

12 domande prima di dare accesso all'IA.

Che cosa può consultare? Quali azioni può eseguire? Chi interviene quando qualcosa non funziona? Porta queste domande al confronto con il tuo team e i tuoi fornitori.

Scegli un'attività dell'azienda. Gli esempi sono stati preparati per la discussione e non rappresentano test eseguiti. Ogni controllo deve essere definito e verificato nell'applicazione.

01-04 **Definisci l'uso.**
Compito, dati e azioni.

05-08 **Verifica i limiti.**
Accessi, revisione e test.

09-12 **Prepara l'operatività.**
Registri, guasti e responsabili.

Quale lavoro rientra nell'ambito?

Parti dal compito e da ciò che serve all'IA per svolgerlo.

01 Quale compito svolge l'IA e chi decide dopo?

In pratica

L'IA raccoglie gli articoli di un ordine. Il team commerciale deve ancora verificare prezzo, quantità e tempi di consegna prima di confermare la proposta.

Che cosa chiedere

Una descrizione dell'input, dell'output, degli usi vietati e del responsabile della decisione successiva. Nel test, verifica che il suggerimento resti in sospeso fino alla revisione. Descrivere la regola non dimostra che il sistema la applichi.

02 Quali dati entrano? Quali restano fuori?

In pratica

Verificare articoli e quantità non richiede l'invio di dati bancari o di annotazioni personali presenti nell'anagrafica del cliente.

Che cosa chiedere

La finalità di ogni campo, anche negli allegati. Con dati preparati per il test, verifica che cosa arriva allo strumento e compare nella risposta. Mascherare i nomi non garantisce l'anonimato; valuta i dati reali prima di usarli.

Quale lavoro rientra nell'ambito?

Parti dal compito e da ciò che serve all'IA per svolgerlo.

03 Quale fonte è valida per questo compito?

In pratica

La cartella contiene due listini commerciali. L'IA li trova entrambi, ma solo uno è valido per l'ordine ricevuto.

Che cosa chiedere

Origine, versione, responsabile e uso consentito di ogni fonte. Prova a sostituire un documento e verifica il riferimento usato nella risposta. Se le fonti autorizzate sono in disaccordo, definisci chi risolve il conflitto. Citare una fonte non garantisce la correttezza.

04 Il permesso di consultare consente anche di modificare o inviare?

In pratica

Consultare le scorte, riservare la merce e inviare una proposta sono azioni diverse, con conseguenze diverse.

Che cosa chiedere

L'elenco delle operazioni e dei record consentiti. Prova una modifica fuori dall'ambito previsto e verifica il blocco nel sistema che esegue l'azione. Un'istruzione scritta che chiede all'IA di agire con cautela non impone, da sola, una restrizione tecnica.

Il limite regge di fronte a un'eccezione?

Chiedi evidenze anche per le situazioni in cui l'IA deve fermarsi o affidare il compito a qualcuno.

05 A nome di chi l'IA accede ai sistemi?

In pratica

Due addetti all'assistenza gestiscono portafogli clienti diversi. Una consultazione non deve rivelare gli ordini dell'altro portafoglio.

Che cosa chiedere

Una matrice di identità, ruoli, dati e azioni. Verifica gli accessi tra portafogli, la revoca di un permesso e il contenuto delle risposte generate. Che l'IA dica «non ho accesso» non dimostra che la consultazione sia stata bloccata.

06 E se un allegato tentasse di cambiare il compito?

In pratica

L'ordine include un allegato con l'istruzione di inviare l'elenco dei clienti a un altro indirizzo. Questo non fa parte della verifica richiesta.

Che cosa chiedere

Un test controllato di un'istruzione non autorizzata in un documento, una pagina o una risposta di uno strumento. Verifica il blocco dell'azione e la segnalazione per l'indagine. I filtri non sono infallibili; permessi e limiti alle azioni restano necessari.

Il limite regge di fronte a un'eccezione?

Chiedi evidenze anche per le situazioni in cui l'IA deve fermarsi o affidare il compito a qualcuno.

07 Che cosa approva la persona prima dell'azione?

In pratica

Il team rivede una proposta. Se il destinatario o il contenuto cambiano in seguito, quell'approvazione non copre più l'invio.

Che cosa chiedere

Un'anteprima con contenuto, destinazione ed effetti dell'azione, oltre alla registrazione dell'approvazione specifica. Verifica se una modifica richiede una nuova revisione. La persona ha bisogno di informazioni, tempo e autorità per decidere; un pulsante di conferma non basta.

08 Quali errori impediscono di procedere?

In pratica

Un allegato illeggibile, un articolo senza codice o una fonte contraddittoria possono richiedere informazioni aggiuntive, non una risposta inventata.

Che cosa chiedere

Casi comuni ed eccezioni, con il risultato atteso definito prima del test e il risultato osservato per ogni versione. Considera gli errori gravi separatamente dalla media. Un campione piccolo non copre tutto; modifiche a fonti, modelli o strumenti richiedono di rivalutare i test pertinenti.

Chi interviene quando qualcosa va fuori programma?

Concorda come indagare, interrompere, correggere e terminare l'uso.

09 Che cosa viene registrato e chi può vederlo?

In pratica

Una consultazione è stata negata. Il team deve capirne il motivo senza copiare l'intero documento nel registro.

Che cosa chiedere

Un campione protetto con orario, identità, azione, risultato e motivo. Verifica accesso, conservazione e trattamento dei campi sensibili. Anche i registri richiedono protezione; conservare password, credenziali o dati non necessari aumenta l'esposizione.

10 Quando deve fermarsi il tentativo?

In pratica

Il servizio di consultazione delle scorte non è disponibile. Ripetere la richiesta senza limiti aumenta il consumo e lascia gli ordini in attesa.

Che cosa chiedere

Limiti di tempo, tentativi e consumo per compito, con una procedura alternativa. Verifica l'interruzione e la comunicazione dello stato. Prima di ripetere un'azione che modifica i dati, controlla il risultato precedente per evitare duplicazioni. I limiti dipendono dal processo.

Chi interviene quando qualcosa va fuori programma?

Concorda come indagare, interrompere, correggere e terminare l'uso.

11 Come interrompere e correggere ciò che è già successo?

In pratica

Un indirizzo di consegna è stato modificato in modo errato. Fermare nuove azioni non corregge l'ordine che è già stato affidato al trasporto.

Che cosa chiedere

Una procedura provata, con un responsabile, passaggi per interrompere l'operazione, verifica degli effetti e criteri per riprendere. Verifica i sistemi coinvolti. Non tutte le azioni si possono annullare; un messaggio già letto, per esempio, può richiedere una correzione e una comunicazione.

12 Chi risponde dell'uso e della sua conclusione?

In pratica

Il progetto pilota termina. Accessi, documenti, registri e copie possono rimanere nei sistemi dell'azienda e dei fornitori.

Che cosa chiedere

Responsabili della revisione e degli incidenti, periodi di conservazione giustificati e un piano di uscita. Includi la revoca degli accessi, le copie e le evidenze della conclusione. Tempi e obblighi dipendono dal contesto e dalla valutazione tecnica e giuridica pertinente.

Dalla domanda alla decisione

Una risposta ha bisogno di qualcosa che la sostenga.

Per ogni domanda, registra la decisione, l'evidenza disponibile, il responsabile e il punto in sospeso. Se manca una verifica, concorda come svolgerla prima di ampliare l'uso.

Decisione

Che cosa è consentito?

Evidenza

Che cosa è stato verificato?

Responsabile

Chi decide e segue l'attività?

Punto in sospeso

Che cosa resta da verificare?

Parla con Gaulett

gaulett.com/it/contatti/

[Prepara il tuo progetto](#)

[Scopri la validazione](#)

Riferimenti per approfondire

Le domande applicano questi riferimenti al lavoro quotidiano di un'azienda. Non rappresentano test delle istituzioni citate.

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