

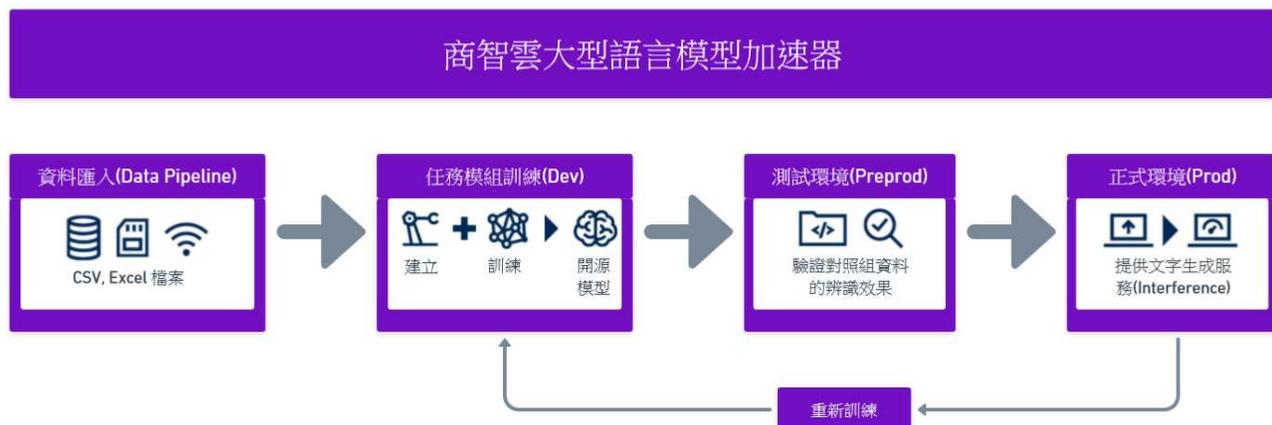
# 地端大型語言模型運作平台



提供地端大型語言模型以及內嵌模型之呼叫功能。

支援 google/gemma-2-9b-it, MediaTek-Research/Breeze-7B-Instruct-v1\_0, taide/TAIDE-LX-7B-Chat 等支援中文之大型語言模型。

具備有訓練與載入地端大型語言模型 LoRA Adapter (專屬任務模組)之能力。



大型語言模型在這兩年內取得各項突破性的進展，本公司將持續支援開源且支援中文的地端大型語言模型，

並且提供網頁介面訓練 LoRA Adapter 之能力，使機關內可以逐漸累積有的資料集，並且儲存封閉式地端部署之生成式 AI 模型。減少上傳雲端造成之資安疑慮。

支援之作業系統:

Windows 2022 Server. Ubuntu 22.04, Ubuntu 24.04

客戶硬體環境需求:

1. 24 cores (每個核心必須至少 2.6 GHz 或更快)
2. Memory: 32 GB 或以上之記憶體
3. NVIDIA 繪圖卡, 具備 24GB 或以上之 VRam 記憶體

商智雲股份有限公司

eMail: [info@ibcloud.com.tw](mailto:info@ibcloud.com.tw)

電話: 02-2700-9460

地址: 臺北市大安區敦化南路 2 段 164 號 11 樓之 1

