

Allucinazioni e come difendersi

*Perché Claude può sbagliare con sicurezza,
e come non portarlo in atto.*

TRE COSE DA SAPERE

- 1 Cos'è un'allucinazione**
prevede il plausibile, non il vero
- 2 Le forme che ti colpiscono**
citazioni, norme e dottrina inventate
- 3 La difesa**
verifica sempre sulla fonte

salva ora • sfoglia con calma →

Il problema

*Il rischio non è che Claude non sappia.
È che ti risponda sicuro anche quando inventa.*

Sembra una citazione vera

Numero, sezione, anno: la forma è perfetta.
Solo che quella sentenza non esiste.

**La sicurezza del tono
non misura la verità.**

Cos'è un'allucinazione

Claude non consulta una banca dati:
prevede la parola più probabile, una dopo l'altra.

Quando il dato preciso non c'è, riempie il vuoto
con qualcosa di plausibile — che a volte è falso.

IL PUNTO

Non mente: non ha modo di sapere
che in quel punto sta sbagliando.

Plausibile non vuol dire vero.

Le 4 forme che ti colpiscono

Quelle che rischiano di finire in un atto.

1

La sentenza inesistente

numero e anno verosimili, pronuncia mai emessa.

2

La massima distorta

sentenza vera, principio attribuito sbagliato.

3

L'articolo sbagliato

norma reale, numero o comma errato.

4

La dottrina inventata

autore plausibile, scritto mai pubblicato.

Quattro modi diversi di sembrare attendibile.

Perché proprio sul diritto

Il linguaggio giuridico è molto regolare.

« Cass. civ., sez. III, n. _ _ _ _ / 20 _ _ »

uno schema facilissimo da imitare.

Più la forma è prevedibile,
più è facile produrne una falsa che sembra vera.

**La regolarità della forma
è la sua trappola.**

Chi trova l'errore?

Esempio • dati finti

OUTPUT DI CLAUDE

In materia di disdetta della locazione commerciale, la giurisprudenza è consolidata (cfr. Cass. civ., sez. III, n. 18472/2021, secondo cui il preavviso semestrale decorre dalla ricezione...).

Il principio fila. Un dettaglio no.

**Quale parte non puoi copiare in atto
a occhi chiusi?**

Lo smontaggio

Il principio è ragionevole. Ma il numero è l'esca:
apri la banca dati — quella pronuncia non risulta.

LA CITAZIONE SOSPETTA

~~Cass. civ., sez. III, n. 18472/2021~~



Non risulta in banca dati. Non va in atto.

Finché non l'hai vista,
la citazione non esiste.

Difesa 1 — il guardrail

Nel profilo tieni un'istruzione fissa:

DA COPIARE NEL PROFILO

« Non inventare riferimenti normativi o giurisprudenziali. Se non sei sicuro, scrivilo e segnala [DA VERIFICARE]. »

Riduce molto il problema. Ma non lo azzerà.
È un airbag, non un freno: ti aiuta, non ti dispensa.

Il guardrail abbassa il rischio.
Non lo annulla.

Difesa 2 — àncora alla fonte

Tre strade, in ordine di solidità.

1

Skill Normattiva

link diretto alla norma sul portale ufficiale: fonte primaria, in un clic.

2

Banca dati via MCP

Claude interroga la banca dati (es. BuddaLaw, Simpliciter), non la memoria.

3

Perplexity

cita le sue fonti web. Non azzera l'errore: lo sposta sull'attendibilità del sito.

Nessuna citazione in atto
senza la fonte aperta sotto gli occhi.

Da qui puoi

La verifica anti-allucinazione, ogni volta.

- ☐ Chiedi: « quali citazioni hai ricostruito a memoria? »
- ☐ Pretendi il [DA VERIFICARE] sulle fonti
- ☐ Norma → aprila con la skill Normattiva
- ☐ Giurisprudenza → banca dati (MCP)
- ☐ Con Perplexity, guarda quale sito sta citando
- ☐ In dubbio, niente atto.

Trenta secondi di verifica, zero citazioni fantasma.

I 3 errori da evitare



Fidarsi del tono sicuro

La certezza del tono non misura la verità.

Fix: verifica la fonte, non l'aplomb.



Controllare solo le citazioni « strane »

Le verosimili sono le più pericolose:
convincono e passano il controllo.

Fix: verifica tutte, anche quelle che ti convincono.



Pensare che basti il guardrail

O che basti Perplexity perché « cita le fonti ».
Riducono, non eliminano.

Fix: la fonte primaria la apri comunque tu.

L'errore più caro è quello che sembrava perfetto.

Verifica, poi firma.

*La prossima volta: la ricerca giuridica fatta bene —
Claude da solo vs ibrido, e la verifica a livelli.*

Puntata 13 — Ricerca giuridica.

Per ogni dubbio o se hai bisogno di aiuto, scrivimi.

Avv. Giuseppe Girlando
giuseppe.girlando@orbitalaw.it

OrbitaLaw — *l'agenda legale che ti fa risparmiare tempo*

orbitalaw.it