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

精神疾病数字疗法产品人因设计分析评价研究报告

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

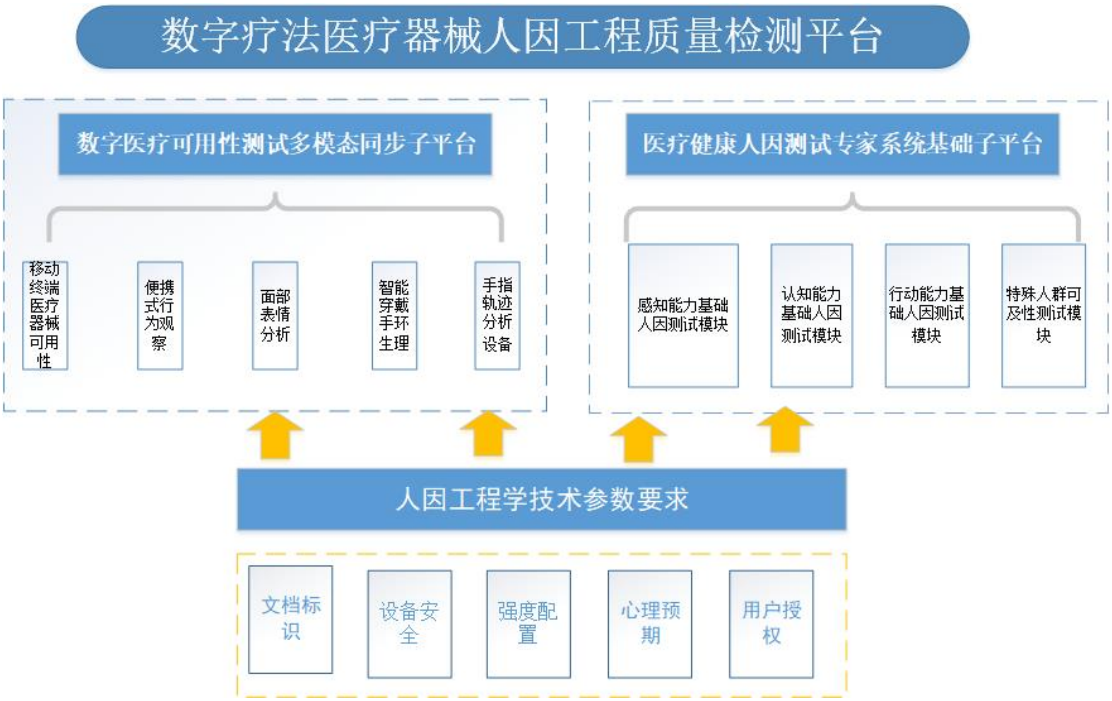
一、检测平台的构建..... 1
二、产品测试过程要求..... 6
三、测试结果和结论..... 12
四、总结与分析..... 17

本文件为北京市科技计划项目《精神疾病数字疗法产品测试验证技术服务平台构建及应用》的研究成果，可作为产品设计开发和成品检验的技术参考，不具有强制性。

数字疗法人因工程质量评价主要是通过产品的使用过程中用户的可用性角度开展研究，研究产品人因工程的设计风险和使用风险，结合构建产品使用的典型场景，从使用者角度考虑语言、年龄、认知水平、健康水平等因素，从产品设计角度考虑用户界面、输入方式、响应速度、巡航时间等，从使用配置角度考虑心理预期、依赖成瘾、剂量和强度、可及性等构建危险源识别方法，面向用户界面、文档标识、设备安全、强度配置及可及性等提出相关要求，建立人因工程质量评估模型。

一、 检测平台的构建

本课题通过构建数字疗法医疗器械人因工程质量检测平台（以下简称平台）实现课题的相关任务目标，平台主要由数字医疗可用性测试多模态同步子平台和医疗健康人因测试专家系统基础子平台组成，其中数字医疗可用性测试多模态同步子平台由移动终端医疗器械可用性测试仪、便携式行为观察仪、面部表情分析仪、智能穿戴手环生理设备组成，医疗健康人因测试专家系统基础子平台由感知能力基础人因测试模块、认知能力基础人因测试模块、行动能力基础人因测试模块、特殊人群可及性测试模块组成。具体架构如下:



具体来讲，各模块的功能如下:

数字医疗可用性测试多模态同步平台能够在实验室、虚拟现实、真实现场等环境下进行神经、心理、行为多维度数据同步采集与分析，包括眼动追踪、主观评价、生理测量、脑电、近红外脑成像、行为、面部表情等。平台采用专利时钟同步算法，保证多维度数据实时同步的精准性，为科学研究及应用提供完整的数据指标。平台覆盖完整的实验流程，包括项目管理、实验设计、同步采集、信号处理、数据分析、人工智能应用、可视化报告等，详细功能如下:

1）移动终端医疗器械可用性测试仪采用基于信息系统原型的人机交互测评技术，可对移动终端 APP 原型或桌面终端 WEB 原型进行以用户为中心的定量化可用性测试。结合眼动追踪技术，眼动采样频率高达 250Hz，并配备双眼动传感器，可对移动终端信息产品及原型进行以用户为中心的定量化

可用性测试。设备配有移动终端可用性测试台，其距离、角度均可调节，设有防疲劳手臂支架，保证测试过程中不同被试使用的舒适性；测试台支持手机、平板测试与横屏、竖屏切换。设备配套眼动分析软件，支持眼动状态分析、眼动可视化分析、眼动兴趣区统计与序列分析、WEB/APP 原型眼动交互分析等功能。

2）便携式行为观察仪是可用于记录、分析个体或群体在自然环境中的行为表现，其通过行为编码体系，采用一种定量的观察记录方式来实现行为观察的结构化和系统化，可以对人员的行为数据进行采集、编码、分析与统计，配套的行为观察分析软件可进行行为编码、基础行为统计、交互行为数据分析、行为顺序效应分析等。设备通过移动终端与便携性手持设备的有机结合，可以在户外或者作业现场进行实时行为编码记录，研究人员可以使用设备在不同实验环境下实时采集并编码人员行为，可在人员出现特定行为时进行实时编码，减少后续数据处理工作。

3）手指轨迹分析设备能够在人员与移动终端 APP 原型进行交互的过程中，采集手指交互行为等详细的人机交互数据，如手指点击、滑动等。设备能够自动识别原型中的所有组件，并以像素位置进行精确计算，确保在测试过程中各阶段刺激材料不失真，客观还原所有交互行为，支持手指交互轨迹分析、手指交互热图分析、手指交互兴趣区（AOI）序列分析、

手指交互 AOI 序列轨迹相似度分析等，为深入分析人员的交互行为模式和界面使用习惯提供数据支持。

4）智能穿戴手环生理设备采用手环式穿戴与采集的设计，可以真正摆脱对人员双手的束缚。设备采用智能组合传感器融合技术及算法，支持采集 PPG 血容量脉搏、EDA 皮肤电、SPO2 血氧、SKT 皮肤温度、HR 心率以及九轴人体姿态数据等多通道数据指标，适合于实验室、虚拟现实、真实现场等多种环境下多通道生理指标信号的采集。设备配套生理信号分析软件，能够对 EDA 皮电、HRV 心率变异性等生理信号进行专业化的处理与分析，内嵌信号处理与分析算法，支持多被试数据批量处理。

医疗健康人因测试专家系统基础平台，医疗健康人因测试专家系统基础平台是专业用于“以人为中心”的测评与训练系统，也是标准化的人因工程科研云平台。平台融合了广泛的适用于医疗领域的标准化问卷量表与行为实验任务，以及大数据采集常模计算等功能。平台基于心理学领域专业研究和十余年年行业科技创新经验，结合人工智能、生物反馈、大数据分析、虚拟现实与仿真、脑认知神经科学与电生理、视觉追踪、行为分析、面部表情与状态识别等技术，一方面可用于对人员认知能力的测试评价、训练提升以及选拔工作，另一方面也可为科研提供完整的主客观数据指标。平台的测试实验内容结合国家行业标准，均具有理论支撑、且经过专

家评审以及专业化的情境任务设计，充分涵盖不同领域人员的心理认知图谱。详细功能如下:

1）感知能力基础人因测试模块通过问卷、量表、行为实验范式,对个体的感知力进行测评,包括 PANAS 情绪量表、PAD 情绪量表、情绪 Stroop 实验、速度知觉、运动追踪、反应时实验等几十种测试模板。模块支持电脑端、Pad 端、手机端等显示终端测试，支持手机、网页等多种形式填写的多人同时施测。

2）认知能力基础人因测试模块通过问卷、量表、行为实验范式，对个体的注意力、记忆力等基础认知能力进行测评，包括精神运动警戒任务、运动追踪任务、多目标追踪任务、N-back 记忆负荷、持续任务注意反应范式等几十种测试模板。模块支持电脑端、Pad 端、手机端等显示终端测试，支持手机、网页等多种形式填写的多人同时施测。

3）行动能力基础人因测试模块通过问卷、量表、行为实验范式，对个体的行为决策、反应速度、操作能力等进行测评，包括爱荷华博弈实验、气球风险任务实验、对象选择实验、感知速度测验、双手协调实验等几十种测试模板。模块支持电脑端、Pad 端、手机端等显示终端测试，支持手机、网页等多种形式填写的多人同时施测。

4）特殊人群可及性测试模块旨在保障特殊人群能够便捷、顺畅地使用相关产品或服务。该模块运用问卷、量表以

及行为实验范式等方式，对特殊人群的注意力、记忆力等基础认知能力进行测评，包含精神运动警戒任务、运动追踪任务、多目标追踪任务、N-back 记忆负荷、持续任务注意反应范式等几十种测试模板。此模块支持在电脑端、Pad 端、手机端等多种显示终端上开展测试，同时支持手机、网页等多种填写形式，可实现多人同时进行施测，致力于为特殊人群打造无障碍的测试体验 。

通过各功能测试模块构建出数字疗法医疗器械人因工程质量检测平台，能够实现产品的安全、强度配置、心理预期和用户授权的评测，完成数字疗法产品人因工程质量评估模型的构建。

二、 产品测试过程要求

按照课题任务书要求，开展产品可用性测试，目标聚焦于产品的安全、强度配置、心理预期和用户授权的评测，以下分别从测试角度、测试类型和测试对象进行展开

1）测试角度，分为有效性、使用效率和满意度

a)有效性

软件能否助力患者达成既定的治疗或健康改善目标。由于数字疗法 App 常以游戏或任务形式开展治疗，有效性就体现在这些游戏或任务能否切实对患者病情产生积极影响。这是影响软件实际应用价值的核心要点。

将有效性进行细化，实现二级指标

• 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

b)使用效率

患者能否以高效、便捷方式借助软件实现治疗目标。对于以游戏或任务为载体的数字疗法 App，游戏关卡设置与任务引导需清晰合理。同时，要关注游戏或任务环节及难度设计，应围绕治疗目标科学设计，结合患者身心状况确定难度，且通过设置每日游戏时长上限、定时提醒休息等机制。

将使用效率细化，实现二级指标

• 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

c)满意度

软件满足有效性和使用效率的基础上，患者对使用过程的综合感受和认可程度。数字疗法 App 若在游戏画面、任务难度设置、隐私保护、广告处理等方面存在不足，即便具备一定治疗效果，也会降低患者的满意度。

将满意度细化，实现二级指标

• 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

2）从测试的类型上分为形成性测试和总结性测试两个方面:

a)形成新测试，通过收集定性数据提供及时的反馈来改

进学习，使用小样本定性调查，数据以问题描述和设计建议的形式输出

- 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

b) 总结性测试，通过收集定量数据评估学习的效果，基准测试和比较测试

- 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

3）从测试对象上分为用户界面、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、个性化对象

a) 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？

b) 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。

c) 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？

d) 辅助功能

提供辅助功能，如语音输入、放大字体等

e) 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）?

f) 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?

g) 个性化

允许用户自定义设置，如主题、壁纸和布局等

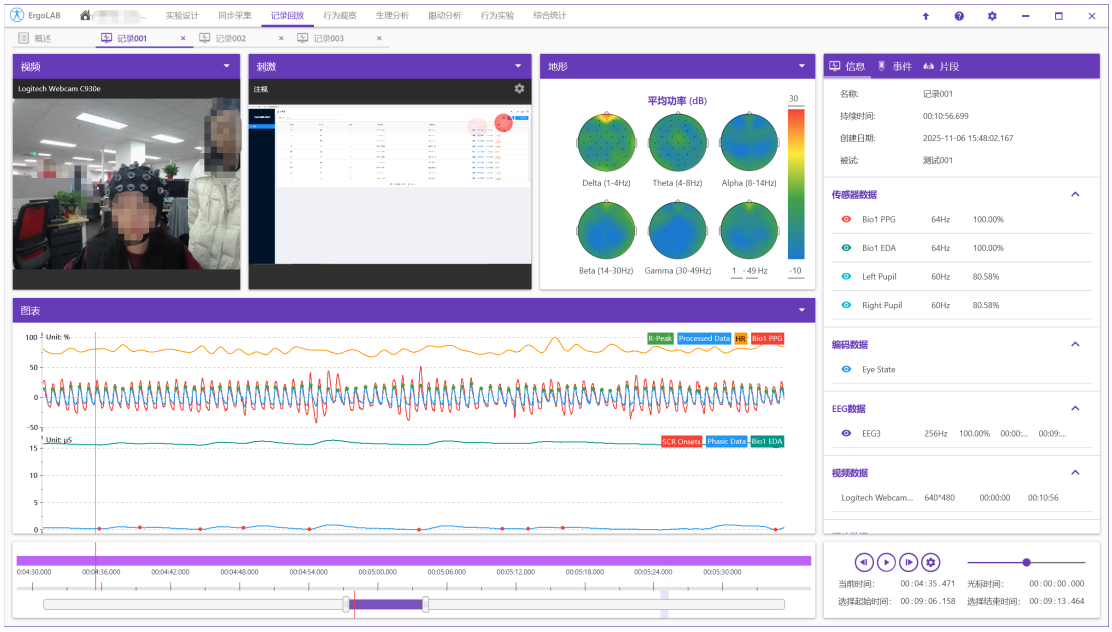
三、 测试结果和结论

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	评估情绪唤

		应（SCR）峰值、SCL 值、心率 LF/HF 比值	醒（应激/挫败）、认知负荷、心理压力水平等

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。以下是各维度的数据分析结果：



1）能力测试结果

感知能力

知觉错误率：测试中出现 2 例图表误读、1 例图标识别错误，总知觉错误率 2 %。表明界面设计满足感知基础需求。

认知能力

注意力：典型任务时长内，有效专注指标（SMR 波功率）

变化如下图所示，医护人员在完成任务过程的不同阶段下，除编辑患者信息外，SMR 指标均在基线值以上。SMR 值越高表示注意力越集中，表明医护人员注意力持续性表现较好。但在编辑患者信息任务下存在注意力分散问题。

记忆力：短期记忆准确率（回忆功能菜单位置、操作路径）平均 86.2%，表明界面功能布局的记忆负担相对较低。

2）高效性测试结果

操作易用性

任务完成程度：5 项核心任务中成功 5 项，任务成功率 100 %；所有任务平均完成时间为预期时间的 115%。表明整体操作正确率较好但存在操作时间的优化空间。

使用错误：测试中仅出现 1 例轻微操作错误（非数据丢失或错误决策类）。表明操作容错性良好。

感知评估

易感知：主要功能按钮等关键信息的 AOI 首次注视时间平均为 1.6 秒，表明医护人员能很快关注到重要信息，视觉搜索效率较高。

易理解：无需指导完成任务的正确率达 91%，表明系统操作的流程逻辑易被理解。

易学习：所有参与者均在 3 次学习试次内掌握核心操作，系统学习成本较低。

3）用户体验测试结果

舒适度与信任度

主观感受：舒适度主观问卷平均得分 4.3 分（5 分制），医护人员对系统操作的心理舒适度较高。

生理佐证：α 波（8-13Hz）主要与放松、平静的心理状态相关，是衡量舒适性的核心脑电指标。SCL（皮肤电导水平）反映情绪唤醒度和心理压力，数值越高代表应激反应越强。瞳孔直径变化与认知负荷、情绪状态直接相关，负荷越低、情绪越平静，瞳孔直径越稳定且偏小。操作期间脑电 α 波最高较基线下降 23%，皮电 SCL 最高上升 10%，瞳孔直径上升 15%，表明在操作过程中总体上用户处于较低的应激状态，舒适性较高。

满意度

感官满意度：主观问卷整体满意度 4.2 分，功能满意度 4.1 分，操作满意度 4.3 分。

生理佐证：皮电与心率数据显示，操作过程中情绪唤醒度低，无明显应激反应，客观验证了主观满意度结果。

安全性

认知负荷：脑电 θ/α 功率比、心率 LF/HF 比值较基线上升 15%，皮电 SCR 上升 13%；NASA-TLX 量表平均评分与基线相比上升 17%，表明系统带来的认知负荷处于低水平。

疲劳度：任务全程瞳孔直径变化率为 16%，长时间操作下的疲劳累积程度较轻。

4）使用环境测试结果

拓展性

多场所适用性：在不同光照（强光、弱光）、轻度环境干扰（人声、设备噪音）模拟场景下，任务成功率均维持在91%。

环境满意度

主观评分：环境因素满意度平均 4.1 分，设备使用环境满意度 4.0 分。

5）关键发现总结

整体达标情况：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。

存在的优化点：部分复杂测评任务的完成时间较长，参与者对特定测评报告图表的理解需主试简要提示，视觉引导可进一步强化；在编辑患者信息过程下，结合医护人员的操作完成时间及多模态生理指标变化，提示仍有优化空间；另外，选择测评任务主观报告结果显示，被试会找不到入口，软件中名称为“测评详情”，被试会认为是已经测评过的内容，而不是新增的测试任务，建议进行修改。

数据有效性：结合主观问卷、行为数据、多模态生理数据的三角验证，结果具有可信度；虽实际测试用户较少，但核心指标表现稳定，可为初步优化提供依据，后续建议扩大

样本量进一步验证。

四、 总结与分析

通过对数字疗法产品的人因工程分析评价研究发现，人因工程学在医疗领域具有较高的应用价值，其核心在于“医-患-器械”协同的优化，本研究采用的数字疗法产品操作界面存在生态效度局限，基于二维显示器的交互方式可能导致的操作临场感确实，研究构建了真实的人机交互评估平台，集成了可穿戴式生物信号采集模块，同步获取操作者的眼动轨迹、肌电信号等人因数据，通过平台分析和模式识别技术，实现产品操作疲劳度的量化评估，最终实现了数字疗法产品的人因分析设计的质量评价。