MCP e OSINT: come l'AI verifica le fonti nel 2026

Maria Cattini | 06/08/2026 | Intelligenza Artificiale

Ho chiesto a un assistente AI di verificare un profilo sospetto. Ecco cosa è cambiato nel 2026

Immagina di ricevere un messaggio da un profilo che dice di lavorare per un'azienda che conosci. Foto professionale, bio credibile, qualche post plausibile. Vuoi capire se è reale prima di rispondere. Fino a poco tempo fa il percorso era manuale: copiare il nome utente, controllarlo su una decina di piattaforme, incollare il dominio del sito citato in bio su un paio di strumenti di analisi, confrontare i risultati a mano. Mezz'ora buona, se sei allenato.

Oggi, con gli strumenti giusti collegati, puoi scrivere la stessa richiesta in una chat con un assistente AI e ottenere in pochi secondi un primo quadro: dove esiste quel nome utente, che reputazione ha il dominio citato, se l'indirizzo email associato è comparso in una violazione di dati nota. Il cambiamento non è che l'intelligenza artificiale abbia imparato a "indagare" da sola. È che adesso può usare gli stessi strumenti che usa un analista, in modo diretto e controllato, grazie a un protocollo chiamato **MCP**.

Vale la pena capire come funziona, perché apre possibilità reali per chiunque faccia verifica delle fonti, ma nasconde anche una trappola precisa in cui è facile cadere.

Cos'è l'MCP e perché conta per la verifica digitale

MCP è la sigla di **Model Context Protocol**, uno standard che permette a un assistente AI di collegarsi direttamente a strumenti esterni: un motore di ricerca IP, un database di domini, un archivio di violazioni di dati. Prima di questo standard, ogni collegamento tra un modello linguistico e un servizio esterno andava costruito su misura, pezzo per pezzo. Con l'MCP, chi sviluppa un servizio pubblica un "connettore" e qualsiasi assistente compatibile può usarlo, senza reinventare l'integrazione ogni volta.

Per capirlo con un paragone semplice: prima l'AI poteva solo raccontarti cosa sapeva, sulla base di ciò che aveva letto durante l'addestramento. Con l'MCP, invece, può alzare la cornetta e chiamare davvero un servizio esterno, leggere la risposta e usarla. La differenza è enorme. Un modello che "ricorda" un dato su un dominio dal proprio addestramento è, nella migliore delle ipotesi, un'informazione vecchia. Un modello che interroga in tempo reale un servizio di reputazione dei domini attraverso MCP ottiene un dato verificabile, con una fonte precisa dietro.

Questa distinzione è il cuore di tutto il discorso. Non riguarda solo l'OSINT: riguarda chiunque usi un assistente AI per prendere decisioni basate su informazioni che devono essere vere, non plausibili.

Come funziona in pratica: la parte operativa

Ecco cosa succede, passo per passo, quando un assistente collegato tramite MCP gestisce una verifica come quella del profilo sospetto di prima.

1. Ricevi la richiesta e la scomponi in domande verificabili. Non "questo profilo è falso?", ma

tre domande concrete: il nome utente esiste altrove con la stessa combinazione di foto e bio? Il dominio citato ha una reputazione pulita o è segnalato come sospetto? L'indirizzo email, se disponibile, è comparso in una violazione di dati nota? Scomporre la domanda serve a evitare che l'assistente risponda con un'impressione generica invece che con dati puntuali.

2. L'assistente richiama gli strumenti collegati, uno per uno. Un connettore verifica la presenza del nome utente su più piattaforme in parallelo. Un altro interroga un servizio di reputazione per il dominio. Un altro ancora controlla se l'email compare in archivi noti di violazioni. Ogni chiamata produce un risultato tracciabile: puoi sempre chiedere "da dove viene questo dato" e ottenere il nome del servizio interrogato, non una frase vaga come "secondo le mie fonti".

3. L'assistente legge i risultati e li mette a confronto, non li inventa. Qui sta il punto che va sempre controllato: il risultato deve provenire da una chiamata reale a uno strumento esterno, non da una risposta generata "a memoria". Se chiedi conferma e l'assistente non riesce a indicarti quale strumento ha usato per un'affermazione specifica, quell'affermazione va trattata come non verificata.

4. Tu, non l'assistente, decidi cosa significano i dati raccolti. Trovare lo stesso nome utente su cinque piattaforme non prova che il profilo sia legittimo: prova solo che quel nome è stato registrato in più posti, cosa che fa anche chi costruisce un'identità falsa con cura. Un dominio senza segnalazioni negative non è automaticamente sicuro: significa solo che nessuno lo ha ancora segnalato. L'interpretazione resta un lavoro umano.

Esempio pratico generico: supponi di dover verificare un annuncio di lavoro sospetto che ti è arrivato per email, con un link a un sito "aziendale" che non conosci. Un flusso di questo tipo interrogherebbe la reputazione del dominio del link, controllerebbe se il mittente dell'email corrisponde al dominio ufficiale dichiarato e verificherebbe se il nome dell'azienda citata risulta in registri pubblici accessibili. Nessuno di questi controlli richiede competenze avanzate: richiede sapere quali domande porre e quali strumenti hanno la risposta.

Risultato atteso: non una sentenza definitiva ("è una truffa" o "è autentico"), ma un quadro di segnali concreti, ciascuno con la propria fonte, su cui basare una decisione informata. Se i segnali sono discordanti, la cautela resta la scelta giusta.

Gli errori più comuni

Il primo errore è scambiare la velocità per affidabilità. Un assistente che risponde in tre secondi con un elenco ordinato sembra autorevole, ma la forma non garantisce la sostanza. Chiedere sempre la fonte di ogni singola affermazione è il modo più semplice per smontare questa illusione.

Il secondo errore è fidarsi di un risultato singolo. Un solo strumento che dice "dominio pulito" non basta: la logica della verifica digitale, con o senza AI, è sempre la stessa da anni: confrontare più fonti indipendenti prima di trarre una conclusione. L'MCP rende più veloce interrogare più fonti, non elimina il bisogno di farlo.

Il terzo errore è dare per scontato che collegare uno strumento significhi cedergli il controllo. Chi configura questi collegamenti dovrebbe sempre poter vedere, prima che una chiamata parta, quale strumento sta per essere usato e con quali dati. Se stai lavorando su informazioni sensibili — le tue o quelle di terzi — un collegamento che agisce senza mostrarti cosa sta interrogando è un rischio di riservatezza, non solo di accuratezza.

Perché conta, anche se non fai l'analista OSINT

Per chi lavora nella comunicazione o nel giornalismo, questa evoluzione significa poter dedicare il tempo risparmiato nella raccolta dati alla parte che davvero richiede giudizio umano: capire cosa significano le informazioni, non solo raccoglierle. Per chi gestisce piccole organizzazioni, significa poter fare un primo controllo su un fornitore o un contatto sospetto senza dover imparare mezza dozzina di strumenti specialistici. Per l'utente comune, significa che la stessa tecnologia che rende più facile creare contenuti falsi rende anche più accessibile verificarli — a patto di sapere come chiederlo e cosa controllare nella risposta.

Il rovescio della medaglia esiste ed è onesto dirlo: gli stessi strumenti collegati in questo modo sono a disposizione di chi li usa per scopi ostili, dalla raccolta di informazioni su obiettivi specifici alla preparazione di truffe più credibili. La differenza tra uso difensivo e uso offensivo non sta nello strumento, ma nell'intento e nel contesto di chi lo usa — un motivo in più per restare dentro i confini della verifica pubblica e responsabile, evitando di trasformare una ricerca legittima in un pedinamento digitale di una persona specifica.

In sintesi

Il valore dell'MCP nella verifica digitale non è che l'AI "sa" di più. È che può controllare di più, in tempo reale, mostrando da dove viene ogni informazione. Il criterio pratico da portarsi via è semplice: prima di credere a un'affermazione prodotta da un assistente AI, chiedi sempre quale strumento è stato interrogato per produrla. Se la risposta è vaga, il dato è vago. Se la risposta indica una fonte precisa, hai qualcosa su cui puoi davvero lavorare.

Ho chiesto a un assistente AI di verificare un profilo sospetto. Ecco cosa è cambiato nel 2026

Immagina di ricevere un messaggio da un profilo che dice di lavorare per un'azienda che conosci. Foto professionale, bio credibile, qualche post plausibile. Vuoi capire se è reale prima di rispondere. Fino a poco tempo fa il percorso era manuale: copiare il nome utente, controllarlo su una decina di piattaforme, incollare il dominio del sito citato in bio su un paio di strumenti di analisi, confrontare i risultati a mano. Mezz'ora buona, se sei allenato.

Oggi, con gli strumenti giusti collegati, puoi scrivere la stessa richiesta in una chat con un assistente AI e ottenere in pochi secondi un primo quadro: dove esiste quel nome utente, che reputazione ha il dominio citato, se l'indirizzo email associato è comparso in una violazione di dati nota. Il cambiamento non è che l'intelligenza artificiale abbia imparato a "indagare" da sola. È che adesso può usare gli stessi strumenti che usa un analista, in modo diretto e controllato, grazie a un protocollo chiamato **MCP**.

Vale la pena capire come funziona, perché apre possibilità reali per chiunque faccia verifica delle fonti, ma nasconde anche una trappola precisa in cui è facile cadere.

Cos'è l'MCP e perché conta per la verifica digitale

MCP è la sigla di **Model Context Protocol**, uno standard che permette a un assistente AI di collegarsi direttamente a strumenti esterni: un motore di ricerca IP, un database di domini, un

archivio di violazioni di dati. Prima di questo standard, ogni collegamento tra un modello linguistico e un servizio esterno andava costruito su misura, pezzo per pezzo. Con l'MCP, chi sviluppa un servizio pubblica un "connettore" e qualsiasi assistente compatibile può usarlo, senza reinventare l'integrazione ogni volta.

Per capirlo con un paragone semplice: prima l'AI poteva solo raccontarti cosa sapeva, sulla base di ciò che aveva letto durante l'addestramento. Con l'MCP, invece, può alzare la cornetta e chiamare davvero un servizio esterno, leggere la risposta e usarla. La differenza è enorme. Un modello che "ricorda" un dato su un dominio dal proprio addestramento è, nella migliore delle ipotesi, un'informazione vecchia. Un modello che interroga in tempo reale un servizio di reputazione dei domini attraverso MCP ottiene un dato verificabile, con una fonte precisa dietro.

Questa distinzione è il cuore di tutto il discorso. Non riguarda solo l'OSINT: riguarda chiunque usi un assistente AI per prendere decisioni basate su informazioni che devono essere vere, non plausibili.

Come funziona in pratica: la parte operativa

Ecco cosa succede, passo per passo, quando un assistente collegato tramite MCP gestisce una verifica come quella del profilo sospetto di prima.

1. Ricevi la richiesta e la scomponi in domande verificabili. Non "questo profilo è falso?", ma tre domande concrete: il nome utente esiste altrove con la stessa combinazione di foto e bio? Il dominio citato ha una reputazione pulita o è segnalato come sospetto? L'indirizzo email, se disponibile, è comparso in una violazione di dati nota? Scomporre la domanda serve a evitare che l'assistente risponda con un'impressione generica invece che con dati puntuali.

2. L'assistente richiama gli strumenti collegati, uno per uno. Un connettore verifica la presenza del nome utente su più piattaforme in parallelo. Un altro interroga un servizio di reputazione per il dominio. Un altro ancora controlla se l'email compare in archivi noti di violazioni. Ogni chiamata produce un risultato tracciabile: puoi sempre chiedere "da dove viene questo dato" e ottenere il nome del servizio interrogato, non una frase vaga come "secondo le mie fonti".

3. L'assistente legge i risultati e li mette a confronto, non li inventa. Qui sta il punto che va sempre controllato: il risultato deve provenire da una chiamata reale a uno strumento esterno, non da una risposta generata "a memoria". Se chiedi conferma e l'assistente non riesce a indicarti quale strumento ha usato per un'affermazione specifica, quell'affermazione va trattata come non verificata.

4. Tu, non l'assistente, decidi cosa significano i dati raccolti. Trovare lo stesso nome utente su cinque piattaforme non prova che il profilo sia legittimo: prova solo che quel nome è stato registrato in più posti, cosa che fa anche chi costruisce un'identità falsa con cura. Un dominio senza segnalazioni negative non è automaticamente sicuro: significa solo che nessuno lo ha ancora segnalato. L'interpretazione resta un lavoro umano.

Esempio pratico generico: supponi di dover verificare un annuncio di lavoro sospetto che ti è arrivato per email, con un link a un sito "aziendale" che non conosci. Un flusso di questo tipo interrogherebbe la reputazione del dominio del link, controllerebbe se il mittente dell'email corrisponde al dominio ufficiale dichiarato e verificherebbe se il nome dell'azienda citata risulta in registri pubblici accessibili. Nessuno di questi controlli richiede competenze avanzate: richiede sapere quali domande porre e quali strumenti hanno la risposta.

Risultato atteso: non una sentenza definitiva ("è una truffa" o "è autentico"), ma un quadro di segnali concreti, ciascuno con la propria fonte, su cui basare una decisione informata. Se i segnali sono discordanti, la cautela resta la scelta giusta.

Gli errori più comuni

Il primo errore è scambiare la velocità per affidabilità. Un assistente che risponde in tre secondi con un elenco ordinato sembra autorevole, ma la forma non garantisce la sostanza. Chiedere sempre la fonte di ogni singola affermazione è il modo più semplice per smontare questa illusione.

Il secondo errore è fidarsi di un risultato singolo. Un solo strumento che dice "dominio pulito" non basta: la logica della verifica digitale, con o senza AI, è sempre la stessa da anni: confrontare più fonti indipendenti prima di trarre una conclusione. L'MCP rende più veloce interrogare più fonti, non elimina il bisogno di farlo.

Il terzo errore è dare per scontato che collegare uno strumento significhi cedergli il controllo. Chi configura questi collegamenti dovrebbe sempre poter vedere, prima che una chiamata parta, quale strumento sta per essere usato e con quali dati. Se stai lavorando su informazioni sensibili — le tue o quelle di terzi — un collegamento che agisce senza mostrarti cosa sta interrogando è un rischio di riservatezza, non solo di accuratezza.

Perché conta, anche se non fai l'analista OSINT

Per chi lavora nella comunicazione o nel giornalismo, questa evoluzione significa poter dedicare il tempo risparmiato nella raccolta dati alla parte che davvero richiede giudizio umano: capire cosa significano le informazioni, non solo raccoglierle. Per chi gestisce piccole organizzazioni, significa poter fare un primo controllo su un fornitore o un contatto sospetto senza dover imparare mezza dozzina di strumenti specialistici. Per l'utente comune, significa che la stessa tecnologia che rende più facile creare contenuti falsi rende anche più accessibile verificarli — a patto di sapere come chiederlo e cosa controllare nella risposta.

Il rovescio della medaglia esiste ed è onesto dirlo: gli stessi strumenti collegati in questo modo sono a disposizione di chi li usa per scopi ostili, dalla raccolta di informazioni su obiettivi specifici alla preparazione di truffe più credibili. La differenza tra uso difensivo e uso offensivo non sta nello strumento, ma nell'intento e nel contesto di chi lo usa — un motivo in più per restare dentro i confini della verifica pubblica e responsabile, evitando di trasformare una ricerca legittima in un pedinamento digitale di una persona specifica.

In sintesi

Il valore dell'MCP nella verifica digitale non è che l'AI "sa" di più. È che può controllare di più, in tempo reale, mostrando da dove viene ogni informazione. Il criterio pratico da portarsi via è semplice: prima di credere a un'affermazione prodotta da un assistente AI, chiedi sempre quale strumento è stato interrogato per produrla. Se la risposta è vaga, il dato è vago. Se la risposta indica una fonte precisa, hai qualcosa su cui puoi davvero lavorare.