並非所有模型都同時支援聊天（chat）與嵌入（embedding）的 API 功能，這主要取決於模型本身的設計與實現目標。以下提供一些判斷與查詢方法：

1. **查閱模型文件或說明**
    - 大部分模型在發布時會附帶文件或說明，明確標示其支援的功能。例如，某些模型可能標榜「專為對話設計」或「優化用於生成嵌入向量」，這些說明可以幫助使用者判斷是否支持 chat 或 embedding。
2. **查看模型的配置或元數據**
    - 在 Ollama 或其他管理工具中，模型可能會有一個配置文件或元數據描述其功能。這裡通常會列出該模型是否實現了對應的 API endpoint（例如 `/v1/chat` 或 `/v1/embeddings`）。
    - 透過查閱這些配置文件或元數據，可以快速知道模型是否包含對應功能。
3. **測試 API 呼叫**
    - 另一種方法是直接通過 API 嘗試發送相應的請求，若模型支援則會返回預期結果；若不支援，則可能返回錯誤信息或提示不支持該功能。
    - 例如，可以先對某模型進行 chat API 的呼叫測試，若返回錯誤或提示找不到對應 endpoint，則可能表示該模型不支持聊天功能。同理，嵌入功能也可以用類似的方法進行驗證。
4. **參考 Ollama 官方文檔和公告**
    - Ollama 的官方文檔往往會詳細說明各個 API 的使用方法以及支持的模型列表或模型特性。通過官方文檔，可以獲得最新的、準確的模型功能支持信息。
5. **社群與論壇反饋**
    - 有時候，其他使用者或開發者在社群、論壇或 GitHub issues 中也會分享哪些模型支持哪些功能，這也是一個了解模型能力的途徑。

總結來說，是否支援 chat 或 embedding 主要取決於模型本身的實現，使用者可以通過閱讀模型相關文件、查閱配置與元數據、實際測試 API 呼叫，以及參考官方文檔來確定具體模型支援哪些 API 服務。