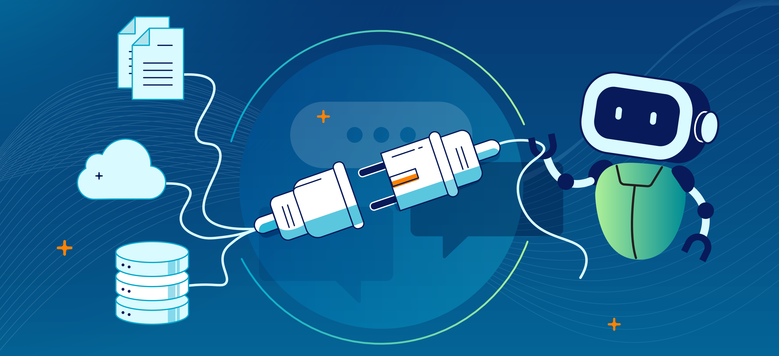
What are the security implications of MCP (Model Context Protocol)?

MCP (Model Context Protocol) is an open standard for data sharing with, and between, AI applications, as well as a point of integration for AI agents. You can think of it as the dominant "language" of AI. MCP enables organizational data to be used for productive AI workflows—with minimal friction or manual integrations.



As with any mechanism that enables data access, MCP requires careful consideration for data security.

Where is MCP being used?

While MCP isn't the only protocol or language for connecting AI tools, it is rapidly emerging as the dominant standard. It's already being adopted by major players across the AI and enterprise software ecosystem.

Salesforce	Microsoft	GitHub
OpenAI	Slack	PayPal
Anthropic	Servicenow	MongoDB
Google	Atlassian	Asana

What should security professionals be looking to do in order to mitigate the risks of data exfiltration via MCP?

- Identify MCP servers and clients in use, with **visibility** into attributes including name, ID, URL, version, host, data source, and protocol
- Assign **risk scores** to MCP servers in the same manner as for cloud applications
- **Detect and monitor traffic** in MCP servers
- **Log MCP events**, including sessions, initializations, tool requests and responses, and deployments
- Identify sensitive data in use with MCP tools
- **Manage access** using granular, context-based policy controls (including a default block option for MCP traffic) and real-time prevention of data leaks

Netskope One provides the ability to detect and monitor MCP traffic, inventory and evaluate the risk of MCP servers, and provide real-time prevention of data leakage to enable your organization to accelerate AI adoption securely.

Read more about the new data security challenges that AI is bringing to your organization, and discover how Netskope can help mitigate them in our AI Security Playbook .

Get the Playbook