

測試與評估表	
Ollama	
1. 工具概述	
姓名：	Ollama
類別：	AI
目的：	在個人電腦上本地運行大型語言模型，以獲得私密的離線 AI 輔助。
測試日期	2025年5月6日
文件翻譯日期：	2025年8月15日
地位：	<div>已部署<ul style="list-style-type: none"><li><input checked="" type="checkbox"/> 營運中 - 積極運作/維護</li><li><input type="checkbox"/> 測試中 - 目前正在評估或試行</li><li><input type="checkbox"/> 不活躍/棄用 - 不再維護或運行</li><li><input type="checkbox"/> 公民社會組織 - 經一個或多個公民社會組織驗證的採用情況</li></ul></div>
部署架構：	<div><ul style="list-style-type: none"><li><input checked="" type="checkbox"/> 獨立軟體 - 完全在本地運行(例如，在電腦上運行並且不依賴外部伺服器)</li><li><input checked="" type="checkbox"/> 具有獨立伺服器和用戶端元件的本機主機服務 - 自行執行後端/前端(例如，後端可以在本機網路上，也可以在雲端自行託管)</li><li><input type="checkbox"/> 由第三方託管的本機用戶端服務 - 您在裝置上安裝用戶端，但它連接並依賴遠端伺服器(例如，Signal:安裝應用程式(用戶端)，但 Signal 的伺服器處理訊息中繼等)</li><li><input type="checkbox"/> 由第三方託管但也可以自行託管的服務</li></ul></div>
版本：	v0.6.8
2. 安裝和設定	
作業系統相容性	macOS (英特爾和 Apple Silicon)、Windows、Linux (基於 Ubuntu/Debian)

安裝手冊：	Yes: <a href="https://github.com/ollama/ollama">https://github.com/ollama/ollama</a>
安裝步驟：	<ol style="list-style-type: none"> <li>1. 訪問<a href="https://ollama.com">https://ollama.com</a></li> <li>2. 根據作業系統下載安裝程序</li> <li>3. 安裝: 運行安裝程式並按照螢幕上的指示進行操作</li> <li>4. 模型設定: 使用命令列介面拉取所需模型, 例如 <code>ollama pull llama3</code></li> <li>5. 運行: 啟動 Ollama 服務並透過 CLI 或整合應用程式與模型交互</li> </ol>
提及是否需要命令列設定或特殊配置	一些舉例: <ul style="list-style-type: none"> <li>- CLI 指令:               <ul style="list-style-type: none"> <li>- <code>curl -fsSL https://ollama.com/install.sh   sh</code></li> </ul> </li> <li>- 執行:               <ul style="list-style-type: none"> <li>- <code>ollama run llama3</code></li> </ul> </li> </ul>
常見安裝問題及修復：	「GPU 不受支援」- 修復: 使用 ”僅 CPU” 設置 「模型無法啟動」- 檢查 RAM 可用性 (某些模型需要 8GB–16GB+) 防火牆阻止下載 - 從官方 repo 手動下載模型
使用者文件：	有 ( <a href="https://ollama.com/library">https://ollama.com/library</a> , <a href="https://github.com/ollama">https://github.com/ollama</a> )
所需的技術知識	初級到中級 <ul style="list-style-type: none"> <li>- 熟悉基本的命令列介面是有益的, 但不是強制性的。</li> </ul>

### 3. 測試與評估

類別	細節	分數
操作功能：	功能 <ul style="list-style-type: none"> <li><input type="checkbox"/> 該工具基本上無法使用, 存在許多損壞的功能和缺陷。</li> <li><input type="checkbox"/> 一些功能損壞或出現錯誤</li> <li><input type="checkbox"/> 小錯誤或問題</li> <li><input type="checkbox"/> 基本功能正常, 幾乎沒有錯誤或沒有錯誤</li> <li><input checked="" type="checkbox"/> 功能齊全, 無任何錯誤</li> <li>- 執行核心功能, 包括模型載入、提示互動和回應生成。</li> </ul> 網路依賴: <ul style="list-style-type: none"> <li>● 安裝後無需互聯網 (模型 100% 離線運行)</li> <li>● 低頻寬效能: 在 2G/3G 網路上測試; 設定後效能保持穩定。</li> </ul> 在地化和語言支持	3.3

	<ul style="list-style-type: none"> <li>● 取決於所使用的型號, 但大多數支援多種語言, 包括英語、中文、日語、西班牙語和韓語。</li> <li>● 社群貢獻: 社群積極參與在地化工作。</li> </ul> <p>移動無障礙</p> <ul style="list-style-type: none"> <li>● 行動應用程式: 沒有專用的行動應用程式可用。</li> <li>● 行動瀏覽器存取: 可透過行動瀏覽器存取; 但是, 效能可能因裝置功能而異。</li> </ul>	
非技術用戶的可用性	<p>易於安裝和部署</p> <ul style="list-style-type: none"> <li>● 安裝過程簡單, 提供清楚的說明</li> <li>● 基於 CLI, 但只需一行命令即可安裝; 設定時間約 10 分鐘</li> </ul> <p>使用者入門體驗</p> <ul style="list-style-type: none"> <li>● 沒有應用程式內指導; 使用者依賴外部文件。</li> </ul> <p>所需的技術經驗水平</p> <ul style="list-style-type: none"> <li>● CLI 中級, 但正在改進</li> </ul>	3.7
安全和隱私強度	<p>加密標準</p> <ul style="list-style-type: none"> <li>● 沒有提到具體的加密標準; 依賴系統級安全性。</li> <li>● 不太適用 – 離線運行, 不傳輸資料</li> </ul> <p>已知強度彈性</p> <ul style="list-style-type: none"> <li>● 抗審查: 離線運行, 使其能夠抵禦基於網路的審查。</li> <li>● 非常適合離線/受審查使用; 透過避免使用網路來繞過監視</li> </ul> <p>與已知標準的比較</p> <ul style="list-style-type: none"> <li>● 大型模型會消耗大量 RAM/CPU</li> <li>● 出色的本地隱私——在安全環境中比 ChatGPT 或 Gemini 更好</li> <li>● 透過最大限度地減少資料傳輸和儲存來符合最佳實踐</li> </ul> <p>資料最小化</p> <ul style="list-style-type: none"> <li>● 不收集或傳輸任何資料; 所有操作都在本地進行</li> </ul> <p>隱私權政策的可近性和清晰度</p> <ul style="list-style-type: none"> <li>● 沒有找到正式的隱私權政策; 然而, 該工具的本地特性本身就支援用戶隱私</li> </ul>	4.0
維護/永續性	<p>社群支持</p> <ul style="list-style-type: none"> <li>● 活躍的 GitHub 儲存庫和社群論壇提供協助和更新</li> </ul> <p>開發活躍狀態</p> <ul style="list-style-type: none"> <li>● 更新頻率: 定期更新, 最新版本 0.6.8 於 2025 年 5 月 3 日發布 (截至 2025 年 5 月 6 日)</li> </ul>	4.7

	<ul style="list-style-type: none"> <li>● 開發人員回應能力:積極參與社群回饋和問題解決。</li> </ul> <p>資金和贊助</p> <ul style="list-style-type: none"> <li>● 社群驅動</li> <li>● 由開源貢獻者和私人資助者支持</li> </ul>	
性能/有效性和可靠性	<p>測試環境設定:</p> <ul style="list-style-type: none"> <li>● 設備:Macbook Pro (14 英寸, M4 晶片), 10 核心 CPU, 24 GB RAM</li> <li>● 作業系統:15.2 Sequoia</li> <li>● 網路:Wifi</li> </ul> <p>使用者體驗觀察</p> <ul style="list-style-type: none"> <li>● LLaMA 2/3 流暢、快速</li> </ul> <p>速度和反應能力:</p> <ul style="list-style-type: none"> <li>● 速度(token 產生): <ul style="list-style-type: none"> <li>○ CPU(無 GPU 加速):~2000 ms/token</li> <li>○ GPU(消費級 GPU):~500 ms/token</li> </ul> </li> <li>● 推理時間:對於 10 個 token 的回應: <ul style="list-style-type: none"> <li>○ CPU:約 20 秒</li> <li>○ GPU:約 5 秒</li> </ul> </li> <li>● 高效率的模型載入和執行時間。</li> </ul> <p>資源使用:</p> <ul style="list-style-type: none"> <li>● 運行 LLaMA 3 8B 時佔用 10–12 GB</li> <li>● 中等使用;資源消耗根據模型大小和複雜性而變化。</li> </ul> <p>網路效能:</p> <ul style="list-style-type: none"> <li>● 初始模型下載後無需網路</li> </ul> <p>可靠性</p> <ul style="list-style-type: none"> <li>● 離線測試正常運作時間為 100%</li> </ul>	4.5
部署注意事項:	<p>開源與透明度:</p> <ul style="list-style-type: none"> <li>● 原始碼託管在 GitHub 上, 允許任何人檢查、審核或修改。</li> <li>● 核心元件和模型載入邏輯是公開維護的, 儘管透過 Ollama 提取的單一模型(如 Meta 的 LLaMA)可能有單獨的許可或限制。</li> </ul> <p>雲端部署與本地部署:</p> <ul style="list-style-type: none"> <li>● 完全在地化。無需雲端基礎設施</li> </ul> <p>依賴項:</p> <ul style="list-style-type: none"> <li>● 除二進位檔案外, 無需任何依賴。可選:Docker、Make 等用於開發建置。</li> </ul> <p>部署後維護</p> <ul style="list-style-type: none"> <li>● 維護: <ul style="list-style-type: none"> <li>○ 一旦安裝, 它就可以在本地運行, 無需任何外部依賴。</li> </ul> </li> </ul>	

	<ul style="list-style-type: none"> <li>○ 透過 GitHub 提取最新版本或透過其安裝腳本重新安裝來套用更新(例如, 新模型版本或錯誤修復)。</li> <li>○ 日誌是本地的;除非使用者明確建立 Web UI 或將 Ollama 整合到更大的系統中, 否則無需維護後端伺服器。</li> <li>● 監控: <ul style="list-style-type: none"> <li>○ 最低要求—主要關注可用的系統資源(RAM/GPU)和作業系統更新後的偶爾相容性檢查。</li> </ul> </li> <li>● 分岔: <ul style="list-style-type: none"> <li>○ 使用 GitHub 非常便捷。此專案支援透過拉取請求 (Pull Request) 進行社群貢獻。</li> <li>○ 模型和系統行為的配置透過 CLI 進行管理, 並且可以透過編輯設定檔或命令邏輯進行擴充。</li> </ul> </li> </ul> <p>合併/可持續性:</p> <ul style="list-style-type: none"> <li>● GitHub 儲存庫包括建置和貢獻說明。</li> <li>● 問題和討論顯示了開發人員的積極回應, 因此那些分叉專案的人可以獲得社群的幫助。</li> </ul>	
<b>4. 測試場景</b>		
● 場景 1	<ul style="list-style-type: none"> <li>● 用例:運行 LLaMA 3 8B, 使用外部整合來總結長 PDF <ul style="list-style-type: none"> <li>○ 結果:成功。透過 CPU 約 30 秒產生摘要(使用 ollama-python 測試)</li> <li>○ 注意:與 LangChain、ollama-python 或 LM Studio 的整合對於管道來說效果很好。</li> </ul> </li> </ul>	
● 場景 2	<ul style="list-style-type: none"> <li>● 使用案例:隔離環境測試—安裝後無法存取互聯網</li> <li>● 結果:100%成功</li> <li>● 注意:確認沒有外部呼叫;Ollama 完全離線運行, 非常適合審查環境。</li> </ul>	
<b>5. 見解和建議</b>		
主要發現	<p>優勢:</p> <ul style="list-style-type: none"> <li>● 完整的離線功能</li> <li>● 適合初學者的 CLI</li> <li>● 快速、本地的 LLM 表現</li> <li>● 隱私保護</li> <li>● 跨平台支援</li> <li>● 活躍且樂於助人的開源社群</li> </ul> <p>弱點:</p>	

	<ul style="list-style-type: none"> <li>● 預設無 GUI (僅限 CLI)</li> <li>● 尚無行動應用程式</li> <li>● 大型模型需要強大的硬體</li> <li>● 如果安裝失敗, 錯誤訊息會受到限制</li> </ul>
建議的改進	<ul style="list-style-type: none"> <li>- 官方 GUI 應用程式</li> <li>- 更好的應用程式內引導或使用指導</li> <li>- Windows 安裝程式可以改善模型拉取錯誤處理</li> </ul>
替代工具:	<ul style="list-style-type: none"> <li>- LM Studio: 使用 Ollama 後端為本機 LLM 提供 GUI</li> <li>- GPT4All: 同樣是本地模型, 但略顯不夠完善</li> <li>- LocalAI: 完全開源, 但設定更複雜</li> <li>- 開放 WebUI: 整合 Ollama 的本地 LLM 前端</li> </ul>
授權	<ul style="list-style-type: none"> <li>- MIT 授權 (Ollama 是開源的: GitHub repo)</li> </ul>
成本/資源影響	<p>總成本:</p> <ul style="list-style-type: none"> <li>● 免費使用, 無高級方案</li> <li>● 硬體成本: 8B 型號需要 8-16 GB RAM</li> <li>● 維護成本: 最低 – 透過 CLI 處理更新</li> <li>● 隱性成本: 無; 完全透明的開源工具</li> </ul>
為什麼這對威權環境中的公民社會有用?	<p>Ollama 為在專制政權下運作的公民社會組織提供了獨特的優勢, 原因如下:</p> <ul style="list-style-type: none"> <li>- 完全離線功能: 安裝後, 所有 AI 推理均在本地進行, 無需網路連線。即使在信號燈昏暗或監控密集的環境中也能確保操作的穩定性。</li> <li>- 抗審查: 該工具不依賴 DNS、API 或外部雲端服務, 因此不受 IP 封鎖或 DPI (深度包檢測) 等常見審查方法的影響。</li> <li>- 透過設計實現隱私和匿名: 不會傳輸、遠端儲存或收集任何資料, 這使得它比需要持久雲端連接的集中式 AI 平台 (例如 ChatGPT、Bard) 更安全。</li> <li>- 開源且可修改: 公民團體可以檢查、定製或自行託管 Ollama 的元件, 以滿足其區域需求或與其他隱私保護工具整合。</li> <li>- 實地工作的適應力: 可以安裝在偏遠地區或災後復原地區使用的筆記型電腦上, 而無需依賴西方基礎設施。</li> <li>- 增強本地能力建設: 活動家和開發人員可以在本地微調或擴展翻譯、法律分析或媒體事實核查等任務的模型, 而無需訪問外國雲端工具。</li> </ul> <p>這使得 Ollama 成為數位主權的關鍵基礎設施工具, 尤其具有以下價值:</p> <ul style="list-style-type: none"> <li>- 記者躲避監視</li> </ul>

	<ul style="list-style-type: none"><li>- 線下工作的法律團隊</li><li>- 在受審查地區進行實地考察的非政府組織</li><li>- 人權監督員和檢舉人</li></ul>
--	--