



慧穩科技股份有限公司
AIWin Technology Co., Ltd

聯絡資訊

04-2252-5580

service@aiwin.com.tw

04-2252-5570

407 台中市西屯區市政路500號7樓之6

www.aiwin.com.tw (官網)

winhub.ai/zh-Hant (產品網站)

Hub.AI



WinHub. AI 融合式 AI 解決方案

一站式AI落地整合方案



慧穩科技股份有限公司
AIWin Technology Co., Ltd



關於慧穩

以創新 AI 驅動落地應用，成就全球智慧解決方案領導品牌



慧穩科技成立於 2016 年，為專注於 AI 落地應用的解決方案提供者，我們已成功交付超過 150 個 AI 專案，深入半導體、智慧製造、智慧水務等 30 多種產業，累積了豐富的實戰經驗與軟硬體整合能力，並致力於協助客戶建立可持續運作的 AI 正循環。

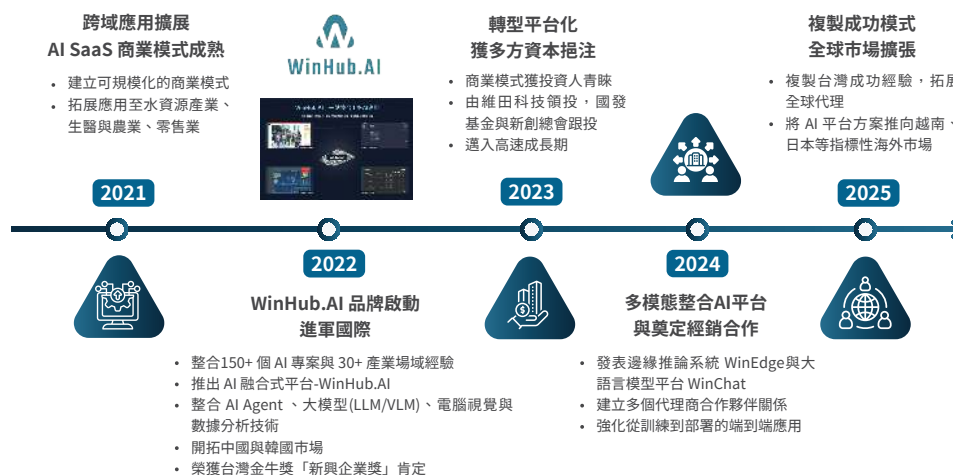
我們推出的 WinHub.AI 融合式 AI 平台，整合 AI Agent、大模型（LLM/VLM）、電腦視覺、數據分析（AutoML）與專家系統，為企業提供一站式的 AI 開發與部署解決方案，實現開箱即落地，並建立可持續的 AI 正循環。

我們以成為全球 AI 解決方案領導品牌為願景，致力提供全面的 AI 導入服務，解決客戶實際痛點。憑藉多年實戰經驗與成功案例，慧穩科技提供彈性的專案式與訂閱式服務模式，協助企業降低 AI 導入門檻，穩健銜接智慧升級。

慧穩科技，您值得信賴的智慧夥伴 —— Be your iPartner.



公司沿革



全球 AI 解決方案的信賴之選
與您共創智慧未來

目錄

服務產業與合作經銷夥伴

p.2

WinHub.AI 融合式平台

p.3

WinHub.AI 產品全系列

p.4

AIWinOps | 專利級AI影像

p.5-p.10

NumOps | 智慧數據 AI

p.11-p.16

WinEdge | 邊緣 AI 推論

p.17-p.22

WinChat | 企業知識中樞

p.23-p.28

服務產業



半導體與高科技業



水資源與環境產業



精密與傳產製造業



消費性與零售服務業



影像與數據監控



生醫與農業

合作與經銷夥伴



基士德科技股份有限公司
GSD Technologies Co., Ltd.



WinHub.AI 融合式平台

整合五大 AI 核心，引領產業智慧革新

生成式與鑑別式 AI Generative & Discriminative AI

整合生成式與鑑別式 AI，提供企業智慧生成與精準判斷能力

- 文字生成與彙整
- KM 知識管理
- 智慧客服助理

自動化機器學習 AutoML

自動建模與最佳模型選擇，加速 AI 導入流程

- 訂單預測
- 設備異常預測
- 庫存預測



專利級機器視覺 Machine Vision

透過專利級 AI 影像與演算法進行自動辨識與分析

- 瑕疵檢測與 AOI 複判
- 影像分類與物件辨識
- 工安與場域安全監控

數據 AI 與專家系統 Data AI & Expert System

整合專家經驗與規則邏輯，快速建立 AI 決策系統

- 設備異常與壽命預測
- 製程參數與節能最佳化
- 老師傅經驗數位化

邊緣運算與 AI Agent Edge Computing & AI Agent

於資料產生端即時分析，結合 AI Agent 自動執行任務

- 對話式操作，不需寫程式
- 跨系統任務自動執行
- 自動分析建議與報表生成

產品全系列

一站整合・彈性賦能

AIWinOps | 專利級 AI 影像

- 五大影像分析工具
- 無需程式碼模型建立
- AI 自動標註減人力
- 多場域穩定應用



NumOps | 智慧數據 AI

- 異常偵測與預測
- 即時數據監控
- 設備狀態分析
- 支援智慧水務
- 多場域穩定應用



WinEdge | 邊緣 AI 推論

- 即時邊緣推論
- 快速現場部署
- 降低系統延遲
- 彈性設備整合
- 穩定長期運行



WinChat | 企業知識中樞

- LLM+RAG 核心
- 整合文件與流程
- 即時查詢可追溯
- 降低依賴加速決策
- 私有部署 AI Agent



專利級AI影像應用 | AIWinOps

■ 跨產業應用，影像 AI 快速落地每個產業場域



交通控管

人工查看時間成本高、無法精確紀錄



解決方案

AI 影像辨識+多目標追蹤

- 多 ROI 區域統計與方向判定
- 交通量統計與列車通過自動偵測
- 支援企業內部私有部署



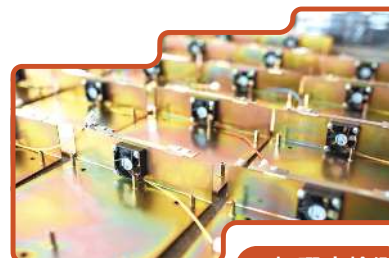
成車樣式檢查

樣式檢查複雜、依賴特定人員經驗

解決方案

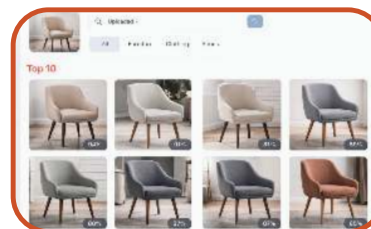
AI 模型訓練+ WinEdge整合

- 無程式碼模型建立
- 推論後處理過濾 (ROI、信心值、面積)
- 運用AI 達到標準化檢驗流程



面板瑕疵檢測

人工查看時間成本高、無法精確紀錄



款式比對與搜尋

相似商品比對困難、品質控管不一致



織帶生產監控

人工瑕疵檢測不完整、缺乏品質紀錄

解決方案

深度學習瑕疵辨識

- 裂痕刮傷自動辨識
- 智慧標註減少人力
- 檢出率達95~98%

解決方案

AI 圖像檢索(以圖搜圖)

- AI 特徵向量比對技術
- 即時回傳 Top-K 相似結果
- 不需重新訓練模型
- 亦可客製化模型



解決方案

AIWinOps+WinEdge 整合

- 更完善的瑕疵檢測和帶寬量測
- 三色燈與API 串接管理系統整合
- 完整品質紀錄

AIWinOps

適用各產業的影像型 AI 訓練平台

五大影像分析核心，滿足多樣化 AI 需求



影像分類
Image Classification

- 大量圖片
- 自動分類



物件偵測
Object Detection

- 瑕疵檢測
- 特徵辨識



影像分割
Image Segmentation

- 精準量測
- 邊界偵測



異常偵測
Anomaly Detection

- 正樣本訓練
- 自動辨識應用



瑕疵生成
Defect Generation

- 生成式AI瑕疵生成
- 極少數樣本學習

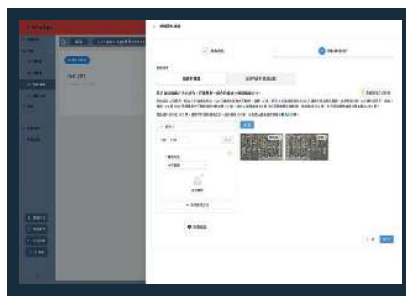
強大的進階功能，加速您的 AI 開發流程

資料擴增 (Data Augmentation)

圖資蒐集不易、數量較少時，可透過資料擴增功能，將原始資料以不同角度旋轉或翻轉方式增加資料量。

自動標註 (Auto Labeling)

圖資數量龐大時，可先用少量圖資訓練模型進行自動標註，經人員複檢調整後再訓練下一版模型。如此循環可在短時間內以倍數速度完成大量標註工作。



智慧標註 (Smart Labeling)

內建視覺語言模型 (VLM)，只需在圖片上手動標註一次，系統即可自動偵測同一圖片上的類似物件並完成標註。

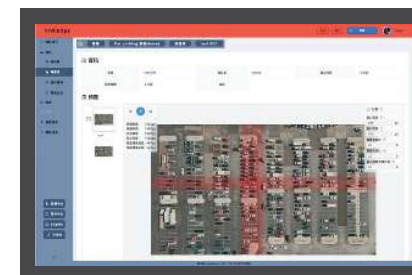
自動分割 (Auto Segmentation)

此為另一VLM應用，僅需框選一個圖例，即可自動精準辨識目標物件的影像輪廓，並完成分割與建立標記。



資料前處理 (Image Pre-processing)

針對高解析度圖片，透過 ROI 裁切與分切小圖功能，減少資料量但不損失關鍵細節，節省運算資源並強化準確率。



模型庫 (Model Zoo)

慧穩科技經長期測試驗證，精選多種影像式AI演算法與模型，上架於 AIWinOps 供各產業依需求選用。同時提供進階功能，開放高階 AI 開發團隊上架自訂的演算法與模型，作為訓練與驗證之用。



專利級AI影像 | AIWinOps

從零到一，輕鬆打造專屬 AI 模型

1. 新增專案

一鍵建立專案，多人協作與資料獨立互不干擾。

01



2. 建立資料集與標註

匯入圖資並進行標註，支援 VOC、COCO、YOLO 格式直接匯入。

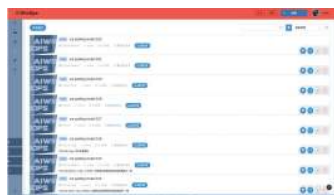
02



3. 訓練模型

選擇資料集與參數設定，啟動模型訓練，多專案自動化排程。

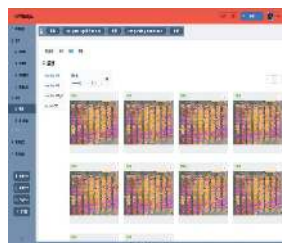
03



4. 模型驗證與決策

檢視 loss/accuracy 圖表，設定推論門檻與過濾條件。

04



5. 模型部署落地

下載模型部署至邊緣運算設備，實現 AI 應用落地。

05



系統要求

硬體需求

項目	最低規格	建議規格
CPU	12 cores Intel® Core™ i7-12700 Processor	20 cores / 40 threads or higher 2 × Intel® Xeon® Silver 4210 Processors
GPU	NVIDIA GPU VRAM 12GB CUDA Computing capability 7.5	NVIDIA GPU VRAM 16GB or more CUDA Computing capability 7.5 or higher
記憶體	64GB	128GB or higher
儲存裝置	512GB SSD + 1TB HDD	1 TB SSD + 4TB HDD or higher
連接埠	USB 2.0 or 3.0 (required for security dongle)	
推薦相容硬體	APLEX AVS-553、APLEX AVS-P710、NVIDIA GB10 Chip	



APLEX AVS-553



APLEX AVS-P710



NVIDIA GB10

軟體需求

項目	版本
作業系統	Ubuntu Focal Fossa 22.04 LTS (64-bit) 或更高版本
NVIDIA GPU 驅動程式	535.154.05 或更高版本
Docker Engine - Community	20.10.17 或更高版本

採購資訊

AIWinOps

落地永久授權系統方案

- 永久使用授權
- 完整 GPU 伺服器配置
- 私有雲落地部署
- 資料完全自主掌控
- 終身技術支援

適合對象：
長期使用、資料敏感企業

AIWinOps

落地年授權系統方案

- 年度授權使用
- 完整 GPU 伺服器配置
- 私有雲落地部署
- 定期系統更新
- 年度技術支援

適合對象：
長彈性預算、階段性需求

AIWinOps

落地年授權模型方案

- 年度授權使用
- 完整 GPU 伺服器配置
- 私有雲落地部署
- 搭配SDK 與 API (依電腦數量授權)

適合對象：
長彈性預算、階段性需求

智慧數據 AI 應用 | NumOps

■ 數據 AI 應用案例，快速優化生產場域



製程異常偵測

人工定義異常門檻僵化、調整成本高



解決方案

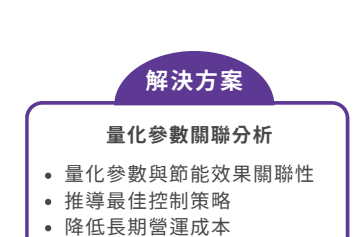
數據分析整合

- 僅需正樣本即可建模
- 連續行為模式異常判斷
- 異常檢出率達 97.5%
- 判斷速度提升 60 倍
(人工: 30分鐘 → 導入 AI: 30秒)



冰機節能控制

冰機控制策略效率不足、高營運成本



解決方案

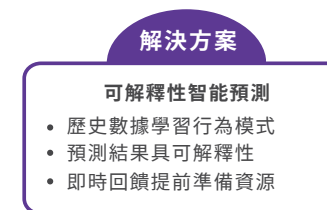
量化參數關聯分析

- 量化參數與節能效果關聯性
- 推導最佳控制策略
- 降低長期營運成本



銷售預測

預測受人主觀判斷影響、缺乏量化依據



解決方案

可解釋性智能預測

- 歷史數據學習行為模式
- 預測結果具可解釋性
- 即時回饋提前準備資源



製程烤箱溫控

控制依賴人工經驗、人才培訓成本高



解決方案

智能控制策略

- 具回饋機制自動控制，全自動化執行
- 低試錯次數策略搜尋
- 控制結果穩定在範圍內
- 優於資深人員操作表現

NumOps

適用各產業的數據型 AI 訓練平台

六大數據分析核心，滿足多元化 AI 需求



專家系統 Expert System

- 無需數據訓練
- 規則庫推論



分類 Classification

- 類別自動判斷
- 風險評估應用



回歸 Regression

- 連續數值預測
- 因果關係建模



時間序列 Time Series

- 時間趨勢預測
- 週期性分析



異常偵測 Anomaly Detection

- 異常發生偵測
- 詐騙犯罪預警



分群 Cluster

- 無需定義自動分組
- 顧客市場區隔

雙軌並進：專家系統與數據 AI 的產出與落地

專家系統 – 規則庫的產出與落地

NumOps 專家系統模仿人類專家的知識經驗與決策方式。無需過去累積的數據資料，僅需定義輸入輸出特徵、設定單位及條件範圍，即可產出專家系統規則庫，依據輸入條件推論輸出結果，成為解決問題的實用工具。

數據 AI 模型的產出與落地

針對分類、回歸、時間序列及異常偵測等需求，遵照以下程序執行，即可產出符合自家需求的專屬數據型 AI 模型，實現應用落地目標：

數據 AI 模型的落地步驟

1. 資料上傳與分析

資料支援 .csv 或 .xls 格式，系統自動檢查空值及格式問題並發出示警。上傳後，自動分析敘述統計量、相關性，可自建統計圖表或執行異常值分析 (PCA、Isolation Forest 等)，於訓練前充分掌握資料特性。

2. 預處理

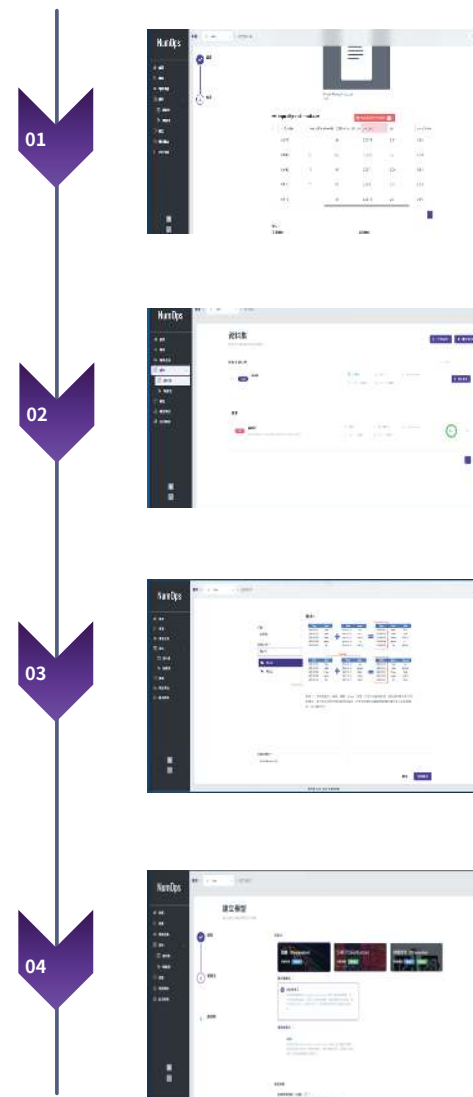
提供簡單易用的預處理功能介面，可輕鬆修正或豐富原始資料，確保模型訓練成效。

3. 合併版本

提供合併多個資料集進行全面分析，提供更深刻洞見，幫助決策制定。

4. 模型訓練

選擇演算法、設定參數、挑選資料集並設定預測目標與輸入特徵，即可啟動訓練。基於 AutoML 自動採用多種演算法產出多個模型。



模型自動化建模與落地

模型分析

訓練完成後，自動完成模型排行(基於 5 折交叉驗證)顯示關鍵性能指標 (MSE、R² RMSE、準確度等)及特徵重要性分析。提供多種圖表幫助評估模型是否符合預期：



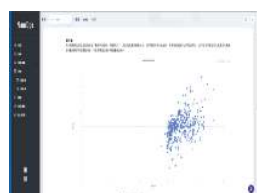
(資訊)



(模型排行)



(特徵重要性)



(殘差圖)



(測試集分析)



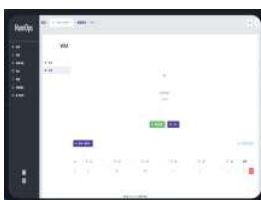
(混淆矩陣)

模型部署與落地

確認 AI 模型符合實際需求後，遵循模型授權程序，即可下載模型並部署至 WinEdge。WinEdge 提供儀表板呈現、部署控制與推論資料蒐集，實現 AI 應用落地，達成提升效率、減少錯誤、降低人力需求及產業升級等目標。

批次預測

模型訓練完成後，可進入批次預測頁面進行模型驗證。釋出模型並選擇新資料集進行測試，若資料集包含實際結果可即時比對。



完整的資料預處理功能，提升模型準確度

缺失值處理

可透過平均數、中位數、眾數等策略對空值做補值處理，或刪除空值

產生新特徵

透過公式運算、使用者經驗制定的規則或條件判斷，快速產生新欄位

標籤編碼

將類別型資料轉換為數值型資料

規則式過濾

透過使用者設計的條件判斷句進行資料刪除

欄位操作

將欄位類型轉換成數值型、類別型、時序型或重新命名欄位

剔除異常值

運用 PCA 或 Isolation Forest 方法進行異常值偵測與剔除

硬體需求

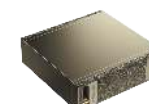
項目	最低規格	建議規格
CPU	Intel 9th Gen. Core i5/i7 Processor	Intel 13/14th Gen. Core i5/i7 Processor
CPU架構	AMD64/x86	AMD64/x86
記憶體	16GB	128GB
儲存裝置	HDD 1TB	SSD 1TB + HDD 2TB
USB 接口	1個 USB 2.0 或 3.0 端口，用於連接加密鎖	
推薦相容硬體	APLEX AVS-530、NVIDIA GB10 Chip	

軟體需求

項目	最低規格
作業系統	Ubuntu 20.04或更高的版本
Docker Engine - Community	20.10.17或更高的版本



APLEX AVS-530



NVIDIA GB10

採購資訊

NumOps

落地永久授權系統方案

- 永久使用授權
- 私有雲落地部署
- 資料完全自主掌控
- 終身技術支援

適合對象：
長期使用、資料敏感企業

NumOps

落地年授權模型方案

- 年度授權使用
- 私有雲落地部署
- 定期系統更新
- 年度技術支援

適合對象：
長彈性預算、階段性需求

NumOps

落地永久授權模型方案

- 永久使用授權
- 私有雲落地部署
- 搭配SDK與API (依電腦數量授權)

適合對象：
長彈性預算、階段性需求

邊緣 AI 推論應用 | WinEdge

■ 跨產業應用，影像 AI 快速落地每個產業場域



工地安全監控

危險行為難即時發現、安控依賴人力



解決方案

邊緣推論管理與集中部署

- Edge架構簡易部署
- 支援模型一鍵更新與版本管理
- 多裝置即時監控與管理
- 儀表板集中顯示推論結果



玻璃瑕疵檢測

人工目檢不穩定，易漏檢微小缺陷

解決方案

影像型 AI 邊緣推論部署

- 支援多相機取像與AI模型部署
- 本地端即時推論，無須仰賴雲端
- 推論後處理與結果輸出整合
- NG 影像儲存與訊號回傳



幼兒園安全監控

師生比例有限，無法全面監看

解決方案

物件追蹤與區域計數推論

- 支援提示詞與框選設定追蹤目標
- ROI 區域設定與進出計數
- 推論後處理與事件結果輸出
- 支援多種取像設備整合



製程異常偵測

異常情況難提前預警，停機影響產能

解決方案

即時數據整合與邊緣監控

- 串接資料庫與工業通訊協定
- 即時儀表板顯示關鍵數據
- 支援資料查詢與警戒設定
- 設備部署後可即時監控運行狀態



WinEdge

滿足各產業 AI 落地的邊緣推論系統

七大核心功能架構，驅動 AI 邊緣運算全面落地



AI 模型推論工具 AI Inference Engine

- 影像分類
- 物件偵測
- 影像分割



數據智能分析 Data Analytics

- 回歸 / 分類
- 時間序列
- 異常偵測



物件追蹤與計數 Object Tracking & Counting

- 提示詞 / 框選設定
- ROI 區域計數
- 路徑追蹤



邊緣部署與管理 Edge Deployment & Management

- 簡易模型部署架構
- 模型部署與更新
- 多裝置集中監控



即時儀表板與整合 Real-Time Dashboard & Integration

- 多元圖表顯示
- SQL 查詢與篩選
- 多資料源串接



推論後處理與輸出 Post-Processing & Output

- NMS 與自定義閾值設定
- 邏輯判斷與路徑追蹤
- 即時顯示 / 資料輸出 / I/O 觸發



系統整合擴充 Integration & Extensibility

- 多 I/O 串接能力
- 模型流程串接
- 開發者自訂擴充模組

強大的進階功能，加速您的AI部署流程

智慧儀表板設計 (Smart Dashboard)

內建彈性儀表板功能，可自動生成基本儀表板並支援自由拖曳排列。提供趨勢圖、圓餅圖、表格、文字等多種顯示模式，滿足不同情境的監控需求。



多元資料源整合 (Data Integration)

支援關聯型資料庫 (MySQL、MSSQL、PostgreSQL)、时序型資料庫 (InfluxDB)、工業通訊協定 (Modbus TCP)，以及各類影像設備的無縫整合。



邊緣運算架構 (Edge Computing)

採用邊緣運算設計，無須依賴雲端資源或網際網路即可穩定運行。

模型管理系統 (Model Management)

整合 NumOps 與 AIWinOps 平台，提供完整的 AI 模型生命週期管理。支援模型自動部署、版本更新、參數配置，以及 Recipe 管理功能。

設備運行監控 (Equipment Monitoring)

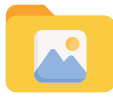
在串接設備部署完成後，可對其進行實時監控與管理，確保在實時數據顯示正常狀況下順利運行。提供 CPU、GPU、記憶體使用率等硬體資源監控，以及連線狀態、IP 位址、埠號等詳細資訊顯示。



多元的取像裝置支援

因應不同產業中各類應用場景的多樣性需求，WinEdge 支援以下取像設備：

- Basler Camera
- Line Scan Camera
- IP Camera (需支援RTSP通訊協定)
- Webcam
- File

	Basler Camera	Webcam	200萬畫素	資料夾
				
型號	支援全系列工業相機	依型號支援	依型號支援	Windows、Linux
解析度	依型號支援	依型號支援	1920 px × 1080 px	—
取像速度	依型號支援	依型號支援	—	—
介面與感光元件	GigE 依型號支援	GigE 依型號支援	GigE 1/2.8"	—

系統要求

軟體需求

項目	版本
作業系統	Ubuntu Focal Fossa 22.04 LTS (64-bit) 或更高版本
NVIDIA GPU 驅動程式	580.x 或更高的版本
Docker Engine - Community	24.x 或更高的版本

硬體需求

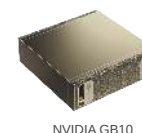
項目	建議規格
CPU	6核心 / 12執行緒以上
GPU(可選)	VRAM 8GB 以上 / 計算能力 7.5 或更高
記憶體	16 GB (最低) / 32 GB 以上 (建議)
儲存裝置	SSD 1 TB 以上

CPU與GPU配置推薦

輕量需求

適合教學與初期的簡單影像辨識任務，如分類和物件偵測。

- CPU：Intel Core i5-12400 或 AMD Ryzen 5 5600X
- GPU：NVIDIA GeForce RTX 4060 或 RTX 4070
- 推薦硬體：APLEX AVS-530或NVIDIA GB10 Chip



中階需求

適合於影像分割、深度學習模型的訓練與推論，如多物件偵測和影像分類。

- CPU：Intel Core i7-12700 或 AMD Ryzen 7 5700G
- GPU：NVIDIA GeForce RTX 4080 或 RTX 4090
- 推薦硬體：APLEX AVS-553、APLEX AVS-P710、NVIDIA GB10 Chip



高階需求

大型深度學習模型訓練，或高效能的影像辨識與推論項目。

- CPU：Intel Xeon 系列伺服器級處理器
- GPU：NVIDIA A100 或 H100 等級顯示卡
(需搭配高效能運算機架式伺服器提供穩定性)
- 推薦硬體：客製化處理

採購資訊

WinEdge

落地永久授權系統方案

- 永久使用授權
- 私有雲落地部署
- 資料完全自主掌控
- 終身技術支援

適合對象：
長期使用、資料敏感企業

企業知識中樞應用 | WinChat

■ AI 驅動知識管理，內外溝通暢行無阻



WinChat

適用各產業的大語言模型應用平台

六大核心能力，企業級 RAG 對話 AI 知識助理



深度文件理解能力 Document Intelligence

- 從非結構化資料提取關鍵資訊
- 支援多種文本格式解析



無需程式碼開發 No-Code Development

- 直觀的網頁式操作介面
- 整合資源、權限與專案管理



多知識庫架構 Multi Knowledge Base

- 建立組織專屬獨立知識庫
- 支援多種資料類型



減少幻覺 Hallucination Reduction

- 關鍵引用快照，內容可追溯
- 大幅降低 AI 幻覺風險



資訊系統串接 System Integration

- 提供易用的 Web API
- 支援彈性擴充與客製化整合



私有雲落地部署 Private Cloud Deployment

- 確保機敏資訊不外流
- 強化資料隱私與存取控管

從知識問答到智慧決策，AI 應用全面落地

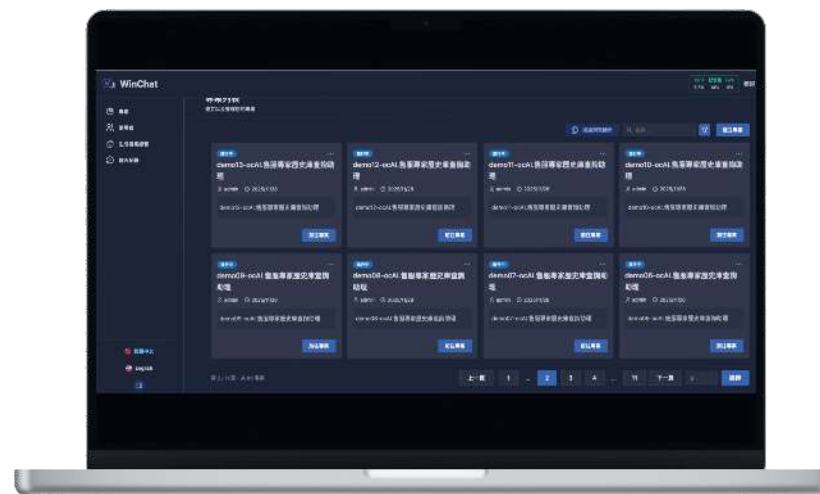
RAG 對話系統 | 企業知識應用落地

WinChat 結合 LLM 與 RAG 架構，將企業內部規範、流程與知識文件轉化為即時對話式知識服務。透過深度文件理解與來源引用機制，確保回應內容精準且可追溯，有效降低幻覺風險，強化知識應用與決策效率。

AI Agent 延伸 | 智慧助理與決策輔助落地

在 RAG 對話應用基礎上，WinChat 結合企業既有資訊系統與慧穩科技於影像及數據鑑別式 AI 的技術能力，發展為具備感知、回應與決策輔助能力的 AI Agent。透過私有雲部署與彈性整合架構，協助企業打造專屬、可控且可擴充的智慧應用平台。

賦予 AI 專屬知識，打造企業級智慧核心



專案管理 | 系統核心入口

此區塊為 WinChat 系統的核心入口，負責建立、管理與分類各專案。用戶可透過此介面快速進行專案切換與設定。

主要功能：

- 建立專案專屬知識庫
- 生成檔案與聊天記錄
- 依時間、知識庫、狀態篩選
- 修改或刪除既有專案

知識庫管理 | 企業知識中樞

支援多種文件格式，讓 AI 精準理解並回應企業知識。透過 RAG (Retrieval-Augmented Generation)，系統可將企業文件轉化為可檢索知識，提升 AI 回答的準確度與可信度。

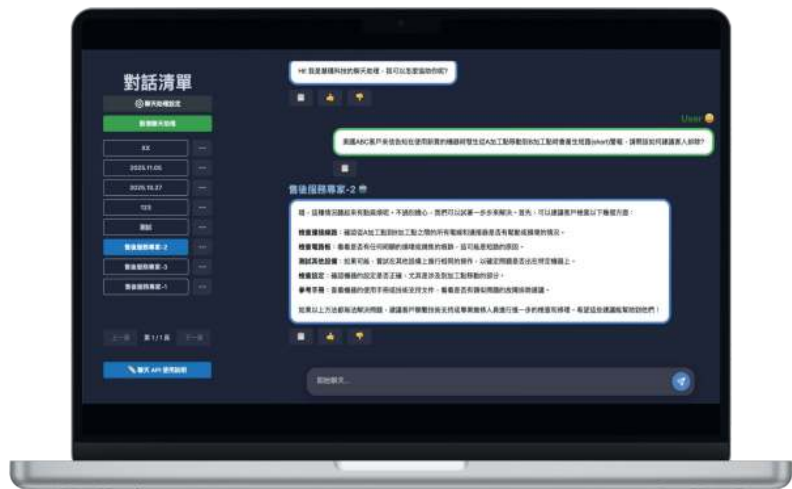
支援格式：

PDF | Word | Excel

主要功能：

- 上傳、建立、編輯與管理知識庫內容
- 支援知識分段、關鍵字與 Top-K 設定
- 即時檢視檢索結果與知識命中效果
- 持續優化 AI 回答品質，降低錯誤風險

專屬 AI 助理，提供精準可靠的即時問答



聊天助理 | 專屬 AI 助理建構

依需求預先建立專屬聊天助理，確保每一次回答都準確可靠。使用者可在對話前設定助理角色、選擇知識庫，並依應用情境打造專屬 AI 助理。

主要功能：

- 符合產業特性，確保回覆精準度
- 有效避免 AI 幻覺發生
- 讓 AI 回覆具備個性與專業度
- 展現適切的情緒表達

問答對話 | RAG 智慧互動介面

整合企業知識庫，讓 AI 成為即時回應的最佳助手。透過 RAG+LLM 架構，系統能即時檢索知識並生成精準回答，提升問答效率與可靠性。

主要功能：

- 顯示完整對話與問答紀錄
- 支援檔案上傳與文件查詢
- 快速生成即時回應，降低人工負擔

特色功能：

- 依專案分類管理
- 附上來源文件索引，提供可追溯的依據
- 依使用回饋持續優化回應品質

系統要求

硬體需求

項目	最低規格	建議規格
CPU	12核心或以上 3.5GHz或以上	20核心或以上 3.5GHz或以上
GPU	VRAM 24GB或以上 NVIDIA GeForce RTX 4090 (24GB)	VRAM 96GB或以上 NVIDIA RTX Pro 6000 Blackwell (96GB)
記憶體	128GB或以上	256GB或以上
儲存裝置	1TB SSD + 4TB HDD或以上	2TB SSD + 8TB HDD或以上
連接埠	USB 2.0或3.0 (需插入安全鎖)	
支援模型(LLM)	可支援 gpt-oss-20b 或同等參數量級模型推論	可支援 gpt-oss-120b 或同等參數量級模型推論
推薦相容硬體	APLEX AVS-553、APLEX AVS-P710、NVIDIA GB10 Chip	



APLEX AVS-553



APLEX AVS-P710



NVIDIA GB10

軟體需求

項目	版本
作業系統	Ubuntu 24.04 LTS (64-bit)或更高版本
NVIDIA GPU 驅動程式	535.154.05或更高版本
Docker Engine - Community	20.10.17 或更高版本

採購資訊

WinChat

落地永久授權
系統方案

- 永久使用授權
- 完整 GPU 伺服器配置
- 私有雲落地部署
- 資料完全自主掌控
- 終身技術支援

適合對象：
長期使用、資料敏感企業

WinChat

落地年授權
系統方案

- 年度授權使用
- 完整 GPU 伺服器配置
- 私有雲落地部署
- 定期系統更新
- 年度技術支援

適合對象：
長彈性預算、階段性需求