

ANNA PEIRATS\*  
FRANCISCO ARTEAGA\*\*

# VERDAD, PRUDENCIA Y MEDIACIÓN TÉCNICA: FUNDAMENTACIÓN TEOLÓGICA DE LA RESPONSABILIDAD ANTE LA INTELIGENCIA ARTIFICIAL

Fecha de recepción: 26 de enero de 2026

Fecha de aceptación: 03 de marzo 2026

**RESUMEN:** La expansión de la inteligencia artificial (IA) en la producción y mediación del conocimiento plantea desafíos críticos y oportunidades sin precedentes para la teología moral. Este artículo analiza el fenómeno desde una perspectiva sistemática y propositiva, superando tanto la tecnofobia como el reduccionismo normativo. El estudio sostiene que el desarrollo de la IA puede integrarse en una fundamentación teológica de la responsabilidad que permita al ser humano ampliar las posibilidades del conocimiento y de la acción, siempre bajo el gobierno de la prudencia. Este artículo establece una distinción operativa entre tres dimensiones de la verdad: informativa (datos), práctica (juicio prudencial) y revelada (fe). Basándose en la tradición tomista y en la reciente nota *Antiqua et Nova* (2025), se desarrolla un modelo en el que la IA desempeña una función legítima y potente en la fase de deliberación (*consilium*), mientras que el mandato de la acción (*im-*

---

\* Universidad Católica de Valencia San Vicente Mártir: anna.peirats@ucv.es; ORCID: <https://orcid.org/0000-0001-6957-262X>

\*\* Universidad Católica de Valencia San Vicente Mártir: francisco.artega@ucv.es; ORCID: <https://orcid.org/0000-0002-9492-6108>

*perium*) se reafirma como una facultad vinculada de modo intransferible al sujeto humano. Este marco teórico se articula con la reciente regulación europea sobre inteligencia artificial (AI Act) para ofrecer una síntesis en la que el desarrollo tecnológico se integra como un instrumento de esperanza y justicia, al servicio de la dignidad personal y del bien común.

PALABRAS CLAVE: inteligencia artificial; teología moral; prudencia; verdad; responsabilidad; *Antiqua et Nova*; tomismo; Ley de IA; antropología digital.

### ***Truth, Prudence, and Technical Mediation: A Theological Foundation of Responsibility in the Age of Artificial Intelligence***

ABSTRACT: The expansion of Artificial Intelligence (AI) in the production and mediation of knowledge poses critical challenges and unprecedented opportunities for moral theology. This article analyzes the phenomenon from a systematic and proactive perspective, moving beyond both technophobia and normative reductionism. The research promotes the enhancement of AI development through a theological foundation of responsibility that enables humans to achieve superior levels of knowledge and efficiency, always under the governance of prudence. The study establishes an operational distinction between three dimensions of truth: informational (data), practical (prudential judgment), and revealed (faith). Drawing on the Thomistic tradition and the recent *Antiqua et Nova* instruction (2025), a model is developed in which AI fulfils a legitimate and powerful role in the deliberative phase (*consilium*), while the command of action (*imperium*) is reaffirmed as a faculty untransferably linked to the human subject. This theoretical framework harmonizes with recent European regulation (AI Act) to offer a synthesis in which technological development is integrated as an instrument of hope and justice at the service of personal dignity and the common good.

KEY WORDS: artificial intelligence; moral theology; prudence; truth; responsibility; *Antiqua et Nova*; Thomism; AI Act; digital anthropology.

## 1. INTRODUCCIÓN: EL «CAMBIO DE ÉPOCA» Y LA CUESTIÓN DE LA VERDAD

La publicación de la nota *Antiqua et Nova* el 28 de enero de 2025, fruto de la colaboración entre el Dicasterio para la Doctrina de la Fe y el Dicasterio para la Cultura y la Educación, consolida la reflexión de la Iglesia sobre el desafío digital<sup>1</sup>. La expresión *Antiqua et Nova* («Antiguo

---

<sup>1</sup> Cf. Dicasterio para la Doctrina de la Fe y Dicasterio para la Cultura y la Educación. *Antiqua et Nova. Nota sobre la relación entre la inteligencia artificial y la inteligencia humana*. Ciudad del Vaticano, 2025, n. 1.

y Nuevo») evoca el pasaje evangélico del escriba que extrae de su arca cosas nuevas y viejas (Mt 13,52), imagen que sugiere que la tradición cristiana posee un caudal sapiencial capaz de iluminar incluso las novedades tecnológicas más disruptivas. El documento fue aprobado por el papa Francisco el 14 de enero de 2025 y publicado el 28 de enero del mismo año, coincidiendo con la memoria de santo Tomás de Aquino.

Este documento identifica en la inteligencia artificial (IA) el epicentro de un «cambio de época» cuya urgencia antropológica trasciende la mera innovación técnica. El concepto de «cambio de época» (distinto a una mera «época de cambios») es una clave hermenéutica constante en el magisterio de Francisco para describir la transformación irreversible que la tecnología digital ha operado en la cultura y la autocomprensión humana. La IA se sitúa en el centro de este cambio<sup>2</sup>. La singularidad de este momento histórico radica en la naturaleza mimética de la tecnología: las variantes generativas actuales producen artefactos textuales, visuales y sonoros que, al resultar indistinguibles de la acción humana, precipitan una crisis de veracidad en el debate público. Ante esta capacidad de simulación, la reflexión teológica debe acoger el desafío evitando lecturas funcionales que reduzcan la dignidad de la persona a su mera capacidad de procesamiento de datos, desde una métrica en la que la máquina siempre tenga ventaja. La distinción entre operatividad funcional y pensamiento (reflexión, conciencia, intencionalidad) es el núcleo de la argumentación antropológica de la nota. El texto subraya que definir la «inteligencia» solo por resultados externos es un error categorial que compromete la dignidad humana<sup>3</sup>.

Para abordar la responsabilidad moral en este contexto, cabe precisar el estatuto ontológico de la tecnología más allá del uso equívoco del término «inteligencia». El papa Francisco insta a recordar que la IA no es una forma de inteligencia, es un producto de ella<sup>4</sup>. La filosofía de la tecnología y el reciente magisterio convergen al señalar que la realidad técnica actual, desde la IA estrecha (IAN) hasta los modelos generativos de propósito general (GPAI), opera mediante inferencias estadísticas y

---

<sup>2</sup> Cf. *Antiqua et Nova*, n. 4.

<sup>3</sup> Cf. *Antiqua et Nova*, n. 12.

<sup>4</sup> Cf. Dicasterio para la Doctrina de la Fe y Dicasterio para la Cultura y la Educación. *Antiqua et Nova*, n. 35; Francisco. *Mensaje para la 58.ª Jornada Mundial de las Comunicaciones Sociales* (2024).

patrones sintácticos ajenos a la comprensión semántica. El Reglamento (UE) 2024/1689 (Ley de IA) define los modelos de IA de uso general por su capacidad para realizar una amplia gama de tareas distintas. Autores como Cortina y Bender describen estos sistemas como «loros estocásticos» que, aunque gestionan con solvencia la probabilidad de sucesión léxica, carecen de la intencionalidad necesaria para la agencia moral<sup>5</sup>. Otros investigadores, sin embargo, han explorado la posibilidad de formas emergentes de representación conceptual moral<sup>6</sup> en arquitecturas avanzadas de aprendizaje automático. El presente artículo pretende examinar sus implicaciones desde la teología moral y la antropología teológica.

De igual modo, la hipótesis de la «inteligencia artificial general» (AGI) se presenta a menudo como una proyección transhumanista que la teología identifica como una forma contemporánea de absolutización tecnológica al confundir la acumulación cuantitativa de potencia de cálculo con el salto cualitativo hacia el espíritu o la conciencia. *Antiqua et Nova* advierte que buscar en la IA un «Otro» mayor es una forma de idolatría: «no es la IA quien es divinizada, sino el ser humano... esclavo de su propia obra»<sup>7</sup>.

Este debate ha sido abordado también en la reflexión actual sobre ética de la tecnología y filosofía de la información. Diversos autores han subrayado que la expansión de los sistemas algorítmicos constituye una transformación profunda de las condiciones epistemológicas y sociales del conocimiento. Se ha señalado que los algoritmos automatizan operaciones técnicas y, además, median procesos sociales, económicos y políticos, por lo que influyen en cómo se interpretan los datos y en qué decisiones se consideran plausibles en cada situación. En este sentido,

---

<sup>5</sup> Sobre la crítica al «loro estocástico» (*stochastic parrot*), que une palabras por probabilidad estadística sin intención comunicativa ni acceso al significado (semántica), véase: Adela Cortina. *¿Ética o ideología de la inteligencia artificial? El eclipse de la razón comunicativa en una sociedad tecnolozada*. Barcelona: Paidós, 2024, cap. 4.

<sup>6</sup> Cf. Emily M. Bender, Timnit Gebru, Angelina McMillan-Major y Margaret Mitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?". En *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. Nueva York: Association for Computing Machinery, 2021. En este trabajo los autores argumentan que los grandes modelos de lenguaje generan textos plausibles mediante la detección de patrones probabilísticos en grandes corpus lingüísticos, sin poseer comprensión semántica ni intencionalidad.

<sup>7</sup> Cf. *Antiqua et Nova*, n. 105.

B. Mittelstadt et al. analizan el papel de los algoritmos como mediadores de procesos sociales, económicos y políticos en la sociedad actual. Estos sistemas participan cada vez más en procesos de orientación y toma de decisiones que antes correspondían exclusivamente a agentes humanos. De esta mediación se derivan riesgos éticos asociados al uso de algoritmos, como la opacidad, el sesgo, la responsabilidad y el desplazamiento del juicio humano en procesos de decisión<sup>8</sup>.

En este contexto, se ha destacado la necesidad de desarrollar marcos normativos y virtudes intelectuales capaces de orientar el uso de estas tecnologías. Luciano Floridi, por ejemplo, ha desarrollado una de las propuestas más influyentes en la filosofía contemporánea de la información, al interpretar el entorno digital como una nueva condición estructural del conocimiento humano y analizar sus implicaciones éticas y políticas<sup>9</sup>. La reflexión teológica puede dialogar con estos enfoques al ofrecer una antropología de la persona que permite interpretar la inteligencia artificial dentro del horizonte de la responsabilidad moral, la dignidad humana y el bien común.

Bajo estas premisas, el presente artículo sostiene que la inteligencia artificial constituye una hipertrofia instrumental de la *ratio*, entendida como la dimensión discursiva y calculadora del entendimiento humano. La operatividad algorítmica amplifica esta capacidad de procesamiento, pero permanece ajena al *intellectus* en sentido tomista, es decir, a la captación inmediata del ser y de los primeros principios, así como a la voluntad que orienta el conocimiento hacia la acción moral. En efecto, santo Tomás distingue el *intellectus*, que implica la íntima penetración de la verdad y la captación de los principios, de la *ratio*, que se refiere a la investigación y al discurso. La IA emula la función discursiva de la *ratio* pero carece del hábito de los primeros principios que funda todo

---

<sup>8</sup> Cf. B. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, y L. Floridi. "The Ethics of Algorithms: Mapping the Debate". *Big Data & Society* (2016): 1-21.

<sup>9</sup> Véase Luciano Floridi. *The Ethics of Information*. Oxford: Oxford University Press, 2013; Id. *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford: Oxford University Press, 2019; Luciano Floridi et al. "AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations". *Minds and Machines* 28, n.º 4 (2018): 689-707. Sobre los riesgos asociados al desarrollo de inteligencias artificiales avanzadas y los problemas de gobernanza tecnológica, cf. Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

conocimiento<sup>10</sup>. La máquina puede generar productos funcionalmente correctos o verosímiles, pero carece de la *veritas formaliter*, la adecuación consciente con la realidad que requiere intencionalidad. Al carecer también de la «sabiduría del corazón», que integra el conocimiento con el afecto y el límite, la IA permanece ontológicamente inhabilitada para ser sujeto de responsabilidad moral o titular de la prudencia. En este sentido, Francisco advierte que «la sabiduría del corazón» es la virtud que permite entrelazar el todo con las partes, integrando el conocimiento con el afecto, algo inaccesible para la utilidad algorítmica<sup>11</sup>.

Este artículo analiza el fenómeno desde una perspectiva sistemática, evitando tanto la tecnofobia como el reduccionismo normativo. Asimismo, se insiste en la conveniencia de potenciar el uso de la IA mediante una arquitectura de responsabilidad que permita al ser humano aprovechar esta «hipertrofia de la *ratio*» para alcanzar niveles de conocimiento y eficiencia inimaginables, siempre bajo el gobierno de la prudencia.

Metodológicamente, el estudio articula un análisis hermenéutico de la instrucción *Antiqua et Nova* con una aplicación sistemática de la teología moral tomista. Este enfoque coincide con la llamada de *Antiqua et Nova* a utilizar «antigua y nueva sabiduría» (Mt 13,52)<sup>12</sup>. A través del tratado de la prudencia (*STh* II-II, qq. 47-56), se contrastan las categorías clásicas de la razón práctica con la operatividad algorítmica. La aportación central del trabajo consiste en superar la descripción fenomenológica de los riesgos para ofrecer una fundamentación estructural basada en las virtudes y proponer que la solución ética reside en fortalecer la prudencia humana como instancia de control indelegable. De este modo, se busca clarificar qué procesos pueden apoyarse en la potencia del algoritmo (ámbito del *consilium*) y cuáles pertenecen de modo intransferible al juicio humano (*imperium*).

---

<sup>10</sup> Cf. Tomás de Aquino. *Summa Theologiae*, II-II, q. 49, a. 5, ad 3; *Antiqua et Nova*, n. 14. En adelante citaremos la Suma por su abreviatura común: *STh*.

<sup>11</sup> Cf. Francisco. *Mensaje para la 58.ª Jornada Mundial de las Comunicaciones Sociales* (2024); *Antiqua et Nova*, n. 114.

<sup>12</sup> Cf. J. P. Rivero González. “Del silicio al espíritu. Tomás de Aquino, la inteligencia artificial y la pregunta por lo verdadero”. *Almogaren* 76 (2025): 17-18, quien propone una lectura crítica desde tres ejes: la tradición tomista, el magisterio de *Antiqua et Nova* y el debate filosófico-jurídico contemporáneo.

## 2. IMAGO DEI FRENTE AL PARADIGMA TECNOCRÁTICO: LA IRREDUCTIBILIDAD DE LA CONDICIÓN HUMANA

El debate sobre la inteligencia artificial, tal como lo plantea la instrucción *Antiqua et Nova*, desborda el marco de la ingeniería para situarse en las coordenadas fundamentales de la antropología teológica. La comprensión cristiana del ser humano rechaza cualquier dualismo que conciba la inteligencia como un *software* independiente de su sustrato biológico; por el contrario, siguiendo la síntesis tomista del *anima forma corporis*, la inteligencia humana se afirma como una realidad constitutivamente encarnada. Esta condición establece una barrera ontológica insalvable para la máquina: el hombre no «tiene» un cuerpo, ya que «es» su cuerpo en unidad sustancial, lo que constituye un antídoto frente al dualismo digital que pretende reducir la conciencia a meros procesos computacionales<sup>13</sup>.

Esta naturaleza encarnada determina la génesis misma del conocimiento. El axioma aristotélico según el cual todo saber comienza en la experiencia sensible (*nihil est in intellectu quod non prius fuerit in sensu*) implica la necesidad de la «conversión a los fantasmas» o imágenes sensibles para la abstracción intelectual. La IA, carente de cuerpo y de inserción biológica, opera sobre representaciones matemáticas y unidades léxicas (*tokens*) que carecen de anclaje empírico directo<sup>14</sup>. Esta carencia de «mundo» determina la naturaleza de su operación cognitiva: mientras que los modelos de lenguaje funcionan bajo una lógica estrictamente sintáctica y probabilística, la inteligencia humana es semántica, afectiva y situacional. La máquina puede simular la verosimilitud, pero permanece ciega al significado profundo de lo que procesa.

La reducción funcionalista de la inteligencia, ejemplificada históricamente en el test de Turing<sup>15</sup>, tiende a equiparar el pensamiento con el

<sup>13</sup> Cf. *Antiqua et Nova*, n. 16. El documento cita el *Catecismo de la Iglesia Católica* (n. 365) para reafirmar que la unión de espíritu y materia constituye una única naturaleza. Véase también: Rivero González, 18-19, sobre la doctrina hilemórfica (*STh* I, q. 76, a. 1).

<sup>14</sup> Sobre el proceso de abstracción y la necesidad de la «conversión a los fantasmas» (*conversio ad phantasmata*), véase: Tomás de Aquino, *STh* I, q. 84, a. 7; y Rivero González, 19 y 34-35.

<sup>15</sup> El conocido test de Turing, propuesto por Alan Turing como criterio para evaluar si una máquina puede imitar el comportamiento lingüístico humano de manera

*output* o resultado externo, ignorando que el acto humano de entender es una experiencia consciente e intencional. Desde la teología moral, evaluar la dignidad o la inteligencia solo por la eficacia de las respuestas constituye un error categorial que compromete la singularidad de la persona. La imposición de esta lógica computacional sobre la vida social conlleva el riesgo de la «datificación», un fenómeno que amenaza la dignidad en sus dimensiones ontológica, moral y existencial al pretender reducir la biografía del sujeto a una métrica de fiabilidad estática basada en datos históricos<sup>16</sup>.

Esta deriva tecnocrática alcanza su punto más crítico en la negación de la libertad y la apertura al futuro. Los sistemas predictivos, al encasillar a la persona en perfiles inmutables, ignoran teológicamente la posibilidad de la conversión (*metanoia*) y de la gracia que rompe toda estadística. Delegar decisiones éticas en algoritmos representa una abdicación de la responsabilidad personal que transforma al sujeto moral en un mero objeto de gestión automatizada<sup>17</sup>.

Esta irreductibilidad antropológica se manifiesta con especial fuerza en la dimensión afectiva del conocimiento moral, un ámbito donde la mediación técnica encuentra un límite absoluto. La síntesis tomista evita el intelectualismo puro al afirmar que la prudencia, más que una operación de la razón, requiere la rectitud del apetito para inclinarse hacia el fin verdadero. Mientras que la inteligencia artificial opera bajo una lógica de objetividad, la tradición teológica sostiene que el juicio moral depende del conocimiento por connaturalidad (*cognitio per connaturalitatem*).

---

indistinguible, se convirtió en una referencia clásica en el debate sobre la inteligencia artificial. Véase A. Turing. "Computing Machinery and Intelligence". *Mind* 59 (1950): 433-460. *Antiqua et Nova*, n. 11, critica la perspectiva funcionalista del test de Turing, advirtiendo que evaluar la inteligencia solo por el *output* es un reduccionismo que ignora la experiencia consciente y la dimensión moral.

<sup>16</sup> La declaración *Dignitas Infinita* (2024), del Dicasterio para la Doctrina de la Fe, distingue cuatro dimensiones de la dignidad. Esta distinción es clave para diagnosticar los riesgos de la tecnología. Cf. S. Umbrello. "Human Dignity in the Digital Age". *Journal of Ethics and Emerging Technologies* 35 (2025): 1.

<sup>17</sup> *Antiqua et Nova* establece que la responsabilidad ética recae exclusivamente en los humanos. La pretensión de que un sistema predictivo sentencie el futuro de una persona niega la libertad. Cf. *Antiqua et Nova*, nn. 39, 44 y 55.

El juicio humano es, por tanto, una «verdad sentida» que integra la totalidad de la persona. La afectividad puede entenderse como una forma de sensibilidad hacia la verdad práctica que permite al sujeto reconocer el valor singular de la vida humana en cada situación concreta. La IA, al carecer de *pathos*, de deseos y de una historia afectiva encarnada, puede procesar valores como si fueran variables matemáticas, pero es estructuralmente incapaz de experimentar la atracción por el bien o el peso moral de la injusticia. Delegar el juicio en una máquina implica el riesgo de desterrar la compasión y la intuición del horizonte de la justicia, reduciendo el obrar humano a una lógica de optimización instrumental que ignora la «lógica del don» y la gratuidad, dimensiones esenciales de la *imago Dei*<sup>18</sup>.

El conocimiento por connaturalidad actúa como una salvaguarda contra la deshumanización técnica. Mientras que el algoritmo optimiza resultados basándose en la probabilidad del pasado, el corazón humano, movido por la caridad, es capaz de intuir el bien posible incluso allí donde la estadística predice el fracaso. Esta integración de la razón y el afecto es lo que permite que el juicio humano sea un acto de sabiduría que reconoce que la verdad del hombre solo se descubre plenamente en el amor y en la entrega personal, dimensiones que permanecen siempre fuera del alcance de la sintaxis computacional. Francisco recuerda que:

«No podemos esperar esta sabiduría de las máquinas. Aunque el término *inteligencia artificial* ha suplantado al más correcto utilizado en la literatura científica, *machine learning*, el uso mismo de la palabra “inteligencia” es engañoso. Sin duda, las máquinas poseen una capacidad inconmensurablemente mayor que los humanos para almacenar datos y correlacionarlos entre sí, pero corresponde al hombre, y sólo a él, descifrar su significado. No se trata, pues, de exigir que las máquinas parezcan humanas»<sup>19</sup>.

Finalmente, la distinción antropológica se manifiesta en la esfera de la vulnerabilidad y el cuidado. La inteligencia humana es una «sabiduría

---

<sup>18</sup> Sobre la importancia de la afectividad en la estructura del acto moral y su distinción de la frialdad operativa de los sistemas expertos, véase: Rivero González, 28-29. El autor subraya que la IA carece de la *vis cogitativa* que une la razón con la sensibilidad.

<sup>19</sup> Cf. Francisco. *Mensaje para la 58.ª Jornada Mundial de las Comunicaciones Sociales; y Antiqua et Nova*, n. 114.

del corazón», que integra el conocimiento con el afecto y la fragilidad compartida; solo quien es vulnerable y consciente de su finitud puede ejercer la *cum-passio* (padecer con el otro). La IA, por su naturaleza inmaterial e invulnerable, es estructuralmente incapaz de esta reciprocidad. Delegar el cuidado de los vulnerables en sistemas automatizados corre el riesgo de cosificar la relación y generar un aislamiento deshumanizador, pues la presencia consoladora y la interpelación del rostro del otro son facultades intransferibles de la condición humana que ningún código puede replicar.

### 3. EL NÚCLEO MORAL: LA INTELIGENCIA ARTIFICIAL BAJO EL EXAMEN DE LA VIRTUD DE LA PRUDENCIA

Para determinar si la inteligencia artificial puede integrarse en la acción del ser humano, es indispensable someterla al escrutinio de la prudencia (*prudentia*). La tradición tomista entiende esta virtud como la «regla recta de la acción» (*recta ratio agibilium*) que conecta el conocimiento universal con la situación singular. Su acto principal es el precepto o mandato, mediante el cual la razón ordena la aplicación del juicio a la obra<sup>20</sup>. El análisis de la estructura de la prudencia y debe tener en cuenta sus ocho partes integrales: memoria, inteligencia, docilidad, solercia, razón, providencia, circunspección y precaución. Solercia (*solertia*) designa en la tradición tomista una agudeza del entendimiento práctico que permite descubrir rápidamente lo que conviene hacer en situaciones complejas o imprevistas. Tomás de Aquino la describe como una prontitud del juicio para encontrar los medios adecuados cuando la deliberación completa no es posible<sup>21</sup>. Esta disección analítica del funcionamiento práctico de la prudencia permite evaluar con precisión qué dimensiones del acto moral pueden ser potenciadas por la tecnología y cuáles permanecen como una reserva antropológica exclusiva<sup>22</sup>.

<sup>20</sup> Cf. Tomás de Aquino, *STh* II-II, q. 47, a. 8. Véase: B. Kemple. "The Preeminent Necessity of Prudence". *Studia Gilsoniana* 6, n.º 4 (2017): 557.

<sup>21</sup> Cf. *STh* II-II, q. 49, a. 4.

<sup>22</sup> Sobre la división de las ocho partes integrales de la prudencia, véase: *STh* II-II, q. 49. Cf. J. F. Sellés. "La virtud de la prudencia según Tomás de Aquino". *Cuadernos de Anuario Filosófico* 90 (1999): 581-582.

### 3.1. EL CONSEJO (*CONSILIUM*) Y LA ILUSIÓN DE LA DELIBERACIÓN

El primer acto de la razón práctica es la indagación en los medios (*inquisitio*). En esta fase, la IA demuestra una potencia instrumental sin precedentes: procesa volúmenes masivos de datos para ofrecer correlaciones que escapan a la cognición humana. Sin embargo, es vital distinguir teológicamente entre la deliberación moral y la optimización técnica. La virtud que perfecciona este acto es la *eubulia* (buen consejo), la cual exige rectitud en la búsqueda y un deseo del bien. La IA opera bajo la lógica de la *eustochia* (conjetura veloz): ofrece resultados inmediatos basados en probabilidades estadísticas sin una verdadera indagación de los fines<sup>23</sup>. Aunque la máquina posee una «memoria» enciclopédica, esta es un archivo estático de datos, carente de la vivencia temporal necesaria para la verdadera deliberación.

Para profundizar en la fase deliberativa, es necesario recurrir a la *eubulia*, la parte potencial de la prudencia que perfecciona el acto de aconsejar bien (*STh* II-II, q. 51, a. 1). La IA representa, en términos tomistas, una *eustochia* técnica: una conjetura que alcanza el medio sin el laborioso proceso de la indagación racional. Mientras que la *eubulia* exige una *inquisitio* (indagación) pausada y atenta a las circunstancias singulares, la operatividad algorítmica implica una inmediatez estadística que anula la dimensión temporal de la deliberación.

La *eubulia* requiere que el consejo se ordene a un fin bueno y que los medios sean honestos. La IA puede optimizar rutas para fines iniciosos con la misma eficacia que para fines nobles, pues carece de la rectitud de la voluntad que define a la virtud. Al delegar el consejo en la máquina, el ser humano corre el riesgo de sustituir la deliberación moral por la optimización funcional. La verdadera potencia de la IA en esta fase debe ser ampliar el horizonte de los medios posibles para que la *eubulia* humana, enriquecida por el dato, pueda discernir cuál de ellos conduce verdaderamente a la vida virtuosa. La virtud humana se fortalece en la duda y la ponderación; la máquina, al no dudar, no puede perfeccionarse moralmente. Tomás de Aquino distingue la *eubulia* de la *eustochia* señalando que la segunda es una «felicidad en la conjetura» que se da

---

<sup>23</sup> La distinción entre la *eubulia* (buen consejo) y la *eustochia* (conjetura veloz) se encuentra en *STh* II-II, q. 51, a. 1, ad 3. Cf. Admethics, “Prudence in Thomas Aquinas” (2020).

sin el discurso de la razón, mientras que la virtud exige una indagación que sea «buena», esto es, orientada al fin último<sup>24</sup>.

### 3.2. EL JUICIO (*IUDICIUM*): LA NORMALIDAD ESTADÍSTICA

Una vez deliberado, el intelecto debe valorar el resultado. La *synesis* es el juicio recto sobre las cosas que suceden según las reglas comunes (*STh* II-II, q. 51, a. 3). Aquí la IA exhibe una competencia sobrehumana en la *synesis* (sensatez), entendida como la capacidad de juzgar según las reglas comunes. Al entrenarse detectando patrones recurrentes, los algoritmos son excepcionalmente hábiles diagnosticando situaciones estándar, pero con el riesgo de imponer patrones mayoritarios que ignoren la singularidad de la persona o generen formas de injusticia automatizada como el *social scoring*<sup>25</sup>.

El límite ontológico de la IA reside en su carencia estructural de *gnome* (perspicacia). La *gnome* importa un juicio recto sobre situaciones en las que conviene apartarse de las leyes comunes para salvaguardar la justicia superior (*STh* II-II, q. 51, a. 4). Santo Tomás explica que, ante situaciones excepcionales, es necesario juzgar «al margen de las reglas comunes» para evitar la injusticia. Esta destreza adaptativa al caso singular para mantener la justicia se denomina *epikeia*. La IA es ciega a la excepción: al depender exclusivamente de datos pasados, no puede reconocer la singularidad irrepitable. Mientras la *synesis* algorítmica normaliza, la *gnome* humana discierne la dignidad única que exige romper la norma estadística mediante la «sabiduría del corazón»<sup>26</sup>.

La incapacidad de la IA para manejar lo que técnicamente se denomina *out-of-distribution* (datos fuera de rango) es la manifestación empírica de su carencia de la virtud de la *gnome*. Fundamentar la responsabilidad significa, por tanto, potenciar el uso de la IA para gestionar lo estándar (función técnica), liberando al humano para que ejerza su capacidad insustituible de discernir la excepción y la misericordia (función moral)<sup>27</sup>.

<sup>24</sup> Cf. *STh* II-II, q. 51, a. 1, ad 3; y J. F. Sellés, “La virtud de la prudencia...”, 581.

<sup>25</sup> Sobre el riesgo de imponer patrones mayoritarios, véase *Antiqua et Nova*, nn. 93-94.

<sup>26</sup> Cf. *Antiqua et Nova*, n. 70.

<sup>27</sup> Sobre la *gnome* como la capacidad de recurrir a principios superiores cuando la ley común falla, véase: A. Freddoso. “Treatise on Prudence” (2021). Cf. *Antiqua et Nova*, n. 32.

### 3.3. EL IMPERIO (*IMPERIUM*): LA RESERVA DE LA DECISIÓN

La ruptura definitiva acontece en el mandato o *imperium*, considerado el motor principal de la prudencia, que aplica el conocimiento a la acción transformadora (*STh* II-II, q. 55, aa. 3-4). Este acto presupone un movimiento de la voluntad (*electio*) del que la máquina carece por completo. Tal como advierte la instrucción *Antiqua et Nova*, es imperativo diferenciar entre la «elección técnica» de la máquina, una selección lógica entre variables, y la «decisión humana»<sup>28</sup>. La IA carece de apetito racional; puede ejecutar procesos automatizados, pero no puede «impe- rar» moralmente porque no asume la responsabilidad del acto ni se ve afectada por sus consecuencias. Delegar el *imperium* en sistemas autó- nomos constituye una abdicación ética que deja el acto moral huérfano de sujeto responsable.

### 3.4. LA IA COMO «ASTUCIA» ARTIFICIAL

Si analizamos la IA desde los vicios opuestos a la prudencia, su operatividad tiende estructuralmente hacia la falsa prudencia. La velocidad de los modelos generativos favorece la precipitación, el salto a la conclusión sin deliberación real, y la inconsideración, al juzgar sin atender al contexto vital cualitativo del sujeto. La precipitación es un «descenso desordenado de la razón a la acción», mientras que la inconsideración es el fallo por falta de atención a las circunstancias (*STh* II-II, q. 53, aa. 3-4). Asimismo, la IA actual encarna con frecuencia la astucia (*astutia*): el uso de vías simuladas para alcanzar un fin (*STh* II-II, q. 55, aa. 3-4). Al simular empatía, razonamiento o autoría humana, la máquina incurre en un engaño fenomenológico que el magisterio califica como una violación ética de la veracidad comunicativa<sup>29</sup>.

---

<sup>28</sup> Cf. *Antiqua et Nova*, nn. 46 y 39.

<sup>29</sup> Cf. Francisco. *Mensaje para la 58.ª Jornada Mundial de las Comunicaciones Sociales*.

#### 4. LA RESPUESTA JURÍDICA: LA VALORACIÓN DE LA PRUDENCIA Y LA GARANTÍA DEL *IMPERIUM*

La constatación de que la inteligencia artificial carece de las estructuras ontológicas necesarias para ejercer la virtud de la prudencia obliga a desplazar el análisis hacia el ámbito del derecho positivo. Ante un agente técnico que posee autonomía operativa, pero carece de juicio ético, el ordenamiento jurídico instaura un sistema de «prudencia externa». El Reglamento (UE) 2024/1689 (Ley de IA) trasciende el establecimiento de estándares técnicos para articular un mecanismo de control extrínseco destinado a suplir el déficit de discernimiento de los sistemas automatizados, imponiendo la intervención humana como condición de legitimidad en los procesos críticos<sup>30</sup>.

##### 4.1. EL ENFOQUE BASADO EN EL RIESGO Y LA PROTECCIÓN DE LA DIGNIDAD HUMANA

El marco europeo estructura la regulación mediante un enfoque basado en el riesgo (*risk-based approach*) que establece una jerarquía de salvaguardas proporcional a la amenaza que el sistema supone para los derechos fundamentales. En la cúspide de esta pirámide, el Artículo 5 del Reglamento define categorías de «Riesgo Inaceptable» que funcionan como un veto moral. La prohibición del uso de sistemas para la evaluación social (*social scoring*) responde a una exigencia ética fundamental: impedir la reducción de la biografía personal a una métrica de fiabilidad estática, protegiendo la capacidad humana de cambio y redención<sup>31</sup>. Asimismo, para los sistemas de «Alto Riesgo», la normativa impone requisitos de gobernanza de datos (Art. 10) destinados a purificar la «memoria» del sistema de sesgos discriminatorios e históricos que atentan contra la dignidad social de la persona<sup>32</sup>.

---

<sup>30</sup> Sobre el enfoque basado en el riesgo y la estructura del Reglamento (UE) 2024/1689, véase: J. Laux, y H. Ruschemeier. “Automation Bias in the AI Act: On the Legal Implications of Attempting to De-Bias Human Oversight of AI”. *European Journal of Risk Regulation* 16, n.º 4 (2025).

<sup>31</sup> El Artículo 5 prohíbe el *social scoring*. La instrucción *Antiqua et Nova* condena estas prácticas en los nn. 93-94, por erosionar la confianza y la libertad. Cf. *Dignitas Infinita*.

<sup>32</sup> Sobre la gobernanza de datos para evitar sesgos discriminatorios, véase Art. 10 del Reglamento. Cf. B. C. Cheong. “Transparency and accountability in AI systems...”. *Frontiers in Human Dynamics* 6 (2024): 4.

#### 4.2. LA SUPERVISIÓN HUMANA EFECTIVA Y LA PARADOJA DE LA DECISIÓN

El núcleo de la tensión entre la eficacia técnica y la responsabilidad moral se encuentra en la regulación de la supervisión humana (*human oversight*). El Artículo 14 del Reglamento (UE) 2024/1689 constituye la traducción jurídica de la reserva del *imperium* prudencial, al exigir que los sistemas se diseñen de manera que las personas físicas retengan la autoridad real para interpretar los resultados y, crucialmente, la capacidad de intervenir o anular el funcionamiento del sistema<sup>33</sup>.

Sin embargo, la implementación efectiva de este mandato se enfrenta al «sesgo de automatización» (*automation bias*), definido en la norma (Art. 14.4.b) como la tendencia psicológica a confiar excesivamente en el resultado del sistema. Si el operador valida acríticamente la propuesta del algoritmo por temor a contradecir la estadística, se produce una abdicación de la responsabilidad. La teología moral advierte aquí que la mera presencia física de un humano en el bucle de decisión es insuficiente, si no va acompañada de un ejercicio activo del juicio crítico que valide la propuesta técnica<sup>34</sup>.

#### 4.3. LA CRISIS DE LA IMPUTABILIDAD Y EL RECHAZO DE LA PERSONALIDAD ELECTRÓNICA

La opacidad inherente a los sistemas de aprendizaje profundo, a menudo descritos como «cajas negras» (*black boxes*<sup>35</sup>), dificulta el establecimiento de nexos causales claros entre la operación de la máquina y el daño producido. Ante este desafío, tanto la doctrina jurídica como la antropología teológica rechazan la creación de una «personalidad electrónica» para los robots. Atribuir derechos a una entidad sin conciencia y capacidad de sufrimiento carece de fundamento ontológico y podría instrumentalizarse para diluir la responsabilidad de fabricantes y usuarios. Por el contrario, el marco europeo refuerza los regímenes de

---

<sup>33</sup> Cf. *Antiqua et Nova*, n. 46.

<sup>34</sup> El «sesgo de automatización» se define en el Art. 14.4(b) del Reglamento. Cf. Lauren Kahn et al. "AI Safety and Automation Bias", 5-6.

<sup>35</sup> Cf. F. Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press, 2015, esp. cap. 1-2.

responsabilidad objetiva y la presunción de que existe un posible defecto para asegurar que el riesgo técnico no derive en impunidad<sup>36</sup>.

#### 4.4. TRANSPARENCIA Y ALFABETIZACIÓN: CONDICIONES EPISTEMOLÓGICAS

Para que el ejercicio del juicio prudencial sea viable, el ordenamiento jurídico establece condiciones de transparencia (Art. 50) destinadas a preservar la veracidad en las interacciones digitales. La obligación de informar sobre el uso de *chatbots* y de etiquetar los contenidos generados artificialmente (*deepfakes*) busca evitar el engaño fenomenológico y la manipulación de la conciencia, y asegura que la distinción entre la persona y el artefacto permanezca nítida en la esfera pública<sup>37</sup>.

El Artículo 4 introduce la exigencia de alfabetización en IA (*AI Literacy*) para los proveedores y usuarios profesionales. La competencia técnica deja de ser una habilidad opcional y se convierte en un requisito de legitimidad operativa. Sin una comprensión adecuada de la lógica probabilística y las limitaciones del sistema, el operador carece de los elementos necesarios para ejercer un juicio responsable.

Desde la perspectiva de la Doctrina Social de la Iglesia, la exigencia de transparencia y el derecho a una explicación (Art. 86 del Reglamento) adquieren una dimensión de justicia distributiva. La opacidad de los sistemas de «caja negra», más que un déficit técnico, supone una amenaza para el bien común, ya que impide el ejercicio de la rendición de cuentas (*accountability*). Una autoridad que decide mediante algoritmos inescrutables se sitúa fuera de la interpelación del ciudadano, vulnera el principio de participación y el derecho a la verdad. La fundamentación de la responsabilidad exige que la transparencia técnica se convierta en una «transparencia antropológica»: el sujeto afectado tiene el derecho a comprender los criterios que rigen su vida, a fin de evitar que la tecnología se convierta en un instrumento de dominación burocrática anonimizada.

---

<sup>36</sup> Sobre el rechazo a la «personalidad electrónica», véase: J. Valls Prieto. «Sobre la responsabilidad penal por la utilización de sistemas inteligentes». *RECPC* 24-27 (2022): 19-20 y 43; *Antiqua et Nova*, nn. 39 y 111.

<sup>37</sup> El Artículo 50 obliga al etiquetado de *deepfakes*. Cf. Francisco. *Mensaje para la 58.ª Jornada Mundial de las Comunicaciones Sociales*. Cf. *Antiqua et Nova*, n. 85.

#### 4.5. EL BIEN COMÚN Y LA JUSTICIA ECOLÓGICA EN EL ENTORNO DIGITAL

La fundamentación de la responsabilidad ante la IA no puede obviar la dimensión material del desarrollo técnico. Existe una tendencia arraigada en el imaginario actual a concebir la tecnología digital como una entidad etérea o inmaterial (metáforas como «la nube» refuerzan esta percepción). Sin embargo, desde la perspectiva de la ecología integral propuesta en *Laudato si'*, debe remarcar que la IA posee un metabolismo sociotécnico de enorme impacto físico. Su despliegue exige una infraestructura masiva de centros de datos, sistemas de refrigeración que consumen millones de litros de agua dulce y una demanda energética que, en el caso del entrenamiento de grandes modelos de lenguaje (LLM), equivale al consumo eléctrico de cientos de hogares durante años.

Desde el examen de la prudencia, el impacto medioambiental es una circunstancia esencial del acto moral. Santo Tomás enseña que la prudencia exige la *circumspectio* (circunspección) o atención a las circunstancias que rodean la acción, y la *providentia* (providencia) o previsión de las consecuencias futuras (*STh* II-II, q. 49, aa. 6-7). Ignorar la huella de carbono o el consumo hídrico de un sistema de IA en nombre de la eficiencia administrativa constituiría una forma de «imprudencia ecológica». Por tanto, la responsabilidad teológica integra hoy la sostenibilidad como una condición de legitimidad técnica: no todo lo que es funcionalmente posible es moralmente aceptable, si compromete la integridad de la «casa común»<sup>38</sup>.

Esta responsabilidad adquiere tintes de justicia distributiva al analizar la cadena de suministro de la tecnología. El desarrollo de la IA descansa a menudo sobre nuevas formas de extractivismo digital en el Sur Global. La obtención de minerales críticos (como el litio o el cobalto) necesarios para el *hardware* de computación ocurre a menudo en condiciones de vulneración de derechos humanos y degradación ambiental. La instrucción *Antiqua et Nova* (n. 54) advierte que el paradigma tecnocrático tiende a invisibilizar estos costes sociales en favor

---

<sup>38</sup> Sobre la infraestructura material de la IA y su impacto hídrico y energético, véase: S. Umbrello. "Human Dignity in the Digital Age". *Journal of Ethics and Emerging Technologies* 35 (2025): 8-10. El autor destaca que la «invisibilidad» de la infraestructura favorece la desresponsabilización ética del usuario final.

de una optimización que beneficia solo a las regiones más desarrolladas. Una fundamentación ética de la responsabilidad debe actuar ante este «colonialismo digital», y exigir que la IA sea un factor de cohesión y no una herramienta que exacerbe la deuda ecológica entre naciones. El papa Francisco describe en *Laudato si'*, n. 106, cómo la racionalidad técnica asume que «el mercado por sí mismo garantiza el desarrollo humano integral y la inclusión social», una falacia que en el ámbito digital se traduce en la marginación de quienes proveen la materia prima del sistema<sup>39</sup>.

Asimismo, el Reglamento (UE) 2024/1689 (Ley de IA) refleja esta preocupación en sus consideraciones (específicamente los artículos 161 y 162), al imponer obligaciones de transparencia sobre el consumo energético y el impacto ambiental de los modelos de IA de uso general (GPAI) que presentan riesgos sistémicos. El marco europeo reconoce que la potencia de cómputo (medida en FLOPS) tiene una correlación directa con la huella medioambiental, por lo que sienta las bases para una normatividad técnica sostenible. Teológicamente, esta positivización de la ecología digital se alinea con el principio de solidaridad intergeneracional: el despliegue de la IA hoy no debe comprometer el acceso a los recursos naturales de las generaciones futuras.

El «bien común digital» exige que la mediación tecnológica se entienda como un instrumento para el desarrollo humano integral. Esto implica evolucionar desde una IA centrada únicamente en la productividad hacia una IA para el cuidado de la vida. La responsabilidad prudencial impulsa a diseñar sistemas que sean «inteligentes» en su capacidad de procesamiento, y también «sabios» en su respeto a los límites de la creación. Solo una tecnología que reconozca su dependencia de la fragilidad del ecosistema<sup>40</sup> podrá calificarse como una herramienta al servicio de la *imago Dei* y de la fraternidad universal.

---

<sup>39</sup> Cf. Francisco. *Discurso al G7 sobre Inteligencia Artificial*.

<sup>40</sup> *Antiqua et Nova*, n. 105, advierte contra la «idolatría del objeto», donde el ser humano termina adorando su propia capacidad técnica mientras descuida la creación que le ha sido confiada como «custodio» (*custos*). La responsabilidad ecológica es, por tanto, una dimensión de la fidelidad al mandato divino de cuidar el jardín del mundo.

## 5. PROPUESTA TEOLÓGICA: HACIA UNA FUNDAMENTACIÓN TEOLÓGICA DE LA RESPONSABILIDAD

La respuesta cristiana ante el avance de la IA debe articular una fundamentación de la responsabilidad que integre la potencia del cálculo algorítmico dentro de la estructura orgánica del acto humano, para alcanzar niveles de conocimiento y eficiencia, siempre bajo el gobierno de la prudencia.

### 5.1. LA JERARQUÍA DE LAS MEDIACIONES: EL *CONSILIUM* INSTRUMENTAL Y LA RESERVA DEL *IMPERIUM*

En continuidad con el análisis previo sobre la estructura de la prudencia, la IA puede desempeñar una función legítima en el ámbito del *consilium*, es decir, en la fase deliberativa en la que la razón práctica examina los medios disponibles para la acción. En este nivel, los sistemas algorítmicos pueden ampliar significativamente la capacidad humana para organizar y analizar información compleja. Sin embargo, esta mediación técnica no puede extenderse al ámbito del *imperium*, que constituye el acto propiamente decisorio de la prudencia y presupone la intervención de una voluntad personal responsable (*STh* II-II, q. 47, a. 8). La inteligencia artificial puede auxiliar en la *inquisitio* (indagación), pero el *imperium* constituye el acto propio y perfecto de la prudencia<sup>41</sup>.

Esta distinción tiene una raíz profundamente encarnacional. El axioma *caro cardo salutis* (la carne es el eje de la salvación) nos recuerda que la responsabilidad moral está ligada a la corporeidad y la vulnerabilidad. El sujeto humano responde de sus actos porque su propia existencia está en juego; la IA, al carecer de finitud y de «cuerpo moral», permanece en una impunidad ontológica que la inhabilita para ser sujeto de una decisión responsable. Por tanto, el *imperium* debe permanecer como una reserva antropológica exclusiva: el ser humano debe ser el autor comprometido de la acción.

---

<sup>41</sup> Cf. Brian Kemple. "The Preeminent Necessity of Prudence". *Studia Gilsoniana* 6 (2017): 571.

## 5.2. EL CONFLICTO ENTRE LA ESTADÍSTICA Y LA GRACIA: EL DERECHO A LA METANOIA

Un pilar fundamental de esta arquitectura es la defensa de la apertura existencial frente al determinismo de los datos. La IA funciona como un sistema de extrapolación de patrones del pasado que predice el futuro mediante regresiones estadísticas. Desde una perspectiva teológica, esta lógica plantea un conflicto con la doctrina de la *metanoia* (conversión). Si permitimos que el juicio sobre una persona sea dictado por un algoritmo, estamos negando su capacidad de cambio radical bajo la acción de la gracia.

El ser humano es *capax Dei*, es decir, constitutivamente abierto a una novedad que no puede reducirse a su pasado ni agotarse en ninguna proyección estadística derivada de sus datos. Desde esta perspectiva, el «derecho teológico a la *metanoia*» exige que la decisión humana ejerza siempre la *epikeia* (equidad), entendida en la tradición tomista como la capacidad prudencial de corregir la aplicación mecánica de una norma general cuando esta no hace justicia al caso singular. Reconocer esta posibilidad de conversión significa afirmar que la persona es una realidad trascendente que no puede ser plenamente capturada por ninguna métrica algorítmica. La intervención humana constituye el espacio donde la justicia permanece abierta a la esperanza y a la misericordia, dimensiones estructuralmente inaccesibles para el cálculo automático<sup>42</sup>.

## 5.3. LA ASCÉTICA DEL DISCERNIMIENTO

La integración de la IA plantea el riesgo del *moral deskilling* o descalificación moral<sup>43</sup>. La virtud es un hábito (*habitus*) que se adquiere por la repetición de actos (*STh* I-II, q. 51, a. 3), mientras que el *habitus* de la prudencia se fortalece con el ejercicio de la deliberación difícil. Así como el uso excesivo del GPS debilita la orientación espacial, la delegación sistemática del juicio en un sistema experto debilita el discernimiento. En

---

<sup>42</sup> La antropología cristiana defiende que el hombre es *capax Dei*. El determinismo algorítmico incurre en un «pelagianismo técnico». Cf. *Antiqua et Nova*, nn. 39 y 55.

<sup>43</sup> Sobre el *moral deskilling*, véase S. Vallor. *Technology and the Virtues*. Oxford: Oxford University Press, 2016.

este punto proponemos una «ascética digital»: el uso de la IA debe ser dialéctico, funcionar como un «adversario de datos» que obligue al sujeto a refinar sus propios argumentos, nunca como un sustituto que le exima de la fatiga de pensar y decidir.

#### 5.4. FENOMENOLOGÍA DE LA DECISIÓN: UN ESTUDIO DE CASO SOBRE LA JUSTICIA Y LA CARIDAD

Para ilustrar este modelo, analicemos la gestión de recursos en una institución de caridad. Un sistema de IA recomienda denegar ayuda a una familia basándose en un patrón de «alto riesgo» financiero derivado de datos históricos.

1. La visión técnica: el sistema estima una probabilidad de fracaso del 80%. Si el responsable acepta el dato por eficiencia, se produce una abdicación del imperio y una injusticia automatizada.
2. La visión prudencial: el responsable usa el dato como advertencia (*consilium*), pero ejerce la *gnome* (perspicacia) para mirar más allá. Mediante la presencia encarnada, descubre una red de apoyo comunitario o un deseo de cambio que el algoritmo no puede registrar. Al otorgar la ayuda, el humano ejerce una justicia superior que integra el riesgo y la esperanza. La IA «puede ver» la probabilidad, pero solo el ser humano «ve» al prójimo con caridad, ya que la *epikeia* permite apartarse de la letra de la norma para seguir el bien de la persona (*STh* II-II, q. 120, a. 1).

#### 5.5. SISTEMATIZACIÓN DE CRITERIOS

Como síntesis, se presentan dos propuestas destinadas al examen de la conciencia y la gobernanza operativa de los sistemas de IA. La primera (tabla 1) se centra en el proceso interno del sujeto, y analiza cómo la mediación técnica afecta a las virtudes potenciales de la prudencia (*eubulia*, *synesis*, *gnome*). Su objetivo es identificar los puntos donde la eficiencia técnica podría reducir la integridad del juicio moral<sup>44</sup>.

---

<sup>44</sup> Estas matrices sintetizan las partes potenciales de la prudencia (*eubulia*, *synesis*, *gnome*). Cf. *STh* II-II, q. 51.

TABLA 1. DISCERNIMIENTO PARA LA INTEGRACIÓN ÉTICA DE LA IA

DIMENSIÓN	FUNCIÓN DE LA IA	PREGUNTA CLAVE	RIESGO LATENTE	CRITERIO TEOLÓGICO
Verdad	Síntesis de datos masivos	¿Confundo la exactitud del dato con la verdad moral del bien?	Reduccionismo	Verdad informativa vs. práctica
Prudencia	Apoyo al <i>consilium</i>	¿He ponderado yo la excepción o he validado el cálculo por comodidad?	Apatía burocrática	Ejercicio de la <i>gnome</i> (perspicacia)
Autoridad	Generación de opciones	Si el resultado es lesivo, ¿puedo justificar por qué tomé yo la decisión?	Anonimato moral	Reserva del <i>imperium</i> humano
Formación	Mediación cognitiva	¿El uso de esta IA me hace más capaz de decidir o me atrofia el juicio?	Atrofia del hábito	Ascética del discernimiento
Justicia	Optimización de recursos	¿Esta decisión respeta la dignidad y el derecho a la conversión?	Determinismo	Apertura a la <i>metanoia</i> y la gracia

En la segunda tabla (tabla 2) se propone una distribución de competencias ontológicas y se establecen las líneas rojas que la técnica no debe cruzar para evitar la deshumanización de los procesos de decisión.

TABLA 2. FUNDAMENTACIÓN FUNCIONAL DE LA RESPONSABILIDAD ANTE LA MEDIACIÓN TÉCNICA

NIVEL DE REALIDAD	AGENTE PRINCIPAL	OPERACIÓN LEGÍTIMA	LÍMITE CRÍTICO (LÍNEA ROJA)	FINALIDAD DE LA MEDIACIÓN
Informativo	IA (algoritmo)	Clasificación y predicción estadística	No puede asignar valores o sentidos vitales	Servicio a la verdad informativa
Deliberativo	Humano + IA	Diálogo entre dato y experiencia vital	El humano no debe renunciar a la duda crítica	Servicio al discernimiento
Decisorio	Humano (sujeto)	Juicio práctico y mandato ( <i>imperium</i> )	El sistema nunca debe decidir «por» el humano	Servicio al bien moral
Existencial	Humano (persona)	Asunción de responsabilidad y cuidado	La máquina no puede sustituir la presencia real	Servicio a la dignidad y a la caridad

### 5.6. LA PRESENCIA CRÍTICA FRENTE AL BUCLE ALGORÍTMICO

El ordenamiento jurídico europeo, a través del concepto de «supervisión humana» (*human oversight*), busca garantizar que los sistemas de IA no operen en un vacío de control; sin embargo, desde la teología moral, este mandato legal corre el riesgo de degradarse a un formalismo administrativo si no se fundamenta en una antropología de la presencia crítica. La verdadera responsabilidad implica algo más que la simple presencia humana en el proceso de decisión (*human-in-the-loop*); exige el ejercicio de un juicio prudencial que permita al sujeto situarse críticamente frente al algoritmo.

En la era de la mediación técnica, la presencia crítica se define como la capacidad del sujeto para romper la inercia del bucle algorítmico. El fenómeno del «sesgo de automatización» revela un problema de la razón práctica: la tendencia a considerar la propuesta de la máquina como una verdad neutra y superior, lo que conduce a una reducción del juicio personal. Frente a esto, la arquitectura de la responsabilidad exige que

el operador humano mantenga una alteridad dialéctica con la herramienta. La IA ofrece la potencia de la *ratio*, la velocidad del cálculo y la detección de patrones, pero solo el humano aporta la luz del *intellectus* y la rectitud de la voluntad. La supervisión, por tanto, va más allá de un trámite de cumplimiento (*compliance*), implica un acto de discernimiento donde el humano debe ser capaz de decir «no» a la recomendación de la máquina en nombre de una justicia que la sintaxis algorítmica es incapaz de comprender.

Este acto de interrupción o «negativa prudencial» es, en realidad, el momento de mayor densidad moral de la acción asistida por tecnología. En esta fase se manifiesta la virtud de la *gnome* o perspicacia, que permite al sujeto reconocer cuándo el caso desborda la norma estadística. Decir «no» al algoritmo es un ejercicio de *epikeia*: la fidelidad al espíritu de la justicia que la letra del código no puede capturar. En este sentido, la fundamentación de la responsabilidad cambia la «eficiencia», valor supremo del paradigma tecnocrático, por la «rectitud». La IA optimiza el proceso, pero solo la presencia crítica humana garantiza que el proceso no se convierta en una estructura de descarte anonimizada. El humano es, en última instancia, el garante del misterio de la persona, que puede garantizar que el sujeto sea tratado como un prójimo interpelante. Francisco enfatiza que el ser humano es el único que puede otorgar «sentido» al mundo, una facultad que ninguna máquina puede simular. La presencia crítica es la defensa de este sentido frente a la pura funcionalidad<sup>45</sup>.

Bajo el gobierno de la prudencia, la IA experimenta una transformación en su estatuto ontológico-funcional: deja de ser una amenaza de sustitución y se convierte en un don de la inteligencia humana. La teología cristiana, que reconoce en la técnica una participación en la actividad creadora de Dios (*participatio in creatio*), valora la IA como una mediación que permite al hombre actuar con una lucidez mayor. La IA ofrece la «verdad de los datos» para que el ser humano pueda ejercer, con mayor libertad y precisión, la «verdad del bien». Se produce así una armonía de la acción donde la máquina aporta la fuerza de la correlación y el ser humano responde con el sentido y el fin común. *Antiqua et Nova*, n. 46, insiste en que el uso de la IA debe estar orientado a «servir al pleno desarrollo del hombre», lo que exige que el diseño de los sistemas favorezca

---

<sup>45</sup> Cf. Francisco. *Mensaje para la 58.ª Jornada Mundial de las Comunicaciones Sociales*.

y no anule la capacidad de juicio del usuario. La IA como «don» es una categoría que conecta la técnica con la gracia y el servicio<sup>46</sup>.

La tecnología, por muy sofisticada que sea su mimesis del pensamiento, permanece siempre en el orden del *instrumentum*. La responsabilidad personal, con el peso de sus riesgos e incertidumbres, constituye una prerrogativa intransferible de la *imago Dei*. La propuesta de estas páginas invita, por tanto, a un optimismo esperanzado pero vigilante: la IA es la herramienta que puede ayudarnos a construir una sociedad más eficiente, siempre que el ser humano mantenga la lucidez necesaria para ser su arquitecto moral. El futuro de la inteligencia reside en el espíritu que sabe usar la técnica como un medio para amar más y mejor la realidad, sin renunciar jamás al deber de responder por cada una de sus decisiones ante la historia y ante el Creador. La responsabilidad es, según Berdiaev, citado en *Antiqua et Nova* (n. 111), la marca de la verdadera nobleza humana. Transferir la responsabilidad a un ente impersonal es una forma de degradación ontológica. La fundamentación teológica propuesta busca, precisamente, evitar esta caída mediante el fortalecimiento del hábito de la virtud.

## 6. CONCLUSIONES

El desafío de la IA radica en su potencia técnica y en la grandeza de la oportunidad que ofrece para redefinir las acciones del ser humano en el cambio de época actual. Lejos de las visiones que enfrentan dialécticamente al hombre con la máquina, este estudio concluye que la técnica y la virtud pueden articularse en una síntesis fecunda que potencie la dignidad de la persona. De este análisis se desprenden cuatro conclusiones:

### 1. La IA como aliada de la deliberación y el conocimiento.

La IA constituye una de las mayores conquistas de la *ratio* humana. Su capacidad para organizar la «verdad informativa» es un don que expande los límites de la deliberación. Al situar a la IA en el ámbito del *consilium* (consejo), la teología moral reconoce en la técnica una mediación providencial que libera al ser humano de la carga del procesamiento masivo y le permite centrar su energía en lo más noble: el discernimiento

---

<sup>46</sup> Cf. *Antiqua et Nova*, n. 115.

del bien. La IA no sustituye al humano; lo informa con una precisión sin precedentes para que su juicio sea más lúcido y profundo.

2. La responsabilidad humana como condición de éxito tecnológico.

La reserva del *imperium* (mandato) al sujeto humano es la condición de la legitimidad y éxito de la IA a largo plazo. Una tecnología que opere sin un sujeto responsable detrás derivaría en una burocracia anónima e ineficaz. Al proponer que el humano mantenga el control final de la decisión, esta arquitectura de responsabilidad dota a la potencia algorítmica de un «corazón moral». La sinergia resultante es una autoridad aumentada: una técnica que ofrece los mejores medios informativos y un humano que les otorga un sentido ético y una finalidad trascendente.

3. La resistencia de la gracia frente al determinismo algorítmico.

El estudio afirma la necesidad de defender el «derecho a la *metanoia*» como un imperativo ético frente al dato estadístico. La IA, al ser una tecnología orientada al pasado, tiende a clausurar el futuro del ser humano; sin embargo, la teología moral aporta una perspectiva clara: ningún dato puede agotar la identidad de un ser capaz de ser transformado por la gracia. La responsabilidad en la era digital implica el deber de ejercer la *epikeia* para romper la estadística en nombre de la esperanza. Integrada en este horizonte, la IA se convierte en la herramienta definitiva para el diagnóstico social.

4. Hacia una «Inteligencia Aumentada» por la virtud.

El futuro del desarrollo tecnológico es, en esencia, un camino hacia una humanidad más consciente y responsable. La IA, gobernada por la prudencia, obliga al ser humano a elevar su propio nivel de discernimiento mediante una «ascética digital». El progreso consiste en fomentar una presencia crítica donde la IA aporte la eficiencia y el humano adquiera la capacidad de orientar el desarrollo técnico hacia el bien común mediante un ejercicio responsable de la prudencia.

En última instancia, la fundamentación de la responsabilidad aquí propuesta recuerda que la técnica es una expresión del mandato de cuidar y perfeccionar la creación. La IA, integrada en el orden de la virtud, implica un nuevo y más ambicioso comienzo. Es la herramienta que, bien gobernada, permitirá al ser humano actuar con una lucidez mayor, para transformar la potencia del cálculo en un servicio genuino a la dignidad de la persona y del Creador.

## REFERENCIAS

- Admethics. “Prudence in Thomas Aquinas”. (2020).
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major y Margaret Mitchell. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”. En *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. New York: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.
- Cheong, Ben Chester. “Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making”. *Frontiers in Human Dynamics* 6 (2024): 1-13.
- Cortina, Adela. *¿Ética o ideología de la inteligencia artificial? El eclipse de la razón comunicativa en una sociedad tecnolozada*. Barcelona: Paidós, 2024.
- Dicasterio para la Doctrina de la Fe. *Declaración “Dignitas Infinita” sobre la dignidad humana*. Ciudad del Vaticano: Libreria Editrice Vaticana, 2024.
- Dicasterio para la Doctrina de la Fe y Dicasterio para la Cultura y la Educación. *Antiqua et Nova: Nota sobre la relación entre la inteligencia artificial y la inteligencia humana*. Ciudad del Vaticano, 2025.
- Floridi, Luciano. *The Ethics of Information*. Oxford: Oxford University Press, 2013.
- Floridi, Luciano. *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford: Oxford University Press, 2019. <https://doi.org/10.1093/oso/9780198833635.001.0001>
- Floridi, Luciano, Josh Cows, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke y Effy Vayena. “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations”. *Minds and Machines* 28, n.º 4 (2018): 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Francisco. *Carta encíclica Laudato si’ sobre el cuidado de la casa común*. Ciudad del Vaticano: Libreria Editrice Vaticana, 2015.

- Francisco. *Carta encíclica Fratelli tutti sobre la fraternidad y la amistad social*. Ciudad del Vaticano, 2020.
- Francisco. “Discurso del Santo Padre Francisco en la sesión de trabajo del G7 sobre Inteligencia Artificial”. Borgo Egnazia, 14 de junio de 2024.
- Francisco. “Inteligencia artificial y sabiduría del corazón: para una comunicación plenamente humana”. Mensaje para la 58.ª Jornada Mundial de las Comunicaciones Sociales. Ciudad del Vaticano, 24 de enero de 2024.
- Freddoso, A. “Treatise on Prudence” (2021).
- Kahn, Lauren, Emelia Probasco y Ronnie Kinoshita. *AI Safety and Automation Bias. The Downside of Human-in-the-Loop*. (Issue Brief. CSET Center for Security and Emerging Technology). Washington, DC: Georgetown University, 2024.
- Kemple, Brian. “The Preeminent Necessity of Prudence”. *Studia Gilsoniana* 6, n.º 4 (2017): 549-572.
- Laux, Johann, y Hannah Ruschemeier. “Automation Bias in the AI Act: On the Legal Implications of Attempting to De-Bias Human Oversight of AI”. *European Journal of Risk Regulation* 16, n.º 4 (2025): 1519-1534. <https://doi.org/10.1017/err.2025.10033>
- Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter y Luciano Floridi. “The Ethics of Algorithms: Mapping the Debate”. *Big Data & Society* 3, n.º 2 (2016): 1-21. <https://doi.org/10.1177/2053951716679679>
- Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge-Londres: Harvard University Press, 2015. <https://doi.org/10.4159/harvard.9780674736061>
- Rivero González, Juan Pedro. “Del silicio al espíritu. Tomás de Aquino, la inteligencia artificial y la pregunta por lo verdadero”. *Almogaren* 76 (2025): 15-37.
- Sellés, Juan Fernando. “La virtud de la prudencia según Tomás de Aquino”. *Cuadernos de Anuario Filosófico* 90 (1999): 541-582.
- Tomás de Aquino. *Summa Theologiae*. Madrid: Biblioteca de Autores Cristianos, 2001.
- Turing, Alan. “Computing Machinery and Intelligence”. *Mind* 59, n.º 236 (1950): 433-460. <https://doi.org/10.1093/mind/LIX.236.433>

- Umbrello, Steven, y Paul O'Hara. "Human Dignity in the Digital Age". *Journal of Ethics and Emerging Technologies* 35, n.º 1 (2025): 1-22. <https://doi.org/10.55613/jeet.v35i1.165>
- Unión Europea. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial). *Diario Oficial de la Unión Europea* L, 12 de julio de 2024.
- Vallor, Shannon. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press, 2016. <https://doi.org/10.1093/acprof:oso/9780190498511.001.0001>
- Valls Prieto, Javier. "Sobre la responsabilidad penal por la utilización de sistemas inteligentes". *Revista Electrónica de Ciencia Penal y Criminología* 24 (2022): 1-35.