

**《数字时代的科学与技术》科研训练项目表**

**（2024版）**

项目名称：AI数字化角色在心理健康支持中的可行性与伦理挑战研究

项目负责人： 　　 牛子豪

项目成员： 樊珂鸣，王诗蕊，谭清灿，黄子谦

项目负责人电话： 　　　18118132223

指导老师： 王希廷

指导老师所在单位： 高瓴人工智能学院

**中国人民大学明理书院**

**填表日期：　 2024　年　 12　月　06 　日**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 项目名称 | | AI数字化角色在心理健康支持中的可行性与伦理挑战研究 | | | |
| 课程主题及主讲教师 | | 《拥抱人工智能时代》  窦志成 | | | |
| 学生　科研  团队 | 姓名及学号 | | 专业 | 项目贡献 | |
| 樊珂鸣  2021202188 | | 人工智能 | 积极参与小组讨论，贡献了关于AI数字化角色在心理健康支持中的正面作用的见解，并协助完善项目表，确保信息的准确性和完整性。深入分析了AI伴侣、AI亲人、AI咨询师和AI治疗师在不同场景中的具体应用和潜在益处，为项目的整体方向和实施策略提供了有力支持。 | |
| 王诗蕊  2024201640 | | 人工智能 | 总结概括研究内容，设计了一个简化的AI角色模型，包括情感识别和个性化反馈模块，为进一步研究奠定基础。通过实验和案例分析，证明AI数字化角色可以有效地为用户提供心理健康支持，如情绪管理和压力缓解。针对隐私保护和算法偏见等常见问题，明确了需要解决的伦理挑战。提出AI角色在心理健康领域未来发展的重点，如优化用户体验和加强数据安全。 | |
| 谭清灿  2024201569 | | 信息管理与信息系统 | 关键技术问题探索，政策和法规评估，提出解决方法，案例分析，哲学和伦理学的探讨，参与小组讨论和问题检查，提供报告，AI数字人的实际开发和应用提供了指导，帮助实践者在设计和实施过程中考虑到伦理问题。 | |
| 黄子谦  2024201641 | | 人工智能 | 进行数据分析，撰写伦理分析部分报告，项目表的完善与填写，与其他小组成员、老师和外部资源进行联系，积极参与讨论，提供支持和鼓励，帮助建立一个积极的工作环境。 | |
| 牛子豪  2024201900 | | 数学学院 | 对AI数字化角色在心理健康支持中的可行性的分析负责设计问卷或访谈大纲，收集有关公众对AI心理健康支持服务的看法和需求的数据。进行数据分析并形成报告，提供数据支持的小组决策，帮助团队了解目标用户的需求。 | |
| 指导  教师 | 姓名 | | 王希廷 | 职称 | 准聘副教授 |
| 电话 | | 13810366562 | E-mail | xitingwang@ruc.edu.cn |
| 研究方向 | | 大模型解释和对齐 | | |
| **一.想法来源** | | | | | |
| 随着科技的不断发展，AI在各个领域都有非常广泛的应用，特别是在心理健康的支持方面。AI数字化角色（例如聊天机器人和虚拟助手）开始被用于提供心理咨询、情感支持以及相关的健康建议。在如今生活节奏日益加快的当今社会，心理健康的问题越来越受到全体社会人士的重视，如何利用新技术来有效地提供心理支持，逐渐成为了一个紧迫的话题，而在这一领域的研究中，不仅要关注技术的可行性，其背后蕴藏的伦理挑战也值得我们深思。 | | | | | |
| **二.研究问题介绍** | | | | | |
| 1. 基本内容：   运用自然语言处理（NLP），情感计算，深度学习模型等核心技术，探讨AI数字化角色的可行性，并结合AI心理健康支持产品与数字化角色应用实例，进行伦理挑战分析，最后从AI在各方面应用的未来展望，为构建负责任的AI心理健康支持体系提供了理论基础和实践方向。   1. 研究方法：   基于法律的规范性和可行性的情况下，以为人类提供价值为中心，对用户开展情绪方面的价值调查和对AI数字人的模型展开实验。因此，我们可以结合新兴的人工智能数字人模型，具体分析AI的语言逻辑性和情绪性的感知，以及用户对AI数字人的交互评价，结合具体案例，如医疗、家庭、教育等领域，探索数字人可以为我们带来那些便利，又会带来那些边缘违法行为的隐患，进行思考改进，并给出应对策略。并从哲学和伦理学的角度出发，探讨人工智能是否具备情感和道德判断能力，以及这一点是否在伦理层面会引发了新的问题。制定AI心理健康应用的伦理准则，明确责任边界。引入第三方监督，确保AI产品的公平性和透明性。  （三）AI数字化角色的正面效益与研究价值：  AI数字化角色在心理健康支持中具有巨大潜力，能够提供个性化、即时性和持续性的支持。根据特定的场景，我们分析了AI数字化角色带来的正面效果：   1. AI伴侣:   AI伴侣可以提供无条件的情感支持和陪伴，帮助用户缓解孤独感，调节情绪，引导健康行为，并在危机时刻提供干预。   1. AI亲人:   用AI模拟已故亲人的特征，可以为用户提供情感寄托，保存珍贵回忆，给予生活指导，帮助用户缓解丧亲之痛。   1. AI咨询师:   AI咨询师可以提供即时、个性化的在线心理支持，分析用户数据以识别潜  在问题，降低心理咨询成本，让更多人获得帮助。  （四）伦理挑战分析与展望：  我们发现，AI智能体通过分析用户习惯，可以形成“过滤气泡”，即向用户推荐与其观点一致的内容，或者在交流过程中不断模仿用户思路和语言风格，某种程度上可能导致用户陷入信息茧房，用户在这种交流中不仅难以接触不同的观点和信息，在数字化角色错误的情感支持下，产生自己是大多数的错觉。同时，AI智能体的决策和行为往往基于算法模型，而算法模型的设计过程中不可避免地存在人的主观意识，对境外敌对势力而言可能利用这一机制进行舆论引导和意识形态渗透，从而对社会稳定和国家安全构成威胁。此外，算法数据来源也可能存在歧视和偏见，在用户与这种数字化角色的交流中，这种歧视会给用户带来不好的体验与错误的价值导向，甚至导致社会不平等现象加剧。  除此之外，AI智能体在提供服务的过程中，往往会收集和处理用户的个人信息，这些信息包括但不限于身份信息、浏览历史、位置信息、购买记录等。这些数据为AI提供了训练基础，但也带来了用户隐私暴露的风险。因此，AI智能体的信息保护至关重要，我们认为，在AI智能体信息保护中，责任方的明确十分关键：   1. 制造商与开发者：   制造商和开发者在AI智能体的设计和生产过程中，应承担起产品质量与安全责任。他们需要确保产品在正常使用条件下不会泄露用户隐私，并为用户提供清晰的隐私声明和告知，采取一系列的技术和管理措施，如数据最小化原则、去识别化和匿名化处理、加密处理等。数据最小化原则要求AI系统仅收集完成任务所需的最少数据，避免不必要的数据收集。去识别化和匿名化处理则通过技术手段移除个人数据中的识别信息，使得数据无法与特定个人直接关联。加密处理则是在数据传输和存储过程中，使用加密技术保护数据的机密性。   1. 应用方与平台方：   应用方和平台方在使用AI智能体时，应遵守相关法律法规，确保数据的合法收集和使用。他们需要对数据进行妥善保护，防止泄露和滥用。   1. 用户：   作为最终使用者，用户也应承担一定的责任。他们需要提高自我保护意识，了解并同意数据的使用方式和潜在风险。同时，用户还应妥善保管自己的个人信息，如仔细阅读隐私政策、谨慎授予权限、定期清理浏览记录和缓存信息、使用强密码和双重验证等，避免泄露给不法分子。 | | | | | |
| **三.可行性及团队分工** | | | | | |
| 一.可行性：  （一）.国家层面  许多国家已经制定了关于心理健康的法律法规，政策导向的支持为AI在心理健康支持中的应用提供更广泛的认可和支持  （二）.技术本身   1. 便利性：   AI数字化角色可以突破时间和空间的界限，尤其是对因为时间、经济等原因而无法接受传统心理治疗的人来说更具有吸引力。   1. 匿名性和舒适性：   AI所提供的服务往往是匿名的，很多人对于面对面咨询存在一定的羞耻而不安，而通过AI进行交流可以在一定程度上带来更大的心理舒缓   1. 数据分析与个性化服务   AI能够分析用户的交互数据，从而提供更为个性化的建议与支持，数据驱动的服务可以提高支持的精准度。  二.团队分工：  1.牛子豪（小组长）：负责整体协调，确保小组讨论有序进行，制定讨论议程，编制小组预算，确定小组目标，并负责AI数字化角色中的理论可行性的探究   1. 谭清灿（文献研究员）：   负责收集和分析相关的文献资料，总结AI在心理健康支持中的应用现状。   1. 王诗蕊（数据分析师）：   设计实验，分析AI角色支持的实际效果。设计问卷，收集与分析公众对AI心理健康支持服务的看法，负责数据的收集、处理和分析。收集用户情感反馈和心理健康数据，并发现模型中的潜在问题（如偏见、误差等）。   1. 黄子谦（伦理与社会研究员）：   职责：负责伦理挑战的识别与对策研究。  具体任务：研究隐私保护、算法公平性和伦理争议，提出解决方案并编写伦理合规指南。调研AI智能体潜在的信息保护风险并分析责任分担问题。   1. 樊珂鸣（技术支持）：   负责AI数字化角色的模型开发与实现。包括搭建AI角色模型，包括情感识别和响应生成模块；进行模型训练、测试和优化；开发和集成隐私保护机制。 | | | | | |
| **四.学生签名** | | | | | |
| **项目组负责人： 牛子豪**  **项目组其他成员：王诗蕊 黄子谦 谭清灿 樊珂鸣**  **2024 年　12　月06　　日** | | | | | |
| **五.指导教师意见** | | | | | |
| 学生科研团队经过独立思考，聚焦于AI数字化角色这一热点议题，具备广泛社会价值及科研创新潜力。团队进行了合理分工，从文献调研、数据分析、伦理问题探讨、技术细节分析等角度对该科研问题进行了拆解，具备可行性，可考虑作为未来大创或小创项目选题。未来如果有机会实现该选题，可以考虑从如下几个方面进行深入研究：1）结合社会热点问题，进行深入访谈与社会调研，选定具备社会影响力的具体任务（如青少年心理问题干预），深化社会价值；2）利用现有大语言模型进行具体角色扮演，通过调整提示词、模型训练等方式进行初步实验结果收集与分析；3）剖析获得的结论及其社会价值，结合实验结论，探讨对社会、国家政策的可能建议。  **签名：**  **年　　月　　日** | | | | | |
| **六.明理书院创新实验室考核结果** | | | | | |
| **盖章：**  **年　　月　　日** | | | | | |