

## 《心理科学进展》审稿意见与作者回应

题目：人工智能服务对消费者道德行为的双刃剑效应

作者：彭晨明 屈奕帆 郭晓凌 陈增祥

---

### 第一轮

审稿人意见：

主要意见：

论文选题属于目前的研究前沿，有比较大的学术贡献。论文整体结构合理，论述清晰，研究推演充分，有一定的学术价值。

回应：

非常感谢您们对我们论文的评审与宝贵意见。您的建议使我们对研究内容有了更加深入的理解，获益良多！针对评审意见，我们已仔细修改了文章，但仍可能存在不足之处，恳请您们继续指正，帮助我们进一步完善论文。

本次我们共收到 4 条修改建议，针对审稿人提出的具体问题，我们逐一进行了回应。为了便于评审，在论文中，我们已将修改内容用蓝色标出，而在修改说明中，原文和修改后的部分均用蓝色标注。

**意见 1：**虽然作者界定了人工智能，但是，就目前的选择来看，研究比较宽泛，服务型 AI 的类别和功能划分实际上对于研究有着重要的意义。所以，建议作者进行一定的思考。特别是在做调节变量的选择时，可以进一步考虑。

回应：

非常感谢您的建议。通过回顾市场营销领域和信息系统领域重要期刊，我们补充了与本文理论最相关的三个 AI 服务特征作为调节变量，即人机协作、服务风格、任务类型。因此，我们在“3.3 研究三：AI 服务对消费者道德行为影响的边界条件研究”中补充了以下内容：

在 AI 服务中，可能存在人机协作，即 AI 服务与人类服务相互分工协作，形成一个整体为客户提供服务(Fügener et al., 2021; Gnewuch et al., 2023; Makarius et al., 2020)。本文提出，人机协作可调节 AI 服务对消费者禁止性道德行为和倡导性道德行为的影响。已有研究

发现，人类与 AI 的分工协作，如 AI 和人类教练的组合训练，可能会取得超越人类和 AI 单独工作的成果(Fügener et al., 2021; Luo et al., 2021; Sowa et al., 2021)。消费者通常认为，机器在认知方面有优势，比如处理复杂计算或海量数据，而人类则在社会和情感技能方面有优势(Longoni & Cian, 2022; Xu & Mehta, 2022; Pittman, 2016)。据此本文推测，对于 AI 服务，如果采取人机协作，消费者会感知到人类在情感方面的优势，从而弥补 AI 在情绪方面的劣势，减轻 AI 服务对于道德行为的负面影响。基于此，提出如下假设：

**假设 H10a:** 人机协作会调节 AI 服务 (vs. 人类服务) 对消费者倡导性道德行为的影响。当采用人机协作时，AI 服务提高消费者的内疚和恐惧情绪，进而减弱对禁止性道德行为的负面影响。

**假设 H10b:** 人机协作会调节 AI 服务 (vs. 人类服务) 对消费者禁止性道德行为的影响。当采用人机协作时，AI 服务提升消费者的同情和自豪情绪，进而减弱对倡导性道德行为的负面影响。

从 AI 服务风格角度看，情感型或认知型 AI 服务可以调节 AI 服务对消费者禁止性和倡导性道德行为的影响。AI 服务可划分为情感型和认知型两类，前者指具备识别、理解、模拟和回应用户情感能力的 AI 系统，后者指 AI 通过模拟人类的学习、推理、问题解决和决策来满足用户需求(Castelo et al., 2019; Huang & Rust, 2018; Yu et al., 2023)。以虚拟客服为例，当用户表现出愤怒时，情感型 AI 会通过诸如“换我我也生气”的话语共情，抚慰用户情绪；认知型 AI 则会直接提供详细解决方案，不会进行情感回应。现有研究表明，认知能力是机器人基本的、固有的可预期特征(Kim et al., 2019)，而情感能力通常是区分人类和机器人的关键所在(Grazzini et al., 2023)。因此，情感型 AI 服务能够通过模拟人类情感来缩小消费者在使用 AI 服务和人类服务之间的差异感知。当 AI 具备情感表达能力时，消费者会倾向于将其视为更具人性化的服务提供者，提升对 AI 的信任感和接受度(Huang & Rust, 2024)。据此，本文推测，更具情感能力的 AI 服务会促使消费者同等对待 AI 服务与人类服务，从而减弱 AI 服务在道德情绪方面的劣势及对道德行为的负面影响。基于此，本研究假设如下：

**假设 H11a:** 情感型 (vs. 认知型) AI 服务会调节 AI 服务 (vs. 人类服务) 对消费者禁止性道德行为的影响。情感型 AI 服务提高消费者的内疚和恐惧情绪，进而减弱对禁止性道德行为的负面影响。

**假设 H11b:** 情感型 (vs. 认知型) AI 服务会调节 AI 服务 (vs. 人类服务) 对消费者倡导性道德行为的影响。情感型 AI 服务提升消费者的同情和自豪情绪，进而减弱对倡导性道德行为的负面影响。

最后，从 AI 服务的任务类型看，客观性任务可以调节 AI 服务对消费者禁止性道德行为和倡导性道德行为的影响。客观性任务指可量化任务，通常需要逻辑和规则，与“认知”能力相关，如帮助用户计算五险一金缴费金额。相比之下，主观性任务指那些可解释的、依赖于个人意见或直觉的任务。这类任务通常需要直觉判断，结果往往受到个人经验和情感因素影响(Castelo et al., 2019; Inbar et al., 2010)，例如根据用户偏好推荐电影。在处理客观性任务时，个体通常依赖较慢且更多思虑的推理过程，即理性认知系统（vs. 感性系统）。道德认知，包括对被监控的意识、对不道德行为可能受到惩罚的感知以及对公平的判断，与认知处理相关(Suhler & Churchland, 2011; Greene et al., 2004; Young & Dungan, 2012)。因此，当 AI 服务用于完成客观任务时，消费者会更多地参与理性认知系统（vs. 感性系统）的认知推理，思考道德后果，提高对道德规范的敏感性，由此增强 AI 服务对于道德认知的正面影响。基于此，提出以下假设：

**假设 H12a:** 客观性（vs.主观性）任务会调节 AI 服务（vs. 人类服务）对消费者禁止性道德行为的影响。面对客观性任务时，AI 服务提高消费者监测可能性感知和惩罚公正性感知，进而增强对禁止性道德行为的正面影响。

**假设 H12b:** 客观性（vs.主观性）任务会调节 AI 服务（vs. 人类服务）对消费者倡导性道德行为的影响。面对客观性任务时，AI 服务提高消费者行为自主性感知和服务公正性感知，进而增强对倡导性道德行为的正面影响。

意见 2：“双刃剑”效应其实对于大多数的科技来说都是存在的，关键还是讨论形成机理，尤其是对于中介的调节更为重要。所以也建议作者可以从这方面加大假设的提出。另外，“双刃剑”效应可能同时出现，这一点作者可能也需要思考。

回应：

衷心感谢您的建议。我们对论文进行了如下修改：

第一，我们修改了“3.3 研究三：AI 服务对消费者道德行为影响的边界条件研究”中的假设，强调了调节变量如何通过影响中介而最终影响自变量和因变量的关系，具体如下：

**假设 H9a:** AI 拟人化程度会调节 AI 服务（vs. 人类服务）对消费者禁止性道德行为的影响。AI 拟人化程度高时，AI 服务提高消费者的内疚和恐惧情绪，进而减弱对禁止性道德行为的负面影响。

**假设 H9b:** AI 拟人化程度会调节 AI 服务 (vs. 人类服务) 对消费者倡导性道德行为的影响。AI 拟人化程度高时, AI 服务提升消费者的同情和自豪情绪, 进而减弱对倡导性道德行为的负面影响。

**假设 H10a:** 人机协作会调节 AI 服务 (vs. 人类服务) 对消费者倡导性道德行为的影响。当采用人机协作时, AI 服务提高消费者的内疚和恐惧情绪, 进而减弱对禁止性道德行为的负面影响。

**假设 H10b:** 人机协作会调节 AI 服务 (vs. 人类服务) 对消费者禁止性道德行为的影响。当采用人机协作时, AI 服务提升消费者的同情和自豪情绪, 进而减弱对倡导性道德行为的负面影响。

**假设 H11a:** 情感型 (vs. 认知型) AI 服务会调节 AI 服务 (vs. 人类服务) 对消费者禁止性道德行为的影响。情感型 AI 服务提高消费者的内疚和恐惧情绪, 进而减弱对禁止性道德行为的负面影响。

**假设 H11b:** 情感型 (vs. 认知型) AI 服务会调节 AI 服务 (vs. 人类服务) 对消费者倡导性道德行为的影响。情感型 AI 服务提升消费者的同情和自豪情绪, 进而减弱对倡导性道德行为的负面影响。

**假设 H12a:** 客观性 (vs. 主观性) 任务会调节 AI 服务 (vs. 人类服务) 对消费者禁止性道德行为的影响。面对客观性任务时, AI 服务提高消费者监测可能性感知和惩罚公正性感知, 进而增强对禁止性道德行为的正面影响。

**假设 H12b:** 客观性 (vs. 主观性) 任务会调节 AI 服务 (vs. 人类服务) 对消费者倡导性道德行为的影响。面对客观性任务时, AI 服务提高消费者行为自主性感知和服务公正性感知, 进而增强对倡导性道德行为的正面影响。

**假设 H13a:** 消费者权力距离会调节 AI 服务 (vs. 人类服务) 对消费者禁止性道德行为的影响。当消费者的权力距离信念低时, AI 服务提高消费者的惩罚公正性感知, 进而增强对禁止性道德行为的正面影响。

**假设 H13b:** 消费者权力距离会调节 AI 服务 (vs. 人类服务) 对消费者倡导性道德行为的影响。当消费者的权力距离信念低时, AI 服务提高消费者的服务公正性感知, 进而增强对倡导性道德行为的正面影响。

**假设 H14:** 捐金钱 (vs. 捐时间) 会调节 AI 服务 (vs. 人类服务) 对消费者倡导性道德行为的影响。在考虑捐金钱的情景下, AI 服务提高消费者服务公正性感知和行为自主性感知, 进而增强对倡导性道德行为的正面影响。

**假设 H15:** 不道德行为的危害程度会调节 AI 服务 (vs. 人类服务) 对消费者禁止性道德行为的影响。当不道德行为的危害程度高时, AI 服务提高消费者监测可能性感知和惩罚公正性感知, 进而增强对禁止性道德行为的正面影响。

第二, 我们修改了“图 1 整体理论框架图”和“图 4 研究三的理论框架图”, 对应假设的提出, 我们在图中对调节变量的影响路径和方向进行了标注, 具体如下:

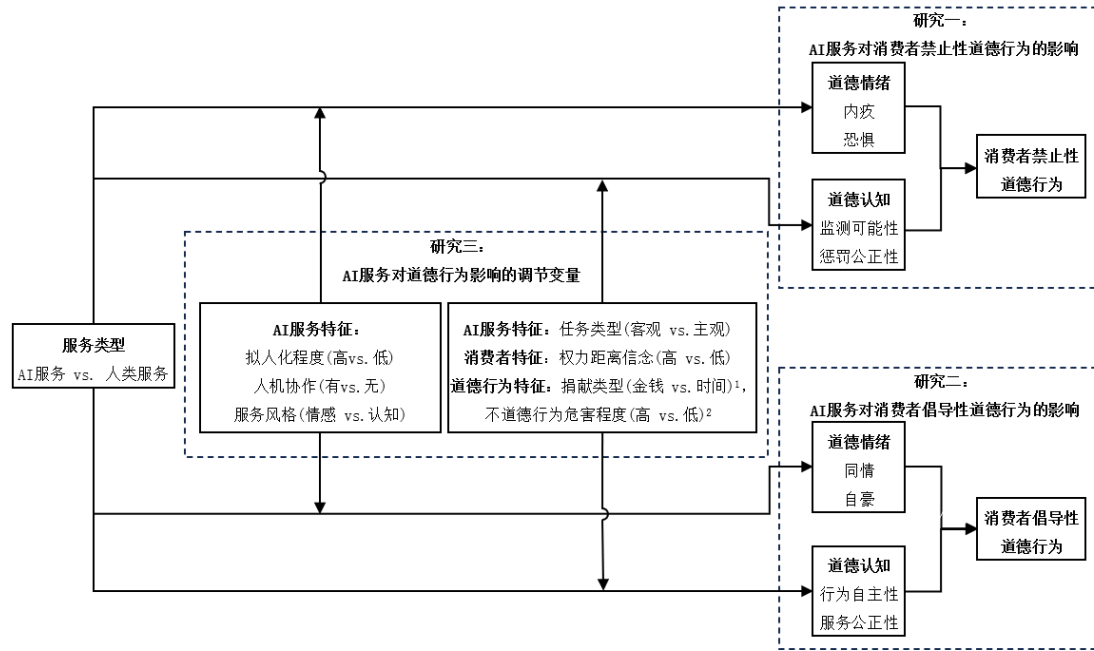


图 1 整体理论框架图

注：1 仅调节倡导性道德行为, 2 仅调节禁止性道德行为

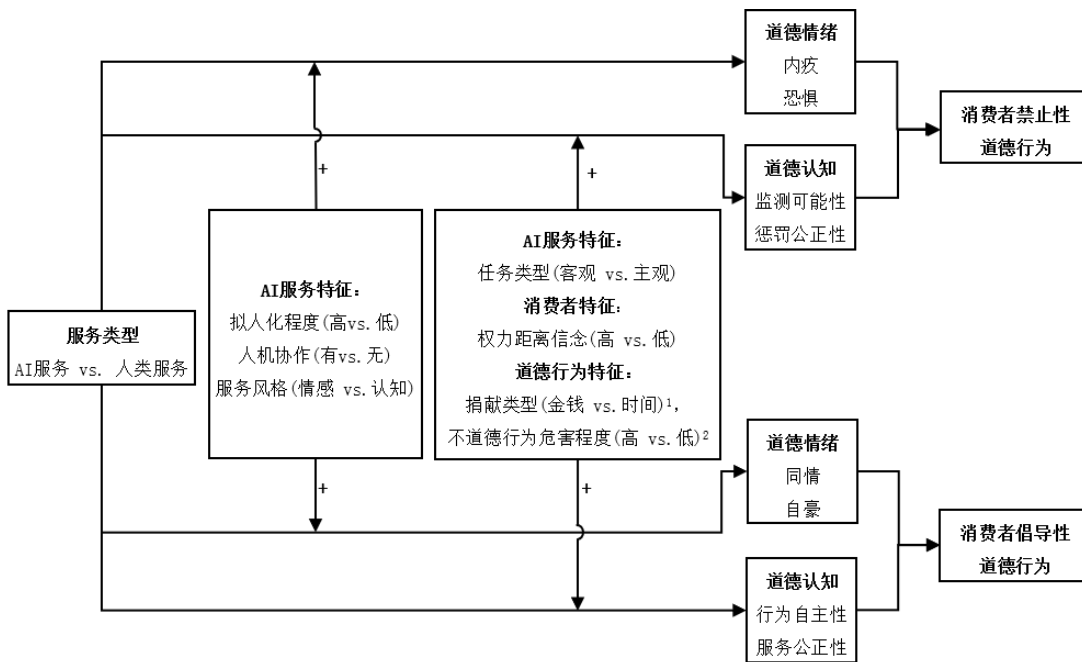


图 4 研究三的理论框架图

注：1 仅调节倡导性道德行为，2 仅调节禁止性道德行为

第三，我们在 3.1 章节“研究一：AI 服务对消费者禁止性道德行为的双刃剑效应机制研究”和 3.2 章节“研究二：AI 服务对消费者倡导性道德行为的双刃剑效应机制研究”中更加突出了双刃剑可能同时存在的论述，具体如下：

研究一从道德情绪（包括内疚和恐惧）和道德认知（包括监测可能性和惩罚公正性）两个角度解释了 AI 服务如何同时通过道德情绪和道德认知两条路径对禁止性道德行为产生影响。预期内疚和恐惧这两个道德情绪因素可解释 AI 服务对消费者禁止性道德行为的负面影响，而监测可能性和惩罚公正性这两个道德认知因素则可解释 AI 服务的正面影响。这些因素主要影响禁止性道德行为，而非倡导性道德行为。

研究二从道德情绪（同情和自豪）和道德认知（行为自主性和服务公正性）两个角度解释了 AI 服务如何同时通过道德情绪和道德认知两条路径对倡导性道德行为产生影响。同情和自豪这两个道德情绪因素能够解释 AI 服务对消费者倡导性道德行为的负面影响，而行为自主性和服务公正性这两个道德认知因素则可解释 AI 服务的正面影响。这些因素同时影响倡导性道德行为，而非禁止性道德行为。

意见 3：作者在画图的时候，最好把正向或负向的影响标注清楚，便于读者阅读。

回应：

非常感谢您细致的评审。我们在“图 2 研究一的理论框架图”、“图 3 研究二的理论框架图”和“图 4 研究三的理论框架图”标记上了正向或负向符号，具体如下：

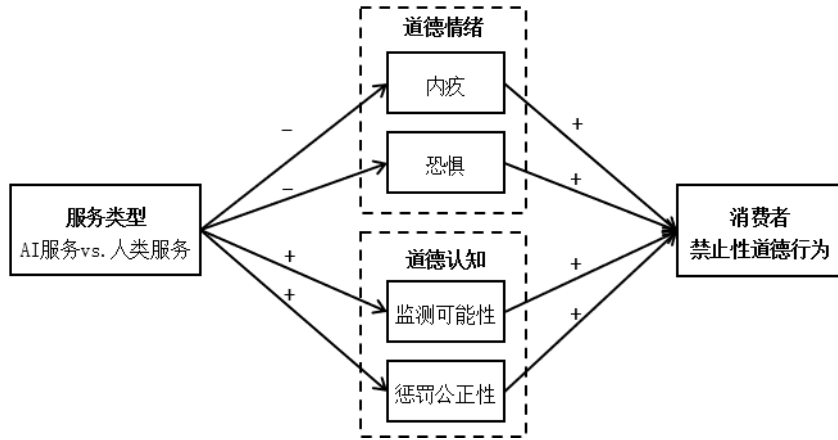


图 2 研究一的理论框架图

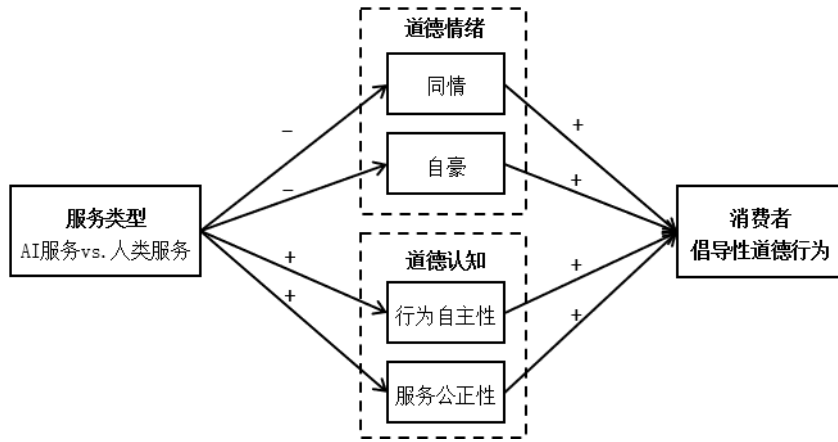


图 3 研究二的理论框架图

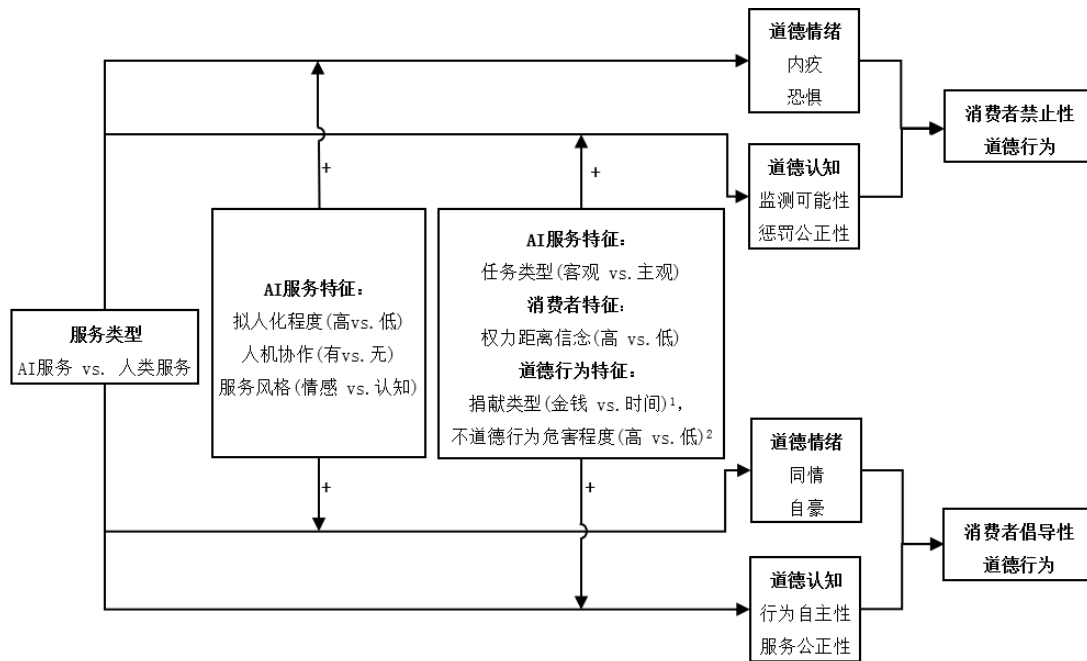


图 4 研究三的理论框架图

注：1 仅调节倡导性道德行为，2 仅调节禁止性道德行为

意见 4：作者是否可以在最后对未来研究做一些展望？

回应：

衷心感谢您的建议，我们在文末增加了 4.2 未来研究方向，具体如下：

#### 4.2 未来研究方向

本文提出了未来可供进一步研究的方向。本研究识别出 AI 服务、消费者心理和道德行为三方面特征的边界调节作用，特别是对 AI 服务所涉及的多个变量进行了深度剖析，试图较为全面地揭示 AI 服务相比人类服务对道德行为影响的复杂性。未来研究可以从国家层面识别调节变量，从而更深入地理解 AI 服务在全球情境中的道德影响。国际商务研究的制度理论指出国家层面的变量可分为经济、技术、文化、政治等方面(North, 1990; Berry et al., 2010; Kozlenkova, Lee, Xiang, Palmatier et al., 2021)。其中，我们认为技术、文化、社会因素可能是重要的调节变量，这些因素在不同国家背景下对 AI 服务的道德效应可能会产生显著影响。

从技术角度看，技术发展将使 AI 服务更为高效和智能，探究消费者的道德认知和道德情绪将如何受到影响(Dietvorst et al., 2015; Gursoy et al., 2019)，具有重要的学术和实践意义。另外，随着科技的发展，消费者对算法性能的认识也逐步深化，算法偏好不断增强(Castelo et al., 2019)，这将如何影响 AI 服务引发的道德认知和道德情绪，以及后续的道德行为，未来值得进一步探讨。



在文化层面，个体主义和集体主义文化差异对个人和组织行为的影响一直受到关注(Hofstede, 1984)。在个体主义文化中，消费者更注重自我利益和个人选择，倾向于基于个人判断进行道德决策，因此对 AI 的接受度有限；而在集体主义文化中，消费者则更注重群体责任，更容易接受 AI，这可能增强 AI 服务对群体性道德行为的影响(Shavitt et al., 2006; De Freitas et al., 2023)。因此，探讨不同的文化价值观（如个人主义和集体主义）如何影响 AI 服务对道德的作用，可以更进一步深化对于 AI 服务与道德行为关系的认识。

在社会层面，研究表明感知腐败水平可能会影响消费者对 AI 的态度。在高腐败环境中，消费者往往认为 AI 算法决策更加公正和透明，因此对 AI 的信任度更高，政府也会运用 AI 工具自下而上反腐以提高政务透明度和公正性(Castelo, 2024; Köbis, 2022)。可推测，腐败社会相较于清廉社会，AI 服务在道德认知方面的优势将更为突出，AI 可能更有效地促进消费者道德行为。未来研究可以重点关注 AI 在腐败或清廉的社会环境中对消费者道德行为的影响有何差异，从而为打击腐败、提高全社会及全人类道德水准提供洞察。

综上，我们呼吁未来研究关注跨国层面的技术、文化和社会因素，这将有助于理解 AI 服务如何在全球不同的宏观环境中发挥作用。此类研究不仅可以为中国企业“走出去”战略提供支持，还可为中国在全球市场实现可持续发展、文化融合以及国际合作提供指导。

再次感谢您付出宝贵时间与心血为我们提出建设性意见。我们诚挚希望此次的修改和回复能够令您满意，如有任何疑问，请不吝赐教，我们一定继续认真修改，使论文不断完善。

祝您工作顺利，万事如意！

---

## 第二轮

### 审稿人意见：

论文聚焦前沿，结构合理，论述清晰，有一定的理论贡献，对于未来研究有一定的指引作用。

### 编委意见：

该文整体质量较好，达到发表要求，但篇幅较长，建议尽可能精简。

### 回应：

非常感谢编委会对我们稿件的悉心复审与指导。在不改变内容要义的前提下，我们对全文内容进行精简凝练，最终共删减字数 1296 字，目前全文字数为 18887 字。