

VenusAI 人工智能开放平台

DeepSeek 大模型

(学科知识库增强版)

使用指南

中国科学院计算机网络信息中心

人工智能发展部

2025年2月19日

1、背景概述

中国科学院计算机网络信息中心的人工智能部基于国产异构智能基础软硬件，满足材料计算、生命科学、金融、能源等学科科研需求，构建了服务多学科交叉融合的人工智能开放平台。2023年6月，该人工智能平台获批科技部“国家新一代开放算力创新平台（筹）”，标志着平台在国家科技创新体系中地位进一步提升，支撑国家人工智能前沿研究和产业化应用。

人工智能开放平台全新上线：**基于 AnythingLLM 框架的 DeepSeek-R1 思维链大模型（学科知识库增强版）**，结合深度学习与智能推理，精准支持自定义学科领域知识库，优化跨学科知识整合与分析，显著提高推理精度与响应速度。本手册旨在指导科研人员高效使用该功能。

2、主要功能

- **自定义学科领域知识库：**支持多学科融合，用户可上传各类研究数据、实验成果、文献资料（支持 PDF, WORD, EXCEL, PPT 和文本等格式），系统自动构建嵌入向量，形成向量库。
- **深度语义理解与推理：**基于 DeepSeek-R1 等大模型，实现复杂科研问题的智能解析，提供前沿领域的精准解读和推理能力，支持多轮会话。
- **智能化学科知识管理：**面向科研人员，通过新建目录/标签，支持建立、优化、管理多层次、多领域的智能知识库，构建专属学术知识体系。
- **高级知识库权限管理：**支持团队协作，实现知识库在团队内部共享；提供多层次的访问权限配置，确保科研数据安全与合规。
- **多模型支持：**目前对接 DeepSeek-R1（32B），可支持 DeepSeek-R1（671B）、DeepSeek-V3（671B）多版本。另外，支持 LLAMA、Qwen 等多种大模型接入，可灵活切换以适应不同科研需求。目前算力资源有限，如有大量需求，可租用更多算力资源。

3、使用流程

3.1 注册与登陆

推荐使用 Chrome 浏览器访问人工智能平台：<http://data.aicnic.cn/>



如果有**中国科技云通行证**，直接使用中国科技云通行证登录。如果没有账户，请点击右上角的登陆按钮，进入注册页面，填写好个人信息后进行注册登录。

The image shows the login page of the AI Open Platform. At the top, there is a header with the text '中国科技云通行证' and a '注册' button. The main content area contains a message: '您正在使用中国科技云通行证登录 人工智能数据对外开放平台，一键通行更轻松'. Below this message, there are two input fields: '账号' (Account) with a user icon and the placeholder text '邮箱/手机号/用户名', and '密码' (Password) with a lock icon and the placeholder text '密码'. To the right of the password field, there is a red error message '密码不能为空'. At the bottom of the form, there is a blue '登录' (Login) button.

3.2 试用申请

登录成功后,在首页左上角/首页横幅/个人中心均可进入 DeepSeek 对话入口,第一次使用的用户会显示“点击此处申请试用”。



按照弹出的页面,填写相关申请信息,提交即可。



此时，状态会变为“审核中，请稍后”，等待状态变为“前往使用”，即可开始探索使用 DeepSeek（学科知识库）增强版的功能。

DeepSeekR1大模型（学科知识库增强版）已上线! 审核中，请稍后

平台全新上线DeepSeek思维链大模型，重点支持自定义学科知识库功能。该模型结合智能推理与深度学习，优化知识检索、逻辑推演和跨学科整合能力，显著提升知识库的智能化水平与响应速度，为科研领域提供更精准高效的知识服务。

Hi 欢迎使用AI+人工智能科学开放平台,构建您的AI世界

AI+致力于协助AI领域的科研工作者，将数据、模型、训练和算力资源管理等功能深度整合，打造即开即用的交互式云端开发环境，应用到教研、论文复现、科学数据集和模型研究等多个场景。

[探索科学数据](#) [探索科学模型](#) [探索科学论文](#) [探索开发环境](#) [大规模计算](#) [开源大赛](#)

DeepSeekR1大模型（学科知识库增强版）已上线! NEW 前往使用

平台全新上线DeepSeek思维链大模型，重点支持自定义学科知识库功能。该模型结合智能推理与深度学习，优化知识检索、逻辑推演和跨学科整合能力，显著提升知识库的智能化水平与响应速度，为科研领域提供更精准高效的知识服务。

Hi 欢迎使用AI+人工智能科学开放平台,构建您的AI世界

AI+致力于协助AI领域的科研工作者，将数据、模型、训练和算力资源管理等功能深度整合，打造即开即用的交互式云端开发环境，应用到教研、论文复现、科学数据集和模型研究等多个场景。

[探索科学数据](#) [探索科学模型](#) [探索科学论文](#) [探索开发环境](#) [大规模计算](#) [开源大赛](#)

3.3 学科知识库功能使用教程

在左侧以账号命名的 workspace 区域，点击 “创建知识库”。

人工智能
DeepSeek

新工作区

wanmeng

default

Thread

+ New Thread

- 使用随机生成的温度数据作为示例（实际应用中请替换为真实气象数据）。
- 数据预处理包括归一化和滑动窗口处理，以便 LSTM 网络可以学习时间序列模式。

02. 模型构建:

- LSTM 层包含 50 个神经元，输入形状为 [batch_size, time_steps, features]。
- 输出层使用 Dense(1) 进行回归预测。

03. 训练与预测:

- 使用均方差 (MSE) 作为损失函数，Adam 优化器进行训练。
- 预测结果通过反归一化还原到原始数据范围。

04. 可视化:

- 绘制实际值和预测值的对比图，方便观察模型性能。

注意事项:

- 如果使用真实气象数据，请确保数据格式正确，并替换示例数据部分。
- 可以调整 look_back (时间步长度)、神经元数量、训练轮数等超参数以优化模型性能。
- LSTM 对于时间序列问题表现出色，但需要足够的数据和合理的特征工程。

:

Show Citations >

上传文献、论文、实验数据等相关材料，点击“创建”，系统自动构建嵌入数

据库。



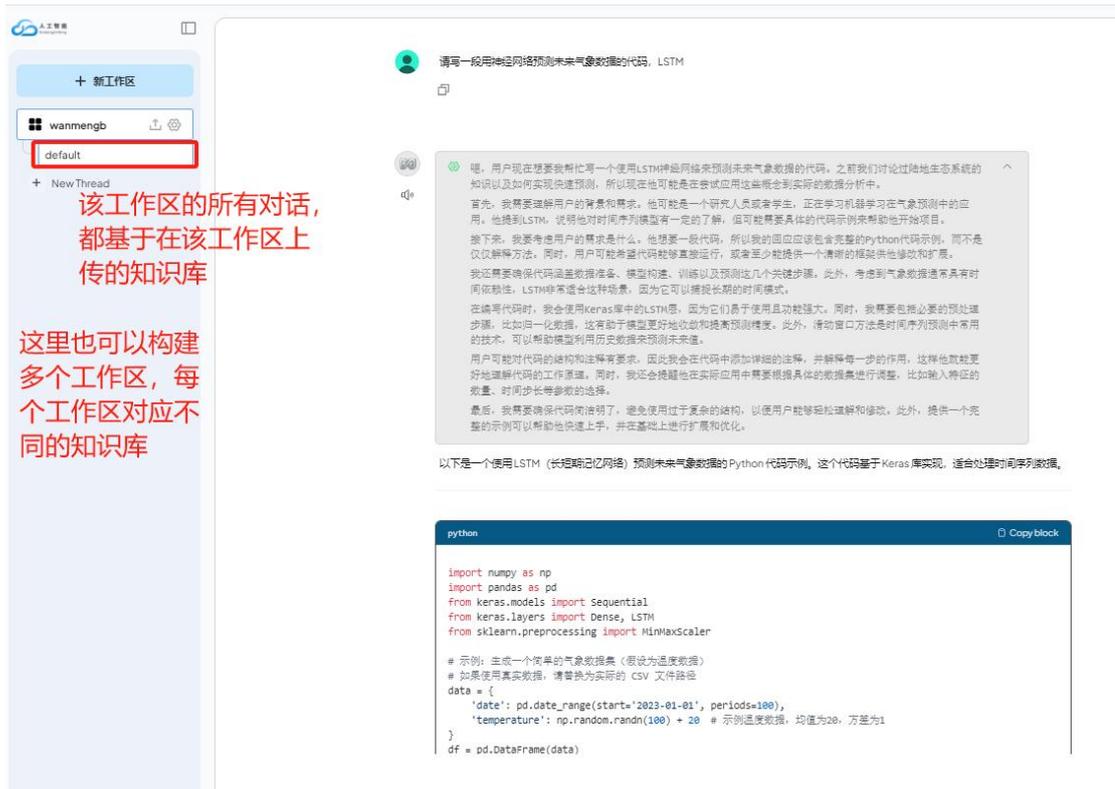
通过对每个用户空间（workspace）的个性化构建，可以实现模型对特定领域知识的增量学习和特定化知识融合检索，并基于此扩展创新。

例如人工智能视觉知识库（上传 200 篇视觉论文、100 个视觉技术文档、50 个视觉代码仓库、20 个技术网页）；

例如生态环境知识库（上传 100 篇生态论文、50 个生态数据集、20 个技术报告）；

3.4 DeepSeek 对话使用教程

在上一小节已经构建了领域特定的知识库，这里以生态环境知识库为例，我已经上传了 30 篇生态环境的论文到 wanmeng 的工作空间。该空间下的所有会话，都会基于该知识库进行知识增强，模型能够更懂这个学科领域的个性化知识，回答出更偏向的答案。（更多偏好值、推理速度等参数设置，可以自行探索）



3.5 常见问题

1、知识库无法解析上传的文件？

确保文件格式为 PDF、TXT、CSV 或 JSON。

确保文档结构清晰，避免特殊字符。

2、推理结果不精准？

检查知识库数据是否全面，尝试补充更多数据。

联系我们申请参数量更大的高级版本。

3、系统运行/回答问题速度缓慢？

由于目前是基于 Nvidia P100 部署的推理版本，用户量较多，请耐心等待，也可以联系我们申请参数量更大更快的 A100 高级版本。

4、如何批量导入知识库数据？

使用批量上传功能，直接全选拖拽即可。

确保数据格式正确，以免解析失败。

5、如何删除或清空知识库？

进入 [知识库管理](#) 页面，找到目标知识库，选择 [删除](#)。

请注意，删除后数据无法恢复。

3.6 更多扩展功能

您可能希望更改新文档在插入到矢量数据库之前拆分和分块的默认方式。只有在了解文本拆分的工作原理及其副作用时，才应修改此设置。



API 密钥允许持有者以编程方式访问和管理此 AnythingLLM 实例。

