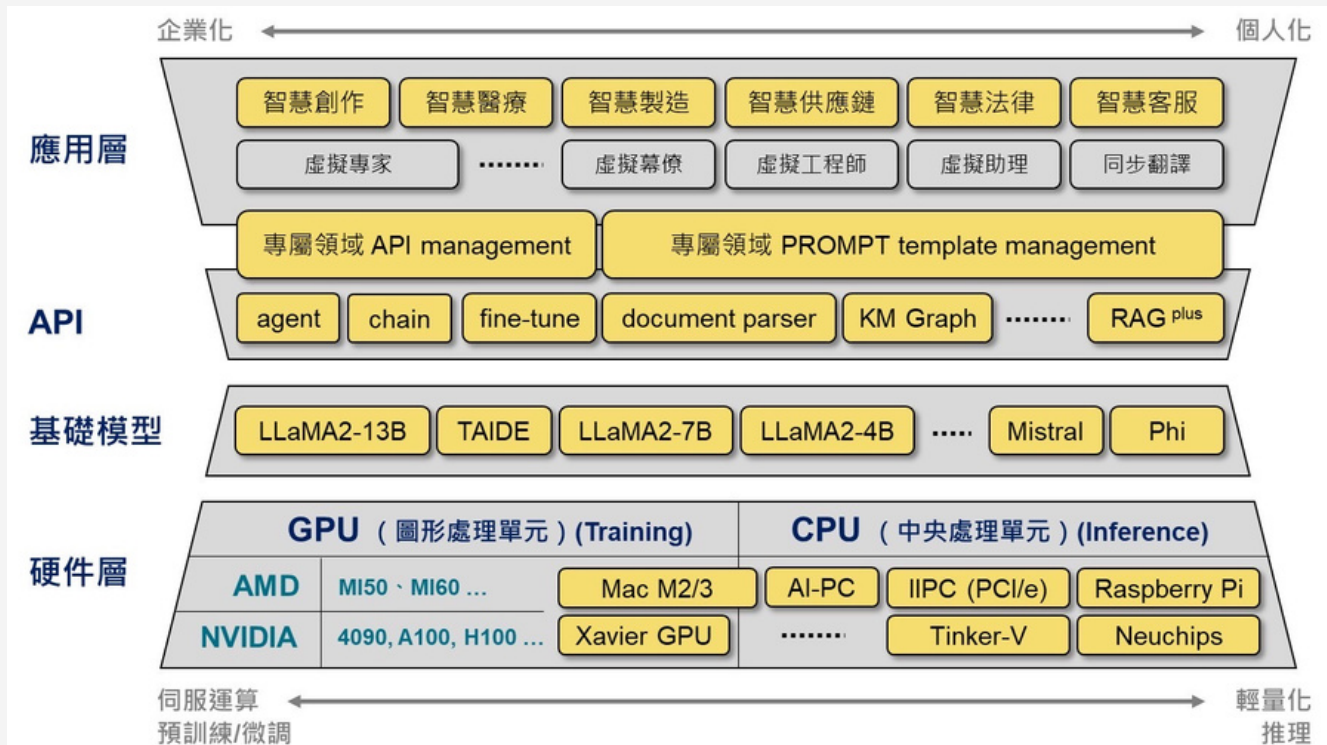


## 邊緣運算賦能生成AI：探索未來的技術架構



邊緣運算與生成AI模型的結合，已成為推動智能技術向前邁進的重要力量。從數據中心到終端用戶，AI的應用正在不斷拓展。今天，我們來解析一種大型神經網絡技術架構，並探討其如何充分利用邊緣運算能力，以支持生成AI的發展。

多樣化的硬件基礎 從高性能的AMD和NVIDIA GPU到適合邊緣運算的RASPERRY PI等輕量級設備，這種多元化的硬件支持確保了不同計算需求和應用場景下的靈活性與適應性。對於模型訓練的重任，AMD的MI系列或NVIDIA的A100可以提供強大支持；而對於需要在本地快速進行推斷的應用，如智慧家居或自動駕駛車輛，RASPERRY PI等裝置則顯得更為合適。

基礎模型層的創新 基礎模型層集中了多個預訓練的大型神經網絡模型，如LLAMA2-13B等。這些模型經過廣泛的訓練，具備了強大的學習和適應能力，能夠在各種應用中進行微調（FINE-TUNE）以達到特定的性能要求。

靈活的**API**層 **API**層作為模型與應用之間的橋樑，提供了一系列工具和接口，如文件解析器、知識管理圖和連鎖反應代理

在實際應用中，生成式**AI**正在改變語音識別、自然語言處理、影像生成和行為預測等多個領域。透過邊緣運算，這些應用可以直接在用戶的設備上運行，降低對雲端的依賴，提升響應速度和用戶體驗。

隨著邊緣運算與生成式**AI**技術的結合，**LLM**進一步加強了這一結合，使企業能夠在離數據源更近的地方進行更複雜的數據分析和內容生成。這不僅意味著速度更快、效率更高，還意味著更高水平的個性化和安全性。企業現在可以利用這些集成的系統，進行服務自動化、個性化內容創建等，為企業提供了一個競爭優勢，提供更精準、更有針對性的服務。