

Latest Agentic Harness Trends

Agent 執行框架的最新趨勢

Harnesses, token routing and evaluation

執行框架、Token 路由與評估

Bruce Huang

Read the research

掃碼閱讀技術分享



My recent PolyU lecture

最近在香港理工大學的分享



Dario Amodei
Anthropic

DARIO'S ESSAY

We Must Pace the Frontier

我們必須掌握前沿 AI 的發展節奏

Dario Amodei · September 2026

His essay is why today starts with people.
這篇文章，讓今天的話題從人出發。

RESPONSES



Sam Altman
OpenAI
Supports pacing the frontier



Elon Musk
xAI
“Dario is right”

EARLIER RELATED PROPOSAL



Demis Hassabis
DeepMind cofounder
Independent testing
JUL 14, 2026

When AI Can Do the Work, What Will We Do?

當 AI 開始接手工作，人該做什麼？

People steer AI

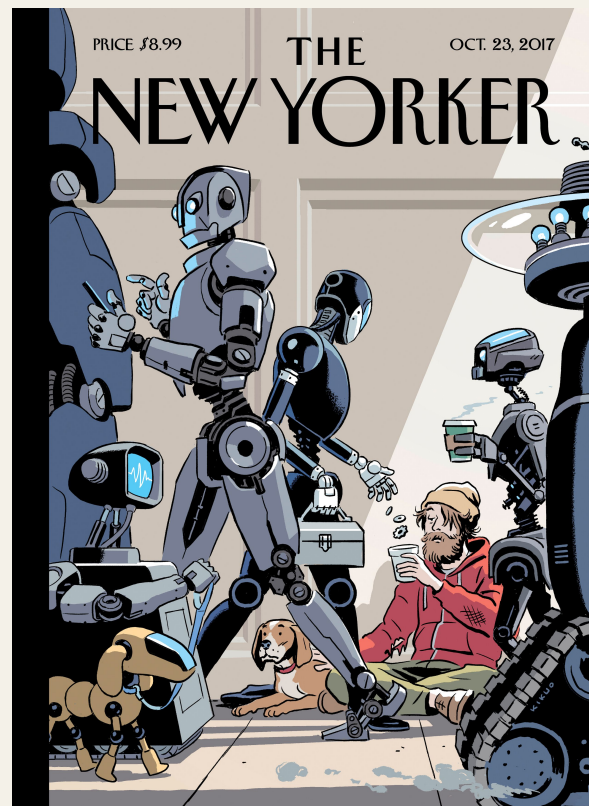
人類駕馭 AI

ClawCon Macao @ BEYOND Expo 2026



People get left behind

人類被邊緣化



Ask a question

問一句話



Delegate a task

交一件事

A goal, actions and a result you can check

目標、執行與可檢查的成果

We can move from doing tasks to directing work

從親自做每一步，走向組織和判斷工作

What should AI take on?

哪些工作可以交給 AI？

What should remain our responsibility?

哪些判斷與責任需要由人承擔？

Top AI trends, then our choices at work

先理解變化，再討論人的選擇

Product & Community

產品與社區

ClawCon HK

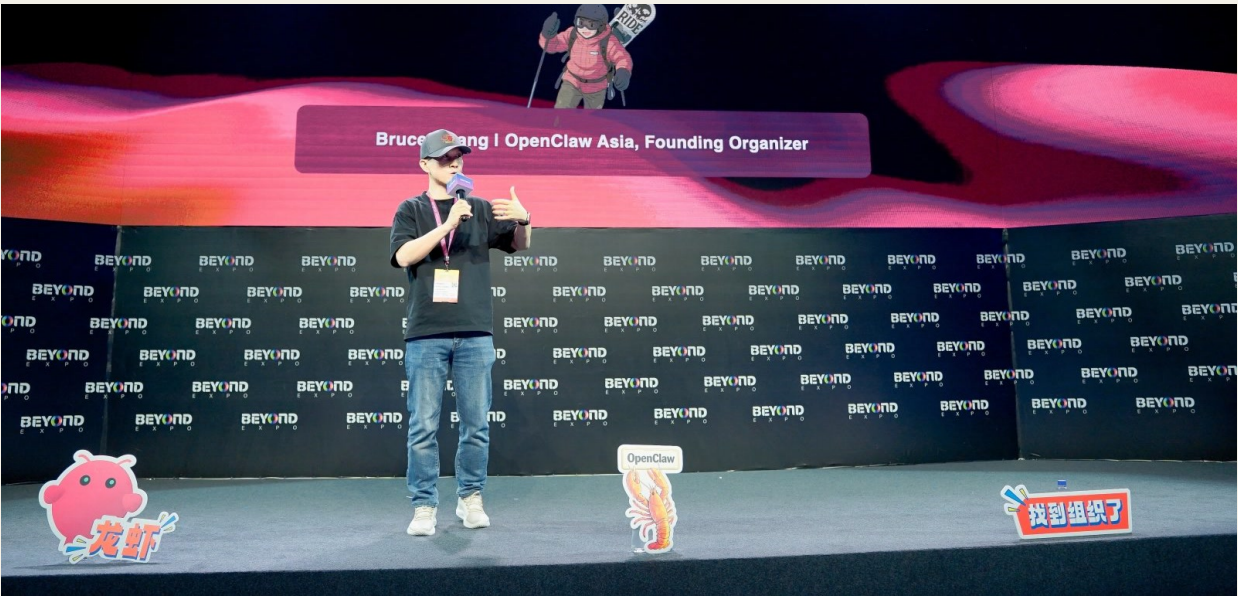
February · 2 月



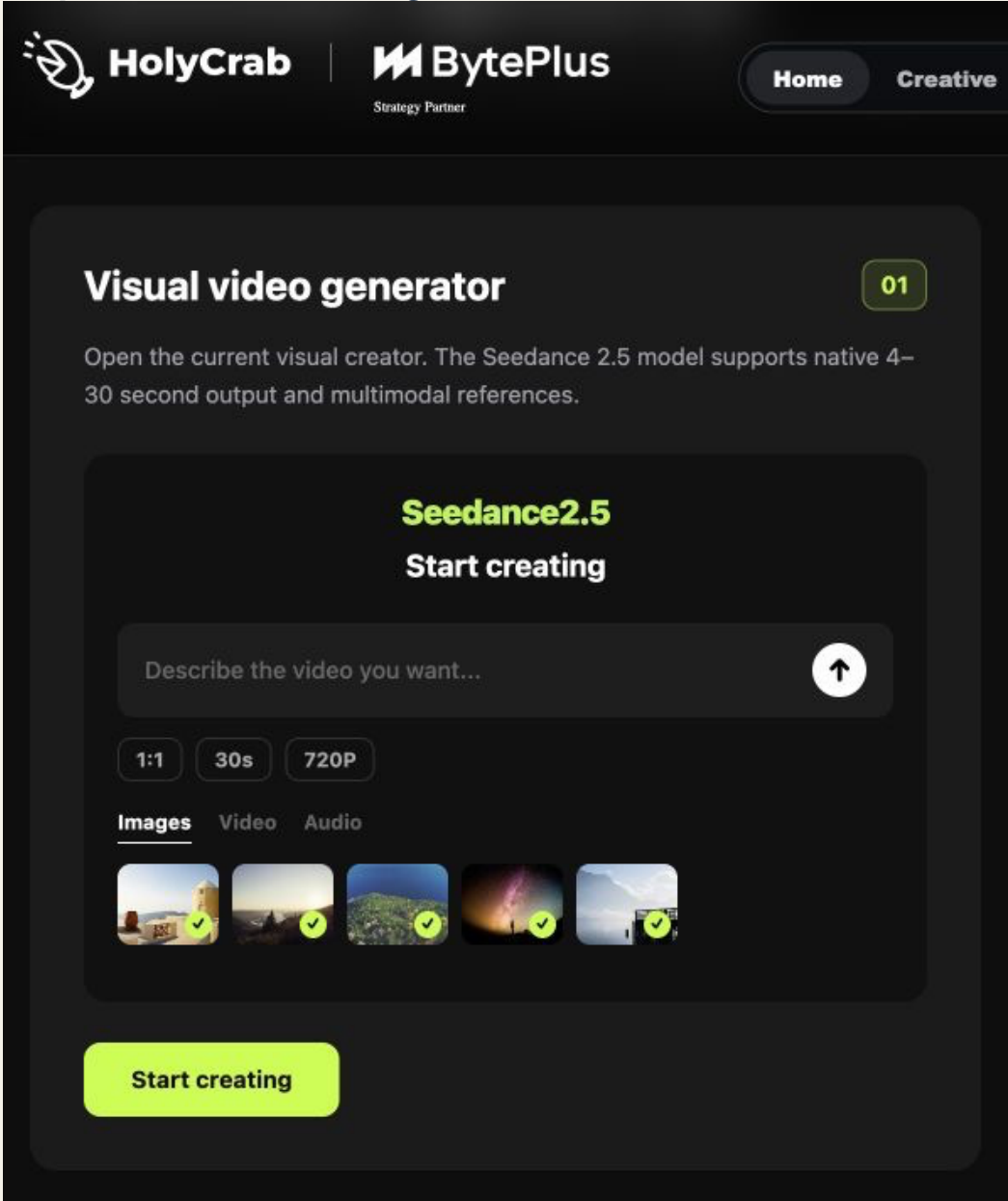
ClawCon Macao

@ The Venetian Macao

Late May · 5 月底

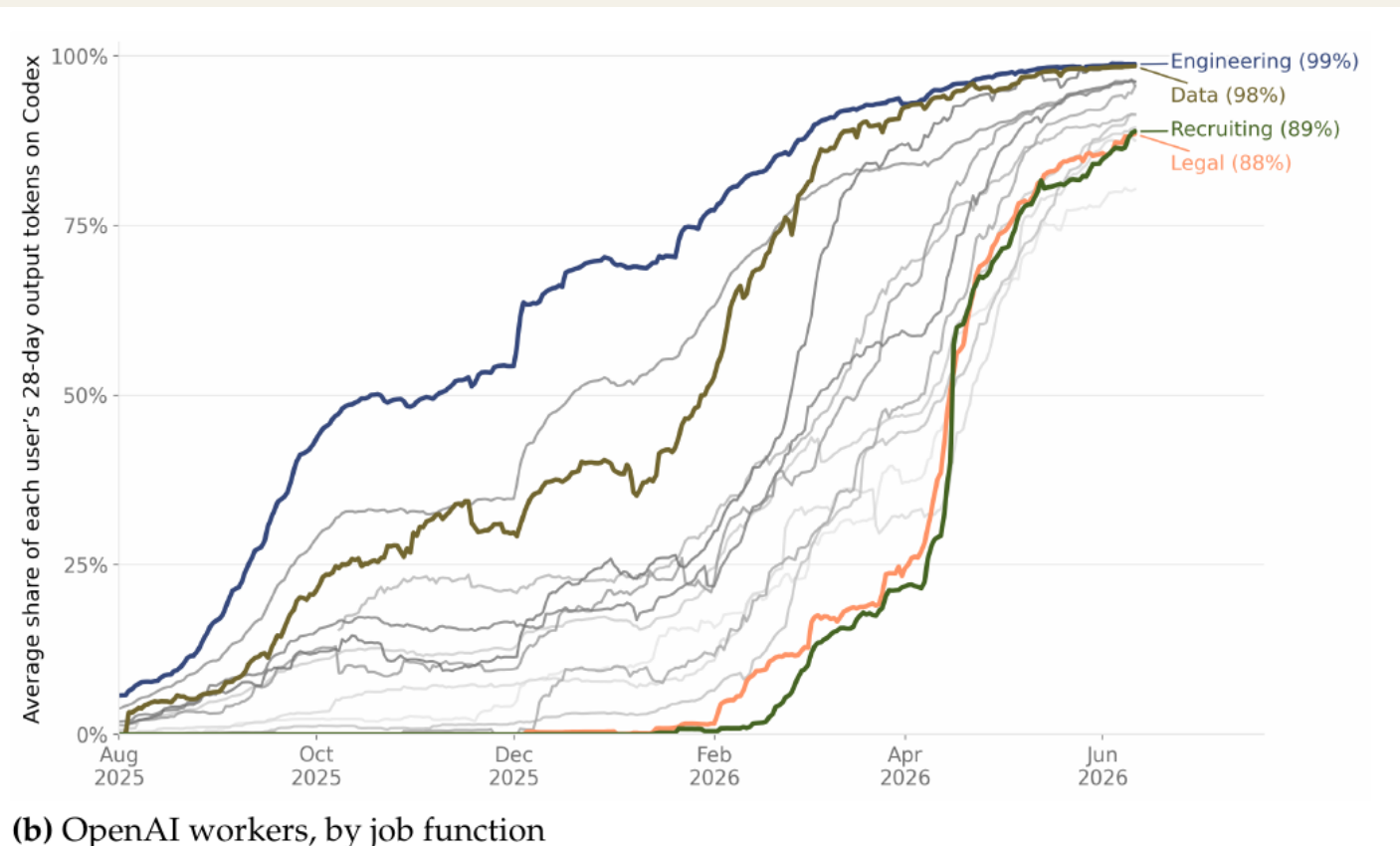


HolyCrab / AI video generation



Codex use across functions at OpenAI

OpenAI 內部不同職能的 Codex 使用份額



AI is moving into real work

模型能力正在進入工作流

Codex for project execution

Codex 接住開發與項目任務

Claude Code and Cowork

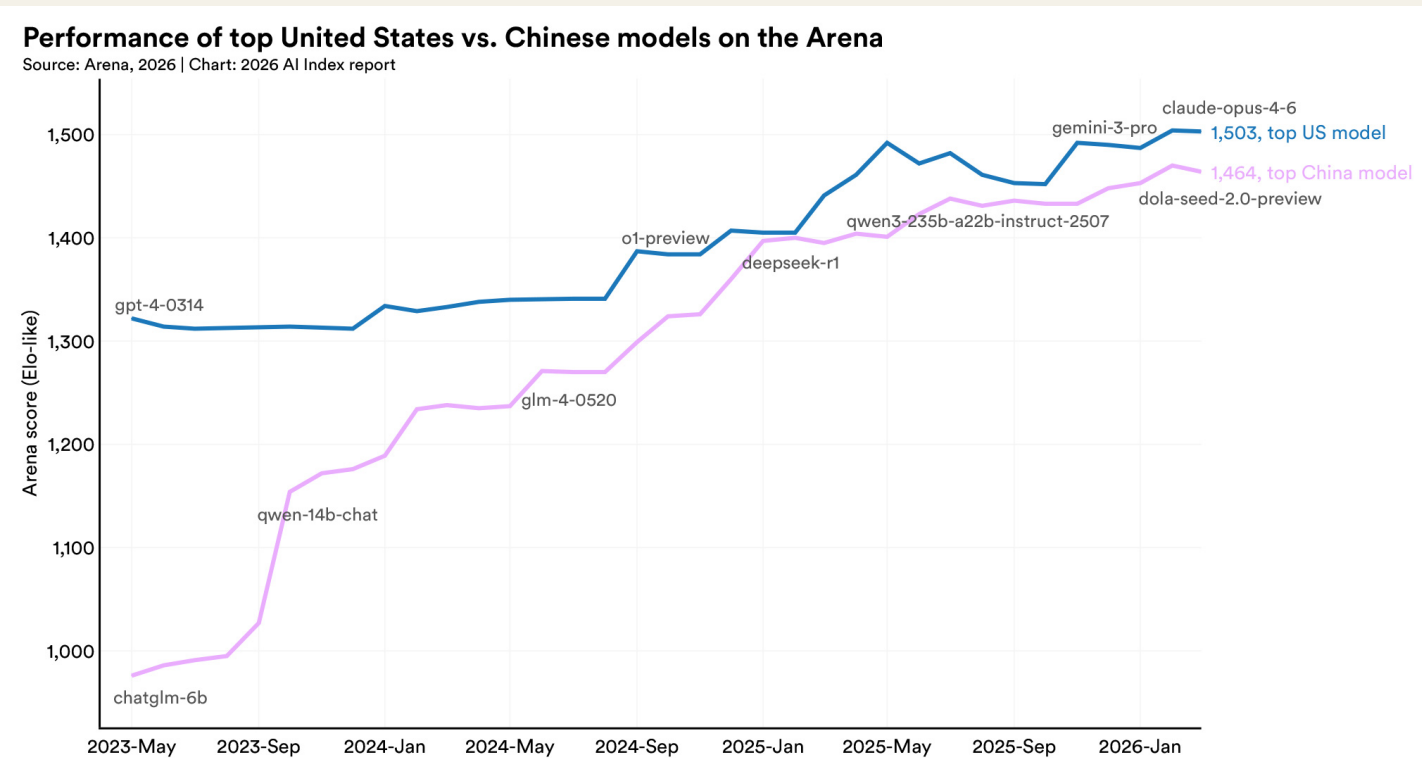
Claude 接住代碼與知識工作

Look at the work a product can take on

看產品能接住什麼任務

Leading models on Arena

Stanford AI Index：頭部模型表現



More models are near the frontier

更多模型正在接近能力前沿

Watch the gap between the curves

先看曲線之間的距離

Then ask what the product can deliver

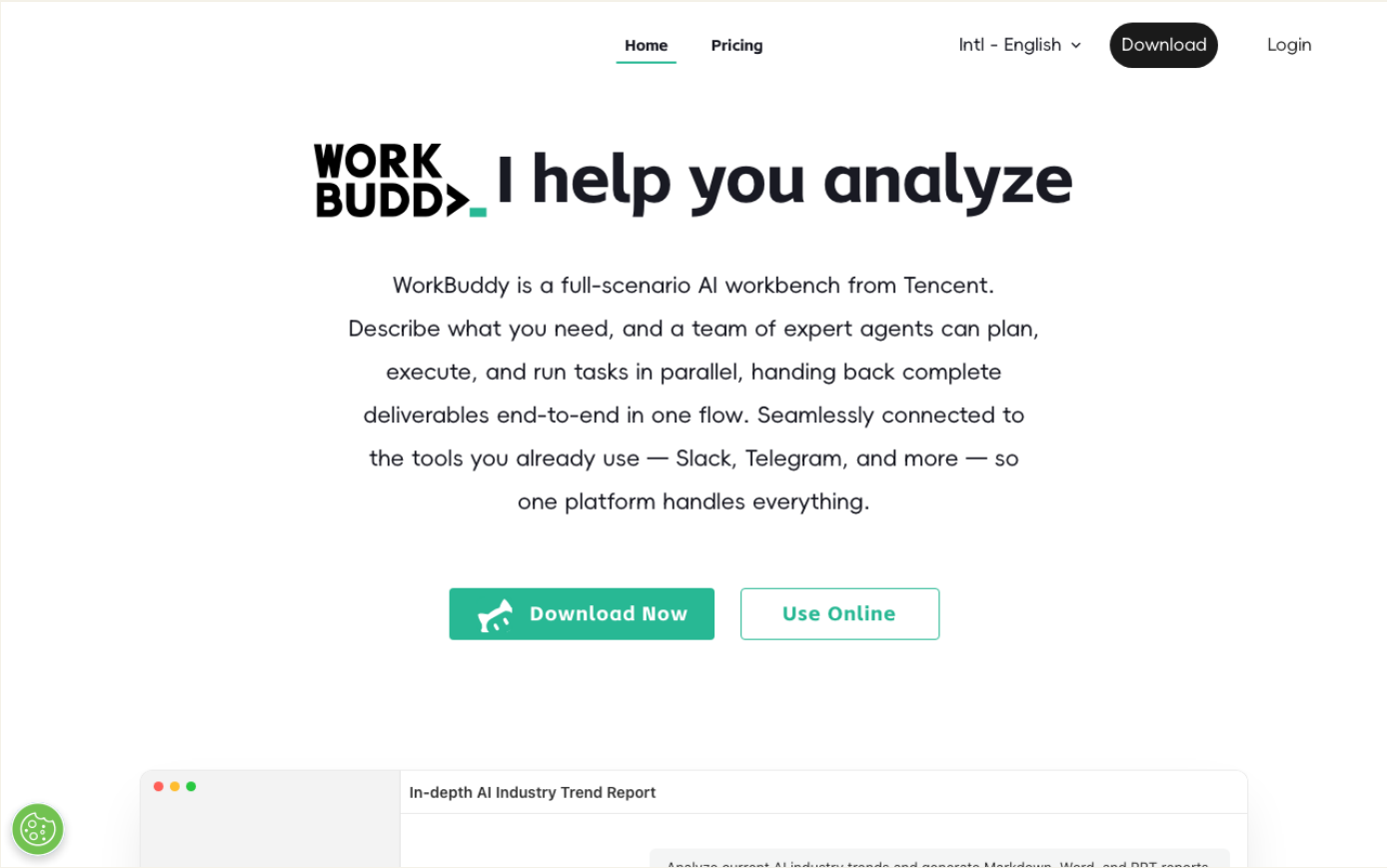
再看產品能否完成工作

Arena data through March 2026

歷史趨勢背景，不是即時榜單

WorkBuddy: a workspace for tasks

WorkBuddy 官方工作台與任務入口



AI is coming to the desktop

WorkBuddy、豆包，把 AI帶進桌面。

WorkBuddy works with local files

WorkBuddy 處理本地文件和辦公成果

Doubao reads files and pages

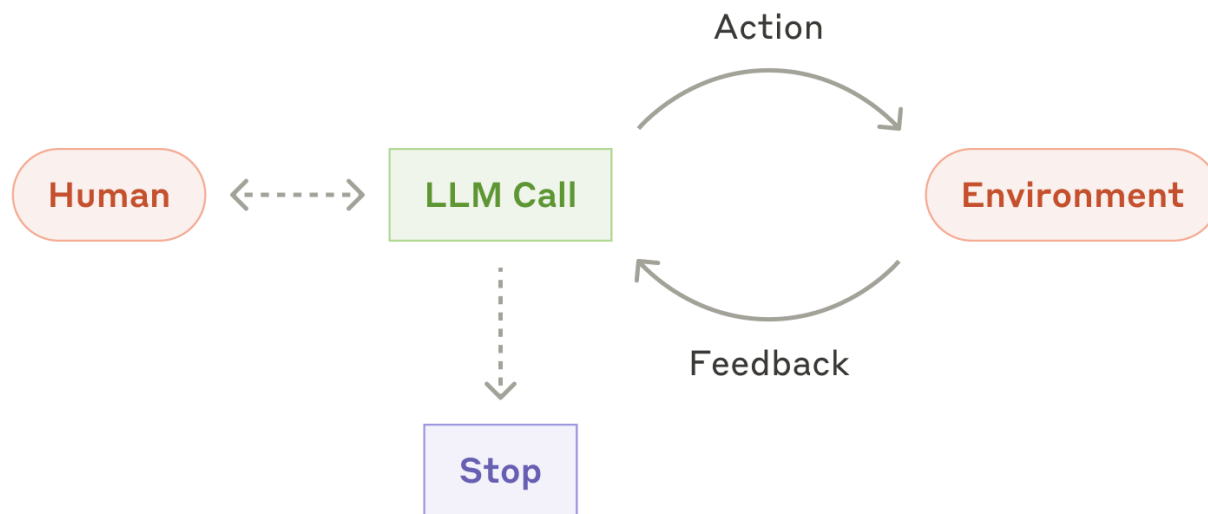
豆包輔助理解文件與頁面

Available actions depend on the product

實際入口與權限決定執行深度

The action and feedback loop

Anthropic · Agent 的行動、反饋與停止循環



An agent acts and learns from the result

Agent 帶着目標行動，再根據反饋調整

It uses tools to make progress

使用工具推進工作

Feedback shapes the next step

根據結果決定下一步

The harness gives it a place to work

Harness 是配套的執行環境

Choose by the work you need done

按工作入口選擇；產品能力會重疊

Build software

編程與項目執行



Codex



Claude Code



Cursor



Grok Build



GitHub Copilot



Qoder

Finish office work

辦公文件與跨應用任務



Claude Cowork



WorkBuddy



Doubao Work



QwenWork

Complete an assignment

從目標到完整交付



Manus



ChatGPT Work



Genspark Super Agent



Perplexity Computer

Keep a teammate

記住背景，持續協作



Grok Bot



OpenClaw



Hermes

Memory · routines · handoffs

Agents now span more of our work

Agent 正在覆蓋更多種工作

From coding to office tasks

從編程走向辦公任務

From one task to ongoing work

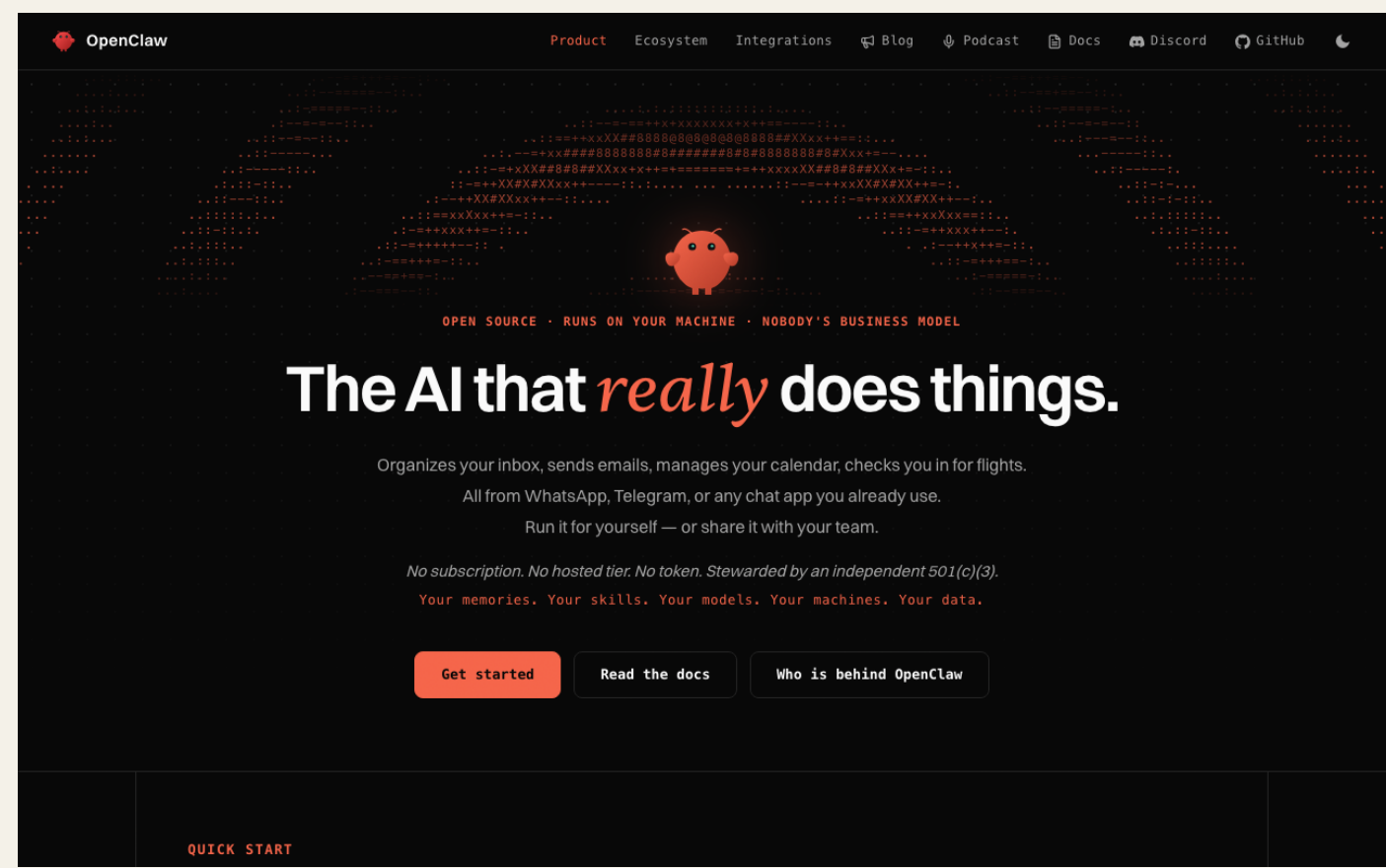
從一次交付走向持續協作

Choose by task. The categories overlap.

先選工作，再選入口；各類能力正在融合。

OpenClaw: an open personal assistant

OpenClaw 官方主頁 · 開源個人助手



OpenClaw gives recurring work a home

OpenClaw：把持續任務，放進日常入口。

A familiar messaging entry point

從熟悉的消息入口開始

Try one daily briefing

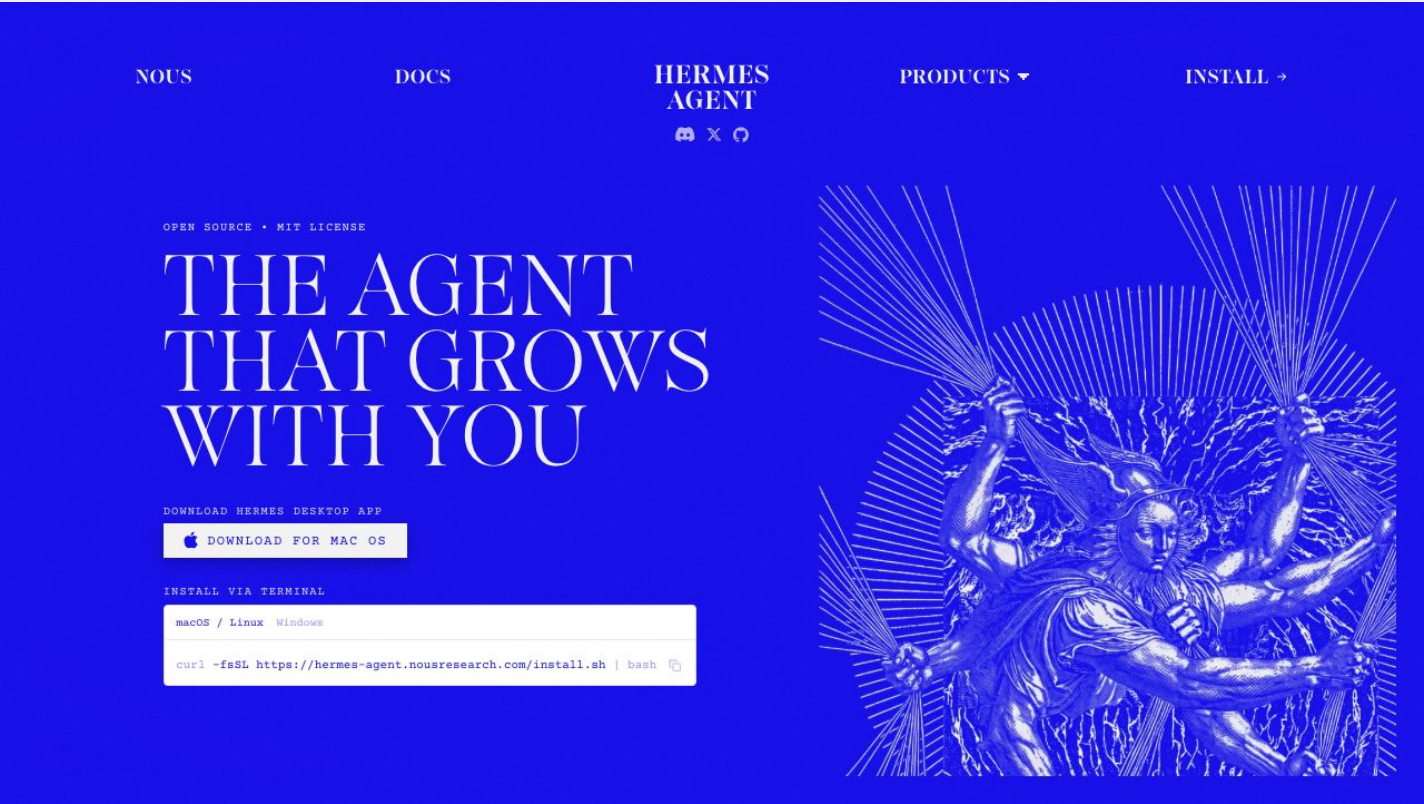
先試一份每天要看的簡報

Start with one recurring task

先接一件事，再擴大可用工具

Hermes Agent: memory and Skills

Hermes Agent 官方主頁 · 持續記憶與 Skills



Hermes can carry experience into the next task

Hermes：把這次經驗，留給下次用。

Persistent memory and reusable Skills
持續記憶與可複用 Skills

Turn a correction into a rule
把一次修正寫成下一次的規則

Saved experience still needs checking
記住流程，仍要核對結果

DeepSeek Harness: everything is a plugin

DeepSeek Harness 官方主頁 · Everything is a plugin



DeepSeek Harness lets developers assemble agents

DeepSeekHarness：讓能力可組裝。

Tools, sessions and execution loops
工具、會話與執行循環

Capabilities connect as plugins
按需要組合插件能力

Developer preview: use a sample project
開發者預覽，適合隔離的試驗項目

Sample CSV to a review list

從樣例記錄到待複審鏡頭

Request

需求

Find failed shots and explain why

篩出未通過的鏡頭與原因

Execute

執行

Read, filter and write the result

讀文件，完成篩選並寫出結果

Check

檢查

Verify a few known records

用已知記錄逐項核對

Codex and Claude can return finished work

把需求交出去，帶回可檢查的成果

Code and files become deliverables

項目執行可以交回文件與工具

A clear check defines completion

先寫清怎樣知道任務完成

Ask for an output you can open and check

把“做完了”寫成可觀察的行為

01

AI Foundation

研究模型與訓練

Models and new capabilities

02

Agent Builders

搭建能執行的系統

Tools, execution and recovery

03

AI in Business

把 AI 用進業務

Useful outcomes in a real context

White-collar work now has only three categories

白領工作，只剩三類

Research the models

研究模型

Build agents or apply AI in business

搭建 Agent，或把 AI 用進業務

Which kind of work will you own?

你要負責哪一類工作？

01

Research

研究模型

A question and a reproducible experiment

問題、實驗與可復現結果

02

Build agents

搭建 Agent

Execution that recovers and checks

執行、恢復與驗證

03

Apply AI

業務應用

A useful outcome you can judge

業務目標、判斷與成果

Your value comes from the problem you solve

你的價值，體現在你能解決哪種問題。

Research proves a capability

研究要證明新能力

Builders and users make it useful

開發與業務把能力變成結果

Choose the responsibility you want to own

從職責出發，決定下一步練什麼



A real introduction. In Japanese.

用日語，做真實的商務自我介紹

Used in client meetings in Japan

在日本客戶會面中使用的自我介紹



Start with one real conversation.

從一個真實的溝通任務開始

iPhone Duo / AI concept advertisement



Two clips. One 30-second ad.

兩段影片，合成一支 30 秒廣告

Made with HolyCrab

Seedance 2.0 Fast · 2 tasks

同一人物、產品與銜接末幀



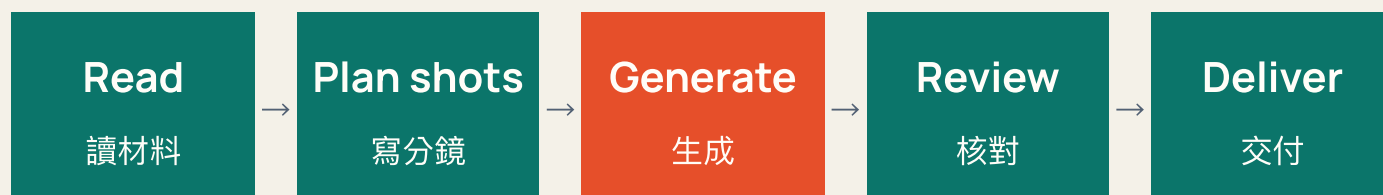
See the workflow

下午一起實作

Review the product and the join.

A workflow with a review loop

一條可以檢查、修正的製作流程



A failed shot goes back with a reason

發現問題，回到對應鏡頭修改

People approve direction, spending and release

方向、預算與最終成片由人確認

Keep task IDs, inputs and review reasons

保留任務編號、輸入參數與評審原因

Agents move the work forward. People judge it.

讓 Agent 推進流程，讓人判斷成片

Review each shot against the brief

逐鏡頭對照要求核驗

Return problems to the right step

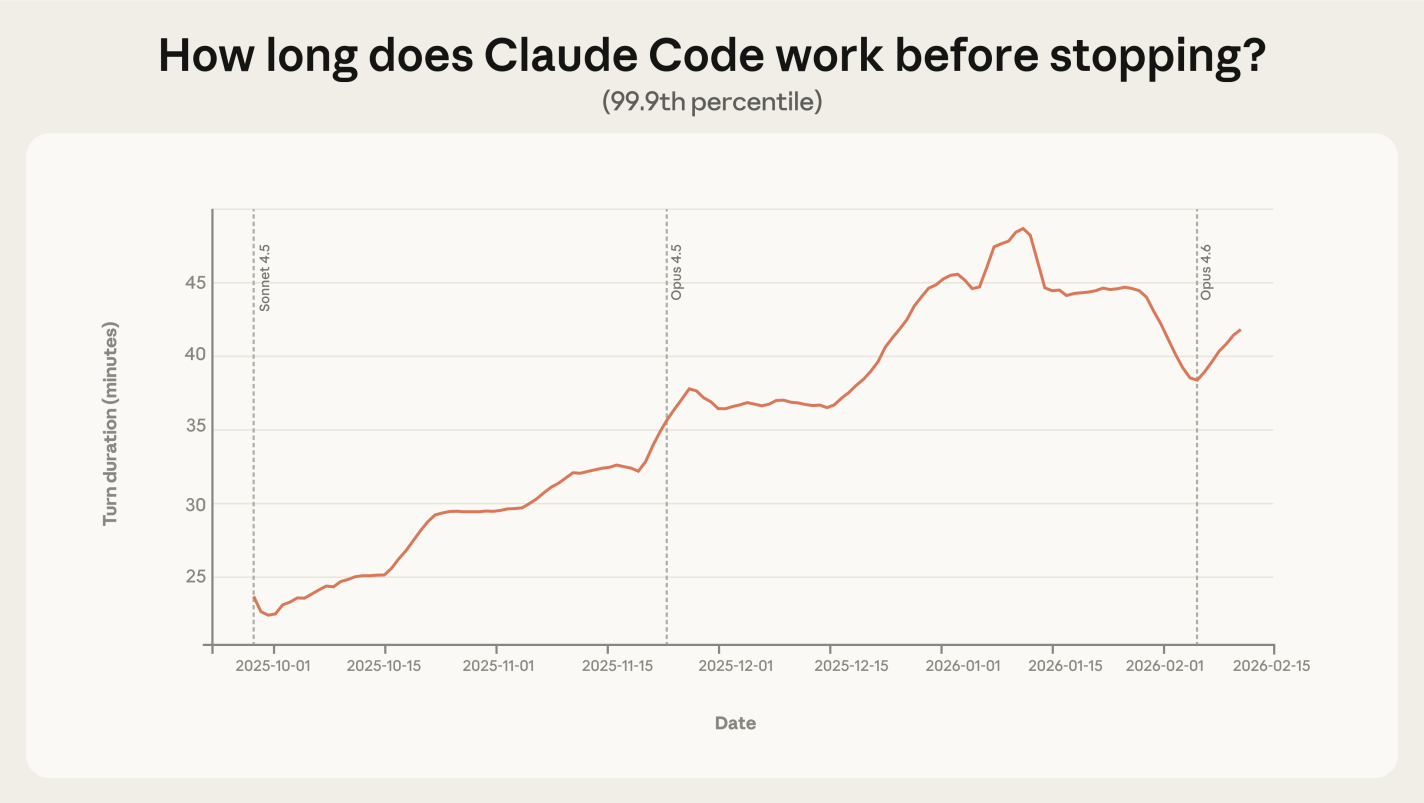
發現問題，回到對應環節修改

People approve direction, spending and release

方向、預算與最終成片由人確認

How long Claude Code works before stopping

Claude Code 單次連續執行時長的 99.9 分位數



Long tasks are getting longer

長任務，開始能持續更久了。

This is the longest-running tail

圖中展示運行最久的一小部分

Typical tasks are much shorter

日常任務仍然短得多

99.9th percentile; median around 45 seconds

不能把最長尾部當成日常水平

An advanced practice goal

一次進階實踐

60

minutes

分鐘

A task with enough context and clear limits

任務清楚、材料夠用、邊界明確

Act, read feedback and check again

執行、讀反饋，再檢查

Could an agent advance an hour of useful work?

你能讓 Agent 自主推進一小時工作嗎？

Give it a clear task and materials

任務清楚，材料夠用

Define checks and stopping points

知道怎樣檢查，也知道何時停下

Useful progress matters more than runtime

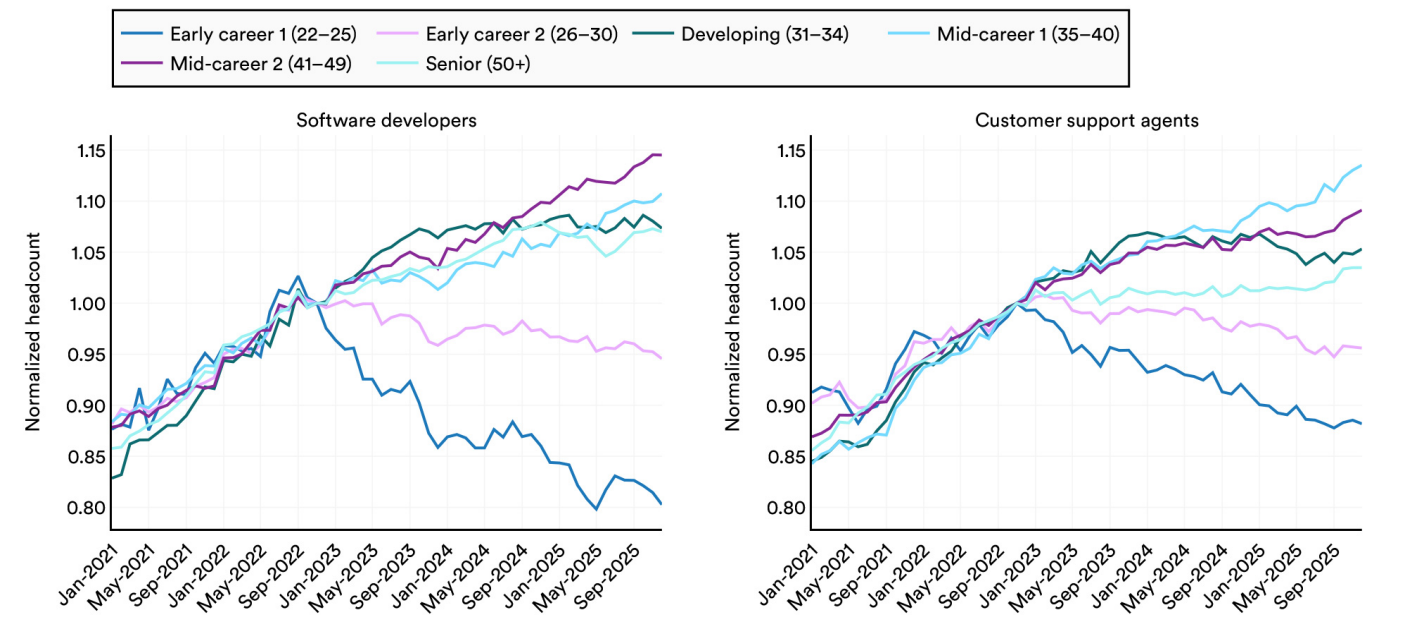
成果質量比運行時長更重要

Employment trends by age and occupation

Stanford · 按年齡區分的開發與客戶支持就業趨勢

Normalized headcount trends by age group for software developers and customer service agents, 2021–25

Source: Brynjolfsson et al., 2025 | Chart: 2026 AI Index report



Career effects will be uneven

職業變化，不會平均地落到每個人。

Look at tasks and experience levels

看具體任務與從業階段

Ask which responsibilities are changing

看哪些責任正在改變

This chart alone cannot establish causation

就業觀察不能單獨證明 AI 因果

Result

結果

What should I receive?

最後要拿到什麼？

Materials

材料

What may it read?

允許讀取哪些材料？

Actions

操作

What may it do?

允許執行哪些動作？

Checks

檢查

How will I judge the output?

怎樣知道結果合格？

Help

求助

When should it come back to me?

什麼情況要交回給人？

A clear result
needs a
clear check

先寫清結果，再寫清檢查方法。

Take 45 seconds to write two lines
現場用四十五秒寫兩句話

Use a task you really need to do
選一件你真的要做的事

What will you receive? How will you check it?
結果是什麼？怎樣知道它合格？

0–10 min

Prepare

準備

One folder and one task card

一個目錄，一張任務卡

10–40 min

Run

運行

Read the materials, then deliver

先讀材料，再交完整輸出

40–60 min

Review

檢查

Revise once and save the rule

修改一次，把做法保存下來

One small task is enough to begin

今晚先完成一個小任務，再保存做法。

Run it, check it and revise once

運行一次，檢查，再修改一次

Save the useful rule for next time

把有用的規則保存下來

Download the brief, prompts and review rubric

資料頁有完整實操包

01

Research contribution

研究貢獻

A question and reproducible evidence

02

A working agent

可用 Agent

Execution, recovery and checks

03

A business result

業務成果

An output, decisions and cost records

Our role is to set direction and judge the result

人的主動權，來自決定方向與判斷結果

Use AI to take on a useful piece of work

用 AI 承擔一件完整、有用的工作

Show the result and the decisions you
made

展示成果，也展示自己的判斷

brucehuang.ai / Slides and practice materials

演講與練習資料已開放下載

AI-generated satire / AI 生成諷刺短片



Direct your own AI video

親手導演你的 AI 短片

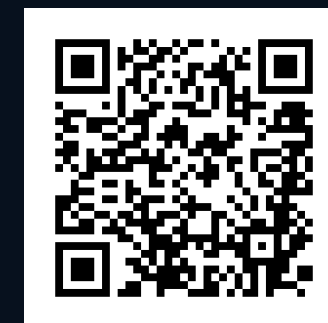
Vibe Filming

Codex + HolyCrab CLI

One idea. Two connected clips.

帶上電腦與一個產品點子

WhatsApp group



掃碼加入 HolyCrab 群

Workshop materials



掃碼開啟講義

Research

研究模型

How can we demonstrate a new capability?

怎樣證明有新能力？

Build agents

搭建 Agent

How can a system deliver reliably?

怎樣讓系統穩定交付？

Apply AI

業務應用

Which part of my work should I begin with?

我的工作從哪裏開始？

Which task would you start with?

你準備從哪一件工作開始？

Tell us the task and where it gets stuck

說說具體工作，以及最卡的一步

We can work backward from the result

從想要的成果倒推做法

brucehuang.ai / Bruce Huang

OpenClaw Asia / HolyCrab.ai