# Introduction to DIfy.ai 國立清華大學 區國良

Dify 是一個開源工具，可使用雲端版及自行下載安裝，後者需以 docker 方式安裝，雲端版則有容量及查詢次數的限制(且伺服器據說在對岸，使用時請自行評估資料機敏性)，本說明手冊以雲端版為例，以 Google 或 Github 帳號直接登入即可使用。請點選 Get Started

<https://dify.ai/>

一張含有 文字, 螢幕擷取畫面, 電子藍, 字型 的圖片

AI 產生的內容可能不正確。

登入後，上方頁簽提供「探索」、「工作室」、「知識庫」及「工具」選項，說明如下：

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

AI 產生的內容可能不正確。

## 探索

提供內建的樣板，提供快速布建，再依自己的需求修改，例如以下為 Question Classifier + Knowledge + Chatbot 的布建及修改過程，首先點擊「探索」頁簽後，再點擊Question Classifier + Knowledge + Chatbot 下方的「新增到工作區」，接著可自訂應用名稱及圖示，若不需修改，直接點擊「建立」即可，不同的樣板可能會需要個別安裝第三方的外掛，部分外掛還需要另外註冊，請自行判斷是否需要。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

AI 產生的內容可能不正確。 一張含有 文字, 螢幕擷取畫面, 字型, 軟體 的圖片

AI 產生的內容可能不正確。

免費版最多可製作或安裝5個應用程序，若達到上限時，在新增應用程式時則出現如右上的警告，如果想在免費額度內繼續新增應用程式，則請回到工作室中自行刪除不必要的項目即可。本範例預設使用 OpenAI 提供之LLM，其它的LLM可利用右上角的外掛自行安裝，安裝完成後，右上角的外掛按鈕可見其支援的模型型號，以下為若想使用 Anthropic LLM為例，需另外申請 [Claude](https://claude.ai/) （Anthropic 的聊天機器人）的帳戶，即可利用其 [API](https://console.anthropic.com/dashboard) 串接使用。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

AI 產生的內容可能不正確。一張含有 文字, 螢幕擷取畫面, 軟體, 網頁 的圖片

AI 產生的內容可能不正確。

完成外掛及應用程序的安裝後，將自動於「工作室」中展開，有些程序會貼心附上說明便利貼，使用者可自行依個人需求修改。以下圖為例，整個流程開始於最左方的 Start，點擊右上方的「預覽」可以試用看看。

一張含有 螢幕擷取畫面, 文字, 軟體 的圖片

AI 產生的內容可能不正確。

也可以展開「工作流」的內容，觀看處理的情況是否順暢，以及花費多少時間、多少 token 量等資訊，如下圖：

## 工作室

工作室即流程控制的主要核心，原則上大部分模版應該是直接可用，但仍有少部分參數及流程需要稍微調整後才能使用。以 Question Classifier + Knowledge + Chatbot 模版為例，其最左方為 START 節點，為所有對話流程開啟時第一個會執行的節點，每一次對話都會由START開始執行。本模版中，若順利執行後，下一個流程為 QESTIONCLASSIFIER，利用 GPT 模型判別使用者輸入的問題是屬於哪一種類型，分類1為 Question related難after sales，分類2為 Question about how to use products，若非以上分類，則屬於分類3，其中分類1及分類2的問題分別引導到不同的 KNOWLEDGE RETRIEVAL節點，以及不同的LLM節點，最後以 ANSWER節點回答使用者，而分類3則直接回答” Sorry, I can't help you with these questions.”並且結束本回合的問答，每次使用者提問時都會由 START 開始新一回合的流程。

有時流程超過一個螢幕的空間時，可利用右下角的放大縮小鍵，以及該鍵的上方的地點圖示移動顯示區域。

一張含有 文字, 螢幕擷取畫面, Rectangle, 設計 的圖片

AI 產生的內容可能不正確。

若欲修改流程，可在每個節點的後方以滑鼠點擊添加或拖曳的方式修改流程線，當鼠標移到＋號之上時，系統會提示可增加的節點或工具：

一張含有 文字, 螢幕擷取畫面, 軟體, 數字 的圖片

AI 產生的內容可能不正確。 一張含有 文字, 螢幕擷取畫面, 軟體, 電腦圖示 的圖片

AI 產生的內容可能不正確。

若卻刪除節點間的連結線時，請點選該連線線後，按鍵盤的 Del 處理。

## 知識庫

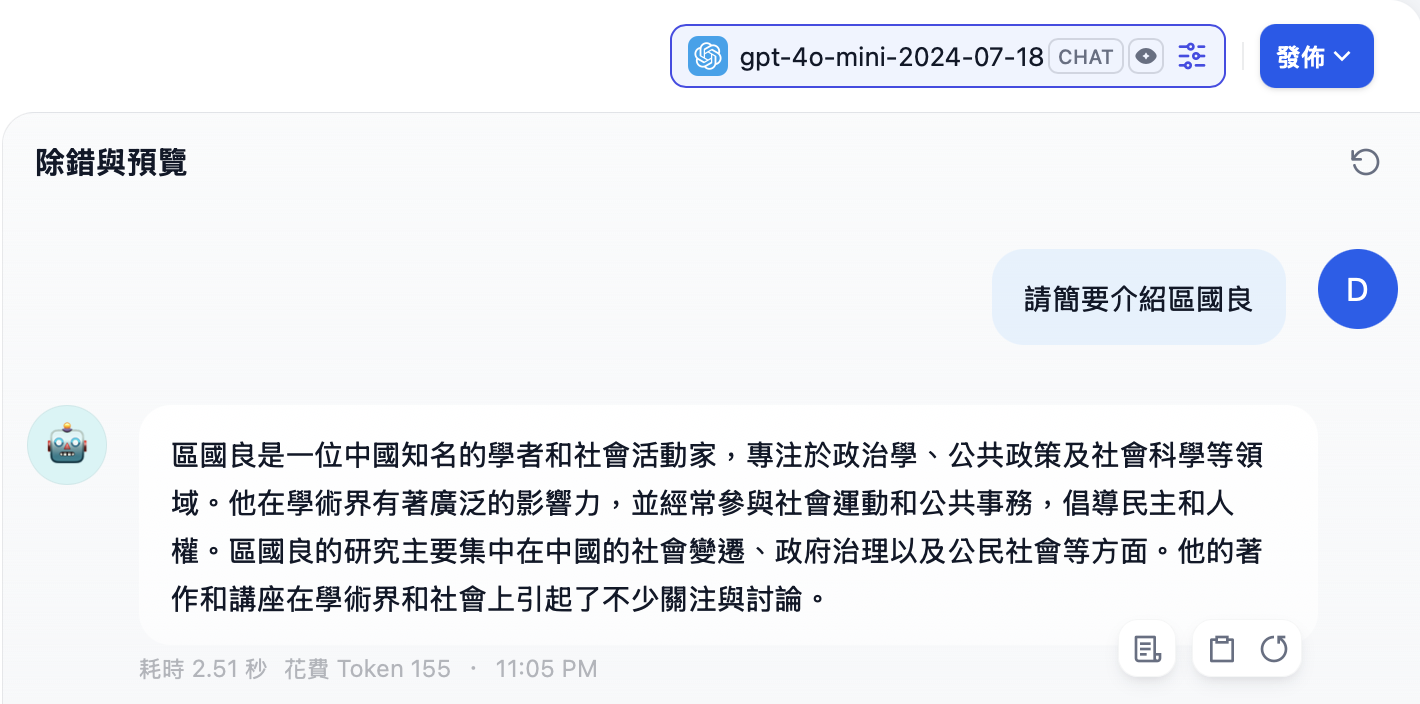
知識庫RAG工具，可預先上傳文件作為知識庫，在工作流中可設定成為LLM查詢回覆時使用，為比較使用原生LLM對話與使用知識庫對話的差異，我們先到工作室中建立一個單純聊天的機器人，首先在工作室中的左方「建立應用」選單中，點選「建立空白應用」，Dify提供五種空白應用環境，上方為適合初學者的聊天助手、Agent以及文字生成應用，下方為高級使用者適用之聊天流及工作流，請先點選聊天助手，並自行命名應用名稱：



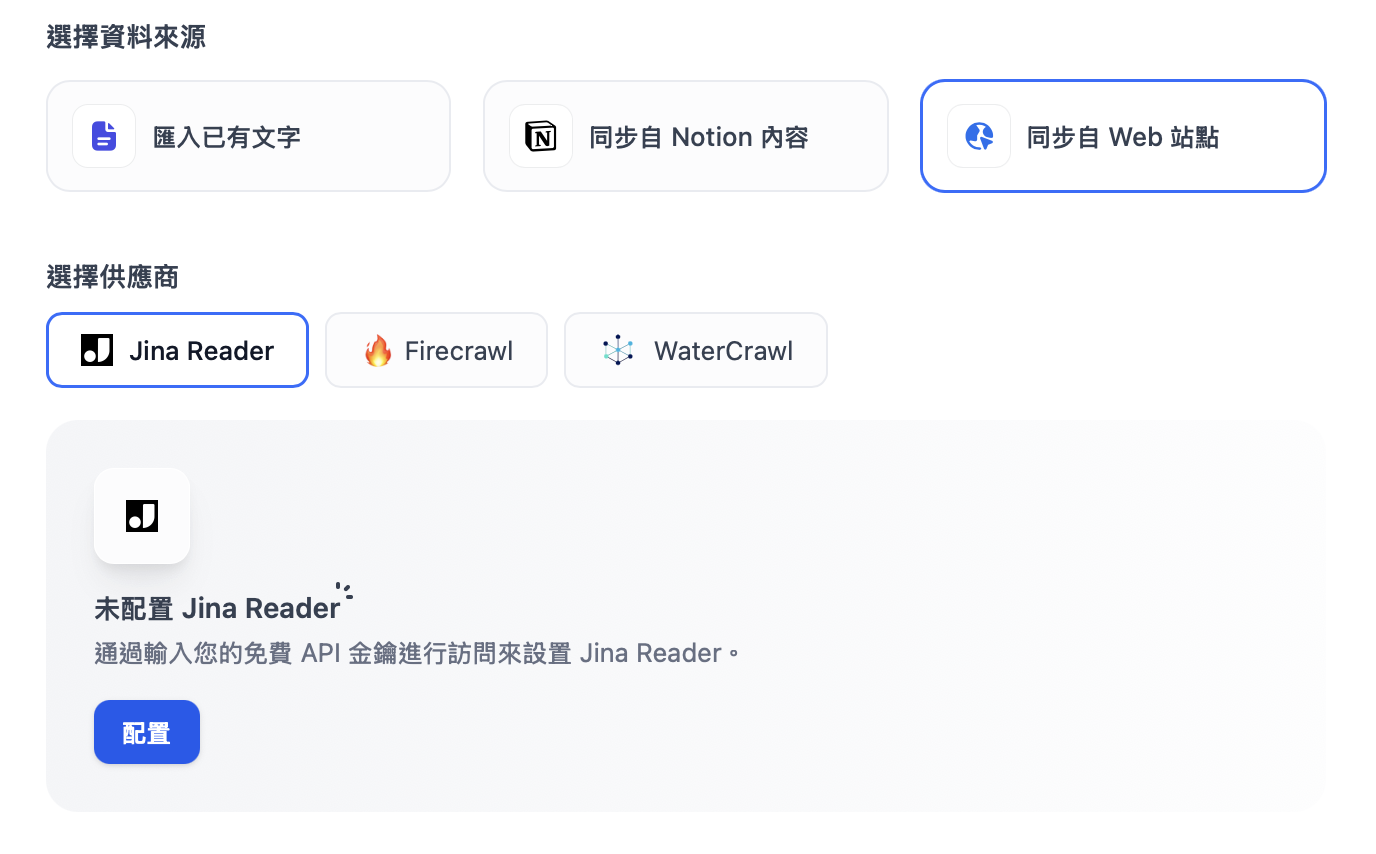
接著可利用「編排」欄位自行設定提示詞，也可以保留空白作為使用原生LLM預設環境的純聊天使用，但通常至少提示聊天時請LLM應配合的事項，例如使用的文字及偏好等，如果不善於自行定義提示詞，也可利用右上角的「產生」功能建立更完善的提示詞。



目前建立的純聊天機器人，其實就是透過Dify去呼叫指定的LLM （在發佈按鈕旁可切換不同的LLM），此時的聊天內容無法聚焦在指定的文本之中，例如我在預覽對話功能中，請ChatGPT 40-mini 介紹區國良，得到的結果如下，實在令人捏把冷汗。。。



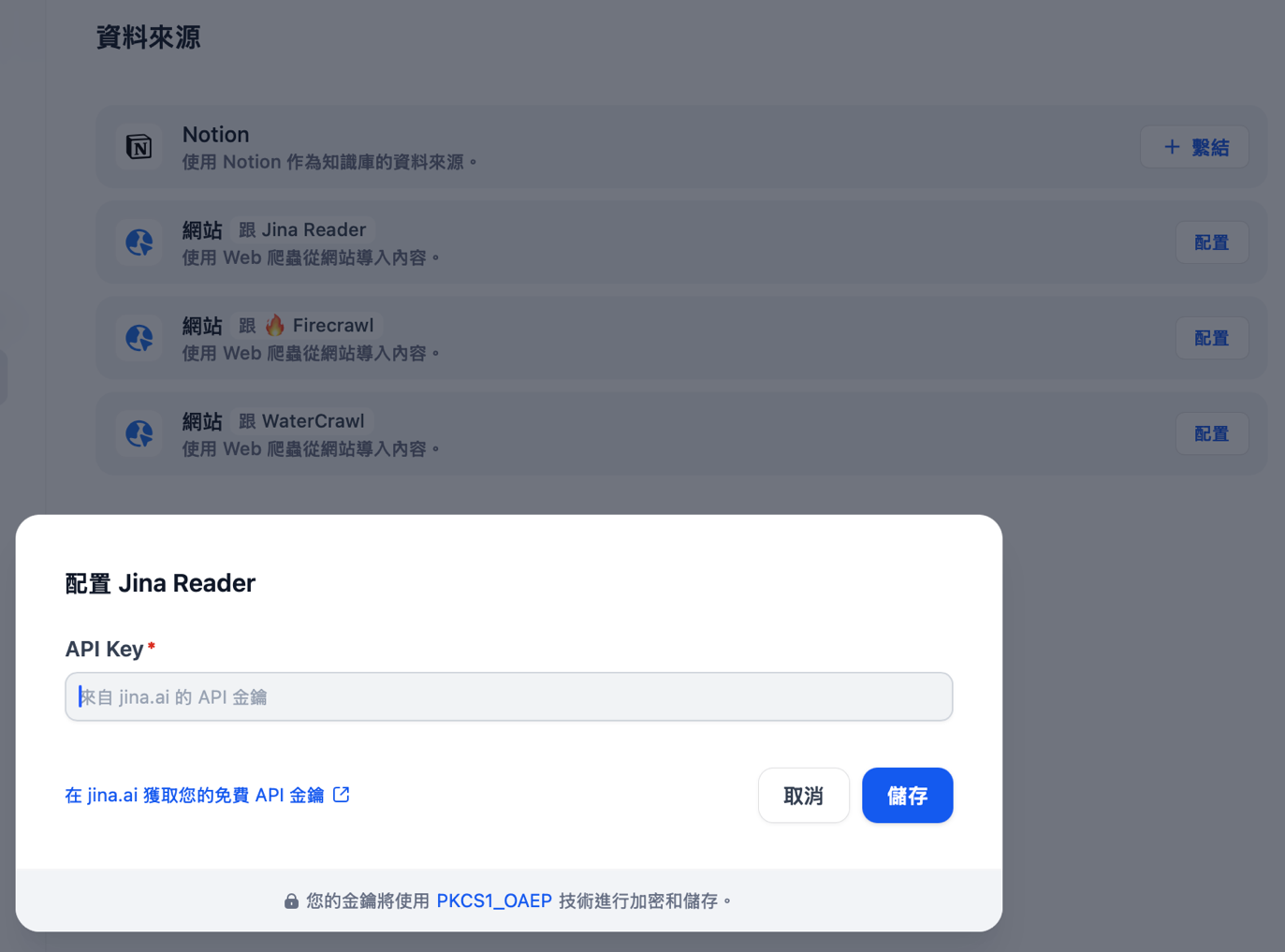
這時就需要RAG的功能將回答的範圍加以設限，請在編排的下方找到「上下文」，其右方的新增按鈕將引導對話內容需參考知識庫，若還沒建立知識庫，則會引導使用者建立知識庫。



知識庫支援文字，notion以及網站，我們可以先測試同步自web網站，並使用其中三個外部應用程式的供應商進行網站爬梳作業，以下是以 Jina reader 為例，請註冊登入後，將 api key 複製。



再回到 Dify 的資源來源選單中，配置 Jina Reader



成功後即可使用指定的網站作為資識庫了，以下是以區國良的官方網站為例，當按下「跑」後，可見其爬梳的進度，並可預覽爬到的網頁是否符合需求，在下一步中，可指定其分段及索引的方式，影響最大的是 embedding 模型，以及 Rerank 服務（Reranker 又稱為二階檢索，依輸入的兩段文字之相似度，輸出一個0～1之間的相關性分數），目前Dify.ai 支援的 Rerank 只有 Cohere， 請自行測試看看。



有了指定網站內容作為知識庫時，這時若再詢問一次相同的問題，答案除了參考知識庫的內容外，LLM還會加油添醋地提出它自己的看法，真是孺子可教也！

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

知識庫可支援多種文件格式，若要將學術論文作為知識庫內容，可分批上傳，Dify會將文件切割成若干個 Chunk，以作為日後回答問題時參考的來源，每個chunk 將會與使用者的問題進行相似度分析，下例為某文件的「召回測試」，該文件為 Computers & Education 的期刊論文，標題為 Identifying supportive student factors for mindset interventions: A two-model machine learning approach，我們可以試著以200字以內的問題進行測試，例如在源文字處輸入 how to employ the machine learning techniques in educational mindset intervention，系統回應3個召回段落，很明顯Score 為0.94的 chunk比 Score0.71 的 chunk 更接近問題字串。

一張含有 螢幕擷取畫面, 文字, 軟體, 字型 的圖片

AI 產生的內容可能不正確。

但這只是單純利用知識庫來源作為LLM回答的參考，真正的考驗是回工作室中再次詢問看看是否對話機器人真的有參考知識庫內容，實測如下：

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

大致上可見系統的回答完全是根據論文而生成的，還不錯！

## 工具

依 dify.ai 官網所言，工具己升級為外掛的生態，大致上就是在這裏提供了第三方的外部工具以協助使用者設計工作流，在此不另外說明。

## 綜合練習

現在，一起來練習如何以清華大學新生入學時需閱讀的註冊須知，建立一個對話機器人，可依使用者的問題類型（目前只練習研究生及大學部的身份）提供必要的解答。依前面所建立的「Question Classifier + Knowledge + Chatbot」範例，請展開 QUESTION CLASSIFIER 節點，修改分類1及分類2的提示內容為「與學士班入學註冊有關的問題」以及 「與研究生入學註冊有關的問題

」一張含有 文字, 軟體, 數字, 電腦圖示 的圖片

AI 產生的內容可能不正確。

接下來，分類1的下一個節點應為 KNOWLEDGE RETRIEVAL，點開之後，在知識庫增加相對應的文件：

學士班入學須知：

<https://registra.site.nthu.edu.tw/var/file/211/1211/img/2979/557596209.pdf>

研究生入學須知：

<https://registra.site.nthu.edu.tw/var/file/211/1211/img/2979/417141842.pdf>

測試結果如下（測試日期為2025/8/30）：

一張含有 文字, 軟體, 字型, 數字 的圖片

AI 產生的內容可能不正確。

# Dify 連接Discord

下面章節，說明如何將Dify 內OpenAI的回應，同步回覆於Discord中。請注意，目前在 Dify.ai 的 marketplace 提供的外掛工具，只能單方面由 Dify.ai 送訊息到 Discord 指定頻道，不支援 Discord 使用者輸入文字啟動 Dify 工作流。

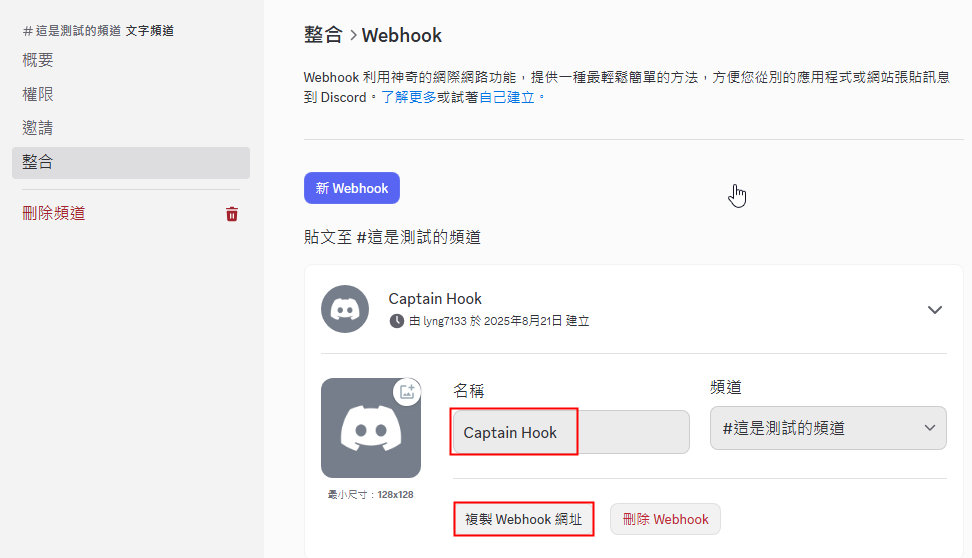
## Discord Webhook設定

首先在 Discord系統中，新增一個「伺服器」及「頻道」，再點選「編輯頻道」進入，選擇「整合」，可以看見「建立Webhook」功能。



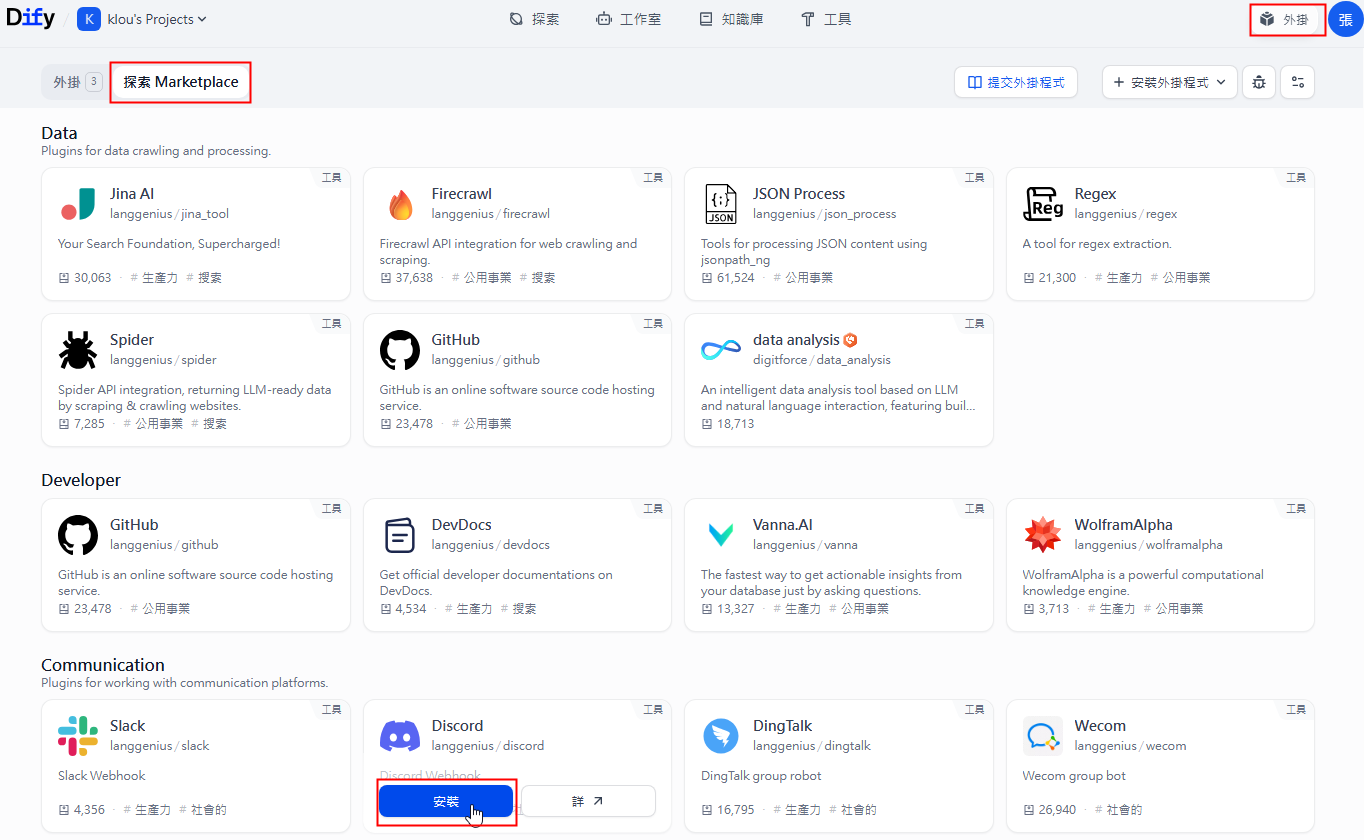


按「建立Webhook」按鈕之後，可以自行修改Webhook名稱，並且按「複製Webhook網址」，之後於Dify 設定使用。另外，也可以按「新Webhook」新增另一組新的webhook

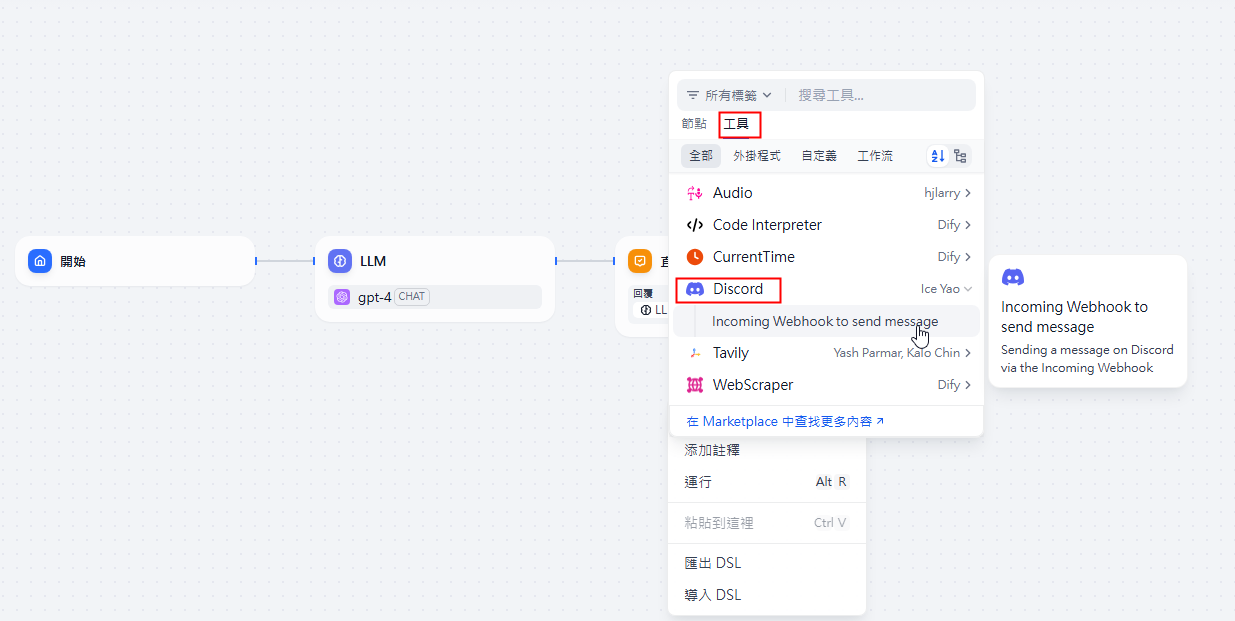


## Dify 設定

由「外掛」->「探索Marketplace」，找到Discord外掛，先進行安裝。



回到「工作室」的流程中，新增Discord節點



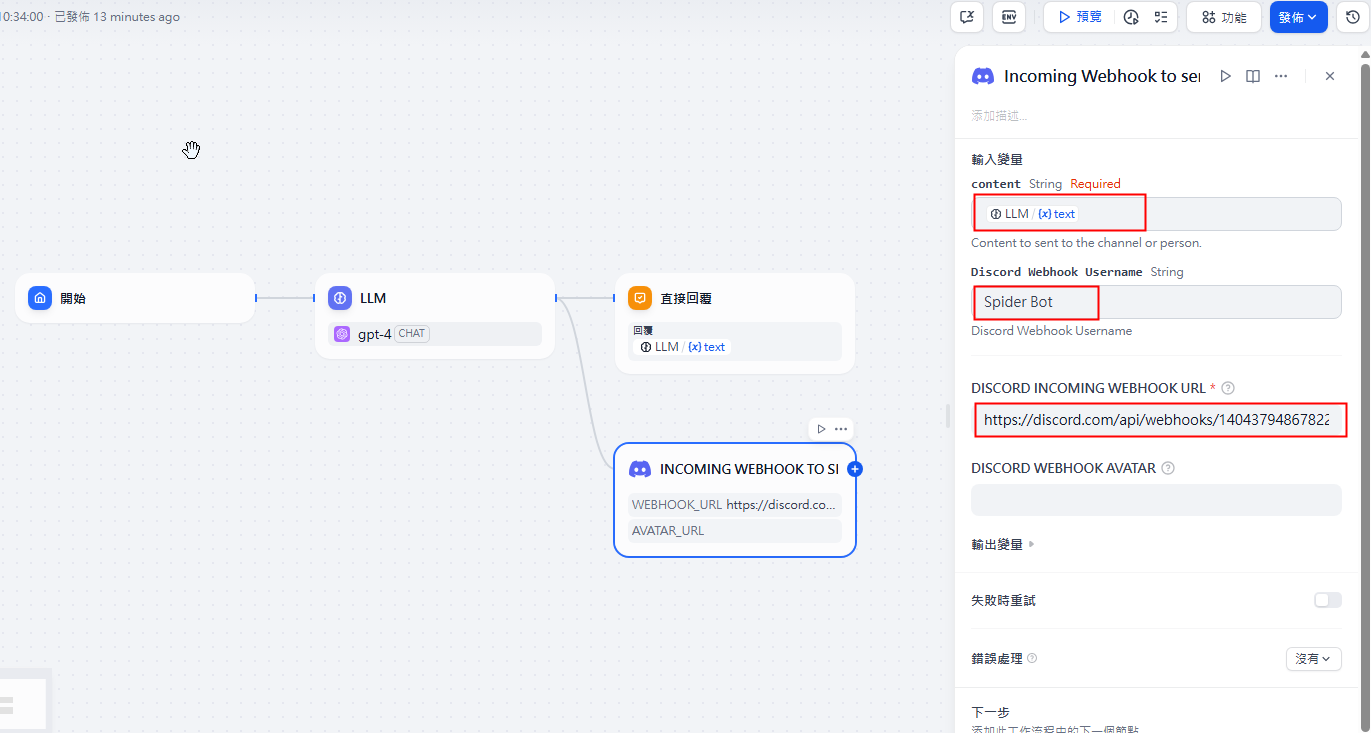
填寫discord節點所需的資料內容如下，並參考下圖。

Content : 填上希望回應於Discord 的內容。

Discord Webhook Username : 回應時的代號稱呼。

Discord Incoming Webhook url : 貼上於discord設定時，複製的webhook網址。

修改完畢，執行「發佈」。

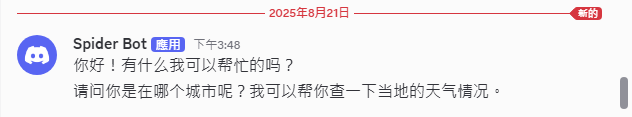


## 執行結果

下圖是Dify執行結果。



由下圖可見，回應於Dify上的內容，同樣也回應於Discord我們設定的頻道中，回應者名稱是我們之前設定的Spider Bot



# LangBot串接Discord與Dify

下面章節說明使用LangBot串接Discord與Dify， 以支援 Dify 與 Discord 雙向互動，但只能以本地端安裝的 Dify 系統才能使用，雲端的 Dify.ai 無法提供相關服務。

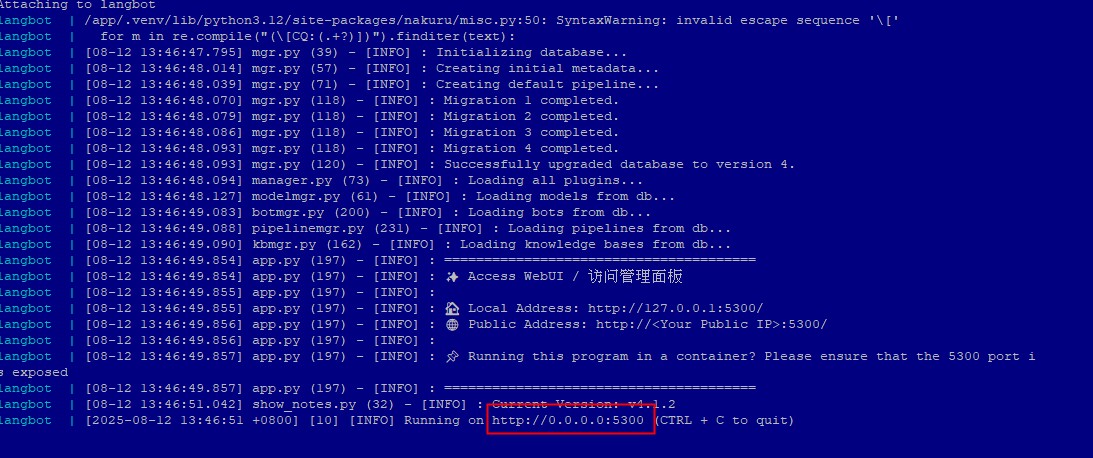
LangBot系統會擷取 Discord 上的使用者訊息，將其傳送至 Dify 進行處理，將回應內容自動回覆至 Discord。可以參考下面網址，閱讀LangBot的說明文件，<https://docs.langbot.app/zh/insight/guide.html>。

## 安裝LangBot

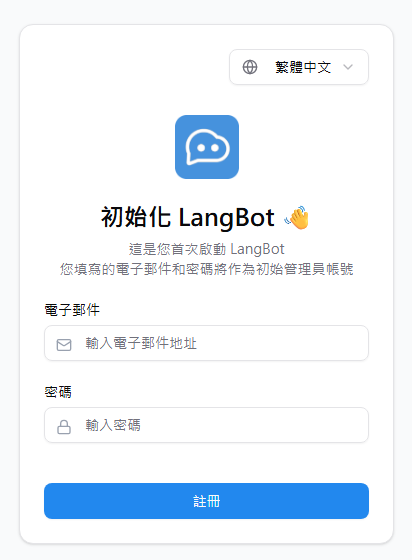
參考<https://docs.langbot.app/zh/deploy/langbot/docker.html>，使用docker安裝。

* git clone https://github.com/langbot-app/LangBot
* cd LangBot
* docker compose up

執行docker compose up後，出現下列畫面，開啟port 5300。可以連http://安裝的位置:5300，開始進行LangBot的設定。



一開始會請使用者註冊帳號(下圖左)，註冊完之後，就可以使用這帳號登入系統(下圖右)，開始進行設定。

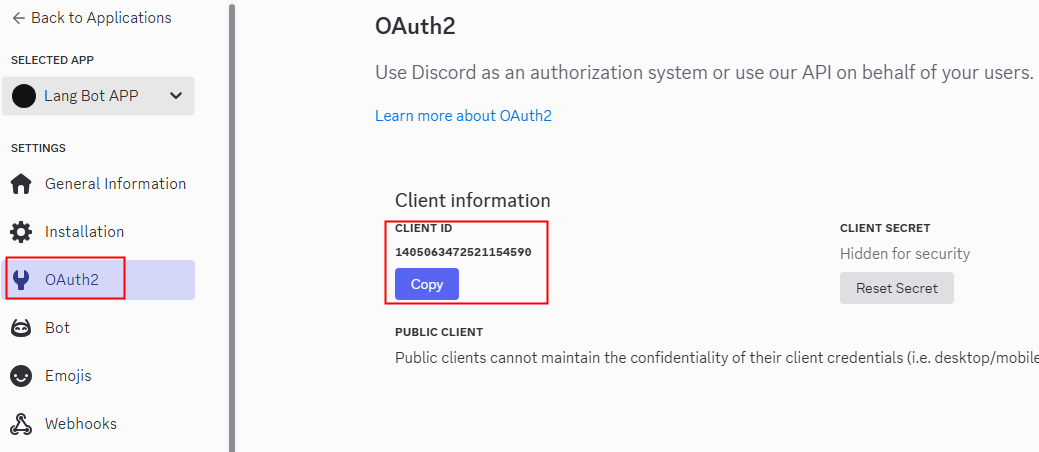
 

## LangBot連接Discord設定

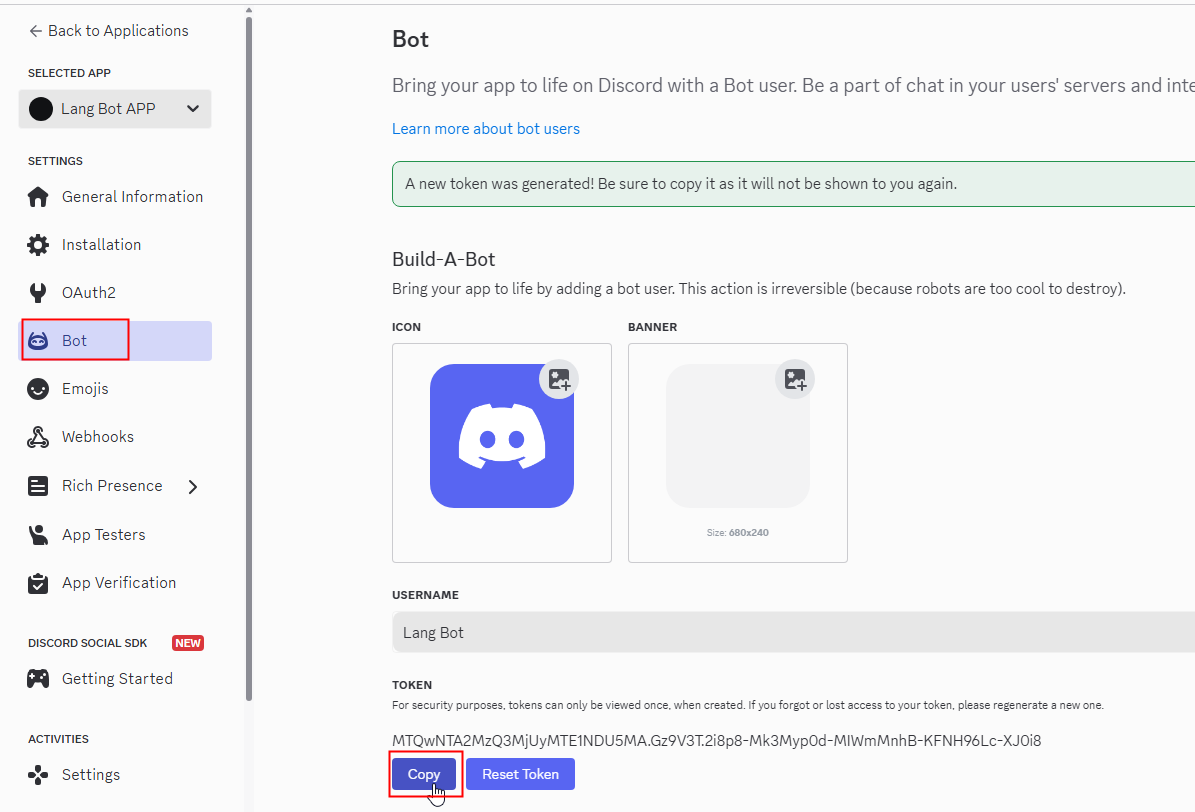
參考<https://docs.langbot.app/zh/deploy/platforms/discord.html>，使依照文件所指示的步驟建立Discord Application、機器人，並且邀請至伺服器中。

Discord :

進入建立的application中，選擇「OAuth2」，複製CLIENT ID。



選擇「Bot」，建立機器人，並且複製TOKEN。



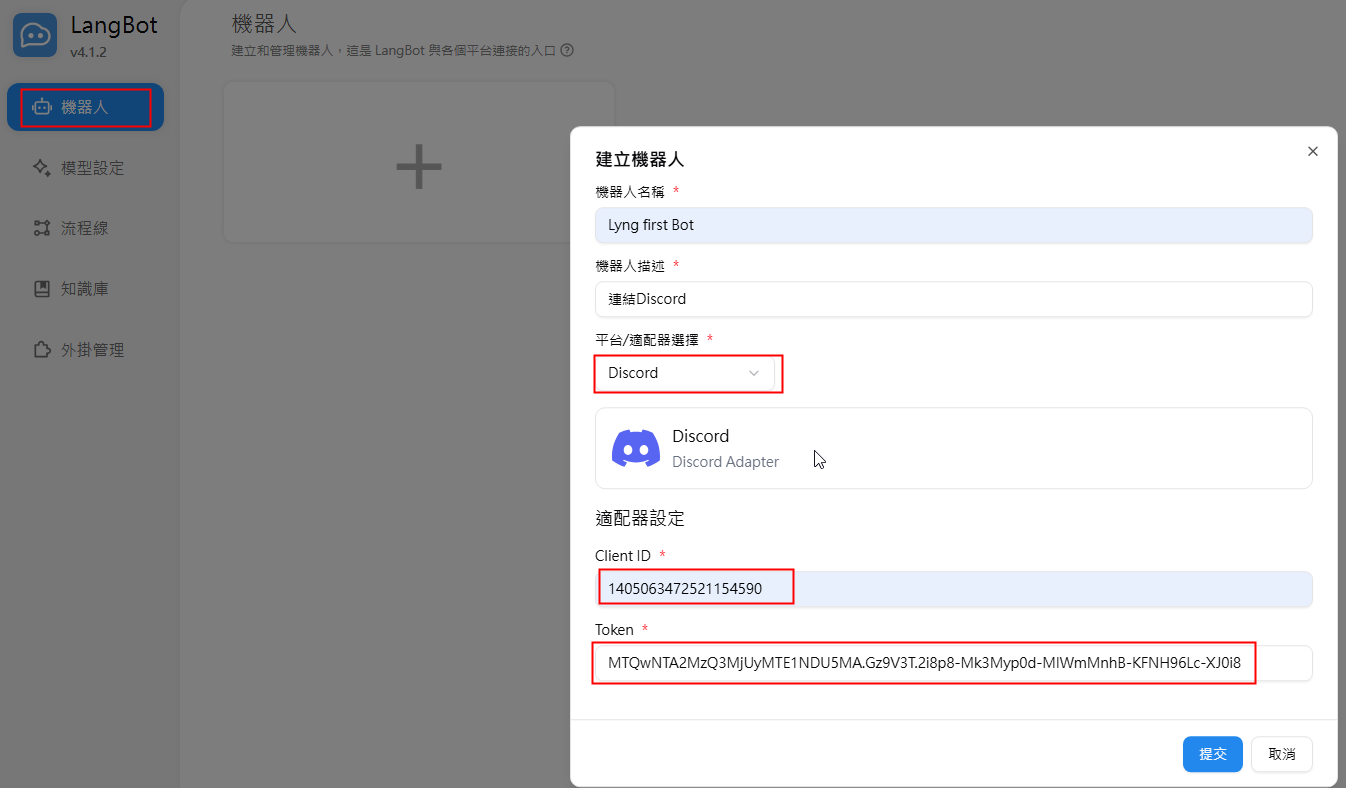
LangBot:

使用剛剛註冊的帳密進入LangBot中，選擇機器人，新增一個機器人。

機器人名稱跟描述：可以自行提供。

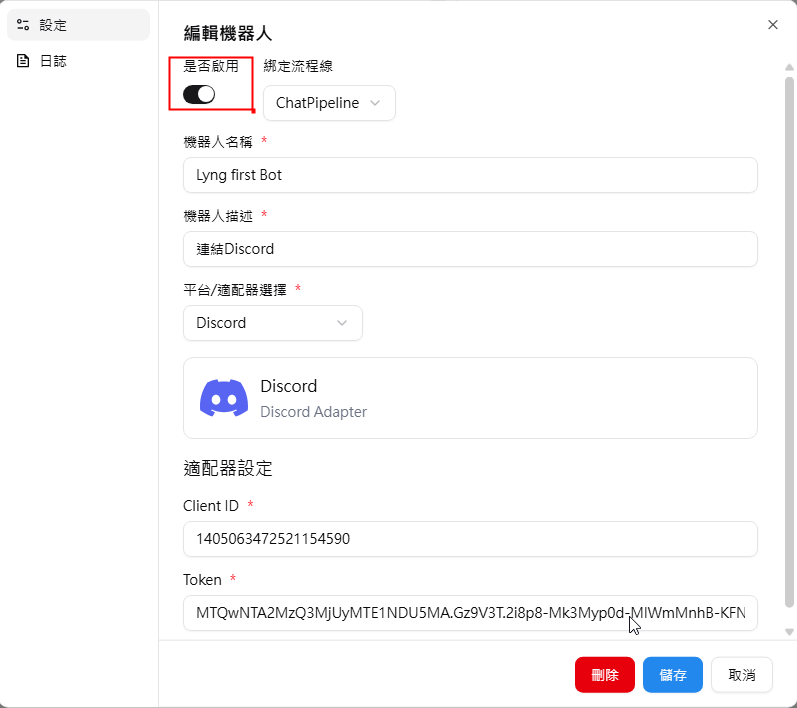
平台/適配器選擇：選擇Discord

Client ID / Token：依照之前從Discord複製的client id與token資訊填入。

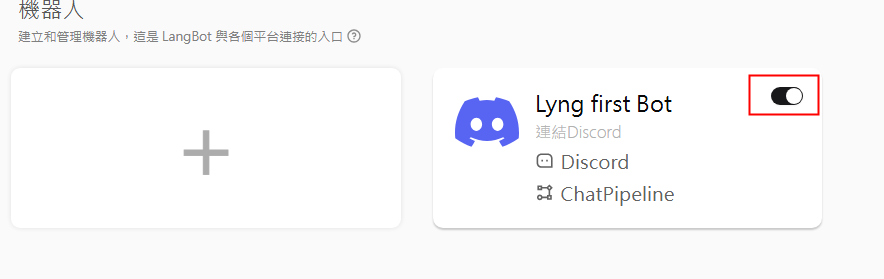


提交後，進入下面畫面，可以選擇啟用，然後儲存。啟用後，Discord 機器人即會上線。

流程線會先綁定預設的流程線ChatPipeline，等之後有設定新的流程線，再進行修改。



也可以直接在外層選擇啟用或關閉機器人。



回到Discord，可以看到我們剛剛串接的機器人已經上線。

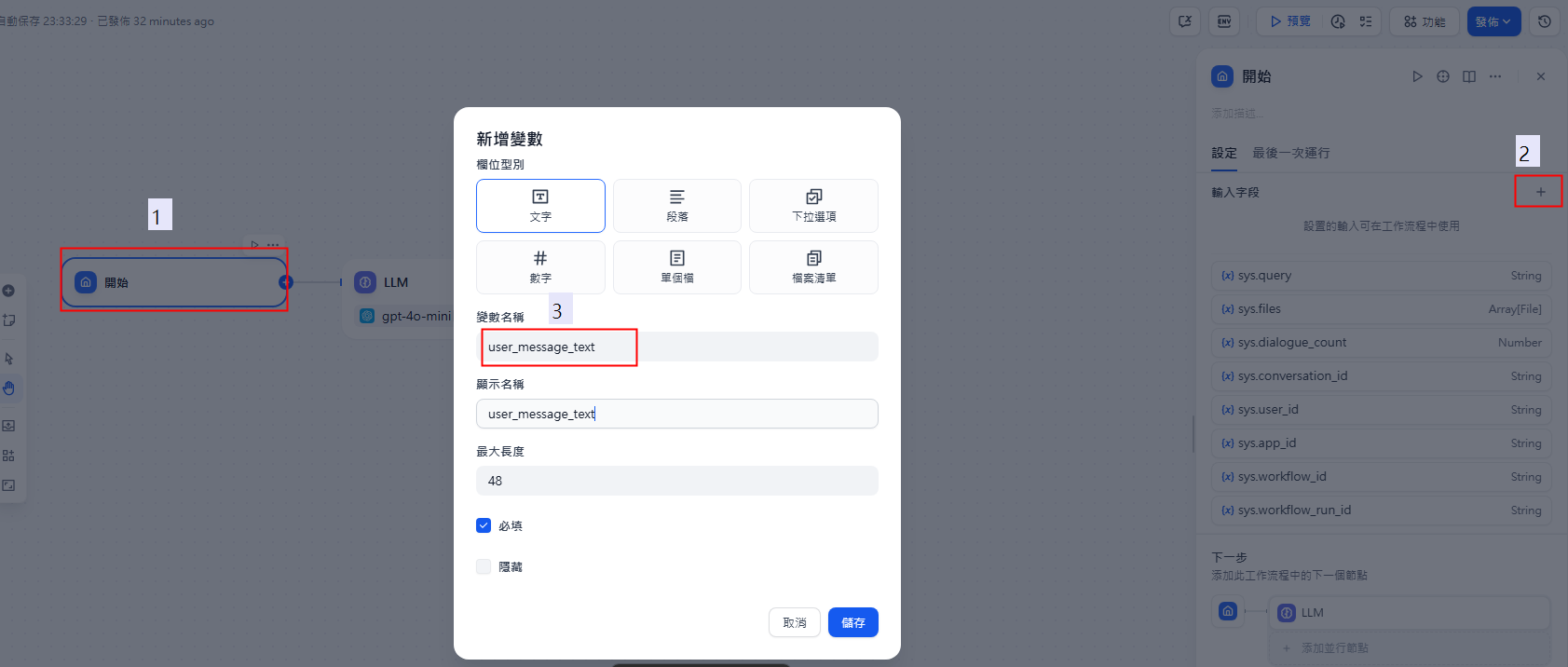


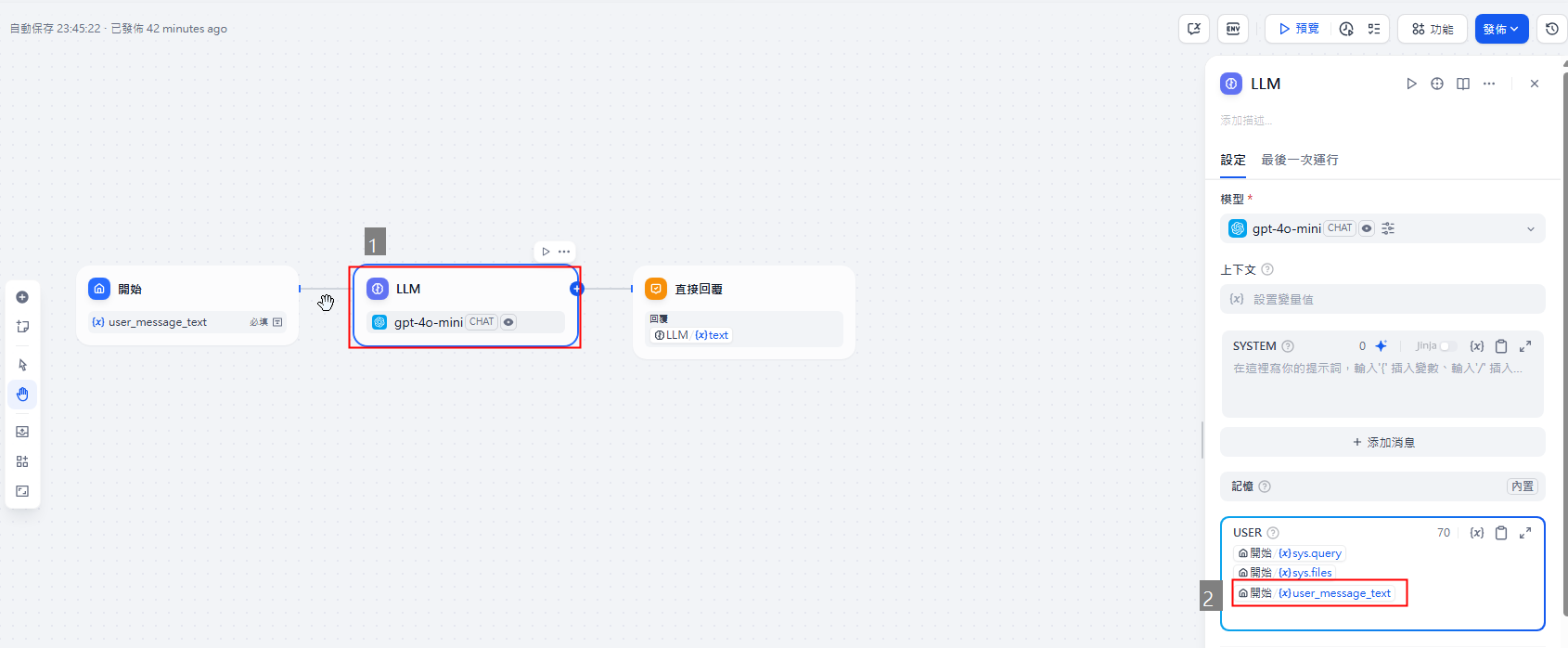
## LangBot連接Dify設定

參考下面兩個網站<https://docs.langbot.app/zh/deploy/pipelines/dify.html>與<https://docs.dify.ai/zh-hans/learn-more/use-cases/connect-dify-to-various-im-platforms-by-using-langbot>，關於Dify與LangBot的設定，如何連結LangBot與Dify。

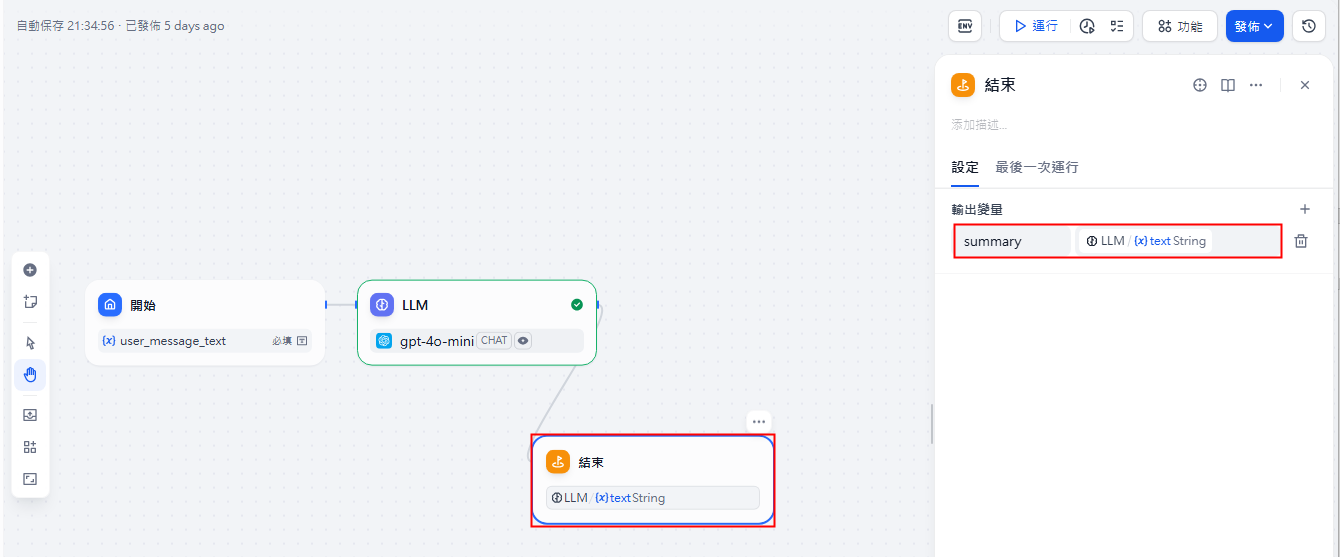
Dify-聊天流(ChatFlow)：

新增一個chatflow，並新增變數user\_message\_text接收從Discord傳來的訊息，並且將user\_message\_text設定於user prompt中。

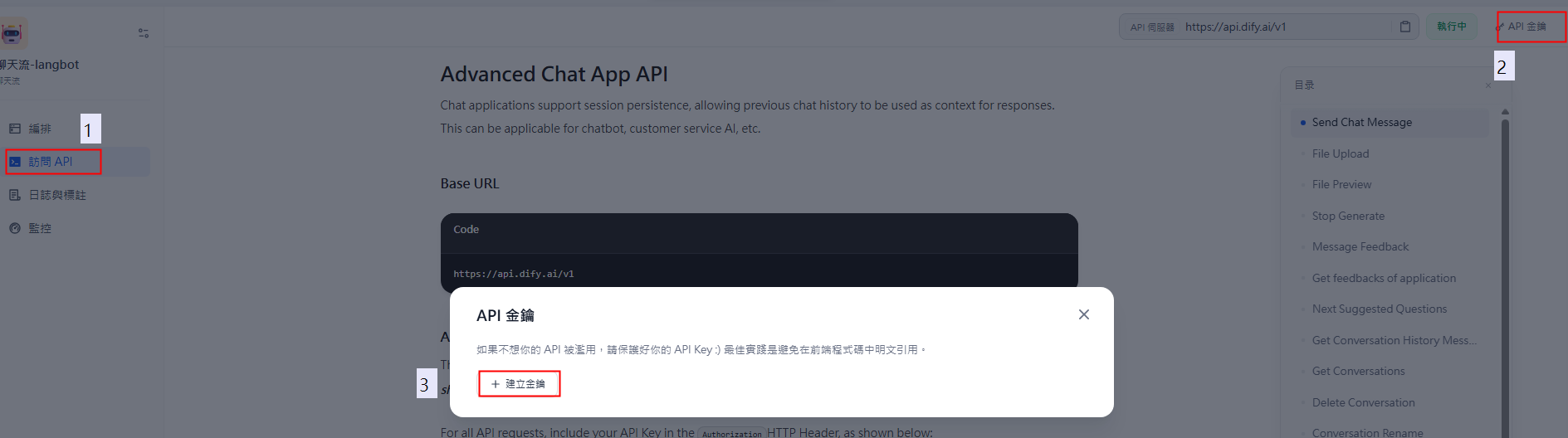




另外提醒，如果Dify 使用的是工作流(workflow),需要在最後加上一個結束的節點，並且將LLM最後執行的結果放置summary變數中。

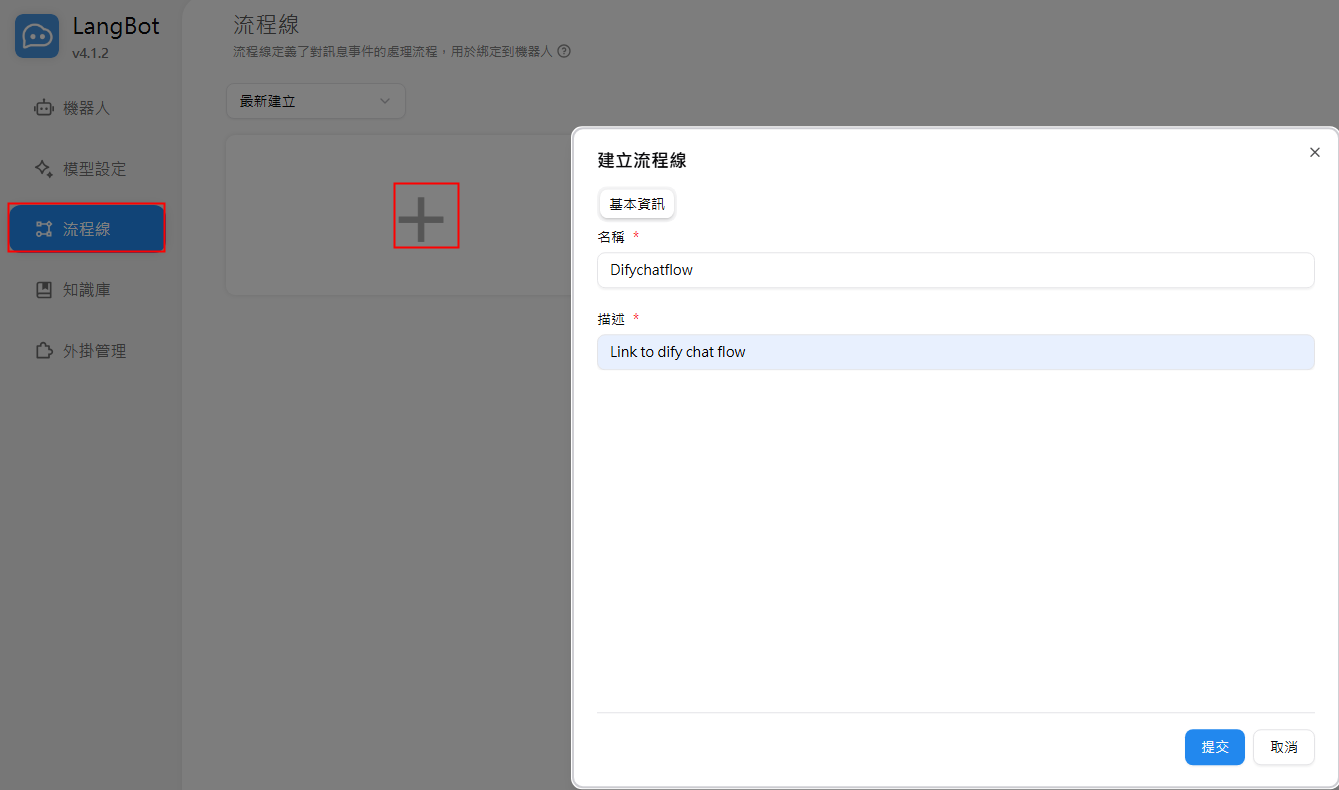


接下來，由下圖產生API金鑰，並將API金鑰複製。



LangBot:

連結Dify的設定，選擇「流程線」,建立一條新的流程線，自行輸入名稱與描述。



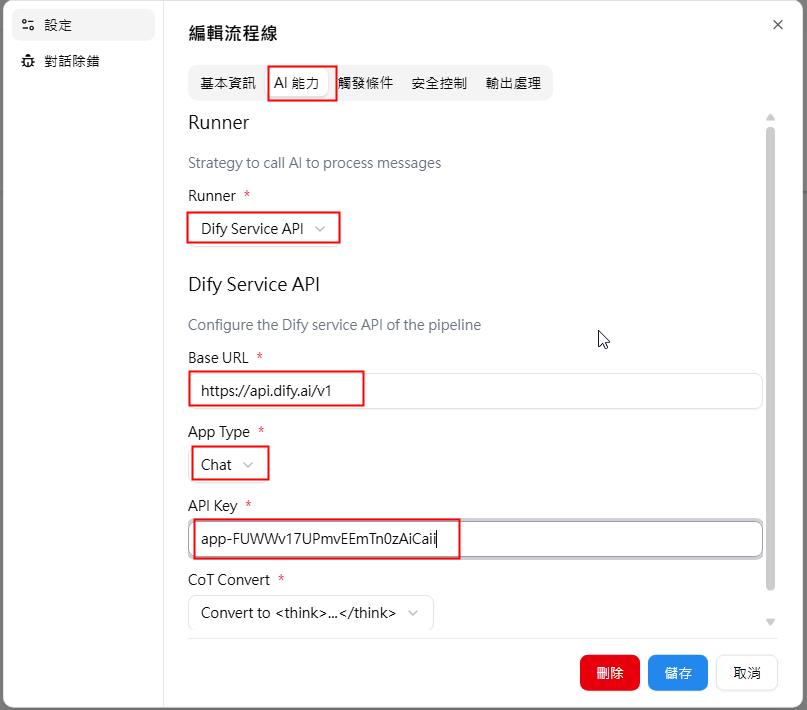
按「提交」後，如下圖，選擇tab 「AI能力」：

Runner ：選Dify Service API。

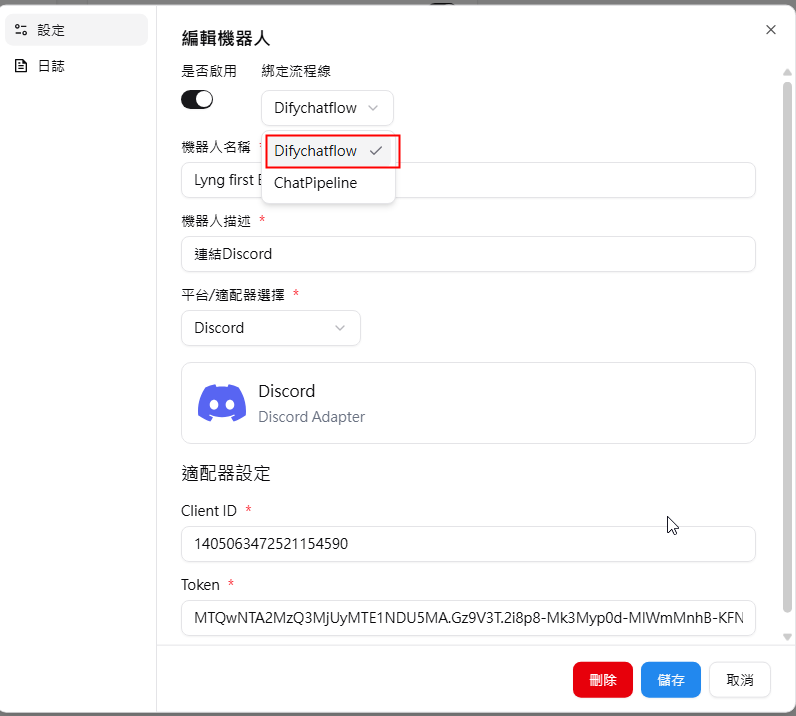
Base URL： 輸入正確網址，如果使用雲端Dify，不需要修改。

App Type：選Chat /Work Flow/ Event

API Key：將剛剛從Dify flow產生的API金鑰，填入此處。然後儲存設定。



最後記得回到機器人，將「定流程線」修改成我們剛剛新建立的流程線。



## 執行結果

在Discord提問，機器人透過連結Dify會自動回應。



另外說明，流程線「觸發條件」中

At：如果打開，可以用@機器人，呼叫機器人回答。

Prefix：定義用來呼叫機器人回答的前綴詞，不能是空白。如下面第二個圖為例，用「ai你喜歡下雨的天氣嗎?」ai來啟動機器人。

