

La arquitectura conceptual de la inteligencia artificial: de la computación mecánica a la predicción estadística

Paulo Vélez León

§1. Génesis y origen histórico de la inteligencia artificial

LA INTELIGENCIA ARTIFICIAL COMO CAMPO DE INVESTIGACIÓN TIENE UN ORIGEN AMPLIO QUE combina campos de conocimiento como la lógica, la matemática, la teoría de la computación o la filosofía de la mente. Mucho antes de que el término “inteligencia artificial” fuera acuñado en el siglo XX, el proyecto de mecanizar procedimientos de cálculo y, por extensión, ciertas formas de razonamiento, ya había tomado forma en el siglo XIX con el diseño de máquinas de propósito general. En este contexto, los trabajos de Charles Babbage sobre la *Difference Engine* y, sobre todo, sobre la *Analytical Engine* resultan decisivos porque plantean una arquitectura computacional programable que separa (i) la representación de operaciones y (ii) la ejecución mecánica de dichas operaciones, anticipando la idea de computación general¹.

El trabajo de Turing sobre máquinas abstractas estableció las bases formales para concebir el cálculo como un proceso mecánico generalizable². Esta concepción permitió articular, por primera vez, la hipótesis de que procesos tradicionalmente considerados “mentales” podían, en principio, ser implementados mediante procedimientos computacionales. El célebre artículo de Turing sobre la posibilidad de

¹ Charles Babbage, *On the Economy of Machinery and Manufactures* (London: Charles Knight, 1832).

² B Jack Copeland, ed., *The Essential Turing: The Ideas That Gave Birth to the Computer Age* (Oxford: Oxford University Press, 2004).

P. Vélez León (✉)
Universidad Técnica Particular de Loja, Ecuador
e-mail: pevelez@utpl.edu.ec

Disputatio. Philosophical Research Bulletin
Vol. 14, No. 29, Dec. 2025, pp. 185–193
ISSN: 2254–0601 | [ES] | **ARTÍCULO**

que las máquinas “piensen” no solo introdujo un criterio conductual de evaluación (el denominado *imitation game*), sino que también inauguró una tradición metodológica orientada a evaluar la inteligencia artificial en términos de desempeño observable más que de estructura interna³.

El acto fundacional del campo suele situarse en la propuesta para el *Dartmouth Summer Research Project on Artificial Intelligence*, redactada en 1955 por McCarthy, Minsky, Rochester y Shannon, en la que se plantea explícitamente la hipótesis de que “cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, describirse de manera tan precisa que una máquina pueda ser construida para simularlo”⁴. Esta formulación no solo da nombre al campo, sino que expresa un programa fuerte de investigación: la reducción de la inteligencia a procesos formalizables y computables.

Durante las décadas siguientes, este programa se desarrolló principalmente bajo enfoques simbólicos, basados en la manipulación explícita de representaciones formales. La denominada “IA clásica” concibió el razonamiento como búsqueda en espacios de estados y aplicación de reglas lógicas, bajo el supuesto de que la inteligencia podía ser modelada como un sistema formal de inferencia (Newell y Simon 1976). Esta concepción, profundamente influida por la lógica matemática y la psicología cognitiva temprana, se refleja en trabajos como los compilados por Minsky (1968), donde el procesamiento simbólico aparece como el núcleo del proyecto de la IA.

Sin embargo, las limitaciones empíricas de estos enfoques —particularmente frente a problemas de percepción, lenguaje natural y aprendizaje a gran escala— condujeron, desde finales del siglo XX, a un desplazamiento progresivo hacia métodos estadísticos y conexionistas. El desarrollo del aprendizaje automático (*machine learning*) marcó un punto de inflexión, al priorizar la inducción de regularidades a partir de datos sobre la codificación explícita de reglas⁵. Este giro metodológico se consolidó con el auge del aprendizaje profundo, que introdujo arquitecturas de redes neuronales profundas capaces de modelar relaciones complejas entre grandes volúmenes de datos⁶.

³ Alan M Turing, “Computing Machinery and Intelligence,” *Mind* 59, no. 236 (1950): 433–60, <https://doi.org/10.1093/mind/LIX.236.433>.

⁴ John McCarthy et al., “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” 1955.

⁵ Tom M Mitchell, *Machine Learning* (New York: McGraw-Hill, 1997).

⁶ Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning* (Cambridge, MA: MIT Press, 2016).

Esta genealogía histórica es conceptualmente relevante para el análisis constitucional, pues muestra que la inteligencia artificial contemporánea no es la continuación lineal del proyecto simbólico original, sino el resultado de un cambio profundo en el modo de concebir el razonamiento: de un ideal de inferencia explícita y trazable a un paradigma de predicción estadística y optimización matemática. Esta transformación tiene consecuencias directas para la manera en que debe evaluarse, desde el derecho constitucional, la pretensión de utilizar sistemas de IA en funciones que tradicionalmente se asocian con el ejercicio de la razón práctica institucionalizada

§ 2. Delimitación preliminar del concepto de inteligencia artificial

El uso de inteligencia artificial en diversos marcos, como el razonamiento jurídico y el análisis constitucional exige, como punto de partida, una delimitación conceptual precisa del objeto de estudio. El término “inteligencia artificial” se utiliza habitualmente para designar un conjunto heterogéneo de técnicas computacionales orientadas a ejecutar tareas que, si fueran realizadas por seres humanos, se asociarían con capacidades cognitivas como el aprendizaje, la clasificación o la resolución de problemas⁷. Esta definición funcional, aunque extendida en la literatura técnica, resulta insuficiente para el análisis jurídico si no se especifican los tipos de sistemas involucrados y los modos de inferencia que los caracterizan.

Desde una perspectiva analítica, la inteligencia artificial no constituye una entidad unitaria, sino un campo compuesto por enfoques metodológicos diversos, con supuestos cognitivos y consecuencias normativas diferenciables⁸. La falta de distinción entre estos enfoques ha generado confusiones conceptuales que afectan directamente a los debates constitucionales sobre automatización decisoria, responsabilidad y control del poder público.

§ 3. IA simbólica e IA estadística

Históricamente, la investigación en inteligencia artificial se desarrolló inicialmente bajo paradigmas simbólicos, basados en la representación explícita del conocimiento mediante reglas, ontologías y sistemas lógicos formales⁹. En estos modelos, el

⁷ Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (Hoboken, NJ: Pearson, 2021).

⁸ Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know*, 2nd ed. (New York: Oxford University Press, 2019).

⁹ Allen Newell and Herbert A. Simon, “Computer Science as Empirical Inquiry: Symbols and Search,” *Communications of the ACM* 19, no. 3 (1976): 113–26.

razonamiento se concebía como manipulación de símbolos conforme a reglas explícitas, lo que permitía una trazabilidad relativamente clara de los procesos inferenciales¹⁰.

A partir de finales del siglo XX, este enfoque fue progresivamente desplazado por métodos estadísticos y conexionistas, centrados en el aprendizaje automático a partir de grandes volúmenes de datos¹¹. El aprendizaje profundo (*deep learning*) constituye la expresión más desarrollada de esta transición, al emplear redes neuronales artificiales de múltiples capas para modelar relaciones complejas entre datos de entrada y salida¹².

Esta distinción no es meramente técnica. Mientras que los sistemas simbólicos aspiran a reproducir estructuras lógicas explícitas, los sistemas estadísticos operan mediante la optimización de funciones matemáticas orientadas a maximizar la precisión predictiva. Desde una perspectiva conceptual, ello implica que los segundos no “razonan” en sentido normativo, sino que infieren correlaciones probabilísticas sin referencia interna a significado, verdad o justificación¹³.

§ 4. Modelos de lenguaje de gran escala como paradigma dominante

Los modelos de lenguaje de gran escala (*large language models*, LLM) representan el estado actual del arte en inteligencia artificial aplicada al lenguaje natural. Estos sistemas se entrenan para estimar distribuciones de probabilidad sobre secuencias lingüísticas, de modo que, ante una entrada dada, generan la continuación estadísticamente más plausible en función de los datos de entrenamiento¹⁴.

Desde el punto de vista conceptual, los LLM no operan mediante reglas semánticas ni poseen representaciones intencionales del mundo. Su funcionamiento se basa en la predicción de tokens a partir de patrones estadísticos, sin acceso a criterios internos de verdad o falsedad¹⁵. Esta característica explica tanto su notable capacidad para

¹⁰ Marvin Minsky, ed., *Semantic Information Processing* (Cambridge, MA: MIT Press, 1968).

¹¹ Mitchell, *Machine Learning*.

¹² Goodfellow, Bengio, and Courville, *Deep Learning*; Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep Learning,” *Nature* 521 (2015): 436–44.

¹³ Christopher M Bishop, *Pattern Recognition and Machine Learning* (New York: Springer, 2006).

¹⁴ Tom B Brown et al., “Language Models Are Few-Shot Learners,” *Advances in Neural Information Processing Systems* 33 (2020); Ashish Vaswani et al., “Attention Is All You Need,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

¹⁵ Emily M Bender et al., “On the Dangers of Stochastic Parrots,” *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 2021.

producir textos coherentes como su tendencia estructural a generar errores plausibles (*hallucinations*) sin mecanismos internos de corrección epistémica.

La literatura coincide en que el éxito empírico de estos modelos no implica comprensión en sentido fuerte, sino una forma altamente sofisticada de correlación estadística¹⁶. Esta constatación resulta central para cualquier análisis jurídico que pretenda evaluar su idoneidad para funciones decisorias.

§ 5. Comprensión, semántica e intencionalidad

La distinción entre manipulación sintáctica y comprensión semántica ha sido ampliamente desarrollada en la filosofía de la mente. El argumento de la “habitación china” formulado por Searle ilustra de manera paradigmática que la ejecución correcta de reglas formales no implica, por sí misma, comprensión del significado¹⁷. Este argumento conserva relevancia en el contexto contemporáneo, pues los modelos de lenguaje actuales, pese a su complejidad, siguen operando en el plano sintáctico.

Desde una perspectiva intencional, la comprensión supone la capacidad de que los estados mentales se refieran a algo como algo, esto es, que posean contenido semántico dirigido¹⁸. Los sistemas de inteligencia artificial carecen de esta propiedad: no mantienen creencias, no evalúan razones ni ajustan sus inferencias a estándares normativos de corrección¹⁹.

Estas limitaciones no constituyen deficiencias contingentes susceptibles de ser superadas mediante incrementos de capacidad computacional, sino rasgos estructurales derivados del tipo de inferencia que estos sistemas realizan. En consecuencia, la atribución de “razonamiento” a la inteligencia artificial debe entenderse, en el mejor de los casos, en un sentido metafórico o instrumental²⁰.

§ 6. Opacidad algorítmica y explicabilidad

Otro rasgo estructural de los modelos contemporáneos de inteligencia artificial es su opacidad. Los procesos internos mediante los cuales un modelo de aprendizaje profundo produce una salida determinada no son directamente accesibles ni

¹⁶ LeCun, Bengio, and Hinton, “Deep Learning.”

¹⁷ John R Searle, “Minds, Brains, and Programs,” *Behavioral and Brain Sciences* 3, no. 3 (1980): 417–57.

¹⁸ Daniel C Dennett, *The Intentional Stance* (Cambridge, MA: MIT Press, 1987).

¹⁹ David J Chalmers, *The Conscious Mind* (Oxford: Oxford University Press, 1996).

²⁰ Kaplan, *Artificial Intelligence: What Everyone Needs to Know*.

reconstruibles en términos comprensibles para observadores humanos²¹. Esta opacidad no se reduce a la falta de interfaces adecuadas, sino que deriva de la complejidad matemática y distribuida de los modelos.

La investigación en inteligencia artificial explicable (*explainable AI*) ha propuesto diversas técnicas para mitigar este problema, pero la literatura coincide en que dichas técnicas ofrecen, en el mejor de los casos, aproximaciones parciales y contextuales a la comprensión del funcionamiento del modelo²². Desde una perspectiva conceptual, ello implica que la opacidad no puede eliminarse sin alterar radicalmente el tipo de sistemas empleados.

§ 7. Uso de la IA en el análisis constitucional

Las características descritas —predicción estadística, ausencia de comprensión semántica, falta de intencionalidad y opacidad estructural— constituyen el marco conceptual mínimo para evaluar el uso de inteligencia artificial en contextos jurídicos. Equiparar estos sistemas con agentes racionales o con sujetos capaces de razonar jurídicamente implica un error categorial que distorsiona el análisis normativo.

Desde una perspectiva constitucional, esta delimitación conceptual no pretende negar la utilidad instrumental de la inteligencia artificial, sino establecer las condiciones bajo las cuales su uso puede ser evaluado sin comprometer exigencias básicas de racionalidad, responsabilidad y control del poder decisorio. Un marco conceptual así, constituye el presupuesto analítico necesario para cualquier discusión posterior sobre inteligencia artificial y razonamiento jurídico.

²¹ Jenna Burrell, “How the Machine Thinks,” *Big Data & Society* 3, no. 1 (2016).

²² Finale Doshi-Velez and Been Kim, “Towards a Rigorous Science of Interpretable Machine Learning,” *ArXiv Preprint ArXiv:1702.08608*, 2017; Andreas Holzinger and others, “From XAI to Trustworthy AI,” in *Explainable AI* (Springer, 2022), 3–18.

REFERENCIAS

- Babbage, Charles. *On the Economy of Machinery and Manufactures*. London: Charles Knight, 1832.
- Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots." *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 2021.
- Bishop, Christopher M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- Brown, Tom B, Benjamin Mann, Nick Ryder, and others. "Language Models Are Few-Shot Learners." *Advances in Neural Information Processing Systems* 33 (2020).
- Burrell, Jenna. "How the Machine Thinks." *Big Data & Society* 3, no. 1 (2016).
- Chalmers, David J. *The Conscious Mind*. Oxford: Oxford University Press, 1996.
- Copeland, B Jack, ed. *The Essential Turing: The Ideas That Gave Birth to the Computer Age*. Oxford: Oxford University Press, 2004.
- Dennett, Daniel C. *The Intentional Stance*. Cambridge, MA: MIT Press, 1987.
- Doshi-Velez, Finale, and Been Kim. "Towards a Rigorous Science of Interpretable Machine Learning." *ArXiv Preprint ArXiv:1702.08608*, 2017.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. Cambridge, MA: MIT Press, 2016.
- Holzinger, Andreas, and others. "From XAI to Trustworthy AI." In *Explainable AI*, 3–18. Springer, 2022.
- Kaplan, Jerry. *Artificial Intelligence: What Everyone Needs to Know*. 2nd ed. New York: Oxford University Press, 2019.
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning." *Nature* 521 (2015): 436–44.
- McCarthy, John, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," 1955.
- Minsky, Marvin, ed. *Semantic Information Processing*. Cambridge, MA: MIT Press, 1968.
- Mitchell, Tom M. *Machine Learning*. New York: McGraw-Hill, 1997.
- Newell, Allen, and Herbert A Simon. "Computer Science as Empirical Inquiry: Symbols and Search." *Communications of the ACM* 19, no. 3 (1976): 113–26.
- Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed. Hoboken, NJ: Pearson, 2021.
- Searle, John R. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, no. 3 (1980): 417–57.
- Turing, Alan M. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433–60. <https://doi.org/10.1093/mind/LIX.236.433>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, and others. "Attention Is All You Need." In *Advances in Neural Information Processing Systems*, Vol. 30, 2017.



The Conceptual Architecture of Artificial Intelligence: From Mechanical Computation to Statistical Prediction

This article reconstructs the conceptual architecture of artificial intelligence by tracing its intellectual genealogy and the methodological shift that leads from early projects of mechanical and programmable computation to the contemporary paradigm of statistical prediction. The analysis begins with nineteenth-century antecedents associated with the Analytical Engine and Ada Lovelace's "Notes" and then situates the mid-twentieth-century foundational turn marked by the theory of computation and the research programme inaugurated at Dartmouth. On this historical basis, the article draws a structural distinction between symbolic artificial intelligence—grounded in explicit rules and inferential traceability—and statistical artificial intelligence, whose performance relies on the optimisation of mathematical functions over large volumes of data. The central section characterises large language models as systems of probabilistic inference over linguistic sequences and identifies three conceptually decisive features: the absence of semantic understanding and intentionality, dependence on statistical correlations, and structural opacity. The contribution is limited to the conceptual level. Its purpose is to establish a precise analytical framework that avoids conflating linguistic performance with reasoning and that enables subsequent normative—including constitutional—assessments of the use of artificial intelligence in decision-making functions.

Keywords: Artificial Intelligence · Conceptual Analysis · Machine Learning · Large Language Models · Opacity · Semantics.

La arquitectura conceptual de la inteligencia artificial: de la computación mecánica a la predicción estadística

Este artículo reconstruye la arquitectura conceptual de la inteligencia artificial a partir de su genealogía intelectual y de la transformación metodológica que conduce desde los primeros proyectos de computación mecánica y programable hasta el paradigma contemporáneo de predicción estadística. El análisis parte de los antecedentes decimonónicos vinculados a la Analytical Engine de Babbage y a las "Notes" de Ada Lovelace, para luego situar el giro fundacional de mediados del siglo XX asociado a la teoría de la computación y al programa de investigación inaugurado por la propuesta de Dartmouth. Sobre esta base histórica, el artículo distingue estructuralmente entre la inteligencia artificial simbólica —orientada a reglas explícitas y trazabilidad inferencial— y la inteligencia artificial estadística, cuyo rendimiento se funda en la optimización de funciones a partir de grandes volúmenes de datos. La sección central caracteriza a los modelos de lenguaje de gran escala como una tecnología de inferencia probabilística sobre secuencias lingüísticas, subrayando tres rasgos conceptualmente decisivos: ausencia de comprensión semántica e intencionalidad, dependencia de correlaciones estadísticas y opacidad estructural. Este trabajo de carácter conceptual establece un marco analítico para evitar equívocos entre desempeño lingüístico y razonamiento, y para habilitar evaluaciones normativas posteriores —incluidas las constitucionales— acerca del uso de IA en funciones decisorias.

Palabras Clave: Inteligencia Artificial · Análisis conceptual · Aprendizaje automático · Modelos Lingüísticos de gran tamaño · Opacidad · Semántica.

P. VÉLEZ LEÓN es Profesor Titular en el Departamento de Filosofía, Artes y Humanidades de la Universidad Técnica de Loja. Coordina el grupo de investigación «Laboratorio de Filosofía - FiloLab UTPL», dirige la Red de Investigación «Studia Humanitatis» y es Investigador Acreditado ante la SENESCYT (Ecuador). Antes de incorporarse a la UTPL, tuvo cargos académicos en las universidades de Salamanca, Autónoma de Madrid (España), y de Cuenca (Ecuador). Fue Titular (Director) de la Cátedra UNESCO de Ética y Sociedad en la Educación Superior. Realizó ampliación de estudios en el Consejo Superior de Investigaciones Científica [CSIC] y estancias de investigación en universidades de Brasil, Argentina y España. Ha obtenido becas y ayudas de investigación, entre otras, de la Organización de Estados Americanos (USA), Banco Santander y Ministerio de Educación (España). Trabaja principalmente en metafísica, epistemología y estética. También está interesado en la historia del pensamiento filosófico así como en los estudios CTS/STS, especialmente lo relativo a políticas científicas. En la UTPL su investigación actual explora sobre los fundamentos de la ontología y sus implicaciones en la estructura ontológica de los mundos regionales (p.e. el mundo del arte).

Contact: Universidad Técnica Particular de Loja, Facultad de Ciencias Sociales, Educación y Humanidades, Departamento de Filosofía, Artes y Humanidades, Loja. e-mail (✉): pevelez@utpl.edu.ec · [iD: https://orcid.org/0000-0002-5133-5041](https://orcid.org/0000-0002-5133-5041)

HISTORIA DEL ARTÍCULO | ARTICLE HISTORY

Received: 30-junio-2025; Accepted: 15-julio-2025; Published Online: 30-December-2025

COMO CITAR ESTE ARTÍCULO | HOW TO CITE THIS ARTICLE

Vélez León, P. (2025). «La arquitectura conceptual de la inteligencia artificial: de la computación mecánica a la predicción estadística». *Disputatio. Philosophical Research Bulletin* 14, no. 29: pp. 185–193.

© Studia Humanitatis – Universidad de Salamanca 2025