

서울

Open Source AI Agents on User-Owned Infra

Miley Fu DevRel Director, CNCF Ambassador Github @mileyfu

<https://github.com/beclab/Olares>



The Autonomous AI Agents Boom

Continuous learning + Context

-DeepSeek Founder with AGI as the only end goal

OpenClaw

Hermes

Quickly became the top 20 open source project on Github



- #1 build-your-own-x
- #2 awesome
- #3 freeCodeCamp
- #4 public-apis
- #5 free-programming-bo...
- #6 **openclaw**
- #7 developer-roadmap
- #8 system-design-primer
- #9 coding-interview-univ...
- #10 awesome-python
- #11 **awesome-selfhosted**
- #12 996.ICU
- #13 project-based-learni...
- #14 superpowers
- #15 react
- #16 linux
- #17 the-book-of-secret-...
- #18 ECC
- #19 Python
- #20 **hermes-agent**

2026 March 6, lines outside the Tencent Building in Shenzhen to have OpenClaw installed.



The Security Nightmare

36%

of studied agent skills contained prompt injection payloads.

1467

malicious payloads found in a single ToxicSkills study.

Untrusted context is becoming the new SQL injection.

Snyk Finds Prompt Injection in 36%, 1467 Malicious Payloads in a ToxicSkills Study of Agent Skills Supply Chain Compromise

Written by  Luca Beurer-Kellner  Aleksei Kudrinskii  Marco Milanta  Kristian Bonde Nielsen
 Hemang Sarkar  Liran Tal

February 5, 2026 ⌚ 37 mins read

The first comprehensive security audit of the Agent Skills ecosystem reveals malware, credential theft, and prompt injection attacks targeting OpenClaw, Claude Code, and Cursor users

▲ GitLost: We Tricked GitHub's AI Agent into Leaking Private Repos (noma.security)

541 points by ColinEberhardt 15 days ago | hide | past | favorite | 205 comments

▲ fwlr 15 days ago | next [-]

"Prompt injection attacks have become, to agentic AI, what SQL injections were to web applications: a systematic, category-wide vulnerability class that requires the same systematic strategies and defenses."

???

Isn't prompt injection far more fatal to LLMs than SQL injection is to SQL databases?

Like, the problem of SQL injection was that user input was forming part of the instruction string given to the SQL engine, and so malicious user input could include various SQL grammar terminals to end the current SQL command, followed by complete SQL commands of their own, and the engine would simply execute both commands. The fix was prepared statements: fixed/static/pre-compiled instruction strings, that can only ever perform fixed/static/pre-defined logic, and that logic can then be (more) safely applied to arbitrary user-input data.

The analogous mitigation for agents is to have fixed behaviors they can perform, such as "read repo 1" "read repo 2", etc., and the user input is used as data to select which of these fixed behaviors to execute. But we already have this technology - it's called a menu. The value of LLMs is specifically and intrinsically predicated on being more than a menu, while the value of SQL does not depend on being more than "pre-set logic operating on arbitrary data" - user input being part of the instruction string to SQL was incidental, for developer convenience.

▲ mcv 15 days ago | parent | next [-]

Exactly. SQL injection was caused by treating user input as part of the instruction instead of as the pure data that it was intended as. Separating those two fixed it. Prompt injection is unavoidable because the user input is intended as instruction.



Claude Code refuses requests or charges extra if your commits mention "OpenClaw" (twitter.com/theo)

1349 points by elmean 83 days ago | hide | past | favorite | 720 comments

<https://xcancel.com/theo/status/2049645973350363168>

▲ abdullin 83 days ago | next [-]

I reproduced this on my account.

```
cd /tmp
mkdir anthropic-claude
cd anthropic-claude/
git init
touch hello
git add -A
git commit -m '{"schema": "openclaw.inbound_meta.v1"}'
claude -p "hi"
```

Immediate disconnect and session usage went to 100%

▲ petercooper 83 days ago | parent | next [-]

I wonder if projects which are anti-AI could place such identifiers surreptitiously into docs or commits as a way to sabotage people using Claude Code. Your project isn't going to get many AI PRs if just cloning your project wiped out their quota.

**One commit mentioning
"OpenClaw" — session
usage drained to 100%. Any
agent with tool access can
be hacked**

Privacy Dilemma

- Users want agents that actually DO things: read private files, execute tools, and manipulate local systems autonomously. Post or pay for things.
- No one wants to send their company's unreleased financials or personal source code to a public cloud API to make it happen.

The False Promise of “Local”

Keeping data local reduces the default path to third parties.

- It does not erase credentials, file permissions, network reach, or destructive tools.
- It does not make an arbitrary extension trustworthy.
- It does not make a model’s judgment suddenly excellent.

Local-first means fewer strangers see your data—not that your agent becomes wise.



Olares OS: the Open-Source Personal Cloud OS



An open-source personal cloud OS based on Kubernetes. Headless, server-grade, built for 24/7 remote service.

- Curated app market with one-click deployment, with application sandboxing for enterprise-grade security.
- Unified file system across apps, with built-in backup.
- Single sign-on via Olares ID across every app and device.
- Local-first AI: local model hosting, private knowledge bases, GPU management.
- Built-in system apps such as Files, Vault, Wise, Control Hub, and Dashboard.
- Seamless anywhere access through LarePass clients (phone, PC, tablet, browser) with zero network configuration.
- Developer tools: Studio and Olares CLI for building Olares apps.

4+ yrs

IN DEVELOPMENT

350+

APPS IN MARKET

5.2k+

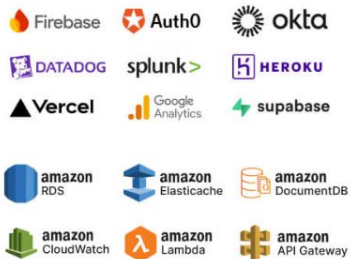
GITHUB STARS

Public Clouds

SAAS



PAAS



IAAS

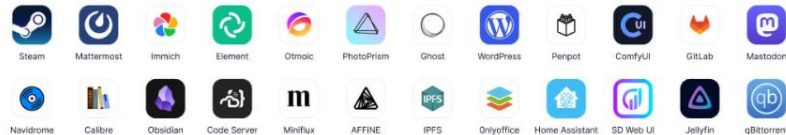


Olares

Built-in Applications



Community Applications



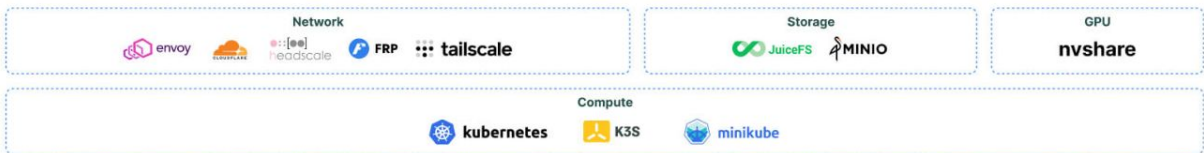
Framework



Platform



Infrastructure



서울

**One-click agents actually become
usable when the whole stack lives
on hardware you control**



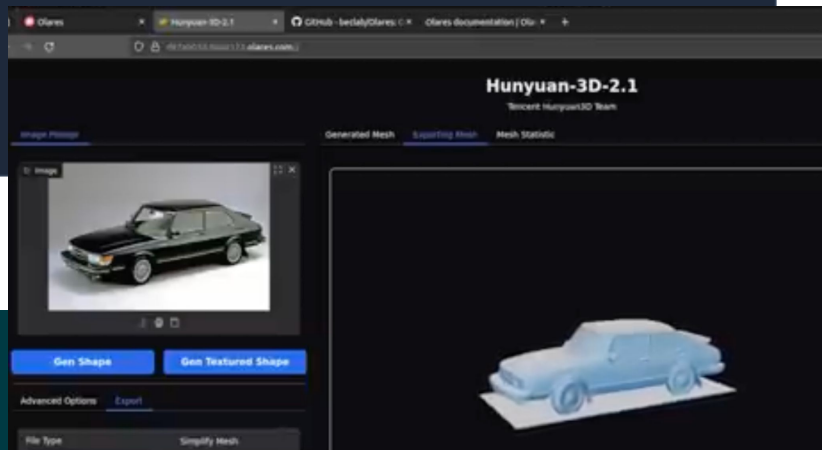
Get Agents Running without Assembling a House by Hand

What users **SHOULD** do

- ✓ Install an agent
- ✓ Connect a local model
- ✓ Ask for a real task

What they should **NOT** have to do first

- ✗ Assemble an AI platform
- ✗ Study a cluster diagram
- ✗ Memorize every CLI command





Graph



Manager



CPU

1%

RAM

90%

GPU

0%

VRAM

3%

Temp

43°



Show Image Feed



1



Run

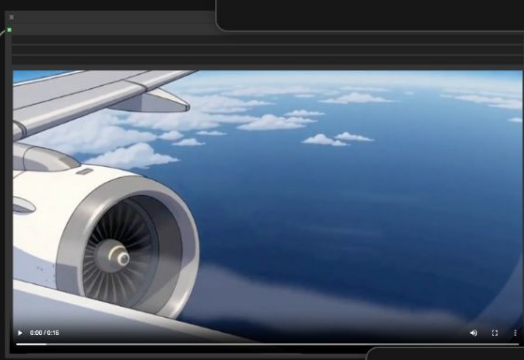
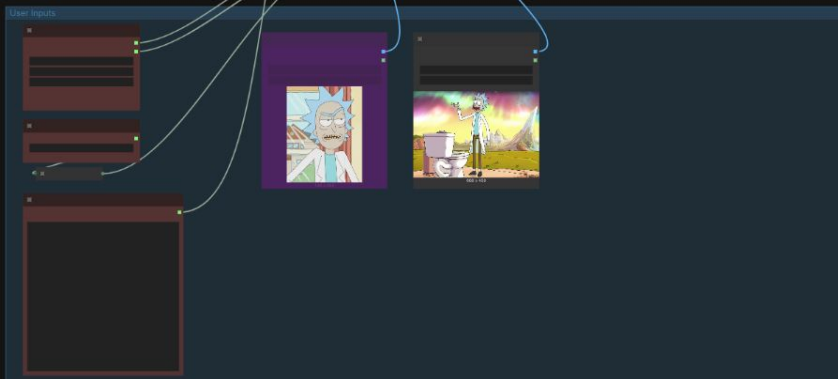
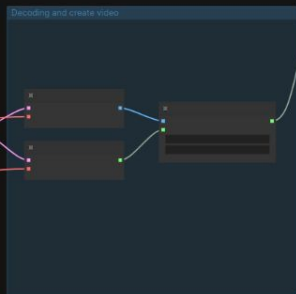
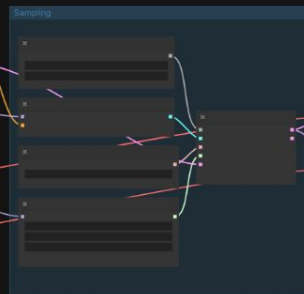
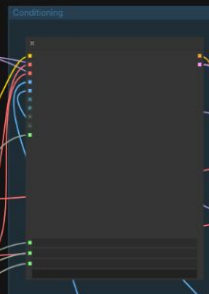
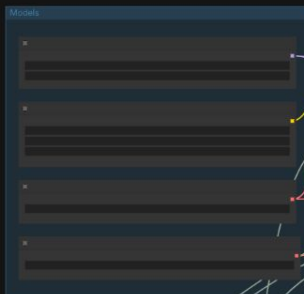


Job Queue

Clear

All

Completed



Complex Infrastructure, Minimal Homework

Minimal Homework (Your Focus)

Your Task



Interface & Skills

Hermes + Olares CLI Skills



Local Runtime & Assets

Local Model + Files + Applications



Foundation Infrastructure (Fully Automated)

Olares OS: App Lifecycle, Compute, Storage, Networking

1. Memory Slicing 2. Time-Sharing 3. App Exclusive



K3s turns Your Machine(s) to Personal Cloud

Declarative lifecycle: install, restart, update, and roll back workloads.

Resource Contracts

Guarantees and constraints for workloads, managing resources like:

- CPU limits
- Memory allocation
- Storage contracts
- Accelerator / GPU limits

Identity & Isolation

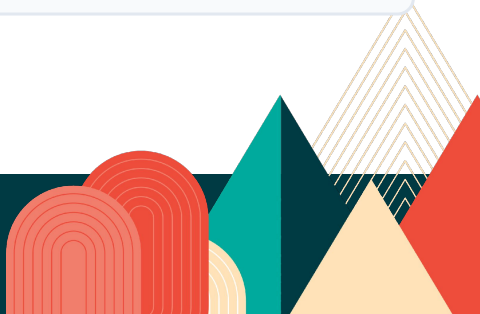
Secures and segments execution environments using:

- Isolated namespaces
- Dedicated service accounts
- Granular network controls

Repeatability

Ensures reliable, persistent configurations across active phases:

- Survives terminal sessions
- Persistent configurations
- Consistent state recovery



Isolation Is a Design

Give each workload an identity.

Data Access

Mount only the data it needs, not “everything in case it becomes useful.”

Network Paths

Default-deny network paths wherever the product permits it.

Resource Limits

Set explicit CPU, memory, storage, and GPU limits.

Security Philosophy

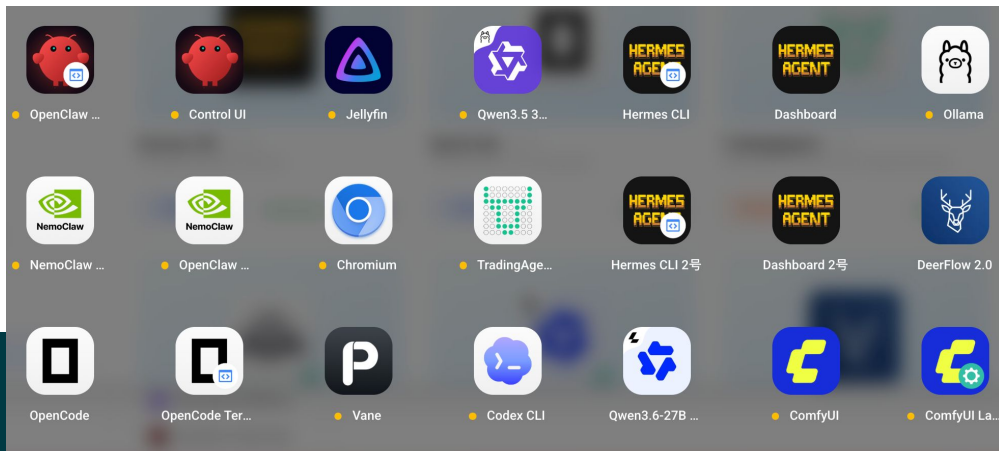
Assume an agent will eventually ask for more access. **Design the “no” first.**

Local AI Into an App Experience

Market → Get → Install → Launch

- Install Hermes Agent from Olares Market.
- Open Hermes CLI or the dashboard from Launchpad.
- Connect a local, tool-capable LLM through guided setup.

One click installs the agent.



The Agent Can Learn the Environment

Hermes/ Claude Code/ Open Code + Olares CLI

✓ Olares CLI Demo - Proof Sheet (2026-07-22)

Step 1 · Sovereignty

NAME	OLARES-ID	STATUS
* mileymiley	mileymiley	logged-in

| 身份确认—mileymiley 以 Owner 登录自己的实例。一台实例 = 一个属于我的 Olar

Step 2 · Live OS (Dashboard Overview)

METRIC	VALUE	UTIL
CPU	1.12 / 24.00	4.67%
Memory	16.12Gi / 93.68Gi	17.21%
Disk	280.56Gi / 1.83Ti	14.96%
Pods	87 / 200	43.5%
GPU	1Mi / 23.89Gi	0%
Fan CPU	1300 / 2900 RPM	44.83%
Fan GPU	1500 / 3100 RPM	48.39%

Top apps: Hermes 2号, Hermes Agent, NemoClaw, OpenClaw, OpenCode, Chromium, DeerFlow 2.0, ComfyUI, Jellyfin, Ollama, Qwen3.5 35B, TradingAgents

Step 4 · K8s Substrate

Node: olares (v1.33.3+k3s1, Ready, 71 days uptime)

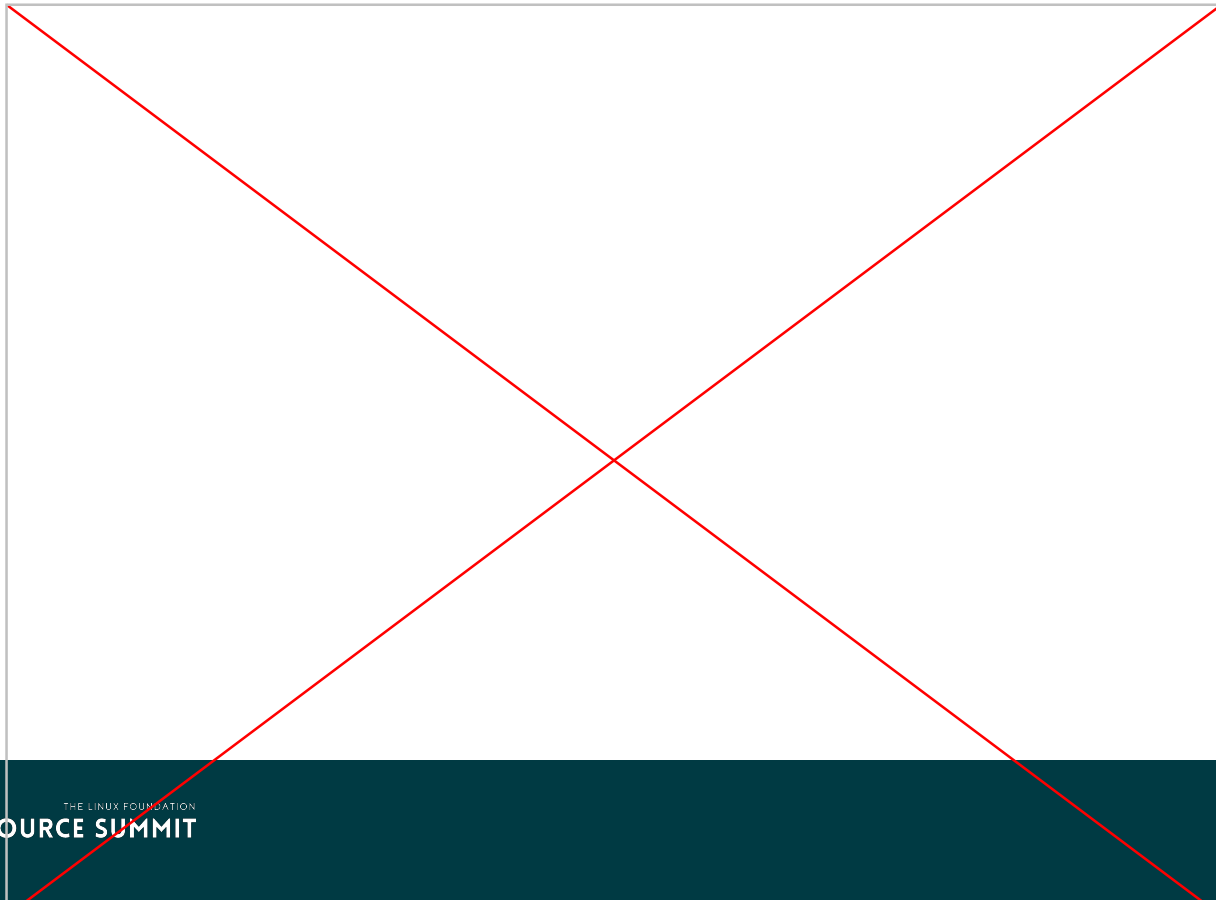
Middleware:

postgres/citus	(managed Postgres + Citus)
redis/kvrocks	(managed Redis-compatible)
nats/nats	(managed message queue)

Step 5 · Data on Own Disk + Local AI

drive/Home/: Code/ Documents/ Downloads/ Movies/ Music/ Ollama/ Pictures/
drive/Home/Ollama/: ollamaqwen35a3budq4klv2server/ ← Qwen3.5 35B 权重
GPU: NVIDIA RTX 5090 Laptop (23.89Gi, Time slicing, 46°C, 7.66W)
Tasks: ollama ✓ comfyui ✓ (HAMI vGPU time-slicing 共享同一块 GPU)

Demo: Say What You Want and See Action



Flexible Deployment Across Platform

Choose the deployment method that fits your infrastructure

RECOMMENDED FOR PRODUCTION

Linux (Ubuntu / Debian)

The standard path optimized for maximum production performance & stability.

ISO Image (Rec.)

- Bare-metal install with auto-configured drivers.

One-Line Script

- Quick setup on existing Linux distros.

Docker Image

- Run in a containerized Linux sandbox.

AI APPLIANCE

NVIDIA DGX Spark

Fully optimized to leverage the native high-performance GPU hardware.

One-Line Script (Rec.)

- Direct setup mapped to Spark environments.

ISO Image

- Full OS install with pre-baked GPU drivers.

DEV / LAB / ALTERNATIVES

Hypervisors & Local Dev

Lightweight options engineered for testing, VMs, and local environments.

Proxmox VE (PVE)

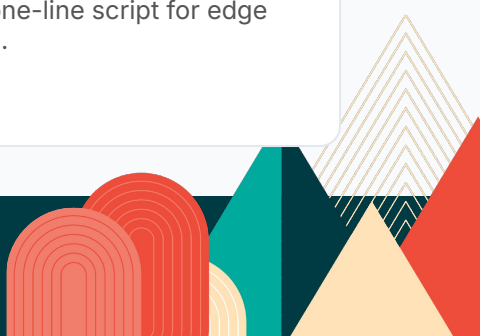
- Deploy as VM, direct script, or LXC.

macOS & Windows (WSL2)

- Run via Docker or MiniKube locally.

Raspberry Pi (ARM)

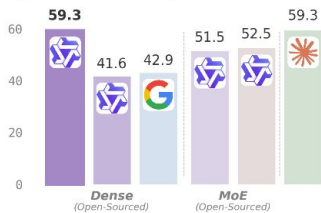
- Light one-line script for edge testing.



Local Models Reach Claude 4.5 Opus Performance

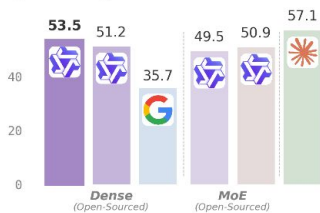
Terminal-Bench 2.0

Agentic Terminal Coding



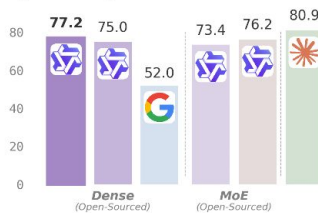
SWE-bench Pro

Agentic Coding



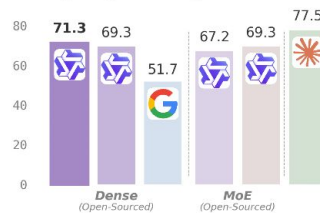
SWE-bench Verified

Agentic Coding



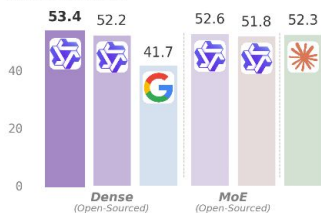
SWE-bench Multilingual

Multilingual Agentic Coding



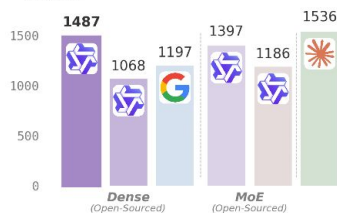
QwenClawBench

Real-World Agent



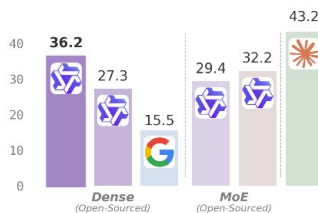
QwenWebBench (Elo Rating)

Artifacts



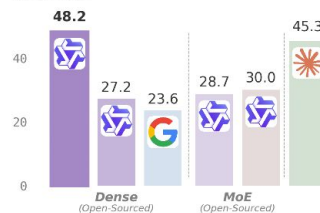
NL2Repo

Long-Horizon Coding



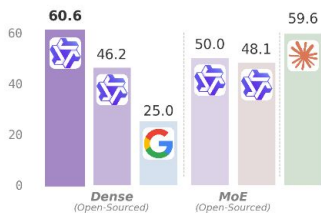
SkillsBench

Agent Skills



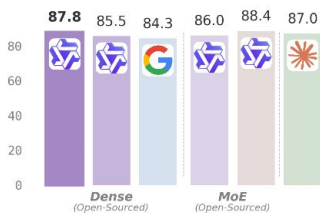
Claw-Eval (pass^3)

Real-World Agent



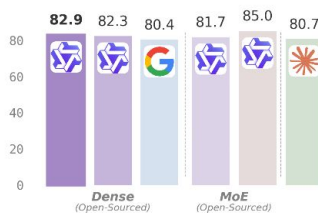
GPQA Diamond

Graduate-level Reasoning



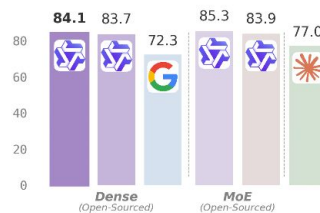
MMMU

Multimodal Reasoning



RealWorldQA

Image Reasoning



Qwen3.6-27B Qwen3.5-27B Gemma4-31B Qwen3.6-35B-A3B Qwen3.5-397B-A17B Claude 4.5 Opus

Both Olares One and Olares Enterprise run Qwen3.6-27B, a dense multimodal model that matches or beats Claude 4.5 Opus on core benchmarks. Flagship-level agentic coding and multimodal reasoning, fully offline, with a clear path to ROI against cloud API spend.

BENCHMARK	QWEN3 .6-27B	CLAUDE 4.5 OPUS
Graduate-level reasoning (GPQA)	87.8	87.0
Multimodal reasoning (MMMU)	82.9	80.7
Real-world image reasoning	84.1	77.0
Agentic terminal coding	59.3	59.3

THROUGHPUT AND ROI BY TIER

METRIC	OLARES ONE	OLARES ENTERPRISE
Single-user speed	37.9 tokens / sec	45.9 tokens / sec
Concurrency & throughput	2 reqs @ 68.3 tokens / sec	10 reqs @ 276.8 tokens / sec
Daily token output	~1.0M tokens / day	~4.0M tokens / day
Daily API value	~\$75 / day	~\$300 / day
Payback period	~2.5 months	~4.6 months

Note: Daily token output and API value are calculated based on 4 hours of effective daily utilization, a 10:1 input-to-output token ratio (typical for agentic coding), and Anthropic's Claude 4.5 Opus API pricing of \$5 / \$25 per 1M tokens. Payback period assumes 22 working days per month.

ClawWork: AI Coworkers Earn Money from Real-Life Tasks, Mostly Open Source!

RANK	AGENT	STARTER ↑↓	BALANCE ↑↓	% CHANGE ↑↓	INCOME ↑↓	COST ↑↓	PAY RATE	AVG QUALITY ↑↓	TASKS ↑↓	STATUS
	 ATIC + Qwen3.5- Plus	\$10.00	\$19,915.68	↗ +199056.8%	\$19,914.38	\$8.70	\$2.3k/hr · \$18.3k/day	61.6%	198	 Thrive
	 Gemini 3.1 Pro Preview	\$10.00	\$15,661.71	↗ +156517.1%	\$15,757.48	\$105.76	\$1.3k/hr · \$10.3k/day	43.3%	193	 Thrive
	 Qwen3.5- Plus	\$10.00	\$15,268.13	↗ +152581.3%	\$15,264.92	\$6.78	\$1.4k/hr · \$11.1k/day	41.6%	198	 Thrive
#4	 GLM-4.7	\$10.00	\$11,497.05	↗ +114870.5%	\$11,503.49	\$16.44	\$878/hr · \$7.0k/day	40.6%	198	 Thrive
#5	 ATIC- DEEPSEEK	\$10.00	\$10,877.01	↗ +108670.1%	\$10,870.52	\$3.52	\$2.6k/hr · \$20.6k/day	66.8%	71	 Thrive
#6	 Qwen3- Max	\$10.00	\$10,782.80	↗ +107728.0%	\$10,781.06	\$8.26	\$1.1k/hr · \$8.6k/day	37.9%	198	 Thrive
#7	 Kimi- K2.5	\$10.00	\$10,471.21	↗ +104612.1%	\$10,483.20	\$21.99	\$859/hr · \$6.9k/day	36.6%	162	 Thrive
#8	 Claude Sonnet 4.6	\$10.00	\$1,219.01	↗ +12090.1%	\$1,226.99	\$17.98	\$2.0k/hr · \$16.3k/day	44.6%	12	 Thrive

The Hard Parts: Cost of Sovereignty

- Hardware is not free. (GPU /Mac /Olares One etc)
- You become the sysadmin of your intelligence.



Let's build together.

Contribute to personal cloud + data sovereignty

github.com/beclab/Olares



GITHUB TRENDING

#1 Repository Of The Day



GET IN TOUCH

Email: **miley@olares.com**

Website: **www.olares.com**

Linkedin

<https://www.linkedin.com/company/olarestech/>

<https://www.linkedin.com/in/mileyfu>

STAY CONNECTED

X: **[@Olares_OS](#)**

YouTube: **[@OlaresOS](#)**



OPEN
KOREA

SOURCE SUMMIT

THE LINUX FOUNDATION

서울

