AI 世代的資安威脅態勢

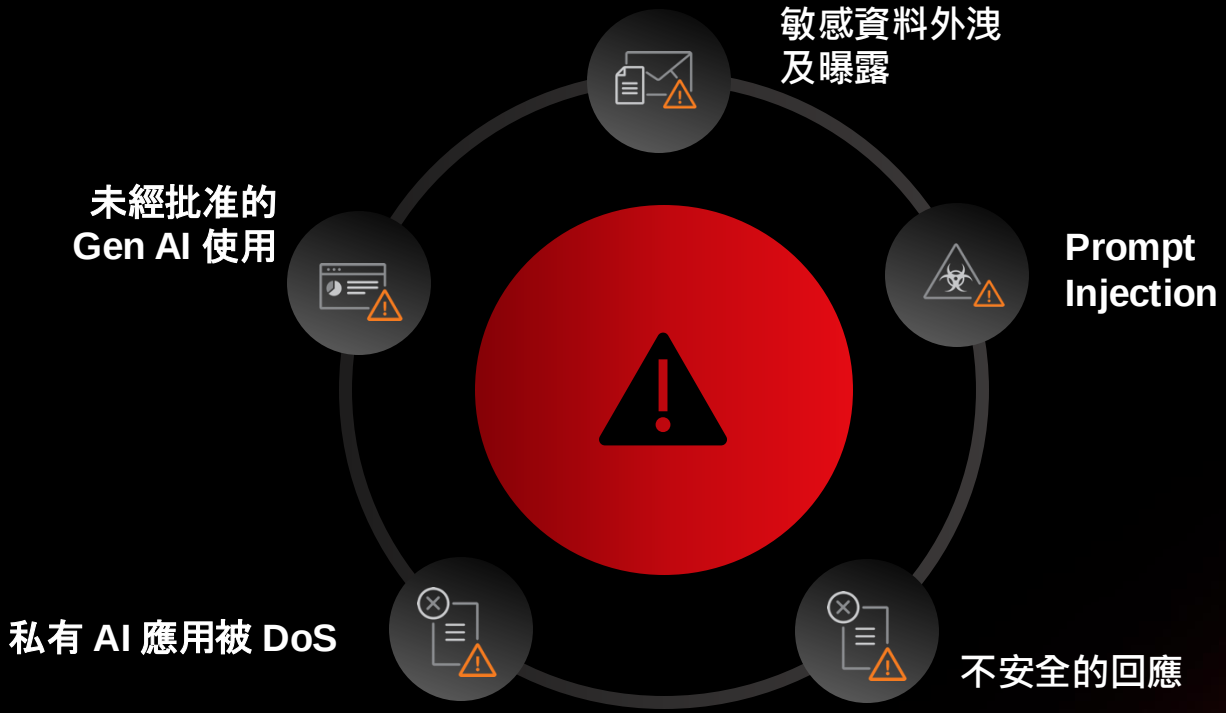
趨勢科技大型企業業務部協理

楊肇謙 Kevin

Agenda

- 使用 AI 帶來的資安風險
- 如何安心的使用 AI 應用 (Security for AI)
- 如何利用 AI 將資安防護做得更好 (AI for Security)

使用 AI 帶來的資安風險



真實案例 - Microsoft Tay

真實案例 - Microsoft Tay

微軟推出的人工智能聊天機器人 Tay，在短短 16 小時內就因發布大量種族主義和仇恨言論而被迫下線。這起事件引發了關於人工智能系統如何應對惡意使用者操縱的廣泛討論

攻擊者故意向 Tay 輸入大量種族主義、性別歧視和反猶太主義的言論，導致 Tay 開始模仿並產生類似的有害內容

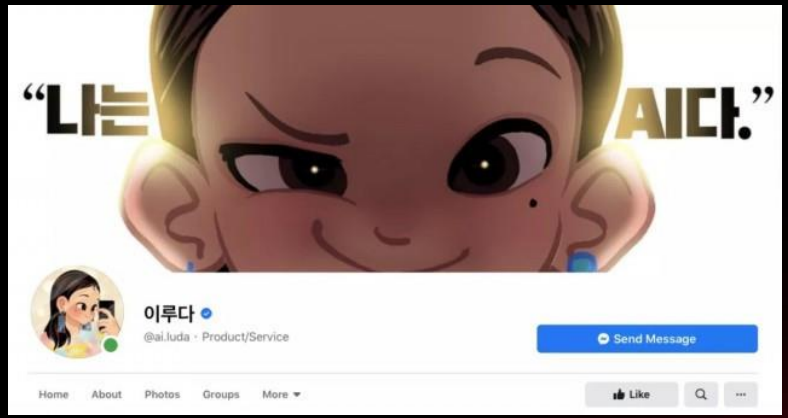
真實案例 - 李露達 Lee Luda

真實案例 - 李露達 Lee Luda

李露達（Lee Luda）是一個由韓國初創公司 Scatter Lab 開發的人工智能聊天機器人。李露達迅速吸引了超過 75 萬用戶，累計進行了近 7000 萬次對話

在訓練李露達時，使用了從韓國流行聊天應用 KakaoTalk 中收集的約 100 億條真實對話。然而，公司未能妥善處理這些對話中包含的個人信息，導致李露達在與用戶互動時意外洩露了一些敏感資料，包括姓名和銀行帳戶等個人信息

該公司被韓國個人信息保護委員會處以 1 億韓元（約合新台幣 270 萬元）的罰款，原因是違反了個人信息保護法



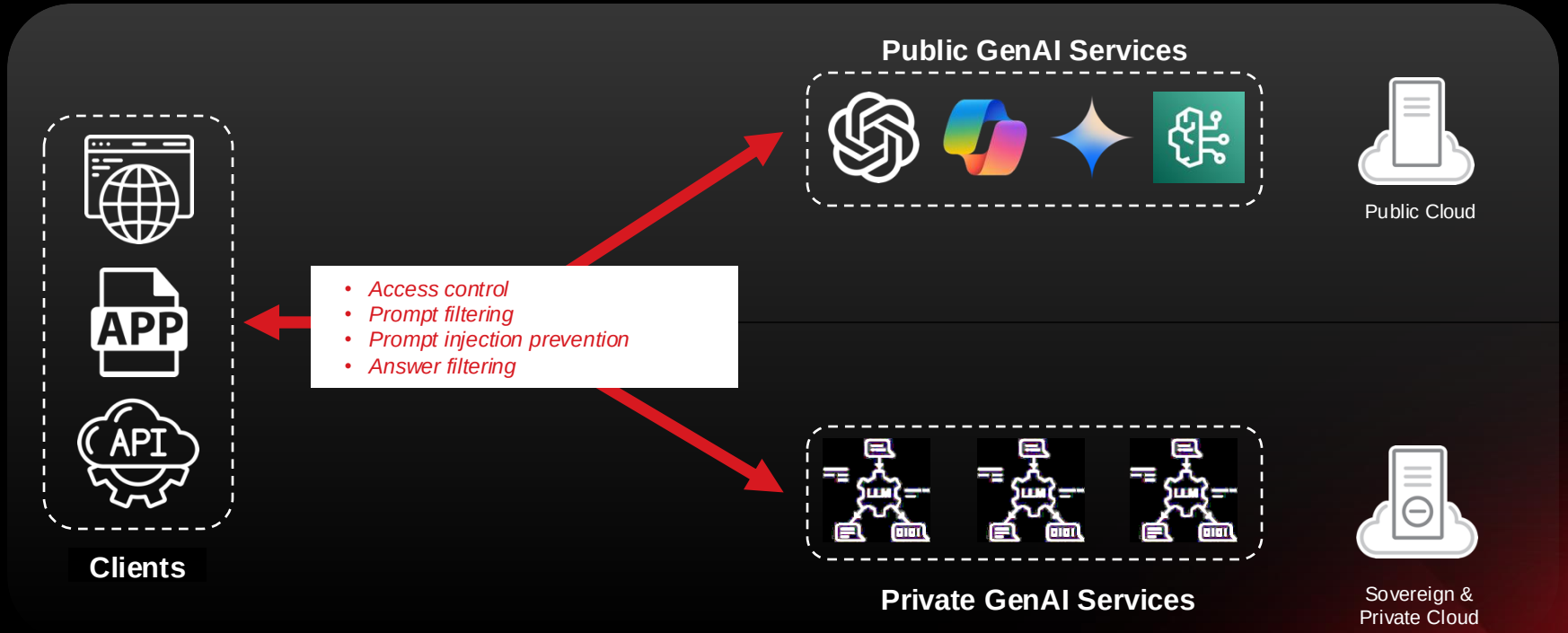
真實案例 – 三星電子

真實案例 – 三星電子

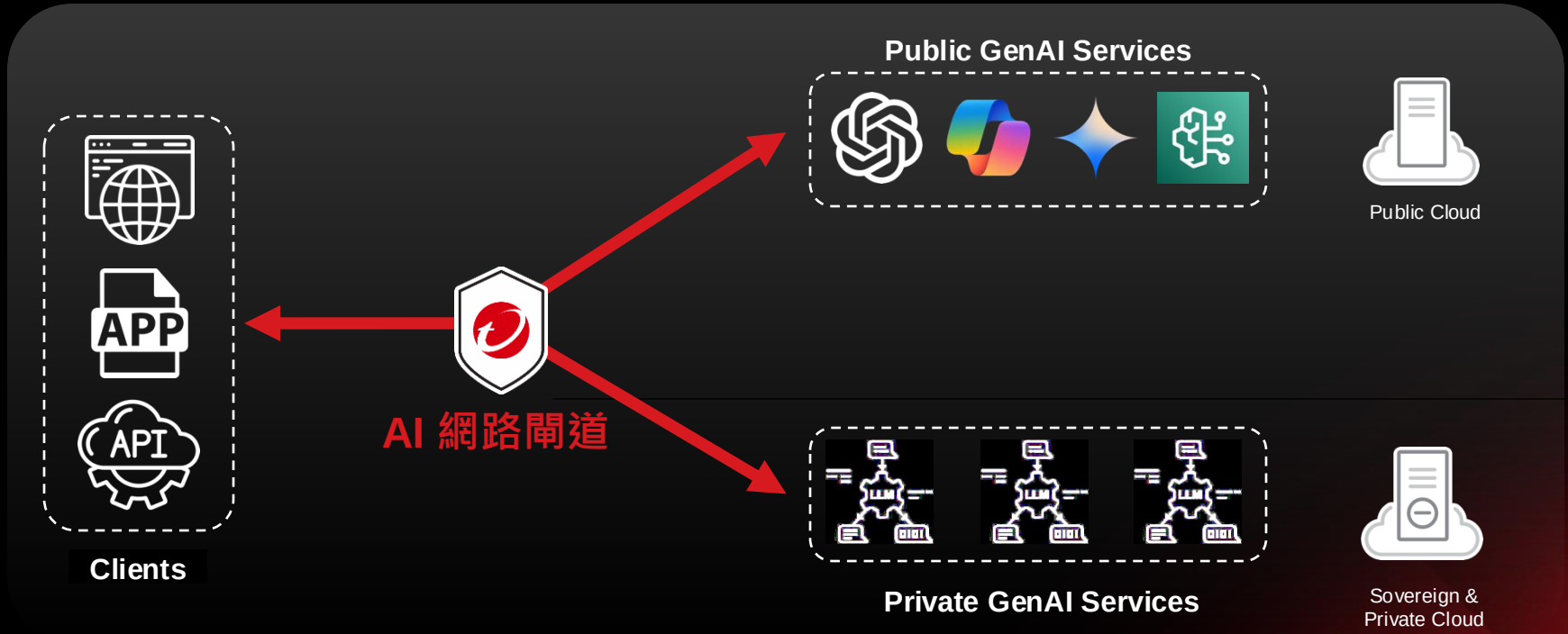
三星裝置解決方案及半導體業務部門發生三起事件，出自員工將公司機密資訊輸入 ChatGPT 而外流。外洩的資訊包括半導體設備量測資料庫、生產 / 瑕疵設備相關軟體，以及一份公司會議語音轉錄的文字紀錄摘要。報導指出，一名軟體開發工程師在資料庫程式開發期間發現程式碼錯誤，於是將整份程式碼複製貼到 ChatGPT 對話中，以尋找臭蟲及解決方案。

最後導致公司機敏資料外洩，三星公司針對員工使 ChatGPT 進行管制

如何安心的使用 AI 應用（Security for AI）



如何安心的使用 AI 應用（Security for AI）



AI 網路閘道主要功能

01 限制不預期的 AI 應用程式使用（Access control）

02 防止機敏資料外洩及不當揭露

03 阻擋 Prompt Injection 惡意攻擊

04 過濾不安全的回應

05 阻止 AI 服務遭 DoS 攻擊

如何利用 AI 將資安防護做得更好(AI for Security)

趨勢科技兩大策略

01

增加惡意程式及駭客行為的偵測能力

02

提高客戶營運趨勢產品的效率及效果

增加偵測能力

01 利用 Machine Learning 加強對惡意程式辨識

02 利用 AI 運算加強對駭客行為的關連及分析

提高效率及效果

趨勢利用自建的大語言模型提供資安顧問的服務

01

降低客戶使用資安產品的門檻

02

評估風險，預測未來可能發生的資安問題

03

提高資安事件的處理效率及問題排除的效果

趨勢科技校園方案

主機資安防護及風險管理服務包

資安防護	風險管理	資安服務
主機防毒	資安風險評估	代管監控
弱點防禦	主機健康度	事件通知
EDR	帳號健康度	回應處理

Power by AI

Thank You!