《心理科学进展》审稿意见与作者回应

题目：从“拟人归因”到“联盟建立”：人与聊天机器人关系对参与度的影响
作者：磨然，方建东，常保瑞

第一轮

审稿人 1 意见：

论文阐述了用户如何和聊天机器人通过发展关系从而提高用户的参与度，对人与聊天机器人的互动提出了 4 个阶段并给出了相应的机制解释，对已有的研究做了整合的工作。文献丰富，结构清晰，写作流畅。

意见 1：题目“人与聊天机器人关系的内在机制”有点抽象，本文主要是说在 ISIs 中人与聊天机器人的互动过程。
回应：感谢审稿专家的意见，我们已将原题目《人与聊天机器人关系的内在机制及其对用户参与度的影响》调整为《从“拟人归因”到“联盟建立”：人与聊天机器人关系对参与度的影响》，希望使同行们能够更明晰地从题目中把握文章的核心思想，具体修改详见文中红色字体及标注。

意见 2：论文的目的是“将在 Skjuve 等人(2021a)的基础上将 HCRs 模型进一步完整化、具体化，并使之适应数字心理咨询/心理治疗”，但在论证过程中，许多论述好像不仅局限在数字心理咨询/心理治疗的背景下，讨论的范围可以更明确。
回应：感谢审稿专家的意见，我们已细致检查了文中每个部分的观点及证据，发现有些部分确实存在偏移，因此有必要作出调整以使文章更紧扣主题。具体而言，我们使每一阶段的讨论与总结更聚焦于 ISIs 以及 HCRs 对参与度的影响，删减了不当的论述，并补充、更替了更合适的观点及文献。具体修改详见文中红色字体及标注。

意见 3：论文在已有理论上提出了人与聊天机器人不断深入互动的几个阶段，不过也有文献提到人工智能机器人的恐怖谷效应，人和聊天机器人的互动过程中会不会也存在呢？
回应：感谢审稿专家的提醒，在人机交互的过程中，用户的体验的确有可能会受到恐怖谷效应的影响。鉴于时下聊天机器人的设计标准并未统一，又随着 3D 建模技术、GPT 等自然语言生成模型(如基于 GPT-3 的 Replika, 基于 GPT-3.5 的 ChatGPT)的迅速发展，因此很容易在人机交互时出现类似情况。例如，聊天机器人虽已表明其机器人的身份，却又在语言话术上展现出高度的智能；亦或是使用逼真的 3D 人类形象/真人照片，但对话时却又表现出机械、愚钝的反应，而用户对这种差异的感知则很有可能会使认知失调(它十分像人类，但它又不是真实的人类)亦或者说恐怖谷效应产生，进而对用户的参与度产生负面影响。基于此，恐怖谷效应及其影响需在文中阐述，聊天机器人的似人程度并非越高越好，其外在特征及对话能力需在符合用户期望的前提下进行合理地平衡，具体修改详见文中红色字体及标注。

意见 4：阶段 3：发展依恋关系这部分目前的论据有些薄弱，有不少人与人工智能机器人形成情感联结或依恋关系的研究可以参考。
回应：感谢审稿专家的指出，经过对全文的再次审视，我们认为文章所提出的 4 阶段的论证

均需加强。因此，我们基于意见 6 中的思路——“整合两部分”删除了“参与度部分”，并利用该部分的证据及新补充的文献来加强“4 阶段部分”的论证，同时也进一步梳理、优化了整合部分的论证逻辑，使之更有助于文章目的的体现，具体修改详见文中红色字体及标注。

意见 5：阶段 4：建立数字治疗联盟，先需要对数字治疗联盟有些界定，说明其内涵，这样才能更清晰地说明如何建立 DTA 的机制，以及 DTA 与前面阶段 2 和阶段 3 的关系。是不是前面两个阶段正好是 DTA 的认知和情感两个成分？

回应：感谢审稿专家的意见。首先，前三阶段的确是促进 DTA 的认知(拟人归因、功利价值判断)和情感(依恋关系发展)部分。其次，基于论证逻辑，需在引出 DTA 的作用和影响前，先对其概念内涵进行阐述，而后再在此基础之上厘清其与此前阶段的关联。最后，基于上述的考量，我们从 3 个方面进行了修改：第一，讨论了依恋关系发展的弊端，并引出 TA 概念；第二，阐述了 DTA 的内涵，讨论了建立 DTA 的原因，并从认知、情感两个维度分析了 DTA 与此前阶段的关联；第三，阐述了建立 DTA 的思路，具体修改详见文中红色字体及标注。

意见 6：目前感觉前面的 4 个阶段和后面的用户参与度两部分关联不是太紧密，而且用户参与度的 4 个指标其实不那么容易区分开。因此，可否考虑了将后面参与度的这些研究整合到前面的 4 个阶段，既可以增强 4 个阶段的论证（目前稍显薄弱），又可以更直接体现两者的关系（研究的主要目的）。或者用户参与度下面的内容重新按研究的性质和主题来组织，也许可以和前面的阶段更加对应。或者先介绍聊天机器人对提升参与度的研究，然后再提出理论解释。这些思路供作者参考，都是希望两部分的关系更紧密。

回应：感谢审稿专家独到且周全的建议。首先，我们也认同参与度的指标确实不那么容易分开(如配合度与出席率)，这也会导致在论证时观点出现重合，因此我们将直接讨论参与度本身。此外，我们也发现“4 阶段部分”的观点与“参与度部分”的证据实际上存在较高的契合度。因此，在权衡之下，我们认为您提出将“参与度部分”整合至“4 阶段部分”的思路更有利于提升文章质量。一方面，整合有助于“4 阶段部分”的观点强化，弥补论证薄弱的短板；另一方面，我们还可以借此机会对文章目的作进一步聚焦，将合适的观点、证据调整至更恰当的位置，使此前的行文逻辑更为流畅，同时也尽力避免这两部分“相互独立”之嫌，具体修改详见文中红色字体及标注。

意见 7：样本量的符号斜体。

回应：感谢审稿专家的意见，我们已检查文中的样本量符号，并按要求规范了格式，具体修改详见文中红色字体及标注。

……

审稿人 2 意见：这篇论文总体不错，有助于对该领域研究者了解相关进展。

意见 1：主要的问题在于，其所提出的阶段 1-拟人化，与其他阶段区分度还不够。客观上，拟人化其实伴随所有的阶段，可能是并行的，拟人化程度越高，效果会越好，所以这一点还需要作者再斟酌。

回应：感谢审稿专家的指出，此意见发人深省，经过再次检索、阅读相关文献及认真思考，我们认为导致此问题的原因，是我们对“拟人化”的理解不够深入，进而误出现了表述不准确、阐述不全面两个问题。

首先，拟人化指的是个体将似人特征(形象、言语)、动机、意图或情感归赋予非人对象的过程。这个过程包含了三个协同作用的因素：一是诱发主体知识(Elicited Agent

Knowledge)，即个体被非人对象激活后，调整并运用知识来推知其特征；二是效能动机(Effectance Motivation)，即个体寻求与其所处环境互动，并增进对其理解、预测和掌控的需要；三是，社会动机(Sociality Motivation)，即个体对社会接触、社会联结及社会支持的需要(Epley et al., 2007; 许丽颖 等, 2017)。基于此，我们认为“拟人化”诚如您所指出的，是贯穿于整个人机关系之中的，无论是从人类感知到聊天机器人后拟人化被启动，到与聊天机器人进行简单的社会互动，再到与聊天机器人进行情感交互并发展出依恋关系，这整个过程均是聊天机器人逐渐具备“人性”的过程。因此，只有人类将聊天机器人逐步拟人化，每个关系阶段才得以顺利推进。

其次，我们得到了新的启发：若结合拟人化的因素及关系发展的目的，从需求满足的角度进行分析，拟人化中的几个因素在每个关系阶段的侧重可能有所不同。在人机交互之初，拟人化的认知过程以拟人归因为主(形象、言语等初级归因)，动机驱动为辅，目的是使后续的需求满足有良好的认知基础。而在后续的关系发展过程中，逐渐以动机驱动为主，拟人归因为辅，目的是使得个体的需求能得到更好的满足，因此，在效能、社会动机的推动下，关系将得到发展，而拟人的归因也将更为深入(情感、动机等高级归因)，并进一步促进关系的发展。可以说，拟人化为关系发展奠定了基础，并随着关系的发展得到加强，而其效应的增强也会反过来服务于关系的发展及稳定。

再次，我们设计阶段 1 的目的，是为了补充、强调人机交互初期的认知加工过程，只有建立在拟人的归因、认知之上，关系才能得到初步的发展。因此，“诱发主体知识”相对于另外两个动机因素在此阶段扮演了更为重要的角色，即用户由于基本的需要，在受到聊天机器人的外在形象或是言语等初级线索的刺激时，无意识地将其认知为一个社会行动者，进而愈加倾向、习惯于使用人际策略与之进行社会交互——这不但是拟人化的开端，也是本文所讨论的人机关系发展的重要前提(Epley et al., 2007; Nass & Moon, 2000; Nass et al., 1994)。在 Pentina 等人(2023)的最新研究中，也验证了聊天机器人外在形象的似人程度(Human-Likeness)与逼真程度(authenticity)作为依恋关系发展模型起点的合理性。被试越认为聊天机器人看起来是一个可以进行社会交互的对象，他们与之进行人际互动的次数也越频繁，而依恋关系也因此得到了发展。

最后，我们根据此意见得到的启示是，应将阶段 1 更清晰化、具体化(更准确)，并在拟人化的三因素理论(Epley et al., 2007)的基础上，使拟人化的影响融入后续的 HCRs 阶段中(更全面)。具体修改思路是：第一，我们将阶段 1 的命名更改为“拟人归因”，希望强调拟人化之初的效应，即个体对非人对象推理、归因及带来的影响。第二，基于新的思考与观点，调整阶段 1 中的相关论述与总结，使观点更聚集于这一具体过程。第三，将拟人化的影响融入于后续的 HCRs 阶段中，使其在本文的理论假设中得到更为完整体现。具体修改详见文中红色字体及标注。

许丽颖, 喻丰, 邬家骅, 韩婷婷, 赵靓. (2017). 拟人化：从“它”到“他”. *心理科学进展* , *25*(11), 1942-1954.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81−103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*(pp. 72–78). Boston: ACM.

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, *140*, 107600. https://doi.org/10.1016/j.chb.2022.107600

意见 2：“用户的事后接受期望相对于事前接受期望”是什么意思？

回应：感谢审稿专家的指出。此处应存在表述错误以及表述不清晰两个问题。首先，“事后接受期望”应为“使用后的期望证实”。其次，这句话应具体解释为：在用户使用产品后，他们会根据使用后的实际体验来与使用前的期望进行对比，来确认他们的期望是否得到了证实，而这种期望得到证实与否，也将决定用户对产品的满意度，并影响其后续的实际使用行为。基于此，我们已经将对应内容更正，具体修改详见文中红色字体及标注。

意见 3：人化效应与发展依恋关系之间的区别和联系是什么呢？

回应：感谢审稿专家的指出，基于意见 1 中的思考，我们以拟人化的三因素理论(Epley et al., 2007)来解释两者的区别与联系，具体如下：

两者的区别：

a) 拟人化效应，也即拟人化的产生及其所带来的影响，是个体在动机因素(效能动机、社会动机)的推动下，将似人特征赋予非人对象，并无意识地将它们当作社会行动者，个体也因此更容易对它们产生积极的反应，进而使个体与环境互动、获得社会支持等基本需求能得到更好地满足(Epley et al., 2007; Nass & Moon, 2000; Nass et al., 1994)。

b) 发展依恋关系的重点，是个体试图与依恋对象建立稳固的情感联系，使其更深层次情感需求(被爱)得到满足——维持紧密联系、拒绝分离(Bowlby, 1977)。

c) 因此，他们二者的不同体现为三点：第一，机制与目标不同。拟人化主要为认知过程，目标是为后续社会交互及情感需求的满足提供良好的基础，具有间接性。发展依恋关系更侧重情感过程，目标是使情感需求得到直接、稳定地满足，具有直接性；第二，触发的难易不同。其中，拟人化更容易被触发，强调个体的自发性，即个体会因内在动机的推动下自然而然地将周遭的非人对象拟人化。而依恋则需要非人对象具有特定的功能(安全基地：值得信任且能提供支持，避风港：缓解痛苦)才能得到发展，因此其效应相对不那么容易产生(Rabb et al., 2022)。第三，情感需求的满足程度存在差异。其中，拟人化效应虽可使个体的社会需要得到一定满足，但它更多是作为一种认知基础，非人对象需有所“作为”，个体的需求才可得到更好的满足，因此，具备特定功能的依恋对象能够更好地满足个体的情感需求。

两者的联系：

a) 在人机交互之初，拟人化的重点在于个体将非人对象作拟人的归因，以期为其后续的需求满足(降低不确定性、增加社会接触)提供良好的认知基础，这也是依恋关系发展的前提(Epley et al., 2007)。

b) 而在后续的关系发展过程中，拟人化主要由动机主导，为个体的需求满足服务。由于个体具有寻求社会接触、社会支持的基本需要，因此，这些动机将会推动依恋关系的发展，并增进拟人化的程度(许丽颖 等, 2017)。究其原因，则是为了使情感需求的满足变得更容易、稳定，因此，个体对非人对象的拟人归因将会从外在特征的初级归因，而逐渐深入至动机、情感方面的归因(Pentina et al., 2023)。

c) 总之，拟人化为依恋关系的发展奠定了基础，并依赖依恋关系的发展使之得到加强，而其效应的增强也反过来服务于依恋关系的发展及稳定。

基于此，我们将此处的部分观点补充至文章中，使这两个概念具有逻辑关联，具体修改详见文中红色字体及标注。

许丽颖, 喻丰, 邬家骅, 韩婷婷, 赵靓. (2017). 拟人化：从“它”到“他”. *心理科学进展* , *25*(11), 1942-1954.

Bowlby, J. (1988). A Secure Base: Clinical Applications of Attachment Theory. Routledge: London, UK.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56 (1), 81−103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In P*roceedings of the SIGCHI Conference on Human Factors in Computing Systems*(pp. 72–78). Boston: ACM.

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, *140*, 107600. https://doi.org/10.1016/j.chb.2022.107600

Rabb, N., Law, T., Chita-Tegmark, M., & Scheutz, M. (2022). An attachment framework for human-robot interaction. *International Journal of Social Robotics*, *14*(2), 539–559.

意见 3：表述问题，如“HCI”的缩写问题、“交互效应”的意思不明确。

回应：感谢审稿专家的意见，我们已认真检查文中的相关表达并作出了相应调整，具体修改详见文中红色字体及标注。

第二轮

审稿人 1 意见：

作者对人与聊天机器人关系的发展过程进行了深入探讨，本次修改也补充了不少内容，还有一些意见供参考。

意见 1：行文逻辑方面：

1) 一些段落太长，包含多个主题，可以考虑分段，每段话写一个主题更好，这样也能清楚看到写作的逻辑。

2) 不少地方的写作逻辑还可以更清晰一些。比如，2.4 阶段 4：建立数字治疗联盟部分，第一段首先讲不安全依恋（不清楚为何以这个开头，以及与后续内容的关系），紧接着用 1156 个字的一段话分别介绍治疗联盟 TA、数字治疗联盟 DTA、DTA 与参与度、然后再说为何要建立 DTA、再说 DTA 与之前阶段的关系，最后一段说如何促进 DTA 的发展（没有明白利用多模态技术与 DTA 之间的关系）。

3) 内容和逻辑可以更加简洁清晰，突出主题。其他部分的写作也有类似的情况。将参与度与阶段合并后，每个阶段下的内容较丰富，有个清晰的结构来体现变量的关系或论证思路更好。比如，每个阶段下面可以先介绍概念理论、机制，再说与参与度的关系，最后再说如何提升的方法。现在很多都是混在一起的，思路不是很清晰。

回应：感谢审稿专家的具体意见。文章内容在整合后逻辑较为庞杂，在结构性上也较为欠缺，这将不利于清晰地呈现文章思路。对此，我们在横向对比各阶段后将此次修改分为三部分：第一，规整各阶段的内部逻辑，尽量以合理且一致的结构进行论述。具体而言，将基于下述结构对 4 阶段进行统一调整：核心概念介绍→机制阐述→与参与度关系→小结。第二，确保主题间的逻辑衔接，调整、删减关联性弱的内容，并根据新结构适当增补论述。第三，避免长段落，使用分段讨论使思路得到更清晰的展现。具体修改详见文中绿色字体及标注。

意见 2：DTA 既有认知也有情感部分，但在模型图中，只有情感部分促进，没有体现认知部分的影响。

回应：感谢审稿专家的意见，我们已调整模型图，具体修改详见文中绿色字体及标注。

意见 3：ChatGPT、恐怖谷、多模态技术这些可以考虑放到展望部分。
回应：感谢审稿专家的意见，我们已将对应内容调整至展望部分，具体修改详见文中绿色字体及标注

第三轮

审稿人 1 意见：同意发表，无进一步修改意见。

编委 1 意见： 该文经数轮修改，已达发表水平，推荐发表。

编委 2 意见：文章思路清晰、逻辑性好，文笔流畅。文中观点新颖，对读者有启发意义。达到了发表水平！

主编意见：同意发表。