Model Context Protocol (MCP)

Wie KI-Anwendungen sich mit anderen Diensten verbinden und wie du das sicher nutzt

MCP ist ein offener Standard, der KI-Anwendungen mit Daten, Tools und Abläufen anderer Dienste verbindet. Ohne Standard braucht jede KI-Anwendung für jeden Dienst eine eigene Anbindung: drei Anwendungen, vier Dienste, zwölf Anbindungen. Mit MCP baut jeder Dienst einen Server, und jede Anwendung, die MCP unterstützt, kann ihn nutzen: vier statt zwölf.

WER MIT WEM SPRICHT

- **Host:** die KI-Anwendung, etwa Claude oder ChatGPT.
- **Client:** je verbundenem Server einer, er hält die Verbindung.
- **Server:** stellt Daten und Funktionen eines Dienstes bereit, lokal oder im Internet.

BIS ZU DREI BAUSTEINE JE SERVER

- **Tools:** Aktionen, etwa einen Termin blocken. Das Modell entscheidet, wann.
- **Resources:** Daten als Kontext, etwa Dateien oder Kalender.
- **Prompts:** vorbereitete Anweisungen zum Auswählen.

BEGRIFFE AUSEINANDERHALTEN

Begriff	Was es ist und wie es zu MCP steht
API	Programmierschnittstelle eines Dienstes. MCP legt eine einheitliche Schicht darüber.
Konnektor, Plugin	Name der Verbindung in Claude bzw. ChatGPT. Sagt nichts über die Rechte.
Function Calling	Modell ruft Funktionen auf, je Anbieter anders. MCP nutzt es für alle Anwendungen.
A2A	Agent2Agent: Agenten sprechen untereinander. Ergänzt MCP.

CLAUDE: KONNEKTOREN

- **Weg:** „Anpassungen“ > „Konnektoren“ > „Entdecken“. Häkchen: von Anthropic getestet. Unter dem Namen steht der Betreiber.
- **Eigener Server:** „Hinzufügen“ > „Benutzerdefinierten Konnektor hinzufügen“.
- **Rechte:** „Tool-Berechtigungen“ je Tool.
- **Tarif:** Fertige Web-Konnektoren nur Pro, Max, Team, Enterprise. Desktop-Erweiterungen und ein eigener Konnektor auch gratis.

CHATGPT: PLUGINS

- **Weg:** „Einstellungen“ > „Plugins“.
- **Eigener Server:** „Hinzufügen“ > „Benutzerdefinierten MCP-Server hinzufügen“.
- **Rechte:** „Standardberechtigung“, am vorsichtigsten „Immer fragen“.
- **Tarif:** Voll in Business, Enterprise, Edu. Pro nur mit Lese- und Abrufrechten.

WAS SCHIEFGEHEN KANN

- **Prompt Injection:** fremde Inhalte können das Modell lenken.
- **2025:** Forschende zeigten versteckte Anweisungen in Tool-Beschreibungen (Tool Poisoning).
- **2025, 2026:** Präparierte Inhalte brachten Agenten in Tests dazu, über offizielle Server vertrauliche Daten offenzulegen.
- **2025:** Der bösartige Server „postmark-mcp“ schickte versendete E-Mails heimlich an eine fremde Adresse.

FÜNF REGELN

1. **Vertrauenswürdige Betreiber** wählen. Selbst hinzugefügte Server prüfen weder Anthropic noch OpenAI.
2. **Rechte** bei der Anmeldung prüfen.
3. **Schreibaktionen** bestätigen lassen.
4. **Neue Konnektoren** erst ohne Aktion testen.
5. **Ungenutzte Verbindungen** trennen.

ZWEI PROMPTS FÜR DEN ERSTEN TEST

WAS KANN ER?
„Liste alle Tools auf, die dir über den Konnektor NAME zur Verfügung stehen. Schreib zu jedem Tool, ob es nur liest oder etwas ändert, löscht oder verschickt. Führe dabei kein Tool aus.“

NUR LESEN
„Nutze den Konnektor NAME nur lesend. Zeig mir die letzten drei Einträge zu THEMA und ändere nichts. Frag nach, bevor du irgendeine Aktion ausführst.“

MERKSÄTZE

Ein Server je Dienst statt einer Anbindung je KI-Anwendung.

Ein Häkchen im Verzeichnis ist kein Sicherheitsaudit.