

**VenusAI 人工智能开放平台**

**DeepSeek 大模型**

**(学科知识库增强版)**

**学科知识库使用指南**

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

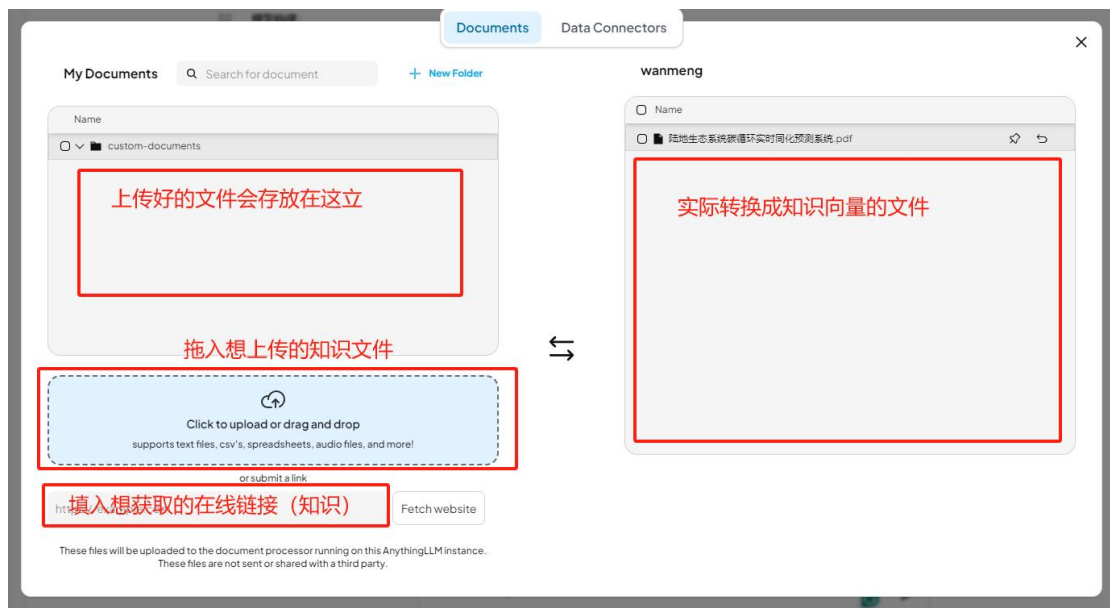
- **自定义学科领域知识库：**支持多学科融合，用户可上传各类研究数据、实验成果、文献资料（支持 PDF, WORD, EXCEL, PPT 和文本等格式），系统自动构建嵌入向量，形成向量库。
- **深度语义理解与推理：**基于 DeepSeek-R1 等大模型，实现复杂科研问题的智能解析，提供前沿领域的精准解读和推理能力，支持多轮会话。
- **智能化学科知识管理：**面向科研人员，通过新建目录/标签，支持建立、优化、管理多层次、多领域的智能知识库，构建专属学术知识体系。
- **高级知识库权限管理：**支持团队协作，实现知识库在团队内部共享；提供多层次的访问权限配置，确保科研数据安全与合规。
- **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。

### 3.3 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”。



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）

新工作区  
 wanmengb  
 default  
 + New Thread

该工作区的所有对话，  
 都基于在该工作区上  
 传的知识库

这里也可以构建  
 多个工作区，每个  
 工作区对应不同  
 的知识库

请写一段用神经网络预测未来气象数据的代码。LSTM

嗯，用户现在想要我帮忙写一个使用LSTM神经网络来预测未来气象数据的代码。之前我们讨论过陆地生态系统的知识以及如何实现快速预测，所以现在他可能是在尝试应用这些概念到实际的数据分析中。  
 首先，我需要理解用户的需求和背景。他可能是一个研究人员或者学生，正在学习机器学习在气象预测中的应用。他提到LSTM，说明他对时间序列模型有一定的了解，但可能需要具体的代码示例来帮助他开始项目。  
 接下来，我要考虑用户的需求是什么。他想要一段代码，所以我的回答应该包含完整的Python代码示例，而不是仅仅解释方法。同时，用户可能希望代码能够直接运行，或者至少能提供一个清晰的框架供他修改和扩展。  
 我还需要确保代码包含数据准备、模型构建、训练以及预测这几个关键步骤。此外，考虑到气象数据通常具有时间相关性，LSTM非常适合这种场景，因为它可以捕捉长期的时间模式。  
 在编写代码时，我会使用Keras库中的LSTM层，因为它们易于使用且功能强大。同时，我还需要包含必要的预处理步骤，比如归一化数据，这有助于模型更好地收敛和提高预测精度。此外，滑动窗口方法是时间序列预测中常用的技术，可以帮助模型利用历史数据来预测未来。  
 用户可能对代码的结构和注释有要求，因此我会在代码中添加详细的注释，并解释每一步的作用，这样他就能更好地理解代码的工作原理。同时，我还会提醒他在实际应用中需要根据具体的数据集进行调整，比如输入特征的维度、时间步长等参数的选择。  
 最后，我要确保代码清晰明了，避免使用过于复杂的结构，以便用户能够轻松理解和修改。此外，提供一个完整的示例可以帮助他快速上手，并在基础上进行扩展和优化。

以下是一个使用 LSTM (长短期记忆网络) 预测未来气象数据的 Python 代码示例。这个代码基于 Keras 库实现，适合处理时间序列数据。

```

python Copy block

import numpy as np
import pandas as pd
from keras.models import Sequential
from keras.layers import Dense, LSTM
from sklearn.preprocessing import MinMaxScaler

# 示例：生成一个简单的气象数据集（假设为温度数据）
# 如果使用真实数据，请替换为实际的 CSV 文件路径
data = {
    'date': pd.date_range(start='2023-01-01', periods=100),
    'temperature': np.random.randn(100) * 20 # 示例温度数据，均值为20，方差为1
}
df = pd.DataFrame(data)
  
```

## 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。