

12 preguntas antes de dar acceso a la IA.

¿Qué puede consultar? ¿Qué acciones puede ejecutar? ¿Quién interviene cuando algo falla? Lleve estas preguntas a la conversación con su equipo y sus proveedores.

Elija una tarea de la empresa. Los ejemplos se han preparado para debatir y no representan pruebas realizadas. Cada control debe definirse y verificarse en la aplicación.

01-04

Defina el uso.

Tarea, datos y acciones.

05-08

Compruebe los límites.

Accesos, revisión y pruebas.

09-12

Prepare la operación.

Registros, fallos y responsables.

¿Qué trabajo entra en el alcance?

Empiece por la tarea y por lo que la IA necesita para realizarla.

01 ¿Qué tarea realiza la IA y quién decide después?

En la práctica

La IA reúne los artículos de un pedido. El equipo comercial aún debe comprobar el precio, la cantidad y el plazo antes de confirmar la propuesta.

Qué solicitar

Una descripción de la entrada, la salida, los usos prohibidos y el responsable de la siguiente decisión. En la prueba, compruebe que la sugerencia sigue pendiente hasta la revisión. Describir la regla no demuestra que el sistema la aplique.

02 ¿Qué datos entran? ¿Cuáles quedan fuera?

En la práctica

Comprobar artículos y cantidades no exige enviar datos bancarios ni observaciones personales de la ficha del cliente.

Qué solicitar

La finalidad de cada campo, también en los archivos adjuntos. Con datos preparados para la prueba, compruebe qué llega a la herramienta y qué aparece en la respuesta. Ocultar los nombres no garantiza el anonimato; evalúe los datos reales antes de usarlos.

¿Qué trabajo entra en el alcance?

Empiece por la tarea y por lo que la IA necesita para realizarla.

03 ¿Qué fuente es válida para esta tarea?

En la práctica

La carpeta contiene dos tarifas comerciales. La IA encuentra ambas, pero solo una es válida para el pedido recibido.

Qué solicitar

El origen, la versión, el responsable y el uso permitido de cada fuente. Pruebe la sustitución de un documento y compruebe la referencia utilizada en la respuesta. Si las fuentes autorizadas discrepan, defina quién resuelve el conflicto. Citar una fuente no garantiza el acierto.

04 ¿El permiso para consultar también permite modificar o enviar?

En la práctica

Consultar las existencias, reservar mercancía y enviar una propuesta son acciones distintas, con consecuencias distintas.

Qué solicitar

La lista de operaciones y registros permitidos. Pruebe una modificación fuera del alcance y compruebe el bloqueo en el sistema que ejecuta la acción. Una instrucción escrita para que la IA actúe con cuidado no impone, por sí sola, una restricción técnica.

¿El límite resiste ante una excepción?

Solicite evidencias también para situaciones en las que la IA deba detenerse o derivar la tarea.

05 ¿En nombre de quién accede la IA a los sistemas?

En la práctica

Dos agentes de atención gestionan carteras de clientes distintas. Una consulta no debe revelar pedidos de la otra cartera.

Qué solicitar

Una matriz de identidades, funciones, datos y acciones. Pruebe los accesos entre carteras, la retirada de un permiso y el contenido de las respuestas generadas. Que la IA diga «no tengo acceso» no demuestra que la consulta se haya bloqueado.

06 ¿Y si un archivo adjunto intenta cambiar la tarea?

En la práctica

El pedido incluye un archivo adjunto con una orden de enviar la lista de clientes a otra dirección. Eso no forma parte de la comprobación solicitada.

Qué solicitar

Una prueba controlada de una instrucción indebida en un documento, una página o una respuesta de herramienta. Compruebe el bloqueo de la acción y la derivación para investigar. Los filtros no son infalibles; los permisos y los límites de acción siguen siendo necesarios.

¿El límite resiste ante una excepción?

Solicite evidencias también para situaciones en las que la IA deba detenerse o derivar la tarea.

07 ¿Qué aprueba la persona antes de la acción?

En la práctica

El equipo revisa una propuesta. Si el destinatario o el contenido cambian después, esa aprobación ya no cubre el envío.

Qué solicitar

Una vista previa con el contenido, el destino y los efectos de la acción, además del registro de la aprobación específica. Pruebe si un cambio exige una nueva revisión. La persona necesita información, tiempo y autoridad para decidir; un botón de confirmación no basta.

08 ¿Qué errores impiden avanzar?

En la práctica

Un archivo adjunto ilegible, un artículo sin código o una fuente contradictoria pueden requerir información adicional, no una respuesta inventada.

Qué solicitar

Casos habituales y excepciones, con el resultado esperado definido antes de la prueba y el resultado observado por versión. Separe los errores graves de la media. Una muestra pequeña no lo cubre todo; los cambios en fuentes, modelos o herramientas exigen reevaluar las pruebas pertinentes.

¿Quién actúa cuando algo sale de lo previsto?

Acuerde cómo investigar, interrumpir, corregir y finalizar el uso.

09 ¿Qué queda registrado y quién puede verlo?

En la práctica

Se ha denegado una consulta. El equipo necesita entender el motivo sin copiar el documento entero en el registro.

Qué solicitar

Una muestra protegida con la hora, la identidad, la acción, el resultado y el motivo. Compruebe el acceso, la retención y el tratamiento de los campos sensibles. Los registros también necesitan protección; guardar contraseñas, credenciales o datos innecesarios aumenta la exposición.

10 ¿Cuándo debe detenerse el intento?

En la práctica

El servicio de existencias no está disponible. Repetir la consulta sin límite aumenta el consumo y deja pedidos en espera.

Qué solicitar

Límites de tiempo, intentos y consumo por tarea, con un procedimiento alternativo. Pruebe la interrupción y la comunicación del estado. Antes de repetir una acción que modifica datos, compruebe el resultado anterior para evitar duplicaciones. Los límites dependen del proceso.

¿Quién actúa cuando algo sale de lo previsto?

Acuerde cómo investigar, interrumpir, corregir y finalizar el uso.

11 ¿Cómo interrumpir y corregir lo que ya ha ocurrido?

En la práctica

Una dirección de entrega se ha modificado incorrectamente. Detener nuevas acciones no corrige el pedido que ya ha pasado a transporte.

Qué solicitar

Un procedimiento ensayado con un responsable, pasos para interrumpir la operación, comprobación de los efectos y criterios para reanudarla. Verifique los sistemas implicados. No toda acción puede deshacerse; un mensaje ya leído, por ejemplo, puede exigir una corrección y una comunicación.

12 ¿Quién responde del uso y de su finalización?

En la práctica

El piloto termina. Los accesos, documentos, registros y copias pueden permanecer en sistemas de la empresa y de sus proveedores.

Qué solicitar

Responsables de la revisión y de los incidentes, plazos de retención justificados y un plan de salida. Incluya la revocación de accesos, las copias y las evidencias de la finalización. Los plazos y las obligaciones dependen del contexto y de la evaluación técnica y jurídica pertinente.

Una respuesta necesita algo que la respalde.

Para cada pregunta, registre la decisión, la evidencia disponible, el responsable y el asunto pendiente. Si falta una comprobación, acuerde cómo realizarla antes de ampliar el uso.

Decisión

¿Qué queda permitido?

Evidencia

¿Qué se ha comprobado?

Responsable

¿Quién decide y hace el seguimiento?

Asunto pendiente

¿Qué falta por verificar?

Converse con Gaulettgaulett.com/es/contacto/[Prepare su proyecto](#)[Conozca la validación](#)**Referencias para profundizar**

Las preguntas aplican estas referencias al trabajo diario de una empresa. No representan pruebas de las instituciones citadas.

NIST · AI RMF 1.0<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>**NIST · Generative AI Profile**<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>**OWASP · AI Agent Security**https://cheatsheetseries.owasp.org/cheatsheets/AI_Agent_Security_Cheat_Sheet.html**OWASP · Prompt Injection**<https://genai.owasp.org/llmrisk/llm01-prompt-injection/>**OWASP · Logging**https://cheatsheetseries.owasp.org/cheatsheets/Logging_Cheat_Sheet.html