iThome
SURVEY RESEARCH

iThome大調查系列分享｜KubeSummit 2026

從 Cloud Native 到 AI Native，當維運對象從容器變成 Agent

王宏仁 iThome副總編輯 ray@mail.ithome.com.tw

© iThome 2026

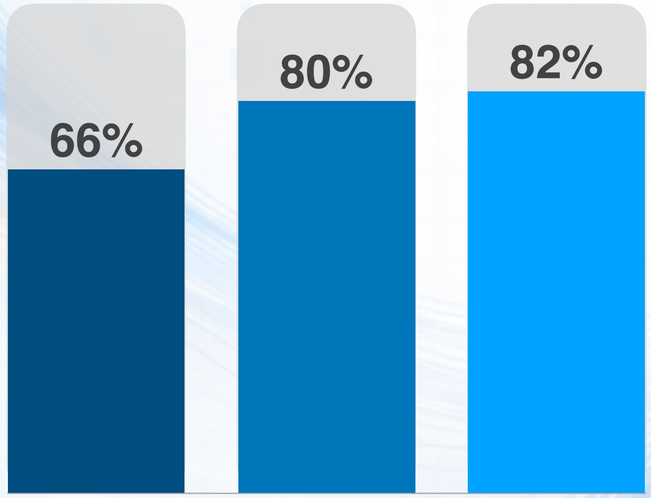
本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示—非商業性—禁止改作 4.0國際]授權。
可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢

iThome SURVEY RESEARCH

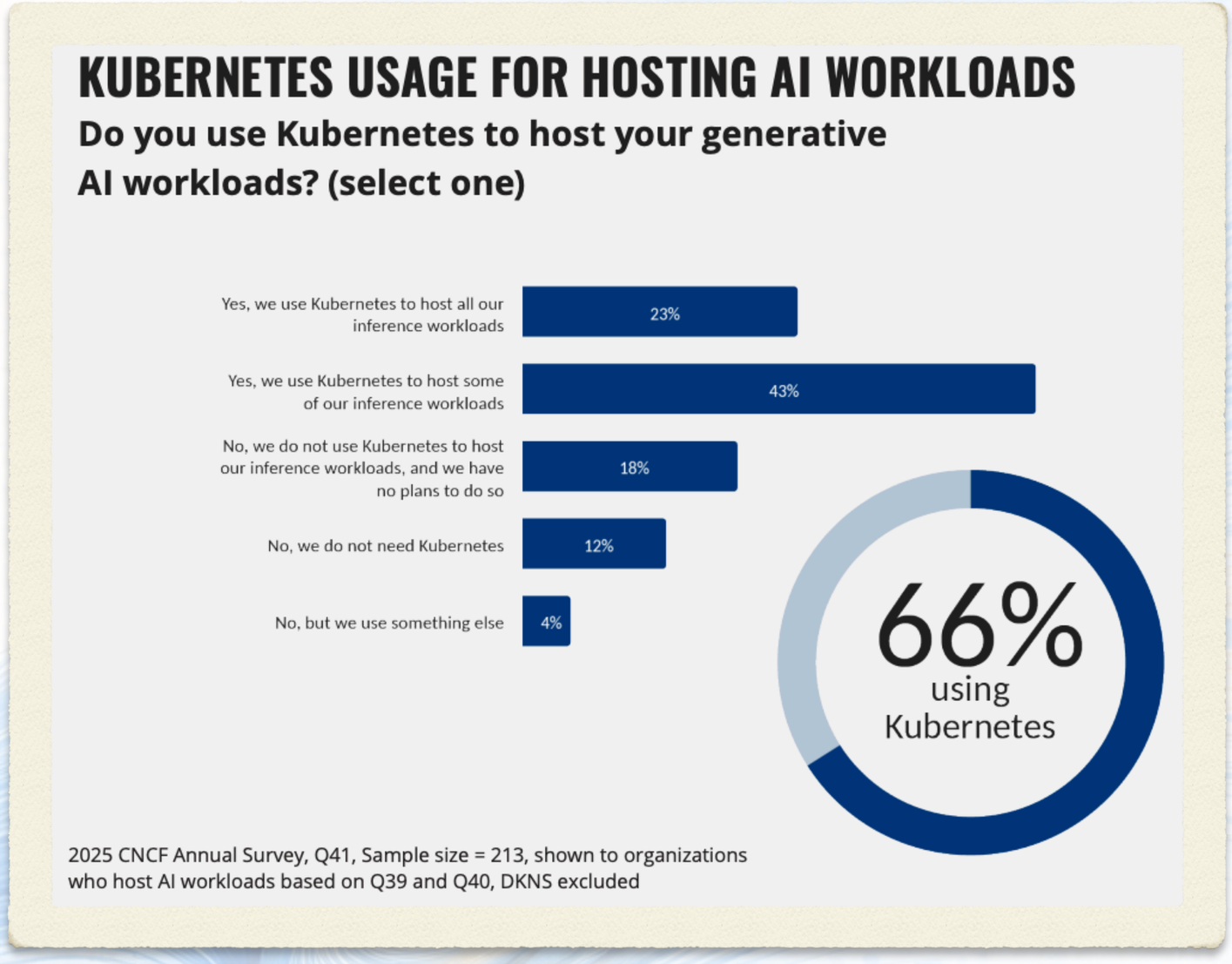
多少AI Workload跑在K8s上？

全球K8s採用率達到82%

66%企業的AI負載跑在K8s上



資料來源： CNCF 2025年度大調查，CC BY-ND 4.0



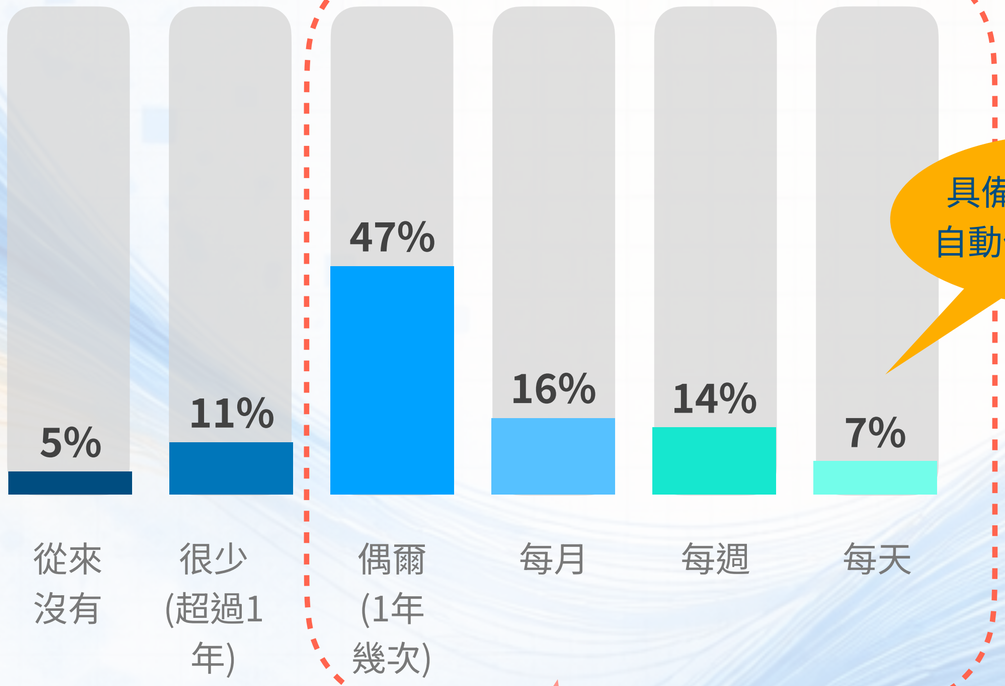
© iThome 2026

iThome SURVEY RESEARCH

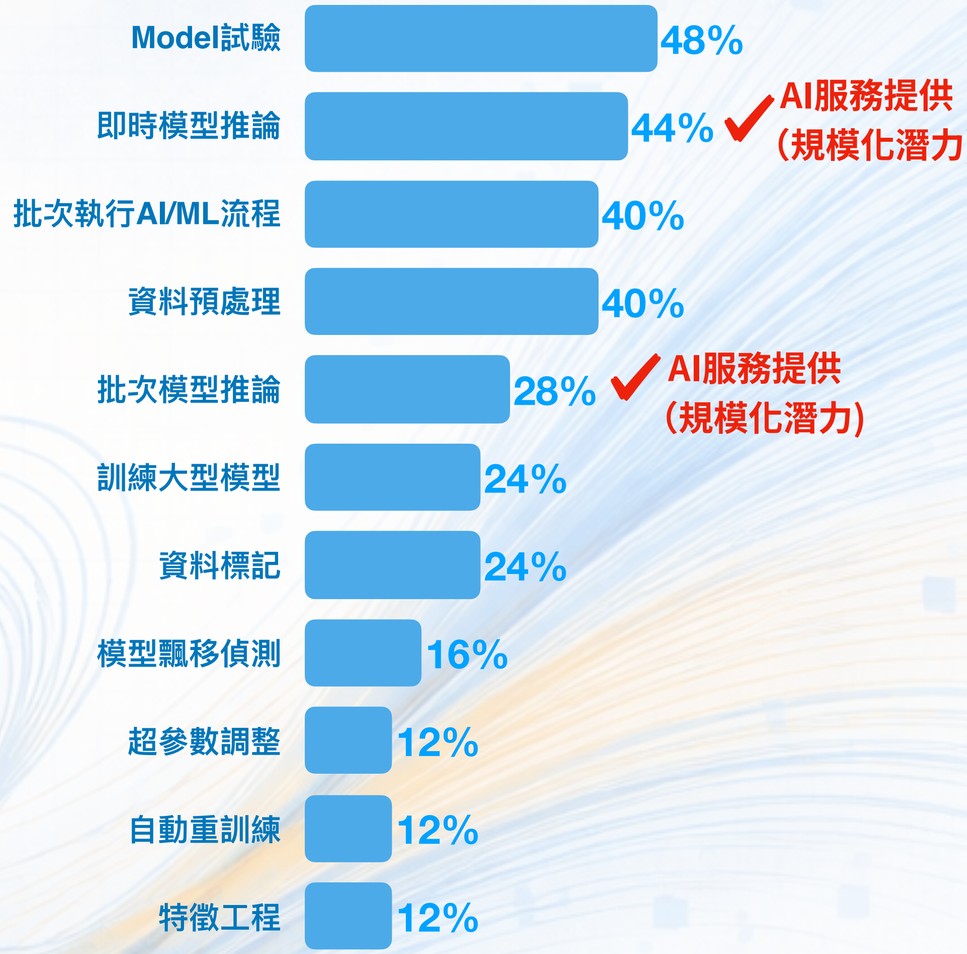
AI Model部署頻率？用來跑什麼AI負載？

全球企業AI Model部署頻率

47%每年幾次，7%天天部署



全球企業AI/ML工作負載用途



資料來源： CNCF 2025年度大調查，CC BY-ND 4.0

© iThome 2026

“每一個員工都有自己的個人化AI助手，每個流程都由AI代理驅動，每一次顧客體驗都搭配AI禮賓服務”

Agent規模化課題

JPMorgan執行長 傑米·戴蒙

LLM Suite/25萬名員工

- 2023年底推出，每兩個月改版，2025年底時半數人天天用。轉向Agentic AI架構可以支援企業內部複雜、多步驟的自主型AI代理。
- 個人助理，員工可以自己定義助理的能力，包括角色描述、角色可用工具、可用知識

1萬人理專個人化AI教練

- 2025年推出，到2026年2月有8千人活躍用戶。一年回答了1百萬個問題，提供3萬個理財點子。
- 內建50個AI功能和代理式流程，在對話中就能提供各種理財分析圖表和工具

AI研究Agent:ASK David

- 2025年中推出，可提供量身打造的AI投資報告
- 整合數十年的結構化數據和多模態資料：文件，郵件，會議記錄，內部影音檔和研究資料
- 媲美真人投資研究團隊的Agent研究團隊

資料來源：iThome 2026 CIO & CISO 大調查，2026年

調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份

本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示—非商業性—禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。

© iThome 2026

Example: Why was a fund terminated?

J.P.Morgan

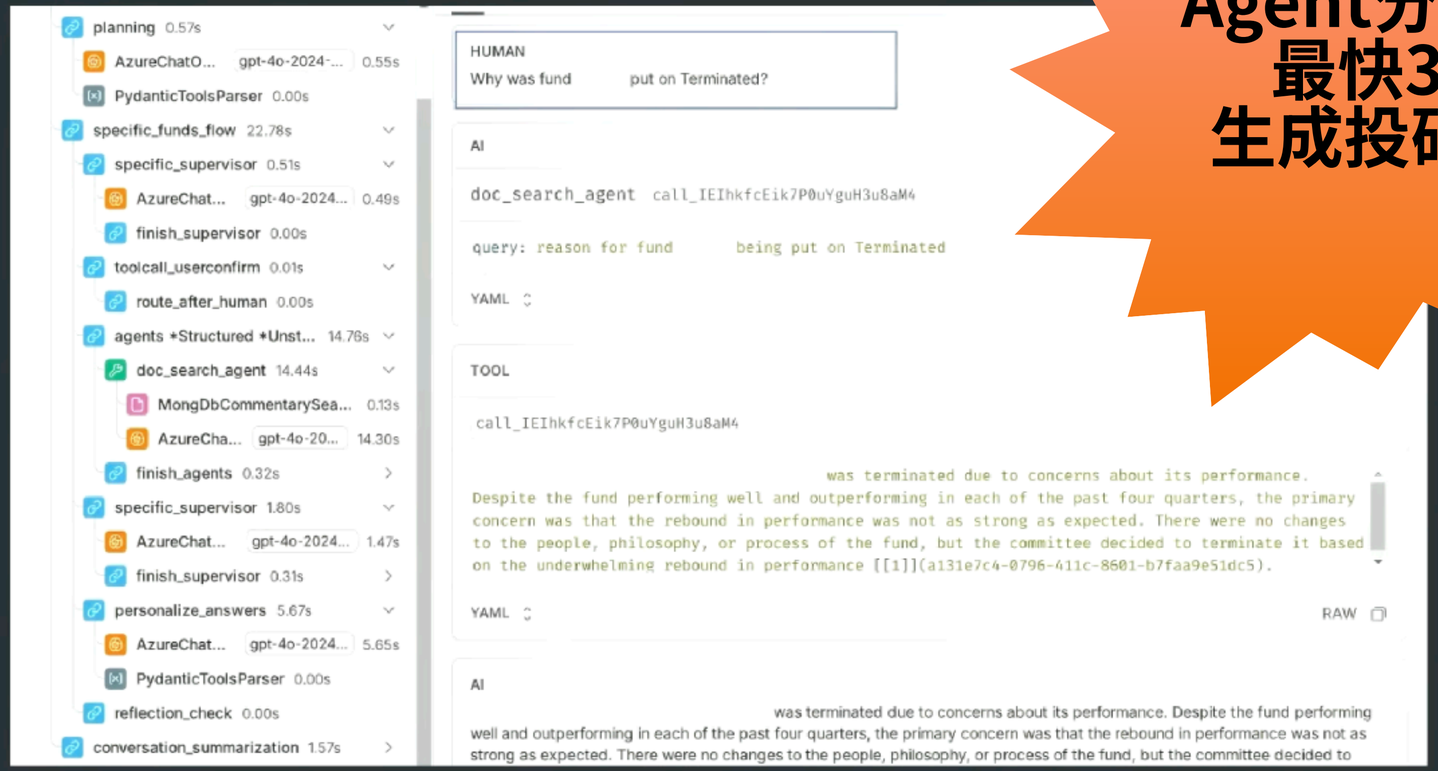
圖片來源：J.P.Morgan

Planning
- Determines the intention
- Invokes the corresponding subgraph

Subgraph - Specific Funds Flow
- delegates to doc_search_agent

Doc Search Agent
- Conducts mongoDB search to retrieve relevant information

Supervisor
- Personalization based on the role of human
- Reflect to ensure the accuracy and relevancy of answers
- Summarize conversation, update memory, and return answer



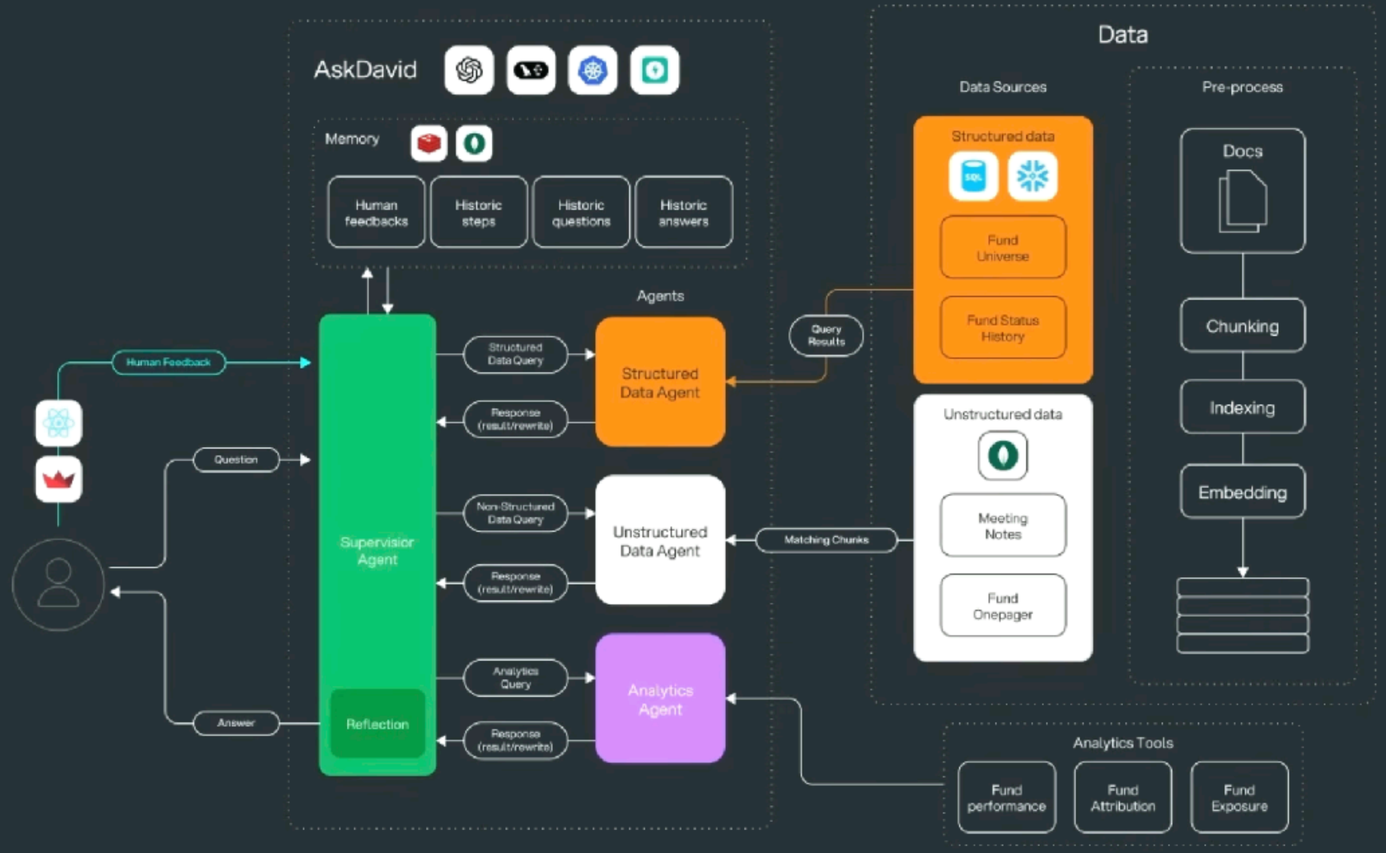
Agent分析團隊
最快30秒
生成投研簡報

資料來源：iThome 2026 CIO & CISO 大調查，2026年
調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份
本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示－非商業性－禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。
© iThome 2026

The Approach

J.P.Morgan

圖片來源：
J .P.Morgan



Supervisor Agent:

- Acts as the orchestrator
- It receives questions from users and determines a plan to provide answers
- This may involve engaging one or more of the specialized sub-agents, depending on the nature of the query
- It incorporates human in the loop to ensure the accuracy and reliability
- It also has access to long term memory

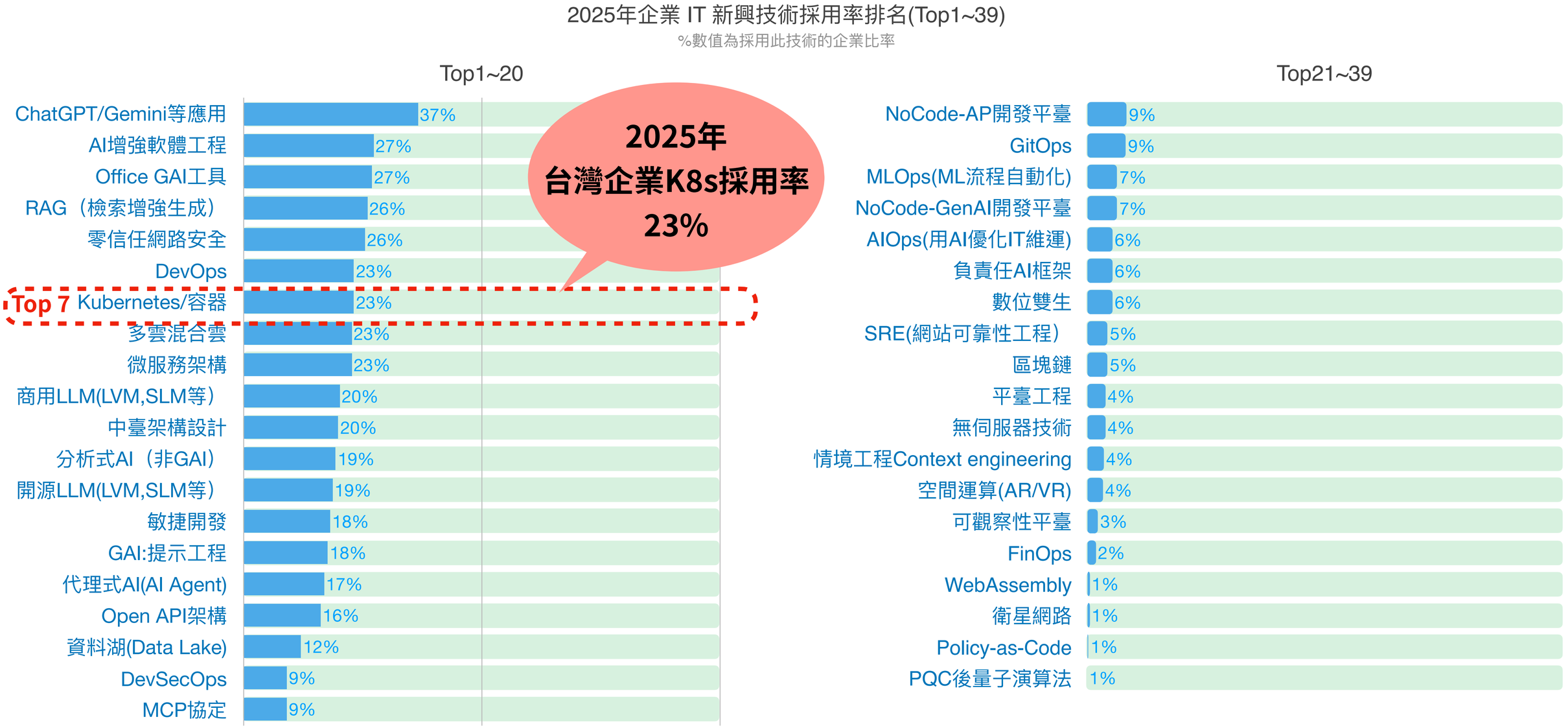
Structured Data AI Agent:
This agent translates natural language queries into SQL or GraphQL to retrieve answers from databases or APIs, and then uses a large language model (LLM) to summarize the results.

Unstructured Data AI Agent:
This RAG agent utilizes a vector database and hierarchical index retrieval to store documents and search for relevant information chunks, followed by summarization using an LLM.

Analytics Agent:
ReAct agent with access to analytics APIs as tools.

資料來源：iThome 2026 CIO & CISO 大調查，2026年
調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份
本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示—非商業性—禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。
© iThome 2026

2025企業 IT 新興技術採用率排行榜 Top 39



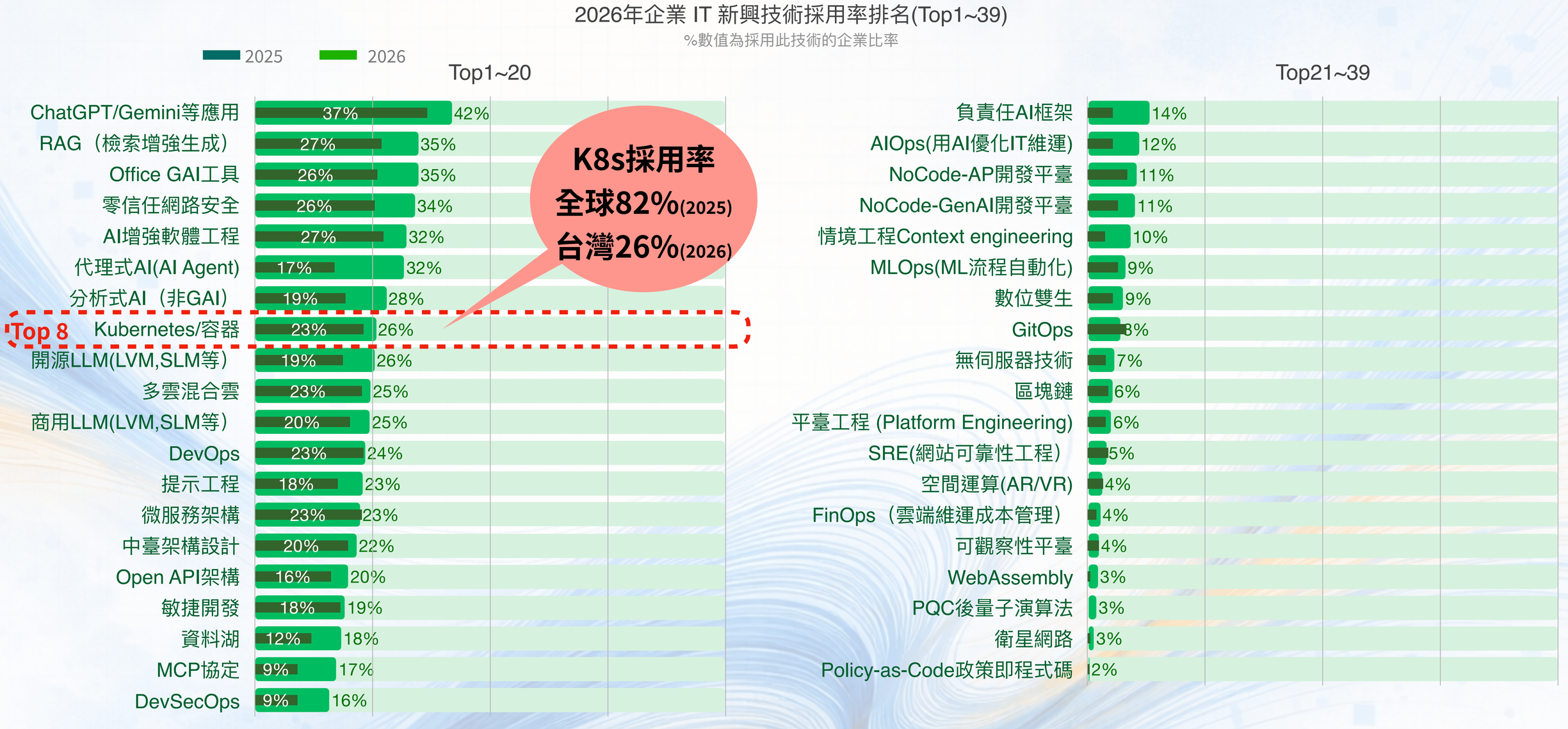
資料來源：iThome 2026 CIO & CISO 大調查，2026年

調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份

本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示—非商業性—禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。

© iThome 2026

2026企業 IT 新興技術採用率排行榜 Top 39

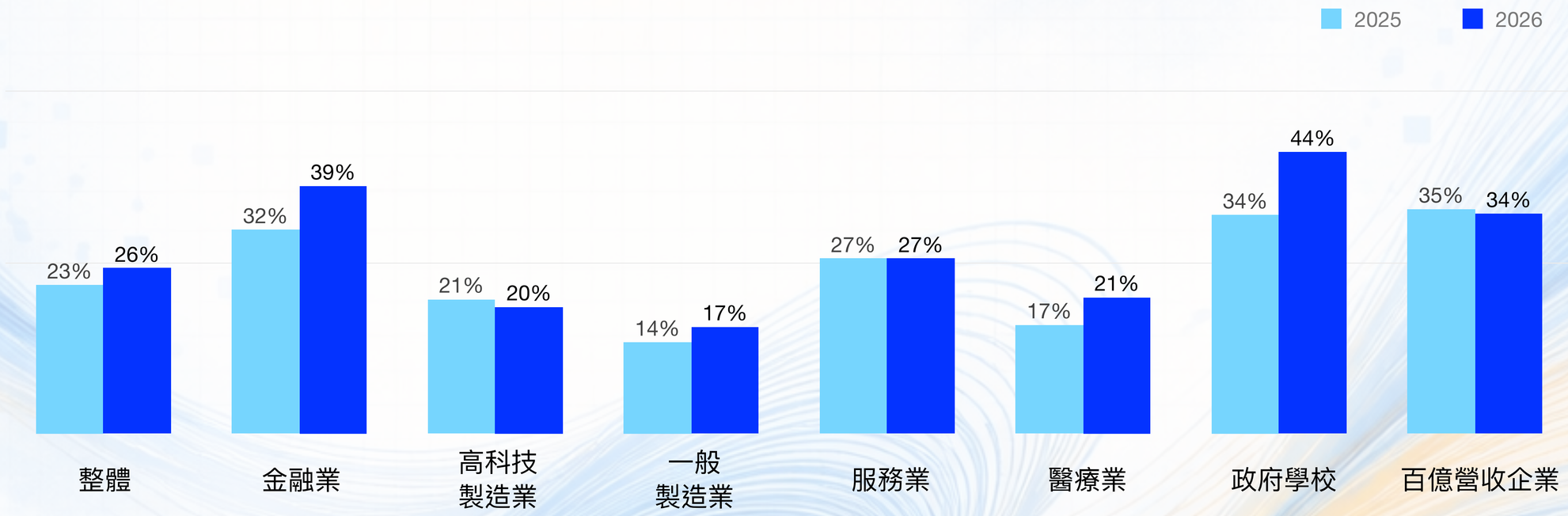


© iThome 2026

K8s採用率緩慢增加，政府學校最積極，百億企業和高科技反而些微下滑

雲原生核心：K8s容器技術採用率年度比較

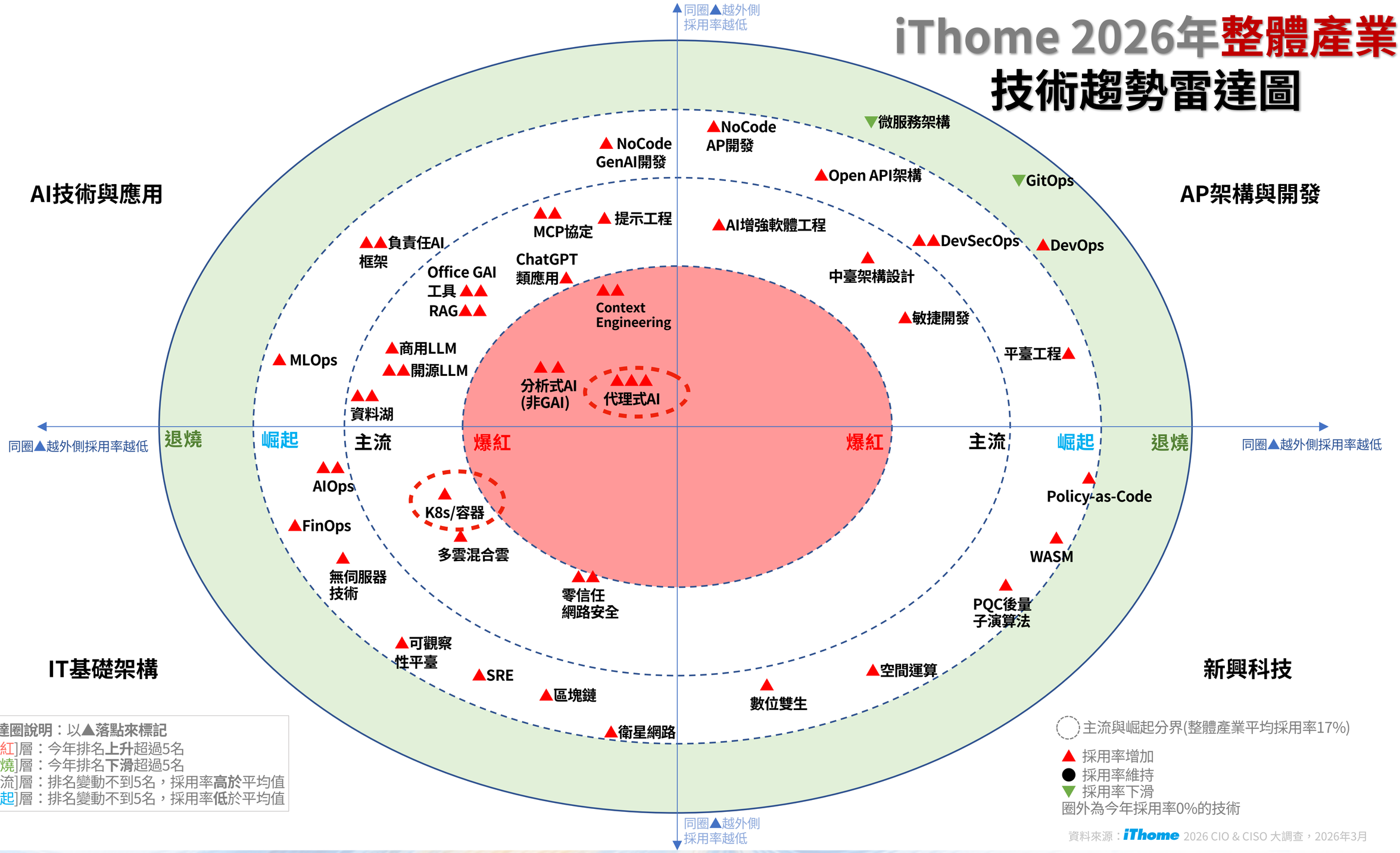
%數值為該產業在該年的採用率



	整體	金融業	高科技製造業	一般製造業	服務業	醫療業	政府學校	百億營收企業
採用率增減幅度	+3%	+7%	-1%	+3%	0%	+4%	+10%	-1%

© iThome 2026

iThome 2026年整體產業技術趨勢雷達圖



雷達圈說明：以▲落點來標記
[爆紅]層：今年排名**上升**超過5名
[退燒]層：今年排名**下滑**超過5名
[主流]層：排名變動不到5名，採用率**高於**平均值
[崛起]層：排名變動不到5名，採用率**低**於平均值

主流與崛起分界(整體產業平均採用率17%)

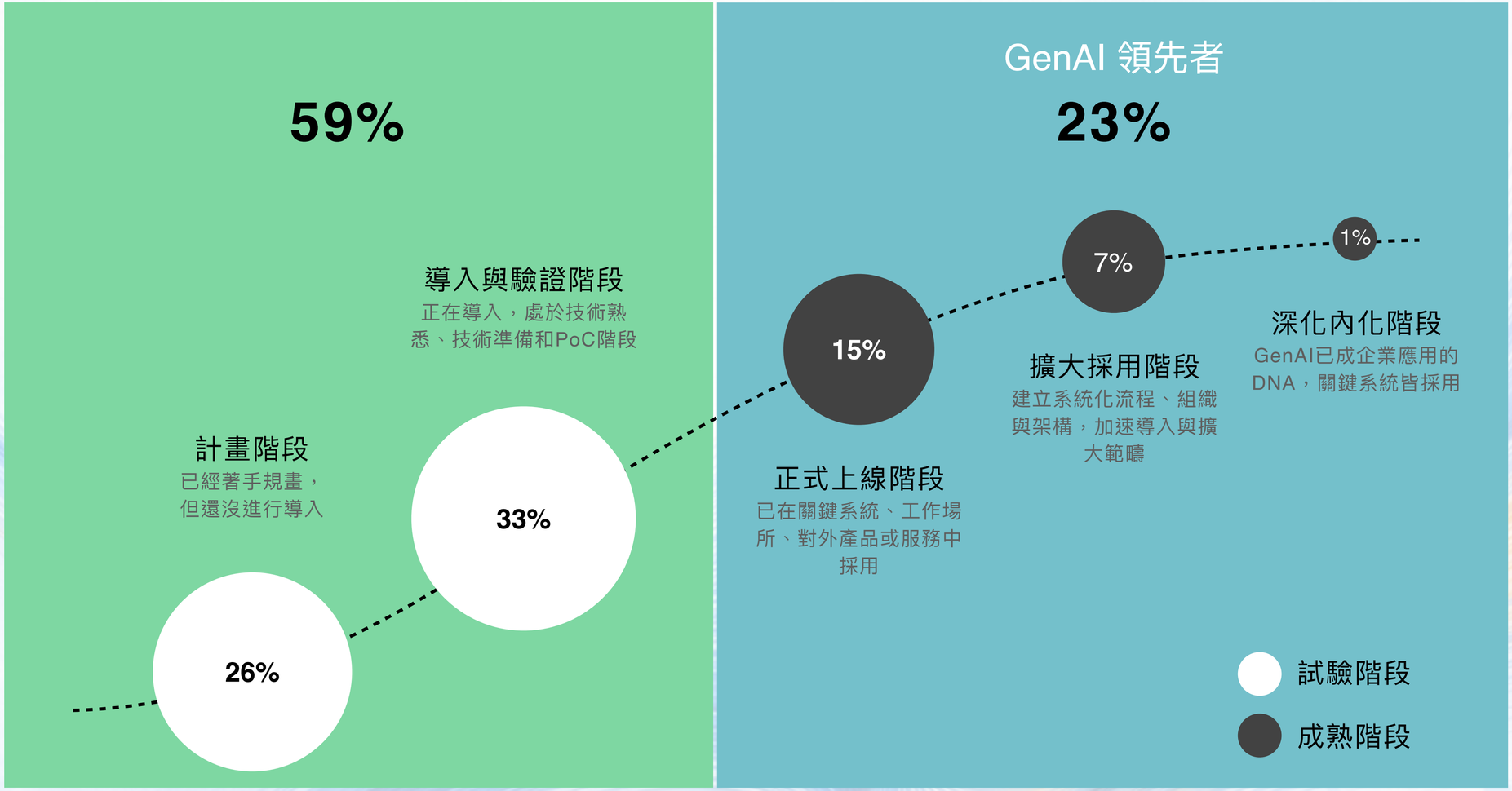
▲ 採用率增加
● 採用率維持
▼ 採用率下滑
圈外為今年採用率0%的技術

資料來源：iThome 2026 CIO & CISO 大調查，2026年3月

企業生成式 AI 應用成熟度分析顯示，GenAI 領先者的比率已達23%

企業生成式 AI 應用成熟度分布

%數值為該成熟度階段的企業比率，如33%的企業組織處於生成式 AI 導入與驗證階段



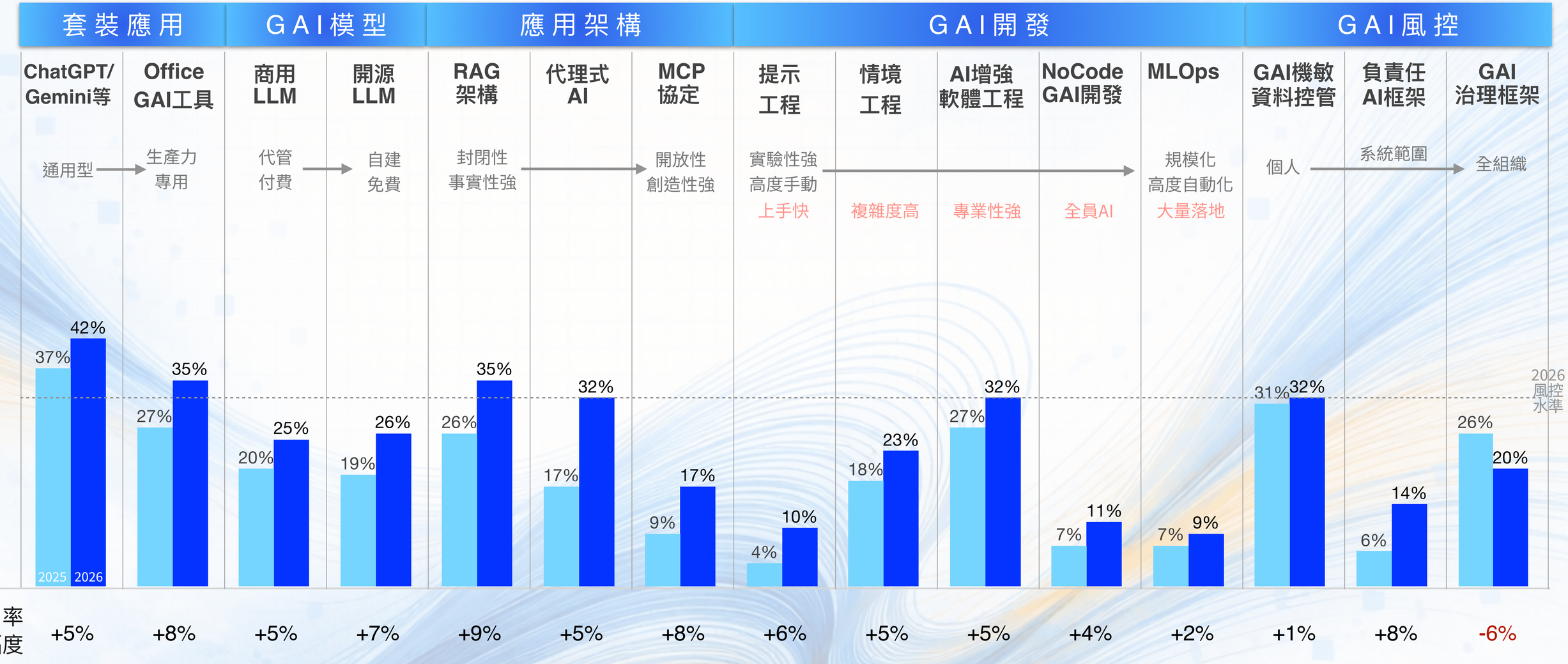
參考Gartner的AI發展成熟度模型，分成7個成熟度階段：「計畫階段」「導入與驗證階段」「正式上線階段」「擴大採用階段」「深化內化階段」「有需要但沒有計畫」及「不需要或不知道」。處於正式上線、擴大採用與深化內化階段的企業歸類為Gen AI領先者。

© iThome 2026

整體企業：代理式AI和RAG採用率大增，17%企業要用MCP

整體企業2026 GenAI採用策略的變化

%數值為該產業採用率



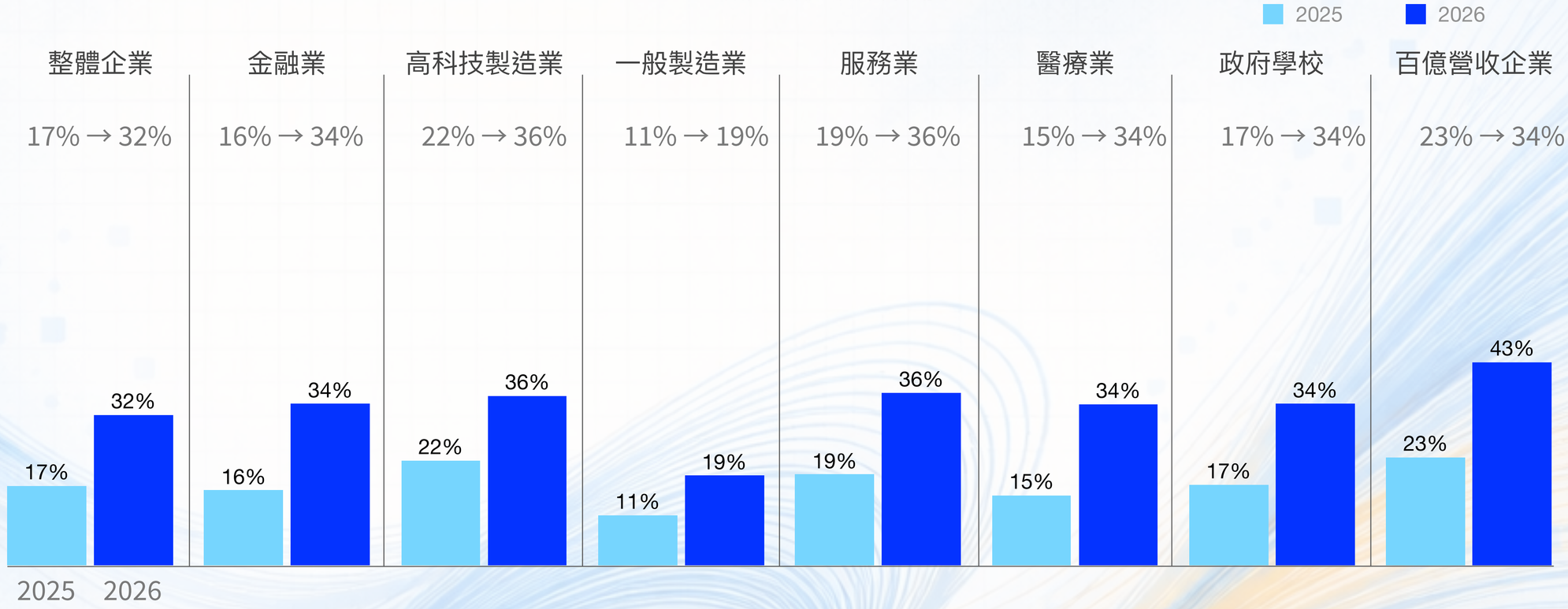
註：代理式AI採用率以企業自行採用的代理式AI技術為主，不含廠商產品內建AI代理功能或工具

© iThome 2026

各產業今年超積極擁抱代理式AI技術，百億營收企業最積極

各產業代理式AI技術採用趨勢

%數值為該產業該年採用率



	整體企業	金融業	高科技製造業	一般製造業	服務業	醫療業	政府學校	百億營收企業
採用率增減幅度	+15%	+18%	+14%	+8%	+17%	+19%	+17%	+11%

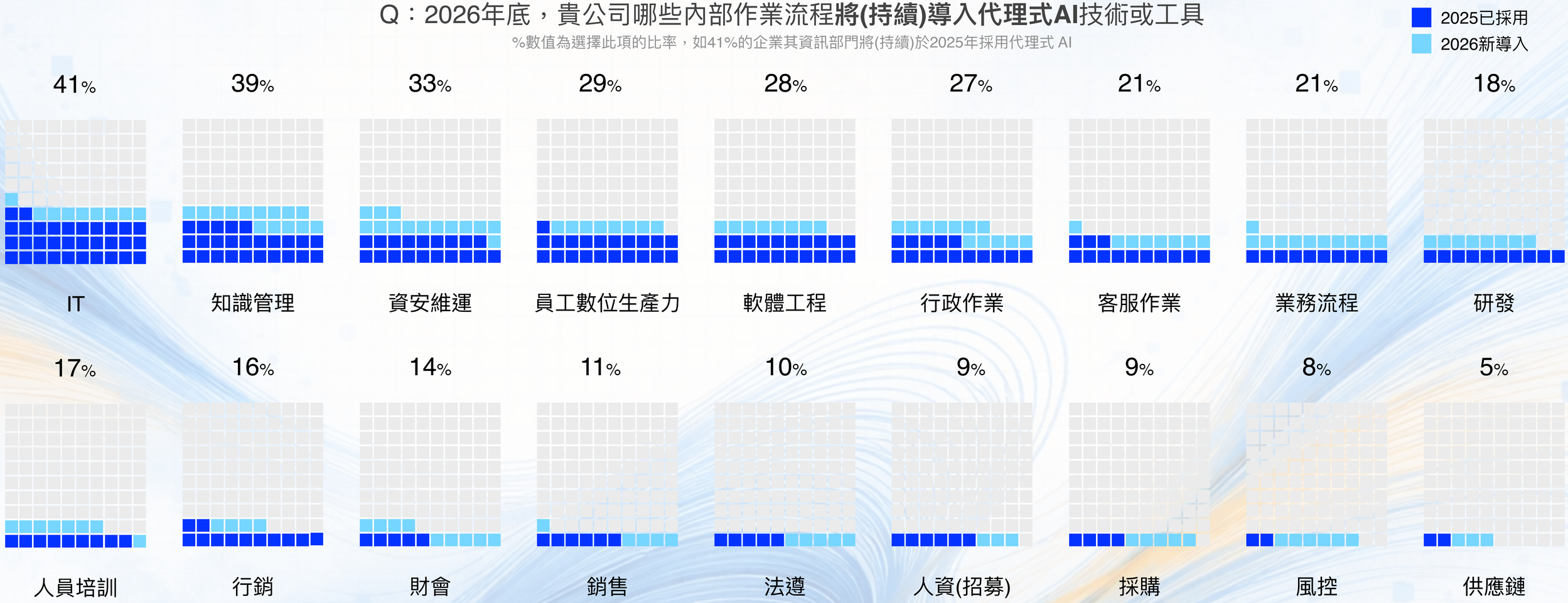
註：代理式AI採用率以企業自行採用的代理式AI技術為主，不含廠商產品內建AI代理功能或工具

© iThome 2026

AI代理開始擴散，從AI助手邁向日常作業和企業營運

Q：2026年底，貴公司哪些內部作業流程**將(持續)導入代理式AI**技術或工具

%數值為選擇此項的比率，如41%的企業其資訊部門將(持續)於2025年採用代理式 AI



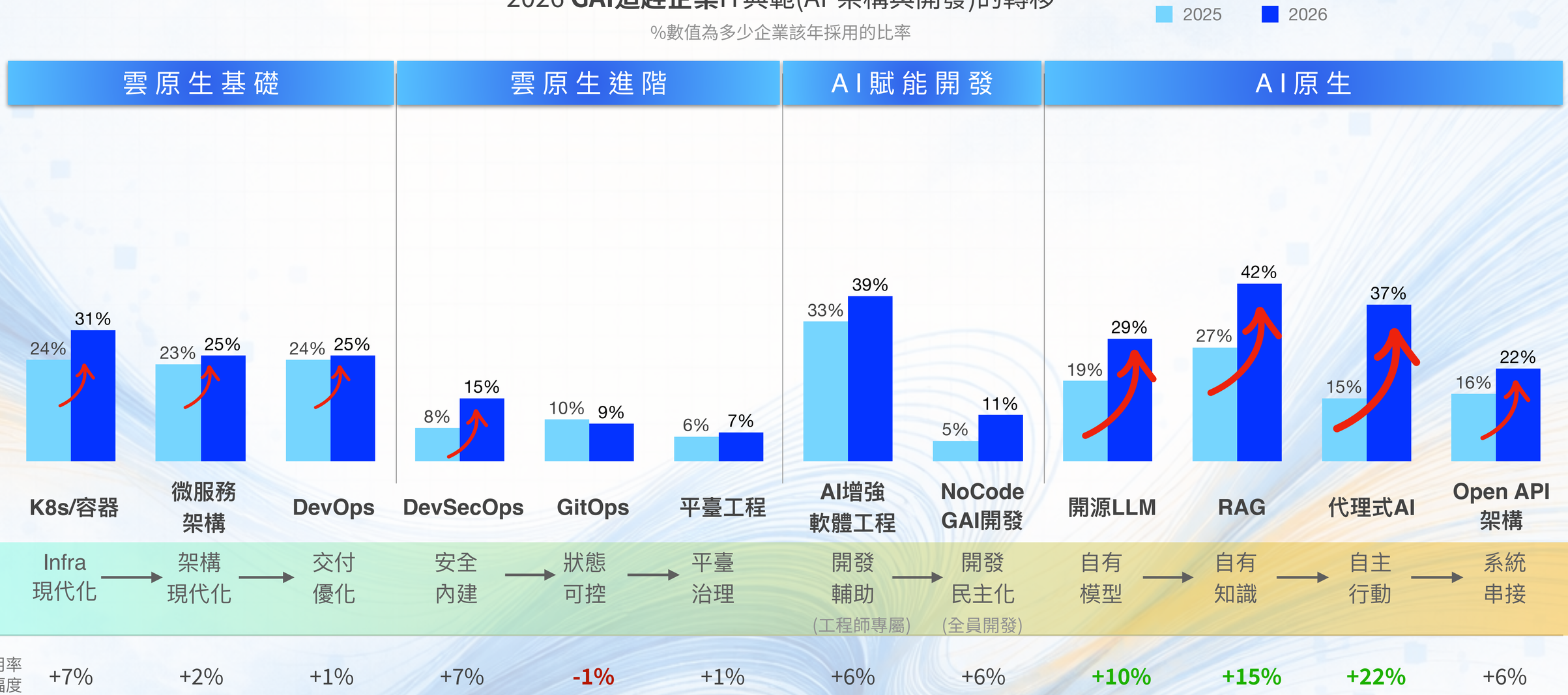
註：代理式AI技術或工具涵蓋企業自建代理式AI應用和廠商提供的AI Agent工具，通稱AI代理

© iThome 2026

計畫和PoC中的GAI追趕企業，持續發展雲原生，但更積極擁抱AI原生

2026 GAI追趕企業IT典範(AP架構與開發)的轉移

%數值為多少企業該年採用的比率



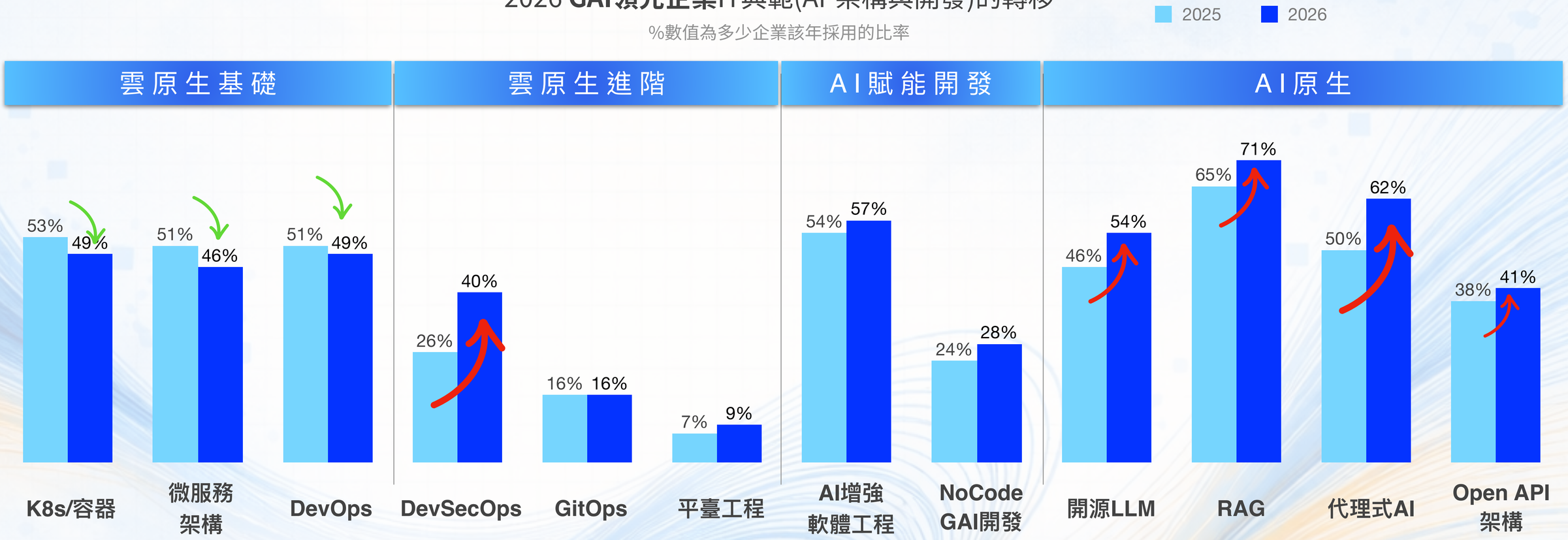
註：代理式AI採用率以企業自行採用的代理式AI技術為主，不含廠商產品內建AI代理功能或工具

© iThome 2026

GAI領先企業：雲原生基礎採用率下滑，AI原生成為新焦點

2026 GAI領先企業IT典範(AP架構與開發)的轉移

%數值為多少企業該年採用的比率

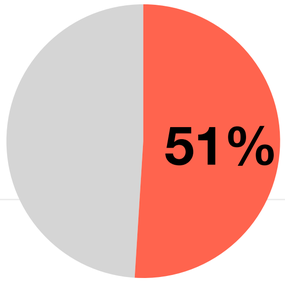


註：代理式AI採用率以企業自行採用的代理式AI技術為主，不含廠商產品內建AI代理功能或工具

© iThome 2026

主權AI意識高漲，政府學校和高科技積極自建開源LLM

採用開源LLM的台灣企業
半數部署了K8s



AI負載跑在K8s的企業
全球66%(2025)
台灣13%(2026)

開源LLM採用率年度比較

開源LLM也包括LVM,SLM等，%數值為該產業在該年的採用率

	整體	金融業	高科技製造業	一般製造業	服務業	醫療業	政府學校	百億營收企業
採用率增減幅度	+7%	+7%	+10%	+6%	+6%	0%	+12%	+7%

資料來源：iThome 2026 CIO & CISO 大調查，2026年
調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份

本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示—非商業性—禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。
© iThome 2026

在台灣境內建立台灣主權AI算力中心，明年將出現更多大規模Agent部署案例

算力提供單位		2026年6月 GPU 數量 (僅包含 nVidia H100 等級以上的 GPU)	2027年6月 預估GPU 數量 (僅包含 nVidia H100 等級以上的GPU)
政府	國科會國網中心	2,420	4,773
政府	數發部數產署算力池	40	140
民間	數發部促參 算力BOO案	0	> 20,000



- 數產署算力池已有40片H100等級以上GPU卡，未來一年計畫擴充到上百片
- AI算力中心BOO案一年投資額達到臺幣1,000億元，提供2萬片GPU
- 保留一定比例的GPU算力，提供政府和學研單位免費或優惠使用。
- 林宜敬部長：「兩家廠商申請，年底前動工，最慢明年第一季開始提供算力。」

台灣企業擁抱AI代理的課題

“傳統系統(Legacy Systems)並非為 AI 代理設計，缺乏可供 AI 代理即時呼叫的現代化 API，導致自動化鏈結斷裂，系統如何建立API呼叫機制及良好的權限控管將是重大的挑戰。”

技術債

iThome 醫學中心資訊主任

資料來源：iThome 2026 CIO & CISO 大調查，2026年
調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份
本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示－非商業性－禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。
© iThome 2026

台灣企業擁抱AI代理的課題

認知債

“在 AI 代理時代，IT 面臨的挑戰已從單點技術問題，升級為「人 × AI × 系統」的整體協作與治理能力。

首先是開發協作的重構。AI 代理已能參與需求拆解、寫程式、測試與文件產出，IT 團隊需重新設計開發流程，明確劃分人與 AI 的責任邊界，否則容易出現效率提升但品質失控的情況。

二是AI生成程式碼的資安與風險控管，包含是否引入不明授權碼、潛在漏洞、後門風險，企業必須建立自動化 Code Review、資安掃描與合規檢查機制，讓 AI 成為可被信任的開發夥伴。

三是AI 協作的監控與判斷機制，IT 不只要「讓 AI 能做」，更要「確保 AI 做對」，包含行為監控、決策紀錄、可解釋性與回溯能力，才能確保系統設定在實際營運中具可行性。

最後是商用化與規模落地。

在我們所屬產業，AI 代理必須能跨系統整合數據、支援大量客戶與即時投放需求，IT 的關鍵價值在於打造可複製、可擴展、可治理的 AI 架構，讓 AI 從實驗工具，真正成為企業成長的核心引擎。”

架構債

iThome 網路平台研發主管

資料來源：iThome 2026 CIO & CISO 大調查，2026年

調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院[illegible]296份

本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示—非商業性—禁止改作 4.0國際]授權。可用於非商業性使用，需[illegible]據。商業授權另詢。

© iThome 2026

台灣企業擁抱AI代理的課題

“四大AI代理時代的挑戰

1.IT架構更分散，更複雜。
2. Shadow AI。需建立「AI身份驗證、行為審計、零信任」機制。
3. 如何落地 AI治理框架，確保代理行為可追溯、可控。

4. 人才與技能缺口 IT團隊需懂 AI模型、Prompt工程、API整合、安全策略”



iThome 證券業IT協理

資料來源：iThome 2026 CIO & CISO 大調查，2026年
調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份

本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示—非商業性—禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。
© iThome 2026

台灣企業擁抱AI代理的課題

“ITSM/DevOps/SecOps流程要被重寫、從模型治理到代理行為治理改變、攻防節奏加速”

既有流程
砍掉重練

iThome 大型金控 資訊長

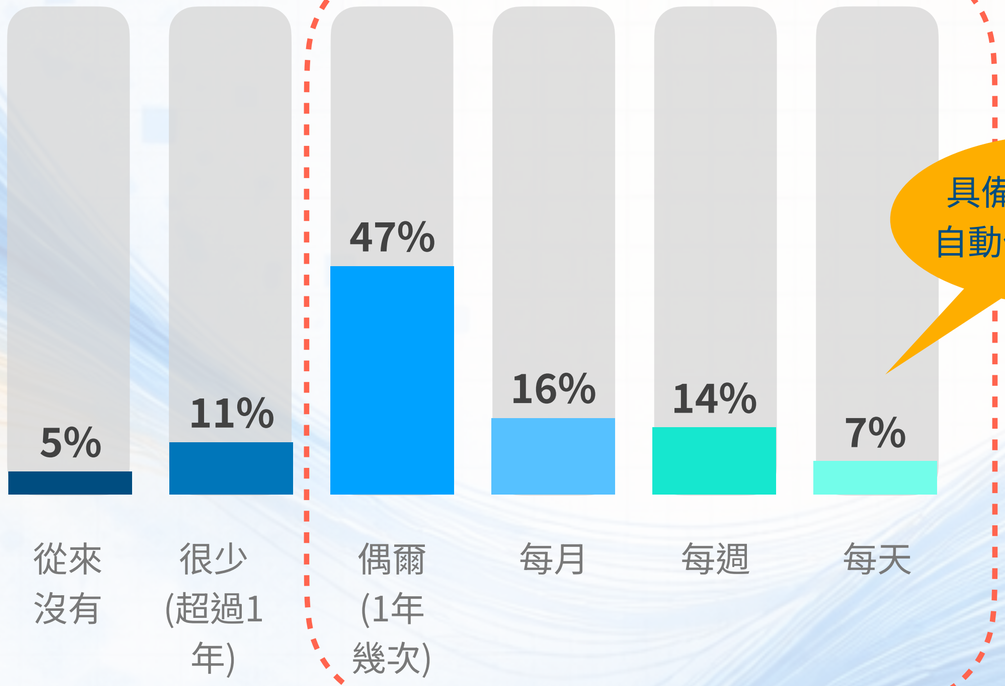
資料來源：iThome 2026 CIO & CISO 大調查，2026年
調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份

本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示－非商業性－禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。
© iThome 2026

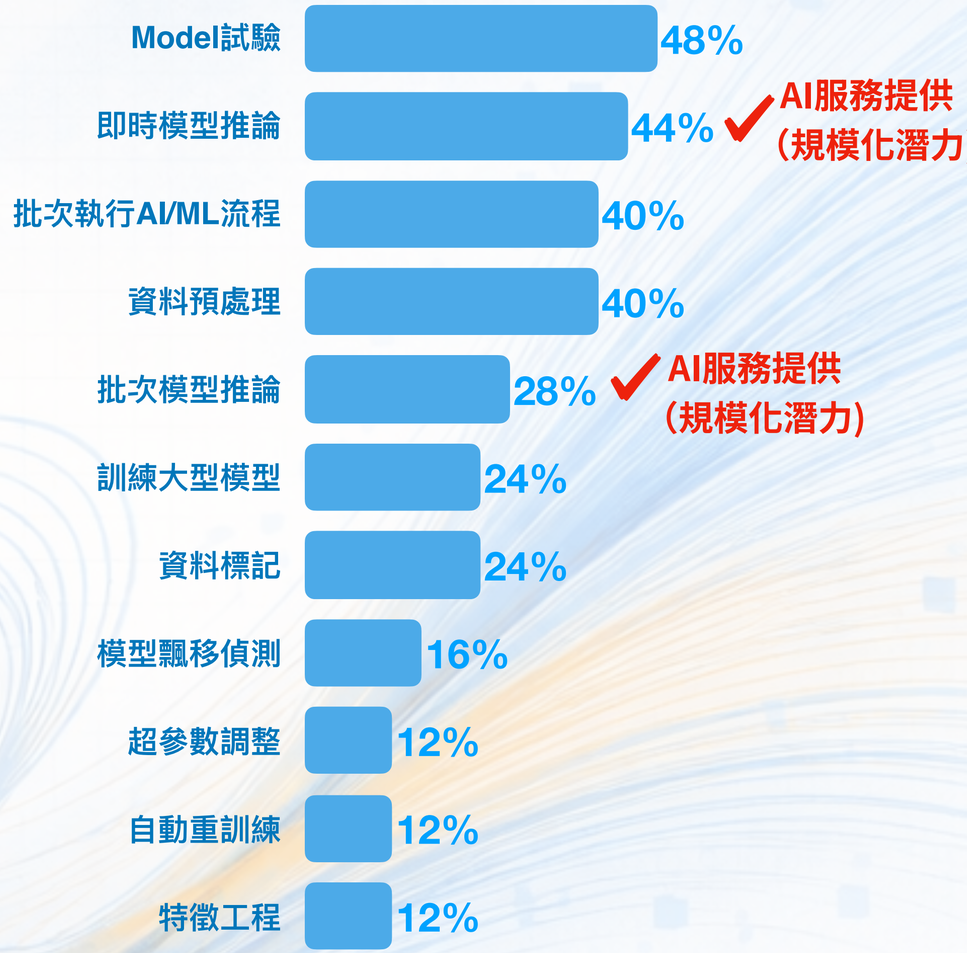
AI Model部署頻率？用來跑什麼AI負載？

全球企業AI Model部署頻率

47%每年幾次，7%天天部署



全球企業AI/ML工作負載用途



資料來源： CNCF 2025年度大調查，CC BY-ND 4.0

© iThome 2026

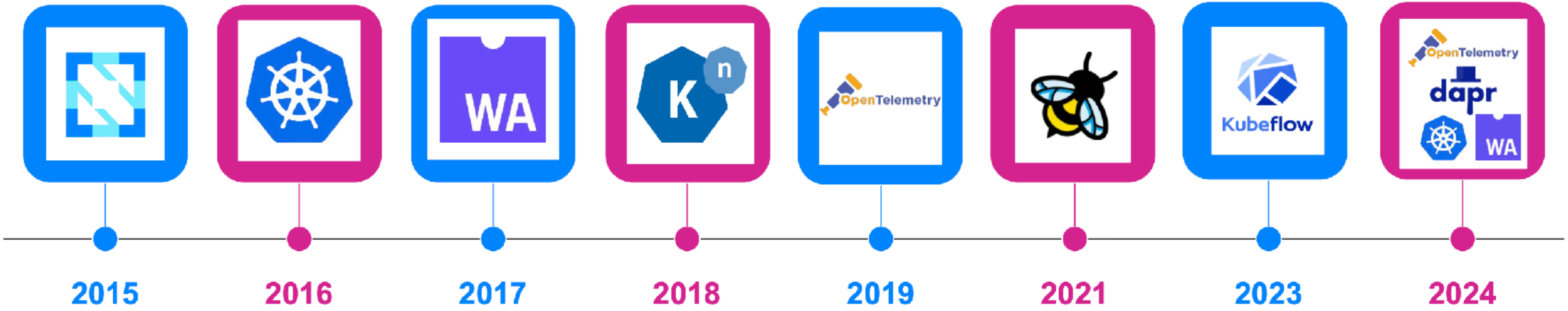
Cloud Native生態圈2024年開始聚焦AI工作負載

2024: CNCF Focuses on AI Workload Enablement

2025年9月
K8s 1.34版
GA

- **Kubernetes 1.32 DRA Highlights**
  - Structured parameter support got prom
  - Standardized network interface data for resource claim status
- **Cloud Native AI Paper Released**
  - https://www.cncf.io/blog/2024/03/19/announcing-the-ai-working-groups-new-cloud-native-artificial-intelligence-whitepaper/

- **WASM + LLMs** – A powerful pairing for sandboxing AI workloads using WasmEdge, with Nvidia showcasing this in Dec 2024.
- **AI Telemetry & Routing** – OpenTelemetry is emerging as a key player for AI workload observability, while projects like kgateway tackle intelligent AI request routing.



圖片來源： CNCF

資料來源：iThome 2026 CIO & CISO 大調查，2026年
調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份
本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示—非商業性—禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。
© iThome 2026

「無聊」的KubeCon，2025年底北美場，變成IT圈最夯、最潮的AI技術創新大會

熱門關鍵字：代理式工作流程（Agentic Workflow）、MCP協定、A2A協定 、 AI原生工作負載

1. 支援**鬆散耦合**微服務架構的K8s，轉變成可以適合**大規模、緊密耦合**型態的AI工作負載
2. K8s成為AI服務底層：**因應AI推理的資源分配，如何動態調度GPU**， 來改善成本效益和效能
3. 開發者體驗和平臺工程越來越重要，AI增強開發課題：**開發速度、治理、合規要求和成本優化的平衡**
4. AI應用的可觀察性挑戰：AI應用長期維運的挑戰



圖片來源： CNCF

© iThome 2026

從傳統 SRE 到 AI SRE：系統可靠性的範式轉移



背景概述

傳統 SRE 專注於 API 呼叫與系統可用性，但 AI 應用本質上是一條隨機的「推理鏈」。追要求維運患維從筆緘的服務運行轉向對模型語意、真實性及高昂運算成本（如 GPU、Token）的全面控管。

傳統SRE實踐做法，無法解決AI系統的可靠性挑戰。



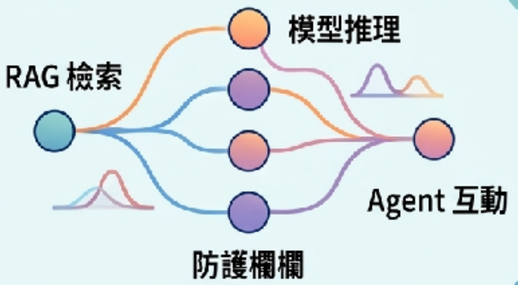
ING集團SRE團隊負責人Ehsan

核心思維與運作模式的轉變

從「API 呼叫」轉向「機率推理鏈」



AI 流程涵蛋 RAG 檢索、模型推理、Agent 互動及防護欄欄等複韓環節。





偵測難度的質變

傳統工具灥以甭覺隱蔽的語意錯誤、資料漂移或檢索調度的異常。

監控目標與經濟成本挑戰



新興的監控指標：語意與安全

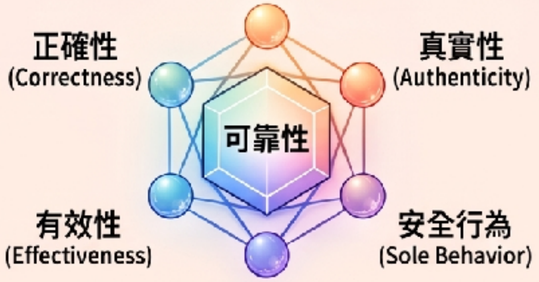
維運目標延伸至語意真實性、安全政策連僭及 AI 行為控制。

可靠性定義的維度擴張



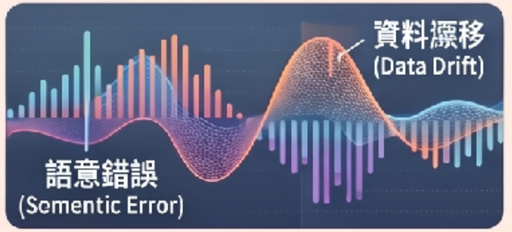
可用性 (Availability)

除了服務可用性，更強調結果的正確性、真實性、有效性及安全行為。





AI 驅動的語意監控能夠深入洞察隱蔽的語意錯誤、資料漂移或檢索調度的異常，提供更全面的保護。



比較項目	傳統 SRE	AI SRE
核心關注點	網站可用性、擴充性	語意真實性、安全性、有效性
主要挑戰	API 故障、系統負載	模型漂移、語意錯誤、機率性失敗
成本維度	運算資源、網路頻寬	Token 數、GPU 成本、資料搜集成本

圖片來源：CNCF



資源效率與經濟效益

需嚴格控管 Token 消耗、GPU 成本、API 支出及向量資料庫儲存成本。

資料來源：ING

Gemini Notebook

AI原生時代的崛起：雲端原生的下一步

iThome 1271期 封面故事

新挑戰：數十億AI代理的時代

我們正邁向一個由數十億AI代理驅動的新世界。



挑戰從「打造」轉為「大規模維運」

如何部署、維運及管理成千上萬的AI代理是當前最大課題。





AI工作負載需要全新優化的基礎架構

GenAI需要高算力、高效能儲存且具彈性的IT架構。





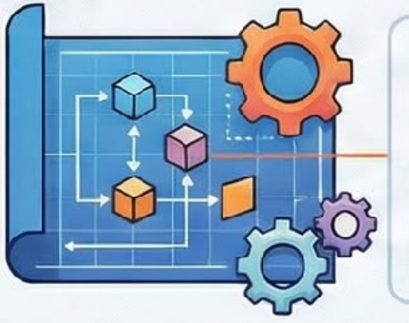
雲端原生對策：AI原生時代崛起

CNCF技術長Chris Aniszczyk

> 「AI代理就是新型態的微服務！」

CNCF借鏡微服務經驗，來管理、觀察並保護大規模AI代理。

CNCF提出雲原生AI藍圖與白皮書

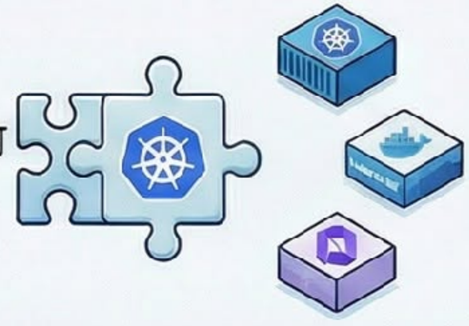


CNAI 分類

新增「CNAI」分類，涵蓋超過90個原生AI相關開源專案。

建立統一標準：Kubernetes AI合規計畫

確保AI工作負載在不同K8s環境中的可移植性與互適性。



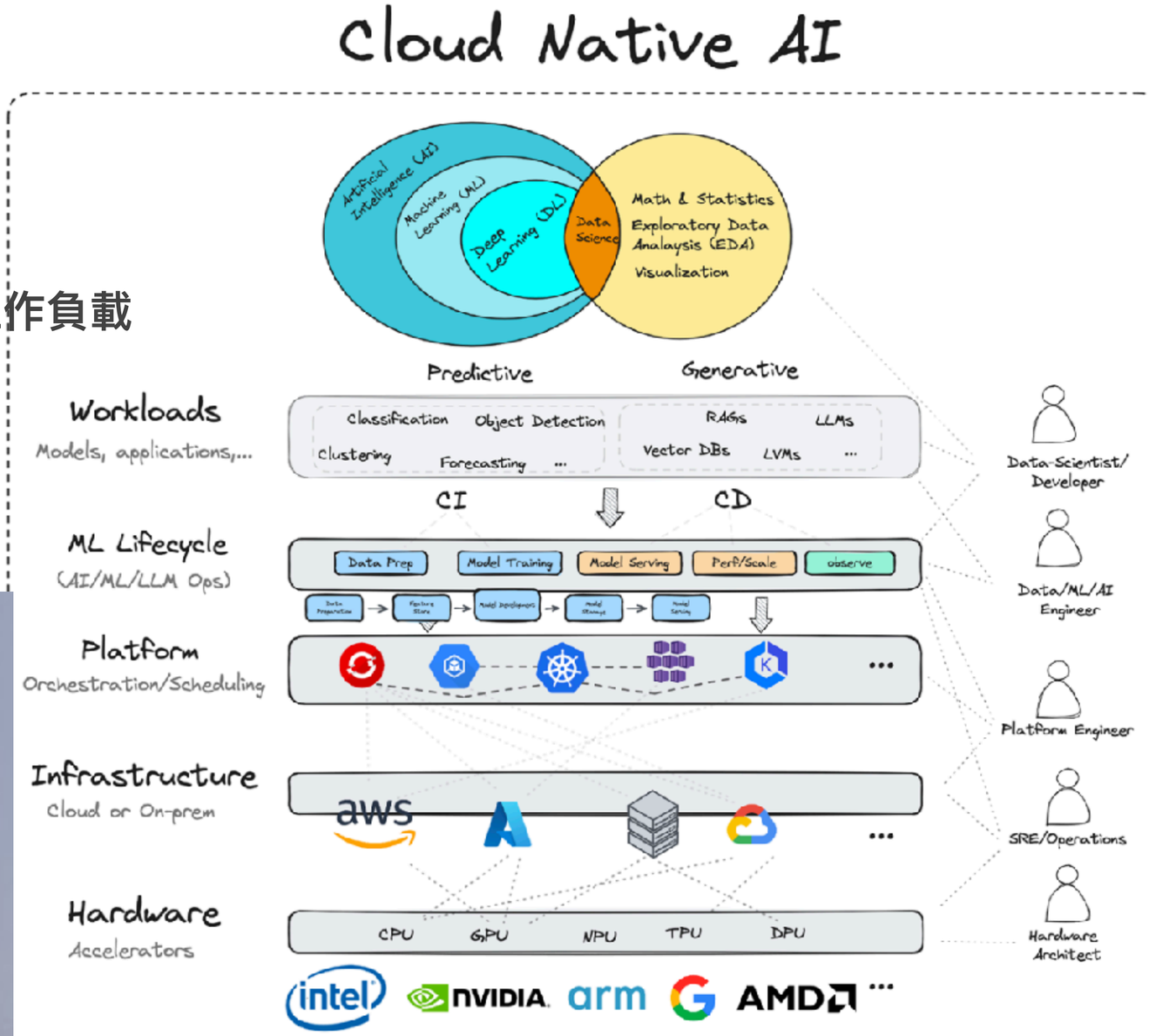
NotebookLM

2025年，雲原生積極融合AI，CNCF力推雲原生AI架構(CNAI)發展方向

- 多模型抽象層抽象架構，類似K8s的抽象層架構
- K8s對GPU支援度更高，甚至可達1萬顆GPU叢集
- 在雲端基礎架構，開發、部署、執行、擴充和監控AI工作負載
- 雲原生生態圈成為「雲原生AI」生態圈
- 2025年11月北美Kubecon發表K8s AI合規計劃 1.0

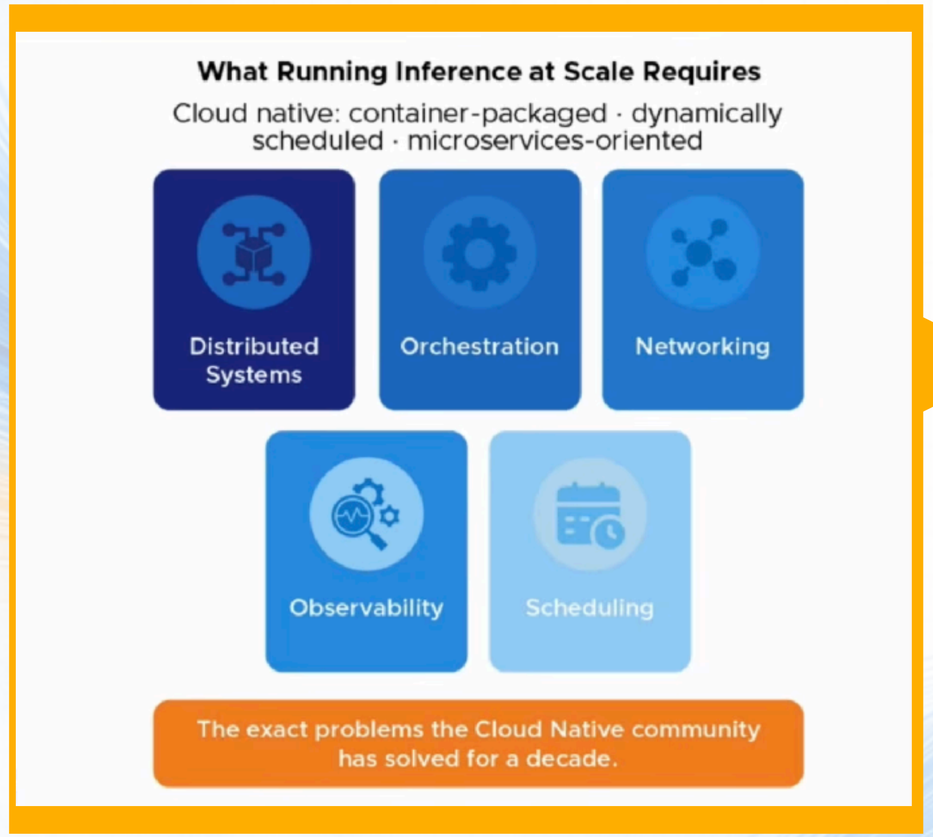


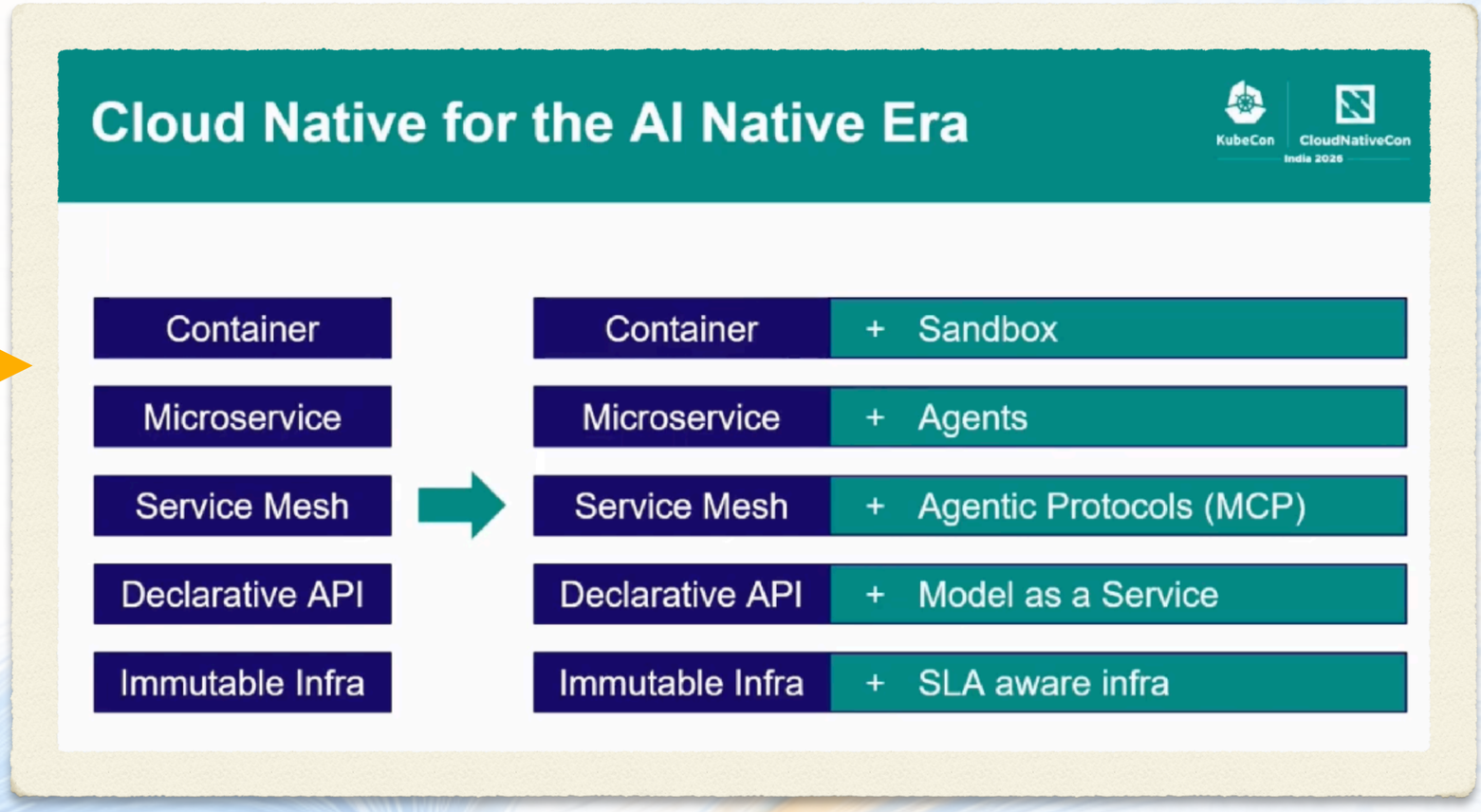




圖片來源： CNCF

Cloud Native 架構，如何支援AI Native應用需求？



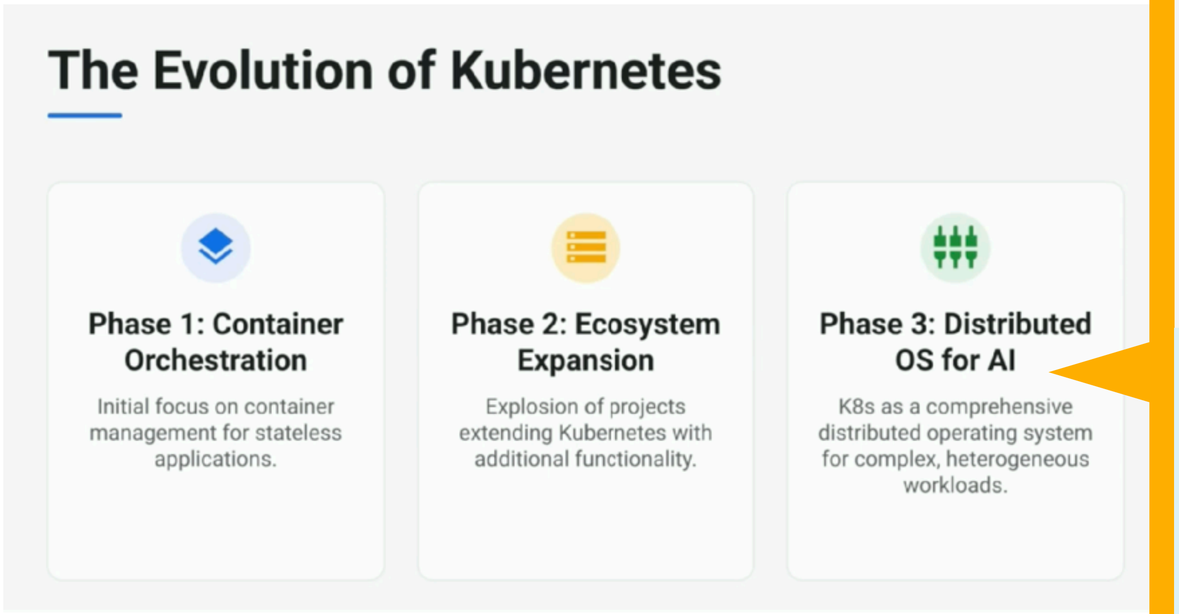


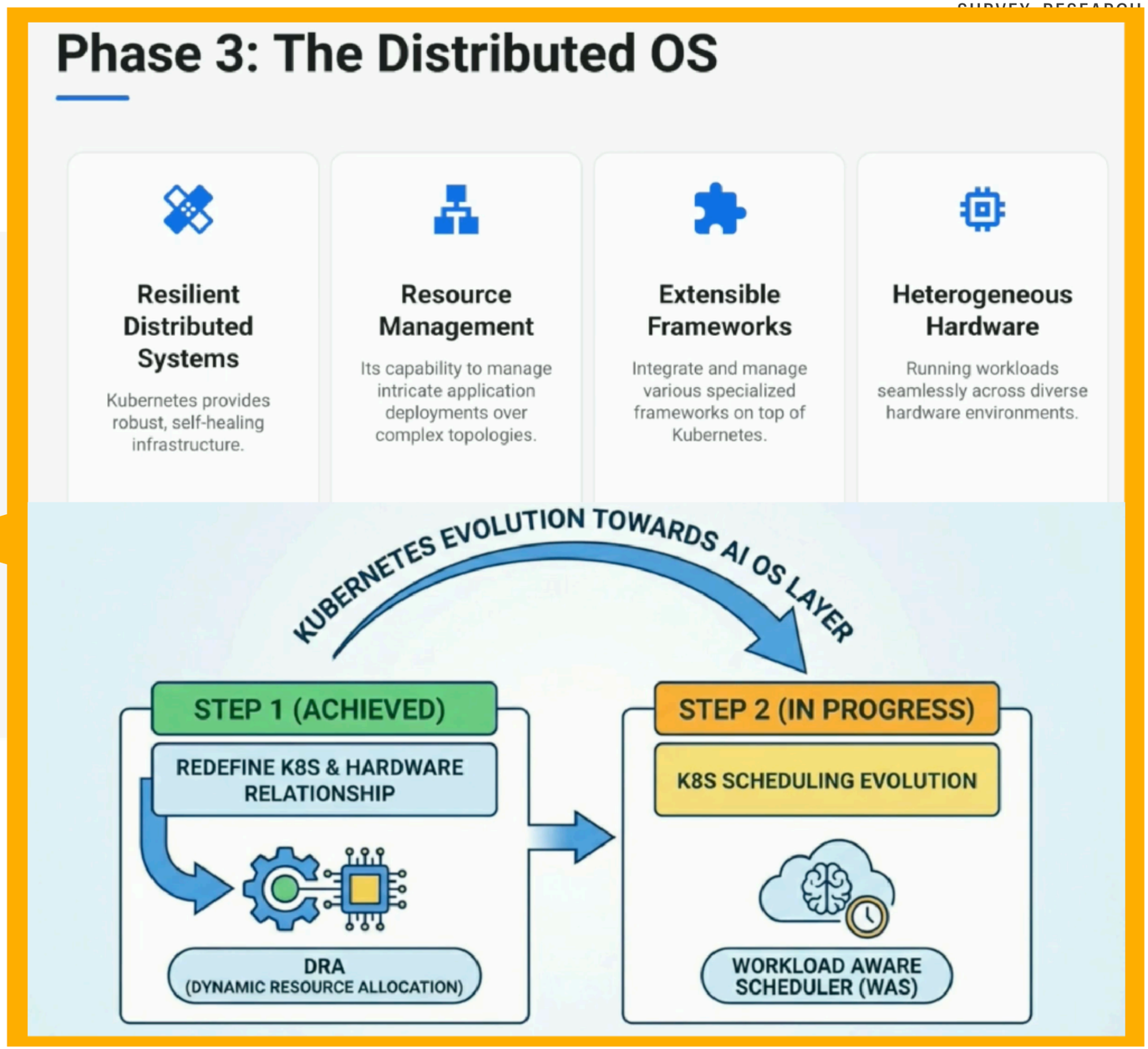
圖片來源：CNCF

© iThome 2026

2026年底全球算力消耗：67%推論+33%訓練

K8s要變身AI用的分散式OS





圖片來源：Google

資料來源：iThome 2026 CIO & CISO 大調查，2026年

調查說明：自2026年1月1日至1月29日執行，對臺灣大型企業、iThome歷屆CIO大調查企業、政府機關及大學院校的資訊主管進行線上問卷調查，有效問卷數396份

本報告採CC BY-NC-ND 4.0授權釋出 [姓名標示—非商業性—禁止改作 4.0國際]授權。可用於非商業性使用，需標註來源是iThome，不可修改圖表內容與數據。商業授權另詢。

© iThome 2026

K8s未來發展方向

支援安全的大規模Agent部署需求

Core Investment: Advanced Infrastructure Primitives

Workload & Topology Awareness

Enhances scheduling by making underlying MA nodes. with local latency

Dynamic Resource Allocation (DRA)

Allows discovery and allocation of specialized hardware beyond CPU/memory. Efficiently schedules workloads requiring accelerators like GPUs or FPGAs.

2025年9月GA
(K8s 1.34版中)

Agent Sandbox

Declarative, standardized API for managing long-running, stateful workloads with a stable identity, built on Kubernetes primitives.

2026年8月 1.0版

Workload Aware-Scheduler (WAS)

EVOLVING THE KUBERNETES CORE

The solution **isn't another workaround.** It's to **evolve the core.** We are building the **foundational primitives** for workload scheduling directly into Kubernetes.

WORKLOAD API
THE FOUNDATION OF WORKLOAD-AWARE SCHEDULING

GANG SCHEDULING
Coordinated scheduling of related tasks.

WORKLOAD-AWARE PREEMPTION
Priority-based eviction for critical workloads.

TOPOLOGY-AWARE SCHEDULING
Optimizing placement based on hardware layout.

2026年9月
發布Beta
(K8s 1.37版中)

圖片來源：Google

© iThome 2026

Kubernetes AI 合規計畫：打造標準化的雲原生 AI 生態系

CNCF 在 2025 年 KubeCon 北美大會宣布「Kubernetes AI 合規計畫」，旨在為 AI 基礎架構建立統一標準。透過定義加速器、網路、安全等領域的最低要求，確保 AI 工作負載在不同 K8s 環境中具備可移植性與互通性，加速企業擁抱雲原生 AI 技術。

計畫核心目標與認證流程

建立 AI 基礎架構的統一標準

確保 AI 工作負載在不同 K8s 環境間具備可移植性與互通性。



三階段認證：自評、審查與產品測試



廠商須先取得原有的 K8s 合規認證，並於 2026 KubeCon 前提交自評。

仿效 K8s 認證軟體計畫的成功模式

透過標準化工具生態系，降低企業導入門檻並加速產業發展。



資料來源： CNCF

合規範疇與技術要求



9 項必備要求 (Must)

23 項建議要求 (Should)

作為產品設計與架構規劃的參考。

涵蓋 6 大核心技術領域

- 加速器 (GPU)
- 網路
- 安全
- 調度
- 觀測性
- 算子 (Operators)

封裝技術與安全性整合：利用標準化 OCI 映像檔封裝模型，提升 AI 應用的分派效率與安全性。

三大核心 AI 應用場景與技術支援重點

超大規模訓練與微調



高效能加速器 (GPU)、高吞吐網路、分組排程 (Gang Scheduling)

高效能推論



進階流量管理、標準化監控矩陣、存取遮罩加速晶片

MLOps 流程



批次任務系統、資源管理佇列、可靠的 CRD 與鷹子支援

Gemini Notebook

CNAI 全景圖演進：2025 vs 2026 版本對比

呈現雲原生 AI (CNAI) 生態系的快速擴張與分類架構的細緻化演進。

2025 年版

2026 年版

規模與結構重大躍遷



90+

專案總數

從 2025 年的 90 餘個專案...

分類架構由扁平轉向層級化



由 13 個單一類別進化為...



開源專案數量增長

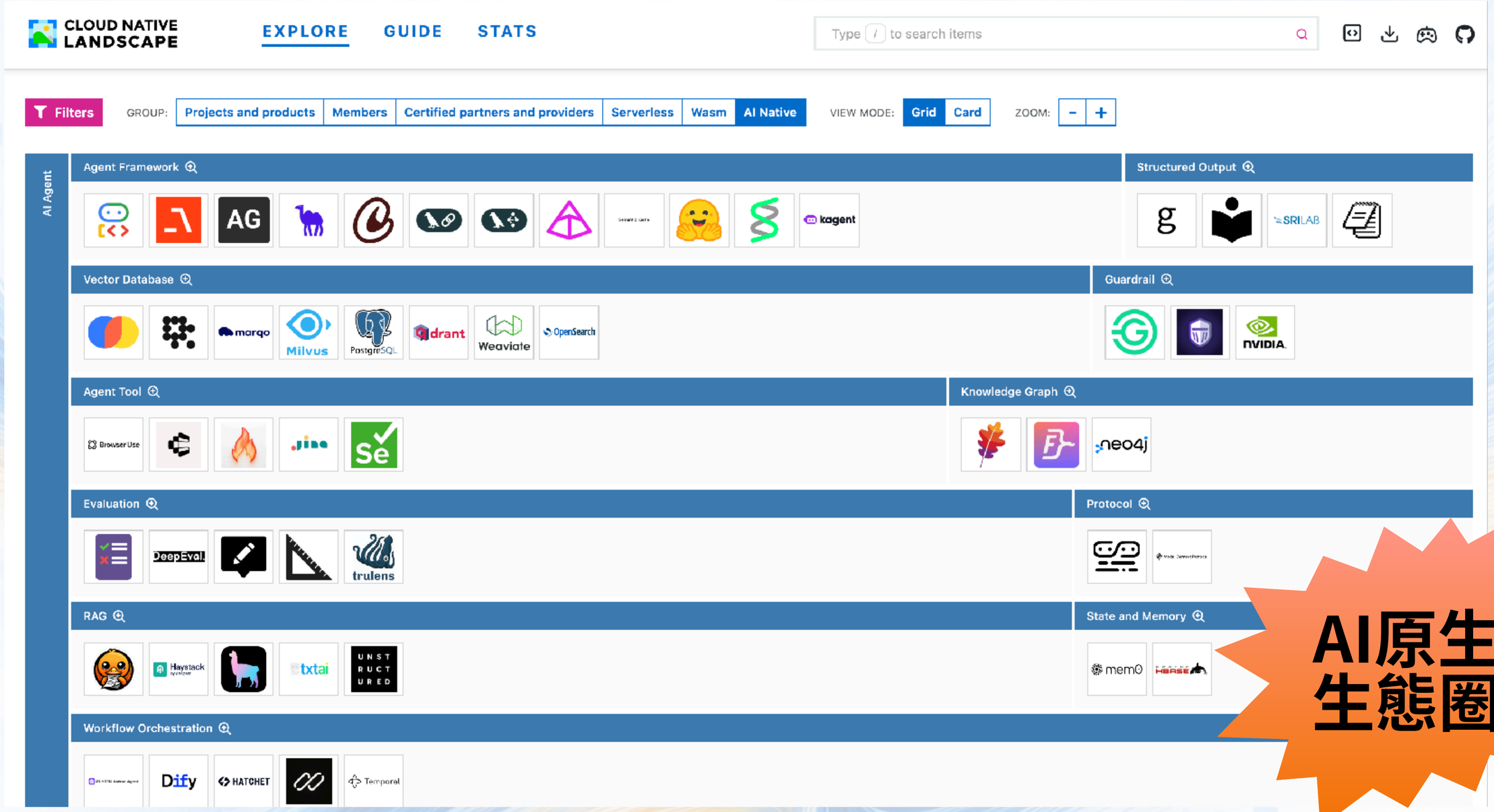
... 大幅擴張至 2026 年的 192 個。



... 5 大核心領域及 27 個細分小類。

資料來源： CNCF

CNCF的AI原生專案全景圖：2025年 90個專案 vs. 2026年 192個專案

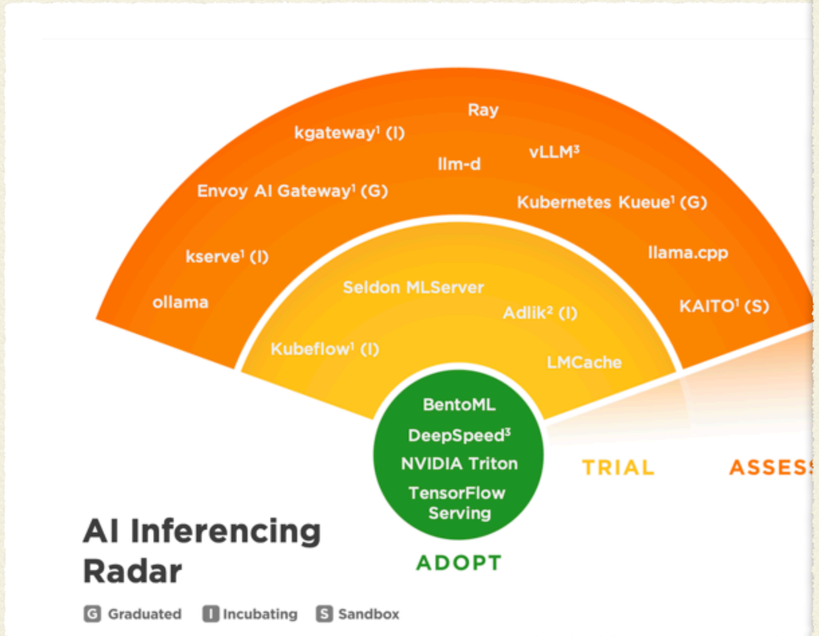


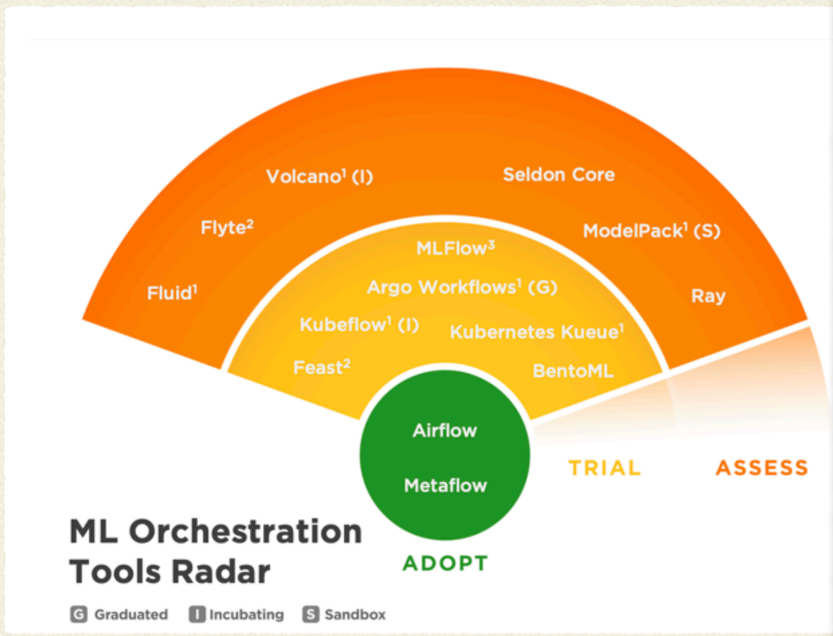
AI原生生態圈

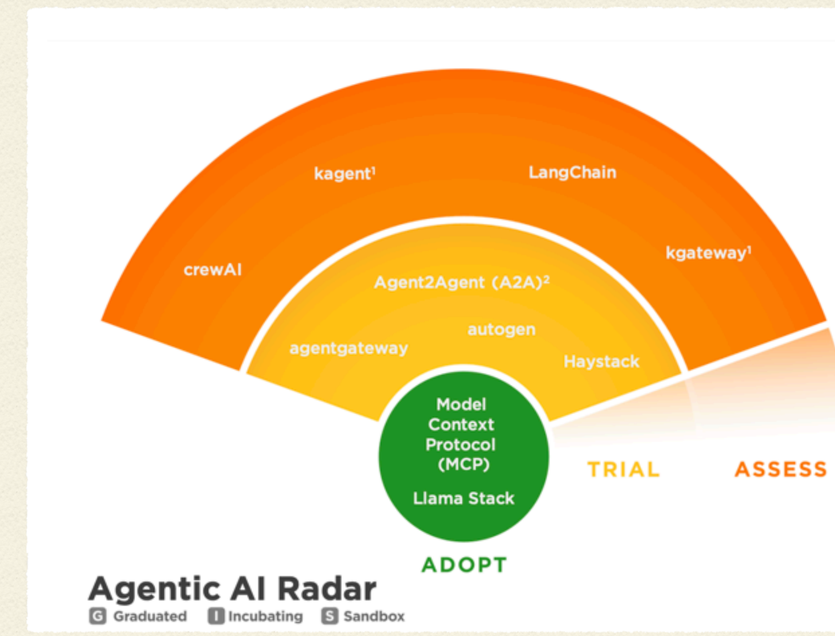
© iThome 2026

怎麼挑選AI開源專案？CNCF技術雷達圖









© iThome 2026

AI原生時代的崛起：雲端原生的下一步

iThome 1271期 封面故事

新挑戰：數十億AI代理的時代

我們正邁向一個由數十億AI代理驅動的新世界。



挑戰從「打造」轉為「大規模維運」

如何部署、維運及管理成千上萬的AI代理是當前最大課題。





AI工作負載需要全新優化的基礎架構

GenAI需要高算力、高效能儲存且具彈性的IT架構。





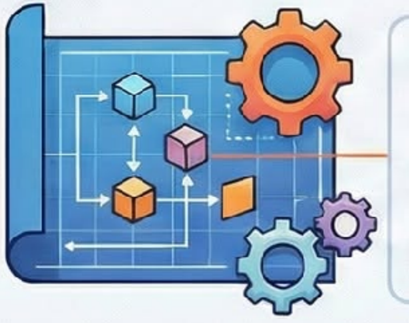
雲端原生對策：AI原生時代崛起

CNCF技術長Chris Aniszczyk

> 「AI代理就是新型態的微服務！」

CNCF借鏡微服務經驗，來管理、觀察並保護大規模AI代理。

CNCF提出雲原生AI藍圖與白皮書

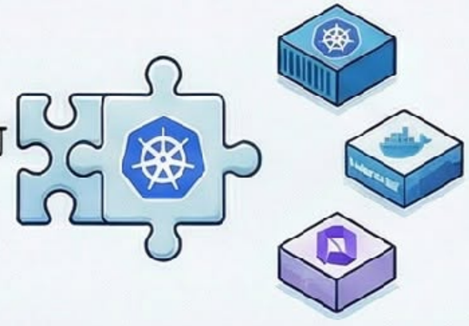


CNAI 分類

新增「CNAI」分類，涵蓋超過90個原生AI相關開源專案。

建立統一標準：Kubernetes AI合規計畫

確保AI工作負載在不同K8s環境中的可移植性與互適性。



NotebookLM