Un sito agent-ready è un sito che un programma riesce a leggere e a usare al posto di una persona, ovvero un sito in cui i contenuti sono ordinati, le informazioni importanti stanno scritte in chiaro e i passaggi per compiere un'azione si capiscono anche senza guardare la pagina.

Sempre più spesso a navigare, confrontare e comprare è un agente AI che lavora per conto di una persona. Per chi ha un sito la questione diventa allora pratica più che teorica, perché prepararsi a essere letti da un programma richiede scelte concrete sui contenuti, sui dati e sui moduli, mentre il tema si muove in fretta e diverse proposte in circolazione sono ancora giovani.

In questa guida raccontiamo che cosa fanno gli agenti quando arrivano su un sito, che cos'è il file llms.txt di cui si parla molto, quali dati strutturati servono, come si riconoscono i crawler degli assistenti e che cosa si può controllare da soli, senza promettere numeri che al momento nessuno ha.

Che cosa fa un agente quando arriva sul tuo sito

Un agente AI è un programma che compie azioni per conto di una persona, ovvero cerca, confronta, prenota, compra. Legge i siti in un modo tutto suo, ovvero contenuti, struttura e dati, senza badare a colori e animazioni, quindi una pagina bellissima ma povera di informazioni scritte gli dice molto poco.

Quello che gli agenti fanno già oggi si può riassumere in tre attività: confrontano prodotti e prezzi, compilano moduli e prenotano, scelgono le fonti da cui comprare. Naturalmente non tutti si comportano allo stesso modo, perché alcuni arrivano sul sito nel momento in cui una persona sta chiedendo qualcosa, mentre altri lo hanno letto tempo prima e lo citano a memoria, quindi la leggibilità conta in entrambi i momenti.

Per chi vende, la parte interessante sta nel fatto che un agente si ferma dove si fermerebbe una persona: se la disponibilità non è scritta, se il modulo chiede dieci campi, se le condizioni di reso stanno dentro un'immagine, il percorso si interrompe. Le stesse cure che rendono un sito comodo lo rendono anche usabile da un programma: da qui viene il fatto che il lavoro tecnico e il lavoro sui testi, su questo tema, si toccano.

Che cos'è il file llms.txt e com'è fatto

Il file llms.txt è un file di testo in formato Markdown che si mette nella radice del sito e

presenta i contenuti a chi arriva con un modello linguistico, ovvero un indice ragionato che dice di che cosa si occupa il sito e dove stanno le pagine che contano. La proposta risale al 2024 ed è di Jeremy Howard: la versione aggiornata ad agosto 2026 si legge su llmstxt.org.

La struttura è semplice e regolare: l'unica parte obbligatoria è il titolo in H1 con il nome del sito, seguito da un riassunto breve in blockquote, eventuali informazioni di contesto e poi sezioni in H2 con elenchi di link, ciascuno con una riga di descrizione. Il formato è pensato per essere leggibile da una persona e insieme processabile da una macchina, quindi si scrive bene con le stesse parole con cui si presenterebbe il sito a un collega.

Dove si mette e come si genera su WordPress

L'indirizzo è sempre lo stesso, www.iltuosito.it/llms.txt, così chi lo cerca lo trova senza doverlo chiedere. Su WordPress la generazione automatica è arrivata dentro gli strumenti che già si usano per la SEO, Yoast e AIOSEO fra gli altri, mentre le piattaforme di documentazione come GitBook e Mintlify lo producono da tempo per conto proprio.

Chi lo scrive a mano ottiene di solito un file migliore, perché la scelta delle pagine da segnalare è una scelta di strategia e non un'operazione meccanica: un file che elenca tutto assomiglia alla mappa del sito e non aggiunge niente, mentre un file che indica le venti pagine che spiegano l'offerta racconta qualcosa.

L'altro formato di cui si parla, l'Open Knowledge Format

Accanto a llms.txt circola l'Open Knowledge Format, annunciato da Google Cloud il 12 giugno 2026 e arrivato alla versione 0.1: cartelle di file Markdown con un frontmatter in YAML, un file per concetto invece di un file per pagina, così un agente prende soltanto il pezzo che gli serve. La specifica sta sul blog di Google Cloud.

La documentazione serve soprattutto a capire il perimetro, che è diverso da quello che si immagina: l'OKF nasce per la conoscenza interna delle organizzazioni, ovvero schemi di tabelle, significato delle metriche, procedure, mentre il sito pubblico resta fuori dal suo scopo. Per adesso, quindi, è un formato da guardare con curiosità se in azienda ci sono documenti e dati da mettere in ordine per lavorare con l'AI, mentre sul sito il lavoro utile resta un altro.

Chi pubblica llms.txt e chi dice di non usarlo

Sul file convivono due fatti che sembrano contraddirsi. Da una parte llms.txt si è diffuso: migliaia di siti lo pubblicano, i laboratori che fanno i modelli pubblicano il proprio, mentre Lighthouse, lo strumento di analisi dentro Chrome, lo controlla in un gruppo di verifiche chiamato Agentic browsing, insieme agli strumenti per gli agenti e all'accessibilità delle pagine.

Dall'altra parte Google dice che per comparire nelle funzioni AI della ricerca non serve nessun file speciale e nessuna marcatura dedicata, come si legge nella sua documentazione sulle funzioni AI. A giugno 2026 John Mueller ha aggiunto che il file esiste da anni senza che i sistemi AI lo usino, quindi ogni previsione sul suo effetto resta una supposizione.

L'incastro fra le due cose lo ha spiegato bene Marie Haynes: llms.txt è per gli agenti che usano il sito, non per il motore di ricerca, quindi le due affermazioni parlano di due mondi diversi. La nostra indicazione, di conseguenza, è misurata: pubblicarlo costa poco e non fa danno, tenerlo aggiornato è una buona abitudine, mentre chi promette risultati misurabili in cambio di un file di testo sta vendendo qualcosa che oggi non si può dimostrare.

I dati strutturati, la parte che funziona da anni

I dati strutturati sono marcature invisibili a chi legge, ovvero righe di codice nel vocabolario schema.org che dicono a una macchina che quel numero è un prezzo, che quella data è l'inizio di un evento e che quel testo è una domanda con la sua risposta. Sono la parte meno nuova di questa guida ed è probabilmente per questo che vengono trascurate, mentre sono l'unica marcatura che le macchine leggono da anni.

Le tipologie che rendono di più sono poche e si scelgono in base a quello che si fa: prodotto con prezzo e disponibilità per chi vende, sede e orari per chi riceve clienti, articolo per i contenuti informativi, evento per i corsi e gli appuntamenti, domande frequenti dove esistono già in pagina. La regola del buon senso qui coincide con quella tecnica, perché una marcatura che dichiara una cosa diversa da quella scritta nel testo crea confusione e viene ignorata.

Sulle funzioni AI della ricerca, come si è visto, Google dichiara di non aspettarsi nessuno schema dedicato. I dati strutturati restano utili per quello che sono sempre stati, ovvero far capire senza equivoci che cosa contiene una pagina: l'aiuto arriva quindi anche ai programmi che leggono il sito per conto di qualcuno.

Le informazioni che un agente riesce a verificare

Un agente si comporta come un compratore diffidente e informato, quindi cerca le cose che si possono confrontare: prezzi, disponibilità, tempi di consegna, condizioni di reso, coperture geografiche. Se queste informazioni stanno scritte nel testo della pagina vengono lette, mentre se vivono soltanto dentro un'immagine, in un PDF del listino o in una risposta che si ottiene scrivendo a un indirizzo email, per lui non esistono.

Torna poi la vecchia osservazione sui moduli, che qui torna con un significato nuovo: quello che è facile per una persona lo è anche per un agente, quindi un modulo con i campi essenziali e un percorso di acquisto con pochi passaggi si lasciano completare da entrambi. I passaggi che chiedono di interpretare un'immagine o di indovinare dove cliccare, viceversa, fermano tutti e due.

Su questo terreno si stanno muovendo anche i protocolli per gli acquisti fatti dagli agenti, come l'Universal Commerce Protocol, che Google descrive come uno standard in evoluzione a cui i negozi accedono da lista d'attesa, con Merchant Center e il pagamento già configurati. Per adesso riguarda una minoranza di siti e si può seguire da lontano, perché il presupposto rimane sempre lo stesso, ovvero dati corretti e aggiornati sulle proprie pagine.

Quando l'agente non legge ma agisce, ovvero WebMCP

WebMCP è una proposta che permette a un sito di dichiarare i propri moduli e le proprie funzioni come strumenti che un agente può usare, invece di lasciargli indovinare l'interfaccia guardando la pagina. Il funzionamento parte da Chrome 149 in fase di prova pubblica, come raccontato sul blog di Chrome for Developers il 9 giugno 2026.

Le due strade previste sono di difficoltà molto diversa: si possono annotare i moduli in HTML con attributi che spiegano a che cosa servono, oppure si possono registrare strumenti veri e propri con qualche riga di JavaScript, dichiarando nome, descrizione e dati richiesti. Lighthouse verifica se in pagina ci sono strumenti registrati e segnala i moduli che potrebbero dichiararsi. Lo stesso Mueller, nella risposta di giugno su llms.txt, ha detto di trovare più convincente questo approccio.

Trattandosi di una prova pubblica, WebMCP non è una voce da mettere in un preventivo di rifacimento, cosa che va detta con chiarezza perché in giro si legge il contrario. Nel frattempo, però, la parte che serve comunque si può fare adesso: moduli con etichette scritte per esteso, pulsanti che dicono che cosa fanno, campi con nomi comprensibili, ovvero

le stesse cose che rendono un sito accessibile alle persone che navigano con la tastiera o con un lettore di schermo.

I crawler e gli agenti che passano dal tuo sito

I programmi che leggono i siti per conto delle aziende di AI si dividono per scopo, ed è questa la distinzione che serve quando si decide chi lasciare passare. Alcuni raccolgono contenuti che possono servire ad addestrare i modelli, altri costruiscono l'indice con cui l'assistente risponde alle domande, altri ancora visitano una pagina nel momento esatto in cui una persona ha chiesto qualcosa che la riguarda.

I nomi si trovano nelle documentazioni ufficiali e cambiano numero di versione, quindi nei registri si guarda il nome. OpenAI dichiara GPTBot per l'addestramento, OAI-SearchBot per la ricerca dentro ChatGPT, OAI-AdsBot per il controllo delle pagine promosse e ChatGPT-User per le visite che partono da una richiesta di una persona, mentre Anthropic distingue ClaudeBot, Claude-User e Claude-SearchBot con la stessa logica. Google, dal suo lato, tiene separato Google-Extended per l'uso dei contenuti nei suoi modelli.

Le regole si scrivono nel file robots.txt, dove la conseguenza va valutata prima di scriverla: bloccare i programmi che raccolgono per l'addestramento è una scelta legittima che molti editori hanno fatto, mentre bloccare quelli che costruiscono le risposte significa uscire dalle risposte, quindi rinunciare a essere citati. Le visite avviate da una persona, infine, non seguono necessariamente robots.txt, proprio perché somigliano a una persona che apre una pagina.

Come si misura, per ora

La misurazione di questo canale è la parte più giovane di tutte, però qualcosa si vede. Da giugno 2026 Search Console ha un rapporto dedicato alle prestazioni nell'AI generativa, che mostra le impressioni dentro AI Overviews e AI Mode divise per pagina, paese, dispositivo e data, come si legge sul blog di Google Search Central.

Quello che il rapporto non dà è altrettanto importante da sapere, perché evita conclusioni affrettate: mancano i clic, le query, il tasso di clic e la posizione, quindi si osserva la presenza e non il rendimento. Su Analytics, in parallelo, le visite che arrivano dagli assistenti si riconoscono dalle sorgenti, tenendo conto che una quota di sessioni resta senza provenienza dichiarata e che i numeri, oggi, sono piccoli quasi per tutti.

Gli ordini di grandezza, per adesso, sono questi. Sul nostro sito, che è piccolo, le visite arrivate dagli assistenti valgono attorno a mezzo punto percentuale in dodici mesi; sui siti che seguiamo la quota va da un decimo di punto a circa un punto, senza che la dimensione del sito basti a spiegare la differenza. In tutti i casi la parte grossa arriva da ChatGPT, con Gemini, Perplexity, Copilot e Claude molto più indietro.

La direzione, poi, non è la stessa per tutti. Confrontando gli ultimi sei mesi con i sei precedenti, su alcuni siti le visite dagli assistenti quasi raddoppiano mentre il traffico complessivo resta uguale o scende un poco, su altri si dimezzano senza che il totale si muova. Da qui viene la prudenza di questa guida, che è una scelta tecnica: la percentuale di qualcun altro non dice niente sulla tua, quindi la cosa utile è guardare i propri dati per qualche mese e decidere con quelli.

Questo è anche il motivo per cui su questa guida non troverai promesse: il commercio via agenti è agli inizi e nessuno può garantire risultati, mentre prepararsi costa poco e si fa una volta. La distanza fra quello che gli agenti sanno già fare e quello che i siti offrono loro, a dire il vero, sembra ancora larga, quindi il lavoro utile sta più nel mettere in ordine le informazioni che nell'inseguire le novità.

Le verifiche che puoi fare oggi sul tuo sito

La prima verifica è il file robots.txt, che si apre scrivendo il proprio indirizzo seguito da /robots.txt: lì si legge chi è ammesso e chi no, con blocchi che capita spesso di trovare messi anni fa per ragioni che non valgono più. Il nostro, per dire, è aperto a tutti e llms.txt non c'è ancora, quindi anche noi abbiamo una cosa da fare.

La seconda è una lettura delle pagine che contano con l'occhio di un programma, ovvero controllando che prezzi, disponibilità, condizioni e contatti siano testo e non immagini, che ogni pagina risponda a una domanda dichiarata nel titolo e che le informazioni di base non stiano soltanto nel PDF allegato. La terza è la prova diretta: chiedere a un assistente informazioni sulla propria azienda e guardare che cosa risponde, da dove pesca e che cosa sbaglia.

Alla domanda che a questo punto arriva sempre, ovvero se serva rifare il sito, la risposta è quasi sempre no: quello che serve è mettere in chiaro le informazioni, sistemare i dati strutturati e tenere i moduli semplici, tre lavori che si fanno sul sito che c'è. Un rifacimento si giustifica quando la struttura stessa impedisce di scrivere le informazioni dove servono, mentre in quel caso il tema diventa più ampio degli agenti.

Il sito agent-ready dentro il gioco del cliente

Nel gioco del cliente, l'esercizio con cui si disegna il percorso di chi compra da te, questo tema sta nella terza fase, quella dell'acquisto, dove il luogo è fisso ed è il tuo sito. Al posto degli strumenti, in quella fase, ci sono le nove cure del sito, fra cui una si chiama proprio «Dati per gli agenti AI».

La descrizione che si legge nel gioco è la stessa cosa raccontata in una riga: informazioni ordinate e leggibili da un programma, ovvero prezzi, disponibilità, tempi e condizioni, così gli assistenti che comprano al posto delle persone trovano quello che serve. Accanto sta l'infrastruttura tecnica, che nel gioco chiede che il sito sia leggibile in modo chiaro e veloce dai motori e dagli assistenti, quindi le due cure si tengono per mano.

Giocare quella fase serve anche a vedere una cosa che dai dati non si vede: se le altre otto cure sono scoperte, i dati per gli agenti da soli non portano da nessuna parte, perché un agente che trova prezzi impeccabili dentro un sito lento e senza recensioni arriva alla stessa conclusione di una persona. La sequenza con cui si mettono a posto le cose, insomma, resta un ragionamento umano.

Da dove continuare

Questa guida basta per capire il tema e fare le prime verifiche da soli. Chi vuole sapere come sta il proprio sito trova l'analisi della leggibilità per gli agenti dentro l'audit SEO, che comprende il controllo del robots.txt, le pagine che rispondono a una domanda in chiaro e le impressioni dentro le funzioni AI di Search Console.

Chi invece vuole lavorare sulla visibilità dentro le risposte, ovvero farsi nominare quando qualcuno chiede consiglio a un assistente, trova il lavoro descritto nella pagina Visibilità AI e GEO. Sono due cose diverse e le teniamo distinte: qui si parla di essere leggibili e usabili da un programma, là di essere scelti come fonte.