Comprender la inteligencia artificial y los LLM

Comprender, evaluar y encuadrar los modelos de lenguaje

© 2026 ASCALEA. Tous droits réservés.

Comprender la inteligencia artificial y los LLM

Comprender, evaluar y encuadrar los modelos de lenguaje

ASCALEA — MYGASPOZ

© 2026 ASCALEA. Tous droits réservés.

Table des matières

© 2026 ASCALEA. Tous droits réservés.

CHAPITRE 1

Objeto y alcance del libro

El término inteligencia artificial (IA) se emplea en contextos muy diferentes: investigación, software de consumo, automatización empresarial, análisis de datos o creación de contenidos. Esta diversidad puede generar una confusión inicial: una aplicación que responde a preguntas, un sistema que clasifica imágenes y un programa que recomienda un producto pueden calificarse todos como «IA», sin basarse en el mismo modelo ni perseguir el mismo objetivo.

En este libro, la IA designa un conjunto de sistemas informáticos capaces de ejecutar tareas asociadas con la percepción, la predicción, el razonamiento práctico, la clasificación o la generación de contenido. El foco principal recae en los sistemas basados en el aprendizaje automático —también llamado machine learning (ML)— y, más particularmente, en los modelos de lenguaje utilizados para comprender o producir texto. El vocabulario del ML distingue, entre otros elementos, los datos, las características, los modelos, el entrenamiento y la inferencia: estos elementos constituyen el marco básico necesario para describir el funcionamiento de un sistema aprendido a partir de datos [GOOGLE-ML-GLOSSARY].

Por lo tanto, el alcance abarca:

- las nociones que permiten distinguir la IA de un software clásico;
- el papel de un modelo y de sus datos de entrenamiento;
- las especificidades de un modelo de lenguaje;
- los grandes modelos de lenguaje, comúnmente designados por el acrónimo inglés Large Language Model (LLM);
- el paso de un modelo a una aplicación utilizable;
- los límites de comprensión que deben mantenerse al interpretar una respuesta producida por un LLM.

© 2026 ASCALEA. Tous droits réservés.

El objetivo no es presentar los LLM como interlocutores dotados de comprensión humana, ni reducir la IA a una simple interfaz conversacional. Consiste en construir un vocabulario preciso para evaluar qué hace realmente un sistema, de qué depende su resultado y qué preguntas deben plantearse antes de confiarle

> Referencia de lectura — tres niveles que no deben confundirse - La IA es el ámbito general y el conjunto de enfoques implicados. - El modelo es un componente aprendido o parametrizado que produce una salida a partir de una entrada. - La aplicación es el producto o servicio que organiza el uso del modelo: interfaz, instrucciones, datos proporcionados, herramientas conectadas y reglas de funcionamiento.

Glosario inicial de términos esenciales

Las siguientes definiciones no sustituyen los detalles técnicos de los capítulos posteriores. Establecen referencias de vocabulario estables.

Término	Definición útil para este libro	No confundir con
Inteligencia artificial (IA)	Término general que abarca enfoques que permiten a sistemas informáticos realizar determinadas tareas de predicción, decisión, reconocimiento o generación.	Un único producto, un único método o solo los asistentes conversacionales.
Aprendizaje automático (ML)	Subconjunto de métodos en los que un sistema aprende relaciones a partir de datos, en lugar de estar enteramente descrito mediante reglas escritas una por una. Las nociones de conjunto de datos, entrenamiento y predicción son centrales en él [GOOGLE-ML-GLOSSARY].	Toda la IA, incluidos los sistemas basados puramente en reglas explícitas.
Aprendizaje profundo	Familia de métodos de aprendizaje automático basada en redes neuronales que cuentan con varias capas. Esta familia se asocia habitualmente con los modelos modernos que procesan texto, imágenes o audio [GOOGLE-ML-GLOSSARY].	Un sinónimo de toda forma de aprendizaje automático.
Datos	Información utilizada para entrenar, evaluar o alimentar un sistema. Sus características, su cobertura y su calidad influyen directamente en los resultados obtenidos [GOOGLE-DATA-CHARACTERISTICS].	Una garantía de calidad o de neutralidad.
Modelo	Representación computacional aprendida a partir de datos o configurada para transformar entradas en salidas. En ML, el modelo produce, entre otros resultados, predicciones durante la inferencia [GOOGLE-ML-GLOSSARY].	Una aplicación completa o una base de datos.
Entrenamiento	Fase durante la cual un modelo ajusta sus parámetros a partir de datos para reducir una medida de error o mejorar una tarea determinada [GOOGLE-ML-GLOSSARY].	El uso cotidiano del modelo por parte de un usuario.
Inferencia	Fase de uso de un modelo entrenado para producir una predicción o una salida a partir de una nueva entrada [GOOGLE-ML-GLOSSARY].	El entrenamiento del modelo.
Modelo de lenguaje	Modelo que procesa secuencias lingüísticas y produce, entre otros resultados, probabilidades o continuaciones de texto.	Una gramática formal, un diccionario o un motor de búsqueda.
Gran modelo de lenguaje (LLM)	Modelo de lenguaje de gran tamaño, entrenado con vastos corpus textuales y utilizado para generar o transformar texto. Los LLM modernos se basan con frecuencia en la arquitectura Transformer [GOOGLE-LLM-INTRO].	Toda aplicación de mensajería, todo motor de búsqueda o toda IA.
Transformer	Arquitectura de red neuronal introducida en torno al mecanismo de atención, diseñada para procesar secuencias sin apoyarse en la recurrencia como mecanismo principal [GOOGLE-TRANSFORMER-PAPER].	Un LLM en sí mismo: un Transformer es una arquitectura, mientras que un LLM es un tipo de modelo construido según determinadas elecciones de arquitectura y entrenamiento.
Token o token	Unidad manipulada por el modelo durante el procesamiento del texto. Un proceso de tokenización convierte el texto en identificadores de tokens que el modelo puede procesar [HF-TOKENIZER].	Necesariamente una palabra completa: un token puede corresponder a una palabra, una parte de palabra, un signo u otra unidad según la segmentación utilizada.
Instrucción o prompt	Texto y contexto proporcionados al modelo para orientar su respuesta. La formulación, el contexto y la estructuración de la instrucción influyen en el resultado [ANTHROPIC-PROMPT-BEST-PRACTICES].	Una orden determinista que garantice por sí sola una respuesta exacta.
Generación	Producción secuencial de nuevos tokens a partir del contexto disponible. Los parámetros de generación pueden modificar la manera en que se seleccionan las salidas [HF-GENERATION-CONFIG].	Una prueba de que el contenido generado es verdadero, está respaldado por fuentes o es adecuado para un uso determinado.

© 2026 ASCALEA. Tous droits réservés.

Un mapa conceptual mínimo

La siguiente cadena ofrece una primera representación del funcionamiento general. Está simplificada deliberadamente: sirve para situar los objetos, no para describir cada detalle matemático.

``text Intelligence artificielle Approches à base de règles Apprentissage automatique Données > entraînement > modèle inférence
Apprentissage profond ¼ sortie produite Transformers Modèles de langage LLM Application : interface, instruction, outils, données et règles d'usage ``

Los datos no son un simple material accesorio. Participan en la definición efectiva de un sistema aprendido: su calidad, su representatividad y los desequilibrios que contienen pueden afectar a los comportamientos observados. Los cursos de Google destacan, en particular, que las características de los datos pueden crear dificultades de generalización y que los sesgos pueden introducirse o amplificarse en distintas etapas de un sistema de ML [GOOGLE-DATA-CHARACTERISTICS] [GOOGLE-FAIRNESS-BIAS].

Del mismo modo, la expresión «el modelo ha respondido» es práctica, pero incompleta. En un producto real, la salida visible depende por lo general de un conjunto más amplio: el modelo elegido, las instrucciones del sistema o del usuario, el texto de la conversación, los parámetros de generación, los documentos que puedan haberse recuperado y, a veces, herramientas externas. Las bibliotecas de procesamiento del lenguaje ofrecen así pipelines distintos para diversas tareas, lo que ilustra que un modelo puede integrarse según diferentes configuraciones de uso [HF-PIPELINES] [HF-TASKS-EXPLAINED].

Distinciones conceptuales que deben mantenerse a lo largo de la lectura

IA, aprendizaje automático y aprendizaje profundo: una relación de inclusión

Un error frecuente consiste en utilizar estas tres expresiones como si fueran intercambiables. Sin embargo, designan niveles diferentes.

- La IA es la categoría más amplia en el marco de este libro.
- El aprendizaje automático es una forma de construir determinados sistemas de IA a partir de datos.
- El aprendizaje profundo es una familia particular de métodos de ML, basada en redes neuronales de varias capas [GOOGLE-ML-GLOSSARY].

Esta relación puede retenerse de forma sencilla: todos los sistemas de aprendizaje profundo estudiados aquí pertenecen al ML; los sistemas de ML pertenecen al amplio campo de la IA; pero lo contrario no es sistemáticamente cierto.

Modelo de lenguaje y LLM: una relación de especialización

Un modelo de lenguaje trabaja con el lenguaje en forma de secuencias de tokens. Según su diseño, puede estimar la continuación probable de un texto, generar una continuación, reformular, resumir, traducir o contribuir a otra tarea lingüística.

Un LLM es un modelo de lenguaje caracterizado por una escala importante de datos, parámetros y capacidad de cálculo. Sin embargo, no debe deducirse de ello que un LLM esté automáticamente adaptado a cada

© 2026 ASCALEA. Tous droits réservés.

tarea ni que su resultado sea exacto por defecto. La generación textual corresponde a una selección progresiva de tokens, cuyo comportamiento depende del contexto recibido y de los parámetros de generación [GOOGLE-LLM-INTRO] [HF-TEXT-GENERATION] [HF-GENERATION-CONFIG].

Por lo tanto, las dos formulaciones siguientes no tienen el mismo significado:

- «Esta aplicación utiliza un LLM» describe una elección de componente principal.
- «Esta aplicación es una IA» describe una categoría general, mucho menos precisa.

Arquitectura, modelo y producto: tres objetos diferentes

La palabra «modelo» puede designar, en una conversación corriente, un producto completo. Para analizar correctamente un sistema, deben separarse al menos tres objetos.

Objeto	Pregunta que debe plantearse	Ejemplo de función
Arquitectura	¿Cómo está estructurada la familia de modelos?	El Transformer proporciona una arquitectura basada en la atención [GOOGLE-TRANSFORMER-PAPER].
Modelo entrenado	¿Qué ha aprendido a hacer a partir de sus datos y de su entrenamiento?	Producir una salida a partir de una secuencia de tokens.
Aplicación	¿Cómo se hace útil el modelo en un contexto determinado?	Añadir una interfaz, instrucciones, controles, documentos o herramientas.

Esta distinción evita diagnósticos imprecisos. Si una respuesta es insatisfactoria, la causa no se encuentra necesariamente solo en el modelo: puede proceder de la instrucción, del contexto transmitido, de los documentos disponibles, de la parametrización de la generación o de la integración de la aplicación.

Datos de entrenamiento, datos de contexto y datos de salida

También deben separarse tres categorías de datos.

1. Los datos de entrenamiento sirven para ajustar el modelo antes de su despliegue o uso.
2. Los datos de contexto son la información proporcionada al modelo durante una interacción: una pregunta, un documento, una conversación anterior o instrucciones.
3. Los datos de salida son los contenidos producidos durante la inferencia.

El hecho de que una información figure en el contexto de una solicitud no significa que modifique el entrenamiento del modelo. A la inversa, el hecho de que un modelo haya sido entrenado con grandes volúmenes de texto no permite, por sí solo, determinar el origen preciso de una respuesta. Esta separación es esencial para razonar posteriormente sobre la confidencialidad, la calidad de las respuestas, la evaluación y la gobernanza de los usos.

© 2026 ASCALEA. Tous droits réservés.

Generar no es ni buscar ni verificar

Un LLM genera texto a partir de los tokens y del contexto que recibe. Su capacidad para producir una respuesta fluida no constituye una verificación independiente de los hechos. Los riesgos asociados a la inteligencia artificial generativa incluyen, en particular, la producción de contenido confabulado, es decir, contenido presentado de forma plausible pero erróneo o engañoso; el perfil del National Institute of Standards and Technology (NIST) identifica este riesgo en su marco dedicado a la IA generativa [NIST-AI-600-1].

Esta distinción conduce a una regla de lectura sencilla:

> Una respuesta bien formulada es una salida lingüística. Su fiabilidad para una decisión depende de elementos adicionales: fuentes disponibles, método de verificación, contexto, nivel de riesgo y control humano adecuado.

Lo que pretende una comprensión operativa de los LLM

Comprender los LLM no significa memorizar el conjunto de su arquitectura interna. Una comprensión operativa consiste más bien en saber responder a cinco preguntas.

1. ¿Qué tarea se busca realizar?

La solicitud puede consistir en una redacción, una síntesis, una extracción de información, una clasificación, una traducción, una generación de código o una interacción conversacional. Las bibliotecas de procesamiento del lenguaje distinguen por sí mismas tareas y pipelines específicos, lo que recuerda que un mismo modelo puede movilizarse de formas diferentes [HF-TASKS-EXPLAINED] [HF-PIPELINES].

2. ¿Qué información recibe el sistema?

Es necesario identificar el texto transmitido al modelo, las instrucciones que lo encuadran, el idioma utilizado, los documentos que puedan añadirse y las herramientas a las que la aplicación puede dar acceso. La calidad de una respuesta depende en parte de la claridad de la instrucción y del contexto proporcionado [ANTHROPIC-PROMPT-BEST-PRACTICES].

3. ¿Qué componente produce la salida?

Una respuesta puede generarse directamente mediante un LLM, extraerse de una base documental, calcularse mediante una herramienta o construirse por la aplicación a partir de varios componentes. Nombrar este componente evita atribuir indebidamente todo el valor o todo el error al modelo de lenguaje.

4. ¿Cómo debe controlarse la salida?

El control esperado varía según las consecuencias de un error. Una idea de reformulación puede requerir una revisión ligera; una información destinada a un contrato, una decisión financiera, una publicación o un procedimiento sensible debe verificarse a la luz de fuentes y reglas apropiadas. El NIST presenta la gestión de riesgos como un proceso que permite integrar la fiabilidad y la gestión de riesgos en el diseño, el desarrollo, el uso y la evaluación de los sistemas de IA [NIST-AI-RMF].

© 2026 ASCALEA. Tous droits réservés.

5. ¿Qué límites siguen siendo pertinentes pese a una buena respuesta aparente?

Incluso una salida convincente puede ser incompleta, sesgada, obsoleta, inadecuada para el contexto, demasiado general o basarse en una interpretación errónea de la solicitud. Los fenómenos de sobreajuste ilustran, de forma más amplia, que un buen rendimiento con ciertos datos no garantiza la generalización a situaciones nuevas [GOOGLE-OVERFITTING].

Límites del contenido y cuestiones fuera del alcance

Este capítulo establece el lenguaje común del libro; no pretende resolver, desde el principio, todas las cuestiones relacionadas con la IA.

Los siguientes temas no se tratan aquí de manera exhaustiva:

- la demostración matemática detallada de las redes neuronales y del mecanismo de atención;
- la comparación cuantificada de modelos particulares;
- el análisis jurídico aplicable en un país o sector determinado;
- la auditoría completa de seguridad, confidencialidad o conformidad de un proveedor;
- la evaluación de una herramienta comercial concreta;
- los debates filosóficos sobre la conciencia, la intencionalidad o la inteligencia humana.

Estos límites no significan que estas cuestiones sean secundarias. Simplemente indican el orden de progresión adoptado: antes de juzgar una herramienta, es necesario distinguir el ámbito, el modelo, los datos, la aplicación y el uso real.

Para recordar

- La IA es un ámbito amplio; los LLM solo constituyen una familia particular de herramientas.
- Un modelo no es una aplicación completa: la experiencia del usuario también resulta de las instrucciones, el contexto, los parámetros, los datos y las integraciones.
- Un LLM procesa texto en forma de tokens y genera una salida secuencial; no debe confundirse con un mecanismo autónomo de verificación de hechos [HF-TOKENIZER] [HF-TEXT-GENERATION].
- Los datos influyen en los comportamientos observados, incluidas sus carencias y sus posibles sesgos [GOOGLE-DATA-CHARACTERISTICS] [GOOGLE-FAIRNESS-BIAS].
- Para analizar un uso, deben examinarse la tarea, las entradas, el componente movilizado, el control requerido y las posibles consecuencias de un error.

© 2026 ASCALEA. Tous droits réservés.

CHAPITRE 2

Un modelo de inteligencia artificial no «comprende» un ámbito como lo haría una persona. Se construye a partir de datos, se ajusta durante un entrenamiento, se examina mediante evaluaciones y, posteriormente, se utiliza para producir resultados durante la inferencia. Distinguir estas etapas es esencial: modificar un mensaje dirigido a un modelo ya disponible no lo reentrena; solo modifica las condiciones de su respuesta en ese momento.

En el caso de los grandes modelos de lenguaje (Large Language Models, LLM), este ciclo explica tanto sus capacidades aparentes como parte de sus limitaciones. Manipulan texto en forma de unidades llamadas tokens (tokens) y, por lo general, generan una secuencia prediciendo progresivamente los siguientes tokens en un contexto dado. [HF-TOKENIZER] [GOOGLE-LLM-INTRO]

© 2026 ASCALEA. Tous droits réservés.

Una visión general: construir, verificar, utilizar

El ciclo puede representarse de forma sencilla:

1. Definir la necesidad y el alcance: precisar la tarea esperada, las personas implicadas, el contexto y lo que constituiría un resultado aceptable.
2. Preparar los datos: seleccionar, estructurar, limpiar y describir los datos pertinentes para ese alcance.
3. Entrenar el modelo: ajustar sus parámetros para que reduzca un error definido sobre ejemplos.
4. Evaluar el modelo: medir su comportamiento sobre datos o escenarios previstos para ello, con respecto a criterios explícitos.
5. Desplegar e inferir: proporcionar una entrada al modelo entrenado y aprovechar su salida en una situación real.
6. Supervisar y mejorar: examinar las desviaciones, los errores y los cambios de contexto que pueden hacer insuficiente la evaluación inicial.

> Referencia esencial — el aprendizaje y el uso son dos operaciones distintas. El entrenamiento actúa sobre los parámetros internos del modelo a partir de datos y de un objetivo de optimización. La inferencia utiliza estos parámetros ya fijados para calcular una salida a partir de una nueva entrada. Una conversación, una instrucción o un ejemplo proporcionado en una solicitud puede orientar una respuesta, sin constituir por sí solo un nuevo entrenamiento del modelo.

Esta representación es deliberadamente simplificada. En un proyecto real, las fases suelen solaparse: una evaluación revela una carencia de datos, una nueva versión del modelo exige una nueva evaluación y la retroalimentación del uso puede llevar a redefinir el alcance.

Datos: función, preparación y alcance

Los datos son los materiales a partir de los cuales un sistema de aprendizaje automático (machine learning) establece regularidades. Sin embargo, su presencia no garantiza ni la pertinencia ni la fiabilidad del modelo: las características de un conjunto de datos —en particular su completitud, su distribución, su calidad y su relación con la tarea prevista— influyen directamente en lo que un modelo puede aprender y generalizar. [GOOGLE-DATA-CHARACTERISTICS]

Qué pueden contener los datos

Según el sistema, los datos pueden comprender:

- textos, imágenes, sonidos o datos estructurados;
- entradas que describen una situación;
- salidas esperadas, a veces llamadas etiquetas, cuando la tarea requiere ejemplos anotados;
- metadatos que permiten conocer el origen, el período, el formato o las condiciones de recopilación;
- instrucciones y ejemplos de diálogos, en el caso de algunos sistemas conversacionales.

© 2026 ASCALEA. Tous droits réservés.

En un LLM, el texto se convierte en tokens antes de su procesamiento. Esta conversión tiene consecuencias prácticas: una misma frase puede segmentarse de manera diferente según el segmentador léxico (tokenizer) utilizado, y los límites de longitud se expresan en tokens en lugar de palabras. [HF-TOKENIZER]

Preparar no significa solamente «limpiar»

La preparación de los datos abarca mucho más que eliminar duplicados o errores de formato. Supone decidir qué entra dentro del alcance y qué queda excluido. Estas decisiones deben ser coherentes con el uso previsto.

Pregunta de preparación	Por qué es importante
¿Los datos corresponden a las situaciones reales de uso?	Un modelo puede funcionar bien con los ejemplos disponibles y, aun así, fallar ante casos ausentes o muy diferentes. [GOOGLE-DATA-CHARACTERISTICS]
¿Algunas categorías, lenguas, períodos o poblaciones están poco representados?	Una cobertura desigual puede introducir o amplificar diferencias de comportamiento entre grupos o contextos. [GOOGLE-FAIRNESS-BIAS]
¿Se conocen la fuente, el período y las transformaciones?	Esta documentación permite interpretar los resultados e identificar una discrepancia entre los datos y el uso. [NIST-AI-600-1]
¿Los datos contienen información sensible, confidencial o innecesaria?	La gestión de los datos forma parte de los riesgos que deben considerarse en el ciclo de vida de los sistemas de inteligencia artificial generativa. [NIST-AI-600-1]

La representatividad no significa necesariamente que un conjunto de datos deba reproducir a toda la población. Significa, más bien, que su composición debe examinarse en relación con la finalidad declarada. Un modelo destinado a asistir en la redacción de contenidos en español, por ejemplo, debe examinarse con respecto a las variedades lingüísticas, los registros y los tipos de documentos realmente previstos; no basta con constatar que procesa «español» de manera general.

De los datos a los tokens

Para procesar una secuencia textual, un LLM recibe una representación numérica de los tokens. Las arquitecturas de tipo Transformer utilizan, entre otros, mecanismos de atención para procesar las relaciones entre los elementos de una secuencia. [GOOGLE-TRANSFORMER-PAPER] Esta mecánica no exime de una pregunta fundamental: ¿los datos disponibles aportan los elementos pertinentes para la tarea, en un formato y un contexto que el sistema pueda aprovechar?

Entrenamiento: ajustar un modelo a un objetivo

El entrenamiento es la fase durante la cual los parámetros de un modelo se ajustan a partir de datos y de un objetivo medible. En una formulación general, el sistema produce una predicción, compara esta predicción con un objetivo o una señal esperada, calcula una medida de error y, a continuación, modifica

© 2026 ASCALEA. Tous droits réservés.

sus parámetros para reducir ese error a lo largo de las iteraciones. Las nociones de función de pérdida, ejemplos de entrenamiento y optimización forman parte del vocabulario central del aprendizaje automático. [GOOGLE-ML-GLOSSARY]

Para un modelo de lenguaje generativo, una representación simplificada es la siguiente: a partir de una secuencia de tokens, el modelo estima qué tokens son plausibles a continuación. La generación consiste entonces en seleccionar sucesivamente tokens según la distribución calculada, hasta alcanzar una condición de parada o un límite. [GOOGLE-LLM-INTRO] [HF-TEXT-GENERATION]

Esta explicación no debe confundirse con una lista de reglas redactadas a mano. El comportamiento final resulta de un gran número de parámetros ajustados durante el entrenamiento. Puede producir formulaciones útiles, resumir, transformar o continuar un texto, pero no constituye, por naturaleza, una prueba de que el contenido generado sea exacto, actual o apropiado para el caso tratado.

Objetivo, datos y parámetros: una analogía prudente

Una analogía útil consiste en comparar el entrenamiento con el ajuste de un conjunto muy amplio de configuraciones:

- los datos proporcionan los ejemplos a partir de los cuales se realiza el ajuste;
- el objetivo precisa lo que se busca mejorar;
- los parámetros son las configuraciones internas modificadas durante el entrenamiento;
- la evaluación verifica si el resultado sigue siendo satisfactorio en casos que no se reducen a los ejemplos utilizados para el ajuste.

La analogía tiene un límite importante: no describe una comprensión humana ni una base de conocimientos estable y verificada explícitamente. Sirve únicamente para distinguir el mecanismo de construcción del modelo de su uso posterior.

El riesgo de sobreajuste

Un modelo puede llegar a estar demasiado estrechamente adaptado a los datos utilizados para entrenarlo. Este fenómeno, denominado sobreajuste, se manifiesta cuando un buen rendimiento sobre los datos de entrenamiento no se traduce en un buen rendimiento sobre datos nuevos. [GOOGLE-OVERFITTING]

El sobreajuste recuerda que una respuesta convincente en unos pocos ejemplos no basta para concluir que un modelo sea fiable dentro del alcance anunciado. Es necesario examinar su capacidad para tratar casos no vistos, cercanos pero no idénticos a los empleados durante su construcción.

Evaluación: verificar una capacidad definida, en un contexto definido

La evaluación responde a una pregunta distinta de la del entrenamiento: no «¿ha reducido el modelo su error en los ejemplos utilizados para ajustarlo?», sino «¿su comportamiento satisface los criterios seleccionados para el uso previsto?».

© 2026 ASCALEA. Tous droits réservés.

Las prácticas de aprendizaje automático distinguen habitualmente los datos de entrenamiento de los datos reservados para validación y prueba, con el fin de estimar el comportamiento del modelo más allá de los ejemplos de ajuste. [GOOGLE-ML-GLOSSARY] Esta separación es una precaución metodológica, no una garantía absoluta: solo cubre las situaciones representadas por el protocolo de evaluación.

Definir qué se evalúa

Una evaluación útil comienza con una descripción explícita de su objeto. Para un asistente de redacción en español, los criterios pueden referirse, por ejemplo, a:

- la fidelidad a una fuente proporcionada en la solicitud;
- el cumplimiento de una instrucción de tono o formato;
- la calidad lingüística para una variedad definida de español;
- la capacidad de señalar una incertidumbre en lugar de afirmar un elemento no establecido;
- el comportamiento ante solicitudes fuera de alcance o ambiguas.

Estos ejemplos no constituyen una lista universal de medidas. Ilustran el principio según el cual una medida solo tiene sentido en relación con una tarea, unos usuarios y unas consecuencias identificadas.

Las herramientas de evaluación pueden comparar una salida con una respuesta de referencia, aplicar criterios estructurados o examinar casos de prueba representativos. Las evaluaciones dedicadas a sistemas basados en modelos pueden organizarse en forma de conjuntos de datos y ejecuciones documentadas. [OPENAI-EVALS] Sin embargo, una puntuación agregada no debe sustituir el examen de los casos importantes: dos sistemas con un resultado medio comparable pueden fallar en casos diferentes.

Documentar el protocolo, no solo la puntuación

Para interpretar un resultado, es útil documentar como mínimo:

Elemento que se debe documentar	Ejemplo de pregunta
Alcance	¿Qué tipos de solicitudes y qué lenguas están incluidos?
Versión evaluada	¿Qué modelo, qué configuración y qué instrucciones se utilizaron?
Conjunto de casos	¿De dónde proceden los casos y qué situaciones siguen ausentes?
Criterios	¿Qué se considera una salida correcta, segura o utilizable?
Resultados detallados	¿Qué errores aparecen, en qué casos y con qué gravedad?
Límites	¿Qué conclusiones no permite extraer la evaluación?

Esta documentación facilita la comparación entre versiones y evita presentar un rendimiento local como una propiedad general del modelo. El marco de gestión de riesgos de inteligencia artificial del National Institute

© 2026 ASCALEA. Tous droits réservés.

of Standards and Technology (NIST) insiste en la necesidad de tener en cuenta los riesgos de la inteligencia artificial generativa a lo largo de todo su ciclo de vida. [NIST-AI-RMF] [NIST-AI-600-1]

Inferencia: utilizar un modelo ya entrenado

La inferencia es el momento en que un modelo entrenado recibe una entrada y calcula una salida. En un sistema de generación de texto, la entrada puede incluir una instrucción, un documento, un historial de conversación y restricciones de formato. A continuación, el modelo produce una continuación según su arquitectura, sus parámetros y la configuración de generación seleccionada. [HF-TEXT-GENERATION] [HF-GENERATION-CONFIG]

En la práctica, una cadena de inferencia puede incluir varias etapas: preparación de la entrada, conversión en tokens, llamada al modelo, generación y, después, procesamiento de la salida. Las bibliotecas técnicas ofrecen pipelines que agrupan tales etapas para tareas determinadas. [HF-PIPELINES] [HF-TASKS-EX-PLAINED]

Qué influye en una respuesta sin reentrenar el modelo

Durante la inferencia, varios elementos pueden hacer variar la salida:

- el contenido y la claridad de la instrucción;
- la información incluida en el contexto;
- el orden de los elementos en la solicitud;
- los ejemplos que puedan proporcionarse en dicha solicitud;
- los parámetros de generación, como los mecanismos que controlan la diversidad o la longitud de la generación. [ANTHROPIC-PROMPT-BEST-PRACTICES] [HF-GENERATION-CONFIG]

Estos elementos son importantes para el uso, pero no equivalen a un aprendizaje permanente. Una buena formulación de la solicitud puede mejorar la precisión operativa de una respuesta; no transforma automáticamente los parámetros del modelo ni corrige las limitaciones de sus datos de entrenamiento.

> Consecuencia práctica: una salida producida durante la inferencia es una propuesta calculada en un contexto dado. Cuando sirve para informar una decisión, un documento profesional, una traducción sensible o una acción de impacto, su adecuación debe verificarse según el nivel de riesgo y las reglas aplicables al contexto de uso.

Limitaciones relacionadas con los datos, el contexto y la evaluación

Comprender el ciclo de vida permite identificar varias familias de limitaciones. No conciernen únicamente al modelo: pueden aparecer en cada etapa, desde la definición del problema hasta el uso de la respuesta.

1. Limitaciones de los datos

Los datos incompletos, obsoletos, mal etiquetados, desequilibrados o insuficientemente relacionados con la tarea pueden afectar a los resultados. Por tanto, las características y la distribución de los datos son factores determinantes del rendimiento observado. [GOOGLE-DATA-CHARACTERISTICS]

© 2026 ASCALEA. Tous droits réservés.

Los sesgos también pueden provenir de elecciones realizadas incluso antes del entrenamiento: definición de las categorías, método de recopilación, anotaciones, exclusiones o calidad variable de la información. Por ello, la identificación de sesgos exige examinar los datos, el problema formulado y los resultados para los grupos o casos afectados. [GOOGLE-FAIRNESS-BIAS]

2. Limitaciones del contexto proporcionado a la inferencia

Un modelo responde únicamente a partir de la entrada y del contexto efectivamente disponibles en el momento de la inferencia, dentro de los límites de su mecanismo de procesamiento. Una pregunta vaga, una fuente incompleta o instrucciones contradictorias pueden conducir a una salida mal adaptada, incluso si el modelo es técnicamente capaz de tratar la tarea. Las prácticas de formulación de instrucciones recomiendan hacer explícitos el contexto, las restricciones y el formato esperado. [ANTHROPIC-PROMPT-BEST-PRACTICES]

3. Limitaciones de la evaluación

Una evaluación no demuestra una fiabilidad universal. Establece un resultado en las condiciones del protocolo elegido: versión del modelo, datos de prueba, configuración, criterios y escenarios cubiertos. Los cambios de población, lengua, documentos, instrucciones o condiciones de uso pueden hacer que una evaluación anterior sea menos informativa. Por lo tanto, la gestión de riesgos debe tener en cuenta el contexto de uso y evolucionar con el sistema. [NIST-AI-600-1]

4. Limitaciones de la propia generación

La generación textual comporta una parte de variabilidad ligada a los parámetros de decodificación. Por tanto, una misma solicitud puede no producir siempre una formulación idéntica, según la configuración utilizada. [HF-GENERATION-CONFIG] La calidad aparente de un texto no es, por sí sola, una prueba de su exactitud factual, de la exhaustividad de su razonamiento ni de su conformidad con una necesidad profesional.

Una guía de lectura para analizar un resultado

Ante una respuesta producida por un modelo, cuatro preguntas permiten mantener una lectura estructurada:

1. ¿A partir de qué datos y con qué objetivo se construyó el modelo?
2. ¿En qué contexto preciso se generó la respuesta?
3. ¿Qué cubrió realmente la evaluación disponible?
4. ¿Qué verificaciones humanas o procedimentales son necesarias antes de utilizarla?

Esta guía no busca oponer la inteligencia artificial a la experiencia humana. Clarifica los roles: el modelo produce una salida a partir de mecanismos estadísticos y de un contexto; las personas y las organizaciones definen la finalidad, valoran las consecuencias, controlan la información importante y deciden el uso final.

Para recordar

- Los datos definen en gran medida lo que un modelo puede aprender; su calidad y su adecuación a la necesidad deben examinarse, no darse por supuestas. [GOOGLE-DATA-CHARACTERISTICS]
- El entrenamiento ajusta los parámetros de un modelo; la inferencia utiliza un modelo ya entrenado para producir una salida a partir de una nueva entrada. [GOOGLE-ML-GLOSSARY] [HF-TEXT-GENERATION]

© 2026 ASCALEA. Tous droits réservés.

• La evaluación debe estar vinculada a un alcance, unos criterios y unos casos de prueba documentados explícitamente; una puntuación aislada no resume todos los riesgos. [NIST-AI-600-1]

• Las limitaciones pueden provenir de los datos, de la formulación de la solicitud, del contexto ausente, de la configuración de generación o del propio protocolo de evaluación. [GOOGLE-FAIRNESS-BIAS] [HF-GENERATION-CONFIG]

• Una respuesta generada debe valorarse en función del uso previsto y de sus consecuencias, especialmente cuando interviene en una situación sensible o de alto impacto. [NIST-AI-RMF]

© 2026 ASCALEA. Tous droits réservés.

CHAPITRE 3

Después de distinguir el modelo, sus datos de entrenamiento y su uso en inferencia, es necesario disponer de una guía sencilla para orientar una necesidad hacia una familia de modelos. Esta guía no constituye una taxonomía universal: sirve para relacionar tres elementos operativos:

1. la modalidad de entrada: la forma de la información proporcionada al modelo;
2. la tarea esperada: lo que el sistema debe hacer con esa información;
3. el tipo de resultado: una clase, una puntuación, un texto, una imagen, una señal de audio o una respuesta estructurada.

Un mismo modelo puede a veces cubrir varias tareas, mientras que dos modelos que procesan la misma modalidad pueden perseguir resultados muy diferentes. Las bibliotecas técnicas de modelos distinguen así pipelines para tareas de texto, imagen, audio y tareas que combinan texto e imagen. [HF-PIPELINES] [HF-TASKS-EXPLAINED]

> Punto de referencia — La elección de un modelo no comienza por su nombre comercial ni por su tamaño. Comienza por la decisión que se debe respaldar, los datos realmente disponibles y el resultado verificable esperado.

Criterios de clasificación de los modelos

Una familia de modelos puede describirse según varios ejes complementarios.

© 2026 ASCALEA. Tous droits réservés.

Eje de clasificación	Pregunta que se debe plantear	Ejemplos de respuestas posibles
Modalidad procesada	¿En qué forma llega la información?	Texto, imagen, audio, datos estructurados, combinación de modalidades
Naturaleza de la tarea	¿Qué transformación se espera?	Clasificar, extraer, responder, resumir, generar, traducir, detectar
Forma de la salida	¿Qué resultado será consumido por una persona o un sistema?	Etiqueta, puntuación, fragmento de texto, transcripción, imagen, coordenadas, respuesta estructurada
Alcance funcional	¿El modelo se orienta a una tarea precisa o a muchos usos?	Especializado, generalista, multimodal
Modo de integración	¿El resultado es leído, validado o ejecutado por otro sistema?	Asistencia humana, búsqueda, automatización supervisada

Esta clasificación evita una confusión frecuente: una modalidad no es una finalidad. Un modelo «de texto» puede servir para clasificar un mensaje, extraer una información, traducir una frase, producir un resumen o generar un borrador. Del mismo modo, un modelo «de imagen» puede asignar una categoría a una imagen, localizar objetos, segmentar zonas o producir una descripción textual. [HF-PIPELINES] [HF-TASKS-EXPLAINED]

También se puede distinguir entre los sistemas que producen principalmente una predicción restringida —por ejemplo, una categoría o una puntuación— y los que producen un contenido nuevo. Los modelos generativos presentan riesgos propios: su producción puede parecer plausible sin ser adecuada al contexto o veraz. El marco del National Institute of Standards and Technology (NIST) dedicado a la Inteligencia artificial (IA) generativa invita, por tanto, a considerar los riesgos asociados al contenido generado en el diseño, el despliegue y la evaluación de un sistema. [NIST-AI-600-1] [NIST-AI-RMF]

Modelos orientados al texto

Los modelos orientados al texto procesan secuencias de símbolos divididas en unidades llamadas tokens. La división en tokens depende del tokenizer, componente que transforma el texto en representaciones utilizables por el modelo y que luego permite reconstruir texto. [HF-TOKENIZER]

Entre ellos, los grandes modelos de lenguaje, o Large Language Models (LLM), se entrenan para manipular y generar lenguaje. Los LLM modernos se basan habitualmente en la arquitectura Transformer, introducida en la publicación Attention Is All You Need. [GOOGLE-LLM-INTRO] [GOOGLE-TRANSFORMER-PAPER]

Principales finalidades textuales

© 2026 ASCALEA. Tous droits réservés.

Finalidad	Entrada típica	Resultado esperado	Ejemplo de uso
Clasificación de texto	Un mensaje, una reseña, una solicitud	Una o varias categorías	Dirigir una solicitud hacia una cola de procesamiento
Clasificación de tokens	Un texto dividido	Una etiqueta por porción de texto	Identificar entidades en un documento
Pregunta-respuesta	Una pregunta y un contexto	Una respuesta extraída o formulada	Encontrar una información en un corpus proporcionado
Resumen	Un documento o un conjunto de fragmentos	Una versión condensada	Preparar una síntesis de lectura
Traducción	Un texto de origen	Un texto en el idioma de destino	Producir una primera versión de traducción
Generación de texto	Una instrucción y, eventualmente, contexto	Una continuación de texto	Redactar un borrador, reformular o estructurar un contenido

Las herramientas documentadas para los modelos Transformer cubren, en particular, la clasificación de texto, la clasificación de tokens, la pregunta-respuesta, el resumen, la traducción y la generación de texto. [HF-PIPELINES] [HF-TASKS-EXPLAINED]

La generación textual no se reduce a la elección de un modelo. Sus parámetros influyen en el comportamiento de salida, en particular en la longitud producida y la estrategia de selección de tokens. Por lo tanto, una configuración de generación debe formar parte de la definición del sistema, al igual que la instrucción proporcionada al modelo. [HF-GENERATION-CONFIG] [HF-TEXT-GENERATION]

> Para recordar — Un LLM puede ser muy útil para producir, reformular u organizar lenguaje. No debe deducirse de ello que sea, por naturaleza, un sistema de búsqueda fiable, una base de datos actualizada o una autoridad de decisión.

Cuándo priorizar una tarea textual específica

Una tarea específica suele ser preferible cuando la salida esperada es limitada y directamente medible: una categoría, una extracción, una respuesta a partir de un contexto impuesto o una traducción. En este caso, la especificación debe definir sin ambigüedad:

- las categorías, campos o formatos de salida aceptados;
- los casos ambiguos y la conducta que se debe adoptar en caso de incertidumbre;
- el corpus o el contexto que el sistema tiene derecho a utilizar;
- los criterios que permiten verificar la calidad del resultado.

© 2026 ASCALEA. Tous droits réservés.

Para una generación libre, la instrucción, el contexto y el formato de respuesta se convierten en una parte esencial de la interfaz. La documentación de buenas prácticas para la formulación de instrucciones recomienda, en particular, el uso de plantillas de prompts y variables para hacer que las solicitudes sean reutilizables y coherentes. [ANTHROPIC-PROMPT-BEST-PRACTICES]

Modelos orientados a imagen, audio y otras modalidades

Una modalidad designa la forma en la que una información está representada o es percibida por el sistema. El texto, la imagen y el audio son tres modalidades habituales. Las familias de tareas asociadas no deben confundirse con la capacidad humana correspondiente: reconocer una clase de imagen, por ejemplo, no significa necesariamente comprender su contexto, su intención o su alcance de negocio.

Imagen

Las herramientas dedicadas a los modelos Transformer enumeran, en particular, las siguientes tareas para imagen:

- clasificación de imágenes: asignar una o varias categorías a una imagen;
- detección de objetos: identificar objetos y su localización;
- segmentación de imágenes: asociar zonas o píxeles a categorías;
- imagen a texto: producir texto a partir de una imagen;
- generación imagen a imagen y texto a imagen en los pipelines que las admiten. [HF-PIPELINES] [HF-TASKS-EXPLAINED]

Estas finalidades producen resultados diferentes. Una clasificación puede ser suficiente cuando se deben clasificar contenidos. Una detección es adecuada cuando se debe localizar un elemento. Una segmentación responde a una necesidad de precisión espacial más fina. Una salida de imagen a texto, por su parte, es contenido generado y debe evaluarse como tal.

Audio

Para el audio, las tareas documentadas incluyen, en particular:

- clasificación de audio: asociar una grabación a una categoría;
- reconocimiento automático del habla: transformar habla en texto;
- síntesis de voz: transformar texto en habla;
- audio a audio en ciertos pipelines. [HF-PIPELINES] [HF-TASKS-EXPLAINED]

El reconocimiento automático del habla puede servir para crear una transcripción que pueda ser utilizada por una cadena textual: búsqueda, resumen, clasificación o extracción. Sin embargo, esta cadena incluye varias etapas. Un error de transcripción puede propagarse a los procesamientos de texto posteriores; por tanto, la evaluación debe centrarse en el resultado final y no solo en cada componente considerado de forma aislada.

Otras formas de datos

Los sistemas de aprendizaje automático también pueden trabajar con datos estructurados o numéricos. En este caso, la clasificación por modalidad es menos reveladora que la descripción de las variables

© 2026 ASCALEA. Tous droits réservés.

disponibles, su calidad y el objetivo que se debe predecir. La documentación de Google recuerda que las características de los datos influyen directamente en la calidad y la generalización de un modelo. [GOOGLE-DATA-CHARACTERISTICS] [GOOGLE-OVERFITTING]

Para todas las modalidades, la cuestión central sigue siendo la misma: ¿están disponibles los datos representativos de las condiciones reales de uso, suficientemente definidos y evaluados? Los sesgos presentes en los datos o en su recopilación pueden afectar a los resultados y deben identificarse en el proceso de desarrollo. [GOOGLE-FAIRNESS-BIAS]

Enfoques multimodales

Un enfoque multimodal procesa o relaciona varias modalidades. Es pertinente cuando el significado buscado depende de su combinación: una pregunta sobre una imagen, una descripción de imagen guiada por una instrucción textual, o el análisis de un contenido que incluye texto y elementos visuales.

Los pipelines documentados incluyen, en particular, la tarea de pregunta-respuesta visual, que asocia una imagen y una pregunta para producir una respuesta. [HF-PIPELINES] [HF-TASKS-EXPLAINED]

Combinación	Ejemplo de finalidad	Punto de vigilancia
Texto + imagen	Responder a una pregunta sobre una imagen	Verificar que la respuesta esté respaldada por los elementos visibles y el texto proporcionado
Audio + texto	Transcribir y luego resumir un intercambio	Medir el efecto acumulado de los errores de transcripción y resumen
Imagen + texto generado	Describir un elemento visual	Controlar las descripciones sin fundamento o excesivamente interpretativas
Varios documentos y medios	Buscar o sintetizar información	Definir las fuentes autorizadas, la trazabilidad y el formato de restitución

La ganancia potencial de la multimodalidad es una representación más completa del caso tratado. Su coste es una superficie de evaluación más amplia: es necesario probar la calidad de cada entrada, el comportamiento ante señales contradictorias o incompletas, y la adecuación de la salida a la decisión prevista.

> Principio de prudencia — Añadir una modalidad no mejora automáticamente un sistema. Una modalidad adicional es útil si aporta información necesaria y si su calidad puede controlarse en condiciones reales de uso.

Modelos generalistas y modelos especializados

Un modelo generalista se orienta a una amplia gama de tareas, a menudo mediante instrucciones en lenguaje natural. Los LLM se utilizan habitualmente con esta lógica: un mismo modelo puede emplearse

© 2026 ASCALEA. Tous droits réservés.

para resumir, traducir, extraer o redactar, según el contexto y la instrucción. [GOOGLE-LLM-INTRO] [HF-TEXT-GENERATION]

Un modelo especializado se orienta hacia una tarea, una modalidad o un ámbito más restringido. Puede, por ejemplo, estar diseñado para una clasificación determinada, una transcripción o una tarea de visión. Los pipelines de tareas ilustran esta especialización funcional. [HF-PIPELINES] [HF-TASKS-EXPLAINED]

Dimensión	Modelo generalista	Modelo especializado
Amplitud de usos	Varias tareas posibles	Tarea o alcance más delimitado
Interfaz habitual	Instrucción, contexto y formato de salida	Entrada y salida a menudo fuertemente definidas
Ventaja principal	Flexibilidad de adaptación	Definición precisa de la función esperada
Riesgo de diseño	Esperar una fiabilidad idéntica para todo tipo de solicitud	Subestimar los casos fuera de alcance
Evaluación	Conjuntos de casos que cubren los usos reales y las instrucciones	Medidas directamente vinculadas a la tarea objetivo

La elección correcta no es sistemáticamente la más generalista ni la más especializada. Depende de la variabilidad de la necesidad.

- Cuando las solicitudes cambian con frecuencia, la salida sigue siendo redactada y se prevé una revisión humana, un modelo generalista puede ofrecer una interfaz flexible.
- Cuando el resultado debe respetar un conjunto reducido de categorías, campos o reglas verificables, un enfoque más específico puede ser más sencillo de especificar y evaluar.
- Cuando una necesidad combina una tarea estable y una comunicación en lenguaje natural, una arquitectura híbrida puede ser pertinente: un componente especializado proporciona un resultado controlado, mientras que un componente generativo le da forma para el usuario. Esta combinación debe preservar la distinción entre los datos calculados y el texto redactado.

En todos los casos, la evaluación no puede sustituirse por una impresión de fluidez. Las prácticas de evaluación deben basarse en casos definidos, resultados esperados y criterios de juicio adaptados al uso. Las herramientas de evaluación documentadas por OpenAI organizan, en particular, ejecuciones de evaluaciones y sus resultados; ilustran la importancia de tratar la evaluación como una función explícita del ciclo de desarrollo. [OPENAI-EVALS]

Criterios de comparación adaptados a una necesidad

Comparar modelos a partir de un único indicador rara vez es suficiente. Una comparación útil parte de un caso de uso formulado y luego evalúa los candidatos con las mismas entradas, las mismas restricciones y los mismos criterios de aceptación.

© 2026 ASCALEA. Tous droits réservés.

Una guía práctica de comparación

Criterio	Pregunta operativa	Elementos que se deben observar
Adecuación a la tarea	¿El modelo realiza precisamente la transformación buscada?	Exactitud, pertinencia, respeto de las categorías o del formato
Modalidades admitidas	¿Acepta las entradas realmente disponibles?	Texto, imagen, audio, combinaciones necesarias
Calidad en casos representativos	¿Tiene éxito en ejemplos habituales y difíciles?	Casos frecuentes, ambigüedades, datos incompletos, casos límite
Fiabilidad de la salida	¿Puede controlarse el resultado antes de su uso?	Fuentes proporcionadas, restricciones de formato, validación humana o automática
Sensibilidad a las instrucciones	¿El comportamiento se mantiene estable cuando varía la formulación?	Robustez ante solicitudes similares, papel del contexto, plantillas de instrucciones
Datos y equidad	¿Los datos y resultados crean diferencias indeseables?	Representatividad, poblaciones afectadas, errores diferenciales
Riesgos de uso	¿Qué consecuencias hay en caso de una salida errónea, incompleta o inadecuada?	Nivel de supervisión, reversibilidad, registro, procedimientos de escalamiento

La calidad de los datos, el sobreajuste y los sesgos son factores que deben examinarse antes de concluir que un modelo es adecuado. Un modelo puede obtener buenos resultados en los datos utilizados para probarlo sin generalizar a situaciones reales; este es precisamente el problema que describe el sobreajuste. [GOOGLE-OVERFITTING] [GOOGLE-DATA-CHARACTERISTICS]

Para los usos de alto impacto, el análisis también debe incluir los riesgos de la IA generativa, las personas que podrían verse afectadas y los mecanismos de gobernanza. El marco de gestión de riesgos de IA del NIST y su perfil dedicado a la IA generativa proponen un marco para estructurar este enfoque. [NIST-AI-RMF] [NIST-AI-600-1]

Árbol de orientación inicial

1. ¿Cuál es el material de entrada principal?

• Texto: examinar las familias de clasificación, extracción, pregunta-respuesta, resumen, traducción o generación.

• Imagen: examinar la clasificación, la detección, la segmentación, imagen a texto o la generación según el resultado esperado.

© 2026 ASCALEA. Tous droits réservés.

• Audio: examinar la clasificación, la transcripción, la síntesis de voz o una cadena que combine audio y texto.

• Varias modalidades: examinar un enfoque multimodal solo si su combinación es indispensable para la necesidad.

1. ¿Qué resultado debe entregarse?

• Categoría, puntuación o campo estructurado: priorizar una tarea explícitamente definida y medible.

• Texto o medio nuevo: prever restricciones de salida, controles y una evaluación de la calidad generada.

• Respuesta a partir de un corpus: definir con precisión el contexto accesible y las reglas de validación.

1. ¿Cuál es la consecuencia de un error?

• Consecuencia baja: puede considerarse una asistencia con validación ligera según el contexto.

• Consecuencia significativa: reforzar las pruebas, la supervisión, la trazabilidad y los mecanismos de detención o corrección.

1. ¿Cómo se demostrará la calidad?

• Constituir un conjunto de casos representativos.

• Definir resultados esperados o criterios de juicio.

• Comparar los modelos en condiciones idénticas.

• Repetir la evaluación cuando cambien los datos, las instrucciones, el modelo o el contexto de uso.

Síntesis

Las familias de modelos se entienden al relacionar modalidades, tareas y formas de salida. Los modelos textuales cubren tanto tareas restringidas como generación; los modelos de imagen y audio responden a otras transformaciones; los enfoques multimodales vinculan varias fuentes de información. Por tanto, la distinción más útil no es entre tecnologías percibidas como competidoras, sino entre necesidades claramente formuladas y expectativas imprecisas.

Un modelo generalista aporta flexibilidad. Un modelo especializado aporta un alcance más explícito. En ambos casos, la adecuación a la necesidad depende de datos representativos, instrucciones o interfaces bien definidas, una evaluación adaptada y controles proporcionales a las consecuencias de un error.

© 2026 ASCALEA. Tous droits réservés.

CHAPITRE 4

Comprender los grandes modelos de lenguaje

Un Large Language Model (LLM), o gran modelo de lenguaje, es un modelo diseñado para procesar y producir lenguaje. En una interacción habitual, recibe un texto, lo transforma en unidades que el modelo puede manipular, tiene en cuenta el contexto que se le proporciona y, a continuación, genera una secuencia de texto. Los LLM modernos se basan con frecuencia en la arquitectura Transformer, introducida como una arquitectura basada en mecanismos de atención. [GOOGLE-LLM-INTRO] [GOOGLE-TRANSFORMER-PAPER]

La idea esencial es sencilla, pero sus consecuencias son importantes: un LLM no «lee» una solicitud como lo haría una persona que comprendiera espontáneamente la intención, el mundo real y las implicaciones. Produce texto a partir de regularidades aprendidas y del contexto disponible. Por tanto, una formulación convincente, gramaticalmente correcta o muy detallada no basta para establecer que una respuesta sea exacta, adecuada para la situación o segura de utilizar.

> Referencia útil — Un LLM es un sistema de generación y transformación del lenguaje. Puede ser muy eficaz para reformular, resumir, traducir, estructurar o continuar un texto, sin que ello le otorgue automáticamente un conocimiento verificado de cada hecho enunciado.

Características generales de un LLM

© 2026 ASCALEA. Tous droits réservés.

Un LLM se distingue, en particular, por su capacidad para manipular largas secuencias de texto y adaptarse a instrucciones expresadas en lenguaje natural. Según la interfaz o la aplicación que lo emplee, puede realizar distintas tareas: generación de texto, clasificación, resumen, extracción de información, respuesta a una pregunta o traducción. De hecho, las bibliotecas técnicas que rodean estos modelos suelen organizar su uso en «pipelines», es decir, en cadenas de procesamiento asociadas a una tarea determinada. [HF-PIPELINES] [HF-TASKS-EXPLAINED]

Deben diferenciarse varias características:

- El modelo: el sistema estadístico que produce o transforma texto.
- El contexto: el contenido transmitido al modelo en el momento de la solicitud.
- La instrucción: lo que se pide al modelo que haga.
- La configuración de generación: los ajustes que influyen en la manera en que se produce la salida.
- La interfaz o la aplicación: la capa que presenta el modelo al usuario y puede añadir reglas, documentos, herramientas o controles.

Esta distinción evita una confusión frecuente: atribuir únicamente al modelo todo lo que sucede en una aplicación conversacional. Una respuesta puede depender del texto introducido por el usuario, de instrucciones añadidas por la aplicación, de un historial de conversación, de documentos adjuntos y de parámetros de generación. Además, las prácticas de diseño de prompts recomiendan utilizar plantillas de instrucciones y variables para estructurar las entradas de forma coherente. [ANTHROPIC-PROMPT-BEST-PRACTICES]

Texto de entrada, contexto y texto de salida

Una interacción puede representarse mediante una cadena sencilla:

``text Demande de l’utilisateur “ Instructions + contexte disponible + texte de la demande “ Préparation du texte en unités de traitement “ LLM “ Texte généré “ Présentation éventuelle par l’application ``

El texto visible en un área de conversación no es necesariamente la única información procesada. Una aplicación puede, según su diseño, asociar a la solicitud instrucciones de rol, el historial de los intercambios, extractos de documentos o los resultados de una herramienta. Por tanto, es preferible hablar de entrada efectiva: el conjunto de elementos realmente proporcionados al modelo para producir la salida.

La tokenización: del texto a las unidades procesables

Antes de ser procesado, el texto se divide en tokens —también se emplea habitualmente el término inglés tokens. Un token no es necesariamente una palabra completa: puede corresponder a una palabra, una parte de una palabra, un signo de puntuación u otra unidad dependiente del tokenizador utilizado. La tokenización es, por tanto, una etapa determinante entre el texto legible por el ser humano y la representación aprovechable por el modelo. [HF-TOKENIZER]

Esta precisión ayuda a comprender dos fenómenos prácticos:

1. el tamaño del texto no se evalúa únicamente por el número de palabras;
2. la capacidad del modelo para tener en cuenta una solicitud depende de la cantidad total de tokens presentes en su contexto.

© 2026 ASCALEA. Tous droits réservés.

El contexto incluye así no solo la pregunta final, sino también los elementos que la preceden o acompañan. Cuando un contexto es demasiado largo, está mal estructurado o es contradictorio, la interpretación de la solicitud se vuelve más difícil de controlar. La calidad de una interacción depende entonces tanto de la selección de la información proporcionada como de la formulación de la pregunta.

Una instrucción no es un simple tema

Comparar las dos solicitudes siguientes permite ver qué cambia:

Formulación	Efecto esperado
«Explica la inteligencia artificial.»	Solicitud amplia; el nivel, la extensión, el enfoque y el formato siguen sin determinarse.
«Explica en español, en cinco frases, la diferencia entre un modelo de lenguaje y un motor de búsqueda, para un principiante.»	Solicitud más controlable; precisa el idioma, el formato, el público y el tema.

Una instrucción útil puede precisar:

- el objetivo que se debe alcanzar;
- el público destinatario y el nivel de tecnicidad;
- el idioma de la respuesta;
- el contenido que se debe incluir o excluir;
- el formato esperado;
- los criterios de prudencia, por ejemplo, pedir al modelo que señale sus incertidumbres.

La estructuración explícita de las instrucciones, el uso de variables y la separación clara entre las diferentes partes de un prompt forman parte de las prácticas documentadas de diseño de instrucciones. [ANTHROPIC-PROMPT-BEST-PRACTICES]

Representación pedagógica de la generación de texto

La generación de texto puede entenderse como una producción secuencial. A partir de la entrada disponible, el modelo estima qué tokens podrían constituir una continuación plausible, selecciona un token según su configuración, lo añade al texto ya producido y vuelve a empezar. Las herramientas de generación exponen precisamente mecanismos de control como la longitud máxima de nuevos tokens, el muestreo, la temperatura y los umbrales o métodos de selección de tokens. [HF-TEXT-GENERATION] [HF-GENERATION-CONFIG]

Puede representarse así:

``text Entrée : « Résume ce texte en espagnol : ... » “ Le modèle traite le contexte disponible “ Il produit un premier jeton “ Ce jeton rejoint le texte de sortie “ Le modèle produit le jeton suivant “ La boucle continue jusqu’à une condition d’arrêt ``

© 2026 ASCALEA. Tous droits réservés.

Esta representación es deliberadamente simplificada. Sin embargo, permite comprender un punto fundamental: la respuesta se construye progresivamente. El modelo no recupera necesariamente una frase preexistente de una base de conocimientos identificable; genera una continuación en función de la entrada y de los ajustes de generación.

Por qué el contexto es decisivo

El mismo LLM puede dar respuestas muy diferentes según la formulación de la solicitud y la información recibida. Por ejemplo, la instrucción «traduce este texto» no proporciona ni el idioma de destino ni el registro deseado. En cambio, «traduce este texto al español de España, con un registro profesional y sin modificar los nombres de productos» reduce varias ambigüedades.

El contexto también puede contener tensiones: una instrucción solicita una respuesta breve, mientras que un documento adjunto contiene numerosas excepciones; una solicitud exige una respuesta basada en los documentos proporcionados, pero ningún documento aborda realmente la cuestión. En este tipo de situación, la fluidez del texto producido no resuelve la falta de información inicial.

Instrucciones, contexto y parámetros: conceptos que se deben precisar

Para interactuar con un LLM de manera controlada, es necesario distinguir lo que corresponde al contenido de la solicitud de lo que corresponde a la configuración técnica.

Elemento	Pregunta que se debe plantear	Ejemplo
Instrucción	¿Qué debe hacer el modelo?	«Clasifica estos mensajes por tema.»
Contexto	¿En qué información debe basarse?	Los mensajes que se deben clasificar y una lista de temas permitidos.
Formato de salida	¿Cómo debe presentarse el resultado?	Una tabla con una fila por mensaje.
Parámetro de longitud	¿Qué extensión de respuesta se debe permitir?	Limitar el número de nuevos tokens generados.
Parámetro de diversidad	¿Debe priorizarse una salida estable o explorar más formulaciones?	Ajustar los mecanismos de muestreo.
Criterio de parada	¿Cuándo debe terminar la generación?	Alcance de un límite o aparición de una secuencia de parada.

Los parámetros de generación no sustituyen una instrucción clara. Actúan sobre el comportamiento de producción del texto; no transforman por sí solos una solicitud vaga en un resultado fiable. Las documentaciones técnicas de generación distinguen, en particular, las estrategias de decodificación y los parámetros asociados a la longitud, al muestreo y a la selección de tokens. [HF-GENERATION-CONFIG]

Ajustes y resultado: una relación que no se debe simplificar

© 2026 ASCALEA. Tous droits réservés.

Sería engañoso presentar un ajuste como una garantía. Aumentar o disminuir la diversidad de generación puede modificar el estilo o la variedad de las formulaciones, pero no verifica los hechos. Aumentar la longitud máxima puede permitir una respuesta más desarrollada, pero no mejora automáticamente su razonamiento ni su adecuación a la necesidad.

El control más sólido combina más bien:

1. una instrucción explícita;
2. un contexto seleccionado y organizado;
3. un formato de salida controlable;
4. parámetros adaptados a la tarea;
5. una verificación humana proporcionada a las consecuencias del uso.

Límites de fiabilidad, interpretación y control

Un LLM puede producir una respuesta plausible incluso cuando la información necesaria está ausente, es ambigua o es errónea. Este riesgo es especialmente importante cuando el usuario espera información verificable, una recomendación de alto impacto o una interpretación precisa de un documento.

El marco de gestión de riesgos de inteligencia artificial del National Institute of Standards and Technology (NIST) destaca, para la inteligencia artificial generativa, riesgos que incluyen, en particular, la producción de contenido confabulado, es decir, contenido falso o incoherente presentado de manera convincente. [NIST-AI-600-1]

> Principio de prudencia — La calidad de redacción de una salida no constituye una prueba de veracidad. Una respuesta debe evaluarse según su uso: una reformulación interna sin consecuencias no requiere el mismo nivel de verificación que una información médica, jurídica, financiera u operativa.

Los principales límites que se deben tener en cuenta

1. La plausibilidad no es la exactitud. El modelo busca generar una continuación textual coherente con su entrada y su configuración. Por tanto, puede presentar una explicación fluida que contenga errores, omisiones o referencias imprecisas. Los riesgos de confabulación identificados para la inteligencia artificial generativa obligan a no confundir una apariencia de seguridad con validación. [NIST-AI-600-1]

2. La interpretación depende del contexto proporcionado. Una solicitud incompleta puede conducir a una respuesta que parece responder a la pregunta, pero se basa en una interpretación no deseada. El usuario debe explicitar las restricciones importantes en lugar de suponer que se deducirán correctamente.

3. Los sesgos presentes en los datos y en la formulación pueden reflejarse en las salidas. Los datos pueden contener sesgos históricos, de medición, de selección o de agrupación; estos sesgos pueden influir en un sistema de aprendizaje automático y en los resultados que produce. [GOOGLE-FAIRNESS-BIAS] Por tanto, para una solicitud relativa a personas, grupos o decisiones sensibles, es necesario examinar no solo el texto final, sino también las categorías empleadas, los ejemplos proporcionados y las posibles consecuencias de su uso.

© 2026 ASCALEA. Tous droits réservés.

4. El modelo no sustituye una fuente primaria ni una validación especializada. Cuando una tarea exige información actualizada, una norma precisa, una decisión vinculante o un análisis de caso, una respuesta generada debe contrastarse con las fuentes y las personas competentes adecuadas. El marco del NIST recomienda una gestión de los riesgos de inteligencia artificial basada en prácticas de gobernanza, medición y gestión adaptadas al contexto de uso. [NIST-AI-RMF] [NIST-AI-600-1]

5. El control de la herramienta también depende de la aplicación utilizada. Los controles relativos a los datos, las aplicaciones conectadas o los permisos no son propiedades universales de cualquier LLM: dependen del proveedor, de la oferta y de la configuración. Las documentaciones de las plataformas distinguen, en particular, los controles de datos y los controles administrativos relacionados con las aplicaciones y los conectores. [OPENAI-DATA-CONTROLS] [OPENAI-CONNECTED-APPS-CONTROLS]

Un método de lectura crítica de una respuesta

Ante una salida producida por un LLM, resulta útil examinar cinco preguntas antes de utilizarla:

1. ¿La solicitud era suficientemente precisa?

¿Se explicitaron el objetivo, el idioma, el público, las restricciones y el formato?

1. ¿El contexto proporcionado es suficiente y pertinente?

¿Estaban presentes los documentos, datos o extractos necesarios? ¿Se introdujo información contradictoria?

1. ¿La salida responde realmente a la solicitud?

Una respuesta elegante puede eludir una restricción, olvidar una excepción o responder a una pregunta distinta de la planteada.

1. ¿Qué afirmaciones requieren verificación?

Los hechos, cifras, referencias, recomendaciones y citas atribuidas deben verificarse cuando la importancia lo justifique.

1. ¿Cuál es el riesgo si la respuesta es errónea?

Cuanto mayor sea el impacto potencial, más rigurosos deben ser la validación humana, la trazabilidad de las fuentes y los controles.

Este enfoque no consiste en rechazar los LLM. Pretende emplearlos por lo que son: herramientas potentes de procesamiento y generación del lenguaje, cuyas producciones deben encuadrarse, revisarse y validarse según el contexto.

Puntos clave

- Un LLM procesa una entrada textual y genera una salida de manera secuencial, a partir de tokens y del contexto disponible. [HF-TOKENIZER] [HF-TEXT-GENERATION]
- La entrada efectiva puede incluir mucho más que la pregunta visible: instrucciones, historial, documentos y reglas añadidas por la aplicación.
- Una instrucción clara precisa como mínimo el objetivo, el contexto útil y el formato esperado.

© 2026 ASCALEA. Tous droits réservés.

• Los parámetros de generación influyen en la producción del texto, pero no garantizan ni exactitud ni pertinencia. [HF-GENERATION-CONFIG]

• Una respuesta fluida puede ser incompleta, errónea o inadecuada; las salidas con consecuencias deben verificarse y situarse en un proceso de control de riesgos. [NIST-AI-600-1] [NIST-AI-RMF]

© 2026 ASCALEA. Tous droits réservés.

CHAPITRE 5

Un gran modelo de lenguaje, o Large Language Model (LLM), no es una funcionalidad en sí misma: es un componente que puede integrarse en experiencias muy diferentes. Una misma tecnología puede impulsar una conversación, redactar un texto, reformular un documento, extraer elementos estructurados, buscar en una base documental o desencadenar una acción en otro sistema.

Para comprender una solución basada en un LLM, resulta útil separar tres niveles:

1. la solicitud del usuario: pregunta, instrucción, texto, archivo o dato estructurado;
2. la función esperada: responder, generar, transformar, resumir, clasificar, extraer o actuar;
3. el dispositivo en torno al modelo: interfaz, reglas de contexto, documentos proporcionados, herramientas conectadas, memoria y controles.

Esta distinción evita confundir la capacidad lingüística del modelo con las funciones añadidas por el producto que lo emplea. Por ejemplo, un asistente que «responde a partir de documentos internos» no se apoya únicamente en el modelo: también depende del mecanismo que selecciona los documentos, los transmite al modelo y presenta la respuesta.

> Punto de atención — Una salida formulada con seguridad no demuestra ni la exactitud de su contenido, ni la exhaustividad de las fuentes consultadas, ni la ejecución efectiva de una acción externa. Estas propiedades deben diseñarse, controlarse y verificarse en el sistema que rodea al LLM. El National Institute of Standards and Technology (NIST) recomienda identificar y gestionar los riesgos propios de los sistemas de inteligencia artificial generativa durante todo su uso. [NIST-AI-600-1]

© 2026 ASCALEA. Tous droits réservés.

Mapa de las funciones posibles

Función	Entrada habitual	Salida esperada	Pregunta principal de control
Conversación	Mensajes sucesivos	Respuesta en lenguaje natural	¿Está bien delimitado el contexto del intercambio?
Generación	Instrucción, plan o datos	Contenido nuevo	¿Respeta el resultado las restricciones solicitadas?
Transformación	Texto o contenido existente	Versión modificada	¿Se preservan el sentido y la información importante?
Síntesis	Documento o conjunto de contenidos	Resumen, puntos clave, esquema	¿Qué información se ha retenido u omitido?
Clasificación	Texto, imagen o dato según el sistema	Categoría, etiqueta o prioridad	¿Están definidas y probadas las categorías y los criterios?
Extracción	Documento semiestructurado o no estructurado	Campos estructurados	¿Puede vincularse cada valor con su fuente?
Búsqueda documental	Pregunta y corpus de referencia	Respuesta respaldada por documentos	¿Qué documentos se han consultado y con qué derechos?
Llamada a una herramienta	Instrucción y parámetros	Resultado de un servicio externo o acción	¿Está la acción autorizada, es trazable y reversible?

Interacción conversacional

La interacción conversacional es la forma más visible de uso de un LLM. El usuario formula una solicitud, recibe una respuesta y, posteriormente, precisa o corrige su intención a lo largo de varios turnos. El sistema puede entonces utilizar los mensajes anteriores como contexto de la conversación.

Esta modalidad es adecuada, en particular, para explorar un tema, ayudar en la redacción, asistir en un proceso o explicar un contenido. Sin embargo, no debe interpretarse como un diálogo con una fuente de autoridad. El LLM produce texto a partir de la instrucción y del contexto disponible; la conversación no transforma automáticamente sus respuestas en hechos verificados.

© 2026 ASCALEA. Tous droits réservés.

Diseñar una solicitud utilizable

La calidad de una interacción depende en gran medida de cómo se formule la solicitud. Una instrucción puede precisar:

- el papel esperado, por ejemplo, «explicar como a un principiante»;
- el objetivo que debe alcanzarse;
- los elementos de contexto útiles;
- el formato de respuesta deseado;
- los límites que deben respetarse, como un idioma, una longitud o información que no debe inventarse;
- los criterios de verificación o los elementos que deben señalarse como inciertos.

Las documentaciones de desarrollo recomiendan utilizar instrucciones explícitas y variables o plantillas de solicitudes cuando estas deban repetirse de forma coherente. [ANTHROPIC-PROMPT-BEST-PRACTICES]

Una conversación es especialmente sensible a la gestión del contexto. El sistema debe decidir qué mensajes anteriores, instrucciones permanentes y documentos asociados se transmiten al modelo. Esta selección influye directamente en la respuesta. Un intercambio largo también puede acumular formulaciones ambiguas, hipótesis erróneas o instrucciones incompatibles.

> Práctica de control — Para una decisión importante, reformular explícitamente el objetivo, los datos de partida y el resultado esperado en el mensaje actual. No asumir que todo elemento mencionado anteriormente sigue presente, es prioritario o se ha interpretado correctamente.

Generación y transformación de contenido

La generación produce contenido nuevo a partir de una instrucción: un borrador, una lista de ideas, un esquema, una respuesta, un título o una descripción. La transformación parte de un contenido existente y lo modifica: traducción, reformulación, corrección estilística, simplificación, cambio de tono, formato o adaptación a un público.

La frontera entre ambas resulta útil: en una generación, deben controlarse principalmente las restricciones de creación y la validez de las afirmaciones; en una transformación, también debe controlarse la fidelidad al contenido original.

Las bibliotecas técnicas dedicadas a los modelos de lenguaje describen la generación como una producción de secuencias de texto a partir de entradas y parámetros de generación. [HF-TEXT-GENERATION] Estos parámetros pueden influir en el comportamiento de producción y deben considerarse ajustes que deben probarse, más que una garantía de calidad o fiabilidad. [HF-GENERATION-CONFIG]

Generar con un marco claro

Una instrucción de generación se vuelve más controlable cuando distingue:

- los hechos proporcionados al modelo;
- las hipótesis autorizadas;

© 2026 ASCALEA. Tous droits réservés.

- los elementos prohibidos, en particular la información no confirmada;
- la estructura de salida esperada;
- el nivel de prudencia deseado cuando faltan datos.

Por ejemplo, una solicitud puede imponer la creación de un borrador precisando que todo dato ausente debe señalarse, no completarse mediante una invención. Esta regla es importante cuando el resultado está destinado a una publicación, a una comunicación institucional o a un uso profesional.

Transformar sin deformar

Una transformación suele parecer menos arriesgada que una generación, ya que se basa en un texto proporcionado. Sin embargo, puede modificar matices, omitir una salvedad, simplificar excesivamente una condición o introducir una interpretación. Esta dificultad es especialmente importante para los textos contractuales, normativos, médicos, financieros o técnicos.

Para reducir este riesgo, una interfaz puede pedir al sistema que proporcione, además del texto transformado:

- los pasajes modificados;
- los elementos que no ha podido interpretar con certeza;
- una comparación estructurada entre la entrada y la salida;
- una indicación clara cuando la solicitud va más allá de una simple reformulación.

La verificación humana sigue siendo necesaria cuando la fidelidad semántica o la exactitud de un documento tiene consecuencias importantes.

Síntesis, clasificación y extracción: perímetro que debe precisarse

Un LLM puede emplearse para organizar información, pero estas funciones abarcan objetivos distintos.

- La síntesis reduce o reorganiza un conjunto de contenidos para destacar las ideas principales.
- La clasificación asigna una entrada a una o varias categorías definidas de antemano.
- La extracción identifica en un contenido información específica y la devuelve en una forma estructurada.

Las herramientas de software para el procesamiento del lenguaje presentan estas tareas como canalizaciones distintas, con entradas y salidas adaptadas a cada objetivo. [HF-PIPELINES] Por tanto, el diseño debe partir del resultado operativo buscado, y no de la sola pregunta «¿se puede utilizar un LLM?».

Síntesis: reducir sin ocultar

Un resumen nunca es una copia completa. Selecciona elementos, establece una jerarquía y deja de lado detalles. Antes de utilizarlo, es necesario precisar:

- el tamaño o el nivel de detalle esperado;
- el público destinatario;
- los temas que deben conservarse obligatoriamente;
- la posible necesidad de distinguir los hechos, las opiniones, las decisiones y las incertidumbres;

© 2026 ASCALEA. Tous droits réservés.

- el documento o corpus exacto que sirve de base.

Cuando hay mucho en juego, resulta útil exigir referencias a los pasajes originales o conservar un vínculo explícito entre cada punto sintetizado y su material fuente. Una síntesis sin trazabilidad es difícil de auditar.

Clasificación: definir las clases antes de automatizar

Clasificar mensajes, tickets, documentos o solicitudes puede ayudar a priorizar un tratamiento. Pero una categoría solo es útil si su definición es estable y compartida. Antes de cualquier automatización, conviene documentar:

- la lista de categorías posibles;
- los criterios que las distinguen;
- el tratamiento de los casos múltiples o ambiguos;
- el umbral a partir del cual se requiere una revisión humana;
- las consecuencias asociadas a cada clasificación.

Los datos y las decisiones de categorización también pueden introducir o reproducir sesgos. Los recursos pedagógicos de Google subrayan la importancia de identificar las fuentes de sesgo en los datos y en las decisiones de modelización. [GOOGLE-FAIRNESS-BIAS]

Por tanto, una clasificación que influye en el acceso a un servicio, la prioridad de un expediente o la evaluación de una persona requiere una atención reforzada: pruebas con casos representativos, seguimiento de los errores y procedimiento de escalamiento humano.

Extracción: hacer verificable la salida

La extracción tiene como objetivo, por ejemplo, identificar nombres, fechas, importes, referencias o campos presentes en documentos. Su valor depende tanto de la estructura de salida como de la propia respuesta.

Un diseño robusto prevé:

1. un esquema de salida definido, con los campos esperados;
2. un valor explícito para la información ausente o incierta;
3. el texto fuente o la posición del pasaje que ha servido para extraer cada valor;
4. controles de formato y coherencia después de la respuesta del modelo;
5. una revisión humana para los campos cuyo error tiene una consecuencia significativa.

El LLM puede proponer una extracción; por sí solo no garantiza que el documento contenga realmente la información ni que la interpretación sea correcta.

Uso de documentos y bases de conocimiento

Muchas interfaces permiten adjuntar un archivo, consultar un conjunto documental o conectar una base de conocimiento. Esta capacidad responde a una limitación práctica: el modelo solo conoce los elementos transmitidos en su contexto y no dispone necesariamente de la información propia de una organización, un proyecto o una versión reciente de un documento.

© 2026 ASCALEA. Tous droits réservés.

El funcionamiento general comprende varias etapas:

1. el usuario o el sistema pone documentos a disposición;
2. un mecanismo selecciona los pasajes considerados pertinentes para una pregunta;
3. estos pasajes se proporcionan al LLM junto con la instrucción;
4. el modelo redacta una respuesta a partir de este contexto;
5. la interfaz puede mostrar los documentos o extractos utilizados.

Algunas interfaces de programación de aplicaciones, o Application Programming Interfaces (API), distinguen explícitamente las llamadas de búsqueda de archivos en el desarrollo de una respuesta. [OPENAI-TOOLS-REFERENCE]

Lo que no garantiza la incorporación de documentos

El hecho de cargar o conectar un documento no garantiza que:

- se seleccione el documento correcto;
- se recupere el pasaje decisivo;
- la respuesta reproduzca fielmente el documento;
- se hayan examinado todas las fuentes pertinentes;
- los derechos de acceso al documento sean adecuados para cada usuario.

Por tanto, es preferible presentar una respuesta documental como una respuesta basada en contenidos seleccionados, y no como una lectura exhaustiva o una validación automática de todo el corpus.

Controles previstos para un corpus documental

© 2026 ASCALEA. Tous droits réservés.

Tema de control	Pregunta que debe plantearse antes del uso
Perímetro	¿Qué documentos pueden consultarse y cuáles están excluidos?
Actualización	¿Cómo se gestionan las versiones, eliminaciones y actualizaciones?
Acceso	¿Puede un usuario obtener únicamente los contenidos para los que está autorizado?
Trazabilidad	¿Indica la respuesta los documentos o extractos utilizados?
Calidad	¿Son los documentos legibles, completos y suficientemente estructurados?
Confidencialidad	¿Qué datos se transmiten al proveedor del modelo o a la herramienta conectada?
Verificación	¿Quién controla la respuesta antes de que se utilice para una decisión o se difunda?

Los controles de datos propuestos en una plataforma pueden variar según el punto de acceso o el tipo de servicio utilizado; deben verificarse en la configuración realmente seleccionada. [OPENAI-DATA-CONTROLS]

Llamadas a herramientas e integraciones: conceptos que deben verificarse

Un LLM puede conectarse a herramientas externas. En este caso, ya no se limita a redactar una respuesta: puede solicitar a un sistema circundante que realice una búsqueda, consulte un dato, calcule un resultado, cree un elemento o transmita una instrucción a otro servicio.

Deben distinguirse dos etapas:

- proponer una llamada: el modelo identifica una herramienta y formula parámetros;
- ejecutar la llamada: la aplicación autoriza la acción, contacta con el servicio externo y recupera su resultado.

El modelo no debe considerarse la autoridad que decide por sí sola la ejecución. La aplicación que lo integra debe controlar las herramientas accesibles, las autorizaciones, los parámetros aceptados y las acciones que requieren una confirmación humana.

Los mecanismos de administración de aplicaciones conectadas pueden abarcar, en particular, la activación, la gestión y los requisitos de seguridad o cumplimiento asociados a las conexiones externas. [OPENAI-CONNECTED-APPS-CONTROLS] Los datos asociados a aplicaciones conectadas también requieren controles específicos sobre su acceso y uso. [OPENAI-GOOGLE-APP-DATA]

© 2026 ASCALEA. Tous droits réservés.

Niveles de riesgo de las integraciones

Una integración puede ser únicamente informativa o producir un efecto en un sistema externo. La diferencia es esencial.

Tipo de integración	Ejemplo de resultado	Control mínimo recomendado
Lectura	Mostrar un estado o buscar información	Limitar los datos accesibles y registrar las consultas
Cálculo	Producir un cálculo o una conversión	Verificar los parámetros y el resultado antes de un uso crítico
Preparación	Crear un borrador de mensaje o ficha	Prever una validación humana antes del envío o la publicación
Escritura	Modificar un registro o crear un objeto	Restringir los derechos, mostrar un resumen y conservar un registro
Acción irreversible o sensible	Enviar, eliminar, comprometer un gasto o modificar un derecho	Exigir una autorización explícita y un mecanismo de control reforzado

Las evaluaciones son un elemento útil para verificar un sistema antes y después de su despliegue, definiendo casos de prueba y criterios adaptados a la función correspondiente. [OPENAI-EVALS] Para las integraciones, los conjuntos de pruebas deben cubrir no solo la calidad del texto, sino también los rechazos esperados, los parámetros incorrectos, los casos ambiguos y los límites de autorización.

> Regla de diseño — Cuanto más pueda una herramienta modificar el mundo exterior, más deberá separarse la decisión de ejecución de la generación textual del LLM.

Memoria, personalización y contexto: nociones que deben enmarcarse

Los términos memoria, personalización y contexto se emplean con frecuencia juntos, aunque designan mecanismos diferentes.

- El contexto de sesión corresponde a la información proporcionada para responder en un intercambio determinado: mensajes anteriores, instrucción del sistema, documento adjunto o datos de tarea.
- La memoria designa información conservada más allá de un intercambio, con el fin de reutilizarla posteriormente.
- La personalización adapta el comportamiento o la presentación del sistema a un usuario, un rol, un equipo o una preferencia.

Estos mecanismos pueden mejorar la continuidad de la experiencia, pero también aumentan las cuestiones de gobernanza: ¿qué datos se memorizan, durante cuánto tiempo, con qué finalidad, quién puede consultarlos, modificarlos o eliminarlos?

© 2026 ASCALEA. Tous droits réservés.

Principios de delimitación

Antes de activar una memoria o una personalización, resulta útil definir:

1. la finalidad: ¿qué problema concreto resuelve la conservación de información?
2. el perímetro: ¿qué categorías de información pueden conservarse?
3. las exclusiones: ¿qué información nunca debe memorizarse?
4. la duración: ¿cuándo y cómo se elimina o revisa la información?
5. la visibilidad: ¿puede el usuario ver, corregir o eliminar lo que se conserva?
6. la separación: ¿permanecen aisladas de las demás las informaciones de un usuario, un equipo o un expediente?
7. la verificación: ¿puede una preferencia o un hecho memorizado estar obsoleto, ser erróneo o inadecuado para el contexto actual?

La personalización no debe transformar una información antigua en una verdad permanente. Una preferencia declarada puede evolucionar; un dato memorizado puede estar incompleto; un contexto profesional puede cambiar. Por tanto, el sistema debe permitir corregir y reevaluar la información retomada en una respuesta.

Elegir una funcionalidad en función del control necesario

La pregunta central no es únicamente «¿qué puede hacer el LLM?», sino «¿qué función puede utilizarse con un nivel de control proporcional a sus efectos?».

Una función de generación de borradores puede tolerar una intervención humana importante después de la producción. Por el contrario, una extracción que alimenta una base de datos, una clasificación que influye en una prioridad o una integración capaz de escribir en un sistema exige controles más estructurados.

El marco de gestión de riesgos del NIST invita a integrar la gobernanza, la cartografía de riesgos, su medición y su gestión en el ciclo de vida de los sistemas de inteligencia artificial. [NIST-AI-RMF] Aplicado a las funcionalidades descritas en este capítulo, este principio lleva a documentar:

- la función exacta proporcionada al usuario;
- los datos utilizados como entrada;
- las fuentes de contexto y las herramientas conectadas;
- las limitaciones conocidas de la salida;
- las verificaciones humanas y técnicas previstas;
- las acciones posibles a partir de esta salida;
- los responsables del seguimiento y la corrección.

Una interfaz de LLM bien diseñada no promete una inteligencia general sin límites. Explicita sus capacidades, sus fuentes de contexto, sus controles y los momentos en que el usuario debe verificar, decidir o autorizar una acción.

© 2026 ASCALEA. Tous droits réservés.

CHAPITRE 6

Un modelo de lenguaje grande (LLM, Large Language Model) no recibe una intención implícita: produce una respuesta a partir de los elementos presentes en el intercambio. Por tanto, la calidad de una solicitud no garantiza una respuesta exacta, pero hace que el resultado sea más específico, más fácil de revisar y más sencillo de corregir.

© 2026 ASCALEA. Tous droits réservés.

Este capítulo propone un método de trabajo aplicable a numerosos usos: redacción, síntesis, traducción, explicación, preparación de un plan o análisis de un contenido proporcionado. No se trata de encontrar una formulación «mágica». Las recomendaciones de formulación dependen del modelo, de la herramienta y de la tarea. La documentación técnica recomienda, en particular, aclarar las instrucciones, separar los elementos variables y probar las formulaciones en el contexto real de uso [ANTHROPIC-PROMPT-BEST-PRACTICES].

> Principio de prudencia — Una solicitud mejor formulada mejora las condiciones de producción de una respuesta; no transforma al LLM en una fuente de autoridad. Toda respuesta importante debe evaluarse según su uso, sus consecuencias y las fuentes disponibles.

Objetivo, contexto, restricciones y formato esperado

Una solicitud utilizable puede construirse en torno a cuatro elementos:

1. El objetivo: lo que se quiere obtener o decidir.
2. El contexto: la información necesaria para interpretar correctamente la solicitud.
3. Las restricciones: los límites que deben respetarse.
4. El formato esperado: la forma concreta de la respuesta.

Esta estructura reduce los elementos implícitos. Resulta especialmente útil cuando un mismo término puede designar varias realidades, cuando una respuesta debe adaptarse a un público específico o cuando debe poder integrarse en un documento existente.

1. Definir el objetivo

El objetivo describe el resultado buscado, no solo el tema general.

Solicitud demasiado amplia	Solicitud orientada a un objetivo
«Explícame los LLM.»	«Explica el papel de un LLM con un lenguaje accesible para una persona principiante, distinguiendo entre generación de texto y verificación de hechos.»
«Corrige este texto.»	«Revisa este texto en español, corrige los errores gramaticales e indica por separado las formulaciones cuyo significado siga siendo incierto.»
«Haz un resumen.»	«Resume el texto proporcionado en un máximo de cinco viñetas, sin añadir información ausente del texto.»

Un objetivo puede precisar:

- el público destinatario;
- la decisión o la acción que la respuesta debe respaldar;
- el nivel de detalle;

© 2026 ASCALEA. Tous droits réservés.

- el idioma de salida;
- la distinción entre hechos, hipótesis, sugerencias y preguntas abiertas.

2. Proporcionar el contexto necesario

El contexto es la información que permite al modelo producir una respuesta adecuada. Puede incluir un extracto de texto, una consigna editorial, un público, un vocabulario impuesto, un estado de avance o los datos que se deben analizar.

El nivel adecuado de contexto no es necesariamente el mayor volumen de información. Debe ser suficiente para la tarea y pertinente para la respuesta esperada. Cuando se transmiten documentos o datos a una herramienta, también conviene verificar las normas de tratamiento de datos aplicables al servicio utilizado, en particular si el contenido es confidencial, personal o estratégico [OPENAI-DATA-CONTROLS].

Ejemplo de contexto explícito:

``text Contexte : ce texte sera lu par des apprenants francophones qui commencent l’espagnol. Matériau à traiter : [texte fourni dans le message] Vocabulaire à conserver : « modelo de lenguaje », « instrucción », « verificación ». ``

Una práctica útil consiste en distinguir sin ambigüedad:

- las instrucciones, es decir, lo que el modelo debe hacer;
- el material, es decir, el texto, los datos o los ejemplos que se deben tratar;
- los criterios de éxito, es decir, las propiedades esperadas de la respuesta.

La separación visual de las secciones, mediante títulos o delimitadores, ayuda a que la solicitud sea legible y mantenible. Las guías de formulación de instrucciones recomiendan utilizar una estructura coherente y separar los elementos susceptibles de variar, en particular mediante plantillas de solicitudes y variables [ANTHROPIC-PROMPT-BEST-PRACTICES].

3. Enunciar las restricciones útiles

Las restricciones definen lo que la respuesta debe hacer, no hacer o señalar. Son más eficaces cuando pueden observarse durante la revisión.

Ejemplos de restricciones operativas:

- «Utiliza únicamente la información contenida en el extracto.»
- «Si falta información, formula una pregunta en lugar de suponerla.»
- «Distingue claramente las citas del texto y tus explicaciones.»
- «No propongas asesoramiento jurídico, médico o financiero personalizado.»
- «Utiliza un español de nivel principiante a intermedio.»
- «Limita la respuesta a 250 palabras.»

Evitar una instrucción no garantiza que el resultado no contenga nunca el elemento en cuestión. Por tanto, una restricción también sirve como criterio de control después de la generación.

© 2026 ASCALEA. Tous droits réservés.

4. Definir el formato de salida

El formato de salida reduce el trabajo de transformación posterior. Puede precisar la estructura, los apartados, la longitud, el idioma, el tono o un esquema de datos.

Necesidad	Formato que puede solicitarse
Preparar una decisión	Tabla: opción, ventajas, limitaciones, información faltante
Revisar un texto	Dos partes: «versión revisada» y después «puntos por confirmar»
Aprender vocabulario	Lista de tres columnas: término en español, traducción al francés, ejemplo breve
Preparar una síntesis	Cinco viñetas, seguidas de tres preguntas no resueltas
Utilizar una respuesta en un sistema	Estructura con campos nombrados y reglas explícitas de cumplimentación

En un uso automatizado, la forma de generación puede verse influida por parámetros técnicos. La documentación de Hugging Face describe, en particular, configuraciones de generación que controlan ciertos aspectos de la producción textual [HF-GENERATION-CONFIG]. No obstante, estos ajustes no sustituyen ni una instrucción clara ni un control del resultado.

Modelo de instrucción estructurado

```text
Objectif
Produire une explication courte de [sujet].

Contexte
Le lecteur est [public].
Le matériau de référence est : [données ou texte fourni]

Contraintes
- Utiliser uniquement le matériau fourni pour les affirmations factuelles.
- Signaler explicitement les informations absentes ou incertaines.
- Employer un vocabulaire adapté à [niveau].

Format attendu
1. Réponse principale en [nombre] puces maximum.
2. Termes importants définis en une phrase.
3. Questions ou limites à vérifier séparément.
```

Este modelo es un punto de partida, no una obligación. Para una solicitud muy sencilla, unas pocas líneas bien enfocadas son preferibles a una estructura pesada.

© 2026 ASCALEA. Tous droits réservés.

Aclaración de términos ambiguos

Las ambigüedades son frecuentes en las solicitudes habituales. Una palabra como «modelo», «resumen», «simple», «reciente», «seguro» o «profesional» puede recibir varias interpretaciones.

Antes de solicitar una respuesta definitiva, resulta útil precisar:

- el sentido elegido de un término;
- el período correspondiente, cuando este sea determinante;
- el territorio, la organización o el público afectados;
- el nivel de precisión deseado;
- la diferencia entre un hecho establecido, una interpretación y una recomendación.

Por ejemplo, la solicitud «redacta un texto sencillo en español» puede precisarse así:

``text Rédige un texte en espagnol destiné à des apprenants de niveau A2. Utilise des phrases courtes. Évite les expressions idiomatiques. Le texte doit expliquer la différence entre une instrucción et una respuesta generada. Ajoute ensuite un glossaire français-espagnol de cinq termes. ``

Cuando una ambigüedad no puede resolverse a partir del contexto, una respuesta fiable en su proceso debe hacerla visible. Entonces son posibles dos estrategias:

1. solicitar una aclaración antes de continuar;
2. proponer una respuesta condicional anunciando explícitamente la hipótesis adoptada.

La primera opción es preferible si la ambigüedad cambia considerablemente la naturaleza de la respuesta. La segunda puede ser adecuada para una tarea exploratoria, siempre que la hipótesis sea claramente identificable durante la revisión.

> Formulación útil — «Antes de responder, indica la información faltante que impediría una respuesta fiable. Si propones una hipótesis, sepárala de los elementos proporcionados.»

Iteración y reformulación

La primera respuesta suele ser un borrador de trabajo. La iteración consiste en examinar este borrador, identificar con precisión la diferencia con respecto a la necesidad y, después, formular una solicitud de corrección específica.

En lugar de decir únicamente «mejora», resulta más útil nombrar el cambio esperado:

- «Reduce esta respuesta a tres ideas principales.»
- «Conserva los ejemplos, pero elimina las afirmaciones que no estén respaldadas por el texto proporcionado.»
- «Transforma los párrafos en una tabla comparativa.»
- «Explica el segundo punto con un ejemplo en español y su traducción al francés.»
- «Enumera por separado los elementos que no puedes verificar.»

© 2026 ASCALEA. Tous droits réservés.

La documentación sobre formulación de instrucciones recomienda un enfoque empírico: definir un objetivo, probar las salidas y ajustar la instrucción según los resultados observados, en lugar de suponer que una única formulación será adecuada para todos los casos [ANTHROPIC-PROMPT-BEST-PRACTICES].

Un ciclo de trabajo breve

1. Formular una solicitud inicial con un objetivo y un formato claros.
2. Leer la respuesta en función de un criterio preciso: exactitud, cobertura, tono, estructura o idioma.
3. Identificar la discrepancia: información ausente, añadido no justificado, ambigüedad, longitud excesiva, error de formato.
4. Solicitar una revisión específica recordando el criterio correspondiente.
5. Verificar de nuevo antes de cualquier reutilización importante.

Este enfoque evita dos errores opuestos: aceptar de inmediato una respuesta atractiva o rehacer por completo una solicitud cuando basta una corrección limitada.

Conservar un registro de las versiones útiles

En un marco profesional o repetitivo, puede ser pertinente conservar:

- la instrucción utilizada;
- el contexto y los datos transmitidos;
- la versión del resultado seleccionada;
- los controles realizados;
- las correcciones humanas aplicadas.

Esta trazabilidad facilita la comparación entre formulaciones y la comprensión de los errores recurrentes. Los enfoques de gestión de riesgos en inteligencia artificial (IA) del National Institute of Standards and Technology (NIST) insisten en la gobernanza, la medición y la gestión de los riesgos relacionados con los sistemas de IA, incluida la inteligencia artificial generativa [NIST-AI-RMF] [NIST-AI-600-1].

Criterios de revisión y verificación

La verificación debe adaptarse al resultado y a su uso. Una reformulación interna de bajo impacto no exige el mismo nivel de control que un contenido destinado a un cliente, a una publicación, a una decisión o a un ámbito regulado.

Una tabla de control práctica

© 2026 ASCALEA. Tous droits réservés.

Criterio	Pregunta de revisión	Acción posible si no se cumple el criterio
Pertinencia	¿La respuesta responde realmente al objetivo?	Reformular el objetivo o solicitar una respuesta más específica.
Fidelidad al material proporcionado	¿La respuesta añade elementos ausentes de los datos?	Eliminar el añadido, solicitar referencias o marcar el elemento como no verificado.
Exactitud	¿Pueden confirmarse las afirmaciones importantes mediante una fuente competente?	Verificar en una fuente primaria o solicitar una validación cualificada.
Exhaustividad	¿Falta algún elemento esencial?	Añadir el contexto faltante o solicitar una sección complementaria.
Coherencia	¿La respuesta se contradice o contradice los datos disponibles?	Identificar los pasajes en conflicto y solicitar una revisión específica.
Formato	¿Se respetan la estructura, el idioma y la longitud?	Solicitar una reformulación sin modificar el contenido no verificado.
Tono y destinatario	¿El nivel de lenguaje está adaptado al lector?	Ajustar el público, el registro y los ejemplos.
Confidencialidad	¿Se han integrado o expuesto datos sensibles de manera inapropiada?	Suspender la difusión y aplicar las normas de la organización.

El control no consiste únicamente en buscar errores de detalle. También deben examinarse las omisiones, las generalizaciones excesivas, las formulaciones demasiado afirmativas y los vínculos de causalidad no demostrados.

Verificar las afirmaciones en lugar de su apariencia

Un texto fluido, estructurado y convincente puede contener información inexacta o imposible de establecer a partir del contexto proporcionado. Los perfiles de riesgo de la inteligencia artificial generativa publicados por el NIST identifican, en particular, los riesgos de contenidos confabulados y subrayan la importancia de evaluar las salidas según el contexto de uso [NIST-AI-600-1].

Para una afirmación importante, la revisión puede seguir este orden:

1. aislar la afirmación que debe verificarse;
2. determinar si procede del material proporcionado, de una fuente citada o de una producción sin fuentes;
3. consultar una fuente adecuada para la importancia del asunto;
4. corregir, matizar o eliminar la afirmación si no puede confirmarse;

© 2026 ASCALEA. Tous droits réservés.

5. conservar la distinción entre lo que está establecido y lo que sigue siendo incierto.

Cuando una respuesta incluye cifras, fechas, obligaciones, definiciones técnicas o citas, estos elementos merecen una atención reforzada. No basta con que se mencione una referencia: hay que verificar que realmente respalda la afirmación formulada.

Evaluar un uso repetitivo

Cuando una misma instrucción debe utilizarse a gran escala —por ejemplo, para clasificar mensajes, producir descripciones o resumir documentos— una revisión ocasional es insuficiente. Entonces resulta pertinente preparar un conjunto de casos representativos, incluidos casos ordinarios, casos límite y casos susceptibles de causar daño si se tratan incorrectamente.

Las herramientas de evaluación pueden ayudar a comparar sistemáticamente las salidas con criterios definidos; la documentación de OpenAI presenta las evaluaciones como un mecanismo que permite medir los resultados de un sistema [OPENAI-EVALS]. Sin embargo, una medición automatizada no debe confundirse con una garantía de calidad: los criterios elegidos, los datos de prueba y el examen humano siguen siendo determinantes.

Casos que requieren validación humana o especializada

El uso de un LLM debe suspenderse, limitarse o someterse a una validación competente cuando el coste de un error es elevado, cuando la información es insuficiente o cuando el resultado va más allá de un simple trabajo de redacción o exploración.

Una validación humana o especializada es particularmente necesaria en las siguientes situaciones:

- decisiones relativas a una persona, especialmente cuando pueden afectar a sus derechos, a su acceso a un servicio, a su empleo, a su salud o a su reputación;
- asesoramiento jurídico, médico, financiero o de seguridad, cuando el contenido pudiera guiar una acción con consecuencias significativas;
- comunicaciones externas que comprometen a una organización, como un anuncio regulatorio, una respuesta contractual, una información de crisis o un documento oficial;
- tratamiento de datos sensibles, confidenciales o personales, cuando las normas de la organización o del servicio no permiten claramente este uso;
- contenido cuya exactitud depende de hechos recientes o de fuentes especializadas;
- resultados que sirven de entrada para una acción automatizada, una clasificación o una decisión sin control suficiente;
- situaciones en las que la respuesta contiene incoherencias, incertidumbres no resueltas o afirmaciones imposibles de fundamentar con fuentes.

Los marcos del NIST dedicados a la gestión de riesgos de la IA recomiendan integrar la gobernanza y la gestión de riesgos a lo largo de todo el ciclo de vida de los sistemas, prestando especial atención a los posibles efectos negativos de la inteligencia artificial generativa [NIST-AI-RMF] [NIST-AI-600-1].

© 2026 ASCALEA. Tous droits réservés.

Cuándo detenerse en lugar de reformular

Es razonable detener el intercambio con el modelo y solicitar la intervención de una persona competente cuando:

- no se dispone de la información necesaria;
- la respuesta varía mucho sin una razón identificable;
- es imposible verificar una afirmación esencial;
- la solicitud implica una decisión que el usuario no está autorizado a tomar;
- el tratamiento previsto incumple potencialmente las normas internas, los requisitos de confidencialidad o las condiciones de la herramienta.

Por tanto, la buena práctica no consiste ni en delegar ciegamente ni en renunciar a todo uso. Consiste en asignar al LLM un papel proporcional: producir un borrador, organizar ideas, reformular un contenido, proponer vías o preparar una estructura, mientras que las personas responsables conservan la verificación, el juicio y la decisión.

Síntesis operativa

Antes de utilizar una respuesta generada, verifique los siguientes puntos:

- el objetivo de la solicitud es explícito;
- se proporciona el contexto útil sin exponer indebidamente información sensible;
- los términos ambiguos están definidos o señalados;
- las restricciones y el formato permiten una revisión concreta;
- los hechos importantes se verifican en fuentes adecuadas;
- las incertidumbres, hipótesis y la información faltante son visibles;
- se realiza una validación humana o especializada cuando lo exige la importancia del asunto.

Una instrucción estructurada y una verificación proporcionada forman un mismo proceso. La primera organiza la producción; la segunda decide si el resultado puede utilizarse realmente.

© 2026 ASCALEA. Tous droits réservés.

CHAPITRE 7

Evaluar para tomar una decisión, no para designar el «mejor» modelo

Un modelo o una herramienta no es eficaz en términos absolutos: es más o menos adecuado para una tarea, una población de usuarios, un entorno técnico y un nivel de riesgo determinados. Una respuesta muy fluida puede ser inutilizable si omite información esencial; a la inversa, una respuesta factualmente prudente puede parecer menos natural y, aun así, ser preferible para una tarea sensible.

© 2026 ASCALEA. Tous droits réservés.

Por lo tanto, la elección debe partir del caso de uso y, a continuación, hacer corresponder criterios observables con una decisión. El National Institute of Standards and Technology (NIST) propone, en su marco de gestión de riesgos de la inteligencia artificial, integrar la medición, la documentación y la gestión de riesgos a lo largo de todo el ciclo de vida de un sistema de inteligencia artificial. Su perfil dedicado a la inteligencia artificial generativa destaca, en particular, la necesidad de evaluar las capacidades, las limitaciones y los riesgos en su contexto de uso. [NIST-AI-RMF] [NIST-AI-600-1]

> Principio rector — No se pregunte «¿Cuál es el mejor modelo?», sino: «¿Qué modelo o herramienta satisface mejor nuestros criterios de éxito, con limitaciones conocidas y aceptables, para este caso de uso concreto?»

Este capítulo propone un protocolo reutilizable. No produce una clasificación universal de productos o modelos: tal comparación exigiría un alcance, versiones, tarifas, parámetros y pruebas fechadas.

Definir el caso de uso y los criterios de éxito

Antes de comparar resultados, describa con precisión lo que el sistema debe lograr. Una fórmula como «usar un LLM para la atención al cliente» sigue siendo demasiado vaga: responder una pregunta sencilla, resumir un expediente, clasificar una solicitud, redactar un borrador o buscar información en una base documental son tareas distintas, con criterios distintos.

La ficha mínima de definición

Una ficha de caso de uso puede caber en una página si responde claramente a las siguientes preguntas:

© 2026 ASCALEA. Tous droits réservés.

Elemento	Preguntas que deben documentarse
Objetivo	¿Qué resultado concreto se desea obtener?
Usuario	¿Quién recibe o utiliza el resultado: colaborador, experto, cliente, sistema informático?
Entrada	¿Qué documentos, datos, instrucciones e idiomas se proporcionan?
Salida esperada	¿Qué formato, qué longitud, qué idioma y qué nivel de detalle se necesitan?
Decisión o acción	¿Qué permite hacer esta salida? ¿Se publica, controla, envía o ejecuta?
Consecuencia de un error	¿Un error es solo molesto, costoso, engañoso o potencialmente perjudicial?
Restricciones	¿Cuáles son los plazos, el presupuesto, las reglas de acceso, los requisitos de integración y los requisitos de confidencialidad?
Control humano	¿Quién verifica la salida, en qué momento y con qué posibilidad de corrección?

Para un libro o un proyecto destinado a contenidos en español, el idioma merece una línea específica: variante lingüística buscada, público objetivo, registro, terminología profesional, conservación o no de las citas y los nombres propios, y tolerancia a los regionalismos. Una evaluación realizada únicamente en francés no permite concluir sobre la calidad de una salida en español.

Transformar el objetivo en criterios observables

Un criterio es útil cuando una persona diferente puede, a partir de elementos definidos de antemano, observarlo y evaluarlo. Por ejemplo:

- «la respuesta es buena» es demasiado impreciso;
- «la respuesta responde a todas las preguntas obligatorias de la solicitud» es verificable;
- «las afirmaciones factuales están respaldadas por los documentos proporcionados» es verificable;
- «el resumen no supera las 150 palabras y conserva las tres decisiones principales» es verificable.

Los criterios deben estar vinculados a las consecuencias. Si una herramienta prepara borradores internos, la rapidez de revisión puede ser un criterio central. Si alimenta una decisión importante, la exactitud, la justificación, la trazabilidad y el control humano adquieren mayor peso.

© 2026 ASCALEA. Tous droits réservés.

Definir los umbrales antes de las pruebas

Fijar un umbral después de haber visto las respuestas favorece una comparación oportunista. Por ello, antes de las pruebas, defina:

1. los criterios eliminatorios, es decir, las condiciones sin las cuales se descarta la solución;
2. los criterios puntuados, que diferencian las soluciones restantes;
3. los umbrales de aceptación;
4. las personas autorizadas para juzgar los resultados;
5. la decisión que se tomará al finalizar la evaluación.

Un umbral no tiene que ser matemáticamente sofisticado. Puede adoptar la forma de: «ninguna respuesta debe divulgar datos no autorizados», «cada respuesta debe respetar el formato solicitado» o «las respuestas insuficientes deben poder ser detectadas por un revisor». Lo esencial es que la regla se conozca antes de la prueba.

Calidad del resultado: dimensiones que deben definirse

La calidad de una salida de modelo no es una medida única. Es preferible evaluar varias dimensiones por separado, con el fin de distinguir las fortalezas y las debilidades de una solución.

© 2026 ASCALEA. Tous droits réservés.

Dimensión	Pregunta de evaluación	Ejemplo de indicador o regla
Pertinencia	¿La salida aborda realmente la solicitud?	Presencia de los elementos solicitados; ausencia de divagaciones determinantes
Exactitud	¿Los enunciados pueden confirmarse mediante una fuente autorizada o una referencia de control?	Número de afirmaciones incorrectas, no respaldadas o contradictorias
Exhaustividad	¿Están presentes todos los elementos indispensables?	Porcentaje de criterios obligatorios cubiertos
Fidelidad a las fuentes	¿La respuesta reproduce correctamente los documentos proporcionados?	Ausencia de añadidos sin respaldo; remisión correcta al pasaje fuente cuando se requiere
Cumplimiento de las instrucciones	¿Se respetan las restricciones de idioma, formato, tono y longitud?	Tasa de conformidad con una rúbrica explícita
Claridad	¿El lector puede comprender y utilizar el resultado?	Evaluación por el público objetivo o por revisores designados
Utilidad operativa	¿La salida ahorra tiempo sin crear una carga desproporcionada de corrección?	Tiempo de revisión, número de correcciones sustanciales
Robustez	¿El resultado sigue siendo aceptable ante formulaciones o entradas representativas diferentes?	Resultados en un conjunto de casos variado
Equidad	¿Algunas categorías, variantes lingüísticas o formulaciones reciben un trato sistemáticamente menos favorable?	Comparación de los resultados entre grupos de casos pertinentes

La equidad no se reduce a una impresión general. Los sesgos pueden proceder de elecciones de datos, de la definición del objetivo, del muestreo o de variables sustitutivas; por ello, un análisis útil comienza por identificar los grupos, las decisiones y los efectos que importan en el caso de uso. [GOOGLE-FAIRNESS-BIAS]

Elegir una unidad de juicio adecuada

El resultado puede evaluarse en diferentes niveles:

- a nivel de la respuesta, por ejemplo: ¿una traducción es comprensible y fiel?
- a nivel de una afirmación, por ejemplo: ¿se confirma cada elemento factual?

© 2026 ASCALEA. Tous droits réservés.

- a nivel de la tarea, por ejemplo: ¿la síntesis permite realmente a un agente tramitar la solicitud?
- a nivel del flujo de trabajo, por ejemplo: ¿la herramienta reduce el tiempo total, incluidos el control y las correcciones?

Esta distinción evita una trampa habitual: medir únicamente la calidad textual cuando la necesidad se refiere a una acción completa. Un texto elegante no es necesariamente un resultado operativo.

Crear un conjunto de prueba representativo

El conjunto de prueba reúne las entradas que sirven para comparar las soluciones. Debe cubrir el trabajo esperado, pero también los casos que revelan las limitaciones: solicitudes ambiguas, documentos largos, instrucciones contradictorias, términos especializados, formatos inusuales, datos incompletos y casos poco frecuentes pero importantes.

Las características de los datos influyen considerablemente tanto en el aprendizaje como en la evaluación: la representatividad, la exhaustividad, la calidad de las etiquetas y las diferencias entre los datos de prueba y las condiciones reales pueden modificar la validez de las conclusiones. [GOOGLE-DATA-CHARACTERISTICS]

Mantenga una separación entre:

- los ejemplos utilizados para desarrollar las instrucciones o la configuración;
- los ejemplos utilizados para comparar las soluciones;
- los ejemplos recopilados posteriormente en condiciones reales para supervisar la diferencia entre la prueba y el uso.

Reutilizar continuamente los mismos casos para ajustar las instrucciones y luego anunciar una puntuación sobre esos mismos casos sobreestima la calidad práctica. Esta lógica se asemeja al sobreajuste: un sistema o un proceso puede parecer exitoso con ejemplos ya conocidos sin generalizar satisfactoriamente a situaciones nuevas. [GOOGLE-OVERFITTING]

Prever una rúbrica de puntuación

Una rúbrica reduce los juicios implícitos. Puede emplear una escala sencilla, por ejemplo de 0 a 2:

Puntuación	Interpretación genérica
0	No conforme, erróneo, inutilizable o riesgo inaceptable
1	Parcialmente conforme; se necesita una corrección o un control sustancial
2	Conforme con el criterio; se requiere una corrección menor o ninguna corrección

Cada puntuación debe completarse con una definición específica de la tarea. Para la fidelidad a un documento, «2» puede significar que las afirmaciones verificadas son compatibles con los pasajes proporciona-

© 2026 ASCALEA. Tous droits réservés.

dos; para el formato, «2» puede significar que todas las secciones solicitadas están presentes y correctamente estructuradas.

Fiabilidad, reproducibilidad y trazabilidad: conceptos que deben verificarse

Una respuesta satisfactoria aislada no basta para establecer que un sistema es fiable. También hay que examinar en qué condiciones se obtuvo y si puede explicarse, reproducirse y controlarse.

Fiabilidad: la calidad en una serie de casos

En este capítulo, la fiabilidad designa la capacidad práctica de producir resultados aceptables con suficiente regularidad dentro del alcance definido. Por lo tanto, se mide sobre una serie de casos, no sobre una única demostración.

Es útil registrar los fallos tanto como los éxitos: tipo de solicitud, gravedad, causa aparente, facilidad de detección y posible acción. Una salida errónea pero claramente incierta y fácil de detectar no tiene las mismas consecuencias que una salida errónea, persuasiva y difícil de controlar.

El perfil del NIST para la inteligencia artificial generativa identifica, entre otros, riesgos relacionados con la confabulación, la confidencialidad, los contenidos nocivos y la gestión de la información; estos riesgos deben examinarse según el contexto y la gravedad del caso de uso. [NIST-AI-600-1]

Reproducibilidad: repetir una prueba en condiciones descritas

La reproducibilidad operativa no significa necesariamente obtener palabra por palabra el mismo texto. Significa poder repetir una prueba con condiciones suficientemente documentadas para interpretar las diferencias observadas.

Para cada campaña de evaluación, registre como mínimo:

- el identificador del modelo o de la herramienta, así como la versión mostrada cuando esté disponible;
- la fecha y la hora de las pruebas;
- el modo de acceso utilizado, como una interfaz, una interfaz de programación de aplicaciones o un entorno local;
- las instrucciones del sistema y del usuario;
- los documentos, los datos de referencia y las herramientas conectadas autorizadas;
- los parámetros de generación pertinentes;
- el conjunto de prueba y su versión;
- la rúbrica de puntuación, los evaluadores y las reglas de arbitraje;
- las salidas sin procesar, las puntuaciones y las justificaciones de las puntuaciones.

En los sistemas de generación de texto, los parámetros pueden modificar el comportamiento de generación, en particular el carácter más o menos determinista o diverso de las salidas. Por lo tanto, la configuración de generación debe formar parte del protocolo, en lugar de dejarse implícita. [HF-GENERATION-CONFIG]

© 2026 ASCALEA. Tous droits réservés.

Trazabilidad: hacer que la salida sea controlable

La trazabilidad responde a una pregunta práctica: «¿Qué debe consultarse para comprender cómo se produjo esta salida y en qué se fundamenta?»

Es especialmente importante cuando una herramienta utiliza documentos internos, llama a funciones o accede a aplicaciones conectadas. Los controles de administración, los derechos de acceso y los parámetros de conexión deben examinarse antes del uso, ya que determinan qué datos pueden consultarse o compartirse en el flujo de trabajo. [OPENAI-CONNECTED-APPS-CONTROLS] [OPENAI-GOOGLE-APP-DATA]

Una salida trazable puede incluir, según el sistema:

- las referencias a las fuentes utilizadas;
- el historial de la solicitud y las instrucciones;
- la identificación de las herramientas llamadas;
- el estado de éxito o fallo de cada etapa;
- la identidad o la función de la persona que validó la salida;
- la versión del resultado efectivamente transmitido o publicado.

> Punto de atención — Una cita generada en una respuesta no es, por sí sola, una prueba de fiabilidad. Debe remitir a una fuente accesible, pertinente y efectivamente compatible con la afirmación que se supone que respalda.

Coste, plazo, integración y accesibilidad: criterios que deben documentarse

Un modelo puede obtener los mejores resultados en una rúbrica de calidad y seguir siendo inadecuado si es demasiado lento, demasiado costoso, inaccesible para los usuarios implicados o difícil de integrar en el proceso real.

Coste total, y no solo coste de acceso

Documente por separado los costes que influyen en la decisión:

- acceso al modelo o a la herramienta;
- desarrollo, configuración e integración;
- preparación y actualización de los contenidos de referencia;
- tiempo de control humano;
- formación y asistencia a los usuarios;
- seguimiento, evaluación periódica y gestión de incidentes;
- costes derivados de los errores, los retrasos o las repeticiones de trabajo.

© 2026 ASCALEA. Tous droits réservés.

A menudo resulta más instructivo medir el coste por resultado aceptable que el coste por solicitud. Una solución muy económica de utilizar puede generar un coste elevado si una gran proporción de sus salidas debe rehacerse.

Plazo y rendimiento

El plazo pertinente es el del proceso, desde la solicitud inicial hasta el resultado validado. Puede incluir la preparación de las entradas, la espera de respuesta, las llamadas a herramientas, la revisión humana y las correcciones.

Cuando la aplicación muestra los resultados progresivamente, el tiempo hasta el primer elemento visible puede mejorar la experiencia percibida, sin garantizar que el resultado final sea más rápido o más fiable. Las interfaces de programación de aplicaciones pueden ofrecer eventos de transmisión continua; es necesario distinguir este mecanismo de visualización de la duración total y de la calidad del procesamiento. [OPENAI-TOOLS-REFERENCE]

Integración en el trabajo existente

Evalúe la integración en casos concretos:

- ¿pueden los usuarios proporcionar fácilmente las entradas correctas?
- ¿el resultado llega en un formato reutilizable?
- ¿la herramienta se incorpora a las validaciones ya necesarias?
- ¿puede utilizar únicamente los datos y las conexiones autorizados?
- ¿los errores son visibles y recuperables?
- ¿el equipo puede cambiar de modelo, instrucción o proveedor sin reconstruir todo el proceso?

En las bibliotecas técnicas, una misma familia de modelos puede movilizarse mediante canalizaciones adaptadas a tareas diferentes. Este hecho recuerda que debe evaluarse el sistema completo —modelo, instrucciones, datos, herramientas e interfaz— y no únicamente el modelo. [HF-PIPELINES] [HF-TASKS-EXPLAINED]

Accesibilidad y condiciones reales de uso

La accesibilidad no se refiere únicamente al acceso técnico. También incluye el idioma de la interfaz, la legibilidad de las salidas, la compatibilidad con los equipos disponibles, la comprensión de los mensajes de error, los derechos de usuario y la capacidad de las personas para verificar el resultado.

Para una producción en español, pruebe con los contenidos, los registros y las variantes que realmente se utilizarán. Haga revisar los resultados por personas capaces de identificar las ambigüedades, los falsos amigos, los cambios de registro o las formulaciones culturalmente inadecuadas para el contexto previsto.

© 2026 ASCALEA. Tous droits réservés.

Protocolo de comparación que debe adaptarse al contexto

Un protocolo sencillo y explícito es preferible a una impresión general tras unas cuantas demostraciones. Las herramientas de evaluación propuestas en algunas plataformas permiten estructurar evaluaciones y registrar sus ejecuciones; independientemente de la herramienta elegida, el protocolo debe seguir siendo legible e independiente de un proveedor. [OPENAI-EVALS]

Paso 1 — Formular la decisión que debe tomarse

Enuncie la decisión final antes de la prueba. Por ejemplo:

- seleccionar una herramienta para un piloto limitado;
- mantener dos soluciones para una fase de integración;
- elegir una configuración de modelo para una tarea concreta;
- decidir que ninguna solución es lo suficientemente madura para el alcance previsto.

Este paso evita recopilar puntuaciones sin saber qué arbitraje deben esclarecer.

Paso 2 — Constituir un corpus de prueba

Seleccione casos representativos, incluidos los casos normales y las situaciones difíciles. Clasifíquelos, por ejemplo, según su tipo, complejidad, idioma, sensibilidad e impacto potencial.

No mezcle sin distinción los datos autorizados para la prueba y los datos que no lo están. Los parámetros de control de datos y las reglas de uso pueden variar según el endpoint, el producto o la configuración; verifique las condiciones aplicables al entorno elegido antes de transmitir información. [OPENAI-DATA-CONTROLS]

Paso 3 — Estabilizar las condiciones de comparación

Para comparar las soluciones, mantenga, en la medida de lo posible, constantes:

- las mismas entradas;
- los mismos documentos de referencia;
- la misma solicitud;
- el mismo formato de salida;
- el mismo plazo de respuesta esperado;
- el mismo método de puntuación.

Si también compara configuraciones, cree series separadas: comparación de los modelos con instrucciones constantes y, después, comparación de las instrucciones o los parámetros con un modelo constante. De lo contrario, resulta imposible saber qué explica la diferencia observada.

Paso 4 — Evaluar sin favorecer una solución

Cuando sea practicable, oculte el nombre de la solución a los evaluadores. Haga que varias personas evalúen las mismas salidas para los criterios importantes y, a continuación, examine las divergencias. Los

© 2026 ASCALEA. Tous droits réservés.

desacuerdos no son un fracaso: a menudo revelan que un criterio o una regla de puntuación debe precisarse.

Asocie los comentarios cualitativos a las puntuaciones. Una tabla de cifras muestra una diferencia; los ejemplos de errores permiten determinar si esa diferencia es aceptable, corregible o prohibitiva.

Paso 5 — Documentar las limitaciones

Toda comparación debe mencionar lo que no permite afirmar. Como mínimo, registre:

- el tamaño y la composición del conjunto de prueba;
- las categorías de casos ausentes o infrarrepresentadas;
- los parámetros, las versiones y las fechas de las pruebas;
- los criterios no medidos;
- las posibles diferencias con el uso real;
- los arbitrajes de ponderación;
- los incidentes, fallos y resultados no concluyentes;
- las dependencias de una integración, un conector o datos de referencia específicos.

El marco del NIST fomenta una gestión estructurada de los riesgos que tenga en cuenta el contexto, la documentación y las compensaciones. Por tanto, una conclusión responsable formula a la vez los beneficios observados, los riesgos residuales y las condiciones que deben respetarse para desplegar o continuar la evaluación. [NIST AI RMF] [NIST GENAI PROFILE]

Rúbrica de comparación reutilizable

La matriz siguiente sirve de apoyo para la decisión. Las ponderaciones y los umbrales deben adaptarse antes de las pruebas; no se proporcionan aquí como valores universales.

Criterio	Condición o pregunta	Peso definido por el proyecto	Solución A	Solución B	Prueba o comentario
Pertinencia	¿Responde a la tarea y al público previstos?	Por definir			
Exactitud y fidelidad	¿Los elementos verificables son correctos y están respaldados?	Por definir			
Cumplimiento de las instrucciones	¿Se respetan el formato, el idioma, la longitud y las restricciones?	Por definir			
Robustez	¿Resultados aceptables en los casos difíciles?	Por definir			
Riesgos y controlabilidad	¿Los fallos son detectables y gestionables?	Por definir			
Trazabilidad	¿Pueden recuperarse las entradas, las fuentes, las etapas y las versiones?	Por definir			
Plazo de extremo a extremo	¿El resultado validado está disponible dentro del plazo requerido?	Por definir			
Coste total	¿El coste incluye control, integración y operación?	Por definir			
Integración	¿Compatibilidad con los procesos, los datos y los derechos existentes?	Por definir			
Accesibilidad	¿Puede utilizarla el personal implicado en sus condiciones reales?	Por definir			
Reversibilidad	¿Sigue siendo viable cambiar de configuración o de proveedor?	Por definir			

© 2026 ASCALEA. Tous droits réservés.

Una puntuación global puede ayudar a sintetizar, pero no debe ocultar un criterio eliminatorio. Una solución que obtenga una buena media y, sin embargo, falle en un requisito esencial —confidencialidad, exactitud mínima, control humano o integración indispensable— no debería seleccionarse basándose únicamente en una puntuación agregada.

Concluir una evaluación de manera utilizable

Una conclusión útil no se limita a anunciar un ganador. Responde a cuatro preguntas:

1. ¿Para qué alcance se considera adecuada la solución?
2. ¿Qué resultados y qué pruebas respaldan esta conclusión?
3. ¿Qué limitaciones, riesgos e incertidumbres persisten?
4. ¿Qué condiciones deben reunirse antes del uso: control humano, restricciones de datos, seguimiento, formación o nueva prueba?

El resultado esperado es una decisión condicional y documentada: una solución puede seleccionarse para un piloto, solo para determinadas tareas, con validación humana obligatoria, o descartarse mientras sus limitaciones no estén controladas. Este enfoque transforma la comparación de modelos y herramientas en un proceso de selección profesional, anclado en la necesidad real y no en la demostración más convincente.

© 2026 ASCALEA. Tous droits réservés.

CHAPITRE 8

Emplear una Inteligencia Artificial (IA), y en particular un modelo de lenguaje de gran tamaño (LLM, Large Language Model), en un contexto operativo no consiste únicamente en obtener una respuesta útil. También es necesario decidir qué puede hacer la herramienta, sobre qué información, con qué nivel de control humano y según qué reglas de escalamiento.

La pregunta central no es, por tanto: «¿El resultado parece convincente?» Es: «¿Puede utilizarse este resultado para esta decisión, por esta persona, en este contexto, con controles proporcionales a las consecuencias de un error?» El marco de gestión de riesgos de IA del National Institute of Standards and Technology (NIST) propone abordar los riesgos relacionados con la IA de manera estructurada, identificándolos, evaluándolos, gestionándolos y documentándolos. Su perfil dedicado a la IA generativa señala, en particular, riesgos de contenido confabulado, confidencialidad, seguridad, sesgos y uso indebido. [NIST-AI-RMF] [NIST-AI-600-1]

> Principio rector — proporcionalidad del control respecto al impacto. Una reformulación de texto público y una recomendación que pueda influir en una decisión individual, financiera, médica, jurídica o de seguridad no justifican el mismo nivel de verificación, autorización y trazabilidad.

Una lectura práctica del riesgo

Antes de desplegar un uso, resulta útil examinar simultáneamente cuatro dimensiones:

© 2026 ASCALEA. Tous droits réservés.

Dimensión	Pregunta que formular	Ejemplo de medida que adaptar
Fiabilidad del contenido	¿Puede un error provocar un daño, una mala decisión o información engañosa?	Verificación humana, fuentes de referencia, prohibición de uso autónomo para decisiones críticas
Datos y confidencialidad	¿Los datos transmitidos están autorizados, son necesarios y están correctamente protegidos?	Minimización, anonimización cuando sea posible, reglas de acceso y configuración validada
Equidad e impactos	¿Puede el resultado perjudicar injustificadamente a personas o grupos?	Pruebas con casos variados, revisión de criterios, mecanismo de recurso o escalamiento
Seguridad y uso indebido	¿Puede la herramienta ser manipulada, utilizar accesos excesivos o desencadenar una acción no deseada?	Principio de mínimo privilegio, validación de acciones, registro y supervisión

Esta matriz no sustituye ni el análisis de negocio ni el análisis jurídico. Sin embargo, crea un lenguaje común entre los equipos que diseñan el uso, las personas que lo ejecutan, los responsables de seguridad y las funciones de cumplimiento.

Errores, contenidos inadecuados y límites de fiabilidad

Un LLM produce una salida a partir de patrones aprendidos y del contexto que se le proporciona; no es, por naturaleza, un sistema de prueba ni una autoridad competente. Una respuesta puede ser fluida, aparentemente precisa y, no obstante, errónea, incompleta, obsoleta, incoherente con el expediente real o mal adaptada a una excepción importante. El perfil del NIST para la IA generativa identifica, en particular, la confabulación —a menudo llamada «alucinación»— entre los riesgos que deben tenerse en cuenta. [NIST-AI-600-1]

Los errores pueden adoptar varias formas:

- invención de hechos, referencias o detalles que no figuran en los elementos proporcionados;
- omisión de una condición importante, de una excepción o de un caso particular;
- confusión entre hipótesis y hecho establecido;
- interpretación incorrecta de una instrucción ambigua;
- respuesta inadecuada para el público, el tono esperado o el nivel de sensibilidad del tema;
- variación de la respuesta cuando cambian la solicitud, el contexto o los parámetros de generación.

Los parámetros de generación pueden influir, en particular, en la diversidad y la selección de las salidas producidas; por tanto, conviene considerarlos un elemento del comportamiento del sistema que debe probarse, y no un simple detalle técnico. [HF-GENERATION-CONFIG]

© 2026 ASCALEA. Tous droits réservés.

Distinguir entre asistencia y decisión

Un uso se vuelve más riesgoso cuando la salida se utiliza directamente para decidir, actuar o comunicar sin revisión. Es útil distinguir tres niveles:

1. Asistencia de bajo impacto: generación de un borrador, reformulación, clasificación provisional o síntesis de contenido no sensible. Sigue siendo necesaria una revisión antes de la difusión.

2. Asistencia de impacto moderado: preparación de un análisis, priorización de expedientes, respuesta destinada a un cliente o elaboración de un documento interno. La persona responsable debe controlar los hechos, las hipótesis y la adecuación al caso tratado.

3. Uso de alto impacto: salida susceptible de afectar significativamente a una persona, desencadenar una acción irreversible, producir una opinión especializada o comprometer a la organización. El proceso debe definir un control humano cualificado, criterios de rechazo o escalamiento y una validación previa del marco aplicable.

La noción de «control humano» no debe reducirse a un clic de aprobación. Un control solo es útil si la persona que lo ejerce dispone de:

- las competencias necesarias para detectar un error;
- el tiempo, la información y la autoridad para corregirlo o rechazarlo;
- un procedimiento explícito para escalar los casos dudosos;
- acceso a las fuentes o los elementos que permitan verificar la respuesta.

Reducir el riesgo de contenido erróneo

Las medidas más eficaces dependen del caso de uso, pero varias prácticas suelen ser pertinentes:

- definir con precisión la tarea autorizada y las salidas esperadas;
- proporcionar un contexto limitado a lo necesario e indicar las fuentes de referencia que pueden utilizarse;
- pedir al sistema que señale la información ausente, las incertidumbres y las hipótesis en lugar de completarlas;
- verificar las afirmaciones importantes frente a fuentes independientes y competentes;
- probar el uso con casos normales, límite, ambiguos y contradictorios;
- prohibir presentar una salida no verificada como un hecho, un consejo especializado o una decisión final.

Las evaluaciones deben diseñarse en función del uso previsto, en lugar de basarse únicamente en la impresión general proporcionada por algunos ejemplos. La documentación de OpenAI dedicada a las evaluaciones ilustra la utilidad de definir y ejecutar evaluaciones estructuradas para seguir la calidad de un sistema. [OPENAI-EVALS]

© 2026 ASCALEA. Tous droits réservés.

Confidencialidad y tratamiento de la información

La primera protección consiste en considerar que la información introducida en una herramienta externa no es neutra: puede transmitirse, almacenarse, tratarse, conectarse con otros servicios o conservarse según la configuración, el producto y el contrato aplicables. Los parámetros de control de datos difieren según las plataformas y los servicios; deben examinarse en la documentación del proveedor y validarse dentro del marco adoptado. [OPENAI-DATA-CONTROLS]

Categorizar antes de transmitir

Antes de cualquier introducción de datos o envío de documentos, conviene clasificar la información al menos según las siguientes categorías:

Categoría de información	Pregunta de atención	Postura prudente antes de validar el marco
Pública	¿La información está realmente publicada y actualizada?	Puede utilizarse verificando la fuente y el contexto
Interna no sensible	¿Está autorizada su difusión a un proveedor o a una herramienta?	Verificar las reglas internas y la configuración del servicio
Confidencial	¿Su divulgación podría perjudicar a la organización, a un socio o a una persona?	No transmitir sin autorización explícita y medidas de protección adecuadas
Datos personales	¿El uso, la finalidad, el proveedor y las transferencias están permitidos por el marco aplicable?	No transmitir sin validación de los responsables competentes
Información altamente sensible	¿Una divulgación o un uso no controlado crearía un perjuicio grave?	Excluirla del uso mientras no se aprueben un análisis y controles específicos

La última columna no establece una regla jurídica universal. Refleja una regla operativa prudente: en ausencia de un marco aprobado, es preferible no exponer información sensible a una herramienta de IA.

Minimizar, separar, controlar

Una organización puede reducir la exposición aplicando tres reflejos:

- Minimizar: proporcionar únicamente la información necesaria para la tarea. A veces puede bastar un resumen o un extracto anonimizado.
- Separar: evitar mezclar, en una misma solicitud, información procedente de expedientes, personas o proyectos que no deberían asociarse.
- Controlar: identificar quién puede utilizar qué herramientas, con qué datos y según qué configuraciones.

© 2026 ASCALEA. Tous droits réservés.

Las integraciones y aplicaciones conectadas merecen especial atención. Pueden aumentar el valor de un uso al dar acceso a documentos o servicios, pero también incrementan el perímetro de acceso y las posibles consecuencias de una instrucción maliciosa o errónea. Por ello, los controles de administración, las autorizaciones y las modalidades de conexión deben examinarse antes de la activación. [OPENAI-CONNECTED-APPS-CONTROLS] [OPENAI-GOOGLE-APP-DATA]

> Para recordar: un documento puede parecer inocuo por sí solo y volverse sensible una vez combinado con un nombre, un calendario, un identificador, información de ubicación u otros elementos de contexto. La evaluación debe abarcar el conjunto de lo que se transmite y el uso previsto.

Sesgos, equidad e impactos potenciales

Un sistema puede producir efectos inequitativos incluso si no se formula ninguna intención discriminatoria. Los sesgos pueden proceder de los datos, de su recopilación, de las etiquetas utilizadas, de variables indirectas, de las elecciones de diseño, del contexto de despliegue o de la manera en que los usuarios interpretan los resultados. El curso de Google dedicado a la identificación de sesgos subraya, en particular, la importancia de examinar la representatividad de los datos, los sesgos históricos, los sesgos de medición y los sesgos vinculados al proceso de selección. [GOOGLE-FAIRNESS-BIAS]

Para un LLM, el riesgo no se limita a un cálculo de clasificación. Puede aparecer en:

- el vocabulario utilizado para describir a una persona o un grupo;
- los estereotipos reproducidos en una respuesta;
- la calidad desigual de las respuestas según la lengua, la variedad lingüística o el nivel de formulación;
- las recomendaciones o resúmenes que ignoran situaciones minoritarias;
- la automatización de una práctica ya cuestionable o insuficientemente explicada.

Preguntas de revisión antes de ponerlo en uso

Una revisión de equidad útil parte de las decisiones y de las personas potencialmente afectadas, y no solo del modelo. Las siguientes preguntas estructuran esta revisión:

1. ¿Quién se beneficia de la herramienta, quién la padece y quién puede quedar excluido por su funcionamiento?
2. ¿Qué grupos o situaciones corren el riesgo de estar peor representados en los casos de prueba?
3. ¿Qué criterios, palabras o señales pueden servir como sustitutos indirectos de características sensibles?
4. ¿Afecta un error de la misma manera a todos los públicos?
5. ¿Puede la persona afectada comprender, impugnar o solicitar la revisión de un resultado que la perjudica?
6. ¿El uso mantiene una decisión humana responsable cuando las consecuencias son importantes?

© 2026 ASCALEA. Tous droits réservés.

Los propios datos pueden reflejar problemas de cobertura, calidad, recopilación o distribución. Estas características deben examinarse porque influyen en los límites del sistema construido a partir de dichos datos. [GOOGLE-DATA-CHARACTERISTICS]

Las pruebas deben incluir formulaciones diversas, casos atípicos, lenguas pertinentes para el contexto de uso y escenarios en los que el modelo debería reconocer su incertidumbre o negarse a concluir. Un resultado medio aceptable no garantiza un rendimiento aceptable para cada población o cada situación.

Seguridad y usos indebidos: perímetro que definir

La seguridad de un uso de IA no depende únicamente del modelo. Depende de todo el dispositivo: interfaces, cuentas, documentos, conectores, herramientas invocadas, derechos otorgados, personas autorizadas y procedimientos de respuesta.

El perfil del NIST relativo a la IA generativa señala riesgos como ataques dirigidos a las entradas, la fuga de información, los abusos por parte de usuarios o actores maliciosos y las vulnerabilidades introducidas por los componentes del sistema. [NIST-AI-600-1]

Definir claramente lo que puede hacer la herramienta

El perímetro de autorización debería explicitar:

- las tareas autorizadas y las tareas prohibidas;
- los tipos de datos admitidos y excluidos;
- los usuarios, roles y niveles de acceso;
- las aplicaciones, bases documentales o servicios a los que la herramienta puede conectarse;
- las acciones que la herramienta solo puede proponer y aquellas que puede ejecutar;
- las validaciones humanas requeridas antes de una acción con efecto externo;
- las condiciones para suspender el uso en caso de incidente.

Cuando un sistema puede invocar una herramienta, buscar en archivos, acceder a una aplicación conectada o desencadenar una acción, una instrucción recibida en un documento o una página puede intentar influir en su comportamiento. Este riesgo exige no confundir el contenido procesado con una instrucción de confianza. Es prudente separar las instrucciones del sistema, los datos no fiables y las autorizaciones de acción.

Controles de seguridad proporcionales

Los controles que deben considerarse, según el perímetro, incluyen:

- una autenticación adecuada y una gestión de accesos basada en roles;
- el principio de mínimo privilegio: conceder solo los accesos necesarios;
- una separación entre entorno de pruebas y entorno de producción;
- una validación humana antes de cualquier envío, modificación, pago, publicación o eliminación;
- límites de volumen, frecuencia o coste para reducir el impacto de un abuso;

© 2026 ASCALEA. Tous droits réservés.

- la revocación rápida de accesos y conectores cuando se detecte un riesgo;
- el registro de acciones significativas y la revisión de anomalías.

Los detalles de implementación deben definirse con los equipos competentes en seguridad, arquitectura y operaciones. Dependen en gran medida del proveedor, de las integraciones realmente activadas y de la criticidad del proceso en cuestión.

Supervisión humana, gobernanza y trazabilidad

Una gobernanza útil asigna responsabilidades explícitas. Evita que la responsabilidad se diluya entre el usuario que formula la solicitud, el equipo que configura la herramienta, el área de negocio que utiliza la salida y el proveedor de tecnología.

El marco del NIST presenta la gestión de riesgos de IA como una actividad continua que abarca la gobernanza, la cartografía de contextos y riesgos, la medición y la gestión de los riesgos identificados. [NIST-AI-RMF]

Roles mínimos que aclarar

Rol	Responsabilidad que definir
Responsable del caso de uso	Justifica el objetivo, el perímetro, el valor esperado y el nivel de riesgo aceptable
Referente de negocio	Verifica que las salidas sean útiles, comprensibles y compatibles con el proceso real
Responsable de datos	Evalúa los datos admitidos, su calidad, su sensibilidad y las condiciones de acceso
Referente de seguridad	Analiza los accesos, las integraciones, las amenazas y las medidas de protección
Funciones jurídicas, de cumplimiento o de ética	Verifican los requisitos aplicables y los impactos sobre las personas, según el contexto
Usuario autorizado	Respeta las reglas de uso, verifica las salidas y comunica los incidentes o anomalías
Instancia de decisión o escalamiento	Dirime los usos sensibles, las excepciones y los incidentes significativos

En una estructura pequeña, una misma persona puede desempeñar varios roles. Esto no exime de documentar las responsabilidades ni de prever una mirada independiente cuando el impacto lo justifique.

© 2026 ASCALEA. Tous droits réservés.

Diseñar una cadena de escalamiento

Un procedimiento de escalamiento debe indicar qué hacer cuando una respuesta parezca peligrosa, errónea, discriminatoria, confidencial o contraria al perímetro autorizado. Puede prever:

1. la suspensión del uso de la salida en cuestión;
2. la conservación de los elementos necesarios para el análisis, respetando las reglas de seguridad y confidencialidad;
3. la notificación al punto de contacto designado;
4. el análisis del incidente y de su perímetro;
5. la corrección de la configuración, las instrucciones, los datos o el proceso;
6. la decisión de reanudar, limitar o retirar el uso.

Registrar lo que importa

La trazabilidad no implica necesariamente conservarlo todo, indefinidamente. Consiste en preservar, según una política definida, los elementos necesarios para comprender cómo se autorizó un uso, cómo se produjo una salida y cómo se validó una decisión.

Un registro de casos de uso puede incluir:

- el nombre y el objetivo del caso de uso;
- el propietario responsable;
- el público o proceso afectado;
- el modelo, la herramienta, la versión y las integraciones utilizadas;
- las categorías de datos admitidas y prohibidas;
- los riesgos identificados y los controles asociados;
- los criterios de aceptación y los resultados de las pruebas;
- las reglas de supervisión humana y escalamiento;
- las decisiones de aprobación, revisión o retirada;
- los incidentes, las limitaciones conocidas y las acciones correctivas.

Esta documentación también permite reexaminar un uso cuando evolucionan el modelo, el proveedor, los datos, las integraciones o el proceso de negocio. Una validación inicial no es suficiente si el contexto cambia de manera significativa.

Aspectos jurídicos y regulatorios que deben verificarse según la jurisdicción y el sector

Los requisitos jurídicos, regulatorios, contractuales y sectoriales no pueden deducirse de un principio general aplicable a todos los países y todos los usos. Deben analizarse en la jurisdicción correspondiente, para el sector, la organización, las categorías de datos y la finalidad específica del tratamiento.

Antes de ponerlo en producción, los responsables competentes deberían verificar, en particular:

© 2026 ASCALEA. Tous droits réservés.

- las reglas aplicables a los datos personales, confidenciales y a las posibles transferencias de datos;
- las obligaciones de seguridad, conservación, archivo y notificación aplicables al contexto;
- los contratos con proveedores, subcontratistas, socios y clientes;
- los derechos sobre los datos proporcionados, los documentos utilizados y los contenidos producidos;
- los requisitos de transparencia, información, consentimiento o control humano que puedan ser aplicables;
- las reglas específicas de los sectores regulados o de las profesiones sujetas a obligaciones particulares;
- las reglas relativas al empleo, la no discriminación, la protección de los consumidores, la propiedad intelectual o la prueba, cuando el uso les concierna;
- las condiciones en las que una recomendación o una salida automatizada puede contribuir a una decisión que afecte a una persona.

Ninguna de estas verificaciones puede sustituirse por las condiciones de uso de una herramienta o por una afirmación generada por un LLM. Los documentos del proveedor pueden informar sobre las funcionalidades y los parámetros disponibles; no determinan por sí solos la conformidad de un uso concreto. Las decisiones deben ser tomadas por personas que dispongan de la competencia y la autoridad adecuadas.

> Decisión de puesta en uso: una regla sencilla No desplegar un caso de uso mientras su objetivo, sus datos admitidos, sus límites, su control humano, sus responsabilidades, sus medidas de seguridad y su marco de validación aplicable no estén suficientemente definidos.

Síntesis operativa

Un uso responsable de la IA se basa en una disciplina de diseño y operación:

- no tratar la fluidez de una respuesta como una garantía de veracidad;
- limitar los datos transmitidos y excluir la información sensible mientras no se haya validado el marco;
- buscar impactos inequitativos sobre las personas y las situaciones menos visibles;
- reducir los accesos, regular las integraciones y controlar las acciones;
- designar responsables, organizar el escalamiento y conservar una trazabilidad proporcional;
- hacer que se examinen las cuestiones jurídicas, regulatorias, contractuales y sectoriales en su contexto real.

El objetivo no es eliminar todo riesgo —lo que rara vez es posible—, sino hacer que los riesgos sean visibles, debatibles y gestionables antes de que se conviertan en incidentes.

© 2026 ASCALEA. Tous droits réservés.

CHAPITRE 9

Un caso de uso no se reduce a la idea de «poner un modelo de lenguaje grande» — Large Language Model (LLM) — delante de una tarea. Describe un servicio preciso, prestado a usuarios identificados, a partir de entradas definidas y bajo controles explícitos. Esta descripción es indispensable para distinguir una demostración convincente de un uso que pueda encuadrarse de forma duradera.

Los ejemplos de este capítulo son deliberadamente genéricos. No expresan una recomendación para una organización, un sector o una herramienta: el contexto de aplicación, las obligaciones aplicables, los datos disponibles y el nivel de riesgo aceptable no se definen en la presente obra.

© 2026 ASCALEA. Tous droits réservés.

> Principio rector: no se despliega un modelo; se organiza un uso. El modelo no es más que un componente entre los usuarios, los datos, las reglas, las interfaces, los controles y los responsables.

El marco de gestión de riesgos del National Institute of Standards and Technology (NIST) estructura este enfoque en torno a las funciones de gobernanza, mapeo, medición y gestión de riesgos. El perfil dedicado a la inteligencia artificial (IA) generativa completa este enfoque para las particularidades de los sistemas que producen contenido. [NIST-AI-RMF] [NIST-AI-600-1]

1. Describir un caso de uso antes de elegir una solución

La primera pregunta no es «¿qué modelo elegir?», sino «¿qué trabajo se busca apoyar, para quién y dentro de qué límites?». Una descripción suficientemente concreta evita dos errores frecuentes: evaluar una herramienta sobre ejemplos demasiado favorables o esperar de un sistema generativo una exactitud que no ha sido organizada por el proceso.

Un caso de uso puede formularse como una cadena simple:

1. Una necesidad observable: una tarea repetitiva, lenta, difícil de homogeneizar o que requiere explorar un volumen importante de contenido.
2. Un usuario y una decisión: la persona que utiliza el resultado, lo que hace con él y aquello que sigue siendo la única habilitada para decidir.
3. Entradas autorizadas: textos, documentos, datos estructurados, instrucciones o bases de conocimiento efectivamente accesibles al sistema.
4. Una transformación esperada: sintetizar, clasificar, extraer, redactar, traducir, buscar, explicar o asistir otra etapa de trabajo.
5. Una salida utilizable: formato, idioma, nivel de detalle, fuentes eventuales, criterios de calidad y destinatario.
6. Controles y un límite de alcance: lo que el sistema no debe hacer, lo que debe verificarse y las situaciones en las que el uso debe interrumpirse o reorientarse.

Las bibliotecas de software dedicadas a los modelos de tipo Transformer proponen, entre otras, tareas de generación, clasificación y otras transformaciones de texto. Esta diversidad de tareas no debe ocultar que cada una exige sus propias entradas, medidas y modalidades de control. [HF-TASKS-EXPLAINED] [HF-PIPELINES]

2. La plantilla descriptiva de un caso de uso

La plantilla siguiente sirve para hacer debatible un uso antes de hacerlo operativo. Puede completarse para un prototipo y enriquecerse a medida que avanzan las pruebas.

© 2026 ASCALEA. Tous droits réservés.

Sección	Preguntas que documentar	Ejemplo de formulación genérica
Problema tratado	¿Qué dificultad concreta se busca reducir?	Reducir el tiempo de primera lectura de un conjunto de documentos.
Usuarios	¿Quién solicita el sistema? ¿Quién lee o utiliza el resultado?	Unos analistas preparan una síntesis destinada a un responsable.
Decisión apoyada	¿El resultado informa, prepara o desencadena una decisión?	El resultado prepara una revisión humana; no decide.
Entradas	¿Qué datos, documentos e instrucciones se admiten?	Documentos seleccionados, instrucción de síntesis y formato de entrega.
Tratamiento esperado	¿Qué operación se supone que debe efectuar el sistema?	Extraer los temas, las divergencias y los elementos que deben verificarse.
Salida	¿Qué entregable debe producirse?	Una nota estructurada, con remisión a los pasajes de origen cuando el dispositivo lo permita.
Criterios de calidad	¿Cómo se reconoce un resultado útil o insuficiente?	Fidelidad a los documentos, cobertura de los puntos importantes, legibilidad y señalización de las incertidumbres.
Controles	¿Quién verifica qué y en qué momento?	El usuario controla los pasajes importantes antes de cualquier difusión.
Límites	¿Qué casos se excluyen o exigen una vía diferente?	Expedientes incompletos, solicitud ambigua, datos no autorizados o consecuencias elevadas.
Responsables	¿Quién responde por el contenido, la configuración y el seguimiento?	Se identifican un responsable de negocio, un responsable técnico y un responsable del control.

Esta plantilla pone de manifiesto una distinción esencial: la capacidad de producir una respuesta no establece la pertinencia de esa respuesta para un proceso determinado. Por lo tanto, la evaluación debe partir de los resultados esperados en el trabajo real, y no solo de una impresión de fluidez o de calidad de redacción.

3. Cuatro formas de casos de uso genéricos

Los casos siguientes ilustran estructuras de uso, no recetas de despliegue. Una misma herramienta puede ser adecuada para varias categorías, pero los criterios de aceptación cambian según la tarea.

© 2026 ASCALEA. Tous droits réservés.

3.1 Asistencia a la redacción

Finalidad. Producir una primera versión de un texto a partir de una instrucción, un esquema y elementos de contexto proporcionados por el usuario.

Entradas esperadas. Un objetivo de comunicación, una audiencia, restricciones de forma, información factual validada y, si es necesario, ejemplos de estilo.

Salida esperada. Un borrador modificable, presentado explícitamente como tal.

Principal punto de atención. La calidad estilística puede ocultar información ausente, errónea o insuficientemente matizada. Por lo tanto, la revisión debe centrarse en las afirmaciones, las referencias, las cifras, las formulaciones sensibles y la adecuación al objetivo, y no solo en la corrección lingüística.

Condición de éxito. El sistema reduce el esfuerzo de preparación sin eliminar la responsabilidad editorial de la persona que valida el texto.

Las buenas prácticas para formular solicitudes hacen hincapié en la claridad de las instrucciones, el uso de variables o plantillas cuando el contexto se repite, y una estructuración que haga el resultado más controlable. [ANTHROPIC-PROMPT-BEST-PRACTICES]

3.2 Síntesis y exploración documental

Finalidad. Ayudar a recorrer un corpus y preparar una lectura humana más focalizada.

Entradas esperadas. Una colección de documentos cuyo origen, completitud y autorización de uso se conocen; una pregunta precisa; un alcance temporal o temático, si es necesario.

Salida esperada. Una síntesis, una lista de temas, divergencias o cuestiones que profundizar. Cuando la arquitectura lo permite, la salida puede incluir remisiones a los documentos consultados para acelerar la verificación humana.

Principal punto de atención. Una síntesis no prueba ni que se hayan tenido en cuenta todos los documentos pertinentes ni que se hayan conservado los matices. El proceso debe determinar qué resultados exigen volver a las fuentes y cómo señalar las zonas no cubiertas.

Condición de éxito. El usuario puede rastrear los elementos importantes hasta el contenido de origen y distinguir la información extraída de una interpretación generada.

3.3 Extracción y estructuración de información

Finalidad. Localizar en textos elementos definidos de antemano y devolverlos en una estructura utilizable: campos, categorías, tablas o listas.

Entradas esperadas. Documentos suficientemente legibles, una definición no ambigua de los campos que se deben extraer, reglas para los valores ausentes o inciertos y un formato de salida determinado.

Salida esperada. Datos estructurados acompañados, cuando resulte útil, de un nivel de confianza operativo o de una indicación «por verificar».

© 2026 ASCALEA. Tous droits réservés.

Principal punto de atención. La estructura de salida puede dar una impresión de precisión superior a la de los documentos de entrada. Los casos incompletos, contradictorios o atípicos deben anticiparse en lugar de forzarse dentro de un campo.

Condición de éxito. La proporción de resultados correctamente utilizables se mide sobre un conjunto de casos representativo, incluidos documentos difíciles y excepciones.

3.4 Asistencia conversacional basada en un perímetro de conocimiento

Finalidad. Responder a preguntas recurrentes u orientar a un usuario dentro de un conjunto definido de contenidos.

Entradas esperadas. La pregunta del usuario, documentos de referencia seleccionados, instrucciones de respuesta y una regla para tratar las solicitudes fuera de alcance.

Salida esperada. Una respuesta concisa, una orientación hacia una fuente, una solicitud de precisión o una negativa explícita a responder cuando el sistema no dispone de un fundamento suficiente.

Principal punto de atención. El perímetro de conocimiento debe ser explícito. Sin un mecanismo claro de recuperación, selección o actualización de los contenidos, una respuesta puede parecer segura y, sin embargo, ser inadecuada para el contexto buscado.

Condición de éxito. Los usuarios saben qué cubre el asistente, qué no cubre y cómo obtener una respuesta humana o una fuente de referencia cuando sea necesario.

4. De la necesidad a la funcionalidad: criterios que confirmar

La elección de un modelo, una interfaz o una funcionalidad se produce después de delimitar el trabajo que se debe realizar. Los criterios siguientes no constituyen una clasificación universal; sirven para organizar las confirmaciones necesarias.

© 2026 ASCALEA. Tous droits réservés.

Criterio	Lo que debe confirmarse	Consecuencia si el punto sigue sin estar claro
Adecuación a la tarea	¿El sistema trata realmente el tipo de entrada y salida previsto?	Las pruebas pueden centrarse en una capacidad impresionante pero irrelevante.
Calidad esperada	¿Qué defectos son tolerables, cuáles bloquean y cómo detectarlos?	El equipo no puede decidir objetivamente si el resultado es aceptable.
Controlabilidad	¿Pueden encuadrarse las instrucciones, el formato de salida, los accesos y los usos autorizados?	Los resultados y las prácticas se vuelven difíciles de estabilizar.
Trazabilidad	¿Pueden conservarse los elementos necesarios para analizar un resultado o un incidente?	Resulta difícil comprender, corregir o auditar un comportamiento.
Integración en el trabajo	¿En qué momento consulta el usuario el resultado y cómo lo corrige?	La ganancia teórica no se traduce en beneficio operativo.
Gestión de datos	¿Qué datos pueden entrar, transitar o ser accesibles a través del dispositivo?	El uso puede exceder el perímetro de autorización definido.
Continuidad	¿Qué sucede si evolucionan el servicio, el modelo, los documentos de referencia o los parámetros?	Las prestaciones observadas durante la prueba pueden dejar de ser representativas.

En un sistema generativo, los parámetros de generación intervienen en el comportamiento de salida: las opciones de generación permiten, en particular, configurar la forma en que un modelo produce texto. Por lo tanto, deben documentarse en un protocolo de evaluación cuando su modificación pueda hacer evolucionar los resultados. [HF-GENERATION-CONFIG] [HF-TEXT-GENERATION]

Del mismo modo, la segmentación del texto en unidades tratadas por el modelo —la tokenización— y los límites asociados pueden tener una incidencia práctica en el procesamiento de contenidos extensos. Por lo tanto, la composición de los documentos de prueba debe representar los volúmenes y formatos realmente previstos. [HF-TOKENIZER]

5. Diseñar los controles como una parte del servicio

Un control útil no es una etapa decorativa añadida después de la generación. Responde a una pregunta concreta: ¿qué defecto se busca detectar, antes de qué consecuencia, por qué persona o qué mecanismo?

Los controles pueden repartirse en cuatro niveles complementarios.

Controles sobre las entradas

© 2026 ASCALEA. Tous droits réservés.

Definen los datos admitidos, las fuentes autorizadas, el nivel de preparación de los documentos y las instrucciones utilizables. También pueden prever una orientación hacia otro procedimiento cuando una solicitud es ambigua, incompleta o está fuera de alcance.

La calidad, la cobertura y la representatividad de los datos influyen en los resultados de un sistema de aprendizaje automático; los sesgos también pueden introducirse o revelarse mediante los datos y las decisiones de medición. [GOOGLE-DATA-CHARACTERISTICS] [GOOGLE-FAIRNESS-BIAS]

Controles sobre el comportamiento esperado

Se refieren al formato de la salida, las formulaciones prohibidas, las solicitudes de aclaración, las situaciones de negativa y las instrucciones de prudencia. En un uso conversacional, las plantillas de instrucciones coherentes facilitan la repetibilidad de la interacción. [ANTHROPIC-PROMPT-BEST-PRACTICES]

Controles sobre las salidas

Organizan la revisión humana: muestreo, verificación sistemática de determinados campos, contraste con las fuentes, validación antes de la difusión o tratamiento particular de las respuestas señaladas como inciertas. Su intensidad debe adaptarse a las consecuencias de un error en el contexto real.

Controles sobre el acceso y las conexiones

Cuando un asistente puede acceder a aplicaciones o datos externos, el perímetro de las conexiones, los derechos concedidos y los mecanismos de administración forman parte integrante del caso de uso. Las documentaciones de las plataformas distinguen, en particular, los controles vinculados a las aplicaciones conectadas y los parámetros de control de datos; estos elementos deben examinarse en la configuración realmente prevista. [OPENAI-CONNECTED-APPS-CONTROLS] [OPENAI-DATA-CONTROLS] [OPENAI-GOOGLE-APP-DATA]

> Para recordar: la automatización de una acción y la asistencia a una acción no presentan la misma necesidad de control. Cuanto más directamente produzca un efecto la salida, más explícitas deben ser las condiciones de autorización, verificación y reversión.

6. Asignar las responsabilidades sin diluirlas

El vocabulario de la IA puede crear una dilución de responsabilidad: el modelo «respondió», la herramienta «decidió» o el sistema «encontró». En un uso encuadrado, las responsabilidades siguen siendo humanas y organizacionales.

Una distribución mínima puede adoptar la siguiente forma:

© 2026 ASCALEA. Tous droits réservés.

Rol	Responsabilidad principal
Responsable de la necesidad	Define el objetivo, los límites de uso y los criterios de valor.
Usuario	Formula la solicitud dentro del marco autorizado, examina el resultado y aplica las reglas de validación.
Responsable del contenido o de la decisión	Asume la validación del resultado cuando se utiliza en un entregable o una decisión.
Responsable técnico	Configura, integra, supervisa y hace evolucionar el dispositivo de acuerdo con el perímetro convenido.
Función de control competente	Examina los riesgos, los datos, los accesos y los requisitos aplicables según el contexto de la organización.

Esta distribución debe adaptarse a la organización real. No sustituye ni al análisis de las obligaciones que eventualmente sean aplicables ni a los procedimientos internos de seguridad, calidad, cumplimiento o gestión documental.

7. Pasar de una experimentación a un uso encuadrado

Una experimentación suele responder a la pregunta: «¿el sistema parece útil en algunos ejemplos?». Un uso encuadrado debe responder a una pregunta más exigente: «¿puede el servicio funcionar de manera repetible, controlada y comprensible en las condiciones previstas?».

La transición entre estos dos estados puede organizarse en cinco secuencias.

1. Delimitar el alcance. Definir la tarea, los usuarios, las entradas, las salidas, las exclusiones y las consecuencias de un resultado defectuoso.
2. Diseñar el dispositivo. Determinar las fuentes, las instrucciones, el modelo o la funcionalidad que se debe examinar, la interfaz, los accesos y los puntos de control.
3. Evaluar sobre casos representativos. Constituir un conjunto de casos que cubra las situaciones ordinarias, los casos límite, las ambigüedades y las entradas insuficientes. Comparar las salidas con los criterios definidos antes de la prueba.
4. Decidir las condiciones de uso. Formalizar quién puede utilizar el servicio, para qué tareas, con qué datos, según qué validación y con qué procedimiento de escalamiento.
5. Hacer seguimiento y revisar. Reexaminar el uso cuando cambien los datos, los documentos de referencia, el modelo, los parámetros, los usuarios o el contexto.

Existen herramientas y enfoques de evaluación para estructurar conjuntos de pruebas, ejecutar evaluaciones y analizar los resultados; sin embargo, no eximen de definir qué constituye una salida correcta en el contexto de negocio estudiado. [OPENAI-EVALS]

© 2026 ASCALEA. Tous droits réservés.

El sobreajuste — overfitting — ilustra la necesidad de no concluir a partir de un conjunto de ejemplos demasiado estrecho: un rendimiento observado en los casos de prueba disponibles no garantiza un rendimiento equivalente en casos nuevos. [GOOGLE-OVERFITTING]

8. Una matriz de decisión para la transición

Antes de ampliar el uso, un equipo puede examinar la siguiente matriz. No concede automáticamente una autorización; hace visibles las cuestiones no resueltas.

Ámbito	Pregunta de transición	Estado que documentar
Valor	¿El caso de uso aporta un beneficio observable respecto al proceso actual?	Beneficio esperado y modo de constatación.
Calidad	¿Los resultados alcanzan los criterios definidos en casos representativos?	Resultados de las pruebas, defectos conocidos y umbrales adoptados.
Seguridad y datos	¿Los datos, accesos y conexiones se ajustan al perímetro autorizado?	Reglas de acceso, restricciones y responsables.
Uso	¿Los usuarios saben interpretar los límites y aplicar los controles?	Instrucciones de uso y modalidades de acompañamiento.
Responsabilidad	¿Se identifica a una persona o función para cada decisión importante?	Distribución de roles y procedimiento de escalamiento.
Operabilidad	¿Pueden detectarse y tratarse los cambios, incidentes y anomalías?	Proceso de seguimiento y revisión.
Contexto aplicable	¿Las funciones competentes han examinado los requisitos internos, contractuales, sectoriales o jurídicos?	Conclusión del examen o límites explícitos del perímetro.

Un resultado «no establecido» en uno de estos ámbitos no implica siempre abandonar el caso de uso. Puede llevar a reducir el alcance, reforzar una validación humana, excluir determinados datos, aplazar una automatización o mantener el dispositivo en la fase experimental.

9. Los límites que deben mantenerse visibles

Incluso dentro de un perímetro bien descrito, persisten varios límites:

- un resultado convincente no es necesariamente exacto;
- una respuesta adecuada a un ejemplo no demuestra su solidez frente a un caso nuevo;
- una salida estructurada no elimina la ambigüedad de los documentos de entrada;

© 2026 ASCALEA. Tous droits réservés.

- un control humano meramente nominal no es suficiente si no cuenta con el tiempo, las fuentes o la autoridad necesarios para cuestionar el resultado;
- una configuración aceptable en un momento dado puede requerir una nueva revisión después de un cambio de modelo, datos, conexión o proceso.

Por lo tanto, el enfoque más sólido consiste en limitar el uso a aquello que está efectivamente demostrado, controlado y comprendido. Prefiere una promesa operativa acotada pero verificable a una promesa general que no podría cumplirse.

Conclusión

Poner la IA y los LLM en perspectiva consiste menos en buscar una herramienta universal que en vincular una capacidad técnica con un trabajo preciso, datos identificados, usuarios formados, reglas de control y una responsabilidad clara. El caso de uso se convierte entonces en una unidad de decisión: hace posible evaluar el valor, identificar los límites y definir una transición progresiva hacia un uso encuadrado.

© 2026 ASCALEA. Tous droits réservés.

Síntesis: disponer de un mapa en lugar de una definición única

Al término de este recorrido, el reto no es retener una definición aislada de la inteligencia artificial (IA), sino saber situar una herramienta, un modelo y un resultado en una misma cadena de decisión. Esta capacidad de orientación evita dos errores simétricos: atribuir al modelo capacidades que no ha demostrado, o subestimar las condiciones necesarias para un uso útil.

Las siguientes distinciones forman un mapa de lectura operativo.

© 2026 ASCALEA. Tous droits réservés.

Elemento	Pregunta orientadora	Referencia clave
Inteligencia artificial (IA)	¿De qué ámbito general se habla?	La IA designa un conjunto de métodos y sistemas; no designa ni un producto único ni una garantía de calidad. El marco de gestión de riesgos del National Institute of Standards and Technology (NIST) aborda, entre otros aspectos, los riesgos asociados a los sistemas de IA generativa. [NIST-AI-RMF] [NIST-AI-600-1]
Aprendizaje automático	¿Cómo aprovecha un sistema los datos?	Se trata de una familia de métodos dentro de la IA, con su vocabulario propio: entrenamiento, generalización, sobreajuste y evaluación. [GOOGLE-ML-GLOSSARY] [GOOGLE-OVERFITTING]
Modelo	¿Qué mecanismo produce la salida?	Un modelo es el artefacto entrenado que transforma una entrada en una salida según su arquitectura, sus parámetros y su modo de uso.
Modelo de lenguaje de gran tamaño (Large Language Model, LLM)	¿Qué familia de modelos manipula principalmente el lenguaje?	Un LLM es un modelo de lenguaje a gran escala; las arquitecturas de tipo Transformer constituyen una referencia técnica importante para comprender esta familia. [GOOGLE-LLM-INTRO] [GOOGLE-TRANSFORMER-PAPER]
Herramienta o aplicación	¿Cómo accede el usuario al modelo?	La herramienta reúne una interfaz, instrucciones, ajustes y, en ocasiones, archivos, búsquedas o aplicaciones conectadas. Estas capas influyen en el resultado tanto como el modelo por sí solo. [OPENAI-CONNECTED-APPS-CONTROLS]
Resultado generado		Una salida es una propuesta que debe controlarse con respecto a un objetivo, unas fuentes y un nivel de riesgo definidos. Una evaluación explíci-

© 2026 ASCALEA. Tous droits réservés.

	¿Qué debe decidirse a partir de la respuesta?	ta permite hacer repetible este control. [OPENAI-EVALS]

> Referencia central — Un LLM no es sinónimo ni de IA ni de herramienta conversacional. La IA es el ámbito general; el modelo es un componente entrenado; el LLM es una familia particular de modelos; la herramienta es el entorno concreto en el que una persona formula una solicitud y aprovecha una respuesta.

Este mapa conduce a una regla sencilla: cuando un resultado parece sorprendente, no preguntar solamente «¿es bueno el modelo?». Examinar también la tarea solicitada, los datos o documentos proporcionados, la instrucción, los parámetros de generación, las herramientas conectadas y el método de verificación.

Referencias de vocabulario consolidadas

El vocabulario es útil cuando sirve para formular preguntas precisas. Por tanto, los términos siguientes no son un diccionario que memorizar, sino un conjunto de referencias para dialogar con un equipo técnico, comparar ofertas o delimitar un uso.

Referencias relacionadas con los datos y el aprendizaje

• Conjunto de datos: conjunto de datos utilizado en una etapa de entrenamiento, validación o evaluación. Las características de los datos, su representatividad y sus posibles desequilibrios influyen en lo que un modelo puede aprender y en su comportamiento. [GOOGLE-DATA-CHARACTERISTICS] [GOOGLE-FAIRNESS-BIAS]

• Entrenamiento: fase durante la cual los parámetros de un modelo se ajustan a partir de datos.

• Inferencia: fase de uso durante la cual el modelo produce una salida a partir de una nueva entrada.

• Generalización: capacidad buscada de un modelo para producir resultados útiles sobre datos que no son simplemente los ejemplos encontrados durante el entrenamiento. [GOOGLE-OVERFITTING]

• Sobreajuste: situación en la que un modelo se ajusta demasiado estrechamente a los datos de entrenamiento y puede comportarse peor ante casos nuevos. [GOOGLE-OVERFITTING]

• Sesgo: efecto sistemático susceptible de producir resultados injustos, desequilibrados o inadecuados según los grupos, los datos o el contexto considerados. La identificación del sesgo requiere examinar la tarea, los datos y los resultados. [GOOGLE-FAIRNESS-BIAS]

Referencias relacionadas con los LLM

• Token: unidad procesada por el modelo durante la preparación y la generación del texto; la tokenización determina cómo se segmenta y representa el texto. [HF-TOKENIZER]

• Contexto: información disponible para orientar la respuesta en un momento dado: instrucción, texto proporcionado, ejemplos, historial o documentos accesibles según la herramienta.

• Instrucción: solicitud que define el papel esperado, el objetivo, las restricciones, el formato y los criterios de salida. La estructuración de las instrucciones y el uso de variables pueden mejorar su reutilización. [ANTHROPIC-PROMPT-BEST-PRACTICES]

© 2026 ASCALEA. Tous droits réservés.

• Generación: producción progresiva de una salida textual. Los parámetros de generación contribuyen al equilibrio entre regularidad, diversidad y longitud de la respuesta. [HF-GENERATION-CONFIG] [HF-TEXT-GENERATION]

• Transformer: arquitectura introducida en el trabajo de investigación Attention Is All You Need, basada, entre otros elementos, en mecanismos de atención. [GOOGLE-TRANSFORMER-PAPER]

Referencias relacionadas con la implementación

• Tarea: resultado operativo buscado, por ejemplo resumir, clasificar, extraer, redactar, traducir o responder a una pregunta. Las bibliotecas de software pueden ofrecer mecanismos que asocian tareas con componentes de procesamiento. [HF-TASKS-EXPLAINED] [HF-PIPELINES]

• Evaluación: proceso consistente en comparar las salidas con criterios, casos de prueba o referencias definidos de antemano. [OPENAI-EVALS]

• Trazabilidad: capacidad de conservar los elementos necesarios para comprender cómo se ha producido y controlado una salida: versión de la herramienta, instrucción, entradas, fuentes, fecha, evaluador y decisión tomada.

• Conexión o aplicación conectada: mecanismo mediante el cual una herramienta puede acceder a servicios o datos externos. Esta posibilidad debe examinarse como una decisión de alcance, autorización y seguridad, y no como una simple comodidad funcional. [OPENAI-CONNECTED-APPS-CONTROLS] [OPENAI-GOOGLE-APP-DATA]

Un mapa de profundizaciones para elegir

La continuación adecuada del recorrido depende menos del nivel de curiosidad técnica que del uso previsto. Es más útil profundizar en el tema que reduce una incertidumbre real que acumular nociones aisladas.

© 2026 ASCALEA. Tous droits réservés.

Perfil o intención	Prioridades de profundización	Preguntas que abordar primero
Usuario individual	Formulación de instrucciones, verificación de respuestas, gestión de información sensible	¿Qué beneficio concreto se espera? ¿Qué salidas deben verificarse antes de reutilizarlas?
Responsable de negocio	Definición de la necesidad, criterios de calidad, conjunto de casos representativos, supervisión humana	¿Qué decisión o actividad debe apoyar la herramienta? ¿Qué criterios permitirán concluir que el uso es útil?
Equipo de producto o técnico	Arquitectura, integración, generación, evaluación, observabilidad	¿Dónde se sitúan el modelo, las herramientas y los datos en el flujo de trabajo? ¿Cómo probar los casos esperados y los fallos importantes?
Responsable de datos, seguridad o cumplimiento	Circulación de datos, derechos de acceso, conexiones, conservación, riesgos	¿Qué datos entran, salen o se vuelven accesibles? ¿Quién puede activar, consultar o modificar las conexiones?
Responsable de decisiones o líder de transformación	Gobernanza, priorización, medición de valor, gestión de riesgos	¿Qué problema justifica la adopción? ¿Quién asume la responsabilidad del resultado y de su control?
Lector que desea comprender los fundamentos	Datos, entrenamiento, inferencia, arquitectura Transformer, tokenización	¿Cómo relacionar una salida visible con los mecanismos que la hacen posible y con sus límites?

Itinerario de profundización recomendado

1. Estabilizar el caso de uso. Describir una tarea, sus usuarios, su resultado esperado y sus criterios de aceptación antes de elegir un modelo.
2. Construir algunos casos representativos. Prever ejemplos fáciles, ambiguos, incompletos y sensibles para no evaluar la herramienta en un único caso favorable.
3. Distinguir las fuentes de variación. Modificar por separado la instrucción, el contexto, los parámetros de generación y el modelo para identificar qué explica una diferencia de resultado.
4. Organizar la verificación. Definir qué debe verificarse, por quién, con qué fuentes y antes de qué decisión.
5. Documentar el alcance. Conservar las reglas de uso, los límites conocidos, los datos autorizados y las condiciones de revisión.

Esta progresión vincula comprensión y práctica: evita confundir una prueba convincente con una capacidad demostrada en condiciones reales.

© 2026 ASCALEA. Tous droits réservés.

Verificaciones que deben realizarse antes de utilizar una herramienta o un modelo

La adopción no debería comenzar con la pregunta «¿qué herramienta es la más impresionante?», sino con un control de compatibilidad entre una necesidad, un contexto de datos y un nivel de riesgo aceptable.

1. Aclarar la finalidad

- [ ] La tarea está formulada en términos observables: producir, comparar, extraer, clasificar, resumir o asistir una decisión identificada.
- [ ] Se conocen el destinatario del resultado y el momento en que lo utilizará.
- [ ] El resultado esperado tiene un formato y criterios de aceptación explícitos.
- [ ] Está claro que la herramienta asiste una actividad definida, en lugar de sustituir indistintamente el juicio necesario.

2. Examinar las entradas y los datos

- [ ] Se han identificado las informaciones que se introducirán, importarán o harán accesibles.
- [ ] Los datos sensibles, confidenciales o sujetos a restricciones internas están sujetos a reglas específicas antes de cualquier prueba.
- [ ] Las fuentes proporcionadas al modelo se separan de las hipótesis, los borradores y los contenidos no verificados.
- [ ] Se han examinado las características de los datos útiles para la tarea, en particular las carencias, los desequilibrios y la representatividad. [GOOGLE-DATA-CHARACTERISTICS]
- [ ] Los riesgos de sesgo pertinentes para el caso de uso forman parte de los criterios de control. [GOOGLE-FAIRNESS-BIAS]

3. Verificar el funcionamiento de la herramienta

- [ ] Se identifican el modelo específico, la interfaz utilizada y las funciones activadas.
- [ ] Se documentan los ajustes que influyen en la generación cuando la herramienta los expone. [HF-GENERATION-CONFIG]
- [ ] Las instrucciones reutilizables precisan el contexto, las restricciones y el formato deseado. [ANTHROPIC-PROMPT-BEST-PRACTICES]
- [ ] Las conexiones con aplicaciones, archivos o servicios externos son conocidas, justificadas y limitadas a lo necesario. [OPENAI-CONNECTED-APPS-CONTROLS]
- [ ] Los controles de datos propuestos por la plataforma se examinan en el marco del uso elegido. [OPENAI-DATA-CONTROLS]

4. Probar la calidad y los límites

- [ ] Un pequeño conjunto de casos de prueba cubre los casos habituales, los casos difíciles y los casos que deben excluirse.

© 2026 ASCALEA. Tous droits réservés.

- [ ] Cada caso incluye un resultado esperado, una fuente de referencia o una regla de juicio.
- [ ] Los errores importantes se clasifican: error factual, omisión, formato no respetado, contenido inapropiado, sesgo o respuesta inutilizable.
- [ ] Los resultados se evalúan en varias ocasiones cuando la tarea o la configuración pueden producir variaciones.
- [ ] Se define una regla de detención: situaciones en las que el resultado no debe utilizarse sin un control adicional.

5. Asignar las responsabilidades

- [ ] Una persona o un equipo es responsable del alcance de uso.
- [ ] El nivel de revisión humana es proporcional a las consecuencias de un error.
- [ ] Los usuarios saben a quién informar de un resultado inesperado o un comportamiento preocupante.
- [ ] Las decisiones, los incidentes y las mejoras se registran para revisar el dispositivo.

> Decisión de puesta en uso Puede seleccionarse una herramienta cuando su uso está definido, los datos y las conexiones están controlados, se han probado casos representativos, se conocen los límites importantes y se ha asignado una responsabilidad de control. Si falta alguno de estos elementos, el enfoque más prudente consiste en volver a una fase de exploración limitada en lugar de generalizar el uso.

Recursos y referencias que deben constituirse

Una biblioteca útil no es una lista indiscriminada de enlaces. Asocia cada recurso a una pregunta de trabajo. Las siguientes referencias pertenecen al corpus documental de esta obra; pueden servir como punto de partida para una vigilancia estructurada.

Comprender los conceptos y el vocabulario

- Glosario del aprendizaje automático: para estabilizar las definiciones y distinguir nociones técnicas próximas. [GOOGLE-ML-GLOSSARY]
- Introducción a los LLM: para revisar el lugar de los grandes modelos de lenguaje en el aprendizaje automático. [GOOGLE-LLM-INTRO]
- Publicación sobre el Transformer: para remontarse a una fuente primaria relativa a la arquitectura Transformer. [GOOGLE-TRANSFORMER-PAPER]
- Documentación sobre la tokenización: para comprender el papel de los tokens en el procesamiento del texto. [HF-TOKENIZER]

Profundizar en el aprendizaje, los datos y los sesgos

- Características de los conjuntos de datos: para examinar las propiedades de los datos que influyen en el aprendizaje. [GOOGLE-DATA-CHARACTERISTICS]
- Sobreajuste: para profundizar en la diferencia entre el rendimiento observado durante el entrenamiento y el comportamiento ante casos nuevos. [GOOGLE-OVERFITTING]

© 2026 ASCALEA. Tous droits réservés.

• Identificación de sesgos: para estructurar el análisis de los riesgos de sesgo en una tarea determinada. [GOOGLE-FAIRNESS-BIAS]

Diseñar interacciones y tratamientos

• Buenas prácticas para formular instrucciones: para construir modelos de instrucciones reutilizables y parametrizables. [ANTHROPIC-PROMPT-BEST-PRACTICES]

• Generación de texto y parámetros de generación: para relacionar los ajustes de generación con el comportamiento observado. [HF-TEXT-GENERATION] [HF-GENERATION-CONFIG]

• Tareas y pipelines: para comprender cómo pueden organizarse los procesos de lenguaje en una cadena técnica. [HF-TASKS-EXPLAINED] [HF-PIPELINES]

Evaluar, gobernar y proteger los usos

• Marco de gestión de riesgos relacionados con la IA: para disponer de un lenguaje de gobernanza de riesgos. [NIST-AI-RMF]

• Perfil dedicado a la IA generativa: para profundizar en los riesgos específicos de los sistemas generativos. [NIST-AI-600-1]

• Documentación sobre las evaluaciones: para estructurar un enfoque de prueba y medición. [OPENAI-EVALS]

• Controles de datos y de aplicaciones conectadas: para estudiar los parámetros disponibles en la plataforma correspondiente antes de activar datos o servicios externos. [OPENAI-DATA-CONTROLS] [OPENAI-CONNECTED-APPS-CONTROLS] [OPENAI-GOOGLE-APP-DATA]

Mantener un expediente de referencia vivo

Para cada recurso seleccionado, el expediente de referencia puede contener:

• su título y su editor;

• la pregunta a la que responde;

• la fecha de consulta;

• el alcance afectado;

• el nivel de fiabilidad o la naturaleza del documento;

• las decisiones internas a las que contribuye;

• la fecha prevista de reexamen.

Esta última etapa es esencial: los documentos técnicos, las capacidades de las herramientas y los parámetros de control evolucionan. Antes de una decisión de despliegue o una publicación, deben verificarse de nuevo la disponibilidad, la versión y la pertinencia de cada recurso.

© 2026 ASCALEA. Tous droits réservés.

Conclusión: aprender a hacer las preguntas adecuadas

Comprender la IA y los LLM no consiste en predecir el futuro de cada herramienta. Se trata de desarrollar un razonamiento disciplinado: distinguir el ámbito, el modelo, la interfaz y el resultado; relacionar cada uso con datos y una tarea; controlar lo que se genera; y organizar las responsabilidades.

Por tanto, la pregunta más productiva para continuar el aprendizaje es la siguiente: ¿en qué contexto preciso una salida producida por este sistema sería lo suficientemente fiable, útil y controlada como para utilizarse? La respuesta no se encuentra solo en el nombre de un modelo. Se construye mediante la delimitación de la necesidad, la evaluación de los resultados, el control de los datos y la revisión regular de las condiciones de uso.

© 2026 ASCALEA. Tous droits réservés.

• Anonimización — Medida mencionada para reducir los riesgos relacionados con los datos y la confidencialidad, cuando sea posible. Se integra, junto con la minimización de datos, las reglas de acceso y una configuración validada.
• Aprendizaje automático (machine learning, ML) — Familia de métodos de IA en los que un sistema se construye a partir de datos. Su vocabulario comprende, en particular, los datos, las características, los modelos, el entrenamiento y la inferencia.
• Clasificación — Tarea que consiste en asignar una categoría, una etiqueta o una prioridad a una entrada. Puede aplicarse, en particular, a un texto, una imagen o un dato, según el sistema.
• Configuración de generación — Conjunto de ajustes que influyen en la forma en que se produce la salida de un modelo. Debe distinguirse del propio modelo y de la interfaz utilizada.
• Contexto — Contenido transmitido al modelo en el momento de la solicitud. Contribuye a orientar la respuesta producida.
• Entrenamiento — Etapa en la que los parámetros internos de un modelo se ajustan a partir de datos y de un objetivo de optimización. Es distinta del uso del modelo en una conversación o una solicitud.
• Extracción de información — Función que consiste en identificar y producir elementos estructurados a partir de un contenido proporcionado. Forma parte de los posibles usos de un LLM en una aplicación.
• Inferencia — Uso de un modelo cuyos parámetros ya están fijados para calcular una salida a partir de una nueva entrada. Tiene lugar cuando el modelo se despliega en una situación real.
• Instrucción — Indicación dada al modelo sobre lo que debe hacer. Forma parte de los elementos que estructuran una solicitud dirigida a un LLM.
• Inteligencia artificial (IA) — Conjunto de sistemas y métodos capaces de ejecutar tareas asociadas, en particular, a la percepción, la predicción, el razonamiento práctico, la clasificación o la generación de contenido. La IA designa un ámbito general, y no un producto único.
• Modalidad — Forma en la que la información llega al modelo o es producida por él, por ejemplo, texto, imagen, audio, datos estructurados o combinación de modalidades. Una modalidad no define por sí sola la finalidad del modelo.
• Modelo — Componente aprendido o parametrizado que produce una salida a partir de una entrada. Su comportamiento depende, en particular, de su arquitectura, sus parámetros y su modo de uso.
• Modelo de lenguaje de gran tamaño (Large Language Model, LLM) — Modelo diseñado para procesar y producir lenguaje. Genera texto a partir de regularidades aprendidas y del contexto que se le proporciona.
• Pipeline — Cadena de procesamiento asociada a una tarea determinada en bibliotecas técnicas de modelos. Los pipelines pueden organizarse para tareas de texto, imagen, audio o que combinen texto e imagen.
• Principio de mínimo privilegio — Medida de seguridad que consiste en limitar los accesos para evitar permisos excesivos. Puede asociarse con la validación de acciones, el registro y la supervisión.

© 2026 ASCALEA. Tous droits réservés.

• Registro — Registro de las acciones o los eventos de un sistema con fines de control y supervisión. Se menciona entre las medidas útiles frente a una seguridad insuficiente o al uso indebido.

• Token (jetón) — Unidad de texto que puede ser manipulada por un LLM. Los LLM generalmente generan una secuencia de texto prediciendo progresivamente los siguientes tokens en un contexto dado.

• Transformer — Arquitectura utilizada con frecuencia por los LLM modernos. Se basa en mecanismos de atención.

© 2026 ASCALEA. Tous droits réservés.

[ANTHROPIC-PROMPT-BEST-PRACTICES] Prompting best practices — Documentation technique officielle d'Anthropic. — https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/prompt-templates-and-variables

[GOOGLE-DATA-CHARACTERISTICS] Datasets: Data characteristics — Cours technique officiel Google Machine Learning Crash Course. — https://developers.google.com/machine-learning/crash-course/overfitting/data-characteristics

[GOOGLE-FAIRNESS-BIAS] Fairness: Identifying bias — Cours technique officiel Google Machine Learning Crash Course. — https://developers.google.com/machine-learning/crash-course/fairness/identifying-bias

[GOOGLE-LLM-INTRO] LLMs: What's a large language model? — Cours technique officiel Google Machine Learning Crash Course. — https://developers.google.com/machine-learning/crash-course/llm/transformers?authuser=0

[GOOGLE-ML-GLOSSARY] Machine Learning Glossary — Documentation pédagogique technique officielle de Google for Developers. — https://developers.google.com/machine-learning/glossary

[GOOGLE-OVERFITTING] Overfitting — Cours technique officiel Google Machine Learning Crash Course. — https://developers.google.com/machine-learning/crash-course/overfitting/overfitting

[GOOGLE-TRANSFORMER-PAPER] Attention is All You Need — Publication de recherche primaire des auteurs du Transformer, publiée par Google Research. — https://research.google/pubs/attention-is-all-you-need/

[HF-GENERATION-CONFIG] Generation — Documentation technique de Hugging Face Transformers. — https://huggingface.co/docs/transformers/main_classes/text_generation

[HF-PIPELINES] Pipelines — Documentation technique de Hugging Face Transformers. — https://huggingface.co/docs/transformers/main_classes/pipelines

[HF-TASKS-EXPLAINED] How ➢Transformers solve tasks — Documentation technique de Hugging Face Transformers. — https://huggingface.co/docs/transformers/main/tasks_explained

[HF-TEXT-GENERATION] Text generation — Documentation technique de Hugging Face Transformers. — https://huggingface.co/docs/transformers/main/llm_tutorial

[HF-TOKENIZER] Tokenizer — Documentation technique de Hugging Face Transformers. — https://huggingface.co/docs/transformers/main_classes/tokenizer

[NIST-AI-600-1] Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile — Publication officielle du National Institute of Standards and Technology (NIST), organisme public américain. — https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

[NIST-AI-RMF] AI Risk Management Framework — Cadre officiel du NIST. — https://www.nist.gov/itl/ai-risk-management-framework

[OPENAI-CONNECTED-APPS-CONTROLS] Admin controls, security, and compliance for plugins and apps — Centre d'aide officiel OpenAI. — https://help.openai.com/en/articles/11509118

© 2026 ASCALEA. Tous droits réservés.

[OPENAI-DATA-CONTROLS] Data controls in the OpenAI platform — Documentation officielle de l'API OpenAI. — https://platform.openai.com/docs/models/default-usage-policies-by-endpoint

[OPENAI-EVALS] Evals — Documentation officielle de l'API OpenAI. — https://platform.openai.com/docs/api-reference/evals/deleteRun?lang=python

[OPENAI-GOOGLE-APP-DATA] Google app data controls FAQ — Centre d'aide officiel OpenAI. — https://help.openai.com/en/articles/10408842-google-app-data-controls-faq

[OPENAI-TOOLS-REFERENCE] Streaming events — Référence officielle de l'API OpenAI. — https://platform.openai.com/docs/api-reference/responses-streaming/response/file_search_call/completed?lang=javascript

© 2026 ASCALEA. Tous droits réservés.

© 2026 ASCALEA. Tous droits réservés.