

## 第二轮修改

### 专家 1 给作者的意见:

作者对第一轮审稿意见中提出的问题给出了较为详细且有针对性的回应,并在文中补充了相应内容,基本解决了之前提到的问题。但还可以进一步优化。

**回应:** 非常感谢您对本文的肯定!下面我们将针对您的各项意见作出回应。

1.大模型模拟人类被试,仅仅是节省数据采集成本?文章开篇这样立意似乎把文章写小了, AI 作为一种新的 agents & products,不仅具有类人性,而且也具有相对独立性,后者作为被试似乎具有更加深远的理论和方法意义。请作者慎重考虑。

**回应:** 非常感谢专家的意见!您的这个建议真是太深刻了,一针见血地指出了我们文章开篇立意的不足之处,对我们提升全文的理论高度起到了画龙点睛的作用!

我们完全采纳了您的建议。正如您所指出的,仅仅将大模型看作节省成本的工具确实把文章“写小了”。大模型作为一种兼具类人性与相对独立性的非人类智能体,其本身作为研究对象的意义确实更为深远。根据您的启发,我们对全文的立意和框架都做了相应的提升和修改:

(1) 在引言部分,我们重写了文章的开篇。新的引言不再局限于“节省成本”的工具视角,而是开篇就点明了大模型作为模拟器和非人类智能体的双重角色,从而为全文奠定了更宏大、更深刻的理论基调。

(2) 为了让这个核心思想贯穿全文,我们还在正文的关键地方做了呼应。比如,在 4.4.2,我们将模型的“超精度扭曲”这一缺陷进一步解读为模型的“独立特性”。

(3) 最后,我们重点对“5 建议与展望”进行了修改和升华。在 5.2 部分,我们再次回应了引言提出的双重视角,并明确展望了未来研究可以并行的两条路径(深化“工具”角色与开拓“主体”角色),使全文的逻辑在结尾处形成完美的闭合。

再次由衷地感谢您提出的宝贵建议,这确实让我们的文章有了一个质的飞跃!为了与第一轮修改内容相区别,本轮修改之处,我们在修改稿中用绿色字体标出,以便您审阅。

2. 3.1 部分“由于大模型主要从互联网文本中学习”这句话不准确，因为大模型最初并没有接入互联网。建议修改，并增加参考文献。

**回应：**感谢您的细致指正！您指出的这个问题非常关键，我们之前的表述“主要从互联网文本中学习”确实不够准确，有失严谨。我们完全同意您的看法，大模型并非直接接入互联网动态学习。因此，我们根据您的宝贵建议，对这句话进行了修改，明确了“文本语料库”这一核心概念，并增加了相应的参考文献。修改如下：

由于大模型的知识与观点主要源于其训练所用的海量文本语料库——其中绝大部分数据来自互联网(Abdurahman et al., 2024)——其知识与观点分布...

再次感谢您的严谨与细致，让我们的文章在细节上更加完善！

3.正如专家 2 提出的第 2 条审稿意见，作者将“4 大模型模拟人类被试的方法路径与准确性评估”与“5 优化大模型模拟人类被试的方法”分在两处进行讨论，但是部分论点又相互联系甚至有些矛盾，让人感觉写作逻辑有所欠缺。例如，在本次修改稿中，作者新增了“4.1 方法路径”，总结了大模型模拟人类被试的两种方法，作者指出第一种方法存在“拟人化”陷阱，那么“5.1 简单提示词”中提到的为大模型赋予“人格”无法避免这个问题，又如何能够优化大模型模拟人类被试呢？类似的，第二种方法的不足也与“5.2 上下文学习”中提出的优化方法存在矛盾之处。当然，作者已经在 5.1 和 5.2 部分的最后加入了这些优化方法本身的不足，一定程度缓解了这些矛盾。但是写作逻辑上的缺陷在于，作者首先总结并评价了两种模拟人类被试的方法路径，并且指出了其缺陷，然后又在后文中说这些被指有“缺陷”的方法可以优化大模型模拟人类被试，会给读者造成一定阅读困难，建议认真思考这两部分的写作逻辑。

**回应：**非常感谢！您再次提出了一个非常关键的逻辑结构问题。我们反复阅读了您的意见，深刻认识到我们之前将“方法路径”与“优化方法”分在两章论述，确实造成了您所指出的逻辑矛盾和阅读困难。我们之前的设计是想先介绍“是什么”，再讨论“怎么做得更好”，但正如您指出的，这两者在实践中是密不可分的，强行拆分反而破坏了论述的整体性和逻辑的连贯性。

根据您的建议，我们对文章的第四、五部分进行了结构重组，将两者合并为全新的第四部分——“大模型模拟人类被试的方法、实现与评估”。为了从根本上解决您指出的逻辑问题，我们在新章节中根据技术干预的深度，建立了一个全新的分类框架：

(1) 新的 4.1 节 专门论述“基于提示工程的模拟路径”，涵盖所有不改变模型参数的方法。在这一节内部，我们再分别探讨“个体行动者模拟”和“群体水平模拟”两种具体方法，并将各自的实现技术（如简单提示词、上下文学习）、优点与风险集中阐述。

(2) 随后的 4.2 节 则专门论述“基于模型微调的模拟路径”，即直接修改模型参数的方法。

(3) 最后，在完整介绍了这两大技术路径后，我们再用专门的一节（新 4.4 节）来讨论如何对这些方法的准确性进行评估。

新的第四部分已经全部标为绿色。通过这样的重组，我们相信文章的逻辑变得更加流畅和严密，彻底解决了您指出的“先指出缺陷，又用有缺陷的方法来优化”的矛盾感，希望您能为您带来更好的阅读体验。再次感谢您的深刻洞见！

4.作为对意见 6 的回应，作者在 6.2 部分的最后增加了一段。但是在 691 行，第三点创新应用的总起句说的是大语言模型在敏感性议题上的应用，但是后面的例子却是大语言模型用于社会模拟研究。需要对这一论点进行修改，明确其与论据的对应关系。

回应：非常感谢您的意见！您敏锐地指出了我们这段论述中存在论点与论据不匹配的问题。我们原来的总起句聚焦于敏感性议题，但后续的例证（如斯坦福虚拟小镇）确实是更宽泛的“社会模拟”研究。

为了解决这个问题，我们重写了这一段（相关修改内容已经标为绿色）。在新的版本中，我们将大规模社会模拟作为核心论点，用斯坦福小镇等研究作为支持论据。然后，我们再将研究敏感性议题阐述为社会模拟平台的一个重要优势和具体应用方向进行阐述，从而确保了论点与论据的紧密对应和逻辑上的层层递进。再次感谢您的细致审阅！

5.一些笔误。如：（1）摘要的首句“随着大语言模型能力的飞速发展”，还是“随着大语言模型技术的飞速发展”？（2）本文讨论的是大语言模型模拟人类被试，而不是模拟研究参与者，因此在文中，应该注意保持措辞的准确性和一致性，例如，3.1 部分的第一句话，“大模型作为研究参与者的首要局限”建议改成“大模型模拟人类被试的首要局限”。（3）第 674 行“Wang

P 等人(2024)指出”应改为“Wang 等人(2024)指出”, 690 行“Qin 等(2024)的用 ChatGPT”应改为“Qin 等(2024)用 ChatGPT”。

**回应:** 非常感谢您的火眼金睛, 指出了我们文中的多处笔误和不一致之处! 我们已根据您的意见, 对全文进行了仔细的检查 and 修改。

(1) 关于摘要首句, 我们认为您的建议非常好, “技术”的表述确实比“能力”更严谨、更宏观。我们已经将其修改为“随着大语言模型技术的飞速发展”。

(2) 关于术语一致性的问题, 您说得完全正确。本文的核心是模拟“人类被试”。我们已经将您指出的 3.1 节中的“研究参与者”修改为“人类被试”, 并以此为原则, 对全文进行了筛查, 确保了术语的统一。

(3) 关于具体的笔误, 我们已将“Qin 等(2024)的用 ChatGPT”修正为“Qin 等(2024)用 ChatGPT”。但关于“Wang P 等人”的修改, 一般来说 APA 格式中不应包含首字母缩写, 但本文所引的参考文献恰好有两篇发表于 2024 年且第一作者姓氏同为 Wang (Wang, P. et al., 2024 与 Wang, X. et al., 2024)。根据 APA 格式的消歧规则, 必须在文内引用中加入第一作者的名字首字母以作区分。因此, 我们在文中保留了“Wang P 等 (2024)”的表达。

最后, 我们对全文的语言表达和格式进行了全面的检查与润色, 并根据修改内容调整了中英文摘要。再次感谢您的深刻洞见与细致审阅!

## **专家 2 给作者的意见:**

作者此次修改非常认真, 显著地提升了论文的质量与说服力, 我认为此篇文章已经基本符合了发表标准, 还有一些小建议请作者在发表前考虑。

**回应:** 非常感谢您对我们的肯定! 下面, 我们将对您的意见给出具体回应。

1. 在大语言模型的缺陷方面。大模型可以在文本上一定程度模拟人类心智, 并且贡献于关注外显表现的心理学研究, 但是对于那些关注反应时或者生理指标的心理学研究者来说, 大模型如何起到帮助作用, 可能还需要进一步探究。

**回应：**非常感谢您提出的这个极其重要的观点！您敏锐地指出了当前大模型作为研究被试的一个核心能力边界，这对于明确我们文章的适用范围至关重要。

我们完全同意，对于依赖反应时和生理指标的研究领域，纯文本大模型的应用价值确实是一个需要审慎探讨的问题。根据您的建议，我们在“**5.1 认清大语言模型的缺陷**”的第一段中，增加了相关论述（**相关修改内容已经标为绿色**）。在新增的内容中，我们明确指出，正是由于缺乏具身认知，大模型尚无法为注重反应时和生理指标的认知心理学及认知神经科学研究提供直接帮助，其在这方面的应用价值仍有待探索。

再次感谢您的宝贵意见，它让我们的论述更加全面和严谨！

2. 另外，在大模型能力不断提高的当下，可能更要额外提醒研究者们遵守伦理规范，避免出现利用大模型伪造数据生成论文的情况发生。

**回应：**您提出的这个伦理风险问题非常及时且至关重要！我们完全赞同您的观点，随着大模型能力的提升，防范其被用于伪造数据的风险，是学术界必须正视的紧迫议题。

根据您的建议，我们在新的“4.3 方法论风险与共同局限”中，紧接着对“研究者自由度”的讨论，增加了一小段关于数据伪造风险的论述：

另外，随着大模型模拟能力的日益成熟，我们必须高度警惕并坚决杜绝一种更为严重的伦理风险：即利用大模型凭空伪造研究数据。这种行为严重违背了科学诚信的基本原则，学术界必须对此建立明确的规范，确保这项技术被负责任地用于科学研究，而非学术欺诈。例如，在未来的研究报告中，研究者必须明确表明数据是来自真实人类被试还是 AI 模拟，并详细披露数据收集的时间、所用平台与工具、原始数据等关键信息，以确保数据收集和使用过程的可追溯性和可验证性。

3. 最后，文中还存在一些语病，比如，在“6.2 明确方法论原则与应用路径”中，“虽然大语言模型在一些心理学任务和模拟群体意见方面表现类似人类的反应”，应该是“表现出类似人类的反应”；“更关键的是要为解读模拟结果提供理论分析视角”，多了一个“理”字，请作者在发表前仔细检查全文。

**回应：**非常感谢您的耐心和细致！逐字逐句地阅读并为我们指出这些语病和笔误，真的让我们非常感动。我们已经根据您的指正，将文中的两处错误进行了修正。

同时，遵循您的宝贵建议，在完成所有内容和结构性修改后，我们又通读了一遍全文，力求消除所有可能存在的笔误和语病。再次由衷地感谢您为提升本文质量付出的辛勤努力！

## 第一轮修改

### 专家 1 给作者的意见：

文章探讨了大语言模型模拟研究被试的原理、局限、建议等重大问题，对人工智能时代的心理学研究具有重要理论和方法学意义。但作者对问题的基础逻辑论证还不够清晰，研究路径的介绍不够具体，行文中也还有一些小问题需要改进。具体建议如下：

1、本文主标题“大语言模型能在心理学研究中模拟研究参与者吗”使用问好，将能不能作为一个问题有失偏颇，因为能不能的问题已经有很多人回答过了。然而，大语言模型模拟人类被试的原理很多人不清楚，其背后的算法也在不断变化；同时，大语言模型究竟是模拟研究参与者、还是模拟人类被试，二者相关，但是两个不同的问题，其指向的研究对象也根本不同，前者指向一个独立的群体或物种，而后者指向的是一个类人的 agent。

**回应：**感谢您提出的宝贵意见。我们根据您的建议修改了本文的题目。现在的题目已经改为：《大语言模型在心理学研究中模拟人类被试的原理、局限与建议》。感谢您提供的建议！

另外需要说明的是，我们根据您与专家 2 的意见，对本文进行了大篇幅的修改。为回应二位专家意见而修改的部分，我们在正文中用蓝色标记出来了。但因为修改的地方实在太多，为了不影响您阅读，我们其余的文字调整和润色部分，就没有再标蓝了。再次感谢您的各项意见，让我们对自己的文章又有了新的理解！

2、本文试图聚焦“心理学研究”，但是许多论点和论据其实指向更一般的社会科学，例如社会民意调查。心理学研究的方法和范式与其他社会科学相比有共通之处，但是也有许多独特性，

例如量表法、实验法等。建议对心理学研究方法进行更细致的区分，并补充更多大语言模型在心理学研究中的应用，探讨其表现和适用性？

**回应：**感谢您的宝贵反馈！根据您的建议，我们重新撰写了文章的第一部分。我们用三个自然段，依次论述了大模型在实验研究、问卷研究、访谈研究中对人类研究参与者的模拟情况。具体修改内容已标蓝。

3、大语言模型的局限是本文的三个主要问题之一，但正文仅在不同的部分分析了其代表性偏差、模型缺陷，缺乏系统的论证。另外还有两点局限，请作者参考：（1）与社交媒体类似的虚假信息、观点极化问题；（2）大语言模型只是模仿了人类的语言、而非人类全部，并且人类语言本身也存在很多问题。

**回应：**非常感谢您一针见血地指出了本文的不足，并提出了两个非常好的补充建议。这为提升我们的文章质量提供了很好的契机！我们重写了第三部分（3.大模型模拟人类研究参与者的局限）。具体来说，就是把第三部分分成了两个部分。第一部分讨论的是此前提及的代表性偏差，这一部分我们稍微调整了语言，主要是压缩篇幅，变成了现在的 3.1 部分。然后我们增加了 3.2 部分（3.2 数据内容的潜在风险），增加了两段新内容，就是大模型造成的虚假信息风险，以及观点极化与偏见固化风险。当然，第三部分的总结段也根据新内容做了适当调整。修改部分已经标记为蓝色。再次感谢您提出的宝贵建议！

4、本文对具体方法层面的讨论较少，仅在第 5 节介绍了部分优化大语言模型模拟人类参与者的方法，但是对大语言模型如何应用于心理学研究缺少总结。目前大语言模型模拟人类参与者的方式主要有两种，一种是要求其模拟群体，即推测某个群体在某些问题上的平均反应，如本文引用的 Simmons & Savinov (2024)；另一种是要求其模拟个体，即作为某个群体中的个体或者具有某种特质的个体，对某些问题进行反应，如本文引用的 Aher 等(2023)。这两种模拟方式有何适用范围和优劣？建议对具体方法进行介绍和评价。

**回应：**您提出的意见和参考文献都非常有价值，对我们启发很大！我们在修改稿中将第四部分由原来的准确性评估扩展为两个方面：模拟方法与准确性评估。在“4.1 方法路径”部分，我们增加了两段内容，专门探讨这一重要但之前被我们忽略的内容。新增部分已标蓝。

5、本文在第6节“建议与展望”中的第2小节提出，“在有效性评估标准上，研究者应致力于发展超越简单平均值比较的方法根据上述思路”，虽然当前似乎没有研究提出具体的评估标准，但是有一些研究在更细致的层面进行了探索。例如，Wang 等人(2024)指出，虽然大语言模型在一些心理学任务和模拟群体意见方面表现类似人类的反应，但是从心理测量学的角度对项目层面的相似性进行检验，却发现大语言模型反应与人类反应存在明显差异，大语言模型能复现整体变异性，却无法捕捉个体差异模式。建议本文纳入更多的对大语言模型代替人类参与者的相关研究，并对其进行更全面的探讨。

参考文献：Wang, P., Zou, H., Yan, Z., Guo, F., Sun, T., Xiao, Z., & Zhang, B. (2024, September 23). Not Yet: Large Language Models Cannot Replace Human Respondents for Psychometric Research. <https://doi.org/10.31219/osf.io/rwy9b>

回应：感谢您的宝贵意见！意见和参考文献都非常棒！我们已经将该意见整合到“6.2 明确研究方向”的第一段中，相关部分已标蓝。感谢！

6、第6节对大语言模型的优势和劣势以及社会科学研究自动化有较为宏观的讨论，但缺乏可操作性的研究路径。例如，Qin 等(2024)建议将大语言模型模拟人类参与者作为一种快速且低成本的研究方法，用于预研究、敏感问题研究、构建“玩具模型”等方面。建议在结论与展望中，列出3~5个具体可行的研究方向，提出更具体的研究建议。

参考文献：Qin, X., Huang, M., & Ding, J. (2024). AITurk: Using ChatGPT for social science research. <http://dx.doi.org/10.2139/ssrn.4922861>

回应：非常感谢您的这个建议，让我们的思考落实到具体的未来研究思路的展望上。我们在6.2最后增加了一段，依次介绍了大模型作为预研究的工具、复现已有研究，以及模拟研究的三个具体方向。相关部分已标蓝。相信新增加的这一段能让读者有更具体的收获！

7、目前大语言模型的偏见问题已经积累了大量研究，并且随着大语言模型的发展，其偏见和对齐失败问题已经得到改善，例如Bai等(2025)发现，大语言模型可以通过显性的社会偏见测验，但是仍然存在隐性偏见。本文中仅引用了Santurkar等(2023)和Durmus等(2024)的两篇研究，略显单薄，建议纳入较新的研究，进一步探讨该问题。

参考文献：Bai, X., Wang, A., Sucholutsky, I., & Griffiths, T. L. (2025). Explicitly unbiased large

language models still form biased associations. *Proceedings of the National Academy of Sciences*, 122(8), e2416228122.

**回应：**非常感谢您的这个建议。这让我们将宏观的、外显的代表性偏差扩展到了微观的、内隐的层次，非常有启发！在修改稿中，我们在 3.1 代表性偏差这部分末尾增加了一段内容，就是现在 3.1 部分的最后一段。这一段专门论述大模型的内隐偏见问题。相关内容已标蓝。

8、当前第 4 节仅介绍了简单提示词、上下文学习和微调三种策略，而大语言模型正逐步向多模态（图像、声音等）方向发展。建议指出在一些心理学研究（如视觉认知或情感表达）中，如果仅以文本为输入，模拟效果的局限性，并简要讨论未来可以借助“多模态大模型”来更真实地模拟人类参与者的可能路径。

**回应：**非常感谢您提出的宝贵建议！我们在“5. 优化大模型模拟人类研究参与者的方法”介绍完三种基于文本大模型的优化方法后，在本部分的结尾新增了一段对这些方法的局限做出了说明：都是基于文本的，而没有考虑到多模态问题，新增内容如下：

除了上述方法论层面的风险，还需要指出的是，本节讨论的所有优化策略都存在一个共同的技术局限：它们都是基于纯文本的大模型。然而，大模型技术的发展正迅速超越单一文本模态。我们将在 6.3 部分阐明多模态的大模型如何为模拟更真实、更丰富的心理现象带来新的机遇。

然后，我们在“6. 建议与展望”将原先仅探讨研究自动化的 6.3 节更新为：“6.3 大模型技术迭代可能产生的新方向”，探讨大模型技术发展带来的两个方向：多模态与研究自动化。我们新增了一段内容，专门探讨多模态大模型的发展及其对模拟人类研究参与者的贡献。位置在 6.3 的第二段。相关部分已标蓝。

9、修改文中的笔误、小标题数字序号等错误。例如，第 2 节第一段，“大模型模拟(图 2 右侧蓝色部分)”应为“大模型模拟(图 1 右侧蓝色部分)”；第 5 节的标题应删去“地”；第 4 节之后的二级标题序号与一级标题序号不一致。

回应：非常感谢您这么认真地阅读我们的初稿，感谢！图 1 排版有误，现在已经改正！第五节标题有误，也已改正。文章的标题序号出错的地方也都改正了。我们还对全文进行了检查。感谢您的意见！

## 专家 2 给作者的意见：

本文从利用 AI 来模拟甚至替代人类参与者的潜在用途出发，结合当前大语言模型发展阶段，分析总结了使用大语言模型模拟人类参与者的原理、潜在偏差以及提升准确性的方法。本文叙述清楚简洁、证据充足，文章内容在当下具有重要意义。不过，还有一些有助于提升文章质量的小建议：

1. 在“3.1 基于算法保真度的评估标准”部分，作者总结了评估算法保真度的四个标准，在这些标准的基础上，是否有研究表明当前大模型发展水平在多大程度上符合了这些标准。如果有相关研究结果，能够让读者更直观地理解当前大模型的发展阶段。

回应：感谢您提出的这个意见，这启发我们去思考如何将评估标准与具体的实证研究结合起来。我们找到了一些研究来说明当前大模型对于这四个标准的满足情况。我们在 4.2 最后一部分增加了一段（4.2 的最后一段），相关部分已标蓝。

另外需要说明的是，我们根据您与专家 1 的意见，对本文进行了大篇幅的修改。回应二位的意见修改的部分，我们在正文中用蓝色标记出来了。但因为修改的地方实在太多，为了不影响您阅读，我们其余的文字调整和润色部分，就没有再标蓝了。再次感谢您的各项意见，让我们对自己的文章又有了新的理解！

2. 在“4.1 简单提示词”部分，作者介绍可以通过简单的提示词为大模型赋予特定属性，但是在“3 大模型的代表性偏差”部分，作者提到通过提示词来引导模型观点靠近目标国家的转变往往依赖于简单化甚至刻板的文化印象，而非深层理解。那么，这是否表明通过提示词为大模型赋予特定属性也存在过于简单化和刻板化的风险呢。

**回应：**非常感谢您提出的这个深刻见解！您提到的这两个点确实是相通的！因为简单提示词与代表性偏差的内在逻辑是一致的：就是大模型被赋予一个笼统而抽象的标签时（国别或者某种性格特点），它就会从海量的训练数据生成的神经网络模型中检索并激活与这个标签最相关、最具普遍性的统计模式。从另一方面看，这就是刻板印象了。仔细想想，这与人类的社会认知加工的过程具有相似之处。所以，我们在 5.1 最后增加了一段，用来表达上述内容，并提示读者注意简单提示词方法的局限所在。相关内容见 5.1 部分最后一段，已标蓝。

3. 作者提出两种优化大模型模拟人类研究参与者的方法，即“4.2 上下文学习”以及“4.3 大模型微调”。不过这两种方法都需要研究者为大模型提供部分数据或者嵌入特定信念，那么是否会存在研究者人为操纵数据结果的风险呢？

**回应：**感谢您的意见。您提到的这一点确实非常重要，一种方法越是强大，其被研究者人为操纵的风险也越高。我们在第五部分增加了一段（第五部分的倒数第二段），探讨上下文学习和大模型微调带来的研究者自由度的新问题。新增内容已标蓝。

4. 作者基于自然科学领域的科学发现自动化，提出了“社会科学研究自动化转型”的构想。然而，要注意社会科学研究与自然科学研究的一个主要区别就在于，自然科学的研究对象大多是客观存在并且不受主观意识转移的，但是社会科学的研究对象主要是真实人类的主观感受。因此，要强调尽管大模型可能会通过自动化一定程度上加快社会科学研究效率，但是无法完全替代真实人类参与者在社会科学研究中的核心角色。

**回应：**感谢您的提醒！虽然自然科学与社会科学的研究对象的不同是众所周知的，但是放在本文的语境中确实很容易让读者造成误解——很容易让读者误以为社会科学研究也能像自然科学一样实现完全的自动化。我们在 6.3 的第三段的最后增加了一小段来提醒甚至警醒读者：

然而，我们必须警醒的是，社会科学与自然科学存在一个根本差异：自然科学的研究对象是客观实体，其规律不以研究者的意志为转移。但是，社会科学，尤其是心理学，研究的却是拥有自由意志和内在主观世界的人类。人类有自由意志，因此是不可能被完全预测的。这一本质差异决定了研究的自动化在社会科学中永远存在一个无法逾越的边界：技术可以无限逼近地“模拟”，却永远无法真正地“成为”或“取代”研究的参与者，乃至研究者本人。

5. 文章存在一些格式错误，比如在“4. 评估大模型模拟人类研究参与者的准确度”下面的两个小标题为 3.1 和 3.2，应该是 4.1 和 4.2，后文也存在类似问题，请调整。另外，文章还存在一些语病，比如“5. 优化大模型地模拟人类研究参与者的方法”，其中“地”应该被删去，请作者全文检查并调整。

**回应：**感谢您的耐心和细致的阅读。我们已经修改了您指出的这些问题，并对全文进行检查和调整。再次感谢您的意见！