



# 大模型公司 调研

---

## Table of Contents

- [大模型公司 调研](#)
  - [Table of Contents](#)
  - [Minimax](#)
    - [技术能力](#)
      - [MoE LLM abab-6.5](#)
      - [文本大模型](#)
      - [语音大模型](#)
    - [海螺AI](#)
    - [产品定价](#)
- 

## Minimax


















[minimaxi 官网](#)

MiniMax 是一家专注于人工智能大模型的创业公司，成立于2021年11月，由前商汤科技副总裁、通用智能技术负责人闫俊杰创立

# 合作伙伴

诚邀开发者、企业和机构成为MiniMax开放平台合作伙伴  
商务合作邮箱 [open-platform@minimaxi.com](mailto:open-platform@minimaxi.com)

# 技术能力

## MoE LLM abab-6.5

	英文综合	中文综合	知识	基础算数	数学解题	逻辑推理	编程	指令遵从
	MT-Bench	AlignBench	MMLU	GSM8K	MATH	BBH	HumanEval	IFEval
abab-6.5-2024 0415	8.77	7.42	78.7	91.7	51.3	83.1	78.0	78.9
abab-6.5s-2024 0415	8.69	7.18	73.0	87.5	42.6	76.8	68.3	72.8
GPT-4	8.96	7.53	86.4	92.0	52.9	83.1	67.0	85.6
GPT-3.5	8.21	6.08	70.0	57.1	34.1	66.6	48.1	70.5
Gemini 1.5 Pro	8.93*	7.33*	81.9	91.7	58.5	84.0	71.9	82.3*
Claude 3 Opus	9.00*	7.62*	86.8	95.0	61.0	86.8	84.9	91.9*
Claude 3 Sonnet	8.47*	6.70*	79.0	92.3	40.5	82.9	73.0	87.8*
Claude 3 Haiku	8.39*	6.42*	75.2	88.9	40.9	73.7	75.9	85.4*

2024 年 4 月 17 日，MiniMax 正式推出 abab 6.5 系列模型

abab 6.5 包含万亿参数，支持 200k tokens 的上下文长度

abab 6.5s 使用了同样的训练技术和数据，更高效，支持 200k tokens 的上下文长度

## 文本大模型

多轮记忆 | 信息提取 | 代码生成 | 知识问答

高并发 支持最高 10万+ RPM (requests per minute)

低门槛微调 仅需百条训练数据

 <b>对话闲聊</b> 与设定的背景人物进行对话聊天	 <b>办公效率</b> 在文档、PPT、表格等场景下 协助用户提升办公效率	 <b>评论/文案创作</b> 根据给定的主题和格式进行内容创作
 <b>舆情/情感分析</b> 根据内容判断文本内容的情绪类型	 <b>数字人/IP复刻</b> 创造自定义虚拟人或复刻IP人物 用IM方式进行沟通	 <b>客服场景</b> 根据用户的提问提供与产品相关的解答
 <b>会议助理</b> 根据给出内容形成总结摘要	 <b>内容摘要</b> 根据给出内容形成总结摘要	 <b>搜索优化</b> 集合搜索引擎的知识精确回答问题

### 客户案例

1. 办公 - 为金山办公WPS提供大模型能力，提供文本生成、要点总结、任务分发、Excel表格处理及PPT生成等能力
2. 社交 - 为小红书提供生成式搜索，用户可通过自然语言表达个人意图以获得大模型提炼和总结各搜索结果后的精准回答，并对答案进行多轮追问
3. 医疗 - 为高济健康搭建医生及药师AI助手，对医患进行定期回访并协助回答病患专业问题，极大提升服务效率与专业度

## 语音大模型

多品质音色复刻 | 语音在线合成 | 自动理解文本



**有声书合成**  
支持批量生成高质量有声书音频



**语聊播客**  
高保真、超自然情感表达



**资讯快报**  
适配各领域头条信息语音播报



**教育教学**  
教培行业真人老师音色复刻  
深入理解每一堂课



**智能硬件**  
优化硬件设备语音交互  
提升用户体验



**语音客服**  
自动化客服24小时倾心提供语音帮助

# 海螺AI

类似 KIMI 的 AI 助手

海螺AI

# 产品定价

产品定价

计费项是token数，以1000个token为单位，1000 tokens约对应750个字符文本（包括标点等字符），计费token数包括输入和输出

## ChatCompletion Pro & ChatCompletion V2 & ChatCompletion

计算项	单价
abab6	0.1 元/千tokens
abab5.5	0.015 元/千tokens
abab5.5s	0.005 元/千tokens

## Embeddings 向量化

接口为Embeddings，使用MiniMax自研的embo-01模型，仅针对token数进行统计计费

计算项	单价
embo-01	0.0005 元/千tokens

**Retrieval 知识库**

知识库向量化产生的embedding token费用 和 向量化后的 数据存储 费用

计算项	单价
embedding token	0.001 元/千tokens
存储 - 向量数据	1.5 元/GB/天

**Finetune 微调**

微调接口支持用户构造定制数据集对预训练模型进行多样化微调

训练计费 = 每千tokens的价格 \* 输入文件中的千tokens数量 \* 训练的 epoch 数

计费类型	模型	单价
训练	abab5.5-chat-240123	0.06 元/千tokens
训练	abab5.5-chat-240119	0.06 元/千tokens
推理	abab5.5-chat-240123	0.05 元/千tokens
推理	abab5.5-chat-240119	0.015 元/千tokens