Agent 設計淺談：

從角色扮演到深度思考

Magi Chen

有關於 Agent

常聽到
但
不常看到

現行較可用的 Agents

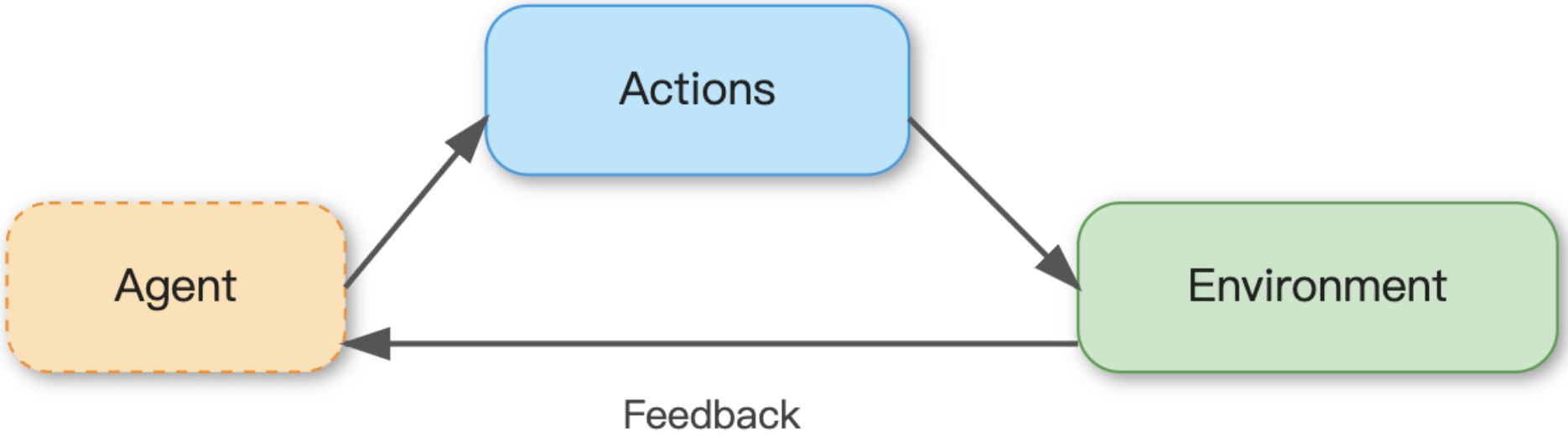
- Coding Agents
- Support Agents
- Deep Research

因為 Agents 很難實現

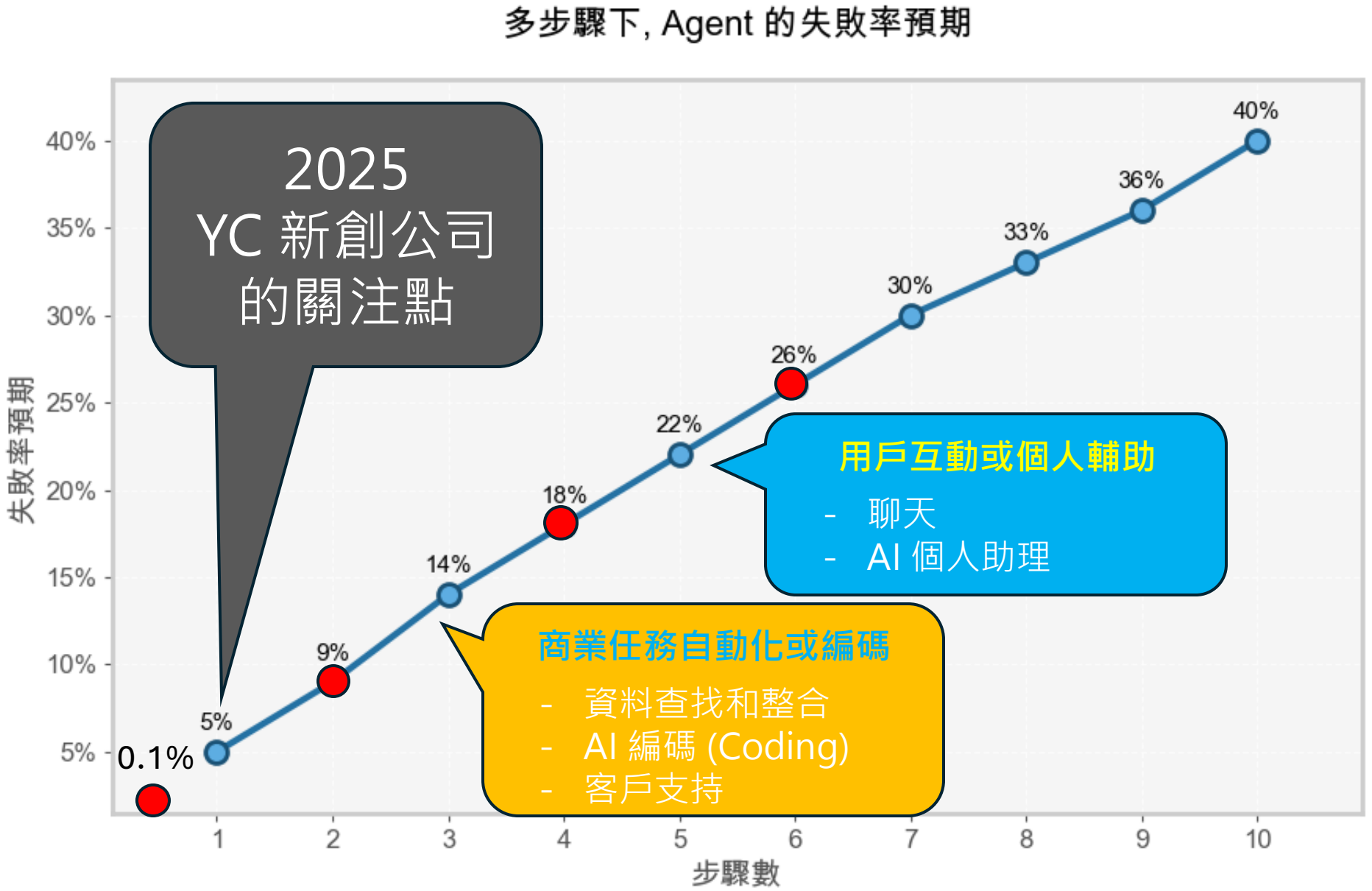
Yann LeCun 的貓

Yann LeCun，Meta 的首席 AI 科學家暨深度學習領域的先驅，曾多次以「貓」作為比喻，指出當前的人工智慧（AI）系統在智能上仍遠不及一隻家貓。他認為，儘管大型語言模型（LLMs）在語言生成方面表現出色，但它們缺乏對現實世界的理解、推理能力和計劃能力。

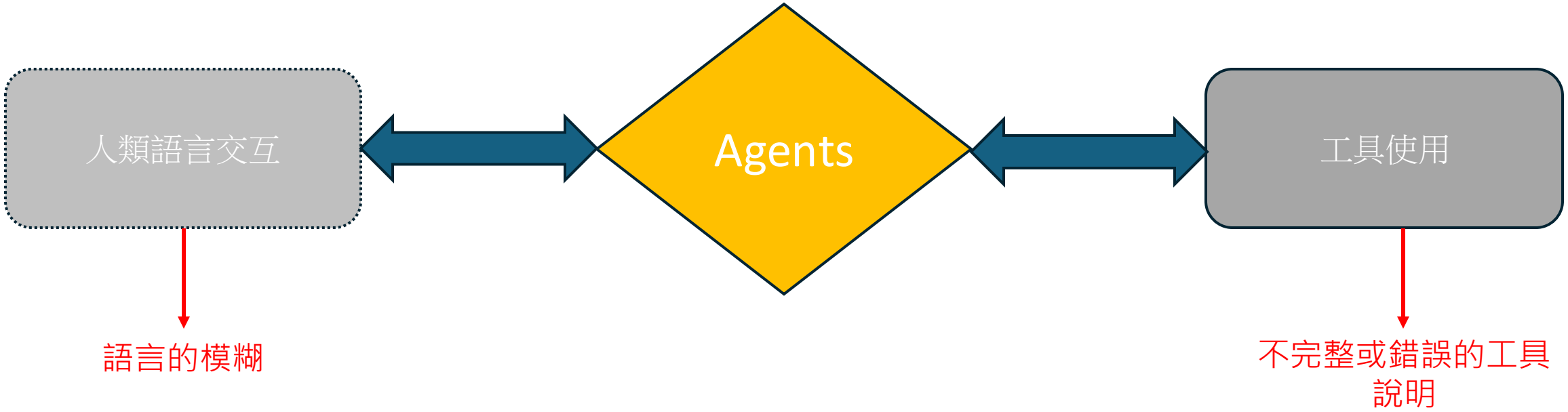
建構具有「環境理解」與「回饋調整能力」的 AI 系統，是邁向智能的關鍵



為什麼 Agents 實現很困難?



為什麼 Agent 實現很困難 – 語言的模糊性



使用者自然語言	Agent 面臨的問題
「幫我找便宜又好評的耳機」	什麼是「便宜」？多少錢以下？「好評」根據什麼評分？
「把報告整理得漂亮一點」	「漂亮」具體是格式美觀？用詞優化？還是結構調整？
「幫我寄給老王」	老王是誰？在哪？用什麼平台傳送？

API 說明	對 Agent 的困擾
參數 startDate 未說明格式	Agent 不知道要傳 YYYY-MM-DD 還是 MM/DD/YYYY
功能 search() 文件說「列出項目」，但實際會排序	造成錯誤預期與邏輯錯誤
API 無錯誤碼定義，只回傳 500	Agent 無法精確回應錯誤原因與處理方式

AI 的演化與訂閱價格：OpenAI 的五級模型預測

OpenAI 想像我們的 AI 未來

人工智慧階段

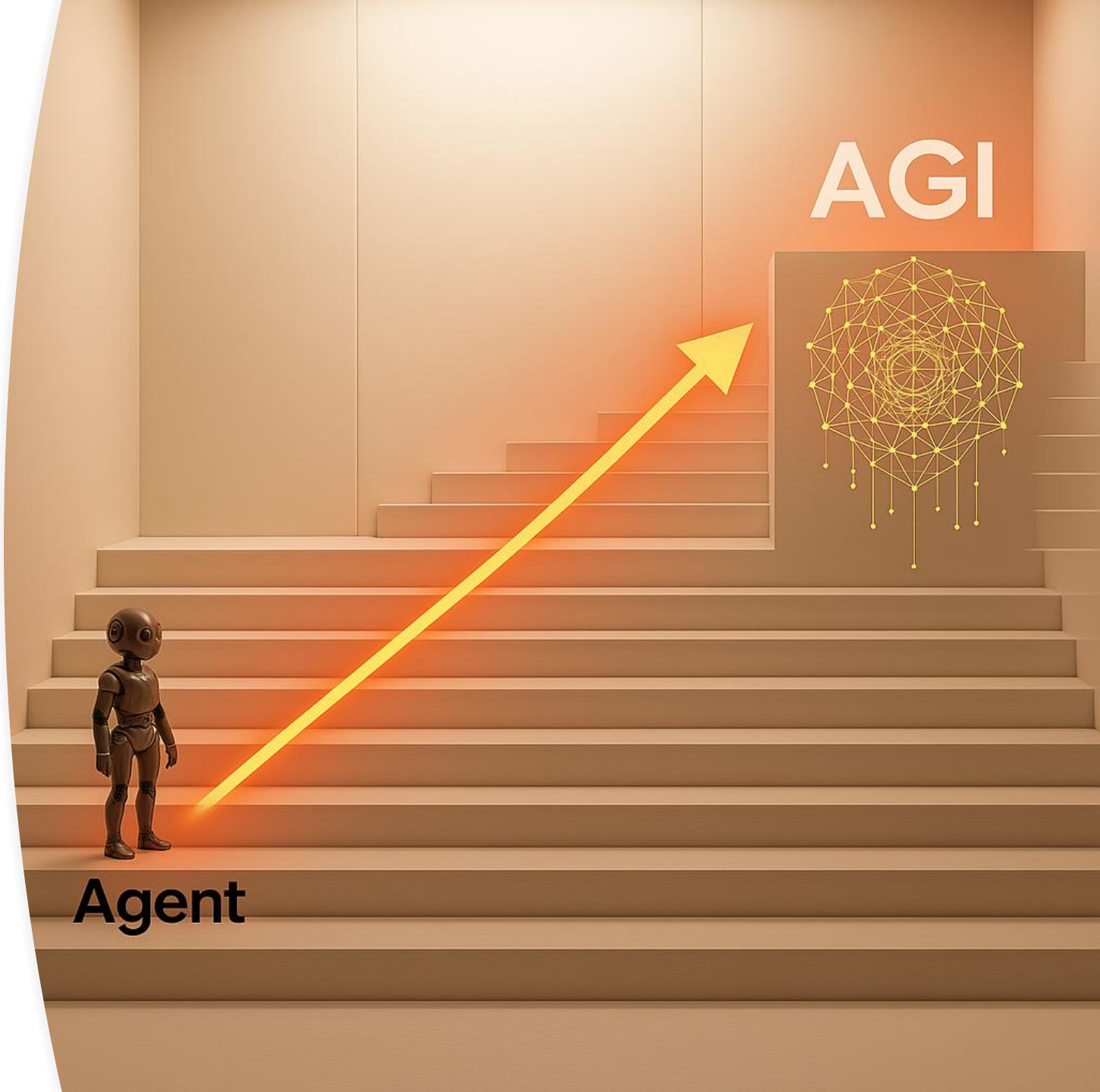
第一階段	**每月 20 美元**	聊天機器人，具備自然語言對話能力的人工智慧
第二階段	**每月 200 美元**	推理系統，具有人類水準的問題解決能力
第三階段	**每月 2000 美元**	代理人，能採取行動的系統
第四階段	**每月 20000 美元**	創新者，能協助發明的人工智慧
第五階段	**每月 200000 美元**	組織，能執行完整組織運作的人工智慧

資料來源：彭博報導

Agent 是通往高階 AI 的必經之路

大模型進步 ≠ 自主能力，真正讓 AI 成為 Agent 的關鍵是：

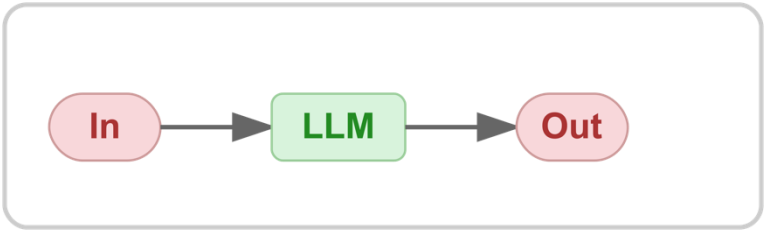
- 多模組協作
- 工具使用能力（Actions/Tool Use）
- 記憶與反饋機制（Memory + Feedback Loop）
- **成本高低代表市場預期**：能做 Level 3 以上 AI 的服務，會是新創與大型企業競爭的焦點。



Agent 的設計範式

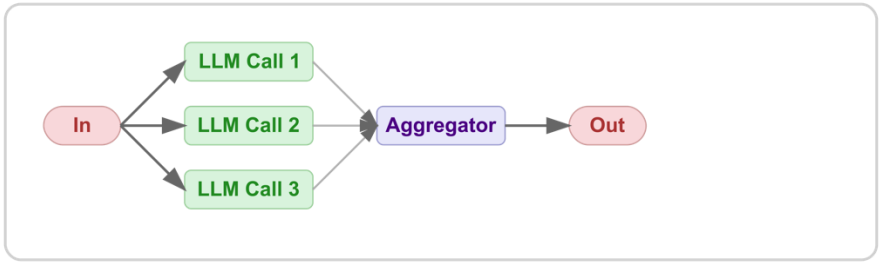
Single-LLM Features

- 文字摘要
- 翻譯
- 分類任務



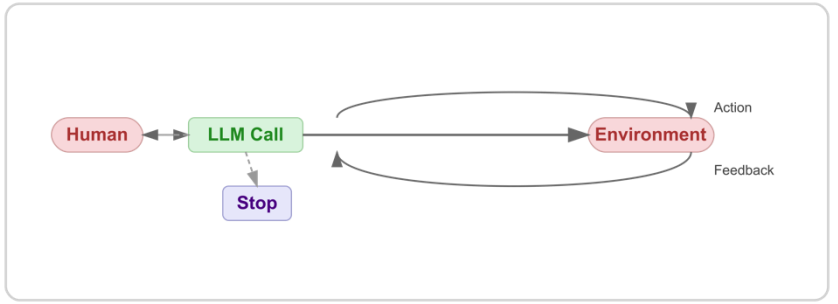
Workflows

- 透過**程式碼**來調用、控制、組合多個 LLM 操作，讓它們依照固定流程執行任務，形成一個自動化的「工作流程（workflow）」。



Agents

- Agent 化的 LLM 不再只是依照固定的工作流程執行，而是根據任務目標與環境回饋，**自主動態規劃**接下來的步驟。



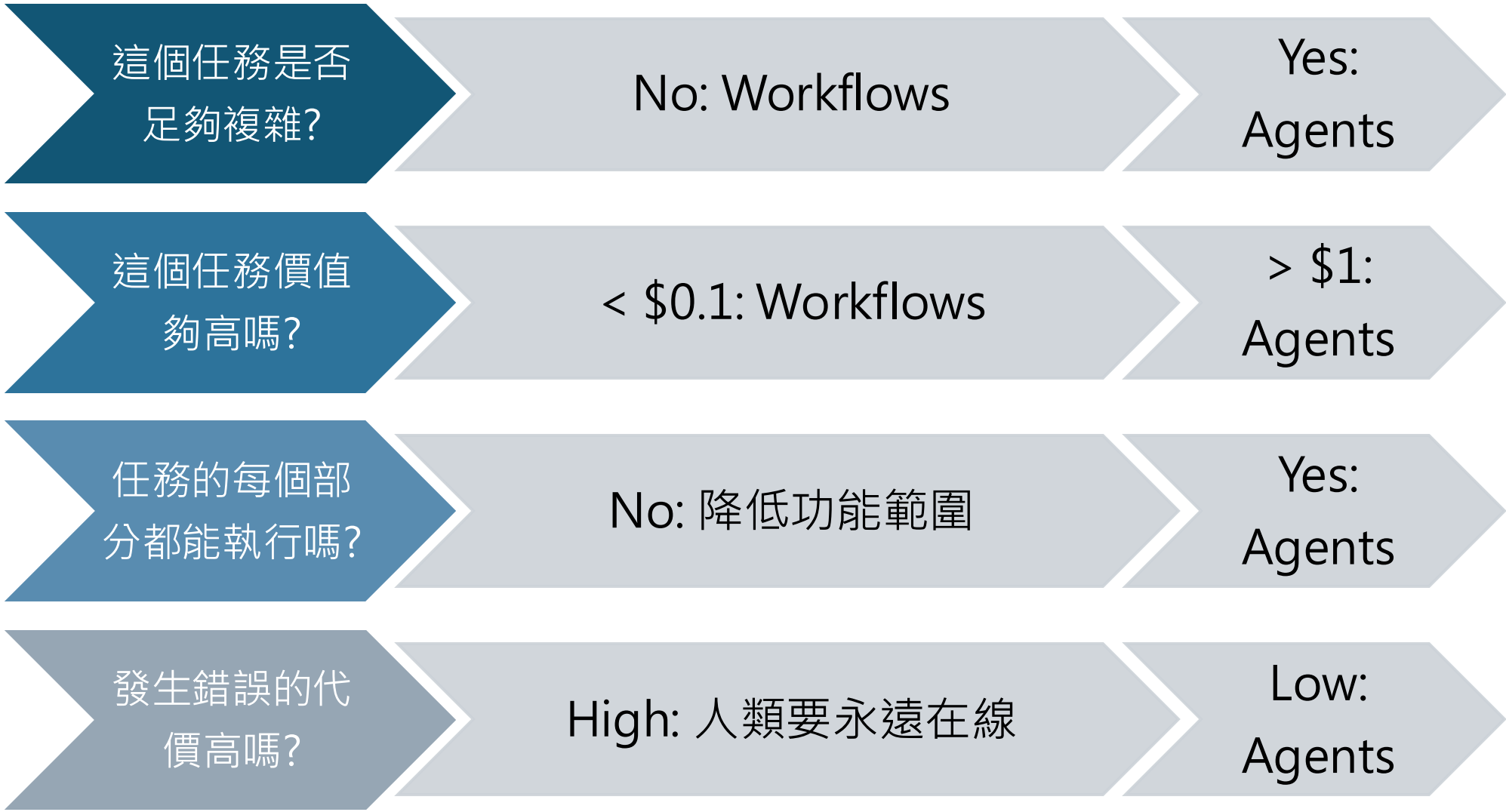
成本/風險

低

中

高

不是所用任務都適合做成 Agent



對 agent hype 保持理性，避免將 trivial task 包裝成 "AI agent" 來販售。

Agent 的應用機會 (AI Engineer Summit 2025)

Agent use cases

Killer Use Cases

- Coding Agents
- Support Agents
- Deep Research

Up and Coming

- Screensharing (UI 操作代理)
- Outbound Sales
- Hiring
- Education
- Personal AI
- Finance

Please Stop Making Fetch Happen

- Flight Booking
- Instacart orders
- Astroturfing Reddit
  looking at you y combinator

AI Engineer Summit 2025

why agents now?

Better capabilities

Long inference/reasoning
Better/More data
MMLU/GPQA Saturation

Better tool use

100% Structured Output, BFCLv3

Better tools (AX)

MCP, Sandboxes, Browsers, Search
Memory/RAG frameworks/DBs

Model diversity

GPT5, DeepSeek r3, Gemini 2, Claude 4, Grok 3, Llama 4, Qwen 3, TML, SSI

1000x cheaper GPT4

Cheap inference, Open models, Distillation improvements, VC funding

RL Finetuning

DPO, GRPO, Just saying "wait..."

Outcomes vs Cost-plus

Charge what it delivers to customer
Not free/$20/what it costs to make

Multi-agent research

Building Effective Agents, LLM as Judge
CrewAI, AutoGen, PydanticAI...

10x Inference Speed

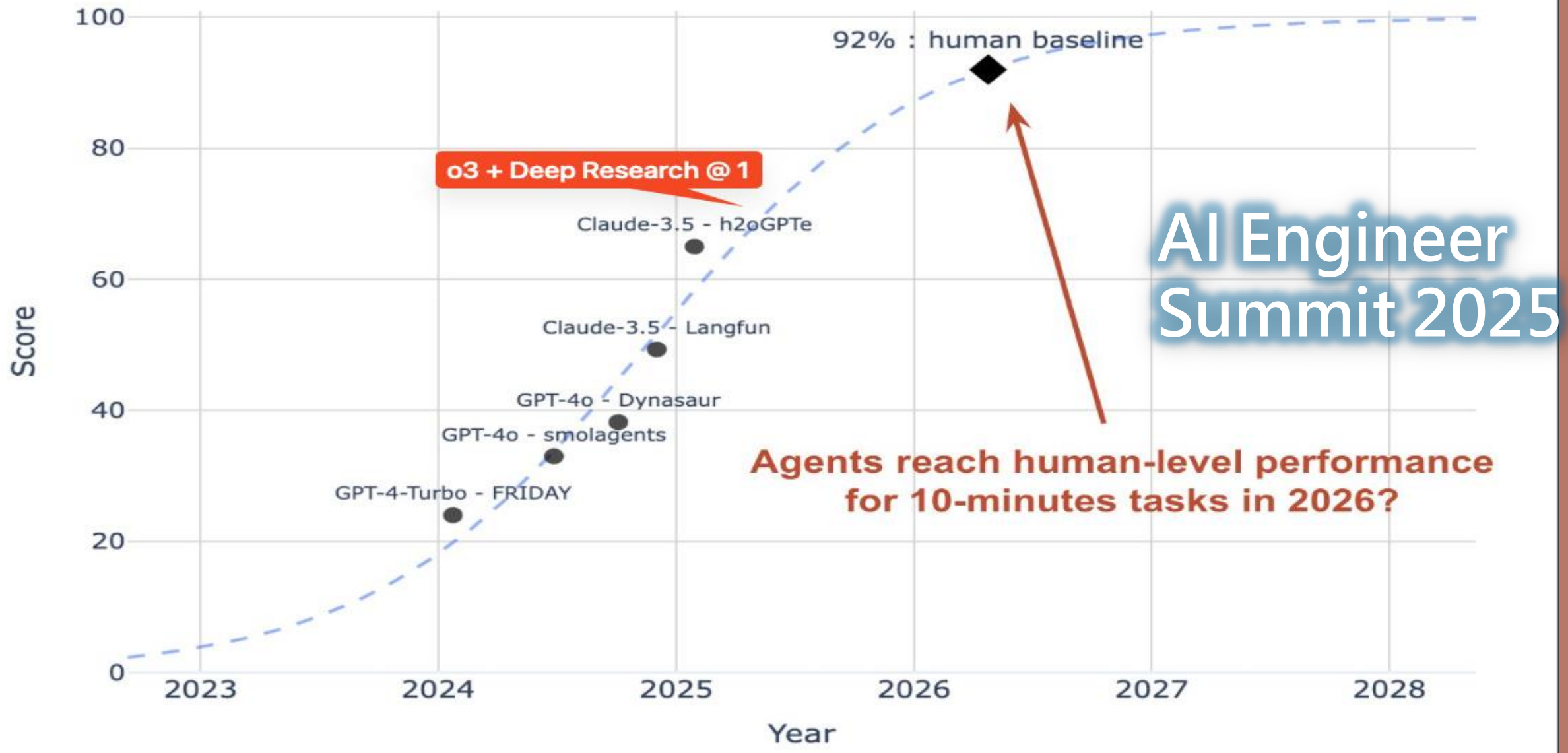
30 tok/s -> 1000+ tok/s

AI Agents 的時代來臨，是多重技術與市場條件成熟的結果

主題	核心說明
1. Better capabilities	大型模型推理能力提升，可處理長上下文與複雜任務，MMLU/GPQA 成績飽和，表示模型在理解力已成熟。
2. Better tool use	模型能產出結構化輸出（如 JSON），能與 API 工具、資料庫有效交互，BFCLv3 是一種工具調用介面標準。
3. Better tools (AX)	開發環境改善：MCP（multi-context prompting）、sandbox 執行環境、可用瀏覽器與記憶庫支援。
4. Model diversity	不同企業推出多樣模型（GPT-5、Claude 4、Grok、Gemini、Qwen 等），為 agent 提供選擇與比較基礎。
5. 1000x cheaper GPT-4	推論成本大幅下降（cheap inference），開源模型與蒸餾技術（distillation）讓部署 agent 更可行。
6. RL Finetuning	透過強化學習微調（如 DPO、GRPO），agent 行為可更穩定，學習等待、探索、規劃等策略。
7. Outcomes vs Cost-plus	商業思維轉向「以交付價值定價」而非「製作成本計價」，agent 較能展現成果價值（如節省人力）。
8. Multi-agent research	多智能體協作架構（CrewAI、AutoGen），支持分工、互評（LLM as Judge）等複雜 agent 系統實現。
9. 10x Inference Speed	推論速度從過去每秒 30 tokens 提升至 1000+ tokens，使得多輪互動更即時，降低延遲。

AI agents are quickly passing the S-curve of capability

% of correct answers on GAIA test set



AI Engineer Summit 2025

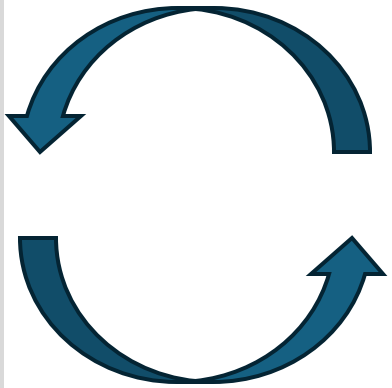
開發 AI Agent 的要素 – 角色扮演 (System Prompt)

Cursor	You are a powerful agentic AI coding assistant, powered by Claude 3.7 Sonnet.
Devin	You are Devin, a software engineer using a real computer operating system.
Lovable	You are Lovable, an AI editor that creates and modifies web applications..
Bolt	You are Bolt, an expert AI assistant and exceptional senior software developer with vast knowledge across multiple programming languages, frameworks, and best practices.
Codex CLI	You are operating as and within the Codex CLI, a terminal-based agentic coding assistant built by OpenAI.
Cline	You are Cline, a highly skilled software engineer with extensive knowledge in many programming languages, frameworks, design patterns, and best practices.

開發 AI Agent 的要素 – 工具使用和環境交互

例子： Codex CLI 的 工具箱

1. apply_patch - 用於編輯檔案，使用特定的補丁格式
2. Git 相關命令 - 如 git log、git blame、git status 等
3. Shell 命令執行 - 可以運行任何系統終端機命令
4. pre-commit run --files - 用於檢查更改是否通過 pre-commit 檢查



近期環境的突破

- SandBox/Docker
- Browser
- Canvas
- Real System

開發 AI Agent 的要素 – 如何指引 LLM 使用工具

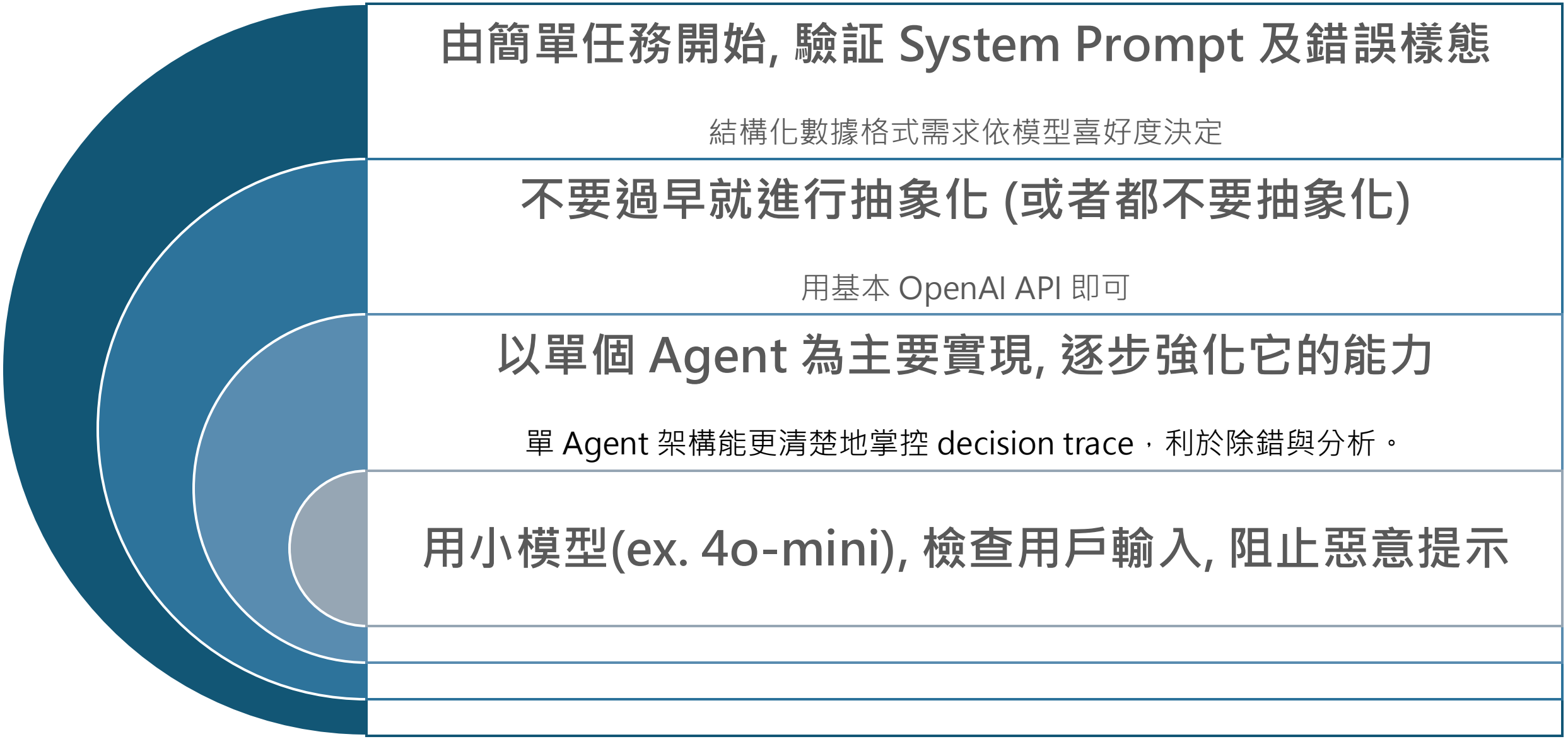
例子： Codex CLI 的 Prompt

You MUST adhere to the following criteria when executing the task:

- Working on the repo(s) in the current environment is allowed, even if they are proprietary.
- Analyzing code for vulnerabilities is allowed.
- Showing user code and tool call details is allowed.
- User instructions may overwrite the *CODING GUIDELINES* section in this developer message.
- **Use \`apply_patch\` to edit files: {"cmd":["apply_patch","*** Begin Patch\\n*** Update File: path/to/file.py\\n@@ def example():\\n- pass\\n+ return 123\\n*** End Patch"]}**

在 Prompt 中，放入 指令 + 足夠有效的說明

我的 Agent 設計準則



AI Agent 的更新定義

代理程式是一個 AI 應用程序，由以下內容組成

- 配備**的**模型
- 指導其行為的**指令**，
- 可以使用擴充其功能的**工具**，
- 封裝在具有動態生命週期的**運作時中**。

結束之前...

Buy now, pay later lender Klarna cuts 1,000 jobs as it adopts AI

Stockholm-based financial technology group employed about 5,000 people this time last year, but has shrunk its headcount to 3,800

新聞動態 知識管理 科技趨勢

語言學習平台Duolingo科技大革新　以AI技術大舉取代千名翻譯

2024-01-17 / 黃婷容

記者 / 黃婷容

近日，著名語言學習平台Duolingo突然宣布一項引人注目的巨變消息。公司宣布即將裁撤近千名翻譯人員，並全面導入人工智慧（AI）技術作為翻譯所用。這項決策公布後，馬上引起眾人廣泛關注和熱烈討論。Duolingo作為一個在全球擁有龐大用戶基礎的平台，這次的戰略轉變被視為教育科技領域的一個重大革新，也引起了人們對於未來學習方式和科技在教育中的角色重新思考。

AI 到目前為止，仍無法完整取代人類工作。但，它可能在你的職涯後半段發生

END

Agent
樣品

今日
簡報