

Sentinel Monitor 哨兵專業版 — 使用手冊

一套桌面 AI 監控軟體：把監視器 (CCTV) 攝影機的影像交給 AI 即時分析，自動數人數、判斷畫面是否有異常或威脅，並把結果即時顯示在畫面右側的日誌欄。

本手冊以「一般使用者」為對象，會一步一步帶你從下載安裝到看到攝影機畫面。看不懂的專有名詞，會在當下用白話解釋。

目錄

1. 這個系統在做什麼？（先看懂再用）
2. 使用前需要準備什麼
3. 步驟一：下載與安裝 (Windows / Mac)
4. 步驟二：第一次開啟 — 輸入授權金鑰
5. 步驟三：認識主畫面
6. 步驟四：選擇 AI 分析引擎 (Inference Backend)
7. 步驟五：設定攝影機 RTSP (最重要的一步)
8. 步驟六：開始監控 (初始化系統)
9. 步驟七：看懂日誌、對攝影機下指令
10. 進階：在本機跑 vLLM (需要 NVIDIA 顯卡)
11. 進階：連接遠端 GPU 推論伺服器 (SSH 通道)
12. 常見問題與疑難排解
13. 名詞解釋 (白話版)

1. 這個系統在做什麼？（先看懂再用）

簡單講，整個流程是這樣：

攝影機(RTSP) ———> Sentinel Monitor ———> AI 模型分析 ———> 右側日誌即時顯示結果
(人數 / 是否異常 / 威脅評估)

- 你提供：每一支攝影機的「RTSP 網址」（影像來源，下面會教怎麼拿）。
- 軟體做的事：把影像抓下來，每隔一秒擷取一張畫面，送給 AI。
- AI 回傳：這個畫面有幾個人、發生什麼事、是否需要警戒（畫面會變紅）。

它最多可以同時監看 8 路攝影機（畫面排成 4 欄 × 2 列）。

你不需要懂 AI，也不需要寫程式。你主要做兩件事：① 告訴它攝影機在哪 (RTSP)、② 選一個 AI 引擎。

2. 使用前需要準備什麼

| 項目 | 必要性 | 說明 |
|---------------------|---------|---|
| 一台 Windows 或 Mac 電腦 | 必要 | 程式有 Windows 版與 Mac 版 |
| 授權金鑰 (License Key) | 必要 | 第一次開啟必須輸入，向供應商索取，格式類似 <code>XXXX-XXXX-XXXX-XXXX</code> |
| 第一次啟用時要有網路 | 必要 | 金鑰要連線驗證一次；之後可離線使用 |
| 攝影機的 RTSP 網址 | 必要 | 沒有它就看不到任何畫面 (第 7 節教你怎麼找) |
| AI 引擎 | 必要 (擇一) | 雲端的 Gemini (只要一把 API Key) 或本機/遠端 vLLM |
| NVIDIA 顯示卡 + Docker | 選用 | 只有你想在「自己電腦上」跑 vLLM 才需要 |

如果你只是想趕快用起來：準備好「授權金鑰」+「攝影機 RTSP」+「一把 Gemini API Key」，這三樣就夠了。

3. 步驟一：下載與安裝 (Windows / Mac)

軟體是透過 GitHub Releases 發佈的。每次發佈會自動產生三個平台的安裝檔。

在哪裡下載

到專案的 Releases 頁面，找最新版本，依你的系統下載對應檔案：

| 你的系統 | 下載這個檔 | 副檔名 |
|---------|---|------------------------|
| macOS | <code>Sentinel Monitor-x.x.x.dmg</code> | <code>.dmg</code> |
| Windows | <code>Sentinel Monitor Setup-x.x.x.exe</code> | <code>.exe</code> |
| Linux | <code>Sentinel Monitor-x.x.x.AppImage</code> | <code>.AppImage</code> |

macOS 安裝

1. 點兩下下載的 `.dmg` 檔。
2. 把 Sentinel Monitor 圖示拖進 Applications (應用程式) 資料夾。
3. 到「應用程式」裡點兩下開啟。
4. 第一次開啟若出現「無法驗證開發者 / 來自不明開發者」的警告：
 - 到 系統設定 → 隱私權與安全性，找到被擋下的 Sentinel Monitor，按「仍要打開」；
 - 或在「應用程式」中對著圖示按右鍵 → 打開。

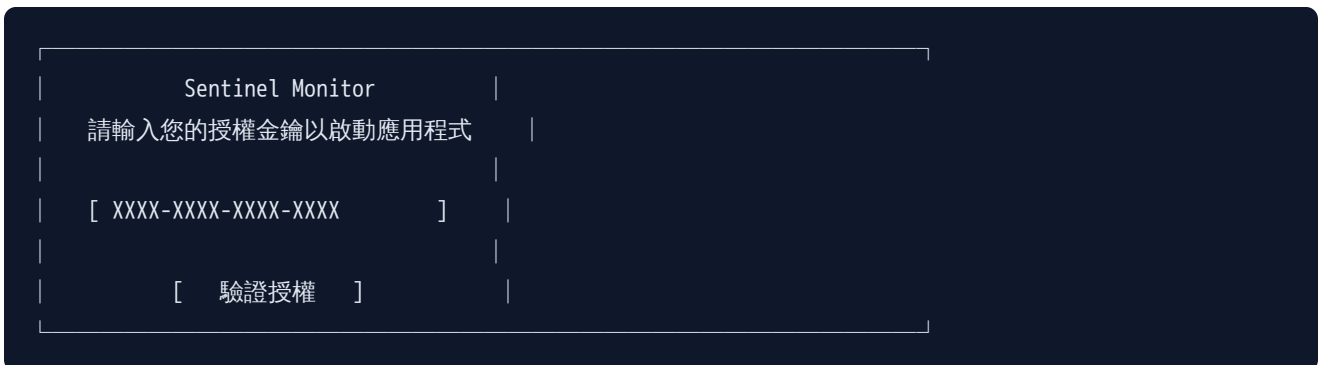
Windows 安裝

1. 點兩下 `Sentinel Monitor Setup-x.x.x.exe`。
2. 若出現「Windows 已保護您的電腦 (SmartScreen)」的藍色視窗，點「其他資訊」→「仍要執行」。
3. 依精靈指示完成安裝，桌面或開始功能表會出現 Sentinel Monitor 捷徑。

安裝完不需要再額外安裝 Python 或後端程式 —— AI 分析所需的後端服務已經內建在安裝檔裡，開啟 App 時會自動在背景啟動。

4. 步驟二：第一次開啟 — 輸入授權金鑰

第一次 (或金鑰失效時) 開啟，會先跳出一個小視窗「Sentinel Monitor — 授權驗證」：



1. 把供應商給你的授權金鑰貼進輸入框。
2. 按「驗證授權」，按鈕會顯示「驗證中...」。
3. 成功 → 顯示「驗證成功 ✓」，視窗自動關閉，進入主程式。

驗證時電腦要能上網，因為金鑰會送到授權伺服器 (Keygen.sh) 核對，並把這台電腦登記為授權裝置。

驗證後會發生的事 (你不用管，自動完成)：

- 金鑰存在本機，下次開啟不用再輸入。
- 之後即使沒有網路也能驗證 (用加密簽章離線檢查)。
- 程式每 24 小時會自動在背景重新驗證一次；若金鑰被停用或過期，會跳出提示並關閉程式。

驗證失敗時的訊息對照

| 畫面訊息 | 意思 | 怎麼辦 |
|-----------------|----------|----------|
| 授權已過期，請續訂後重試 | 金鑰到期 | 聯絡供應商續約 |
| 授權已被暫停，請聯繫客服 | 金鑰被停用 | 聯絡客服 |
| 此裝置未獲授權，請聯繫客服 | 這台電腦未被允許 | 聯絡客服綁定裝置 |
| 無效的授權金鑰 | 金鑰打錯或不存在 | 檢查有沒有打錯字 |
| 連線失敗，請確認網路連線後重試 | 沒網路 | 接上網路再試一次 |

5. 步驟三：認識主畫面

進入後，畫面分成三大區塊：



頂部列重點：

- 左上「哨兵專業版」：軟體名稱。
- 中間狀態燈：系統待命（灰） / 即時監控中（綠色閃爍）。
- READY / LOADING：右上角的小圓點代表「後端服務」狀態。
- READY = 後端已就緒，可以開始監控。
- LOADING = 後端還在啟動，請等它變綠（第一次啟動可能要等一兩分鐘）。
- vLLM 開關（只在 Windows / Linux 出現）：用來在本機開啟 vLLM，詳見第 10 節。

- 最右邊的大按鈕：
- **初始化系統** = 開始連線所有攝影機並啟動監控。
- 連上後會變成 **中斷連接**。

6. 步驟四：選擇 AI 分析引擎 (Inference Backend)

畫面中間有一塊 INFERENCE BACKEND 面板，這裡決定「畫面要交給哪個 AI 分析」。有兩個選擇：

選項 A：Gemini（雲端，最簡單，推薦新手）

- 用 Google 的雲端 AI，不需要顯卡、不需要架伺服器。
- 點「Gemini」按鈕，下面會出現一格輸入框：填入你的 Gemini API Key。
- 缺點：影像會送到雲端、需要網路、可能有用量費用。

選項 B：Local vLLM（本機 / 遠端，影像不外流）

- 影像交給「自己的」AI 伺服器處理，不送雲端。
- 點「Local vLLM」，會出現兩格：
- Base URL：vLLM 伺服器網址，預設 <http://127.0.0.1:8080/v1>（代表「本機 8080 埠」）。
- API key：選填，沒有就留空。
- 需要先把 vLLM 伺服器跑起來（見第 10、11 節）。

測試連線（強烈建議先按一次）

填好之後，按面板上的「Test Connection」：

- 成功會顯示綠色 **OK: ...**，代表 AI 引擎接得通。
- 失敗會顯示紅色 **FAILED: ...**，請回頭檢查網址 / 金鑰。

新手最快路線：選 Gemini → 貼上 API Key → 按 Test Connection 看到綠色 OK，就可以了。

7. 步驟五：設定攝影機 RTSP（最重要的一步）

這是整個系統的核心：你必須告訴軟體每支攝影機的「RTSP 網址」，它才有畫面可以分析。

什麼是 RTSP？

RTSP 是監視器產業常用的「即時影像串流網址」，長這樣：

```
rtsp://帳號:密碼@攝影機IP:埠號/路徑  
|       |       |       |  
└─ 串流路徑（每家廠牌不同）
```


沒有真的攝影機，想先試玩？你可以用測試影像產生器自建一條 RTSP 測試串流（用 `mediamtx + ffmpeg`，詳見專案的 `rtsp.md`），把網址填成 `rtsp://127.0.0.1:8554/test` 即可。

8. 步驟六：開始監控（初始化系統）

確認以下三件事都就緒：

1. 頂部右上角後端狀態是 `READY`。
2. AI 引擎已選好（Gemini 或 vLLM），最好按過 Test Connection 看到 OK。
3. 已上傳 Camera Config（攝影機網址）。

然後：

- 按頂部最右邊的白色大按鈕「初始化系統」。
- 軟體會同時去連每一支攝影機。連線中該格會顯示「連線中」，成功後狀態變綠。
- 中間狀態列變成「即時監控中」，底部顯示「N/8 路在線」。
- 右側日誌開始一筆一筆出現 AI 的分析結果。

畫面變紅 = 警戒：當某一路 AI 判定為「危急 / 威脅」狀態時，該區塊背景會泛紅光，代表需要你注意。

要停止：按同一個按鈕（此時顯示為「中斷連接」）即可全部停止。

9. 步驟七：看懂日誌、對攝影機下指令

右側日誌欄

每一筆是一次 AI 分析結果，包含時間、來自哪一路攝影機（用顏色標籤區分 CAM 1~8），以及 AI 的描述。日誌分顏色：

- 一般資訊（白 / 灰）：例如「正在建立監控工作…」 「系統進入待命狀態」。
- 警示（紅）：例如連線失敗、或畫面被判定為危急。

對某一路攝影機下指令（進階互動）

日誌欄下方有一個指令輸入框，可以即時「問」或「指示」某一路攝影機的 AI：

1. 在「指令目標」那排，點選你要對話的攝影機編號（CAM 1~8）。
 - 該攝影機必須是已連線狀態才能下指令；未連線時輸入框會顯示「CAM X 未連線」。
2. 在輸入框輸入你的指令，例如「特別注意門口有沒有人徘徊」「畫面中有幾個人？」。
3. 按 Enter 送出。AI 會把你的指令納入分析，並把回覆顯示在日誌。

這讓你可以針對特定鏡頭給 AI 「臨時的觀察重點」，而不影響其他鏡頭。

10. 進階：在本機跑 vLLM（需要 NVIDIA 顯卡）

這一節是選用的。如果你用 Gemini，或用第 11 節的遠端伺服器，可以跳過。

如果你想在「自己這台電腦」上跑開源視覺模型（影像完全不外流），需要：

- 一張 NVIDIA 顯示卡（含驅動）。
- 安裝 Docker Desktop（vLLM 是用 Docker 容器跑的）。
- 平台限制：此功能在 Windows / Linux 提供，macOS 不支援（Mac 沒有 NVIDIA GPU）。

開啟方式：

1. 先確定 Docker Desktop 已啟動。
2. 在頂部列點「vLLM」開關，讓它變綠（亮起 emerald 綠點）。
3. 滑鼠移上去會提示「重啟後生效」——請關閉並重新開啟 Sentinel Monitor。
4. 重新開啟後，程式會在背景用 Docker 啟動 vLLM 服務（第一次會下載模型，需要時間與硬碟空間）。
5. 在 INFERENCE BACKEND 選 Local vLLM，Base URL 用預設的 <http://127.0.0.1:8080/v1>，按 Test Connection 確認。

11. 進階：連接遠端 GPU 推論伺服器（SSH 通道）

這一節給有一台遠端 GPU 主機的進階使用者（例如實驗室 / 公司的伺服器）。一般使用者可略過。

系統可以把運算放在遠端伺服器上，本機只負責顯示。遠端會跑兩個服務：

| 服務 | 用途 | 埠 |
|-----------------|----------------------|------|
| vLLM (Qwen3-VL) | 視覺語言模型，負責「看懂畫面、描述場景」 | 8080 |
| YOLO11x | 物件偵測，負責「精準數人數」 | 8081 |

YOLO 數到的人數會被塞進給 vLLM 的提示中，讓場景判斷更準。

大致流程（完整指令見專案 [README.md](#)）：

1. SSH 登入伺服器，建立環境並安裝 `vllm`、`ultralytics`、`fastapi` 等套件。

2. 啟動 vLLM (埠 8080) 與 YOLO 伺服器 (埠 8081, 首次會自動下載 ~109MB 權重)。

3. 在你的本機用 SSH 通道把遠端的埠映射到本機：

```
bash ssh -L 8080:localhost:8080 -L 8081:localhost:8081 <使用者>@<伺服器IP> -p <SSH埠>
```

這個視窗要保持開著。它的作用是把遠端的 8080 / 8081 「假裝成」你本機的 8080 / 8081。

4. 回到 Sentinel Monitor, INFERENCE BACKEND 選 Local vLLM, Base URL 填 <http://127.0.0.1:8080/v1> (透過通道就連到遠端了), 按 Test Connection。

12. 常見問題與疑難排解

Q：右上角一直是紅色 LOADING, 不會變 READY?

後端正在啟動, 第一次可能要 1~3 分鐘。把滑鼠移到 READY/LOADING 上會顯示目前進度訊息。若長時間卡住或顯示「啟動失敗 / 逾時」, 請完全關閉再重開程式。

Q：按了「初始化系統」, 某一路顯示連線失敗 / 紅色警示?

多半是 RTSP 網址有問題。請：

- 先用 VLC 或 `ffplay` 測試該 RTSP 網址能不能播放；
- 檢查攝影機帳號密碼、IP、埠 (554)、串流路徑是否正確；
- 確認電腦與攝影機在同一個網路、防火牆沒擋。

Q：日誌都沒有 AI 分析結果出現?

代表 AI 引擎接不通。回到 INFERENCE BACKEND 按 Test Connection：

- Gemini → 檢查 API Key 是否正確、有沒有網路；
- vLLM → 檢查 Base URL, 本機 vLLM 是否已啟動、遠端的 SSH 通道是否還開著。

Q：上傳 Camera Config 失敗?

- 確認檔案是合法 JSON (逗號、引號、括號要完整)；
- 內容必須是「字串陣列」或含 `cameras` 欄位的物件；
- 數量不能超過 8 組；
- 每個網址要是有效、非空白的 RTSP URL。

Q：Mac 看不到 vLLM 開關?

正常。vLLM (需 NVIDIA GPU) 只在 Windows / Linux 提供。Mac 請改用 Gemini 或遠端伺服器。

Q：換了一台新電腦, 金鑰還能用嗎?

授權會綁定電腦。換機若提示「此裝置未獲授權」, 請聯絡客服協助綁定新裝置。

Q：要去哪裡看詳細的錯誤紀錄?

程式會把後端與啟動過程寫進記錄檔 (`sentinel-main.log`, 位於系統的應用程式記錄資料夾)。回報問題時附上它會很有幫助。

13. 名詞解釋 (白話版)

| 名詞 | 白話解釋 |
|--------------------|---|
| RTSP | 監視器的「即時影像網址」，軟體靠它拿到攝影機畫面。 |
| 授權金鑰 / License Key | 你買軟體後拿到的啟用密碼，格式像 <code>XXXX-XXXX-XXXX-XXXX</code> 。 |
| Inference / 推論 | 就是「讓 AI 看畫面、給出判斷」的動作。 |
| Inference Backend | 負責做 AI 判斷的引擎，可選雲端 Gemini 或本機/遠端 vLLM。 |
| Gemini | Google 的雲端 AI，只要一把 API Key 就能用，最省事。 |
| vLLM | 在自己（或遠端）伺服器上跑的開源 AI 服務，影像不外流，但要顯卡。 |
| YOLO | 專門用來「精準數人 / 抓物件」的 AI 模型，搭配 vLLM 提高準確度。 |
| 後端 / Backend | 內建在軟體裡、在背景默默幹活的服務；READY 就代表它準備好了。 |
| Docker | 一種把軟體連同環境打包好的「容器」工具，本機跑 vLLM 時會用到。 |
| SSH 通道 (Tunnel) | 一條安全的連線，把遠端伺服器的埠「接到」你本機，讓你像連本機一樣連遠端。 |

最短上手清單：下載安裝 → 輸入授權金鑰 → 等右上角變 `READY` → `INFERENCE BACKEND` 選 Gemini 並貼 API Key (按 Test Connection 看到 OK) → 上傳 Camera Config (你的攝影機 RTSP) → 按「初始化系統」→ 看右側日誌開始跑。完成！