

讓 AI 真正聽懂台灣: Twinkle AI 深耕在地語料, 一週年推出全新 T1 模型

在這個演算法主導資訊的時代, 語言不僅是溝通工具, 更是生活的核心。新興 AI 社群 Twinkle AI 近日慶祝成立一週年, 回顧這一年在推動台灣在地化大型語言模型(LLM)上的成果。從早期的「Llama Taiwan」到如今的規模, Twinkle AI 始終秉持一個簡單卻堅定的願景: 如果未來的數位工具無法理解台灣在地的語言細節, 我們熟悉的說話方式與文化脈絡, 終將在通用的演算法中慢慢淡化。

從語料基礎出發: 用實戰經驗串起產學合作

Twinkle AI 的起點源於一個共同的觀察: 主流模型多經由大量外來資料訓練, 往往難以準確回應台灣在地的需求。為了打造一個真正「聽得懂我們」、對答自然順手且貼切的模型, 團隊從一開始就投入大量心力在蒐集與過濾台灣在地語料。

正是因為具備了這些蒐集與清洗資料的實戰經驗, Twinkle AI 得以在 2025 年 3 月與台灣大哥大展開深度合作, 共同投入「Fineweb-edu-zhtw 計畫」, 為高品質的教育語料建置奠定基礎。這項合作並非偶然, 而是建立在社群對在地資料處理的深厚理解之上。

Formosa-1 與 T1: 紮實基礎上的技術展現

在更為紮實的資料集經驗支持下, 團隊於同年 4 月與 APMIC 合作並獲國網中心技術支援, 推出了 3B 級別的在地模型 Formosa-1。這款模型不只是一個技術展示, 後續更被奧義智慧(CyCraft)整合, 應用於 XecGuard 商業資安產品中, 證明了在地模型的實用價值。

榮獲 IT Matters 開源貢獻獎: 看見由下而上的社群力量

Twinkle AI 這一年來的努力, 不僅展現在模型效能上, 更建立了一套完整的開源協作模式。這樣的投入在 2025 年底開花結果——榮獲 IT Matters 開源社群貢獻獎。這份獎項證明了, 即便不是大型科技巨頭, 由社群自發集結的力量, 也能在 AI 技術領域中發揮關鍵影響力。這份榮耀屬於每一位參與語料校正、測試與開發的夥伴, 是大家一起讓在地的聲音被聽見。

T1 模型問世: 邁向下一階段的里程碑

帶著這份獎項給予的信心, 社群在 2026 年 1 月的一週年聚會中, 正式公開了集大成之作——T1 模型。

T1 模型選用了 Google 最新的 Gemma 3 4B 架構為基底。受惠於這一年來累積的資料處理技術, T1 注入了更強化的台灣在地資料庫, 不僅反應速度更迅速, 對答也更具備「台灣味」。這也代表我們離當初那個「想讓 AI 聽懂台灣話」的梦想, 又更近了一步。