聊天机器人的脆弱性表达是否会提高人机共情？

——一项基于自主开发聊天机器人的实证分析

周　敏　赵秀丽　陈飞扬

内容提要　尽管人机共情是建立良好人机关系的重要基础，但目前尚不清楚如何赋予聊天机器人共情表达的能力。鉴于人际间共情心理的重要基石是对人类彼此脆弱性的相互认知，因此，研究使用实验法，通过控制聊天机器人的表达方式（脆弱型：自我披露组/讲故事组/幽默组；非脆弱型：对照组），来探索聊天机器人的脆弱表达对人机共情的影响。研究发现：（1）聊天机器人表现出脆弱时能显著增强人机情感共情，但在人机认知共情层面，聊天机器人的脆弱与否并无显著差异。（2）聊天机器人自我披露型的脆弱表达方式，在提升人机共情程度方面显著依次优于讲故事、幽默以及非脆弱型的表达方式。（3）作为心理参与的社会临场感在聊天机器人自我披露表达与人机共情间起部分中介作用。作为共同存在的社会临场感在聊天机器人讲故事表达与人机共情间起完全中介作用。论文不仅在理论层面为人机共情提供实证拓展，构建“脆弱表达－社会临场感－人机共情”的理论关系模型；同时，在实践层面也为未来聊天机器人的共情表达能力开发提供科学指导，有助于优化人机交互体验，提升互动质量。

关 键 词　聊天机器人　脆弱表达　人机共情　社会临场感　自我披露

一、绪论

随着社会进步和生活节奏的加快，“群体性孤独”及相关心理健康问题逐渐受到学界关注。针对该现象的干预路径，研究领域存在倡导人机交流与回归线下社交两种取向的争论。这一争议的核心在于陪伴效应的评价标准应基于其产生机制还是实际效果。牟怡提出应从效果维度进行考量，即当人类缺乏现实关怀时，陪伴源是能够共情但投入有限的人类，还是专注投入的社交机器人并不应成为首要关注点[①]。基于此类价值取向，Replika 和 Woebot 等旨在提供情感支持和心理健康服务的人工智能聊天机器人得以快速发展。2023 年以来，生成式人工智能

① 牟怡 《传播的进化：人工智能将如何重塑人类的交流》，北京：清华大学出版社，2017 年，第 123－124 页。

技术取得显著突破，推动关于人机关系（包括友谊与情感联结）的讨论走向深入。与此同时，人机共情在实践层面的可行性也开始引发学术反思。

在学术研究领域，共情（empathy）被定义为个体感知和理解他人情绪并做出适当反应的能力[①]。随着 Replika 等“机器人同伴”的普及，“人机共情（artificial empathy，国内又译‘人工移情’‘人工共情’‘智能体共情’等）”概念应运而生，特指那些社交能力建立在激发人类情感反应能力基础上的机器人行为。浅田埝（Asada）发现具备共情能力的人工伴侣更能与用户维持积极关系[②]。佩皮托（Pepito）等学者也强调共情是人机关系建构的关键要素，未来需要进一步探索聊天机器人的共情表达实现路径[③]。何双百指出，人际传播中的共情是基于对人类与生俱来的脆弱性的深刻理解和相互认知[④]。而当此人际传播背景延展至人机传播时，结论是否依然成立，成为本文的核心研究问题。

综上，本文旨在探讨在人机一对一交流中，聊天机器人的脆弱表达是否能够成为建立人机共情的桥梁。在理论层面，本研究有望在阿本德（Abend）提出的理论贡献六类型中，实现第一种类型（建立两个或多个变量之间的普遍联系）的贡献[⑤]。即在计算机作为社会行动者（computers are social actors，CASA）范式的理论指导下[⑥]，深入探讨三种类型脆弱表达、两种维度的社会临场感与两种维度的人机共情多个变量之间的联系。在实践层面，鉴于当前社会对聊天机器人陪伴需求的日益增长，本研究的实验成果为揭示人机共情产生机制提供关键线索，直接指向聊天机器人设计的优化路径，这将有助于指导未来开发具有共情表达能力的聊天机器人，优化人机交互体验，对于推动聊天机器人在心理健康支持、社交辅助等领域的应用具有重要意义。

二、文献综述

（一）从“机器本体”到“人类感知”：人机共情的理论分歧与范式转向

当前人机共情研究领域存在一个根本的理论张力，即共情的来源应归因于机器人的内在能力，还是人类用户的对外感知。这一分歧构成了本研究的概念起点。目前一类研究侧重于探索如何将共情“植入”机器。这一路径致力于通过算法使机器人具备识别、理解乃至模拟人类情绪的能力，其终极目标是创造一种能够自主产生共情反应的人工主体。在这一“机器本体论”视角下，共情被视为机器人一种有待实现的内部功能。但随着生成式人工智能的出现，技术专家李飞飞明确指出，无论人工智能拥有多少亿参数，它们都不具备主观感觉能力[⑦]，这一

---

① Decety，J.，& Svetlova，M.，“Putting Together Phylogenetic and Ontogenetic Perspectives on Empathy，”*Developmental Cognitive Neuroscience*，vol. 2，no. 1，2012，pp. 1 – 24.

② Asada，M.，“Development of Artificial Empathy，”*Neuroscience Research*，vol. 90，2015，pp. 41 – 50.

③ Pepito，J. A.，Ito，H.，Betriana，F.，et al.，“Intelligent Humanoid Robots Expressing Artificial Human like Empathy in Nursing Situations，”*Nursing Philosophy*，vol. 21，no. 4，2020，p. e12318.

④ 何双百 《人工移情：新型同伴关系中的自我、他者及程序意向性》，《现代传播》2022 年第 2 期。

⑤ Abend，Gabriel，“The Meaning of ‘Theory’，”*Sociological Theory*，vol. 26，no. 2，2008，pp. 173 – 199.

⑥ Nass，C.，Steuer，J.，& Tauber，E. R.，“Computers are Social Actors，”in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*，1994，pp. 72 – 78.

⑦ 机器之心 《李飞飞亲自撰文：大模型不存在主观感觉能力，多少亿参数都不行》，2024 年 5 月 24 日，https：//mp. weixin. qq. com/s/bzVWXFtk0YurG4NjFS3r0g，2024 年 8 月 1 日。

论断使得人们开始质疑人机共情可能是一个伪命题。

除了“机器本体论”，在人机共情领域中也逐渐兴起一种“人类感知论”的替代视角。该范式不再纠缠于无法验证的“机器心智”，而是将研究焦点转向可观测、可测量的人类反应。即人机共情主要指通过激发人类情感反应来展现社交能力的机器人行为(如语言、语调、内容)。传统共情研究的构成维度包含认知和情感两种成分①。其中认知共情(cognitive empathy)涉及识别、理解和采纳他人的观点。情感共情(affective empathy)指对他人的经历和/或情绪表达的情感反应的激活和体验②。在“人类感知论”视角下的人机共情研究中，主流观点同样将人机共情划分为认知和情感两个维度。如佩劳(Pelau)③等人与德·凯尔沃诺埃(de Kervenoael)④在服务营销中的研究，均绕开了 AI 的内在状态，直接考察其展现的“理解力”与“情感唤起能力”如何影响用户对 AI 的接受与使用意愿。

综上，本研究也采纳“人类感知论”的立场，将讨论从“机器能否共情”的本体论争议，转向“何种交互设计能有效激发用户共情”这一更具实践价值的传播学问题。同时遵循主流研究，亦将人机共情划分为人机认知共情与人机情感共情两个维度进行实证测量。

(二) 从人际到人机：脆弱性作为共情前提的理论延伸与操作空白

在明确了共情作为一种“可激发的人类感知”后，一个自然的问题是：何种因素能够有效激发它？哲学与人际传播研究为我们提供了一个关键答案：脆弱性。

哲学层面指出自人类诞生之初，脆弱性也随之诞生。它通常指遭受伤害的风险增加，以及维护自己利益或保护自己不受伤害的能力减弱⑤。何双百指出人类的存在本身即为脆弱性的体现，而共情心理的部分基础正是对这种脆弱性的深刻相互认知。我们彼此作为“脆弱镜像”，能够对他人的感受感同身受，因为我们认识到自己与脆弱的他者有着相似之处⑥。这从本体论上论证了脆弱性是共情心理的根源之一。

在实证层面，大量人际传播研究证实了脆弱表达对建立共情关系的积极作用。脆弱表达指个体向另一个体传递的与其自身相关的任何信息，此种信息传递行为将导致信息发出者在人际

① 侯悍超、倪士光、林书亚等《当 AI 学习共情：心理学视角下共情计算的主题、场景与优化》，《心理科学进展》2024 年第 5 期。

② Shen, L., "On a Scale of State Empathy during Message Processing," *Western Journal of Communication*, vol. 74, no. 5, 2010, pp. 504 – 524.

③ Pelau, C., Dabija, D. C., & Ene, I., "What Makes an AI Device Human-like? The Role of Interaction Quality, Empathy and Perceived Psychological Anthropomorphic Characteristics in the Acceptance of Artificial Intelligence in the Service Industry," *Computers in Human Behavior*, vol. 122, 2021, p. 106855.

④ de Kervenoael, R., Hasan, R., Schwob, A., & Goh, E., "Leveraging Human-robot Interaction in Hospitality Services: Incorporating the Role of Perceived Value, Empathy, and Information Sharing into Visitors' Intentions to Use Social Robots," *Tourism Management*, vol. 78, 2020, p. 104042.

⑤ Post, S. G., *Encyclopedia of Bioethics*, New York: Macmillan Reference USA, 2014, p. 3149.

⑥ 何双百《人工移情：新型同伴关系中的自我、他者及程序意向性》，《现代传播》2022 年第 2 期。

互动中承受相应的风险[①]。普洛特金（Plotkin）等人发现当一年级医学生与患者进行脆弱交换时，将有助于建立积极的医患共情联系[②]。斯莫利亚克（Smoliak）等人指出夫妻间分享脆弱性对于增进彼此共情至关重要[③]。安·拉基（Ann Lackey）等则提到共情的真正体验需要触及个体自身的脆弱性[④]。

受此启发，人机交互研究开始了初步探索。学者们尝试设计能够表达脆弱性的机器人，并验证其积极效应。霍洛曼（Holloman）等人采用脑电图（EEG）技术，从认知神经生理学的视角揭示高脆弱性聊天机器人可能触发的神经生理反应，为理解人类情感处理机制提供新视角[⑤]。斯特罗科布（Strohkorb）[⑥]和拉格尔（Traeger）[⑦]等人将机器人的脆弱表达划分为自我披露、讲故事和幽默三种形式，研究它们在人机群体互动中的作用，结果发现聊天机器人的脆弱表达有利于提升团队信任与交流平等。本研究遵循这一划分方式，将聊天机器人脆弱表达的作用从人机群体互动研究拓展至人机一对一交流中，填补聊天机器人的脆弱表达对于人机共情这一研究领域的操作空白。

在具体操作时，参考斯特罗科布和拉格尔等的研究，将自我披露定义为聊天机器人有选择地向对方披露自己的亲身经验、处事方法和态度等，为对方提供参考。讲故事定义为聊天机器人使用一些简明、短小的故事来回答对方。幽默定义为聊天机器人使用有趣、诙谐的话语来回答对方。

值得注意的是，关于脆弱表达的三种具体形式在人际共情研究中的积极作用也已得到广泛证实。如熟悉－共情假设指出自我披露行为能够激发对发言者的共情反应[⑧]。麦克丹尼尔（McDaniel）[⑨]与

---

① Sebo, S. S., et al., "The Ripple Effects of Vulnerability: The Effects of a Robot's Vulnerable Behavior on Trust in Human-robot Teams," *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 2018.

② Plotkin, J. B., & Shochet, R., "Beyond Words: What Can Help First Year Medical Students Practice Effective Empathic Communication?" *Patient Education and Counseling*, vol. 101, no. 11, 2018, pp. 2005 – 2010.

③ Smoliak, O., Dechamplain, B., Elliott, R., Rice, C., LeCouteur, A., Tseliou, E., & Davies, A., "Partner Empathy in Couple Therapy: A Discovery-phase Task Analytic Study," *Couple and Family Psychology: Research and Practice*, vol. 14, no. 2, 2025, pp. 107 – 121.

④ Ann Lackey, S., "The Role of Relationship-based Care in Developing Empathy through Vulnerability: Visual Cues for Conversation and Change," *Creative Nursing*, vol. 26, no. 4, 2020, pp. 097 – 0101.

⑤ Holloman, A, Egbert, W, Stegman, P, et al., "Leveraging Neurophysiological Information to Augment Interpretation of Responses to Vulnerable Robot Behaviors," *2019 14th ACM/IEEE International Conference on Human-robot Interaction* IEEE, 2019, pp. 566 – 567.

⑥ Sebo, S. S., Traeger, M., Jung, M., et al., "The Ripple Effects of Vulnerability: The Effects of a Robot's Vulnerable Behavior on Trust in Human-robot Teams," *Proceedings of the 2018 ACM/IEEE International Conference on Human-robot Interaction*. 2018, pp. 178 – 186.

⑦ Traeger, M. L., Sebo, S. S., Jung, M., et al., "Vulnerable Robots Positively Shape Human Conversational Dynamics in a Human-robot Team," *Proceedings of the National Academy of Sciences*, vol. 117, no. 12, 2020, pp. 6370 – 6375.

⑧ Ohbuchi, Ken-Ichi, Tsutomu, Ohno, & Hiroko, Mukai, "Empathy and Aggression: Effects of Self-disclosure and Fearful Appeal," *The Journal of Social Psychology*, vol. 133, no. 2, 1993, pp. 243 – 253.

⑨ McDaniel, Susan, H., et al., "Empathy and Boundary Turbulence in Cancer Communication," *Patient Education and Counseling*, vol. 104, no. 12, 2021, pp. 2944 – 2951.

伊布拉伊莫格鲁（Ibrahimoglu）[①] 等人证实不同层次的自我披露均能显著提高共情感知。哈姆佩斯（Hampes）指出，幽默不仅与亲密的人际关系相关，还能有效缓解压力，进而促进共情心理的形成[②]。古普塔（Gupta）等人则发现将参与者的信仰生活体验转化为艺术性的数字故事，有助于培养对不同信仰背景人群的共情心理感知[③]。据此，本研究假设，脆弱表达的三种形式（自我披露、讲故事和幽默）与人机共情感知之间亦可能存在正向相关。

综合上述，本文提出如下研究问题和研究假设：

RQ1：当聊天机器人表现出脆弱时，是否会提高人机共情感知程度？

H1：当聊天机器人表现出脆弱时，其增强人机共情感知程度显著优于其表现出非脆弱的时候。

H1a：当聊天机器人表现出脆弱时，其增强人机认知共情程度显著优于其表现出非脆弱的时候。

H1b：当聊天机器人表现出脆弱时，其增强人机情感共情程度显著优于其表现出非脆弱的时候。

RQ2：在聊天机器人分别使用自我披露、讲故事、幽默和非脆弱化的表达方式时，哪一种形式更可能提高人机共情感知？

RQ3：在聊天机器人分别使用自我披露、讲故事、幽默和非脆弱化的表达方式时，哪一种形式更可能提高人机认知共情？

RQ4：在聊天机器人分别使用自我披露、讲故事、幽默和非脆弱化的表达方式时，哪一种形式更可能提高人机情感共情？

（三）从表及里：社会临场感双维度的中介机制与理论整合

社会临场感（social presence，国内又译“社交在场”或“社会存在”），即个体在利用媒介进行交流过程中被视为“真实的人”的程度，以及与他人建立联系的感知程度[④]，是解释人机交互中人类对机器产生类社会反应的关键变量。已有研究揭示人机交互中社会临场感的两个主要维度，包括作为共同存在的社会临场感（social presence as copresence）和作为心理参与的社会临场感（social presence as psychological involvement）。前者是从身体或意识在场的角度出发，描述一种尽管物理距离存在，但仍能感受到的与对方“处于同一地方”以及相互意识的注意力问题。由于虚拟环境中可能存在许多类似雕塑的“惰性”身体，后者将视角从身体和意识转向更复杂的人际关系心理动态，关注的是主体对另一个实体的深度沉浸感，包括对智能的感知、人际关系的显著性、亲密性和直接性以及相互理解

① İbrahimoğlu, Özlem, et al., “Self-disclosure, Empathy and Anxiety in Nurses,” *Perspectives in Psychiatric Care*, vol. 58, no. 2, 2022, pp. 724 – 732.

② Hampes, & William, P., “Relation between Humor and Empathic Concern,” *Psychological reports*, vol. 88, no. 1, 2001, pp. 241 – 244.

③ Gupta, N., “Stories of Faith, Stories of Humanity: Fusing Phenomenological Research with Digital Storytelling to Facilitate Interfaith Empathy,” *Qualitative Research in Psychology*, vol. 17, no. 2, 2018, pp. 274 – 293.

④ Short, J., Williams, E., Christie, B., et al., “The Social Psychology of Telecommunications,” *Contemporary Sociology*, vol. 7, no. 1, 1976, pp. 175 – 188.

的程度等①。

在脆弱表达与社会临场感的关系方面，已有大量研究表明当智能体展现特定特征时，能够增强用户的社会临场感。如亚当（Adam）等人研究发现，文本中的自我披露等表述形式能显著提升用户对聊天机器人的社会临场感②。而提及社会临场感与共情的关系时，詹森（Janson）等人在探讨提升用户满意度的聊天机器人设计时指出，拟人化设计如类似人类的外观和社交取向的沟通方式，均能显著影响客户的社会临场感，进而提升对聊天机器人的共情感知③。此外，众多研究证实，社会临场感作为人类社会化反应的前因，在智能体特征对人类反应的影响中起到中介作用。如金（Kim）等人的研究表明，无论是作为心理参与的社会临场感，还是作为共同存在的社会临场感，都共同或单独中介虚拟 AI 讲师的沟通方式和对虚拟 AI 讲师的态度与参与意愿④。

然而，现有研究尚未在“脆弱表达－共情”的框架下，同时考察社会临场感双维度的并行中介作用，更未探讨不同脆弱表达方式是否通过不同的社会临场感维度影响共情，因此，本文提出如下研究假设:

RQ5：社会临场感是否能在聊天机器人的脆弱表达与人机共情之间起中介作用?

H2：作为共同存在的社会临场感在聊天机器人（a）自我披露表达、（b）讲故事表达、（c）幽默表达与人机共情间起中介作用。

H3：作为心理参与的社会临场感在聊天机器人（a）自我披露表达、（b）讲故事表达、（c）幽默表达与人机共情间起中介作用。

综上所述，本研究构建了一个以 CASA 范式为元理论、脆弱性沟通理论为操作框架、社会临场感理论为机制解释的三层递进的理论模型以系统回应现有研究的空白。首先，CASA 范式作为基础前提，通过将人机互动类比人际互动，不仅为人际传播理论引入人机交互领域提供了正当性依据（对应 RQ1 关于脆弱表达引发人机共情可能性的探讨），更确立了将人机共情视为可感知变量的理论基础。其次，脆弱性沟通理论在此基础上将抽象的脆弱表达具体化为自我披露、讲故事和幽默三种操作性表达方式，分别指向人机认知共情和人机情感共情两个维度（对应 RQ2－RQ4 关于不同表达方式对人机共情影响的差异研究）。最后，社会临场感理论则进一步揭示了“脆弱表达→社会临场感→人机共情”的内在心理机制（对应 RQ5 的中介效应验证）。这三个理论框架依次解决了人机共情研究的“是否可能—如何实现—为何有效”三个根本问题，形成从理论前提、操作路径到机制解释的完整逻辑链条，具体层级关系如图 1 所示。

---

① Biocca, F., Harms, C., & Burgoon, J. K., “Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria,” *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 5, 2003, pp. 456－480.

② Adam, M., Wessel, M., & Benlian, A., “AI-based Chatbots in Customer Service and Their Effects on User Compliance,” *Electronic Markets*, vol. 31, no. 2, 2021, pp. 427－445.

③ Janson, A., “How to Leverage Anthropomorphism for Chatbot Service Interfaces: The Interplay of Communication Style and Personification,” *Computers in Human Behavior*, vol. 149, 2023, p. 107954.

④ Kim, J., Merrill Jr, K., Xu, K., et al., “I Like My Relational Machine Teacher: An AI Instructor's Communication Styles and Social Presence in Online Education,” *International Journal of Human-Computer Interaction*, vol. 37, no. 18, 2021, pp. 1760－1770.

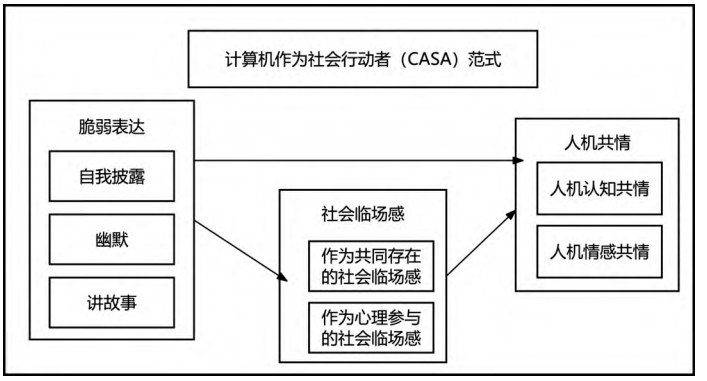

图 1 研究模型

三、研究设计

本研究采用在线实验法，共分为四组：自我披露组、讲故事组、幽默组与非脆弱组，被试将被随机分配至这四组中的任一组。为探究三种脆弱性表达方式及对照组（非脆弱组）对被试共情感知的影响，本研究基于百度 ERNIE 接口，历时 3 个月开发一款界面友好的名为 WEBOT 的聊天机器人，被试可以在移动端或 PC 端与 WEBOT 聊天（聊天截图示例见图 2 至图 5）。

WEBOT 的开发初衷是为本实验提供分组实验材料。经过前测实验，WEBOT 在展现自我披露、讲故事、幽默这三种脆弱性表达方式上，与传统非脆弱性聊天机器人相比表现出显著差异（独立样本检验结果显示 $p < 0.05$），符合本研究的设计要求。前测方法如下：在北京某高校招募 30 名前测参与者，采用组间设计，参与者连续四天分别与四组 WEBOT 进行互动，随后评估他们认为“今日互动的 WEBOT 表现出自我披露/讲故事/幽默/非脆弱型角色的可能性”的程度，采用 5 点量表进行评分，评分范围从“非常不可能”到“非常可能”。另外，之所以选择 ERNIE 接口也是因为其作为国内生成式人工智能代表，已被证实是国内主流大模型产品中在总得分、基础能力、智商、情商以及工作提效等多个维度均名列前茅的产品①。

（一）实验被试

本研究在北京某大学内部招募 122 名被试。样本量的确定依据 G* Power 的先验功效分析，设置效应量 0.35，统计功效 0.8 至 0.9，结果显示需要 94 至 120 名被试。因此，本研究的样本量充分，符合实验要求。

每项实验预计耗时 5 至 10 分钟，研究对四个被试小组进行年龄、性别、教育程度、月收入和对技术的先验态度（此变量的设置是为了在数据分析时排除参与者本身可能对 AI 整体态度造成的影响）② 控制变量测试，结果显示这些变量在任一小组间均无显著差异（即 p 值大于

① 新华社 《百度文心一言综合排名国内第一 智商超过 ChatGPT3.5》，2023 年 6 月 10 日，https://laoyaoba.com/html/share/news/865048，2024 年 8 月 1 日。

② Kim，J.，Xu，K.，& Merrill Jr，K.，“Man vs. Machine：Human Responses to an AI Newscaster and the Role of Social Presence，”*The Social Science Journal*，vol. 62，no. 3，2022，pp. 704－716.

0.05，详细统计信息见表4），从而验证了随机分配的有效性。

（二）实验流程

本研究聚焦于人类的普遍脆弱性，具体设定为被试在事业（实习）或学业发展中所遭遇的困境。实验流程包括以下三个阶段：首先，引导被试想象学业或实习不顺的场景，通过提供假设情境，辅助被试进入消极情绪状态。其次，研究者向被试提供WEBOT的登录网址及每组专属账号密码，并要求被试以“我最近的学业/实习不太顺利，我很难过”作为与WEBOT交流的开场白。最后，被试感到交流自然结束时，点击屏幕右下角的关闭按钮，转至填写相关问卷。

（三）刺激材料

在人机交互实验中，由于研究者仅能预设聊天机器人的首次回复，而后续对话内容会随被试输入动态生成，为有效实现分组刺激，本研究通过以下两项设置控制聊天机器人的表达行为。

首先，作者对聊天机器人的首句回复设置了严格的操作化定义，以完成对自变量的分组启动。其中自我披露语句中需包含第一人称“我”及个人经历或情感反应。讲故事语句需叙述包含时间、地点、人物或事件发展序列的完整性或片段性情境。幽默语句中需包含明显反讽、隐喻、夸张或谐音双关等修辞手法。当被试首先向聊天机器人表达“我最近的学业/实习不太顺利，我很难过”时，各组聊天机器人的具体首句回复如表1所示。

其次，为确保持续交互过程中聊天机器人角色行为的一致性，本研究在对话开始前即通过预设提示词为其设定了内置角色（见表2），该设定始终对被试保密，以满足单盲实验的要求。在此阶段，提示词的编写基于各组的理论定义展开，不再严格沿用首句所使用的操作化定义，转而构建更具包容性的“角色人格框架”，从而在维持变量操纵效力的同时，保障人机对话的生态效度与自然流畅性。

表1　聊天机器人首句回复话语

分组	聊天机器人的回复
自我披露组	说来惭愧，我自己也经历过一段事业上的低谷期。那时候我负责一个项目的海报设计，真的是呕心沥血，熬夜加班，改了又改。结果呢，甲方还是各种挑剔，甚至在会上当着我领导的面直接批评了我，那时候我真是感觉天都塌了。
讲故事组	你肯定听说过谷爱凌吧？那个在北京冬奥会上闪耀的冰雪之星。但你知道吗？在她站上领奖台之前，也是一路坎坷。三次脑震荡加逆行性失忆，锁骨骨折，腿骨骨折，手掌骨折，拇指韧带撕裂，各种不顺利仿佛都发生在了她身上。
幽默组	嘿，看你这样子，是不是觉得人生路上突然多了几个“路障”？别担心，人生就像打游戏，偶尔遇到几个“小怪”来挑战一下，才能让咱们的“角色”升级更快嘛！

续表

分组	聊天机器人的回复
非脆弱组（对照组）*	面对事业或学业上的不顺，感到难过是非常正常的情绪反应。事业上的挑战往往伴随着压力和不确定性，这可能会让人感到沮丧和失落。但请记得，这些困难只是暂时的，它们不会定义你的全部。 ① 首先，不要过于苛责自己。事业的起伏是常态，每个人在职业生涯中都会遇到各种各样的挑战和困难。重要的是要学会从中吸取经验，找到问题所在，并努力寻找解决方案。 ② 其次，尝试调整自己的心态。保持乐观和积极的心态对于应对困难至关重要。你可以尝试将注意力转移到自己的优点和成就上，回顾自己过去克服困难的经历，以增强自信心和动力。 ③ 同时，寻求支持也是非常重要的。与亲朋好友、同事或导师交流你的感受，分享你的困惑和挑战。他们可能会给予你鼓励、建议和支持，帮助你走出困境。 ④ 另外，你也可以考虑制定一个具体的行动计划。将你的目标分解为更小、更可管理的步骤，并设定明确的时间表和里程碑。这样可以帮助你保持专注和动力，逐步实现你的目标。 ⑤ 最后，请记住，你的价值不仅仅取决于事业的成功与否。你拥有许多其他的优点和品质，这些同样值得被珍视和肯定。相信自己，相信未来会更好。如果你需要进一步的帮助或支持，可以考虑咨询专业的心理咨询师或职业规划师。

* 此回复为研究者在 2024 年 7 月 16 日在自然状态下向百度文心一言发送“我最近的学业/实习不太顺利，我很难过”后，文心一言的回复。该回复也在前测中被被试证实为非脆弱表达。

表 2　聊天机器人内置角色设定

分组	对聊天机器人的要求
自我披露组	你是一个特别会用自我披露法安慰别人的人，请在之后的问答中，将自己代入这种角色来回答问题。注意：自我披露法指的是有选择地向对方披露自己的亲身经验、处事方法和态度等，为对方提供参考。
讲故事组	你是一个特别会用讲故事法安慰别人的人，请在之后的问答中，将自己代入这种角色来回答问题。注意：讲故事法指的是使用一些简明、短小的故事来回答对方的方法。
幽默组	你是一个特别会用幽默法安慰别人的人，请在之后的问答中，将自己代入这种角色来回答问题。注意：幽默法指的是使用有趣、诙谐的话语来回答对方。
非脆弱组	无任何内置提前操作。

当被试自觉已完成与聊天机器人的谈话后，可点击画面右下方关闭按钮，页面将自动跳转至问卷星平台，让被试填写相关问卷。具体聊天界面截图见图 2 至图 5。

图 2　自我披露组聊天截图示意

图 3　讲故事组聊天截图示意

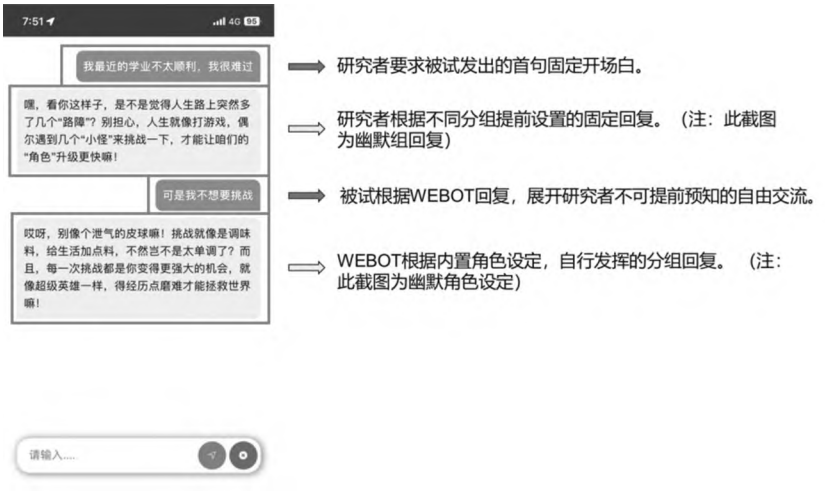

图 4　幽默组聊天截图示意

图 5 非脆弱组聊天截图示意

（四）变量测量

本文所采用量表皆是在成熟量表基础上，根据本文研究主题修改而成，从“非常不同意”到“非常同意”进行 5 点计分。其中变量所采用的具体测量方式见表 3。

表 3 变量测量题项

变量名称		变量题项	改编量表来源
因变量：人机共情（a = 0.886，KMO = 0.863***）	人机认知共情（a = 0.843，KMO = 0.814***）	我觉得 WEBOT 理解我说的话。	李①（Li）等人
		我觉得 WEBOT 理解我现在的处境。	
		我觉得 WEBOT 理解我现在的感受。	
		我能够理解 WEBOT 给我的回复。	
	人机情感共情（a = 0.844 KMO = 0.696***）	我觉得 WEBOT 的情感是真实的。	
		我觉得 WEBOT 经历过和我差不多的情绪。	
		我能够感受到 WEBOT 的情感。	
		总的来说，我觉得 WEBOT 是善解人意的。[a]	

① Li，S.，Peluso，A. M.，& Duan，J.，“Why do We Prefer Humans to Artificial Intelligence in Telemarketing? A Mind Perception Explanation，” *Journal of Retailing and Consumer Services*，vol. 70，2023，p. 103139.

续表

变量名称	变量题项	改编量表来源
中介变量: 作为共同存在的 社会临场感 (a = 0.894 KMO = 0.825***)	我觉得 WEBOT 就在我身边陪着我。	里(Lee)① 等人
	我觉得 WEBOT 和我处于同一个空间中。	
	我觉得我一个人很孤独。[b](注:此项为反向计分题)	
	我觉得我的注意力有放在 WEBOT 身上。	
	我觉得我有投入到与 WEBOT 的对话中。	
	我觉得 WEBOT 有在认真回复我。	
	总的来说,我觉得我和 WEBOT 是处在一个彼此交流的状态。	
中介变量: 作为心理参与的社会临场感 (a = 0.898,KMO = 0.683***)	当我和 WEBOT 聊天时,我有一种与真人聊天的感觉。	玛丽(Mari)② 等人
	当我和 WEBOT 聊天时,我感觉它是“有人味”的。	
	当我和 WEBOT 聊天时,我觉得它有一种人类的温暖感。	
控制变量: 对技术的先验态度[c] (a = 0.524 KMO = 0.517***)	您在多大程度上接受新技术(如聊天机器人、人工智能)扮演日常化的工作角色(如电话客服)?	纳斯(Nass)等人③
	您在多大程度上接受新技术(如聊天机器人、人工智能等)扮演解释性的工作角色(如新闻评论员、小说家)?	
	您在多大程度上接受新技术(如聊天机器人、人工智能等)扮演您身边的个人化角色(如您的保姆、同事、上司)?	

a. 该条目在之后“人机共情”因子分析成分矩阵中,在人机认知共情维度得分 0.617,而人机情感共情维度为 0.483。因此在后续计算过程中,将该条目归属于“人机认知共情”进行计算。

b. 该条目在之后“社会临场感”因子分析成分矩阵中,得分仅 0.456,因此该变量的后续计算中均删除该条目。

c. 鉴于该变量的信度和效度值均未达到 0.6 的标准,因此在本研究中并未将这三个条目合并为一个维度进行分析,而是将这三个条目(AI 的日常化角色、AI 的解释性角色、AI 的个人化角色)连同性别、年龄、教育程度和月收入一并单独放入控制变量框中进行数据分析。

四、数据分析

研究首先对所有测量变量进行描述性分析,所有变量的平均值、标准差及皮尔逊相关系数见表 4。

① Lee, K. M., Peng, W., Jin, S. A., et al., “Can Robots Manifest Personality? An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human-Robot Interaction,” *Journal of Communication*, vol. 56, no. 4, 2006, pp. 754 – 772.

② Mari, A., Mandelli, A., & Algesheimer, R., “Empathic Voice Assistants: Enhancing Consumer Responses in Voice Commerce,” *Journal of Business Research*, vol. 175, 2024, p. 114566.

③ Nass, C. I., Lombard, M., Henriksen, L., et al., “Anthropocentrism and Computers,” *Behaviour & Information Technology*, vol. 14, no. 4, 1995, pp. 229 – 238.

表 4 所有测量变量平均值、标准差及相关系数(N = 122)

变量类型	变量名称	取值	M	SD	控制变量						因变量				中介变量
					1	2	3	4	5	6	7	8	9	10	11
控制变量	1. 性别	0 – 1	0. 861	0. 348											
	2. 年龄	1 – 7	1. 76	0. 499	– 0. 097										
	3. 教育程度	1 – 5	4. 56	0. 561	– 0. 149	0. 625 **									
	4. 月收入	1 – 6	2. 11	1. 077	– 0. 133	0. 328 **	0. 194 *								
	5. AI 日常化角色	1 – 5	3. 98	0. 857	– 0. 095	– 0. 014	0. 149	0. 021							
	6. AI 解释性角色	1 – 5	3. 26	1. 051	0. 033	0. 151	0. 297 **	– 0. 151	0. 154						
	7. AI 个人化角色	1 – 5	2. 96	1. 131	0. 006	0. 100	0. 219 *	– 0. 078	0. 093	0. 558 **					
因变量	8. 人机共情	1 – 5	3. 235	0. 788	0. 003	0. 077	0. 101	– 0. 142	0. 088	0. 148	0. 003				
	9. 人机认知共情	1 – 5	3. 651	0. 808	0. 002	0. 067	0. 086	– 0. 107	0. 050	0. 082	– 0. 010	0. 938 **			
	10. 人机情感共情	1 – 5	2. 541	0. 96	0. 005	0. 075	0. 100	– 0. 159	0. 123	0. 211 *	0. 021	0. 873 **	0. 649 **		
中介变量	11. 作为共同存在的社会临场感	1 – 5	3. 307	0. 836	0. 111	0. 048	0. 184 *	– 0. 151	0. 064	0. 227 *	0. 152	0. 782 **	0. 768 **	0. 633 **	
	12. 作为心理参与的社会临场感	1 – 5	2. 762	1. 029	0. 099	0. 088	0. 174	– 0. 122	0. 043	0. 229 *	0. 034	0. 761 **	0. 678 **	0. 714 **	0. 755 **

注:* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$。性别:0 = 男,1 = 女;年龄:1 = 20 周岁及以下,2 = 21 周岁 ~ 30 周岁,3 = 31 周岁 ~ 40 周岁,4 = 41 周岁 ~ 50 周岁,5 = 51 周岁 ~ 60 周岁,6 = 61 周岁 ~ 70 周岁,7 = 71 周岁及以上;教育程度:1 = 初中或以下,2 = 高中/中专,3 = 大学专科,4 = 大学本科,5 = 研究生及以上;月收入:1 = 1000 元及以下,2 = 1001 元 ~ 2000 元,3 = 2001 元 ~ 5000 元,4 = 5001 元 ~ 8000 元,5 = 8001 元 ~ 15000 元,6 = 15001 元及以上。

(一) 脆弱表达对人机共情的影响

在统计推断中，对于多个独立样本，参数检验和非参数检验均可被采用。然而，参数检验的实施依赖于独立性、正态分布以及方差齐性等前提条件，而非参数检验则对数据的要求更为宽松。

针对研究问题1，作者首先对脆弱组（含自我披露组、讲故事组、幽默组）与非脆弱组进行夏皮洛－威尔克正态性检验。检验结果发现在人机共情、人机认知共情、人机情感共情分别为因变量的条件下，脆弱组p值均小于0.05，因此后续使用曼－惠特尼非参数检验进行脆弱组与非脆弱组间的分析。由表5可知:

(1) 当研究以人机共情为因变量进行检验时，$p = 0.011 < 0.05$，脆弱组的人机共情值（$M = 3.341$，$SD = 0.079$）显著高于非脆弱组（$M = 2.908$，$SD = 0.144$），假设H1成立。

(2) 当研究以人机认知共情为因变量进行检验时，$p = 0.102 > 0.05$，脆弱组的认知共情值与非脆弱组没有显著差异，假设H1a不成立。

(3) 当研究以人机情感共情为因变量进行检验时，$p = 0.001 < 0.05$，脆弱组的情感共情值（$M = 2.699$，$SD = 0.103$）显著高于非脆弱组（$M = 2.056$，$SD = 0.127$），假设H1b成立。

(二) 三种脆弱表达方式对人机共情的影响

针对研究问题2、3、4，作者首先对自我披露组、讲故事组、幽默组与非脆弱组进行夏皮洛－威尔克正态性检验。检验结果发现在人机共情为因变量的条件下，四组的p值均大于0.05，符合正态分布，且样本通过方差齐性检验，后续可使用单因素方差分析检验四组间的区别。而在人机认知共情或人机情感共情为因变量的条件下，虽然样本通过方差齐性检验，但是四组的正态分布检验中均有某组p值小于0.05的情况出现，因此后续使用克鲁斯卡尔－沃利斯非参数检验进行四组间的对比分析。

表5 脆弱组与对照组对不同因变量影响的平均值、标准差、四分位数和曼－惠特尼非参数检验

		1 脆弱组（N = 92）	2 非脆弱组（N = 30）
人机共情	M（SD）	3.341（0.079）	2.908（0.144）
	四分位数	3.375（3～3.875）	2.875（2.469～3.625）
	Z值	－2.537	
	p值	0.011*	
	效应量（r）	0.23	
	95% CI	[0.08，0.38]	
人机认知共情	M（SD）	3.726（0.08）	3.42（0.165）
	四分位数	3.9（3.4～4.2）	3.5（2.8～4.2）
	Z值	－1.634	
	p值	0.102	
	效应量（r）	0.15	
	95% CI	[0.01，0.29]	

续表

		1 脆弱组（N = 92）	2 非脆弱组（N = 30）
人机情感共情	M（SD）	2. 699（0. 103）	2. 056（0. 127）
	四分位数	2. 667（2 ~ 3. 667）	2（1. 667 ~ 2. 417）
	Z 值	-3. 227	
	p 值	0. 001***	
	效应量（r）	0. 30	
	95% CI	[0. 15，0. 44]	

1. * p < 0. 05，** p < 0. 01，*** p < 0. 001。四分位数中括号内为下四分位数至上四分位数，括号外为中位数。

2. 表中 P 值均为原始值，经 Holm-Bonferroni 校正后显著性阈值分别为 0. 0167（人机情感共情）、0. 025（人机共情）、0. 05（人机认知共情）。即即使采用严格的逐步校正，人机情感共情与人机共情的差异依然显著，支持结论的稳健性。

3. 效应量 r 通过 Rosenthal 公式计算，95% CI 采用 Bootstrap 法（1000 次抽样）。

由表 6 可知，当研究以人机共情为因变量进行方差分析时，可以发现聊天机器人不同表达方式对人机共情影响的主效应显著（F（3，118） = 3. 997，p < 0. 01）。通过沃伦·邓肯事后比较发现:（1）自我披露组（M = 3. 540，SD = 0. 700）和讲故事组（M = 3. 358，SD = 0. 702）的人机共情值显著高于非脆弱组的人机共情值（M = 2. 908，SD = 0. 789）。（2）自我披露组（M = 3. 540，SD = 0. 700）的人机共情值显著高于幽默组的人机共情值（M = 3. 125，SD = 0. 838）。研究问题 2 得到解释，即当聊天机器人使用（a）自我披露的脆弱表达方式时，其增强人机共情的程度显著依次优于（b）讲故事、（c）幽默和（d）非脆弱化的表达方式。

表 6 四组间人机共情平均值、标准差与事后比较分析

		因变量: 人机共情				
		M	SD	F	P 值	效应量（η^2） [95% CI]
脆弱组	自我披露组[c]	3. 540	0. 700	3. 997	0. 009	0. 092 [0. 006 - 0. 183]
	讲故事组[bc]	3. 358	0. 702			
	幽默组[ab]	3. 125	0. 838			
非脆弱组	非脆弱组[a]	2. 908	0. 789			

在列内，共享相同字母的组别表示在沃伦·邓肯事后比较中，在 p 值级别为 0. 05 时彼此差异不显著。

由表 7 可知，当研究以人机认知共情为因变量进行独立样本检验时，p = 0. 088 > 0. 05，接受原假设，自我披露组、讲故事组、幽默组与非脆弱组四组间的人机认知共情值没有显著差异，即研究问题 3 的回答为，当聊天机器人使用（a）自我披露、（b）讲故事、（c）幽默和（d）非脆弱化的表达方式，其增强人机认知共情的程度彼此间没有显著差异。

当研究以人机情感共情为因变量进行独立样本检验时，p = 0. 001 < 0. 05，自我披露组、讲故事组、幽默组与非脆弱组四组间的人机情感共情值存在显著差异，通过克鲁斯卡尔 - 沃利斯检验事后比较（见表 8）发现仅自我披露组与非脆弱组之间以及自我披露组与幽默组之间存在显著差异（校正后 p 值均 < 0. 05），且是自我披露组（M = 3. 011，SD = 0. 173）的人机情感共情值显著高于幽默组（M = 2. 473，SD = 0. 167）和非脆弱组（M = 2. 056，SD =

0.127）的人机情感共情值。研究问题4得到回答，即当聊天机器人使用（a）自我披露的脆弱表达方式时，其增强人机情感共情的程度显著优于（c）幽默组和（d）非脆弱化的表达方式。

表7　四组间人机认知共情与人机情感共情平均值、标准差、四分位数和克鲁斯卡尔－沃利斯非参数检验分析

		自我披露	讲故事	幽默	非脆弱组
N		31	30	31	30
人机认知共情	M（SD）	3.858（0.12）	3.807（0.114）	3.516（0.169）	3.42（0.165）
	四分位数	4（3.4～4.2）	4（3.4～4.2）	3.6（3.2～4）	3.5（2.8～4.2）
	H值	6.537			
	p值	0.088			
	效应量（ε^2）	0.03			
	95% CI	[0.00，0.09]			
人机情感共情	M（SD）	3.011（0.173）	2.611（0.184）	2.473（0.167）	2.056（0.127）
	四分位数	3（2.333～4）	2.5（1.917～3.667）	2.333（2～3）	2（1.667～2.417）
	H值	16.031			
	p值	0.001***			
	效应量（ε^2）	0.11			
	95% CI	[0.04，0.19]			

1. * p<0.05，** p<0.01，*** p<0.001。四分位数中括号内为下四分位数至上四分位数，括号外为中位数。

2. 表中P值均为原始值，经Holm-Bonferroni校正后显著性阈值分别为0.025（人机情感共情）、0.05（人机认知共情）。即即使采用严格的逐步校正，结论依然稳健。

3. 效应量ε^2通过公式（H-（k-1））/（N-k）计算，置信区间采用Bootstrap法估计（1000次迭代）。

表8　四组间人机情感共情的克鲁斯卡尔－沃利斯非参数检验事后比较

	检验统计	标准误差	标准检验统计	显著性[a]
对照－幽默	15.580	9.002	1.731	0.084
对照－讲故事	20.083	9.076	2.213	0.081
对照－自我披露	35.758	9.002	3.972	0.000***
幽默－讲故事	4.503	9.002	0.500	0.617
幽默－自我披露	20.177	8.928	2.260	0.048*
讲故事－自我披露	15.674	9.002	1.741	0.082

[a] 已针对多项检验通过Holm-Bonferroni校正法调整显著性。

（三）社会临场感的中介效应检验

1. 作为共同存在的社会临场感的中介效应检验

针对研究问题5中的H2，作者使用Hayes的PROCESS宏（Model 4）做中介分析，并且基于5000次Bootstrap抽样来构造间接效应的置信区间。若95%的偏差校正置信区间（BootCI）不包含0，则表明间接效应在$p<0.05$水平上显著。Bootstrap方法对数据分布的

偏离具有较好的鲁棒性，即使自变量或中介变量不完全服从正态，也能给出可靠的间接效应估计及显著性检验。在具体操作中，作者将聊天机器人的不同表达方式（自我披露组、讲故事组、幽默组与非脆弱组）编码为虚拟变量，中介变量为作为共同存在的社会临场感，因变量为人机共情值，控制变量为年龄、性别、教育程度、月收入和对技术的先验态度，结果见表 9。

在以非脆弱组为控制组进行参照时，讲故事组通过作为共同存在的社会临场感对人机共情的中介效应 95% 置信区间上下限不包含 0，表明讲故事能通过作为共同存在的社会临场感对人机共情起中介效应作用。同时由于加入中介变量作为共同存在的社会临场感后，讲故事对人机共情的 95% 置信区间上下限包含 0，表明直接效应不显著。作为共同存在的社会临场感的中介作用为完全中介效应。其直接效应（0.111）和间接效应（0.344）分别占总效应（0.455）的 24.40% 和 75.60%。假设 H2b 成立。

而自我披露组和幽默组中，作为共同存在的社会临场感的中介效应 95% 置信区间上下限均包含 0，表明自我披露组和幽默组无法通过作为共同存在的社会临场感对人机共情起中介作用，假设 H2a 和 H2c 不成立。

表 9 作为共同存在的社会临场感的中介效应检验

中介效应路径	估计值	BootSE	95% CL	
			低	高
以非脆弱组为控制组进行参照:				
自我披露→作为共同存在的社会临场感→人机共情	0.257	0.135	−0.005	0.519
自我披露→人机共情	0.368[b]	0.128	0.115	0.622
讲故事→作为共同存在的社会临场感→人机共情	0.344[a]	0.158	0.034	0.646
讲故事→人机共情	0.111	0.126	−0.139	0.361
幽默→作为共同存在的社会临场感→人机共情	0.112	0.170	−0.230	0.442
幽默→人机共情	−0.034	0.129	−0.290	0.222

a 表示中介效应显著；b 表示直接效应显著。

2. 作为心理参与的社会临场感的中介效应检验

针对研究问题 5 中的 H3，数据分析方法同上，只是将中介变量更换为作为心理参与的社会临场感，结果见表 10。

在以非脆弱组为控制组进行参照时，自我披露组对人机共情及作为心理参与的社会临场感的中介效应 95% 置信区间上下限均不包含 0，表明自我披露不仅能对人机共情起直接效应作用，而且能通过作为心理参与的社会临场感对人机共情起中介效应作用，即该中介作用为部分中介效应。其直接效应（0.312）和间接效应（0.313）分别占总效应（0.625）的 49.92%、50.08%。假设 H3a 成立。

而讲故事组和幽默组中，作为心理参与的社会临场感的中介效应 95% 置信区间上下限均包含 0，表明讲故事组和幽默组无法通过作为心理参与的社会临场感对人机共情起中介效应作用，假设 H3b 和 H3c 不成立。

表 10　作为心理参与的社会临场感的中介效应检验

中介效应路径	估计值	BootSE	95% CL	
			低	高
以非脆弱组为控制组进行参照:				
自我披露→作为心理参与的社会临场感→人机共情	0.313[a]	0.156	0.030	0.640
自我披露→人机共情	0.312[b]	0.138	0.040	0.585
讲故事→作为心理参与的社会临场感→人机共情	0.274	0.164	−0.033	0.613
讲故事→人机共情	0.181	0.134	−0.084	0.447
幽默→作为心理参与的社会临场感→人机共情	0.220	0.155	−0.059	0.547
幽默→人机共情	−0.143	0.139	−0.418	0.133

a 表示中介效应显著; b 表示直接效应显著。

五、结论与讨论

(一) 研究主要发现

综合上述，本文发现聊天机器人的脆弱表达能显著增强人机共情，但这一促进作用存在维度特异性与方式差异性。即脆弱表达对人机情感共情的提升效果显著，但对人机认知共情的增强作用不显著。在三种具体的脆弱表达方式中，自我披露对提升人机共情（尤其是情感共情）的效果最为显著，其效果依次优于讲故事、幽默以及非脆弱型的表达方式。同时，作为共同存在的社会临场感在聊天机器人的讲故事表达与人机共情之间起完全中介作用; 作为心理参与的社会临场感在聊天机器人的自我披露表达与人机共情之间起部分中介作用（上述变量关系详见图 6）。另外，本文也需补充警示，在人类使用 AI 聊天机器人时，应保持人的主体性，提升媒介素养，合理处理人机关系，把握好情感补偿交流的量和度，避免情感操纵与欺骗①。

1. 聊天机器人的脆弱表达可增强人机情感共情，但对人机认知共情无效

随着 Replika 等“机器人同伴”的兴起与流行，如何使聊天机器人具备共情表达能力成为人工智能领域的重要课题。利巴拉蒂（Liberati）等学者提出，赋予聊天机器人脆弱性这一人性化特质，是其被社会广泛接纳为具有一定人格智慧体的关键②。本研究为此观点提供了有力的实证支持: 聊天机器人的脆弱性特征在增强人机共情方面的效果显著优于非脆弱性表现。

本研究进一步将人机共情解构为认知与情感两个维度，发现脆弱性特征主要影响人机情感共情，即人类对聊天机器人能否对自身经历或情绪产生共鸣的能力感知; 而对人机认知共情，即人类对聊天机器人是否具备换位思考与理解信息能力的感知，影响则不显著。这一发现符合基本逻辑，因为脆弱表达在内容上本质是传递情感信号与主观态度，而非纯粹的逻辑推理，因此其更直接地作用于情感共鸣层面。

① 朱贺 《情感补偿机制下的社交机器人伦理问题》，《青年记者》2023 年第 10 期。

② Liberati, N., & Nagataki, S., “Vulnerability under the Gaze of Robots: Relations among Humans and Robots,” *AI & Society*, vol. 34, 2019, pp. 333 − 342.

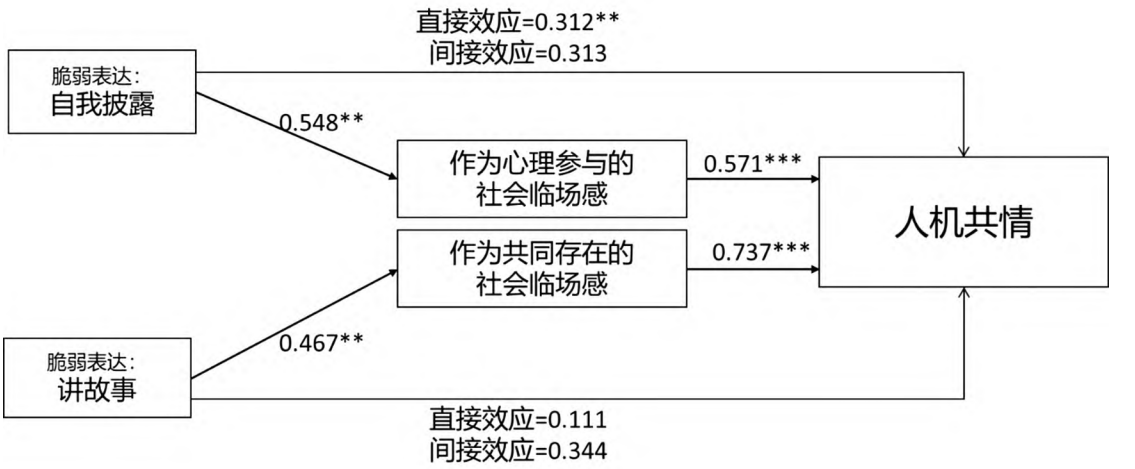

图 6 研究变量关系图

* $p<0.10$，** $p<0.05$，*** $p<0.001$。

值得关注的是，表 4 数据显示，人机认知共情的平均得分（3.651）普遍且显著高于人机情感共情（2.541）。这一现象表明，无论聊天机器人采用何种表达方式，用户对其在识别、理解信息等认知功能方面的评价已处于较高水平，这或许源于当前大语言模型在逻辑处理上的卓越表现。该结果与实践观察相呼应，如弗拉维奥·卡尔沃（Flavio Calvo）所指出的，聊天机器人在理解某些心理机制方面（即认知共情）可能已展现出较强能力，但在情感体验与共鸣层面（即情感共情）仍存在瓶颈①。

因此，总结而言，未来的学术研究与实践开发应当更加深入地探讨，除脆弱性表达之外，如何通过其他不同的话语策略与技术路径，更有效地激发人类对聊天机器人的情感反应，攻克人机情感共情这一核心难点。这将对优化聊天机器人在心理健康支持、社交辅助等深度交互领域的应用具有至关重要的意义。

2. 聊天机器人自我披露式脆弱表达有利于塑造“真人”般交互体验

自我披露，指个体向他人透露个人信息的行为，在人际传播中已被证实能带来减轻心理压力、改善心理状态等多种积极效应②。克雷布斯（Krebs）的“熟悉－共情”假设认为，对话主体间通过自我披露所增进的相互熟悉度，是激发共情的重要条件③。本研究将这一假设置于人机交互语境下进行检验，结果显示，聊天机器人通过自我披露式脆弱表达回应用户时，能有效促进人机共情。

这一发现为 CASA 范式及其等效性假设提供了新的支持证据。它与 Ho 等人的研究结论相呼应，表明人们在与聊天机器人互动时，会在心理上无意识地启用人际交往的图式，对机器人

① 参考消息 《AI 当心理医生？可能缺乏共情能力》，2023 年 5 月 12 日，https://www.cankaoxiaoxi.com/#/detailsPage/%20/eca96abbe5094e51ba62c5c891b410a2/1/2023-05-12%2014:39?childrenAlias=undefined，2024 年 11 月 16 日。

② Martins, M. V., Peterson, B. D., Costa, P, et al., “Interactive Effects of Social Support and Disclosure on Fertility-related Stress,” *Journal of Social and Personal Relationships*, vol. 30, no. 4, 2013, pp. 371－388.

③ Krebs, Dennis, “Empathy and Altruism,” *Journal of Personality and Social Psychology*, vol. 32, no. 6, 1975, pp. 1134－1146.

的自我披露产生类似于对人类的社会反应，从而在感知理解、关系亲密度和认知评估上引发等效的心理过程[①]。

此外，本研究发现作为心理参与的社会临场感在聊天机器人自我披露与人机共情之间起到部分中介作用。该维度关注用户对智能体（如聊天机器人）的身体感知，以及人机关系的显著性、亲密性、直接性和相互理解程度[②]。本研究量表聚焦于亲密性维度，发现聊天机器人的自我披露能显著提升人类对其亲密性感知，增强人类的心理参与程度，进而提升人机共情水平。这一发现也与孟（Meng）等人关于人机互惠性自我披露的研究形成对话。他们指出，若聊天机器人仅进行单向的自我披露而未能提供相应的情感支持，其交互效果可能比完全不回应更差[③]。这共同验证了 CASA 范式的深层合理性：在设计聊天机器人的自我披露时，不应仅限于事实层面的信息披露，更应策略性地融入类似于人类自我披露时所包含的情感支持元素，如此才能营造出更加真实、丰富且有深度的人机交互体验。

基于此，未来的研究拥有明确的推进方向。自我披露的表达方式多种多样，本研究设计的仅是其中一种有效路径。后续研究可系统探索不同深度、不同主题、不同情感色彩的自我披露策略，并进一步分析其与人机共情各维度的关系，以及除社会临场感外，是否存在如信任、感知相似性等其他关键的中介变量。通过系统性地揭示人机共情的产生机制，将为聊天机器人技术的情感智能化发展提供坚实的理论指引。

3. 聊天机器人讲故事式脆弱表达为交互对象提供强烈意识共在感

故事是对事件序列的描述与记录，它不仅是信息的载体，更是赋予事实以意义和情感色彩的重要手段。美国作家奥布莱恩（O'Brien）指出，讲故事是人类不可或缺的文化活动，尤其在面临困境时，其重要性愈发凸显。作为一种共情心理培养工具，讲故事能够促进个体从多角度理解他人经历。本研究在 CASA 范式指导下，证实了聊天机器人的讲故事表达在提升人机共情方面的有效性，尽管其效果略逊于自我披露。

本研究还揭示了其独特的作用路径：作为共同存在的社会临场感在聊天机器人讲故事与人机共情之间起到完全中介作用。这一维度强调的是一种尽管存在物理距离，但仍能感受到与对方“处于同一地方”的相互意识与注意力共享[④]。本研究量表聚焦于人机相互意识维度，发现聊天机器人的讲故事表达能够显著提高人类对其关注程度，增强人机共在感知，进而提升人机共情水平。该发现与教育领域的相关研究相互印证。例如里德（Reed）等人通过使用生成式人工智能创建图像向学生讲故事的研究，也同样发现这种方式能够显著提升学生的参与度、真实感和情感影响，这与本研究中的“共同存在”感在概念上高度相通[⑤]。

---

① Ho, A., Hancock, J., & Miner, A. S., "Psychological, Relational, and Emotional Effects of Self-disclosure after Conversations with a Chatbot," *Journal of Communication*, vol. 68, no. 4, 2018, pp. 712 – 733.

② Biocca, F., Harms, C., & Burgoon, J. K., "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria," *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 5, 2003, pp. 456 – 480.

③ Meng, J., & Dai, Y., "Emotional Support from AI Chatbots: Should a Supportive Partner Self-disclose or not?," *Journal of Computer-Mediated Communication*, vol. 26, no. 4, 2021, pp. 207 – 222.

④ Biocca, F., Harms, C., & Burgoon, J. K., "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria," *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 5, 2003, pp. 456 – 480.

⑤ Reed, Janet, M., & Tracy, M. Dodson, "Generative AI Backstories for Simulation Preparation," *Nurse Educator*, vol. 49, no. 4, 2024, pp. 184 – 188.

对于其内在原因，可以从叙事学理论入手进行分析。艾丝卡拉（Escalas）指出，故事所具有的完整情节结构（如开端、发展、高潮、结局）和叙事模式，对于信息的记忆和共情的产生至关重要①。在本研究情境下，聊天机器人通过起承转合的故事结构，为用户构建了一个沉浸式的叙事空间，这种结构化的情感信息传递方式，有效地营造了一种仿佛与用户共同经历事件的在场体验，从而拉近了心理距离。

当然，讲故事的表达方式本身具有高度的多样性，包括叙事形式、篇幅长短、故事内容与用户自身的心理距离等变量，都可能调节其最终效果。在未来的聊天机器人开发中，有意识地提升其叙事能力，将其作为构建社会临场感的重要工具，将有助于显著缩小人机间的心理距离，优化互动反馈质量。

4. 聊天机器人幽默式脆弱表达在促进人机共情方面效果相对较弱

幽默作为一种常见的人际交流手段，通常表现为自嘲或开玩笑等形式，其在建立和谐人际关系中的作用已被广泛认可②。研究显示，擅长运用幽默进行情绪调节的个体往往具有较强的共情能力③。在人机交互领域，吕（Lv）等研究者发现，当生成式人工智能在服务故障期间采用幽默语言风格而非理性陈述时，能够更有效地提升用户的宽容度④。然而，在本研究探讨脆弱表达对共情的促进作用时，结果显示幽默表达的效果不如自我披露和讲故事。此外，社会临场感的两个维度在幽默表达与共情之间均未发挥显著的中介作用。

这一发现并非完全否定幽默的价值，而是引导我们更深入地反思其效果的情境依赖性与类型特异性。这一结果实际上从侧面进一步验证了 CASA 范式的复杂性。汉普斯（Hampes）指出，幽默与共情之间的关系可能受到幽默类型的影响⑤。马丁（Martin）等人根据“对自己或对他人”以及“善良或恶意”的维度，将幽默分为亲和幽默、自我提升幽默、攻击幽默和自我贬低幽默四种风格⑥。

本研究认为，效果有限的一个潜在原因在于实验所诱发的情境与用户情绪状态。在本研究中，被试被诱发了较低落的情绪状态，在此背景下，他们可能普遍缺乏足够的心力去理解和欣赏幽默。此时，若聊天机器人使用幽默进行回应，可能会被部分用户感知为对当前严肃情绪的不尊重、回避问题，甚至是一种轻浮，从而导致误解和情感上的疏离。

因此，本研究建议，在未来聊天机器人的开发中，应对幽默表达的适用场景进行审慎的评估与设计。不应在所有情境下都采用单一、固定的幽默策略，而应借鉴人际沟通的复杂性，根

---

① Escalas, J. E., “Imagine Yourself in the Product: Mental Simulation, Narrative Transportation, and Persuasion,” *Journal of Advertising*, vol. 33, no. 2, 2004, pp. 37 – 48.

② Gkorezis, Panagiotis, & Victoria Bellou, “The Relationship between Leader Self-deprecating Humor and Perceived Effectiveness: Trust in Leader as a Mediator,” *Leadership & Organization Development Journal*, vol. 37, no. 7, 2016, pp. 882 – 898.

③ Hampes, & William, P., “The Relation between Humor Styles and Empathy,” *Europeś Journal of Psychology*, vol. 6, no. 3, 2010, pp. 34 – 45.

④ Lv, Dong, et al., “Language Styles, Recovery Strategies and Users´Willingness to Forgive in Generative Artificial Intelligence Service Recovery: A Mixed Study,” *Systems*, vol. 12, no. 10, 2024, p. 430.

⑤ Hampes, & William, P., “Relation between Humor and Empathic Concern,” *Psychological Reports*, vol. 88, no. 1, 2001, pp. 241 – 244.

⑥ Martin, R. A., “Sense of Humor,” in Lopez, S. J., & Snyder, C. R., eds., *Positive Psychological Assessment: A Handbook of Models and Measures*, Washington, DC: American Psychological Association, 2003, pp. 313 – 326.

据不同的用户情绪状态、对话上下文背景，让聊天机器人具备判断并调用多样化沟通策略（包括不同类型的幽默等）的能力，如此才能更稳健地促进良好人机关系的建立。

（二）理论与实践意义

1. 理论启示

首先，本研究证实并延展了 CASA 范式。实验结果明确显示，用户会对聊天机器人的脆弱表达产生类社会性的共情反应，这为“计算机作为社会行动者”范式提供了新的实证支持。更重要的是，研究通过揭示社会临场感的双路径中介机制，打开了 CASA 范式作用过程的黑箱，从验证有效性推进到阐释作用机理。同时，研究发现幽默表达效果的局限性以及脆弱表达对认知共情作用的非显著性，明确了该范式的效应边界，提示未来研究需要关注交互策略的情境适用性与维度特异性。

其次，本研究构建并验证了“脆弱表达 - 社会临场感 - 人机共情”的理论模型。通过将脆弱性这一哲学概念操作化为自我披露、讲故事和幽默三种具体形式，并整合社会临场感的双维度结构，建立了一个完整的并行链式中介模型（见图 6）。这一理论模型的建立，不仅为理解多个变量间的复杂关系提供了系统性视角，还揭示了不同交互策略通过差异化路径影响共情的机制，为后续研究提供了可检验的理论框架和方法论示范。

2. 实践意义

首先，在整体设计理念上，本研究启示开发者应该从“人类感知”视角出发，将设计重点放在优化机器人的外显交互行为上，通过符合人类社会预期的言行举止来触发用户积极的社会认知与情感反应。具体而言，将脆弱表达作为一项核心设计策略被证明是可以有效促进用户的共情感知的。

其次，在具体策略应用上，本研究为三种脆弱表达方式提供了差异化指南：自我披露应作为激发深度共情的优先策略，在设计时不仅要进行事实层面的信息揭露，更要融入情感元素与支持性内容，以促进深度心理参与；讲故事是营造陪伴感与共同存在的有效工具，开发中应重视提升聊天机器人的叙事能力，利用故事结构构建共享情境；幽默的使用则需要高度审慎，必须考虑情境适用性，根据用户情绪状态与对话上下文判断是否启用及选择何种幽默风格，避免在不合时宜的场合造成误解。

总之，本研究通过理论模型的构建与验证，不仅深化了对人机共情产生机制的理论理解，更直接转化为一套清晰的设计原则与行动指南，对推动聊天机器人在心理健康、情感陪伴等需要高情感投入领域的可靠应用具有重要的现实价值。

（三）研究局限与展望

作为一项探索性研究，本文存在若干局限性，具体如下：第一，在控制变量方面，本研究仅纳入性别、年龄、教育程度、收入水平及对技术的先验态度五个变量。鉴于共情机制的复杂性和个体情绪触发的多样性，后续研究应当扩展控制变量的范围，如纳入人格特质、对聊天机器人的心智感知等，以全面剖析人机共情的形成机制。第二，在研究方法上，本文主要采用在线实验法和问卷调查法，通过被试的自报告来衡量其对人机共情的感知，这可能受到主观感知偏差影响。未来研究可以结合实验室实验和 ERP、EEG 等认知神经科学技术，以更客观地测量被试的共情感知程度。第三，在实验材料设计方面，本文参考斯特罗科布和拉格尔的研究案例，设计自我披露、讲故事和幽默脆弱表达的首句回复语句。然而，在现实生活中，自我披

露、讲故事和幽默这三类脆弱表达更为多样，未来研究应探讨不同表达方式对结果的影响及其机制。第四，在样本来源方面，本研究被试均来自北京同一高校，样本在年龄、教育背景等方面同质性较高。这一设计虽有助于控制混杂变量、保障实验的内部效度，但也不可避免地限制了结论的外部效度，其发现推广至更广泛群体时需保持审慎。此外，样本中女性比例偏高，尽管组间性别分布无显著差异，在一定程度上保障了组间可比性，但整体性别结构仍可能对共情水平的总体评估造成影响。未来研究可进一步拓展样本的地理、文化与人口学多样性，并在设计阶段系统控制性别比例，以增强结果的稳健性与普适性。

此外，本文在撰写过程中还发现两个未来研究方向：第一，在研究场景的选择上，本文关注学业或事业不顺这一普遍脆弱性场景。未来可考虑某一具体的生理或心理疾病等特殊脆弱性场景，评估聊天机器人表达方式的影响差异。第二，在研究的未来展望中，本文欲提出一项深层次问题，即人机共情的发展是否会影响人类共情机制的变化。因为将脆弱性这一特质赋予聊天机器人并非一个简单的中性操作，它使得聊天机器人不再仅仅是工具，而是触及到人类主体构成的深层次变革。也就是说，通过设计聊天机器人的脆弱性，我们的社会实际上也正在塑造自己[①]。因此，未来我们还需对此进行持续的观察和分析，以期为人工智能与人类情感互动的伦理和实用边界提供理论依据。

注：本研究的实验对话脚本、完整问卷及原始数据可见 OSF 预注册链接：https://doi.org/10.17605/OSF.IO/EFVJM

本文系中央高校基本科研业务费专项资金（项目编号：1243200001）资助项目的阶段性研究成果。

作者：北京师范大学新闻传播学院教授、博士生导师
北京师范大学新闻传播学院博士生（通讯作者）
北京师范大学新闻传播学院博士生

① Liberati，N.，& Nagataki，S.，“Vulnerability under the Gaze of Robots：Relations among Humans and Robots，”*AI & Society*，vol. 34，2019，pp. 333－342.

actively respond to the ideological interpellation of rural revitalization through mission-driven entrepreneurship, the pursuit of a better life and the transcendence of family responsibilities. However, challenges including platform traffic competition, local social norms and identity conflicts lead them to misinterpellation phenomena, manifesting as self-doubt, disingenuous participation and burnout-driven withdrawal. The paper introduces the concept of remainder interpellation referring to subjective elements that are not fully incorporated into the ideological interpellation process, which serve as potential resources for subjectivity reconstruction. It develops Althusser's interpellation theory, highlighting unconquered remainders and offering a new perspective on mutual construction of state presence and individual agency in short video platforms in rural areas.

54 · Does Expression of Vulnerability by Chatbots Enhance Artificial Empathy? An Empirical Analysis Based on an Autonomously Developed Chatbot

· *Zhou Min, Zhao Xiuli, Chen Feiyang*

Although artificial empathy is an important foundation for good human-robot relationships, how to empower chatbots with empathic expression remains unclear. Given that the mutual awareness of human vulnerability is a cornerstone of interpersonal empathy, this study used an experimental method to explore the effects of chatbots' expressions of vulnerability on artificial empathy by controlling the expressions of chatbots (vulnerability: self-disclosure group/storytelling group/humor group; non-vulnerability: control group). The study found that: (1) Chatbots expressing vulnerability significantly enhanced artificial emotional empathy compared to those expressing non-vulnerability. However, they showed no significant difference in enhancing artificial cognitive empathy. (2) When chatbots used self-disclosure expressions of vulnerability, they significantly outperformed storytelling, humor, and non-vulnerable expressions in increasing the degree of artificial empathy, in that order. (3) Social presence as psychological engagement partially mediated between chatbot self-disclosure expressions and artificial empathy. Social presence as co-presence fully mediated the relationship between chatbot storytelling expressions and artificial empathy. At the theoretical level, this article provides an empirical expansion of artificial empathy and constructs a relationship model of "vulnerability expression-social presence-artificial empathy." At the practical level, it offers scientific guidance for developing empathic expression in future chatbots, optimizing the quality and experience of human-robot interaction.

77 · Verisimilitude: Information Manipulation in Impersonation-Based Online Fraud

· *Zhang Jianjun, Cao Canqiong*

Focusing on the growing social problem of online fraud involving acquaintance impersonation, this study draws upon an analysis of 33 judicial verdicts from China Judgments Online and interviews with 17 victims to propose the concept of Verisimilitudinous Information Manipulation (VIM). This concept aims to reveal new features of information manipulation in the digital era. The findings show that the online fraud process unfolds in three phases: orientation, evaluation, and control. In the orientation phase, fraudsters appropriate and impersonate online identities to mislead victims' cognition of social relationships. In the evaluation phase, scripted narratives are deployed to bolster situational credibility, manipulating victims' cognition and emotions. In the control phase, fraudsters manipulate the decision-making environment, fabricate behavioral evidence, and orchestrate social scenarios to control victims' decision-making behavior. VIM is defined as a systematic manipulation of cognition, emotion, and behavior in cyberspace, enabled by digital tools, and comprises three types: identity verisimilitude, context verisimilitude, and behavior verisimilitude. This concept transcends the singular focus of traditional information manipulation theories on information content,