

基于人工智能双模态交互的无损检测作业管理系统研究

吴洋¹, 刘洋², 贾新³, 张鹏⁴, 焦传飞¹

(1. 中安检测集团(湖北)有限公司, 武汉 430000;

2. 宜昌市计量检定测试所, 湖北 宜昌 443000;

3. 武汉明臣焊接无损检测有限公司, 武汉 430000;

4. 中城绿脉(湖北)检测有限公司, 武汉 430000)

摘要: 针对 NDT 作业流程效率低、交互受限及专业知识依赖度高等问题, 设计并实现了融合语音-文本双模态交互与 RAG 技术的智能作业管理系统; 系统采用分层架构, 集成自然语言处理引擎与双知识库框架, 通过意图识别、结构化数据提取及多策略知识检索, 实现检测数据实时语音录入、标准条款查询与智能报告生成等, 并通过便携式语音终端实现高效人机交互; 实验结果显示, 该系统现场数据录入效率较传统方式提升约 6.1 倍, 在高噪声环境下识别成功率达 88%, 办公环境下报告生成效率提升 63.8%; 用户满意度调查表明, 系统在易用性、功能性及交互体验方面获得一致认可; 该系统有效降低了 NDT 现场操作与报告编制的技术门槛, 提高了作业效率与数据可靠性, 满足了工业检测智能化与现场化的应用需求。

关键词: 无损检测; 作业管理系统; 双模态交互; 大语言模型; 检索增强生成

AI-Powered Dual-Modal Interaction System for Non-Destructive Testing Operations

WU Yang¹, LIU Yang², JIA Xin³, ZHANG Peng⁴, JIAO Chuanfei¹

(1. Zhong An Testing Group (HuBei) Co., Ltd., Wuhan 430000, China;

2. Yichang Institute of Metrology Verification and Measurement, Yichang 443000, China;

3. Wuhan Mingchen Welding Non-destructive Testing Co., Ltd., Wuhan 430000, China;

4. Zhongcheng Greenway (Hubei) Testing Co., Ltd., Wuhan 430000, China)

Abstract: This research addresses the issues of low efficiency, limited interaction, and high dependence on specialized knowledge in NDT operations. It designs and implements an intelligent operation management system integrating dual-modal (speech-text) interaction and RAG technology. The system employs a layered architecture, integrating a natural language processing engine and a dual knowledge base framework. Through intent recognition, structured data extraction, and multi-strategy knowledge retrieval, it enables real-time voice input of inspection data, querying of standard clauses, and intelligent report generation. Efficient human-machine interaction is achieved via a portable voice terminal. Experimental results show the system improves on-site data entry efficiency by approximately 6.1 times compared to traditional methods, achieves an 88% recognition success rate in high-noise environments, and enhances report generation efficiency by 63.8% in office environments. User satisfaction surveys indicate consistent approval of the system's usability, functionality, and interactive experience. This system effectively lowers the technical barriers for NDT field operations and report compilation, enhances operational efficiency and data reliability, and meets the application demands for intelligent and on-site industrial inspection.

Keywords: non-destructive testing; operational management system; dual-modal interaction; large language models; retrieval-augmented generation

收稿日期:2025-07-24; 修回日期:2025-08-13。

作者简介:吴洋(1987-),男,大学本科,工程师。

引用格式:吴洋,刘洋,贾新,等. 基于人工智能双模态交互的无损检测作业管理系统研究[J]. 计算机测量与控制, 2025, 33(12):230-236.

0 引言

无损检测（NDT，non-destructive testing）是现代工业质量控制体系中极为重要的一个环节，它在航空航天、能源电力及高端制造等安全敏感领域中的作用无可替代。伴随工业检测方面的需求日益复杂，NDT 作业流程管理的效率情况及其规范化的程度，已经成为影响检测可靠性以及企业运营成本的关键因素。

当前 NDT 现场作业流程存在不少问题。检测人员通常得先在工程现场完成数据采集与记录工作，之后再返回办公环境进行后续的数据录入及报告编制工作。这样的工作模式不仅增加了作业的时间成本，更重要的是，现场操作人员在碰到疑难问题时往往缺乏即时有效的专家支持，检测结果的判定在很大程度上要依赖于人员自身的经验积累和主观判断。这种对个人经验的过度依赖，再加上现场与后台处理环节的分离，不仅效率低下，还特别容易因人为操作失误而导致数据错误。传统的 B/S（Browser/Server）系统虽实现基础数据电子化^[1]，不过其交互依靠图形界面与键盘输入，在移动或双手占用的场景中的应用存在一定的局限性。此外，基于 B/S 架构的传统数字化系统往往要求操作者既需掌握专业检测知识，又得具备操作复杂系统的能力^[2]，如此双重技能门槛严重制约了该技术的普及应用。

近年来，大语言模型（LLM，large language models）技术飞速发展，它为重构 NDT 作业流程管理系统开拓出全新的技术路径。LLM 在自然语言理解以及逻辑推理方面具有卓越的表现，结合多模态人机交互技术^[3]，可构建真正契合现场作业需求的智能管理系统。在此基础之上，本研究提出一种融合 NDT 知识的双模态交互的作业管理系统，其核心创新包括两方面：其一，基于 LLM 构建自然语言处理（NLP，natural language processing）引擎^[4]，实现 NDT 标准查询、数据记录、专业问题咨询等“无感化”操作；其二，运用检索增强生成（RAG，retrieval augmented generation）技术把 NDT 标准、缺陷特征库等领域知识嵌入到语言模型当中^[5]，以此保证系统输出的专业可靠性。

为满足 NDT 工程现场实际需求，本系统提供语音与文本双模交互能力^[6]，通过集成便携式语音交互终端并优化语音交互流程，显著提升人机交互效率，实现用户双手的解放。同时系统还具备高精度的数据检索功能，其能够在现场检测期间，为相关决策给予数据支撑，并在办公环境下依据采集数据自动生成标准化检测报告。在知识处理层面，运用 RAG 技术来构建 NDT 专业知识库，有效解决了 LLM 在专业领域应用中存在的精准性问题。这条技术路线不但提升了系统的易用性、实用性与工作效率，而且还大幅缩短了检测报告生成周期，降低了使用门槛，使检测人员能够更加专心地

投身于现场检测任务当中，也为工业 NDT 领域的智能化转型贡献了一套可复用的方法。

1 系统架构

如图 1 所示，系统采用分层架构设计，实现功能模块的解耦与协同。交互层是用户与系统进行沟通的桥梁，主要提供语音和文本两种交互方式^[7]；NLP 引擎是平台的核心智能模块，通过 LLM 解析用户输入的自然语言，并生成相应的响应^[8]；业务逻辑层负责实现核心业务功能，该层与 NLP 引擎紧密协作，根据用户的意图和输入，调用相应的功能模块，完成业务操作；数据层是平台的数据存储中心，负责存储和管理各种类型的数据^[9]。各层通过定义良好的接口进行通信，形成端到端的智能处理流程。

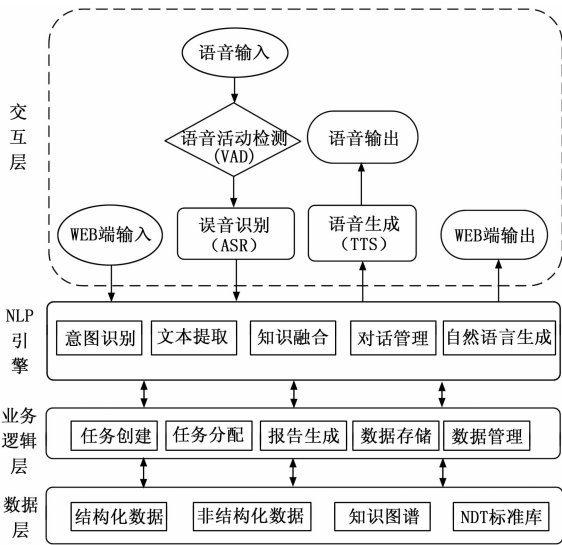


图 1 系统架构示意图

2 系统实现与验证

2.1 技术选型

本系统在技术选型方面综合考量了性能、成本、可扩展性以及社区生态支持等多项因素。在 LLM 选型策略上，选用 DeepSeek 官方提供的通用 LLM 进行混合部署。

轻量级任务由 DeepSeek-V3-0324（660B）模型处理。该模型是一个拥有高达 660B 参数的混合专家（MoE，mixture of experts）模型，通过动态激活部分专家模块实现高效推理，显著降低资源消耗，能够在高并发访问场景下保持低延迟、高吞吐的服务性能。

复杂任务处理则交由具备更强计算能力的 DeepSeek-R1-0528（685B）模型负责。该模型在多项开放集和闭集语义理解基准评测中均展现出领先性能，具备深度语义理解的能力，能够显著提升复杂场景下的文本生成准确率与语义一致性。尤其在处理多轮交互、跨段落乃至跨文档的信息抽取与知识融合任务时，该模型表现

稳定且高效。

语音交互模块采用本地部署的语音识别（ASR, automatic speech recognition）模型及语音生成（TTS, text to speech）微调模型，这样不仅能够支持方言识别优化，还能对输出的音色展开定制化操作，并对语调加以调节等，从而满足各式各样的个性化需求，进而保证语音交互的自然流畅性。

为有效规避通用 LLM 在专业领域应用中存在的知识幻觉风险，本研究对主流技术路径进行了评估^[10]。NDT 领域相关标准更新迭代极为频繁，若采用微调（Fine-tuning）技术，将无可避免地陷入持续、高昂的时间与算力资源消耗，这与系统的即时性要求相悖。基于此，我们最终确立了以 RAG 为核心的技术框架，该框架将模型的推理能力与外部知识库进行连接，实现了灵活性与专业性的统一。

2.2 核心模块实现

2.2.1 NLP 引擎

本系统 NLP 引擎的核心，是一套旨在实现精准语义理解与高效任务执行的协同 workflow。该机制摒弃了由单一模型处理全流程任务的传统范式，创新性地构建了基于双模型分工的二级处理架构^[11]。在该架构中，提示词工程（Prompt Engineering）作为核心纽带，精确引导各类 LLM 聚焦其最擅长的子任务，保障整个系统在复杂语义处理与任务调度中的精确性与鲁棒性。

系统第一阶段的核心任务在于高精度、低延迟的用户意图识别与分类。当用户输入文本（包括语音转换后的文本）时，系统首先启动意图分类流程。该流程中，LLM 被严格限定为“意图分类器”角色。为精确引导模型聚焦于分类任务并约束其功能边界，我们设计并应用了定制化的提示模板^[12]。该模板融合了明确的分类体系定义与少样本学习（Few-shot Learning）示例，从而能够将用户请求准确映射至预定义的意图类别（例如，“无损检测数据记录”“标准条款查询”或“专业咨询”等）。模型输出强制采用标准化的 JSON 格式，确保结果的结构化与系统可读性，为下游任务流程的自动化触发奠定了坚实的基础。

在意图完成分类后，系统进入第二阶段，即基于意图的任务分发与执行。若用户意图属于“无损检测数据记录”，系统将启动结构化数据提取 workflow。该模块接收原始文本输入，并通过提示词引导的专属 LLM 提取关键字段，如“检测部位”“检测方法”“检测结果”等。提示词中明确设定了数据模式（Schema）与输出规范，从而实现对现场非结构化描述的高效结构化处理，最终将结果存入数据库。

另一方面，若识别结果为“标准条款查询与专业咨询”，系统将转向 RAG 模块所支持的专业咨询 workflow。

在该流程中，模型在回答用户问题前，先从内置的专业知识库中检索相关条款或技术内容，并基于检索结果进行语言生成。这种“先检索、后生成”的模式，确保了模型提供的解决方案或技术解释不仅在语言上流畅自然，更在专业性与权威性上有所保障，为现场遇到的复杂问题提供了可靠的决策依据。

2.2.2 语音交互

语音交互模块设计的关键在于主控芯片选型。如表 1 所示，研究团队对比分析了当前主流芯片，综合考量运算能力、连接性、存储与扩展性、功耗等条件后，选定 ESP32-S3 作为核心处理单元^[13]，进行语音交互的嵌入式部署。

表 1 主流芯片特点

芯片型号	ESP32-S3	ESP32-C3	STM32WL33	RL78
处理器架构	Xtensa® 32 位 LX7 双核	RISC-V 32 位单核	Arm® Cortex®-M0+	16 位 CISC
主频	最高 240 MHz	最高 160 MHz	最高 64 MHz	最高 32 MHz
内存配置	384 kB ROM+ 512 kB SRAM	384 kB ROM+ 400 kB SRAM	最高 256 kB Flash+ 32 kB SRAM	最高 768 kB Code Flash+ 8 kB Data Flash+ 最高 48 kB SRAM
AI 加速能力	支持向量指令用于神经网络和信号处理加速	无专用 AI 硬件加速, 但可通过软件库实现基础 AI 运算	无专用 AI 硬件加速	无专用 AI 硬件加速, 但提供 e-AI 解决方案

在该芯片上，本系统采用了乐鑫 ESP-Skainet 框架。该框架集成了先进的声学前端算法（AFE, acoustic front end），提供回声消除（AEC, acoustic echo cancellation）、自动增益控制（AGC, automatic gain control）、噪声抑制（NS, noise suppression）、语音活动检测（VAD, voice activity detection）以及麦克风阵列语音增强（MASE, microphone array speech enhancement）等关键功能^[14]。ASR、NLP 和 TTS 等复杂任务则交由云端服务器执行。

为验证 ESP32-S3 在本应用场景中的性能表现，研究团队开展了专项测试。在实时性方面，系统表现出色的响应能力至关重要。测试结果表明，从用户发出唤醒词至系统完成激活并进入指令接收就绪状态，平均唤醒时间约为 500 ms。在系统激活状态下，本地 AFE 模块的处理延迟稳定控制在 50 ms 以内，有效保障了用户指令下达过程中的交互流畅度，避免了可感知的卡顿现象。关于系统资源占用情况，得益于 ESP32-S3 的 LX7 双核架构优势，在仅执行唤醒词监听的待机状态下，

CPU 平均负载保持在 15% 以下, 对应极低功耗水平。即使在执行实时语音采集、前端处理及 Wi-Fi 数据传输等高负载任务时, 峰值工作状态下的 CPU 总体负载也仅为 65%, 系统运行始终平稳无卡顿, 且仍保有显著的算力冗余, 为应对突发任务提供了坚实的稳定性保障。内存资源管理方面, 系统固件及 ESP-Skainet 框架的整体 SRAM 占用约为 280 kB, 剩余超过 200 kB 的可用内存空间, 为高吞吐量的音频数据流缓冲以及复杂的应用逻辑处理提供了充裕的资源, 整个测试过程中未出现内存溢出的风险。

研究团队对硬件架构加以优化, 把麦克风阵列以及音频输出模块高度集成到了独立主板之上, 以此构建出轻量化的边缘计算终端。终端采用低功耗设计, 能够实时对语音数据展开处理, 借助无线通信模块与云端服务器建立安全连接, 可高效传输经过压缩的语音数据与指令。电路设计运用模块化布局, 如图 2 所示, 在缩减设备体积的同时也确保了功能完整性。

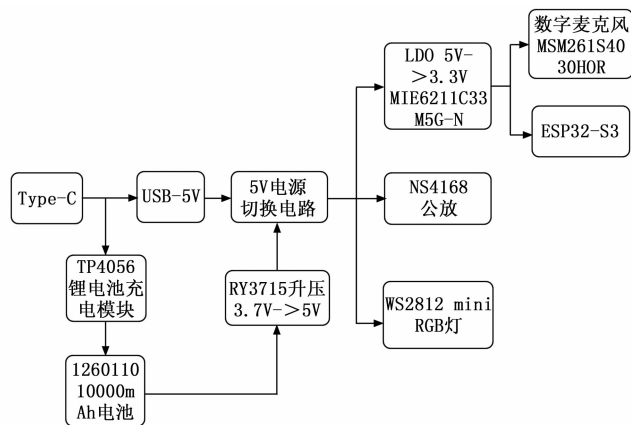


图 2 电路模块

为精确控制系统工作状态, 研究团队设计了基于双关键词触发的唤醒机制。该机制通过持续对音频进行监听, 只有在识别出预先设定好的启动关键词的时候, 才会将系统激活; 而一旦检测到停止关键词, 系统便会马上切换到休眠状态。这种双阈值触发策略有效杜绝了工业现场由设备噪音、人声干扰等导致的误激活风险, 并且还能保证在检测任务中断之际, 资源可以即时得到释放。

针对 NDT 检测现场复杂的声学环境, 系统集成先进的声学特征建模技术^[15]、语音增强技术、流式语音识别技术等, 实现高效、低延迟、高精度的实时交互。流式语音识别技术把音频流实时分割为短时帧, 送入并行处理线程。在语音增强技术方面, 系统采用基于 WebRTC (Web Real Time Communication) 音频处理模块的多级预处理流水线^[16], 其 VAD 模块实时区分语音/非语音段, 这种增量式多线程机制有效避免了批处理延迟, 支持对任意长度音频流的持续分析。

2.2.3 专业知识增强

本研究构建的 RAG 框架采用异构双库架构, 以满足用户多层次专业知识需求。该架构包含两个互补知识库: 动态标准知识库与静态专家知识图谱。

动态标准知识库以现行 NDT 标准为核心, 基于 LangChain 框架实现模块化知识管理^[17], 使系统具备高度灵活性。管理人员仅需编辑相应文件, 就可以让知识库实时得到更新。这样的设计极大地降低了维护方面的门槛, 与此同时, 还切实保证了标准知识能够具备时效性。

静态专家知识图谱则是对主流专业书籍、核心教材以及详细缺陷特征等这些深层理论资源加以整合。对于这类稳定知识, 研究团队采用 GraphRAG 技术方案^[18], 将非结构化文本转化为富含语义关联的结构化知识图谱。和常规向量检索相比, 该方案能够更加直观地把专业概念之间那种复杂的关联网络展现出来, 在处理需深度理解概念内在联系的复杂问题时优势明显, 有效弥补了传统检索系统在知识深度挖掘方面的不足。

为确保检索知识片段与用户问题具备高度的相关性与准确性, 本框架创新性地引入了基于意图分析的动态混合检索策略。该策略的核心在于首先对用户提问进行深度意图解析, 精准识别其内在的查询倾向。例如, 系统能够有效区分用户是寻求具体标准条款的精确匹配 (如询问“XX 缺陷在 XX 标准中等级评定为几级”), 还是意图进行深层次的因果或关联关系探索 (如探究“焊接速度过快可能产生什么缺陷”)。基于实时的意图分析结果, 系统继而动态地调整其检索策略并分配不同知识库的权重。针对标准查询类问题, 优先启用基于语义向量的相似度检索, 重点在动态更新的标准知识库中定位和匹配最相关的文本片段。而对于需要深度知识探索的问题, 系统则侧重于在构建完善的静态专家知识图谱中进行知识路径遍历: 即将问题中的核心概念映射为图谱中的起始节点, 运用图算法高效发掘与之紧密关联的实体及蕴含复杂逻辑的关系路径, 最终生成一个聚焦于问题本质的深层知识子图。最终, 系统通过并行执行双知识库检索任务实现多路信息召回, 将异构的检索结果 (标准文本片段与图谱路径) 进行初步整合。一个本地部署的重排序模型 (Re-ranking model) 继而对这些上下文信息进行二次评估, 依据其与问题的相关性、信息本身的重要性以及知识覆盖的全面性进行综合打分与排序, 从而筛选并输出与用户查询意图最匹配、内容最精准的知识组合。经过上述策略筛选出的最优上下文被输入 LLM, 并通过精细化 Prompt 设计约束生成范围, 从而保证输出的专业性与准确性。

以“CSK-IIA-2 试块直探头设置灵敏度时, 除可以调校 10, 20, 30, 40 深度外, 还可以调校哪些? (选

项: A. 不能设置其他孔深; B. 60、80、100; C. 50、70、100; D. 50、80、100) ”这一道专业选择题为例进行对比分析, 未引入 RAG 机制的通用 LLM 的回答仅停留在选项的简单罗列层面, 缺乏必要的技术解释和数据支撑。而在引入 RAG 机制后, LLM 基于检索到的权威技术资料, 明确指出除标定深度外, 还可调校深度为 50、70、100 mm (对应选项 C), 并进一步给出接近实际检测需求的精确参考值 (50.98、70.98、101.60 mm), 同时结合试块结构与孔间距原理进行说明。由此可见, RAG 增强后的回答不仅验证了选项的正确性, 还补充了设计依据与应用解释, 显著提升了答案的可靠性与权威性。

2.3 实验验证

为客观评估本系统性能, 本研究采用双场景对比实验设计, 分别在工业管道检测现场和企业办公环境下, 与传统 B/S 架构系统进行性能对比。实验核心聚焦于数据采集效率、环境适应性、报告生成效率及用户满意度 4 个关键性能指标。

2.3.1 检测现场实验: 数据采集效率与环境适应性评估

在工业管道检测现场, 研究团队委托具备相应资质的持证 NDT 人员, 在执行环焊缝超声检测任务时, 同步使用本系统的语音交互功能实时口述包含缺陷坐标、波形特征等关键专业参数的检测指令。为进行有效对比, 检测人员在相同任务中还需使用移动端传统 B/S 系统进行手动数据录入。

实验过程中, 研究团队详细记录了每条语音指令从口述开始到系统完成识别并记录的平均耗时。同时, 采用包含 NDT 专业术语的 100 条数据构建的噪声测试集, 在不同背景噪声等级下测试了语音识别准确率。此外, 还统计了使用移动端 B/S 系统手动录入相同内容每条指令的平均耗时, 以进行对比分析。实验结果表明, 本系统语音交互模式下平均单条数据录入耗时 ($M=9.05$, $SD=1.28$) 显著低于传统 B/S 系统手动录入 ($M=55.00$, $SD=3.34$)。独立样本 t 检验结果显示, 二者差异具有统计学意义 [$t(118)=-99.38$, $p<0.001$], 效率提升约 6.1 倍。这种效率突破主要归功于语音交互技术对双手操作的解放, 使检测人员能够同步进行检测作业与数据记录, 有效规避了传统模式下设备操作与数据录入频繁切换导致的时间损耗。

如表 2 所示, 在环境适应性方面, 系统在不同背景噪声条件下展现出良好的稳定性和鲁棒性。当语音识别文本存在误差时, NLP 引擎凭借深度优化的提示词模板精准解析用户意图, 在高达 85 dB 的背景噪声下, 首次识别成功率仍保持在 88%, 充分满足工业现场复杂作业环境下的可靠性要求。

表 2 不同背景噪声下语音识别任务成功率评估

背景噪声/dB	首次识别成功数/次	首次识别成功率/%	三次内识别成功数/次	三次内识别成功率/%
55	97	97	100	100
60	95	95	100	100
65	94	94	100	100
70	94	94	100	100
75	93	93	100	100
80	91	91	100	100
85	88	88	91	91

2.3.2 办公环境实验: 报告生成效率评估

在企业办公环境中, 研究团队模拟了 NDT 报告编制的完整流程。邀请了 10 名代表目标用户群体但不具备 NDT 专业知识的操作人员, 在实验中被要求利用本系统的智能引导功能完成检测报告所需信息的采集。随后基于现场采集的结构化数据, 利用 LLM 自动进行数据逻辑校验与完整性检查, 在用户确认信息无误后, 生成符合 NB/T 47013《承压设备无损检测》格式规范的检测报告。作为对照组, 同一批操作人员需在传统 B/S 系统中, 先人工录入现场采集的检测数据, 然后手动编制相同的报告。

实验结果表明, 在报告生成效率方面, 本系统组 ($M=6.32$ min, $SD=0.61$) 显著优于传统 B/S 架构系统组 ($M=17.45$ min, $SD=0.64$)。独立样本 t 检验结果显示, 两组平均耗时差异具有高度统计学意义 [$t(198)=-113.44$, $p<0.001$]. 具体而言, 非专业背景人员使用本系统完成标准化检测报告的平均耗时较传统手动操作模式缩短约 11.1 min, 效率提升幅度达 63.8%。这一结果充分验证了系统智能引导、结构化数据采集和 LLM 辅助报告生成功能在降低专业技术门槛、提升报告编制效率方面的有效性。

2.3.3 用户满意度评估

为全面评估系统的用户满意度, 研究团队在实验结束后采用科学化的问卷调查方法展开评估。在调查设计与实施层面, 本次评估覆盖全部 16 名完成实验的目标用户群体 (含 6 名一线现场检测人员及 10 名办公环境操作人员)。问卷基于国际通用的 5 分 Likert 量表 (1=非常不满意, 5=非常满意) 构建核心指标^[19], 并借鉴成熟的系统可用性量表 (SUS, system usability scale) 框架, 结合本研究场景特性从易用性、功能性、交互体验及总体满意度 4 个维度展开测评。为确保评估工具的有效性, 问卷初稿经人机交互专家审核修订, 并通过匿名发放形式鼓励参与者提供真实反馈。针对问卷的可靠性验证, 团队对回收数据进行内部一致性信度分析, 获得的克隆巴赫系数 (Cronbach's α) 为 0.87, 显著高于学界 0.7 的基准阈值, 证实该评估工具具有优异的信度

水平。

由表 3 可知, 系统获得全体参与者的积极认可, 各维度总体平均分均高于 4.4 分。尤为值得注意的是, 现场检测人员在易用性、功能性及总体满意度三项均给出满分评价 (5.0 分), 而办公操作人员对功能性的评分亦达 4.8 分。

表 3 用户满意度评估结果

评估维度	现场检测 人员平均分	办公操作 人员平均分	总体平均分
易用性	5	4.2	4.6
功能性	5	4.8	4.9
交互体验	5	3.8	4.4
总体满意度	5	4.3	4.7

为探究两组用户群体在评价上是否存在显著差异, 研究团队采用了独立样本 t 检验进行分析。结果显示, 在“交互体验”维度上, 现场检测人员的平均分 ($M=5.0$) 显著高于办公操作人员 ($M=3.8$), 差异具有统计学意义 ($p<0.001$)。这表明现场人员对“语音录入结合移动端复核”的交互模式满意度极高。而在易用性、功能性和总体满意度维度上, 两组评分差异未达到统计学显著水平, 表明系统在这几方面获得了两类用户的一致认可。

定性分析结果与定量数据形成了有效互补, 进一步佐证了评估结论。现场检测人员普遍强调, 相较于传统纸质记录或纯界面操作, 混合交互模式显著提升了“边检测边记录”的现场工作效率; 办公操作人员则高度认可系统功能的完备性与操作逻辑的清晰度。综合严谨的问卷设计、信度验证、统计检验及质性反馈, 尽管样本规模存在局限, 但评估结果一致表明: 本系统在目标用户群体中获得了高度认可, 其革新性的现场交互范式被验证具备显著的实用效能与应用价值^[20-21]。

3 结束语

本研究致力于提升 NDT 作业效率与管理水平, 提出一种融合语音—文本双模态交互与 RAG 技术的智能作业管理系统。该系统整合 NLP 与 RAG 技术, 给工业检测场景带来智能化解决方案。检测现场, 语音交互技术实现了参数的实时录入, 检测人员得以同步执行检测任务与数据采集, 彻底改变传统模式下人工记录与现场操作分离的状况。数据处理以及报告生成阶段, 系统通过引导用户逐步填写检测报告参数, 实现了标准化报告的快速生成, 大幅缩短了报告出具时间并降低了错误率。除此之外, 系统集成了 NDT 领域知识库, 结合其独特的 RAG 框架, 为现场检测人员提供决策辅助和标准查询, 同时为非专业人员提供专业解答, 从而降低了技术的理解难度。

针对 NDT 智能化发展需求, 后续研究将重点推进 3 个方面的工作: 1) 融合多模态 LLM 来强化目标检测能力, 实现高精度缺陷识别的效果, 并构建支持视觉、语音与文本协同的多模态信息采集体系。在实际操作过程中, 通过拍摄缺陷部位照片或导入检测仪器截图, 系统自动识别缺陷类型并记录相关数据, 减少人工口述和手动录入环节; 2) 达成与主流 NDT 仪器的数据接口对接, 支持检测波形、图像等原始数据的自动读取与关联, 并与企业管理系统 (ERP, enterprise resource planning) 对接, 实现检测任务自动接收、检测结果/报告自动推送回传, 融入企业生产管理闭环; 3) 优化模型部署策略, 借助模型压缩和知识蒸馏技术, 将通用 LLM 替换成可本地部署的小参数量模型。重点开发适配普通服务器的轻量化部署方案, 这样就可以在降低系统硬件门槛与部署成本的同时提升响应速度与运行效率。

参考文献:

[1] 陆建平, 张小龙, 周承华, 等. 基于 B/S 架构的无损检测作业管理系统设计与实现 [J]. 化工装备技术, 2024, 45 (4): 24-27.

[2] 苏 岩. 无损检测数据管理系统的设计与实现 [D]. 沈阳: 东北大学, 2012.

[3] 张晴雪, 张欣亮, 武凌杰. 多模态 AI 智能体在工业缺陷检测中的应用 [J]. 机器人技术与应用, 2025 (1): 37-40.

[4] 林毅炜. 基于深度学习的中文自然语言生成复杂 SQL 语句技术研究 [D]. 武汉: 华中科技大学, 2021.

[5] 祖红月, 于 惠, 刘洪运, 等. 中医本体及知识库的构建与应用研究 [J]. 中国数字医学, 2024, 19 (9): 37-43.

[6] 苟晓茹. 基于多模态的 AI 语音识别及人机交互系统研究 [J]. 自动化与仪器仪表, 2024 (12): 159-162.

[7] 张 斌, 陈 莉, 唐 琳, 等. 基于 AI 语音交互和专业分类的安规题库快速查找系统 [J]. 电力安全技术, 2023, 25 (12): 27-29.

[8] 杨卓琳. 联合意图识别和语义槽填充的口语理解方法研究 [D]. 西安: 西安工程大学, 2023.

[9] 高翌飞, 李 阳, 李震波. 非结构化无损检测数据管理系统的研究 [C] //2015 远东无损检测新技术论坛——基于大数据的无损检测论文集. 北京: 中国船级社实业公司无损检测实验室; 北京数码易知科技发展有限责任公司, 2015: 635-640.

[10] 魏 炜, 张 坤, 徐哲淇. 从通用到垂直: 大模型赋能管理学研究的新路径 [J]. 管理学报, 2025, 22 (1): 1-11.

[11] 曾 亮. 大模型在媒体行业的落地——从对话机器人到智能体工作流 [J]. 现代电视技术, 2024 (9): 56-60.

[12] 孔文强. 基于 LLM Prompt 技术的领域知识图谱构建框

架研究 [D]. 成都: 电子科技大学, 2024.

[13] 边玉亮, 毛 润, 李启航, 等. 基于 ESP32 的 ChatGPT 语音交互系统设计与实现 [J]. 黄冈师范学院学报, 2024, 44 (5): 20-25.

[14] KURNIAWAN A. Internet of things projects with ESP32: Build exciting and powerful IoT projects using the all-new Espressif ESP32 [M]. Packt Publishing Ltd., 2019.

[15] 朱 洋. 基于深度学习的连续语音识别 [D]. 桂林: 桂林电子科技大学, 2024.

[16] 李 智. 基于 WebRTC 的浏览器实时音视频通信技术研究 [J]. 电脑编程技巧与维护, 2024 (6): 135-138.

[17] 姜嘉伟. 基于 Langchain-LLMs 框架的智能问答系统的

然科学版), 2025, 46 (2): 76-84.

通过滤波器降低噪声干扰; 在预处理基础上将 RBF 神经网络、自适应 PID 算法、模糊算法深度融合, 显著提升了模型对机器人的控制能力, 通过参数的自适应调节提升了模型的匹配度, 同步改善了模型对双轮自平衡机器人倾角的控制精度。实验结果显示, 提出的融合人工智能算法具有良好的机器人倾角控制效果, 在有障碍物和坡度的全过程采样点标准差仅为 0.852, 且在锁定机器人倾角的条件下能够最先完成指定任务。人工智能和机器学习是未来机器人控制的主要方向, 通过尝试多种算法的融合能够实现各控制算法取长补短和递进优化, 以进一步提升模型的控制效果。

参考文献:

[1] 吴雄伟, 周云成, 刘峻淳, 等. 面向温室移动机器人的无监督视觉里程估计方法 [J]. 农业工程学报, 2023, 39 (10): 163-174.

[2] 张 宇, 李宇翔, 张贺飞, 等. 撼地者: 基于遥操作与自主导航的移动救援机器人 [J]. 中国科学技术大学学报, 2023, 53 (1): 32-43.

[3] 陈艺文, 江文松, 杨 力, 等. 基于运动约束的移动机器人路径规划 [J]. 计算机集成制造系统, 2023, 29 (4): 1186-1193.

[4] 叶长龙, 田 松, 于苏洋, 等. 基于降阶 ESO 的全方位移动机器人滑模控制 [J]. 信息与控制, 2025, 55: 1-11.

[5] 孙鹏宇, 李 君, 刘宇飞, 等. 基于车臂协同策略的移动机械臂柔顺开门技术 [J]. 机器人, 2025: 1-14.

[6] 文 浩, 陈小龙, 李静波, 等. 面向大型空间结构建造维护的爬行机器人研究综述 [J]. 振动工程学报, 2025, 38: 1-20.

[7] 陈 刚, 匡 鹏, 曾红兵, 等. 自平衡机器人控制系统的时滞稳定性分析 [J]. 控制工程, 2025, 32 (4): 738-744.

[8] 李小彭, 李雪东, 樊 星, 等. 基于模糊 PID 控制的输电线路巡检机器人自平衡机理分析 [J]. 东北大学学报 (自

设计与实现 [D]. 延吉: 延边大学, 2024.

[18] 王冰芬, 王国鑫. 基于大语言模型的 BIM 模型智能审查系统 [J]. 建设科技, 2025 (3): 57-59.

[19] PESCAROLI G, VELAZQUEZ O, ALCÁNTARA-AY-ALA I, et al. A likert scale-based model for benchmarking operational capacity, organizational resilience, and disaster risk reduction [J]. International Journal of Disaster Risk Science, 2020, 11 (3): 404-409.

[20] 赵小芬, 彭 朋. 基于多模态视听融合的 Transformer 语音识别算法研究 [J]. 传感器与微系统, 2025, 44 (2): 48-52.

[21] 李 通, 羊红光, 刘 康, 等. 大模型知识蒸馏方法研究进展 [J]. 河北省科学院学报, 2025, 42 (2): 94-96.

然科学版), 2025, 46 (2): 76-84.

[9] 董慧芬, 姚勤燃, 王俊峰. 基于动力学参数辨识的两轮机器人平衡控制 [J]. 控制工程, 2025, 32 (2): 345-353.

[10] 杜巧玲, 薛成泽, 郑 伟. 基于 MPC 的仿生尺蠖机器人运动控制 [J]. 沈阳工业大学学报, 2025, 47 (1): 124-129.

[11] 韩子航, 陈 力, 韩大鹏. 基于变论域模糊补偿的空间机器人自适应控制 [J]. 南京航空航天大学学报, 2024, 56 (3): 424-431.

[12] 吴以童, 李 斌, 刘 春, 等. 融合 GMM 和 PSO 的储罐超声检测机器人避障研究 [J]. 计算机仿真, 2023, 40 (11): 408-414.

[13] 张智超, 闵淑辉, 廖 凯, 等. 基于改进模糊 PID 果园双轮移动机器人运动控制 [J]. 农机化研究, 2024, 46 (7): 14-20.

[14] 周 余, 吴功平, 杨 松, 等. 巡检机器人塔上自主充电控制方法研究 [J]. 机械设计与制造, 2023 (3): 271-276.

[15] 于力率, 苏晓杰, 孙少欣, 等. 基于分层控制策略的六轮滑移机器人横向稳定性控制 [J]. 自动化学报, 2023, 49 (7): 1421-1432.

[16] 杨贤昭, 黄晶晶, 刘惠康, 等. 多轴隧道机器人非线性环形耦合协同控制 [J]. 机械设计与制造, 2023 (9): 277-281.

[17] 杨继荣, 黄卫华, 付亮亮, 等. 立方体机器人自适应滑模容错控制系统设计 [J]. 组合机床与自动化加工技术, 2023 (4): 100-103.

[18] 陈睢帅, 刘宇琪, 汤 卿. 机器人工具坐标系标定中的非线性问题求解方法研究 [J]. 机床与液压, 2023, 51 (9): 9-15.

[19] 商德勇, 黄云山, 黄欣怡, 等. 基于奇异摄动的刚柔耦合 Delta 机器人非线性混合控制 [J]. 机械工程学报, 2024, 60 (5): 95-106.

[20] 易成群, 吴佳晔, 李嘉莉. 基于扰动环境下的轮式机器人路径跟踪控制 [J]. 组合机床与自动化加工技术, 2023 (4): 23-26.