**Enhancing UX Evaluation Through Collaboration with Conversational AI Assistants: Effects of Proactive Dialogue and Timing**

**读书报告**

**1.引言**

**1.1.研究背景**

随着数字化产品的普及，用户体验（UX）的重要性日益凸显。可用性测试作为提升UX的关键方法，通过分析用户在交互系统中的行为来识别潜在问题。然而，分析可用性测试视频是一项复杂且资源密集的任务，需要评估者同时观察用户行为和音频信号，并迅速识别多个任务中的可用性问题。在工业环境中，时间和资源的限制可能导致关键信息的遗漏或误解。尽管多人协作分析可以提高分析的可靠性和完整性，但协作成本较高，这在一定程度上阻碍了UX评估者采用这一方法。鉴于AI技术的快速发展，研究者开始探索如何利用AI驱动的分析为UX评估者提供补充视角，特别是通过自然语言接口实现的人机协作分析成为研究热点。

**1.2.研究目的和重要性**

本研究聚焦于主动式对话AI助手在UX评估中的应用，探讨其自动建议在不同时间节点（问题发生前、同步、问题发生后）对UX评估者分析行为的影响。研究此问题的重要性在于，理解最佳建议时机将指导未来工具的设计和功能，从而增强AI辅助的UX评估决策。

**2.相关研究**

**2.1.使用AI检测可用性问题的研究**

**2.1.1.自动化方法：**

自动化方法包括机器学习、模式识别、音频和视频分析以及自然语言处理。这些方法在一定程度上减轻了人工分析的负担.

但自动化方法的局限性在于难以完全替代人类评估者的直觉和上下文理解能力。ChatGPT在识别可用性问题方面的有效性仍存在认识不足。我们利用 ChatGPT 生成可用性问题建议的策略需要对其输出的质量进行评估。

**2.1.2.人机协作方法：**

AI通过算法循环（“algorithm-in-the-loop”）过程向人类提供信息，支持而非取代人类决策。

然而，现有的UX评估工具主要提供非交互式可视化，限制了评估者提问和寻求解释的能力。

**2.2.主动式对话的研究**

主动式对话可以提醒用户注意可能遗漏的信息，并提供建议以辅助决策。在多个领域（如教育、健康管理和决策支持）中，主动式对话已被证明能提高用户满意度和信任度。

但是，不恰当的主动交互可能导致用户分心或产生不信任感。因此，需要谨慎设计交互时机和内容，确保消息时机合适且内容合理。

**3.研究方法**

常用的研究方法中，本文用了，实验法、调查法和文献调研法，这里主要对前两者的适用性和局限性讨论，实验法中有AI自动生成问题建议方法和Wizard of Oz方法：

前者适用于需要快速生成大量问题建议的场景，如用户体验（UX）研究中的可用性问题识别。能够处理自然语言文本，分析转录内容，为研究人员提供潜在的问题点。但是GPT只能访问用户的口头内容，无法获取视频中的视觉信息，可能导致建议不够准确或完整。而且对复杂或语境依赖的问题可能识别困难，需要人工校对和编辑。

后者适用于需要处理参与者即兴问题的研究，如用户体验研究中的实时交互。能够确保响应的一致性，专注于调查特定变量（如自动建议的时机）的影响。但是引入了人为干预因素，可能影响研究的自然性。并且对主持人的要求较高，需要预设脚本和分类参与者可能提出的问题。而且实验偶然性大（因为实验的基数，或者说人太少了）。

实验设计上三个方法的流程都很严谨，收集了大量数据。所有的指标都有定量的计算而且是基于前人的理论改进，考虑到前人忽略的变量，针对性强。但是参与者样本量相对较小（24名参与者），可能影响结果的普适性。而且他自己也说由于在时间、资源和精力方面存在相关成本，评估人员对协作实践的采用有限。因此，未来的努力应致力于在效率和稳健性之间取得平衡。

**4.实验结果**

**4.1.已有的实验结果和发现**

研究发现，建议出现的时间对评估者识别的问题数量没有显著影响，但问题发生后的建议显著提高了信任度和效率。大多数评估者偏好问题发生后的建议，因为这有助于验证他们的分析。

另外，评估者普遍认同ChatGPT生成的建议，但认为这些建议不够全面。对于高一致性的建议，评估者倾向于直接确认；对于低一致性的建议，评估者会寻求澄清或忽略。

尽管ChatGPT能够识别一些可用性问题，但它错过了评估者识别的大部分问题（58.8%）。这主要是由于ChatGPT仅依赖于用户口述的文本信息，缺乏视频中的视觉和交互信息。

**4.2.影响以及讨论**

本研究结果强调了人机协作在UX评估中的重要性，特别是在利用AI辅助提高评估效率和可靠性方面。同时，研究结果也指出了当前AI工具（如ChatGPT）在UX评估中的局限性，提示未来研究需要关注如何结合多模态数据（如视频和音频）来提高AI的分析能力。

研究者通过实证研究验证了主动式对话AI助手在UX评估中的潜力，特别是在建议时机对评估者行为的影响方面提供了有价值的见解。然而，研究也揭示了当前AI工具的局限性，特别是它们在处理复杂UX评估任务时的不足。未来研究应进一步探索如何结合多模态数据来增强AI的分析能力，并开发更加智能化和用户友好的人机协作工具。

**4.3.未来可能的发展方向和挑战**

根据这篇研究，未来可以在以下方面进行研究：

（1）多模态数据分析：结合视频、音频和文本数据，利用深度学习技术提高AI对UX问题的识别能力。

（2）更个性化的建议：根据评估者的经验和偏好，提供个性化的建议时机和内容，以提高协作效率。

（3）考虑可扩展设计：适应更长的可用性测试视频

（4）从效率和稳健性中取得平衡：为AI建议提供详细的解释，帮助评估者理解建议背后的原因，从而增强信任度。提供对评估者的培训，帮助他们更好地理解和利用AI建议，并收集反馈以不断优化工具。

**5.感悟**

除了ai本身技术的问题的优良，其回答问题的时机居然也对建议的可信度有如此大的影响，这篇研究并没有为提升ai的性能直接出力，但是却从人机交互的视角给出了优化ai的一条思路。另外，这个研究的复杂度，全面程度超出了我的预料，他涵括了心理学，统计学，计算机科学等诸多学科，在前人的研究上做出诸多改进，对调查数据的详尽分析，计算让人信服。

在ai快速发展的当下，其涉及的诸多领域也在经历快速的更新，涉及的研究越来越多，人机交互是ai和人之间很重要的一环，因此ai的人机交互研究也在蓬勃发展。相信将来人工智能和人之间的交互将能达到人和人之间交互的水平，甚至超过。