

WeAgentChat 配置指南

本指南将帮助你完成 WeAgentChat 的核心配置。以下配置值基于系统的**最佳默认配置**。

提示：请保持默认配置以获得最佳体验，仅在明确需要时进行修改。

0. 注册硅基流动

- 点击链接: <https://cloud.siliconflow.cn/i/jD3fSdmD>，注册硅基流动账号
- 进行实名认证，会赠送 2000 万 Tokens。
- 进入“API 密钥”，点击“新增 API 密钥”，并复制密钥。

你也可以选择其他平台，硅基流动可以一站式搞定很多模型。

1. 访问设置界面

- 下载并安装 WeAgentChat: <https://github.com/WeAgentChat/WeAgentChat/releases>
- 启动 WeAgentChat 应用程序。
- 在左侧侧边栏底部，点击「设置」图标（通常是一个齿轮形状）。
- 设置弹窗将打开，包含「LLM 设置」、「向量化设置」、「记忆设置」和「聊天设置」。

2. LLM 设置 (大语言模型)

后端默认配置如下：

字段名称	默认值 / 最佳配置	说明
API Base URL	https://api.siliconflow.cn/v1/	使用官方 OpenAI 时留空。如果使用中转服务，请填写对应地址。
API Key		请填写硅基流动的 API 密钥。
Model Name	Pro/zai-org/GLM-4.7	基础对话模型。你也可以选择别大模型，但是大模型必须支持函数调用。

3. 向量化设置 (Embedding)

字段名称	默认值 / 最佳配置	说明
Provider	openai	默认提供商协议。
API Base URL	https://api.siliconflow.cn/v1	Embedding 服务地址。
API Key		硅基流动的密钥。
Model Name	BAAI/bge-m3	系统优选的向量模型。

字段名称	默认值 / 最佳配置	说明
Dimension	1024	必须与模型匹配。系统默认采用 1024 维。
Max Token Size	8000	单次处理的最大长度。

4. 记忆管理设置

系统默认启用了最佳的记忆召回参数。

基础配置

- 会话过期时间: 30 分钟。
(超过此时间未交互，系统将自动归档会话并生成记忆)

召回设置 (Recall)

字段名称	默认值 / 最佳配置	说明
启用记忆召回	开启 (True)	允许 AI 参考历史记忆。
相似度阈值	0.5	语义匹配的最低门槛。
搜索轮数	3	记忆检索的深度。
事件召回 TopK	5	单次召回的记忆条数。

注意事项

- 保存配置: 修改任何内容后，务必点击右下角的「保存」按钮。
- 测试连接: 在保存前，建议点击「测试」按钮确保连接畅通。