

人工智能命题技术研习报告

汪鹏

国能铁路装备有限责任公司，内蒙古 包头 014060

DOI:10.61369/ECE.2025080020

摘 要： 随着人工智能的发展，深度学习与自然语言处理大模型相继成熟，使用大模型的成本也显著降低，使利用基于人工智能的试题编写技术通过深度学习与自然语言处理能力进行试题编制成为了现实，通过人工智能技术，混合模型的使用，人机结合的试题编写模式，显著提升了考试、技能认定等需要大量试题编制场景的时间，同时提高了命题效率与质量。

关 键 词： 人工智能；DeepSeek；试题编写

Research Report on Artificial Intelligence Proposition Technology

Wang Peng

Guoneng Railway Equipment Co., Ltd., Baotou, Inner Mongolia 014060

Abstract： With the development of artificial intelligence, deep learning and large models of natural language processing have matured one after another, and the cost of using large models has also significantly reduced. This makes it possible to compile test questions using artificial intelligence-based test question writing technology through deep learning and natural language processing capabilities. Through artificial intelligence technology, the use of hybrid models, and the man-machine combined test question writing mode, the time required for scenarios that need a large number of test questions to be compiled, such as examinations and skill certification, has been significantly shortened, and the efficiency and quality of proposition have also been improved.

Keywords： artificial intelligence; DeepSeek; test question writing

一、项目背景

当前，装备公司教培中心对于持续更新内部各工种技能水平题库的需求日益迫切^[1]。然而，传统的命题方式效率较低，难以适应快速更新的需求。试题主要用于单位职工职业技能水平的认定，因此，在《技规》、《厂规》、《段规》、作业指导书等政策指导性文件发生更新或修订时，对变更知识点范围的试题进行初步确定、审核和修订等工作显得尤为重要。这需要大量的人力和物力投入。通过引入人工智能技术，可以实现自动化命题和持续更新题库，显著提高工作效率，确保题库的更新速度和准确率。

鉴于上述需求，装备公司教培中心于2025年3月27日至4月2日组织了2025年技能大赛试题专家校审会议，并同时进行了人工智能编写试题的探索性研究。本次人工智能试题编写测试利用人工智能技术为12个不同工种生成了1150道试题，试题类型涵盖了四种：单选题、多选题、判断题和填空题^[2]。测试结束后，通过收集和汇总专家意见，整理出11条问题，主要包括试题选项描述错误、多选题设计混乱、知识点引用不完整以及题型间存在重复等。针对上述问题反馈，本文提出了相应的改进方案，包括模型选择、加强题型平衡性检查等措施，以提高人工智能自动化命题的质量和可靠性。

二、测试概况

1、模型选型

为满足针对特定知识点的试题编制需求，我们对市场主流 AI 模型的评估工作。受限于技术环境与数据安全要求，本次评估未能部署私有化模型进行测试^[3]。在此基础上，我们选取了腾讯 IMA、豆包、通义千问、DEEPSEEK-V3及讯飞星火五款云端部署的主流模型展开了出题测试。

此次测试聚焦于试题生成数量、响应时效、内容合理性、排版规范性、答案准确性及解析完整性六大维度。经综合评估，豆包模型凭借其出题数量和多模态排版适配度上的突出表现，最终决定选择豆包工具作为此次试题编制测试的辅助工具^[4]。

2、测试方式

本次测试工作主要围绕人工智能技术在专业试题生成领域的应用展开。首先，我们上传了多个专业方向的系统性知识要点，这些知识点经过严格筛选和分类。在测试过程中，人工智能技术对这些专业知识内容进行深度解析与多维度分析^[5]。系统不仅能够识别知识点之间的内在逻辑关系，还能精准把握每个知识点的核心要素和考查重点。根据预先设定的需求数量及多样化的题型要求，人工智能系统展现出强大的自动生成能力。

为确保试题质量达到预期标准，我们特别组建了由各专业领

域专家组成的评审团队。试题编制完成后，每个专业领域的三位专家需现场独立作答，并对试题进行核对。他们不仅需要验证试题内容的准确性，还要评估知识点覆盖的全面性和考查方式的科学性^[9]。通过多轮严格评审和反复校对，确保人工智能所出试题既符合专业标准，又能有效检验学习者的真实水平。整个测试过程严谨规范，为人工智能技术在教育领域的应用提供了可靠的数据支持和实践依据。

3、专家反馈结果分析

在本次评估产生的 1150 道试题，其题型分布情况如下：

题型	单选题	多选题	判断题	填空题	合计
数量	290	290	280	290	1150

经过专家对人工智能编撰试题的审阅，此次共收集到 11 条专家反馈，可归纳整理为以下几类问题：

1. 试题选项设计问题^[7]。涉及知识点描述错误、多选题选项混乱且缺乏严谨性、选题选项存在分歧，以及多选题中答案中含 AB 选项较多。
2. 知识点引用不完整，准确性不足，题干表达模糊，表现为引用一半知识点，专业性表述不合理。
3. 题型平衡性问题，在不同题型间存在知识点重复及答案选项比例不均衡的情况。
4. 答案设计问题，主要体现在填空题答案过于复杂，选择题选项不符合题意。
5. 适用性不足，存在知识点分布不均匀，题型难易程度不均衡^[8]。
6. 试题缺乏深度逻辑与创造性，模式化输出，容易生成重复率高、结构模式化的题目，难以突破传统框架，缺乏创新性和灵活性。
7. 语义理解不足，在理解复杂知识点方面存在一定的局限性，人工智能不能准确捕捉到题目背后的实际意图和上下文联系，导致生成的试题与预期目标有所偏离。

三、问题分析

上述问题的成因可归结为以下几点：

1. 人工智能算法需经过训练，目前其理解能力尚存在不足，尤其在处理复杂知识点时易出现偏差；其次，数据训练集的覆盖面有限，导致部分知识点引用不完整；再者，模型在题型设计和选项平衡方面缺乏精细调优，影响了试题的整体质量^[9]。
2. 模型选择不当，导致人工智能在理解专业领域知识时可能出现偏差，从而引起表述不准确，知识点引用不完整等问题此外，模型训练过程中缺乏对复杂知识点的深度学习，导致其在生成试题时难以把握细微差别，进而影响试题的严谨性和准确性。
3. 输入指令的不合理性，导致生成的结果无法满足既定要求^[10]。
4. 投入不足，人工智能未经过专业数据训练，尚未形成完善的认知体系，缺乏对复杂知识点的深度理解，难以精准把握试题设计的核心要求。

四、问题改进

1. 更换人工智能模型，引入更专业的领域语言模型进行微调，训练铁路相关知识，提高人工智能专业水平^[11]。
2. 综合考量模型的知识理解能力、生成可控性、领域适应性，选择恰当的大模型。
3. 不断调整语言命令，提高人工智能理解能力，让试题更符合实际需求。在生成试题后，新增人工智能自动进行“知识点完整性验证”，在试题生成节点之后，增加重复率识别的判断，若存在重复率不满足条件的试题，则再次进行补充生成。调整知识点分布，改进选项生成逻辑^[12]。
4. 加大人工智能模型训练投入，优化算法，扩大训练集，进一步优化模型结构，增强其对专业知识的理解和应用能力，确保试题生成的全面性和精确性，并加强题型设计及选项平衡的调优，以提高试题生成质量^[13]。同时，增加专家反馈机制，持续修正 AI 理解偏差，确保试题准确性和专业性。针对现有问题，团队已着手调整算法参数，扩充数据集，并引入更多专家参与评审，力求在下一轮测试中显著提升试题质量。通过不断迭代优化，预期 AI 辅助编制的试题将更加精准、全面，满足高标准测试需求^[14]。

五、下一步方向

为了解决上述存在的问题，可以使用混合大模型，利用不同大模型的优势，减少错题错误，同时建立动态优化机制，将每次评审中发现的易错点、歧义表述及知识盲区同步反馈至大模型训练库，通过持续迭代知识图谱与认知模型，提高出题质量。

采用向量库 + 大模型微调的方式，例如，利用 DeepSeek MoE 或 Qwen 模型进行微调，加强模型的推导能力和文本理解能力，利用 Mistral 加对抗生成，实现生成干扰选项，利用 FAISS 检索试题，避免重复。通过相关技术的利用能够实现精准知识定位与可控题目生成，实现专业知识与工作实际进行深度结合，生成高质量试题^[15]。

1. 采用向量库，存储对文件材料进行向量化存储，从而实现通过简单指令快速召回相关手册的内容。
2. 对使用的大模型进行微调，并训练铁路货车相关专业知识，让大模型对铁路相关专业知识的更加“熟悉”。
3. 使用大模型的自然语言处理能力，根据提供的手册内容及试题生成的要求，生成试题。
4. 增加试题校验判断，针对重复性试题或不符合要求的试题，自动进行删除，并根据删除试题的知识点自行补充试题。
5. 当学习文件进行更换后，自动删除相关知识点试题，并重新生成试题，确保使用试题符合最新规范要求。
6. 建立优化机制，对评审中出现的问题反馈至大模型进行不断训练和更新。

六、结论

目前，人工智能命题仍需通过人机协同机制确保试题质量达标。采用人工智能生成初稿、专家审核删除的出题方式，相较于传统的试题编写审核流程，有效缩短了40%以上的试题编制时间。本次测试虽暴露出阶段性技术瓶颈，但已充分验证该系统在考试命题领域的应用价值。通过实施系统性优化方案，重点推进

算法架构优化、知识图谱迭代升级、多维度质量评估体系构建等关键环节的专项提升措施，可有效解决命题效率与质量的平衡问题。当前测试尚处于人工智能技术在试题编写领域的探索性实践阶段，随着自然语言处理模型的持续演进与升级、领域知识库的动态完善以及全流程质控机制的健全，有望实现命题工作的智能化转型，逐步构建人机协同的新型命题模式。

参考文献

[1] 穆肃,周德青,胡小勇,崔冠利. 人工智能技术赋能下乡村教育精准帮扶的实施模式与对策 [J]. 中国电化教育, 2023(9): 18-26.DOI:10.3969/j.issn.1006-9860.2023.09.003.

[2] 孟天广. 智能治理：通用人工智能时代的治理命题 [J]. 高等学校文科学术文摘, 2023, 40(3): 210-210.

[3] 马赫,郭南明,倪小明,等. 一种基于人工智能识别的辅助命题审题检索办法及系统 :202311730244[P][2025-04-30].

[4] 刘庆峰. 用人工智能破解“时代命题”[J]. 国企管理, 2021, 000(004): P.82-83, 3.

[5] 张晓君女,周正,王琪瑶女. 面向人工智能的命题动态逻辑及其扩展研究 [M]. 浙江大学出版社, 2022.

[6] 唐邦顺,丁继达,袁汉邦. 利用生成性人工智能技术提升英语七选五选项的命题质量 [J]. 教学考试, 2024, (30): 54-60.

[7] 刘博佳. 人工智能技术在网络化教育考试领域的应用价值研究 [J]. 电脑知识与技术, 2022, 18(29): 83-85.DOI: 10.14004/j.cnki.ckt.2022.1902.

[8] 方兴. 方兴：如何使信息设计的产出更有效地为受众服务是首要命题 [J]. 设计, 2021, 34(14): 52-55.

[9] 黄荣怀,王运武,焦艳丽. 面向智能时代的教育变革——关于科技与教育双向赋能的命题 [J]. 中国电化教育, 2021, (07): 22-29.

[10] 惠志斌,李佳. 人工智能时代公共安全风险治理 [M]. 上海社会科学院出版社 :202105.226.

[11] 陈汝东,李净. 论智能技术发展的伦理向度与规制框架 [J]. 新闻爱好者, 2025, (05): 4-9.DOI:10.16017/j.cnki.xwzhz.2025.05.003.

[12] 马磊. 人工智能作文的理性审视与教学对策 [J]. 语文教学通讯, 2025, (13): 24-26.

[13] 丁文丽. 人工智能时代意识形态话语权建构研究 [D]. 中共北京市委党校, 2025.

[14] 徐颖. 人工智能赋能教育研究范式转型的路径研究 [D]. 湖北师范大学, 2025.

[15] 李佳,盘延明,梁爱丽,等. 人工智能共生园林职业教育的嬗变演进、胁迫共生与进化拟态 [J]. 中南农业科技, 2025, 46(04): 213-218.