ROBOTS.TXT CONTRO L'ADDESTRAMENTO AI, SENZA SPARIRE DA GOOGLE E CHATGPT

Il file pronto per una PMI: le righe che valgono oggi, crawler per crawler, e il segnale comune annunciato il 30 ottobre 2026, che finora ha cominciato a riconoscere solo Anthropic. Fonti lette l'11 novembre 2026.

IL FILE

```
# robots.txt per una PMI: fuori dall'addestramento dei modelli AI, dentro Google, Bing e le risposte AI.
# Righe verificate l'11 novembre 2026 sulle pagine dei proprietari e sulla bozza IETF.
# Guida e fonti: itria.io/addestramento-ai-robots-txt
# Il file esprime una preferenza: chi lo legge sceglie di rispettarlo. Va messo in ogni sottodominio.

# Tutti gli altri, Googlebot e Bingbot compresi: si entra.
# Non bloccarli mai per motivi AI: toglierebbero il sito da Google, da Bing e da Copilot.
User-agent: *
Allow: /

# ADDESTRAMENTO, RIGHE CHE VALGONO GIA' OGGI (una per azienda).
# OpenAI: il sito resta nella ricerca di ChatGPT, che usa OAI-SearchBot.
User-agent: GPTBot
# Anthropic: valgono per i contenuti futuri. Il sito resta nella ricerca di Claude.
User-agent: ClaudeBot
# Apple: Applebot continua a leggere per Siri, Spotlight e Safari.
User-agent: Applebot-Extended
# Common Crawl: l'archivio aperto del web.
User-agent: CCBot
# Google: niente addestramento di Gemini, ma anche niente citazioni con link
# nelle app Gemini (grounding). AI Overview e AI Mode non dipendono da qui.
# Chi vuole restare citato nelle app Gemini cancella questa riga e scrive
# un gruppo a parte: User-agent: Google-Extended / Allow: /
# (e allora Google puo' anche addestrare).
User-agent: Google-Extended
Disallow: /

# RICERCA E RISPOSTE AI: dentro. Le righe Allow rendono la scelta leggibile.
User-agent: OAI-SearchBot
User-agent: Claude-SearchBot
User-agent: Claude-User
Allow: /

# SEGNALE COMUNE CONTRO L'ADDESTRAMENTO, annunciato il 30 ottobre 2026.
# Non e' il nome di un crawler: dice a chi lo riconosce "non addestrare".
# Anthropic: riconosciuto dal 30 ottobre 2026, con estensione graduale.
# Microsoft: dal 1 febbraio 2027. Google: nel 2027. OpenAI: "nei prossimi mesi".
# Non sostituisce i gruppi sopra: si tengono tutti e due.
User-Agent: __usage_ai-training
Disallow: /

Sitemap: https://www.tuosito.it/sitemap.xml
```

I NOMI DEL FILE E COSA SUCCEDE SE LI BLOCCHI

Nome	Azienda	Cosa fa	Se lo blocchi	Da quando vale
GPTBot	OpenAI	Addestramento	Testi fuori dall'addestramento; resti nella ricerca di ChatGPT	Già oggi
OAI-SearchBot	OpenAI	Ricerca di ChatGPT	Esci dalle risposte di ricerca di ChatGPT: da non bloccare	Già oggi, circa 24 ore dopo il cambio
ClaudeBot	Anthropic	Addestramento	Contenuti futuri fuori dall'addestramento	Già oggi
Claude-SearchBot	Anthropic	Ricerca di Claude	Meno visibilità nelle risposte di Claude: da non bloccare	Già oggi
Google-Extended	Google	Addestramento e grounding di Gemini	Fuori da entrambi; Ricerca e AI Overview restano	Già oggi

Googlebot	Google	Ricerca, AI Overview, AI Mode	Esci da Google: da non bloccare	Già oggi
Applebot-Extended	Apple	Addestramento (solo uso)	Fuori dall'addestramento; Applebot resta	Già oggi
CCBot	Common Crawl	Archivio aperto del web	Esci dall'archivio	Già oggi
__usage_ai-training	Segnale comune	No all'addestramento, per chi lo riconosce	Nessuna perdita nella ricerca	Anthropic dal 30/10/2026; Microsoft dal 1/2/2027; Google nel 2027; OpenAI nei prossimi mesi

IL CALENDARIO DEL SEGNALE COMUNE

Data	Cosa succede
30 ottobre 2026	Annuncio di Google, OpenAI e Microsoft; bozza IETF -00; Anthropic comincia a riconoscere il segnale
15 dicembre 2026	Bing Webmaster Tools apre la preferenza temporanea di grounding per chi usa NoArchive o NoCache
1 febbraio 2027	Microsoft rispetta le nuove preferenze man mano che ripassa sulle pagine
1 marzo 2027	NoArchive e NoCache non valgono più per addestramento e grounding su Bing
3 maggio 2027	Scade la bozza IETF -00, se non viene rinnovata
30 settembre 2027	Scadono le preferenze temporanee di Bing Webmaster Tools
Nel 2027	Google comincia a leggere il segnale (la documentazione dice inizio 2027)
Nei prossimi mesi	OpenAI lo implementa, con un aggiornamento dei dettagli annunciato

*User-agent: * con Disallow: / non esprime il no all'addestramento (bozza IETF) e toglie il sito da tutti i motori. Il segnale vale per i contenuti prelevati da quando è scritto: non toglie quelli già usati. Microsoft scrive che nessuna delle due scelte garantisce inclusione, traffico o pagamento.*

La guida completa, con le fonti:
itria.io/addestramento-ai-robots-txt

File libero, costruito da Itria sulle pagine pubbliche di Google, OpenAI, Microsoft, Anthropic, Apple, Common Crawl e sulla bozza IETF, lette l'11 novembre 2026. Le date sono annunci: ricontrollate le fonti prima di pubblicare il file.

Mattia Esposito lavora sul posizionamento delle aziende, con analisi di mercato per capire dove i concorrenti sono scoperti. Due righe su WhatsApp al 350 574 3578 o una mail a info@itria.io.