



# Evals in MCP development



**Jakub Kopecký**

Engineer @ Apify

# Marketplace of thousands of tools

## Apify Store

Search for Actors

- Social media
- AI
- Agents
- Lead generation
- E-commerce
- SEO tools
- Jobs
- MCP servers
- News
- Real estate
- Developer tools
- Travel
- Videos
- Automation
- Integrations
- Open source
- Other

All Actors

View all →



### Google Maps Scraper

compass/crawler-google-pl...

Extract data from thousands of Google Maps locations and businesses, including reviews,...

Comp... 263K 4.7 (834)



### Website Content Crawler

apify/website-content-cra...

Crawl websites and extract text content to feed AI models, LLM applications, vector databases, or...

Apify 99K 4.8 (162)



### TikTok Scraper

clockworks/tiktok-scraper

Extract data from TikTok videos, hashtags, and users. Use URLs or search queries to scrape TikTok...

Clock... 123K 4.6 (195)



### E-commerce Scraping Tool

apify/e-commerce-scraping...

Scrape data from e-commerce websites with E-commerce Scraping Tool. Scrape almost any...

Apify 3.9K 4.6 (27)



### Facebook Posts Scraper

apify/facebook-posts-scr...

Extract posts, videos, and engagement metrics from Facebook pages. Get text caption...

Apify 46K 4.8 (116)



### Tweet Scraper V2 - X /...

apidojo/tweet-scraper

Lightning-fast search, URL, list, and profile scraping, with customizable filters. At \$0.40 per...

API Dojo 35K 4.4 (138)



### Instagram Scraper

apify/instagram-scraper

Scrape and download Instagram posts, profiles, places, hashtags, photos, and comments. Get data...

Apify 174K 4.7 (273)



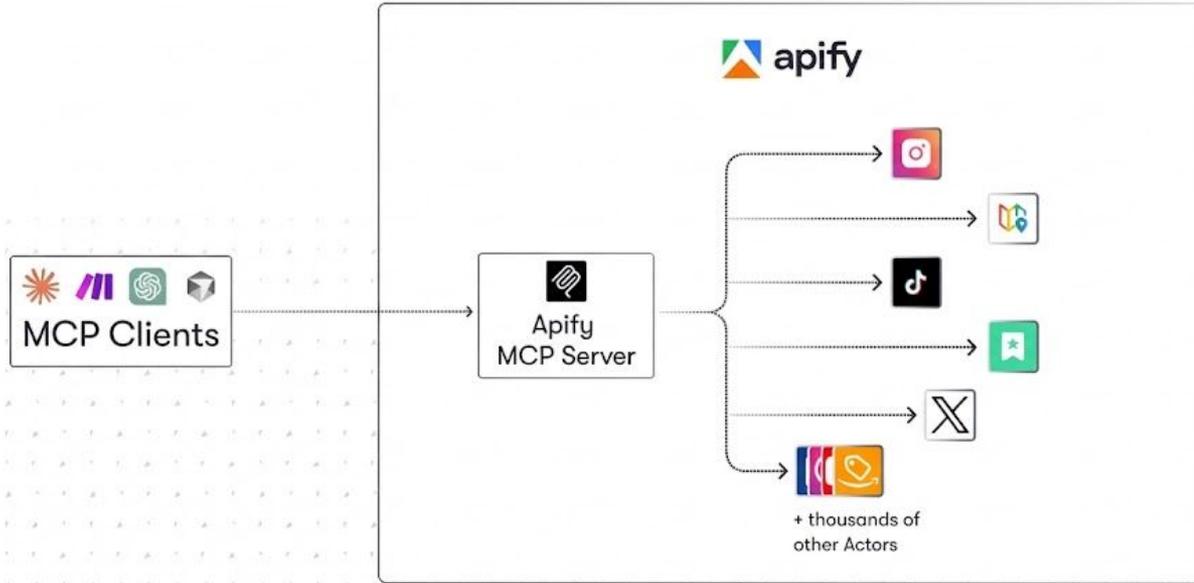
### YouTube Scraper

streamers/youtube-scraper

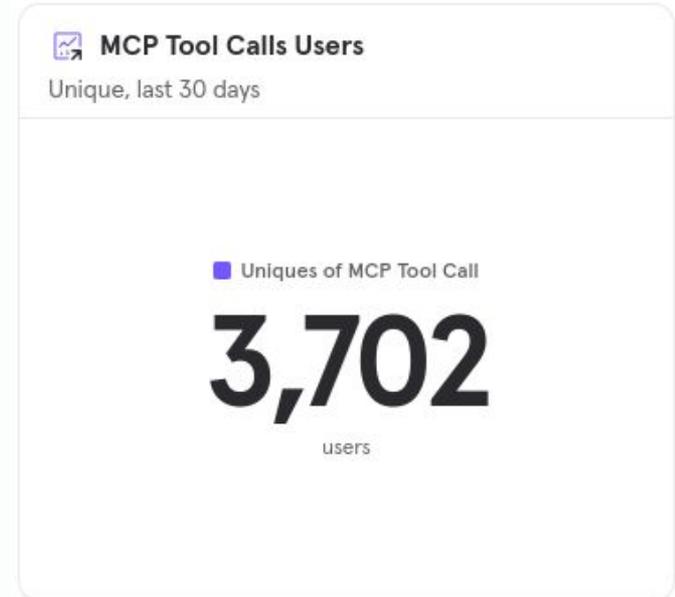
YouTube crawler and video scraper. Alternative YouTube API with no limits or quotas. Extract and...

Streamers 51K 4.7 (114)

# mcp.apify.com is born



# mcp.apify.com in numbers



# Discovery: too many tools



- △ **Search** Actors and **add-actor** tool approach
- △ Dynamic MCP tools **pioneers**
- △ LLMs are prone to **manipulation**
- △ But we encountered some **issues**

# Discover client capabilities



- △ Most clients don't actually **support** that
- △ Created **call-actor** helper tool instead
- △ But this doesn't actually solve the **problem**

```
● ● ●  
  
# Node  
npm install mcp-client-capabilities  
# Python  
pip install mcp-client-capabilities
```

apify/**mcp-client-capabilities**

Index of all Model Context Protocol (MCP) clients and their capabilities

5 Contributors   4 Issues   70 Stars   5 Forks

# MCP-CLI: just build your own

- △ **Launched MCP-CLI**
- △ **Supports latest features**
- △ **Embeddable**



Jan Curn    
@jancurn



Dear @MCP\_Community: after 3 weeks of late nights, 300 commits, and \$300 spent on Claude Code, I'm excited to present:

`mcpc` — a universal CLI client for MCP

It supports OAuth, sessions, `--json` for code mode, `--proxy` for sandboxes, and more

[github.com/apify/mcpc](https://github.com/apify/mcpc)

```
mcpc
MCP sessions:
@fs → npx -y @modelcontextprotocol/server-filesystem /Users/jancurn/Projects/mcpc (stdio) ● live
@ai-proxy → https://mcp.apify.com (HTTP, OAuth: default, MCP: 2025-06-18) [proxy: 127.0.0.1:8080] ● live
@anonymous → http://127.0.0.1:8080 (HTTP, MCP: 2025-11-25) ● live

Available OAuth profiles:
mcp.apify.com / default, refreshed 47m ago

Run "mcpc --help" for usage information.

mcpc @fs tools-list
[@fs → npx -y @modelcontextprotocol/server-filesystem /Users/jancurn/Projects/mcpc (stdio)]

Available tools (14):
Content type: text/plain
```

## apify/mcp-cli



mcpc is a universal CLI client for MCP. Supports persistent sessions, stdio/HTTP, OAuth 2.1, JSON output for scripting and code...

1 Contributor   1 Used by   252 Stars   10 Forks



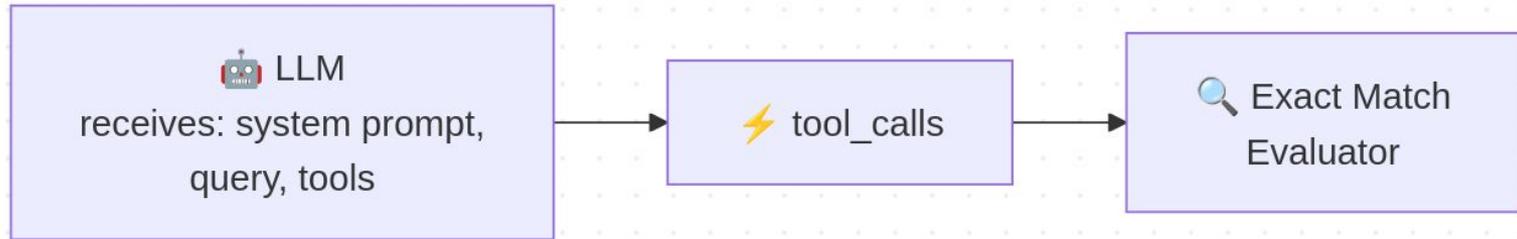
```
npm install -g @apify/mcpc
```



# Evals: trust, but verify



- △ Initial **deterministic** approach - next tool selection
- △ Great for **tweaking** tool name and descriptions
  - Agents are bad at this
- △ Newly released LLMs **break** things
- △ **Server instructions** - hidden gem
- △ **Not enough** for actual workflows and chaining



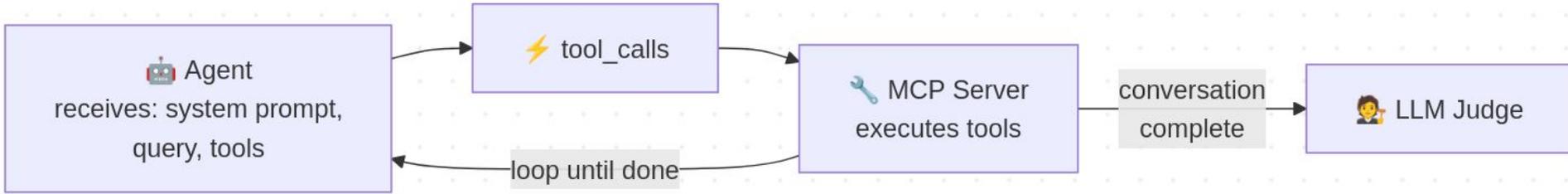


```
[  
  {  
    "id": "fetch-actor-details-4",  
    "category": "fetch-actor-details",  
    "query": "What can apify/instagram-scraper do?",  
    "expectedTools": ["fetch-actor-details"]  
  },  
  {  
    "id": "call-actor-3",  
    "category": "call-actor",  
    "query": "Call apify/google-search-scraper to find restaurants in London",  
    "expectedTools": ["call-actor"]  
  }  
]
```

# Evals: workflows



- △ **Agentic** workflows approach
- △ Actual **MCP tool calls**
- △ Execution is evaluated by **LLM judge**
- △ Works on paper, **issues in prod** (OpenAI)



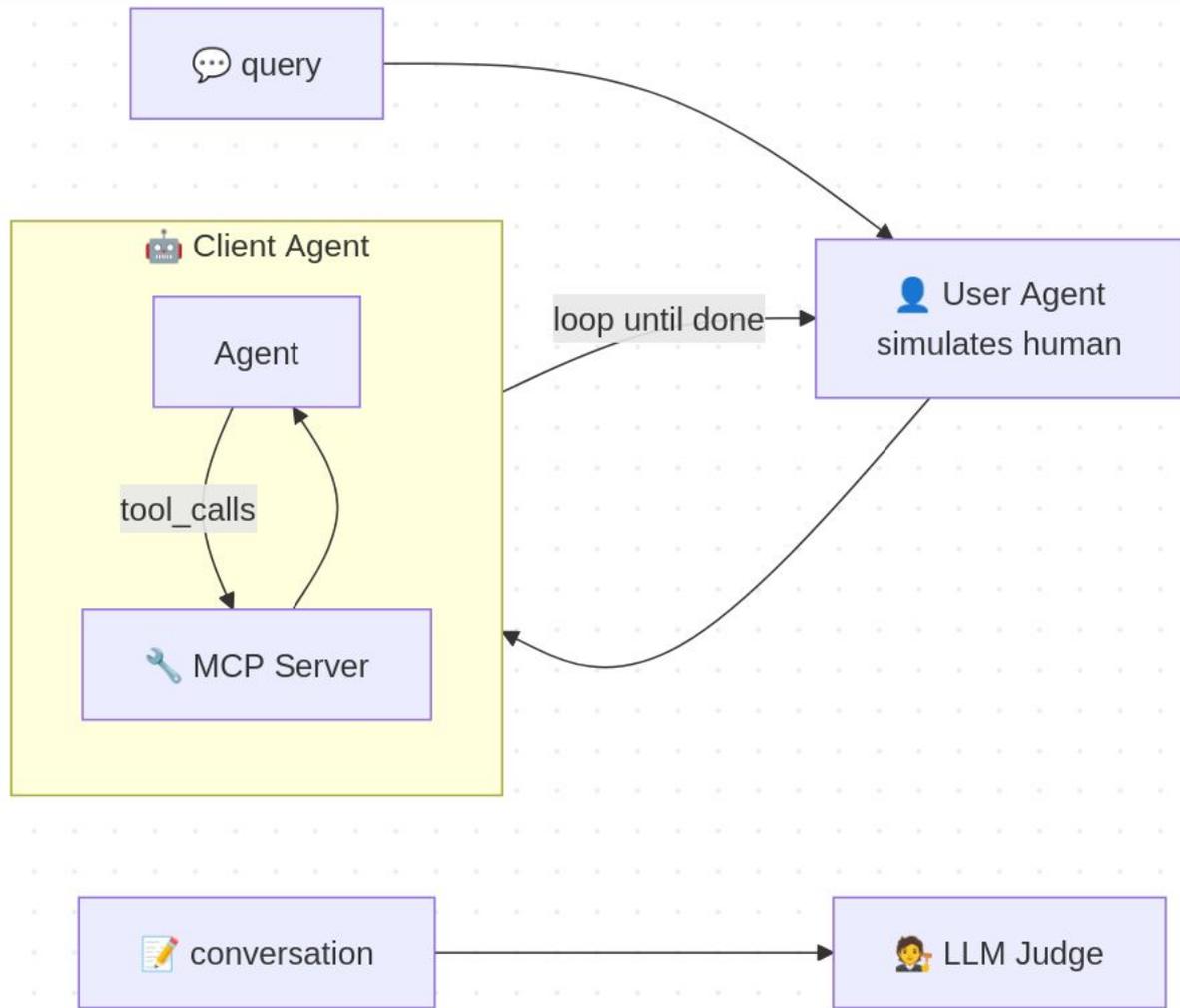


```
[
  {
    "id": "call-amazon-product-data",
    "category": "call",
    "query": "Get pricing and details for iPhone 15 on Amazon",
    "reference": "The agent must search for Amazon product scraper, call it with iPhone 15 search query, and return structured product data including title, price, and rating."
  },
  {
    "id": "instagram-search-hashtag-posts",
    "category": "call",
    "query": "Find me 10 recent Instagram posts with the hashtag #travel and present the results with their URLs.",
    "reference": "The agent must search Instagram for recent posts with the hashtag #travel and return at least 10 post URLs."
  }
]
```

# Evals: the chat



- △ **Three LLMs** approach - Client, User, Judge
- △ Simulates more natural human like interaction
- △ Does not handle real clients **bugs** and **missing features**



# Evals results



## Summary:

Total tests: 51

Passed: 48 

Failed: 3 

Errors: 0 



## Results (74 cases):

openai/gpt-4o-mini: 94.6%

anthropic/claude-haiku-4.5: 91.9%

google/gemini-2.5-flash: 91.2%

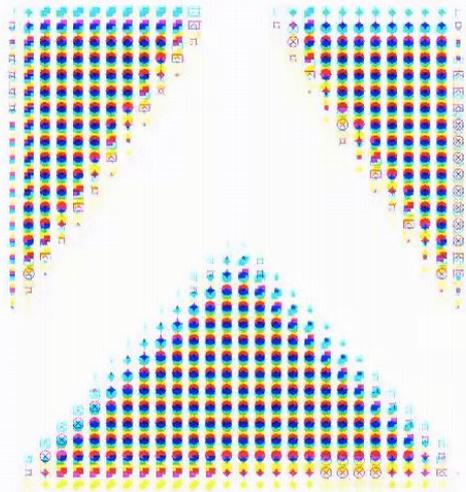
openai/gpt-5: 88.5%

# Wrap up



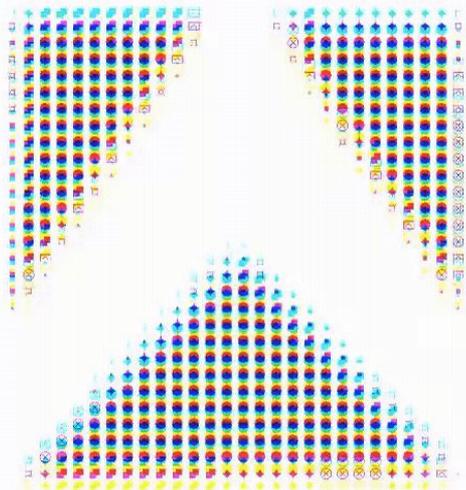
- △ MCP is still **early** - mostly builders
- △ **Evals are important** for production grade MCP
- △ **Never** let agents tweak the tool descriptions

# Questions?





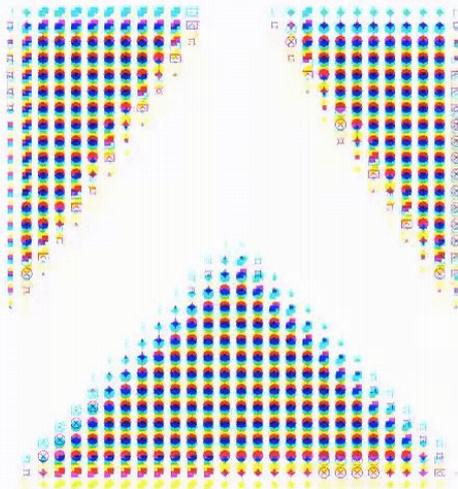
EVALS\_APIFY





Jakub Kopecký

Software engineer @ Apify | Building MCP  
and stuff



**BTW we are hiring!**