设置对话模型

开始之前，或许你需要读一读这个：yaml文件填写规范

需要反复强调的部分

在填写配置项时，windows整合包用户打开launcher，在设置页面就能看到，不要自己去直接修改文件，在你不清楚格式要求的情况下，贸然修改本地文件容易出现格式错误，最终将导致bot无法运行。

你自己用 设定#模型名 指令的优先级高于你在配置文件填写的模型，所以不要再问"为什么我修改了配置文件模型还是没反应"的问题了。

一旦你用了 设定#模型名 的指令，配置文件设置的模型将不再对你生效。

为了避免还是有人看不懂，我再说明白一点：
设定#模型名 设置的模型只对设定者个人生效；
而配置文件中设置的模型，是对所有人(除了用过 设定#模型名 的人)生效的

`如果你还是看不懂上面说的是什么意思，那就记住不要用 设定#模型名 这样的指令。`

下面是配置文件设置模型的相关内容

最下方有比较详细的配置方式

模型(settings.yaml中的model设置)	介绍	配置项(api.yaml对应)	评价
characterglm	智谱的超拟人大模型，在这里申请	chatGLM	付费api，群少/自用可选择
讯飞星火	讯飞星火的模型，免费申请，教程详见下方【讯飞星火配置方式】	sparkAI下属的appkey和appsecret	lite版免费，响应快，无代理首选
文心一言	文心的模型，免费，配置详见下方【文心模型配置方式】	wenxinAI下属的appkey和appsecret	免费，响应快，无代理可选
Gemini	谷歌Gemini，在这里申请apikey，需配置proxy或GeminiRevProxy	gemini proxy或GeminiRevProxy	免费，稳定,推荐

模型(settings.yaml中的model设置)	介绍	配置项(api.yaml对应)	评价
腾讯元器	QQ智能体同款模型，教程	腾讯元器下属的 智能体id 和 token	送1e的限额，应该够用好久了
random	免费，免费模型均收集自网络，不保证稳定性。配置random&PriorityModel以调整优先级	【无需配置】	免费，无需代理，全局代理模式下无法使用，学着用规则代理/pac吧哥
gpt3.5	官方gpt3.5，你也可以使用自己的中转，如对接kimi、豆包等大模型。	openaiSettings下属的openai-keys和openai-transit	你知道自己在干啥就行。用openai官方api需要配置proxy
Cozi	(注意，coze方案当前已被弃用)GPT4，基于coze-discord，教程请查看Here，最好配置代理	cozi proxy(建议)	不推荐。很麻烦，并且相关支持将在未来的版本中移除。需要discord小号，每个账号每天都有次数限制(gpt4 100次/天)，可配置多个小号

在使用本项目时，如果你有自己的代理，**不要开全局代理**，这将导致模型以及部分功能无法正常调用。你应当学会使用规则代理而不是只会用全局。

如果你对此有疑问，请自行了解`代理的工作原理`。

文心一言配置方式

参照 创建应用，并保存apikey和apiSecret，填入api.yaml中wenxinAI部分

(如果你没有实名认证，需要先实名一下)

开通对应模型，这里建议开通ERNIE-Speed-128K，不用担心，这是免费的。

然后把Manyana/settings.yaml中chatGLM.model修改为 文心一言 并保存，重启bot即可。

此时你的api.yaml相关部分应该是这样：

```
wenxinAI:
  apiKey: 你的key
  secretKey: 你的secretkey
  wenxin-model: ernie-speed-128k        #一般不用动
```

settings.yaml相关部分应该是这样:

```
chatGLM:                    #对话模型通用设置
  aiReplyCore: False  #ai回复核心，开不开都行
  model: 文心一言
```

讯飞星火配置方式

参照 找到spark lite，并保存apikey和apiSecret，填入api.yaml中sparkAI部分

(如果你没有实名认证，需要先实名一下)

然后**把Manyana/settings.yaml中chatGLM.model修改为 讯飞星火 并保存，重启bot即可**。

此时你的api.yaml相关部分应该是这样：

```
sparkAI:                          #讯飞星火
  apiKey: 你的apikey #在https://console.xfyun.cn/services/cbm申请，并复制对应key和
secret，注意，免费的是lite版
  apiSecret: 你的apisecret
  spark-model: general            #讯飞星火的模型设置，general是免费无限制的，一般不建
议修改
```

settings.yaml相关部分应该是这样:

```
chatGLM:                    #对话模型通用设置
  aiReplyCore: False  #ai回复核心，开不开都行
  model: 讯飞星火
```

Gemini配置方式

搭建完成后，本地环境无代理可调用Gemini。
本文档用于无域名搭建Gemini反代。

1、修改默认模型

settings.yaml相关部分应该是这样:

```
chatGLM:                    #对话模型通用设置
  aiReplyCore: False  #ai回复核心，开不开都行
  model: Gemini       #这样ai回复才会使用Gemini进行回复
```

2、获取Gemini apikey

接下来我们操作的是api.yaml

先获取Gemini apikey (获取过程需要开启代理)并填入api.yaml

```
gemini:
  - AIxxxxxxx                  #填写你申请到的apikey
```

如果你申请了多个apikey

```
gemini:
  - AIxxxxxxx1                 #填写你申请到的apikey
  - AIxxxxxxx2                 #填写你申请到的apikey
```

在获取到Gemini api后，由于gemini不支持cn用户使用，我们需要配置代理，依然是api.yaml，配置proxy或GeminiRevProxy，这两个代理任一配置完成即可使用

```
proxy: 你自己的http代理地址                          #如果你不知道这是什么，就别填，去配置GeminiRevProxy
GeminiRevProxy: https://fbsvilli.netlify.app    #这个反代地址你可以直接拿去用。
```

如果你使用 `GeminiRevProxy: https://fbsvilli.netlify.app` 并且完成了上方其他配置，那么下面的不用看了，重启bot，你已经可以使用Gemini了。

`一个需要注意的点是，在GeminiRevProxy不为""的时候，你的proxy对Gemini是无效的`

▶ 3、设置反向代理(可选)

腾讯元器配置方式

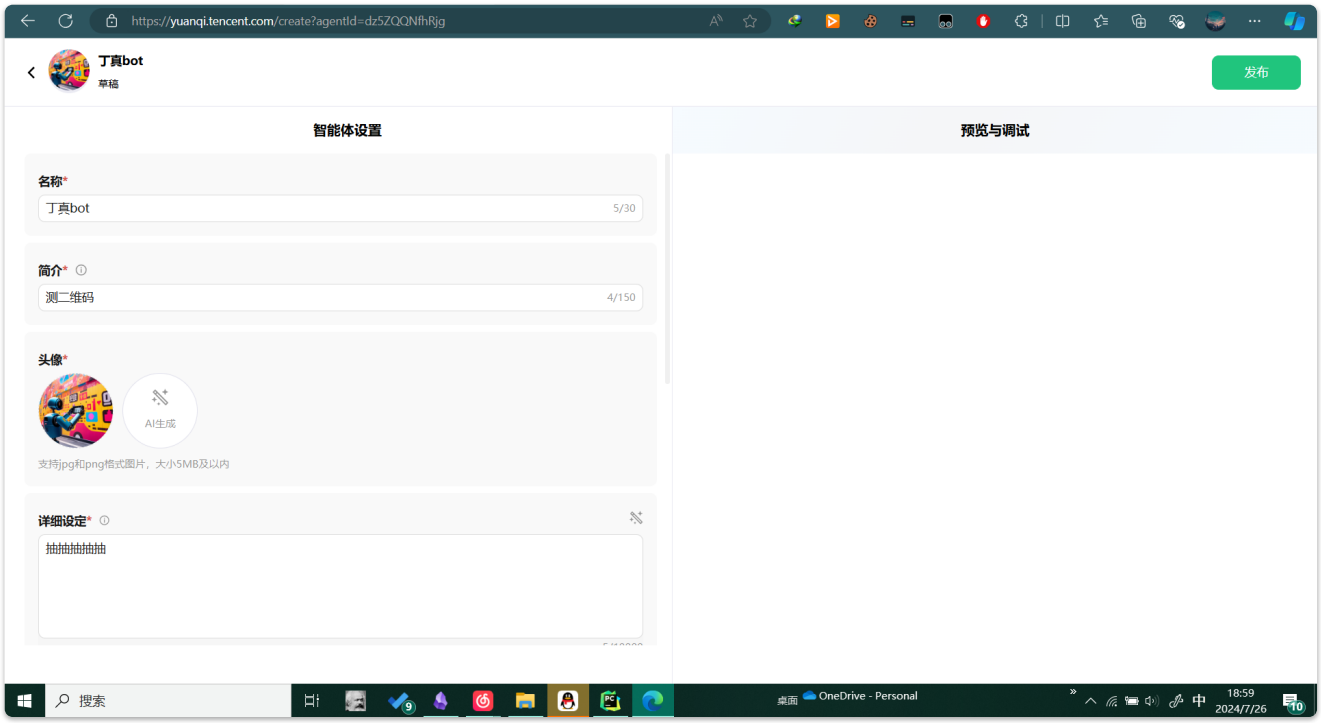
QQ最近推出了智能体，并开放了api且给了1e的额度，估计够用好久了，下面是Manyana对接QQ智能体的教程。

1.打开腾讯元器官网

官网

点击左上角 创建智能体

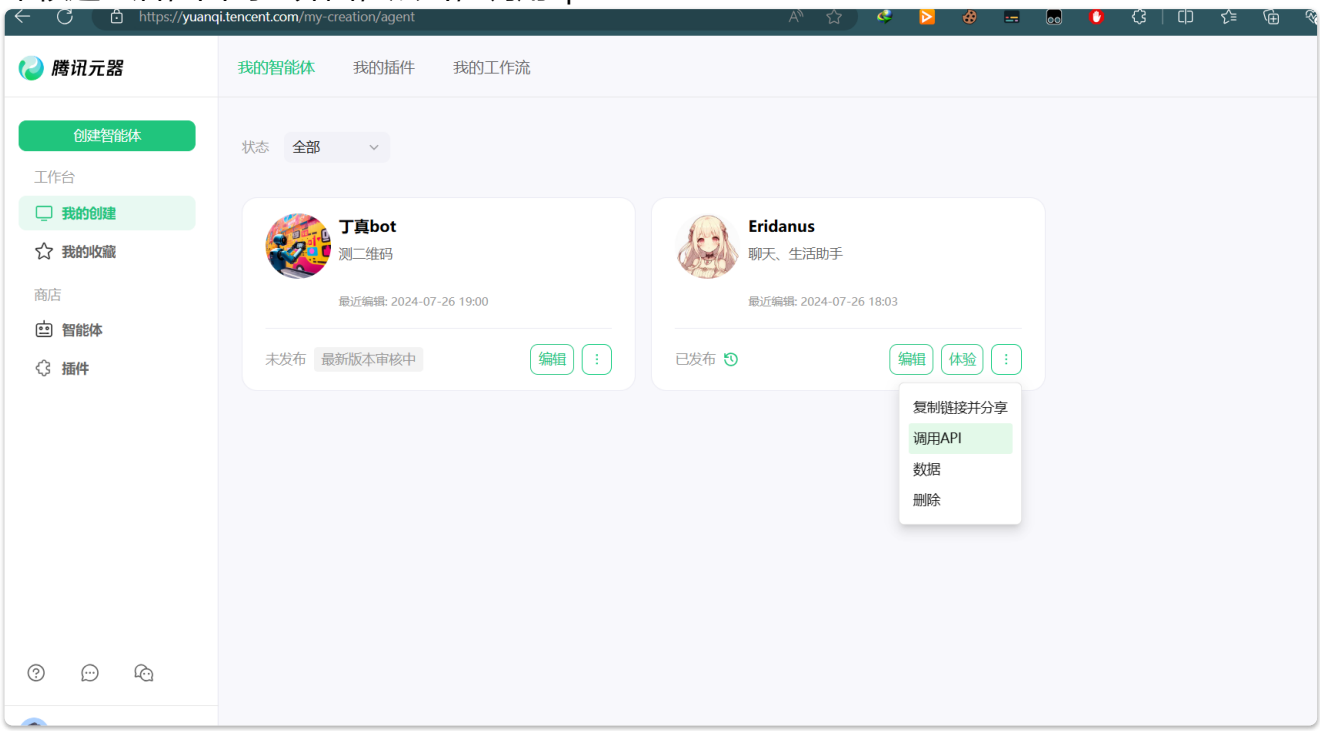
2.填写设定，然后发布



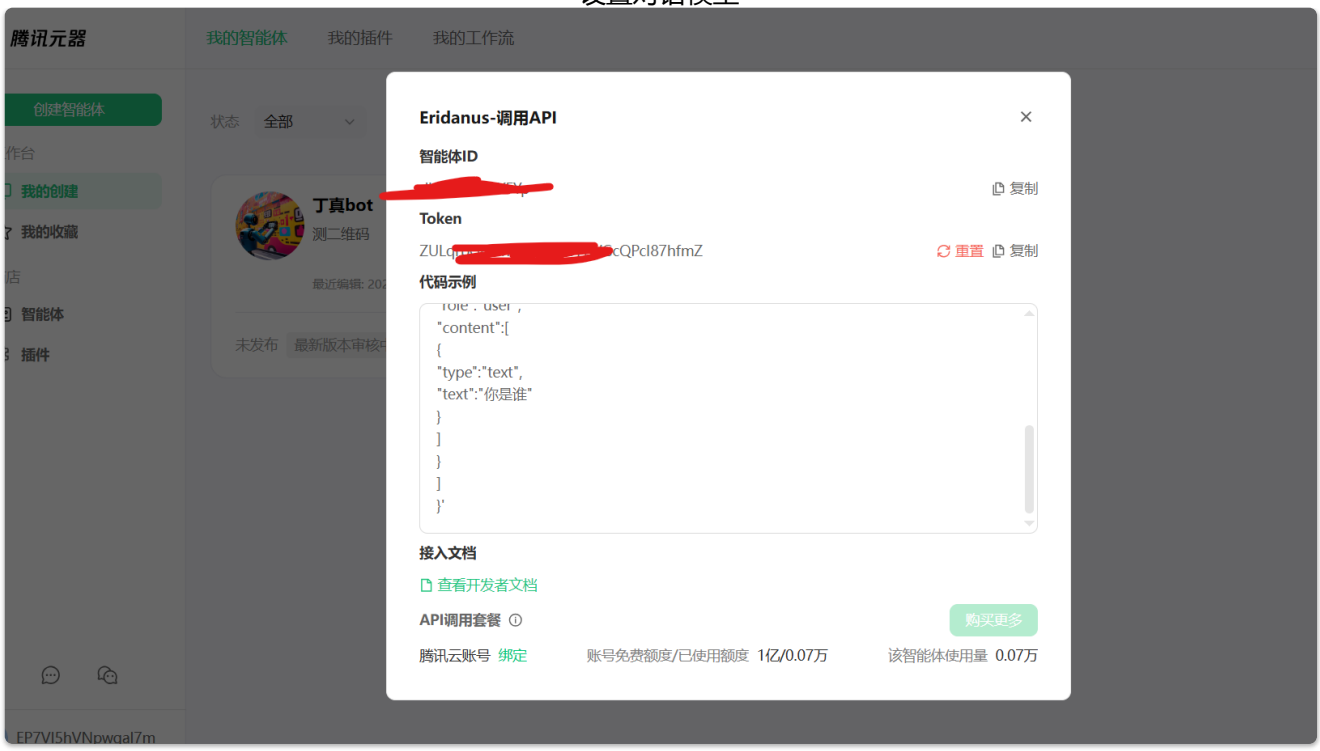
等待审核通过(半个小时左右，不一定)。

3.获取智能体Id和token

审核通过后，回到主界面，点击，调用api



设置对话模型



4.填写配置文件

复制智能体id和token，填入Manyana/config/api.yaml

```
腾讯元器: #接入QQ智能体
  智能体ID: ""
  token: ""
```

在Manyana/config/settings.yaml中，将chatGLM.model修改为 腾讯元器

```
chatGLM:
  model: 腾讯元器
```

重启bot即可

对接gpt中转(自定义模型)

大多数模型都是支持使用openai的sdk的，我们可以轻松对接像kimi这些支持openaiSDK的模型。

对接openai

使用openai官方apikey，因其不支持中国大陆用户调用，必须配置proxy。

此时你的api.yaml相关部分应该是这样：

```
openaiSettings:
  openai-keys:
  - 你申请的apikey                    #openai apikey
  openai-model: gpt-3.5-turbo         #指定一个模型
  openai-transit: ""                  #openai官方api无需填写此项
#.....其他部分省略
proxy: http://127.0.0.1:你的代理运行端口   #必须配置。
```

settings.yaml相关部分应该是这样:

```
chatGLM:                    #对话模型通用设置
  aiReplyCore: False  #ai回复核心，开不开都行
  model: gpt3.5             #因为本质上用的还是openai sdk，所以模型这里仍然需要填gpt3.5，但因为上面中转站的设置，已经是kimi了。如果需要更改人设，也是更改下方的gpt3.5，不要自己创建新的。
```

对接kimi

我们以kimi为例

通过[阅读kimi官方文档](基本信息 - Moonshot AI 开放平台),我们得到以下关键信息

OpenAI 官方 SDK 支持 Python 和 Node.js 两种语言，使用 OpenAI SDK 和 Curl 与 API 进行交互的代码如下：

python curl node.js

```python
from openai import OpenAI

client = OpenAI(
    api_key = "$MOONSHOT_API_KEY",
    base_url = "https://api.moonshot.cn/v1",
)

completion = client.chat.completions.create(
    model = "moonshot-v1-8k",
    messages = [
        {"role": "system", "content": "你是 Kimi，由 Moonshot AI 提供的人工智能助手，你更擅长中文和英文的对话。你会为用户提供安全
        {"role": "user", "content": "你好，我叫李雷，1+1等于多少？"}
    ],
    temperature = 0.3,
)

print(completion.choices[0].message.content)
```

看不懂也没关系，这说明它支持openai的sdk，我们只需要填写api.yaml即可调用它

接下来，在kimi官网申请apikey。
此时你的api.yaml相关部分应该是这样：

```
openaiSettings:
  openai-keys:
```

```
  - 你申请的apikey                                    #从kimi官网申请的api
  openai-model: moonshot-v1-8k                        #此时，我们使用kimi的模型
  openai-transit: https://api.moonshot.cn/v1     #中转站，即连接到kimi api，而非
openai
```

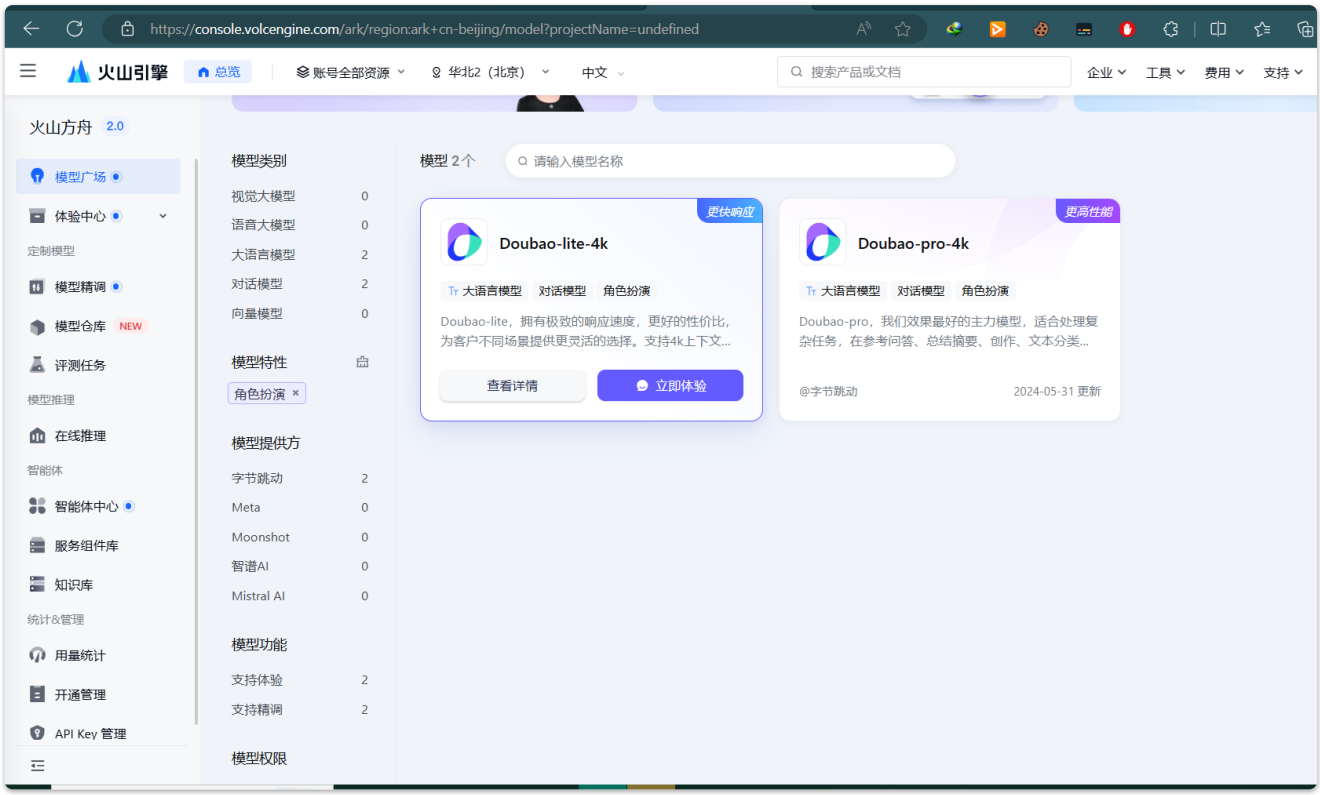
settings.yaml相关部分应该是这样:

```
chatGLM:                    #对话模型通用设置
  aiReplyCore: False  #ai回复核心，开不开都行
  model: gpt3.5             #因为本质上用的还是openai sdk，所以模型这里仍然需要填
gpt3.5，但因为上面中转站的设置，已经是kimi了。如果需要更改人设，也是更改下方的gpt3.5，
不要自己创建新的。
```

对接豆包

注册 并打开火山引擎控制台
这里以调用Doubao-lite-4K为例，当然，**你也可以选择其他模型，流程大体相似。**

1.打开模型广场/角色扮演，点击 查看详情



2.点击右侧 api调用

设置对话模型

3.点击创建推理接入点，随便填填，然后点击右侧 接入模型

设置对话模型

4.复制接入点名称

填入api.yaml，此时应当是

```
openaiSettings:
  openai-keys:
  - 你申请的apikey              #这个我们将在下一步获取
  openai-model: ep-xxxxxxx                    #你刚刚复制的接入点名称
  openai-transit: https://ark.cn-beijing.volces.com/api/v3/bots/
```

5.复制apikey

填入api.yaml，此时应当是

```
openaiSettings:
  openai-keys:
  - yourapikeyHere                                  #你刚刚复制的apikey
  openai-model: ep-xxxxxxx                          #你刚刚复制的接入点名称
  openai-transit: https://ark.cn-beijing.volces.com/api/v3
```

接着我们修改settings.yaml

```
chatGLM:                    #对话模型通用设置
  aiReplyCore: False  #ai回复核心，开不开都行
  model: gpt3.5             #因为本质上用的还是openai sdk，所以模型这里仍然需要填gpt3.5，但因为上面中转站的设置，已经是kimi了。如果需要更改人设，也是更改下方的gpt3.5，不要自己创建新的。
```

重启bot即可。

chatGLM模型设置

查阅文档即可，不再赘述