



中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences



中国科技云MaaS平台

大模型API开放平台、AI工作台

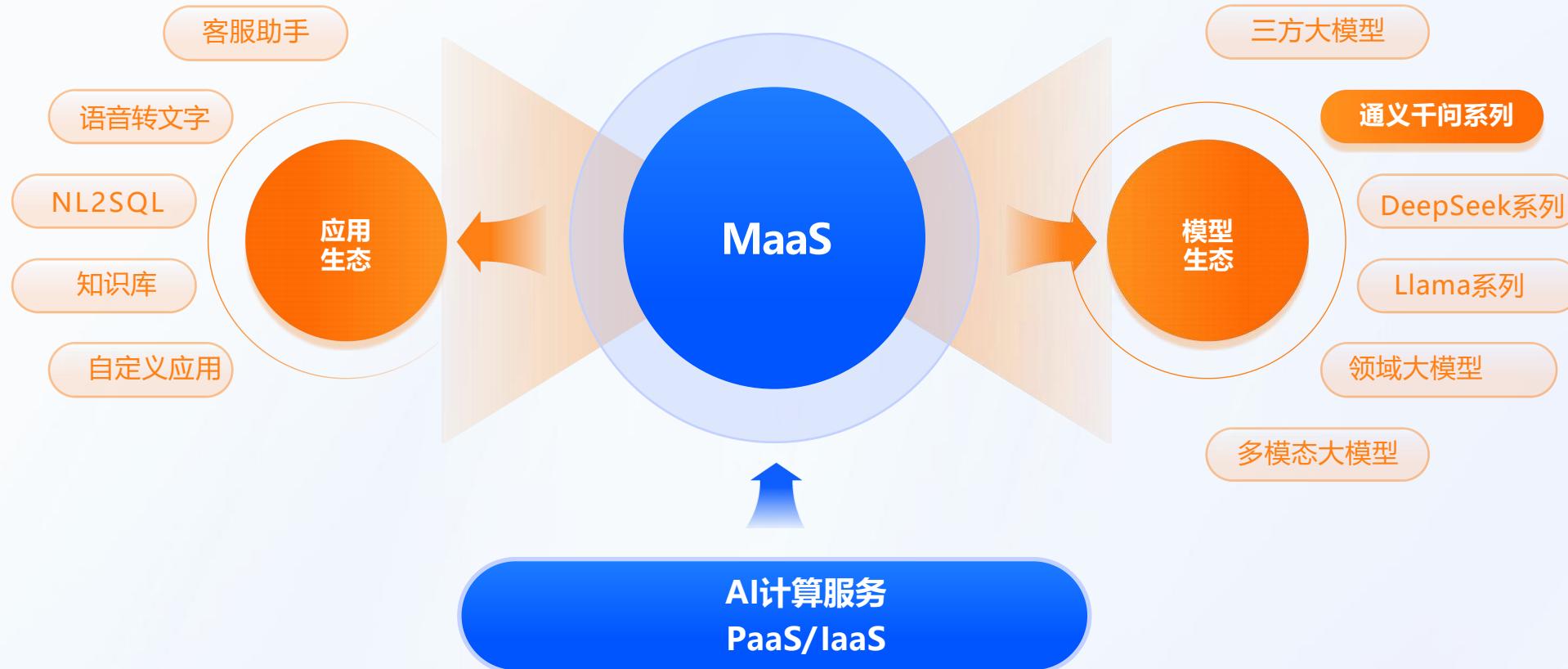
中国科学院计算机网络信息中心

裴祖华

2025年6月10日

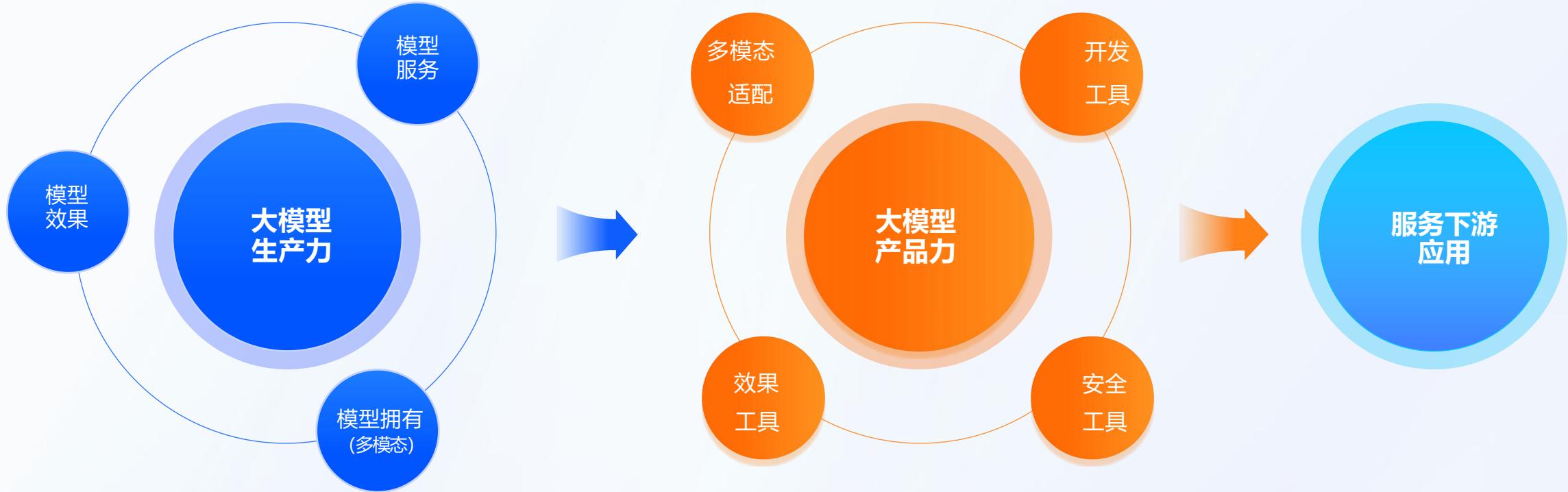
MaaS平台定位

基于大模型，面向单位和个人开发者用户，打造一站式大模型服务和大模型应用构建平台



大模型服务及应用已逐步进入“深水区”

用户对效果需求带来对大模型平台能力，从要求**生产力**到**产品力**的转变

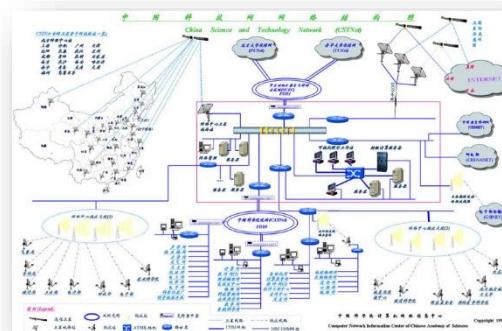


始终坚守服务科研创新的初心，不断前行

30年前实现中国全功能接入国际互联网，这标志着中国互联网时代的开端。



中国科技网
China Science & Technology Network



1994

下一个30年，奋楫扬帆，勇立时代潮头

从“网”到“云”演进



逐步实现“云网融合”

正在迈向“云智网融合”



中国科技云
China Science & Technology Cloud



2018

中国科技云本地部署DeepSeek，提供三类服务

构建API开放平台助力300+项目推进，研发对话服务，上线学术助手、团队知识库

AI 应用层

智能化公文写作与审查系统
攻坚专项研发实施进度和风险
监控智能绩效评估系统

管理大数据中心 继续教育平台
网上办公平台 先导平台智能化
ARP平台智能化提升 ScienceDB
版纳植物园 植物DNA数据库 大化所
微生物数据库 水生所 空间中心等300+项目

对话服务
团队知识库
学术助手

AAI
统
一
身
份
认
证

AI 能力层



Deepseek R1和V3 QwQ 星火

操作系统 + CUDA + 驱动 + Ollama + vLLM + SGLang

科技云镜
代码仓库

资源层

中国科技云一体化云平台 / 统一管理、统一调度、统一监控与告警

英伟达
16块 H100

英伟达
3090/4090/H20

华为
910B

曙光
K100

阿里
PPU

高速、安全、稳定的中国科技网

基于本地部署大模型，中国科技云提供一体化服务

针对不同科研应用场景，定制本地化部署和应用方案

01

API 接口定制开发

02

模型微调

03

RAG 知识库

04

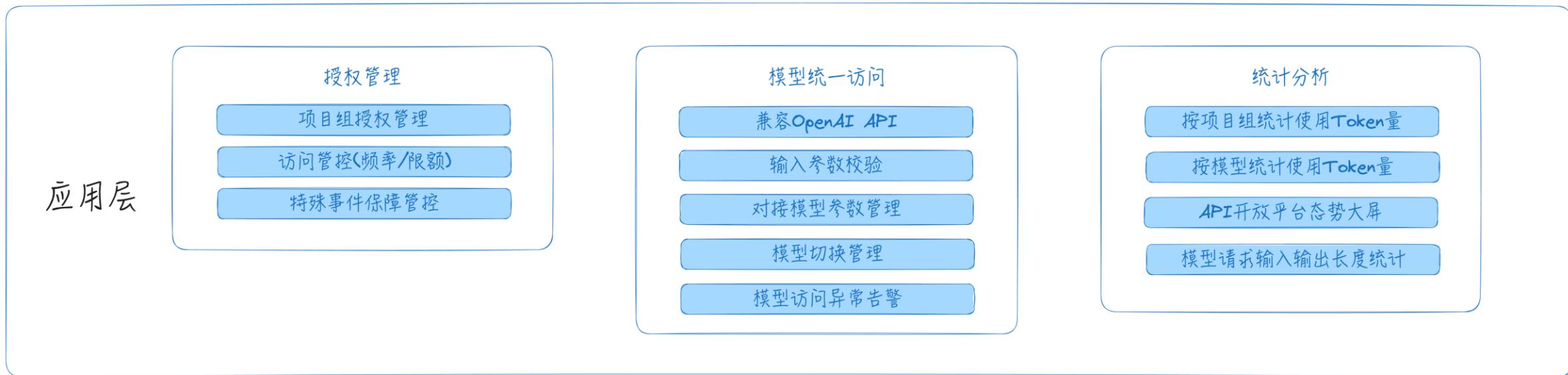
Agent 开发

05

深度适配

中国科技云大模型 API 开放平台

API 开放平台



大模型 API 服务



完全兼容 OpenAI API 格式，方便各类应用的直接接入 (如 ragflow, dify, openwebui等)



提供不同版本的模型选择 (deepseek-v3:671b, deepseek-r1:671b, deepseek-r1:32b, 嵌入模型, QwQ-32b) , 后续会增加 其他模型



使用API key进行认证，方便自研应用的接入，模型请求接口均返回请求消耗的token数，方便应用层对用户行为进行计量

已支持 300+ 课题组使用

API.md

2025-02-16

中国科技云 Deepseek 和相关模型服务 API 使用文档

API 文档是为开发者提供的接口调用说明,开发者可以通过 API 文档了解模型使用相关接口的调用方式, 参数说明, 返回结果等信息.

本文档中的全部 API 均为 OpenAI-API-Compatible, 除文档中描述的直接请求外, 均支持通过 openai sdk 调用. 只需用我们告知的 Token 替换 OPENAI_API_TOKEN 的值, 修改 BASE_URL 即可.

基本说明

BASE_URL: <https://uni-api.cstcloud.cn/v1>

Token: 联系 liuyude@cstnet.cn 获得

Models

获取模型列表

请求方式: GET

请求地址: /models

请求参数: 无

```
curl -X GET "https://uni-api.cstcloud.cn/v1/models" -H "Authorization: Bearer {Token}"
```

返回结果: 模型列表

```
{
  "object": "list",
  "data": [
    {
      "id": "model-id-0",
      "object": "model",
      "created": 1686935002,
      "owned_by": "CSTCloud"
    },
    {
      "id": "model-id-1",
      "object": "model",
      "created": 1686935002,
```

支持模型列表

- DeepSeek-R1:671b – 0528版
- DeepSeek-V3:671b - 0324版
- DeepSeek-R1:32b
- 千问推理模型 QwQ:32b
- 千问 Qwen3:235b
- 星火推理模型 Spark-X1:70b
- 嵌入模型 gte-Qwen2-7B(嵌入维度 3584, 上下文长度 32K)
- 嵌入模型 bge-large-zh:latest
- Rerank模型 bge-reranker-v2-m3
- Web Search API / AI Search API
- 计划近期上线:
 - 千问多模态 qwen2.5-vl:72b
 - OCR 模型

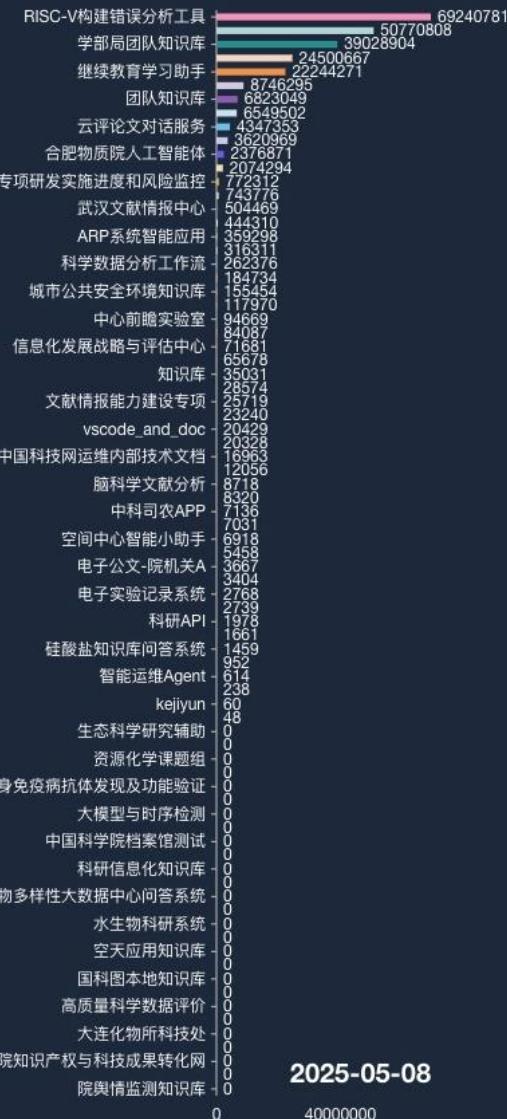
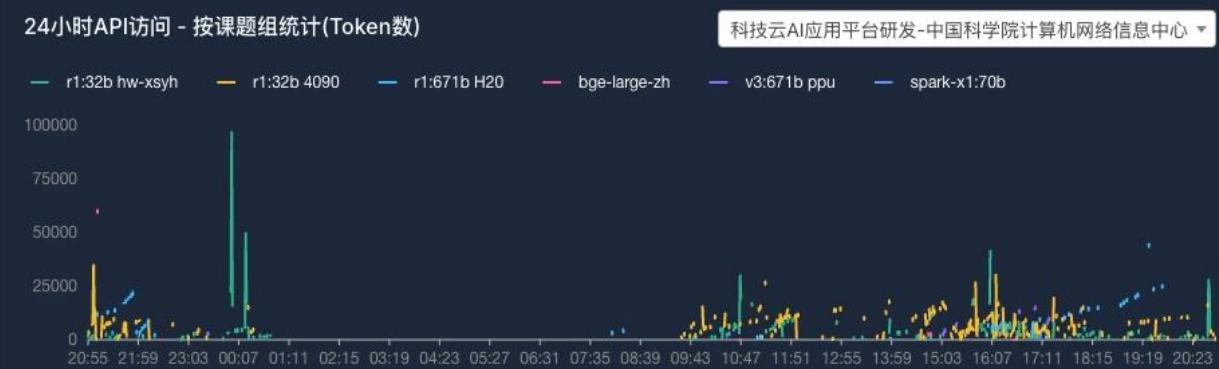
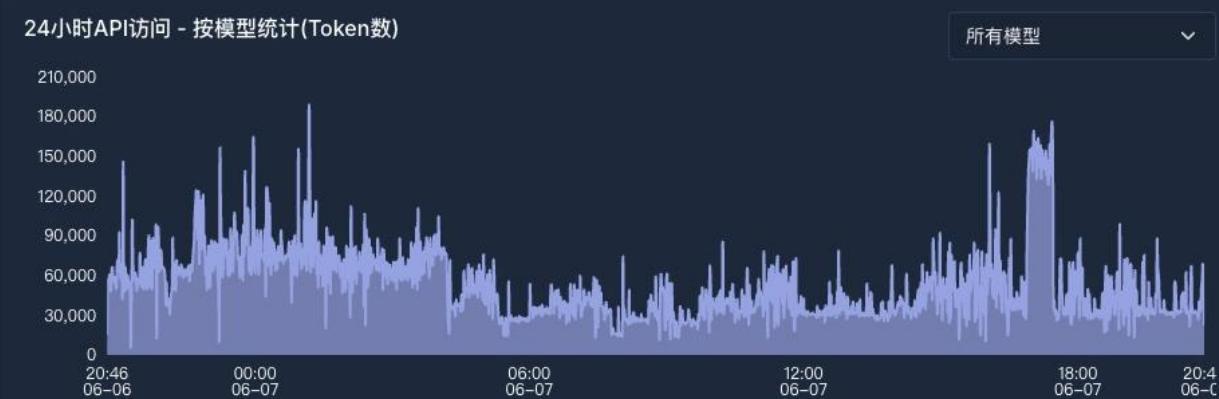
API请求平均输入和平均输出Token数

中国科技云 Deepseek API 使用态势

每日 Token 使用量排名



2025-02-16 以来通过 API 请求 Deepseek 模型消耗 Token 总量



2025-05-08

如何申请使用

- 浏览器访问 <https://uni-api.cstcloud.cn>

大模型 API 开放平台

中国科技云大模型 API 开放平台

创建 & 管理 大模型 API Key

陈灿

输入申请信息

名称(用于简短标识本 API Key 用途) *

联系人姓名 *

联系人 Email *

所在单位 *

联系电话 *

用途说明(描述大模型 API 将用于做什么研究) *

选择需要授权的模型 *

deepseek-r1:32b-16k
bge-large-zh:latest
deepseek-v3:67lb
deepseek-r1:67lb-64k
bge-reranker-v2-m3
qwen:32b
deepseek-r1:32b
spark-xt:70b
gte-qwen2:7b
qwen3:235b
deepseek-r1:67lb

确认提交

Copyright © 2025 CNIC. All rights reserved. · v1.0.0 · 京ICP备09112257号-130 · 京公网安备11010802040559号

获取API Key & 如何使用 & 追加授权模型申请

The screenshot shows a web page from the 'Chinese Technology Cloud Large Model API Interface Usage Manual'. The left sidebar has a navigation menu with items like '说明', 'BASE_URL', 'Models', 'Chats' (which is selected), 'Embeddings', 'Rerank', '技术支持', and '模型发布记录'. The main content area is titled 'Chats' and includes a search bar at the top. Below the title, it says '约 1539 字 大约 5 分钟' and '2025-04-25'. A sub-section titled '和大模型对话' is shown, with a note that the request method is 'POST' and the address is '/chat/completions'. It also lists '请求参数' (Request Parameters) which include 'model' (必选), 'messages' (必选), 'role' (必选), 'content' (必选), 'stream' (可选), 'presence_penalty' (可选), 'frequency_penalty' (可选), 'repetition_penalty' (可选), 'temperature' (可选), and 'top_p' (可选). A green callout box highlights that the 'deepseek-r1' model only supports 'user' type messages. The right sidebar contains links for '此页内容', '和大模型对话', and '特别提示'.

如何使用 – 获取已授权的模型接口

获取模型列表

请求方式: GET

请求地址: /models

请求参数: 无

```
1 curl -X GET "https://uni-api.cstcloud.cn/v1/models" -H "Authorization: Bearer {Token}" bash
```

返回结果: 模型列表

```
1 {
2     "object": "list",
3     "data": [
4         {
5             "id": "model-id-0",
6             "object": "model",
7             "created": 1686935002,
8             "owned_by": "CSTCloud"
9         },
10        {
11            "id": "model-id-1",
12            "object": "model",
13            "created": 1686935002,
14            "owned_by": "CSTCloud",
15        },
16        {
17            "id": "model-id-2",
18            "object": "model",
19            "created": 1686935002,
20            "owned_by": "CSTCloud"
21        },
22    ]
23 }
```

对话服务 <https://deepseek.ai4cas.cn>

- 快速研发对话服务(更新两个大版本，几十个小版本)，基于中国科技云认证联盟进行身份认证，提供多轮对话、对话列表保存、联网搜索、文档上传等功能。<https://deepseek.ai4cas.cn>

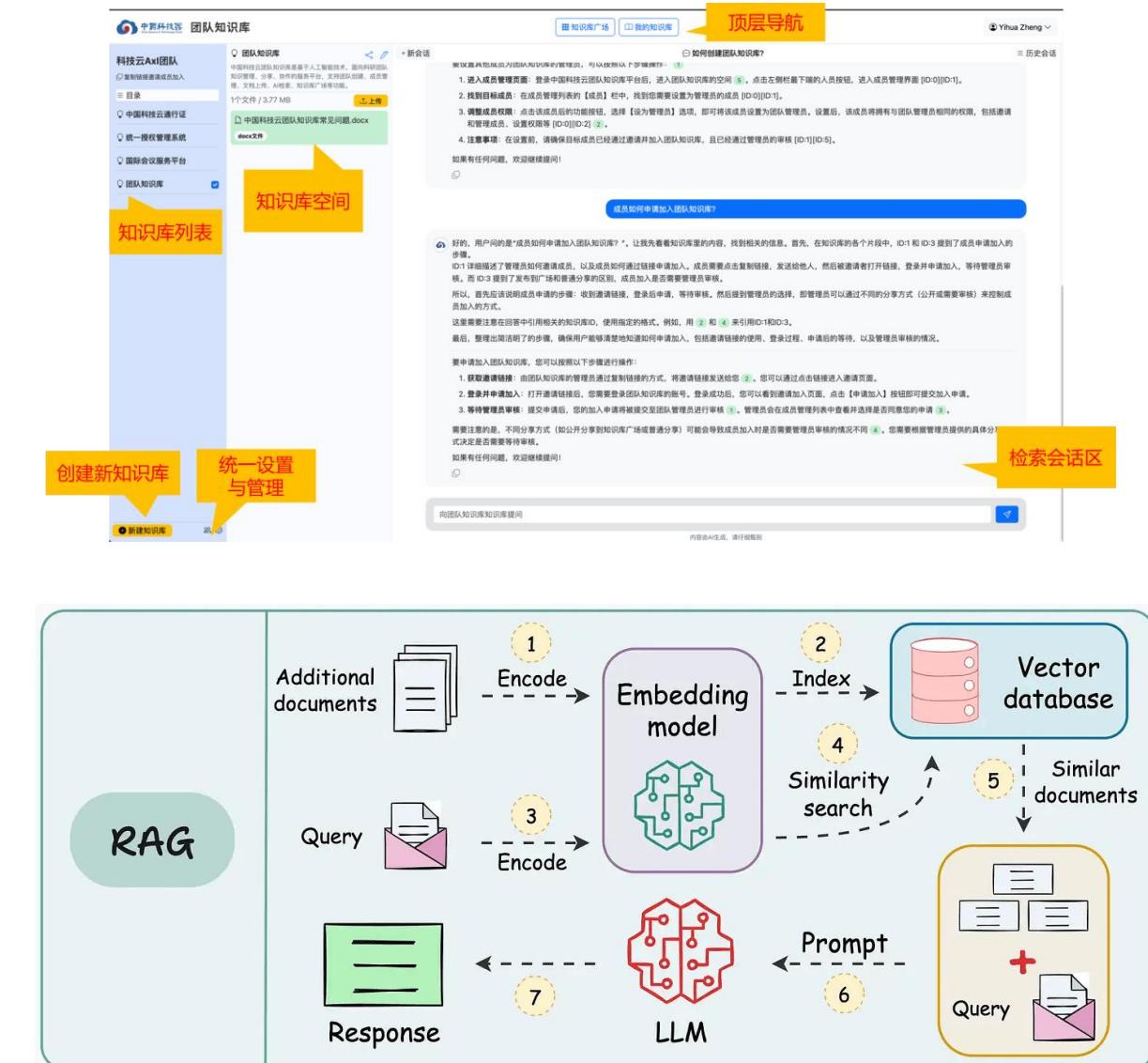
The screenshot displays the DeepSeek AI4CAS service interface, featuring several conversational windows and search results.

- Top Left Window:** "介绍一下中国科技云" (Introduction to China's Science and Technology Cloud). It shows a history of messages:
 - hi
 - 为了更好地回答【如何修改密...】
 - SDG2在非洲落实的进展
 - 联合国可持续发展目标2030最...
 - 请介绍一下中国科技云
 - 小红有两个姐姐和四个妹妹, ...
 - 请总结一下这篇论文
 - 请介绍一下你自己
- Top Middle Window:** "介绍一下中国科技云" (Introduction to China's Science and Technology Cloud). It shows a history of messages:
 - hi
 - 为了更好地回答【如何修改密...】
 - SDG2在非洲落实的进展
 - 联合国可持续发展目标2030最...
 - 请介绍一下中国科技云
 - 小红有两个姐姐和四个妹妹, ...
 - 请总结一下这篇论文
 - 请介绍一下你自己
- Top Right Window:** "请总结一下这篇论文" (Please summarize this paper). It shows a history of messages:
 - hi
 - 为了更好地回答【如何修改密...】
 - SDG2在非洲落实的进展
 - 联合国可持续发展目标2030最...
 - 请介绍一下中国科技云
 - 小红有两个姐姐和四个妹妹, ...
 - 请总结一下这篇论文
 - 请介绍一下你自己
- Middle Left Window:** "④ 思考过程" (Thinking Process). It shows a history of messages:
 - hi
 - 为了更好地回答【如何修改密...】
 - SDG2在非洲落实的进展
 - 联合国可持续发展目标2030最...
 - 请介绍一下中国科技云
 - 小红有两个姐姐和四个妹妹, ...
 - 请总结一下这篇论文
 - 请介绍一下你自己
- Middle Middle Window:** "④ 找到8个网页" (Found 8 web pages). It lists the following search results:
 1. 中国科技云 - 首页 →
 2. 中国科技云的建设进展及特色----中国科学院 →
 3. 中国科技云_百度百科 →
 4. 中国科技云 →
 5. [科技日报] 我国第一个科技云上线----中国科学院 →
 6. 我国第一个科技云上线 - 人民网·频道 →
 7. 中国科技云：一朵为科学家量身打造的“云”|阿里云[云计算]大 ... →
 8. 中国科技云：助力中国跑出科技创新加速度-新闻-科学网 →
- Middle Right Window:** "④ 思考过程" (Thinking Process). It shows a history of messages:
 - hi
 - 为了更好地回答【如何修改密...】
 - SDG2在非洲落实的进展
 - 联合国可持续发展目标2030最...
 - 请介绍一下中国科技云
 - 小红有两个姐姐和四个妹妹, ...
 - 请总结一下这篇论文
 - 请介绍一下你自己
- Bottom Left Window:** "④ 深度思考" (Depth Thinking). It shows a history of messages:
 - hi
 - 为了更好地回答【如何修改密...】
 - SDG2在非洲落实的进展
 - 联合国可持续发展目标2030最...
 - 请介绍一下中国科技云
 - 小红有两个姐姐和四个妹妹, ...
 - 请总结一下这篇论文
 - 请介绍一下你自己
- Bottom Middle Window:** "④ 深度思考" (Depth Thinking). It shows a history of messages:
 - hi
 - 为了更好地回答【如何修改密...】
 - SDG2在非洲落实的进展
 - 联合国可持续发展目标2030最...
 - 请介绍一下中国科技云
 - 小红有两个姐姐和四个妹妹, ...
 - 请总结一下这篇论文
 - 请介绍一下你自己
- Bottom Right Window:** "④ 深度思考" (Depth Thinking). It shows a history of messages:
 - hi
 - 为了更好地回答【如何修改密...】
 - SDG2在非洲落实的进展
 - 联合国可持续发展目标2030最...
 - 请介绍一下中国科技云
 - 小红有两个姐姐和四个妹妹, ...
 - 请总结一下这篇论文
 - 请介绍一下你自己

自主研发团队知识库 <https://kb.ai4cas.cn>

● AI时代的团队知识库

- 面向团队，支持团队创建、邀请成员，协同管理团队文档
- 团队内支持多知识库创建
- 团队内支持多知识库联合对话
- 支持多种方式的知识库分享
 - 知识库广场公开分享
 - 私密分享，授权加入
- DeepSeek智能支持
- 简洁友好，文件上传即刻具有智能能力



如何使用 – 使用 langchainjs

```
js test_langchain.js > [e] client > ⚡ mcpServers > ⚡ "timezone-toolkit" > ⚡ restart
26 const model = new ChatOpenAI({
27   temperature: 0,
28   configuration: {
29     baseURL: "https://uni-api.cstcloud.cn/v1",
30   },
31 },
32 );
33
34 const agent = createReactAgent({
35   llm: model,
36   tools,
37 });
38
39 try {
40   const mathResponse = await agent.streamEvents({
41     messages: [{ role: "user", content: "现在北京什么时间" }],
42   },
43   {
44     version: 'v2',
45   },
46   for await (const chunk of mathResponse) {
47     console.log(chunk.data);
48   }
49 } catch (error) {
50   console.error("Error during agent execution:", error);
51 // Tools throw ToolException for tool-specific errors
52 if (error.name === "ToolException") {
53   console.error("Tool execution failed:", error.message);
54 }
55
56 await client.close();

```

如何使用 – 对接工具

The screenshot shows the ModelLinker application interface. On the left is a sidebar with various settings and model service options. The main area displays the configuration for the 'cstcloud' service.

API 密钥: (redacted)

API 地址: https://uni-api.cstcloud.cn
https://uni-api.cstcloud.cn/v1/chat/completions

模型:

- deepseek-r1
 - deepseek-r1:32b-16k
 - deepseek-r1:32b-32k
 - deepseek-r1:32b-ppu
- deepseek-v3
 - deepseek-v3:671b
- qwen3
 - qwen3:235b
- qwq
 - qwq:32b
- spark-x1
 - spark-x1:70b

中国科技云 AI 工作台

- MCP Servers 共享平台 / MCP Hub / MCP 在线使用平台
- 大模型支持的提示词生成/优化/共享平台
- 论文深度解读

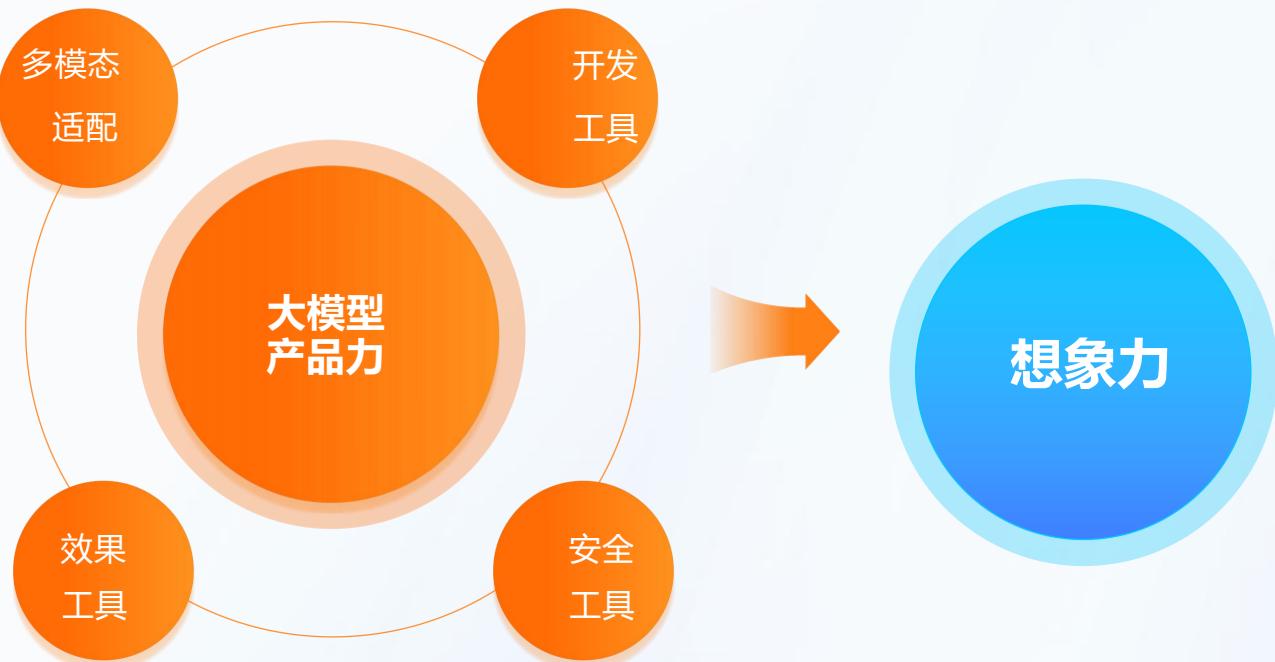
MCP

提示词

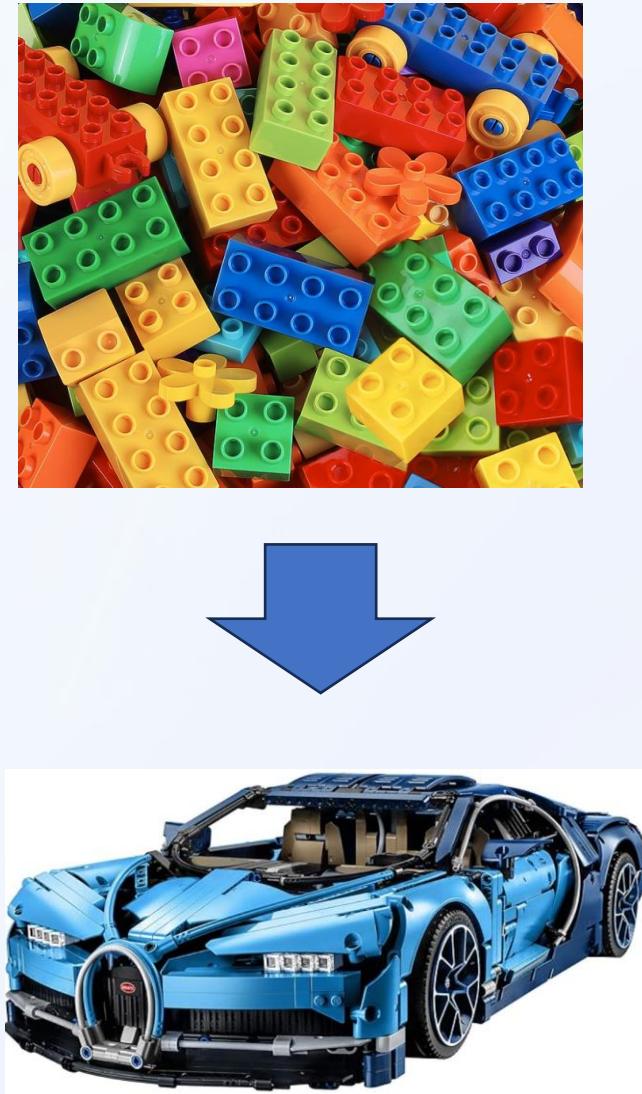
论文解读

19

大模型服务及应用已进入“乐高化”



除了提供MCP server，更多的是通过MCP server组合，售卖“**想象力**”





中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences



欢迎大家试用！

谢谢！

