

# 《心理科学进展》审稿意见与作者回应

题目：当 AI 学习共情：心理学视角下共情计算的主题、场景与优化

作者：侯悍超，倪士光，林书亚，王蒲生等

---

## 第一轮

### 审稿人 1 意见：

这篇综述介绍了以 GPT-4 为代表的超级人工智能技术对共情研究带来的新机遇和挑战。

共情计算作为一个新交叉领域，其研究不仅可以测量一个心理概念，还可以为人类探索未知生产新知识、新方法和新应用。这一交叉领域的发展不仅有助于拓展和加深对共情心理机制的理解，还可以将研究成果运用于智能社会的发展，提升个人和社会的幸福。因此，共情计算领域值得心理学研究者关注，并在心理理论、可解释性数据集和应用实践方案等方面做出贡献。

作者对主题从内容选取、结构框架、对当前研究内容的总结和未来洞见都进行了良好地组织，是一篇有洞见的综述论文。

仅有几个意见供作者一起思考：

意见 1：目前这一领域有知名的 empathy computing lab，但是文章中对来自该 lab 的工作都未多引用，请说明原因或者增加引用。

例如：

Gumilar, I., Sareen, E., Bell, R., Stone, A., Hayati, A., Mao, J., ... & Billinghamurst, M. (2021). A comparative study on inter-brain synchrony in real and virtual environments using hyperscanning. *Computers & Graphics*, 94, 62-75.

Hart, J. D., Piumsomboon, T., Lawrence, L., Lee, G. A., Smith, R. T., & Billinghamurst, M. (2018, October). Emotion Sharing and Augmentation in Cooperative Virtual Reality Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 453-460). ACM.

回应：非常感谢评审专家对本文的肯定和建议！我们有注意到这个实验室的工作，不过该实验的多聚焦于技术本身，其所发表的论文中“共情”一词出现的频率不高，因此我们未给予足够的重视。例如，在 Hart 等人（2018）的全文中，“empath\*”仅在参考文献中出现一次，加上篇幅的限制，我们未能在此前的稿件中引用该实验室的研究。这是我们的疏漏，感谢指正。

考虑到该实验室在共情计算领域的工作和贡献，综合研究代表性和本文篇幅后，我们在修改版文稿中引用了一项该实验室 Jing 等人（2022）的工作。

Jing, A., Gupta, K., McDade, J., Lee, G. A., & Billinghamurst, M. (2022). Comparing Gaze-Supported Modalities with Empathic Mixed Reality Interfaces in Remote Collaboration. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 837-846). IEEE.

Jing et al. (2022) 是该实验室最近发表的文章，探讨了混合现实技术如何增强人类共情，是该实验室代表成果之一，并且与本文的主题高度相关。通过引用这篇文章，可以为读者了

解共情计算领域的前沿发展提供参考，起到抛砖引玉的效果。具体修改内容在 11 页，内容如下（以下修改部分我们用绿色标出）：

此外，一些新兴的研究也正在涌现，如通过混合现实（Mixed Reality, MR）促进人与人之间的共情(Jing et al., 2022)，以及脑机接口提升计算机共情表现(Roshdy et al., 2021) 等方面的初步探索。

**意见 2：** 目前作者引用了 chatGPT 相关预印论文，在自检报告和正文却未曾引用与主题相关的预印综述，请自检报告说明原因或者增加引用。

Cao, S., Fu, D., Yang, X., Wermter, S., Liu, X., & Wu, H. (2021). Can AI detect pain and express pain empathy? A review from emotion recognition and a human-centered AI perspective. arXiv. <https://doi.org/10.48550/arXiv.2110.04249>

**回应：**非常感谢专家指出的问题。在 Cao 等人(2021)的工作中，他们使用了‘artificial empathy’来称呼共情计算，这与本文主题高度相关。因此，在修改版的 1.2 共情计算概念小节中，我们引用了 Cao (2021) 的这一称呼，以便为读者提供更全面的背景和相关研究。

此前未引用的说明如下：如您所注意到的，本文初稿中引用了 4 篇预印本论文，这其实是尽可能删减预印本后的结果。作者过去认为期刊稿件并不鼓励引用预印本，因此只引用了 2022-2023 年最新发布的非常必要的预印本，因为这些工作代表了该领域的前沿，但尚未有足够的时间正式发表为期刊或会议论文。由于这个思维误区，我们遗漏了 Cao 等人 (2021) 稍早的重要工作。感谢审稿专家的提示。

通过专家的意见，我们进一步认识到在共情计算这一迅速发展的领域，引用预印本可以帮助读者更好地把握前沿动态。因此，在修改版中，我们增加了一些最新的、具有代表性的、高质量研究的预印本。这些引用主要集中在 5.2 节中，用以展现心理学家在这一领域可能的贡献。具体内容参加审稿人 2 意见 3。

**意见 3：** 对于文章的摘要部分

“共情计算是指使用人工智能和大数据技术来预测、识别、模拟和生成人类的共情，是一个心理学与计算机科学交叉的新兴研究领域。本研究构建了一个数据层、模型层与任务层的普适性研究框架，总结了一个包括个体共情测评、共情内容分类、共情回应系统和共情对话生成的四个新主题分析框架，建立了面向心理健康、教育学习、商业服务和公共管理等心理应用的场景创新。未来研究有必要建构高整合的共情计算理论模型、建设高可信的共情心理行为特征数据集、并通过以人为中心的评价体系验证并改进共情计算的研究效率。共情计算有益于将当前人际共情研究扩展到智能社会新型人-AI 关系研究中，心理学为这个新兴领域提供了理论基础和研究方法的支持。”

对综述内容的概括还有待加强，实际上内容是传统人类共情研究概述，到共情计算的提出总结，再到框架和应用的展望。

另外，摘要最后一句也“言之无物”，作者要组织的内容应为：计算机科学为这一领域提供哪些支持，心理学为这个新兴领域提供哪些理论和方法的支持。因为严格来说，计算机科学也能为这个新兴领域提供理论和方法的支持。

**回应：**感谢审稿专家指出摘要中的问题并提供修改建议。对两部分的建议回复分别如下。

其一，专家指出摘要的内容应包括“传统人类共情研究概述，到共情计算的提出总结，再到框架和应用的展望”，我们深表认同。本文确实在第一版的稿件中未将“传统人类共情的研究概述”的内容充分体现在摘要中。考虑到这部分并非本文的核心重点，而是作为理论背景的一部分，我们决定将其凝练在第一句对共情计算的描述中。

因此，我们将摘要第一句后半部分进行了相应的修改，从共情计算“是一个心理学与计算机科学交叉的新兴研究领域”扩展为“是传统心理学共情研究与计算机科学交叉的新兴研究领域。”这样修改体现了共情计算与传统人类共情研究之间的联系，同时也保持内容简洁。

第二，有关摘要的最后一句，审稿意见指出需要更具体明确地描述心理学在共情计算领域的作用和价值，同时描述计算机科学的作用。感谢专家的宝贵意见。

为此，我们在摘要处进一步细化了心理学家在共情计算领域引领、评估和改进的重要角色，并在文章 5.2 小节详细阐述了心理学如何为共情计算界定问题、提供理论支撑和评估方法（详见审稿人 2 意见 3，和文稿 5.2 小节）。考虑到本文长度的限制，我们更多聚焦心理学视角，对计算机科学的贡献仅做必要的说明。

综合以上两个建议修改摘要如下（英文摘要的相应修改见正文文稿）：

共情计算是指使用人工智能和大数据技术来预测、识别、模拟和生成人类的共情，是传统心理学共情研究与计算机科学交叉的新兴研究领域。本研究构建了一个数据层、模型层与任务层的普适性研究框架，总结了一个包括个体共情测评、共情内容分类、共情回应系统和共情对话生成的四个新主题分析框架，建立了面向心理健康、教育学习、商业服务和公共管理等心理应用的场景创新。未来研究有必要建构高整合的共情计算理论模型、建设高可信的共情心理行为特征数据集、并通过以人为中心的评价体系验证并改进共情计算的研究效率。共情计算有益于将当前人际共情研究扩展到智能社会新型人-AI 关系研究中。心理学家在该领域承担引领、评估和改进等角色，并与计算机科学家紧密合作，共同推动共情计算理论基础更坚实，效果检验更可靠，应用创新更落地。

意见 4：关于文章的场景，选择的“心理健康、教育学习、商业服务和公共管理”四个领域或者场景，这些选择有何依据？彼此间似乎有共性或者重叠关系。

回应：感谢审稿专家的意见，这些问题让我们更深入地阐述选择这四个领域的依据和它们之间的关系。我们选择这四个领域是因为他们具有代表性，是现阶段研究比较丰富的领域。这是由共情计算的领域特征决定的。共情计算本质上还是研究人与人、人与机器之间交互和沟通的领域，而心理健康、教育学习、商业服务和公共管理都涉及大量的人际交往、情感理解和回应。因此这些领域也是共情计算应用最多的场景。其他一些人工智能的常见应用场景，如金融、交通等并不具备这样的特点，因此不在选择范围内。

同时，由于这 4 个领域具备的共性，他们并非彼此完全独立。为此我们在修改版文稿中做了说明，同时强调这些并不是全部，提到了未来出现新应用场景的可能性。

具体修改如下，我们在第 3 节开始处增加了以下的说明内容：

尽管尚处于起步阶段，但现有研究已经展现出共情计算具有广泛的应用前景，特别是在心理健康、教育学习、商业服务和公共管理等领域积累了较多相关研究。这些领域涉及大量人际互动，在人工智能日益普及的背景下，也将成为人机交互最频繁的领域，因而是共情计算的主要应用场景。由于都涉及通过共情计算，提升人与人、人与机器互动过程中对情感的理解和回应能力，这些领域并非完全独立，而是相互关联却各有侧重的。随着技术的进步，未来还可能出现新的应用场景，如将共情计算应用于游戏提升娱乐体验等。

意见 5：共情计算与共情学习之间的关系如何，是否是相同的概念？如果不是，问题是否应围绕共情计算修改？

回应：感谢您的问题。在回顾全文后，我们注意到本文中并未使用“共情学习”一词。我们理解这个问题可能来源于“共情”与“学习”两个词在标题及第 7 页中并列使用所产生的疑惑。为此我们表示歉意，并澄清如下

本文使用“学习”一词主要有两个用法，第一个是用于“机器学习”等专有技术用语。第二个也就是造成误解的双关用法。一方面它仍然指向机器学习这一技术用语，另一方面，我们希望用“学习”来形象地描述计算机分析和模拟人类共情的过程。这样做是为了增加一点文章的可读性，使描述更加生动。为了避免混淆，我们在第7页使用“学习”一词时加上了双引号（原文：“在模型层，研究者建立共情的计算模型，使计算机可以通过算法‘学习’共情。”），以明确表示其比喻含义。而对于标题中的学习一词我们原本的目的是在冒号前使用双关语，起到引起读者兴趣的效果。考虑到这是一种常见的用法，因此标题中没有额外使用双引号表示双关，这样也可以让标题更简洁。

同时，我们也完全尊重编辑和审稿专家的意见。如果认为在标题中加上双引号更为合适，我们愿意根据您的建议进行修改。我们的首要目的是确保读者对文章内容的准确理解。

**意见 6:** 文中“6. 结语”部分实际与“5 展望和讨论”展望讨论部分逻辑重叠，建议 5 改为“5 讨论和展望”组织合并 6。

**回应:** 非常感谢您的建议。我们已根据您的建议修改了文稿，使得文章内容更紧凑。具体内容见正文相应部分。

.....

#### 审稿人 2 意见:

该论文从工程应用角度，提出了一个包含数据层、模型层与任务层三块内容的共情计算研究框架，总结了共情计算相关的四个主题研究，并展望了该领域的未来发展趋势以及心理学在其中的作用。

该论文聚焦当前研究热点，介绍系统全面，有助于读者全方位了解共情计算的当前发展阶段，具有一定的启发价值。作为一篇一般意义上的科研前沿普及综述，或者计算机领域的研究综述，该论文具有发表价值；然而，对于心理学期刊而言，其在心理学学术理论上的意义似乎不够明确。

**意见 1:** 首先，关于“共情”的概念，作者需要站在心理学立场上进行更精准地界定和辨析。有一些地方存在比较模糊地表述，容易误导作者。例如，p5 中正文第一段，开门见山给出了共情地心理学定义，强调需要理解他人情绪；然而，同一页提到 ChatGPT 等语言生成模型，使用了“表现出了基础的共情能力”、“回复更加共情”等表述。目前，尚无证据表明语言生成模型可以真正理解他人情绪，只是有能力输出看起来像在共情的语言，行为输出上表现类似并不等于有共情能力（以情绪理解为核心）。这些地方表述不谨慎，不作明确辨析，容易造成误解。类似的情况在全文还有不少（包括后面没有分析，共情计算提高心理咨询效果，究竟是因为有了更好的语言技巧，还是别的）。

**回应:** 感谢审稿专家的意见。我们使用了表现出共情“能力”来描述计算机，可能给读者留下人工智能已经具备与人一样的共情能力的印象，这是由于我们未能明确区分计算机的能力（capability）和人的能力（ability）造成的，非常感谢您指出这一问题。在计算机领域，计算机的能力（capability）是指通过软件、硬件等系统，执行特定任务或功能的能力。我们尝试检索了其他翻译方式，用以区别人类的能力，但在 CNKI 学术翻译中 85% 左右的文献将 capability 翻译作“能力”，因此我们保留这一翻译。为了说明区别，首次用到这个用法时，我们在正文中增加了说明性内容，指出 AI 表现出共情能力的研究依据——在情绪识别任务中达到普通人的平均水平(Kosinski, 2023; Wang et al., 2023)。同时，使用注脚进一步说明，

与人类的能力加以区分。相信这样的说明可以减少造成读者误解的可能性。修改内容如下：与过去冷冰冰的机器形象不同，最新研究表明，基于大语言模型的 AI 已经表现出一定的共情能力 (capability)<sup>1</sup>，可以在情绪识别任务中达到普通人的平均水平(Kosinski, 2023; Wang et al., 2023)。

<sup>1</sup> 这里以及后文中提及计算机和 AI 所具备的“能力”，是指计算系统执行特定任务或功能的能力，英文译作 capability。具体而言，文中计算机的“共情能力”，指计算机通过模拟人类语言和行为，输出与人类相似的共情特征的能力 (capability)。尽管同样使用了“能力”一词，但其内涵与人类具有的共情能力 (ability) 并不等同。

其次，“回复更加共情”部分修改为“回复表现出更高的共情水平”。

最后，审稿人提到文中没有分析共情计算提高心理咨询效果的起作用机制。有关这一问题，我们认为提高心理咨询效果的心理机制分析并非本文重点，所以并未直接针对这个问题做出修改。然而，鉴于专家在意见 3 中提到，在共情计算应用场景中应增加心理学角度分析，我们在 3.1 节增加了共情提升数字治疗联盟进而提升数字心理干预效果的内容：“研究者认为，增加数字干预系统表现出的共情水平，有助于促进形成数字治疗联盟 (digital therapeutic alliance)，进而提升干预效果(Tong et al., 2022)。”

意见 2：其次，作者给“共情计算”下了一个宽泛的定义 (p7)，使其能包括所有与共情相关的计算研究。太泛的定义带来的问题是，符合该定义的许多计算过程，是有本质差别的。类似于强、弱人工智能的区别，一些所谓“共情计算”，使用海量数据训练个神经网络，使其能对输入的内容在情绪维度上分分类，完成某个具体场景下的应用，这跟人的共情过程并无关系，无非是神经网络算法在共情领域的应用而已。这类研究，开展的过程完全无需借鉴心理学，其结果对于心理学也毫无参考价值。只有类人的共情计算模型的开发，才需要心理学作为研究基础，其开发过程中所发现的计算特性，才对于心理学的共情研究有启发价值。所以，混在一起谈，无益于揭示其语心理学的关系。同样的问题也发生在作者所构建的框架上。这是计算科学领域的常见研究框架图，罗列了所有可用的数据输入、计算模型，以及可以实现的任务和使用场景，但是缺乏信息量，既不明确用什么样的数据采用什么样的算法能实现何种任务，也不清楚心理学为何能成为研究基础以及如何促进大范围研究和应用。尤其是后者，毕竟这是一篇投给心理学期刊的综述。

回应：感谢审稿专家指出的问题。我们非常认同专家的观点，作为投稿给心理学期刊的文章，需要展现共情计算领域与心理学的紧密联系。具体来说，我们注意到这条意见中涉及以下问题，我们分别回复：

共情计算的宽泛定义。如专家指出，文中对共情计算使用了较为宽泛的定义，涉及不同的计算过程。我们这样做是为了给读者提供对这一领域的整体认知。在新兴领域中，研究的内涵和外延常在发展变化中，因此我们认为对其定义也必然存在一定的模糊性，并且也应接纳这种模糊性的存在。其次，在信息技术快速演变的当下，相较于计算过程，通过目的或任务来定义一个领域更为稳健。因此，我们在定义中关注计算机分析和模拟共情这一研究目标，而没有对具体计算过程做出界定。尽管在定义中我们没有区分不同的计算过程，本文第 2 节中对具体的技术手段以及不同手段间的区别与联系做了说明（见审稿人 3 意见 1）。共情计算的理论基础。我们认同专家指出的一些用大量数据训练神经网络的共情计算研究存在理论基础较弱的问题。因此，本文多次强调了心理学对于这一领域研究的引领和评价作用，如第 4 节和第 5.2 节。同时，我们也希望说明，本文引用的绝大多数研究都涉及心理学对于共情的定义或分类，并非没有借鉴心理学。有关心理学如何成为共情计算的研究基础，我们根据您的意见 3 做了修改，详细内容见意见 3。

对心理学的参考价值。专家指出某些基于神经网络算法应用的研究对心理学没有参考价值，我们也部分认同，因此我们尽力筛选了该领域内对心理学有参考价值的文献。另一方面，我们认为即使是侧重于技术、数据驱动的研究对心理学也具有一定借鉴意义。首先，自动化识别共情的方法为研究者提供了有价值的工具，使他们能够在更大规模的数据尺度上分析共情特征。其次，基于神经网络算法应用的数据驱动型研究，可能有助于发现当前研究方法未曾识别出的模式。有关这些内容，我们结合专家的意见 3，在 5.1 小结中做了修改，进一步阐明了共情计算在心理学中的参考价值，详见下文的意见 3。

概念框架的信息量。针对专家提到的框架图信息量不足的问题，我们认为框架图的目的是为了给读者提供一个整体概览，而详细内容则在正文中体现。鉴于本文旨在为心理学背景的读者介绍心理学与计算机交叉领域的内容，因此并未在文中详细展开技术细节的说明。尽管如此，文中还是包含了对个别代表性研究技术路径的简要说明，以期在不过度分散注意的前提下，为读者提供更多的信息。如 1.3 小节中提到：“一般做法是基于共情的心理学理论(例如 Sharma et al., 2020)或采用数据驱动的方式(例如 Rashkin et al., 2019)抽取数据中有关共情的特征（如表达共情的词汇、缓和的语调等），随后用这些特征训练计算模型，训练后的模型可用于预测新数据或生成新内容。”

意见 3：第三，论文在一些细节方面，也未能很好体现共情计算和心理学的关系。例如 p6-7 提到了一些神经研究，这些研究对于计算有何启示？P10-12 提到了很多共情计算的应用，但在心理学方面缺少深入的分析。P13-14 中 5.1、5.2 两节应是论文重点，因为这是真正跟心理学相关的内容，但是很遗憾，并未看到究竟有什么洞见（或者未来的研究可以有什么洞见），只谈了有什么新工具；心理学家究竟扮演什么角色、能起什么作用，也不明确，泛泛地讲心理学家可以提出问题、设计实验没有太大价值，这句话可以用在任何一个心理学交叉领域。

回应：感谢审稿专家指出的问题。我们在第一版文稿中未能充分展现共情计算与心理学的联系。为了更好地体现这两者之间的关系，我们根据您的建议做了如下修改：

1. 在 1.1 小节传统共情研究部分增加一段阐释传统共情研究对共情计算的基础和启发价值。

这些传统研究成果为共情计算的发展奠定基础并提供启示。共情概念和成分的研究为共情计算提供了理论框架，明确研究范畴。传统自评量表作为校标，可以评估共情计算的准确性和可靠性。人类共情的神经机制研究可能为设计脑机交互提供启发(Roshdy et al., 2021)。以往的应用研究也为共情计算潜在应用场景指明方向

2. 在第 3 节应用场景部分，我们的目的主要是展示共情计算可能的应用领域，并非分析其中的心理学过程，因此相关的分析内容较少。我们在尽量保持文章简洁的同时，增加了若干内容，以提升这部分内容与心理学的联系。如 3.1 中增加数字治疗联盟的分析（见意见 1），3.2 中增加情绪对学业影响的分析：学习不仅仅是认知过程，情绪在其中同样扮演重要角色(Camacho-Morles et al., 2021)。
3. 正如专家所指出的，5.1 和 5.2 两部分确实在本文中有重要价值。为了展现共情计算跟心理学的关系，我们仔细修改了 5.1、5.2 两个小节，具体内容如下（由于修改内容较多，所有的内容均标为绿色）

### 5.1 共情计算为心理学带来新的洞见

共情计算不仅为心理学研究提供新工具，更在理论层面深化和拓展对共情的理解。通过计算机自动分析和模拟共情互动的数据，研究者可以在更大的数据尺度上验证和迭代先前通过传统方法（观察、问卷和实验等）获得的研究成果，也可以识别尚未发现的模式，如有研

究使用共情计算来发掘对话中的潜在共情意图(Chen et al., 2022)。这有助于加深对共情影响因素和内在机制的理解。

另一方面，共情计算也拓展共情研究的理论视角。先前的研究表明，共情普遍存在于人类甚至动物中(de Waal, 2012)，是从亲密关系到大规模合作等一切社会互动基础(Zaki, 2014)。然而，在未来的智能社会中，人们的面对面互动正逐渐演化为面向网络和人工智能的新型互动，以至于人们有时不能直接看到对方的表情和肢体动作，不能直观地感受沟通对象的情绪。在这样的情景下，共情是否仍有普遍性，会发生哪些演变，以及如何促进人机交互中的共情等，都是亟待深入研究的问题（如图 3 所示）。共情计算将为我们理解智能社会互动过程中的共情现象提供基础，为建立包括人-人关系、人-机关系等在内的整体性、普适性的共情理论提供启示。

## 5.2 心理学家在共情计算研究中的角色

心理学家在共情计算领域发挥不可或缺的作用。编程等技术门槛曾是许多心理学研究者参与这一领域的阻碍，不过越来越多介绍计算方法的文章(苏悦 et al., 2022)正在消除这一障碍。大语言模型进一步降低了研究过程对编程技术的依赖，使得研究者可以通过 AI 辅助编程、提示词工程等方法开展研究。这种背景下，心理学家在共情领域积累的丰富理论和研究方法的重要性正在上升。心理学家的首要任务是提出问题。早期个体共情测量研究起源于改善咨询师共情水平的需求，未来心理学家需要持续关注共情计算的创新研究问题。其次，心理学家为设计和提升共情计算提供理论支撑，如心理治疗理论已被用于共情内容改写算法的开发(Lin et al., 2023)，未来心理学理论也可用于优化提示词工程(Li et al., 2023)，提升大语言模型在共情任务中的表现。最后，随着 AI 的行为和心理表现越来越接近人类，研究者开始依据心理测量原理，制定科学有效的评估方法，以测量机器的共情水平(Kosinski, 2023; Wang et al., 2023)。总之，心理学家在共情计算研究和应用过程中承担引领、评估和改进等角色，将与计算机科学家一起，使共情计算的理论基础更坚实，效果检验更可靠，应用创新更落地。

### 审稿人 3 意见：

文章探讨了“共情”及其在计算领域的应用——“共情计算”。共情是指感受和理解他人情绪的能力，它在人类社交中起到关键作用，与亲社会行为、人际关系和幸福感有关。随着 AI 技术的发展，特别是 ChatGPT 的推出，AI 开始展现出基础的共情能力。共情计算是一个新兴的跨学科领域，结合心理学和计算机科学，旨在自动化地分析和模拟共情。文章详细描述了共情计算的理论背景，首先回顾了传统的共情研究，包括共情的概念、测量方法、神经基础和在心理咨询中的应用。接着，文章定义了共情计算，强调了它不仅包括使用计算方法测量共情，还包括使 AI 或机器人展现共情能力。文章还提出了一个共情计算的研究框架，包括数据、模型和任务三个层面。总的来说，文章为我们提供了一个深入而全面的视角，探讨了共情在现代计算领域的应用和意义，强调了心理学和计算机科学之间的交叉合作在推动这一领域发展中的重要性。

意见 1: 不足之处: 1) 文章着重描述了不同方式的共情计算，但这段的逻辑略显杂乱，因为共情计算从对象上可以分为人类以及计算机，从分析手段上可以分为心理学与计算机评价等，在第二章节中，将这些平行排列，很难看到其中的逻辑关系。希望作者在修改稿中能够根据这些计算的目的以及手段等方面，阐述清楚之间的联系与区别，让读者不要从一个主题快速的跳到另外一个主题。

回应: 感谢审稿专家对本文的肯定和建议。此前的稿件中没有充分说明当前研究主题间的关

系，针对这一问题，我们在第 2 节开始处增加了一段概括性的描述，从目的和手段的维度重新组织了当前的四个研究主题，希望可以更清楚的展示他们之间的区别与联系。内容如下：

作为一个新兴领域，共情计算的研究内容在不断发展变化中。当前主要研究，可以根据目的和手段的不同分为四个主题。一方面，共情计算的首要任务是计算机分析和理解共情，分为（1）个体共情测量和（2）共情内容分类，前者侧重于分析个体的共情特质，后者聚焦于分析文本中的共情特征，而非真实的人。另一方面，这一研究领域致力于使计算机模拟和表现共情，包括（3）设计共情性回应系统和（4）开发生成式共情对话系统，前者设计有限的回应方式并根据规则给用户反馈以表现共情，后者则利用 AI 自动生成共情性的对话。这些主题相对独立但互补，随着研究的深入，也会产生新的主题。

意见 2: 2) 文章中提供了大量的研究例子和引用，但没有对这些研究的重要性或相关性进行解释。这可能导致读者感到混淆或不知所云。

回应：感谢专家提出的问题。我们确实未能对研究例子进行充分的解释说明。在修改版中，考虑到篇幅限制的条件下，我们对部分重要研究例子增加了解释，例子见下文。同时，我们结合意见 3 和意见 4，在每个主题后加入了整体的总结，其中也对这些研究的重要性和相关性有整体性的描述，详见意见 3 的回应。

对部分例子的重要性或相关性的解释：

个体共情测评是共情计算较早发展的领域，起初是为了评估心理咨询师，如通过文字和语音评价心理咨询师的共情能力(Xiao, Imel, Georgiou, et al., 2015)，随后扩展到其他情景，如通过社交媒体信息识别可能损害用户健康的致病性共情（pathogenic empathy）(Abdul-Mageed et al., 2017)。

Terzis 等(2012)根据计算机自动面部识别和人工识别相结合的方法，识别了学生的六种基本情绪（开心、生气、伤心、惊讶、害怕、厌恶），并根据情绪给出相应的鼓励或安抚，是这一领域最早的尝试。

如 Rashkin 等(2019)建立了一个包含约 25000 对共情对话的数据集

（EmpatheticDialogues），并利用这个数据集微调预先用 17 亿条论坛日常对话训练的模型，结果表明微调模型生成的对话比预训练模型更具共情性。这个数据集成为此后众多研究的基准。

意见 3: 3) 文章缺乏总结, 描述每个研究主题后，可以加入一个简短的总结或提炼，帮助读者更好地理解该主题的核心内容和重要性。

回应：感谢专家的建议。在之前的稿件中我们的总结不够明确和凝练，未能突出每个研究主题的核心内容和重要价值。根据您的意见 3 和意见 4，我们重新修改了第 2 小节每个研究主题后的评述部分内容。增加了有关每个主题的重要价值、以及优缺点的评述。

如 2.1 个体共情测评部分，增加内容：通过计算的方法测量个体的共情特质，为自动化、大规模评估和研究共情提供了有效的工具。……此外，这些研究也为设计能够展现共情能力的 AI 或机器人提供参考，如训练机器人在对话、语调、语速等方面表现得更接近高共情特质的人。

2.2 共情内容分类增加内容：聚焦于文本内容的共情计算研究不依赖个体心理测量数据，便于更大范围收集数据，补充并扩展了前文中个体共情特质的计算，为分析社交网络对话中的共情提供工具，也为生成共情对话提供参考。……然而，当前研究局限在于从第三人视角评估文本内容，未能真实反应内容作者和读者的个人体验。为了弥补这一局限，研究者正致力于收集真实对话情境中发言者和听众的第一人视角的共情体验数据，以更全面地分

析文本中的共情内容(Barriere et al., 2022; Omitaomu et al., 2022)。未来研究需要整合不同视角的数据, 继续提高算法准确性, 并进一步验证和拓展模型的跨场景适用性。

2.3 共情回应系统增加内容: 基于规则的共情系统设计在早期推动了共情计算领域的发展, 也在实际应用中取得了较好的效果, 如提升答题正确率(Guo & Goh, 2016), 缓冲消极情绪对于创造力的影响(Groh et al., 2022)等。这类研究技术门槛相对较低, 反馈逻辑清晰, 便于理解, 因而至今仍被研究者采用。然而, 其预先设计的回应内容, 难以应对复杂和细微的情绪的变化, 使得它们更适用于任务边界清晰的场景, 而在可迁移性方面存在一定局限。

2.4 共情对话生成增加内容: 共情对话生成研究正在蓬勃发展, 这些研究为人机交互提供更加人性化的体验, 也为其他应用场景奠定了基础, 如改写网络回复内容以增加共情等。以 ChatGPT 为代表的大语言模型的突破性进展, 使得生成共情内容越来越简单, 进一步加速了该领域的发展。尽管如此, 当前研究主要集中于单轮或较短的对话, 这与现实任务中所需的复杂多轮对话相比, 仍存在明显差距。未来的研究应关注将共情对话有机融入各种对话场景, 以满足实际任务需求。例如, 如何在共情和任务目标之间灵活切换并维持平衡, 确保在支持用户情感表达的同时, 避免偏离对话的主要目标。

意见 4: 4) 文章主要是描述性的, 没有对这些研究主题的优缺点或可能的挑战进行深入的分析 and 讨论。

回应: 这部分内容已结合意见 3 一起修改, 具体内容见上文

---

## 第二轮

审稿人 1 意见: 该文经过修改, 已经符合我本人的发表要求。

编委 1 意见: 同意发表。

编委 2 意见: 同意发表。

主编意见: 文章有创新性。同意发表。