《心理学报》审稿意见与作者回应

题目：AI 观点发散性对团体创造表现的影响及其认知神经机制

作者：周志豪，乔熙诺，张文宇，童帅，郝宁

第一轮

审稿人 1 意见：

该稿件围绕生成式人工智能嵌入团体创造过程这一前沿问题，考察 AI 观点发散性对团体创造表现的影响及其认知神经机制。研究采用 2(AI 观点发散性：高、低)×2(创造性任务类型：科学、日常)混合实验设计，以双人小组为分析单位，结合 AI 调用策略、观点采择策略、语义特征以及 fNIRS 超扫描指标，分析 AI 观点发散性对团体创意新颖性和实用性的影响。研究选题具有较强的时代性和理论价值，尝试将团体创造力、人智共创和社会神经科学结合起来，具有一定创新性。有以下意见供作者参考。

意见 1：

当前引言同时涉及人工智能、团体创造力、团体动机性信息加工模型、创造力双路径模型、语义特征和 fNIRS 超扫描等多个理论与方法线索，内容较为丰富，但主线略显分散。建议作者进一步明确全文的核心理论问题，以增强论文的理论集中度。

回应：

感谢您建设性的意见！我们对引言部分进行了重新梳理。修改后的引言明确以团体动机性信息加工模型作为全文的主导理论框架，并围绕“AI 观点发散性如何改变团队所面对的外部信息结构，以及这种外部信息结构的变化如何在不同任务约束下，经由团体信息加工过程影响团体创意的新颖性与实用性”这一核心问题逐层展开。围绕这一核心问题，我们对原稿进行了三方面修改。

首先，我们保留生成式 AI 与人智共创的相关文献，但将其功能限定为引出研究问题，即团体创造研究需要从探讨“是否使用 AI”转向分析“AI 输出观点特征如何影响团体创造表现”。其次，我们弱化创造力双路径模型在引言中的地位，不再将其作为与团体动机性信息加工模型并列的理论框架，而是将新颖性与实用性的分化主要纳入团体动机性信息加工模型

下进行解释。具体来说，高 AI 观点发散性提供了更多异质信息，可能有利于发散性信息加工和创意扩展，但也可能增加筛选、验证和收敛整合负担，进而使新颖性与实用性出现分化。再次，我们将科学创造性任务和日常创造性任务的差异纳入团体动机性信息加工模型的理论逻辑中进行说明，即不同任务在知识约束、问题开放性、目标明确性和评价标准上的差异，会影响团队将异质信息转化为创造性产出的效果。因此，任务类型不再被视为一个外加的材料变量，而是被定位为 AI 观点发散性影响团体信息加工过程的重要边界条件。

同时，我们对观点语义特征和 fNIRS 超扫描相关内容进行了重新定位。修改后，AI 调用策略、观点采择策略和观点语义特征被表述为团体信息加工过程的行为和产出表征，前额叶脑间同步与单脑功能连接则被表述为理解该过程的神经关联线索，而不再作为与核心理论问题并列的独立主线。通过上述修改，引言的理论层级更加清楚，全文主线也更加集中于团体动机性信息加工模型在人智共创情境中的拓展。

正文修改内容:

(1)因此，相较于团队是否使用 AI，更值得关注的科学问题是，AI 输出观点的特征如何改变团队的信息加工过程，并进一步影响团体创造表现？详见第 5 页第 1 段。

(2)团体动机性信息加工模型(Motivated Information Processing in Groups Model, MIP-G)为理解 AI 观点发散性影响团体创造这一过程提供了理论框架。该理论指出，团体创造并非信息数量的简单累加，而取决于成员在面对多样信息时，是否进行开放、系统而精细的比较、整合与重组(De Dreu et al., 2008; Nijstad & De Dreu, 2012)。详见第 5 页第 2 段。

(3)AI 观点发散性对团体创造的作用也需放在具体任务情境中理解。根据 MIP-G 理论，信息多样性能否促进团体创造，并不取决于信息数量本身，而取决于团队能否围绕任务目标对多样信息进行充分加工(De Dreu et al., 2008; Nijstad & De Dreu, 2012)。详见第 5 页第末段。

(4)MIP-G 理论提出，外部信息能否促进团体创造，关键取决于团队如何围绕这些信息进行搜索、比较、采择、整合与重表征。从过程角度看，团队在接收并利用 AI 观点时，通常需要经历提问、筛选、采择、整合与重表征等环节(Ivcevic & Grandinetti, 2024)，这些正是本研究考察 AI 调用策略、观点采择策略和观点语义特征的理论依据。详见第 6 页第 2 段。

(5)由此，将前额叶脑间同步与单脑功能连接纳入同一研究框架，有助于从神经关联层面补充说明 AI 观点发散性影响团体信息加工过程的可能方式。详见第 7 页首段。

(6)综上，本研究以团体动机性信息加工模型为理论框架……并通过上述行为变量进一步关联团体创造表现。详见第 7 页第 2 整段。

意见 2：

题目和摘要中多次使用“认知与神经机制”“阐明机制”等表述，虽然脑、行为整合分析虽然具有启发性，但尚不能严格证明脑间同步、单脑功能连接与策略选择之间的因果方向。因此，建议在讨论中更加谨慎地区分因果证据、关联证据和探索性发现。

回应：

非常感谢您的建议！我们对相关表述进行了修改。首先，我们将题目中的“认知与神经机制”调整为“认知加工基础与神经关联”，以避免过强的机制性表述。其次，我们将摘要中“阐明了……认知与神经机制”的表述调整为“本研究为理解……认知加工基础及其神经关联提供了实证证据”。再次，我们在讨论部分增加了对证据类型的说明，明确区分 AI 观点发散性实验操纵所支持的效应证据，以及脑间同步、单脑功能连接与行为策略之间主要体现的关联性证据。通过上述修改，本文对脑-行为整合结果的解释更加谨慎，也更符合本研究设计所能支持的推论范围。

正文修改内容:

(1)标题改为“AI 观点发散性影响团体创造表现的认知加工基础与神经关联”。

(2)本研究为理解 AI 观点发散性影响团体创意新颖性与实用性的认知神经机制提供了实证证据，进一步拓展了团体动机性信息加工模型在人智共创情境中的应用范围。详见摘要末句。

(3)本研究采用双人 fNIRS 超扫描范式，操控 AI 观点发散性并设置科学与日常两类创造性任务，系统考察了 AI 观点发散性对团体创造表现的影响，分析了相关认知加工基础和神经关联。详见第 20 页末段。

(4)本研究将团体创造力的考察从传统人际协作情境拓展到人智共创情境，并为理解 AI 观点发散性如何通过认知加工基础和神经关联特征影响团体创造表现提供了新的证据。详见第 21 页首段。

(5)需要强调的是，本研究中 AI 观点发散性为实验操纵变量，因此其对团体创造表现的影响可以在实验设计范围内理解为实验因果效应。相比之下，脑间同步、单脑功能连接与 AI 调用策略、观点采择策略、观点语义特征之间的中介路径，主要反映脑活动指标、认知加工基础与团体创造表现之间的关联模式。也即，脑-行为整合分析有助于揭示人智共创过程中可能存在的神经关联线索，但尚不能严格证明脑间同步或单脑功能连接会导致特定策略选择。因此，本文在解释相关结果时，将其视为神经关联证据，而不是严格意义上的神经因果机制。详见第 23 页末段。

意见 3：

本文采用 ChatGPT-4o 和 Deepseek V3.2 对创意的新颖性和实用性进行评分，并抽取 15%创意进行人工评分验证，该方法具有创新性。但请补充 LLM 评分所使用的具体提示词和评分标准，保证研究的可重复性；能否提高人工评分比例，或至少在人工评分子样本中重复核心分析，以检验结果稳健性。这样可以增强因变量测量的可信度。

回应：

感谢您的宝贵意见！我们认同您关于提高因变量测量可信度和研究可重复性的建议。我们进行了两方面修改。首先，我们在方法部分补充了 LLM 评分和人工评分所使用的具体提示词和评分标准。评分时，评分者首先看到具体创造性任务题目和被试生成的创意，然后分别对创意的新颖性和实用性进行 5 点评分。ChatGPT-4o、Deepseek V3.2 和人工评分者均采用相同的任务说明和评分标准，以保证不同评分来源之间的可比性。

其次，结合专家建议，我们将人工评分比例由原来的 15%提高至 30%。创造力主观评分研究通常需要在评分成本和测量效度之间取得平衡，既有研究曾提出 Top-2 评分、Snapshot 评分等部分评分方法，以降低评分负担并检验其信度和效度(Silvia et al., 2008; Shaw, 2021)。因此，在本研究中，我们将人工评分比例提高至 30%，以更充分地检验 LLM 评分与人工评分的一致性。需要说明的是，现有文献并未规定人工复核必须达到某一固定比例，因此我们并不将 30%解释为通用标准，而是将其作为在评分工作量和效度验证之间取得平衡的改进方案。

关于专家建议的“在人工评分子样本中重复核心分析”，我们经过文献查阅和慎重讨论后认为，使用部分创意的人工评分子样本重复核心混合方差分析，可能并不适合作为本文主结果的稳健性检验。原因在于，本文的团体创造表现以每个小组在每类任务中生成的全部 8 条创意的平均新颖性和实用性为指标，反映的是小组在完整任务过程中的整体创造表现。如果仅从 8 条创意中抽取部分创意，并以这些局部创意的均值替代完整任务表现，就会改变因变量的测量含义。发散性思维序列位置效应表明，创意质量可能随生成顺序发生变化，后期创意与早期创意并不一定具有相同的新颖性水平和实用性水平，这一序列位置效应意味着部分创意未必能够稳定代表完整创意集合(Beaty & Silvia, 2012)。同时，创造力主观评分研究也将均值评分、Top-2 评分和 Snapshot 评分视为不同的评分指标，而不是可以简单互相替代的同一指标(Silvia et al., 2008; Shaw, 2021)。因此，如果用部分人工评分创意重新计算小组得分并进行混合方差分析，其结果很可能受到抽样位置和局部创意波动的影响。若其结果与主分析一致，并不能充分证明完整任务层面的主结果稳健。若其结果与主分析不一致，也不能

据此否定基于全部创意的主分析结果。

基于上述考虑，我们没有使用部分创意的人工评分均值替代完整任务得分并进行混合方差分析，而是将人工评分子样本用于检验评分工具的可靠性和评分来源的一致性。具体而言，我们在方法部分补充报告了人工评分者与 AI 评分者之间的一致性，以及人工评分与 AI 评分之间的相关结果。通过提高人工评分比例、补充评分提示词和评分标准，并检验人工评分与 AI 评分的一致性和相关性，本文进一步增强了因变量测量的透明度、可重复性和可信度。

正文修改内容:

(1)为保证评分过程的可重复性，本研究说明 AI 评分和人工评分所使用的提示词。评分时，评分者首先看到具体创造性任务题目和被试生成的创意，然后分别对创意的新颖性和实用性进行评分。以科学创造性任务中的透明陶瓷材料题目为例，新颖性评分提示词为“某课题组开展了一项创造力实验，要求被试围绕题目‘透明陶瓷材料还可以用于哪些场合’这一任务生成创意。透明陶瓷材料通过选用高纯原料，并通过工艺手段排除气孔获得。具有温和透光、耐高温、电绝缘性高、红外线穿透性强、化学稳定性高等特点。该实验收集了多个创意，现在请你从挑剔、严肃的视角来评价这些创意的新颖性。请采用 5 点评分，1 = 一点也不新颖，5 = 非常新颖”。实用性评分时，将上述提示词中的“新颖性”替换为“实用性”。日常创造性任务的评分提示词保持相同结构，仅替换其中的任务题目。ChatGPT-4o、Deepseek V3.2 和人工评分者均采用相同的任务说明和评分标准。详见第 8 页第 3 整段。

(2)为进一步检验大语言模型评分的可靠性，本研究从创意池中随机抽取 30%的创意交由人工评分。创造力主观评分研究通常需要在评分成本和测量效度之间取得平衡，已有研究也曾采用部分评分方式检验创造性评分的信度和效度(Silvia et al., 2008; Shaw, 2021)。本研究据此将人工评分子样本定位为 AI 评分可靠性的验证样本，而非替代完整创意集合的主要因变量。人工评分样本采用分层随机抽样方式获得，以保证抽取创意覆盖 AI 观点发散性条件、创造性任务类型和创意生成顺序。两名具有 3 年以上创造力评分经验的研究生根据与 AI 评分相同的任务说明和评分标准，对抽取创意的新颖性和实用性进行独立评分。两个 AI 评分者与两名人工评分者在新颖性和实用性维度评分一致性良好，ICC 分别为 0.84 和 0.78。人工评分与 AI 评分在新颖性($r = 0.81$, $p < 0.001$)和实用性($r = 0.75$, $p < 0.001$)上均呈显著正相关。上述结果表明，AI 评分与人工评分具有较好一致性(Zhou et al., 2026)。详见第 8 页末段。

参考文献:

Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the

serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(4), 309–319. https://doi.org/10.1037/a0029171

Shaw, A. (2021). It works…but can we make it easier? A comparison of three subjective scoring indexes in the assessment of divergent thinking. *Thinking Skills and Creativity*, *40*, 100789. https://doi.org/10.1016/j.tsc.2021.100789

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 68–85. https://doi.org/10.1037/1931-3896.2.2.68

意见 4：

稿件通过 AI 观点之间的语义离散度检验高、低发散性操纵是否成功，这一做法是合理的。但高温度和发散性较高/较低的提示词可能同时影响 AI 输出的字数、观点数量、具体性、可行性和表达质量。因此，建议作者补充分析高低发散条件下 AI 输出长度、观点数量、AI 观点本身的新颖性和实用性是否存在差异，并在必要时将这些变量作为协变量纳入稳健性分析，以证明核心效应确实来源于观点发散性，而非信息量或输出质量差异。

回应：

非常感谢您的提醒！这是我们之前未考虑到的问题。我们将更清楚地阐明本研究中 AI 观点发散性操纵与 AI 输出过程性特征之间的关系。

通过回顾实验设计和多次讨论后，我们认为，本研究中的 AI 输出长度、观点数量以及 AI 观点本身的新颖性和实用性，不宜简单作为独立于 AI 观点发散性之外的协变量处理。原因在于，本研究采用的是自由交互式人智共创范式，而非固定 AI 材料呈现范式。在实验过程中，研究者控制的是实验期间 AI 系统整体的观点发散倾向，即通过温度参数和提示词控制 AI 回复的发散水平，但并未强制限制每次 AI 回复的长度、观点数量或内容质量。被试可以根据任务进展和对已有 AI 回复的判断，通过主动调整 AI 调用策略来改变 AI 输出内容呈现的数量与质量。例如，当被试对已有 AI 观点不满意时，可能要求 AI 生成更多观点(如 2.2 节 AI 调用策略中所涉及到的广泛性生成策略)；当被试感到 AI 观点过多、难以筛选时，也可能要求 AI 围绕特定方向生成少量观点(如方向性生成策略)；被试还可能直接要求 AI 生成更新颖或更实用的观点(如完善性探索策略)。因此，AI 输出的信息量和质量并不是独立于团体信息加工过程之外的背景变量，而是 AI 观点发散性、团队 AI 调用策略和团队信息加工需求之间共同作用的过程性结果(Zhou et al., 2026)。

基于这一实验设计逻辑，如果将 AI 输出长度、观点数量或 AI 观点本身的新颖性、实用性作为协变量纳入核心分析，可能会带来过度控制问题。具体而言，这些变量本身可能是

团队 AI 调用策略、观点采择过程和语义加工过程的一部分，而本文的研究目的正是考察团队如何在不同 AI 观点发散性条件下主动调用、筛选、采择和整合 AI 信息。将这些过程性产物作为协变量控制，可能会削弱甚至剔除本研究希望解释的人智交互过程。

因此，我们没有将 AI 输出长度、观点数量、AI 观点本身的新颖性和实用性作为协变量纳入核心稳健性分析，而是在修订稿中进一步明确这一设计逻辑，并在局限与展望部分补充说明。未来研究若希望更严格地区分 AI 观点发散性、输出数量和输出质量的独立作用，可以采用固定 AI 回复材料、限制每次 AI 输出观点数量、控制输出长度，或将 AI 观点发散性与 AI 观点质量正交操纵的实验设计。与此同时，我们也在正文中强调，本文并未忽略 AI 交互过程中信息数量和内容质量变化的问题，而是通过 AI 调用策略、观点采择策略和观点语义特征等指标，将其置于人智共创的过程性分析框架中加以考察。

正文修改内容:

(1)需要说明的是，本研究采用自由交互式人智共创范式，研究者控制的是 AI 系统在整个任务过程中的观点发散倾向，而非每次 AI 回复的固定材料内容。因此，实验过程中并未强制限制每次 AI 回复的字数、观点数量或观点质量。被试可以根据任务进展和对 AI 回复的判断，自主决定如何向 AI 提问、是否要求 AI 生成更多或更少的观点，以及是否要求 AI 围绕特定方向、已有创意或评价标准进行回复。由此产生的 AI 输出差异被视为人智共创过程的一部分，并通过后续 AI 调用策略、观点采择策略和观点语义特征等指标加以刻画。详见第 12 页第 1 段。

(2)此外，鉴于本研究采用自由交互式 AI 协作范式，AI 输出长度、观点数量及输出内容质量可能受到被试 AI 调用策略和任务进展的共同影响，因此本研究未将这些过程性产物作为独立协变量纳入核心分析，而是通过 AI 调用策略、观点采择策略和观点语义特征等指标分析团队对 AI 信息的主动加工过程。详见第 12 页第 4 段。

(3)第二，本研究采用自由交互式人智共创范式，有助于提高实验情境的生态效度，但也使 AI 输出的具体信息量和内容质量具有一定动态性。虽然本研究通过温度参数和提示词控制了 AI 系统整体的观点发散倾向，并通过语义离散度验证了高、低 AI 观点发散性操纵的有效性，但在实际交互过程中，被试仍可根据任务进展和对 AI 回复的判断主动调整 AI 调用策略，从而影响 AI 输出的长度、观点数量以及观点的新颖性和实用性。因此，这些 AI 输出特征在本研究中更适合被理解为人智交互过程的组成部分，而不是独立于团体信息加工之外的外在混淆变量。未来研究可进一步采用固定 AI 回复材料、限制每次 AI 输出观点数量、控制输出长度，或将 AI 观点发散性与 AI 观点质量进行正交操纵的设计，以更严格地

区分 AI 观点发散性、信息量和观点质量的独立作用。详见第 24 页第 3 段。

参考文献：

Zhou, Z., Xiao, L., Wang, J., Tong, S., Zhang, W., & Hao, N. (2026). Group-AI collaboration enhances creativity performance: The roles of perspective-taking and AI utilization strategies. *Journal of Computer Assisted Learning*, *42*(2), e70223. https://doi.org/10.1002/jcal.70223

意见 5：

结果显示，排除性生成策略既参与新颖性的中介路径，也参与实用性的中介路径，但其方向与主效应之间的关系并不完全直观。建议作者明确区分解释性中介和竞争性中介或抑制性路径。

回应：

感谢您指出这一问题。我们重新梳理了排除性生成策略的中介路径，并在结果和讨论中明确区分解释性中介与竞争性中介或抑制性路径。按照中介分析相关方法学观点，当间接效应方向与总效应方向一致时，可以理解为解释性或互补性中介；而当间接效应方向与总效应方向相反时，则更适合理解为竞争性中介或抑制性路径(MacKinnon et al., 2000; Zhao et al., 2010)。据此，本研究中排除性生成策略的作用并非单纯解释 AI 观点发散性的主效应，而是反映了一条与主效应方向相反的竞争性路径。

具体而言，在新颖性上，高 AI 观点发散性总体上提高团体创意的新颖性，但高 AI 观点发散性也会减少团队使用排除性生成策略，而排除性生成策略本身有助于团队继续主动求异、避开已有创意类型并扩展搜索空间。因此，AI 观点发散性经由排除性生成策略对新颖性的间接效应为负向，与总效应方向相反。该路径表明，高 AI 观点发散性虽然总体上提高新颖性，但它也可能通过减少团队主动调用 AI 继续求异的行为，削弱一部分新颖性增益。因此，这一路径更适合被解释为竞争性或抑制性路径。

在实用性上，高 AI 观点发散性总体上降低团体创意的实用性，但其通过减少排除性生成策略，反而对实用性产生正向间接作用。这表明，较少使用排除性生成策略可能降低团队继续扩展远距离创意的倾向，从而减少额外筛选、协调和评估负担，并在一定程度上缓冲高 AI 观点发散性对实用性的负向影响。因此，排除性生成策略在实用性中的作用同样不是解释性中介，而是一条与总效应方向相反的竞争性路径。

基于上述修改，我们在正文中进一步说明，排除性生成策略具有双重功能。一方面，它是一种主动求异策略，有助于提升新颖性。另一方面，它也可能增加信息发散和筛选负担，

不利于实用性。正因如此，排除性生成策略在新颖性和实用性中的中介方向不同，也与主效应之间形成竞争关系。修订后，结果解释更加清楚地区分了能够解释主效应的中介路径和与主效应相反的竞争性中介路径。

正文修改内容:

(1)在解释中介结果时，进一步依据间接效应与总效应方向的关系区分不同类型的中介路径。当间接效应方向与总效应方向一致时，将其理解为解释性中介路径。当间接效应方向与总效应方向相反时，将其理解为竞争性中介或抑制性路径(MacKinnon et al., 2000; Zhao et al., 2010)。详见第 12 页第 4 段。

(2)在新颖性上，仅有 AI 观点发散性经由排除性生成次数影响新颖性的中介路径达到显著。该路径表现为负向间接效应(b = -0.02, p = 0.048, 95% CI [-0.06, -0.01]，见图 4A)。结合总效应方向可知，高 AI 观点发散性总体上提高新颖性，但其经由排除性生成次数的间接效应方向相反。具体而言，高 AI 观点发散性减少了团队使用排除性生成策略的次数，而排除性生成策略本身与更高的新颖性相关，因此该路径会削弱高 AI 观点发散性对新颖性的促进作用。由此可见，排除性生成策略在新颖性上更适合被理解为竞争性中介或抑制性路径。详见第 17 页第 2 段。

在实用性上，显著中介路径相对更多。在 AI 调用策略上，AI 观点发散性可通过排除性生成次数间接影响实用性(b = 0.02, p = 0.040, 95% CI [0.01, 0.04])。结合总效应方向可知，高 AI 观点发散性总体上降低实用性，但其经由排除性生成次数的间接效应为正向(见图 4B)。具体而言，高 AI 观点发散性减少了团队使用排除性生成策略的次数，而较少的排除性生成进一步对应更高的实用性。因此，该路径并不是解释高 AI 观点发散性为何降低实用性的解释性中介，而是与总效应方向相反的竞争性中介路径，说明较少使用排除性生成策略在一定程度上缓冲了高 AI 观点发散性对实用性的负向影响。详见第 17 页第 3 段。

(3)结果显示，高 AI 观点发散性减少了团队调用 AI 排除性生成策略的频次，而排除性生成策略在新颖性和实用性上呈现出方向不同的竞争性中介作用。在新颖性上，高 AI 观点发散性总体上有利于提高团体创意的新颖性，但其减少排除性生成策略的使用，削弱了团队继续主动求异和避开已有创意类型的加工倾向，从而抵消了部分新颖性增益。在实用性上，高 AI 观点发散性总体上降低团体创意的实用性，但其减少排除性生成策略的使用，又可能降低过度发散带来的筛选、协调与评估负担，从而在一定程度上缓冲实用性损失。因此，排除性生成策略并不是解释 AI 观点发散性主效应的单一中介，而是一条与主效应方向相反的竞争性路径。这表明，AI 排除性生成策略并非单纯有利或有害，而是一种同时具有主动求

异和加工负担双重含义的策略。详见第 21 页末段。

参考文献：

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181. https://doi.org/10.1023/A:1026595011371

Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

.....

审稿人 2 意见：

文章在生成式人工智能参与团队协作的背景下，探讨了 AI 观点发散性对团体创造表现的影响机制，并结合行为、语义特征及脑神经指标进行多层次分析。研究选题具有较强的现实意义与前沿价值，方法整合较为系统，关键变量的界定与操作化也较为清晰。不过，文章在理论框架整合及复杂分析结果的解释方面仍有进一步提升空间。

意见 1：

摘要中“结果”与结论性表述之间的界限尚不清晰。部分表述(如“阐明了……机制”、“拓展了……解释边界”)更偏向理论意义或研究贡献，建议与具体结果加以区分。

回应：

感谢您的细致意见！在之前的修改中，我们已根据审稿专家 1 的建议，对题目、摘要和讨论中可能暗示强因果机制的表述进行了调整，将研究定位为对认知加工基础与神经关联的考察，而不是对严格神经因果机制的证明。

根据您的建议，我们对摘要结尾部分进行了结构性调整。具体而言，我们将摘要中的实证发现集中表述为结果层面的内容，包括 AI 观点发散性对新颖性和实用性的差异化影响，AI 排除性生成策略和观点语义特征的中介作用，以及前额叶脑间同步、单脑功能连接与相关行为变量之间的间接关联。随后，我们将理论意义单独置于摘要最后一句，并使用“提供实证证据”和“拓展应用范围”等表述，使读者能够清楚区分研究结果与基于结果提出的理论贡献。

正文修改内容:

(1)中介分析表明，AI 排除性生成策略及人类观点语义特征在 AI 观点发散性影响团体创造表现的过程中发挥中介作用。脑-行为整合分析显示，前额叶脑间同步与单脑功能连接可通过 AI 调用策略、观点采择策略与观点语义特征等变量，与团体创意新颖性和实用性产生

间接关联。本研究为理解 AI 观点发散性影响团体创意新颖性与实用性的认知神经机制提供了实证证据，进一步拓展了团体动机性信息加工模型在人智共创情境中的应用范围。详见摘要后半段。

意见 2：

引言的第二段，“有研究发现，AI 能够提高.....AI 观点可能诱发创意同质化(Anderson et al., 2024)。”，文献支撑略显不足，“但与此同时，过高的发散性也可能增加信息的筛选.....进而降低实用性。”建议作者进一步补充相关参考文献，以增强论述的充分性与说服力。

回应：

感谢您的宝贵建议！我们进一步补充了两类参考文献。第一类是关于生成式 AI 对创造表现和创意多样性的影响研究。近期研究表明，生成式 AI 可能提升个体层面的创意质量，但同时也可能使不同个体的创意产出更趋相似，从而降低集体层面的创意多样性。例如，Doshi 和 Hauser(2024)发现，AI 辅助能够提高他人对个体创意作品的评价，但也会使不同创作者的作品更加相似。Anderson 等人(2024)发现，当 ChatGPT 作为个体创造过程的支持工具时，不同使用者生成想法间的语义差异性下降。Moon 等人(2025)也发现，尽管 AI 可能生成较高质量内容，但其广泛使用可能削弱集体层面的创意多样性。上述研究能够更充分地支撑本文关于 AI 促进创造性表现与诱发同质化风险并存的论述。

第二类是关于信息过载、有限认知资源和外部刺激固着的研究。信息过载研究表明，当信息处理要求超过个体有限的注意和工作记忆资源时，过多信息会降低信息加工效率并损害决策质量(车敬上 等, 2019)。设计固着研究也表明，外部示例或 AI 生成内容可能并非单纯拓展创意空间，也可能使个体围绕外部刺激进行模仿、锚定或局部调整，从而限制进一步探索(Wadinambiarachchi et al., 2024)。基于这些文献，我们在引言中进一步说明，较高的 AI 观点发散性虽然能够为团队提供更异质的外部刺激，但当这些刺激过于分散、过多或难以与任务目标相匹配时，团队需要投入更多资源进行筛选、比较、验证和收敛整合，因此可能削弱创意的实用性。

正文修改内容:

(1) 另有研究表明，AI 辅助可能提升个体层面的创意质量，但也可能使不同个体的创意产出更趋相似，从而削弱集体层面的创意多样性(Anderson et al., 2024; Doshi & Hauser, 2024; Moon et al., 2025)。因此，相较于讨论团体是否应该使用 AI，更值得深究的科学问题是，AI 输出观点的特征如何改变团体的信息加工过程，进而影响团体创造表现。详见第 5 页第 1

段。

(2)但与此同时，过高的发散性也可能增加信息的筛选、协调与评估成本。信息过载研究表明，当信息处理要求超过个体有限的注意和工作记忆资源时，过多信息可能降低信息加工效率并损害决策质量(车敬上 等, 2019)。设计固着研究也表明，外部刺激并不总是促进创意扩展，AI 生成内容的示例可能诱发锚定和局部模仿，从而限制个体进一步广泛探索(Wadinambiarachchi et al., 2024)。因此，在团体创造过程中，过高的 AI 观点发散性虽然提供了更多异质信息，但也可能要求团队投入更多认知资源进行比较、验证和收敛整合，从而削弱团队围绕任务目标和现实约束形成可行方案的能力，进而降低创意的实用性。详见第 5 页第 2 段。

参考文献:

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th Conference on Creativity & Cognition* (pp. 413–425). ACM. https://doi.org/10.1145/3635636.3656204

车敬上, 孙海龙, 肖晨洁, 李爱梅. (2019). 为什么信息超载损害决策？基于有限认知资源的解释. *心理科学进展*, *27*(10), 1758–1768. https://doi.org/10.3724/SP.J.1042.2019.01758

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Moon, K., Green, A. E., & Kushlev, K. (2025). Homogenizing effect of large language models (LLMs) on creative diversity: An empirical comparison of human and ChatGPT writing. *Computers in Human Behavior: Artificial Humans*, *6*, 100207. https://doi.org/10.1016/j.chbah.2025.100207

Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q., & Velloso, E. (2024). The effects of generative AI on design fixation and divergent thinking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (380, pp. 1–18). ACM. https://doi.org/10.1145/3613904.3642919

意见 3：

本研究样本量为 180 人，女性 140 人，性别分布不均，该样本是否具有足够的统计效能？以及性别比例失衡是否可能对研究结果产生潜在影响？

回应：

感谢您的意见！我们十分认同关于样本量、统计效能和性别比例失衡可能影响研究结果的提醒。我们将更加谨慎地说明本研究样本特征，并进一步明确研究结果的解释边界。

首先，关于统计效能问题，本研究在实验前已采用 G*Power 3.1 进行了先验效能分析。依据 Cohen(2013)提出的中等效应量标准，设定效应量 $f = 0.25$、显著性水平 $\alpha = 0.05$、检验功效 $1 - \beta = 0.80$，计算得到达到检验要求的最小样本量为 120 名被试，即 60 个双人小组。

实际实验中共招募 180 名被试，因 1 名被试未遵守指导语，其所在双人小组数据被整体剔除，最终有效样本为 178 名被试，即 89 个双人小组。因此，虽然样本中女性比例较高，但本研究用于检验主要实验效应的有效样本量超过了先验效能分析所要求的最低样本量，能够满足本研究主要统计分析的效能要求。

其次，关于性别比例失衡是否可能影响研究结果的问题，我们进一步查阅了团体创造力与性别组成相关文献。既有研究并未显示性别组成对团体创造表现存在稳定一致的影响。例如，Lu 等人(2020)比较了女-男、女-女和男-男三类双人小组在团体创造任务中的表现，结果显示三类小组在创造表现上并无显著差异，但女-女小组在合作行为、集体灵活性和部分脑间同步指标上表现更高。Peter 等人(2021)的电子头脑风暴研究也指出，性别多样性对小组创造表现的影响并不稳定，可能取决于任务内容和群体分化线索等情境因素。基于这些研究，我们认为，性别比例失衡不太可能单独解释本研究中 AI 观点发散性和创造性任务类型对团体创造表现的主要影响模式。

不过，我们也认同专家的提醒。性别比例失衡仍可能对团体互动过程、观点采择方式、合作行为或神经同步模式产生潜在影响。尤其是本研究采用双人小组和 fNIRS 超扫描方法，性别组成可能并不一定直接改变最终创造表现，却可能影响成员间互动过程和神经协同特征。因此，我们在研究不足与展望部分补充了对性别比例问题的说明，明确指出本研究样本女性比例较高可能限制研究结果向性别更均衡或男性比例更高的团队情境推广。未来研究可在样本招募阶段进一步平衡性别比例，并系统比较男-男、男-女和女-女小组在人智共创任务中的行为表现、观点采择策略和神经协同差异。

正文修改内容:

(1)第三，本研究样本中女性被试比例较高。虽然本研究在实验前依据中等效应量标准进行了先验效能分析，最终有效样本为 178 名被试，即 89 个双人小组，超过了达到预期统计功效所需的最低样本量，但性别比例不均仍可能限制研究结果的外部推广性。既有团体创造研究表明，性别组成对团体创造表现的影响并不稳定。Lu 等人(2020)发现，女-男、女-女和男-男双人小组在创造表现上不存在显著差异，但女-女小组在合作行为、集体灵活性和部分脑间同步指标上表现更高。Peter 等人(2021)的电子头脑风暴研究也表明，性别多样性对小组创造表现的影响可能取决于任务内容和群体分化线索。因此，本研究中的性别比例失衡不太可能单独解释 AI 观点发散性和创造性任务类型对团体创造表现的主要影响模式，但仍可能对团体互动过程、观点采择方式或神经协同特征产生潜在影响。未来研究可在样本招募阶段进一步平衡性别比例，并系统比较男-男、男-女和女-女小组在人智共创任务中的行为表现、

策略使用和神经协同差异。详见第 24 页末段。

参考文献:

Lu, K., Teng, J., & Hao, N. (2020). Gender of partner affects the interaction pattern during group creative idea generation. *Experimental Brain Research*, *238*(5), 1157–1168. https://doi.org/10.1007/s00221-020-05799-7

Peter, L., Michinov, N., Besançon, M., Michinov, E., Juhel, J., Brown, G., Jamet, E., & Cherbonnier, A. (2021). Revisiting the effects of gender diversity in small groups on divergent thinking: A large-scale study using synchronous electronic brainstorming. *Frontiers in Psychology*, 12, 723235. https://doi.org/10.3389/fpsyg.2021.723235

意见 4：

建议作者对全文进行仔细校对，统一规范符号的使用，特别是中英文标点及半角/全角符号的转换问题

回应：

十分感谢您的提醒！我们已对全文进行了仔细校对和格式检查。修改过程中，我们重点检查了中英文标点、半角和全角符号、括号、逗号、句号、冒号、分号、统计符号和文献引用中的符号使用情况。尤其是对于正文中的括号，我们已将相关位置统一调整为半角括号，并对英文术语、统计结果和参考文献引用中的符号格式进行了统一处理。

意见 5：

本研究采用大语言模型(ChatGPT4o 和 Deepseek V3.2)对创意的新颖性与实用性进行评分，这一做法具有一定创新性，但其作为主要评分工具仍需进一步论证，建议补充更多文献佐证其合理性。

回应：

感谢您的提醒和建议！我们进一步查阅并补充了近年关于 AI 自动评分和大语言模型创造力评分的相关文献。已有研究表明，LLM 评分已经成为创造力测量中的重要新兴方法，尤其适用于开放式创意文本和发散思维任务的自动评分。例如，Luchini 等人(2025)基于 10 项创造性问题解决研究的数据发现，微调后的 LLM 能够较好预测人工评分的原创性和质量评分，且优于语义距离等传统自动评分指标。Saretzki 等人(2025)在约 50,000 条德语 AUT 反应中比较了 OCSAI、CLAUS 和 GPT-4 等 LLM 评分方法，结果显示 LLM 评分与人工创造力评分具有显著相关。Kranzle 和 Sharratt(2025)进一步比较了 GPT 模型与人类专家在创意评价中的一致性，指出在明确评价框架和适当提示词设计下，LLM 可以与专家评价达到较

高程度的一致性。上述研究表明，在开放式创造性产出评价中，采用大语言模型评分具有一定的经验依据。

同时，我们也认同专家关于谨慎使用 LLM 评分的提醒。因此，本研究并未将大语言模型评分视为完全替代人工判断的无条件工具，而是采取了多重验证措施。具体而言，我们采用两个大语言模型分别对全部创意进行评分，统一 AI 评分和人工评分的任务说明与评分标准，并将人工评分验证比例由 15%提高至 30%。结果显示，AI 评分与人工评分在新颖性和实用性上均具有较好一致性。基于这些文献依据和验证结果，我们将大语言模型评分定位为一种经过人工验证的自动化辅助评分方法，而不是完全替代人工创造力评价的唯一标准。

正文修改内容:

(1)本研究从新颖性和实用性两个维度对团体创造表现进行评估(Runco & Jaeger, 2012)。新颖性指小组生成创意的独特程度，实用性指创意的适用程度。近年来，随着自然语言处理和大语言模型的发展，AI 自动评分逐渐被用于开放式创造性产出的评价。已有研究表明，大语言模型对发散思维任务和创造性问题解决任务中创意的新颖性、实用性、精致性和完整性等维度的评分，与人工评分者的评分具有较高一致性(Kranzle & Sharratt, 2025; Luchini et al., 2025; Saretzki et al., 2025)。因此，参考既有创意评估方式和 AI 评分研究(Wan et al., 2025; Zhou et al., 2026)，本研究采用 ChatGPT-4o 和 Deepseek V3.2 对小组生成的全部创意进行新颖性和实用性评分，评分范围为 1 至 5 分(1 = 一点也不新颖/实用，5 = 非常新颖/实用)。对每个小组，在同一任务中生成的全部创意分别计算新颖性和实用性平均分，作为该小组在该任务中的相应得分，分数越高表示创造性表现越高。详见第 8 页第 2 段。

参考文献:

Kranzle, T., & Sharratt, K. (2025). Evaluating creative output with generative artificial intelligence: Comparing GPT models and human experts in idea evaluation. *Creativity and Innovation Management*, *34*(4), 991–1012. https://doi.org/10.1111/caim.70007

Luchini, S. A., Maliakkal, N. T., DiStefano, P. V., Laverghetta, A., Patterson, J. D., Beaty, R. E., & Reiter-Palmon, R. (2025). Automated scoring of creative problem solving with large language models: A comparison of originality and quality ratings. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000736

Saretzki, J., Knopf, T., Forthmann, B., Goecke, B., Jaggy, A.-K., Benedek, M., & Weiss, S. (2025). Scoring German alternate uses items applying large language models. *Journal of Intelligence*, *13*(6), 64. https://doi.org/10.3390/jintelligence13060064

意见 6：

讨论部分的第三段，“与此同时，科学创造性任务中.....独立生成和采择队友观点。”，文中指出科学任务与日常任务在观点采择策略上的差异，但对其背后的认知理论解释仍较为简略，建议进一步展开相关理论说明并补充参考文献。

回应：

感谢您的建议！我们对讨论部分相关内容进行了补充。根据审稿专家 1 的意见，我们将团体动机性信息加工模型明确为全文的主导理论框架(请见对专家 1 意见 1 的回应)。团体动机性信息加工模型强调，团体创造并不取决于信息数量本身，而取决于成员能否围绕任务目标对不同来源的信息进行开放、系统和精细加工。因此，不同创造性任务的知识约束、问题开放性和评价标准，会影响团队如何选择和采择 AI、队友以及自身经验等信息来源。

在科学创造性任务中，任务通常涉及材料属性、功能限制和应用场景匹配，具有更强的知识约束和更明确的评价标准。团队在此类任务中需要更多任务相关知识来降低不确定性，并判断候选创意是否符合科学材料特征和现实应用要求。因此，AI 观点更容易被团队视为可提供外部知识和候选方案的信息来源，团队也更可能通过直接采择 AI 观点来降低科学任务的不确定性。相反，日常创造性任务的问题空间更开放，任务内容更贴近日常生活经验和社会情境，成员自身经验和队友观点更容易被转化为创意资源。因此，团队在日常任务中更倾向于通过独立生成和采择队友观点来扩展问题空间。

通过上述修改，我们进一步说明了科学任务与日常任务在观点采择策略上的差异，并将这一差异纳入团体动机性信息加工模型关于任务目标、信息来源和加工方式相互作用的解释框架中。修改后的讨论既补充了专家所建议的理论说明，也保持了全文理论主线的集中性。

正文修改内容:

(1)与此同时，科学创造性任务中团队更频繁地直接采择 AI 观点，而日常创造性任务中则更倾向于独立生成和采择队友观点。根据团体动机性信息加工模型，团体创造并不取决于信息数量本身，而取决于成员能否围绕任务目标对不同来源的信息进行开放、系统和精细加工(De Dreu et al., 2008; Nijstad & De Dreu, 2012)。从这一视角看，两类任务在观点采择策略上的差异可能反映了团队对任务约束和信息来源适配性的动态调节。科学创造性任务通常涉及材料属性、功能限制和应用场景匹配，具有更强的知识约束和更明确的评价标准(Acar & van den Ende, 2016)。在此类任务中，团队需要更多任务相关知识来降低不确定性，并判断候选创意是否符合科学材料特征和现实应用要求。因此，AI 观点更容易被团队视为可提供外部知识和候选方案的信息来源，团队也更可能通过直接采择 AI 观点来降低科学任务的不

确定性。相比之下，日常创造性任务的问题空间更开放，任务内容更贴近日常生活经验和社会情境，成员自身经验和队友观点更容易被转化为创意资源。既有研究也表明，对他人观点的加工和整合有助于团队利用多样信息并形成创造性产出(Hoever et al., 2012)。因此，日常任务中更多的独立生成和队友观点采择，可能反映了团队在开放性较高、经验基础更丰富的情境中，更倾向于通过自身经验和人际互动扩展问题空间。上述结果表明，人智共创并非对 AI 建议的被动接受，而是团队根据任务目标、任务约束和信息来源可用性，动态调整加工方式的过程(Chen & Chan, 2024; Zhou et al., 2026)。详见第 22 页首段。

意见 7：

科学创造性任务和日常创造性任务是本研究核心因变量材料类型区分，AI 观点发散性对两类创造性任务的影响机制存在显著差异是本文核心研究发现，建议在引言、科学问题提出、研究假设及全文核心论述中，重点强化两类任务的区分，突出该核心研究点，以此凝练研究主旨，为研究成果的应用转化提供依据。

回应：

感谢您的建设性意见！我们对引言、科学问题提出和讨论部分进行了进一步调整。具体而言，我们将科学创造性任务与日常创造性任务明确表述为两类具有不同任务约束结构的创造情境。科学创造性任务具有更强的知识约束、更明确的评价标准和更高的可行性要求，团队需要将 AI 观点与材料属性、功能限制和应用条件进行匹配。日常创造性任务的问题空间更开放，评价标准更灵活，团队更可能将 AI 提供的远距离联想和类别扩展转化为新颖创意。由此，AI 观点发散性是否能够促进团体创造，并不只取决于 AI 观点本身是否多样，还取决于这种多样信息能否与具体任务的知识约束和评价要求相匹配。

同时，我们进一步修改了科学问题的表述，使其不再只是笼统提出“是否受到创造性任务类型调节”，而是直接突出两类任务差异。修改后的科学问题更明确地指向 AI 观点发散性是否会在科学创造性任务和日常创造性任务中产生不同作用，以及这种差异是否通过 AI 调用策略、观点采择策略和观点语义特征等信息加工过程体现出来。讨论部分也相应强化了这一核心发现，即高 AI 观点发散性并非普遍促进团体创造，而是更有利于开放性较高的日常创造性任务中新颖性的提升，却难以在知识约束较强的科学创造性任务中转化为新颖性收益，并在两类任务中均可能削弱实用性。

通过上述修改，本文将两类创造性任务的区分由实验材料层面的设置进一步提升为解释 AI 观点发散性作用边界的核心变量。这有助于更清楚地凝练研究主旨，也为后续教育和组

织场景中的 AI 协作设计提供应用启示。具体而言，在开放性较高、评价标准较灵活的日常创意任务中，可以适当提高 AI 观点发散性以激发新颖想法。而在科学创新、技术方案设计等知识约束较强的任务中，则需要更加重视 AI 输出的领域相关性、准确性和可行性，避免过度发散的信息增加团队筛选和验证负担。

正文修改内容:

(1)AI 观点发散性对团体创造的作用也需放在具体任务情境中理解……而是会在不同任务约束下对新颖性与实用性产生差异化作用。详见第 5 页末段。

(2)其一，AI 观点发散性是否会在科学与日常创造性任务中对团体创意的新颖性和实用性产生不同影响。即，高 AI 观点发散性是否更有利于高开放性日常任务中团体创意新颖性的提升，而对强约束科学任务则效果有限，且可能削弱两类任务中团体创意的实用性？其二，在科学与日常创造性任务中，AI 观点发散性如何通过影响团队对 AI 观点的加工过程（包括 AI 调用策略、观点采择策略及创意语义特征变化）进而影响最终的团体创造表现？其三，在 AI 观点发散性影响团体创造表现的过程中，前额叶脑间同步与单脑功能连接是否呈现出任务依赖性的变化，并经由上述行为变量与团体创造表现产生关联？详见第 7 页第 2 段。

(3)结果表明，科学创造性任务与日常创造性任务的区分，并非仅关乎实验材料的差异，而是理解 AI 观点发散性作用边界的关键任务条件。高 AI 观点发散性更容易在问题空间开放、评价标准灵活的日常任务中转化为新颖性收益，但在知识约束较强、可行性要求更高的科学任务中，这种发散输入并未带来同等的新颖性增益。同时，高 AI 观点发散性降低了两类任务的实用性，提示过度发散的信息可能增加团队筛选、验证和收敛整合负担。详见第 20 页末段。

意见 8：

在明确并突出两类任务类型核心作用的基础上，需进一步结合现有文献与理论依据，深度阐释 AI 观点发散性对科学创造性、日常创造性任务产生差异化影响的内在机制，弥补现有研究机制论证不足的问题:

从任务本身核心特征来看，科学创造性任务具备极强的专业知识约束性、明确且严苛的成果评价标准，问题解决需依托专业领域知识体系，属于高专业性、强规范性的创造性任务;而日常创造性任务问题空间更开放，专业知识门槛低，评价标准更灵活，属于通用型、开放性创造性任务。这一任务本质差异，与当前大语言模型的技术特性、语料库特征高度契合，是导致差异化影响的核心根源：现有大语言模型的通用语料库中，涵盖了海量日常生活场景

的创造性内容与观点，具备充足的日常创造性知识储备与发散观点支撑，因此高 AI 观点发散性能够充分调动通用语料资源，为日常创造性任务提供更多新颖创意，有效提升任务新颖性；但针对科学创造性任务，大语言模型缺乏垂直领域的专业语料投喂，专业科学知识的深度、精准度不足，难以生成符合专业规范、贴合科学研究需求的高质量发散观点，无法对科学创造性任务的新颖性产生正向促进作用。

回应：

感谢您细致且有启发性的意见！我们在修订稿中进一步强化了 AI 观点发散性对科学创造性任务和日常创造性任务产生差异化影响的机制解释。基于团体动机性信息加工模型(MIP-G)，我们提出“任务约束与 AI 信息适配性”的解释。具体而言，MIP-G 理论强调，外部信息能否促进团体创造，关键不在于信息数量或多样性本身，而在于团队能否围绕具体任务目标对多样信息进行开放、系统和精细加工。因此，AI 观点发散性的作用需要同时考虑任务约束结构和 AI 所能提供的信息类型是否匹配。

在日常创造性任务中，问题空间较为开放，专业知识门槛较低，评价标准也相对灵活。此类任务更贴近日常生活经验、社会情境和常见问题解决场景，与通用大语言模型所擅长的广泛语义联想、情境扩展和多样化表达更为匹配。因此，高 AI 观点发散性能够为团队提供更多远距离联想和类别扩展线索，团队也更容易将 AI 生成的多样化观点转化为新颖创意。相反，科学创造性任务通常具有更强的专业知识约束、更明确的评价标准和更高的可行性要求。团队不仅需要生成新想法，还需要判断这些想法是否符合材料属性、功能限制、科学原理和应用条件。已有 LLM 研究也提示，通用大语言模型在领域特定任务中可能受到专业知识不足、知识更新限制或事实可靠性不足的影响，科学创意生成能力也可能需要专门的评价基准和训练策略。因此，在科学任务中，高 AI 观点发散性虽然扩大了信息搜索空间，但这些观点未必能够与科学任务的知识约束和可行性标准充分匹配，反而可能增加团队筛选、验证和收敛整合的负担，从而难以转化为显著的新颖性收益，并可能削弱实用性。

同时，我们在表述上保持谨慎。由于本研究并未直接操纵大语言模型的训练语料、领域知识库或专业微调方式，因此我们没有将“通用语料与垂直领域语料的差异”表述为本文已经直接证明的因果机制，而是将其作为与既有 LLM 研究一致的可能解释和未来研究方向。未来研究可进一步比较通用 LLM、领域知识增强 LLM 和专家反馈增强 LLM 在科学创造性任务中的作用，以直接检验 AI 知识来源与任务约束匹配程度对人智共创表现的影响。通过上述修改，修订稿进一步从任务约束、信息适配和团队信息加工三个层面解释了 AI 观点发散性差异化影响的内在机制，也为不同任务情境下的 AI 协作设计提供了更明确的理论依据。

正文修改内容:

(1)AI观点发散性对团体创造的作用也需放在具体任务情境中理解……AI观点发散性对团体创造表现的影响可能并非单向促进，而是会在不同任务约束下对新颖性与实用性产生差异化作用。详见第5页末段。

(2)这一结果可以从任务约束与AI信息适配性的角度进行理解……关键在于团队能否根据任务约束将AI生成的多样化观点有效转化为兼具新颖性与实用性的创意整合。详见第21页第2段。

(3)从实践层面看，本研究提示，人智共创中的AI协作设计不应以单纯追求高发散性为导向。在高开放性、弱评价约束的日常创意任务中，高发散AI观点可充当激发新颖想法的外部资源。而在科学创新、技术设计和专业方案生成等高约束性任务中，仅提高AI观点发散性可能并不足以促进创造表现，设计者还需要结合领域知识库、检索增强、专家反馈和可行性约束提示，以提高AI输出与专业任务要求之间的匹配程度。也即，人智共创设计的关键未必在于持续提高AI观点的发散程度，而在于根据任务类型、任务阶段和任务约束提供更适配的AI信息支持(Vinchon et al., 2023; Chan et al., 2025)。详见第24页第2段。

参考文献:

Lu, R., Lin, C., & Tsao, H. (2024). Empowering large language models to leverage domain-specific knowledge in E-learning. *Applied Sciences*, *14*(12), 5264. https://doi.org/10.3390/app14125264

Ruan, K., Wang, X., Hong, J., Wang, P., Liu, Y., & Sun, H. (2026). Evaluating LLMs' divergent thinking capabilities for scientific idea generation with minimal context. *Nature Communications, 17*, 3625. https://doi.org/10.1038/s41467-026-70245-1

第二轮

审稿人1意见：

作者很好的回复了提出的问题，论文质量得到了较大的提升。建议题目可以再斟酌一下，是否可以换成AI观点发散性对群体创造性生成的影响及其认知神经机制更通顺一些。

回应：

非常感谢专家对本文修改工作的肯定和对题目表述的建设性建议！我们认真考虑了您的意见，并认同原题目中"AI 观点发散性影响团体创造表现的……"这一表达略显紧凑。根据您的建议，我们对题目进行了进一步斟酌和修改，以增强题目的通顺性和可读性。

具体而言，我们采纳了专家建议中“AI 观点发散性对……的影响及其……”这一表述方式，以更清楚地呈现本文的研究对象和解释重点。同时，考虑到本文的核心因变量为团体创意的新颖性和实用性，主要反映团队最终创造产出的表现水平，因此我们仍使用“创造表现”而非“创造性生成”。相较而言，“创造性生成”更强调创意产生过程，而“创造表现”能够同时涵盖本文所关注的创造性产出结果(新颖性、实用性)和过程，更符合本文因变量设置和结果呈现方式。

此外，考虑到本文以双人小组为分析单位，且全文理论框架主要基于团体动机性信息加工模型展开，正文中也持续使用“团体创造”、“团体创造表现”等术语，因此我们将专家建议中的“群体”调整为“团体”。相较于“群体”，“团体”更能体现本文研究对象具有共同任务目标、互动过程和协作生成特征，也有助于保持题目、理论框架和正文术语的一致性。

关于“认知神经机制”的表述，我们也进行了慎重考虑。本文整合了 AI 调用策略、观点采择策略、观点语义特征以及 fNIRS 超扫描指标，旨在从认知加工过程和神经活动关联两个层面解释 AI 观点发散性影响团体创造表现的可能路径。因此，在题目中使用“认知神经机制”可以更凝练地概括本文的研究重点。与此同时，我们也在正文讨论中继续保持谨慎表述，明确区分 AI 观点发散性实验操纵所支持的效应证据，以及脑间同步、单脑功能连接与行为变量之间所体现的关联性证据，避免将脑-行为整合结果解释为严格意义上的神经因果机制。

基于上述考虑，我们将题目修改为“AI 观点发散性对团体创造表现的影响及其认知神经机制”。

正文修改内容:

标题改为“AI 观点发散性对团体创造表现的影响及其认知神经机制”。

审稿人 2 意见：

作者较好回应上一轮提到的问题，并对论文进行了完善修改。没有进一步意见。

编委意见：

同意审稿人意见，建议录用

主编意见：

看过了，同意发表。