

星火大模型API使用

1. 星火大模型

星火大模型，全称为讯飞星火认知大模型，是科大讯飞公司发布的一款具有强大认知能力的AI模型

星火大模型自发布以来，经历了多次迭代升级，不断提升其性能和功能。以下是其主要的发展历程：

- 2023年5月6日：科大讯飞正式发布讯飞星火认知大模型。
- 2023年6月9日：发布讯飞星火认知大模型V1.5，升级了开放式知识问答、多轮对话等能力。
- 2023年8月15日：发布讯飞星火认知大模型V2.0，进一步升级了代码能力和多模态能力。
- 2023年9月5日：讯飞星火大模型面向全民开放，用户可以在各大应用商店下载使用。
- 2023年10月24日：发布讯飞星火认知大模型V3.0。
- 2024年1月30日：发布讯飞星火认知大模型V3.5。
- 2024年6月27日：发布讯飞星火V4.0，在文本生成、语言理解、知识问答、逻辑推理、数学能力等方面实现对GPT-4 Turbo的整体超越。

2. 注册和使用API

讯飞目前提供多种不同程度的大模型API接口，都有免费的额度可以使用。

星火大模型API矩阵

接口丰富，助力企业构建自己的大模型产品

Spark 4.0 Ultra 全面超越GPT4-turbo

效果极佳

最强大的星火大模型（星火4.0 Turbo），在文本生成、语言理解、知识问答、逻辑推理、数学能力等七大维度全面超越GPT 4-Turbo，优化联网搜索链路，提供更精准回答。

免费领取

立即体验

Spark Max 整体对标GPT4-turbo

效果惊艳/优异

旗舰级大语言模型（基于4.0 Turbo量化而来），具备更强的意图理解和指令跟随能力；适用于结构化抽取、逻辑推理等对效果有更高要求的业务场景。

免费领取

立即体验

Spark Pro 支持128K长文本版本 延时更低

性能高效/强劲

专业级大语言模型，具有百亿级参数，在医疗、教育和代码等场景进行了专项优化，搜索场景延时更低。适用于文本、智能问答等对性能和响应速度有更高要求的业务场景。

免费领取

立即体验

Spark Lite 能力全面 灵活经济

免费开放

轻量级大语言模型，具有更高的响应速度，适用于低算力推理与模型精调等定制化场景，可满足企业产品快速验证的需求。

立即体验

api:

- <https://xinghuo.xfyun.cn/sparkapi#price>

注册和使用：以 Spark Max 为例子：

星火大模型API矩阵

接口丰富，助力企业构建自己的大模型产品

Spark 4.0 Ultra 全面超越GPT4-turbo

效果极佳

最强大的星火大模型（星火4.0 Turbo），在文本生成、语言理解、知识问答、逻辑推理、数学能力等七大维度全面超越GPT 4-Turbo，优化联网搜索链路，提供更精准回答。

免费领取

立即体验

Spark Max 整体对标GPT4-turbo

效果惊艳/优异

旗舰级大语言模型（基于4.0 Turbo量化而来），具备更强的意图理解和指令跟随能力；适用于结构化抽取、逻辑推理等对效果有更高要求的业务场景。

免费领取

立即体验

Spark Pro 支持128K长文本版本 延时更低

性能高效/强劲

专业级大语言模型，具有百亿级参数，在医疗、教育和代码等场景进行了专项优化，搜索场景延时更低。适用于文本、智能问答等对性能和响应速度有更高要求的业务场景。

免费领取

立即体验

Spark Lite 能力全面 灵活经济

免费开放

轻量级大语言模型，具有更高的响应速度，适用于低算力推理与模型精调等定制化场景，可满足企业产品快速验证的需求。

立即体验

购买 Spark3.5 Max

🔔 开放平台“发票管理”上线啦！可以在订单完成后至“财务>发票管理”索取发票，同时支持电子发票，更为便捷！

选择应用

选择一个应用

请选择或搜索应用

如果没有，需要点击+创建一个应用

aiweker

wedotest

选择套餐

<p>用户认证礼包</p> <p>免费包 (个人认证)</p> <p>0.00元</p> <p>tokens: 500万 有效期: 1年</p> <p>注: 个人或企业认证包限领一种</p>	<p>用户认证礼包</p> <p>免费包 (企业认证)</p> <p>0.00元</p> <p>tokens: 5千万 有效期: 1年</p> <p>注: 个人或企业认证包限领一种</p>	<p>简享包</p> <p>90.00元</p> <p>tokens: 300万 有效期: 1年</p>	<p>轻量包</p> <p>435.00元</p> <p>tokens: 1.5千万 有效期: 1年</p>
<p>套餐一</p> <p>1,350.00元</p> <p>tokens: 5千万 有效期: 1年</p>	<p>套餐二</p> <p>2,500.00元</p> <p>tokens: 1亿 有效期: 1年</p>	<p>套餐三</p> <p>23,000.00元</p> <p>tokens: 10亿 有效期: 1年</p>	<p>套餐四</p> <p>105,000.00元</p> <p>tokens: 50亿 有效期: 1年</p>

订单详情

应用名称	购买商品/套餐	商品总额	优惠金额	应付金额
	免费包 (个人认证)	¥ 0.00	¥ 0.00	¥ 0.00

我的应用 > 创建应用

* 应用名称

请设定一个应用名称，少于30个字符

* 应用分类

应用分类标签最多选一个

* 应用功能描述

请简述使用场景、应用特点等信息，不超过300个字符

提交

[返回我的应用](#)

暂无可用优惠券~

| 礼品卡

暂无可用礼品卡~

商品总额:	0.00元
乐享折扣金额:	0.00元
优惠金额:	0.00元
礼品卡抵扣:	0.00元
应付金额:	0.00元

订单备注:

☐ 下单前请仔细阅读[服务协议](#), 勾选即表示同意该协议。

确认下单

上面就完成了注册

就可以在服务管理中看到具体的token使用和API key等信息

星火大模型API

- 旗舰版七大维度全面超越GPT 4-Turbo
- Lite版本全免费接入, Pro/Max/Ultra效果强劲
- 满足各类业务场景, 更提供大模型全流程定制工具

[在线测试](#) [服务管理](#)

Max/Ultra 版本全新升级星火4.0 Turbo引擎, 欢迎体验! [立即前往 >](#)

通用大模型不满足个性化需求? 定制专属大模型 [立即前往 >](#)



接口的文档信息

- https://www.xfyun.cn/doc/spark/Web.html#_1-接口说明

特别注意：不同的模型接口是不一样的，具体的是：

请根据自己申请的模型来修改访问的domain和domain_url



```
wss://spark-api.xf-yun.com/v3.5/chat
```

Spark Pro-128K请求地址，对应的domain参数为pro-128k:

```
wss://spark-api.xf-yun.com/chat/pro-128k
```

Spark Pro请求地址，对应的domain参数为generalv3:

```
wss://spark-api.xf-yun.com/v3.1/chat
```

Spark Lite请求地址，对应的domain参数为lite:

```
wss://spark-api.xf-yun.com/v1.1/chat
```

行业大模型列表如下:

其他AI厂商的API调用可以详细查看各自的API文档和例子

下一节: 【文档】如何通过ollama部署本地大模型-deepseek-r1

下一节