

WeKnora ??????????????

??????
Rerank

WeKnora

DashScope / Qwen / Embedding /

?????

“?????”

”?????”

LLM / Embedding / Rerank

? ?????? WeKnora ??????????

????

DashScope

WeKnora

???	????	WeKnora type
????	qwen-plus / qwen-max / qwen-turbo	KnowledgeQA
????	text-embedding-v2 / text-embedding-3	Embedding
???	qwen-rerank-v2 / gte-rerank-v2	Rerank

?? ?????? DashScope Base URL ?????

??????

<https://dashscope.aliyuncs.com/compatible-mode/v1>

? ??API Key

DASHSCOPE_API_KEY=sk-xxxx

? ??WeKnora

builtin_models.yaml??????????

□□□□□□□□

? 1?? ??????Qwen?

```
builtin_models:
  - id: builtin-qwen-plus
    name: qwen-plus
    type: KnowledgeQA
    source: aliyun
    description: □□□□ Qwen □□□□
    is_default: true
    parameters:
      base_url: https://dashscope.aliyuncs.com/compatible-mode/v1
      api_key: ${DASHSCOPE_API_KEY}
      provider: generic
```

? 2?? Embedding???????

```
- id: builtin-qwen-embedding
  name: text-embedding-v2
  type: Embedding
  source: aliyun
  parameters:
    base_url: https://dashscope.aliyuncs.com/compatible-mode/v1
    api_key: ${DASHSCOPE_API_KEY}
    provider: generic
    embedding_parameters:
      dimension: 1536
      truncate_prompt_tokens: 8192
```

?? ????? text-embedding-3-large?????

```
dimension: 1024 # ???????
```

? 3?? Rerank?Qwen rerank-v2?

```
- id: builtin-qwen-rerank
  name: qwen-rerank-v2
  type: Rerank
  source: aliyun
  parameters:
    base_url: https://dashscope.aliyuncs.com/compatible-mode/v1
    api_key: ${DASHSCOPE_API_KEY}
    provider: generic
```

? ??????????????

? 1. provider ??? openai

```
??????
```

```
provider: generic
```

? 2. base_url ??? api.openai.com

```
????
```

```
dashscope.aliyuncs.com
```

? 3. rerank endpoint ??

/rerank

? ? ? ? ? ? ? ? ? ?

[] .env []

DASHSCOPE_API_KEY=[] []key

? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?

```
builtin_models:  
  - id: qwen-chat  
    name: qwen-plus  
    type: KnowledgeQA  
    source: aliyun  
    is_default: true  
    parameters:  
      base_url: https://dashscope.aliyuncs.com/compatible-mode/v1  
      api_key: ${DASHSCOPE_API_KEY}  
      provider: generic  
  
  - id: qwen-embedding  
    name: text-embedding-v2  
    type: Embedding  
    source: aliyun  
    parameters:  
      base_url: https://dashscope.aliyuncs.com/compatible-mode/v1  
      api_key: ${DASHSCOPE_API_KEY}  
      provider: generic  
      embedding_parameters:  
        dimension: 1536  
        truncate_prompt_tokens: 8192
```

```
- id: qwen-rerank
  name: qwen-rerank-v2
  type: Rerank
  source: aliyun
  parameters:
    base_url: https://dashscope.aliyuncs.com/compatible-mode/v1
    api_key: ${DASHSCOPE_API_KEY}
    provider: generic
```

? ??????????

□□

```
docker compose restart
```

□□□□

```
docker logs WeKnora-app | grep -i model
```

? ??????????????????

? 1. ??????

□□ □□

- yaml □□□
- □□□□□□
- BUILTIN_MODELS_CONFIG □□□

? 2. embedding ??????

□□ 99% □□

```
dimension □□
```


■■■■■■■■■■

? ??????????????????

■■■■

```
models  ✓  ✓  
3  ✓  ✓  
tenant_id = 10000 ✓  
status = active ✓
```

■■■

```
“  builtin_models loader  ■■■■■  
  ■■■■■  
  UI  = ■■■■■
```

? ??????????????????

■■■■

```
type = KnowledgeQA  
type = Embedding  
type = Rerank
```

■■■■■■■■

```
is_default = f
```

■■ ■■ false

? ??WeKnora UI ??????????????

UI “ models”

? ??????????

```
WHERE is_builtin = true  
OR managed_by = 'system'  
OR tenant_id = 0  
OR is_default = true
```

? ??????????????????

```
is_builtin = f 
```

UI

? ????? tenant_id?????????

```
tenant_id = 10000
```

WeKnora UI

```
tenant_id = 0  NULL = system models
```

?? ??????????????????

<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	✓ <input type="checkbox"/>

UI	UI
YAML	✓
UI	
tenant	

? ?????? UI ????????

WeKnora UI

```
SELECT * FROM models
WHERE tenant_id = current_tenant
AND is_builtin = true;
```

UI

- tenant_id = 10000
- is_builtin = false

UI

```
“ UI
```

? ??????????????

? ?? 1????? builtin

UI SQL

```
UPDATE models
SET is_builtin = true
WHERE tenant_id = 10000;
```

? ?? 2????? tenant_id??????

```
UPDATE models
SET tenant_id = 0
WHERE tenant_id = 10000;
```

? ?? 3????????????????

```
UPDATE models
SET is_builtin = true,
    is_default = true
WHERE name = 'qwen-turbo';
```

? ??????????

☐☐☐

```
docker compose restart
```

☐☐☐ UI☐

☐☐ Settings → Models

? ???????? YAML ???“????UI????”

☐☐☐

```
YAML ✓
↓
DB☐☐ ✓
↓
UI☐☐ ☐☐tenant/is_builtin☐
↓
```


□□□□□□□□

□□ “□□□□□□□□”

□□□□□□□□□□□□□□□□

Revision #2
Created 4 July 2026 07:04:53 by Admin
Updated 4 July 2026 07:45:01 by Admin