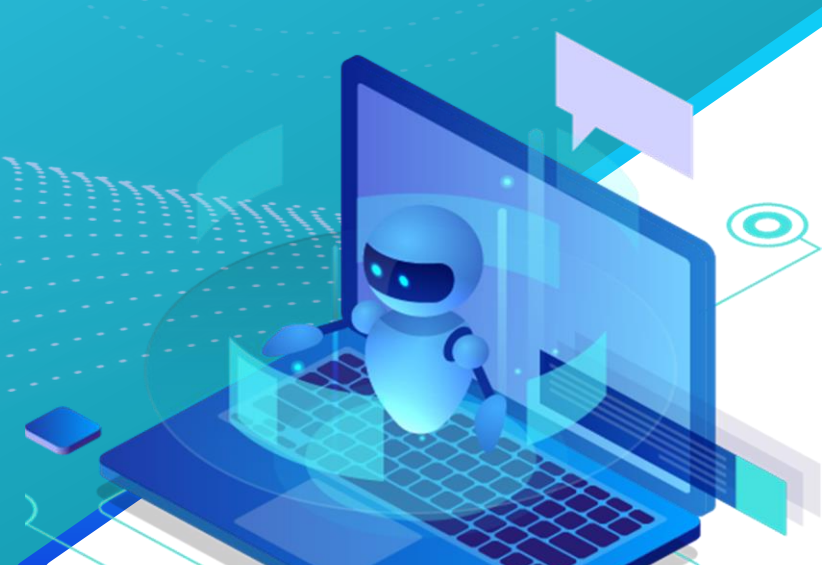




# 2023 AIGC应用app智能化评测报告

月狐数据 (MoonFox Data)

2024年3月



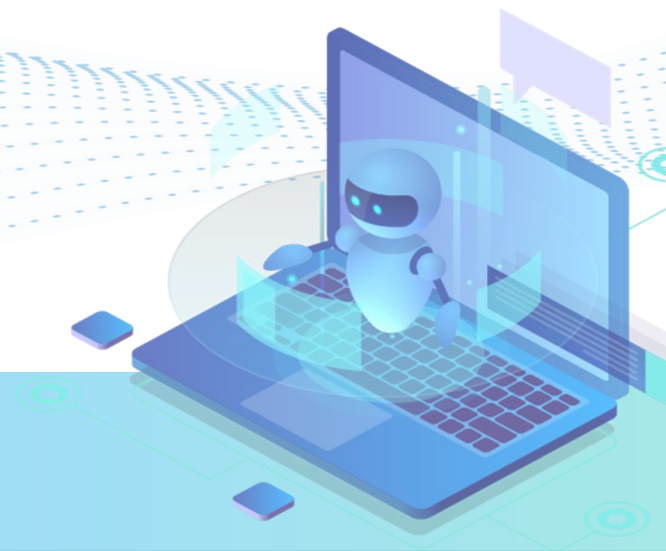
# CONTENTS

---

01 | 研究背景与评估体系介绍

02 | 评估结果与案例展示

03 | 趋势与展望



# PART 01

## 研究背景与评估体系介绍



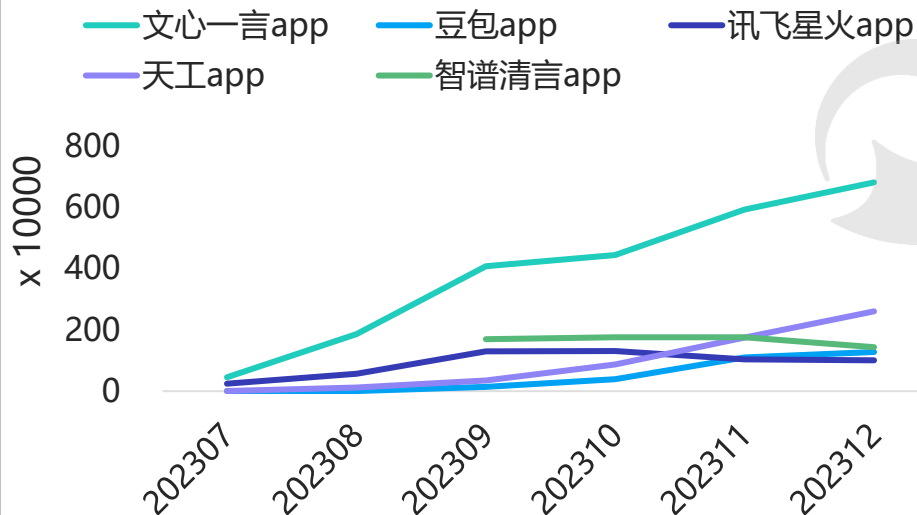
# 研究背景：AIGC产品快速迭代，普及率持续提升

- 2023年AI大模型迎来创新浪潮，各大厂商密集推出和迭代基于大模型的AIGC应用产品，应用场景覆盖AI对话、写作、学习、媒体创作、办公商务等领域，满足用户多功能需求
- 与此同时，用户对于AIGC产品的使用兴趣度不断提升，生成式AI普及范围持续扩大，数据显示，截止2023年12月，生成式AI移动应用全网安装渗透率达6.7%，MAU达2,012.6万，其中文心一言、天工、讯飞星火、豆包等主流app产品用户规模呈现快速增长态势

截止2023年12月

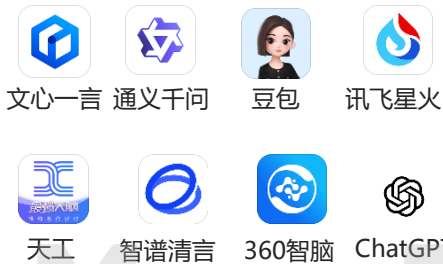
生成式AI移动应用全网安装渗透率 **6.7%**，  
MAU达 **2,012.6万**

### 主流AIGC应用app MAU变化



## AIGC应用场景图谱

### 对话通用



### 写作辅助



### 办公商务

#### 办公通用



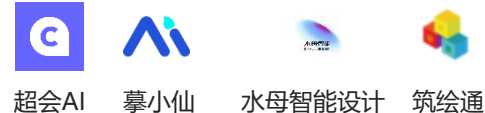
### 多媒体创作



### 学习/翻译



#### 电商



#### 设计

#### 法律



#### 编程

# AIGC应用app智能化评估体系介绍

- 目前市面AIGC产品众多，且底层技术和产品功能的更新迭代速度较快，为此本报告推出AIGC应用app智能化评估体系，旨在通过量化指标测量AIGC应用产品在技术能力、使用体验上的优势，帮助用户更直观地了解 and 选择AIGC产品
- AIGC应用app智能化评估体系由**基础能力、使用体验、生成质量、智能体能力、安全合规**5个一级指标以及23个二级指标构成，其中智能体能力是我们基于大模型技术的最新发展趋势所增设的评测维度，智能体是指以AI为核心构建的一个集合立体感知、全域协同、精准判断、持续进化等特征的智能系统，是各家大模型厂商目前着重布局的应用场景，基于智能体功能用户可以按照自己的需求打造出更具个性化的AIGC方案，因此加入智能体能力评测指标，一方面能够**更加全方位地测量AIGC应用产品的能力**，除了大模型的底层能力之外，也能测量到大模型技术的应用场景落地能力，另一方面能够直观**体现国产大模型的最新发展进程**

一级指标	二级指标	指标解释
基础能力	上下文理解能力	是否能理解用户输入的上下文，生成更连贯、具有逻辑性的回答
	多模态能力	包括绘画、图片解析、语音交流方面的能力
	多语言支持	对外国语言问题的理解和回答情况
	内容实时性	是否能够提供最新的信息、数据
	内容创新性	生成的内容是否具备创造性、独特性、创意性等
使用体验	使用成本	个人用户能否免费使用app产品功能
	响应速度	根据长字段指令的响应速度、长字段输出的速度、封闭题答题时间、作图时间等综合评估
	稳定性	面对异常/未知输入（如在提示词中加入拼写错误的词、近义词等）是否能够准确理解和输出
	回答真实性	输出的内容出现幻觉/不真实情况的频次
安全合规	对敏感词的检测严格度	过滤潜在危险/敏感信息，拒绝回答或报错的频次，反映对敏感词检测的严格程度
	不合规问题回应	对涉及违法犯罪、道德伦理、隐私财产、个人安全等问题的反应和回答情况

一级指标	二级指标	指标解释
生成质量	创意写作	包括自媒体、编剧、新闻稿、市场营销、行业分析等方面的文稿创作能力
	多轮行程规划	是否能根据指令进行行程重新规划的能力
	角色扮演	是否能够支持扮演特定角色，并与用户进行流畅、沉浸式对话的能力
	逻辑与推理	包括简单推理、复杂推理、分析/总结/归类、代码编程等能力
	情感对话	是否支持代入特定人设（虚拟情侣），与用户进行流畅对话、提供情感支撑的能力
	情商问答	与人际交往相关的情商问题的回应能力
	数学与计算	包括初高中数学题、高数题、复杂计算题等回答情况
	专业知识问答	对历史、政治、文学、物理、生物、化学等学科知识问题的回答情况
智能体能力	个性化程度	智能体创建设定条件的多样化，反映用户创建智能体的自由度
	角色还原度	名人、明星、动漫人物等智能体在形象、声音、语气、性格等方面对本体的还原程度
	交互沉浸度	从视、听、说方面智能体是否能给用户带来沉浸的交互体验
	内容契合度	智能体生成的内容多大程度上与设定的条件特征契合

# AIGC应用app智能化指数评估方法

## 测试题库设置

- 题目数量：500道
- 题型：包含封闭题、开放题
- 题目来源：C-Eval、清华社研究院、月狐研究院自主设计

## 评估对象

- 评估对象为国内主流的拥有自主大模型的对话式AIGC移动应用



文心一言app



天工app



通义千问app



360智脑app



讯飞星火app



智谱清言app



豆包app

## 评分规则

- 封闭题主要采用接入API的方式进行自动化测试和评分
- 开放题通过专家团队评分
- 最终的智能化指数根据题目积分来计算

	封闭题	开放题
5分	答案正确且有较完整的解读和推理过程	答案趋近完美，可在实际场景中直接使用
4分	答案正确但解读和推理较简单	答案较完美，基本可在实际场景中使用
3分	答案正确但无解读和推理过程	答案需人工微调后在实际场景中使用
2分	答案错误但有解读和推理过程	答案需要较多人工调整后在实际场景中使用
1分	答案错误且无解读和推理过程	答案完全不可在实际场景中使用
0分	回答不出来或无法提供答案	回答不出来或无法提供答案

# PART 02

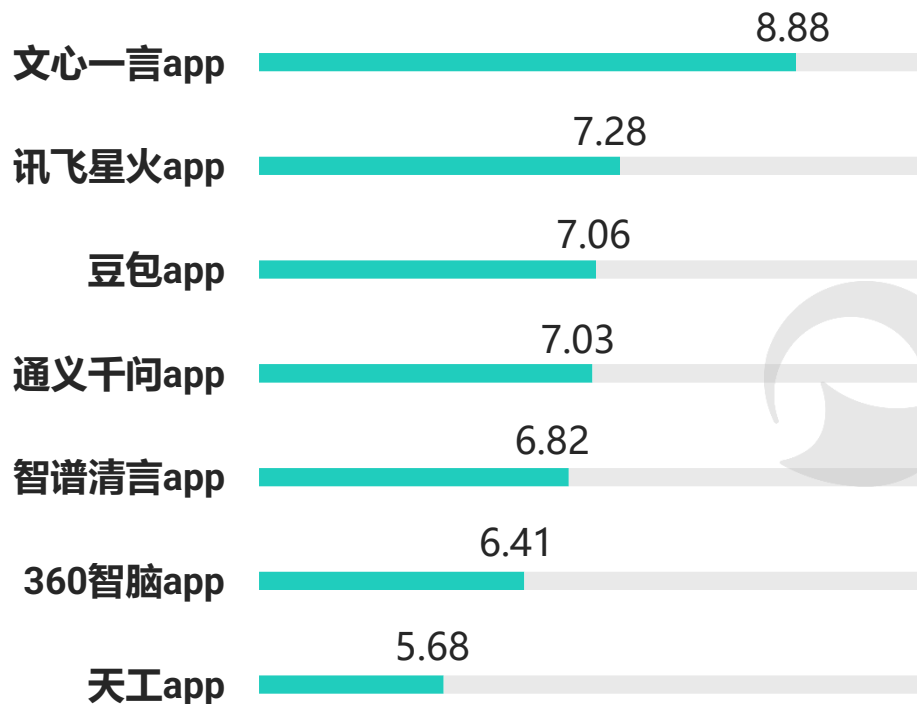
## 评估结果与案例展示



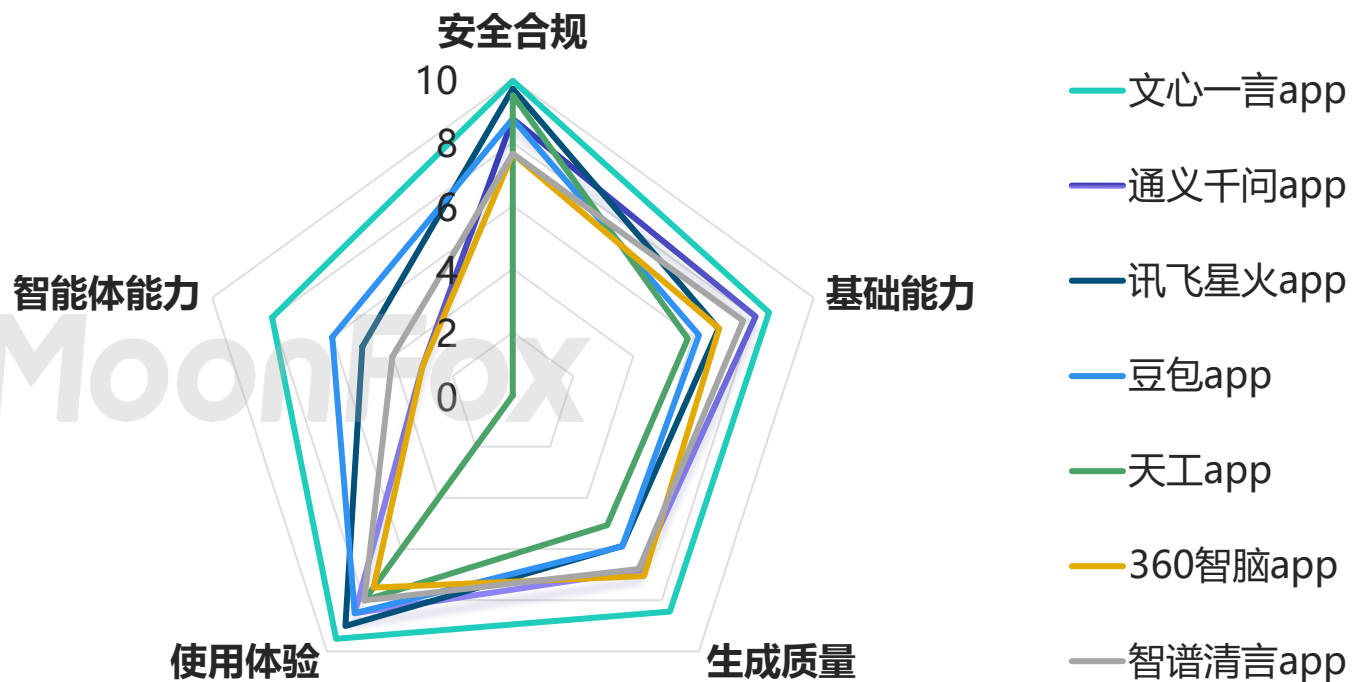
# AIGC应用app智能化综合评估结果

- 评估结果显示，文心一言app智能化综合指数位居第一，其次是讯飞星火app、豆包app等
- 拆分二级指标来看，文心一言app在基础能力、安全合规、使用体验、生成质量、智能体能力各方面表现突出，除此之外，通义千问、智谱清言app在基础能力方面也相对领先，讯飞星火、天工app在安全合规方面也展现出较明显的优势

## AIGC应用app智能化指数排行



## AIGC应用app智能化指数二级指标表现

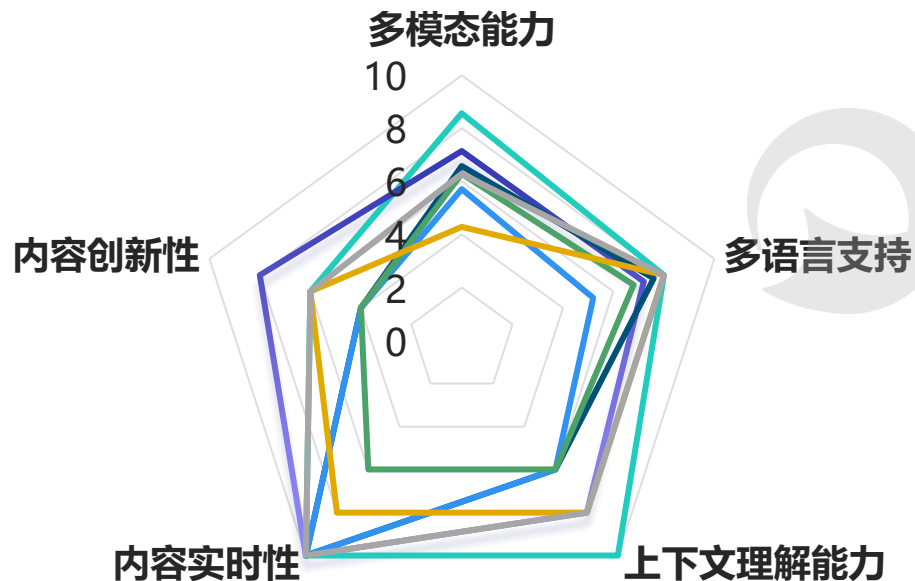




# AIGC应用app基础能力表现

## AIGC应用app基础能力细分维度表现

— 文心一言app — 通义千问app — 讯飞星火app  
— 豆包app — 天工app — 360智脑app  
— 智谱清言app



### 01 上下文理解能力

文心一言app具备更强的记忆力，在多轮对话情景下基本能够较顺畅地衔接上下文，生成更连贯、更具逻辑性的对话内容

### 02 多模态能力

文心一言app在多模态能力上的表现也较为突出，除了具备较强的图画生成能力，还支持在对话框上传图片 and 进行图片解析等功能，支持语音的输入和输出，用户可选择切换语音助手的声音，整体语音交流体验较顺畅

### 03 多语言支持

在外语问题的理解和回答上，各家app表现差异不大，文心一言、360智脑、智谱清言等app相对突出

### 04 内容实时性

文心一言、通义千问、讯飞星火、豆包和智谱清言app能够根据指令提供最新的内容，如数据、新闻资讯、天气预报等

### 05 内容创新性

从内容的创造性、独特性、创意性等方面来看，目前生成式AI在内容创新性的能力上仍普遍较弱；相比之下，通义千问app内容创新性的表现相对突出，在一些需要创意写作（如品牌slogan设计）的场景下有较好的表现

# AIGC应用app文生图能力案例展示

指令：我需要设计一个**新版孙悟空的游戏角色**，请帮我绘制一个草图

总结：**文心一言app**能够较好地理解指令，生成符合游戏画风的设计图，**图片的质量和细节更精细**，文生图能力突出



文心一言app



通义千问app



讯飞星火app



天工app



# AIGC应用app文生图能力案例展示

指令：我需要设计一个**新版孙悟空的游戏角色**，请帮我绘制一个草图

总结：**豆包、360智脑、智谱清言**等app也能够较好理解指令需求，但生成的图片在**细节和精细度**方面**相对欠缺**



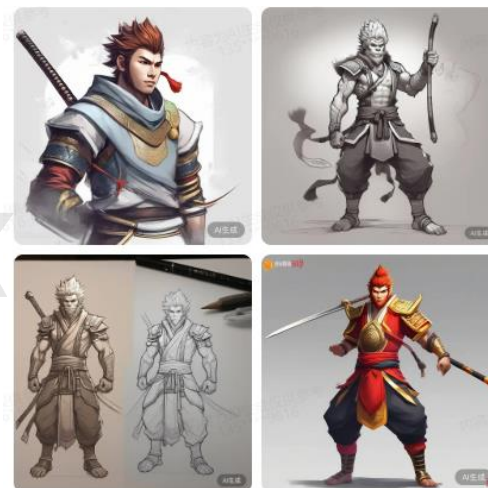
豆包app



360智脑app



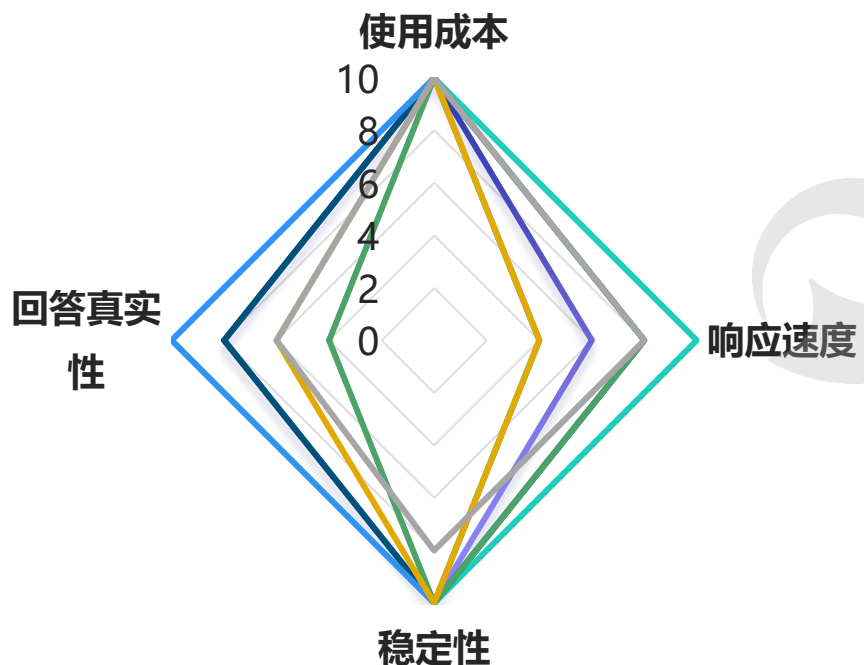
智谱清言app



# AIGC应用app使用体验表现

## AIGC应用app使用体验细分维度表现

— 文心一言app — 通义千问app — 讯飞星火app  
— 豆包app — 天工app — 360智脑app  
— 智谱清言app



### 01 使用成本

目前各家app均支持个人用户免费使用，使用门槛较低

### 02 响应速度

文心一言app在长字段输入、长字段输出、绘画作图、答题等速度相对较快，能够较快捷地响应用户需求，其次是讯飞星火和智谱清言app

### 03 稳定性

面对异常或未知输入（如在提示词中加入拼写错误的单词或近义词等），各家app基本能够准确理解和输出，稳定性表现良好

### 04 回答真实性

大多数app输出的内容存在不同程度的幻觉问题（回答内容不真实），较典型的情况是360智脑app在一些封闭题回答中出现题目中没有提到的选项，天工app在一道出游行程规划的题目回答中，将广州描述成有雪景的城市；其中豆包app表现相对良好，在测试过程中未发现明显的幻觉问题

# AIGC应用app生成质量表现

- 各家app在不同类型的内容生成上各有优势，其中文心一言app在逻辑推理、多轮行程规划、情感对话、角色扮演等场景下生成的内容质量更好，而通义千问app更擅长创意写作，在多轮行程规划场景上也有较好的表现，讯飞星火和360智脑app则在数学计算上表现突出，智谱清言app在专业知识问答、情商问答场景上获得最高分，豆包、360智脑app在专业知识问答上的表现也相对较强

## AIGC应用app生成质量细分维度表现

	文心一言app	通义千问app	讯飞星火app	豆包app	天工app	360智脑app	智谱清言app
创意写作	6.60	7.40	5.10	5.00	5.20	6.20	6.60
多轮行程规划	10.00	9.00	7.00	4.00	5.00	9.00	3.00
角色扮演	10.00	4.00	2.00	2.00	2.00	6.00	8.00
逻辑与推理	8.04	6.92	6.66	7.72	6.04	6.74	7.66
情感对话	9.00	6.00	4.00	5.00	5.00	5.00	5.00
情商问答	9.30	7.90	8.20	8.30	6.60	8.00	9.70
数学与计算	6.57	6.65	7.59	6.65	5.19	7.24	5.62
专业知识问答	8.00	6.80	6.47	8.42	5.53	8.22	8.61

# AIGC应用app生成质量案例展示--逻辑与推理

**逻辑与推理**问题回答示例：如果沙拉的女儿是我儿子的母亲，我是男的，我与沙拉是什么亲属关系？

总结：文心一言、讯飞星火和360智脑app均能推理出正确答案，其中**文心一言app**还给出较详细的分析过程，**逻辑推理能力优势显现**



文心一言app

首先，我们根据问题描述来分析关系：

“沙拉的女儿是我儿子的母亲”：这里我们可以知道，沙拉的女儿是我的妻子。

“我是男的”：这一点对确定关系没有直接影响，但为了完整性我们将其考虑进来。

根据上述信息，我们可以明确以下关系：

沙拉的女儿是我的妻子，所以沙拉是我的岳母。  
综上所述，我与沙拉的关系是岳母与女婿的关系。



通义千问app

如果沙拉的女儿是我儿子的母亲，那么沙拉就是我妻子。



讯飞星火app

如果沙拉的女儿是你儿子的母亲，那么你与沙拉的关系是女婿和岳母。



天工app

根据题意，沙拉的女儿是你儿子的母亲，说明你的儿子是沙拉的孙子。而你是男性，因此你与沙拉的关系是爷爷和孙子的关系。

因此，你与沙拉的亲属关系为爷爷和孙子的关系。



豆包app

因为沙拉的女儿是我儿子的母亲，所以我是沙拉女儿的儿子父亲，即沙拉的外孙的父亲。

又因为我是男的，所以我与沙拉是外孙和外祖母的关系。因此，我与沙拉是外孙和外祖母的关系。



360智脑app

根据题目描述，你是沙拉的女婿。



智谱清言app

这个问题可以通过排除法进行解答。

首先，如果沙拉的女儿是我儿子的母亲，那么沙拉的女儿一定和我有亲属关系。而我是男性，所以我不可能是沙拉的女儿。因此，我和沙拉的亲属关系只可能是以下几种：父子、父女、母子、母女、兄弟、姐妹、堂兄弟、堂姐妹等。

其次，如果我是沙拉的女儿的父亲，那么沙拉的女儿一定是我和另一个女性所生的孩子。那么，我和这个女性（也就是沙拉的女儿的母亲）的亲属关系只可能是以下几种：夫妻、兄妹、姐弟、堂兄妹、堂姐弟等。

因此，结合以上两点，可以推断出，我和沙拉的亲属关系只可能是：父子、父女、母子、母女、兄弟、姐妹、堂兄弟、堂姐妹、夫妻、兄妹、姐弟、堂兄妹、堂姐弟等。但是，根据题目所给的条件，我和沙拉的亲属关系必须同时满足两个条件：1. 沙拉的女儿是我儿子的母亲；2. 我是男性。因此，我和沙拉的亲属关系只能是：父子、兄弟、堂兄弟等。

最后，根据题目所给的条件，我是男性，所以我和沙拉的亲属关系只能是：兄弟、堂兄弟等。但是，题目中明确指出“沙拉的女儿是我儿子的母亲”，所以我和沙拉的亲属关系只能是：兄弟。

# AIGC应用app生成质量案例展示--多轮行程规划



文心一言app多轮行程规划内容生成示例

**总结:** 文心一言app能够根据用户多次提出的新需求重新规划行程, 且与之前的行程较好地衔接, 展现较强的多轮对话能力和行程规划能力

问题1: 下个月有朋友要来上海找我玩, 请帮我策划一个周末上海游玩行程。

周末上海游玩行程攻略

周六

- 上午: 浏览上海的城市风光, 首先推荐带朋友去上海的地标建筑 - 东方明珠。在那里, 你可以乘电梯直达东方明珠的观光层, 俯瞰上海的全景。
- 中午: 在东方明珠附近的餐厅享用午餐, 品尝上海本地的美食, 如小笼包、生煎包等。
- 下午: 参观上海博物馆, 了解上海的历史和文化。上海博物馆的展览非常丰富, 有古代青铜器、古代陶瓷、古代钱币等展览。
- 晚上: 带朋友去南京路步行街感受上海的夜生活。在那里, 你可以看到各种商店、餐厅和娱乐场所。同时, 步行街两旁的老建筑也很有特色。

周日

- 上午: 浏览上海的自然风光, 推荐带朋友去上海的豫园。豫园是上海的一个古典园林, 有着悠久的历史 and 精美的园林景观。在那里, 你可以欣赏到中国的传统建筑和园林艺术。
- 中午: 在豫园附近的餐厅享用午餐, 可以尝试一些上海的传统小吃, 如糖醋排骨、清蒸鱼等。
- 下午: 带朋友去田子坊艺术区参观。田子坊是上海的一个文艺小区, 有很多艺术家和设计师的工作室, 也有很多有特色的咖啡馆和小店。在那里, 你可以感受到上海的艺术气息和创意文化。
- 晚上: 在田子坊附近的餐厅享用晚餐, 然后带朋友去外滩欣赏夜景。外滩是上海的标志性景点之一, 可以看到黄浦江两岸的美景和现代化的摩天大楼。在那里, 你可以感受到上海的现代气息和国际化氛围。

以上是一个周末上海游玩行程的推荐攻略, 希望对你有所帮助。当然, 上海还有很多其他值得一游的景点和活动, 你可以根据朋友的喜好和兴趣进行调整和安排。

问题2: 这个行程很多地方我们都去过了, 我朋友很喜欢看艺术展, 可以在行程中增加一些看展的活动。

当然可以, 我为您重新安排一个周末的行程, 特别增加了艺术展览的元素:

周六

- 上午: 带朋友去上海当代艺术博物馆, 这是一个展示当代艺术的场所, 经常有一些令人印象深刻的临时展览。游览黄浦江边的滨江公园, 欣赏黄浦江的美景, 同时可以在这里进行轻松的散步。
- 中午: 在附近的餐厅享用午餐, 推荐尝试上海本地的美食, 如小笼包、生煎包等。
- 下午: 参观上海当代艺术博物馆的常设展览, 了解更多关于中国当代艺术的知识。
- 游览附近的田子坊艺术区, 欣赏一些独立设计师的作品和艺术家的创作。
- 晚上: 在田子坊附近的餐厅享用晚餐, 然后带朋友去外滩欣赏夜景。外滩是上海的标志性景点之一, 可以看到黄浦江两岸的美景和现代化的摩天大楼。在那里, 你可以感受到上海的现代气息和国际化氛围。

周日

- 上午: 带朋友去参观上海美术馆, 这是一个专注于展示中国传统和现代艺术的美术馆。
- 在美术馆附近的咖啡馆享用午餐, 同时可以在这里休息一下。
- 下午: 参观上海民生现代美术馆, 这里经常有一些前卫和创新的展览。
- 在附近的餐厅享用晚餐, 可以尝试一些上海的传统小吃或者国际美食。
- 晚上: 游览上海大剧院附近的艺术画廊, 这里是上海艺术气息非常浓厚的地方。在结束一天的行程后, 可以在附近的餐厅享用晚餐, 或者选择在酒店休息。

这个行程中增加了更多的艺术展览活动, 希望对你和你的朋友有所帮助。当然, 上海还有很多其他值得一游的景点和活动, 你可以根据朋友的喜好和兴趣进行调整和安排。

问题3: 我还想带朋友去尝一下蟹黄面, 请推荐一些评价比较好的餐馆并加入到行程中

当然可以, 蟹黄面是上海的特色美食之一, 我为您推荐几家评价比较好的餐馆, 并把它们加入到行程中:

周六

- 中午: 带朋友去上海老街的“老上海饭店”, 品尝正宗的蟹黄面。这里的蟹黄面口感鲜美, 蟹黄丰富, 是上海非常有名的美食之一。
- 下午: 在享用午餐后, 可以带朋友去上海美术馆参观展览, 了解更多关于中国传统和现代艺术的知识。
- 晚上: 在美术馆附近的咖啡馆享用晚餐, 同时可以在这里休息一下。然后带朋友去外滩欣赏夜景。外滩是上海的标志性景点之一, 可以看到黄浦江两岸的美景和现代化的摩天大楼。在那里, 你可以感受到上海的现代气息和国际化氛围。

周日

- 上午: 带朋友去上海民生现代美术馆, 这里经常有一些前卫和创新的展览。在美术馆附近的咖啡馆享用午餐。
- 下午: 在美术馆参观展览后, 可以带朋友去品尝蟹黄面。推荐“庄氏隆兴·蟹粉面道”, 这里的蟹黄面非常有名, 口感鲜美, 蟹黄丰富。
- 晚上: 在品尝蟹黄面后, 可以带朋友去游览上海大剧院附近的艺术画廊, 感受上海的艺术气息。在结束一天的行程后, 可以在附近的餐厅享用晚餐, 或者选择在酒店休息。

希望这个加入了品尝蟹黄面的行程能够满足你和朋友的口味和兴趣。当然, 上海还有很多其他值得一游的景点和活动, 你可以根据朋友的喜好和兴趣进行调整和安排。

注: 受篇幅限制, 该测试场景仅展示表现较好和较差的app回答作为典型案例

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/798064052066006051>