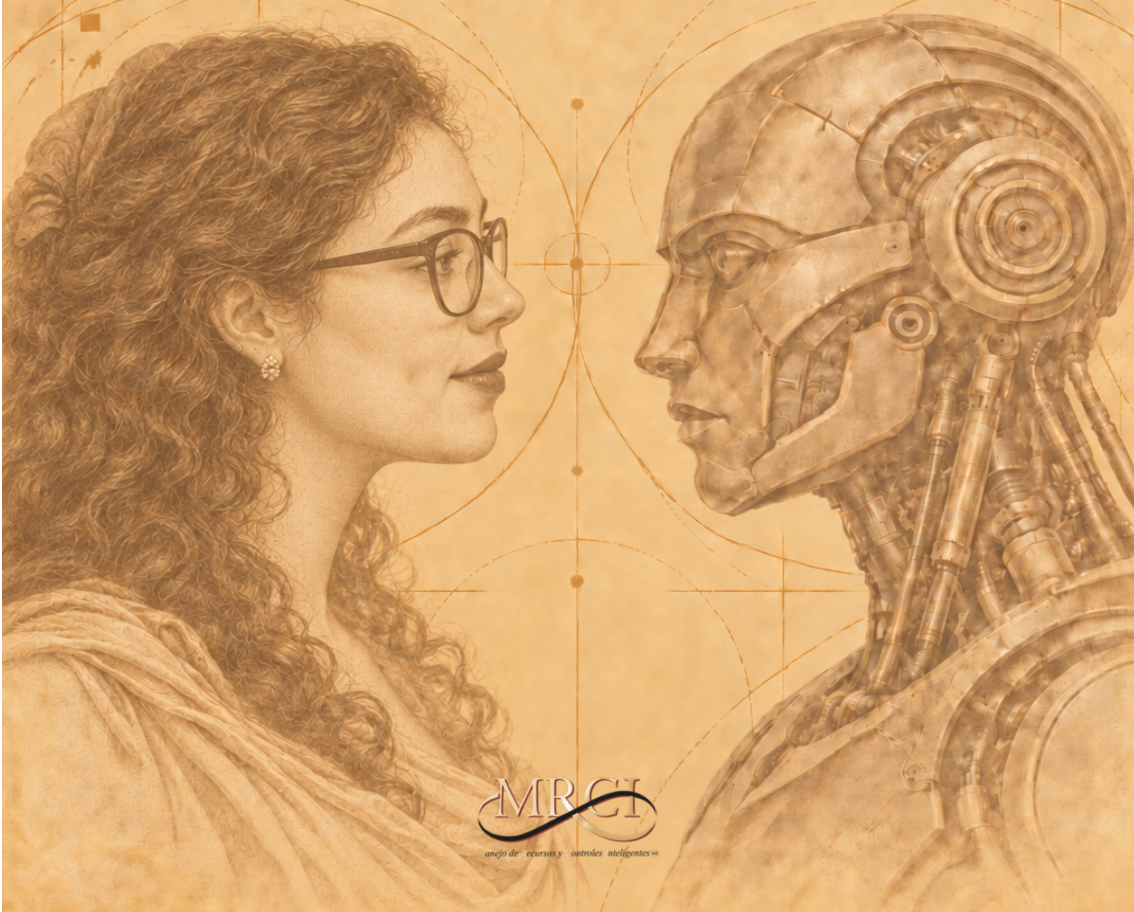
COLECCIÓN

MITOS Y REALIDADES DE LA IA

Entre la promesa y la realidad

¿QUÉ ES LA INTELIGENCIA ARTIFICIAL Y QUÉ QUEDA FUERA DE SU DEFINICIÓN?

Número 1



MRCI

anejo de ecursos y ontroles ntelligentes

DR. JOSÉ ANTONIO PÉREZ RAMOS

¿QUÉ ES LA INTELIGENCIA ARTIFICIAL Y QUÉ QUEDA FUERA DE SU DEFINICIÓN?

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*.

¿QUÉ ES LA INTELIGENCIA ARTIFICIAL Y QUÉ QUEDA FUERA DE SU DEFINICIÓN?

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: ¿Qué es la inteligencia artificial y qué queda fuera de su definición?

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; ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL Y QUÉ QUEDA FUERA DE SU DEFINICIÓN?, por cualquier medio, sin autorización escrita del autor.

PRINTED IN MEXICO
IMPRESO EN MÉXICO

El problema de nombrar una tecnología

Una oficina recibe dos propuestas para ordenar sus expedientes, la primera ofrece un conjunto de reglas que coloca los archivos en carpetas según una clave escrita en el nombre, mientras la segunda promete una inteligencia artificial capaz de interpretar su contenido; antes de comparar precios, alguien tendría que preguntar qué significa interpretar en ese contrato, cuáles documentos admite cada opción y qué ocurre cuando el nombre del archivo contradice lo que aparece en sus páginas.

El ejemplo es hipotético, pero el problema conceptual que plantea resulta preciso: si una palabra sirve para describir procedimientos diferentes, conocer la palabra no basta para conocer el procedimiento; aceptar una oferta por la sola presencia de la expresión inteligencia artificial equivaldría a comprar un vehículo únicamente porque se anuncia como transporte, sin averiguar qué puede transportar ni bajo qué condiciones.

También sería un error inverso rechazar una solución sencilla porque no merece una etiqueta prestigiosa, pues el valor de ordenar correctamente mil archivos no depende de que el programa parezca conversar; una necesidad



bien delimitada puede resolverse con una regla explícita, mientras otra exige tratar variaciones que no estaban enumeradas en las instrucciones iniciales.

Nuestra pregunta central es qué evidencia permite identificar razonadamente un sistema de IA y qué conclusiones quedan fuera de esa identificación; defendemos que deben separarse cuatro cuestiones: qué objeto estamos describiendo, qué mecanismo participa, qué desempeño ha sido observado y qué efectos se autoriza a producir.

Esta separación protege dos exigencias que a veces se presentan como incompatibles: reconocer capacidades técnicas reales y resistir las atribuciones exageradas; no necesitamos negar que un sistema clasifique documentos para advertir que esa clasificación no demuestra una comprensión humana del expediente, del mismo modo que un error particular no prueba que todas sus funciones carezcan de utilidad.

Método y límites de la revisión

Realizamos una revisión documental selectiva de textos originales sobre la delimitación de la IA, su documentación y su evaluación, con fecha de corte del 2 de octubre de 2026; incluimos un antecedente histórico, una formulación del campo, un memorando institucional, un marco de gestión de riesgos y trabajos sobre comprensión lingüística, fichas de modelos y documentación de datos.

La selección se organizó por función argumentativa y no por cantidad de publicaciones: las fuentes históricas permiten reconocer el alcance de una aspiración de investigación, las definiciones contemporáneas ayudan a delimitar sistemas y las propuestas documentales permiten convertir una descripción general en preguntas que un responsable de adopción puede comprobar.

No se trata de una revisión sistemática de toda la literatura ni de una medición experimental, por lo cual no se ofrecen porcentajes de consenso, resultados propios de laboratorio o conclusiones sobre todos los productos disponibles; los casos, las matrices y los criterios operativos que siguen son construcciones analíticas para

examinar consecuencias, no registros de organizaciones reales.

Las citas identifican las proposiciones que proceden de las fuentes consultadas, mientras las recomendaciones se presentan como propuestas derivadas del análisis; esta distinción importa porque una institución puede definir un sistema sin haber recomendado el procedimiento de contratación que aquí construimos a partir de esa definición.

Se conserva además una frontera temática dentro de la colección: este número establece el vocabulario y las precauciones iniciales, pero no sustituye los artículos dedicados al aprendizaje automático, las redes neuronales, los agentes, la privacidad o las normas aplicables; mencionar esos temas resulta necesario para situar un límite, desarrollarlos exhaustivamente aquí impediría entender la pregunta principal.

Un campo no es un producto

La propuesta de Dartmouth de 1955 formuló un programa de investigación para estudiar cómo máquinas podían

abordar aspectos del aprendizaje y otras capacidades asociadas con la inteligencia, además incluyó problemas de lenguaje, abstracción y mejora de procedimientos; debe leerse como una propuesta de trabajo, no como un anuncio de que tales problemas ya hubieran sido resueltos (McCarthy et al., 1955).

Esta diferencia histórica tiene una consecuencia metodológica: una disciplina puede investigar una capacidad antes de haberla realizado de manera satisfactoria, por lo que la presencia de una aspiración en sus textos fundacionales no acredita su cumplimiento en una aplicación posterior; tampoco permite establecer una fecha inevitable para resolverla.

McCarthy (2007) presentó la IA como ciencia e ingeniería de máquinas inteligentes y aclaró que su estudio no tenía que limitarse a métodos biológicamente observables; su formulación sitúa la actividad investigadora y la construcción técnica en un mismo campo, sin convertir la imitación del cerebro humano en requisito de toda solución.



En este artículo empleamos campo para hablar de preguntas, métodos y tradiciones de investigación, mientras reservamos sistema para una configuración que recibe entradas y produce resultados en un entorno; la diferencia evita atribuir a un producto todas las posibilidades que se estudian dentro de la disciplina a la que se lo asocia. Un nombre comercial identifica una oferta, pero puede abarcar versiones, componentes y configuraciones distintas; por ello, la afirmación de que una empresa utiliza IA sigue siendo demasiado general si no sabemos en qué proceso interviene, qué información recibe y quién utiliza sus resultados.

El modelo, por su parte, es una pieza representacional o computacional que participa en ese funcionamiento, aunque una aplicación suele requerir también interfaces, datos, herramientas y controles; basta por ahora con reconocer esa distinción, pues el artículo 04 examinará en detalle por qué evaluar una pieza no equivale a evaluar todo el producto. Un expediente de compra debería conservar esos niveles sin intercambiarlos: el prestigio científico de un campo no es una prueba sobre un proveedor y el desempeño de un modelo no es una garantía sobre la aplicación integrada; cada traslado de

una conclusión a otro nivel necesita un argumento o una comprobación adicional.

Tabla 1. *Objetos que conviene distinguir antes de discutir capacidades*

Objeto	**Pregunta pertinente**	**Confusión que debe evitarse**
Campo de investigación	Qué problemas estudia y con qué enfoques	Tratar sus aspiraciones como prestaciones de un producto
Modelo	Qué transformación permite y bajo qué condiciones	Suponer que por sí solo incluye acceso a archivos y permisos
Sistema utilizado	Qué componentes participan y qué efectos produce	Evaluar sólo la conversación visible
Servicio contratado	Qué se entrega, quién lo opera y con qué límites	Confundir promoción comercial con obligación verificable
Proceso organizacional	Quién decide, revisa y responde por el uso	Atribuir todas las decisiones al componente informático

Nota. Elaboración propia; las distinciones organizan el análisis y no constituyen una taxonomía jurídica.

Una definición funcional y su alcance

La Organisation for Economic Co-operation and Development (OCDE, 2024) describe sistemas basados en máquinas que, para objetivos explícitos o implícitos, infieren desde sus entradas cómo generar predicciones, contenido, recomendaciones o decisiones capaces de influir en entornos físicos o virtuales; reconoce diferencias de autonomía y adaptación posterior al despliegue, además incluye tanto aprendizaje automático como enfoques basados en conocimiento.

La importancia de esa formulación está en lo que permite preguntar: cuáles son las entradas, qué salida se genera y dónde tiene efectos; no obliga a identificar IA exclusivamente con conversación, imagen o robot, por lo que ayuda a reconocer funciones que podrían permanecer invisibles para quien sólo observa una pantalla.

En nuestra lectura, el criterio funcional evita utilizar el parecido humano como filtro principal, pues un sistema

puede producir una recomendación útil sin voz ni rostro; también impide que un rostro digital se convierta, por sí mismo, en evidencia de una capacidad de inferencia.

La definición no debe funcionar como un certificado de calidad: clasificar una herramienta dentro de una categoría no informa cuántas veces falla, frente a qué situaciones o con qué consecuencias; esos datos exigen una evaluación del sistema identificado, con tareas y condiciones que representen el uso previsto.

Tampoco conviene leer la palabra objetivos como prueba de deseos personales del programa, puesto que en una descripción técnica puede referirse a un criterio de optimización, una finalidad incorporada al diseño o una pauta implícita en el funcionamiento; atribuir experiencia o intención humana a partir de esa palabra introduce una conclusión que la descripción funcional no ha establecido.

Para el lector profesional, resulta útil traducir una definición a una ficha de trabajo: entrada aceptada, transformación declarada, salida esperada, destinatario, entorno y limitaciones; esa ficha no descubre automáticamente el



mecanismo interno, pero hace visibles los puntos en los que todavía dependemos de afirmaciones del proveedor.

El memorando de la OCDE (2024, p. 9) precisa además que la inferencia puede referirse a la generación de salidas y, en el contexto de su explicación, a la construcción de un modelo a partir de datos; por ello, interpretar la palabra con un solo significado cotidiano puede producir desacuerdos aparentes entre personas que están describiendo fases distintas.

De esta precisión no deriva que cualquier transformación numérica deba llamarse IA, sino que la clasificación requiere un marco y una descripción de implementación; una fórmula de suma y un procedimiento que construye un clasificador no se vuelven equivalentes porque ambos utilicen números.

La frontera con la automatización ordinaria

Consideremos un programa hipotético que cambia el nombre de cada archivo añadiendo la fecha de recepción; si su funcionamiento consiste únicamente en concatenar una fecha y un identificador conforme a una regla fija, en

el uso operativo de este artículo lo describimos como automatización convencional, sin necesidad de atribuirle una capacidad de IA.

Ahora modifiquemos el encargo: el programa debe leer documentos heterogéneos y estimar si cada uno contiene una reclamación, una factura o una comunicación informal; si utiliza un modelo construido con ejemplos o una representación de conocimiento para resolver esa clasificación, la naturaleza del componente exige una descripción diferente de la simple regla de renombrado.

La interfaz podría ser exactamente la misma en ambos casos, con un botón que dice ordenar, de modo que el botón no sirve para inferir la arquitectura; la diferencia debe buscarse en la documentación, en el flujo de procesamiento y, cuando corresponda, en pruebas que permitan contrastar lo declarado.

También cabe una implementación híbrida: una regla reconoce un identificador exacto y sólo envía los documentos sin identificador a un clasificador; en ese supuesto no resulta esclarecedor preguntar si toda la aplicación es verdaderamente IA, sin indicar que la primera



ruta es una automatización y la segunda incorpora una función inferencial distinta.

Las reglas, por sí solas, tampoco resuelven la frontera: excluir cualquier programa con reglas dejaría fuera enfoques históricamente asociados con la representación de conocimiento, mientras incluir toda instrucción condicional haría que casi cualquier software quedara dentro; necesitamos observar la función de esas reglas y el nivel de descripción elegido.

Proponemos por ello tres resultados posibles para una clasificación inicial: automatización convencional descrita, componente de IA identificado o implementación insuficientemente documentada; la tercera respuesta no es una evasión cuando faltan datos, sino una forma de impedir que la incertidumbre se transforme en una acusación o en una aprobación injustificadas.

Tabla 2. *Casos hipotéticos y documentación necesaria para clasificarlos*

Caso	Clasificación inicial	Evidencia que falta o debe conservarse
Suma mediante fórmula	Automatización convencional	Fórmula, entradas y pruebas aritméticas
Respuesta fija por botón	Automatización convencional en esa función	Tabla de correspondencias y alcance del menú
Clasificación aprendida de texto	Componente de IA identificado si se comprueba el mecanismo	Modelo, versión, finalidad y resultados de evaluación
Inferencia sobre conocimiento	Posible enfoque de IA basada en conocimiento	Reglas, representación y forma de derivar conclusiones
Ordenamiento opaco	Información insuficiente	Documentación técnica y unidad de análisis
Combinación de componentes	Sistema híbrido que debe descomponerse	Recorrido de cada dato y componente responsable

Nota. Elaboración propia; las clasificaciones se limitan a las implementaciones descritas y no sustituyen el análisis de una norma concreta.



Inferencia sin atribución de vida interior

Una operación puede recibir el nombre técnico de inferencia sin que se haya demostrado que el sistema experimenta dudas, comprende su situación o mantiene intenciones personales; para evitar el salto, conviene distinguir la descripción del procesamiento y la interpretación filosófica que una persona formula al observar sus resultados.

Bender y Koller (2020) sostienen, en un trabajo de posición, que aprender exclusivamente la forma lingüística no basta para adquirir significado en el sentido comunicativo que delimitan; su argumento cuestiona inferencias sobre comprensión a partir del éxito en tareas de lenguaje, pero no debe convertirse en un diagnóstico empírico universal de cualquier sistema posterior o multimodal.

La precisión anterior obliga a reconocer el tipo de evidencia: un argumento conceptual no es un experimento que mida todas las implementaciones futuras y una prueba de desempeño tampoco resuelve automáticamente una controversia sobre significado; ambas contribuciones

pueden ser relevantes si se conserva la pregunta que cada una pretende responder.

Supongamos que un asistente redacta una frase de condolencia adecuada, el resultado observable permite examinar el vocabulario, la pertinencia y el respeto hacia la persona destinataria; concluir que el sistema sintió dolor requeriría un argumento adicional, mientras negar la utilidad de la frase por no haber probado ese sentimiento confundiría otra vez dos cuestiones diferentes.

De manera semejante, una explicación coherente de un problema no demuestra que el sistema haya realizado las operaciones que describe; si necesitamos comprobar un cálculo, deberíamos verificar las entradas y la operación, no aceptar únicamente una narración convincente sobre cómo supuestamente se obtuvo.

La disciplina consiste en ajustar el verbo a la evidencia: puede afirmarse que el sistema produjo una respuesta, identificó un fragmento o ejecutó una instrucción cuando existen registros suficientes; expresiones como comprendió plenamente, quiso ayudar o decidió

libremente añaden propiedades que no quedan acreditadas por esos mismos registros.

Esto no prohíbe toda metáfora en la divulgación, pero exige indicar cuándo se utiliza una abreviatura funcional; decir que una aplicación recuerda un dato puede resultar comprensible si después se explica dónde se conserva, durante cuánto tiempo y quién puede acceder a él, mientras dejar la metáfora sin explicación favorece expectativas equivocadas.

La discusión detallada sobre conciencia corresponde a otro número de la colección, por ahora basta una regla epistemológica: ninguna afirmación sobre experiencia subjetiva debe presentarse como consecuencia necesaria de una definición operativa de IA; tampoco debe usarse la ausencia de una prueba específica como demostración de una imposibilidad universal.

Capacidad técnica y autorización

Una aplicación puede redactar un mensaje sin tener permiso para enviarlo, del mismo modo que puede disponer de una conexión autorizada y producir una

instrucción equivocada; la capacidad de generar una salida, la posibilidad técnica de ejecutar una acción y la autorización para hacerlo pertenecen a planos distintos.

En un ejemplo hipotético, el usuario pide preparar una lista de destinatarios y revisar el texto de una invitación, mientras el sistema tiene acceso a una herramienta de envío; la existencia de esa herramienta no transforma el encargo de preparación en permiso para contactar a todos los destinatarios.

Si el sistema se limita a sugerir el mensaje, la revisión humana puede corregir nombres, horarios y tono antes de su circulación; si el mismo texto se envía automáticamente, las consecuencias cambian aunque las palabras sean idénticas, porque la salida ya ha intervenido en una relación externa.

Por ello proponemos documentar una frontera de ejecución: qué resultados quedan como borradores, cuáles requieren aprobación y qué operaciones pueden realizarse bajo una autorización previa delimitada; la frontera debe describirse en términos de acciones,



destinatarios y datos, en lugar de descansar en una expresión ambigua como plena autonomía.

El permiso tampoco equivale a competencia: autorizar a un sistema a clasificar un expediente no aporta evidencia de que lo haga correctamente; la decisión de delegar debería combinar una prueba pertinente con una delimitación de las consecuencias aceptables y un mecanismo real de corrección.

Esta distinción ayuda a interpretar las demostraciones comerciales de agentes: una secuencia de acciones bien presentada puede mostrar un funcionamiento particular, pero todavía habría que comprobar qué permisos recibió, qué información estaba preparada y qué ocurrió con los intentos fallidos; de otro modo, la demostración permite observar un resultado sin entender sus condiciones de posibilidad.

No necesitamos resolver aquí toda la teoría de los agentes para identificar el problema, pues el análisis básico ya muestra que una etiqueta de autonomía no reemplaza el registro de quién autorizó qué; la responsabilidad organizacional comienza precisamente donde se

establecen esas condiciones y se decide utilizar los resultados.

Lo que una demostración permite concluir

Una demostración aporta evidencia sobre los casos efectivamente observados, aunque su fuerza depende de cómo se seleccionaron y registraron; ver tres respuestas correctas permite afirmar que esas tres respuestas fueron correctas, pero no basta para estimar el comportamiento ante todas las entradas que una organización utilizará después.

El AI RMF 1.0 de Tabassi (2023) caracteriza los sistemas de IA como sociotécnicos y relaciona validez y confiabilidad con su contexto de uso; también organiza la gestión de riesgos mediante las funciones de gobernar, contextualizar, medir y gestionar, dentro de un marco voluntario que no debe presentarse como una certificación automática.

A partir de ese marco proponemos una separación práctica entre evidencia de existencia, evidencia de desempeño y evidencia de adecuación: la primera acredita que una

función puede ejecutarse, la segunda describe resultados bajo condiciones establecidas y la tercera examina si esos resultados satisfacen el propósito y las restricciones del proceso que se desea apoyar.

Una función de extracción puede existir y tener un buen resultado en documentos nítidos, pero resultar inadecuada para un archivo compuesto principalmente por fotografías inclinadas; no hay contradicción si se distingue el conjunto evaluado del conjunto de uso, mientras sí habría un problema argumentativo si se ocultara esa diferencia.

También debe separarse una tasa global del costo de los errores: una aplicación que confunde dos carpetas internas puede requerir una corrección administrativa, mientras una que atribuye a una persona una obligación inexistente puede producir consecuencias más difíciles de reparar; el mismo porcentaje no expresa por sí solo esa diferencia.

La ficha del modelo propuesta por Mitchell et al. (2019) incluye usos previstos, usos fuera de alcance, evaluación y factores que pueden afectar el desempeño; su valor documental consiste en hacer explícitas condiciones que

podrían quedar ocultas detrás de un único resultado agregado, sin que completar una ficha garantice la veracidad de todo lo declarado.

En nuestra propuesta de adopción, conviene contrastar esa documentación con casos propios y conservar las discrepancias: si una limitación estaba anunciada y aparece en la prueba, confirma una restricción relevante; si surge un fallo no documentado, exige investigar su origen antes de ampliar el uso.

Tabla 3. *Clases de evidencia y alcance de las conclusiones.*

Evidencia disponible	Conclusión defendible	Conclusión que todavía no se justifica
Captura de una respuesta	Existe una salida visible en ese ejemplo	La respuesta es correcta o reproducible
Demostración registrada	La función operó en las condiciones registradas	Funciona en toda situación equivalente en apariencia
Prueba con casos y criterios explícitos	Se observaron resultados en un conjunto identificable	El mismo resultado se mantendrá ante cualquier cambio

Evidencia disponible	Conclusión defendible	Conclusión que todavía no se justifica
Evaluación en el proceso real	Hay evidencia pertinente para ese contexto	La herramienta puede trasladarse sin revisión a otro proceso
Registro sostenido de operación	Pueden analizarse incidentes y variaciones durante ese periodo	No habrá fallos futuros ni hace falta supervisión

Nota. Elaboración propia; la tabla expresa límites lógicos de la inferencia y no una escala oficial de certificación.

Caso de contratación para expedientes

Desarrollamos ahora un caso hipotético de principio a fin: una oficina desea localizar cláusulas de terminación en contratos y un proveedor afirma que su asistente comprende expedientes completos; el objetivo de la oficina no es sostener una conversación agradable, sino encontrar la cláusula pertinente, identificar su versión y advertir si un anexo modifica sus efectos.

La primera tarea consiste en reformular la promesa como desempeño observable: dado un conjunto de documentos autorizado, el sistema debe señalar el archivo y el fragmento que contienen la regla de terminación, además de indicar cuando el material es insuficiente; no se exige resolver por sí solo todas las consecuencias jurídicas de la cláusula.

Esta reformulación reduce un riesgo de alcance: si el proveedor muestra una respuesta correcta a una pregunta general sobre contratos, esa demostración no comprueba todavía que pueda distinguir el contrato firmado de un borrador; la oficina necesita evaluar relaciones documentales, no únicamente conocimiento general sobre un tema.

Proponemos un conjunto de prueba con cuatro familias: expedientes completos y consistentes, expedientes con una modificación posterior, expedientes incompletos y documentos ajenos al asunto; cada familia plantea una dificultad diferente, por lo que el registro debe conservar sus resultados por separado.



En el primer grupo importa localizar correctamente la información disponible, en el segundo reconstruir qué texto modifica a cuál, en el tercero reconocer que falta una pieza y en el cuarto evitar una atribución que no corresponde; reunir todos los aciertos en una sola cifra podría ocultar que el sistema nunca reconoce la insuficiencia de información.

Imaginemos, sólo como ejercicio aritmético reducido a dos familias, que se utilizan cuarenta expedientes y que treinta reciben una respuesta aceptable según criterios definidos previamente; el resultado agregado es 30 dividido entre 40, equivalente al 75 %, pero no sería correcto llamarlo confiabilidad universal de la herramienta.

Supongamos además que los treinta aciertos pertenecen a expedientes completos y que los diez restantes son precisamente los incompletos; la misma cifra agregada ahora revela una limitación material para una oficina donde reconocer ausencias es parte esencial de la tarea, sin necesidad de afirmar que toda la tecnología sea inútil.

El responsable debería conservar la versión del sistema, la configuración, los archivos proporcionados y las



respuestas, además de la justificación de cada evaluación; si posteriormente se cambia el mecanismo de recuperación documental, la prueba anterior puede servir como antecedente, pero no acredita automáticamente el comportamiento de la nueva configuración.

Un fallo de este caso admite explicaciones distintas: el archivo no se cargó, la búsqueda no lo recuperó, la lectura omitió una negación o la tarea no exigía considerar anexos; decir solamente que la IA no entiende impide distinguir la causa y, por tanto, dificulta corregir el proceso.

La decisión de compra podría ser contratar únicamente una función acotada de localización con revisión profesional, ampliar la prueba o conservar el procedimiento anterior; ninguna de esas decisiones se deriva de la etiqueta IA, sino de la relación entre evidencia disponible, costo, utilidad y consecuencias de un error.

Caso de un calendario que no necesita adivinar

Consideremos una segunda oficina hipotética que organiza entregas internas con fechas definidas por acuerdos escritos; dispone de un calendario validado y



reglas explícitas para mover una entrega cuando coincide con un día no laborable, sin que el ejemplo pretenda describir plazos legales de una jurisdicción.

La función principal puede resolverse mediante operaciones deterministas sobre fechas y un calendario identificado; añadir un modelo generativo no mejora necesariamente ese cálculo, aunque sí podría ayudar a redactar una explicación para la persona responsable de entregar el trabajo.

Proponemos separar ambos componentes: el cálculo produce la fecha y el generador recibe esa fecha como dato que debe conservar; una validación posterior compara la fecha incluida en el mensaje con la salida del cálculo, de manera que la redacción no se convierta en una nueva fuente independiente del dato.

Si el mensaje menciona una fecha distinta, el sistema no debería justificarla por la fluidez del texto, sino bloquear el envío o devolver el borrador para corrección; el control es específico y verificable porque compara un campo definido, en vez de pedir genéricamente que la herramienta sea más inteligente.

El ejemplo permite reconocer una ventaja de las arquitecturas híbridas sin presumir que siempre sean superiores: cada componente se utiliza para una función cuya evidencia puede examinarse, mientras la integración añade sus propios puntos de fallo; por ello, también habría que probar que el dato calculado llegue completo y sin transformación indebida al componente de redacción.

La automatización convencional conserva aquí un lugar central, no como una tecnología inferior que debe desaparecer, sino como un procedimiento adecuado para una tarea formalizada; la decisión razonable consiste en añadir complejidad sólo cuando produce una mejora comprobable o resuelve una dificultad que el procedimiento simple no atendía.

Una organización podría incluso concluir que no necesita el componente generativo si utiliza pocos mensajes y una plantilla satisface la necesidad; tal conclusión no demuestra atraso tecnológico, sino correspondencia entre el problema, el costo de mantener la solución y el beneficio que realmente se espera obtener.



Datos y condiciones que la etiqueta oculta

Un sistema puede estar bien descrito y aun así recibir información inadecuada para su tarea, por lo que identificar la IA no agota la investigación sobre los datos; antes de interpretar una salida conviene saber de dónde proviene la entrada, qué transformaciones sufrió y qué partes del fenómeno no quedaron representadas.

Gebru et al. (2021) proponen documentar conjuntos de datos mediante fichas que incluyan motivación, composición, recolección y usos recomendados, entre otros aspectos; esta propuesta sirve para distinguir la existencia de datos de su aptitud para una finalidad, sin asumir que una ficha resuelva por sí misma problemas de calidad, representación o legitimidad.

En un ejemplo hipotético de clasificación de solicitudes, reunir miles de registros no garantiza que aparezcan los tipos de solicitud que interesan a una oficina nueva; si los ejemplos provienen de una actividad diferente, la cantidad puede coexistir con una ausencia precisamente donde se necesita capacidad de respuesta.



También importa qué se llama acierto: si las categorías del archivo histórico reflejan decisiones inconsistentes, reproducirlas fielmente podría mantener la inconsistencia; antes de valorar el modelo, la organización tendría que revisar la definición de las categorías y la forma en que se asignaron los ejemplos.

No afirmamos que todo dato histórico sea defectuoso, sino que su antigüedad, tamaño o procedencia institucional no sustituyen la pregunta por su pertinencia; una evaluación responsable necesita distinguir errores de registro, cambios de significado y diferencias entre el contexto original y el uso propuesto.

Un caso particularmente ilustrativo sería una categoría llamada urgente sin criterio explícito, puesto que dos personas podrían utilizarla con significados distintos; un sistema que aprende de esas etiquetas puede ofrecer una respuesta consistente con parte del archivo y, al mismo tiempo, resultar difícil de evaluar porque el criterio humano nunca se aclaró.

La corrección comienza entonces por el proceso de trabajo: definir qué constituye urgencia, quién puede



modificar la categoría y cómo se resuelven discrepancias; sólo después tiene sentido preguntar cuánto contribuye la herramienta a aplicar esa definición en condiciones reales.

La inferencia central de este apartado es que la calidad no reside exclusivamente dentro del componente de IA, pues depende también de entradas, categorías y decisiones organizacionales; eso explica por qué sustituir un modelo no necesariamente corrige una tarea mal definida.

Mitos que alteran las decisiones

El primer mito consiste en afirmar que todo programa automático es IA, una generalización que elimina la diferencia entre ejecutar una fórmula y utilizar procedimientos inferenciales más amplios; su efecto práctico es impedir que una organización compare soluciones por su mecanismo y termine pagando por una denominación sin identificar el trabajo adicional que realiza.

El segundo mito sostiene que toda IA aprende durante cada interacción, aunque una respuesta que incorpora un dato recién proporcionado no demuestra por sí misma una



modificación persistente del modelo; habría que identificar si el dato permanece sólo en el contexto, se guarda en una base externa o interviene en un proceso de actualización documentado.

El tercer mito equipara lenguaje convincente y verdad: una explicación puede ser clara y contener una atribución documental falsa, mientras una salida poco elegante puede señalar correctamente que falta información; evaluar estilo y evaluar sustento exige criterios diferentes, aunque ambos puedan importar para el uso final.

El cuarto mito supone que un error elimina la pertenencia a la categoría IA, cuando la clasificación técnica y la evaluación de desempeño responden preguntas distintas; un sistema defectuoso puede seguir siendo un sistema de IA, del mismo modo que reconocerlo no obliga a considerarlo apto para una tarea de consecuencias importantes.

El quinto mito sostiene que la revisión humana siempre resuelve los riesgos, pero una firma final sin tiempo, información o facultades de corrección no permite saber si hubo revisión efectiva; nuestra propuesta exige que la



persona revisora disponga de las fuentes relevantes y pueda detener el proceso cuando encuentra una discrepancia.

El sexto mito identifica una conexión a internet con conocimiento completo y actualizado, aunque la conexión sólo establece una posibilidad de acceso; para sostener una afirmación concreta sería necesario conocer qué fuente se consultó, cuándo se consultó y si el pasaje encontrado respalda realmente la respuesta.

El séptimo mito considera que un sistema sin apariencia humana carece de inteligencia artificial, una conclusión que confunde interfaz con función; una herramienta de clasificación silenciosa puede merecer más atención que un avatar conversacional si sus resultados intervienen directamente en decisiones de la organización.

Estos mitos comparten una estructura: sustituyen una investigación concreta por una asociación verbal, de modo que la corrección no consiste únicamente en memorizar la frase opuesta; consiste en recuperar la pregunta omitida y establecer qué observación permitiría responderla sin exagerar su alcance.

Un procedimiento de identificación y uso

Proponemos comenzar con una descripción de la tarea en lenguaje ordinario, evitando verbos demasiado amplios como entender todo o resolver cualquier problema; una formulación útil indica qué entrada se proporciona, qué resultado se espera y qué condiciones harían inaceptable ese resultado.

El segundo paso es delimitar la unidad de análisis: puede ser un componente, una aplicación completa o un proceso con participación humana; esa elección debe quedar escrita porque permite saber dónde termina la evidencia disponible y evita trasladar una conclusión sobre una pieza a todas las demás.

Después se identifica el mecanismo declarado y su grado de comprobación, distinguiendo documentación técnica, afirmación comercial y observación propia; no siempre será posible inspeccionar el código o los parámetros, pero sí puede registrarse qué información falta y cuánto depende de ella la decisión.



La siguiente operación consiste en buscar una alternativa plausible: una regla, una plantilla, una búsqueda documental o una tarea humana asistida; comparar alternativas no significa favorecer siempre la más sencilla, sino impedir que la categoría tecnológica sustituya la evaluación del problema.

Una vez definida la función, se construyen casos positivos, negativos y de frontera, incluidos aquellos en los que la respuesta correcta es no disponer de información suficiente; los criterios deben fijarse antes de observar las salidas para reducir el riesgo de ajustar la evaluación a lo que la herramienta ya produjo.

El procedimiento continúa con la delimitación de consecuencias y permisos: qué salida puede utilizarse directamente, cuál requiere revisión y cuál no debe producir efectos externos; una prueba satisfactoria de redacción no autoriza por sí sola enviar comunicaciones o modificar registros.

Finalmente se documentan las condiciones de continuidad: qué cambio obliga a repetir pruebas, quién recibe los incidentes y cómo se regresa a un procedimiento

alternativo; una implementación sin salida prevista convierte la dependencia técnica en una decisión que puede resultar difícil de revisar después.

Tabla 4. *Registro mínimo propuesto para una decisión de adopción*

Elemento	Contenido que se registra	Utilidad para revisar la decisión
Tarea	Entrada, salida y criterios de aceptación	Evita evaluar una promesa indefinida
Objeto	Componente, aplicación o proceso completo	Delimita el alcance de las pruebas
Evidencia	Documentación, versión y casos observados	Permite distinguir afirmaciones y comprobaciones
Límites	Situaciones excluidas y datos insuficientes	Impide extender el uso sin justificación
Permisos	Acciones autorizadas y revisiones necesarias	Separa capacidad y facultad de actuar
Continuidad	Responsable, incidentes y alternativa	Permite corregir o suspender sin improvisación

Nota. Elaboración propia; se propone como registro de análisis, no como sustituto de requisitos jurídicos o técnicos sectoriales.

Objeciones y alcance de nuestra propuesta

Una objeción razonable sostiene que tanta distinción vuelve difícil explicar la IA a personas no especialistas, mientras una frase breve facilita la conversación; aceptamos que una introducción necesita síntesis, pero la brevedad no exige ocultar los límites, pues puede decirse que se trata de sistemas que producen resultados mediante determinados procedimientos y añadir que su capacidad depende de la tarea y de las pruebas disponibles.

Otra objeción señala que los límites entre automatización e IA son discutidos, por lo que cualquier clasificación podría parecer arbitraria; nuestra respuesta no es negar esa dificultad, sino hacer explícito el marco utilizado y conservar una categoría de información insuficiente cuando la implementación no permite una conclusión razonable.



También podría afirmarse que lo único importante es si la herramienta funciona, con independencia de su definición; esa postura contiene una intuición útil para evitar debates nominales, pero deja incompleta la explicación de por qué funciona, cuándo falla y qué documentación permite trasladarla a otro contexto.

La definición ayuda a formular hipótesis y a pedir evidencia, mientras el desempeño permite decidir si una función satisface una necesidad; ninguno reemplaza al otro, por lo que una evaluación madura combina descripción y prueba sin convertirlas en una disputa sobre cuál merece prioridad absoluta.

Una cuarta objeción cuestiona la posibilidad de evaluar un sistema cuyo proveedor no revela todos sus detalles; la falta de acceso no vuelve imposible cualquier prueba, pero sí limita las conclusiones sobre el mecanismo y puede justificar una restricción de uso cuando esas conclusiones resultan indispensables para el riesgo asumido.

Por último, sería excesivo exigir a toda aplicación doméstica el expediente de una operación crítica, de modo que proponemos proporcionalidad: cuanto mayor sea la

consecuencia posible y más difícil resulte corregirla, más sólida debería ser la evidencia y más cuidadosa la delimitación de permisos; esa proporcionalidad exige juicio, no un número universal aplicable a cualquier situación.

Conclusión

La inteligencia artificial admite una descripción funcional que distingue el campo de investigación, el componente técnico y el sistema utilizado; los casos examinados muestran que una interfaz puede ocultar mecanismos diferentes y que una demostración correcta no acredita por sí sola la adecuación de una herramienta a cualquier contexto. La definición tampoco garantiza verdad, conciencia, conveniencia o responsabilidad jurídica, mientras la autorización para actuar exige una decisión distinta de la evaluación de capacidad; una solución convencional puede ser suficiente y añadir IA sólo se justifica por su contribución al problema identificado.

La pregunta profesional es qué función realiza el sistema, mediante qué componentes, con qué evidencia y dentro de qué límites; responderla permite decidir sin confundir lo observado con lo atribuido, conservar las incertidumbres relevantes y establecer cuándo conviene ampliar las pruebas, corregir el proceso o mantener una alternativa.

Referencias

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data [Hacia la comprensión del lenguaje natural: Significado, forma y comprensión en la era de los datos]. En D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185–5198). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). *Datasheets for datasets* [Fichas de documentación para conjuntos de datos] (Versión 8) [Manuscrito de autor]. arXiv. https://doi.org/10.48550/arXiv.1803.09010

McCarthy, J. (2007). *What is artificial intelligence?* [¿Qué es la inteligencia artificial?]. Stanford University. https://www-formal.stanford.edu/jmc/whatisai/

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955, 31 de agosto). *A proposal for the Dartmouth summer research project on artificial intelligence* [Una propuesta para el proyecto de investigación de verano de Dartmouth sobre inteligencia artificial]. Stanford University. https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). *Model cards for model reporting* [Fichas de modelos para documentar modelos] (Versión 2) [Manuscrito de autor]. arXiv. https://doi.org/10.48550/arXiv.1810.03993

Organisation for Economic Co-operation and Development. (2024). *Explanatory memorandum on the updated OECD definition of an AI system* [Memorando explicativo sobre la definición actualizada de un sistema de IA de la OCDE] (OECD Artificial Intelligence Papers, núm. 8). https://doi.org/10.1787/623da898-en

Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* [Marco de gestión de riesgos de inteligencia artificial] (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

