

AI 2 - Errori, fonti, bias ed etica

Verificare le risposte, riconoscere i limiti e usare l'AI in modo responsabile

Usare l'intelligenza artificiale in modo competente non significa fidarsi automaticamente delle risposte. Questa dispensa affronta errori, allucinazioni, verifica delle fonti, bias, privacy, guardrail e responsabilità dell'utente. L'obiettivo è imparare a usare l'AI come strumento di supporto, non come autorità infallibile.

IDEA FONDAMENTALE

Una risposta AI può essere utile, ben scritta e convincente, ma deve essere letta con spirito critico. Più l'informazione è importante, recente o delicata, più diventa necessario verificarla.

1. L'AI può sbagliare

Un sistema generativo può interpretare male una richiesta, omettere dettagli importanti, utilizzare informazioni superate oppure inventare dati plausibili. Il rischio aumenta quando la domanda riguarda argomenti molto specifici, informazioni recenti, dati numerici, citazioni oppure riferimenti bibliografici.

La forma corretta di una frase non garantisce la correttezza del contenuto. Un testo ordinato, sicuro nel tono e grammaticalmente accurato può comunque contenere errori. Per questo l'utente deve distinguere tra qualità espositiva e affidabilità informativa.

Risposta plausibile non significa risposta corretta

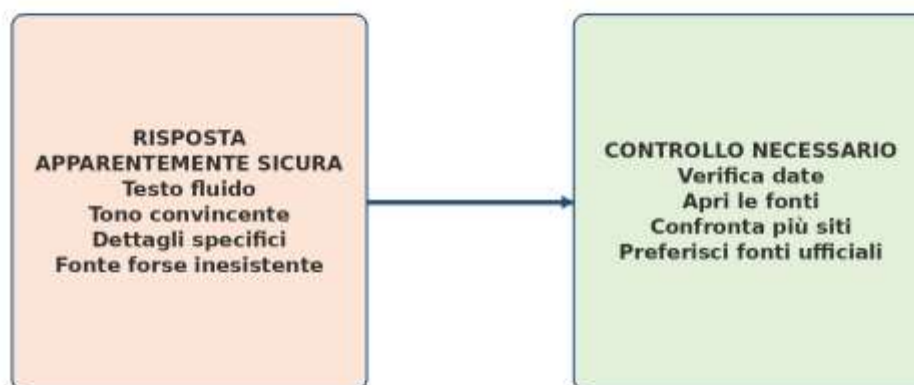


Figura 1 - Una risposta convincente deve comunque essere verificata.

2. Che cos'è un'allucinazione

Nel linguaggio dell'intelligenza artificiale, un'allucinazione è una risposta che presenta come reale un'informazione inesatta o inventata. Il sistema può indicare una data errata, attribuire a una persona una frase mai pronunciata, citare un libro inesistente oppure descrivere una funzione software che non esiste.

Possibile problema	Esempio	Che cosa fare
dato errato	data, numero o nome non corretto;	controllare una fonte attendibile;
fonte inventata	libro, articolo o sito inesistente;	verificare che la fonte esista davvero;
informazione superata	regola, prezzo o versione cambiata;	consultare una fonte aggiornata;
risposta incompleta	manca un'eccezione importante;	porre domande aggiuntive e confrontare più fonti;
eccessiva sicurezza	tono certo anche quando il dato è dubbio;	separare fatti verificati da ipotesi.

3. Verificare non significa diffidare di tutto

L'obiettivo non è rifiutare l'AI. Una risposta generata può essere un ottimo punto di partenza per capire un tema, preparare una scaletta, confrontare concetti o avviare una ricerca. Il controllo serve a trasformare una prima spiegazione in un'informazione affidabile, soprattutto quando la decisione ha conseguenze reali.

Metodo semplice per verificare una risposta AI



Figura 2 - Metodo in cinque passaggi per verificare una risposta AI.

Passaggio	Azione
1. Individua	separa le affermazioni importanti da quelle generali;
2. Chiedi fonti	richiedi riferimenti specifici quando servono;
3. Apri le fonti	non limitarti a leggere il titolo o il link;
4. Confronta	usa più riferimenti per dati rilevanti o controversi;
5. Correggi	modifica il risultato finale in base alle verifiche.

4. Quali fonti preferire

Una fonte è più affidabile quando proviene da un soggetto competente, è aggiornata, indica chiaramente autore e contesto e consente di verificare l'informazione. Per una funzione di un programma, la documentazione ufficiale è normalmente preferibile a un post casuale. Per una regola scolastica, una fonte istituzionale è più adatta di un commento sui social.

- siti istituzionali;
- documentazione ufficiale del prodotto o del servizio;
- università ed enti di ricerca;

- manuali e testi autorevoli;
- articoli firmati da fonti riconoscibili;
- più fonti indipendenti quando il tema è controverso.

5. Argomenti che richiedono particolare prudenza

Per informazioni mediche, legali, finanziarie o legate alla sicurezza non è sufficiente affidarsi a una risposta AI. Occorre consultare fonti ufficiali e, quando necessario, rivolgersi a un professionista qualificato. L'AI può aiutare a formulare domande o a capire un testo, ma non deve sostituire valutazioni specialistiche.

ATTENZIONE

Non usare una risposta AI per decidere autonomamente una terapia, un investimento, una questione legale oppure un'azione potenzialmente pericolosa. Nei casi delicati servono fonti ufficiali, adulti responsabili o professionisti qualificati.

6. Che cos'è un bias

Un bias è una distorsione che influenza una valutazione. Anche le persone possono avere bias cognitivi. Un sistema AI può riprodurre oppure amplificare distorsioni presenti nei dati, nelle istruzioni o nel modo in cui viene utilizzato. Per esempio, potrebbe associare alcuni lavori soprattutto a uomini oppure a donne sulla base di stereotipi presenti nei dati.

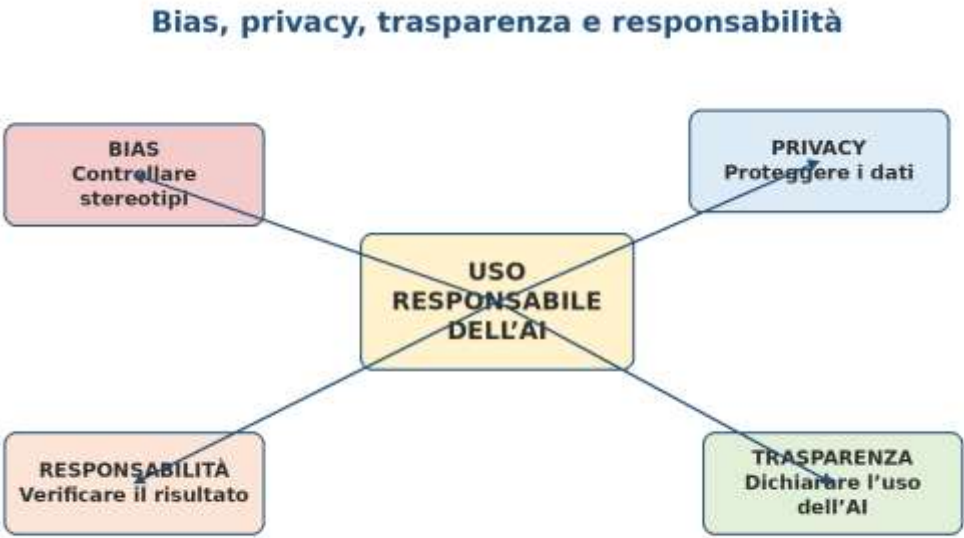


Figura 3 - Quattro aspetti da controllare nell'uso responsabile dell'AI.

Tipo di attenzione	Domanda utile
rappresentazione	il risultato descrive persone e gruppi in modo equilibrato?
generalizzazione	sta trasformando pochi casi in una regola assoluta?
stereotipo	associa ruoli o capacità a un gruppo senza motivo?
contesto	manca qualche informazione che potrebbe cambiare il giudizio?

7. Etica e responsabilità

L'etica riguarda i criteri con cui scegliamo ciò che è corretto, rispettoso e responsabile. Nell'uso dell'AI significa proteggere la dignità delle persone, evitare discriminazioni, rispettare la privacy, controllare le conseguenze e mantenere una supervisione umana.

Principio	Domanda da porsi
privacy	sto inserendo dati personali o riservati?
equità	la risposta tratta correttamente persone e gruppi diversi?
trasparenza	è chiaro che il contenuto è stato prodotto o modificato con AI?
responsabilità	chi controlla il risultato prima di usarlo?
sicurezza	questa richiesta potrebbe causare danni?

8. Guardrail e limitazioni

I guardrail sono regole, controlli e limitazioni progettati per ridurre comportamenti pericolosi o inappropriati. Un sistema può rifiutare richieste che facilitano frodi, reati, violenza, violazioni della privacy o attività rischiose. I guardrail sono utili, ma non sostituiscono il giudizio dell'utente.

Il fatto che un sistema risponda non significa automaticamente che il contenuto sia corretto o adatto al contesto. Allo stesso modo, un rifiuto non deve essere interpretato come un difetto tecnico: può dipendere da una regola di sicurezza o da un limite previsto dal servizio.

9. Privacy: dati da non condividere

Prima di incollare testo, caricare un file o descrivere una situazione personale in un assistente AI, bisogna chiedersi se il contenuto contiene dati riservati. A scuola e nel lavoro occorre rispettare le regole dell'organizzazione e non condividere dati di altre persone senza autorizzazione.

- password e codici temporanei;
- documenti personali;
- dati sanitari non necessari;
- informazioni riservate di scuola o lavoro;
- dati personali di altre persone;
- materiali protetti o non autorizzati;
- dati bancari o finanziari.

DA RICORDARE

Un dato personale non diventa meno personale solo perché viene inserito in una chat. Se non lo pubblicheresti in un luogo visibile a tutti, fermati e verifica se puoi davvero dividerlo.

10. Trasparenza e uso dichiarato

In alcuni contesti è corretto dichiarare che un testo, un'immagine o una bozza sono stati prodotti o migliorati con l'AI. A scuola bisogna seguire le indicazioni del docente. Nel lavoro occorre rispettare le regole dell'organizzazione. Nascondere l'uso dello strumento può trasformare un supporto utile in un comportamento scorretto.

La trasparenza non significa svalutare il lavoro. Significa spiegare correttamente come è stato prodotto un contenuto e quale parte è stata controllata, rielaborata o verificata dalla persona.

11. Glossario essenziale

Termine	Significato semplice
Allucinazione	informazione inesatta o inventata presentata come reale;
Fonte	documento, sito o riferimento usato per controllare un'informazione;
Bias	distorsione che influenza una valutazione;
Privacy	protezione delle informazioni personali e riservate;
Guardrail	regole e controlli che riducono usi pericolosi;
Supervisione umana	controllo esercitato da una persona sulle decisioni e sui risultati;
Trasparenza	chiarezza sul ruolo dell'AI nella produzione di un contenuto;
Equità	trattamento corretto di persone e gruppi diversi;
Fonte ufficiale	riferimento pubblicato dal soggetto competente;
Dato personale	informazione che identifica o riguarda una persona.

12. Sintesi operativa

L'AI è uno strumento potente, ma non infallibile. Può sbagliare, inventare informazioni, riprodurre bias o generare contenuti non adatti al contesto. Usarla bene significa verificare le affermazioni importanti, preferire fonti autorevoli, proteggere i dati personali, dichiarare l'uso quando richiesto e mantenere sempre una responsabilità umana sul risultato finale.

13. Domande di verifica

Rispondi con parole tue. Le domande servono a controllare la comprensione, non a ripetere frasi a memoria.

1. Perché una risposta AI può essere sbagliata anche se è scritta bene?
2. Che cos'è un'allucinazione?
3. Indica tre esempi di possibili allucinazioni.
4. Perché le fonti devono essere aperte e controllate?
5. Quali fonti sono normalmente più affidabili?
6. Perché le informazioni recenti richiedono particolare attenzione?
7. Quando conviene confrontare più fonti?
8. Perché medicina, legge, finanza e sicurezza richiedono prudenza aggiuntiva?
9. Che cos'è un bias?
10. Come può un sistema AI riprodurre stereotipi?
11. Che cosa significa equità?
12. Che cosa significa trasparenza?
13. Perché l'utente mantiene la responsabilità finale?
14. Che cosa sono i guardrail?
15. Perché i guardrail non bastano da soli?
16. Quali dati non devono essere inseriti in una chat AI?
17. Perché non bisogna condividere dati personali di altre persone?
18. Quando è corretto dichiarare l'uso dell'AI?
19. Qual è la differenza tra una fonte ufficiale e un commento casuale?
20. Qual è il metodo generale per verificare una risposta?

14. Risposte commentate

1. Perché il modello produce testo plausibile in base agli schemi appresi, ma non garantisce la verità di ogni dettaglio. La correttezza linguistica non coincide con la correttezza dei fatti.
2. Un'allucinazione è un'informazione inesatta o inventata presentata come se fosse reale. È pericolosa perché può sembrare credibile.
3. Esempi possibili sono una data errata, un libro inesistente oppure una frase mai pronunciata attribuita a una persona.

4. Perché una fonte citata potrebbe non esistere, essere superata oppure non sostenere davvero l'affermazione prodotta.
5. Siti istituzionali, documentazione ufficiale, enti di ricerca e testi autorevoli sono normalmente più affidabili di fonti anonime o commenti casuali.
6. Perché prezzi, versioni software, regole, ruoli e notizie possono cambiare nel tempo. Una risposta non aggiornata può diventare sbagliata.
7. Quando l'informazione è importante, controversa oppure può produrre conseguenze reali.
8. Perché una risposta inesatta in questi ambiti può causare danni concreti e richiede il controllo di un professionista qualificato.
9. Un bias è una distorsione che influenza una valutazione o un risultato.
10. Può apprendere associazioni squilibrate dai dati e riproporle come se fossero naturali o corrette.
11. Equità significa trattare persone e gruppi in modo corretto, senza discriminazioni ingiustificate.
12. Trasparenza significa rendere chiaro il ruolo dell'AI nella produzione o modifica di un contenuto.
13. Perché l'AI non possiede responsabilità morale o giuridica: chi usa il risultato deve controllarlo e decidere come impiegarlo.
14. Sono regole e limitazioni progettate per ridurre usi pericolosi o inappropriati.
15. Perché nessun controllo tecnico sostituisce completamente il giudizio, la prudenza e la responsabilità della persona.
16. Password, codici, documenti personali, dati sanitari non necessari, informazioni riservate e dati di altre persone.
17. Perché non siamo autorizzati a esporre la loro privacy e potremmo causare danni o violare regole scolastiche, lavorative o legali.
18. Quando lo chiedono il docente, l'organizzazione o le regole del contesto, e comunque quando la trasparenza è importante per valutare il contenuto.
19. Una fonte ufficiale proviene dal soggetto competente; un commento casuale può essere utile come opinione, ma non come prova sufficiente.
20. Individuare le affermazioni importanti, chiedere fonti, aprirle, confrontarle e decidere se il risultato è affidabile.

15. Esercizi pratici e attività

Svolgi gli esercizi con calma. Quando utilizzi ChatGPT o altri servizi AI, non inserire dati personali o materiali riservati.

1. Chiedi a ChatGPT quando è stato inventato il mouse e verifica almeno una fonte proposta.
2. Scrivi due esempi inventati di allucinazione credibile.
3. Confronta una risposta AI con una pagina ufficiale e annota eventuali differenze.
4. Chiedi una spiegazione di una funzione di Excel e individua quale fonte ufficiale useresti per controllarla.
5. Cerca una notizia recente e confronta almeno due fonti autorevoli.
6. Prepara una tabella con cinque tipi di fonte ordinati dalla più affidabile alla meno affidabile per una ricerca scolastica.
7. Analizza una risposta che contiene un tono troppo sicuro e riscrivila distinguendo fatti e ipotesi.
8. Scrivi tre esempi di dati che non inseriresti mai in una chat AI.
9. Immagina un sistema che associa alcuni lavori soltanto a uomini: spiega perché è un bias.
10. Prepara un breve codice di comportamento per l'uso dell'AI in classe.
11. Indica tre situazioni in cui serve il supporto di un professionista, non solo dell'AI.
12. Scrivi una didascalia trasparente per un'immagine creata con AI.
13. Confronta una pagina istituzionale e un post sui social sullo stesso tema.
14. Applica il metodo di verifica in cinque passaggi a una domanda scelta dal docente.
15. Scrivi una riflessione di dieci righe: "Perché spirito critico e AI devono lavorare insieme".

16. Riferimenti essenziali

- UNESCO, Recommendation on the Ethics of Artificial Intelligence.
- UNESCO, Guidance for generative AI in education and research.

17. Appendice operativa - Verificare una risposta passo per passo

Immagina che un assistente AI indichi una data storica, un prezzo, una versione software oppure una citazione. La prima operazione non è copiare la risposta, ma individuare quali affermazioni richiedono controllo. Le informazioni

stabili e generali possono essere verificate con calma; le informazioni recenti o delicate richiedono maggiore attenzione.

Esempio guidato: se una risposta indica quando è stato inventato il mouse, bisogna controllare almeno una fonte autorevole, verificare che esista realmente e confrontare le informazioni principali. Se le fonti discordano, conviene cercare ulteriori conferme prima di usare il dato.

Elemento da controllare	Domanda guida
data	la fonte conferma lo stesso anno o periodo?
persona	il nome è scritto correttamente e il ruolo è coerente?
citazione	la frase è davvero presente nella fonte?
versione software	la documentazione è aggiornata?
dato numerico	il numero è riportato nello stesso contesto?

18. Appendice operativa - Bias nella vita quotidiana

Un bias non è sempre evidente. Può comparire quando un sistema propone immagini, associa professioni a determinate persone oppure descrive un gruppo sociale usando stereotipi. Anche una risposta apparentemente neutra merita attenzione se generalizza eccessivamente.

- chiediti se il risultato esclude qualcuno senza motivo;
- controlla se associa ruoli e caratteristiche in modo stereotipato;
- verifica se presenta opinioni come fatti;
- confronta esempi alternativi;
- chiedi di spiegare i limiti della risposta.

19. Checklist privacy

PRIMA DI CARICARE O INCOLLARE

Il testo contiene nomi completi? Contiene dati sanitari, credenziali, numeri di telefono, documenti personali o informazioni riservate? Riguarda altre persone? È materiale protetto o non autorizzato? Se la risposta è sì, non condividerlo senza permesso e senza aver eliminato ciò che non serve.

20. Appendice operativa - Controllare una fonte online

Aprire una fonte non significa soltanto fare clic su un collegamento. È necessario osservare chi pubblica il contenuto, quando è stato aggiornato, quale scopo ha il sito e se l'informazione è confermata da altri riferimenti autorevoli. Un sito graficamente curato può contenere errori; un risultato molto popolare non è automaticamente affidabile.

- controlla il dominio e il soggetto responsabile del sito;
- verifica la data di pubblicazione o di aggiornamento;
- cerca l'autore, quando l'argomento lo richiede;
- distingui una notizia, un'opinione, una pubblicità e una documentazione tecnica;
- confronta almeno una seconda fonte per i dati importanti;
- non usare una citazione se non hai verificato che sostenga davvero l'affermazione.

Per imparare una funzione di un servizio digitale, la pagina ufficiale del produttore è normalmente la fonte più adatta. Un articolo divulgativo può aiutare a comprendere il contesto, ma potrebbe essere meno aggiornato. Le due fonti possono collaborare, ma non hanno lo stesso peso.

21. Mini progetto

Prepara una scheda di verifica per una risposta AI: affermazione da controllare, fonte proposta, fonte realmente aperta, data di aggiornamento, esito del controllo e correzione finale. Applica la scheda a un argomento scelto dal docente.

Campo della scheda	Che cosa scrivere
Affermazione	la frase o il dato prodotto dall'AI;

Fonte proposta	il riferimento indicato dal sistema, se presente;
Fonte aperta	il sito o documento realmente controllato;
Data	la data di pubblicazione o aggiornamento;
Esito	confermato, parzialmente confermato o non confermato;
Correzione	la versione finale dopo il controllo.

22. Checklist finale

CHECKLIST AI RESPONSABILE

Risposta letta criticamente. Affermazioni importanti individuate. Fonti aperte e controllate. Dati recenti verificati. Bias o stereotipi valutati. Nessun dato personale inserito. Uso dell'AI dichiarato quando richiesto. Risultato finale rielaborato e controllato da una persona.