《心理学报》审稿意见与作者回应

题目：观点采纳塑造人智协作中的创造性表现

作者：张曼，彭洋，王雅洁，胡慧清，陈石，赵庆柏

第一轮

回应说明:

根据审稿专家三关于编码一致性的建议，鉴于 PIT 任务中基于 Cohen's κ 方法计算得到的即时观点采纳和延迟观点采纳行为一致性相对较低，为提高编码结果的可靠性，本研究请两名原类别编码者依据编码手册对 PIT 任务中的想法类别重新进行独立编码。重新编码后，PIT 任务想法类别划分的一致性达到良好水平，加权 Cohen's κ 由原先的 0.58 提高至 0.85；基于重新编码后的类别结果计算得到的即时观点采纳和延迟观点采纳行为一致性也达到可接受水平，Cohen's κ 分别由 0.52 和 0.50 提高至 0.76 和 0.74。因此，修改稿中 PIT 任务的观点采纳指标及相关分析均改为基于重新编码后的类别结果。

相较于编码前的结果，重新编码后，PIT 中即时观点采纳(IOC)和延迟观点采纳(DIOC)的组间以及 IOC 的角色间差异保持稳定；DIOC 的角色间比较中，原先“人智协作人类显著高于双人协作人类”的结果不再显著。APIM 分析中，IOC 和 DIOC 对创造性表现的核心预测模式总体稳定，但新颖性维度出现新的角色调节效应，且 AI 的 DIOC 对自身和人类同伴想法流畅性的预测关系由显著变为不显著。下表为重评前和重评后结果比较及结果变动说明。

分析内容	重评前结果	重评后结果	变动说明
组间 IOC 差异	人智协作显著高于双人协作，$p<0.001$	人智协作显著高于双人协作，$p<0.001$	显著性稳定
组间 DIOC 差异	人智协作显著高于双人协作，$p<0.001$	人智协作显著高于双人协作，$p<0.001$	显著性稳定
角色间 IOC 差异	AI 显著高于人智协作中的人类和双人协作中的人类；两组人类差异不显著	结果模式相同	显著性稳定
角色间 DIOC 差异	AI 显著高于两组人类；人智协作中的人类显著高于双人协作中的人类，$p=0.014$	AI 仍显著高于两组人类；但两组人类差异不再显著，$p=0.099$	主要变化：人类组间 DIOC 差异由显著变为不显著
IOC 的 APIM 主效应	IOC 对流畅性的行动者效应和同伴效应显著；IOC 对新颖性的行动者效应显著	上述主效应仍显著	主要显著性稳定

分析内容	重评前结果	重评后结果	变动说明
IOC 的 APIM 调节效应	组别和角色调节作用均不显著	新颖性模型中出现显著角色调节：同伴 IOC × 角色显著，p=0.038	新增显著调节效应
DIOC 的 APIM 主效应	DIOC 对流畅性和灵活性的主要预测作用显著	DIOC 对流畅性和灵活性的主要预测作用仍显著；对新颖性的行动者效应也显著	主要显著性稳定
DIOC 对新颖性的 APIM 效应	未发现显著角色调节	同伴 DIOC × 角色显著，p=0.046；在人智协作中，AI 的 DIOC 负向预测人类新颖性，AI 自身 DIOC 也负向预测 AI 新颖性	新增显著角色调节及简单斜率结果
DIOC 对流畅性的简单斜率	人智协作中，AI DIOC 显著预测人类流畅性；AI 自身 DIOC 显著预测 AI 流畅性	这两个效应降为边缘/不显著，*ps*≈0.06；人类自身 DIOC 预测人类流畅性、人类 DIOC 预测 AI 流畅性仍显著	部分条件效应由显著变为不显著
DIOC 对灵活性的简单斜率	双人协作中同伴 DIOC 显著预测灵活性；人智协作中人类自身 DIOC 显著预测人类灵活性，人类 DIOC 显著预测 AI 灵活性	结果模式相同	显著性稳定
DIOC 对适用性的 APIM 效应	未发现显著观点采纳相关效应	同伴 DIOC × 组别达到显著，p=0.047，但简单斜率均不显著	新增一个交互项显著，但条件效应解释有限

审稿专家 1：

以下意见供作者参考：

意见 1

（1）文章选用了替代用途任务（AUT）和产品改进任务（PIT）作为研究范式。建议作者在文中明确阐述这两种任务在创造性研究领域的权威性与代表性。为了增强说服力，建议引用相关领域的经典文献作为支撑。

回应：衷心感谢您对本文的细致审阅以及所提出的宝贵意见。我们同意原稿中对 AUT 和 PIT 任务权威性与代表性的说明不够充分。修改稿已在引言部分补充相关说明和经典文献依据。具体而言，AUT 是创造力研究中经典的发散思维任务，要求参与者针对常见物品提出尽可能多的新颖用途(Guilford, 1967)；PIT 源自 Torrance 创造思维测验中的产品改进活动，要求

参与者对常见产品提出改进或优化方案，能够较好反映现实情境中的创造性问题解决(Torrance, 2008)。通过同时采用 AUT 和 PIT，本研究能够在经典发散思维任务和现实问题解决任务中比较双人协作与人智协作的创造性表现差异。具体修改请参见修改稿引言部分“1.3 研究目的”小节(p.22)。

意见 2

（2）在进行统计功效分析以确定样本量时，作者将预估效应量设定为 d=0.55。这一数值相对较高（属于中等偏大效应）。请作者详细说明设定该数值的具体理由，例如，是基于前期的预实验数据，还是参考了特定的同类研究？

回应：感谢您的宝贵建议。我们同意原稿中对预估效应量 *d*=0.55 设定的理由不够充分。修改稿已补充该效应量设定的理由。具体而言，本研究参考 Urban 等(2024)在产品改进任务中报告的组间差异效应量，并采用其中较为保守的 *d*=0.55 作为先验功效分析中的预估效应量。该研究比较了使用或不使用 ChatGPT 下的创造性问题解决表现，相关效应量为 *d*=0.55–0.69。修订过程中，我们也注意到近期同类研究同样基于 Urban 等(2024)报告的效应量（*d*=0.55）进行样本量估计(Wong & Qiu, 2026)，这进一步说明该效应量在相关人智协作研究中具有可参考性。具体修改请参见修改稿方法部分“2.1 被试”小节(p.22)。

意见 3

（3）研究将大五人格作为控制变量纳入分析。虽然人格特质确实可能影响创造性表现，但它并非创造力的直接构成要素。建议作者在文献综述或方法部分，对引入该控制变量的理论依据和潜在角色进行更细致的说明。例如，前人研究关于大五人格与创造力的关系有哪些主要发现？为何在本研究中必须将其控制？请补充相应的参考文献。

回应：感谢您的宝贵建议。我们同意原稿中对大五人格作为控制变量的理论依据说明不够充分。修改稿已在方法部分补充相关说明。具体而言，既有研究表明，大五人格特质与个体创造性表现存在关联，其中开放性通常被认为是与个体创造力和发散思维表现关系较为稳定的人格特质(Grajzel et al., 2023)。同时，人格特质也可能影响团体合作过程与协作问题解决表现。团体研究显示，人格特质与团体绩效存在关联(O’Neill & Allen, 2011)；关于大五人格与协作问题解决的系统综述也表明，个人层面和团体层面的责任心、宜人性等特质与协作问题解决表现相关，外向性和神经质也可能影响团体协调互动(Jolić Marjanović et al., 2024)。鉴于大五人格可能影响个体创造性表现及其在协作中的互动方式，因此本研究将其作为潜在控制变量纳入分析。具体修改请参见修改稿方法部分“2.4 测量工具”小节(p.25)。

意见 4

（4）文章提及选取 GPT-5 作为互动对象。这是一个非常关键的实验材料。为了保证研究的透明度和可重复性，建议作者披露更详细的技术信息，包括但不限于具体的模型版本号、API

调用时间、系统提示词（System Prompt）的详细内容以及温度（Temperature）等参数的设置。

回应：感谢您的宝贵建议。我们同意 GPT 模型作为本研究的关键实验材料，需要提供更充分的技术信息，以保证研究透明度和可重复性。原稿中将模型较笼统地表述为 GPT-5，修改稿已进一步明确具体模型版本为 GPT-5-mini，并在方法部分补充 API 调用时间和主要参数设置。具体而言，人智协作条件中的 AI 互动对象通过 API 调用 GPT-5-mini 模型实现。GPT-5-mini 模型发布于 2025 年 8 月 7 日，实验中的 API 调用时间为 2025 年 11 月 8 日至 2025 年 12 月 21 日。模型参数设置为：温度（temperature）=0.70，核采样参数（top p）= 1.00，存在惩罚（presence penalty）=0.00，频率惩罚（frequency penalty）=0.20。该参数设置主要是为了在任务中保留一定输出多样性，减少重复性表达。此外，我们已将 AUT 和 PIT 任务中的人类被试指导语及 AI 系统提示词整理至附录 1，以便读者了解不同任务中 AI 接收到的具体任务要求。具体修改请参见修改稿方法部分“2.2 实验设计与任务”小节(p.23)及附录 1。

意见 5

（5）文章提到由“两名具有创造力研究背景的研究生”对被试的创造性表现进行评分。这两名评分者是如何招募的？他们是否为本研究的合著者？如果是，如何确保评分过程的客观性，避免确认偏误？

回应：感谢您的细心提醒。我们同意原稿中对评分者招募流程、培训过程及客观性控制的说明不够充分。具体而言，两名评分者均为课题组内具有创造力研究背景的研究生，通过本文第一作者邀请参与评分工作，但并非本文合著者。正式评分前，本文第一作者向评分者说明新颖性和适用性的评分标准，并组织其进行练习评分；随后针对评分理解中的分歧进行讨论，待两名评分者对评分标准形成一致理解后再开展正式评分。为避免确认偏误，正式评分时评分者独立完成评分，且不知晓实验条件、被试身份及研究假设；评分材料中亦隐去了相关身份和条件信息。修改稿已在方法部分补充相关信息，具体修改请参见方法部分“2.5 测量指标”小节(p.26)。

参考文献

Grajzel, K., Acar, S., & Singer, G. (2023). The Big Five and divergent thinking: A meta-analysis. *Personality and Individual Differences*, *214*, 112338. https://doi.org/10.1016/j.paid.2023.112338

Guilford, J.P., 1967. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior, 1(1)*:3-14.

Jolić Marjanović, Z., Krstić, K., Rajić, M., Stepanović Ilić, I., Videnović, M., & Altaras Dimitrijević, A. (2024). The big five and collaborative problem solving: A narrative systematic review. *European Journal of Personality*, *38*(3), 457–475. https://doi.org/10.1177/08902070231198650

O’Neill, T. A., & Allen, N. J. (2011). Personality and the Prediction of Team Performance. *European Journal of Personality*, *25(1)*, 31–42. https://doi.org/10.1002/per.769

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, *215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Torrance, E. P. (2008). *Torrance tests of creative thinking: Norms-technical manual, verbal forms A and B.* Scholastic Testing Service.

Wong, S. S. H., & Qiu, S. X. (2026). Think First, ChatGPT Later: Guiding Human–AI Collaboration for Learning Gains in Independent Human Creativity. *Educational Psychology Review*, *38*(1), 45. https://doi.org/10.1007/s10648-026-10118-7

审稿专家 2：

总体来看，本文围绕人智协作情境下的创造性表现，从观点采纳这一互动机制切入，结合行动者–同伴相依模型（APIM）进行分析，选题具有较好的前沿性，也契合当前生成式人工智能背景下心理学研究的发展方向。在方法上，将 APIM 应用于人机互动过程分析，具有一定新意。整体而言，文章已经具备较好的基础，但当前稿件在理论表述、变量说明及分析呈现上仍有很多可以改进的地方。

意见 1

首先，文章目前对“观点采纳不对称作用模式”的表述略显偏强。从结果来看，这种不对称性主要体现在不同任务（AUT、PIT）和不同指标（流畅性、灵活性）上，并非在所有指标上都一致成立。建议作者在摘要和讨论中适当降低结论强度，将表述改为“在特定任务和指标下观察到差异模式”，使结论与结果更加一致。

回应：衷心感谢您对本文的细致审阅以及所提出的宝贵意见。我们同意原稿中关于“观点采纳不对称作用模式”的表述存在一定概括性，容易使读者理解为该模式在所有任务和所有创造性指标上均一致成立。根据专家建议，修改稿已在摘要和讨论部分降低相关结论的表述强度，将原先较为笼统的“不对称作用模式”修改为“在特定任务和指标下观察到的差异模式”或“任务依赖的角色差异模式”，以避免将部分指标上的发现过度推广为整体结论。具体修改请参见摘要(p.19)和讨论“4.3 观点采纳行为塑造双方的创造性表现”(p.40)。

意见 2

其次，在方法部分已经列出创造潜能、大五人格、合作倾向等控制变量，但在后续分析中未明确说明这些变量是否进入模型。建议补充说明这些变量是否被控制，以及是否影响主要结果，或者明确说明未纳入模型的理由，以保证前后一致。

回应：感谢您的宝贵建议。我们同意原稿中虽然列出了创造潜能、大五人格和合作倾向等潜在控制变量，但未明确说明这些变量是否进入后续模型，修改稿已在结果部分补充相关说明。具体而言，本研究在前测中测量创造潜能、大五人格和合作倾向，是因为这些变量可能影响人类个体创造性表现和协作互动过程。但在主要统计模型中未进一步纳入这些变量，理由包括两个方面。第一，在人类被试层面的条件比较中，双人协作条件与人智协作条件的人类被试在上述变量上均无显著差异。第二，本研究的 APIM 分析同时将人类与 AI 作为行动者和同伴纳入模型，而创造潜能、大五人格和合作倾向均为人类个体层面的心理特征，无法在 AI 角色上进行测量；若将其纳入模型，将导致不同角色之间协变量含义不一致。因此，这些潜

在控制变量未纳入后续模型。具体修改请参见修改稿结果部分“3.1 描述性统计”(p.28)。

意见 3

第三，关于观点采纳的测量，文章使用 IOC 指标（基于想法类别一致性）来表示观点采纳行为。该指标定义是清楚的，但建议在方法或讨论中增加一句说明：IOC 反映的是行为层面的趋同，而不完全等同于主观理解他人观点，从而使指标解释更加准确。

回应：感谢您的宝贵建议。我们同意原稿中未对 IOC 指标的解释边界作出充分说明。修改稿已在讨论部分的“不足与展望”一节补充了该指标的局限性。正如您所指出的，IOC 是基于想法类别一致性计算的行为指标，反映行为层面的趋同，而不完全等同于主观理解他人观点。相关说明已写入修改稿讨论部分“4.5 不足和展望”(p.42)。

意见 4

第四，在 APIM 分析中将 AI 与人类共同作为“行动者”和“同伴”进行建模，这一处理在统计上是可行的，但建议在方法部分补充说明：AI 的“观点采纳”是基于对话行为的操作性定义，而非对其内部心理过程的推断，以避免概念上的歧义。

回应：感谢您的宝贵建议。我们同意在人智协作条件下，将 AI 与人类共同作为“行动者”和“同伴”建模时，需要进一步明确 AI 观点采纳行为的操作性定义。修改稿已在 APIM 分析部分补充说明：AI 的观点采纳行为同样通过聚合指数进行计算，即根据 AI 当前生成想法与人类同伴先前想法在类别上的一致性进行操作性定义，并不涉及对 AI 内部心理过程的推断。相关说明已写入修改稿方法部分“2.6 数据分析”(p.27)。

意见 5

第五，结果部分同时呈现了组间比较、角色比较和 APIM 分析，但各部分之间的关系没有明确说明。建议在每一部分开头简单交代分析目的（例如用于回答哪个研究问题），使整体结构更加清晰。

回应：感谢您的宝贵建议。我们同意原稿中结果部分的组间比较、角色比较与 APIM 分析之间的逻辑关系说明不够清晰。根据您的建议，修改稿已在各结果小节开头补充分析目的，明确说明组间比较主要用于回答研究问题 1，即检验双人协作与人智协作在创造性表现和观点采纳行为上的差异；角色比较是在此基础上进一步考察不同角色的表现差异；APIM 分析则用于回答研究问题 2，即检验个体及其同伴的观点采纳行为对个体创造性表现的预测关系，并比较该关系是否因合作类型和角色而发生变化。上述修改已补充至结果部分 3.2.1、3.2.2、3.2.3、3.3.1、3.3.2 和 3.3.3 节。

意见 6

最后，部分结果解释可以更加贴近数据。例如在 AUT 任务中，观点采纳对不同指标的作用

方向并不完全一致，建议在讨论中分别说明不同指标的结果，而不是用单一结论概括全部结果。

回应：我们根据您的建议重新调整了讨论部分的结果解释方式。修改稿不再用单一结论概括观点采纳对创造性表现的作用，而是分别围绕不同任务和不同创造性指标进行讨论。具体而言，在 AUT 任务中，讨论重点放在 AI 即时观点采纳对想法流畅性的正向预测作用，并避免将这一结果泛化到所有创造性指标；在 PIT 任务中，修改稿进一步区分了想法新颖性、流畅性和灵活性等指标，分别讨论 AI 延迟观点采纳对新颖性的负向预测、人类延迟观点采纳对流畅性和灵活性的正向预测，以及 AI 延迟观点采纳对流畅性和灵活性预测作用不显著的相关结果。通过上述修改，修改稿的讨论更加贴近实际数据结果，也更清楚地呈现了观点采纳作用模式在不同任务和指标上的差异。上述修改已补充至“4.3 观点采纳行为塑造双方的创造性表现”（p.40）。

总体而言，稿件已有较好的数据基础，主要问题集中在表述和结构层面，建议作者根据以上几点进行调整。

……

审稿专家 3：

本文围绕人智协作中的观点采纳及其对创造性表现的作用展开研究，关注生成式 AI 从辅助工具转向协作伙伴这一重要前沿议题，选题具有很强的时代意义和理论价值。作者尝试将观点采纳引入人智协作的互动过程分析，并结合行动者-同伴相依模型比较双人协作与人智协作中的作用机制，具有很好的创新性和方法学探索意义。总体来看，研究为理解 AI 作为协作伙伴如何影响创造性互动提供了有价值的经验证据。若作者能够进一步澄清理论框架、关键变量定位、任务设计依据、AI 相关测量效度及统计分析策略等问题，稿件质量将得到显著提升。

主要意见

主要意见 1

1. 文中部分句式较为嵌套，如 1.1.1 节中“需要将其置于。。。中关于。。。脉络之中”等表达，影响论述的简洁性和流畅性。建议作者精炼全文语言，尤其注意核心概念之间的逻辑关系，避免因表述复杂或概念搭配不当削弱理论论证的清晰度。

回应：衷心感谢您对本文的细致审阅以及所提出的宝贵意见。我们已根据建议对全文语言进行了通读和润色，重点修改句式嵌套、表达冗余和概念衔接不够清晰的表述。针对审稿专家指出的 1.1.1 节中“需要将其置于……中关于……脉络之中”这一句的问题，修改稿已删除该句，并重新组织了段落开头，直接引入团体创造理论。在全文的润色工作中，我们将长句和嵌套的句子进行重新组织，删除了冗余表述，确保核心概念之间的逻辑链条前后一致。通过

上述修改，全文语言的简洁性和流畅性得到显著提升。

主要意见 2

2. 研究变量角色的说明不够清晰。观点采纳是本文的核心过程变量，但 2.2 节仅列出创造性表现相关因变量，未明确说明观点采纳在研究设计中的定位。建议作者明确区分协作类型、创造性表现和观点采纳的变量角色，例如将观点采纳界定为关键过程变量，在组间比较中作为比较指标，在 APIM 分析中作为预测变量。

回应：感谢您的宝贵建议。我们同意原稿对研究变量角色的说明不够清晰，实验设计部分未明确交代观点采纳的定位。根据您的建议，修改稿已在“2.2 实验设计与任务”一节补充说明：本研究中协作类型为自变量，创造性表现为主要结果变量，观点采纳为关键过程变量。同时，修改稿在“2.6 数据分析”部分进一步明确了观点采纳在不同分析中的具体角色：在组间比较和角色比较中，观点采纳作为协作过程指标，用于考察不同协作条件和角色的互动行为差异；在 APIM 分析中，观点采纳作为预测变量，用于检验个体及其同伴的观点采纳行为对个体创造性表现的预测作用。上述修改已补充至修改稿方法部分“2.2 实验设计与任务”(p.23)以及“2.6 数据分析”(p.27)。

主要意见 3

3. 引言的理论铺垫与研究问题之间的衔接仍需加强。引言前半部分较多论述传统团体创造中的观点采纳研究，直到 1.2 节才切入人智协作情境，导致主线略显分散。建议作者更早引入人智协作问题，并围绕 1 到 2 个核心理论深入说明“为什么观点采纳在双人协作与人智协作中可能呈现不同作用模式，以及 AI 作为协作对象与人类同伴有何关键差异”。

回应：感谢您的宝贵建议。我们同意原稿引言前半部分较多论述传统团体创造中的观点采纳研究，但与人智协作情境之间的衔接仍不够充分。原稿中这一部分的写作目的在于为本研究提供理论和实证基础，但确实需要进一步说明这些研究如何引出人智协作中的核心问题。根据专家建议，修改稿已在 1.1.2 节结尾增加过渡段。

具体而言，修改稿首先基于视角可供性理论和相关实证研究进一步说明，观点采纳不仅可能影响个体自身的创造性表现，也可能影响互动对象的创造性表现。研究进一步强调，观点采纳本质上是一种关系性和对话性的互动过程，其行为表现和作用模式并非固定不变，而是可能受到互动对象特征的影响。先前研究为这一观点提供了支持，例如成员教育背景(Lu et al., 2021)、团队创造力水平(Lu et al., 2023)和关系类型(Bai et al., 2026)等群体特征均会影响互动中的观点采纳模式。最后，针对审稿专家关于“AI 作为协作对象与人类同伴有何关键差异”的建议，修改稿进一步指出，相比传统双人协作，人智协作中的互动双方成员身份（人类 vs. AI）、创造潜能和想法生成机制等方面存在显著差异(Zhang et al., 2026)。因此，当协作对象由人类同伴转变为 AI 时，观点采纳行为及其对创造性表现的作用关系可能呈现不同于双人协作的模式。相关修改请参见修改稿前言部分“1.1.2 团体创造中观点采纳作用的

实证研究”(p.20)。

主要意见 4

4. 行动者-同伴相依模型的理论铺垫不足。APIM 是本文区分行动者效应和同伴效应的关键分析框架，但目前主要在 1.3 节才被提出，前文对其理论依据、适用条件及其与观点采纳机制的关系说明不足。建议作者更早介绍 APIM 的基本逻辑，并解释双人协作为何可被视为不可区分二元组，而人智协作为何可被视为可区分二元组。

回应：感谢您的宝贵建议。我们同意，APIM 是本文区分行动者效应与同伴效应的关键分析框架，原稿主要在研究目的部分提出该模型，对其理论依据、适用条件及其与观点采纳机制之间关系的说明不够充分。修改稿已在 1.1.2 节介绍 Bai 等人(2026)研究时更早引入 APIM 的基本逻辑。具体而言，我们补充说明：基于视角可供性理论，观点采纳本质上是一种关系性和对话性的互动过程，个体在理解、回应和发展他人观点时，不仅可能影响自身想法的生成，也可能影响同伴后续想法的发展。因此，在二元创造性互动中，当研究问题涉及双方观点采纳行为如何分别影响自身与同伴的创造性表现时，APIM 能够用于区分个体自身观点采纳行为对自身创造性表现的影响(行动者效应)，以及同伴观点采纳行为对个体创造性表现的影响(同伴效应)。Bai 等人(2026)在儿童—成人团体合作研究中采用 APIM 分析了观点采纳对自身和同伴创造性表现的作用，为本研究提供了直接的方法和理论依据。相关修改请参见修改稿前言部分“1.1.2 团体创造中观点采纳作用的实证研究”(p.20)。

此外，修改稿已在 1.3 节进一步说明本研究中二元组类型的界定。在双人协作条件下，两名成员均为随机分配的人类被试，研究设计中不存在理论上可区分的成员身份差异，因此可视为不可区分二元组；在人智协作条件下，互动双方分别为人类与 AI，二者在成员身份和想法生成机制上存在明确差异(Zhang et al., 2026)，因此可视为可区分二元组。相关修改请参见修改稿前言部分“1.3 研究目的”(p.22)。

主要意见 5

5. 创造力任务类型的理论意义和统计处理需进一步说明。本研究同时采用 AUT 和 PIT，两类任务可能在开放性、约束性、现实应用性及对观点采纳的依赖程度上存在差异，建议作者在引言中补充说明选择这两类任务的理论依据，并考虑在统一模型中检验“协作类型 × 任务类型”及其与观点采纳的交互作用。若分任务分析，建议说明理由并谨慎解释任务间差异。

回应：感谢您的宝贵建议。首先，我们同意 AUT 和 PIT 在开放性、任务约束、现实应用性及对观点采纳的依赖程度上可能存在差异，已按照审稿人的建议，在引言中补充说明选择两类任务的理论依据。具体而言，AUT 是创造力研究中经典的发散思维任务，要求参与者针对常见物品提出尽可能多的新颖用途，强调思维的发散性(Guilford, 1967)；PIT 源自 Torrance 创造思维测验中的产品改进活动，要求参与者对常见产品提出改进或优化方案，能够较好反映现实情境中的创造性问题解决(Torrance, 2008)。已有研究已在 AUT 任务中初步揭示观点

采纳的行动者效应和同伴效应(Bai et al., 2026)，但观点采纳在更复杂的现实问题解决任务中是否具有相似作用模式仍有待检验。相较于 AUT，PIT 更强调对已有信息的整合与重组(Cheng & Zhang, 2025)，个体可能更需要将同伴想法作为线索，推动方案向新的方向发展，因此 PIT 中的创造性表现可能更依赖观点采纳。基于此，本研究采用 AUT 和 PIT，以比较双人协作与人智协作在经典发散思维任务和现实问题解决任务中的创造性表现及观点采纳作用模式差异。相关修改请参见修改稿前言部分“1.3 研究目的”(p.22)。

其次，本文将分任务分析作为分析策略，主要是基于以下原因：第一，如前所述，AUT 与 PIT 在任务目标、任务复杂程度和创造性加工要求上存在差异。相应地，本研究也因此根据两类任务的复杂程度设置了不同的互动时长（AUT 为 10 分钟，PIT 为 15 分钟）。若将两类任务直接合并为一个总体模型，模型中的主效应或低阶交互反映的是两个任务的平均效应，可能难以呈现观点采纳在不同任务情境中的作用模式。虽然联合模型可以检验任务类型差异，但需要解释较复杂的交互项，因此本文将分任务分析作为主要结果呈现方式。第二，AUT 和 PIT 在本研究中具有不同的理论功能。AUT 作为创造力研究中最常用的经典发散思维任务，有助于将本研究结果与既有团体创造和观点采纳研究进行关联；PIT 则更接近现实的创造性问题解决任务，能够进一步检验观点采纳在人智协作中是否能够支持对既有方案的改进、整合和拓展。因此，本文将其作为两类具有不同理论功能的创造性任务，来分别检验双人协作和人智协作在两个任务下的创造性表现和观点采纳行为的差异。基于这一研究目的，修改稿以分任务分析作为主要结果呈现方式，以更直观地展示 AUT 和 PIT 中观点采纳作用模式的具体差异。修改稿已在讨论中进一步谨慎表述任务间差异，相关说明已补充至修改稿引言“1.3 研究目的”(p.22)和讨论“4.5 不足与展望”小节(p.42)。

主要意见 6

6. AI 主观后测指标的测量依据和解释边界可以更清晰一些。方法部分仅说明 AI 的想法选择和单一任务后测问卷由主试通过统一提示词获取，但未报告具体提示词、评分规则、重复获取次数和稳定性检验，也未论证人类主观问卷是否适用于 AI。鉴于任务努力、任务兴趣、任务难度和同伴观点采纳倾向本质上属于人类主观体验或心理倾向指标，建议作者谨慎界定其含义。

回应：感谢您的宝贵建议。首先，关于 AI 后测反应的测量依据和解释边界，我们同意原稿说明不够充分，修改稿已对此进行补充。已有 AI 心理测量研究尝试使用人类心理量表考察大语言模型在不同提示词和任务语境下的反应模式，为获取 AI 问卷反应提供了方法依据(Li & Qi, 2025; Salecha et al., 2024)。但 AI 问卷反应可能受到提示词、选项顺序和测量形式等因素影响，其构念含义不能直接等同于人类问卷反应(Serapio-García et al., 2025)。修改稿进一步阐明，本研究获取 AI 对任务努力、任务兴趣、任务难度和同伴观点采纳倾向等项目的反应，主要是为了探索模型在完成任务后的反应模式，而不是推断 AI 具有相应的主观体验或心理状态。相关说明已补充至修改稿方法“2.4.2 单一任务后测”(p.25)。

其次，关于 AI 想法选择和单一任务后测问卷的程序细节，修改稿已补充提示词、评分规则、获取次数。具体而言，在人智协作条件下，每个任务结束后，主试在保留该任务完整对话记录的同一会话中分两步获取 AI 反应：（1）使用预设提示词要求 AI 从该任务中双方生成的所有想法中选出一个其认为最具创造性的想法；（2）使用预设提示词要求 AI 按照与人类被试相同的评分项目作答。上述两个步骤在每个任务结束后各获取一次，未进行重复生成。由于本研究并未对 AI 后测反应进行重复施测和独立效度验证，因此未进行生成稳定性检验。修改稿已在方法部分“2.3 实验程序”小节补充上述说明(p.24)，完整提示词见附录 1。

最后，关于人类主观问卷适用于 AI 的边界，修改稿进行了谨慎说明。我们认为，人类主观问卷可以尝试用于 AI 研究，但其适用性需要被视为尚待验证的探索性问题。已有 AI 心理测量研究尝试采用人类心理量表对大语言模型进行施测，并检验其问卷反应的信度和效度。结果发现，在特定提示词设置下，部分 AI 的问卷反应可以表现出一定的信度和效度(Serapio-García et al., 2025)。但如前所述，AI 的问卷反应会受到提示形式和选项顺序的显著影响，从而削弱 AI 反应的可靠性(Gupta et al., 2024; Li & Qi, 2025)。因此，本研究中的 AI 任务难度、任务兴趣、任务努力和观点采纳倾向问卷，仅用于获取 AI 反应的探索性测量工具，而不是已经充分验证的 AI 心理测量工具。相应地，修改稿已将这些指标解释为辅助性、探索性的 AI 反应指标，并避免将其表述为 AI 的真实主观体验或稳定心理特质。相关说明已补充至修改稿讨论“4.5 不足与展望”小节(p.42)。

次要问题

次要意见 1

1. 文中参考文献括号的全角与半角使用不统一，建议按照学报格式要求统一处理。

回应：感谢您的提醒。我们已根据学报的格式要求，对全文参考文献引用中的括号格式进行了统一，均修改为半角括号。

次要意见 2

2. 部分参考文献引用格式不够规范，例如“Hoever 等(2012)”建议改为“Hoever 等人(2012)”，“Gonzalez & Heidari(2025)”建议改为“Gonzalez 和 Heidari(2025)”。全文类似格式需统一核查。

回应：感谢您的提醒。我们已根据学报的格式要求，对全文参考文献的文内引用格式进行了核查和修订，统一规范了中文语境下“等人”“和”等表述方式。

次要意见 3

3. 2.1 节建议更完整地报告被试的基本信息，包括年龄、性别、专业/学科背景等。若相关信息主要呈现在表 1 中，也建议在正文中作简要说明并与表格呼应。

回应：感谢您的宝贵建议。我们已在修改稿 2.1 节中补充报告被试的基本人口学信息，包括年龄、性别和专业背景，并与表 1 中的统计结果相对应。具体修改请见修改稿 2.1 节(p.23)。

次要意见 4

4. 2.4.3 节中，观点采纳一般倾向量表的题项数需明确报告。AI 使用经验虽为单题测量，也建议报告具体题目或题干内容，以便读者判断其测量含义。

回应：感谢您的细致提醒。我们已在修改稿 2.4.3 节中补充说明观点采纳一般倾向量表的题项数。该量表共 7 个题项。同时，我们也补充报告了 AI 使用经验的具体题干和选项内容。AI 使用经验采用单题测量，参考 Urban 等人(2024)的测量方式，要求被试根据自身实际情况报告对话式人工智能的使用经验，选项从“没有听说过对话式人工智能”到“每天都会使用对话式人工智能”。具体修改请见修改稿 2.4.3 节(p.26)。

次要意见 5

5. 2.5 节中，创造性表现由两名评分者评分，建议报告评分者一致性指标，如 ICC，而不仅是 α 系数。对于想法类别划分及观点采纳行为编码，也建议报告分类一致性或编码一致性指标，例如 Cohen’s κ 等合适指标。

回应：感谢您的宝贵建议。我们已在修改稿 2.5 节中补充报告创造性评分、想法类别划分以及观点采纳行为编码的一致性指标。首先，创造性表现中新颖性和适用性的评分一致性采用 ICC 评估。结果显示，AUT 任务中新颖性和适用性的 ICC 分别为 0.88 和 0.82，PIT 任务中新颖性和适用性的 ICC 均为 0.75。

其次，想法类别划分由课题组两名具有创造力研究背景的研究生依据编码手册独立完成，并采用加权 Cohen’s κ 评估分类一致性。初始编码结果显示，AUT 任务的类别划分一致性为 0.77，PIT 任务为 0.58。随后，基于两名类别编码者的独立编码结果，分别计算即时观点采纳行为和延迟观点采纳行为，并采用 Cohen’s κ 评估由两套类别编码结果计算得到的观点采纳行为指标的一致性。结果显示，AUT 任务中即时观点采纳和延迟观点采纳行为的一致性分别为 0.97 和 0.77，PIT 任务中分别为 0.52 和 0.50。

鉴于 PIT 任务初始类别划分及由此计算得到的观点采纳行为指标一致性未达到理想水平，我们进一步请两名类别编码者对 PIT 任务中的想法类别进行重新独立编码。重新编码后，PIT 任务中想法类别划分的一致性达到良好水平，加权 Cohen’s κ 为 0.85；基于重新编码后的想法类别计算即时观点采纳和延迟观点采纳行为，其一致性也达到可接受水平，Cohen’s κ 分别为 0.76 和 0.74。因此，修改稿中 PIT 任务相关的观点采纳指标及后续分析结果均基于重新编码后的类别结果进行更新，结果变动说明请参考审稿回复第一页的“回应说明”。相关内容已在修改稿“2.5 测量指标”进行补充说明（p.27）。

次要意见 6

6. 全文 p 值格式需统一。例如，目前同时存在“$p < 0.001$”和“$p < .001$”两种写法，建议按照学报规范统一处理。

回应：感谢您的细致提醒。我们已根据学报的格式要求，对全文统计结果中的 p 值格式进行了统一，均修改为“p<0.001”的写法。

次要意见 7

7. 表 3 和表 5 信息量较大，建议采用更规范的三线表形式呈现，并适当优化排版。若模型参数过多，也可考虑将部分模型拟合信息或非核心结果移至补充材料。

回应：感谢您的宝贵建议。根据专家意见，我们已对表 3、表 5 及附表 1 的呈现方式进行调整，统一采用更规范的三线表格式。同时，考虑到原表中模型拟合信息较多，已将表 3、表 5 和附表 1 中的模型拟合结果移至附表 2 统一呈现，从而减少正文表格的信息负荷。具体修改请见修改稿表 3、表 5、附表 1 及附表 2。

次要意见 8

8. 图 3、图 5 和图 6 的图示说明仍可进一步规范。建议删除横轴标题中的“角色类型”，并在图注中明确不同颜色或图形元素所代表的角色。显著性标注可考虑用“*”表示，并在图注中说明显著性水平。精确 p 值可保留在正文或表格中。

回应：感谢您的细致提醒。我们已根据意见对图 3、图 5 和图 6 的图示说明进行了规范化修改。具体而言，上述图示已删除横轴标题中的“角色类型”，并在图注中补充说明不同颜色所代表的角色。同时，显著性标注已统一采用“*”表示，并在图注中说明相应的显著性水平；精确 p 值保留在正文或表格中报告。

次要意见 9

9. 4.4 节关于研究不足与未来展望的论述较为笼统，建议结合本文的关键局限进行实质性展开，例如 AI 指标的测量效度、任务类型的限制、观点采纳指标的操作化、因果解释边界等，并适当补充相关文献支持。

回应：感谢您的宝贵建议。我们同意原稿中“不足与展望”部分对关键局限的展开不够充分。根据专家建议，修改稿已对该部分进行实质性补充，重点围绕四个方面展开说明。第一，进一步明确 IOC 作为观点采纳行为指标的解释边界，指出其优势在于能够相对客观地量化互动中的观点采纳行为，但并不能完全等同于个体主观上站在他人视角理解其观点。第二，补充说明 AI 后测问卷的测量效度问题，强调 AI 任务难度、任务兴趣、任务努力和观点采纳倾向问卷仅用于探索 AI 在任务情境中的反应模式，不能等同于人类意义上的主观体验或稳定心理倾向。第三，补充讨论任务类型的限制，说明 AUT 和 PIT 在任务约束性、任务复杂性和认知加工要求上存在差异，因此本文采用分任务分析以避免掩盖不同任务情境下的具体作用模式；同时也承认该策略尚未在统一模型中直接检验任务类型的调节作用。第四，进一步明确因果解释边界，指出 APIM 结果应被理解为观点采纳行为与创造性表现之间的统计预测关系，而不能直接推断因果作用。修改稿同时补充了相关文献支持，并在未来研究方向中

提出可通过改进测量方式、引入统一任务模型以及实验操纵观点采纳水平或 AI 观点采纳策略来进一步检验相关问题。具体修改请参见修改稿“4.5 不足与展望”部分(p.40)。

次要意见 10

10. 结论部分建议更加简洁，可用数字序号概括主要发现。理论贡献和实践启示建议放入讨论部分集中展开，而不宜在结论中进行较多论述。

回应：感谢您的宝贵建议。我们已根据该建议对结论部分进行修改。修改稿中，已将原结论部分中关于理论贡献和实践启示的论述移至讨论章节 4.4 中；同时，结论部分已根据更新后的结果重新表述，采用数字序号概括主要发现，以突出研究结果本身并使结论更加简洁清晰。

具体修改请参见修改稿第 43 页结论部分及第 41 页“4.4 理论与实践价值”。

参考文献

Bai, H., Chan, L. S., Luo, A., Kroesbergen, E. H., & Christie, S. (2026). Creativity in dialogues: How do children interact with parents vs. strangers for generating creative ideas? *Learning and Instruction*, *102*, 102286. https://doi.org/10.1016/j.learninstruc.2025.102286

Cheng, X., & Zhang, L. (2025). Inspiration booster or creative fixation? The dual mechanisms of LLMs in shaping individual creativity in tasks of different complexity. *Humanities and Social Sciences Communications*, *12*(1), 1563. https://doi.org/10.1057/s41599-025-05867-9

Erwin, A. K., Tran, K., & Koutstaal, W. (2022). Evaluating the predictive validity of four divergent thinking tasks for the originality of design product ideation. *PLOS ONE*, *17*(3), e0265116. https://doi.org/10.1371/journal.pone.0265116

Gupta, A., Song, X., & Anumanchipalli, G. (2024). Self-Assessment Tests are Unreliable Measures of LLM Personality. *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, 301–314. https://doi.org/10.18653/v1/2024.blackboxnlp-1.20

Guilford, J.P., 1967. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior, 1(1)*:3-14.

Li, C., & Qi, Y. (2025). Toward accurate psychological simulations: Investigating LLMs' responses to personality and cultural variables. *Computers in Human Behavior*, *170*, 108687. https://doi.org/10.1016/j.chb.2025.108687

Lu, K., Gao, Z., Wang, X., Qiao, X., He, Y., Zhang, Y., & Hao, N. (2023). The hyper-brain neural couplings distinguishing high-creative group dynamics: An fNIRS hyperscanning study. *Cerebral Cortex*, *33*(5), 1630–1642. https://doi.org/10.1093/cercor/bhac161

Lu, K., Qiao, X., Yun, Q., & Hao, N. (2021). Educational diversity and group creativity: Evidence from fNIRS hyperscanning. *NeuroImage*, *243*, 118564. https://doi.org/10.1016/j.neuroimage.2021.118564

Serapio-García, G., Safdari, M., Crepy, C., Sun, L., Fitz, S., Romero, P., Abdulhai, M., Faust, A., & Matarić, M. (2025). A psychometric framework for evaluating and shaping personality traits in large language models. *Nature Machine Intelligence*, *7*(12), 1954–1968. https://doi.org/10.1038/s42256-025-01115-6

Salecha, A., Ireland, M. E., Subrahmanya, S., Sedoc, J., Ungar, L. H., & Eichstaedt, J. C. (2024). Large language models display human-like social desirability biases in Big Five personality surveys. *PNAS Nexus*, *3*(12), pgae533. https://doi.org/10.1093/pnasnexus/pgae533

Torrance, E. P. (2008). *Torrance tests of creative thinking: Norms-technical manual, verbal forms A and B.* Scholastic Testing Service.

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, *215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Zhang, M., Li, Y., Peng, Y., Sun, Y., Guo, W., Hu, H., Chen, S., & Zhao, Q. (2026). AI delivers creative output but struggles with thinking processes. *Thinking Skills and Creativity*, *61*, 102196. https://doi.org/10.1016/j.tsc.2026.102196

第二轮

审稿专家 1：作者已经在修改稿中比较好的完成了审稿人之前的问题，稿件质量得到显著提升，没有进一步修改意见了，建议录用。

审稿专家 2：我对作者的修改十分满意，没有进一步意见。

审稿专家 3：

以下意见供作者参考：

作者对上一轮审稿意见做出认真回应。修改稿在理论框架、变量定位、APIM 分析依据和任务设计说明等方面均有明显补充，也对 PIT 任务的类别编码和观点采纳指标进行了重新处理。这些修改显著提升了稿件的清晰度、透明度和方法严谨性。还有几个意见建议，请作者思考。

意见 1

1. 论文同时将 AUT 和 PIT 作为核心创造任务，并在结果部分分别展开分析，但引言中对两类任务的理论定位有所不足。鉴于任务类型构成了本研究的核心情境变量，建议作者新建一节（一段），集中说明为何必须同时采用 AUT 与 PIT 而非其他创造任务，二者在开放性、约束性、现实应用性和观点采纳依赖程度上的关键差异是什么，以及这些差异如何与 PAT 理论或 APIM 分析框架发生联系。此外，建议作者进一步解释为何将 AUT 设为 10 分钟、PIT 设为 15 分钟，以及不同时长是否可能影响流畅性、观点采纳频率等关键指标的可比性，补充任务选择和任务时长设置的理论与文献依据，并在讨论中说明分任务分析的解释边界。

回应：感谢您的细致审阅和宝贵建议。接下来，我们分别对任务理论定位，任务实验设置，以及分任务分析的解释边界进行回应。

首先，我们同意，原稿在引言中对 AUT 和 PIT 的理论定位说明不足。根据您的建议，我们在引言中新增一段，集中说明任务类型为何是人智协作中影响观点采纳作用模式的核心情境变量，并进一步阐明 AUT 和 PIT 的任务差异及其与视角可供性理论（PAT）的关系。第一，我们补充说明了同时纳入不同类型创造任务的必要性。已有研究发现，人智协作的创造性收益在不同任务中并不一致：在开放性较高、约束较低的 AUT 中，个体与 AI 合作表现显著高于双人协作；而在更复杂、现实约束更强的 PIT 中，人智协作的优势并不稳定出现(Cheng et al., 2025)。这一结果提示，人智协作的创造性收益可能受到任务类型影响，因此有必要进一步考察人智协作的互动过程是否也具有任务特异性。此外，已有研究在 AUT 中初步揭示了观点采纳对创造性表现的行动者效应和同伴效应(Bai et al., 2026)，但尚不清楚这一作用模式能否推广到更强调现实约束和信息整合的 PIT 任务情境。第二，我们进一步补充说明了 AUT 和 PIT 在开放性、约束性、现实应用性上的差异，并结合 PAT 说明两类任务观点采纳作用模式的可能差异。根据视角可供性理论，观点采纳可以使个体接触并利用同伴视角，从而发现新的行动可能性。然而，不同创造任务提供的问题空间不同，观点采纳的作用模式也可能不同。具体来说，AUT 作为经典发散思维任务，问题空间较为开放，主要强调想法发散(Guilford, 1967)；在这类任务中，观点采纳可能更多体现为对同伴当前想法的即时采纳，即借助同伴视角激活新的联想方向。相比之下，PIT 作为现实问题解决任务，要求个体围绕特定产品提出改进方案，并整合产品功能、使用情境和现实可行性等信息(Torrance, 2008)；在这类任务中，观点采纳可能不仅体现为对当前想法的即时回应，更可能体现为对同伴先前想法的延续、修正和整合，即通过延迟采纳发挥作用。因此，AUT 和 PIT 分别代表低约束、高开放性的发散任务与高约束、强调信息整合的问题解决任务，本研究同时纳入这两类任务有助于分别考察不同任务情境下人智协作的观点采纳作用模式。具体修改请参见修改稿引言部分“1.2 人智协作中的创造性互动”小节(p.28)。

其次，关于任务时长设置的依据。本研究关注的是人智协作的互动过程，因此任务时长需要保证参与者有足够的交替发言轮次，以支持互动过程指标的编码。既有采用 AUT 的团体创造研究中，互动时长设置包括常见的 5 分钟(Lu et al., 2020; Brucks & Levav, 2022)和 10 分钟(Glăveanu et al., 2019)等。本研究选择 10 分钟，以保证双人协作和人智协作条件下均能进行充分的互动。相较于 AUT，PIT 具有更强的任务约束性和现实应用性。被试不仅需要生成新想法，还需要围绕具体产品识别问题、提出改进方案，并考虑方案的可行性和适用性。因此，PIT 涉及更复杂的认知加工和互动协商过程(Erwin et al., 2022)。Howard 等人(2010)关于产品创新的头脑风暴研究表明，在前 30 分钟内产生的有效创意中，75%出现在前 15 分钟，提示 15 分钟是捕获有效产品改进想法的重要时间窗口。因此，基于 PIT 较高的任务复杂度，以及本研究对互动过程进行编码的需要，我们将 PIT 设置为 15 分钟。此外，关于不同时长对关键指标可比性的影响。我们同意您的意见，不同的任务时长可能影响流畅性(Paek

et al., 2021)、观点采纳频率等指标。因此，本研究并不直接比较 AUT 与 PIT 两个任务之间的创造性表现以及观点采纳频率的差异，而是按任务分别进行统计分析。不可否认的是，不同的任务时长设置限制了 AUT 与 PIT 原始数量指标之间的直接比较，未来研究可在统一任务时长或更系统操纵任务时间的条件下进一步检验时长对流畅性和观点采纳过程的影响，我们在局限部分也讨论了这些限制。具体修改请参见修改稿方法部分“2.2 实验设计与任务”小节(p.30)，以及讨论部分“4.5 不足与展望”小节(p. 50)。

最后，关于分任务分析的解释边界，我们根据您的建议在讨论和局限部分作了进一步修改。第一，在讨论部分，我们结合引言中新补充的任务理论定位，对人类在 AUT 和 PIT 任务中的观点采纳频率差异进行了分任务解释，避免用单一结论概括两个任务的结果。第二，我们降低了讨论中关于任务差异的表述强度，避免将 AUT 和 PIT 中的结果差异直接解释为任务类型的调节效应，而是将其表述为不同任务情境下观察到的结果模式差异。第三，我们在局限部分进一步说明，由于 AUT 和 PIT 在任务开放性、约束性、现实应用性和互动时长上均存在差异，本研究不直接比较两类任务之间的创造性表现和观点采纳水平差异，也不能据此直接推断任务类型在观点采纳作用模式上具有调节作用。未来研究仍需在统一任务时长和更大样本基础上，将任务类型纳入同一模型，以进一步检验任务特征如何调节观点采纳对创造性表现的作用。具体修改请参见修改稿讨论部分“4.2 人智协作中的观点采纳行为更频繁”小节(p.46)，以及讨论部分“4.5 不足与展望”小节(p. 50)。

意见 2

2. AUT 任务有效被试为 155 名，PIT 任务有效被试为 164 名，作者似乎是按任务分别剔除未按指导语完成的被试。该处理在分任务分析中具有合理性，但作者需要说明剔除标准、各条件和任务中的剔除人数及原因，并解释为何未完成某一任务的被试其另一任务数据仍可保留。

回应：感谢您的宝贵建议。我们同意原文对数据剔除标准和剔除情况的说明不够充分。根据审稿人的意见，我们已在方法部分补充说明按任务分别进行数据有效性判断的依据，并报告了各任务和各条件下的剔除人数及原因。剔除标准不仅包括不按照提示词要求，还包含技术引起的任务中断，具体来说：第一，未按任务指导语要求在每轮生成一个想法，例如仅复制 AI 回答、仅向 AI 提问，或与人类同伴进行与任务无关的闲聊；第二，因网络故障导致对话中断，无法形成完整互动过程记录。在 AUT 任务中，共保留有效被试 155 名，剔除 14 名。其中，人智协作条件下，因未按指导语完成剔除 4 名，因网络故障剔除 2 名；双人协作条件下，因未按指导语完成剔除 2 组（4 名），因网络故障剔除 2 组（4 名）。在 PIT 任务中，共保留有效被试 164 名，剔除 5 名。其中，人智协作条件下，因网络故障剔除 1 名；双人协作条件下，因未按指导语完成剔除 1 组（2 名），因网络故障剔除 1 组（2 名）。

此外，根据您的建议，文中进一步解释了为何未完成某一任务的被试其另一任务数据仍可保留。由于 AUT 任务和 PIT 任务分别对应不同的创造性任务情境，且本文主要采用分任

务分析策略，因此数据有效性判断也按任务分别进行。若被试在某一任务中因未按指导语完成或网络故障导致该任务数据无效，其另一任务的数据若完整且符合指导语要求，则仍可为相应任务分析提供有效信息。因此，我们仅剔除对应任务中的无效数据，而未将该被试在所有任务中的数据一并删除。具体修改请参见修改稿方法部分“2.1 被试”小节(p.29)。

意见 3

3. 结果部分仍存在若干统计数值、表文对应和报告规范问题，建议作者全面复核。

尤其是表 1 中部分 t 值正负方向似乎与均值呈现顺序不一致，3.3.2 节中个别效应量明显异常，例如 PIT 任务流畅性比较中报告的 d=11.65 和 d=11.53。此外，部分显著交互在表格中出现但正文未加解释，部分术语如“适用性”和“有用性”使用不够统一。请作者系统检查所有表格、正文统计报告、效应量、显著性解释和图注说明，确保前后一致。

回应：感谢您的细致审阅和宝贵建议。针对统计数值、表文对应和报告规范问题，我们已对全文的表格、正文统计报告、效应量、显著性解释、术语使用和图注说明进行了检查，并作如下修改。

首先，表 1 中部分 t 值正负方向确实与均值呈现顺序不一致。修订稿已统一按照表头顺序“人智协作−双人协作”报告组间差异，并对表 1 中相关 t 值的正负符号进行了校正。

其次，关于原稿第 3.3.2 节中部分效应量异常偏高（如 d=11.65、d=11.53）的问题，这些效应量来自角色分析中的两两比较，原稿角色分析的效应量采用混合线性模型并通过 emmeans::eff_size()函数计算，即将模型估计的边际均值差除以模型残差标准差进行标准化 (Lenth, 2023)。在本研究的 PIT 任务流畅性等指标上，模型残差标准差较小，使据此计算得到的标准化效应量数值偏大。为使效应量报告更具可解释性，并符合 Cohen’s d 的经典定义 (Cohen, 1988)，我们改用基于观测均值差和合并标准差的 Cohen’s d，即以两组观测均值之差除以两组观测标准差的合并估计值，重新计算了所有两两比较的效应量。修订后，3.2.2 和 3.3.2 节中 AUT 与 PIT 角色分析的事后比较的效应量均已更新，效应量大小也更符合 Cohen’s d 的常规解释范围。

再次，关于表格中部分显著交互作用的呈现与解释，我们已逐项核查相关模型结果，并在修订稿中补充相应的简单效应分析。例如，在 AUT 任务灵活性和 PIT 任务适用性指标上，部分模型未发现显著主效应，但存在显著交互作用。为避免结果呈现不充分，修订稿已补充报告相应的简单效应分析结果，并对 AUT 任务灵活性指标上的角色差异进行了简要讨论。请参阅结果部分 3.2.3 小节(p.37)、3.3.3 小节(p.41)和讨论部分 4.3 小节(p.47)。

最后，关于术语使用不统一的问题，原稿中确实存在“适用性”和“有用性”混用的情况。修订稿已将相关表述全部统一为“适用性”。同时，我们也全面复核了全文统计报告、表文对应关系、图注说明和显著性标注，以确保正文、表格和图示之间保持一致。

意见 4

4. “理论与实践价值”较为笼统，尚未充分说明本研究对既有理论和分析框架的具体贡献。作者在引言中以 PAT 理论解释观点采纳为何影响创造性表现，并以 APIM 区分行动者效应和同伴效应，但理论贡献部分并未明确说明本研究结果究竟是支持、限定还是拓展了 PAT，也未说明将 APIM 应用于人智协作二元组相较于既有人际互动研究有何方法学增量。建议作者围绕核心发现，具体阐明人智协作中的角色不对称和任务依赖模式如何修正或扩展传统团体创造中的观点采纳机制。

回应：感谢您的宝贵建议。我们同意原稿中“理论与实践价值”部分较为笼统，未能充分说明本研究对既有理论和分析框架的具体贡献。根据您的建议，我们已重写了这一小节，主要修改如下：

首先，我们明确了本研究对 PAT 理论的贡献。在理论层面，本研究为 PAT 理论在人智协作情境中的适用性提供了证据，并进一步限定了其解释边界。PAT 理论认为，个体可以通过采纳他人视角发现新的行动可能性，从而促进创造性想法的发展。既有研究主要基于人类同伴之间的互动，将观点采纳视为一种相对互惠过程(Bai et al., 2026)。本研究结果表明，观点采纳的作用会随协作对象、任务类型和创造性指标而变化。这一结果有助于扩展传统团体创造中关于观点采纳机制的认识。与双人协作不同，人智协作并没有延续这种互惠模式，而是呈现出更明显的角色不对称。AI 更容易顺承和延续已有观点，但这种采纳不一定带来更高质量的创造性表现；相比之下，人类对 AI 观点的主动理解、筛选和整合，更可能推动想法向新的方向发展。因此，人智协作中的观点采纳机制并不是传统双人协作机制的简单延伸，而是由人类和 AI 角色差异共同塑造的非对称互动过程。

其次，我们补充了 APIM 应用于人智协作二元组分析的方法学意义。我们指出，传统组间比较难以揭示双方在互动过程中如何相互影响(Huang et al., 2026)，而 APIM 可以区分行动者效应和同伴效应，从而分析自身观点采纳和同伴观点采纳分别如何预测创造性表现。将 APIM 用于人智协作二元组，有助于研究者从过程层面解释人智协作中创造性表现的形成过程。

此外，我们也修改了实践价值部分，使其更加贴合研究结果。我们强调，与 AI 进行创造性协作时，使用者不宜直接接受 AI 观点，而应根据任务目标进行选择性评估、拓展和整合；AI 系统设计也可围绕观点采纳过程提供支持，例如提示用户比较不同观点、保留多种问题表征，或主动提供替代视角，以促进更加开放的创造性协作。具体修改请参见讨论部分“4.4 理论与实践价值”小节(p. 49)。

参考文献

Bai, H., Chan, L. S., Luo, A., Kroesbergen, E. H., & Christie, S. (2026). Creativity in dialogues: How do children interact with parents vs. strangers for generating creative ideas? *Learning and Instruction*, *102*, 102286. https://doi.org/10.1016/j.learninstruc.2025.102286

Brucks, M. S., & Levav, J. (2022). Virtual communication curbs creative idea generation. *Nature*.

https://doi.org/10.1038/s41586-022-04643-y

Cheng, X., & Zhang, L. (2025). Inspiration booster or creative fixation? The dual mechanisms of LLMs in shaping individual creativity in tasks of different complexity. *Humanities and Social Sciences Communications*, *12*(1), 1563. https://doi.org/10.1057/s41599-025-05867-9

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Lenth, R. (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means. *R package version 1.8. 5*.

Lu, K., Yu, T., & Hao, N. (2020). Creating while taking turns, the choice to unlocking group creative potential. *NeuroImage, 219*, 117025.

Glăveanu, V. P., Gillespie, A., & Karwowski, M. (2019). Are people working together inclined towards practicality? A process analysis of creative ideation in individuals and dyads. *Psychology of Aesthetics, Creativity, and the Arts, 13*(4), 388–401.

Guilford, J.P., 1967. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, *1(1)*:3-14.

Erwin, A. K., Tran, K., & Koutstaal, W. (2022). Evaluating the predictive validity of four divergent thinking tasks for the originality of design product ideation. *PLoS ONE, 17*(3), e0265116.

Howard, T. J., Dekoninck, E. A., & Culley, S. J. (2010). The use of creative stimuli at early stages of industrial product innovation. *Research in Engineering Design, 21*(4), 263–274.

Huang, S., Long, L., Zhu, Y., & Zhu, J. N. Y. (2026). Human–GenAI collaboration across creative phases: Cognitive mechanisms shaping novelty and usefulness. *International Journal of Information Management*, *86*, 102986. https://doi.org/10.1016/j.ijinfomgt.2025.102986

Paek, S. H., Abdulla Alabbasi, A. M., Acar, S., & Runco, M. A. (2021). Is more time better for divergent thinking? A meta-analysis of the time-on-task effect on divergent thinking. *Thinking Skills and Creativity, 41*, 100894.

Torrance, E. P. (2008). *Torrance tests of creative thinking: Norms-technical manual, verbal forms A and B.* Scholastic Testing Service.

第三轮

审稿专家 3：作者很好回应了问题，并针对性修改了论文。没有进一步意见。推荐发表。

编委意见：同意审稿人意见，建议录用。

主编 1 意见：这篇文章我看过了，没有问题，同意发表。

主编 2 意见：同意发表。