人工智能医疗器械创新合作平台
ARTIFICIAL INTELLIGENT MEDICAL DEVICE INNOVATION AND COOPERATION PLATFORM

数字疗法人因工程
质量技术要求

人工智能医疗器械创新合作平台数字疗法工作组

2025 年 10 月

目录

一、评价角度 1
1）有效性 2
2）使用效率 2
3）满意度 2
二、评价类型 3
1）形成新测试 3
三、评价内容 3
1）用户界面 4
2）交互设计/交互流程 4
3）内容和可读性 4
4）辅助功能 4
5）行为和互动 4
6）架构和导航 5
7）个性化 5
四、评价结果 6
1）能力测试结果 6
2）高效性测试结果 6
3）用户体验测试结果 6
4）使用环境测试结果 6
附录 11

本文件为数字疗法产品人因工程质量评价的推荐性要求，仅用于产品设计开发和成品检验的技术参考，不具有强制性。

数字疗法产品人因工程的质量评价是产品安全有效性评价的重要组成，将从评价角度、评价类型、评价内容和评价结果四个方面展开，此外附录中提供测试过程规范举例，供相关人员参考。

一、 评价角度

分别有效性、使用效率和满意度三个方面展开

1）有效性

- 可见性 – 核心内容信息完整可见
- 可理解性 – 信息框架有层次，文案清晰易懂，逻辑清晰明确，是否容易理解
- 可供性 – 主体对效用的主观感知与技术的客观品质之间的相互作用
- 一致性 – 操作结构、界面信息、反馈等统一
- 容错性 – 对操作流程或特殊状态的避错
- 易学性 – 是否简单易用、步骤少、不复杂，第一次使用也能驾驭
- 可访问性 – 检查操作系统是否提供了辅助功能，以满足不同能力水平用户的需求
- 可记忆性 – 用户再次使用产品，是否还记得足够的

信息，以便有效地进行使用

2）使用效率

- 完成度 - 整套测试任务中的完成达标程度
- 易操作 - 在操作中，是否能够快速上手
- 费力度 - 测试任务全局或单项任务的费力度
- 完成时长 - 不同人群操作的时长比较
- 跳出率 - 因某些障碍使得用户放弃或分心
- 犯错率 - 用户犯错误的类型和严重程度，犯错率高低
- 简洁性 - 设计是否简洁，避免不必要的复杂性，让用户能够快速找到所需功能

3）满意度

- 任务满意度 - 进行一项操作任务后对任务过程的满意度
- 整体满意度 - 产品多个维度的满意度
- NPS 推荐度 - 测试使用后的推荐程度
- 体验满意度 - 交互体验、视觉体验、感官体验等多维度的体验满意度
- 实用性/功能满意度 - 功能是否实用，是否满足用户需求，包括基本功能和高级功能

二、 评价类型

分别从形成性评价和总结性评价展开

1）形成新测试

- 3~5名用户参与测试
- 半天时间测试，当天/隔天出报告
- 根据软件迭代周期，每两周进行一次测试
- 必要时可放宽招募标准，经常测试比测试实际用户更重要
- 生活场景或任何独立空间都能进行，甚至可以进行线上测试

2）总结性测试

- 30~100名用户参与测试
- 测试时间一/两周，一/两周撰写报告
- 软件完全成型后，按年/季度进行测试
- 只招募实际用户
- 需要专门的场地进行测试

三、 评价内容

分别从用户界面、交互设计/交互流程、内容和可读性、辅助功能、行为和互动、架构和导航、个性化展开

1） 用户界面

操作系统应提供直观、易用的用户界面包括布局、图标

设计、颜色使用、字体大小和选择等方面

- 是否采用了用户熟悉的界面元素和控件?
- 界面元素和控件的文字、位置、布局、分组、大小、颜色、形状等是否合理、容易识别、一致?
- 界面元素/控件之间的关系是否表达正确?
- 主要操作/阅读区域的视线是否流畅?
- 其他文本（称谓、提示语、提供反馈）是否一致?

2）交互设计/交互流程

评估用户与操作系统之间的交互是否流畅和自然

- 包括但不限于响应时间、触控精度、界面元素的响应性等
- 涉及到触摸响应、动画效果、手势识别等。

3）内容和可读性

- 文字内容的交流对象是用户吗?
- 语言是否简洁、易懂、礼貌?
- 内容表达的含义是否一致?
- 重要内容是否处于显著位置?
- 是否在需要时提供必要的信息?
- 是否有干扰视线和注意力的元素?

4）辅助功能

提供辅助功能，如语音输入、放大字体等

5）行为和互动

- 是否告知、引导用户可以做什么？
- 是否告知需要进行哪些步骤？
- 是否告知需要多少时间完成？
- 是否告知第一步做什么？
- 是否告知输入/操作限制？
- 是否有必要的系统/用户行为反馈？
- 是否允许必要的撤销操作？
- 是否页面上所有操作都必须由用户完成？
- 是否已将操作步骤、点击次数减至最少？
- 是否所有跳转都是必须的（无法在当前页面呈现）？

6）架构和导航

测试系统的导航和结构是否清晰明了，用户能否快速找到需要的功能或信息，是否存在歧义或混淆

- 是否采用了用户熟悉或容易理解的结构？
- 是否能识别当前在网站中的位置？
- 是否能清晰表达页面之间的结构？
- 是否能快速回到首页/主要页面？
- 链接名称与页面名称是否相对应？
- 当前页面的结构和布局是否清晰？

7）个性化

允许用户自定义设置，如主题、壁纸和布局等

四、 评价结果

1）能力测试结果

- 感知能力

认知能力

2）高效性测试结果

- 操作易用性

感知评估

3）用户体验测试结果

舒适度与信任度

满意度

安全性

4）使用环境测试结果

- 拓展性

环境满意度

附录：测试过程规范

（一） 测试目标

本次可用性测试旨在基于多模态人因数据，结合主客观量化该认知测评系统的可用性，具体目标如下：

1、识别使用错误风险：通过任务效能与生理指标，定位可能导致临床决策错误的界面设计缺陷。

2、评估认知工效学表现：量化系统给医护人员带来的认知负荷、视觉交互效率及情绪应激反应。

3、提供设计优化实证：将可用性问题与具体的生理特征值、界面元素关联，为迭代提供精准、可执行的依据。

（二） 测试对象与核心任务

1、测试对象：认知功能测评软件可用性测评报告-测评端

2、系统概述：本产品为认知功能测评软件，包括管理端、测评端。管理端由管理员登录，可创建并管理医生账号。测评端由医生登录，可创建并管理患者账号、完成测评任务。

3、测试重点流程：

①核心工作流：登录系统 → 搜索/新增患者 → 编辑患者信息 → 选择并开始测评任务 → （完成测评：患者端）→ 查看并解读测评报告面。

（三） 测试方法与人因数据采集

本测试采用模拟任务法，在控制环境中同步采集多模态人因数据，以实现主客观数据三角验证。

数据模态	测量设备	核心指标	测评目的
主观数据	NASA-TLX 量表、满意度问卷	主观负荷评分、满意度得分	对生理与行为数据进行补充和解 x 释。
行为数据	屏幕录制、主试观察记录、眼动轨迹图	任务成功率、任务完成时间	量化任务效能，识别行为层面的难点
生理数据	脑电	α 波（8-13Hz）功率、θ/α 功率比等	直接评估认知负荷、情绪状态、疲劳程度与信息处理深度
	眼动	注视点、注视时长、瞳孔直径变化、扫视路径、热点图	评估视觉搜索效率、界面布局合理性与注意力分配

	皮电/心率	皮肤电导反应（SCR）峰值、SCL 值、心率 LF/HF 比值	评估情绪唤醒（应激/挫败）、认知负荷、心理压力水平等

（四） 参与者招募（用户特征）

1、目标用户：在职医生、护士等系统用户。

2、样本量：30 名（满足小样本被试量），性别、年龄均衡。

3、招募标准：有、无使用同类医护系统的经验；视力或矫正视力正常，无色盲；感知、认知能力正常。

注:本次测评报告不涉及用户选择,采用 2 名被试进行测试。

（五） 测试任务设计与流程

1、任务设计需覆盖软件系统功能，并预设观察点。

任务点	引导语
登录系统	请登录账号，准备给患者进行测试
搜索/新增患者	请建立新入院患者的信息录入，并检查是否录入正确
编辑患者信息	刚才患者的年龄信息登记错误，请修改

选择并开始测评任务	请给患者设置测评任务(只有 10 分钟做任务);（第二天）继续完成别的测评任务（搜索昨天的患者）
完成测评	患者完成
查看并解读测评报告面	查看该患者的测评报告

2、测试流程

①准备阶段：准备相关设备与实验项目，准备保密协议和知情同意书。

②指导语：介绍测试目的、流程，强调“测试系统而非测试人”，营造放松氛围。

③任务执行:先采集 2min 睁眼静息数据作为基线数据使用，参与者需按主试指导语要求，按顺序完成核心任务，主持人观察并记录，必要时可以提供指导。

④数据收集：同步录制屏幕、采集主观、客观数据。

(六) 结果分析

本次测试采集参与者在完成核心任务过程中的多模态人因数据，结合主观量表(NASA-TLX、满意度问卷)与行为记录，对系统的可用性进行综合评价。

结合以上结果分析，系统在基础能力、高效性、用户体验、使用环境四大维度的核心量化指标表现均较好，能够支撑医护人员完成认知测评核心工作流。