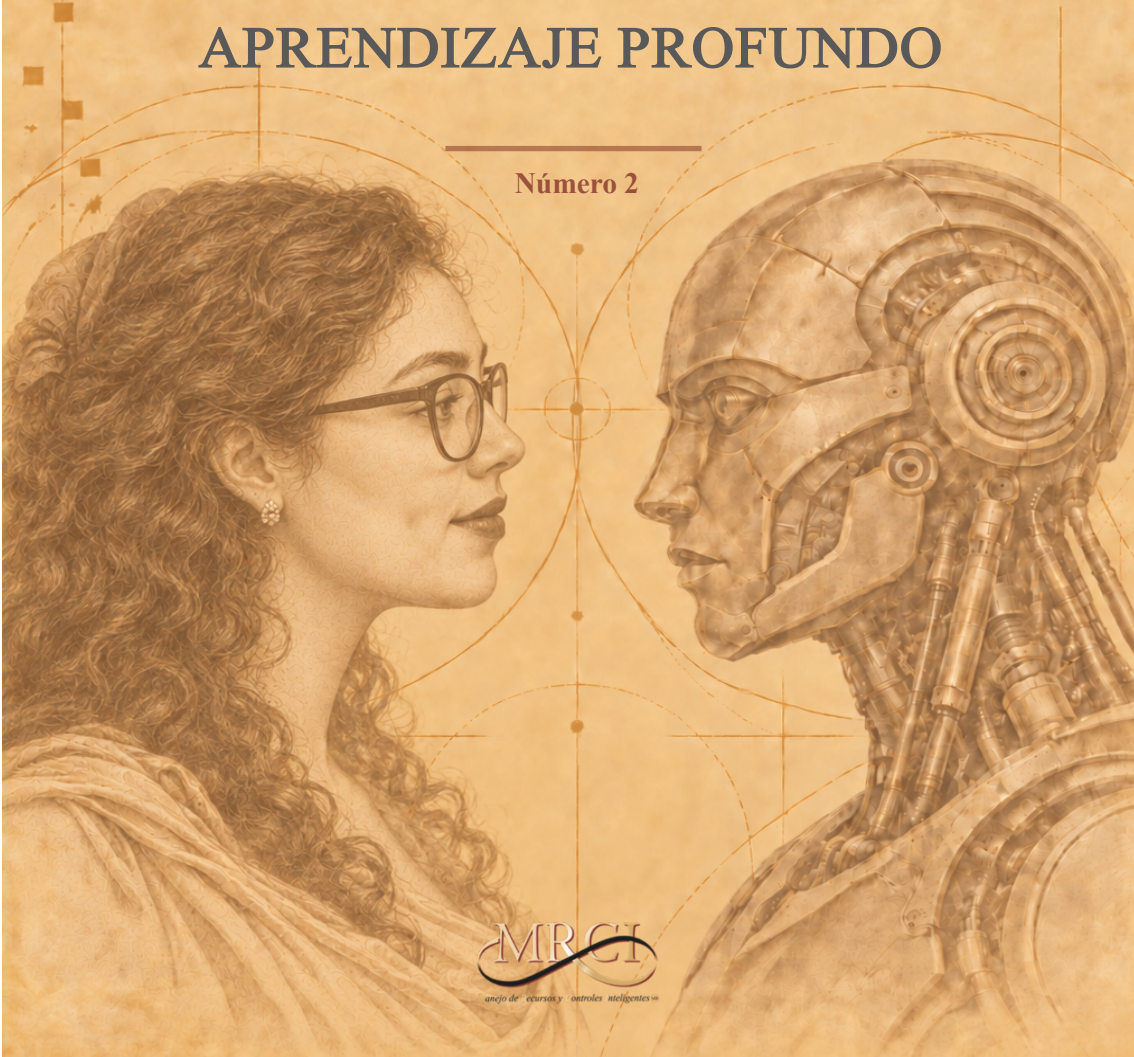
COLECCIÓN

MITOS Y REALIDADES DE LA IA

Entre la promesa y la realidad

DIFERENCIAS ENTRE INTELIGENCIA ARTIFICIAL: APRENDIZAJE AUTOMÁTICO Y APRENDIZAJE PROFUNDO

Número 2



MRCI

anejo de ecursos y ontroles nteligentes

DR. JOSÉ ANTONIO PÉREZ RAMOS

DIFERENCIAS ENTRE INTELIGENCIA ARTIFICIAL: APRENDIZAJE AUTOMÁTICO Y APRENDIZAJE PROFUNDO

MRCI
Manejo de Recursos y Controles Inteligentes SRL

SOBRE EL AUTOR

Doctor en Ciencias de lo Fiscal por el Instituto de Especialización para Ejecutivos. Maestro en Fiscal por la Universidad Autónoma "Benito Juárez" de Oaxaca. Maestro en Estudios Avanzados en Derechos Humanos por la Global Open University. Licenciado en Contaduría Pública por la Universidad Autónoma "Benito Juárez" de Oaxaca y Licenciado en Filosofía por la Universidad Nacionalista de México.

Socio fundador y director general de la Firma Manejo de Recursos y Controles Inteligentes (Firma MRCI). Fiscalista del año 2009 por la revista *Defensa Fiscal*. Doctor honoris causa por la Asamblea General del Claustro Doctoral Iberoamericano (2016). Doctor honoris causa por el Instituto Mexicano de Líderes de Excelencia (2017). Doctor honoris causa por 1Million Startups, Latinomics, Leaderships Forum y la Fundación Humanist World (2018). Doctor honoris causa por el Claustro de Doctores de la AISOL (American International School of Law) (2019).

Autor de diversas obras, entre las que destacan *Ius mutantur o del derecho que cambia* (Tirant Lo Blanch, 2025), *Futurología* (MRCI, 2025) y *Hasta que la muerte nos separe* (MRCI, 2025). Ha participado como coordinador en *Cuestiones tributarias. Problemas y controversias en el México actual* (Tirant Lo Blanch, 2023), *El costo de la justicia* (Apex Iuris, 2019), *Remuneraciones estratégicas inteligentes* (MRCI, 2015) y *Retos de la justicia tributaria* (Tirant Lo Blanch-AMDF, 2025).

Titular de *Pie de Página*, programa para la divulgación de los derechos humanos en México, a través de la plataforma Legal Cloud. Expositor en foros nacionales e internacionales. Colaborador en diversas revistas jurídicas, entre las que destacan *Defensa Fiscal* y *Legal Cloud*.

DIFERENCIAS ENTRE INTELIGENCIA ARTIFICIAL: APRENDIZAJE AUTOMÁTICO Y APRENDIZAJE PROFUNDO

DR. JOSÉ ANTONIO PÉREZ RAMOS

PRIMERA EDICIÓN, ATEMPORAL

Derechos reservados, propiedad de
Dr. José Antonio Pérez Ramos

Comentarios y opiniones: investigacion@mrci.com.mx

Título original: Diferencias entre inteligencia artificial: aprendizaje automático y aprendizaje profundo

Colección mitos y realidades de la inteligencia artificial. Entre la promesa y la realidad.

Autor: Dr. José Antonio Pérez Ramos.

Queda prohibida la reproducción total y parcial de esta obra denominada; DIFERENCIAS ENTRE INTELIGENCIA ARTIFICIAL: APRENDIZAJE AUTOMÁTICO Y APRENDIZAJE PROFUNDO, por cualquier medio, sin autorización escrita del autor.

PRINTED IN MEXICO
IMPRESO EN MÉXICO

Una diferencia que cambia decisiones

Una oficina recibe tres propuestas para ordenar su archivo digital: la primera ofrece inteligencia artificial, la segunda aprendizaje automático y la tercera aprendizaje profundo; si las presenta como tres productos de potencia creciente, la comparación ya comienza con un error, porque las expresiones se refieren a niveles distintos de descripción y una misma solución podría pertenecer a las tres categorías.

El problema importa antes de elegir un proveedor, pues influye en lo que la organización cree estar adquiriendo: un procedimiento que aplica conocimiento previamente formulado, un modelo ajustado con ejemplos, un componente que extrae representaciones de imágenes o una aplicación que reúne varios de ellos; cada posibilidad exige preguntas distintas sobre datos, resultados y responsabilidades.

También importa al explicar una falla: atribuirla genéricamente a la IA no permite saber si el sistema recibió un archivo ilegible, aprendió una asociación inadecuada, utilizó una categoría mal definida o aplicó una condición de



negocio equivocada; la palabra puede identificar el ámbito del problema, pero todavía falta localizar su mecanismo.

La pregunta central es cómo distinguir estos términos sin convertirlos en una jerarquía de superioridad ni perder de vista los sistemas compuestos; la respuesta propuesta combina un mapa conceptual con ejercicios que obligan a separar el nombre del método, la información disponible y la prueba que justificaría su utilización.

El alcance es deliberadamente acotado: no se ofrece un catálogo exhaustivo de técnicas ni una clasificación actual de productos comerciales; tampoco se pretende resolver aquí qué cuenta como inteligencia en todas las disciplinas, explicar en detalle cada red neuronal o comparar el tamaño de los modelos, asuntos que requieren desarrollos propios dentro de esta colección.

El método consiste en contrastar definiciones y descripciones técnicas con investigaciones identificables, examinar las condiciones de estudios empíricos y derivar consecuencias mediante ejemplos expresamente hipotéticos; las cifras de los casos son construcciones didácticas comprobables aritméticamente, no mediciones

de una institución real ni estimaciones sobre un producto disponible.

Una revisión selectiva permite reconstruir distinciones útiles, aunque no demuestra que se hayan agotado las publicaciones existentes; por ello, los resultados de 2012, 2022 y 2025 se conservan en su contexto, y las recomendaciones de evaluación se presentan como propuestas razonadas que requieren adaptación y comprobación local.

El campo y las familias de métodos

La propuesta de Dartmouth de 1955 incluía lenguaje, abstracción, solución de problemas, redes neuronales y mejora de las máquinas dentro de un programa de investigación compartido; su amplitud histórica muestra que la inteligencia artificial no nació como sinónimo de una única técnica, aunque las conjeturas del documento tampoco deben confundirse con capacidades entonces demostradas (McCarthy et al., 1955).

Una definición de trabajo suficiente para este artículo entiende la IA como un campo que investiga y construye sistemas para realizar tareas asociadas con capacidades

como inferir, percibir, planificar o producir lenguaje; esta formulación orienta el análisis, pero no convierte toda automatización en IA ni resuelve por decreto los casos fronterizos.

Dentro de ese campo puede haber procedimientos que exploran alternativas conforme a criterios explícitos, representaciones de conocimiento y métodos que se ajustan con datos; describir una tarea como inteligente no determina de antemano cuál de esos recursos se utiliza, del mismo modo que conocer la finalidad de una herramienta no revela todos sus componentes.

El aprendizaje automático, por su parte, identifica métodos en los que la experiencia interviene en la adquisición o modificación de una capacidad evaluable; Mitchell (2006, p. 1) organiza esa idea mediante tres elementos: tarea, medida de desempeño y experiencia, lo que obliga a precisar qué mejora se espera y respecto de qué evidencia se comprobará.

La conexión con la estadística no desaparece por utilizar el nombre aprendizaje automático; Mitchell (2006, p. 1) sitúa la disciplina en una intersección con la informática y

la estadística, de modo que el mapa de inclusión utilizado en divulgación no debe interpretarse como un reparto exclusivo de propiedad sobre técnicas que participan en varias tradiciones de investigación.

El aprendizaje profundo constituye una familia particular dentro del aprendizaje automático: LeCun et al. (2015) lo caracterizan por modelos con múltiples capas que aprenden representaciones de los datos en diferentes niveles de abstracción; el rasgo relevante es cómo se construye y ajusta esa composición, no que el programa tenga muchos archivos, instrucciones o pantallas.

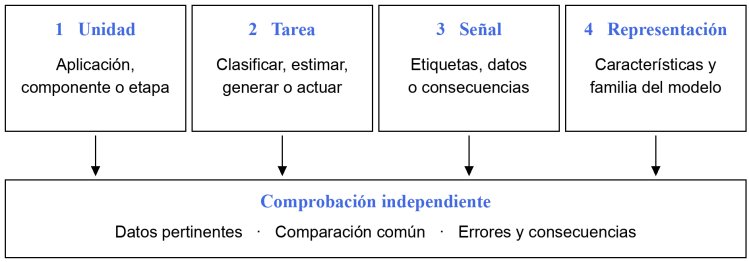
Goodfellow et al. (2016, cap. 1, figs. 1.4–1.5) sitúan entre aprendizaje automático y aprendizaje profundo el aprendizaje de representaciones; esta distinción evita una simplificación frecuente, porque obtener una representación mediante aprendizaje no implica por sí solo que el modelo pertenezca a la familia profunda.

La inclusión conceptual tampoco expresa una escala de mérito: afirmar que un método pertenece a un conjunto más específico informa sobre sus características, pero no demuestra que produzca mejores resultados para

cualquier tarea; esa conclusión necesitaría una comparación que el diagrama, por definición, no contiene.

La Figura 1 resume un modo de lectura complementario: las etiquetas generales se conectan con preguntas sobre tarea, señal y representación, mientras el desempeño permanece como una exigencia de comprobación independiente; las relaciones del esquema son conceptuales y no representan una cadena obligatoria de desarrollo tecnológico.

Figura 1. *Cuatro preguntas para describir un componente de IA*



Nota. Elaboración propia; el esquema organiza preguntas de análisis y no representa resultados experimentales ni una clasificación exhaustiva.

La primera pregunta identifica la unidad descrita: una aplicación completa, un componente de clasificación, una función de búsqueda o un procedimiento de aprendizaje; si esa unidad cambia a mitad de la conversación, expresiones aparentemente contradictorias pueden estar refiriéndose a objetos distintos.

La segunda pregunta define la tarea observable: asignar una categoría, estimar una cantidad, agrupar registros, generar una secuencia o escoger una acción; compartir entrada, como un documento, no vuelve equivalentes sus finalidades, pues reconocer el tipo de archivo exige una prueba diferente de determinar si una obligación se cumplió.

La tercera pregunta examina la señal de aprendizaje, entendida aquí como la información con la que se orienta el ajuste: ejemplos acompañados de respuestas, estructura de los propios datos o consecuencias de acciones; la cuarta pregunta describe la representación y la familia de modelos, sin deducir de ellas una calidad que todavía no se ha medido.



Qué significa aprender en una tarea

Supongamos que un archivo contiene contratos, comprobantes y comunicaciones, y que una organización desea proponer una categoría para cada documento nuevo; formular el problema como aprendizaje requiere identificar qué constituye un ejemplo, cuál es la respuesta esperada y cómo se juzgará una predicción, en lugar de limitarse a reunir todos los archivos disponibles.

La unidad de ejemplo puede ser una página, un documento completo o un expediente integrado por varios archivos; esas alternativas producen problemas diferentes, porque una página aislada puede contener únicamente una firma mientras el expediente incluye el texto que permite determinar su función, por lo que cambiar la unidad modifica la información accesible al modelo.

La respuesta esperada tampoco existe siempre como una propiedad inequívoca del documento: una comunicación puede adjuntar un contrato y un comprobante puede formar parte de una solicitud, de manera que el equipo debe decidir si clasifica el archivo completo, su pieza principal o varias unidades contenidas en él; los



desacuerdos de definición no se corrigen simplemente aumentando la capacidad del clasificador.

Una vez fijada la tarea, el aprendizaje puede ajustar una relación entre características de entrada y categorías; en un ejemplo elemental, las características podrían ser la presencia de ciertas expresiones, la extensión del documento y la existencia de campos estructurados, mientras los coeficientes o divisiones aprendidos establecen cómo combinar esa información.

Este ejemplo permite distinguir una regla impuesta por la persona de una regularidad estimada con datos: si la instrucción ordena clasificar como comprobante todo archivo cuyo campo verificado indique ese tipo, la condición fue especificada directamente; si un procedimiento selecciona condiciones a partir de ejemplos etiquetados, el origen de esas condiciones incluye aprendizaje, aunque su aplicación posterior pueda verse como una secuencia de reglas.

Los bosques aleatorios ilustran una familia que aprende sin ser una red profunda: Breiman (2001) describe conjuntos de predictores de árbol construidos con

aleatoriedad y analiza su generalización, además de su aplicación a clasificación y regresión; por tanto, identificar aprendizaje automático con redes neuronales eliminaría indebidamente una parte importante del campo.

Un árbol aprendido puede preguntar por una variable y dividir los casos según su valor, pero la legibilidad de una condición local no equivale a una explicación completa de cualquier sistema que use árboles; cuando muchas unidades interactúan, comprender cada pieza por separado puede no bastar para reconstruir la salida conjunta, distinción que aquí interesa por sus consecuencias para la documentación.

La medida de desempeño introduce otra decisión: contar todos los aciertos puede servir para ciertas tareas, aunque deja de ser suficiente cuando las categorías tienen frecuencias muy diferentes o cuando un error tiene consecuencias mucho mayores que otro; por eso conviene formular el objetivo en el lenguaje del trabajo, además de expresarlo mediante una cifra.

En el archivo hipotético, perder una comunicación común quizá implique repetir una búsqueda, mientras archivar



equivocadamente el único documento que requiere respuesta urgente podría tener otra consecuencia; no se deduce de ello un peso universal para cada error, sino la necesidad de acordar qué fallas resultan tolerables antes de seleccionar el método.

Aprender una relación tampoco acredita que esa relación sea causal: si ciertos documentos llevan un sello que coincide con su categoría en el material de entrenamiento, el modelo puede utilizarlo como señal, pero ello no demuestra que el sello constituya la naturaleza del documento ni que mantenga esa asociación cuando cambie el procedimiento administrativo.

La experiencia disponible delimita las oportunidades de aprendizaje y sus trampas: un conjunto formado únicamente por archivos bien digitalizados no informa directamente sobre fotografías borrosas, y un archivo histórico que reúne solo asuntos concluidos puede omitir precisamente los casos que más interesan en la operación actual; estos son problemas de diseño del caso, no propiedades que una etiqueta técnica resuelva de antemano.



Por último, utilizar un modelo aprendido no significa que esté aprendiendo durante cada uso: producir una respuesta con parámetros previamente ajustados y modificar esos parámetros son operaciones distinguibles; una aplicación también puede actualizar documentos consultables o preferencias de presentación sin que ese cambio constituya un nuevo entrenamiento del modelo.

Qué añade una representación profunda

Una representación es una forma de codificar información para que un procedimiento pueda operar con ella; en el archivo hipotético, convertir el contenido en un conjunto de indicadores de palabras ya constituye una decisión representacional, aunque dichos indicadores hayan sido diseñados manualmente y aunque la salida final sea tan breve como una etiqueta.

La aportación característica descrita por LeCun et al. (2015) consiste en aprender transformaciones sucesivas, de modo que las representaciones intermedias participan en el ajuste del sistema; la descripción no implica que cada nivel corresponda siempre a un concepto comprensible para una persona ni que pueda asignarse una interpretación única a cada componente interno.



Pensemos, como analogía didáctica limitada, en un documento fotografiado: antes de proponer su categoría, un sistema podría aprovechar regularidades locales, distribución de elementos y relaciones más amplias entre partes; lo relevante es que esa información pueda transformarse y combinarse de manera útil, no suponer que el programa sigue exactamente la secuencia de lectura de una persona.

El nombre profundo se refiere a composición y niveles de procesamiento, aunque no hay un umbral universal que permita clasificar todos los modelos contando de la misma manera; Goodfellow et al. (2016, cap. 1, pp. 5–9) explican que la medición de profundidad depende incluso de cómo se representan las operaciones, por lo que una cifra aislada de capas necesita contexto.

Tampoco debe confundirse profundidad con cantidad de parámetros, tiempo de ejecución o tamaño del archivo: son propiedades relacionadas en algunas implementaciones, pero no definiciones equivalentes; un sistema puede contener numerosos parámetros sin que ese dato determine por sí mismo su profundidad, y dos modelos profundos pueden utilizar recursos muy diferentes.



La comparación con una representación manual permite precisar qué se delega al aprendizaje: en el primer caso, las personas pueden decidir que se cuenten expresiones concretas; en el segundo, el procedimiento puede ajustar transformaciones internas a partir de los datos y del objetivo, aunque siguen existiendo decisiones humanas sobre ejemplos, señales, estructura y evaluación.

Por ello, la frase el sistema descubre sus propias características exige cautela: describe un proceso de ajuste dentro de condiciones establecidas, no una investigación autónoma sin restricciones ni la creación de información inexistente; las posibilidades de representación dependen del diseño y de la experiencia, y la utilidad depende de su comprobación en la tarea relevante.

Un resultado histórico permite apreciar la importancia del enfoque sin convertirlo en una garantía general: Krizhevsky et al. (2012) presentaron una red convolucional profunda para clasificación de imágenes de ImageNet y comunicaron una mejora respecto de alternativas de su evaluación; ese trabajo acredita un avance en condiciones

concretas, no una prueba de que toda tarea posterior requiera redes profundas.

La diferencia entre demostrar una capacidad y extenderla a cualquier entorno es decisiva: reconocer categorías de imágenes en una competencia no equivale a ordenar automáticamente todos los documentos de una oficina, porque cambian entradas, clases, criterios de error y condiciones de uso; la investigación ofrece razones para ensayar un enfoque, mientras la adopción exige evidencia adicional.

Una organización tampoco tiene que reproducir desde cero toda la experiencia utilizada para desarrollar cada componente que emplea: puede partir de representaciones aprendidas previamente y evaluar su adaptación; la existencia de esa posibilidad no elimina la necesidad de revisar procedencia, compatibilidad de las entradas y desempeño, pero impide concluir que pocos ejemplos locales descartan automáticamente cualquier técnica profunda.



Las señales cruzan las familias

Supervisado, no supervisado, autosupervisado y por refuerzo describen aspectos del aprendizaje que no ocupan el mismo eje que profundo; colocarlos como cuatro productos rivales junto a las redes neuronales confunde la fuente de información para aprender con la familia de procedimientos que la utiliza.

En aprendizaje supervisado, los ejemplos incorporan respuestas objetivo que orientan el ajuste; esas respuestas pueden provenir de personas, registros o procedimientos de obtención previamente definidos, pero llamarlas etiquetas no certifica su corrección, pues una clasificación inconsistente seguirá siendo inconsistente aunque el archivo que la contiene sea grande y esté bien organizado.

La etiqueta supervisado tampoco significa que una persona revise cada predicción después de instalar el sistema: supervisión durante el aprendizaje y revisión humana durante la operación son decisiones diferentes; el archivo podría utilizar un modelo supervisado con salidas automáticas o someter todas sus propuestas a revisión, y

ambas configuraciones requieren descripciones separadas.

Mitchell (2006, secc. 2.3) distingue tareas supervisadas, agrupamiento, detección de anomalías y aprendizaje por refuerzo; esa diversidad muestra que la experiencia no se reduce a una colección de preguntas acompañadas de respuestas correctas, aunque cada procedimiento necesita supuestos y un criterio que permita apreciar lo que obtuvo.

En un agrupamiento hipotético, reunir documentos por semejanza puede ayudar a explorar el archivo sin que se hayan fijado categorías previas; sin embargo, el nombre de cada grupo y su utilidad institucional requieren interpretación, porque una partición coherente con ciertas características podría reflejar plantillas o estilos de digitalización en lugar de los temas que interesan a los usuarios.

La ausencia de etiquetas humanas específicas tampoco significa ausencia de una señal de aprendizaje: en una tarea autosupervisada se puede construir un objetivo a partir de partes de los propios datos; el modelo recibe un

problema cuyo resultado de entrenamiento puede recuperarse del material disponible, aunque nadie haya anotado manualmente cada respuesta.

BERT ofrece un ejemplo documentado: Devlin et al. (2019, secc. 3.1) entrenaron representaciones de lenguaje con objetivos que incluían predecir elementos ocultos del texto y posteriormente ajustaron modelos para tareas específicas; el trabajo permite distinguir la etapa que aprovecha texto sin etiquetas de la tarea final del usuario, sin atribuir a ambas el mismo tipo de supervisión.

La distinción afecta el archivo hipotético: aprender regularidades del lenguaje en una gran colección no equivale a haber aprendido sus categorías administrativas locales; una representación reutilizable puede facilitar el trabajo, pero seguirá siendo necesario comprobar si separa los documentos conforme al criterio que la oficina realmente necesita.

El aprendizaje por refuerzo organiza otra relación: las acciones del sistema influyen en una secuencia de resultados y la señal permite valorar sus consecuencias; Watkins y Dayan (1992) desarrollan Q-learning mediante



estimaciones de la calidad de acciones en estados, con un resultado de convergencia sujeto a condiciones explícitas, entre ellas exploración reiterada y representación discreta de los valores.

Ese antecedente basta para descartar la equivalencia entre refuerzo y aprendizaje profundo: una representación tabular de valores puede participar en aprendizaje por refuerzo sin ser una red profunda; utilizar otra familia para aproximar esos valores cambia la implementación, pero no elimina el carácter secuencial del problema ni la importancia de la señal de recompensa.

Generativo pertenece asimismo a otro eje: señala una finalidad o capacidad de producción de datos, no la presencia obligatoria de un tipo específico de supervisión; las redes adversarias generativas de Goodfellow et al. (2014) constituyen un ejemplo de entrenamiento de modelos generativos mediante interacción entre componentes, suficiente para mostrar que el mapa necesita más de una dimensión.

La Tabla 1 concentra estas distinciones para consulta: cada fila identifica qué pregunta permite responder una



expresión y qué conclusión sería indebido obtener únicamente a partir de ella; los ejemplos son ilustrativos y no forman una clasificación exhaustiva de todos los sistemas posibles.

Tabla 1. *Categorías que responden preguntas distintas*

Expresión	**Pregunta que ayuda a responder**	**Lo que todavía debe comprobarse**
Inteligencia artificial	¿En qué campo o tipo de aplicación se sitúa?	Componentes reales y capacidad concreta
Aprendizaje automático	¿Interviene experiencia en la obtención de una capacidad?	Tarea, señal, criterio y generalización
Aprendizaje profundo	¿Se aprenden representaciones mediante composición de niveles?	Utilidad y costo en el contexto previsto

Expresión	Pregunta que ayuda a responder	Lo que todavía debe comprobarse
Supervisado	¿El ajuste utiliza respuestas objetivo?	Calidad de etiquetas y revisión en operación
Autosupervisado	¿La señal se construye a partir de los propios datos?	Relación con la tarea final y su evaluación
Por refuerzo	¿Se valoran acciones mediante sus consecuencias?	Recompensa, exploración y restricciones
Generativo	¿El componente produce nuevas muestras o secuencias?	Fidelidad, validez y uso permitido de la salida

Nota. Síntesis conceptual propia basada en Mitchell (2006), LeCun et al. (2015), Devlin et al. (2019), Watkins & Dayan (1992) y Goodfellow et al. (2014); las preguntas de comprobación son propuestas de este artículo.

Qué enseñan las comparaciones publicadas

Los resultados favorables a una familia no deben convertirse en una ley universal, pero tampoco pueden descartarse porque contradigan una preferencia; el examen exige saber qué modelos se compararon, qué datos se incluyeron, cómo se seleccionaron configuraciones y qué recursos se contabilizaron antes de trasladar la conclusión a otro problema.

Grinsztajn et al. (2022) compararon métodos profundos y basados en árboles mediante un conjunto de 45 bases tabulares y un procedimiento que consideró la selección de hiperparámetros; sus resultados favorecieron los métodos de árboles en el régimen de tamaño medio estudiado, mientras los criterios de selección y preparación excluían o controlaban situaciones como datos faltantes, series temporales y ciertos desequilibrios.

Por esa razón, utilizar el estudio para sostener que los árboles siempre superan a las redes sería exceder su alcance; también sería incorrecto omitirlo al justificar una red únicamente porque se la considera más moderna, pues aporta evidencia de que la estructura de los datos y

las condiciones de comparación pueden modificar la elección.

Hollmann et al. (2025) publicaron resultados favorables a TabPFN en comparaciones de predicción tabular con hasta 10,000 observaciones y 500 características; el método había sido previamente entrenado sobre conjuntos sintéticos y se evaluó frente a alternativas que incluían métodos basados en árboles, con un protocolo que debe distinguirse del trabajo de 2022.

Los dos estudios no obligan a elegir entre versiones incompatibles de la realidad: documentan métodos, momentos y alcances diferentes, además de una investigación que puede cambiar las alternativas disponibles; el resultado posterior cuestiona convertir una ventaja histórica en regla permanente, aunque tampoco prueba superioridad fuera del dominio evaluado.

Existe una consecuencia adicional para la contabilidad de recursos: aplicar rápidamente un modelo previamente entrenado no significa que su desarrollo haya carecido de costo; al comparar opciones para una organización conviene separar el costo de producir el componente, el de



adaptarlo y el de utilizarlo, sin sumar o excluir partidas de manera diferente para favorecer una conclusión.

Este contraste no constituye una clasificación de desempeño vigente en octubre de 2026: se utiliza para examinar dos afirmaciones generales mediante evidencia histórica identificable; una decisión de compra actual necesitaría, además, versiones disponibles, condiciones de servicio y pruebas propias, aspectos que no se deducen del prestigio de los artículos.

El archivo administrativo como prueba

Retomemos la oficina hipotética con una finalidad limitada: facilitar la recuperación de contratos, comprobantes y comunicaciones, manteniendo acceso al documento original y permitiendo corregir la categoría; no se le pide determinar consecuencias jurídicas, acreditar pagos ni sustituir una decisión profesional, por lo que el éxito debe medirse en el trabajo de localizar y organizar información.

El primer paso propuesto es examinar una muestra de archivos y formular un manual breve de categorías, incluyendo documentos mixtos, duplicados y casos

incompletos; dos revisores podrían clasificar una parte por separado y discutir las diferencias, no para fabricar una unanimidad artificial, sino para identificar dónde la tarea admite varias respuestas o requiere información adicional.

Si un revisor clasifica por el título y otro por el contenido principal, entregar sus etiquetas al modelo como si representaran una única norma oculta el desacuerdo; antes de ajustar cualquier método se necesita resolver el criterio o conservar explícitamente la ambigüedad, porque pedir al sistema una respuesta única puede imponer una precisión que el propio encargo no permite.

La oficina debería conservar una alternativa de referencia: por ejemplo, una búsqueda por metadatos verificados combinada con clasificación manual; esa alternativa puede parecer modesta, pero permite saber si el nuevo componente mejora un proceso ya ordenado o si el beneficio atribuido al aprendizaje proviene, en realidad, de haber limpiado el archivo y definido mejor las categorías.

Una primera solución candidata podría utilizar características textuales explícitas y un clasificador aprendido; otra podría aprovechar representaciones

profundas obtenidas previamente, y una tercera combinar reconocimiento de imagen con clasificación, aunque todas deben recibir el mismo objetivo y evaluarse sobre documentos comparables antes de sostener diferencias de utilidad.

El conjunto de prueba necesita respetar la unidad del problema: si una versión de un contrato queda en entrenamiento y otra casi idéntica en evaluación, la facilidad para reconocer esa plantilla puede exagerar la capacidad ante documentos nuevos; se propone agrupar versiones relacionadas y decidir la separación antes de examinar resultados, dejando constancia de cualquier caso fronterizo.

Separar por fecha puede ser pertinente cuando se espera procesar documentos futuros, aunque tampoco constituye una garantía automática: una época puede incluir un tipo de expediente inexistente en otra y la comparación deberá reconocer ese cambio; la elección entre separación temporal, por expediente o por emisor debe responder al uso previsto, en vez de aplicarse como una receta universal.

La evaluación debería conservar al menos tres lecturas: qué documentos se clasificaron correctamente, cuáles fueron remitidos a revisión y qué errores llegaron al usuario final; una solución que acierta mucho porque rechaza casi todo puede resultar poco útil, mientras otra que responde siempre podría trasladar una carga elevada de corrección a las personas.

No existe contradicción en utilizar una etapa profunda y una regla explícita dentro del mismo flujo: el reconocimiento podría extraer información de una imagen y una condición de negocio impedir que se archive sin revisión un documento incompleto; sin embargo, ese límite debe probarse, pues la presencia de una regla en el diseño no demuestra que se aplique correctamente a todas las salidas del componente anterior.

La evaluación de extremo a extremo puede descubrir problemas que no aparecen al medir piezas aisladas: un clasificador quizá funcione bien con texto limpio, pero falle cuando el reconocimiento intercambia un número o pierde un encabezado; atribuir al modelo de clasificación el desempeño obtenido con entradas ideales ocultaría una dificultad de integración.



Considérese un ejemplo aritmético propio: entre 1,000 archivos, 900 se convierten en texto utilizable y, de esos 900, el clasificador acierta en 855; el 95% de acierto sobre entradas utilizables equivale a 855 resultados correctos sobre los 1,000 archivos iniciales si los otros 100 no producen una clasificación correcta, es decir, 85.5% de cobertura correcta total bajo estos supuestos.

El cálculo no presupone independencia entre fallas, porque utiliza conteos encadenados del mismo escenario; tampoco significa que los 100 archivos restantes deban considerarse errores del clasificador, pues pueden enviarse a revisión, pero exige informar ambos denominadores para que la organización sepa cuántos documentos quedan realmente resueltos por el flujo.

Una decisión razonable podría favorecer el método sencillo si ofrece el resultado requerido con menor carga de mantenimiento, o el profundo si mejora justamente los casos que importan; en ambos supuestos, el fundamento sería la prueba local y la comparación completa, no la posición de una palabra dentro de un diagrama.



También podría justificar una solución dividida: automatizar un subconjunto estable y revisar el resto, siempre que se explique cómo se identifica cada grupo y que el criterio de separación no dependa de una certeza inexistente; declarar el alcance operativo resulta más útil que prometer una clasificación universal imposible de verificar con los datos disponibles.

Muchos aciertos y una omisión decisiva

El segundo caso hipotético modifica la tarea: de 10,000 documentos, exactamente 100 requieren una revisión especial, y se busca señalarlos para que una persona los examine; los otros 9,900 no necesitan esa revisión, por lo que responder siempre que no hay nada que revisar produce 99% de acierto global y omite todos los casos de interés.

Supongamos ahora un clasificador que identifica 90 de los 100 documentos relevantes y marca por error 99 de los 9,900 restantes; la Tabla 2 muestra los conteos, que son construidos exclusivamente para explicar la relación entre frecuencia, errores y utilidad, sin atribuirlos a ningún método o institución real.

Tabla 2. *Conteos del caso hipotético de revisión especial*

Resultado	Requiere revisión	No requiere revisión
Señala revisión	90	99
No señala revisión	10	9,801
Total	100	9,900

Nota. Elaboración propia con datos hipotéticos; las filas expresan la salida del clasificador y las columnas la situación establecida para el ejercicio, sin incertidumbre en esas etiquetas de referencia.

El acierto global es de 9,891 documentos sobre 10,000, equivalente a 98.91%; parece inferior al 99% del procedimiento que no señala ningún documento, pero recupera 90 casos relevantes, lo que demuestra que ordenar métodos únicamente por el promedio puede premiar una conducta contraria a la finalidad del trabajo.

La sensibilidad, entendida aquí como la proporción de documentos relevantes identificados, es 90 de 100, o 90%; la precisión de las alertas, entendida como la proporción



de documentos realmente relevantes entre todos los señalados, es 90 de 189, aproximadamente 47.62%, de manera que más de la mitad de las alertas requieren revisión para terminar descartadas.

Ninguna de esas cifras informa por sí sola si el sistema debe adoptarse: el 90% de sensibilidad puede resultar insuficiente si las diez omisiones tienen consecuencias graves, mientras 189 revisiones pueden ser aceptables o excesivas según la capacidad disponible; la decisión necesita vincular las métricas con efectos concretos y con alternativas factibles.

El término precisión suele emplearse coloquialmente como calidad general, pero aquí designa una razón específica; informar su definición evita comparar una precisión de alertas con una exactitud global como si fueran la misma medida, y permite reconocer que dos porcentajes altos pueden describir aspectos diferentes del procedimiento.

La frecuencia del evento cambia también la lectura: si otro conjunto hipotético conserva sensibilidad de 90% y tasa de falsas alertas de 1%, pero ahora contiene 2,000 documentos relevantes y 8,000 restantes, produciría 1,800



alertas correctas, 200 omisiones y 80 falsas alertas; la precisión sería 1,800 de 1,880, aproximadamente 95.74%, sin que hayamos supuesto una mejora de esas dos tasas del clasificador.

Mantener iguales las tasas es una condición del ejercicio, no una predicción de lo que ocurrirá al trasladar un modelo: las nuevas entradas pueden modificar también su desempeño; el ejemplo demuestra aritméticamente que una precisión observada depende de la composición del conjunto y, por ello, no puede transportarse como propiedad invariable del producto.

Se puede añadir una valoración hipotética de consecuencias: si cada omisión recibe un costo de 20 unidades y cada falsa alerta uno de una unidad, la regla que nunca avisa acumula 2,000 unidades, mientras el primer clasificador acumula 299; esos números muestran cómo cambia la comparación, pero las unidades no representan dinero ni un juicio universal sobre el valor de cada documento.

Los pesos deberían ser decididos y justificados por quienes conocen las consecuencias del proceso,

incluyendo la posibilidad de que un daño no sea reducible adecuadamente a una cifra; aun cuando se emplee una medida compuesta, conviene conservar los conteos originales para que la aparente simplicidad del resultado no oculte las decisiones que lo producen.

La lección para distinguir familias de métodos es directa: nada en este cálculo depende de que el clasificador sea profundo, un árbol o una combinación; conocer su familia ayuda a explicar cómo se obtiene la predicción, mientras conocer los denominadores y las consecuencias permite valorar si esa predicción sirve para la finalidad establecida.

Cuando la acción modifica lo que sigue

El tercer caso es una simulación conceptual de gestión de turnos, sin aplicación a personas reales: en cada paso, un sistema puede asignar un recurso a una de dos filas de trabajo, y la decisión modifica cuántas solicitudes quedan pendientes y qué opciones estarán disponibles después; esta dependencia temporal permite apreciar por qué escoger acciones no siempre equivale a clasificar registros independientes.



Una regla fija podría alternar filas, mientras un procedimiento de aprendizaje por refuerzo podría valorar acciones a partir de consecuencias observadas en un entorno simulado; el propósito del ejemplo no es demostrar cuál funcionaría mejor, sino distinguir el problema secuencial del que asigna una etiqueta sin intervenir en la evolución posterior del entorno.

Supongamos que el diseño premia únicamente el número de solicitudes cerradas en el siguiente paso: una política que atiende repetidamente los casos más breves podría obtener buen resultado inmediato y dejar esperando los más largos; la dificultad proviene del criterio de recompensa y de las restricciones, aunque el método consiga optimizar precisamente lo que se le pidió.

Para corregir el planteamiento habría que discutir el tiempo acumulado de espera, límites de postergación y condiciones de atención, además de la capacidad del modelo; añadir capas sin modificar el objetivo mantendría abierta la posibilidad de aprender una política eficaz para una finalidad mal especificada.

El resultado de Watkins y Dayan (1992) no autoriza a prometer que cualquier implementación encontrará una solución óptima: su demostración depende de condiciones matemáticas, y trasladarla a aproximaciones diferentes o entornos donde la exploración está restringida exige argumentos adicionales; citar el nombre del algoritmo no suprime esas condiciones.

En este caso sería pertinente registrar episodios completos, comparar políticas bajo escenarios comunes y observar consecuencias acumuladas; un alto porcentaje de acciones coincidentes con una lista de ejemplos no bastaría para probar una mejora del proceso si la decisión cambia los estados futuros que el sistema tendrá que enfrentar.

Tampoco se justifica explorar libremente en una operación con consecuencias reales porque el método aprende mediante experiencia: la definición técnica del aprendizaje no concede autorización para experimentar sobre usuarios; el ejemplo permanece en un entorno hipotético precisamente para separar la comprensión del mecanismo de una eventual decisión de despliegue.

La información que no está en la entrada

El cuarto caso recupera un límite sencillo y frecuentemente oculto por la fluidez de una respuesta: un expediente contiene un documento que describe la entrega prevista de un bien, pero no incluye constancia de recepción ni otra evidencia de ejecución; se pide a un sistema concluir si la entrega ocurrió, aunque la entrada únicamente expresa lo que debía suceder.

Un clasificador podría identificar el tipo de documento y un modelo de lenguaje podría explicar sus cláusulas, pero esas capacidades no aportan por sí mismas el hecho faltante; reconocer una declaración de entrega prevista y acreditar una entrega realizada constituyen tareas diferentes, con necesidades informativas que tampoco se vuelven equivalentes al usar aprendizaje profundo.

La imposibilidad se aprecia mediante dos escenarios compatibles con la misma entrada: en uno se entregó el bien y falta incorporar la constancia; en el otro nunca se entregó, y tampoco existe la constancia; si el sistema recibe exactamente el mismo documento en ambos, no dispone allí de una observación que permita distinguir esos hechos particulares con certeza.

Podría utilizar regularidades generales para estimar qué situación es más frecuente bajo ciertos supuestos, pero una estimación no equivale a determinar lo ocurrido en ese caso; la salida necesita conservar la diferencia entre probabilidad, afirmación documental y comprobación, en lugar de presentar un resultado plausible como si fuera una observación directa.

El criterio de evaluación debería premiar la identificación de esa insuficiencia cuando la tarea exige evidencia, no únicamente la producción de una respuesta cerrada; si todos los ejemplos de evaluación obligan a escoger sí o no aun cuando faltan datos, el diseño puede penalizar precisamente la conducta más responsable para el propósito declarado.

También importa evitar una inferencia simétrica: la ausencia de constancia en los archivos aportados no demuestra necesariamente ausencia de entrega; una respuesta cuidadosa delimitaría qué se examinó y qué dato permitiría resolver la diferencia, preservando la incertidumbre en lugar de sustituir una conclusión positiva injustificada por otra negativa igualmente excesiva.

Este caso no constituye una opinión jurídica sobre un expediente real, sino un análisis informativo: demuestra que dos situaciones distintas pueden ser indistinguibles a partir de la entrada definida; la búsqueda de documentación adicional cambiaría el problema, mientras cambiar únicamente la etiqueta del modelo mantendría la misma carencia.

Elegir con una hipótesis comprobable

Una selección responsable puede comenzar por una frase operativa: para este conjunto de entradas, se espera que el componente produzca esta salida con estas condiciones mínimas; formularla antes de escoger una familia evita que la tarea se adapte retrospectivamente al modelo preferido y permite reconocer cuándo la alternativa sencilla es suficiente.

Después conviene identificar qué parte requiere aprendizaje: tal vez el problema consiste en reconocer textos variables, mientras las condiciones de archivo son conocidas y estables; separar ambas funciones ayuda a diseñar la comparación y a no exigir que un modelo vuelva

a estimar relaciones que la organización ya puede establecer de manera verificable.

El siguiente paso es determinar la información disponible y la que falta: número y diversidad de ejemplos, procedencia de etiquetas, modalidades de entrada, frecuencia de casos difíciles y posibilidad de reutilizar representaciones; ninguno de esos datos selecciona automáticamente un ganador, pero delimita qué propuestas tienen una justificación plausible para ser probadas.

La prueba debería establecer previamente resultados que permitirían conservar o cambiar la elección inicial: si el componente más complejo no mejora las omisiones importantes ni reduce la carga de revisión, el costo adicional necesita otra justificación; si la alternativa sencilla falla justamente en los documentos que concentran el trabajo, su facilidad de explicación no resuelve esa insuficiencia.

Se propone documentar también el resultado que refutaría la hipótesis de utilidad: una tasa inaceptable de documentos perdidos, una degradación marcada ante



nuevos formatos o una carga de revisión superior a la alternativa; sin una posibilidad real de rechazo, la evaluación corre el riesgo de convertirse en una demostración preparada para confirmar una decisión ya tomada.

El costo debe seguir la misma unidad de comparación que el beneficio: comparar el precio de una predicción con el costo total de un proceso manual produce una asimetría; resulta más informativo contar preparación, revisión, corrección y mantenimiento para un volumen y una calidad equivalentes, explicitando partidas excluidas e incertidumbres.

La documentación mínima propuesta incluye tarea, familia del modelo, forma de aprendizaje, versión, entradas admitidas, origen de las etiquetas, prueba utilizada y límites de la conclusión; el propósito es permitir que otra persona reconstruya por qué se eligió el componente y qué cambio justificaría volver a evaluarlo.

No toda actualización requiere repetir exactamente el mismo trabajo, pero una modificación relevante de categorías, tipos de entrada o consecuencias del error

puede invalidar la comparación anterior; conservar la fecha y el alcance de la evaluación permite distinguir un resultado todavía aplicable de un antecedente que solo orienta la siguiente prueba.

Objeciones al mapa conceptual

Una primera objeción sostiene que las fronteras son convencionales y que, por ello, carece de sentido insistir en los nombres; la crítica identifica una limitación real, porque varias técnicas pertenecen a tradiciones compartidas, pero no elimina la diferencia entre describir la finalidad de un sistema, la experiencia utilizada para ajustarlo y la forma de sus representaciones.

El mapa no pretende resolver una disputa sobre propiedad disciplinaria: su valor consiste en detectar inferencias indebidas, como deducir supervisión humana permanente de la palabra supervisado o capacidad superior de la palabra profundo; las fronteras pueden ser parcialmente discutibles y esas inferencias seguir siendo incorrectas.

Una segunda objeción señala que las combinaciones modernas vuelven insuficiente la clasificación: un

componente preentrenado puede recibir ejemplos durante el uso y convivir con reglas, recuperación de información o decisiones humanas; precisamente por esa razón se propone describir unidades y etapas, en lugar de asignar una única etiqueta al producto completo.

Una tercera objeción advierte que exigir tantas distinciones podría retrasar decisiones prácticas; la respuesta es proporcionalidad: no toda organización necesita estudiar la demostración matemática de cada método, pero sí requiere comprender qué promete el sistema, qué evidencia sostiene la promesa y quién asumirá los errores que la evaluación reconoce como posibles.

Una cuarta objeción defiende elegir siempre la alternativa sencilla para reducir riesgos; la sencillez puede ser una ventaja operativa, pero tampoco garantiza adecuación, pues una regla corta puede omitir variaciones esenciales o una representación demasiado limitada puede excluir la información necesaria para la tarea; el principio defendible es complejidad justificada por evidencia, no preferencia absoluta por una familia.

Finalmente, una buena comparación no agota las razones para adoptar una solución: pueden existir restricciones de acceso, privacidad, mantenimiento o uso permitido que descarten un método aun cuando obtenga mejores resultados en una prueba; clasificarlo correctamente facilita la discusión, pero no reemplaza las demás decisiones que corresponden a las personas responsables.

Conclusión

La inteligencia artificial designa un campo amplio; el aprendizaje automático identifica métodos en los que la experiencia participa en la obtención de capacidades evaluables, y el aprendizaje profundo caracteriza una familia que aprende representaciones mediante composiciones de niveles; la relación es útil siempre que no se confunda inclusión con superioridad. Los casos examinados muestran que una misma tarea admite familias distintas, que las señales de aprendizaje atraviesan esas familias y que los resultados dependen de información y criterios concretos; la exactitud global puede ocultar omisiones decisivas, la rapidez de uso puede descansar en un entrenamiento previo y una entrada insuficiente puede impedir distinguir hechos diferentes cualquiera que sea la complejidad del modelo. La elección técnica debe formularse, por tanto, como una hipótesis de utilidad sometida a contraste: describir el componente, definir la tarea, identificar su experiencia, comprobar sus resultados y conservar los límites de la evidencia; entender las diferencias entre IA, aprendizaje automático y aprendizaje profundo permite hacer mejores preguntas, mientras la respuesta sobre qué utilizar corresponde a una evaluación situada y revisable.

Referencias

Breiman, L. (2001). Random forests [Bosques aleatorios]. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Devlin, J., Chang, M.-W., Lee, K., y Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding [BERT: Preentrenamiento de transformadores bidireccionales profundos para la comprensión del lenguaje]. En J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning* [Aprendizaje profundo]. MIT Press. https://www.deeplearningbook.org/

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., y Bengio, Y. (2014). Generative adversarial nets [Redes generativas adversarias]. En Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 27, pp. 2672–2680). Curran Associates. https://proceedings.neurips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html

Grinsztajn, L., Oyallon, E., y Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? [¿Por qué los modelos

basados en árboles todavía superan al aprendizaje profundo en datos tabulares típicos?]. En S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in neural information processing systems* (Vol. 35, pp. 507–520). Curran Associates. https://doi.org/10.52202/068431-0037

Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S. B., Schirrmeister, R. T., y Hutter, F. (2025). Accurate predictions on small data with a tabular foundation model [Predicciones precisas con pocos datos mediante un modelo fundacional tabular]. *Nature, 637*(8045), 319–326. https://doi.org/10.1038/s41586-024-08328-6

Krizhevsky, A., Sutskever, I., y Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks [Clasificación de ImageNet con redes neuronales convolucionales profundas]. En F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 25, pp. 1097–1105). Curran Associates. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

LeCun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning [Aprendizaje profundo]. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

McCarthy, J., Minsky, M. L., Rochester, N., y Shannon, C. E. (1955, 31 de agosto). *A proposal for the Dartmouth summer research project on artificial intelligence* [Propuesta para el proyecto de investigación de verano de Dartmouth sobre inteligencia artificial]. https://www-

formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Mitchell, T. M. (2006). *The discipline of machine learning* [La disciplina del aprendizaje automático] (Informe técnico CMU-ML-06-108). Carnegie Mellon University. https://www.cs.cmu.edu/~tom/pubs/MachineLearningTR.pdf

Watkins, C. J. C. H., y Dayan, P. (1992). Q-learning [Aprendizaje Q]. *Machine Learning, 8*(3–4), 279–292. https://doi.org/10.1007/BF00992698

