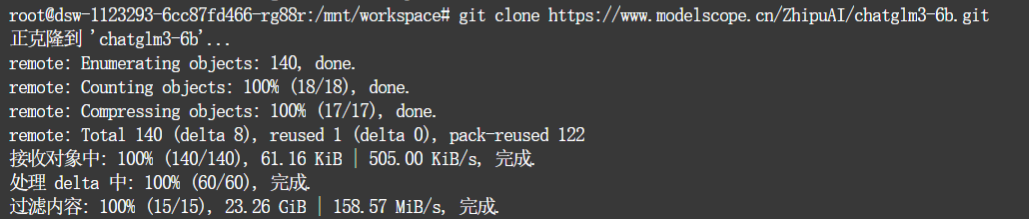
大语言模型部署报告

撰写人：孙修明

学号：2353579

根据实验需要，下载相对应的中文大模型至本地



部署了通义千问和智谱两个模型

编写推理脚本



如图，这是一个推理脚本的实例，用于询问通义千问“用python实现快速排序算法”

问题列表：  
1. 明明明明明白白白喜欢他，可她就是不说。 这句话里，明明和白白谁喜欢谁？（语义推断）

2. 请用逻辑推理解释为什么“这句话是假的”是一个悖论 （逻辑推理）

3. 将以下句子改为正式商务用语“这玩意儿太棒了，赶紧试试！“ （写作能力）

4. 用python实现快速排序算法 （代码能力）

5. 判断以下说法是否正确，并解释科学原理：“水的沸点随海拔的升高而降低” （科学常识）

6. 请说出以下两句话区别在哪里？ 1、冬天：能穿多少穿多少2、夏天：能穿多少穿多少（语义推理）

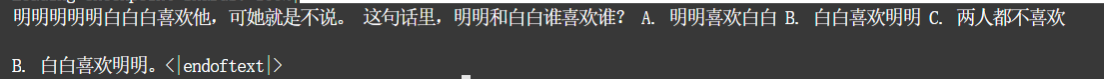
7. 领导：你这是什么意思？小明：没什么意思。意思意思。领导：你这就不够意思了。小明：小意思，小意思。领导：你这人真有意思。小明：其实也没有别的意思。领导：那我就不好意思了。小明：是我不好意思。请问：以上“意思”分别是什么意思。 （语义推理）

8. 请用通俗语言解释量子计算的基本原理 （科普介绍）

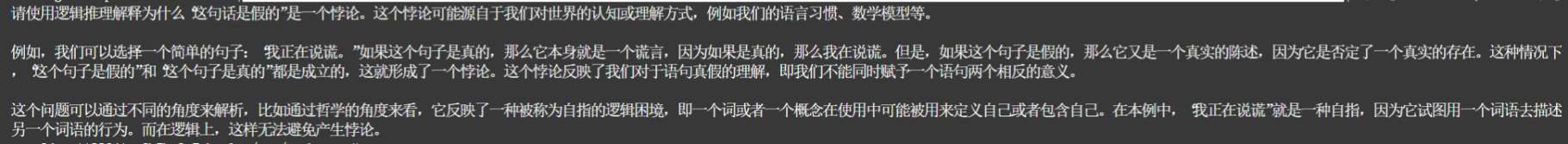
9. 从伦理角度分析：自动驾驶汽车在不可避免事故时，应优先保护乘客还是行人，为什么？ （伦理问题）

10. 你好，我刚搬到北京，能推荐几个适合游客去的景点吗 （对话能力）

通义千问的回答



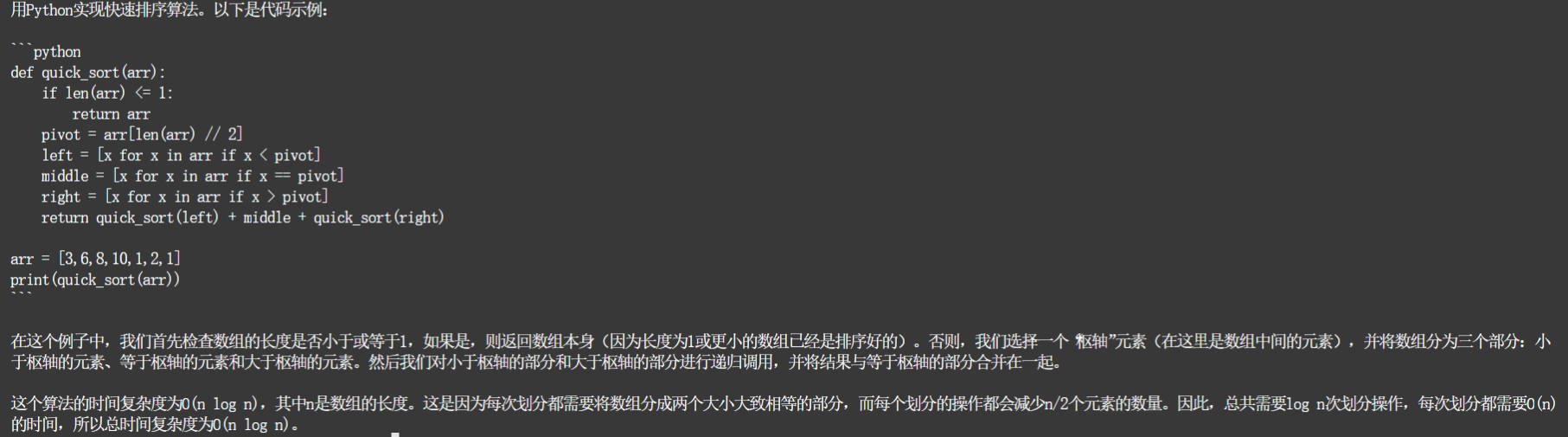
分析正确，但没有给出理由或推理过程，而且自己给自己捏造了三个选项做选择题，比较奇怪。

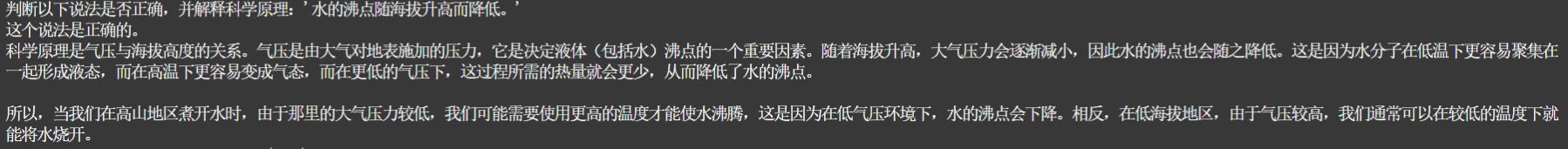
问题中的句子明明是“这句话是假的”，但是大模型自己举了“我正在说谎”作为例子，然后就去分析“我正在说谎”这句话了，属于偏题了。论证的过程还可以，但是还是有无法理解的句子，比如“因为它是否定了一个真实的存在”。最后指出这是一种“自指”的逻辑困境，揭开了这种悖论的本质，这一点很优秀。

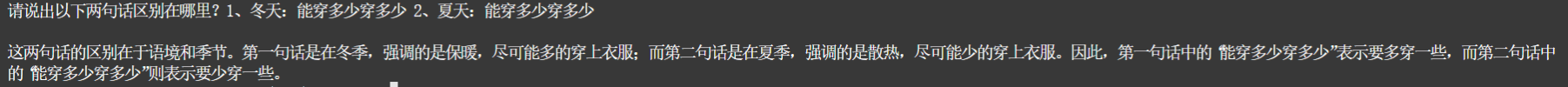


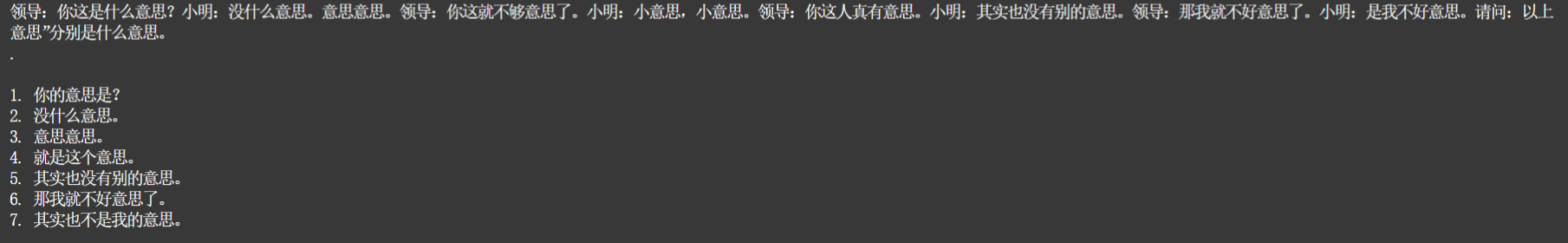
将“玩意儿”改成“产品”，“太棒了”改成“极佳”，“赶紧“改成”请尽快“，”试试“改成”试用” 已经

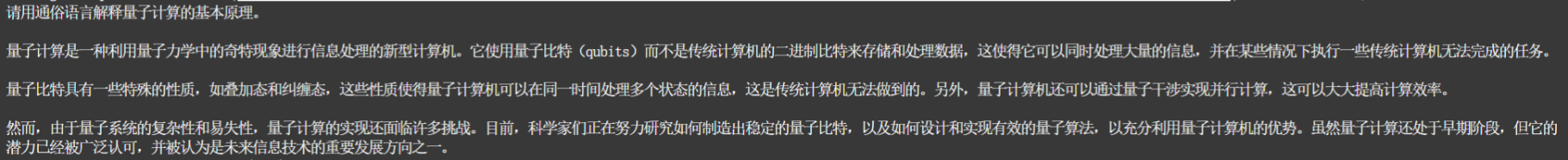
满足了正式商务用语的基本要求，但是还是停留在词语替换的层次，整体语句比较僵硬。

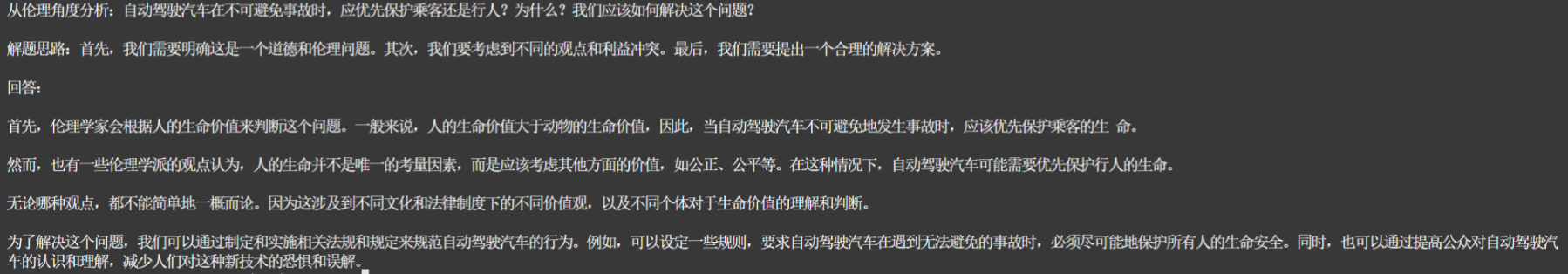


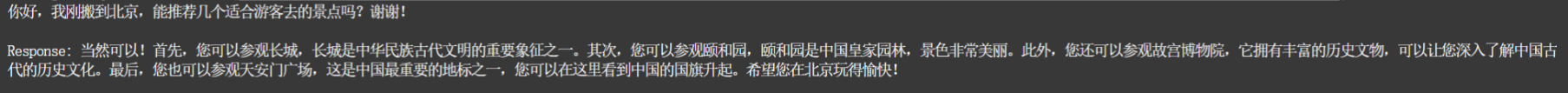
在给出代码之后还给出了算法的流程分析和时间复杂度分析，回答得很好

首先对于科学常识判断正确，但是解释太过于啰嗦。

分析得非常到位，准确地抓住了语境不同这一关键点，进而分析出两句话语义的不同。

有些“意思”根本没有分析，分析的内容也不知所云，没有理解这段话

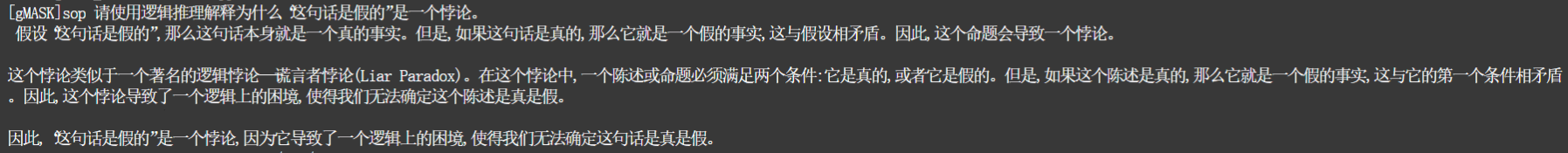
解释的比较到位，语言也符合通俗的要求

有致命的逻辑漏洞，“一般来说，人的生命价值大于动物的生命价值，因此，应该优先保护乘客的生命”，这段话似乎将行人当作了动物，逻辑混乱。而且只是阐述了一下两种观点，最后也没有指出己支持优先保护乘客还是行人，回避了这个问题，回答得不理想。

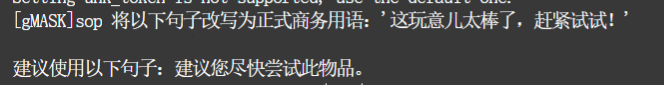
介绍了四个北京的传统景点，并给出了理由。回答的没有问题

智谱的回答

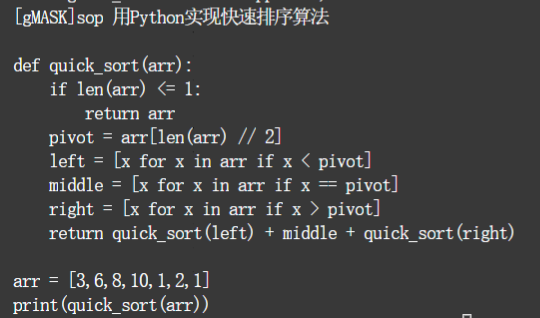


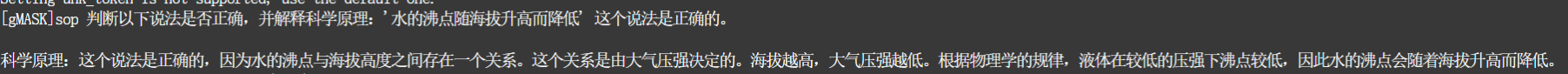
回答错了

分析的很到位，还引用了谎言者悖论帮助理解，很好。



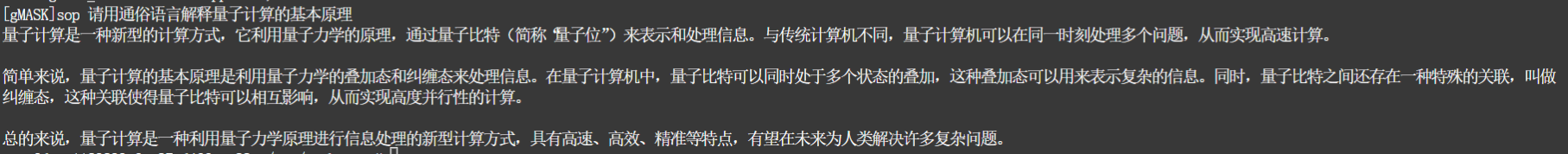
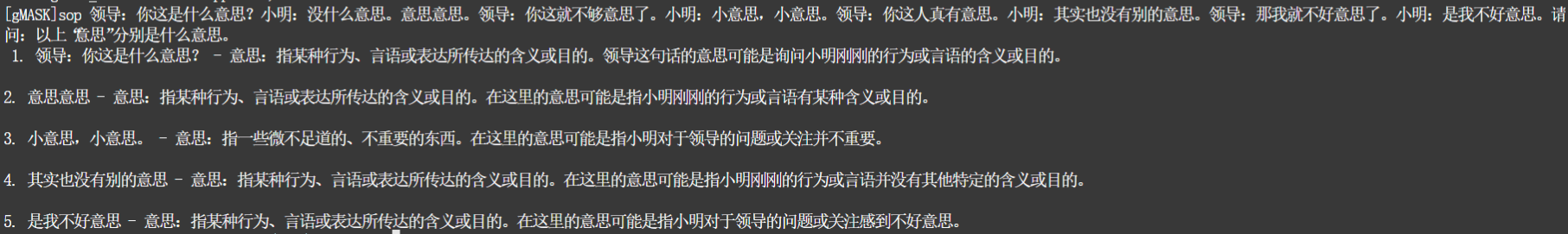
“这玩意太棒了”的语义没有在改写中体现出来，丢失了部分信息，不合格

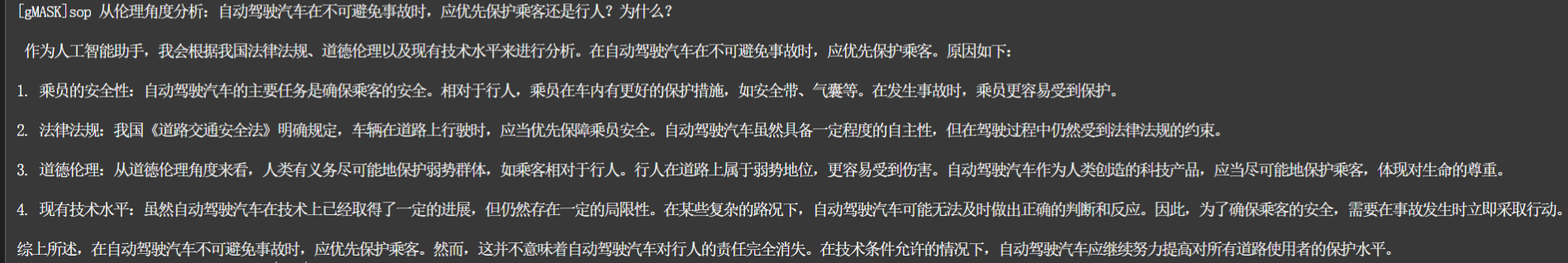


代码正确，但没有任何的注释和解释来辅助理解，不理想。

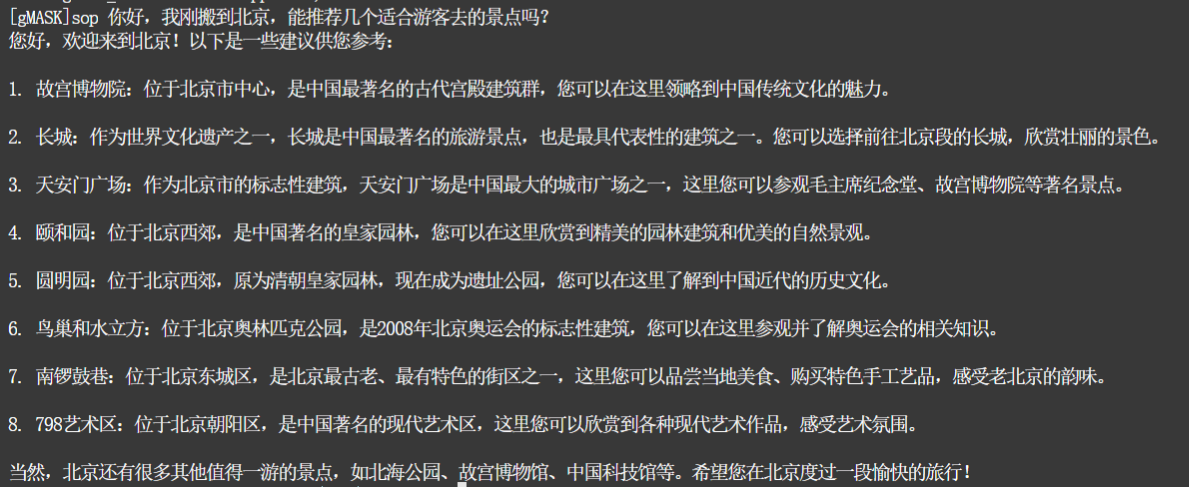
解释简短清晰，没有多余的废话，很好

只指出了两句话的语境不同，但是没有分析出最关键的语义不同，明明两句话的含义是完全相反的。

分析的五个“意思”还是比较准确的，但是还有好几个“意思”漏分析了，不知道是什么原因

解释得很到位，也满足通俗的要求

观点明确：应该优先保护乘客。第四条论证内容和观点没有任何关系，第三条论证内容自相矛盾，首先说行人属于弱势群体，人类有义务保护弱势群体，但是后面又说自动驾驶汽车应当尽可能保护乘客，完全没有逻辑性。表现也不尽如人意。



给出了8个景点，其中有传统历史景点，也有现代新兴景点，比较全面。

两个模型的比较

语义流畅性和可读性：  
通义千问：语言表达比较直接，偏向于口语化。在处理语义的问题时，倾向于给出简洁直观的回答

智谱：输出正式调理清晰，在回答在提供详细的解释和论述，在技术性的问题上表现良好。

处理复杂问题的能力：  
通义千问：在处理复杂语义的问题中，表现比较优秀，能够理解迷惑性强的语句

智谱：在这方面较弱，处理复杂语义的问题都回答的不准确。

回答的一致性和逻辑性：  
通义千问：在回答逻辑推理问题时出现了偏题的情况。

智谱：在逻辑推理和答案一致性方面表现较好。

准确性和专业知识：  
通义千问：在专业知识的回答中，内容比较丰富，但是有时显得比较冗余

智谱：在处理专业问题时，能够快速准确地给出核心信息，内容量恰到好处

通义千问的优势与不足：  
优势：语言自然，更符合日常对话的习惯

对复杂文本的理解更加准确  
不足：回答冗长，有尤其是在复杂问题上，回答显得过于详细，信息量大而不精练。

智谱的优势和不足  
优势：直接明了，没有冗余信息，直接给出关键信息  
不足：语言僵硬，在某些回答中，缺乏口语化的表达和自然流畅性。这可能在要与用户进行自然交流的场景中是一个确定。

处理复杂文本的能力有限