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
charactergIm	智谱的超拟人大模型,在这里 <u>申请</u>	chatGLM	付费api,群少/ 自用可选择
讯飞星火	讯飞星火的模型,免费 <u>申请</u> , 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】	wenxinAl下属的 appkey和appsecret	免费,响应 快,无代理可 选
Gemini	谷歌Gemini,在这里 <u>申请</u> <u>apikey</u> ,需配置proxy或 <u>GeminiRevProxy</u>	gemini proxy或 <u>GeminiRevProxy</u>	免费, 稳定,推 荐

	设置对话模型		
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, <u>教程</u>	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 <u>coze-discord</u> , 教程请查看 <u>Here</u> , 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wenxinAl部分

(如果你没有实名认证,需要先实名一下)

<u>开通对应模型</u>,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sparkAl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

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

gemini:

- AIxxxxxx

#填写你申请到的apikey

如果你申请了多个apikey

gemini:

- AIxxxxxx1
- AIxxxxxx2

#填写你申请到的apikey #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GeminiRevProxyGeminiRevProxy: 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

<u>官网</u> 点击左上角 创建智能体

2.填写设定,然后发布

← C	create?agentId=dz5ZQQNfhRjg		A ^N ☆	4 🔁	# =		0	\$ ∣	())	¢ @	~~	÷ 👻	·· 🍫
< 🥡 丁真bot ^{幸福}												发	布
	智能体设置					-	预览与	调试					
名称*													
丁真bot		5/30											
简介* ①													
测二维码		4/150											
头像・ を持pgfūpng檻式園片、大小SMB及以内													
详细设定* ①		+);+											
hinininini													
● 伊索	H 🛃 🍫 🍐	📑 🖲 🖺 🔁		_{桌面} 📤 Onel	Orive - Person	al			* ^	· 🦟 📾 d) <i>(</i>] =	18:59 2024/7/	26 둸

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

← C 🗗 https://yuanq	ni.tencent.com/my-creation/agent	A [™] ☆	🗳 🔰 🍪	-	0	¢	Cþ	£'≡	¢ ⊕
闷 腾讯元器	我的智能体 我的插件 我的工作流								
创建智能体工作合	状态 全部 🗸								
	J真bot	Eridanus							
高店	Ø□ 最近编辑: 2024-07-26 19:00	· · · · · · · · · · · · · · · · · · ·	8:03						
 	未发布 最新版本审核中 编辑 ::	已发布 🕲	编辑(体验						
			复制链接并 调用API	分享					
			数据删除						

腾讯元器	我的智能体	我的插件	我的工作流		
创建智能体	状态 全部	~	Eridanus-调用API		×
作台			智能体ID		
〕我的创建		Tabot			回复制
了我的收藏	275	测二维码	Token		
TF.				cQPcl87hfmZ	C 重置 🕒 复制
		最近编辑: 202	代码示例		
J 省限科		新版木宙核中	"content":[
3 插件		371122-1-122	type":"text",		
			"text":"你是谁"		
]		
			}		
			}'		
			接入文档		
			查看开发者文档		
			API调用套餐 ①		购买更多
			腾讯云账号 绑定	账号免费额度/已使用额度 1亿/0.07万	该智能体使用量 0.07万

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



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

OpenAl 官方 SDK 支持 Python 和 Node.js 两种语言,使用 OpenAl SDK 和 Curl 与 API 进行交互的代码如下: python curl node.js G 1 from openai import OpenAI 2 3 client = OpenAI(4 api_key = "\$MOONSHOT_API_KEY", 5 base_url = "https://api.moonshot.cn/v1", 6) 7 8 completion = client.chat.completions.create(model = "moonshot-v1-8k", 9 messages = [10 {"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 11 {"role": "user", "content": "你好, 我叫李雷, 1+1等于多少?"} 12 13], 14 temperature = 0.3, 15) 16 17 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
    openai-model: moonshot-v1-8k
    openai-transit: https://api.moonshot.cn/v1
    openai
```

#从kimi官网申请的api
#此时,我们使用kimi的模型
#中转站,即连接到kimi api,而非

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

\leftarrow C \textcircled{D} https://	/console.volcengine.com/ark/regi	on:ark+cn-beijing/model?projectName=undefined	A 🏠 🧳 🔀 🏶 🚍	0 ¢ Ф ¢	œ
😑 🔥 火山引擎	▲ 意览 ● 账号全部资源	× ② 华北2 (北京) × 中文 ∽	Q. 搜索产品或文档 企业	▼ 工具 ▼ 費用 ▼ 支	持~
火山方舟 2.0					
	模型类别	模型 2 个 Q 请输入模型名称			
■ 体验中心 ● ✓	1000人模型 0 语音大模型 0	更快响应 Doubao-lite-4k	更高性能 Doubao-pro-4k		
₩ 模型精调 ●	大语言模型 2 对话模型 2	▼ Tr 大语言模型 对话模型 角色扮演	✓ ↓ 大语言模型 对话模型 角色扮演		
● 模型仓库 NEW	向量模型 0	Doubao-lite,拥有极致的响应速度,更好的性价比, 为客户不同场景提供更灵活的选择。支持4k上下文…	Doubao-pro,我们效果最好的主力模型,适合处理复 杂任务,在参考问答、总结摘要、创作、文本分类…		
▲ 评测任务	模型特性	查看详情	@字节跳动 2024-05-31 更新		
🏛 在线推理	模型提供方				
智能体	字节跳动 2 Meta 0				
📻 服务组件库	Moonshot 0 智谱AI 0				
知识库 统计&管理	Mistral Al 0				
🕥 用量统计	模型功能				
开通管理	支持精调 2				
	模型权限				

2.点击右侧 api调用



3.点击创建推理接入点,随便填填,然后点击右侧 接入模型

基本信息		费用细则
* 接入点名称	test1	接入模型 模型广场/doubao-lite-4k/240328
接入点描述	16161616616 5/300 g	为天儿五 按Unemy要 计器方式 后付要 结道周期 小时
接入配置		费用预估 输入价格 0.0003元/干tokens
* 接入模型	Doubao-lite-4k/240328 更饶模型	输出价格 0.0006元/千tokens *具体费用以账单为准
* 购买方式	按Token付费 按概实际消耗Token最后付费,更加灵活,但可达到的并发上限较低。	接入模型
模型限频 ⑦	10000 RPM 800000 TPM 当前该账号下访问 doubao-lite-4k 的模型频率限制	
接入点跟额	设置单接入点访问频率限制	

4.复制接入点名称

😑 🔥 火山引撃	↑ 总览 參账号全部资源 >	② 华北2 (北京) → 中	文 ~ Q	搜索产品或文档	
火山方舟 2.0 ・ ・ ・ ・ ・ ・ ・ ・ ・	在线推理 ③ 由安全沙箱守护 提供实时的模型推理服务,通过推 + 创建推理接入点 全部	理接入点灵活调整资源并访问相 3 我创建的	莫型,可通过监控查看运行状况		
定制模型	接入点名称	状态 17	接入模型 9	购买方式	创建时
 Ⅰ Ⅰ ● ●	test1 ep-2024cm	2.复制接入点名	Doubao-lite-4k 240328 3称 ^{模型广场}	按Token付费	2024-0
👗 评测任务			¢		
模型推理			à		
1 在线推理	1.打开在线推理页面				
智能体					
₿ 智能体中心 •					
■ 服务组件库					
🔚 知识库					
统计&管理					
🥡 用量统计					

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			
<pre>openai-transit: https://ark.cn-beijing.volces.com/api/v3/bots/</pre>				

		设置对	讨话模型				
	API Key 是您请求火山方舟大模型服务的重要先证。API Key 长期有效,请您不要将密钥信息共享至公开环境,妥善保管并定期轮换密钥,避免因未经授权的使用造成安全风险或资金损失。						
■ 体型中心 ● ◆	当前恋住 至即资源 视图下,为了恋	的数据女主,仅展示 Default (款认项目)下的 API Key, 您可通过平台	在工用切换项目亘着其他项目下的 API Key	6		
₩ 模型精调 ●	+ 创建 API Key						
● 模型仓库 NEW	名称	API Key	复制	创建时间 🔶	权限		
👗 评测任务	api-key-20240811004222	3a8	06e 🗇 🗞	2024-08-11 00:42	all		
模型推理						< [1]	
▲ 在线推理			2.复制apikey, 浏	没有你就创建一个。			
智能体							
🚼 智能体中心 🔍							
■ 服务组件库							
🔚 知识库							
统计&管理							
🕥 用量统计							
■ 开通管理							
😵 API Key 管理	1.找到apikey管理						
Ē							

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