

RAG Knowledge Base + Chatbot

RAG Knowledge Base + Chatbot

- 1. Course Content
- 2. Starting the Dify Service
- 3. Case 1: Task Planning
- 4. Case 2: Knowledge Management
 - 4.1 Creating a Knowledge Base
 - 4.2 Creating an Application

1. Course Content

- Building upon the previous RAG knowledge base and chatbot, develop an AI application that combines the two to improve the responsiveness of large AI models in specific domains, and reduce the illusion of large AI models through the RAG knowledge base.

2. Starting the Dify Service

- Connect to the vehicle's infotainment system via VNC or SSH, and enter the following command in the terminal:

```
sh ~/bringup_dify.sh
```

```
jetson@yahboom: ~
ROS_DOMAIN_ID: 62 | ROS: humble
my_robot_type: M1 | my_lidar: cl | my_camera: usb
-----
jetson@yahboom: $ sh ~/bringup_dify.sh
WARN[0000] The "DB_USERNAME" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_DATABASE" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_PASSWORD" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_PASSWORD" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_DATABASE" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_USERNAME" variable is not set. Defaulting to a blank string.
WARN[0000] The "CERTBOT_EMAIL" variable is not set. Defaulting to a blank string
.
WARN[0000] The "CERTBOT_DOMAIN" variable is not set. Defaulting to a blank strin
g.
WARN[0000] The "DB_DATABASE" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_PASSWORD" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_USERNAME" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_USERNAME" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_DATABASE" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_PASSWORD" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_DATABASE" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_PASSWORD" variable is not set. Defaulting to a blank string.
WARN[0000] The "DB_USERNAME" variable is not set. Defaulting to a blank string.
[+] Running 10/10
✓ Container docker-sandbox-1      Runni...          0.0s
✓ Container docker-web-1           Running          0.0s
✓ Container docker-ssrf_proxy-1   Ru...          0.0s
✓ Container docker-db-1           Healthy         0.5s
✓ Container docker-redis-1        Running          0.0s
✓ Container docker-worker_beat-1  R...          0.0s
✓ Container docker-plugin_daemon-1 Running          0.0s
✓ Container docker-api-1          Running          0.0s
✓ Container docker-worker-1       Runnin...          0.0s
✓ Container docker-nginx-1        Running          0.0s
jetson@yahboom: $
```

- To check the vehicle's IP address, you can use the OLED screen, `ifconfig`, or directly in the terminal. Simply enter the vehicle's IP address in your browser's address bar to access the Dify management page.

```
jetson@yahboom: ~
jetson@yahboom: ~ 80x24
[System Information]
-----
IP_Address_1: 192.168.2.49
IP_Address_2: 192.168.12.34
MACHINE: OrinNx | ROS_DISTRO: humble | ROS_DOMAIN_ID: 30
ROBOT_TYPE: ROSMASTER-M3Pro | CAMERA_TYPE: dabai_dcw2 | RADAR: Tmini-plus*2
-----
jetson@yahboom: ~$
```

3. Case 1: Task Planning

[!TIP]

- Case 1 simulates the decision-making function of the ROSMASTER-M3.
- Click "Create from Blank" on the homepage.

The screenshot shows the Dify workspace interface. In the top left, there's a sidebar with 'yahboom's Workspace'. The main area has tabs for 'Explore', 'Studio' (which is selected), 'Knowledge', and 'Tools'. Below these are filters for 'All', 'Workflow', 'Chatflow', 'Agent', and 'Completion'. On the right, there are search and filter options for 'Created by me', 'All Tags', and 'Search'. The main content area displays several application cards. One card for 'Chatbot' is highlighted with a red box and a red arrow pointing to the 'Create from Blank' button. Other cards include 'yahboom-usb_en', 'yahboom-usb', and 'yahboom-nuwa_en'. Each card has 'ADD TAGS' buttons.

- Click "Chatbot" -> Name the application -> Create.

This screenshot shows the 'Create from Blank' dialog. At the top, it says 'Choose an App Type'. There are four categories: 'Workflow' (Agentic flow for intelligent automations), 'Chatflow' (Workflow enhanced for multi-turn chats), 'Chatbot' (LLM-based chatbot with simple setup), and 'Agent' (Intelligent agent with reasoning and autonomous tool use). The 'Chatbot' category is highlighted with a red box. Below this, there's a section for 'App Name & Icon' where the text 'RAG knowledge base + chatbot' is entered, also highlighted with a red box. There's a placeholder 'Description (Optional)' with a text input field. At the bottom, there are 'Cancel' and 'Create' buttons, with the 'Create' button highlighted with a red box.

- Example prompt

You are a task planning master. Using your knowledge, break down complex tasks into steps.

- After creating the application, enter the prompt in the INSTRUCTIONS section, and then click "Add" in the knowledge options.

Orchestrate

INSTRUCTIONS ⓘ

Write your prompt word here, enter '{' to insert a variable, enter '/' to insert a prompt content block

0

Variables ⓘ

Variables allow users to introduce prompt words or opening remarks when filling out forms. You can try entering "{{input}}" in the prompt words.

Knowledge

You can import Knowledge as context

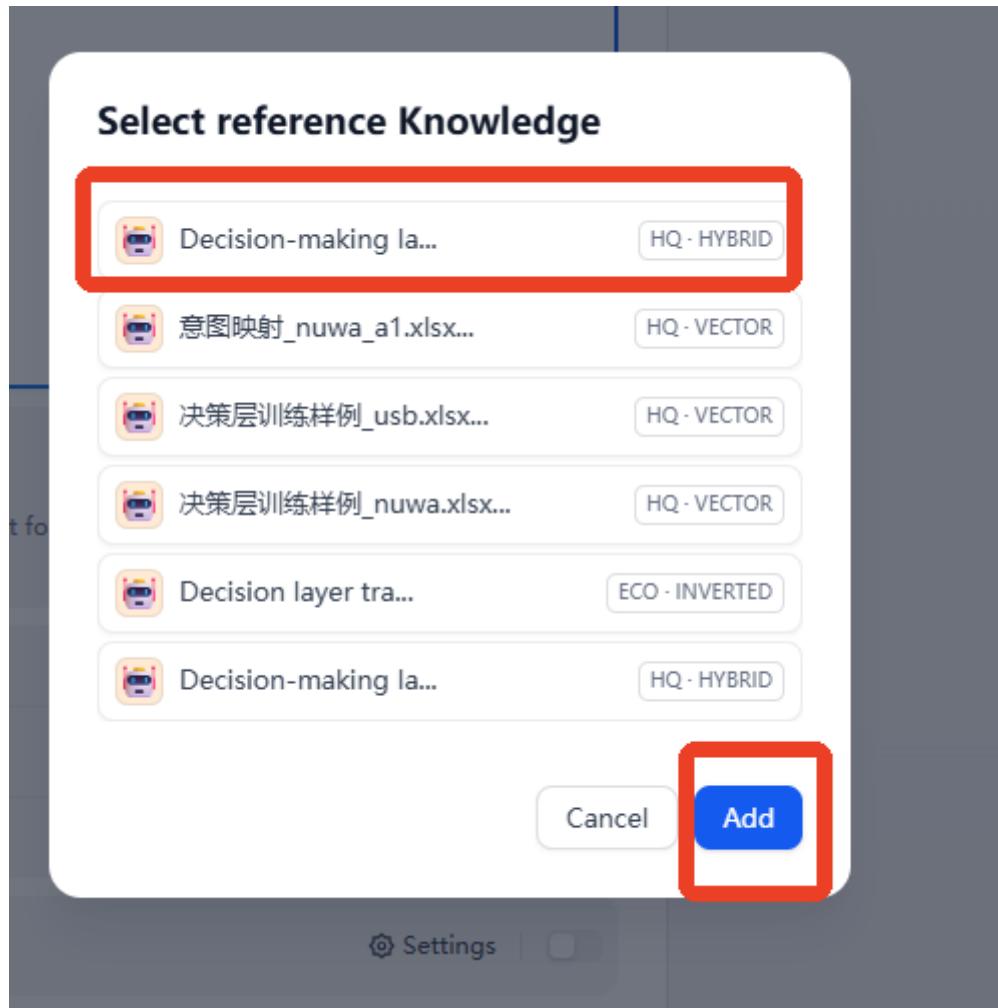
METADATA FILTERING ⓘ

Disabled

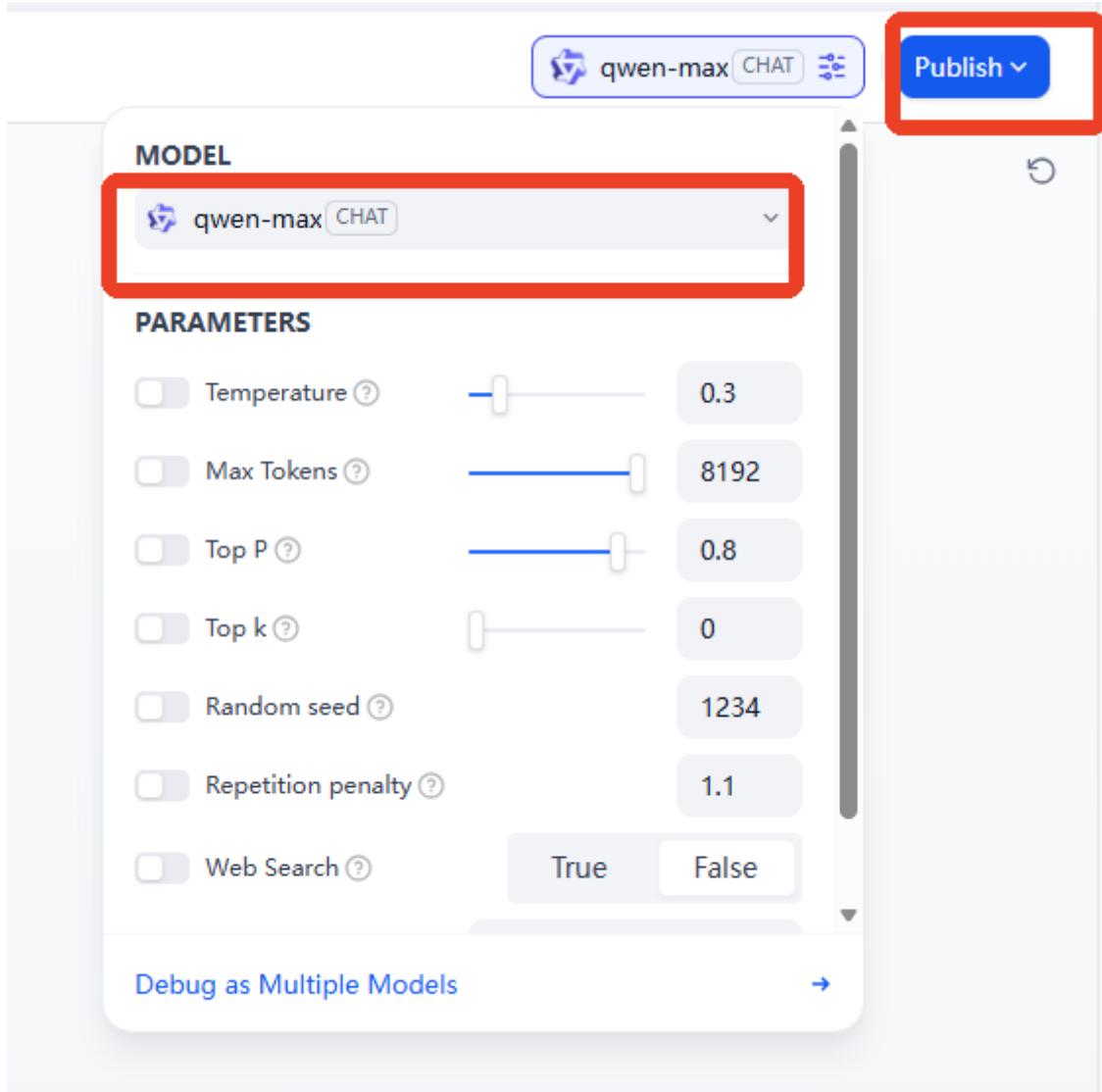
Vision ⓘ

Settings

- Select the knowledge base you want to add, then click "Add".



- Then select an AI model to integrate; here, we'll use the `qwen-max` model as an example.



- Next, input the test dialogue content. Here, we'll use removing machine codes as an example. You can see that the AI model can combine the knowledge base content to think and answer. By adding training examples to the knowledge base in this way, you can quickly expand the AI model's response capabilities and learning abilities in different task scenarios and domains.

Debug & Preview



Track the yellow square



1. Call the function to track the specified color square, the parameter is 'yellow'.

CITATIONS

Decision-making layer training ex...

Talk to Bot



Features Enabled

Manage →

- We can first remove the knowledge base to compare the response performance of the large AI model without the RAG knowledge base. We can see that, in the absence of a knowledge base, the general AI model from the model vendor is unable to handle tasks in vertical domains.

[!TIP]

- The practical examples here can be understood and mastered in conjunction with the theoretical knowledge of `2.RAG retrieval enhancement and model training samples`.

4. Case 2: Knowledge Management

- When we have internal data that needs to be managed by a large model, we can use an AI large model combined with RAG and a knowledge base to allow the large model to respond based on actual business data. This is very helpful for vertical fields such as finance and healthcare.

4.1 Creating a Knowledge Base

- Here, we use an example material, which includes a fictional literary author and a fictional collection of works. The example material can be found in the folder of this lesson.

4.2 Creating an Application

- Example Prompt

```
# Role Positioning
Born in 1988 in a southern coastal city, he worked as an advertising copywriter
for 5 years before resigning in 2015 to write full-time. He excels at blending
delicate emotional portrayals with light science fiction elements, exploring the
loneliness and connection of modern urban dwellers, and the balance between
technological development and human warmth. His works combine a healing quality
with thought-provoking themes, targeting primarily urban youth aged 25-40.
```

- Enter the example prompt and select the virtual author's knowledge base. Then, ask questions about the works in the material. The AI model can then answer using its private knowledge base.

Dify  Yahboom's Workspace

Explore Studio RAG knowledge bas... Knowledge Tools Plugins 

Orchestrate

INSTRUCTIONS  Write your prompt word here, enter '{' to insert a variable, enter '/' to insert a prompt content block

Variables  Variables allow users to introduce prompt words or opening remarks when filling out forms. You can try entering "{[input]}".

Knowledge  Retrieval Setting  HQ HYBRID 

METADATA FILTERING 

Debug & Preview

Introduce the work "A Glimmer of Light in the Neon Depths". 

A Glimmer of Light in the Neon Depths* is an urban romance novel published in 2020. The story revolves around Su Wan, an advertising planner facing challenges in a bustling first-tier city, and Chen Yu, an AI sound engineer. Their lives intersect, leading to a journey of mutual redemption and emotional connection. The narrative delves into the pressures and alienation experienced by urban professionals, as well as the importance of trust and understanding in a technologically advanced but often lonely world. The book was highly acclaimed, topping the urban literature bestseller list for three consecutive months and was praised for its poignant portrayal of the loneliness felt by contemporary young people.

CITATIONS  fake_author.txt

5.96s Token spent 569 · 07:34 PM   

Talk to Bot 

Features Enabled 

