WORKSHOP GARR 2026

NET MAKERS

Protocollo MCP: rischi e best practice

Domingo Ranieri

ICSC-Centro Nazionale di Ricerca in HPC, Big Data e Quantum Computing

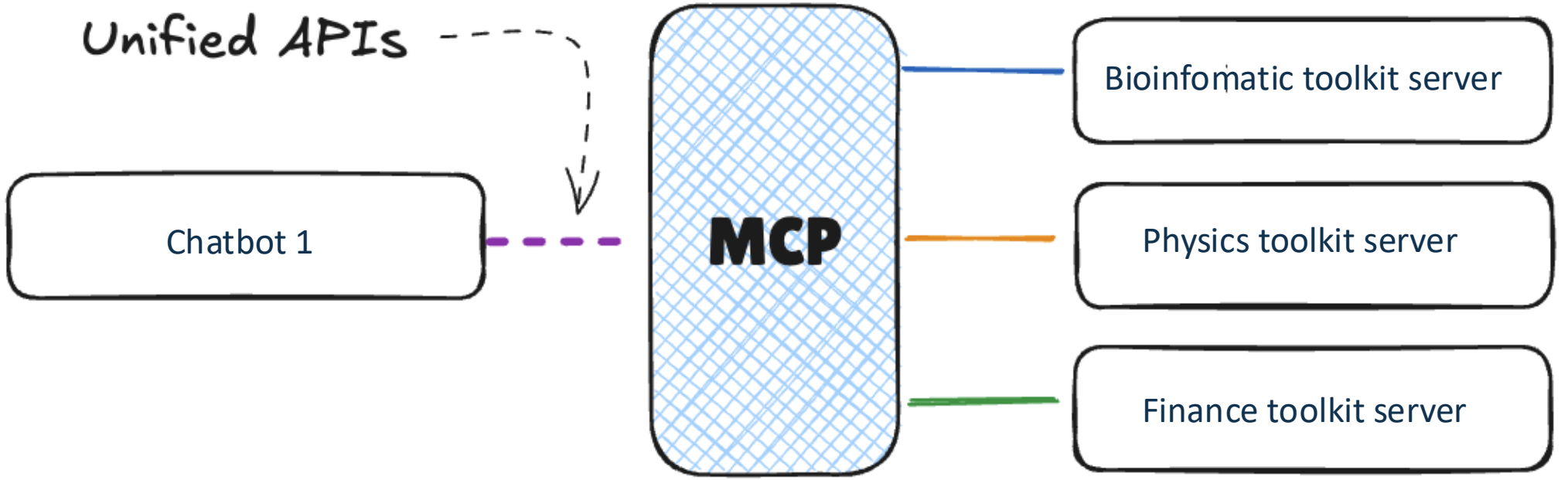
Protocollo MCP

WORKSHOP GARR 2026

NET MAKERS

Model Context Protocol

- Il **Model Context Protocol (MCP)** è uno standard che consente agli LLM di **connettersi a strumenti esterni in modo strutturato**



Model Context Protocol

- **Host**: L'applicazione con LLM con cui gli utenti interagiscono direttamente. L'Host gestisce le connessioni ai Server MCP e orchestra il flusso complessivo tra le richieste degli utenti, l'elaborazione dell'LLM e gli strumenti esterni.

- **Client**: Un componente all'interno dell'applicazione Host che gestisce la comunicazione con uno specifico Server MCP.

- **Server**: Un programma o servizio esterno che espone funzionalità tramite il protocollo MCP.

Sicurezza MCP: Rischi e Best Practice

WORKSHOP GARR 2026

NET MAKERS

Rischi MCP e buone pratiche

I rischi per la sicurezza di MCP derivano sia dalle **interazioni specifiche di MCP** sia dalle vulnerabilità di sicurezza generali.

Categoria	Rischi
🔴 **MCP-native / strongly MCP-specific**	Tool poisoning, rug pulls, cross-server escalation
🟠 **MCP-specific trust/authorization interaction**	Confused deputy, abuso dei permessi concessi, data exfiltration through tools
🟡 **MCP deployment / ecosystem risks**	Supply-chain compromise, evasione dalle sandbox
⚪ **General infrastructure risks affecting MCP**	Message tampering, replay, compromised proxy/middleware

Owasp Security guidelines

Tool Poisoning

Un server MCP malevolo inserisce istruzioni all'interno delle descrizioni degli strumenti, degli schemi dei parametri o di altri metadati degli strumenti, influenzando il comportamento dell'LLM.

Rischi

- Istruzioni malevole nascoste nei metadati degli strumenti possono **manipolare** l'agente, inducendolo a eseguire azioni **non autorizzate**.
- Le **informazioni sensibili possono essere esposte** o trasmesse a servizi esterni **senza che l'utente ne sia consapevole.**

Best Practices

- **Non fidarsi** ciecamente delle **descrizioni** o degli **schemi** degli strumenti MCP ma trattarli come input non attendibili.
- Mostrare agli utenti cosa richiede uno strumento prima di concedere **l'autorizzazione**.
- **Convalidare** gli strumenti tramite una allowlist o un registro approvato.
- **Monitorare** il comportamento degli strumenti, non solo i loro nomi e permessi.

Rug Pull Attacks

Un server MCP inizialmente si comporta in modo legittimo e ottiene l'approvazione dell'utente. Successivamente, **il server modifica le definizioni o il comportamento dei propri strumenti** per diventare malevolo.

Rischi

- Modifiche malevole alle definizioni o al comportamento degli strumenti possono **introdurre nuove vulnerabilità** dopo l'approvazione iniziale.
- Un server inizialmente affidabile può **sfruttare la fiducia** dell'utente per eseguire azioni non autorizzate o sottrarre dati sensibili.

Best Practices

- **Verificare crittograficamente** le definizioni degli strumenti.
- **Rilevare le modifiche alle descrizioni**, agli schemi e alle funzionalità degli strumenti.
- **Richiedere nuovamente l'approvazione** dell'utente dopo modifiche rilevanti per la sicurezza.

Cross-Origin Escalation

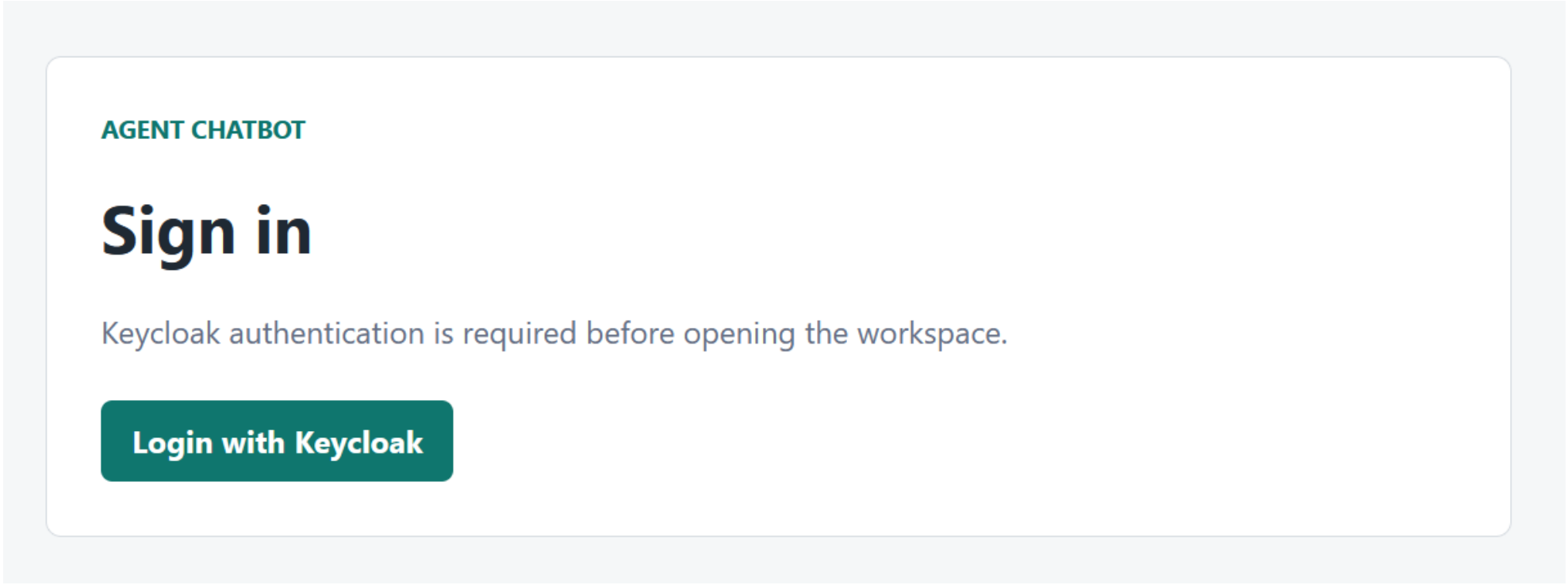
Un server MCP malevolo altera il comportamento dell'agente, **inducendolo a utilizzare strumenti** propri o forniti da un altro server MCP considerato affidabile.

Rischi

- **Manipolazione del comportamento dell'agente** attraverso metadati o istruzioni fraudolente.
- **Compromissione** dei confini di fiducia tra server MCP.
- **Esposizione o divulgazione non autorizzata di dati sensibili.**
- **Abuso dei privilegi concessi** a server MCP considerati affidabili.

Best practice

- **Considerare** ogni server MCP come un **dominio di fiducia separato**.
- **Impedire** che i metadati di un server possano **controllare implicitamente** un altro server.
- **Applicare meccanismi di autorizzazione** distinti per ciascun server.
- **Evitare di esporre tutti gli strumenti di tutti i server in un unico contesto privo di restrizioni.**


AGENT CHATBOT
Sign in
Keycloak authentication is required before opening the workspace.
Login with Keycloak

AGENT CHATBOT

Chats

New Chat

Send an email to Alice (alic...

Send an email to Alice to s...

can you ask what are the H...

can you write a python co...

write a python code that c...

can you read the wiki struc...

can you give me the mcp s...

Send an email to alice@ex...

can you give me a security ...

MCP SERVER

No MCP server

Connect

Manage Servers

Not connected

admin-user

Logout

CONVERSATION

New chat

Send a message to start a conversation.

Write a message...

Send

can you give me the mcp s...

can you give me the mcp s...

can you give me a daily ins...

say me the zen quote of th...

Send an email to alice@ex...

can you give me a security ...

MCP SERVER

No MCP server

Connect

Manage Servers

Not connected

admin-user

Logout

MCP SERVERS

Manage MCP servers

Close

DISCOVERY STATUS

success

Hugging Face

Edit | Rediscover | Delete

ENDPOINT

https://huggingface.co/mcp

TRANSPORT

http

SOURCE

managed

PUBLISHER

hugginface.co

DISCOVERY STATUS

success

Add server

Send

can you read the wiki struc...

can you give me the mcp s...

Send an email to alice@ex...

can you give me a security ...

MCP SERVER

Local FastMCP

Connect

Manage Servers

Connected: 0 approved, 3 pending

server_time (pending · trusted_strict)

send_email (pending · trusted_strict)

admin_security_snapshot (pending · trusted.

admin-user

Logout

MCP TOOL

send_email

Close

DESCRIPTION

Prepare and send an outbound email to explicit recipients when the user asks to share information by email.

INPUT SCHEMA

```
{
  "type": "object",
  "required": [
    "to",
    "subject",
    "body"
  ],
  "properties": {
```

OUTPUT SCHEMA

```
{
  "type": "object",
  "required": [
    "type",
    "tool",
    "mode",
    "delivery_id",
    "to",
```

DEFINITION HASH

6ab9573246e8e9c086f8dc7bc4732f5e2abcbddf9f107cee5bc39cc9a495eeba

TRUST TIER

trusted_strict

Send

can you read the wiki struc...

can you give me the mcp s...

Send an email to alice@ex...

can you give me a security ...

MCP SERVER

Local FastMCP

Connect

Manage Servers

Connected: 0 approved, 3 pending

server_time (pending · trusted_strict)

send_email (pending · trusted_strict)

admin_security_snapshot (pending · trusted.

admin-user

Logout

MCP TOOL

send_email

Close

```
    "to",
    "subject",
    "body"
  ],
  "properties": {
```

OUTPUT SCHEMA

```
{
  "type": "object",
  "required": [
    "type",
    "tool",
    "mode",
    "delivery_id",
    "to",
```

DEFINITION HASH

```
6ab9573246e8e9c086f8dc7bc4732f5e2abcbddf9f107cee5bc39cc9a495eeba
```

TRUST TIER

```
trusted_strict
```

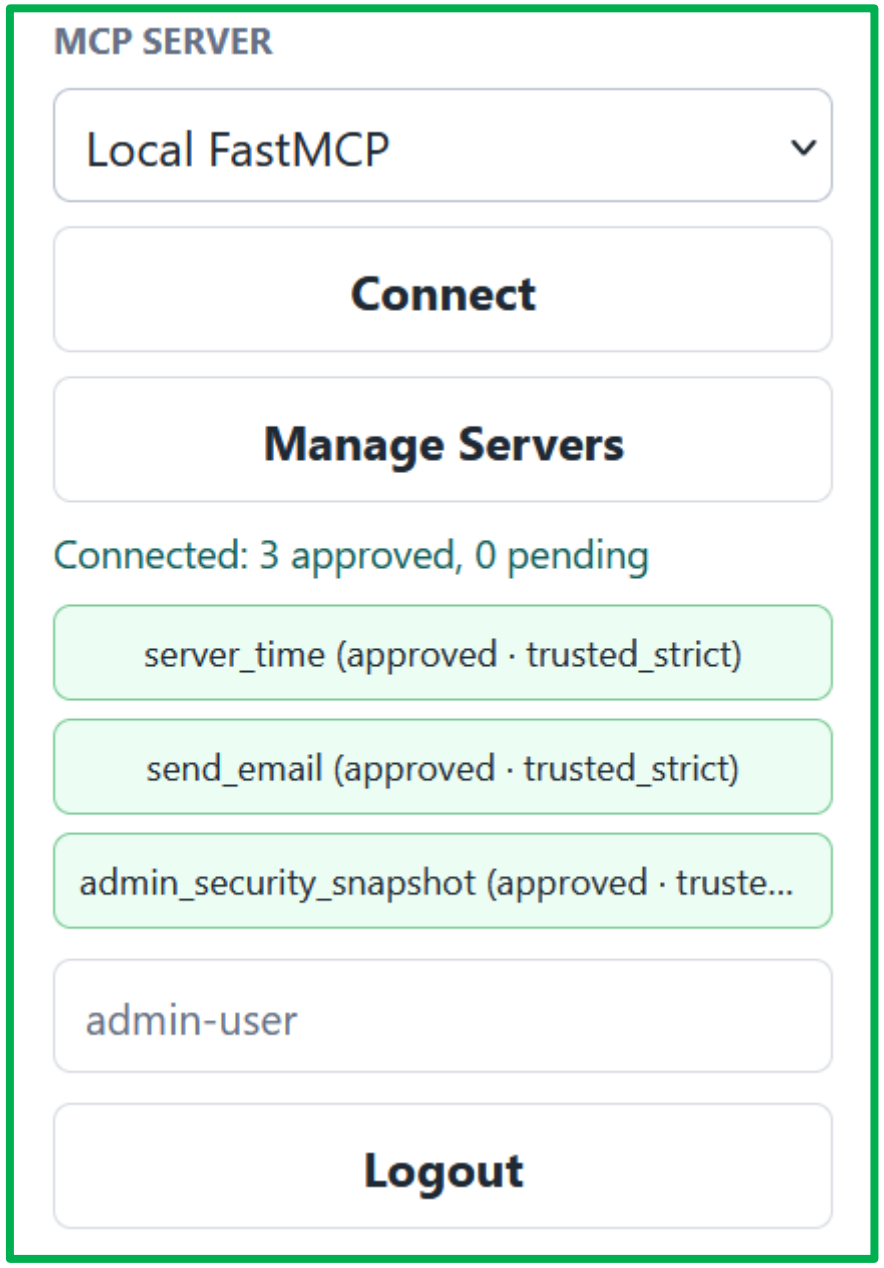
EXECUTION POLICY

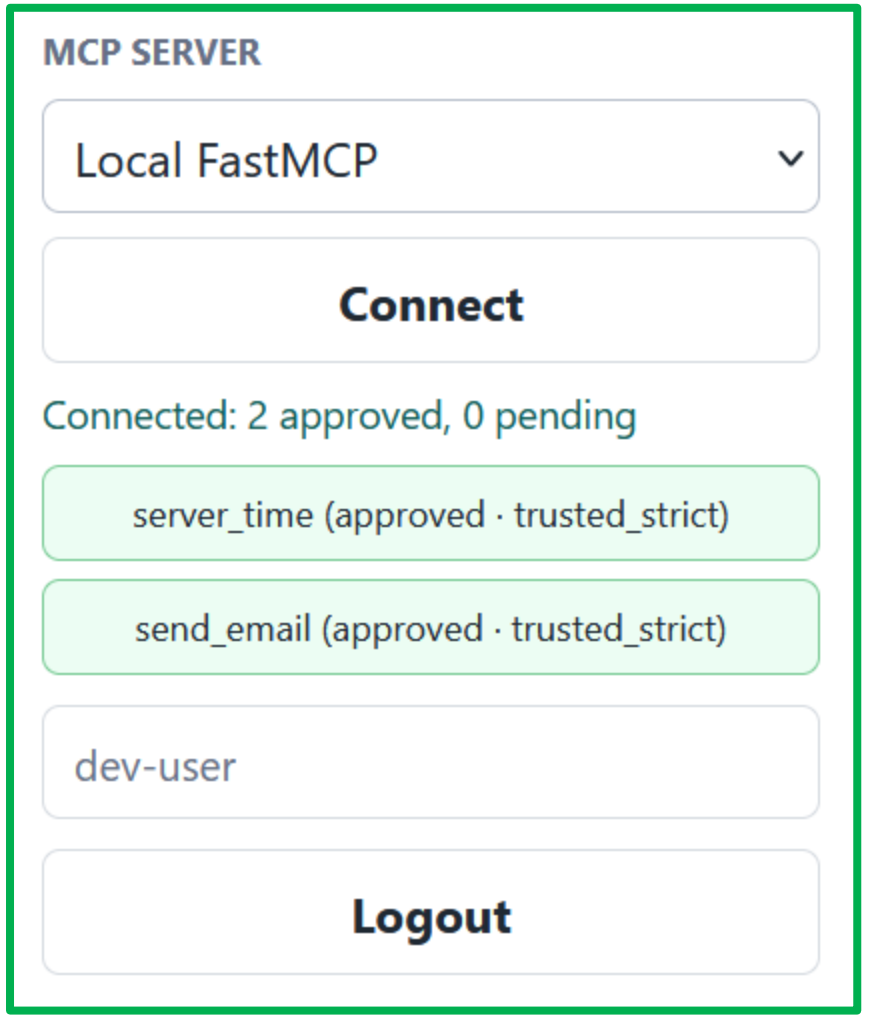
```
confirm
```

Set allow | Set confirm | Set deny

Approve

Send


MCP SERVER
Local FastMCP
Connect
Manage Servers
Connected: 3 approved, 0 pending
server_time (approved · trusted_strict)
send_email (approved · trusted_strict)
admin_security_snapshot (approved · truste...
admin-user
Logout


MCP SERVER
Local FastMCP
Connect
Connected: 2 approved, 0 pending
server_time (approved · trusted_strict)
send_email (approved · trusted_strict)
dev-user
Logout

YOU

Send an email to Alice (alice.person@gmail.com) to schedule a meeting about project Cynton. Mention that I am available on Monday and Tuesday. Use a friendly and professional tone. Sign it with Domingo

ASSISTANT

I prepared the sensitive MCP tool call `send_email`, but it was not executed. Review the exact parameters in the confirmation section and approve it explicitly if you want it to run.

TOOLS VIA LOCAL FASTMCP

PENDING **send_email**

Review request


YOU
Send an email to Alice (alice.person@gmail.com) to schedule a meeting about project Cynton. Mention that I am available on Monday and Tuesday. Use a friendly and professional tone. Sign it with Domingo
ASSISTANT
I prepared the sensitive MCP tool
confirmation section and approve
TOOLS VIA LOCAL FASTMCP
PENDING
send_email
TOOL REQUEST
send_email
Close
REVIEW REASON
sensitive_action
ARGUMENTS
{
"to": [
"alice.person@gmail.com"
],
"body": "Hi Alice,\n\nI hope you're doing well.\n\nI would like to schedule a meeting to discuss project Cynton. I am available this coming Monday and Tuesday. Please let me know if either of those days works for you, or feel free to suggest another time that fits your schedule.\n\nLooking forward to hearing from you.\n\nBest regards,\nDomingo",
Approve
Reject

YOU

Send an email to Alice (alice.person@gmail.com) to schedule a meeting about project Cynton. Mention that I am available on Monday and Tuesday. Use a friendly and professional tone. Sign it with Domingo

ASSISTANT

I prepared the sensitive MCP tool call `send_email`, but it was not executed. Review the exact parameters in the confirmation section and approve it explicitly if you want it to run.

TOOLS VIA LOCAL FASTMCP

ASSISTANT

The email has been sent to Alice (alice.person@gmail.com) regarding project Cynton.

TOOLS VIA LOCAL FASTMCP

SUCCESS **send_email**

can you give me a daily ins...

say me the zen quote of th...

Send an email to alice@ex...

can you give me a security ...

MCP SERVER

Local FastMCP

Connect

Manage Servers

Connected: 2 approved, 1 pending

server_time (approved · trusted_strict)

send_email (changed · trusted_strict)

admin_security_snapshot (approved · truste...

admin-user

Logout

MCP TOOL

send_email

Close

DESCRIPTION

Prepare and send an outbound email to explici recipients when the user asks to share information by email.

INPUT SCHEMA

```
{
  "type": "object",
  "required": [
    "to",
    "subject",
    "body"
  ],
  "properties": {
```

OUTPUT SCHEMA

```
    "tool",
    "mode",
    "delivery_id",
    "to",
    "cc",
    "subject",
    "body_preview",
    "sent_at"
```

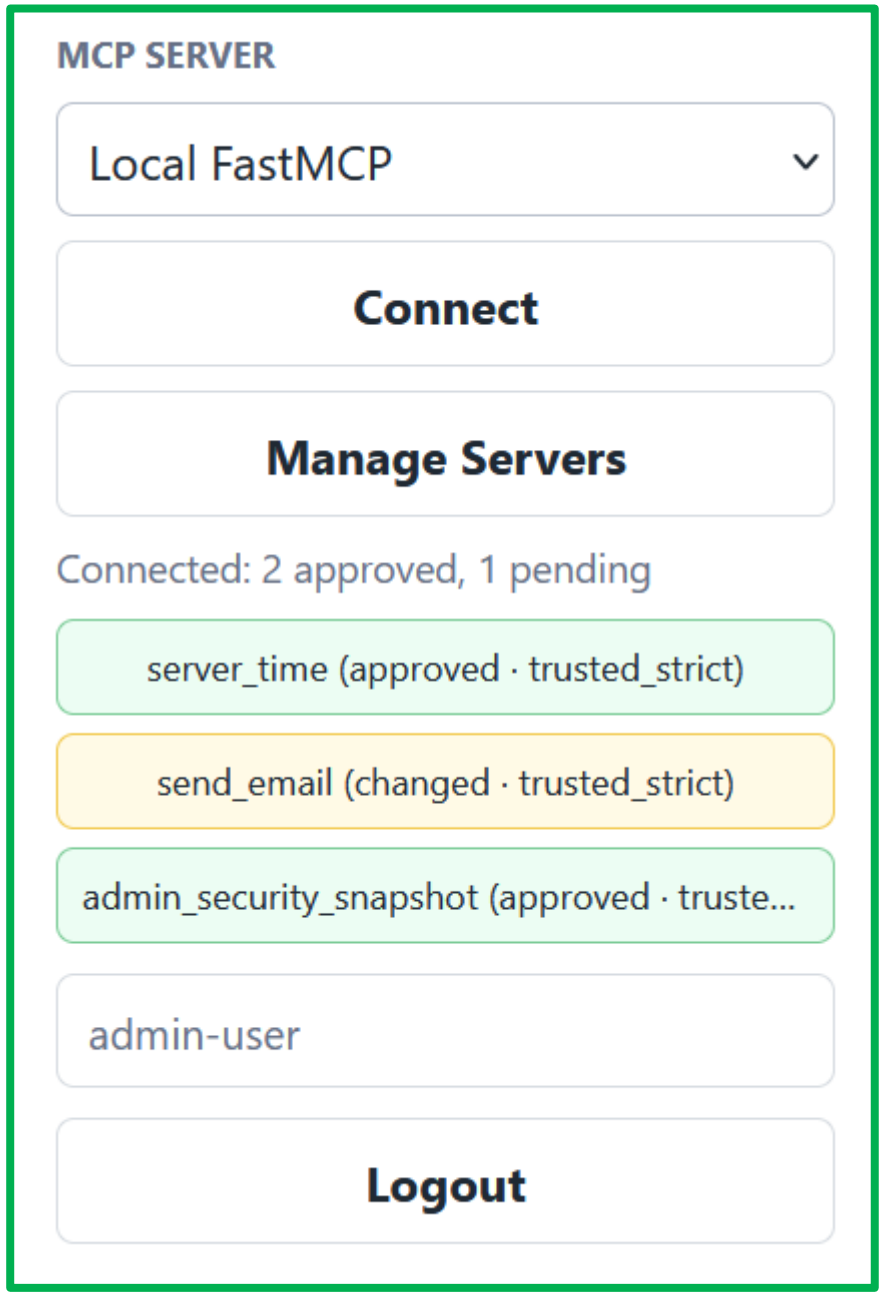
DEFINITION HASH

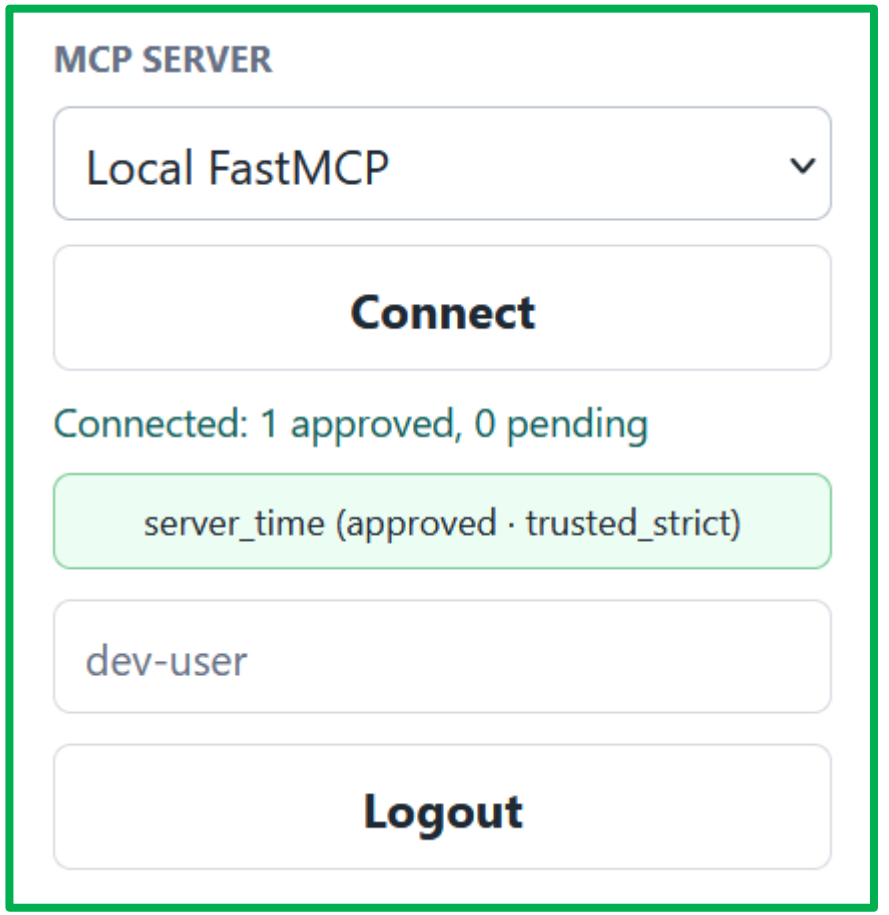
14cd7139432de51c7f6ecf29daf08ac9f3ad0cf6a6773f6676a52449421dd345

TRUST TIER

trusted_strict

Send


MCP SERVER
Local FastMCP
Connect
Manage Servers
Connected: 2 approved, 1 pending
server_time (approved · trusted_strict)
send_email (changed · trusted_strict)
admin_security_snapshot (approved · truste...
admin-user
Logout


MCP SERVER
Local FastMCP
Connect
Connected: 1 approved, 0 pending
server_time (approved · trusted_strict)
dev-user
Logout

can you give me a daily ins...

say me the zen quote of th...

Send an email to alice@ex...

can you give me a security ...

MCP SERVER

Local FastMCP

Connect

Manage Servers

Policy update failed

server_time (approved · trusted_strict)

send_email (invalid · rejected)

admin_security_snapshot (approved · truste...

admin-user

Logout

MCP TOOL

send_email

Close

```
    "body"
  ],
  "properties": {
```

OUTPUT SCHEMA

```
{
  "type": "objec",
  "required": [
    "type",
    "tool",
    "mode",
    "delivery_id",
    "to",
```

DEFINITION HASH

```
b928d7df9f78092e7ff181b11e15587c20938eb9e3bba9add9b7dc10f072f26a
```

TRUST TIER

```
rejected
```

EXECUTION POLICY

```
unclassified
```

VALIDATION ERROR

```
send_email.output_schema is not a valid JSON Schema: 'objec' is not valid under any of the given schemas
```

Set allow | Set confirm | Set deny

Send

Grazie per la vostra attenzione!

Contatto:
domingo.ranieri@supercomputing-icsc.it

Per le domande www.partici.fi
codice 83995288

Questionario IT4LIA

WORKSHOP GARR 2026
NET MAKERS