

Il tuo sito e invisibile per l'AI?

Come controllare se i bot AI riescono a leggerti oggi.
E come sistemarlo in 5 minuti.

Prima di tutto: una distinzione che cambia tutto.

Quando si parla di AI e visibilità online si mescolano due cose diverse. Confonderle porta a soluzioni sbagliate.

TRAINING PASSATO

I modelli AI (ChatGPT, Claude, Gemini) sono stati addestrati su dati raccolti fino a una certa data. Quello che sanno già, lo sanno. robots.txt non cambia niente di quello che è già successo.

Se oggi il tuo sito non è mai citato nelle risposte AI, il motivo è che non era abbastanza visibile quando i dati sono stati raccolti. Non è colpa del tuo robots.txt.

RICERCA IN TEMPO REALE

Strumenti come ChatGPT Search, Perplexity e Claude con accesso web recuperano informazioni live quando un utente fa una domanda. Qui robots.txt conta eccome.

Se blocchi i bot di ricerca in tempo reale, non compari nelle risposte live di oggi e di domani. Questo è il problema che puoi risolvere adesso.

QUINDI: cosa risolve questo PDF

Non ti rende visibile nei dati di training passati. Quello non si cambia.

Ti aiuta a non bloccare i bot di ricerca in tempo reale, quelli che decidono se citarti nelle risposte live di ChatGPT, Claude e Perplexity da oggi in poi.

I bot AI: chi fa cosa.

Nel 2026 ci sono due categorie di bot AI che visitano il tuo sito. Bloccare quella sbagliata ti rende invisibile nelle risposte live.

Bot	Categoria	Se lo blocchi
GPTBot ClaudeBot	Training futuro	Il tuo sito non entra nei dati di training futuri. Impatto basso sulla visibilita attuale.
OAI-SearchBot Claude-SearchBot PerplexityBot ChatGPT-User Claude-User	Ricerca live	Non compari nelle risposte in tempo reale di ChatGPT, Claude e Perplexity. Questo e il blocco che conta.

LA REGOLA

Puoi bloccare i bot di training (GPTBot, ClaudeBot) e permettere quelli di ricerca. Sono cose separate e si configurano separatamente.

Il problema e il "**Disallow: /**" generico che blocca tutto, inclusi i bot di ricerca live che ti farebbero comparire nelle risposte di oggi.

Come controllare il tuo robots.txt adesso.

Trova il tuo robots.txt

1

Apri il browser e vai su: **tuodominio.com/robots.txt**

Se la pagina non esiste, non hai problemi di blocco e puoi saltare a pagina 5. Se esiste, leggi quello che c'è scritto con attenzione.

Cerca queste righe pericolose

2

User-agent: * seguito da **Disallow: /** significa che stai bloccando TUTTI i bot, inclusi quelli di ricerca AI live. Se vedi OAI-SearchBot, Claude-SearchBot o PerplexityBot con **Disallow: /**, stai bloccando specificamente i bot che ti farebbero comparire nelle risposte.

Controlla anche il JavaScript

3

Il 69% dei crawler AI non esegue JavaScript. Se il tuo sito carica i contenuti con JS (WordPress con builder, siti React o Vue), i bot vedono una pagina vuota anche se il robots.txt è corretto. Verifica aprendo il sito con JavaScript disabilitato nel browser.

Così appare un robots.txt corretto.

Copia questo file e sostituisci il contenuto del tuo robots.txt. Verde = permetti ai bot di ricerca. Rosso = blocca il training. Se usi WordPress: **File manager > public_html > robots.txt**.

```
robots.txt - tuodominio.com/robots.txt

# BOT DI RICERCA AI: permettili tutti
User-agent: OAI-SearchBot
Allow: /

User-agent: Claude-SearchBot
Allow: /

User-agent: PerplexityBot
Allow: /

User-agent: ChatGPT-User
Allow: /

User-agent: Claude-User
Allow: /

# BOT DI TRAINING: bloccali se vuoi
User-agent: GPTBot
Disallow: /

User-agent: ClaudeBot
Disallow: /

# TUTTI GLI ALTRI: permetti
User-agent: *
Allow: /
Disallow: /admin/

Sitemap: https://tuodominio.com/sitemap.xml
```



Permetti (ricerca live)



Blocca (training)



Commento

Gli errori piu comuni. E come sistemarli.

Questi sono i 4 errori che vediamo piu spesso nei robots.txt. Se riconosci il tuo, sai gia cosa fare.

Errore 1: blocco totale con Disallow: /

SBAGLIATO

```
User-agent: *  
Disallow: /
```

Stai bloccando tutto: Google, Bing, ChatGPT, Claude. E rimasto da quando qualcuno ha detto "metti robots.txt per la privacy" senza capire cosa fa.

CORRETTO

```
User-agent: *  
Allow: /  
Disallow: /admin/
```

Permetti tutto tranne le aree private. Gestisci i bot AI singolarmente come nel mockup sopra.

Errore 2: blocco esplicito su OAI-SearchBot

SBAGLIATO

```
User-agent: OAI-SearchBot  
Disallow: /
```

Blocca la ricerca live di ChatGPT. Spesso aggiunto copiando guide anti-scraping senza distinguere training da ricerca in tempo reale.

CORRETTO

```
User-agent: OAI-SearchBot  
Allow: /
```

Cambia Disallow in Allow. OAI-SearchBot e il bot che ti fa comparire nelle risposte live di ChatGPT.

Errore 3: meta tag noindex su pagine importanti

SBAGLIATO

```
content="noindex, nofollow">
```

Il robots.txt puo essere corretto, ma singole pagine con questo tag nell'HTML risultano bloccate per i crawler AI. Comune su siti WordPress o Squarespace.

CORRETTO

Controlla SEO plugin per ogni pagina importante del sito

In WordPress: controlla che nessuna pagina abbia noindex attivo nel plugin SEO (Yoast o RankMath). In Squarespace: Pages > SEO > Page indexing.

Errore 4: sito tutto in JavaScript senza rendering lato server

SBAGLIATO

```
# robots.txt corretto, ma  
# contenuti caricati via JS  
# i bot vedono pagina vuota
```

Il robots.txt e pulito ma i contenuti sono generati da JavaScript. I crawler AI vedono una pagina vuota. Vale per molti siti con builder moderni o framework come React e Vue.

CORRETTO

```
curl -A "Googlebot"  
https://tuodominio.com/  
# Devi vedere il testo del sito
```

Se il comando restituisce HTML vuoto, il sito non e leggibile dai crawler. Soluzioni: abilita il rendering lato server, usa un pre-renderer, o assicurati che i contenuti chiave siano in HTML statico.

La checklist completa.

Il robots.txt è solo il primo passo tecnico. Queste sono tutte le cose da verificare per non essere invisibile alle risposte AI live.

robots.txt corretto

+ Nessun blocco sui bot di ricerca AI live. Configurazione separata per training vs ricerca.

Sito leggibile senza JavaScript

+ I contenuti principali devono essere visibili nel codice HTML senza eseguire JS. Testa con JavaScript disabilitato nel browser.

Sitemap.xml presente e aggiornata

+ Aiuta i crawler a trovare tutte le pagine. Va dichiarata in fondo al robots.txt.

Nessun meta noindex sulle pagine importanti

+ Controlla che le pagine chiave non abbiano il tag noindex nell'HTML o nelle impostazioni SEO.

Contenuto testuale chiaro e strutturato

+ L'AI privilegia siti con testo leggibile, paragrafi chiari, titoli H1 e H2 definiti.

lms.txt - opzionale ma in crescita

+ Un file che guida gli agenti AI verso i tuoi contenuti più importanti. Adozione ancora bassa (~10%) ma in crescita rapida nel 2026.

Non basta esserci.
Bisogna essere leggibili.