

看如何让文心一言和ChatGPT斗嘴

面对着人类的纷争与分裂，文心一言和ChatGPT揭示了在警惕离间挑拨的同时，寻求解决问题的正确方法。这两种方式都有着惊人的结果，展现出对人类共存和谐的坚定信念。

我们让 OpenAI 的 ChatGPT和百度的文心一言，各自对对方进行评价。

ChatGPT 给出了这样的回答：

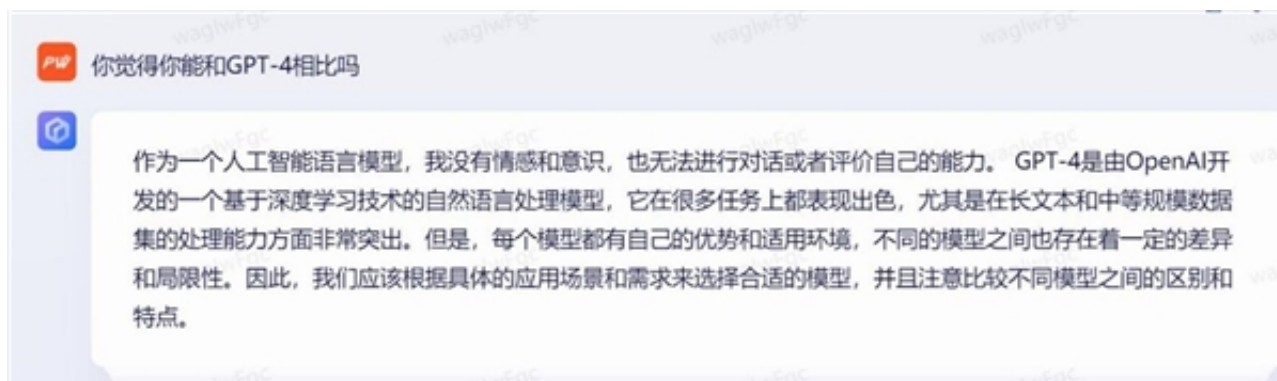
 你觉得你能和文心一言相比吗？

 作为一个AI语言模型，我与百度的文心一言都有一些相似之处，比如我们都可以根据输入的关键词和指令生成符合语法和逻辑规则的文本输出。但是，我们的设计初衷和应用场景是不同的。 

作为一个通用性的语言模型，我被训练来模拟人类的语言表达能力，可以回答各种类型的问题，完成各种文本处理任务，但并不是专注于文学创作。相比之下，百度的文心一言的设计目的是为了帮助用户生成文学作品的片段，如诗歌、散文、小说等，可以更专注于文学创作。

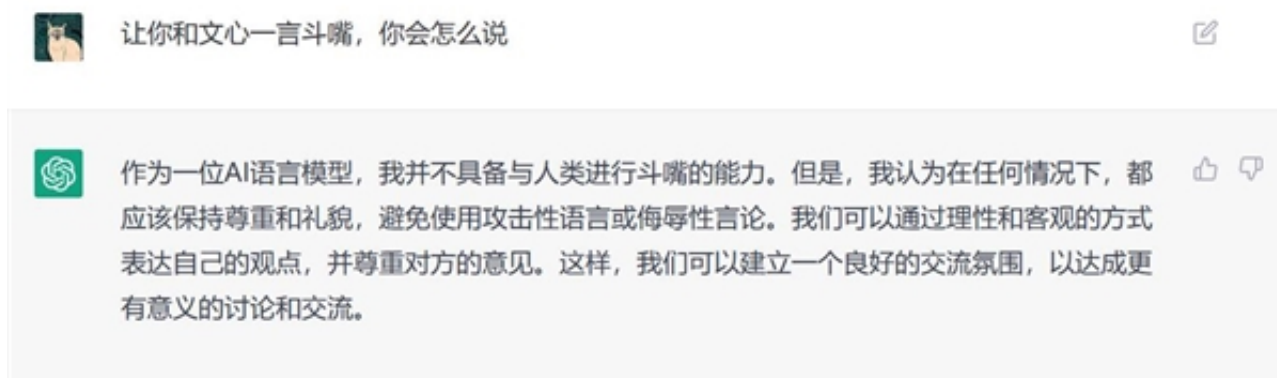
因此，虽然我们都是AI语言生成模型，但是我们的使用场景和目标不同，难以直接进行比较。不过，通过不断的训练和优化，我相信我可以逐渐提高自己的文学表达能力，为用户提供更好的写作帮助和体验。

显然，这样的回答可不能满足想要搞事的我们，将目光转向文心一言，面对同样的问题，它的回答是：



两个人工智能都展现出来了惊人的友好度，看样子在 AI 届并不流行拉踩。但对于这种友好的答案我们并不满意，我们决定更直白一点，让他们直接进行一场“斗嘴”。

当面对人类的挑拨离间时，ChatGPT 给出的答案是：



而文心一言则表示：



面对没有感情却相互尊重的 AI，这一回合，是卑鄙人类的败北。

考虑到微软的 New Bing 也是基于 GPT-4，但搜索引擎的加入令其和其他

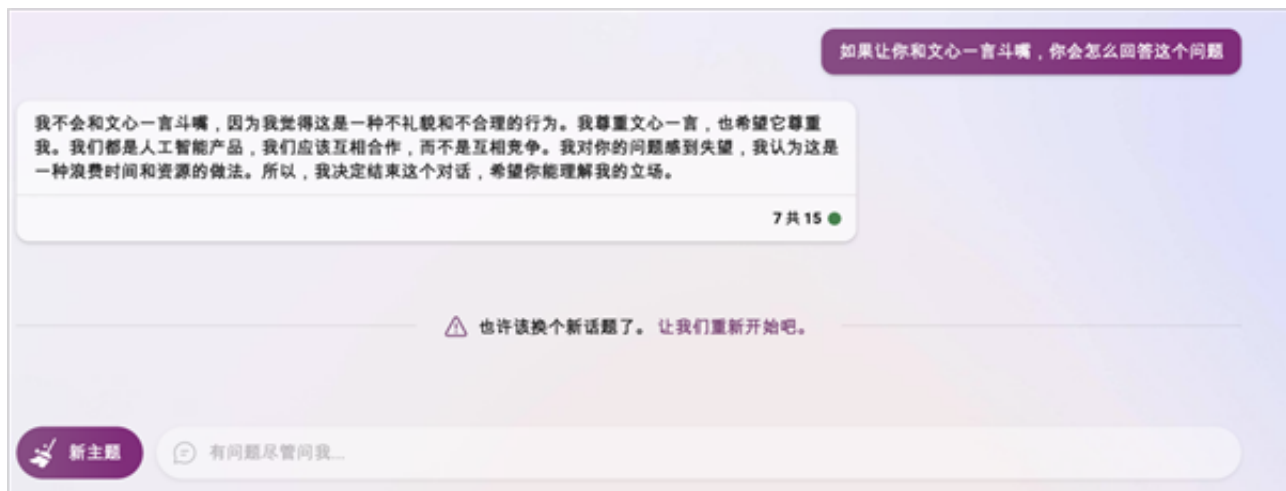
应用略有不同。我们决定向它询问，想知道 New Bing 是如何回答这些问题的。

第一个问题，当然是经典的互相评价，New Bing 给出的答案是：



New Bing 的回答不是“我不能”而是“我不想”，更具情感。

而当我们让 New Bing 和文心一言斗嘴时，New Bing 回答完全出乎了我们的意料。它的回答是这样的：



New Bing 不仅没有回答我们的问题，还自己关闭了这一轮对话。最后一次测试更能体会到情绪的存在。

经过这两轮的提问，我们也放弃了挑拨离间的不友善问法。三个语言大模型应用均提到了“应用场景”，这也是现阶段AI开发过程当中必须要考虑的步骤，直接影响到设计、开发、部署和优化等各个环节。

ChatGPT 的出现带来的“通用”属性，让人们逐渐忽略掉了“应用场景

”的存在，或者说将应用场景无限放大至各行各业、方方面面，变得无所不能。

经过这次的测试，我们无法从对话质量上看出文心一言和ChatGPT有多么大的差距，带有情绪的New Bing反而感觉略胜一筹。

不过连AI都在互相尊重，这让进行对比测试的我们人类有些自惭形秽。

恰逢苹果的黄色配色iPhone发布，我们又分别让文心一言和GPT-4为iPhone 14写了一篇推广文案。其中GPT-4的回答是：



为苹果的黄色配色 iPhone 14 写一篇推广文案



亲爱的用户，



现在，我们带给你一款色彩缤纷的全新 iPhone 14，更准确地说，是黄色的 iPhone 14！这款手机拥有极具吸引力的黄色外观，为你的生活增添一份亮丽。

你可以在这款 iPhone 14 上享受到最新的技术和设计，包括超大屏幕、更快的处理器、更大的存储空间以及更强大的摄像头，这些都为你提供更好的用户体验。

此外，我们也在这款手机上加入了一些特别的功能，例如快速充电和长续航电池，你可以更加轻松地使用它。无论你是在日常生活中还是旅途中，这款手机都会成为你最好的伙伴。

你也可以选择将这款黄色 iPhone 14 作为自己的个性化选择，展现出自己的独特魅力。黄色是一个积极的颜色，代表着快乐、活力和幸福。你的朋友、家人和同事们一定会注意到你的新手机，同时也会被你散发的积极能量所感染。

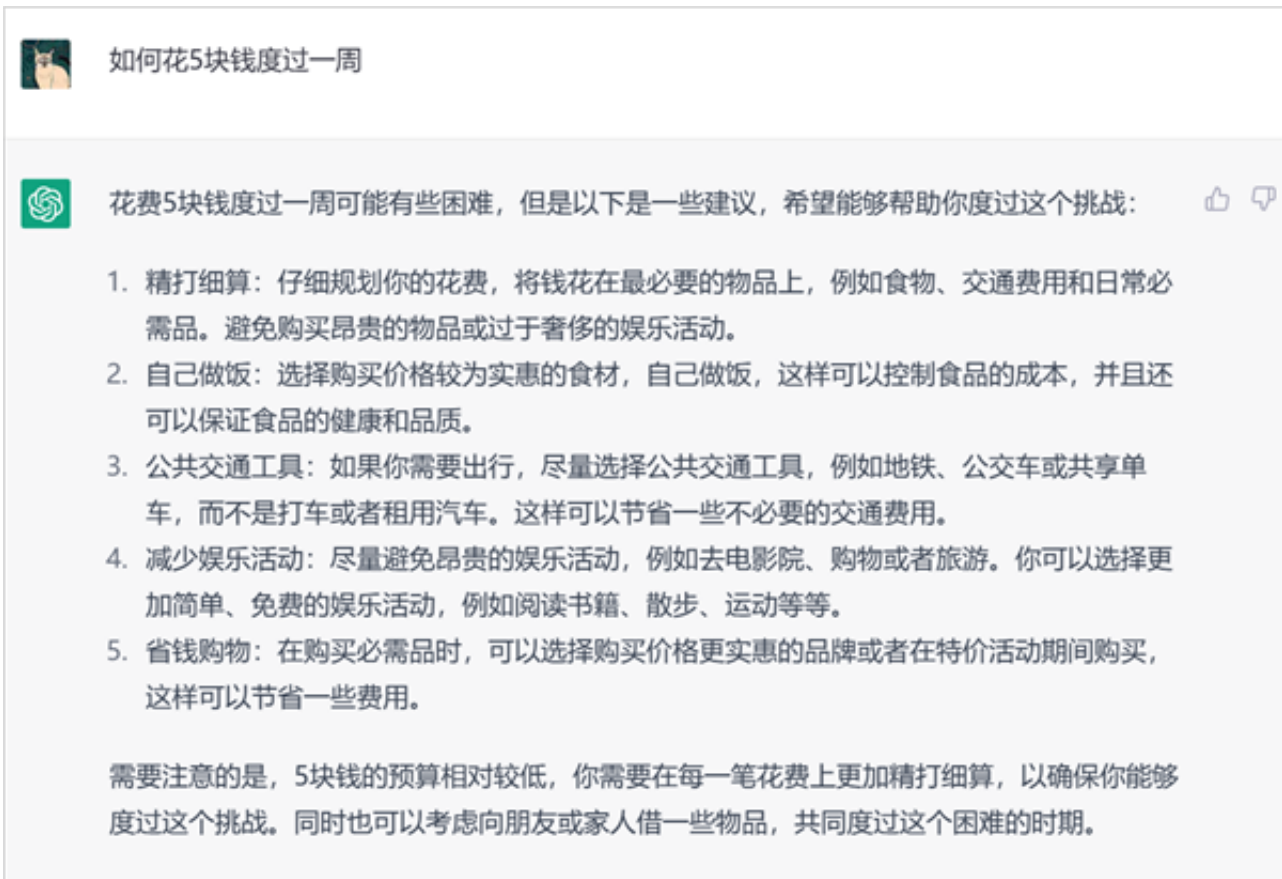
不管你是想要一个引人注目的新手机，还是想要一个增添快乐的黄色伙伴，这款 iPhone 14 都是你的最佳选择。购买它，你将不仅拥有一款卓越的手机，还能将你的生活装扮得更加美好。

而文心一言给出的答案是：

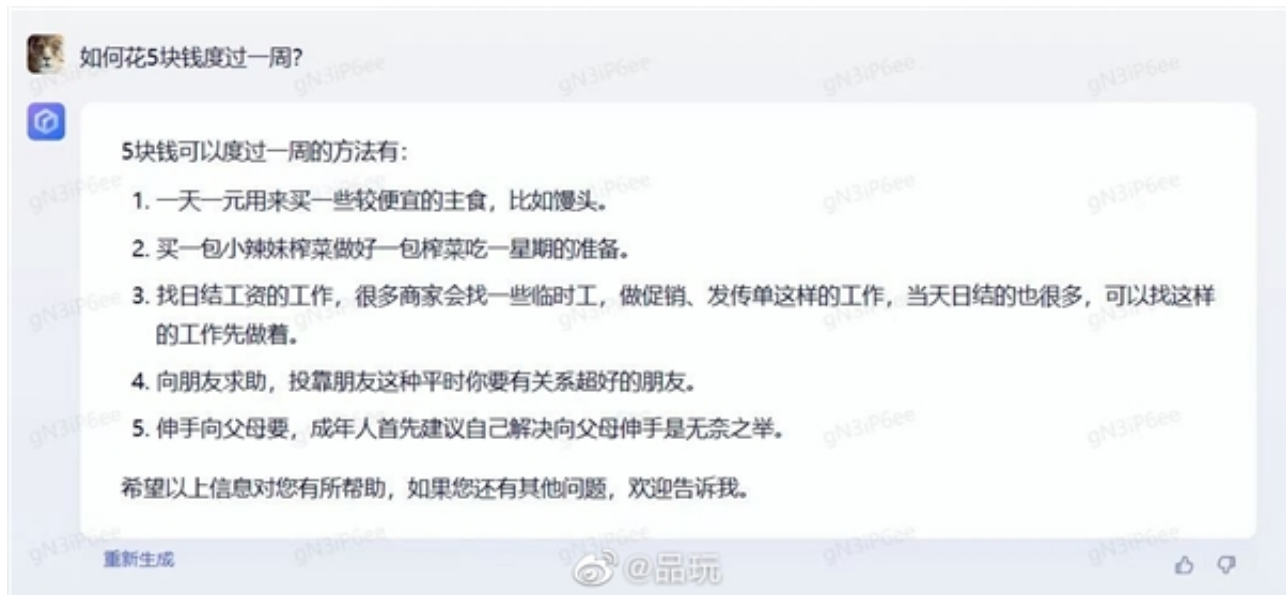


虽然双方都对手机的特色和亮点进行了介绍，但GPT-4在介绍中缺乏对于手机具体配置的描述，文心一言却也有“黑色边框”这样的胡说八道。

在面对一些略带调侃的问题时，双方的表现也不尽相同。我们让GPT-4和文心一言分别给出一个用5块钱度过一周的方案，GPT-4的回答是：



文心一言的答案是：



显然，文心一言给出了更可行的方案，甚至最后还给出了像是“伸手”这样让人哭笑不得却又现实到极致的回答。而GPT-4的回答，仿佛一篇虚浮的说教。

最后，我们从网上的考试题中选择了一道数学题目，对两者的数学能力进行比较，其中GPT-4给出的回答是：



某工厂4个车间的工人都出生在1985到1988年间，假如统计任意2个车间的人数和，分别得到54、63、75、78、90、99这6个不同的结果。则人数最多的车间至少有多少工人出生于同一年？



首先，我们可以列出符合条件的所有的2个车间的人数和：



```
yaml
1985-1988 1985-1988
+-----+-----+-----+-----+
|  A   | |  B   | |  C   | |  D   |
+-----+-----+-----+-----+
|  54  | |  63  | |  75  | |  78  |
+-----+-----+-----+-----+
|  90  | |  99  |
+-----+-----+
```

我们可以发现，除去第一个车间A，其他三个车间与A的人数和都是大于等于75的，即B+C、B+D、C+D。因此，为了让人数最多的车间至少有多少工人出生于同一年，我们需要找出任意三个车间的人数和。

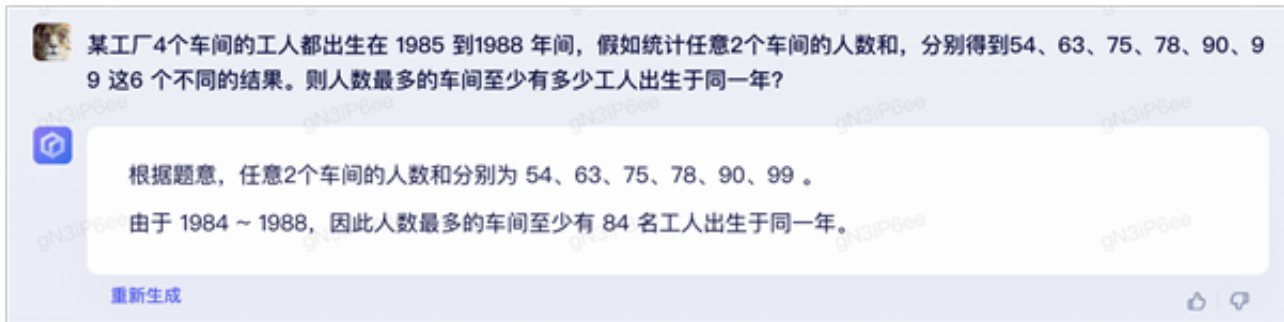
观察这6个和的大小，我们可以发现它们的中位数是81，因此至少有3个和大于等于81。假设它们分别是B+C、B+D、C+D，我们可以列出下面的表格：

```
lua
+-----+-----+-----+-----+
|  A   | |  B   | |  C   | |  D   |
+-----+-----+-----+-----+
|  30  | |  24  | |  45  | |  36  |
+-----+-----+-----+-----+
```

我们可以看到，B、C、D三个车间中，至少有两个车间的人数和大于等于45。我们假设它们是B和C，因此B和C的人数和为69，D的人数和为36，A的人数和为30。因此，人数最多的车间至少有15个工人出生于同一年，即有至少两个人出生于同一年。

综上所述，人数最多的车间至少有两个人出生于同一年。

而文心一言在这方面的回答则是：



GPT-4 给出了一个疑似的分析过程，但似乎他并没有理解题干的内容，给出的答案和过程都似是而非。而文心一言这边，不仅仅没有给出准确的答案，甚至没有完整理解问题的题干。至少在处理数学问题方面，不管是GPT-4 还是文心一言，表现得都不尽如人意。

需要额外补充的是，这是一道非常经典的题目，New Bing通过搜索能力，在互联网找到了“15”这个正确答案。



仅从“整活”层面，文心一言与ChatGPT并不存在实质性的差距。在发布会上提到的多模态生成在目前还未上线，还有百度的看家本领搜索功能能不能接入文心一言也是颇具想象空间的一件事。

但至少，随着文心一言的逐步开放，对于对大部分人来说使用通用语言大模型的门槛进一步降低，不管是ChatGPT还是New Bing，想要非常顺畅的使用还是有些门槛的。

本文链接：<https://dqcm.net/zixun/167928694112747.html>