

OpenSquilla 使用向导

1、下载 opensquilla

连接地址：[OpenSquilla - 提升单位成本的 Agent 智能](#)

目前支持 windows、mac 系统用户安装图形化用户界面



2、安装与配置

- 下载安装包后，双击下一步即可安装；
- 安装完成后，初次进入会开始新手配置引导

2.1 模式

选择配置模式：简单配置 / 详细配置

高级设置：比简单设置多一项关于“路由”的配置引导。

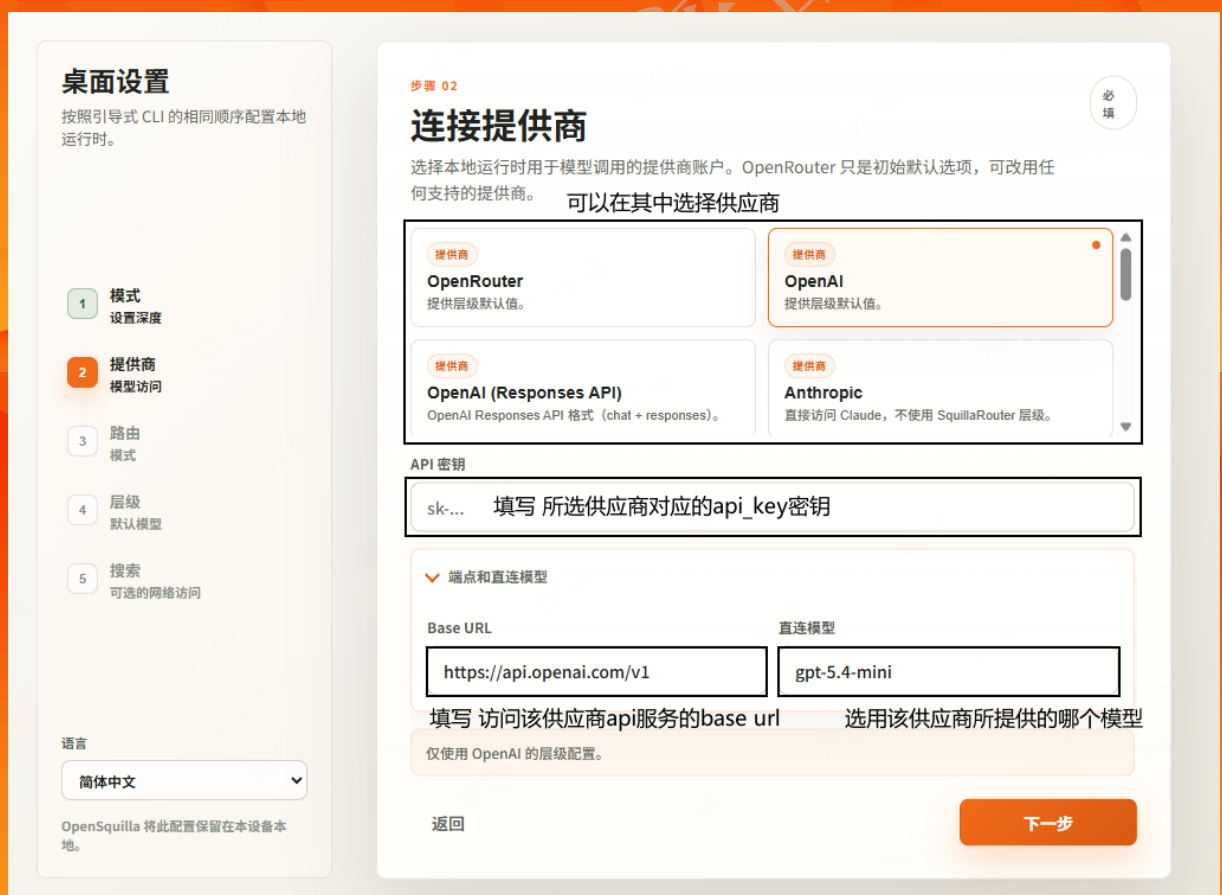
备注：以下内容后续可以在“设置”中重新配置，填写如遗漏或错误不用担心。



2.2 供应商

- 选择供应商
- Base URL：填写访问该供应商 api 服务的 base url
- 直连模型：填写使用该供应商所提供的哪个模型

2.2.1 官方供应商



2.2.2 使用 Ollama 自行部署的模型服务

桌面设置

按照引导式 CLI 的相同顺序配置本地运行时。

1 模式
设置深度

2 提供商
模型访问

3 路由
模式

4 搜索
可选的网络访问

语言
简体中文

OpenSquilla 将此配置保留在本设备本地。

步骤 02

必填

连接提供商

选择本地运行时用于模型调用的提供商账户。OpenRouter 只是初始默认选项，可改用任何支持的提供商。

提供商
Google Gemini
提供层级默认值。

提供商
Moonshot AI
提供层级默认值。

提供商
Ollama (local)
本地直连模型路径。

提供商
Baidu Qianfan
需要填写直连 provider 模型 ID。

个人部署的模型服务

API 密钥
sk-... 如Ollama部署模型服务时，没有设置密钥，则无需填写

端点和直连模型

Base URL
http://localhost:11434

直连模型
部署模型服务时所设置的模型名

访问该服务的url地址
本地直连模型路径。

返回

下一步

2.3 路由（高级设置）

是否开启“智能路由”功能

- 开启后，会根据用户提问、选择合适的模型作答，能够确保质量与成本
- 关闭时，会使用“直连模型”配置的指定模型

桌面设置

按照引导式 CLI 的相同顺序配置本地运行时。

1 模式
设置深度

2 提供商
模型访问

3 路由
模式

4 层级
默认模型

5 搜索
可选的网络访问

语言
简体中文

OpenSquilla 将此配置保留在本设备本地。

步骤 03

高级

选择路由模式

选择 OpenSquilla 使用 Smart Router 层级、直连一个固定模型，还是使用 OpenRouter Ensemble。

Smart Router
使用此提供商现有的 Squilla Router 层级默认值。

直连单模型
每个请求都发送到一个固定模型，不使用层级路由或 Ensemble。

Ensemble
当前 onboarding 中 Ensemble 需要 OpenRouter。

开启“模型路由”功能，可自定义路由配置

选用OpenRouter作为供应商
可以选择“Ensemble”默认配置

返回

下一步

2.4 层级（即模型池配置，可选项）

- 如. 开启智能路由

在此配置模型层级（可选，可以用默认配置，不改就能用）：

当“智能路由”功能开启时，Opensquilla 会根据用户问题、选择合适的模型作答；更难的问题会选择更高的 T 级；

建议：C0 ~> C3，按照 模型能力递增 来配置；

The image displays two screenshots of the OpenSquilla configuration interface, specifically the 'Candidate Model Pool' (候选模型池) section. The interface is in Chinese and shows a sidebar with navigation options: 模式 (Mode), 设置深度 (Set Depth), 提供商 (Provider), 路由 (Routing), 层级 (Level), and 搜索 (Search). The main content area is titled '候选模型池' and includes a subtitle '选择默认文本层级并保留 CLI 默认值，或在启动前自定义模型 id。' (Select the default text level and retain the CLI default value, or customize the model ID before startup). The interface shows a list of model levels (C0, C1, C2, C3) with their respective providers and models. In the top screenshot, C2 is selected as the default. In the bottom screenshot, the '自定义模型' (Custom Model) section is expanded, showing a table for configuring models for each level. The table has columns for '提供商' (Provider) and '模型' (Model). The bottom screenshot also shows a checkbox for '路由到该层级时，使用哪个供应商的哪个模型' (When routing to this level, use which provider's which model).

Level	Provider	Model
C0	deepseek	deepseek-v4-flash
C1	deepseek	deepseek-v4-pro
C2	glm	glm-5.2
C3	claude	claude-opus-4.5

Level	Provider	Model
C0	openrouter	deepseek/deepseek-v4-flash
C1	openrouter	deepseek/deepseek-v4-pro
C2	openrouter	z-ai/glm-5.2
C3	openrouter	anthropic/claude-opus-4.8

2.5 搜索

指定模型使用哪个搜索引擎

其中 DuckDuckGo 是免费的（但是服务不稳定），其余需要去相应的官网注册账户、填

写 API 密钥

桌面设置

按照引导式 CLI 的相同顺序配置本地运行时。

1 模式
设置深度

2 提供商
模型访问

3 路由
模式

4 层级
默认模型

5 搜索
可选的网络访问

语言
简体中文

OpenSquilla 将此配置保留在本设备本地。

步骤 05 可选

选择网络搜索

搜索为可选项。可以不添加其他密钥直接开始，或连接运行时支持的搜索提供商。

DuckDuckGo
No key required. Good default for getting started.

Bocha
Web search with inline summaries and freshness support.

Brave Search
Managed search access with freshness support.

Tavily
Freshness-oriented web search for current research.

Exa
Semantic and content-oriented search for research workflows.

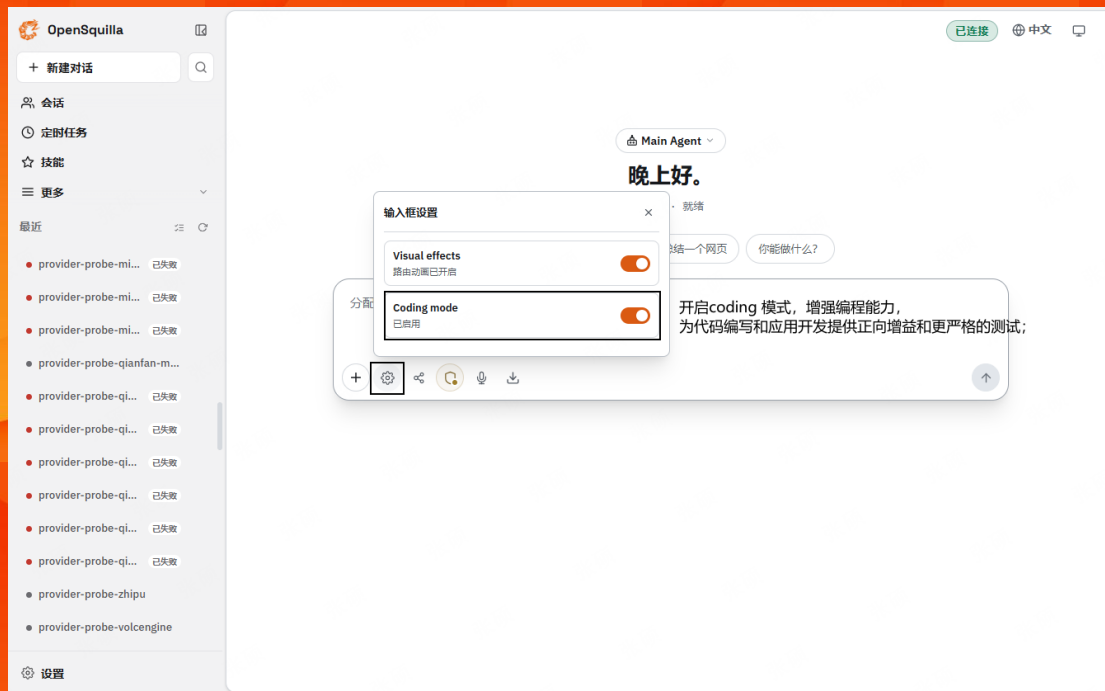
搜索 API 密钥
BOCHA_SEARCH_API_KEY
Web search with inline summaries and freshness support.

返回 启动 OpenSquilla

3、使用

3.1 Coding mode（编程模式）

Coding 模式是专注写代码、软件的模式，在独立的隔离空间写代码、补测试、跑验证。



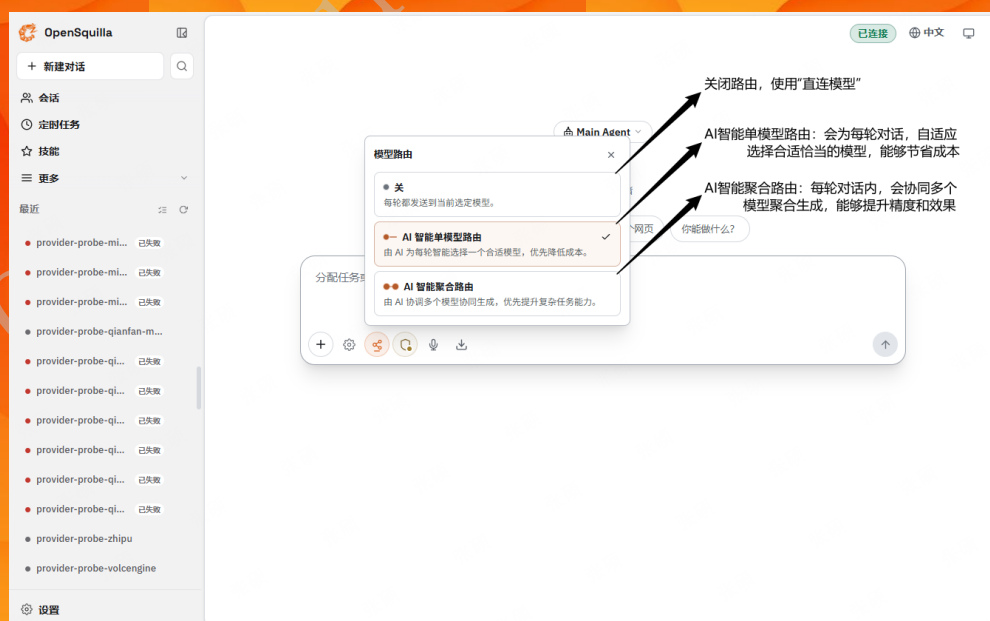
3.2 模型路由

三个选项:

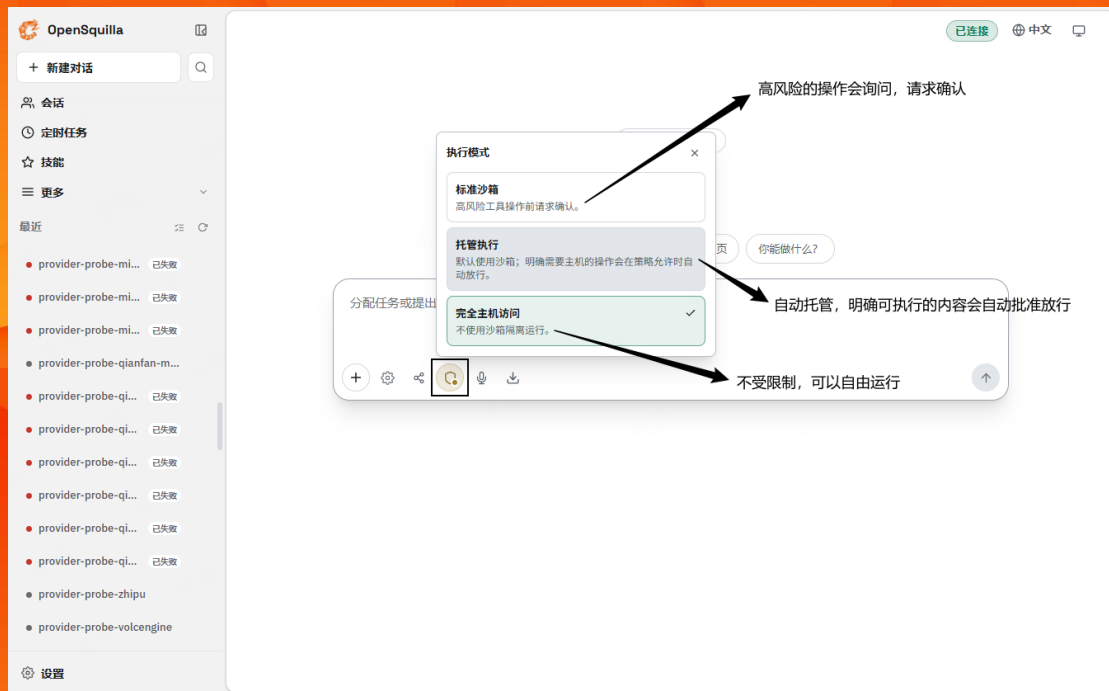
关：关闭路由，使用“直连单模型”

AI 智能单模型路由：每轮对话，自适应选择合适的模型，能够省成本

AI 智能聚合路由：每轮对话，协同多个模型聚合生成，能够提升精度和效果



3.3 执行模式



4、卸载



退出后，在操作系统中卸载应用；