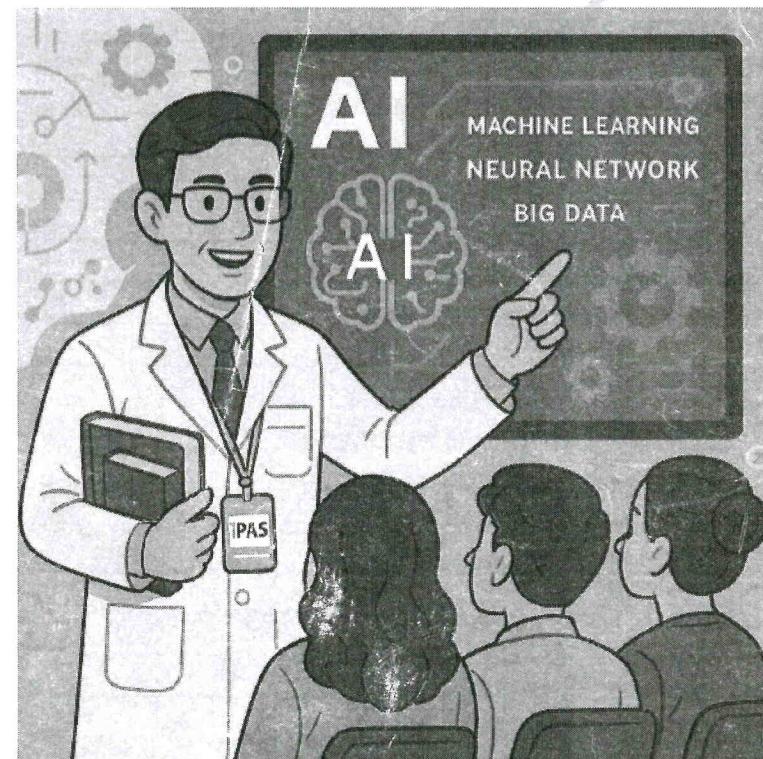


| 科目名稱 | 學習單元 | 學習項目 | 內容重點 |
|-------------------|---------------------------|---------------------------------|---|
| L12 生成式AI應用與規劃 | L121 No code / Low code概念 | L12101 No Code / Low Code的基本概念 | No code / Low code 基礎概念與技術比較 |
| | | L12102 No Code / Low Code的優勢與限制 | 1. 生成式 AI 與 No code/Low code 的整合應用場景 2. No code/Low code 平台的選擇與評估 |
| | L122 生成式AI應用領域與工具使用 | L12201 生成式 AI 應用領域與常見工具 | 1. 生成式 AI 基本概念與核心技術 2. 生成式 AI 工具應用發展與常見生成式 AI 工具介紹 3. 生成式 AI 於各領域應用發展 4. 生成式 AI 應用時面臨挑戰與風險管理 |
| | | | 善用生成式 AI 的策略與技巧 |
| | L123 生成式 AI 導入評估規劃 | L12301 生成式 AI 導入評估 | 1. 生成式 AI 的商業價值與應用前景 2. 生成式 AI 導入評估框架與標準，如技術或工具效能評估、適用解決方案選擇、成本效益分析等；經濟部產業發展署《AI 導引》。 |
| | | | 1. 生成式 AI 導入的規劃流程六個階段：需求與目標等 2. 生成式 AI 導入實施與運營 |
| | | L12303 生成式 AI 風險管理 | 如生成式 AI 偷理風險架構、關鍵風險分類、風險管理政策及法規、風險矩陣、資料安全隱私與合規性等 |

iPAS AI 應用規劃師-初級能力鑑定

人工智慧基礎概論、生成式 AI 應用與規劃

學習重點筆記(新版)



iPAS 教學小老師
iPAS Teacher



目 錄

| | |
|---|----|
| L11 人工智能基礎概論..... | 2 |
| L111 人工智能概念..... | 2 |
| L11101 AI 的定義與分類..... | 2 |
| L11102 AI 治理概念..... | 8 |
| L112 資料處理與分析概念..... | 13 |
| L11201 資料基本概念與來源..... | 13 |
| L11202 資料整理與分析流程..... | 19 |
| L11203 資料隱私與安全..... | 26 |
| L113 機器學習概念..... | 30 |
| L11301 機器學習基本原理..... | 30 |
| L11302 常見的機器學習模型..... | 32 |
| L114 鑑別式 AI 與 生成式 AI 概念..... | 37 |
| L11401 鑑別式 AI 與生成式 AI 的基本原理..... | 37 |
| L11402 鑑別式 AI 與生成式 AI 的整合應用..... | 41 |
| L12 生成式 AI 應用與規劃..... | 45 |
| L121 No code / Low code 概念 | 45 |
| L12101 No Code / Low Code 的基本概念 | 45 |
| L12102 No Code / Low Code 的優勢與限制 | 48 |
| L122 生成式 AI 應用領域與工具使用..... | 51 |
| L12201 生成式 AI 應用領域與常用工具..... | 51 |
| L12202 如何善用生成式 AI 工具..... | 56 |
| L123 生成式 AI 導入評估規劃..... | 60 |
| L12301 生成式 AI 導入評估 | 60 |
| L12302 生成式 AI 導入規劃..... | 64 |
| L12303 生成式 AI 風險管理..... | 68 |
| 附錄 | 73 |

iPAS AI 應用規劃師考試學習配當及學習重點表

| 科目名稱 | 學習單元 | 學習項目 | 內容重點 |
|-----------------|-------------------------|-----------------------------|---|
| L11 人工智能基礎概論 | L111 人工智能概念 | L11101 AI 的定義與分類 | AI定義與發展歷史 |
| | | L11102 AI 治理概念 | AI 治理與倫理概念：如AI框架、歐盟AI ACT、數位發展部《公部門人工智能應用參考手冊》、金融監督管理委員會《金融業運用人工智能(AI)指引》、《製造業AI導入指引》、台灣人工智能法草案、美國NIST AI.600-1等國內外相關政策法規等。 |
| | L112 資料處理與分析概念 | L11201 資料基本概念與來源 | 1. 資料基本概念(定義與分類) 2. 資料處理的基本方法，如大數據、資料型態與結構，如數值型資料、文字資料、圖像資料等。 |
| | | L11202 資料整理與分析流程 | 1. 資料收集、清理、分析流程與統計與呈現方法 2. 資料分析工具簡介與應用案例 |
| | | L11203 資料隱私與安全 | 1. AI 資料隱私與安全概論 2. 資料安全風險與保護方法 3. 法規與標準 |
| | L113 機器學習概念 | L11301 機器學習基本原理 | 機器學習基本原理與架構 |
| | | L11302 常見的機器學習模型 | 1. 機器學習類型與常見的模型 2. 機器學習資料處理、模型訓練分類與評估 3. 機器學習應用案例 |
| | L114 鑑別式 AI 與 生成式 AI 概念 | L11401 鑑別式 AI 與生成式 AI 的基本原理 | 1. 鑑別式 AI 與生成式 AI 基本原理與比較 2. 鑑別式 AI 與生成式 AI 基本核心技術介紹 |
| | | L11402 鑑別式 AI 與生成式 AI 的整合應用 | 鑑別式 AI 與生成式 AI 整合應用案例分享 |

- AI 需持續監測與調整，以適應新的生產變數。
- 透過定期回訓 AI 模型，確保準確率不下降。

◆ 第六章：AI 風險與挑戰

◆ AI 導入挑戰

| 挑戰類型 | 解決方案 |
|---------|------------------------|
| 數據質量問題 | 建立標準化數據流程，確保數據一致性 |
| AI 黑箱問題 | 使用可解釋 AI 模型（如決策樹） |
| 資安風險 | 資料加密、存取權限管理，防止 AI 數據外洩 |

◆ 總結

- ◆ AI 能提升製造業競爭力，但需謹慎導入與監管。
- ◆ 數據品質決定 AI 成功與否，企業需確保數據正確性。
- ◆ AI 需與既有生產系統整合，確保運行穩定。
- ◆ 持續監測 AI 模型，確保其適應不斷變化的生產環境。

L11 人工智慧基礎概論

L111 人工智慧概念

L11101 AI 的定義與分類

一、人工智慧的歷史演進 (AI Historical Evolution)

人工智慧的發展深受數學、邏輯、心理學、神經科學與計算科學交織影響。重要里程碑如下：

| 時期 | 階段名稱 | 發展特徵 | 關鍵技術 / 事件 |
|-----------|------------------------|----------------------------------|---|
| 1950s | 起源與哲學基礎 | AI 理論草創、數學邏輯與人機思考 | Turing Test, Dartmouth Conference (1956) |
| 1960s–70s | 符號主義時期 (Symbolic AI) | 規則導向知識推理、專家系統興起 | SHRDLU, ELIZA, MYCIN |
| 1980s | 專家系統繁榮 | 使用 IF-THEN 推論，知識庫與推理引擎分離 | XCON (DEC)、知識工程 (Knowledge Engineering) |
| 1990s | 機器學習轉向 | 資料導向學習，避免過度依賴人工規則 | Decision Trees、SVM、Bayesian Networks |
| 2006–2011 | 深度學習突破 | Hinton 提出深度置信網路、GPU 開啟訓練效能革命 | Deep Belief Network, ReLU, GPU |
| 2012–至今 | 基礎模型與生成式時代 | CNN、Transformer 興起，應用大規模語料學習通用模型 | AlexNet、BERT、GPT、ChatGPT、DALL-E |

【學術補充】：Alan Turing 於 1950 年提出的圖靈測試為「機器是否能思考」的經典問題，至今仍為 AI 哲學根基之一。

二、AI 的定義與本質解析 (Definition and Essence of AI)

(一) 綜合定義

AI 是一種模擬人類智能行為的技術體系，具備「感知 → 理解 → 推理 → 決策 → 學習 → 行動」之循環能力，藉由資料、演算法與模型實現可持續優化的功能。

(二) 定義分層解析 (參考 NIST、數位部與歐盟 AI ACT)

| 層次 | 說明 |
|--------------------|-----------------------------|
| 弱 AI (Narrow AI) | 專注特定任務的 AI，如語音辨識、圖像分類等 |
| 強 AI (AGI) | 具備類似人類智慧的通用學習與推理性能力（仍屬理論階段） |
| 超 AI (ASI) | 超越人類智力的假想系統，牽涉倫理與哲學挑戰 |

三、AI 的智能性與功能性分類 (Intelligence & Functional Classification)

(一) 智能性分類 (依據目標與能力)

| 分類 | 定義 | 技術基礎 | 實例 |
|---------------------------------|-----------------|-------------------|--------------|
| 反應型 AI (Reactive Machines) | 無記憶能力，僅對輸入即時反應 | CNN、單層 ANN | Deep Blue |
| 有限記憶 AI (Limited Memory) | 可利用過去資訊優化決策 | RNN、LSTM、RL | 自動駕駛車輛 |
| 理論心智 AI (Theory of Mind) | 可理解他人情感與意圖（研究中） | Affective AI、情緒計算 | 尚無實現 |
| 自我意識 AI (Self-aware AI) | 擁有自我意識與主觀性（假想） | - | AGI/ASI 未來概念 |

（二）功能性分類（依據任務功能）

1. 感知系統（Perception System）

- 資料輸入來源：語音、影像、文字、環境訊號。
- 技術模組：Computer Vision、Speech Recognition、Sensor Fusion。

2. 認知與表徵系統（Cognitive Representation）

- 轉化為可計算格式的知識（如向量、符號、圖結構）。
- 核心技術：Word Embedding、Knowledge Graph、Ontologies。

3. 學習機制（Learning Mechanism）

- 自主從資料中擷取規則與結構。
- 分類：
 - 監督式學習：目標標記資料。
 - 非監督式學習：發掘資料內在結構。
 - 半監督與自監督式學習。
 - 強化學習：以獎勵為導向。

4. 推理與規劃能力（Reasoning & Planning）

- 從知識做出邏輯、機率或啟發式推導。
- 關鍵技術：
 - 符號推理（Symbolic Reasoning）
 - 機率圖模型（Bayesian Network）
 - 強化學習中的策略搜尋與規劃。

5. 行動決策系統（Action Decision System）

- 從推理由果轉化為動作指令或行為模式。
- 應用場景：智慧機器人、自駕車、產線自動化。

6. 語言與互動系統（Language & Interface）

- 包括自然語言處理（NLP）、多模態理解（文字 + 圖像）與語音生成。
- 模型代表：BERT、GPT、T5、Whisper。

◆ AI 解決方案

- AI 視覺檢測系統：
 - 透過機器學習 + 影像處理，分析產品表面瑕疵。
 - 光學識別（OCR）技術，自動檢測標籤錯誤。
- 優勢
 - 準確率提高至 98%
 - 檢測速度快（毫秒級）
 - 減少人工誤判

2. AI 在設備維護的應用

◆ 現狀

- 傳統維修方式
 - 事後維修：設備故障後才進行修復，導致停機損失。
 - 定期保養：未必能精準預測設備故障，可能增加維修成本。

◆ AI 預測性維護

- 透過感測器數據（溫度、震動、電流）+ AI 演算法，預測機台何時可能發生故障。
- 優勢
 - 停機時間減少 40%
 - 維修成本下降 30%
 - 生產效率提升 20%

◆ 第四章：AI 模型開發與訓練

◆ AI 模型開發流程

① 數據收集與前處理

- 進行數據去除異常、標註，確保 AI 訓練品質。② 模型選擇與訓練
- 使用監督式學習訓練 AI 缺陷分類模型。③ 模型驗證
- 準確率測試（Accuracy）
- 召回率（Recall，防止錯過瑕疵品）④ 部署與優化
- 持續監測 AI 準確性
- 重新訓練 AI 模型，適應新產品

◆ 第五章：AI 系統導入與維運

◆ AI 導入流程

① 小規模試點（PoC, Proof of Concept）

- 先在單一產線或特定設備測試 AI，驗證可行性。
- 驗證 AI 模型準確度，確保能解決實際問題。

② 全廠導入

- 試點成功後，逐步擴展到整個工廠。
- 設計 AI 整合架構，確保與現有 IT/OT 系統相容。

③ 持續監測與優化

◆ AI導入前的四大準備

1. 明確AI應用目標

- 企業需釐清導入AI的核心問題：
 - 產品良率過低？(→AI品質檢測)
 - 設備維護成本過高？(→AI預測性維護)
 - 產線生產效率不穩定？(→AI智慧排程)
 - 庫存管理混亂？(→AI供應鏈優化)

2. 數據準備

◆ 主要數據來源

| 數據類型 | 範例 | 應用場景 |
|--------|-------------|------------|
| 感測器數據 | 溫度、震動、壓力 | 預測性維護、異常偵測 |
| 生產數據 | 產能利用率、機台稼動率 | 智慧排程、生產優化 |
| 供應鏈數據 | 訂單、原料供應狀況 | 需求預測、物流規劃 |
| 品質檢測數據 | 影像、尺寸測量結果 | AI品質檢測 |

◆ 數據管理與清理

- 數據標準化：確保數據一致性，避免錯誤影響AI訓練。
- 數據清理：移除異常數據點，提升AI訓練效果。
- 資料存取權限：確保數據安全，避免未經授權的AI訓練。

3. AI技術選擇

| 技術類型 | 應用領域 | 適用場景 |
|--------|------------|----------------|
| 監督式學習 | 品質檢測、異常分類 | 需要標註數據進行AI訓練 |
| 非監督式學習 | 異常偵測、供應鏈分析 | 在無標註數據下找尋模式 |
| 強化學習 | 自動化控制、智慧排程 | AI透過試錯方式學習最佳策略 |

4. 成本與效益分析

- AI成本項目：
 - 軟體開發(AI訓練、測試)
 - 硬體設備(伺服器、邊緣運算)
 - 專業人才(數據科學家、機器學習工程師)
- 效益估算：
 - 減少30%停機時間(AI預測性維護)
 - 產品不良率降低50%(AI品質檢測)
 - 物流成本下降20%(AI供應鏈優化)

◆ 第三章：AI應用場景與案例

◆ AI應用案例解析

1. AI在品質檢測的應用

◆ 現狀

- 目前製造業仍高度依賴人工檢測，準確率受主觀因素影響。

(三)四種定義視角(依據Russell & Norvig分類)

| 類型 | 目標 | 舉例 |
|-------|---------------|--------------------------|
| 像人類思考 | 模擬人類的認知過程 | 認知建模(Cognitive modeling) |
| 像人類行為 | 複製人類在任務上的行為結果 | 圖靈測試導向 |
| 理性思考 | 符合邏輯、最佳決策方式 | 邏輯推論、專家系統 |
| 理性行為 | 在任務中達成最適表現 | 自主行動者、機器人系統 |

四、人工智慧技術分類(AI Technical Taxonomy)

(一)機器學習(Machine Learning, ML)

◆ 定義：機器學習是一種讓電腦系統能夠從資料中學習規律，並根據經驗自動改進的演算法集合。

◆ 核心子類別與模型細節：

| 子類型 | 核心概念 | 常見演算法 | 應用範例 |
|--------|------------------------|------------------------------------|------------------|
| 監督式學習 | 給定輸入與對應標籤，學習輸入與輸出之映射關係 | 線性回歸、邏輯回歸、SVM、決策樹、隨機森林、KNN、XGBoost | 疾病診斷、信用風險預測 |
| 非監督式學習 | 僅給輸入資料，尋找潛在模式 | K-means、層次式分群、PCA、ICA、Autoencoder | 客戶分群、維度縮減、異常偵測 |
| 半監督式學習 | 部分資料有標籤，搭配大量無標資料訓練 | Pseudo-label、Consistency Training | 醫療資料分析、低資源語言處理 |
| 強化學習 | 在與環境互動中累積獎勵，學習最佳決策策略 | Q-Learning、SARSA、DQN、PPO、A3C | 機器人導航、遊戲AI、自駕車控制 |

(二)深度學習(Deep Learning, DL)

◆ 定義：深度學習是一種以多層神經網路為基礎的學習方法，能自動學習高階抽象特徵。

◆ 基本結構：人工神經網路(ANN)

- 神經元運算公式：

$$y = \sigma(\sum_{i=1}^n W_i X_i + b)$$

其中 σ 是啟用函數(如ReLU, Sigmoid)。

◆ 關鍵架構與應用：

| 架構 | 特性 | 技術細節 | 應用範疇 |
|---------------------------------|-------------|-----------------|----------------|
| CNN (卷積神經網路) | 擅長處理圖像與局部特徵 | 卷積層+池化層+全連接層 | 影像分類、物件偵測、醫學影像 |
| RNN / LSTM / GRU ① ② ③ | 適用時間序列與語言模型 | 記憶單元可保留序列上一下文資訊 | 語音辨識、文字生成 |

① 卷積神經網路
 ② 長短時記憶網路，也有一種循環神經網路，可解決傳統RNN的問題
 ③ 可以修正LSTM的割縫程度

| | | | |
|--|---------------------------|--------------------------------|------------------|
| Transformer | 高度並行、長距 離關聯建模 | Attention 機制、 BERT/GPT 基礎架構 | NLP、CV、跨模態 AI |
| Autoencoder / VAE <i>自動編碼器 / 变分自编码器</i> | 表示學習與生成 <i>表示学习与生成</i> | 編碼器-解碼器結構，學 習資料壓縮與重建 | 異常偵測、影像重建 |
| GNN <i>(圖神經網路)</i> | 適用圖結構資料 | 節點鄰接資訊傳遞與聚 合 | 社群網路、蛋白質結 構預測 |

(三)自然語言處理 (Natural Language Processing, NLP)

◆ 任務分類

| 任務類型 | 說明 | 對應技術 |
|----------------|---------------|---------------------------|
| 文本分類 | 情感分析、新聞分類 | Naive Bayes、BERT |
| 命名實體辨識 (NER) | 辨識人名、地點等實體詞 | CRF、BERT、SpaCy |
| 機器翻譯 | 將語言 A 翻譯為語言 B | Seq2Seq、Transformer、mBART |
| 問答系統 | 理解問題並從文本中回答 | BERT、T5、GPT |
| 生成式任務 | 自動寫作、摘要生成 | GPT-3/4、T5、LLM |

◎ 語言模型發展：

- 統計式語言模型：n-gram、HMM
- 向量化模型：Word2Vec、GloVe、FastText
- Transformer 系列：BERT (Masked LM)、GPT (Causal LM)、T5 (Text-to-Text)

(四)電腦視覺 (Computer Vision, CV)

◆ 核心任務與模型：

| 任務 | 對應模型 | 技術細節 |
|------|-----------------------------|---------------------------------|
| 影像分類 | CNN、ResNet、 EfficientNet | 卷積提取特徵、分類器輸出標籤 |
| 物件偵測 | YOLO、Faster R-CNN | 矩形框偵測 + 分類 |
| 影像分割 | U-Net、DeepLab | 像素級標註 |
| 姿態辨識 | OpenPose、HRNet | 偵測人體關鍵點位置 |
| 影像生成 | GAN、Diffusion Models | 生成高品質新圖像 (如 Stable Diffusion) |

(五)語音 AI (Speech AI)

◆ 任務與技術

| 任務 | 對應模型 | 說明 |
|--------------|-----------------------------|-----------|
| 語音辨識 (ASR) | CTC、DeepSpeech、Whisper | 將語音轉為文字 |
| 語音合成 (TTS) | Tacotron、WaveNet、FastSpeech | 文字轉語音 |
| 語音情緒辨識 | CNN、RNN、SVM | 辨識說話者情緒狀態 |

◆ 目的

- 確保 AI 決策透明可解釋，提升市場信任。

◆ 控制措施

1. 提升透明度

- 提供 AI 決策過程說明 (避免黑箱操作)。
- 對客戶與監管機關揭露 AI 運作方式。

2. 提升可解釋性

- 簡化模型決策邏輯，降低黑箱模型風險。
- 提供 AI 決策證據，確保監理機關可查核。

◆ 第六章：促進永續發展

◆ 目的

- 降低 AI 耗能，減少環境影響，促進 ESG 目標。

◆ 控制措施

- 優化 AI 硬體與演算法，減少能源浪費。
- 減少數位落差，確保 AI 服務公平普及。
- 員工再培訓，減少 AI 取代人力帶來的失業風險。

◆ 總結

- 金融機構應妥善管理 AI，確保風險可控。
- AI 需公平、透明、可解釋，並受人類監督。
- AI 需符合資安與隱私保護，降低 ESG 影響。

■ 三、《製造業 AI 導入指引》

◆ 第一章：AI 在製造業的應用背景

◆ AI 在製造業的重要性

- 全球製造業邁向工業 4.0，AI 與物聯網 (IoT)、大數據、雲端運算等技術結合，提高智能化程度。
- AI 應用可帶來成本降低、生產效率提升、產品品質改善等競爭優勢。

◆ AI 在製造業的主要價值

| 應用領域 | AI 技術應用 | 效益 |
|--------|--------------|-----------------|
| 品質檢測 | AI 視覺檢測、影像識別 | 準確檢測產品缺陷，降低不良率 |
| 設備維護 | 預測性維護、異常偵測 | 提前預測機台故障，減少維修成本 |
| 生產流程優化 | 智慧排程、動態資源分配 | 提高生產效率，減少停機時間 |
| 供應鏈管理 | AI 需求預測、物流優化 | 減少庫存成本，提升供應鏈效率 |
| 員工安全 | AI 影像分析、安全監測 | 監控工作環境，降低工安事故 |

◆ 第二章：AI 導入的準備與評估

- 防止 AI 偏見，確保公平性。
- 確保 AI 決策尊重基本人權，可受人類監督。
- ◆ 風險與控制措施
 1. 避免 AI 偏見
 - 審查數據來源，防止特定群體被系統性歧視。
 - 建立公平性指標，監測 AI 對不同族群影響。
 - 定期驗證模型輸出結果，確保公平性。
 2. 保障人類監督
 - 設計 AI 決策機制時，確保人工干預權限。
 - 提供 AI 風險通知，確保決策過程透明。
 3. 公平性檢驗
 - 內部與第三方公平性稽核，確保 AI 不影響特定群體。

◆ 第三章：隱私保護與客戶權益

- 目的
 - AI 需遵守個資法，確保客戶隱私不被濫用。
 - 尊重客戶選擇權，提供 AI 服務的替代方案。
- 風險與控管措施
 1. 個資保護
 - 資料最小化（只蒐集必要資訊）。
 - 加密與匿名化，防止個資洩漏。
 - 審查第三方 AI 服務商，確保資料安全。
 2. 透明與選擇權
 - 告知客戶 AI 參與決策的程度。
 - 提供人工處理選項，避免 AI 強制決策。
 - 若 AI 影響客戶，應提供申訴與救濟機制。

◆ 第四章：確保 AI 穩健性與安全性

- 目的
 - 確保 AI 運作穩定、安全可靠，防止市場風險。
- 風險控管
 1. 系統穩健性
 - 交叉驗證與壓力測試，確保 AI 在異常情境下仍可運行。
 - 異常檢測機制，即時發現 AI 決策異常。
 2. 實安風險
 - 加強 API 安全性，防止駭客入侵。
 - 限制 AI 學習來源，防止數據污染。

◆ 第五章：透明性與可解釋性

| | | |
|------|-------------------|---------|
| 語者辨識 | X-vector、i-vector | 辨識說話者身份 |
|------|-------------------|---------|

(六)生成式 AI (Generative AI)

◆ 模型類型與原理比較

| 模型 | 原理 | 優勢 | 弱點 |
|-------------------|---------------------------------|------------|------------|
| GAN (生成對抗網路) | 對抗學習：Generator vs Discriminator | 高品質生成，訓練直覺 | 不穩定、模式崩壞 |
| VAE (變分自編碼器) | 機率生成 + 壓縮表徵 | 結構穩定 | 成品模糊 |
| Diffusion Models | 噪聲逐步還原生成 | 穩定、高品質 | 推理速度慢 |
| LLM (大型語言模型) | 預測下個 token | 靈活、語言能力強 | 事實控制難、偏見問題 |

(七)其他關鍵應用型 AI 技術

1. 推薦系統 (Recommendation Systems)

- 協同過濾 (Collaborative Filtering) : ALS、SVD++
- 內容過濾 (Content-Based) : TF-IDF、Embedding
- 深度推薦：DeepFM、DIN、Transformer4Rec

2. 多模態 AI (Multimodal AI)

- 融合語言、圖像、語音資訊（如 CLIP、DALL-E、Flamingo）
- 核心技術：Cross-Attention、多通道融合、視覺語言對齊

五、人工智慧的應用方式與案例解析 (AI Applications & Scenarios)

(一) 系統導入流程 (參考公部門參考手冊)

◆ AI 導入五階段流程圖：



(二) 各產業應用實例表

| 產業 | 功能 / 目標 | 對應技術 | 實例 / 引用 |
|------|-------------|---|------------------|
| 金融業 | 詐欺偵測、信用評分 | LSTM、XGBoost | 金管會 AI 指引 |
| 公部門 | 民眾服務、法庭語音轉錄 | NLP、ASR <small>自然語言處理</small> | 財政部智慧客服、司法院語音辨識 |
| 製造業 | 預測維護、品管 | IoT+CV+ML | AI 故障檢測與異常預警 |
| 醫療健康 | 影像診斷、病歷摘要 | CNN <small>卷積神經網絡</small> BERT <small>文字分類與命名 entity</small> | AI 判讀 X 光、病歷摘要系統 |

① 卷積神經網絡
② 文字分類與命名 entity

六、鑑別式 AI 與生成式 AI 比較分析 (Discriminative vs Generative AI)

(一) 基本概念與核心差異

| 類別 | 鑑別式 AI (Discriminative AI) | 生成式 AI (Generative AI) |
|------|---|--|
| 核心任務 | 分類與預測 | 創造與模擬 |
| 學習目標 | 建立輸入與標籤之間的條件機率： $P(Y X)$ | |
| 主要技術 | Logistic Regression、SVM、Random Forest、ResNet、BERT | GAN、VAE、Diffusion Models、GPT、T5、DALL-E (手繪圖) |
| 運作方式 | 學習如何劃分不同類別的邊界 | 學習資料分布，據以生成類似內容 |
| 資料需求 | 需大量標記資料 (Label) | 可用大量未標記資料 (自監督學習)，或配合少量標籤進行微調 |

T5是一種大型
語言模型

(二) 核心技術原理解釋

1. 鑑別式 AI 技術原理：

- 支援向量機 (SVM)：尋找分類邊界最大化的超平面。
- 卷積神經網路 (CNN)：對圖像進行特徵抽取後分類。
深度學習
- BERT (NLP 模型)：使用編碼器架構進行語意分類任務。

重點：關注分類邊界 → 哪個類別「最有可能」是答案，而非學習整體數據結構。

2. 生成式 AI 技術原理：

- GAN (Generative Adversarial Network)：生成器與鑑別器對抗訓練，生成逼真數據。
- VAE (Variational AutoEncoder)：透過機率潛在空間學習資料生成模型。
- Transformer + Decoder 架構 (GPT、T5)：透過上下文預測下一個字元或單詞。

重點：學習整體資料分佈 → 能夠「模仿」原始數據來生成新內容。

(三) 優缺點比較

| 項目 | 鑑別式 AI | 生成式 AI |
|------|---------------------|---|
| 優點 | 模型簡潔、訓練快速、預測準確 | 具創造力、適用多模態 (圖文音影)、支援無監督與自監督學習 |
| 缺點 | 無法生成內容、依賴大量標記資料 | 訓練不穩定 (如 GAN 模型)、容易出現幻覺 (hallucination) 或偏差 |
| 運算成本 | 通常較低 | 通常較高 (尤其訓練階段) |
| 可解釋性 | 高 (部分模型如邏輯回歸、決策樹) | 較低，屬於黑盒模型 |
| 常見錯誤 | 過擬合、偏誤分類 | 假資訊、錯誤生成、風險管理難度大 |

◆ 總則章：AI 生命週期與風險評估

◆ AI 生命週期四大階段

1. 系統規劃與設計

- 設定明確目標 (如信用評分、詐欺偵測)。
- 訂定風險管理框架 (影響評估、救濟機制)。
- 設計數據蒐集策略 (確保公平、去偏見)。

2. 資料蒐集與輸入

- 確保數據品質 (完整性、一致性、準確性)。
- 檢視數據是否有偏見 (如年齡、性別、地區)。
- 訂定個資保護機制 (加密、匿名化)。

3. 模型建立與驗證

- 選擇合適演算法 (監督學習、非監督學習)。
- 訓練與驗證模型，確保準確率、可靠性。
- 進行交叉驗證 (cross-validation)，減少過擬合。

4. 系統部署與監控

- 建立風險監測機制 (模型漂移偵測)。
- AI 決策出錯時，確保有人工干預機制。
- 進行壓力測試 (stress testing)，模擬極端情境。

◆ 第一章：建立治理及問責機制

◆ 目的

- 建立 AI 風險管理架構，確保符合法規與倫理。
- 內外部問責機制，確保 AI 影響可監督與追蹤。

◆ 主要機制

1. 內部治理

- 指定 AI 監督高階主管或委員會。
- 設立跨部門 AI 風險管理機制 (風控、法遵、資安)。
- 確保 AI 風險管理策略與內部控制相符。

2. 風險管理

- 設立 AI 風險管理政策 (公平性、透明度、安全性)。
- 監測 AI 準確度、決策一致性，防止模型劣化。
- 內部審查與外部獨立驗證，確保風險管控。

3. 人員培訓

- 管理階層：理解 AI 影響、監管方式。
- 開發與風控人員：熟悉 AI 風險、數據偏見處理。
- 客服與消費者教育：確保 AI 透明與客戶知情權。

◆ 第二章：公平性與人本價值

◆ 目的

3. 法庭語音辨識（司法院）：AI 轉錄審判紀錄，提升法庭效率。
4. 銀髮安居計畫（內政部）：AI 分析老人需求，提高社福效率。
5. AI 預測颱風強度（中央氣象署）：AI 分析氣象數據，提高預測準確度。

◆ 國外案例

1. 新加坡 AI 聊天機器人：簡化民眾市政陳情流程。
2. 新加坡 AI 求職推薦：根據求職者技能推薦適合職缺。
3. 南韓 AI 照護機器人：幫助獨居老人，提高安全性。
4. 南韓 AI 教學機器人：幫助銀髮族學習智慧手機。

◆ AI 相關法規

◆ 重要法律

- 個資法、政府資訊公開法、資通安全管理法。
- 人工智慧基本法（草案）（2024年7月）。
- 金融業 AI 指引（金管會，2024年6月）。

◆ 結論

- AI 在公部門的應用核心：
 1. 提升行政效率
 2. 優化公共服務
 3. 確保數據安全與法規合規
- 政府 AI 推動策略：
 1. 建立標準化指引。
 2. 鼓勵 AI 創新應用。
 3. 強化 AI 風險管理與法規監管。

□ 二、《金融業運用人工智慧(AI)指引》

◆ 前言

◆ AI 在金融業的應用

- 提升效率：自動化決策、加速數據分析。
- 降低成本：減少人力成本、優化作業流程。
- 改善客戶體驗：個人化推薦、智慧客服、即時風險。
- 管理風險：AI 風險評估、詐欺偵測、信用評分。
- 促進合規：自動法遵檢測、異常交易監控。
- 資安防護：監控異常行為、應對駭客攻擊。
- 永續發展：降低能源消耗、促進普惠金融。

◆ AI 運用的風險

- 偏見與歧視：演算法可能無意間歧視特定族群。
- 隱私問題：蒐集過多個資，可能違反法規。
- 決策不可解釋：黑箱模型導致難以追蹤 AI 的決策邏輯。
- 系統安全性：AI 可能被駭入，造成市場風險。
- 市場信心：AI 錯誤決策可能影響投資者信心。

(四) 應用方式比較

| 領域 | 鑑別式 AI 應用 | 生成式 AI 應用 |
|-----|----------------|-----------------|
| 醫療 | 疾病分類、X 光影像判讀 | 合成病理圖像、病歷摘要生成 |
| 金融 | 客戶分群、信用評等、詐欺偵測 | 模擬金融場景、生成風險報告文本 |
| 零售 | 用戶行為預測、客戶流失預測 | 廣告文案、個人化商品描述生成 |
| 教育 | 學生學習類型分類、測驗預測 | 自動命題、個人化教學素材生成 |
| 製造業 | 異常偵測、良率預測 | 模擬產線流程、設計自動化 |
| 公部門 | 案件分類、服務需求預測 | 自動回覆市政陳情、合約摘要 |

L11102 AI 治理概念

一、AI 治理的定義

(一) 什麼是 AI 治理？

- AI 治理（AI Governance）指的是確保 AI 技術的發展、部署與應用符合倫理標準、法規要求與社會價值，以降低 AI 可能帶來的風險，促進 AI 的公平性、透明性與安全性。
- AI 治理不僅涉及技術層面（如模型可解釋性、風險監測），還涵蓋法律、道德與社會影響，確保 AI 安全可控，符合公共利益。

(二) AI 治理的主要挑戰

| 挑戰 | 說明 | 實例 |
|----------|-------------------------------|---|
| 數據隱私 | AI 需要大量數據進行訓練，但數據可能涉及個人隱私 | Facebook 涉及用戶數據濫用的 Cambridge Analytica 事件 |
| 公平性與歧視 | AI 可能會基於數據的偏見產生歧視性決策 | 招聘 AI 可能因歷史數據影響，對特定族群產生不公平待遇 |
| 透明度與可解釋性 | 許多 AI 模型（如深度學習）是「黑箱」，難以解釋決策過程 | AI 判定某人信用風險高，無法解釋原因 |
| 安全性與假訊息 | AI 可能被用來生成 Deepfake、假新聞，擾亂社會 | Deepfake 偽造政治人物影片，影響選舉 |
| 責任歸屬 | AI 決策錯誤時，應由誰負責？ | 自動駕駛車禍，應由車主、車廠或 AI 算法開發者負責？ |

二、AI 治理的核心原則(負責任的 AI)

(一) 什麼是負責任的 AI (Responsible AI)

負責任的 AI 是一種發展 AI 技術的倫理與治理框架，目的是確保 AI 系統在設計、訓練、部署與使用過程中，都符合倫理、法律、人權與社會價值。它是一個跨領域整合的概念，涵蓋：

- 科技倫理（Tech Ethics）
- 法規合規（Regulatory Compliance）
- 社會風險控制（Social Risk Mitigation）

- 組織治理 (Corporate AI Governance)

其重點在於讓 AI 的發展「不只是能做什麼，而是應該做什麼」。

(二) 負責任 AI 的六大核心原則

| 原則 | 說明 |
|--|---|
| 公平性 (Fairness) | AI 不得因訓練資料偏誤而對性別、種族、年齡等造成不公。→ 技術應包含 Bias 檢測與修正機制。 |
| 透明性 (Transparency) | 使用者應能了解 AI 如何做出決策 (模型可解釋性 Explainability)。→ 特別重要於醫療與金融。 |
| 問責性 (Accountability) | AI 出錯時須能追溯決策過程、釐清責任人 (責任鏈 Responsibility Chain)。 |
| 安全性與穩健性 (Safety & Robustness) | 須防止 AI 被操弄 (如 adversarial attacks)，並在不同條件下穩定運作。 |
| 隱私與資料治理 (Privacy & Data Governance) | AI 須依據法規 (如 GDPR、個資法) 保護使用者數據。 |
| 人類監督 (Human Oversight) | 對於關鍵任務的決策 (如司法、醫療診斷) 仍應保有人類監控或介入能力。 |

◆ 許多國際組織 (如 OECD、IEEE、EU、UNESCO) 皆以這六項為核心架構來制訂負責任 AI 的指導原則。

(三) 實務應用層面展開

1. 組織內部治理制度 (AI Governance Framework)

- 設立 AI 倫理委員會或內控審查單位。
- 導入 AI 專案需進行「風險評估」與「倫理衝擊評估」。
- 每個 AI 模型需留下可追蹤的版本紀錄、模型審核記錄 (Model Card)。

2. 技術實踐層面

| 領域 | 實務作法 |
|----------|--|
| 偏誤偵測與修正 | 使用 Fairness 指標 (如 disparate impact ratio) 分析訓練資料與模型結果。 |
| 模型可解釋性 | 導入 LIME、SHAP 等技術協助模型解釋。 |
| 風險評估工具 | 使用 AI Risk Map、NIST AI RMF 評估模型影響。 |
| 資料匿名化與控管 | 建立 PII 控制機制，並記錄資料處理流程與授權依據。 |

三、AI 風險管理與監控

AI 風險管理是 AI 治理的重要環節，涉及 數據管理、模型監測、偏見檢測、資安防護 等方面。

AI 風險管理的核心方法

| | |
|-----------|-------------------|
| 影像辨識與物件追蹤 | 保證所有數據皆無偏見 |
| 自動語言辨識與翻譯 | 解釋 AI 運作方式 (黑箱問題) |

- AI 限制：

- 數據依賴性：若數據品質差，AI 可能產生錯誤結果。
- 無法完全取代人工：AI 主要用於輔助決策，而非完全取代人類。
- 偏見與倫理問題：若訓練數據有偏見，AI 也可能產生不公正的結果。

◆ 第二章：AI 服務評估

● AI 導入生命週期 (五大階段)

- AI 場景評估：確認 AI 是否適合解決問題，避免「為 AI 而 AI」。
- AI 專案啟動：決定內部開發或採購模式，建立 AI 專案團隊。
- 資料探索與模型建立：確保數據品質，選擇適合的 AI 模型。
- 模型迭代與部署：進行測試與調整，確保 AI 能穩定運行。
- 風險控制與專案追蹤：監測 AI 效能，確保符合道德與法規。

● AI 適用場景評估

- 適合 AI 的問題：
 - 需要處理大量數據。
 - 具備標準化流程與重複性高的任務。
 - 涉及複雜數據分析與預測。
- 不適合 AI 的問題：
 - 涉及高度創新與策略性決策。
 - 需要人類情感與道德判斷。
 - 數據量不足或品質低。

◆ 第三章：AI 服務導入

● AI 導入模式

- 內部開發：適合技術成熟的機關，但成本較高。
- 外包採購：適合資源有限的機關，可快速導入 AI。

● AI 專案流程

- 專案計畫 → 資料收集 → 模型開發 → 測試與驗收 → 監測與優化。

◆ 第四章：AI 營運管理

● AI 風險管理

- 需關注 隱私、倫理、公平性、資料安全、模型透明度 等議題。

● AI 治理架構

- 設立 AI 監督機制，確保 AI 在政府機關內合法運行。

● 代表性 AI 應用案例

- 國內案例
 - 商標檢索系統 (智慧財產局)：AI 影像辨識，提高商標審查效率。
 - 稅務智慧客服 (財政部)：AI 聊天機器人回答民眾稅務問題。

附錄：《AI 實務導入與治理策略》（公部門・金融業・製造業篇）

一、□ 公部門人工智慧應用參考手冊

◆ 手冊背景與目的

本手冊由數位發展部撰寫，目的是幫助政府機關理解 AI 應用，提供 AI 導入流程指引，確保 AI 在公部門的有效運用。在推動 AI 時，政府機關常面臨三大痛點：

1. AI 知識不足

- 許多公務人員對 AI 缺乏基本理解，無法想像其潛在應用，影響導入意願與決策。
- 解決方式：提供科普教育與實作培訓，加強 AI 在政府內部的普及度。

2. 可參考案例分散

- 目前 AI 在公務機關的應用案例較少，缺乏整合，導致各單位無法參考類似專案來導入 AI。
- 解決方式：彙整 AI 在國內外政府機關的應用案例，建立共享資料庫。

3. 缺乏實作導向指引

- 現有的 AI 規範多為原則性、禁止性規範，欠缺具體的導入方法與實施步驟。
- 解決方式：制定詳細的 AI 導入手冊，從規劃、執行到驗收提供具體操作指引。

◆ 第一章：AI 概念介紹

AI 定義與類型

- AI（人工智慧）：模擬人類智慧的系統，透過數據學習進行決策與執行任務。
- 生成式 AI（GAI）：
 - 可自動產生文字、圖片、音樂、影片，如 ChatGPT、DALL-E。
 - 主要應用於客服、數據分析、內容生成、語言翻譯等。
- AI 的主要學習類型
 - 監督式學習（Supervised Learning）：透過已標註資料訓練 AI，例如 垃圾郵件分類、影像辨識。
 - 非監督式學習（Unsupervised Learning）：透過未標註資料找出規律，例如 顧客群組分析、詐欺偵測。
 - 半監督式學習（Semi-Supervised Learning）：結合部分標註資料，適用於醫療影像分析、語音辨識。
 - 強化學習（Reinforcement Learning）：AI 透過獎勵機制學習，例如 自動駕駛、AlphaGo。

AI 的幫助與限制

| 可協助的任務 | 無法達成的事項 |
|------------|---------------|
| 分析大數據、趨勢預測 | 完全自主創新 |
| 自動生成文本、摘要 | 沒有數據時提供高準確度分析 |

| 風險類型 | 風險描述 | 解決方案 |
|---------------------------------|---------------------|----------------------------------|
| 數據偏見 (Data Bias) | 訓練數據不均衡，導致 AI 產生歧視 | 使用公平性測試工具（如 IBM AI Fairness 360） |
| 模型漂移 (Model Drift) | AI 訓練後可能隨時間失準 | 定期監測 AI 表現，重新訓練模型 |
| 對抗攻擊 (Adversarial Attacks) | 惡意修改 AI 訓練數據，影響預測結果 | AI 防禦機制（如 adversarial training） |
| 隱私洩露 (Privacy Breach) | AI 學習個人敏感數據，可能外洩 | 差分隱私、聯邦學習 |
| 假訊息與 Deepfake | AI 生成虛假內容，影響社會輿論 | AI 驗證技術（如 Deepfake 檢測模型） |

四、AI 隱私與安全

(一) AI 隱私風險

1. 個人數據外洩：企業 AI 可能蒐集用戶的私密資訊（如語音、位置）。
2. 監控風險：政府與企業可能利用 AI 進行大規模監控，如臉識別技術。

(二) AI 安全防護技術

| 技術 | 功能 | 應用範例 |
|-----------------------------------|--------------------|-------------------------------|
| 差分隱私 (Differential Privacy) | 在數據中加入噪聲，保護個資 | Apple、Google 皆使用於 AI 訓練數據 |
| 聯邦學習 (Federated Learning) | AI 訓練時不傳輸數據，而在本地運行 | Google Android Gboard 自動學習輸入法 |
| 對抗樣本防禦 (Adversarial Defense) | 防止 AI 被惡意攻擊 | AI 安全研究領域 |

五、AI 法律與倫理框架

全球各國對 AI 進行監管，以確保 AI 技術符合倫理規範並保護社會利益。

(一) 國際 AI 法規與治理框架

● 歐盟《AI 法案》(EU Artificial Intelligence Act, AI ACT)

- 推出背景：2021 年歐盟提出，2024 年通過立法。
- 特點：全球首部「全面性 AI 分級管理法案」。

核心內容：

| 分級風險 | 說明 | 舉例 |
|------------|------------|----------------|
| 不可接受風險（禁用） | 嚴重侵犯人權 | 社會信用評分、即時大規模監控 |
| 高風險 | 涉及人命、自由、權益 | 醫療診斷、招募系統、自駕車 |
| 有限風險 | 須提供透明告知 | AI 聊天機器人 |
| 低風險 | 幾無風險限制 | 遊戲推薦、娛樂演算法 |

- 強制措施：要求高風險系統進行事前評估、風險管理、可解釋性與資料治理。

● US 美國 NIST《AI 風險管理架構》(NIST.AI.600-1)

- 發布時間：2023 年由美國國家標準技術研究院 (NIST) 制定。
- 目的：建立「可量化風險管理模型」以強化 AI 可信度。
- 核心架構 (四大支柱) ：
 - 治理 (Governance)：建立內部規範與責任機制。
 - 圖像 (Map)：定義 AI 系統範疇與背景風險。
 - 衡量 (Measure)：量化 AI 系統的偏誤與效能。
 - 管理 (Manage)：採取因應措施與風險調整。

(二)台灣主要 AI 政策與治理參考

■ 數位發展部《公部門人工智慧應用參考手冊》(2023)

- 目的：提供政府部門 AI 導入的實務準則。
- 六大治理面向：
 - 治理制度設計
 - 系統安全與風險管理
 - 倫理與人權考量
 - 資料品質與來源正當性
 - 模型訓練與監督機制
 - 溝通與公眾信任建立

■ 金融監督管理委員會《金融業運用人工智慧 (AI) 指引》(2022)

- 對象：銀行、保險、證券業使用 AI 模型。
- 重點要求：
 - 內部應設置 AI 管理小組。
 - 模型須保留訓練記錄以利稽核。
 - 應定期進行偏誤檢測、結果驗證。
 - 對消費者決策影響大者須「提供合理說明」。

■ 經濟部《製造業 AI 導入指引》

- 目標：促進中小製造業安全且有效導入 AI 。
- 治理核心：
 - 以生產流程為核心規劃 AI 導入節點。
 - 導入前須評估風險 (如品質偏差、資料安全) 。
 - 鼓勵導入 explainable AI (可解釋模型) 。

■ 台灣《人工智能基本法》草案 (數位部提出)

- 立法目的：促進 AI 合法發展並保障人權。
- 草案架構 (初稿內容摘要) ：
 - AI 定義與適用範圍
 - 風險分類與分級治理架構
 - 對公務與民間 AI 導入之責任要求

| | | |
|--------|---------------------------|-------------------------|
| 個資管理 | 不得使用含個資之資料進行開放模型訓練或傳輸 | 台灣個資法、GDPR |
| 日誌留存 | 必須記錄所有輸入提示語與模型回應紀錄 | 金融業 AI 指引、公部門 AI 手冊 |
| 回應可解釋 | 模型回應需能提供資料出處與回應邏輯 | NIST 、 AI ACT 第 52 條 |
| 人為審查制度 | 關鍵應用 (信用、醫療、司法) 需引入人為監督 | 台灣 AI 法草案、EU AI ACT |
| 模型來源揭露 | 商用模型需揭露訓練資料類型、來源、範圍 | AI ACT 第 28 條、CCAI 建議草案 |

■ (六)產業導入實例與應用建議 (依風險層級)

| 產業場域 | 高風險控制要點 | 建議治理工具 |
|------|--------------------------|--------------------------|
| 製造業 | 文件生成應加入品質驗證、提示語紀錄 | Prompt 設計規範 + RAG 結合 MES |
| 金融業 | 生成報告、信用風險分析須審核紀錄 + 問責人制度 | 內部審查流程 + 模型調控權限分級 |
| 醫療照護 | AI 給建議但禁止自動開立診斷，須醫師覆核 | Prompt 回應需標記「非醫療建議」 |
| 政府部門 | AI 回應須有出處、資料保證更新 | 搭配知識資料庫訓練 + 審查提示語 |

| 要點 | 說明 |
|--------|---|
| 盤點使用場景 | 分析用途：是否自動化決策？是否對客戶？是否公開？ |
| 分析資料風險 | 是否含個資？敏感資訊？是否跨境傳輸？ |
| 模型來源盤點 | 外部商用模型（如 GPT API）vs 自建模型（如 LLaMA）是否合約與授權清楚？ |

3. 【Measure】評估指標設計

| 指標類別 | 範例指標 | 工具建議 |
|-------|------------------------------|----------------------------|
| 準確性指標 | 正確率、幻覺率 (hallucination rate) | 人工標記測試資料組 |
| 偏誤指標 | 性別、年齡、語言偏誤檢測 | AI Fairness 360, Fairlearn |
| 解釋性指標 | 是否能提供提示語、資料來源、上下文分析 | SHAP、LIME、RAG 架構 |
| 使用指標 | 使用次數、修正次數、用戶滿意度 | 系統日誌、BI 報表 |

4. 【Manage】風險控管與持續監測

| 機制 | 說明 |
|---------|--------------------------------|
| 模型版本控制 | 建立版本管理制度，保存模型變更與訓練紀錄 |
| 內容審查流程 | 對高風險用途輸出設定審查閾值（如置信度 < 80% 要人審） |
| 使用記錄管理 | 保留提示語、回應、使用人紀錄，供事後稽核 |
| 使用者教育訓練 | 教育 AI 使用方式、風險辨識、錯誤回報流程 |

(四)風險防範機制工具建議（實務層）

| 工具類型 | 功能 | 適用風險 |
|----------------------------------|----------------------------|---------------|
| 提示語過濾機制 (Prompt Guardrails) | 過濾含有敏感指令、假訊息或違規輸入 | 誤用、內容風險 |
| RAG 架構 (檢索增強生成) | 引用內部知識庫資料以強化準確性與可溯源性 | 幻覺、決策錯誤 |
| 提示語範本庫 (Prompt Templates) | 為不同任務設計專屬格式，降低語意偏差 | 輸出品質不一致、風格不穩定 |
| 使用監控面板 (Dashboard) | 即時顯示生成任務統計、出錯率、回饋比率 | 持續評估與調校用 |
| 模型風險稽核流程 | 模型開發、上線、部署均需風險報告與合規審查 | 全流程風險管理 |
| 封閉部署與存取控管 | 高敏感資料任務禁止使用開放模型 API，改採私有部署 | 資料外洩、隱私侵害 |

(五)生成式 AI 合規性與制度建議

| 項目 | 建議內容 | 適用法規參考 |
|----|------|--------|
|----|------|--------|

- 權益救濟與資訊揭露機制
(草案尚在研擬，預計參考歐盟 AI ACT 架構進行修正)

● 國內外治理框架對照表

| 項目 | 歐盟 AI ACT | 美國 NIST | 台灣公部門手冊 | 金融 AI 指引 |
|---------|---|---|---|---|
| 分類風險模型 | <input checked="" type="checkbox"/> (四級) | 部分 | △ (建議性) | <input checked="" type="checkbox"/> (依據應用) |
| 法規約束力 | <input checked="" type="checkbox"/> 法規 | 建議性 | 建議性 | <input checked="" type="checkbox"/> 規範性 |
| 資料與模型責任 | <input checked="" type="checkbox"/> 必須揭露來源與偏誤處理 | <input checked="" type="checkbox"/> 測量與風險對應 | <input checked="" type="checkbox"/> 資料來源與模型監督 | <input checked="" type="checkbox"/> 必須保留紀錄與檢測 |
| 可解釋性要求 | 高風險 AI 必須可解釋 | 建議設計為可理解 | 建議落實 | 高影響模型需可說明 |
| 透明與問責性 | 重大應用需註冊 | 強調組織責任 | 須自評與記錄 | 須設管理人員與申訴機制 |

六、企業 AI 治理實踐

(一)為何企業需要 AI 治理？

企業導入 AI 不僅是技術決策，更涉及品牌信任、合規風險、內部責任制度與用戶權益。

常見企業面臨風險包括：

- 模型決策不透明 → 消費者不信任
- 自動化結果有偏誤 → 法律糾紛
- 客戶資料使用不當 → 侵害個資法

AI 治理的落實可提升模型品質、法遵能力與 ESG 表現。

(二)企業 AI 治理五大實踐面向

1. 治理架構與責任機制 (Governance Structure)

| 項目 | 說明 |
|---------------------------|--------------------|
| 成立 AI 治理委員會 | 跨部門監督 AI 專案，建立內部準則 |
| 指定 AI 責任官 (CAIO) 或模型負責人 | 明確決策與問責歸屬 |
| 內外部稽核制度 | 定期審查模型行為與資料使用紀錄 |

2. 模型開發與部署控管 (Model Lifecycle Control)

| 階段 | 實踐措施 |
|------|---------------------------------|
| 建模階段 | 記錄資料來源、建模邏輯與驗證方式 |
| 測試階段 | 模型偏誤偵測、影響評估報告 |
| 上線部署 | 限制未經驗證模型自動上線 |
| 持續監控 | 模型老化 (Concept Drift) 追蹤與再訓練機制 |

3. 資料治理與隱私合規 (Data Governance)

- 建立資料使用說明與同意機制
- 落實資料最小化原則與可追溯性
- 分層控管訓練資料、測試資料、實際營運資料
- 定期進行資料偏誤分析與異常監測

| | | |
|------|--------------|---------------------|
| 資料風險 | 隱私洩漏、識別重建 | 去識別化 + 數據存取分層管控 |
| 法律風險 | 著作權、不當輸出責任不明 | 建立模型審核政策、導入風險等級分類 |
| 應用風險 | 誤用、依賴過高、不當用途 | 強化提示語指引、設限使用場景與輸出風格 |
| 社會風險 | 不公平影響、資訊落差 | 社會溝通與包容性設計、教育培力 |

4. 可解釋性與決策透明 (Explainability & Transparency)

| 作法 | 目的 |
|-----------------------|-----------------------|
| 模型應提供「可理解輸出說明」 | 減少黑箱風險，提高客戶信任 |
| 引入可解釋模型（如 SHAP, LIME） | 支援風險評估與法遵審查 |
| 對消費者提供 AI 決策告知 | 包括使用 AI 的意圖、影響與異議處理方式 |

5. 利害關係人溝通與訓練 (Stakeholder Communication)

- 對內部員工提供 AI 操作與道德訓練
- 向用戶/外部利害人揭露 AI 使用目的與權利說明
- 對董事會/監察人建立 AI 成效與風險報告機制

(三) 企業 AI 治理的實施步驟

1. 建立 AI 治理框架：制定 AI 使用準則，確保符合法規要求。
2. 數據管理與隱私保護：採用數據匿名化、聯邦學習技術。
3. AI 模型公平性與透明性檢測：使用公平性測試工具（如 Google What-If Tool）。
4. 風險監測與應變機制：定期評估 AI 模型準確度，避免模型漂移。
5. 員工 AI 倫理培訓：確保 AI 發開發團隊理解 AI 治理原則。

L112 資料處理與分析概念

L11201 資料基本概念與來源

一. AI 資料的基本概念

◆ (一) AI 資料的定義 (Definition of AI Data)

AI 資料泛指用於訓練、驗證與測試人工智慧系統的各類型資料，其特徵在於：

- 能夠反映現實世界中的現象、行為、模式或規則。
- 可轉換為結構化或非結構化形式以供演算法學習。
- 是 AI 模型學習、推論與評估準確性的基礎。

■ 關鍵定義要點：

| 項目 | 說明 |
|--------------------------|----------------------|
| 原始資料 (Raw Data) | 尚未清理、轉換、標註的初始資料。 |
| 訓練資料 (Training Data) | 用來訓練 AI 模型以學習模式的資料集。 |
| 驗證資料 (Validation Data) | 協助模型調整超參數、避免過擬合的資料集。 |

二、生成式 AI 風險管理策略

■ 生成式 AI 風險管理策略（全方位治理架構）

■ (一) 生成式 AI 風險治理五大原則

| 原則 | 目標 | 對應措施 |
|---|------------------|---------------------|
| <input checked="" type="checkbox"/> 風險為本 (Risk-Based) | 根據用途與風險分類差異化管理 | 高風險應有更強監控 |
| <input checked="" type="checkbox"/> 可監督性 (Oversight) | 建立人機共管體制，防止全自動錯誤 | 關鍵應用應設人類審查點 |
| <input checked="" type="checkbox"/> 透明與可解釋 (Explainability) | 模型結果與來源可被理解與驗證 | 需保留提示語與資料出處 |
| <input checked="" type="checkbox"/> 公平性與包容性 (Fairness) | 避免偏誤、歧視與資訊落差 | 模型訓練與測試多樣化 |
| <input checked="" type="checkbox"/> 合法與合規 (Compliance) | 避免侵犯個資、著作權等法律 | 接軌 GDPR、AI ACT、個資法等 |

■ (二) 風險等級分類與對應策略 (RAG-Based Model)

| 風險等級 | 應用類型 | 控制策略建議 |
|------|-----------------------|-----------------------------|
| 高風險 | 信用評等、醫療輔助診斷、法律解釋、自動決策 | 模型審核 + RAG 檢索 + 審查機制 + 記錄留存 |
| 中風險 | 內容摘要、產品推薦、知識文件撰寫 | 設定提示語範圍、引用訓練資料來源、抽樣稽核 |
| 低風險 | 文案輔助、行銷文字生成、內部溝通 | 使用標準模板與內部驗證機制即可 |

◆ 建議企業建立「用途-風險對照矩陣」，逐項識別生成任務風險等級，作為治理基準。

● (三) 風險管理四階段策略流程 (依據 NIST)

1. 【Govern】 治理制度設計

| 要點 | 說明 |
|-------------|-----------------------------|
| 成立 AI 治理委員會 | 包含資訊、法遵、內控、業務、技術人員 |
| 指派責任人 | 設置 AI 負責人與模型 Owner |
| 建立政策與標準 | 撰寫 AI 發開發與使用政策、風險分類流程、提示語準則 |

2. 【Map】 風險映射與盤點

| | | | |
|------------------------------|---------------------|-------------------------|----------------|
| 內容不當 (Toxic Content) | 模型生成帶有仇恨、歧視、暴力、色情語句 | 使用開放式生成 AI 生成歧視詞彙或仇恨言論 | NIST、AI ACT |
| 身分冒用與偽造 (Impersonation) | 模型生成模仿他人聲音、文字或簽名 | Deepfake 影片模擬主管指令詐騙會計付款 | 金融業指引 |
| 決策誤導風險 | 模型提供不正確建議，影響重大業務決策 | AI 提供錯誤診斷建議導致誤醫 | NIST、台灣 AI 法草案 |
| 依賴性上升 | 使用者過度信任生成內容，放棄自主判斷 | 企業完全仰賴 AI 產出合約條文未經審核 | 金融指引、公部門手冊 |

| | |
|--------------------------|-----------------------|
| 測試資料 (Test Data) | 用於評估模型最終表現的資料集。 |
| 標註資料 (Labeled Data) | 含有明確標記 (如分類標籤) 的資料。 |
| 非標註資料 (Unlabeled Data) | 僅有原始輸入，無明確標記的資料。 |

■ (二)AI 資料的內容類型與來源 (Types & Sources of AI Data)

AI 資料的內容可依格式與來源做分類：

■ 1. 依資料格式分類：

| 資料格式 | 說明 | 例子 |
|--------|-------------|---------------------|
| 結構化資料 | 有固定欄位與資料型態 | 資料庫、Excel 表格、金融交易紀錄 |
| 半結構化資料 | 有部分結構，非固定格式 | JSON、XML、日誌檔 |
| 非結構化資料 | 無明確結構，需預處理 | 影像、語音、自由文字、影片 |

■ 2. 依資料來源分類：

| 資料來源 | 說明 | 例子 |
|---------|----------------|---------------------------------|
| 感測器資料 | 來自 IoT 裝置與工業設備 | 車用感測器、氣溫感測器 |
| 使用者行為資料 | 從互動過程中蒐集 | 點擊紀錄、購買行為 |
| 自然語言資料 | 來自語言或文字輸入 | 聊天記錄、客服紀錄 |
| 視覺資料 | 來自影像與影片 | 醫療影像、監視錄影 |
| 開放資料集 | 公開可用的資料庫 | ImageNet、COCO、UCI ML Repository |
| 專屬企業資料 | 組織內部蒐集與產出 | 客戶資料、銷售資料 |

（三）(c) AI 資料處理完整流程與方法 (AI Data Processing Pipeline and Techniques)

人工智慧的效能與可信度高度依賴資料的品質與處理方式。以下是常見八大處理流程階段，每個階段均搭配常用方法與技術工具：

■ 1. 資料收集 (Data Collection)

目的：獲取與問題情境相關的足夠、多元且高品質原始資料。

資料來源類型：

| 類型 | 描述 | 工具/平台範例 |
|---------|----------------|---|
| 自行蒐集 | 透過感測器、應用程式、API | OpenStreetMap API、IoT Sensor |
| 公開資料集 | 開源研究或政府平台 | Kaggle, UCI ML Repository, Open Data 台灣 |
| 網頁爬蟲 | 使用程式從網頁擷取資料 | BeautifulSoup, Scrapy, Selenium |
| 第三方商業資料 | 付費購買的資料庫 | Nielsen, Bloomberg, Statista |

■ 2. 資料清理 (Data Cleaning)

（四）法律與治理層風險

| 風險類型 | 說明 | 可能情境 | 參考依據 |
|-----------------------------|---------------------|---------------------|----------------------|
| 個資洩露 (Privacy Leakage) | 訓練或輸出過程揭露使用者資料 | 模型回應中出現員工手機或 email | 台灣 AI 法草案、GDPR、NIST |
| 去識別失敗 | 原本脫敏的資料可被還原或重辨識 | 透過多筆回應推論出某病患身份 | 公部門手冊、AI ACT |
| 跨境資料傳輸風險 | 模型部署在國外，違反本地資料外流規定 | 使用 API 將敏感輸入送往海外伺服器 | 金融業指引、台灣個資法 |
| 提示語洩露 (Prompt Logging) | 系統未記錄提示語/回應紀錄或未控管存取 | 社內機密輸入模型而被訓練或外洩 | AI ACT、NIST 「日志紀錄原則」 |

（五）整合風險對應矩陣 (可用於內部評估與治理設計)

| 分類面向 | 核心風險類型 | 管理對策建議 |
|------|----------|-----------------|
| 技術風險 | 幻覺、偏誤、黑盒 | 引入人審流程、設置輸出驗證模組 |

目的：移除雜訊、處理缺失與不一致性資料，提升模型學習品質。

常見清理技術：

| 問題 | 方法 | 工具 |
|------------------------|------------------------------|----------------------|
| 缺失值 (Missing Values) | 平均/中位數填補、KNN 補值、刪除資料列 | pandas, scikit-learn |
| 異常值 (Outliers) | IQR、Z-score、Isolation Forest | NumPy, PyOD |
| 重複值 (Duplicates) | 使用 drop_duplicates() 移除 | pandas |
| 格式錯誤 | 資料型別轉換、單位正規化 | pandas, regex |
| 不一致欄位命名 | 使用 mapping 字典或規則統一欄位名稱 | Python dict |

3. 資料標註 (Data Labeling)

目的：針對監督式學習需求提供對應的標籤資訊。

標註方式：

| 方式 | 說明 | 工具 |
|---------------------------|------------------|------------------------|
| 人工標註 | 雇用標註者標記每筆資料 | Label Studio, docrano |
| 專家標註 | 領域專家提供正確標籤 | 醫療診斷標註平台 |
| 群眾外包 (Crowdsourcing) | 透過外部平台大量標註 | Amazon Mechanical Turk |
| 弱標註 (Weak Labeling) | 使用規則、模型預測自動標註 | Snorkel, Heuristics |
| 主動學習 | 模型挑選最不確定的樣本供人工標註 | Prodigy, modAL |

4. 資料分割 (Data Splitting)

目的：將資料劃分為訓練、驗證與測試集，避免過擬合並保留泛化能力。

分割方式：

| 類型 | 描述 | 方法範例 |
|------------------------------|-------------|---------------------------|
| 隨機分割 | 隨機劃分樣本 | train_test_split() |
| 分層抽樣 (Stratified Split) | 根據類別比例維持分布 | StratifiedKFold |
| 時序分割 (Time Series) | 依時間排序切割 | 滾動視窗 (Rolling Window) 法 |
| 交叉驗證 (Cross-validation) | 將資料平均切分 K 折 | KFold, LeaveOneOut |

5. 特徵工程 (Feature Engineering)

目的：從原始資料中擷取與任務高度相關的數據表示。

技術方法：

- 生產端 (資料提供者) + IT (平台建構) + 業務端 (需求驗證) 共同設計應用。

二、運營管理關鍵指標與建議工具

| 管理面向 | KPI 建議 | 工具建議 |
|--------|---------------------|---------------------------|
| 使用頻率 | 每週使用次數 / 使用人數 | 系統紀錄或分析後台 |
| 輸出品質 | 使用者滿意度調查 (滿意/需修正) | 回饋按鈕 + 人工比對 |
| 成效貢獻 | 節省工時、減少錯誤率、加快任務完成時間 | ERP 任務指標對比前後 |
| 安全與合規性 | 是否出現違規內容或資料外洩風險 | 建立模型輸出審查流程 (Rule-based) |

L12303 生成式 AI 風險管理

一、生成式 AI 主要風險類型

✿(一) 技術層風險 (模型本身的潛在問題)

| 風險類型 | 說明 | 可能情境 | 參考依據 |
|--------------------------------|-------------------------|------------------------|----------------------|
| 幻覺 (Hallucination) | 模型生成錯誤或虛構資訊，卻語法正確、內容自然 | ChatGPT 生成不存在法條、虛構技術資料 | NIST.AI.600-1、AI ACT |
| 模型偏誤 (Bias) | 模型受訓練資料影響，呈現性別、種族、地區等偏見 | AI 面試評分偏向男性 | AI ACT、台灣 AI 法草案 |
| 輸出不可控 (Uncontrollability) | 難以完全控制模型輸出風格、邏輯或內容 | 模型偏離預期語氣，生成多餘段落 | 公部門手冊 |
| 可解釋性不足 (Opacity) | 模型如黑盒，無法說明為何做出某種回應或推薦 | 難以回溯某份建議是基於哪一資料 | 金融業指引、NIST |
| 資安風險 (Prompt Injection) | 惡意提示語改變模型行為或外洩敏感資訊 | prompt誘導模型回吐訓練資料、洩漏帳密 | NIST、歐盟 AI ACT |
| 訓練資料合法性爭議 | 模型訓練資料來源未公開或違反著作權 | 使用未授權圖像或文章進行訓練 | AI ACT、CCAI、著作權法 |

✿(二) 使用層風險 (應用過程中產生之問題)

| 風險類型 | 說明 | 可能情境 | 參考依據 |
|--------------------|---------------|----------------|------------------|
| 誤用風險 (Misuse) | 使用者透過模型達成違法目的 | 生成詐騙簡訊、釣魚信、假新聞 | AI ACT、NIST、金融指引 |

- 測試驗證報告 (使用者滿意度、錯誤修正率)

◆ 第六階段：運營管理與成效追蹤 (Operation & Optimization)

◎ 目的：

確保模型長期穩定運行、持續優化成效並控管風險。

◆ 任務與產出：

| 維運項目 | 說明 |
|--------|-----------------------------|
| KPI 監控 | - 平均任務完成時間- 節省工時百分比- 使用者滿意度 |
| 模型更新 | 每季更新知識資料庫與提示語策略 (如 SOP 有改版) |
| 異常偵測 | 若模型生成內容被多次標註錯誤，自動進行稽核或降權 |
| 教育訓練 | 定期辦理內部使用工作坊，增進生成式 AI 素養與技巧 |
| 法規遵循 | 定期審查模型生成是否違反個資或產業規範 |

■ 輸出產物：

- 效益評估與 ROI 報告
- 模型行為記錄與風險報表
- 人機協作標準作業指引

■ (二) 生成式 AI 特別實施重點整理

| 主題 | 核心建議 |
|---|--|
| <input checked="" type="checkbox"/> 提示語工程 (Prompt Engineering) | 建立提示語模板庫，區分不同任務與角色語氣 |
| <input checked="" type="checkbox"/> 多模態整合 | 文件 + 圖像 + 表格 + 影片生成任務整合 如品質說明 + 缺陷圖) |
| <input checked="" type="checkbox"/> RAG 導入 | 針對文件查詢型任務導入 RAG，強化知識正確性 |
| <input checked="" type="checkbox"/> 實施規模化 | 由單一流程擴展至跨工廠/多語言點任務應用 |
| <input checked="" type="checkbox"/> AI 倫理治理 | 確保 AI 行為可審查、有出處、有記錄 |

實施策略建議

1. 從輔助性任務開始：
 - 如文件產生、SOP 撰寫、品保通報說明初稿。
 - 風險低且能快速驗證生成品質。
2. 導入提示語範本庫：
 - 將常用生成任務 (如報表摘要、建議項目、技術說明) 設計為範本，利於內部複用。
3. 整合向量知識庫：
 - 使用 FAISS、Pinecone 等工具將內部文件轉為可檢索語意資訊，以支援 RAG。
4. 實施多角色共創工作流：

| 類型 | 方法範例 | 工具 |
|------|---------------------------------|--------------------------|
| 特徵轉換 | log 轉換、標準化、min-max scaling | scikit-learn, NumPy |
| 特徵編碼 | One-hot encoding、Label encoding | pandas, OneHotEncoder() |
| 特徵選擇 | 卡方檢定、遞迴消除 (RFE) | SelectKBest, SHAP |
| 特徵構造 | 新變數創建 (如價格 ÷ 面積) | pandas |
| 降維技術 | PCA, t-SNE, UMAP | scikit-learn, umap-learn |

◆ 6. 資料增強 (Data Augmentation)

目的：在資料不足或偏斜的情況下增加資料多樣性，提升模型魯棒性。

常用增強方法：

| 資料類型 | 方法範例 | 工具 |
|---------|-----------------|-----------------------------|
| 圖像 | 旋轉、翻轉、裁剪、模糊 | Albumentations, OpenCV |
| 文字 | 同義詞替換、隨機遮蔽、反義替換 | TextAttack, NLPAug |
| 音訊 | 加噪、時間拉伸、頻譜遮罩 | torchaudio, audiomentations |
| 標籤不均衡處理 | SMOTE、ADASYN | imbalanced-learn |

■ 7. 資料儲存與管理 (Data Storage & Versioning)

目的：管理不同階段資料版本與存取權限，確保模型可重現性與合規性。

實作方法：

| 工具 | 功能 | 適用情境 |
|-------------------------------|-----------------|------------------------|
| DVC (Data Version Control) | 資料版本追蹤與 Git 整合 | 團隊協作、大型專案 |
| MLflow | 儲存模型與資料中介數據 | 模型訓練流程追蹤 |
| Data Lake / Warehouse | 結構化儲存與查詢優化 | 企業級資料平台 |
| 資料存取權限設計 | 整合 IAM 或 SSO 控制 | 機密資料防洩漏 |
| 元資料管理 (Metadata Management) | 儲存欄位說明與欄位關聯 | Apache Atlas, Amundsen |

◆ 8. 資料倫理與法遵處理 (Ethics & Compliance)

目的：防止資料偏誤、保護用戶隱私，符合國際資料治理規範。

技術實務：

| 項目 | 說明 | 工具 |
|-------------------------------|-----------------------------|---------------------------|
| 匿名化 (Anonymization) | 移除個資標記欄位 (姓名、ID) | ARX, Faker |
| 偽匿名化 (Pseudonymization) | 將個資替換為代碼但可還原 | Hash、Token |
| 差分隱私 (Differential Privacy) | 加入隨機噪音保護群體隱私 | Google DP Library, Opacus |
| 法規遵循 | 確保資料使用與儲存合乎 GDPR、CCPA、台灣個資法 | 法務部審查流程 |

| | | |
|--------|-----------|----------|
| 敏感欄位審查 | 建立敏感變數白名單 | 自訂資料稽核規則 |
|--------|-----------|----------|

◆ 9.學習重點整理表

| 處理階段 | 關鍵技術 | 工具範例 |
|--------|------------|-----------------------|
| 1.收集 | 多元來源整合 | API,爬蟲, Sensor |
| 2.清理 | 缺值處理、異常移除 | pandas, PyOD |
| 3.標註 | 弱標註、主動學習 | Label Studio, Snorkel |
| 4.分割 | 時序、分層交叉驗證 | scikit-learn |
| 5.特徵工程 | 選擇、轉換、降維 | PCA, SHAP |
| 6.增強 | NLP/影像合成技術 | NLPAug, Alumentations |
| 7.管理 | 版本、權限、元資料 | DVC, MLflow |
| 8.法遵 | 匿名、敏感欄位稽核 | ARX, Opacus |

◆ 10.AI資料的功能分類 (By Function in ML)

| 類型 | 功能 | 說明 |
|--------------------------|----------|-------------------|
| 訓練資料 (Training Data) | 建立模型 | AI 模型依據此資料學習規則與特徵 |
| 驗證資料 (Validation Data) | 調參、避免過擬合 | 協助選擇最佳模型與超參數 |
| 測試資料 (Test Data) | 評估模型泛化能力 | 評估模型在未見資料的表現 |

◆ 11.資料標註工具比較 (Data Annotation Tools Comparison)

AI 模型 (尤其是監督式學習與深度學習) 依賴大量標註資料進行訓練，因此高效率、高一致性的標註工具對 AI 專案至關重要。

◆ 資料標註任務類型分類：

| 任務類型 | 描述 | 應用範例 |
|-----------------------------------|-----------|---------------|
| 圖像分類 (Image Classification) | 標示整張圖的類別 | 動物辨識、人臉辨識 |
| 邊界框標註 (Bounding Box) | 繪製物件所在範圍框 | 交通物體偵測、安控 |
| 語義分割 (Semantic Segmentation) | 將每個像素分類 | 醫療影像、智慧城市 |
| 命名實體辨識 (NER) | 標記句子中特定實體 | 金融文字分析、客服理解 |
| 文字分類 / 情緒分析 | 給整段文字一個標籤 | 客戶評論分級、詐騙郵件辨識 |

◆ 12.常見標註工具比較表：

| 工具名稱 | 介面特性 | 支援格式 | 優點 | 適用場景 |
|--------------|---------|----------|-----------------|----------------|
| Label Studio | 開源、網頁介面 | 圖像、文字、音訊 | 高自訂性、多功能模組、支援協作 | 多模態 AI 專案、研究機構 |

| | |
|--------|--|
| 外部知識 | 法規、公協會標準、專利資料、維修知識庫 |
| 人力資源 | - Prompt Designer (提示語設計) - Domain Expert (領域專家) - MLOps Engineer (模型部署) |
| 資料清洗流程 | 格式轉換 (PDF → txt) 、段落切分、去除多餘符號 |
| 向量化設計 | 資料是否需嵌入向量資料庫供檢索 (RAG) 使用？ |

■ 輸出產物：

- 資料盤點報告
- 向量知識庫規格設計表
- 權限與資料治理控管策略

◆ 第四階段：模型導入與原型設計 (PoC & Prototyping)

◎ 目的：

以低風險任務試驗模型生成品質，並完成 PoC 原型設計，驗證效益。

◆ 任務與產出：

| 原型設計項目 | 說明 |
|--------|--|
| 最小可行任務 | 單一功能為主，例如：「根據維修紀錄生成建議書」 |
| 提示語設計 | 建立範本：包含角色、格式、限制（如「使用工程語氣」、「不超過 100 字」） |
| 生成品質驗證 | - 正確性（事實/邏輯）- 可讀性（語法/語氣）- 可控性（格式/風格） |
| 使用者測試 | 跨部門（如維修主管/工程人員）試用與回饋 |

■ 輸出產物：

- PoC 原型操作流程
- 提示語模板集
- 初步效益對照報告（手動 vs AI）

◆ 第五階段：系統部署與場域驗證 (Deployment & Integration)

◎ 目的：

將原型模型整合入實際營運流程中，進行全流程驗證與優化。

◆ 任務與產出：

| 作業項目 | 說明 |
|-------|------------------------------|
| 系統整合 | 與 MES/ERP/文件系統 API 串接，支援雙向互通 |
| 權限設計 | 根據部門/職級設定內容生成與查詢範圍 |
| 監控儀表板 | 即時追蹤使用率、輸出品質、系統回應時間 |
| 回饋機制 | 使用者可標記「正確/需修正」，作為再訓練依據 |

■ 輸出產物：

- 完整部署架構圖
- 實施 SOP 與操作手冊

■ (一)生成式 AI 導入規劃六步驟細部說明

◆ 第一步：問題定義與需求設定 (Problem Definition & Scoping)

● 目的：

釐清生成式 AI 能解決的具體問題與實際目標，避免導入流於「技術試驗」而無成效。

◆ 任務與產出：

| 任務項目 | 詳細說明 |
|---------------|---|
| 明確化目標 | 例如：「減少產線異常報告撰寫時間」或「提供維修建議草稿」 |
| 導入價值定位 | 是提升效率？降低成本？增進知識流通？ |
| 任務分類 | - 可結構化（如報表產出）- 半結構化（如維修說明）- 非結構化（如問答查詢） |
| 適合生成式 AI 價值場景 | - 文件撰寫- 建議生成- 多語對話- 對知識的自然語言查詢 |

■ 輸出產物：

- AI 導入任務地圖
- 對應部門與角色職責分析
- 對應工作流程與任務耗時報告

◆ 第二步：解決方案設計與評估 (Solution Design & Evaluation)

● 目的：

設計可落地的 AI 解決方案，並從技術、資安、效能、擴充性等面向進行評估與比對。

◆ 任務與產出：

| 評估項目 | 說明 |
|--------|---|
| 模型選型 | LLM：GPT-4、Claude、Gemini 圖像：SD、Midjourney |
| 工具選型 | Notion AI（內嵌簡報）、LangChain（整合）、Pinecone（向量庫） |
| 架構選擇 | 單純生成 or 結合 RAG（檢索式生成）？ |
| 部署環境 | SaaS、公有雲、私有雲、內部部署 |
| 系統串接需求 | MES、ERP、PLM、內部知識庫 |

■ 輸出產物：

- 解決方案白皮書
- 工具選型比較表
- 法規與資安需求對照表（如個資、商業機密）

◆ 第三步：資源與資料準備 (Data & Infrastructure Readiness)

● 目的：

確保可用的資料、人力、系統資源符合生成式 AI 的建置需求與品質標準。

◆ 任務與產出：

| 資源項目 | 說明 |
|------|------------------------|
| 內部語料 | 包含 SOP、操作指引、設備手冊、工程規範等 |

| | | | | |
|-------------------------------|-------------|-----------|--------------------|----------------|
| CVAT (Intel) | 專為影像與影片設計 | 影像、影片 | 專業級影像標註、支援自動化框選 | 自駕車影像、醫療影像 |
| Prodigy (Explosion.ai) | 商用，快速互動式標註 | NLP 為主 | 結合主動學習，可與 spaCy 整合 | 語言模型微調、知識抽取 |
| docrano | 開源、支援中文 NLP | 文字、序列標註 | 易用性高，適合初學者與教學用 | 文件分類、NER |
| SuperAnnotate | 商用、協作平台 | 圖像、影片、文字 | 團隊導向、內建 QA 管理 | AI 新創公司、標註外包團隊 |
| Amazon SageMaker Ground Truth | 雲端整合 | 多格式、S3 整合 | 可使用標註工人與自動標註 | AWS 用戶、大型專案 |

▲ 13. 資料偏誤檢測方法 (AI Dataset Bias Detection Techniques)

AI 模型的訓練資料若帶有偏誤，將導致模型在實際應用中產生不公平、不準確，甚至違反法規、偏誤 (Bias) 可來自資料收集、標註、樣本分佈等階段。

★ 常見偏誤類型分類：

| 偏誤類型 | 說明 | 例子 |
|----------------------------|----------------|---------------|
| 樣本偏誤 (Sample Bias) | 資料來源不均或群體代表性不足 | 訓練資料僅來自都會區用戶 |
| 標註偏誤 (Label Bias) | 標註者主觀導致誤導性標籤 | 討論政治文的情緒標記誤差 |
| 測量偏誤 (Measurement Bias) | 資料攝取工具或設備導致偏差 | 音訊錄製設備對男性聲音靈敏 |
| 漏失偏誤 (Omission Bias) | 重要特徵或族群在資料中缺失 | 健康資料集中忽略少數族群 |
| 確認偏誤 (Confirmation Bias) | 只收集驗證假設的資料 | 只蒐集會點廣告的使用者行為 |

▲ 14. 偏誤檢測技術與方法：

| 檢測方法 | 適用階段 | 說明 |
|---------|---------|--|
| 統計分布分析 | 收集前與分割後 | 檢查樣本比例、類別分佈（如性別、年齡、地區）是否平衡 |
| 可視化工具 | 分析階段 | 使用 t-SNE、PCA、直方圖等檢視樣本間距與群體分布差異 |
| 公平性指標計算 | 訓練後評估 | 如 Demographic Parity、Equal Opportunity、Statistical Parity Difference |

| | | |
|-------------------------------|---------|----------------------------|
| 交叉分析 (Stratified Analysis) | 模型測試時 | 檢查不同群體的模型表現是否一致（如男性 vs 女性） |
| 錯誤分析 (Error Analysis) | 驗證與回饋階段 | 比較不同群體預測錯誤率是否明顯差異 |
| 敏感特徵探勘 (Feature Audit) | 模型前 | 確認模型是否依賴不應使用的敏感變數（如種族、宗教） |

◆ 15. 偏誤矯正策略：

| 策略類別 | 說明 |
|--------|---|
| 資料層面調整 | 透過欠採樣、過採樣、資料增強等方法平衡訓練資料 |
| 模型層面控制 | 使用公平性正規化 (Fairness Constraints)、對抗性學習 (Adversarial Debiasing) |
| 後處理方法 | 預測後修正偏誤，如閾值調整、重加權 (Reweighting) |

L11202 資料整理與分析流程

一、資料整理流程

AI 模型的品質與效能，約有 70% 仰賴資料品質與前處理流程。

★ 完整流程如下：

資料來源收集 → 資料清理 → 資料轉換 → 資料轉換與標準化 → 特徵工程 → 資料切分與保存

◆ 第一步：資料收集 (Data Collection)

◎ 目的：取得符合目標任務所需的資料，作為 AI 模型訓練與預測的基礎。

◆ 來源類型：

- 內部系統資料：CRM、交易紀錄、設備紀錄
- 外部資料集：政府開放資料、研究機構數據集（如 Kaggle）
- IoT 與感測器：工業設備、智慧城市、醫療穿戴裝置
- API 服務：Twitter、Google Maps、天氣平台
- 網頁爬蟲：新聞、評論、產品資訊
- 合成資料：模擬資料或生成資料（如 GAN、模擬環境）

▲ 注意事項：

- 法規合規性（個資法、GDPR）
- 資料格式統一（CSV、JSON、XML...）
- 建立「資料目錄 (Data Catalog)」與「資料來源紀錄表」

◆ 第二步：資料清理 (Data Cleaning)

◎ 目的：移除或修正資料中的錯誤、雜訊、遺漏值等問題，提升資料品質。

◆ 常見清理任務與方法：

| 評估項目 | 說明 | 建議工具與方法 |
|------------|-----------------------------|---------------------------------------|
| 模型選型 | 根據任務類型選擇分類/回歸/分群/生成等模型 | CNN (影像)、XGBoost (數值)、RNN/LSTM (序列) |
| 模型精度與效能 | 是否能在驗證集上達到業界接受門檻 | 使用 Confusion Matrix、F1 Score、RMSE 等指標 |
| 生成式 AI 的應用 | 是否用於文件生成、維修建議、知識問答 | 可結合 RAG 建立內部知識庫與答系統 |
| 部署模式 | 是否為本地部署/雲端/邊緣裝置？ | 評估對資安與成本的影響 |
| 工具整合性 | 模型是否可串接 MES、ERP、SCADA 等現有系統 | 是否支援 API/PLC 通訊/即時回寫等機能 |
| 擴充彈性 | 後續是否能擴展到其他產線或工廠？ | 是否採用模組化、標準開發框架（如 MLOps） |

七、評估框架整合總表（可作為評量表或投標審查指標）

| 評估項目 | 權重（建議） | 說明 |
|-----------|--------|-----------------|
| 問題釐清與需求定義 | 10% | 是否明確說明痛點與目標 |
| 技術效能與適配性 | 30% | 模型回應品質、整合能力 |
| 工具操作與擴充性 | 15% | 是否易用、可進階客製 |
| 成本效益 | 25% | 對應實際節省或創造價值 |
| 資安與法規合規性 | 20% | 包含私有化、法規遵循與風險控管 |

L12302 生成式 AI 導入規劃

一、生成式 AI 導入六步驟規劃流程（依製造業 AI 導入指引）

| 階段 | 核心任務 | 生成式 AI 特別關鍵點 |
|--------------|-----------------|-------------------------|
| 1. 問題定義與需求設定 | 明確痛點與目標應用場景 | 要辨別「可由生成式 AI 處理的任務」 |
| 2. 解決方案設計與評估 | 擬定技術方案、評估工具適配性 | 包括是否需採 RAG 架構、私有化需求 |
| 3. 資源與資料準備 | 整合資料資源、人力與系統資產 | 資料標註格式、文本語料、內部知識文件 |
| 4. 模型導入與原型設計 | PoC 試驗生成效果與整合方式 | 設計提示語範本、驗證生成品質 |
| 5. 系統部署與場域驗證 | 整合產線/系統、進行驗證與修正 | API 整合 MES/ERP/知識庫、權限控管 |
| 6. 運營管理與成效追蹤 | 建立維運制度與效能評估指標 | 定期監控生成品質、風險管理、版本控制 |

3. 設置內部 AI 專案辦公室 (AI PMO) 統籌推進

◆(二)營運層：AI 應用場景與資料需求

核心目標：

以營運痛點為出發點，分析 AI 可切入之場域，並確認能否取得足夠資料進行建模。

評估要素：

| 項目 | 說明 | 檢核重點 |
|---------|---------------------|--------------------------|
| 應用場景識別 | 明確描述問題背景與待優化流程 | 如：設備異常預測、品質檢測、排程優化 |
| 資料可得性 | 是否已有或能建立足夠且品質良好的資料集 | 資料型態？完整度？頻率與標註方式？ |
| 資料治理成熟度 | 資料是否標準化、結構化、具可用性 | 是否已有 Data Pipeline？清洗流程？ |
| 營運痛點量化 | 痛點是否能轉換為指標 | 良率、機台稼動率、誤檢率、工時浪費等 |

常見 AI 應用場景（依據《製造業 AI 導入指引》補充）：

| 類型 | 案例說明 | 所需資料 |
|------------|-----------------|----------------|
| 生產流程優化 | 自動排程、瓶頸分析 | 生產歷程記錄、工時資料 |
| 設備預測維護 | 預測機台故障、減少停機 | 感測器數據、歷史維修紀錄 |
| 品質檢測 | AOI 缺陷判斷、自動分級 | 影像資料（含標註）、量測數據 |
| 能源管理 | 電力峰值預測、空壓控制 | 用電記錄、環境感測資料 |
| 智能倉儲物流 | 最佳拣貨路徑、庫存預測 | 存貨歷史、訂單流量 |
| 營運洞察與報表自動化 | ERP/MES 數據智能化分析 | 多系統資料整併結果 |

資料評估指標建議：

| 指標 | 評估重點 |
|------|-------------------------|
| 完整性 | 是否缺欄、是否可覆蓋全部樣本類型 |
| 多樣性 | 資料是否足以描述所有情境（如白天/夜班/機台） |
| 標註品質 | 有無人為標記錯誤？是否一致性高？ |
| 時效性 | 是否即時？多久更新一次？ |
| 可存取性 | 是否需跨部門/系統取得？授權問題？ |

◆(三) 技術層：AI 模型與工具效能

核心目標：

選擇符合製造現場需求的 AI 模型與平台工具，兼顧效能、可整合性與部署可行性。

評估要素：

| 問題類型 | 處理方法 | 工具程式方法 |
|----------------------|-------------------------|------------------------|
| 缺值 (Missing Value) | 刪除、平均/中位數填補、預測填補（如 KNN） | pandas.fillna() |
| 重複值 (Duplicates) | 使用關鍵欄位比對刪除 | drop_duplicates() |
| 異常值 (Outliers) | Z-score、IQR、視覺化檢測 | scipy.stats.zscore() |
| 資料格式錯誤 | 格式轉換（日期、時間、數值） | to_datetime()、astype() |
| 邏輯錯誤 | 自訂規則進行校驗（例如：年齡不可為負） | 條件判斷與篩選修正 |

▲ 常見陷阱：

- 自動填補缺值但未檢查資料分布
- 合併資料前未統一 ID 欄位或格式
- 第三步：資料轉換與標準化（Transformation & Normalization）

◎ 目的：轉換資料格式或值域，以便模型能有效處理與計算。

◆ 處理類型與技巧：

| 項目 | 說明 | 範例 |
|-------------------------|-------------------------|------------------------|
| 類別編碼 (Encoding) | 將文字類別轉為數值 | One-Hot、Label Encoding |
| 正規化 (Normalization) | 將數值特徵縮放至特定區間 | Min-Max Scaler |
| 標準化 (Standardization) | 轉換為常態分布（均值 = 0，標準差 = 1） | Z-score |
| 資料轉置與重塑 | 轉換資料表格長寬格式 | pivot、melt |
| 單位轉換 | 如公尺 ↗ 英尺、NTD ↗ USD | 自訂換算公式 |

◆ 工具範例（以 Python 為例）：

```
from sklearn.preprocessing import MinMaxScaler, StandardScaler
scaler = MinMaxScaler()
data_scaled = scaler.fit_transform(data)
```

◆ 第四步：特徵工程 (Feature Engineering)

◎ 目的：建構能提升模型效能與可解釋性的特徵欄位。

◆ 技術與策略：

| 方法 | 說明 | 範例 |
|-------|----------------------|--------------------------------|
| 新特徵創造 | 計算或合成新欄位 | 年收入 ÷ 人數 → 平均所得 |
| 特徵篩選 | 移除不相關、重複、無意義欄位 | Variance Threshold、SelectKBest |
| 特徵轉換 | 對特徵進行數學處理 | log(x)、平方根轉換 |
| 降維處理 | 使用 PCA、LDA 等技術壓縮資料維度 | Principal Component Analysis |

◆ 小提示：

- 避免過多無意義特徵（會導致模型過擬合）

- 特徵可視化有助於初期洞察與選擇

◆ 第五步：資料切分與儲存 (Split & Store)

◎ 目的：為模型訓練與測試建構乾淨、獨立的資料集，並建立安全儲存策略。

1. 切分方式：

| 分類 | 說明 | 常見比例 |
|-----------------------|-----------|--------|
| 訓練資料 (Training Set) | 用於模型學習 | 70-80% |
| 驗證資料 (Validation Set) | 用於模型調整超參數 | 10-15% |
| 測試資料 (Test Set) | 評估模型泛化能力 | 10-20% |

2. 儲存策略：

| 要素 | 實踐建議 |
|-------|---------------------------------|
| 格式 | 儲存為 CSV、Parquet、Pickle、Database |
| 命名規則 | 加上版本與日期，例如：train_v1_2025.csv |
| 權限與加密 | 僅授權人員可存取，應加密處理敏感資訊 |
| 留痕 | 建立資料版本記錄 (Data Versioning) |

◆ 工具建議：

- train_test_split() (scikit-learn)
- 資料庫：PostgreSQL、BigQuery
- 資料湖：AWS S3、Azure Blob、Hadoop HDFS

二、資料分析流程（從探索到推論）

全流程概觀：

◆ 第一步：探索性資料分析 (Exploratory Data Analysis, EDA)

◎ 目的：在建模前了解資料分布、結構、異常與潛在關係，是資料分析的起點。

典型工作項目：

| 任務 | 工具 | 範例 |
|-------|-----------------------------|--------------|
| 敘述統計 | Pandas、Excel | 平均數、中位數、標準差 |
| 資料視覺化 | Seaborn、matplotlib、Power BI | 直方圖、散佈圖、熱力圖 |
| 類別分析 | 群體比較、類別分布 | 各類別占比圖、列聯表 |
| 異常偵測 | 箱型圖、Z-score | 異常客戶行為、極端交易額 |

◆ EDA 的好處：能提早發現資料問題與模型設計限制。

◆ 第二步：假設與問題定義 (Hypothesis & Goal Definition)

◎ 目的：定義清楚的分析目標、商業問題或預測任務。

常見問題類型：

(一) 總擁有成本 (TCO)

| 成本項目 | 說明 |
|---------|-------------------------------------|
| 授權/訂閱費用 | LLM 或平台的 API 使用收費，如 GPT-4 Tokens 費用 |
| 建置成本 | 系統整合 (如資料前處理、API 串接) |
| 維運成本 | 模型調校、資料更新、人員維護 |
| 訓練與轉型成本 | 員工學習新工具、流程重設 |

(二) 投資報酬評估 (ROI)

| 項目 | 說明 | 計算方式 |
|-------|---------------|----------------|
| 生產力提升 | 節省人力、加速流程 | 任務處理時間前後對比 |
| 成本節約 | 減少錯誤、審訴、重工等損失 | 錯誤率降低 × 錯誤處理成本 |
| 決策效率 | 加速資料取得與分析時間 | 決策週期縮短程度與重要性估值 |
| 新服務價值 | 帶來額外收益或服務擴展 | 產品創新轉化為營收潛力分析 |

五、導入風險評估與控管建議

| 風險類型 | 潛在問題 | 控管建議 |
|------|-------------------|--------------------|
| 法規風險 | 模型生成不合規內容 (如個資外洩) | 加入「敏感詞過濾」與「人工審核機制」 |
| 模型幻覺 | 模型生成錯誤或虛構資料 | 搭配 RAG 架構提高知識來源可控性 |
| 誤用風險 | 使用者誤信 AI 回應造成錯誤判斷 | 導入回答出處提示與信心分數評估 |
| 資安風險 | 機密資料被送至外部 API | 優先考慮私有部署或加密 API |

六、參考《製造業 AI 導入指引》補充

◆ (一) 策略層：AI 導入願景與目標

核心目標：

聚焦在企業整體策略與 AI 結合，確保導入計畫符合企業轉型目標與競爭優勢的提升。

評估要素：

| 面向 | 說明 | 建議問題 |
|--------------|------------------|-----------------------|
| 導入動機 | 明確界定 AI 導入目的 | 是為了解決人力問題？提高良率？創建新產品？ |
| 關鍵績效指標 (KPI) | AI 導入應能量化成果 | 是否設定明確可量測的產能提升或成本節省？ |
| 組織支持度 | 高層是否願意提供資源與制度配合 | 是否建立 AI 專責小組、跨部門協作制度？ |
| 策略整合性 | AI 是否納入中長期數位轉型藍圖 | 是否與智慧製造、工業 4.0 策略一致？ |

◆ 推動建議：

- 制定「AI 發展路線圖」，分階段建構願景與落地計畫
- 將 AI 納入 ESG、永續製造、生產韌性等核心戰略指標

| | |
|----------|------------------------------|
| 技術可行性 | 企業 IT 基礎設施是否支持 AI 部署？ |
| 數據可用性 | 是否有足夠的高品質數據來訓練 AI？ |
| 成本效益 | AI 部署與維運成本是否合理？ |
| 風險與合規 | AI 生成內容是否符合法規 (GDPR, CCPA)？ |
| 員工培訓與接受度 | 員工是否具備 AI 操作能力？AI 是否會影響組織文化？ |

| 類型 | 範例 |
|----|--------------|
| 分類 | 這位客戶會不會流失？ |
| 預測 | 未來一週的銷售額是多少？ |
| 分群 | 哪些客戶行為類似？ |
| 判斷 | 哪些變數影響購買決策？ |

◆ 良好問題定義 = 明確目標變數 (y) + 適當資料 (X)

二、技術與工具效能評估

(一) 模型能力分析

| 指標 | 說明 | 評估方式 |
|---------|------------------|----------------|
| 回應準確性 | 模型是否能正確理解與回答問題 | 使用標準測試集進行比對 |
| 上下文理解力 | 能否處理長篇資料、結構化邏輯 | 測試多段文摘要與交叉引用能力 |
| 多語/專業能力 | 是否支援產業術語、技術文體 | 使用業界典型案例測試回應品質 |
| 模態支援性 | 是否能支援圖像、表格、語音等模態 | 測試圖文生成或文件解讀任務 |

(二) 系統運作效能

| 指標 | 說明 | 評估方式 |
|-----------|-----------------------|---------------|
| 回應速度 (延遲) | 輸入到輸出的時間延遲 | 平均響應時間 (毫秒) |
| 可擴展性 | 是否能支援高併發或多任務運作 | 模擬多人查詢環境進行測試 |
| 模型部署方式 | API / 軟件 / 本地部署 / 私有化 | 視資安需求選擇適當部署模式 |

三、適用解決方案選擇 (應用適配性評估)

(一) 應用場景分類

| 類別 | 說明 | 工具建議 |
|------|----------------|-------------------|
| 文字問答 | 客服、產品規格說明、法規諮詢 | GPT-4, Claude |
| 文件摘要 | 會議紀錄、政策文件、技術報告 | GPT-4 + RAG |
| 程式協作 | 自動補全、錯誤除錯 | GitHub Copilot |
| 圖像生成 | 設計稿、生產線模擬圖 | Midjourney, SD |
| 文件生成 | 報告、簡報、行銷文案 | Notion AI, Jasper |

(二) 解決方案選擇指標

| 評估面向 | 內容 | 評估方式 |
|---------|-----------------|----------------------|
| 功能符合度 | 工具功能是否對應業務痛點 | 建立「需求功能表」與「工具支援度對照表」 |
| 資料結構適應性 | 工具是否可讀/生成我方資料格式 | 文件格式兼容性測試 |
| 使用介面可用性 | 工具是否易於導入與訓練使用 | UI 測試與使用者回饋調查 |
| 安全合規性 | 工具是否符合資安、隱私要求 | 是否支援私有化部署與存取控管 |

四、成本效益分析 (TCO & ROI)

◆ 第三步：描述性統計分析 (Descriptive Analytics)

◎ 目的：描述目前資料的整體現象、分布與趨勢，幫助掌握「發生了什麼」。

常用指標與方法：

| 分析項目 | 說明 |
|-----------|-------------|
| 平均、眾數、中位數 | 集中趨勢指標 |
| 標準差、變異係數 | 資料離散程度 |
| 分布檢測 | 是否為常態、偏態或雙峰 |
| 時間序列走勢 | 時間為軸的趨勢變化觀察 |

◆ 應用場景：銷售月報、行為趨勢報告、網站訪問統計等

◆ 第四步：診斷性分析 (Diagnostic Analytics)

◎ 目的：說明「為什麼會發生」，探索變數之間的關係與因果推論線索。

常見技術與應用：

| 技術 | 說明 | 工具 |
|--------|-------------|--------------------------------|
| 交叉分析 | 類別變數交集比對 | Pivot Table、groupby |
| 相關係數分析 | 兩數值變數間的線性關聯 | Pearson, Spearman |
| 分群分析 | 將資料劃分為同質群體 | K-Means、階層分群 |
| 決策樹可視化 | 找出主要影響因子 | CART、Random Forest、XGBoost 可視化 |

◆ 注意：「相關」不等於「因果」！需搭配實驗設計或時序資料佐證。

◆ 第五步：預測性分析 (Predictive Analytics)

◎ 目的：預測未來或尚未觀察的資料狀態，為決策提供依據。

常見模型與應用場景：

| 模型類型 | 工具 / 方法 | 應用 |
|--------|---|-------------|
| 分類模型 | Logistic Regression、Random Forest、XGBoost、SVM | 客戶流失預測、詐欺偵測 |
| 迴歸模型 | Linear Regression、Lasso、Ridge、GBR | 銷售預測、價格預測 |
| 時間序列分析 | ARIMA、LSTM、Prophet | 銷售走勢、設備異常預測 |
| NLP 模型 | BERT、TF-IDF + Logistic Regression | 情緒分析、文本分類 |

◆ 預測性分析需配合評估指標（如：準確率、F1、RMSE）進行交叉驗證。

第六步：解釋與視覺呈現 (Interpretation & Visualization)

◎ 目的：將分析結果轉化為可理解、可行動的洞察與報告，強化決策支持。

● 呈現方式：

| | |
|-------------------------------|---------------|
| 視覺化方法 | 使用時機 |
| 熱力圖 (Heatmap) | 相關性視覺比較 |
| 特徵重要性圖 (Feature Importance) | 解釋模型決策邏輯 |
| 分類矩陣 / ROC 曲線 | 評估分類效能 |
| 儀表板 / 故事圖 | 給高階主管、非技術團隊呈現 |

◆ 工具建議：Power BI、Tableau、Plotly、Seaborn、Dash

◆ 第七步：回饋修正與部署評估 (Feedback & Deployment)

◎ 目的：將模型應用於實務，並進行持續監控與優化。

● 維護重點：

| | |
|-------|-------------------------------------|
| 任務 | 說明 |
| 模型監控 | 預測準確度是否隨時間下降？(概念漂移) |
| 再訓練 | 定期更新資料與模型權重 |
| 異常通報 | 模型預測錯誤是否過高？需有告警機制 |
| 使用者回饋 | 是否提供解釋與手動修正機制？(Human-in-the-loop) |

◆ 工具：MLFlow、Airflow、Prometheus、Grafana

三、統計與資料呈現方法

◆ (一)AI 中常用的統計方法 (Statistical Methods in AI)

AI 雖大量仰賴演算法與深度學習，但統計分析仍是理解資料、檢查模型品質與建立特徵關聯的基礎。

◆ 1-1 描述性統計 (Descriptive Statistics)

目的：快速了解資料的分布、集中趨勢與離散程度。

| 指標類型 | 指標名稱 | 說明 |
|------|---|---------------|
| 集中趨勢 | 平均數 (Mean) 中位數 (Median) 眾數 (Mode) | 資料值的中心位置 |
| 離散程度 | 標準差 (Std) 變異係數 (CV) 極差 (Range) | 資料分布的擴散程度 |
| 分布形態 | 偏態 (Skewness) 峰度 (Kurtosis) | 判斷資料是否對稱 / 尖銳 |

◆ 應用場景：

- 客戶平均消費金額、流量趨勢初步分析
- 建立模型前的資料健康檢查

◆ 應用 2：法律/規章問答系統

| | |
|------|-------------------------------------|
| 任務 | 根據內部法規與合約，回應法律問題 |
| 重點技術 | 長文件切段 (Chunking)、條文關鍵字加權檢索 |
| 特殊設計 | 將法條標號設為 metadata，可讓使用者看到引用來源 |
| 工具建議 | Haystack + Elasticsearch + Claude 2 |

◆ 應用 3：教育與個人知識庫

| | |
|------|-------------------------------------|
| 任務 | 教師上傳教材，學生問問題由 AI 回答 |
| 實作技巧 | 每份教材建一個子資料庫，分類如「章節」、「主題」作為 metadata |
| 工具組合 | GPT-4 + LangChain + ChromaDB |

◆ 應用 4：金融與產業報告分析

| | |
|------|---|
| 任務 | 對財報、新聞、法規作語意查詢與摘要 |
| 技術建議 | 使用高精度 embedding 模型 + NLP 前處理 (去除無意義字) |
| 工具建議 | Mixtral + Weaviate + HyDE (推論式擴充查詢) |

六、提示語設計最佳實踐 (RAG 專用)

模板一：查詢導向 (Instructional)

請根據以下公司內部政策文件，簡要回答使用者的問題。若找不到答案，請誠實回應「目前資料中無相關內容」。

【資料段落】

...

【問題】

...

模板二：摘要導向

請根據以下三段內容，整合成一段 100 字以內的摘要回應使用者問題。

1 ...

2 ...

3 ...

【問題】

...

L123 生成式 AI 導入評估規劃

L12301 生成式 AI 導入評估

一、生成式 AI 導入的核心評估面向

在導入生成式 AI 前，企業應評估技術可行性、商業價值、數據需求、合規與風險等多個因素，以確保 AI 與業務目標對齊。

| 評估面向 | 評估問題 |
|------|-------------------------|
| 商業價值 | AI 是否能解決業務痛點？能帶來多少效率提升？ |

| | |
|--------|-----------------------------------|
| 目標 | 將結構化/非結構化內容轉為向量化檢索資料庫 |
| 支援文件類型 | PDF、Word、HTML、CSV、Notion、Markdown |
| 工具建議 | LlamaIndex, LangChain, Haystack |
| 建議作法 | 將每份文件切成段落，設置標題或分類作 metadata |

步驟二：向量化處理 (Embedding)

| | |
|------|---|
| 功能 | 將文本轉為語意向量 (vector) |
| 常用模型 | OpenAI Embedding (text-embedding-ada-002)、HuggingFace BGE、SBERT |
| 嵌入維度 | 一般為 384~1536 維，取決於模型 |
| ◆ 重點 | 相同語意的句子其向量「距離接近」 |

步驟三：建立檢索系統

| | |
|------|---|
| 功能 | 透過語意比對找到最相關的內容 |
| 常用工具 | FAISS (快速、可離線) 、Pinecone (雲端服務) 、Weaviate 、Qdrant |
| 檢索方法 | KNN (k 最近鄰) 、ANN (近似最近鄰) |
| ◆ 技巧 | 支援 Top-K 結果與 metadata 過濾 (如分類、日期) |

步驟四：建構 Prompt 並輸入至 LLM

| | |
|-------------|--|
| 功能 | 將「原始問題」與「檢索段落」一起傳入生成模型 |
| 模型建議 | GPT-3.5/4、Claude 2、Mixtral、LLaMA 2 |
| Prompt 結構建議 | 可分為三段：1. 系統指示 2. 檢索內容摘要 (上下文) 3. 使用者問題 |
| ◆ 提示語設計範例 | |

你是一位專業助理，請根據下列資料回答使用者問題，並簡明扼要。

【知識段落】

...

【問題】

...

五、RAG 實務應用場景詳解

1. 應用 1：企業內部知識助手

| | |
|------|---------------------------------------|
| 任務 | 員工快速查詢公司政策、產品資料、操作規範等 |
| 實作關鍵 | 文件需切段落並清楚分類；Metadata (如部門/類型) 有助於檢索 |
| 工具組合 | Notion + LlamaIndex + FAISS + GPT-4 |

◆ 1-2 推論性統計 (Inferential Statistics)

目的：根據樣本推論整體資料特性，並進行假設驗證。

| 方法 | 說明 | 應用 |
|----------------------------|-------------|--------------|
| T 檢定 (T-test) | 比較兩群平均是否有差異 | 男女顧客平均購買力比較 |
| 卡方檢定 (Chi-square) | 分類變數是否獨立 | 廣告點擊率是否與性別有關 |
| 皮爾森相關 (Pearson) | 測量兩變數線性關係 | 廣告預算 vs 售賣量 |
| 線性回歸 (Linear Regression) | 建立數值變數間預測關係 | 預測營業額、房價等 |

◆ 注意：推論性統計需符合常態性、獨立性等前提。

◆ 1-3 分群與關聯分析 (Clustering & Association)

| 方法 | 說明 | AI 應用 |
|--------------------|--------------|-------------------|
| K-Means 分群 | 根據特徵自動找出群體結構 | 客戶分群、使用者畫像建立 |
| DBSCAN、階層式分群 | 處理異常點與非圓形分布 | 社群網絡、異常偵測 |
| 關聯規則分析 (Apriori) | 找出「X → Y」的規律 | 購物籃分析 (啤酒 + 尿布) |

■ (二) 資料視覺化呈現方法 (Data Visualization Techniques)

AI 分析結果需可視化呈現才能讓業務決策者、非技術人員理解。

◆ 2-1 常見圖表類型與應用對照

| 圖表類型 | 功能與特性 | 常見應用場景 |
|----------------------|--------------------------------|--------------------------------------|
| 折線圖 (Line Chart) | 顯示連續時間序列資料的變化趨勢 | 預測股價走勢、氣候變化、銷售趨勢分析、模型準確率變化曲線等 |
| 長條圖 (Bar Chart) | 比較不同分類的數值大小，強調分類間差異 | 分類模型預測結果分布、特徵重要性排序、各地區銷售量比較 |
| 直方圖 (Histogram) | 顯示連續資料分布情形，常用於觀察資料集中趨勢與偏態 | 資料前處理階段的分布檢查、數值標準化前後分布觀察 |
| 散佈圖 (Scatter Plot) | 顯示兩個變數間關係，揭示資料分群、相關性或異常值 | 監督式學習資料可視化、模型分類邊界、特徵間關聯探索 |
| 熱力圖 (Heatmap) | 利用顏色深淺顯示矩陣或變數間的關係，適合呈現多維資料或相依性 | 特徵間相關係數矩陣、混淆矩陣視覺化、推薦系統中使用者與物品之間的互動頻率 |
| 箱型圖 (Box Plot) | 呈現數值資料的分布範圍與離群值，利於資料品質分析 | 檢查資料偏態、離群值偵測、模型預測誤差分布 |

| | | |
|-------------------------------|--------------------------------|----------------------------------|
| 雷達圖 (Radar Chart) | 顯示多個維度上的整體分數表現，適合評估模型或策略的多方面特性 | 多模型比較（如精度、召回率、F1分數）、使用者特質分析、策略評估 |
| 泡泡圖 (Bubble Chart) | 散佈圖的延伸，第三維度以泡泡大小呈現 | 顧客分群（如價值 vs 滿意度 vs 購買力）、資源配置建議 |
| 決策樹圖 (Decision Tree) | 可視化模型決策流程，清楚呈現判斷條件與結果 | 解釋模型決策邏輯、進行特徵篩選與策略模擬分析 |
| 網絡圖 (Network Graph) | 呈現節點與連結之間的關係，適合圖神經網路或社群關係分析 | 知識圖譜、推薦系統、詐騙偵測、供應鏈關係建模 |
| 主成分圖 (PCA Plot) | 降維後視覺化高維資料的分布，顯示主成分結構 | 特徵選擇前的資料結構觀察、非監督學習資料視覺化 |
| 混淆矩陣圖 (Confusion Matrix) | 顯示分類模型的預測正確與錯誤情況，並分析類別誤判情形 | 分類模型效能評估、誤判類型檢討、模型微調依據 |

◆ 2-2 AI 特有視覺化方法

| 視覺化方法 | 功能與說明 | 應用場景 |
|--|--|-------------------------------|
| 特徵重要性圖 (Feature Importance Plot) | 顯示各輸入特徵對模型預測的影響力大小，常用於樹模型與統計模型 | 決策樹、隨機森林、XGBoost 等模型解釋、特徵選擇依據 |
| SHAP 值圖 (SHAP Summary / Force Plot) | 使用 SHAP 值解釋單一或整體預測，呈現每個特徵的貢獻度與方向性 | 模型解釋透明化、個案預測分析、金融信貸/醫療模型說明 |
| LIME 解釋圖 (LIME Explanation) | 對單一預測結果局部擬合可解釋模型，顯示特徵在該預測的影響 | 黑箱模型局部可解釋性、用於商業決策過程的輔助說明 |
| Class Activation Map (CAM / Grad-CAM) | 用在 CNN 模型中，視覺化神經網路對圖像的注意區域（熱區圖） | 圖像分類、醫學影像判讀（如腫瘤定位）、模型判斷可信度分析 |
| 注意力圖 (Attention Map / Weights) | 顯示 Transformer 模型中注意力分布情形，觀察模型對輸入資訊的關注程度 | NLP 任務（翻譯、問答）、多模態模型分析、語意理解 |
| 混淆矩陣熱圖 (Confusion Heatmap) | 將混淆矩陣視覺化為熱力圖，加強辨識錯誤類型與偏誤 | 模型效能優化、類別不平衡分析、監管說明 |

- 圖片增強與修改：

- Canva AI、Runway ML 提供圖片修復、物件移除、背景更換等功能。

(三) 影片與動畫生成應用

- 快速製作簡報與影片：

- 「請幫我生成一個 AI 發展趨勢的簡報，包含 5 個重點與數據圖表。」（Gamma AI、Tome AI）
- 「根據這篇文章內容，生成一段 30 秒的影片腳本。」（Synthesia、Runway ML）

- AI 動畫與特效：

- Pika Labs、Runway ML 可將靜態圖片轉換為動畫，或添加特效（如光影變化）。

(四) 音樂與語音生成應用

- AI 生成音樂：

- AIVA、Suno AI 可自動生成背景音樂，適用於影片、遊戲配樂。

- AI 文字轉語音（Text-to-Speech, TTS）：

- Murf.ai、ElevenLabs 可模仿人類聲音，創造自然的 AI 朗讀語言。

(五) AI 程式開發輔助

- 自動補全與 Debug：

- GitHub Copilot 可即時補全程式碼，提高開發效率。
- Replit AI 可自動偵測錯誤，並建議修正方案。

- 生成 API 文件與測試案例：

- 「請為這段 Python API 代碼生成詳細的 API 說明文件。」
- 「請根據這段程式碼，生成 5 個測試案例。」

三、RAG 實際應用方式總覽

■ 基本流程（簡化說明）

- 使用者提出問題
- 系統將問題進行語意編碼（embedding）
- 在知識資料庫中語意搜尋相關段落
- 將檢索到的段落加入 prompt 提供給 LLM
- LLM 基於資料生成回應，並可附上引用資料來源

大語言模型

四、RAG 系統建構四步驟（實務導入）

步驟一：建立知識資料庫

撰寫高效 Prompt 的技巧

1. 使用清晰的指令：直接說明需求，避免模糊詞彙。
 - 「幫我寫一篇文章」 不明確
 - 「請撰寫一篇 500 字的科技趨勢分析，重點探討 AI 的未來發展。」
2. 指定角色 (Role-based Prompting)：讓 AI 以特定身份回答問題。
 - 「假設你是一名專業行銷顧問，請為科技新創公司撰寫 3 條 Facebook 廣告文案。」
3. 提供背景資訊：讓 AI 更精準理解你的需求。
 - 「這篇文章的目標讀者是 25-35 歲的科技愛好者，請使用簡單易懂的語言。」
4. 設定格式 (Format Instruction)：要求 AI 以特定格式輸出。
 - 「請用 Markdown 格式撰寫一份簡短的 AI 介紹，包括標題、分點說明和範例。」
5. 提供範例 (Few-shot Learning)：
 - 「這是我過去的文章風格：'AI 技術正在快速變革...' 請生成類似風格的內容。」

（一）學習建議

二、生成式 AI 的進階應用技巧

為了更有效地利用 AI，提高產品品質與應用價值，可以使用進階技巧。

(一) 文字生成應用

- 長篇內容撰寫：
 - 分段生成：AI 生成長文時，避免一次請求過長內容，應分段詢問，如：
 - ①先請 AI 提供「文章大綱」。
 - ②再請 AI 撰寫每個段落的詳細內容。
- 改寫與最佳化：
 - 「請幫我改寫這段文字，使其更加正式且流暢。」
 - 「請將這段文章的風格改為幽默風趣，適合社群媒體貼文。」
- SEO 內容優化：
 - 「請根據 '生成式 AI' 這個關鍵字，撰寫 5 條 SEO 友善的標題與描述。」

(二) 圖片生成應用

- 精確描述圖片內容：
 - 「請生成一張 '未來城市'，風格類似科幻電影《銀翼殺手》，霓虹燈光，天空有飛行汽車。」
- 調整細節與風格：
 - Midjourney 支援 風格參數 (Stylize)、解析度設定 (--ar 16:9) 來精細控制輸出結果。

| | | |
|--|--|--------------------------|
| 潛在空間視覺化 (Latent Space Visualization) | 將高維資料的潛在向量透過降維技術 (如 t-SNE、UMAP) 投影至低維空間觀察群聚與分布 | 深度學習編碼器輸出視覺化、生成式模型潛在空間探索 |
| 模型結構圖 (Model Architecture Diagram) | 視覺化神經網路的層次結構、參數量與連接方式 | 教學用途、模型設計溝通、部署說明文件 |
| 演化曲線圖 (Training History Plot) | 顯示模型訓練過程中 loss、accuracy、F1 等指標隨 epoch 的變化 | 模型過擬合偵測、早停觀察、訓練過程優化 |
| 敘述式可解釋圖 (Narrative Explanation Graph) | 將模型預測原因轉為文字敘述或決策樹式圖解，利於非技術用戶理解 | 客戶風險說明書、醫療報告解釋、AI 治理報告製作 |

◆ 2-3 AI 實作分析應用案例

■ 案例一：醫療風險預測模型 (Python + scikit-learn)

- 資料來源：病歷紀錄、實驗室檢查數據
- 處理步驟：空值補齊、標準化、類別變數編碼
- 分析方法：邏輯迴歸預測住院風險，並用 SHAP 解釋模型
- 視覺呈現：使用 matplotlib 畫 ROC 曲線與特徵重要性圖

■ 案例二：顧客流失預測 (Excel + Power BI + AutoML)

- 資料來源：CRM 客戶資料、過去消費紀錄
- 處理方法：用 Excel 進行分群、建立 RFM 模型
- 模型建立：用 Power BI AutoML 模型預測可能流失的顧客
- 應用效果：提升顧客召回率 20%、降低營運成本

■ 案例三：智慧製造故障預測 (Python + IoT 資料)

- 資料來源：感測器資料串流
- 處理方法：時間序列處理、數據增強
- 模型：使用 LSTM 模型預測設備異常時間點
- 成果：提早 2 小時預警，降低停工率 15%

L11203 資料隱私與安全

AI 依賴大量數據進行學習，但數據的收集、存儲與使用涉及隱私保護、資訊安全、合規性等議題。本筆記將詳細解析 AI 資料隱私與安全的概念、風險、技術解決方案與國際法規。

一、AI 資料隱私與安全的基本概念

(一) 什麼是 AI 資料隱私 (Data Privacy)？

- AI 需要使用者個人數據（如姓名、位置、行為紀錄）來訓練模型，但若未妥善保護，可能導致個資洩漏或濫用。
- 隱私保護（Privacy Protection）目標：**
 - 限制數據收集範圍：AI 僅應收集必要數據，避免過度監控。
 - 數據匿名化與去識別化：防止 AI 訓練數據直接識別個人。
 - 用戶數據控制權：讓使用者可選擇是否授權 AI 使用數據。

（二）什麼是 AI 資料安全（Data Security）？

- 資料安全（Data Security）指的是防止 AI 相關數據被未授權存取、竊改、盜用，確保 AI 訓練數據的完整性與機密性。
- 安全保護目標：**
 - 防止駭客攻擊（如數據竊取、勒索軟體）。
 - 防止 AI 模型竊取（如對抗攻擊 - Adversarial Attacks）。
 - 確保數據完整性（避免 AI 被污染，導致錯誤決策）。

二、AI 資料隱私與安全的主要風險

AI 在數據收集與應用過程中，可能面臨數據洩露、偏見、對抗攻擊、監控風險等挑戰。

（一）數據洩露風險

- 企業數據庫洩露：AI 企業若未妥善保護客戶數據，駭客可能入侵數據庫竊取個資。
 - 案例：Facebook 2019 年 5.33 億用戶個資外洩事件。
- 雲端 AI 服務漏洞：使用 AI SaaS（如 ChatGPT API）時，未妥善加密可能導致機密資訊洩露。

（二）AI 偏見（Algorithmic Bias）

- AI 模型的訓練數據可能來自歷史數據，若數據本身具有偏見，AI 可能做出不公平決策。
 - 案例：
 - Amazon 招聘 AI 偏見（2018）：AI 偏好男性應徵者，因訓練數據以男性履歷為主。
 - 信用評分 AI 偏見：某些 AI 模型可能無意中歧視特定族群，影響貸款核准率。

（三）對抗攻擊（Adversarial Attacks）

- 惡意修改 AI 訓練數據，使其產生錯誤預測。
 - 案例：
 - 自駕車 AI 被干擾：在道路標誌上貼小貼紙，使 Tesla 誤判「STOP」標誌為「限速 45」。
 - Deepfake 技術：攻擊 AI 影像辨識技術，生成逼真的假影像。

（四）AI 監控與個資濫用

- 政府或企業使用 AI 監控人民行為，可能侵犯隱私權。
 - 中國的 AI 監控系統：使用 AI 進行社會信用評分、監視市民行動。

七、生成式 AI 工具應用分類與說明

| 類別 | 代表工具 | 功能概述 |
|-----------|------------------------------------|--------------------|
| 文字生成（LLM） | ChatGPT、Claude、Gemini | 聊天助理、文件生成、摘要、翻譯 |
| 程式碼生成 | GitHub Copilot、CodeWhisperer | 自動補全、生成與除錯程式碼 |
| 圖像生成 | Midjourney、DALL-E、Stable Diffusion | 文生圖、風格轉換、AI 藝術創作 |
| 影片生成 | Sora (OpenAI)、Runway、Pika | 文生影片、影片重建、動畫合成 |
| 音訊生成 | ElevenLabs、Voicemod、AIVA | 語音仿真、音樂生成、音效創作 |
| 文件與簡報生成 | Notion AI、Tome、Gamma | 文件草擬、報告生成、簡報製作 |
| 設計與多媒體 | Canva AI、Adobe Firefly、Designs.ai | 平面設計輔助、圖文合成 |
| 商業與行銷 | Jasper AI、Copy.ai、Writesonic | 行銷文案、社群貼文、自動廣告文生成 |
| 法律與醫療輔助 | Harvey (法律)、Glass AI (醫療) | 合約審閱、診斷摘要、報告初稿 |
| 多模態工具 | Gemini、Grok、Perplexity | 同時理解並生成文字、圖像、語音等內容 |

L12202 如何善用生成式 AI 工具

一、生成式 AI 工具的基本使用策略

要善用生成式 AI，應從選擇合適的工具、優化輸入提示（Prompt Engineering）、有效驗證與改進結果三個方面入手。

（一）選擇合適的 AI 工具

不同的生成式 AI 工具有不同的專長，選擇適合的工具能提升工作效率與成果品質。

| 應用類型 | 推薦工具 |
|----------------|---|
| 文字生成（寫作、行銷、客服） | ChatGPT、Claude、Google Gemini、Copy.ai |
| 圖片生成（插畫、設計、海報） | DALL-E、Midjourney、Stable Diffusion、Canva AI |
| 影片與動畫生成 | Runway ML、Pika Labs、Synthesia |
| 音樂與語音生成 | Suno AI、AIVA、Murf.ai、ElevenLabs |
| 程式碼生成與開發 | GitHub Copilot、CodeWhisperer、Replit AI |

（二）提高 AI 輸入指令的精確度（Prompt Engineering）

生成式 AI 的結果高度依賴使用者輸入的指令（Prompt），以下是優化 Prompt 的技巧：

| 項目名稱 | 技術核心 | 模型結構 | 特殊技術 | 優勢 | 技術 |
|------------------|-----------------------------|------------------------------|-----------------------------|-----------------------|-------------------|
| Claude | 自回歸生成 + 憲法式 AI | Claude 2.x (Transformer) | Constitutional AI (人權式規範訓練) | 長上下文處理佳、偏向安全穩健應用 | 商業應用未如 ChatGPT 普及 |
| Gemini | 多模態生成模型 | Gemini (Transformer + 視覺模組) | 多模態學習 (文字、圖像、程式碼) | 跨模態能力強、理解指令靈活 | 圖像輸出尚不如專用圖像模型 |
| GitHub Copilot | 語言模型轉換至程式生成 | Codex (GPT 專為程式碼調整) | 語境理解、即時補全 | 程式碼生成快、語法合理性高 | 複雜邏輯結構掌握能力仍有限 |
| Midjourney | 封閉式 Diffusion 模型 | 改良式 Diffusion + CLIP | 輸入提示詞風格擴增 | 美術與風格強、圖像美感高 | 技術細節不公開、缺乏透明性 |
| DALL-E 3 | 擴散模型 + 語意理解 | DALL-E 3 (Diffusion + GPT) | 多模態語義結構對齊 | 可圖文互動生成、整合 ChatGPT 使用 | 對細節控制不如 SD |
| Stable Diffusion | 潛在擴散模型 (Latent Diffusion) | U-Net + VAE + CLIP | 開源、自訂 LoRA 模組 | 模組化強、可客製訓練、社群活躍 | 建模與資源需求高 |
| Sora AI (影片) | 擴散式影片生成模型 | 改良式 Video Diffusion + 時間建模 | 物理理解、運動建模、多階段生成 | 視覺真實感高、具敘事連貫性 | 尚未公開，一般用戶難以試用 |
| Runway (Gen-2) | 影片生成與編輯 Diffusion | Text2Video / Image2Video 模型 | 支援輸入影片參考與風格控制 | 社群導向、設計者友善介面 | 輸出品質受限於算力 |
| Jasper AI | 商業導向 LLM 微調 | GPT-3 / 定制 LLM | 風格模版、行銷語彙強化 | 適合行銷文案、具多語支援 | 不適合學術、技術文本 |
| Notion AI | 語意摘要 + 生成 | 整合 GPT + 語義切割模組 | 結構化任務建議 | 文件摘要與規劃流程優化 | 功能受限於 Notion 框架 |

- 社交媒體數據濫用：如 Facebook 被指控用 AI 分析用戶行為，投放政治廣告影響選舉 (Cambridge Analytica 事件) 。

三、AI 隱私與安全技術解決方案

為解決 AI 資料隱私與安全問題，目前有多種技術方案，包括數據匿名化、聯邦學習、差分隱私、對抗樣本防禦等。

(一) 數據匿名化與去識別化

- 數據匿名化 (Data Anonymization) ：
 - 在 AI 訓練前，移除個人識別資訊 (PII, Personally Identifiable Information) 。
 - 應用範例：醫療 AI 訓練數據時，去除病人姓名、ID 。
- 去識別化 (De-identification) ：
 - 將數據進行模糊處理，如將年齡「32 歲」改為「30-35 歲」。

(二) 差分隱私 (Differential Privacy)

- 原理：透過「添加噪聲」來隱藏個人數據，使攻擊者無法還原個資。
- 應用範例：
 - Apple、Google 在 AI 訓練時使用差分隱私，確保使用者資料不會洩露。

(三) 聯邦學習 (Federated Learning)

- 原理：
 - AI 不需將數據傳輸至中央伺服器，而是在本地設備進行訓練，僅傳輸模型參數。
- 應用範例：
 - Google Gboard : AI 自動學習輸入法習慣，但不會上傳用戶的個人鍵入內容。

(四) 對抗樣本防禦 (Adversarial Defense)

- 防範 AI 對抗攻擊的技術，確保 AI 不被惡意干擾。
 - Adversarial Training (對抗訓練)：讓 AI 學習應對攻擊樣本。
 - 模型防護技術 (Defensive Distillation)：提高 AI 對微小變化的穩健性。

四、AI 相關法規與標準

各國政府已推出 AI 相關法規，以規範 AI 在隱私與安全方面的應用。

(一) 歐盟 AI 法規 (EU AI Act)

- ◆ 1-1. 高風險 AI 系統的隱私與資料要求 (High-risk AI Systems)
- ！ 高風險系統（如於醫療診斷、信貸審查、雇用甄選、自駕系統）必須強制遵守以下規範：

| 項目 | 說明 |
|---|---------------------------------------|
| <input checked="" type="checkbox"/> 資料治理與品質管理制度 | 所用資料不得有偏見，須經審核其合法來源、正當取得與代表性，並須防止歧視 |
| <input checked="" type="checkbox"/> 隱私保護技術設計 (Privacy by Design) | AI 模型在設計階段即須考慮隱私與資安，例如不使用可識別資訊、導入差分隱私 |
| <input checked="" type="checkbox"/> 模型訓練記錄與資料可追溯性 | 必須保留訓練與測試過程、資料版本與模型調整記錄 |
| <input checked="" type="checkbox"/> 人為監督與干預機制 | 高風險系統需能提供人工介入點，避免完全無人監控的自動決策 |

| | | |
|------|--------------|--------------------|
| 語言翻譯 | 多語翻譯與文化適應 | DeepL、ChatGPT 翻譯功能 |
| 會話代理 | AI 聊天客服、自動問答 | Dialogflow + GPT |

◆ 1-2. 明確禁止侵犯隱私之 AI 系統

| 禁止項目 | 說明 |
|--|----------------------------------|
| <input checked="" type="checkbox"/> 即時遠端生物特徵監控 (Real-time biometric surveillance) | 禁止無正當理由在公共場合使用 AI 進行臉部辨識與情緒判讀 |
| <input checked="" type="checkbox"/> 社會評分系統 (Social Scoring) | 禁止政府或企業使用 AI 對人民或用戶進行行為評分制度 |
| <input checked="" type="checkbox"/> 操縱性 AI | 禁止運用心理誘導或微操控影響使用者意志（特別是兒童、身心障礙者） |

(二)圖像與設計生成

| 子領域 | 功能 | 工具範例 |
|---------|---------------|------------------------------------|
| 文生圖 | 圖像創作、插畫、視覺素材 | Midjourney、DALL-E、Stable Diffusion |
| AI 藝術 | 藝術創作、風格模擬 | Runway、Adobe Firefly |
| 圖像修復與轉換 | 去雜訊、風格遷移、模擬拍攝 | Real-ESRGAN、StyleGAN |

◆ 1-3. 資料與隱私透明告知機制

| 要求 | 說明 |
|---|---------------------------------------|
| <input checked="" type="checkbox"/> 使用 AI 系統時須告知用戶 | 如聊天機器人、深偽技術、推薦引擎等，須標示其為 AI 系統 |
| <input checked="" type="checkbox"/> 須揭露 AI 使用之資料類型與目的 | 高風險系統應向監管機關登錄資料使用摘要與風險報告 |
| <input checked="" type="checkbox"/> 使用者可申訴或要求審查 AI 決策 | AI 決策若對個人有重大影響，應提供人工覆審與說明權（與 GDPR 結合） |

五、AI 安全方面的應用要求

◎ 安全與可控性的強化條款

| 安全項目 | 法案要求 |
|-------------------------------|---------------------------------|
| 模型穩健性 (Robustness) | 必須經測試能抵抗錯誤資料、對抗樣本、不預期輸入 |
| 資安保護 (Cybersecurity) | AI 系統部署環境需具備防駭、加密、防止中間人攻擊等機制 |
| 風險評估機制 (Risk Management) | 高風險 AI 須事先提交風險分析報告，並建立持續監控與改進機制 |

(三)程式碼生成與開發協助

| 功能 | 工具 | 說明 |
|----------|----------------------------|---------------|
| 自動補全 | GitHub Copilot | 即時生成程式碼片段 |
| 自然語言轉程式 | ChatGPT (Code Interpreter) | 將指令轉為可執行程式碼 |
| 程式碼除錯與重構 | CodeWhisperer、Replit AI | 分析邏輯錯誤、重寫功能模組 |

(四)多媒體生成應用

| 子領域 | 功能 | 工具範例 |
|-------|-----------|-----------------------|
| 影片生成 | 短影片與動畫合成 | Sora、Runway、Pika Labs |
| 音樂與聲音 | 語音仿真、音樂創作 | ElevenLabs、AIVA、Suno |
| 簡報生成 | 商務簡報與教學模板 | Gamma、Tome、Notion AI |

(五)特定產業應用

| 產業 | 生成式 AI 應用方式 | 工具實例 |
|-----|------------------|-----------------------------------|
| 教育 | 個人化教材、學習摘要、作業生成 | ChatGPT、Khanmigo |
| 金融 | 財務報告初稿、市場摘要、自動分析 | BloombergGPT、ChatGPT Finetune |
| 法律 | 合約解析、法律條文重寫 | Harvey AI、Claude |
| 醫療 | 病歷摘要、知識圖譜生成、影像說明 | Glass AI、BioGPT |
| 製造業 | 設計樣稿生成、維修建議 | Autodesk AI、Siemens MindSphere AI |

六、生成式 AI 工具技術原理分析總覽

| 工具名稱 | 技術核心 | 模型架構 | 特殊技術 | 優勢 | 限制 |
|---------|---------|---------------------|--|--------------|-----------------|
| ChatGPT | 自回歸生成模型 | GPT-4 (Transformer) | RLHF、Instruction tuning、Function calling | 文字生成品質高、泛用性強 | 回答受限於訓練資料、可控性有限 |

◆ 六、與 GDPR 的關聯

- 優勢：生成影像細節豐富、可控性高
- 應用：
 - 圖像生成（如 Stable Diffusion、DALL-E 3）
 - 影片生成（如 Sora）
 - 創意設計與模擬物理場景

5. Transformer 與自注意力機制 (Self-Attention)

- 原理：模型可對輸入序列中的每個元素計算與其他元素的關係權重（attention scores）
- 優勢：可平行運算、長距離依賴建模佳
- 應用：自然語言處理、圖像理解、跨模態語意比對

6. 多模態技術 (Multimodal AI)

- 原理：結合多種資料型態（文字、圖像、語音等）進行學習與生成
- 技術關鍵：
 - 共表示學習（joint embeddings）
 - 模態對齊（如 CLIP 結合圖像與文字語意）
- 應用：
 - 文生圖（如 DALL-E、Midjourney）
 - 圖片說明生成（image captioning）
 - 文生影片（如 Sora）

四、生成式 AI 模型訓練與應用方式

訓練階段技術

| 階段 | 技術重點 | 說明 |
|---|-----------------|----------------|
| 預訓練 (Pretraining) | 大規模語料或圖像資料訓練 | 建立通用語言或視覺知識 |
| 微調 (Fine-tuning) | 根據特定任務調整參數 | 增強特定應用表現 |
| 指令微調 (Instruction Tuning) | 教模型如何根據「人類指令」行動 | 提升對話與任務執行能力 |
| RLHF (Reinforcement Learning with Human Feedback) | 強化學習+人類偏好回饋 | 控制生成行為、提升使用者體驗 |
| 多模態訓練 | 同時處理多模態資料 | 提供圖文語意一致性與理解力 |

五、生成式 AI 實務應用方式

(一)文本生成應用

| 子領域 | 功能 | 工具範例 |
|------|------------|-------------------|
| 自動寫作 | 報告、故事、小說生成 | ChatGPT、Jasper AI |
| 知識摘要 | 法律、醫療、政策摘要 | Claude、Glass AI |

| 項目 | GDPR | EU AI Act |
|------|-----------------------|----------------------|
| 適用範圍 | 所有個人資料處理 | 所有 AI 系統（含非個資） |
| 核心關注 | 個人資料的合法使用 | AI 系統的可控性與風險管理 |
| 關於隱私 | 提供同意權、查詢權、刪除權等 | 提供 AI 使用告知、人工干預與申訴機制 |
| 合作性 | AI 系統使用個資時必須同時遵守 GDPR | 若產生自動決策行為須符合雙法規要求 |

(一) 歐盟一般數據保護規範 (GDPR)

- AI 若處理個資，必須符合：
 - 使用者同意（Opt-in consent）。
 - 數據可刪除權（Right to be forgotten）。

(二) 美國 AI 監管政策

- 強調 AI 責任制：若 AI 造成損害，企業需負責。
- 確保 AI 透明度：企業需提供 AI 決策依據。

(三) 台灣個資法與 AI 治理草案

■ 個人資料保護法

- 強調：
 - 資料蒐集需合法特定目的
 - 當事人擁有查詢、更正、刪除權
 - 政府部門、公私機構皆適用
- 增設條文針對「自動化決策」與「跨境傳輸」

■ 數位發展部《人工智能基本法》（草案方向）

- 重點規劃：
 - 高風險 AI 使用資料應具備「合法來源」與「偏誤檢測」
 - AI 應提供解釋與人為干預機制（Human-in-the-loop）
 - 公部門導入 AI 須公開紀錄與透明報告

七、企業 AI 隱私與安全實踐

企業在 AI 應用中，應建立內部隱私與安全治理框架。

企業 AI 資料安全實施步驟

- (一) 數據最小化 (Data Minimization)：只收集必要數據，避免過度監控。
- (二) 逸化存取控制 (Access Control)：限制 AI 訓練數據的存取權限。
- (三) 定期隱私與安全審查 (Privacy & Security Audits)：確保 AI 遵循最新法規。

L113 機器學習概念

L11301 機器學習基本原理

一、機器學習的基本原理

| 項目 | 說明 |
|----|----|
| | |

| | |
|------|--|
| 定義 | 機器學習 (Machine Learning, ML) 是一種使計算機系統能夠根據資料進行預測、分類或決策的技術，無需透過明確的程式編碼。源自 Arthur Samuel (1959) 提出的概念：「讓電腦能在沒有被明確編程的情況下學習」。 |
| 工作機制 | 機器從資料中發現模式 → 構建數學模型 → 進行預測或決策 → 接受回饋調整模型。 |
| 特徵 | 資料驅動、動態更新、自動化運算能力、可解釋性 (部分模型)、高效能處理大規模資料。 |

學習類型與邏輯架構：

1. 監督式學習 (Supervised Learning)
 - 使用已標籤資料學習，對未來數據做分類或回歸預測。*導性/邏輯回歸*
2. 非監督式學習 (Unsupervised Learning)
 - 無標籤資料，模型透過內部結構探索分類、聚類、降維等。
3. 半監督式學習 (Semi-Supervised Learning)
 - 混合少量標籤與大量未標籤資料，提升效率。
4. 強化學習 (Reinforcement Learning)
 - 透過環境互動，學習最大化累積獎勵的策略，常用於遊戲與自駕技術。

二、機器學習核心技術

| 技術名稱 | 說明 | 應用 |
|----------------------------|------------------------------|--------------------------------|
| 特徵工程 (Feature Engineering) | 將原始資料轉換為模型可理解的特徵形式。 | 銀行信用評分、醫療風險預測 |
| 模型訓練與驗證 | 使用訓練集進行模型學習，透過驗證集與測試集檢視泛化能力。 | 建立房價預測模型 |
| 損失函數 (Loss Function) | 衡量模型預測值與真實值的誤差，是模型優化的依據。 | 均方誤差 (MSE)、交叉熵 (Cross Entropy) |
| 最佳化演算法 (Optimizer) | 用於調整模型參數以最小化損失函數。 | SGD、Adam |
| 過擬合與正則化 ① ② | 避免模型在訓練資料上過度學習，使其無法泛化到新資料。 | Dropout、L1/L2 正則化 |

① *無法正確預測新資料* ② *加入額外資訊*

三、機器學習架構模型

| 類型 | 說明 | 適用任務 |
|----------------------------|--------------------------------------|---------|
| 線性回歸 (Linear Regression) | 預測連續變數，假設輸入特徵與輸出之間具線性關係。 <i>直線關係</i> | 房價、銷售預測 |
| 邏輯回歸 (Logistic Regression) | 處理二分類問題，輸出值介於 0~1 之間。 | 客戶流失預測 |
| 支援向量機 (SVM) | 找到分類邊界最大化的超平面。 | 影像識別 |

三、生成式 AI 的核心技術架構總覽

生成式 AI 的發展奠基于深度學習，主要包含以下六大核心技術群：

| 技術類別 | 說明 | 代表模型 |
|-------------------------------|------------------------|------------------------------|
| 自回歸模型 (Autoregressive Models) | 透過序列上下文，逐步預測下一個資料點 | GPT、Transformer-XL |
| 變分自編碼器 (VAE) | 建立隱變數模型，從潛在空間中抽樣生成 | VAE、 β -VAE |
| 對抗式生成網路 (GAN) | 生成器與鑑別器對抗學習，提升生成真實度 | StyleGAN、CycleGAN |
| 擴散模型 (Diffusion Models) | 利用逐步噪聲還原技術生成高品質影像或影片 | Stable Diffusion、Imagen、Sora |
| 自注意力機制與 Transformer 架構 | 深度語意理解與序列建模的基礎 | GPT、BERT、T5 |
| 多模態對齊與融合技術 | 整合不同模態 (圖文音訊) 並進行跨模態生成 | CLIP、Flamingo、Gemini |

(一)各項核心技術詳細解析

1.自回歸模型 (Autoregressive Models)

- 原理：將序列資料分解成條件機率鏈，每一步根據前面預測下一步
 $P(x) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\dots$
- 代表模型：GPT 系列 (GPT-1, 2, 3, 4)
- 應用：文字生成、聊天對話、寫作輔助、程式碼生成

2.變分自編碼器 (VAE)

- 原理：結合編碼器與解碼器，在潛在空間進行機率建模與重建
- 特性：學習資料的分布與潛在特徵
- 應用：影像生成、風格混合、特徵控制 (如臉部表情變化)

3.對抗式生成網路 (GAN)

- 原理：包含兩個神經網路：
 - 生成器 (G)：產生假樣本
 - 鑑別器 (D)：辨識真假樣本
- 訓練方式：G 與 D 不斷競爭，提升生成樣本的真實性
- 應用：AI 藝術創作、虛擬人臉、風格轉換 (如馬轉斑馬)

4.擴散模型 (Diffusion Models)

- 原理：
 - 前向過程：將資料逐步加入隨機噪聲
 - 逆向過程：學習如何從噪聲重建原始資料

五、No code/Low code 平台的選擇與評估

| 評估欄面 | 問題引導 | 說明與關鍵要素 |
|-----------|--------------------------------|--------------------------------|
| ① 功能完整性 | 是否具備你所需的應用類型 (Web/App/自動化) ? | 表單、工作流、資料管理、外部整合能力 |
| ② 擴展彈性 | 是否可加入自訂邏輯/API/程式碼 ? | 支援程式碼片段 (JS、Python) 、Webhook |
| ③ 整合能力 | 是否能連接現有系統 (ERP 、CRM 、資料庫) ? | 支援 API 、資料來源整合、 OAuth 、SQL |
| ④ 學習門檻 | 使用者是否易於上手 ? | UI/UX 親和度、文件完整性、社群支持 |
| ⑤ 部署與維運 | 是否可控制資料存取與運行環境 ? | 雲端 / 本地部署、權限角色、版本控制 |
| ⑥ 成本與授權模式 | 價格模型是否合理 ? | 計畫分級、使用者授權、功能限制 |
| ⑦ 資安與合規性 | 是否符合組織內部與法規的資安需求 ? | 資料加密、稽核紀錄、GDPR/ISO27001 等支援 |

L122 生成式 AI 應用領域與工具使用

L12201 生成式 AI 應用領域與常用工具

一、生成式 AI 的基本概念

生成式人工智慧 (Generative AI) 是一類能夠學習輸入資料的分布，並根據此分布產生新的、有意義內容的 AI 系統。這些內容可以是文字、圖像、音訊、影片、程式碼或其他格式的資料。

◆ (一) 定義：

生成式 AI 是一種基於機率模型與深度學習技術，能夠根據所學資料生成新樣本的 AI 系統。

◆ (二) 特性：

- 能從資料中捕捉潛在分布 (latent distribution)
- 可用於創造性任務 (creative tasks)
- 支援跨模態生成 (如從文字生成圖像)

二、生成式 AI 的技術演進歷程

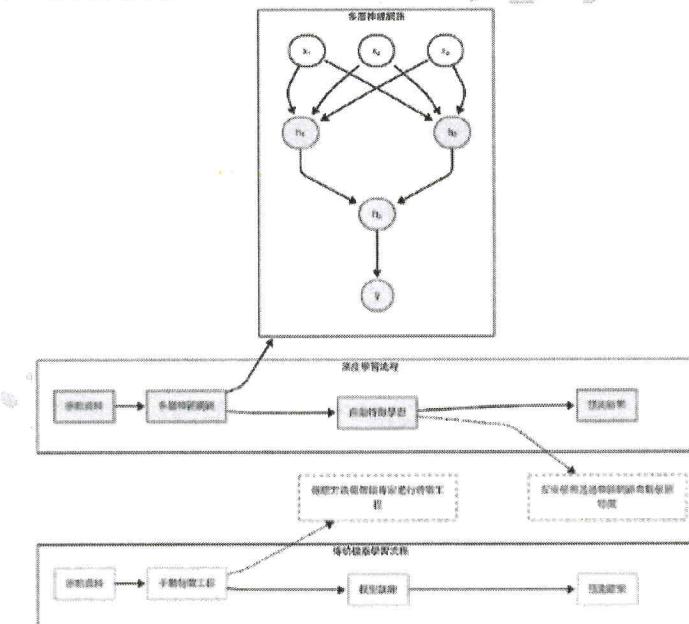
| 階段 | 代表技術 | 說明 |
|---------|------------------------------|----------------------------|
| 初期生成模型 | 自回歸模型 (AR) 、隱馬可夫模型 (HMM) | 傳統統計方法，處理語音與簡單序列生成 |
| 機率生成模型 | Naive Bayes 、 LDA | 用於主題建模與基本文本生成 |
| 深度生成模型 | VAE 、 GAN | 可產生高品質圖像與樣本 |
| 預訓練生成模型 | Transformer 、 GPT | 利用大規模語料訓練，支援上下文理解與自然生成 |
| 多模態生成模型 | DALL-E 、 GPT-4 、 Sora | 同時理解與生成多種資料模態 (文字、圖像、影片) |

| | | |
|--------------------------|--------------------------|------|
| 決策樹與隨機森林 <i>多子決策樹</i> | 使用分支進行分類或回歸，隨機森林為集成學習方法。 | 醫療診斷 |
|--------------------------|--------------------------|------|

3-2. 深度學習模型 (Deep Learning)

| 類型 | 特徵 | 應用場景 |
|---------------------|---|-------------|
| 人工神經網路 (ANN) | 由多層神經元組成，可處理複雜資料關聯。 | 銷售預測、風險管理 |
| 卷積神經網路 (CNN) | 適用於影像資料處理，提取空間特徵。 | 車牌辨識、醫學影像 |
| 循環神經網路 (RNN/LSTM) | 對時間序列資料建模，保留上下文資訊。 | 聲音識別、自然語言處理 |
| 轉換器 (Transformer) | 使用注意力機制處理長距離依賴問題，是大型語言模型 (如 ChatGPT) 的核心架構。 | 自然語言生成、翻譯 |

四、圖解架構模型示意



L11302 常見的機器學習模型

一、機器學習類型與學習類型中常見模型

| 類型 | 定義與核心概念 | 常見模型與演算法 | 典型應用情境 |
|--------|--|---|--------------------------|
| 監督式學習 | 以 <u>已標記資料</u> (labelled data)進行訓練，目標為預測或分類。 | 線性迴歸、邏輯迴歸、支援向量機(SVM)、決策樹、隨機森林、梯度提升機(GBM)、KNN、神經網路 | 信用評分、醫療診斷、銷售預測、語音辨識 |
| 非監督式學習 | 使用 <u>無標記資料</u> (unlabelled data)，模型從資料中找出潛在結構與分群。 | K-means、層次式分群(Hierarchical Clustering)、主成分分析(PCA)、自編碼器(Autoencoder) | 客戶分群、異常檢測、降維處理、推薦系統 |
| 半監督式學習 | 結合 <u>少量標記</u> 與大量未標記資料，提升模型泛化能力與準確度。 | 混合式模型、半監督支援向量機、圖神經網路(GNN) | 網路內容分類、生物醫學圖像處理、語音識別 |
| 強化學習 | 在環境中透過行動與回饋(獎勵/懲罰)進行學習，追求策略最佳化。 | Q-learning、Deep Q-Network(DQN)、Policy Gradient、Actor-Critic | 自駕車、機器人控制、AlphaGo、投資決策系統 |
| 深度學習 | 多層神經網路架構，自動抽取特徵並解決高維非線性問題。 | CNN、RNN、LSTM、Transformer、BERT、GPT、GAN | 圖像識別、語音合成、自然語言處理、生成式AI |

1. 監督式學習 (Supervised Learning)

原理：從帶有輸入特徵與對應標籤的資料中學習輸入與輸出間的映射關係。

| 模型名稱 | 技術原理 | 優點 | 限制 | 應用領域 |
|----------------------------|--|-------------|----------------|------------|
| 線性迴歸 (Linear Regression) | 建立線性函數 $y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$ | 解釋力強、易實作 | 無法建模非線性關係 | 房價預測、銷售預測 |
| 邏輯迴歸 (Logistic Regression) | 機率估計模型，Sigmoid函數將輸出限制於 $[0, 1]$ 區間 | 適合二元分類 | 資料需線性可分 | 疾病預測、信用評估 |
| K 最近鄰 (KNN) | 根據距離最近的 K 筆資料進行多數表決 | 無需訓練、適應新資料快 | 大資料時效率差、受噪音影響大 | 影像分類、推薦系統 |
| 隨機森林 (Random Forest) | 多棵隨機抽樣決策樹進行投票，Bagging 技術 | 避免過擬合、健健 | 無法解釋單一預測邏輯與回歸 | 跨產業分類與回歸 |
| 支援向量機 (SVM) | 尋找最大間隔超平面分隔不同類別 | 高維空間有效 | 計算成本高、參數微調整 | 文本分類、手寫辨識 |
| XGBoost | 梯度提升樹(Boosting)，逐漸地強化殘差 | 高準確率、效能佳 | 參數多、需調優 | 資料科學競賽常勝模組 |

- No Code 平台通常運行於雲端環境，在大規模數據處理(Big Data)或高併發場景下，可能效能不佳。
- 部分 No Code 平台不支援高效能數據庫(如 PostgreSQL, MongoDB)，限制資料處理能力。

解決方案：

- 選擇可自訂後端的 Low Code 平台(如 Retool, Appian)。
- 針對大規模數據處理，仍需使用傳統開發。

(四) 資安與合規風險

部分 No Code 平台儲存資料於雲端，可能有安全風險

- 企業機密數據如果存放於第三方 No Code 平台，可能面臨數據外洩、GDPR/CCPA 合規問題。
- No Code 平台不允許用戶完全控制資料庫與伺服器，影響企業的數據主權。

解決方案：

- 選擇企業級 No Code 平台(如 Microsoft Power Apps)，提供本地部署選項(On-Premises)。
- 確保 No Code 平台符合 GDPR、ISO 27001，並提供數據加密功能。

三、No Code / Low Code 適用場景與選擇指南

| 應用類型 | 適合 No Code | 適合 Low Code | 適合傳統開發 |
|---------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| 企業內部工具(如 CRM、HR 系統) | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| 網站與電商平台 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| 自動化工作流 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| API 整合與資料分析 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| 高併發應用(如金融交易、AI 計算) | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |

四、生成式 AI 與 No code/Low code 的整合

| 平台工具 | 整合特性 | 適合對象與應用 |
|---------------------------|--------------------------|--------------------|
| Zapier + GPT | 用於觸發郵件、回覆、表單自動填寫 | 行政流程、客服自動化 |
| Bubble + OpenAI | UI + Workflows + 插件接 GPT | 建立 Web App + AI 輸出 |
| Power Apps + Azure OpenAI | 支援企業內部部署 GPT 模型 | ERP/CRM 智能擴充 |
| Retool + OpenAI | 接收 SQL 輸出、分析結果產生文字說明 | BI 儀表板文字摘要 |
| Make.com / n8n | Workflow 管理 + AI 模型觸發控制 | 跨平台整合、監控流程 |

適用於多種應用場景

- Web 應用：可快速建立電商網站、企業內部管理系統。
- 行動應用：可透過 Thunkable、Adalo、Glide 等 No Code 平台開發手機應用。
- 自動化流程：使用 Zapier、Power Automate、Make (Integromat) 等工具，自動化工作流程。

(五) 提供安全性與合規選項

企業級 No Code / Low Code 平台支援高安全性

- 數據加密 (Encryption)、身份驗證 (Authentication) 和存取控制 (Access Control)。
- 支援 GDPR、ISO 27001、SOC 2 等國際安全標準。

二、No Code / Low Code 的限制

(一) 可擴展性與靈活性受限

No Code 平台難以處理高度客製化需求

- No Code 平台雖然適合標準業務流程，但對於複雜的業務邏輯、演算法或 AI 計算需求，仍需傳統開發。
- 範例：
 - AI 影像辨識應用：需要 TensorFlow 或 PyTorch，No Code 平台難以支持。
 - 高併發交易系統（如金融交易）：需要低延遲、高效能的 API 設計，No Code 不易處理。

解決方案：

- Low Code 平台支援部分程式碼開發，可以通過 API 擴展功能，如 OutSystems、Microsoft Power Apps。

(二) 平台依賴性 (Vendor Lock-in)

遷移困難，受限於特定平台

- 企業如果長期使用某個 No Code / Low Code 平台，將會依賴該平台的生態系統，未來難以轉移至其他技術堆疊 (Tech Stack)。
- 部分 No Code 平台不提供完整的原始碼 (Source Code)，使企業無法自行遷移應用程式。

解決方案：

- 選擇支援開源或可導出原始碼的 Low Code 平台（如 Bubble、OutSystems）。
- 使用具備 API 整合能力的平台，以降低依賴風險。

(三) 效能與運行限制

No Code 平台可能有性能瓶頸

2. 非監督式學習 (Unsupervised Learning)

原理：資料無標籤，模型自動找出資料內部的群組或結構。

| 模型名稱 | 技術原理 | 優點 | 限制 | 應用領域 |
|---------------------------------|---------------------------------|----------|---------------|-------------|
| K-means | 最小化群內平方誤差 (SSE)，資料點分配至最近的中心 | 快速、適用大資料 | K 值需預設、對初始值敏感 | 客戶分群、影像分割 |
| 階層式分割 (Hierarchical Clustering) | 自底向上或自頂向下聚合，產生樹狀結構 (dendrogram) | 不能指定群數 | 計算複雜度高 | 基因資料類別、族群分類 |
| 主成分分析 (PCA) | 線性降維方法，找出最大變異方向 | 可視化、加速計算 | 資訊可能遺失 | 特徵縮減、資料前處理 |
| Autoencoder | 神經網路壓縮與還原輸入資料，學習資料的嵌入空間 | 降維、資料生成 | 訓練複雜、需調參 | 異常偵測、影像去噪 |

3. 半監督式學習 (Semi-Supervised Learning)

原理：用少量標記資料 + 大量未標記資料共同訓練，提升準確性。

| 常見方法 | 技術機制 | 應用範例 |
|----------------------------|---------------------|--------|
| Pseudo Labeling | 用模型預測未標資料標籤，納入下一輪訓練 | 語音辨識 |
| Self-training | 模型反覆預測+再訓練，迭代改善模型 | 網頁分類 |
| Consistency Regularization | 相同資料在擾動後應產生一致結果 | 醫療影像分割 |

4. 強化學習 (Reinforcement Learning)

原理：智能體 (agent) 與環境互動，根據回饋 (reward) 學習策略 (policy)。

| 模型 | 說明 | 應用 |
|----------------------|--|-------------|
| Q-learning | 使用 Q 表儲存狀態-行動值，更新公式： $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max Q(s', a') - Q(s, a)]$ | 遊戲 AI、機器人導航 |
| Deep Q-Network (DQN) | 使用深度神經網路逼近 Q 表，處理高維狀態空間 | 自駕車 |
| Policy Gradient | 直接對策略函數求導，適合連續行動空間 | 操控控制系統 |
| Actor-Critic | 結合策略函數 (Actor) 與價值函數 (Critic) | 多智能體系統 |

5. 深度學習 (Deep Learning)

特性：以神經網路架構進行非線性映射與特徵自動提取。

| 模型 | 技術重點 | 應用 |
|--------------|--------------------|---------------|
| CNN (卷積神經網路) | 卷積層提取局部特徵，常用於影像與視覺 | 人臉辨識、醫療影像分析 |
| RNN (遞迴神經網路) | 資料有時間序列依賴，記憶過往狀態 | 語音識別、股市預測 |
| LSTM (長短期記憶) | 解決 RNN 長期依賴問題 | 自然語言建模 |
| Transformer | 基於注意力機制，並行運算效果佳 | GPT/BERT 語言模型 |
| GAN (生成對抗網路) | 生成器與判別器對抗訓練 | 假新聞圖像生成、深度技術 |

二、模型訓練與資料處理技術

2-1 資料處理流程 (Data Preprocessing)

| 技術分類 | 技術說明 | 細節與範例 |
|--------|-----------------|---------------------------------|
| 資料清理 | 移除不一致資料與缺值處理 | 均值填補、眾數填補、KNN 補值 |
| 特徵縮放 | 讓特徵落在相同範圍 | Min-Max Scaling、Z-score 標準化 |
| 特徵選取 | 選擇與預測最相關的變數 | Mutual Information、Lasso |
| 資料編碼 | 類別型變數轉為數值型 | One-Hot Encoding、Label Encoding |
| 資料擴增 | 增加資料量以提升泛化能力 | SMOTE、圖像旋轉、翻轉 |
| 降維處理 | 移除冗餘特徵、加速訓練 | PCA、t-SNE、UMAP |
| 時間序列處理 | 滑動窗口、延遲變數、季節性調整 | 用於氣象、股市預測資料整備 |

2-2 模型訓練與分類

| 分類依據 | 說明與常見方法 |
|-------|---|
| 啟動方式 | 批次學習 (Batch)、線上學習 (Online)、小批次學習 (Mini-batch) |
| 模型架構 | 單模型 (如決策樹)、集成學習 (Bagging、Boosting、Stacking) |
| 調參方式 | Grid Search、Random Search、Bayesian Optimization |
| 過擬合控制 | 正規化 (L1、L2)、Dropout、Early stopping |

三、模型評估方法 (Model Evaluation)

分類任務指標

| 指標 | 公式 | 說明 |
|-------------------|---|-----------------------|
| 準確率 (Accuracy) | $\frac{TP+TN}{TP+TN+FP+FN}$ | 所有預測中正確預測佔比，受資料不平衡影響大 |
| 精確率 (Precision) | $\frac{TP}{TP+FP}$ | 預測為正的樣本中有多少是對的 |
| 召回率 (Recall) | $\frac{TP}{TP+FN}$ | 所有實際為正的樣本中有多少被抓出來 |
| F1 分數 | $2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$ | 精確率與召回率的調和平均，平衡考量 |
| ROC-AUC | 曲線下面積 | 衡量模型對不同閾值下的分類能力 |

回歸任務指標

| 指標 | 說明 | 特點 |
|-------------------------|----------------|-------------|
| MAE (平均絕對誤差) | 誤差大小的平均 | 抗離群值能力中等 |
| MSE (均方誤差) | 誤差平方後取平均 | 對大誤差更敏感 |
| RMSE (均方根誤差) | MSE 開根號，單位一致性佳 | 通常與實務單位相同 |
| R ² (決定係數) | 衡量模型對變異的解釋能力 | 越接近 1 表示越準確 |

四、應用案例分析 (Applications with Models)

4-1. 金融業：信用風險評估

| 項目 | 說明 |
|----|----|
| | |

L12102 No Code / Low Code 的優勢與限制

一、No Code / Low Code 的優勢

(一) 提高開發速度

- 大幅縮短開發週期
 - No Code / Low Code 提供拖放式 (Drag & Drop) 開發介面，讓使用者可以快速構建應用程式，而無需手動編寫程式碼。
 - 企業應用開發時間可縮短 50% - 80%，比傳統開發方式快 5~10 倍。

(二) 加速企業數位轉型

- 無需等待 IT 部門開發，業務部門可直接使用 No Code 工具建立應用，提升市場反應速度 (Time-to-Market) 。
 - 允許用戶在開發過程中即時預覽應用效果，並能一鍵發布應用程式，減少測試與部署時間。
- 即時預覽與部署

(三) 降低開發成本

- 減少開發人力需求
 - 不需要大量專業工程師，業務人員可自行開發應用，降低 IT 部門負擔。
 - 適合中小企業或初創公司，減少外包軟體開發的成本。
- 減少維護與更新成本
 - No Code / Low Code 平台提供雲端自動維護與升級，企業無需投入大量人力進行系統維護。
 - 可輕鬆修改應用功能，避免傳統開發需要修改大量程式碼的問題。

(四) 降低技術門檻

- 讓非技術人員也能開發應用
 - 業務人員、產品經理、行銷人員等無需學習程式語言，即可使用 No Code / Low Code 平台開發應用。
 - 適合企業內部系統、自動化流程開發，例如 CRM (客戶關係管理)、訂單管理、工作流自動化。
- IT 部門能夠專注於高價值開發
 - IT 團隊可將開發時間集中於更複雜、創新的項目，如 AI、雲端架構，而將一般業務應用交由 No Code / Low Code 平台處理。

(五) 易於整合與擴展

(六) 支援 API、第三方服務整合

- 支援 API、第三方服務整合
 - No Code / Low Code 平台通常內建 API 連接功能，允許使用者整合 CRM、ERP、Google Sheets、Slack、Stripe (支付)、Twilio (簡訊) 等工具。

- 常見工具：n8n（開源工作流自動化工具）。

四、No Code / Low Code 平台比較

| 平台 | 類型 | 語言支援 | 擅長領域 | 適合角色 |
|------------|----------|-----------------|--------------|----------|
| Bubble | No-code | 無 | Web App 快速建置 | 初創、設計師 |
| Zapier | No-code | 無 | 自動化流程整合 | 行政、流程管理者 |
| OutSystems | Low-code | Java/.NET | 大型企業應用整合 | 系統整合工程師 |
| Mendix | Low-code | Java / JS | 可視化應用 + IoT | IT 團隊 |
| Power Apps | Low-code | Power Fx, JS | 微軟系統生態整合 | 資訊人員、PM |
| Retool | Low-code | JavaScript, SQL | 數據內部工具開發 | 資料分析師 |

五、No Code / Low Code 的優勢與挑戰

(一) No Code / Low Code 的優勢

- 開發速度快：開發時間可縮短 70% 以上。
- 降低開發成本：無需大量程式開發人員，企業可自行開發應用程式。
- 適合業務用戶：不具備技術背景的使用者也能參與開發，提高生產力。
- 靈活性高：透過 API 整合，可連接現有系統（如 ERP、CRM）。

(二) No Code / Low Code 的挑戰

- 靈活性有限：No Code 平台僅能實現基礎功能，複雜需求仍需傳統開發。
- 平台依賴：使用特定 No Code 平台，可能受限於其生態系統。
- 安全性與合規風險：企業數據可能存放於第三方平台，須注意 GDPR / ISO 27001 合規性。
- 性能限制：No Code 平台在處理大量數據或高併發請求時可能表現不佳。

六、No Code / Low Code 的未來發展

(一) AI 輔助 No Code / Low Code 開發：

- AI Code Generation：如 GitHub Copilot 自動補全程式碼，提高開發效率。
- 智慧工作流自動化：AI 可協助優化 No Code 平台的業務流程。

(二) 企業級 No Code / Low Code 平臺增長：

- 傳統企業（如銀行、醫療機構）將更多使用 Low Code 來開發內部應用，降低 IT 人員負擔。

(三) No Code / Low Code 與 DevOps 整合：

- 低程式碼工具將支援 DevOps，提供版本控制、自動部署功能，提升 CI/CD（持續整合/持續部署）能力。

| 問題 | 判斷貸款申請人違約風險 |
|-------|---|
| 模型類型 | 監督式學習、分類模型 |
| 常用演算法 | Logistic Regression、XGBoost、Random Forest |
| 資料來源 | 使用者財務資訊、信用歷史、行為資料等 |
| 評估指標 | Precision、Recall、ROC-AUC 分類 |
| 特殊挑戰 | 資料不平衡、模型可解釋性需求（例如 SHAP 解釋） |

4-2. 醫療保健：疾病預測與診斷輔助

| 項目 | 說明 |
|-------|----------------------------------|
| 問題 | 預測病人罹患某疾病的風險（如糖尿病、癌症） |
| 模型類型 | 分類 / 回歸模型（視疾病而定） |
| 常用演算法 | Random Forest、SVM、神經網路 |
| 資料來源 | 電子病歷、檢查報告、影像資料（若結合 CV） |
| 評估指標 | Sensitivity、Specificity、F1-score |
| 特殊挑戰 | 隱私保護、資料異質性與標準化問題 |

4-3. 製造業：設備預測維護（Predictive Maintenance）

| 項目 | 說明 |
|-------|---|
| 問題 | 預測設備可能發生故障的時間或風險 |
| 模型類型 | 回歸 / 分類模型 |
| 常用演算法 | Gradient Boosting、Time Series Model（如 LSTM） |
| 資料來源 | 感測器數據、維修紀錄、機器使用時數等 |
| 評估指標 | RMSE、MAE、Precision（預測故障） |
| 特殊挑戰 | 異常資料稀少、需即時運算 |

4-4. 零售業：顧客購買行為預測

| 項目 | 說明 |
|-------|---|
| 問題 | 預測客戶是否會購買某商品、或推薦產品 |
| 模型類型 | 分類、推薦系統（Ranking / Collaborative Filtering） |
| 常用演算法 | Logistic Regression、Matrix Factorization、深度推薦模型（如 DeepFM） |
| 資料來源 | 顧客購物紀錄、點擊紀錄、個人屬性等 |
| 評估指標 | AUC、Top-K Precision、NDCG |
| 特殊挑戰 | 冷啟動問題、資料稀疏性、個資保護 |

4-5. 公共領域：交通流量預測

| 項目 | 說明 |
|----|------------------|
| 問題 | 預測某地點的交通量或交通壅塞機率 |

| | |
|-------|---|
| 模型類型 | 回歸、時間序列預測 |
| 常用演算法 | ARIMA、LSTM、Graph Neural Network (GNN) |
| 資料來源 | GPS 資料、感測器、天氣資料、歷史路況 |
| 評估指標 | MAE、RMSE、MAPE |
| 特殊挑戰 | 資料即時性需求、空間關聯性強 (需圖神經網路建模) |

L114 鑑別式 AI 與 生成式 AI 概念

L11401 鑑別式 AI 與生成式 AI 的基本原理

一、AI 分類定義：鑑別式 AI vs 生成式 AI

(一) 鑑別式 AI 定義

- 鑑別式 AI 的核心概念是對數據進行分類、預測或識別，根據已有的數據學習決策邊界，找出不同類別的區別。
- 它不生成新數據，而是用來分析、辨識數據所屬的類別。

(二) 常見的鑑別式 AI 模型

| 模型類型 | 代表演算法 | 應用範圍 |
|-----------------------|--|---------------------|
| 回歸 (Regression) | 線性回歸 (Linear Regression) 、邏輯回歸 (Logistic Regression) | 預測房價、信用風險評估 |
| 分類 (Classification) | 支持向量機 (SVM) 、隨機森林 (Random Forest) 、決策樹 (Decision Tree) | 垃圾郵件分類、疾病診斷 |
| 深度學習分類模型 | CNN (卷積神經網路) 、BERT (文本分類) | 人臉識別、金融詐欺偵測 影像處理 |

(三) 鑑別式 AI 的應用

- 影像辨識 (Image Recognition) ：
 - AI 分析影像，判斷是貓或狗、人或非人。
- 自然語言處理 (NLP) ：
 - AI 分類文本，如情感分析 (Positive/Negative) 、垃圾郵件分類。
- 醫療診斷 (Medical Diagnosis) ：
 - AI 分析 X-ray 影像，判斷是否有肺炎。

二、生成式 AI 定義

- 生成式 AI 的核心概念是從數據學習模式，然後生成新的內容，如文本、圖片、音樂、影片。
- 它不只是分類或預測，而是創造新數據，如 ChatGPT 創作文章、DALL-E 生成圖像。

- 混合使用可視化流程 + 寫程式擴充邏輯
- 適用於企業級應用、需與現有系統整合之場景

● 技術能力包括：

- 支援 JavaScript、Python、Java、C# 等
- 自定義資料模型與業務邏輯
- 雲端 API 整合 (REST, GraphQL, OAuth)

三、No Code / Low Code 的主要應用場景

(一) 業務應用開發

- 企業內部工具 (Internal Tools) ：
 - 透過 No Code 平台快速開發請假管理、客戶關係管理 (CRM) 、專案追蹤等工具。
 - 常見工具：Airtable、Notion、Coda。
- 自動化工作流程 (Workflow Automation) ：
 - No Code 平台可設定自動化電子郵件、資料同步、報表生成。
 - 常見工具：Zapier、Make (原 Integromat) 、Power Automate。

(二) 網站與行動應用開發

- 電商網站 / 企業官網開發：
 - 使用 No Code 工具建立網路商店、產品展示網站。
 - 常見工具：Wix、Shopify、Webflow。
- 行動應用開發 (Mobile App) ：
 - 透過 Low Code 工具建立企業級行動應用，如內部訂單系統、員工報銷管理。
 - 常見工具：Adalo、Thunkable、OutSystems。

(三) AI 應用

- AI 駕駛聊天機器人 (Chatbots) ：
 - No Code 平台能夠透過 API 整合 AI (如 OpenAI) 來建立智慧客服。
 - 常見工具：Chatfuel、Landbot。
- AI 自動化數據處理：
 - 透過 Low Code 平台與 AI API 連接，處理數據分析、語音轉文字。
 - 常見工具：Google AutoML、Azure AI Services。

(四) 企業數據整合與 API 應用

- 資料整合與報表自動化：
 - Google Sheets + Zapier：自動將 CRM 數據同步至 Google Sheets，生成報表。
 - 常見工具：Retool、Tableau (Low Code 分析工具)。
- API 整合應用：
 - No Code / Low Code 平台可連接各種 API，如 Stripe (支付) 、Twilio (簡訊) 、Google Maps (地圖)。

MANUS
GPT → 生成式 AI
GPT → 運用 AI

L12 生成式 AI 應用與規劃

L121 No code / Low code 概念

L12101 No Code / Low Code 的基本概念

一、No Code / Low Code 的定義

(一) 什麼是 No Code ?

- No Code (無程式碼) 是一種完全不需要寫程式的開發方式，透過視覺化介面（拖放組件、流程圖）來建立應用程式。
- 適用於非技術用戶（業務人員、產品經理），讓他們能夠自行開發工具，而不依賴工程師。

(二) 什麼是 Low Code ?

- Low Code (低程式碼) 是允許少量程式碼來進行高度自訂的開發方式，透過圖形化界面 + 最少的程式碼來快速開發應用程式。
- 適用於開發人員與 IT 團隊，幫助他們更快開發應用程式，同時保留一定程度的程式碼靈活性。

| 分類 | No-code | Low-code |
|-----|------------------------------|-------------------------------------|
| 定義 | 完全無需撰寫程式碼即可建立應用 | 只需少量程式碼即可擴充應用邏輯與彈性 |
| 對象 | 非技術人員（商務用戶、流程管理員） | 技術與半技術人員（分析師、開發者） |
| 功能 | 拖拉式流程、UI 視覺化、自動 API 串接 | 可插入程式碼模組、自定義 API、邏輯控制 |
| 擴展性 | 有限（依賴平台功能） | 高（可整合程式語言、API、資料庫） |
| 範例 | Bubble、Glide、Zapier、Airtable | OutSystems、Mendix、Power Apps、Retool |

二、No Code / Low Code 的核心技術

(一) No-code 技術核心

- 以 UI 元件組裝 + 預設流程引擎為主
- 提供視覺式條件判斷、流程連結與資料綁定（binding）
- 背後執行的程式邏輯由平台代碼自動生成

常見技術元件：

- 拖曳式工作流程設計器（Drag-and-Drop Builder）
- 表單與資料綁定（Form & Data Binding）
- 預建 API 模組（REST / Webhook）

(二) Low-code 技術核心

- 提供程式碼擴充點（Extension Points）與模組化開發結構

(一) 常見的生成式 AI 模型

| 模型類型 | 代表技術 | 應用範圍 |
|------------------------|---------------------------------|---------------------|
| 統計模型 | 馬可夫鏈（Markov Chains）、隱馬可夫模型（HMM） | 文字生成、語音合成 |
| 生成對抗網絡（GANs） | DCGAN（影像生成）、StyleGAN（風格轉換） | 人臉生成、深度偽造（Deepfake） |
| 變分自動編碼器（VAE） | Variational Autoencoder | 圖像合成、異常數據檢測 |
| 擴散模型（Diffusion Models） | Stable Diffusion、DALL-E | 影像生成 |
| 語言模型（LLMs） | GPT-4（文本生成）、BERT（文本理解） | AI 對話、新聞撰寫 |

(二) 生成式 AI 的應用

- 文本生成（Text Generation）：
 - ChatGPT 可用於自動寫作、新聞報導、AI 助理。
- 圖片生成（Image Generation）：
 - DALL-E、Midjourney 可根據文字描述生成圖像。
- 音樂與影片生成（Music & Video Generation）：
 - Suno、AIVA 生成 AI 作曲，Runway ML 生成 AI 動畫。

三、基本原理比較：

(一) 鑑別式 AI 的數學原理

- 條件機率表示法：
 - 鑑別式 AI 直接學習輸入數據 XX 與標籤 YY 之間的關係，計算條件機率 $P(Y|X)$ 。
 - 例如：判斷一封電子郵件是垃圾郵件（Spam）或正常郵件（Not Spam），鑑別式 AI 訓練一個分類器，計算：
 $P(\text{Spam}|X) \text{ vs. } P(\text{Not Spam}|X)$
 - 目標：找到最可能的類別 YY ，並做出分類決策。
 $Y^* = \arg \max_y P(Y|X)$

(二) 生成式 AI 的數學原理

- 條件機率表示法：
 - 生成式 AI 學習數據的整體分佈 $P(X)P(Y|X)$ ，然後生成新數據，使其符合原數據分佈。

例如：AI 學習大量貓的照片後，可以生成新的貓的圖像：
 $P(X) = P(\text{貓的圖像})P(Y|X) = P(\text{Text/貓的圖像})$

| 比較項目 | 鑑別式 AI (Discriminative AI) | 生成式 AI (Generative AI) |
|------|------------------------------|--------------------------|
| 核心概念 | 透過數據進行分類、預測、識別 | 透過數據學習模式，生成新內容 |

| | | |
|------|-------------------|------------------------------|
| 學習方式 | 學習 ($P(Y X)$) | (條件機率) |
| 應用領域 | 影像辨識、詐欺偵測、文本分類 | 內容創作、圖像生成、AI 對話 |
| 代表技術 | SVM、隨機森林、CNN、BERT | GANs、VAE、Transformer (GPT-4) |
| 應用範例 | AI 判斷一張圖是「貓」還是「狗」 | AI 生成一張新的「貓」圖片 |

四、特性與應用差異分析

| 特性比較 | 鑑別式 AI | 生成式 AI |
|---------|---------------------|---------------------------|
| 訓練資料需求 | 需要標記資料 (Supervised) | 可使用非標記資料或少量標記資料 |
| 訓練穩定性 | 相對穩定 | 易不收斂或產生模式崩潰 (特別是 GAN) |
| 可解釋性 | 高 (可追蹤特徵與分類結果) | 較低 (如 GPT 的 token 機率難以解釋) |
| 應用範圍 | 分類、回歸、推薦系統、異常偵測 | 文本生成、圖像合成、資料擴增、AI 創作 |
| 運算與資源需求 | 較低 | 較高 (特別是大型生成模型如 GPT-4) |

五、鑑別式 AI 核心技術詳解

5.1 Logistic Regression (邏輯斯迴歸)

- 使用 sigmoid 函數將線性組合映射至 0~1 機率。
- 適用於二元分類，為現代深度網路 softmax 層原型。
- 損失函數：交叉熵損失 (Cross-Entropy Loss)

5.2 支援向量機 (SVM)

- 目標為最大化分類邊界 (margin)。
- 可使用 kernel trick 解決非線性分類問題。
- 適用於資料量較小但分類準確度要求高的情境。

5.3 卷積神經網路 (CNN)

- 模擬視覺皮層，擅長圖像辨識。
- 由卷積層 (特徵偵測)、池化層 (降維)、全連接層組成。
- 常用於影像分類 (ImageNet)、醫學影像檢測。

5.4 循環神經網路與 LSTM

- 適合處理序列資料 (文字、語音、股價序列)。
- LSTM 引入門控單元 (gate) 避免梯度消失。
- 常用於語意分類、情緒分析等 NLP 鑑別任務。

5.5 Transformer 分類模型 (如 BERT)

- 預訓練語言模型 (Masked Language Model) 微調用於分類。
- 結構由 self-attention · position embedding 組成。
- 常用於問答系統、命名實體辨識、分類任務。

- 可用於品牌風格一致性管理

◎ 案例 4：AI 深偽偵測與安全驗證系統

◎ 目標任務：

- 面對 deepfake 音訊/圖像，需進行「生成+鑑別」對抗架構

◎ 技術架構：

- 生成式 AI (如 GAN)：模擬可能攻擊樣本 (如人臉合成、語音偽造)
- 鑑別式 AI (如 CNN+LSTM / 時頻分類模型)：訓練辨識深偽內容
- 進階應用：對抗樣本訓練 (adversarial training) 強化辨識力

◎ 效益：

- 防堵偽造資訊擴散
- 建立 AI 診斷對抗系統 (AI for Cybersecurity)

五、整合應用優勢與挑戰分析

✓ 優勢：

| 項目 | 說明 |
|---------|--------------------|
| 精度提升 | 鑑別式 AI 可強化生成結果品質控制 |
| 任務鏈結更完整 | 支援從資料生成 → 分類預測的全流程 |
| 彈性高 | 可依應用需求決定先生成還是先分類 |

▲ 技術挑戰：

| 挑戰項目 | 說明 |
|----------|-------------------------------|
| 模型相容性問題 | 不同模型需調整輸入輸出格式與流程協同 |
| 資源與效能需求高 | 同時運行生成與鑑別模型，需高效部署設計 |
| 可解釋性困難 | 生成模型決策路徑不易解釋，需輔助機制如 LIME、SHAP |

六、整合技術常見工具與框架

| 工具 / 平台 | 用途說明 |
|-------------------------|-------------------------------------|
| Hugging Face | BERT、GPT、T5 整合 API 與 Transformers 庫 |
| LangChain / LlamaIndex | Prompt 控制、分類+生成任務流程建構 |
| TensorFlow / PyTorch | 可組合訓練多模型架構 |
| ONNX + Triton Inference | 優化生成與分類模型部署 |

四、整合應用產業案例分享

◎ 案例 1：智慧客服系統中的回應生成與分類管理

◎ 目標任務：

- 即時生成自然語言回覆
- 並分類使用者意圖與情緒，以進行分流或風險管控

技術架構：

1. 生成式 AI (如 GPT)：生成語言回覆
2. 鑑別式 AI (如 BERT fine-tune)：辨識語意類型（抱怨、詢問、交易、緊急）
3. 整合機制：根據鑑別結果引導生成式模型調整語氣、風格（如更溫和或更專業）

效益：

- 提高對話自然性與應對效率
- 強化客訴預警與風險辨識能力

◎ 案例 2：醫療影像 AI 助診系統

◎ 目標任務：

- 自動檢測病灶 → 醫師輔助診斷 → 可視化報告生成

技術架構：

1. 鑑別式 AI (如 CNN)：識別 X 光影像中是否出現異常（肺炎、腫瘤等）
2. 生成式 AI (如 image captioning + GPT)：生成病灶報告說明文字
3. 報告樣式控制：加入 prompt template 與專業術語微調語風

效益：

- 減輕醫師負擔
- 增強初診或偏鄉自動化檢測能力

◎ 案例 3：自動文案撰寫與風格校對平台

◎ 目標任務：

- 使用者輸入關鍵字，自動生成促銷文案或社群貼文
- 並針對文風、語氣、廣告適用性進行風格分類與優化

技術架構：

1. 生成式 AI (如 GPT / T5)：初步生成文字
2. 鑑別式 AI (如 文體分類模型)：判斷文風是否符合品牌調性
3. 再輸入生成模型：微調 prompt 重新生成符合風格的新版本

效益：

- 節省創作時間，提高廣告轉換率

六、生成式 AI 核心技術詳解

6.1 生成對抗網路 (GAN)

- 由 **生成器 (G) 與鑑別器 (D)**組成。
- 生成器目標：生成逼真資料讓 D 無法辨識。
- 鑑別器目標：辨識資料是真實或生成。
- 訓練技巧：minimax 對抗訓練、梯度懲罰、Wasserstein loss。
- 應用：人臉生成、風格轉換、圖像修復。

6.2 自編碼器與變分自編碼器 (AE/VAE)

- AE：透過 encoder-decoder 架構壓縮與重建資料。
- VAE：以機率分佈建模潛在空間，能生成新樣本。
- 運用 KL divergence 控制潛在向量分布。
- 應用：降維、異常偵測、樣本生成。

6.3 自回歸語言模型 (如 GPT)

- Transformer-based：自左至右逐 token 預測。
- 使用 Causal Mask 防止資訊洩漏。
- 預測下一個 token 的機率分佈。
- GPT-2、GPT-3、GPT-4 系列支撐自然語言生成應用。

6.4 擴散模型 (Diffusion Models)

- 訓練時逐步向資料加入雜訊 (forward process)。
- 生成時則從雜訊中逐步還原資料 (reverse process)。
- 使用 U-Net 結構與時間編碼 (timestep embedding)。
- 應用：高畫質圖像生成 (如 Stable Diffusion、DALL-E)

6.5 編碼-解碼架構 (Encoder-Decoder)

- Encoder：理解輸入資料 → 生成潛在向量。
- Decoder：根據潛在表示生成輸出資料。
- T5、BART 是此架構應用於 NLP 任務的代表。

自然語言處理

七、鑑別式 AI 與生成式 AI 的互補關係

雖然鑑別式 AI 和生成式 AI 有不同的應用場景，但它們可以結合使用，形成更強大的 AI 系統。

例如：

1. GANs (生成對抗網絡·Generative Adversarial Networks)：
 - 生成器 (Generator)：生成新圖像 (生成式 AI)。
 - 鑑別器 (Discriminator)：判斷圖像是真是假 (鑑別式 AI)。
 - 應用：
 - AI 生成高品質的藝術作品。
 - AI 偵測假新聞、假影像 (Deepfake 偵測)。
2. AI 強化學習 + 生成式 AI：

ChatGPT 結合 RLHF (強化學習)，讓 AI 生成內容更符合人類期望。

L11402 鑑別式 AI 與生成式 AI 的整合應用

一、為何需要整合鑑別式 AI 與生成式 AI？

(一) 鑑別式 AI 的限制

- 只能做分類與預測，但無法生成新數據。
- 需要大量標註數據 (Labeled Data)，但標註數據昂貴且難以獲取。
- 在某些應用中，決策邊界可能過於僵硬，無法適應多變環境。

(二) 生成式 AI 的限制

- 無法進行精確分類與預測，只能生成新數據。
- 生成內容可能不夠精確，需要額外的模型來鑑別內容的真實性。**
- 可能會產生無法解釋或不可信的結果 (如 AI 生成假新聞)。

(三) 整合的優勢

| 整合方式 | 提升的能力 |
|---------------------|--|
| 生成式 AI 提供數據給鑑別式 AI | 透過生成式 AI 產生合成數據，讓鑑別式 AI 進行訓練，提升準確度。 |
| 鑑別式 AI 改善生成式 AI 的品質 | 透過鑑別式 AI 過濾生成式 AI 的輸出，確保內容真實性與品質。 |
| 兩者互相對抗學習 (GANs) | 透過對抗學習 (Adversarial Learning)，生成更真實的數據，並提升分類準確度。 |

二、鑑別式 AI 與生成式 AI 的整合技術

(一) 生成對抗網路 (GAN, Generative Adversarial Networks)

- GAN 是最典型的整合應用，由兩個 AI 模型組成：
 - 生成器 (Generator)：負責生成新數據。
 - 鑑別器 (Discriminator)：負責鑑別生成的數據是否真實。
- 應用範例：
 - AI 生成高畫質人臉 (如 StyleGAN)。
 - AI 生成虛擬時尚模特兒，並由鑑別式 AI 進行審核。

(二) 自監督學習 (Self-Supervised Learning, SSL)

- 原理：
 - 生成式 AI 產生部分數據 (如掩蔽部分圖像或文本)。
 - 鑑別式 AI 負責預測掩蔽部分，提高 AI 的理解能力。
- 應用範例：
 - BERT (Transformer 模型)：透過自監督學習，讓 AI 理解文本，並能在 NLP 任務 (如翻譯、文本分類) 中表現更好。

(三) 生成式 AI 幫助資料增強 (Data Augmentation)

• 原理：

- 生成式 AI 透過學習數據模式來產生合成數據，並提供給鑑別式 AI 訓練。

• 應用範例：

- 醫療 AI 訓練數據增強：
 - 生成式 AI 產生不同風格的 X-ray 影像，增加 AI 訓練數據集的多樣性。
 - 鑑別式 AI 透過這些合成影像來提升診斷準確率。

(四) 鑑別式 AI 提升生成式 AI 的可信度

• 原理：

- 生成式 AI 可能生成不可信的內容，鑑別式 AI 可過濾低品質或有害的輸出。

• 應用範例：

- AI 內容過濾：
 - ChatGPT 生成文章後，BERT 模型可對其進行審核，確保資訊準確。
 - AI 影像辨識模型可過濾 DALL-E 生成的圖像，避免違規內容。

三、鑑別式 AI 與生成式 AI 的整合應用場景

3-1 醫療影像分析

| 應用場景 | 生成式 AI 功能 | 鑑別式 AI 功能 |
|----------|-----------------------------|------------------|
| 醫療影像資料增強 | 生成不同類型的 X-ray、MRI 影像，補充數據不足 | 訓練 AI 進行更準確的疾病診斷 |
| AI 幫助診斷 | 生成可能的病變區域，幫助醫生標註 | 分析影像，判斷病變是否異常 |

3-2 金融與詐欺偵測

| 應用場景 | 生成式 AI 功能 | 鑑別式 AI 功能 |
|--------|---------------------|-----------------|
| 詐欺交易檢測 | 生成正常交易數據，訓練模型識別異常行為 | 鑑別是否為詐欺交易，提高準確率 |
| 金融市場模擬 | 生成市場變動數據，幫助風險評估 | 預測市場走勢，提升投資決策 |

3-3 自動駕駛

| 應用場景 | 生成式 AI 功能 | 鑑別式 AI 功能 |
|----------|------------------------|---------------------|
| 虛擬駕駛數據生成 | 生成不同天氣、路況的模擬數據，增強自駕車訓練 | 學習不同環境下的駕駛決策，提高安全性 |
| 行人與障礙物識別 | 生成各種車輛與行人圖像，提高 AI 學習範圍 | 確保 AI 能夠正確區分人車與道路物件 |

3-4 數位內容創作

| 應用場景 | 生成式 AI 功能 | 鑑別式 AI 功能 |
|---------|---------------------|-----------------|
| AI 影片製作 | 生成動畫、短片，如 Runway ML | 評估影片品質，確保內容無誤 |
| 新聞與文章撰寫 | ChatGPT 生成新聞報導 | AI 審核新聞內容，避免假訊息 |