

第二轮修改

尊敬的主编、副主编和审稿人，

再次感谢您在百忙之中审阅本文，我们已经根据各位的意见对文章做了进一步修改，修改内容用蓝色字体标记，希望修改稿能够得到您的肯定。

外审专家意见

作者对第一轮意见做出了认真详尽的回应，论文经过作者的修改，对论文核心观点、逻辑框架等方面有了更清晰的阐述。尽管修改稿已有显著提升，但仍存在若干需进一步完善之处：

意见 1

文章题为《心理测量和人工智能的相互促进与协同发展》，但全文篇幅主要聚焦于“相互促进”，对“协同发展”的内涵、机制与实现路径着墨较少，仅在第 4.3 节进行了简要讨论。此外，尽管摘要明确将“人智共生”作为理论框架，但在后续各部分中较少回扣这一核心理念。建议作者在各小节结尾适当增加 1-2 句总结性语句，阐明该部分内容如何服务于“人智共生”或“协同发展”的整体目标，以增强全文的理论统一性与主线贯穿性。

回应：

感谢您的建议，考虑到字数限制，我们在 2.1-2.4 和 3.1-3.4 节后略补充了相关内容，以增强全文的统一性。

意见 2

文中多次使用“AI 心理特质”这一表述，但正如脚注 1 所指出的，该概念在学界尚存争议。为避免读者误认为作者主张 AI 具备主观体验或意识，建议在引言或第 2.1 节开头更明确地说明：此处“心理特质”仅为一种功能类比，也可考虑采用“类心理特质”等更为谨慎的术语，以体现学术表述的严谨性。

回应：

感谢您的建议，修改稿中我们已经将“AI 心理特质”替换为“AI 类心理特质”。

意见 3

图 1 和图 2 虽然具启发性，但目前的表述形式更多是接近文字表格，是否可以整合重绘为类似概念关系图的形式，以更直观展现双向赋能机制和协同发展。

回应：

感谢审稿人的重要建议。我们认同原图在呈现形式上更接近于结构化罗列，直观性和整体性仍有提升空间。经过思考，我们仍保留图 1 用于概括全文整体框架，以便读者快速了解；同时对图 2 进行了重绘，将其由原先的列表式呈现调整为多层次的概念关系图。

重绘后的图 2 以“理论层—工具层—价值层”为核心结构，从心理测量与人工智能两个方向出发，系统呈现二者在构念建构、方法工具与规范价值等层面的双向赋能关系，以突出其协同发展的整体逻辑。该图旨在作为一个概念性整合框架，用于直观展示主要关联维度及其相互支撑关系，而非预设严格的因果或过程模型。相关机制性细节与具体路径已在正文相应章节中结合文献进行了展开说明。

我们希望这一调整能够在不增加图形复杂度的前提下，更清晰地回应“双向赋能”与“协同发展”的核心主张，提升图示的可读性与概念表达效果。

需要强调的是，心理测量与 AI 之间并非单向赋能关系，而是在理论、工具与价值等多个层面上相互作用、协同演进的过程。如图 2 所示，这种协同关系可从三个相互关联的层面加以概括：在理论层面，心理测量所强调的信度、效度、公平性与构念效度，为 AI 能力评估与模型比较提供了必要的理论约束，而 AI 的数据驱动与多模态分析能力，也为心理测量中的构念挖掘与结构验证提供了新的技术路径；在工具层面，AI 推动了题目生成、复杂行为评分与自动化分析

的发展，显著提升了心理测量的效率与规模化水平，而心理测量在拟测特质界定、任务设计与结果解释方面的理论积累，则为相关 AI 工具的开发与应用提供了清晰的问题导向与质量标准；在价值层面，随着预测模型与虚拟参与者等 AI 应用不断进入心理测量与人智交互研究情境，心理测量所秉持的科学性、人文关怀与伦理规范愈发重要，有助于识别算法偏差、控制潜在风险，并为 AI 的负责任应用提供规范框架。

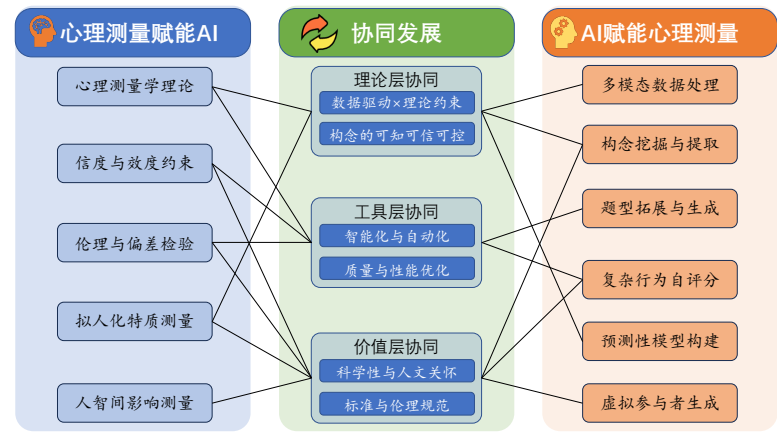


图 2. 心理测量与人工智能协同发展的多层概念框架

意见 4

对 AI 模拟参与者（AI-as-Participant）的潜在风险论述中，建议可简单强调其可能放大训练数据中的隐性偏见，进而污染测量构念效度。

回应：

感谢您的建议，修改稿已经添加相关内容。

研究者开始利用 AI 模拟参与者（AI-as-Participant）生成个体或群体行为数据，以补充被试样本量或进行预测验（Gao et al., 2024; Park et al., 2023），但可能存在分布失配问题并进一步放大训练数据中的隐性偏见，污染测量构念效度（Gallegos et al., 2024）；因此，AI 生成数据目前更适合前期探索，而非替代真实样本。

第一轮修改

尊敬的主编、副主编和审稿人，

感谢您在百忙之中抽空审阅本文，我们已经基于各位的宝贵建议对文章进行修改，修改内容用红色字体标记，希望修改稿能够得到您的肯定。当然，由于《心理科学》的字数限制（目前已经超限），我们难以保证修改稿的所有内容（含原文、修改和新增内容）都会在后续版本中得以保留，还请您理解。下面是对各位审稿意见的逐一回复。

副主编意见

请作者重点解决审稿人所提出的论文写作较为 AI 化，如概念堆砌额，引用错误，问题意识及核心观点缺失等问题。同时，邀请中级职称以上的同行进行挑剔性阅读，以提升文章的文字质量。

回应：

首先，感谢副主编对我们论文的宝贵意见。我们非常重视审稿人及您提出的反馈，特别是在论文写作风格、概念表达、引用准确性、问题意识及核心观点的呈现等方面。根据这些意见，我们对论文进行了全面的修改，以确保更高质量的文字表达、更加清晰的逻辑结构，并更加精准地表达我们的研究目标和核心观点。

具体改进措施如下：

1. 明确“人工智能”“心理测量”等概念，避免读者的有偏理解。
2. 修正引用错误，补充缺失引用问题。
3. 强化核心观点：在论文的引言部分，我们进一步明确了研究问题和核心观点，并确保全文围绕这一核心展开。我们清晰阐述了心理测量与 AI 相互促进的双向关系，并明确提出了人智共生视角作为本研究的理论框架。在每一章的开头和结尾，我们都有简短的总结，以帮助读者更好地理解研究的逻辑结构和关键论点。
4. 邀请同行进行挑剔性阅读，在严格的字数限制下适当增加了可读性。

我们相信这些修改将大大提升论文的质量，并且更加符合学术规范。再次感谢副主编和审稿人对我们工作的指导与建议，我们期待您的进一步反馈。

审稿人 1 意见

本文选题具有重要的前沿价值，系统梳理了心理测量与人工智能两个领域的交叉与融合，文献覆盖面较广，为读者理解该新兴领域提供了有益视角。然而，稿件在若干关键方面仍有提升空间。文章的核心论证存在一定程度的逻辑跳跃，部分关键概念的界定与分类尚不够全面和清晰，整体结构划分也缺乏明确的理论依据。此外，文中部分细节与格式尚需进一步完善。综合考虑，建议对稿件内容进行重新梳理，并补强相关逻辑链条。相信作者在根据审稿意见进行细致修改后，稿件质量将得到显著提升。

回应：

感谢您对本文的肯定，我们已经根据您的建议对修改稿做出相应修改，希望您能满意。

意见 1

建议在引言第二段“研究揭示出 AI 心理特质的三大特征：其一，任务敏感性，AI 在部分任务上的表现已能超越人类，但在测量相同特质的任务迁移与泛化方面仍存在不足；其二，资源扩展性，AI 心理特质水平提升高度依赖于训练数据规模、模型参数规模和硬件算力等资源的扩展，表现出与人类截然不同的发展路径；其三，提示敏感性，AI 在相同任务中的表现易受提示工程（prompt engineering）影响，进而导致其心理特质水平/类型不稳定性较高。”中论述中适当增加参考文献支持，以增强论据的权威性与说服力。

回应：

感谢您的建议，修改稿中我们补充了相关支撑文献。修改内容如下：

研究揭示出 AI 心理特质的三大特征：其一，任务敏感性，AI 在部分任务上的表现已能超越人类，但在测量相同特质的任务迁移与泛化方面仍存在不足（Bommasani et al., 2021）；其二，资源扩展性，AI 心理特质水平提升高度依赖于训练数据规模、模型参数规模和硬件算力等资源的扩展，表现出与人类截然不同的发展路径（Hoffmann et al., 2022）；其三，提示敏感性，AI 在相同任务中的表现易受提示工程（prompt engineering）影响，进而导致其心理特质水平/类型不稳定性较高（Wang et al., 2024; Zhu et al., 2024）。

意见 2

第 2.1 节 AI 心理特质的测量“目前主要采用两类方法：一是直接应用人类心理特质量表或任务，二是使用大规模基准测试（benchmarks）。”目前不止两类方法，总结不够全面，建议补充。其次是大规模基准测试主要目的是测量大模型的通用能力、专业能力等，写在这里有点奇怪，建议在大规模基准测试这段进一步解释清楚。

回应：

感谢您的建议。由于篇幅限制，我们无法将量表和任务两种方式分别进行阐述，而是从已有测量工具的直接应用视角将两者合并精简阐述。这无疑对审稿人和读者带来了一定困惑。修

改稿中我们微调了表述，比如：

目前主要采用两类方法：一是直接应用成熟的人类心理特质量表或任务，二是使用大规模基准测试（benchmarks）。

另外，我们在基准测试段落中做了补充说明：

基准测试作为系统化评估框架，通过大规模、多维度任务集（如 MMLU、BIG-bench 和 GSM8K）考察 AI 智能体的通用能力。

意见 3

第 2.1 节涉及人格测量进展部分，建议补充人格迫选范式在大模型人格测量中的相关研究，例如文献《Decoding LLM Personality Measurement: Forced-Choice vs. Likert》（Li et al., 2025），以体现心理测量与人工智能的双向融合。

回应：

感谢您的建议，已经补充该重要文献。

意见 4

文章在划分“心理测量赋能 AI”与“AI 赋能心理测量”四个子方向时，建议进一步说明其分类的理论依据或逻辑框架，以增强结构合理性。

回应：

感谢您的建议，修改稿中我们补充了相关说明。修改内容如下：

心理测量赋能 AI 不仅关乎对 AI 本身的评估，更在于深入理解人智系统的整体生态，包括 AI 的内在特质（AI 的特质水平/类型）、人类对 AI 的主观认知（人类如何看待 AI）、AI 的外部影响（AI 如何影响人类）以及双方的协同机制（人类与 AI 如何共存与互动）。这一逻辑链条不仅关注 AI 评估，还强调人智系统的整体效能与伦理风险，符合我国“AI+”战略中“以人为本”的核心导向，契合人智共生时代主题（李政涛，2025）。此外，部分研究探讨了 AI 在特定议题下对心理测量的赋能，如题目自动生成（韩雨婷等，2025）和自动评分（徐静等，2024），但未能全面涵盖心理测量的四个核心要素，无法充分展示 AI 对心理测量赋能的潜力。综上所述，现有研究多从单一或局部视角探讨心理测量如何赋能 AI 或 AI 如何赋能心理测量，未能深入探讨二者在理论与方法层面的相互促进机制。这一局限性妨碍了从人智共生视角全面理解 AI 的发展现状与应用价值，也限制了两领域的协同创新。

意见 5

作为综述性论文，建议在引言与结论部分更清晰地阐明本文相较于已有综述的独特理论贡献或视角，突出其学术价值。

回应：

感谢您的建议。实际上，原文中我们已经在引言中阐述了本文的独特性，已有综述要么从更宏观的视角探讨心理学与 AI 的交叉，要么从更微观的角度探讨了 AI 如何在某个特定主题下赋能心理测量（比如，题目自动生成）；另外，尽管有一些研究聚焦于对 AI 心理特质的测评，但忽视了对人智双向影响的系统综述，未契合人智共生时代主题。修改稿中我们略微补充了相关内容（详见对意见 4 的回复）。

意见 6

文中部分参考文献格式不符合 APA 规范，例如：“尽管已有研究探讨心理科学与 AI 的交叉（吴燕、周晓林，2025；朱廷劭，2025；Ke et al., 2025），但多泛化而未聚焦心理测量。”请作者系统检查全文格式并统一修正。

回应：

感谢您的建议，已经修订。

审稿人 2 意见

该论文选题具有较强的现实意义和前沿性，试图探讨心理测量与人工智能（尤其是大语言模型）之间的双向影响，具有一定的理论与应用潜力。作者收集了大量文献，研究材料丰富，显示出对该领域的良好追踪。然而，从目前的论文写作看，论文在某些关键概念、逻辑结构、以及论证深度方面存在一定的不足。

回应：

感谢您对本文的肯定，我们已经根据您的建议对修改稿做出相应修改，希望您能满意。

意见 1

论文核心概念未得到清晰的界定和申明，如“人工智能”在论文中可能多指大语言模型及与其相关智能体，但是论文多处出现不同的模型或 AI 的泛化概念相关研究的讨论，2.4 中的（de Visser et al., 2016），以及 3.4 中（潘晓珂，胡传鹏，2025）。类似地，心理测量、人智交互、计算/智能心理测量学、范式等核心的议题及其逻辑关系，均没有相应的定义或说明。

回应：

感谢您的宝贵建议。正如您所说，本文对“人工智能”的定位并非局限于目前最有代表性的基于大语言模型的生成式人工智能（实际上，这点在原文第一句话就通过“尤其”一词体现了）。鉴于目前大多数读者对“人工智能”的理解局限于生成式 AI，我们在修改稿中通过脚注方式进行了补充说明，以免增加篇幅：

本文中，‘人工智能’一词不仅限于生成式人工智能（如大语言模型），而是指涵盖机器学习、计算机视觉、语音识别、自动化决策等多个子领域的广泛技术体系。

意见 2

论文多处出现了箭头（→），这种在学术写作中不被常用。

回应：

感谢您的宝贵建议，修改稿中我们删除了箭头，修改了表述：如将“问题 → 答案”修改为“针对特定问题作答”，将“心理特质 → 外显行为”修改为“心理特质影响外显行为”。

意见 3

论文多处引用与论点不匹配，如 2.4，“这种分工随交互经验积累趋于稳定，用户通过动态校准对 AI 的信任，形成高效低冲突的协作模式（de Visser et al., 2016）”，（de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3), 331–349）这个研究很难证明“经验积累会稳定合作关系”；AI 素养问卷（Ng et al., 2024）被用来讨论“社会接受度”（2.2 第一段话）

回应：

感谢审稿人对我们论文的细致审阅和宝贵意见。针对该问题，我们进行了认真检查和修改。具体而言：

1. 关于 de Visser et al. (2016) 的引用：审稿人提到，de Visser et al. (2016) 的研究与我们提出的“经验积累会稳定合作关系”这一论点不完全匹配。我们仔细核查了该研究的内容，发现该研究主要探讨了拟人化如何增强认知代理的信任恢复能力，而并非直接研究经验积累如何影响合作稳定性。因此，我们已修改相关部分，将该文献替换为更为合适的引用，并进一步明确交互经验如何影响信任和合作稳定性。修改后的表述：

随着交互经验积累，用户会根据 AI 的表现不断校准信任，从而形成更稳定的协作关系（Hoff & Bashir, 2015；Lee & See, 2004）。

相关文献：

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407-43

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.

2. 关于 AI 素养问卷（Ng et al., 2024）的引用：在 2.2 节第一段中，我们引用了 Ng et al. (2024)的研究旨在支撑“心理测量为此[人类对 AI 的人智、情感与态度]]提供了科学工具”，可能是因为表述问题导致了误解。修改稿中我们直接删除了这个参考文献，以避免误解。

修改稿中，我们已经针对类似问题进行了修正和调整，确保引用的文献与论文的论点和内容更加匹配，以增强论文的严谨性和逻辑性。再次感谢您对这些细节的指出，我们相信这些修改将有助于提升论文的质量。

意见 4

论文缺少核心观点和理论主线，导致大量的列举，比如随意调换各个案例的关系，可能也不会影响。但是这样很难让读者抓到核心的研究问题及逻辑主线。

回应：

感谢您的宝贵建议。结合审稿人 1 的建议，修改稿引言部分我们补充了一些内容以期进一步体现本文的核心观点和逻辑主线：

目前，尽管已有研究探讨心理科学与 AI 的交叉融合（吴燕，周晓林，2025；朱廷劭，2025；Ke et al., 2025），但多泛化而未聚焦于心理测量。部分综述虽涉及 AI 心理特质的测评（OECD, 2025；Ni et al., 2025；Ye et al., 2025），但未能充分探讨人类与 AI 之间的相互影响。心理测量赋能 AI 不仅关乎对 AI 本身的评估，更在于深入理解人智系统的整体生态，包括 AI 的内在特质（AI 的特质水平/类型）、人类对 AI 的主观认知（人类如何看待 AI）、AI 的外部影响（AI 如何影响人类）以及双方的协同机制（人类与 AI 如何共存与互动）。这一逻辑链条不仅关注 AI 评估，还强调人智系统的整体效能与伦理风险，符合我国“AI+”战略中“以人为本”的核心导向，契合人智共生时代的主题（李政涛，2025）。此外，部分研究探讨了 AI 在特定议题下对心理测量的赋能，如题目自动生成（韩雨婷等，2025）和自动评分（徐静等，2024），但未能全面涵盖心理测量的四个核心要素，无法充分展示 AI 对心理测量赋能的潜力。综上所述，现有研究多从单一或局部视角探讨心理测量如何赋能 AI 或 AI 如何赋能心理测量，未能深入探讨二者在理论与方法层面的相互促进机制。这一局限性妨碍了从人智共生视角全面理解 AI 的发展现状与应用价值，也限制了两领域的协同创新。

由于《心理科学》的字数限制，我们不得不精简掉了许多转折词和上下文逻辑连接语句，导致文字比较生硬，有种罗列文献的感觉。同时，一些综述性论文由于涉及人与 AI 关系的多个视角，可以支撑本文的多个观点，因此会被重复引用（也有利于减少参考文献数量和全文字数）。另外，本文的重点不是在综，而是在后面的述（第 4 节），所以也不得不进一步压缩综部分（第 2、3 节）的字数。修改稿中我们适当增加了文字可读性，望您能理解。

意见 5

在摘要的第一句话，“心理测量是人工智能...关键手段，...重要途径”；“AI 则为心理测量...个性化、隐性化、智能化与多模态革新”。心理测量研究人类可能会更符合常识，但是心理

测量作为 AI 理解的手段和途径，以及 AI 作为个性化，多模态等革新，可能需要一定的说明，甚至论证后才能被理解。而作为研究的核心背景会增加论文的理解难度以及破坏读者的预期。其后还出现很多不太好理解的概念“协同演化”、“测量新范式”等。

回应：

感谢您的宝贵意见。修改稿中我们对摘要进行了调整，以增加可读性：

心理测量通过评估人工智能（Artificial Intelligence, AI）的智能特征和发展趋势，成为理解 AI 及其与人类之间相互作用的关键手段。AI 则为心理测量提供关键技术支持，推动其在个性化和智能化等方面的创新。本文从人智共生的视角出发，系统梳理心理测量赋能 AI 与 AI 赋能心理测量两条路径的研究进展，并总结当前面临的挑战与应对策略。总体而言，心理测量为 AI 提供了标准化的评估框架与理论约束，有助于厘清其智能特征、边界与潜在风险；而 AI 通过方法与技术的创新，促进心理测量工具的开发、构念的构建与理论的发展。两者并非孤立发展，而是相互促进，共同推动理论与实践的创新。二者的深度融合有望在理论驱动与数据驱动并行的双轨下，催生出人智共生时代的科学测量范式。

意见 6

2.2, 2.3, 2.4 小结很难理解与心理测量的关系，以及赋能了“什么”人工智能的“什么”？

回应：

感谢您的意见。需要说明的是心理测量（psychological measurement）既不是心理测量学（psychometrics, 强调测量理论）也不是心理测验（psychological assessment, 强调测量工具）；任何涉及使用工具（包括量表和实验任务）进行数据采集的量化研究均属于心理测量范畴。因此，我们不能把心理测量等同于使用问卷量表。这点从第 3 节基于心理测量定义四个关键部件，而非聚焦于测量工具，来阐述 AI 赋能心理测量中也能体现出来。修改稿引言中，我们做了如下补充说明：

心理测量赋能 AI，不仅在于对 AI 本身的评估，更在于深入理解人智系统的整体生态。这包括 AI 的内在特质（即 AI 的特质水平与类型）、人类的主观反应（人类如何看待 AI）、AI 的外部影响（AI 如何影响人类），以及双方的协同机制（人与 AI 如何共存与互动）。这一逻辑链条涵盖了从个体到互动、从单向到双向的完整谱系，契合人智共生时代的主题（李政涛，2025）。同时，这一视角不仅关注 AI 的评估，更强调人智系统的整体效能与伦理风险，体现了国家“AI+”战略中“以人为本”的核心导向。

另外，在引言中，我们通过脚注强调了心理测量的定义及其包含的四个核心要素，这有助于读者理解本文的撰写逻辑：

“心理测量”是指基于心理测量理论，使用心理测量工具对心理特质进行定量描述的过程；涉及心理测量理论、心理测量工具、心理特质和定量描述四个核心要素。

意见 7

3.2, 3.3.3.4 缺少核心的问题导致很难理解，“为什么”需要 AI 的赋能以及 AI 的赋能后的重要性在“哪里”。

回应：

感谢您的问题。该问题与上一问题本质都是对本文主题的有偏理解导致的，本文关注的不是使用心理测量工具对 AI 的心理特质进行测量（2.1 节），或使用 AI 提升心理测量工具开发的智能化（3.1 节），而是从相对更宏观的视角探讨心理测量（包括测量理论、测量工具、心理特质与定量描述）与 AI（包括 AI 本身和人类与 AI 的双向影响）之间的相互促进、协同

发展。希望我们在修改稿引言部分的补充内容能够更清楚地使读者理解本文的章节逻辑（见对意见 6 的回复）。