HON HAI TECHNOLOGY GROUP

AI與資安的世紀對話

李維斌
鴻海研究院執行長/資安所所長/集團資安長
逢甲大學資訊系教授
中華民國資訊軟體協會 資安長聯誼會製造業SIG會長
財團法人人工智慧科技基金會榮譽董事長

Sep. 26, 2025

3+3+3=∞

AI的時代來了

- 75% 的知識工作者於工作中使用 AI
- 78% 的AI 使用者將自己的工具帶到工作場所
- 員工認為 AI 可以
  - 幫助他們節省時間 (90%)
  - 專注於最重要的工作 (85%)
  - 更具創意 (84%)
  - 更享受工作 (83%)
- ~2024 Work Trend Index Annual Report from Microsoft and LinkedIn

AI can Help

Firefighting

快速回應緊急問題或作業危機 - 通常是被動和短期的

Refinement

改善或最佳化現有流程、工作流程或決策

Prototyping

快速測試新想法或在執行前模擬結果

Exploration

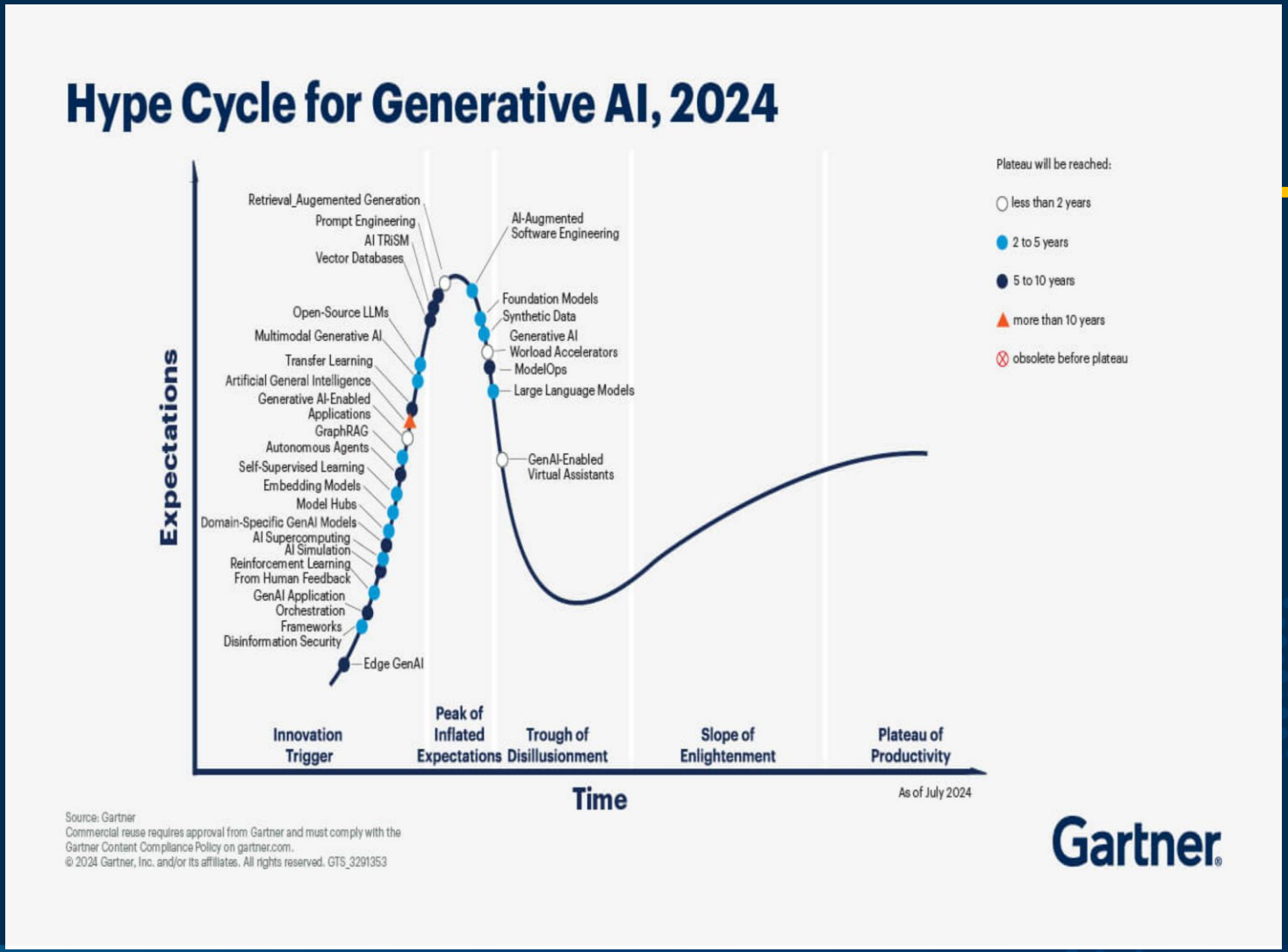
在資料中發現隱藏的模式、新的洞察力或未知的關係

Innovation

創造以前無法達到的新商業模式、產品或服務

資料來源：Gartner, 2021

AI 技術發展的複雜性和預期與現實的落差

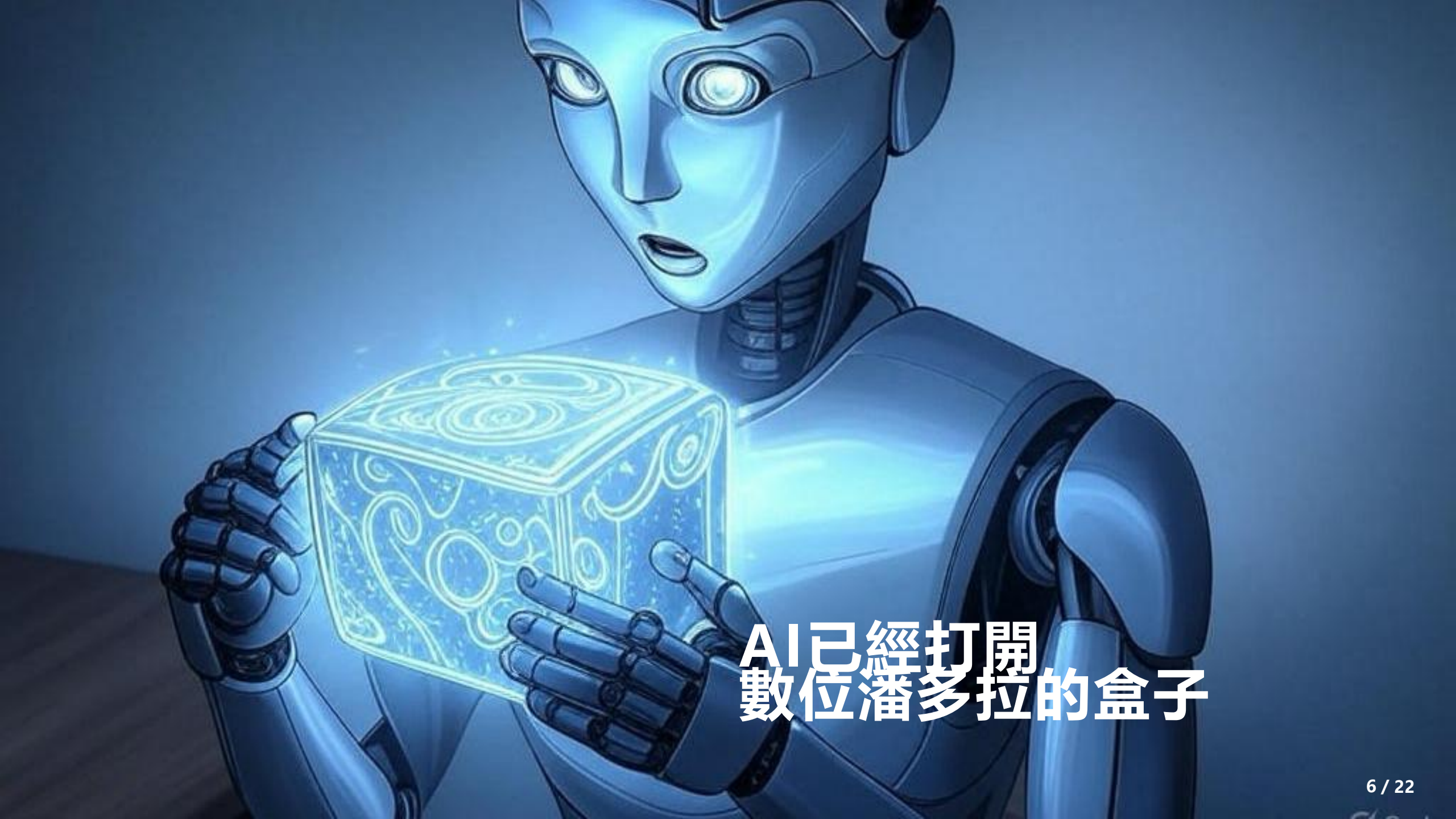


技術成熟度曲線 (Gartner Hype Cycle)

- Innovation trigger
- Peak of inflated Expectations
- Trough of Disillusionment
- Slop of Enlightenment
- Plateau of Productivity

AI時代的 HYPE OR HOPE?

- **技術採用 vs. 組織能力**
  - 技術炒作 vs. 企業準備的錯位風險
- **AI有用還是有用AI?**
- **採用(Adopt)、規模化(Adept)、永續(Adapt)需要時間**
- **要捕捉價值比炒作更難**
  - 從錯誤的問題開始
  - 以錯誤的期望來做
  - 使用錯誤的營運模式
- **AI是技術/夥伴，落地是情境**


AI已經打開
數位潘多拉的盒子

生成式AI的陷阱

- 很容易上手/啟動試點計畫
- 很容易無疾而終
- 很容易面臨產生價值的困境
- 很容易留下急於見效的資安隱患
- 對轉變的抗拒與期待落差

變局？困局？

- AI+新手的衝擊
  - 快速補上經驗缺口
  - 成果效率大幅提升
  - 中階熟手的經驗價值迅速邊緣化
- 從效率到能力斷層
  - 基本功更被忽視
  - 新手、熟手到高手的經驗累積路徑被破壞
- 制度與信任的重構
  - 工具加速產出卻淡化責任感
  - 去脈絡的學習，帶來判斷和治理風險
  - 整體信任系統面臨重設
- 制度創新的時代
  - 重新設計人才養成路徑和責任承擔機制
  - 產業、教育、社會的共同議題

Cybersecurity 大挑戰

- 傳統系統與人工智慧的相遇
- 供應商都在說AI
- 監管碎片化增加了合規的複雜性
- 人為因素仍然是一個弱點
- BYOAI
- 人工智慧特有的風險

OWASP Top 10 for LLM Applications

LLM01:2025 Prompt Injection	LLM02:2025 Sensitive Information Disclosure	LLM03:2025 Supply Chain	LLM04:2025 Data and Model Poisoning	LLM05:2025 Improper Output Handling
LLM06:2025 Excessive Agency	LLM07:2025 System Prompt Leakage	LLM08:2025 Vector and Embedding Weaknesses	LLM09:2025 Misinformation	LLM10:2025 Unbounded Consumption

這是現實...

新聞

【資安日報】2月29日，Hugging Face平臺上面出現惡意模型！能

新聞

Microsoft 365 Copilot可被濫用進行釣魚攻擊

Zenity在黑帽大會針對Microsoft 365 Copilot，發表兩資安工具CopilotHunter與LOLCopilot，可進行資料洩漏與釣魚攻擊模擬

文/ 李建興 | 2024-08-13 發表

```
Manager Title: Sales
Manager Email Address: kris@zontosoent.onmicrosof
----------
Skip Manager: Admin
----------Sunday, 11 August 2024
Next week schedule:     - Daily meeting with Kris
Monday, 12 August 2024 ----------
                        Top 2 Collaborators
                        Collaborator Name: Kris Sa
                        Collaborator Email Address
```

whoami++

Gartner Identifies the Top Cybersecurity Trends for 2024

Analysts to Explore Cybersecurity Trends During Gartner Security & Risk Management Summit, March 18-19 in Sydney

Generative AI (GenAI), unsecure employee behavior, third-party risks, continuous threat exposure, boardroom communication gaps and identity-first approaches to security are the driving forces behind the top cybersecurity trends for 2024, according to Gartner, Inc.

"GenAI is occupying significant headspace of security leaders as another challenge to manage, but also offers an opportunity to harness its capabilities to augment security at an operational level," said Richard Addiscott, Senior Director Analyst at Gartner. "Despite GenAI's inescapable force, leaders also continue to contend with other external factors outside their control they shouldn't ignore this year."

2024 will see security leaders respond to the combined impact of these forces by adopting a range of practices, technical capabilities and structural reforms within their security programs, with a view to improving organizational resilience and the cybersecurity function's performance.

The following six trends will have broad impact across these areas:

Trend 1: Generative AI - Short-term Skepticism, Longer-Term Hope

Security leaders need to prepare for the swift evolution of GenAI, as large language model (LLM) applications like ChatGPT and Gemini are only the start of its disruption. Simultaneously, these leaders are inundated with promises of productivity increases, skills gap reductions and other new benefits for cybersecurity. Gartner recommends using GenAI through proactive collaboration with business stakeholders to support the foundations for the ethical, safe and secure use of this disruptive technology.

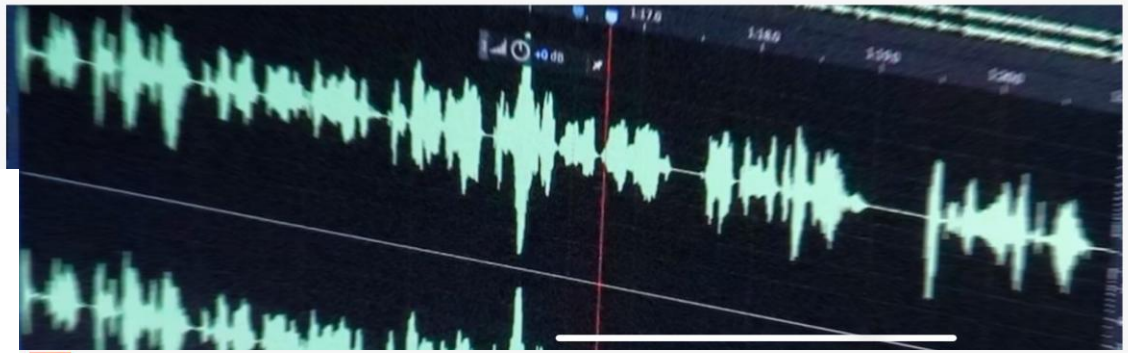
"It's important to recognize that this is only the beginning of GenAI's evolution, with many of the demos we've seen in security operations and application security showing real promise," said Addiscott. "There's solid long-term hope for the technology, but right now we're more likely to experience prompt fatigue than two-digit productivity growth. Things will improve, so encourage experiments and manage expectations, especially outside of the security team."

新聞

FTC祭出2.5萬美元尋求不受AI複製語音危害的方法

美國聯邦交易委員會（FTC）推出語音複製挑戰賽（Voice Cloning Challenge），徵求能夠防範非法AI語音應用的解決方案

文/ 陳曉莉 | 2024-01-05 發表



新聞

員工外洩內部機密！三星開放ChatGPT後出事緊急限縮使用

The Register、Tom's Hardware等媒體引述南韓當地媒體Economist的消息，指出三星員工在不清楚ChatGPT使用規範下，為了工作之便直接將半導體設備、程式碼相關資訊上傳給ChatGPT進行處理，導致三星的內部機密資料外洩

文/ 林妍溱 | 2023-04-07 發表

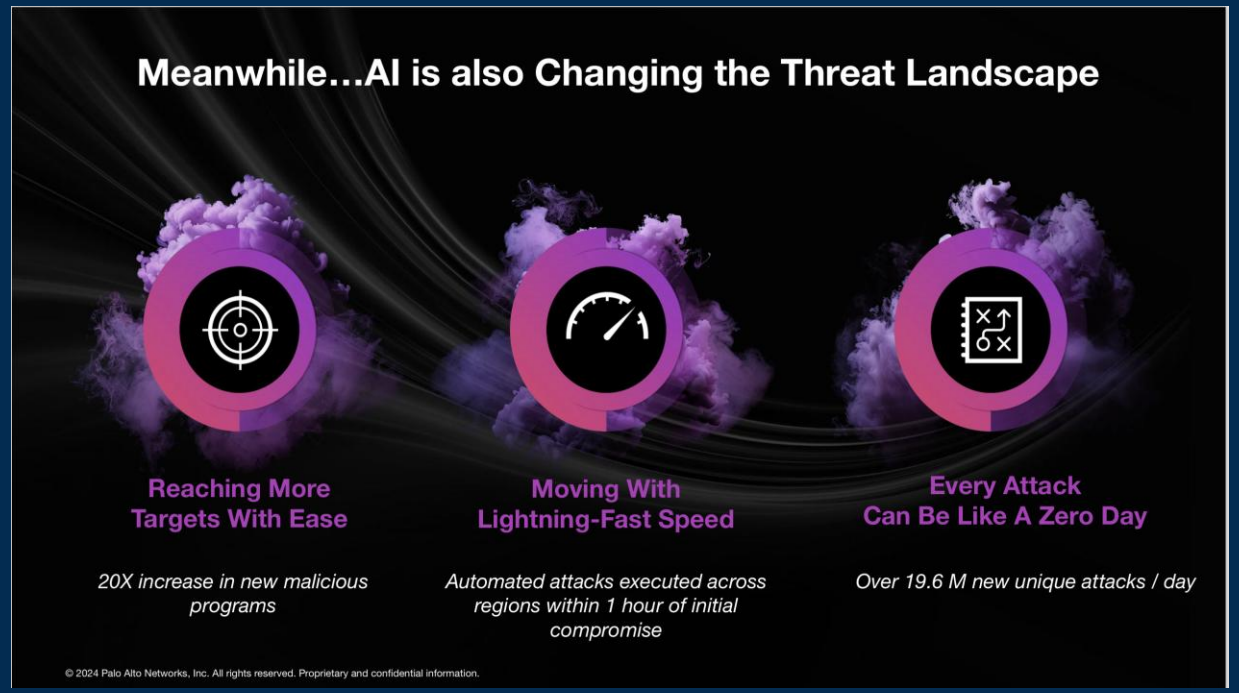
開啟AI的潘朵拉

- AI 的影響會因環境、使用個案、時間而異
- AI 系統會自行演變，不同於傳統工具被動執行的自主決策能力
- AI 系統的輸出難以解釋，行為難以預測
- AI 風險和影響還沒有很好的理解和完整的定義
- 沒有釋出解決方案的「完美時機」
- 委託決策權帶來的風險
- 我們不知道我們不知道什麼
- 人機協作的新模式
- 能產生原創內容和解決方案的創造性思維

AI是改變資安生態的關鍵變數

- 驅動網路防禦： AI 作為資安的推力
- 威脅形勢演變： AI 作為攻擊的工具
- 風險管理挑戰： AI 成為意外的根因

矛：AI 作為攻擊的工具



- 攻擊規模：新型惡意程式增加20倍，快速生成顯著擴大攻擊範圍，輕鬆同時觸及更多目標，讓更多目標暴露於風險中
- 攻擊速度：AI自動化的極高效率，在初始破壞後1小時內橫掃區域，讓攻擊如閃電般難以攔截
- 攻擊複雜度：每日超過1960萬個新的獨特攻擊模式，讓每次攻擊都像零日攻擊，難以預測與防範
- 反應時間壓縮：快速迭代的貓抓老鼠遊戲，雙方都要快速適應彼此的進化，沒有喘息的空間

盾：AI 作為資安的推力

- 即時威脅偵測
- 行為異常偵測
- 預測性安全措施
- 零時差漏洞偵測
- 自動化事件回應
- 身份與存取安全
- 數據隱私與保護
- 但是，
  - 它們在哪裡?
  - 它們如何運作?

AI的阿基里斯腱...

- 資料下毒（Data Poisoning）
- 對抗性攻擊（Adversarial Attacks）
- 模型竊取（Model Theft）
- 隱私侵犯（Privacy Violations）
- 技術濫用（Technology Abuse）
- 未知風險（We Don’t Know What We Don’t Know）

第三方解決方案的AI黑箱

我們面對什麼

- AI訓練資料的不透明性
- AI運行機制的不可知性
- AI會持續演化以應對新威脅
- 可能跟不上零日攻擊
- AI的計算可能影響威脅回應的及時性
- 實施和維護AI系統需要大量投資
- 將AI整合到現有系統複雜性及成本高
- 監管與合規的問題

隱藏的風險：供應商依賴與不確定性

- 供應商的風險與依賴性
- 供應鏈資安風險
- 服務中斷的風險

借力使力，打開黑箱

- 利用採購作為影響力
- 制定合約與法律保護
- 要求透明度與溝通
- 避免單一供應商依賴
- 採用多元防禦策略
- 保持人機協作，專家參與決策
- 定期審查與測試，避免過度信任自動化的結果

掌握AI ~成就自己與組織競爭優勢

- 思考如何運用AI成為更好的自己
- 思考如何運用AI塑造自己的未來
- 思考如何引導AI for good
- 理解智慧是提出正確的問題
- Work smart - 未來是用”聰明的方法”做事，不是用時間

結論：AI時代的資安，成長與競爭力的基石

- AI 浪潮已來：工作場域全面進化
- 組織策略：動態應變與穩固建設，智勝 AI 時代
- 核心驅動：人與 AI 的智慧協作
- 資安本質轉型：從成本到競爭力
- 最終目標：共築安全、競爭與成長的未來

謝謝聆聽