imaginación creativa como una mediación entre lo real y lo posible, capaz de generar nuevas formas de significación simbólica. Maritain (1953) sostiene que la creación artística surge de un acto espiritual que participa de una verdad interior, la cual no se reduce a datos ni a combinaciones sintácticas. Margaret Boden (1990) ha distinguido entre creatividad combinacional y transformacional, pero incluso los casos de mayor innovación computacional carecen de la intencionalidad sintetizadora que caracteriza la imaginación humana.

La ausencia de una intencionalidad estética auténtica en los sistemas artificiales resulta evidente al examinar el estatuto del arte generado por estos medios. Dominic Lopes (2014) ha señalado que el valor estético presupone no solo una forma, sino una autoría situada, es decir, un contexto cultural, histórico y existencial desde el cual la obra cobra sentido. Roger Scruton (2009) refuerza esta idea al sostener que la belleza no es un efecto visual, sino una forma intencional cargada de significado. Ninguna red neuronal puede generar sentido simbólico desde la experiencia, porque carece de vivencia.

El arte genuino, entonces, no es solo una técnica, sino una manifestación de vida interior. León Tolstói (Tolstoy, 1995) afirmó que el arte comunica sentimientos vividos desde el alma del creador, mientras que Gabriel Marcel (1949) identificó en la creación estética un acto existencial que expresa la unicidad del sujeto encarnado. Esta dimensión espiritual de la creatividad pone de relieve la imposibilidad de reducirla a un cálculo algorítmico. Simular el arte no equivale a crear, así como replicar emociones no implica sentir.

En última instancia, la creatividad humana testimonia una interioridad irreductible, capaz de dar forma a lo invisible mediante símbolos compartidos.

En ese sentido, la argumentación desarrollada hasta aquí ha permitido delinear un marco filosófico desde el cual la libertad, la afectividad y la creatividad aparecen como expresiones de una vida espiritual profundamente encarnada, cuya densidad ontológica no puede ser reducida a esquemas artificiales por sofisticados que estos sean. No se trata de una defensa romántica del humanismo clásico, sino de una constatación fenomenológica y estructural: el agente humano se caracteriza por una apertura al sentido, una capacidad de responder éticamente y una potencia de creación simbólica que lo sitúan en una dimensión ontológica distinta de cualquier sistema técnico.

Con estas premisas, resulta indispensable abordar en lo sucesivo una fundamentación metafísica más radical, centrada en los conceptos de vida, alma y muerte. Solo así podrá precisarse en qué consiste verdaderamente la estructura ontológica del viviente y por qué esta no puede ser generada, ni siquiera imitada en su esencia, por entidades artificiales. La siguiente sección se adentrará en este territorio, con el propósito de esclarecer la diferencia entre el habitar humano del mundo y la operatividad funcional de los autómatas. Solo a través de dicha exploración será posible evaluar con rigor los límites y las pretensiones de la inteligencia artificial en relación con la vida humana.

6. Distinciones metafísicas fundamentales entre los seres vivos y los sistemas artificiales

La consideración filosófica de la vida requiere una revisión profunda de los supuestos que, desde las ciencias empíricas, han moldeado la comprensión contemporánea de lo viviente. Frente a las perspectivas mecanicistas o funcionalistas, que tienden a reducir la vida a procesos cuantificables o a patrones de organización material, emerge la necesidad de recuperar una concepción ontológicamente robusta, que permita distinguir con claridad entre lo natural y lo artificial, entre lo sustancial y lo simulado. El punto de partida de esta revisión es la idea de que la vida no puede entenderse meramente como una suma de operaciones bioquímicas ni como el resultado de estructuras autoorganizadas. En cambio, se impone una interpretación en clave de forma sustancial, en la cual la vida se manifiesta como la actualización inmanente de un principio organizador que confiere unidad y finalidad al ser viviente.

En este sentido, la noción aristotélica de alma como acto primero de un cuerpo natural vivo (Aristotle, 1991) proporciona una clave interpretativa decisiva. Lejos de concebirse como un principio extrínseco o añadido desde fuera, el alma es la forma que constituye al ser viviente en su integridad, no un componente ni una función, sino el principio mismo de su ser. Esta visión es reforzada por Tomás de Aquino, quien sostiene que el alma racional es causa formal, no instrumental, del compuesto humano (Aquinas, 1947), lo cual implica que su presencia no es accidental ni contingente, sino esencial para la existencia del viviente como tal. Oderberg (2005) desarrolla esta idea en diálogo con la biología contemporánea, subrayando que la complejidad estructural no basta para explicar la unidad metafísica de lo viviente. La distinción entre estructura funcional y forma sustancial no puede ser ignorada sin incurrir en una confusión categorial que compromete la comprensión misma del fenómeno vital.

La vida, por tanto, no es reducible a organización. Esta afirmación cobra particular relevancia cuando se analiza la jerarquía de las almas, tal como la presenta la tradición aristotélico-tomista. En dicha jerarquía, el paso de lo vegetativo a lo racional no es un mero incremento en complejidad, sino una transición ontológica que introduce una ruptura cualitativa. El alma racional implica operaciones que no se dejan

reducir a la materialidad ni a la sensibilidad. El intelecto, entendido como potencia de lo universal y lo inmaterial, constituye una dimensión irreductible del ser humano (Aristotle, 1991; Aquinas, 1947). Spaemann (1996) vincula esta interioridad racional con la noción de persona, cuyo núcleo consiste precisamente en una subjetividad que no se deja agotar por la descripción empírica. Esta discontinuidad ontológica impide que la racionalidad sea tratada como una simple propiedad emergente, preservando su carácter fundante en la constitución del ser humano.

La vitalidad, por su parte, se expresa en el automovimiento, el cual no debe confundirse con la locomoción física. El automovimiento implica un dinamismo interno dirigido a fines propios, lo cual señala una estructura teleológica inmanente. Aristóteles desarrolla la noción de *entelecheia* como actualización del ser según su propia forma, y Tomás de Aquino prolonga esta intuición al integrar la causalidad final como principio de toda acción natural (Aquinas, 1947). En este marco, Jonas (1966) sostiene que la orientación hacia fines es un rasgo definitorio de la vida, cuya ausencia en los sistemas artificiales evidencia su carácter no-viviente. La actividad de una máquina, por muy compleja que sea, no se orienta espontáneamente hacia ningún fin propio, sino que responde a una programación externa. Esta ausencia de finalidad intrínseca impide hablar con propiedad de autonomía en el ámbito artificial.

Esta diferencia esencial entre lo viviente y lo artificial ha sido objeto de múltiples malentendidos en la filosofía contemporánea de la tecnología. La teoría de la autopoiesis de Humberto Maturana y Francisco Varela (1980), por ejemplo, aunque ofrece un modelo funcionalmente coherente de autoorganización, no capta la unidad sustancial que define a los seres vivos. El sistema autopoiético se mantiene a sí mismo en términos operacionales, pero carece de interioridad y de un principio formal que le confiera unidad ontológica. Terrence William Deacon (2011) reconoce esta limitación al subrayar que los sistemas artificiales carecen de finalidad intrínseca, lo cual los aleja irreversiblemente del estatuto de seres vivos. La organización cibernética, por tanto, no puede confundirse con la vida. La diferencia entre un bucle de retroalimentación y una estructura metafísica es insalvable desde el punto de vista ontológico.

La imposibilidad metafísica de generar vida artificialmente se deduce de esta comprensión sustancial de la vida. Oderberg (2007) afirma que los artefactos, por definición, son entes ontológicamente dependientes, cuya existencia se debe a un principio extrínseco. Simondon (1958), desde otra perspectiva, argumenta que la individuación técnica carece del dinamismo interno propio de la individuación vital. Así, incluso si se lograra replicar todas las características externas de un organismo vivo, se seguiría careciendo del principio metafísico que lo constituye como unidad viviente. La imitación morfológica no basta. La vida, en tanto acto del ser, no puede ser construida ni ensamblada, porque no es una función sino una forma.

La distinción entre operación externa y experiencia interior resulta aún más decisiva cuando se trata de comprender lo viviente desde la perspectiva de la conciencia. Stein (1989) introdujo la noción de *Erlebnis* para designar aquella dimensión afectiva y volitiva que acompaña a todo vivir auténtico. Esta experiencia subjetiva no se deriva de la complejidad ni de la información procesada, sino que constituye una dimensión irreductible del ser animado. Spaemann (1989) vincula esta interioridad con la existencia de un centro en primera persona, sin el cual no hay subjetividad ni vida consciente. En consecuencia, todo intento de inferir conciencia a partir de comportamientos observables está viciado de raíz si no se reconoce esta interioridad como fundamento. La simulación funcional, por muy elaborada que sea, no da lugar a una subjetividad vivida.

Lo anterior cobra su plena significación al considerar la muerte no como un mero cese biológico, sino como un acontecimiento ontológico que revela el carácter único del ser humano. Martin Heidegger (1962) introduce la idea del *Dasein* como ser-para-la-muerte, subrayando que la existencia humana se

estructura en función de la anticipación de su propio fin. Esta conciencia de la muerte no se reduce a un conocimiento abstracto, sino que configura la totalidad de la existencia. Marcel (1951) interpreta la muerte como revelación del ser espiritual, ya que solo quien se sabe finito puede experimentar la muerte como destino. Esta comprensión implica que la muerte es una experiencia que pertenece exclusivamente a quienes poseen una vida interior.

La finitud temporal, lejos de ser una limitación meramente negativa, constituye la condición de posibilidad del desarrollo moral. Ricoeur (1990) sostiene que la identidad narrativa depende de esta conciencia de la muerte, en tanto que permite articular una vida con sentido, sujeta a la responsabilidad y al juicio. Taylor (1989) refuerza esta idea al argumentar que la temporalidad vivida es el fundamento de las evaluaciones morales profundas. Sin esta dimensión, no hay acción significativa, ni posibilidad de imputar responsabilidad. Los sistemas artificiales, al carecer de autoconciencia y de temporalidad vivida, no pueden participar de esta estructura ética. Su operación, por muy avanzada que sea, no está informada por la urgencia moral ni por la necesidad de configurarse como una totalidad significativa.

Por ello, resulta inapropiado hablar de muerte artificial o de experiencia existencial en relación con entidades computacionales. Searle (1992) ha sostenido con firmeza que la conciencia de máquina es una imposibilidad ontológica, y Nagel (1974) ha demostrado que toda experiencia subjetiva posee un carácter cualitativo irreductible, entendido como el carácter fenomenológico de la vivencia o el modo cualitativo de experimentarla desde dentro. La muerte, por tanto, no puede ser simulada, porque presupone una subjetividad capaz de anticipar su fin, de temerlo, de integrarlo en su narrativa vital. Los sistemas artificiales, al carecer de esta interioridad, se mantienen al margen de la finitud y del sentido que de ella se deriva.

La conclusión que se impone tras este recorrido filosófico es que la vida auténtica, la conciencia subjetiva y la experiencia de la muerte no pueden ser replicadas ni emuladas mediante sistemas artificiales, por muy sofisticados que estos sean. Esta irreductibilidad ontológica no es una simple diferencia de grado, sino una distinción categorial que debe informar toda reflexión ética, jurídica y tecnológica. Las consecuencias de esta constatación son múltiples y profundas. Corresponde ahora explorar cómo esta distinción ontológica puede y debe orientar las concepciones contemporáneas de dignidad humana, establecer límites normativos para el desarrollo tecnológico y fundamentar los principios que regulen una interacción justa entre seres humanos y artefactos inteligentes. La siguiente sección se abocará a examinar estas implicaciones con el mismo rigor ontológico, atendiendo a los desafíos concretos que plantea la irrupción de tecnologías emergentes en el entramado ético y jurídico de las sociedades contemporáneas.

7. Implicancias éticas y jurídicas de la distinción ontológica entre el ser humano y los sistemas artificiales

La noción de dignidad humana ocupa un lugar central tanto en la arquitectura ontológica como en la normativa moral de la cultura filosófica occidental. Lejos de reducirse a una categoría jurídico-política o a una convención sociológica mutable, la dignidad se presenta como un principio primero, irreductible y anterior a cualquier forma de positivación. En su raíz más profunda, la dignidad no depende de condiciones externas, sino que remite a la unicidad ontológica del ser humano en cuanto ser racional, libre y espiritual. Esta concepción permite, por un lado, preservar la inviolabilidad de la persona frente a todo intento de cosificación técnica y, por otro, establecer un marco normativo capaz de resistir los embates del relativismo ético y del funcionalismo tecnocientífico.

En este horizonte, resulta particularmente iluminadora la convergencia conceptual entre dos tradiciones filosóficas que, pese a sus diferencias metodológicas y metafísicas, reconocen un mismo núcleo antropológico. Por un lado, la antropología tomista sitúa la dignidad en el apetito racional, en tanto este orienta al ser humano hacia el bien universal, lo cual configura su estructura finalista y le otorga una posición ontológica singular (Aquinas, 1947). Por otro, el pensamiento moral kantiano sostiene que la persona debe ser siempre tratada como un fin en sí mismo y nunca como un medio para otros fines, principio que funda la autonomía moral y el imperativo categórico (Kant, 1997). La síntesis articulada por Spaemann (1989) entre ambas tradiciones subraya que la dignidad no puede derivarse de funciones, capacidades o rendimientos, sino que se encuentra inscrita en la propia estructura de la persona. Esta continuidad entre ontología y normatividad impide que la dignidad quede a merced de criterios extrínsecos y cambiantes, ya sean tecnológicos, jurídicos o culturales.

Definir la dignidad como valor intrínseco conduce a una consecuencia normativa de gran relevancia: la prohibición de toda instrumentalización del ser humano. Nussbaum (2006), al vincular la dignidad con la posibilidad de una vida plenamente humana, refuerza su carácter irreemplazable y no negociable. La ética de la responsabilidad propuesta por Jonas (1984), fundada en la vulnerabilidad y en la irrepetibilidad de la vida humana, aporta un fundamento adicional para oponerse a cualquier lógica que reduzca al ser humano a un nodo funcional dentro de sistemas optimizadores. En esta línea, tratar a los sistemas de inteligencia artificial como herramientas no implica que la persona humana deba ser comprendida bajo los mismos parámetros funcionales. Cualquier intento de establecer una simetría antropomórfica entre lo humano y lo artificial no solo resulta filosóficamente inadecuado, sino también éticamente incoherente, al diluir la distinción fundamental entre funcionalidad técnica y valor ontológico.

La dignidad no solo constituye una categoría metafísica, sino también el fundamento irrenunciable de los derechos fundamentales. Así lo reconoce la Declaración Universal de los Derechos Humanos (Universal Declaration of Human Rights, 1948), que en su artículo primero afirma la igualdad en dignidad y derechos de todos los seres humanos, fundándola en la razón y la conciencia como principios de libertad y fraternidad. La perspectiva histórica ofrecida por Taylor (1994) permite comprender la emergencia de la dignidad como valor secular, dotado de una trascendencia inmanente en las éticas modernas. Por otro lado, Maritain (1994), desde una visión metafísica, sostiene que la dignidad precede a todo derecho positivo y que constituye la condición ontológica de posibilidad del orden jurídico. Esta fundamentación evita confundir la personalidad jurídica, que puede ser atribuida artificialmente, con la persona en sentido ontológico, que posee un valor no delegable ni replicable. Por consiguiente, la dignidad humana se configura como el límite inquebrantable de toda tentativa de equiparación entre la persona y cualquier artefacto inteligente.

Ahora bien, asumir la centralidad de la dignidad humana conlleva inevitablemente una interrogación sobre los límites éticos del desarrollo tecnológico, en particular en lo que respecta a la inteligencia artificial. Dichos límites no pueden establecerse en función de cálculos pragmáticos o de análisis de costo-beneficio, sino que deben derivarse de una antropología normativa que reconozca la irreductibilidad del ser humano a procesos automáticos o cuantificables. En esta perspectiva, Jean-Luc Nancy (2000) advierte sobre el riesgo de homogeneización que impone la tecnociencia al borrar la alteridad y lo inconmensurable. Hubert Dreyfus (1992), por su parte, critica los reduccionismos epistemológicos que intentan modelar la inteligencia humana desde marcos computacionales, desatendiendo sus dimensiones existenciales y encarnadas. De ello se desprende que la singularidad humana no es reductible a lo biológico, sino que se enraíza en lo metafísico. Solo a partir de este reconocimiento puede evitarse la deriva hacia formas de deshumanización estructural.

La tecnología, desde esta óptica, debe ser concebida como instrumento subsidiario de la naturaleza humana y no como agente normativo autónomo. Jacques Ellul (1964) alerta sobre el carácter autorreferencial del sistema técnico, cuyo crecimiento obedece más a su propia lógica interna que a criterios humanos. Jonas (1984), al proponer el principio de precaución, señala que el ejercicio del poder técnico requiere un anclaje ético que le impida sobrepasar límites irreversibles. MacIntyre (1999) enfatiza que la tecnicidad debe subordinarse a prácticas orientadas a la virtud, donde los fines sean modelados por una concepción del bien humano y no por lógicas de eficiencia o rendimiento. Esta concepción de la técnica como subsidiaria exige una normatividad centrada en la persona y no en las capacidades de los artefactos.

No obstante, toda reflexión ética sobre la inteligencia artificial debe traducirse en marcos regulatorios robustos, capaces de responder a la complejidad del fenómeno sin reducirlo a modelos meramente empíricos. Floridi (2013), al desarrollar una ética de la información, ofrece una perspectiva útil, aunque demasiado abstracta para generar normativas con fundamento ontológico. Rafael Capurro (2008) propone una ética intercultural informacional que busca integrar la diversidad de perspectivas sin perder profundidad filosófica. En última instancia, la regulación efectiva de la inteligencia artificial exige una antropología metafísica que oriente la formulación de derechos, deberes y límites. No se trata simplemente de alcanzar mínimos éticos consensuados, sino de sostener una reflexión sustantiva que preserve la centralidad del ser humano frente a la creciente complejidad de los sistemas artificiales.

Lo anterior conduce necesariamente a replantear el modo en que se configura la relación entre el ser humano y la inteligencia artificial, no desde la rivalidad o la subordinación absoluta, sino desde una complementariedad bien ordenada. Esta relación solo puede ser ética si respeta la primacía ontológica de la persona, evitando tanto la sustitución como la idealización tecnológica. En este sentido, el principio de subsidiariedad ofrece una vía fecunda: los sistemas artificiales deben estar diseñados para asistir las capacidades humanas sin pretender simular ni replicar la personalidad. Aunque originalmente formulado en el ámbito social por Pío XI (1931) en *Quadragesimo anno*, este principio resulta igualmente aplicable al campo técnico. Wendell Wallach y Colin Allen (2009) defienden la necesidad de arquitecturas éticas embebidas que remitan siempre al juicio humano, lo cual preserva la centralidad del sujeto moral en la toma de decisiones.

La atribución de juicio moral y responsabilidad sigue siendo un rasgo exclusivo del ser humano. Ningún sistema algorítmico, por sofisticado que sea, posee la conciencia reflexiva y la intencionalidad necesarias para asumir deberes éticos. Michael J. Sandel (2009) critica con agudeza la delegación de decisiones morales a entes no conscientes, señalando los riesgos de despersonalización que tal proceso implica. Searle (1992), en su crítica a la inteligencia artificial fuerte, recuerda que la intencionalidad es inseparable de la conciencia, y que por tanto las máquinas no pueden ser sujetos morales. Afirmar lo contrario supone desdibujar los límites entre agente técnico y sujeto personal, con consecuencias jurídicas y éticas profundamente problemáticas.

Por ende, la gobernanza de la inteligencia artificial debe orientarse al florecimiento humano y no a la optimización sistémica o al crecimiento ilimitado de capacidades técnicas. En este sentido, la propuesta de Amartya Sen y Matha Nussbaum (1993) de vincular el desarrollo de capacidades con el diseño de políticas públicas ofrece un marco valioso para pensar una inteligencia artificial al servicio de la vida humana en todas sus dimensiones. Shannon Vallor (2016) complementa esta visión al proponer el cultivo de virtudes tecno-morales que fortalezcan la agencia en contextos mediados por tecnologías complejas. El objetivo no puede ser otro que la eudaimonía, entendida como realización plena del ser humano. Solo así la interacción entre lo humano y lo artificial podrá desplegarse dentro de un orden ético-político que preserve la dignidad, afirme la libertad y promueva una justicia verdaderamente humana.

8. Conclusión

El análisis llevado a cabo permite sostener que la diferencia entre el ser humano y los sistemas artificiales no puede reducirse a una cuestión de complejidad funcional o nivel de desarrollo computacional. Se trata, en rigor, de una discontinuidad ontológica que afecta de manera decisiva la legitimidad de los marcos normativos aplicables a las tecnologías emergentes. Esta diferencia estructural, lejos de ser contingente o superable por el progreso técnico, remite a la imposibilidad de que una entidad no consciente instancie atributos como libertad, intencionalidad o juicio moral. La conciencia humana, en tanto manifestación de interioridad y no mera representación de estados funcionales, establece un umbral cualitativo que las máquinas, por su propia naturaleza, no pueden cruzar. Por ello, cualquier asimilación normativa entre ambos planos resulta conceptualmente impropia y éticamente riesgosa.

Desde el punto de vista teórico, este trabajo se inscribe en una línea de renovación del diálogo entre la antropología filosófica y los desarrollos recientes en filosofía de la mente. La convergencia entre categorías de la tradición aristotélico-tomista y aportes fenomenológicos permite superar tanto los dualismos que escinden cuerpo y mente como las versiones más crudas del reduccionismo neurocomputacional. Recuperar nociones como forma sustancial, finalidad intrínseca o autotrascendencia no implica un retroceso hacia esquemas premodernos, sino una tentativa crítica por sostener un marco conceptual robusto frente a la erosión metafísica que produce la colonización tecnocientífica del discurso sobre lo humano. Esta recuperación, aunque parcial y perfectible, resulta clave para preservar un núcleo de sentido que resista las confusiones inducidas por el lenguaje técnico y los imaginarios computacionales.

Las consecuencias normativas derivadas de esta distinción ontológica son directas y exigentes. Aceptar que la dignidad humana es irreductible impone la obligación de diseñar políticas y marcos regulatorios que reconozcan la exclusividad del juicio moral humano en las decisiones que comprometen valores fundamentales. En tal sentido, la idea de "zonas de reserva humana" cobra especial relevancia, pues señala aquellos ámbitos donde la delegación algorítmica no puede justificarse éticamente, aunque fuera viable técnicamente. Esta concepción requiere adoptar el principio de subsidiariedad tecnológica, según el cual las herramientas artificiales deben complementar las capacidades humanas sin sustituirlas ni replicarlas. La aplicación concreta de este principio demanda el desarrollo de estándares de diseño, mecanismos de rendición de cuentas y prácticas institucionales que garanticen la supremacía del juicio humano sobre la automatización funcional.

La proyección futura de este enfoque invita a abrir nuevas líneas de indagación que atraviesan tanto el plano filosófico como el jurídico y el técnico. En el primer caso, resulta indispensable explorar aquellas dimensiones de la experiencia humana que, por su cualidad fenomenológica o su densidad simbólica, resisten toda forma de modelización algorítmica. Ámbitos como el sufrimiento, el deseo, la contemplación estética o la experiencia del perdón exigen categorías interpretativas ajenas al cálculo. En el segundo caso, urge construir una arquitectura normativa que sea capaz de evitar tanto el antropomorfismo simplificador como el utilitarismo de mercado. Una vía prometedora radica en integrar la reflexión filosófica en los procesos de diseño tecnológico mediante modelos de gobernanza que conecten el saber humanista con la ingeniería computacional. Esta convergencia, aún incipiente, podría desembocar en una hermenéutica crítica de la técnica que, sin demonizarla, le asigne su lugar en el conjunto de la vida humana.

El problema de fondo que plantea la inteligencia artificial remite, en última instancia, a la cuestión más antigua y persistente de la tradición filosófica: qué significa ser humano. No se trata de oponer un humanismo esencialista a la expansión tecnológica ni de celebrar ingenuamente las promesas del progreso computacional. Se trata, más bien, de elaborar una comprensión de la dignidad humana que

asuma la complejidad del presente sin renunciar a los límites ontológicos que hacen posible la libertad, la responsabilidad y el sentido. Solo sobre esta base puede pensarse un futuro en el cual la inteligencia artificial contribuya efectivamente al florecimiento humano, sin disolver las condiciones mismas de su posibilidad. La tarea es tan urgente como delicada, pues de su acierto depende la configuración de un orden tecnológico que respete la primacía de la vida sobre la eficiencia. En ello nos va no solo la justicia presente, sino la herencia que dejaremos a quienes aún no han nacido.

Declaración de conflictos de intereses:

El autor declara no tener conflicto de interés.

Declaración de uso de IA

Para el presente artículo se usó Chat GPT en su versión 4o para corregir errores ortográficos, de puntuación y mejorar legibilidad. El prompt usado fue "corrige los errores de redacción y ortográficos del texto y de puntuación para mejorar la legibilidad". Luego de la respuesta obtenida, se revisó que se haya mantenido la información proporcionada por el autor. La herramienta no ha sido usada para creación autónoma de contenido.

Referencias bibliográficas

Putnam, H. (1975). *Mind, language and reality* (Vol. 2). Cambridge University Press.

Augustine. (2001). *Confessions* (H. Chadwick, Trans.). Oxford University Press. (Original work published 397).

Aquinas, T. (1947). *Summa Theologica* (Fathers of the English Dominican Province, Trans.). Benziger Bros. (Original work published 1265–1274).

Aristotle. (1991). *De Anima* (R. D. Hicks, Trans.; 25th illus. ed.). Prometheus Books.

Aristotle. (2001). *Physics* (R. P. Hardie & R. K. Gaye, Trans.). In J. Barnes (Ed.), *The complete works of Aristotle: The revised Oxford translation* (Vol. 1, pp. 315–446). Princeton University Press.

Block, N. (1980). *Troubles with functionalism*. In N. Block (Ed.), *Readings in philosophy of psychology* (Vol. 1, pp. 268–305). Harvard University Press.

Block, N. (1981). Psychologism and behaviorism. *Philosophical Review, 90*(1), 5–43. https://doi.org/10.2307/2184371

Boden, M. A. (1990). *The creative mind: Myths and mechanisms*. Basic Books.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Brentano, F. (1995). *Psychology from an empirical standpoint* (L. L. McAlister, Ed., A. C. Rancurello, D. B. Terrell, & L. L. McAlister, Trans.). Routledge. (Original work published 1874).

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4*. arXiv. https://doi.org/10.48550/arXiv.2303.12712

Capurro, R. (2008). Intercultural information ethics. *Ethics and Information Technology, 10*(2–3), 91–101. https://doi.org/10.1007/s10676-008-9179-3

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Deacon, T. W. (2011). *Incomplete nature: How mind emerged from matter*. W. W. Norton.

Dennett, D. C. (1991). *Consciousness explained*. Little, Brown and Company.

Dretske, F. (1981). *Knowledge and the flow of information*. MIT Press.

Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason*. MIT Press.

Ellul, J. (1964). *The technological society* (J. Wilkinson, Trans.). Vintage Books. (Original work published 1954).

Finnis, J. (1980). *Natural law and natural rights*. Oxford University Press.

Floridi, L. (2013). *The ethics of information*. Oxford University Press.

Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Gabriel, M. (2015). *Warum es die Welt nicht gibt* [Why the world does not exist]. Ullstein.

Gehlen, A. (1950). *Der Mensch: Seine Natur und seine Stellung in der Welt* [Man: His nature and place in the world]. Aula-Verlag.

Gilson, É. (1939). *The spirit of Thomism*. Gifford Lectures.

Guardini, R. (1950). *The end of the modern world*. Henry Regnery Company.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42*(1–3), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Harper & Row. (Original work published 1927).

Hui, Y. (2016). *The question concerning technology in China: An essay in cosmotechnics*. Urbanomic.

Husserl, E. (1999). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: First book* (F. Kersten, Trans.). Kluwer Academic. (Original work published 1913).

Jonas, H. (1966). *The phenomenon of life: Toward a philosophical biology*. Harper & Row.

Jonas, H. (1984). *The imperative of responsibility: In search of an ethics for the technological age* (H. Jonas, Trans.). University of Chicago Press.

Kane, R. (1996). *The significance of free will*. Oxford University Press.

Kant, I. (1997). *Groundwork of the metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press. (Original work published 1785).

Kant, I. (1998). *Critique of pure reason* (P. Guyer & A. W. Wood, Trans. & Eds.). Cambridge University Press. (Original work published 1781).

Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT Press.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*(4), 354–361. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x

Lonergan, B. J. F. (1957). *Insight: A study of human understanding*. Longmans, Green and Co.

Lopes, D. M. (2014). *Beyond art*. Oxford University Press.

MacIntyre, A. (1981). *After virtue: A study in moral theory*. University of Notre Dame Press.

MacIntyre, A. (1999). *Dependent rational animals: Why human beings need the virtues*. Open Court.

Marcel, G. (1949). *The mystery of being: Volume I – Reflection and mystery* (G. S. Fraser, Trans.). St. Augustine's Press.

Marcel, G. (1951). *The mystery of being: Volume II – Faith and reality* (R. Rosthal, Trans.). St. Augustine's Press.

Marcus, G. (2023). *The next decade in AI: Four steps towards robust artificial intelligence*. arXiv. https://doi.org/10.48550/arXiv.2301.06545

Maritain, J. (1994). *The person and the common good* (J. J. Fitzgerald, Trans.; Reprint ed.). University of Notre Dame Press. (Original work published 1947).

Maritain, J. (1953). *Creative intuition in art and poetry*. Pantheon Books.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. D. Reidel.

McGinn, C. (1999). *The mysterious flame: Conscious minds in a material world*. Basic Books.

Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. MIT Press.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450. https://doi.org/10.2307/2183914

Nancy, J.-L. (2000). *Being singular plural* (R. D. Richardson & A. E. O'Byrne, Trans.). Stanford University Press.

Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.

Nussbaum, M. C. (2006). *Frontiers of justice: Disability, nationality, species membership*. Harvard University Press.

Oderberg, D. S. (2005). *Real essentialism*. Routledge.

Oderberg, D. S. (2007). Teleology: Inorganic and organic. In A. J. Beckwith (Ed.), *Contemporary perspectives on natural teleology* (pp. 137–164). Cambridge Scholars Publishing.

Picard, R. W. (1997). *Affective computing*. MIT Press.

Pío XI. (1931). *Quadragesimo anno [On reconstructing the social order]*. Vatican Library. https://www.vatican.va/content/pius-xi/en/encyclicals/documents/hf_p-xi_enc_19310515_quadragesimo-anno.html

Polanyi, M. (1966). *The tacit dimension*. University of Chicago Press.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Oxford University Press.

Putnam, H. (1967). *Psychological predicates*. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). University of Pittsburgh Press.

Ricoeur, P. (1965). *De l'interprétation: Essai sur Freud*. Éditions du Seuil.

Ricoeur, P. (1990). *Oneself as another* (K. Blamey, Trans.). University of Chicago Press.

Ross, J. F. (1992). Immaterial aspects of thought. *The Journal of Philosophy, 89*(3), 136–150. https://doi.org/10.2307/2026781

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4.ª ed.). Pearson.

Sandel, M. J. (2009). *Justice: What's the right thing to do?* Farrar, Straus and Giroux.

Scheler, M. (1928). *Die Stellung des Menschen im Kosmos* [The human place in the cosmos]. Francke.

Scheler, M. (1973). *Formalism in ethics and non-formal ethics of values* (M. S. Frings & R. L. Funk, Trans.). Northwestern University Press. (Original work published 1913–1916).

Scruton, R. (2009). *Beauty: A very short introduction*. Oxford University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–457. https://doi.org/10.1017/S0140525X00005756

Searle, J. R. (1992). *The rediscovery of the mind*. MIT Press.

Sen, A., & Nussbaum, M. C. (Eds.). (1993). *The quality of life*. Oxford University Press.

Simondon, G. (1958). *Du mode d'existence des objets techniques* [On the mode of existence of technical objects]. Éditions Aubier.

Spaemann, R. (1989). *Persons: The difference between someone and something* (O. O'Donovan, Trans.). Oxford University Press.

Spaemann, R. (1996). *Glück und Wohlwollen: Versuch über Ethik*. Klett-Cotta.

Stein, E. (1989). *On the problem of empathy* (W. Stein, Trans.). ICS Publications. (Original work published 1917).

Taylor, C. (1985). *Human agency and language: Philosophical papers I*. Cambridge University Press.

Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Harvard University Press.

Taylor, C. (1994). The politics of recognition. In A. Gutmann (Ed.), *Multiculturalism: Examining the politics of recognition* (pp. 25–73). Princeton University Press.

Tolstoy, L. (1995). *What is art?* (R. Pevear & L. Volokhonsky, Trans.). Penguin Books. (Original work published 1897).

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin, 215*(3), 216–242. https://doi.org/10.2307/25470707

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Universal Declaration of Human Rights. (1948). United Nations. https://www.un.org/en/about-us/universal-declaration-of-human-rights

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. Oxford University Press.