模組 1

認識生成式 AI

AI 基本概念・能力與限制・疾管署資安規範

⏱ 片長：5 分鐘

ChatGPT、Claude、Gemini——
您能判斷 AI 什麼時候在說謊嗎？

AI 回答：「台灣 2023 年登革熱確診共 12,847 例……」

⚠ 這個答案看起來很正確——但是錯的。

這支影片會讓您學會辨識 AI 何時出錯，以及安全使用 AI 的完整原則。

看完這 5 分鐘，您將能夠：

1. 用一句話解釋生成式 AI 的運作原理（不需要技術背景）
2. 識別 AI「幻覺」的三個跡象
3. 熟記4條AI使用資安原則

什麼是生成式 AI？

- 生成式 AI（Generative AI）能根據您的提示「生成」新內容

  文字、圖片、程式碼、音樂……各種形式皆可

- 運作邏輯：輸入 Prompt → AI 分析意圖 → 生成回應
- 把它想像成一位「超級助理」：

  讀過海量書籍，但有時也會「腦補」不存在的資訊

- 核心定位：是助理，不是專家；AI 產出仍需人工審核

AI 如何運作？（簡化說明）

1 學習階段

AI 被「餵食」大量文字資料
（書籍、網頁、文章），
學習語言模式與知識。

→

2 理解提示

當您輸入問題時，
AI 會分析您的意圖，
找出最相關的知識。

→

3 生成回應

根據學到的知識，
一個字一個字「預測」
最適合的回應。

⚠ AI 的知識有「截止日期」——不知道訓練資料之後發生的事情。

常見生成式 AI 工具

ChatGPT（OpenAI）

使用最廣泛的聊天 AI
免費版每日限量使用

Claude（Anthropic）

注重安全性與長文處理
免費版每日限量使用

Gemini（Google）

整合 Google 生態系
有 Google 帳號即可用

Copilot（Microsoft）

整合 Office 365 應用
適合已有 M365 授權的機關

AI 擅長的事情 ✓

- 文字撰寫與潤飾（文案、Email、摘要）
- 資料整理與格式轉換
- 翻譯與語言修正
- 腦力激盪與創意發想
- 程式碼撰寫與除錯
- 問答與解釋概念
- 文件摘要與重點提取

AI 的限制 ✗

- 會「幻覺」（Hallucination）：可能編造看起來合理但錯誤的資訊
- 知識有截止日期：不知道訓練資料截止後發生的事
- 無法存取即時資訊：除非有特別功能，否則無法上網查資料
- 數學運算可能出錯：複雜計算不是 AI 的強項
- 缺乏真正的理解：是模式匹配，不是真正「懂」
- 可能有偏見：訓練資料的偏見會反映在輸出中

疾管署情境：AI 可以做什麼？

✓ ✓ AI 可協助

- ✓ 撰寫疫情週報摘要
- ✓ 整理防疫會議紀錄大綱
- ✓ 翻譯 WHO 疫情報告
- ✓ 草擬衛教宣導文案
- ✓ 彙整防疫指引重點

✗ ✗ 不適合使用 AI

- ✗ 處理含個資的疫調報告
- ✗ 查詢即時疫情數據
- ✗ 唯一的醫學診斷依據
- ✗ 處理機密文件內容
- ✗ 發布未審核的疫情資訊

AI「幻覺」（Hallucination）是什麼？

AI 有時會以非常有自信的語氣，描述根本不存在的事實。

AI 說：「台灣 2023 年登革熱確診共 12,847 例，主要集中於南部縣市.....」

⚠ **數字務必至疾管署開放資料平台查證，絕不能直接引用！**

三個辨識幻覺的跡象：

1 給出非常精確的數字，卻無法說明來源

2 引用看起來合理但查不到的文獻或法規

3 面對追問時，答案前後矛盾或越說越不對

不可輸入 AI 的資料類型

資料類型	範例	風險
個人資料	姓名、身分證字號、電話、地址	違反個資法
醫療資訊	病歷、診斷、用藥紀錄	涉及隱私與機密
機密文件	內部簽呈、未公開政策	洩漏公務機密
敏感資料	接觸者名單、疫調細節	影響疫調與隱私

AI 資安原則

🔒 姓名、身分證字號、地址、電話——這些絕對不能輸入 AI。

違規後果：可能違反個人資料保護法，導致法律責任。

AI 資安原則

🔒 尚未公開的疫情數據、內部策略討論、個案追蹤資料，
都不適合使用 AI 處理。

記住：AI 輸入的文字可能用於模型訓練。

AI 資安原則

🔒 已公開的防疫指引、衛教內容、一般性業務文書，
可以安全使用 AI 協助撰寫、整理、翻譯。

原則：官網上找得到的，通常可以輸入 AI。

AI 資安原則

🔒 任何 AI 生成的內容，在正式使用前都必須經過人工確認。

AI 的輸出是「草稿」，不是「定稿」。
最終決策權永遠在您手上。

資安判斷流程卡（可截圖存檔）

這份資料有個人資料？	→ 是 絕不輸入 AI
這份資料是內部機密？	→ 是 絕不輸入 AI
兩者皆否？	→ 可使用，但產出仍需人工審核

最簡單原則：「不確定能不能輸入，那就不要輸入。」

安全使用 AI 的五個好習慣

- 去識別化：移除所有可辨識個人的資訊後再使用
- 使用假資料：練習時使用虛構的範例資料
- 不留存敏感對話：處理完畢後清除對話紀錄
- 遵守機關規定：依照署內資安政策使用 AI 工具
- 審核再發布：AI 產出內容必須經過審核才能對外使用

模組 1 測驗重點回顧

Q1　AI 幻覺是指什麼？

A　AI 可能編造看起來合理但其實不正確的資訊

Q2　哪種資料「可以」輸入 AI？

A　疾管署官網上已公開的疫情週報

Q3　同仁想用 AI 整理疫調報告，應先做什麼？

A　先移除所有個人資料（去識別化）後再使用

Q4　AI 生成的衛教文案，正確做法？

A　經專業人員審核確認後才能發布

您已掌握了：

✓ 生成式 AI 的運作原理
✓ AI 的能力邊界與幻覺現象
✓ 疾管署 AI 資安四條原則

接下來：

模組 2　Prompt 四要素
掌握 RTFC 框架，AI 輸出品質差異立現

完成單元測驗，確認您已掌握資安原則

▶ 前往模組 2：Prompt 四要素