

生成式AI赋能初中数学协同评价理论模型与实现路径研究

江 涛, 陈顺星

黄冈师范学院数学与统计学院, 湖北 黄冈

收稿日期: 2025年11月11日; 录用日期: 2026年1月26日; 发布日期: 2026年2月5日

摘要

在教育数字化与人工智能深度融合的背景下, 传统初中数学评价亟需向过程性、精准化与发展性转型。本研究基于生成式AI, 参考人机协同评价理念, 构建了一个面向初中数学课堂的生成式AI赋能的人机协同评价理论模型。该模型以数学核心素养为导向, 设计“能力层 - 二级维度层 - 具体指标层”三级评价维度体系, 并融合教师、学生与AI三方主体, 形成“准备 - 评价 - 反思”三阶段闭环流程。通过结构化量表与大语言模型的协同, 实现对数学抽象、逻辑推理、建模应用等高阶思维的精细化诊断与个性化反馈。结合“一次函数资源比较”案例, 展示了模型在真实教学场景中的可操作性与应用潜力。研究表明, 该框架有助于推动评价从“结果判定”转向“过程支持”, 为构建“师 - 机 - 生”三元协同的智慧教育新生态提供理论支撑与实践路径。

关键词

生成式AI, 人机协同评价, 初中数学, 数学核心素养

Research on the Theoretical Models and Implementation Pathways of Collaborative Assessment in Middle School Mathematics Empowered by Generative AI

Tao Jiang, Shunxing Chen

School of Mathematics and Statistics, Huanggang Normal University, Huanggang Hubei

Received: November 11, 2025; accepted: January 26, 2026; published: February 5, 2026

Abstract

Against the backdrop of deep integration of educational digitalization and artificial intelligence,

文章引用: 江涛, 陈顺星. 生成式AI赋能初中数学协同评价理论模型与实现路径研究[J]. 创新教育研究, 2026, 14(2): 38-47. DOI: 10.12677/ces.2026.142093

traditional evaluation in junior high school mathematics urgently needs to transform towards process-oriented, precise, and developmental paradigms. Based on generative AI and the concept of human-computer collaborative evaluation, this study constructs a theoretical model of generative AI-enabled human-computer collaborative evaluation for junior high school mathematics classrooms. Oriented by core mathematical competencies, the model designs a three-level evaluation system comprising “capability level-secondary dimension level-specific indicator level”, and integrates three stakeholders—teachers, students, and AI—to form a closed-loop process of “preparation-evaluation-reflection”. By combining structured rubrics with large language models, the framework enables fine-grained diagnosis and personalized feedback on high-order thinking skills such as mathematical abstraction, logical reasoning, and modeling application. A case study on “comparing charging plans using linear functions” demonstrates the model’s operability and application potential in real teaching scenarios. The research indicates that this framework helps shift evaluation from “outcome judgment” to “process support”, providing both theoretical foundations and practical pathways for building a smart education ecosystem of “teacher-machine-student” ternary collaboration.

Keywords

Generative AI, Human-AI Collaborative Assessment, Middle School Mathematics, Mathematical Core Competencies

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

教育部召开的国家教育数字化战略行动 2025 年部署会，标志着我国教育数字化进程进入了新阶段。此次会议以“人工智能与教育变革”为主题，围绕落实《教育强国建设规划纲要(2024~2035 年)》进行了系统部署，体现了国家对人工智能赋能教育强国建设的高度重视。作为推动教育生态深刻变革的关键力量，人工智能技术在教育评价体系革新以及实现规模化教育与个性化培养有机融合方面展现出巨大潜力 [1]。自 2022 年以来，生成式 AI 在文本生成领域迅猛发展，其卓越的自然语言处理能力，使其在教学评估、即时反馈等教育场景中展现出独特优势，为构建“在规模化基础上实现高度个性化”的新型教育形态提供了坚实的技术支撑。

本研究系统梳理生成式 AI 赋能教育评价的研究现状。宛平与顾小清(2024)提出的生成式 AI 支持的人机协同评价实践模式，为构建初中数学协同评价模型提供了实践范式参照[1]；孙丹等(2024)对大学生编程学习行为的分析，揭示了生成式 AI 在解析复杂学习过程、识别思维模式方面的技术潜力[2]；刘邦奇等(2024)从技术框架与应用趋势维度，论证了生成式 AI 重塑未来教育形态的可能性[3]；吴砾等(2023)则强调智能技术对教育评价改革的赋能作用，为模型的技术路径选择提供了政策与实践依据[4]。彭绍东(2023)提出的 AIGC 时代教育创新框架[5]，刘明等(2023)对大语言模型教育应用原理与挑战的剖析，卢宇等(2023)以 ChatGPT 为例的具体应用展望[6]，以及方海光等(2022)对人机协同教育理论的探讨[7]，共同构建了本研究的理论基础与前沿视角。本研究的理论出发点在于，将生成式 AI 的技术优势与建构主义“情境认知”理论、认知负荷理论等经典学习理论深度融合，以数学核心素养为导向，构建“师 - 机 - 生”三元协同的评价新范式。创新点体现为：首创“能力层 - 二级维度层 - 具体指标层”三级评价维度体系，实现数学核心素养的可操作化测量；设计“准备 - 评价 - 反思”闭环流程，推动评价从“结果判定”向“过程支持”转型；创新人机协同机制，通过 AI 初评、教师复核、学生修改的动态交互，实现技术智能与人类智慧的优势互补。

鉴于上述背景，本文尝试构建一个面向初中数学课堂的生成式 AI 赋能的人机协同评价模型。该模型通过细化数学核心素养导向的评价维度，整合结构化规则与大语言模型的语义推理能力，优化“学生作答 - AI 初评 - 教师复核 - 动态反馈”的闭环流程，实现评价从“结果判定”向“过程支持”转变，提升课堂评价的精准性、及时性与发展性。

2. 生成式 AI 应用于初中数学的协同评价

2.1. 人机协同评价的内涵与特征

人机协同评价是指在教育评价过程中，人类教师与人工智能系统基于各自优势分工协作、动态互补，共同完成对学生学习过程与成果的诊断、判断与反馈的一种新型评价范式[5]。在初中数学教学语境下，该模式强调将教师对数学思维逻辑、学生认知发展规律及课堂情境的深刻理解，与生成式 AI 在数据处理、模式识别、多轮交互和即时反馈等方面的计算智能有机融合。

其核心特征体现为三点：一是协同性，即教师主导评价目标与价值判断，AI 承担重复性高、规则性强的分析任务；二是过程性，聚焦学生解题思路、推理链条、错误类型等动态学习证据，而非仅关注答案正误；三是适应性，通过 AI 对个体作答行为的持续追踪与建模，支持教师实施差异化反馈与精准干预。这种“教师智慧 + 机器智能”的深度融合，为人大共育提供了评价层面的方法论支撑。

2.2. 人机协同评价的优势

(1) 提升评价效率，释放教师专业精力

在初中数学教学中，日常练习、作业和小测数量庞大，传统人工批改耗时耗力。生成式 AI 能够快速处理大量学生作答，自动完成格式检查、答案比对、常见错误识别等基础性任务，显著提升评价效率。教师由此从重复性劳动中解放出来，将更多时间投入到教学设计、个别辅导和高阶思维引导等更具教育价值的工作中。

(2) 增强诊断能力，实现精准学情分析

生成式 AI 不仅判断答案正误，更能基于对数学语言(如代数表达、几何推理、函数描述)的理解，深入分析学生的解题思路、逻辑链条与概念运用情况。例如，系统可识别“步骤跳跃”“概念混淆”“模型误用”等典型问题，并生成结构化诊断报告，帮助教师精准把握班级整体与个体学生的认知薄弱点。

(3) 支持个性化反馈，促进因材施评

依托对学生历史作答数据的持续追踪，生成式 AI 可为不同水平的学生提供差异化、适切性的反馈建议，如针对性提示、追问引导或变式练习推荐。这种“一人一策”的评价方式，使反馈更具发展性和激励性。教师在此基础上融入情感关怀与教学经验，进一步优化反馈质量，真正实现“以评促学”。

(4) 推动评价理念转型，落实素养导向

人机协同评价强调过程性、发展性和综合性，契合《义务教育数学课程标准(2022 年版)》提出的“教 - 学 - 评一体化”与核心素养导向要求。通过多维度采集学生在问题解决、推理表达、模型建构等方面的表现，评价从单一的结果甄别转向支持学生数学素养持续进阶的学习支架，助力育人目标落地。

3. 生成式 AI 赋能的人机协同评价模型构建

基于当前生成式人工智能技术的发展趋势与初中数学教育评价改革的现实需求，本研究参考宛平与顾小清提出的人机协同评价实践模型，构建了一个基于生成式 AI 的初中数学人机协同评价模型(如图 1 所示)。该模型从评价理念、评价维度、评价方式、评价技术与评价流程五个方面系统开展评价实践，旨在实现人工智能与数学核心素养导向评价的有机融合，推动初中数学评价向过程性、精准化与

协同化方向发展。

3.1. 评价理念

人机协同教育强调, 在由人类教师、学生与智能机器共同构成的复杂三元教育空间中, 人工智能不再仅是工具或辅助手段, 而是被赋予认知主体地位的协同参与者[1]。本研究在人机协同教育理论指引下, 将生成式 AI 深度融入初中数学评价体系, 旨在构建一种既能发挥技术智能、又彰显教育人文性的新型评价范式。

此外, 数学学习本质上并非静态的知识接收过程, 而是一个高度动态、循环迭代的认知建构过程。正如现代数学教育理论所强调, 学生的数学理解是在问题解决、推理验证、反思修正等活动中逐步形成的, 具有显著的过程性、社会性与创造性特征。在此过程中, 即时、精准、发展性的评价反馈至关重要, 它不仅帮助学生识别思维盲区, 更能引导其主动调整解题路径, 实现从“会做题”到“会思考”的跃迁。

鉴于此, 以数学学习的过程观为基础, 将生成式 AI 技术有机嵌入初中数学评价的全过程构建的评价模型覆盖了“任务准备 - 解题实施 - AI 初评 - 教师复核 - 学生修改 - 再评估 - 总结提升”等关键环节, 以实现评价流程与认知过程的无缝对接。

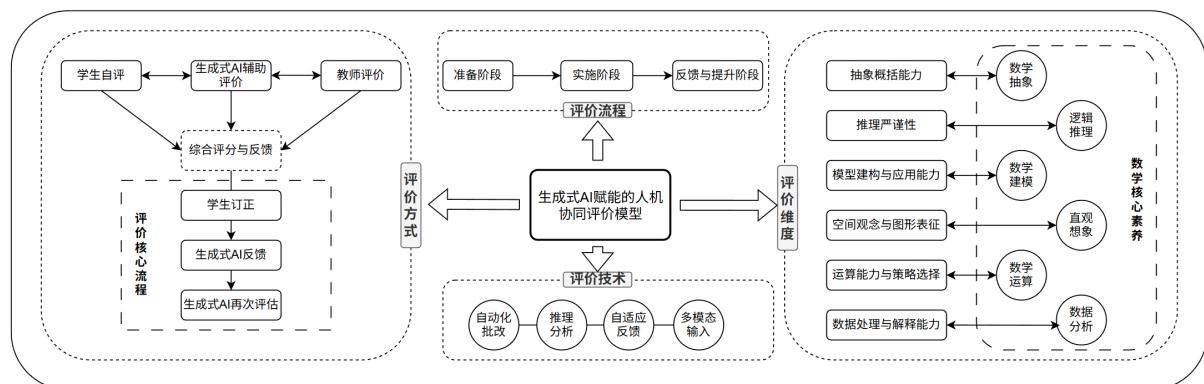


Figure 1. Illustration of a generative AI collaborative assessment model

图 1. 生成式人工智能协同评价模型示意图

3.2. 评价维度

见表 1 所示, 本模型的评价维度采用三级框架结构——即能力层、二级维度层和具体指标层。其中, 能力层以《义务教育数学课程标准(2022 年版)》所确立的数学核心素养为根本依据, 凝练为六大关键能力: 数学抽象、逻辑推理、数学建模、直观想象、数学运算与数据分析。这六大能力不仅体现了初中阶段数学学习的本质要求, 也构成了本评价体系的价值锚点与顶层架构。

Table 1. Evaluation dimensions of the human-AI collaborative assessment model

表 1. 人机协同评价模型的评价维度

数学核心素养	能力维度	二级维度	具体指标
数学抽象	抽象概括能力	情境识别能力	能从实际问题中提取关键数量关系或结构特征
		符号化表达能力	正确使用字母、符号表示变量、公式或规律
		一般化归纳能力	能从特例中总结通式, 区分本质与非本质属性

续表

		步骤连贯性	解题步骤环环相扣, 无逻辑跳跃
逻辑推理	推理严谨性	前提充分性	所有推导均有明确依据
		结论合理性	最终结论与问题要求一致, 未扩大或缩小适用范围
数学建模	建模建构与应用能力	问题转化准确性	正确识别已知量与未知量, 建立符合题意的表达式
		模型假设合理性	对现实情境做出合理简化
		结果解释适切性	能结合实际背景解释数学结果的意义
		图形识别与绘制能力	能准确画出函数图像、几何图形, 并标注关键元素
直观想象	空间观念与图形表征	变换与位置关系理解	理解平移、旋转、对称等变换对图形的影响
		数形结合运用能力	能通过图像辅助分析代数问题
数学运算	运算能力与策略选择	计算准确性	数值计算、符号运算无错误
		步骤规范性	书写格式规范, 单位、等号对齐, 过程完整
		策略灵活性	能根据问题特点选择最优算法
		信息提取能力	能从表格、统计图中准确读取有效数据
数据分析	数据处理与解释能力	统计量计算能力	正确计算平均数、中位数、众数、频率等
		数据推断与决策能力	基于数据分析得出合理结论并提出建议

首先, 数学抽象与逻辑推理共同构成学生高阶思维的基础。数学抽象强调从具体情境中提炼数量关系与空间形式的能力, 而逻辑推理则关注论证过程的严密性与因果链条的完整性。鉴于此, 能力层将二者分别对应为“抽象概括能力”与“推理严谨性”两大维度。在此基础上, 二级维度层是在深入分析初中生数学解题认知过程与常见错误类型的基础上, 对能力层各项要素进行的操作化分解。例如, “逻辑推理”能力被细化为推理的连贯性、前提的充分性、结论的合理性三个二级维度; “数学建模”则分解为问题转化的准确性、模型假设的合理性、结果解释的适切性; “数学运算”进一步拆解为计算的准确性、步骤的规范性、策略的灵活性等。具体指标层则是依据二级维度, 进一步提炼出可观察、可测量、可由生成式AI辅助识别的评价要点。例如以“建模建构与应用能力”中的“问题转化的准确性”为例, 其具体指标包括: ① 能否正确识别问题中的已知量与未知量; ② 是否建立符合题意的数学关系(如方程、函数或不等式); ③ 变量定义是否清晰、单位是否统一。这些具体指标既为教师提供了清晰的评价参照, 也为生成式AI的语义理解与反馈生成提供了结构化输入, 从而实现人机在评价尺度上的共识与协同。

综上, 其通过“能力层-二级维度层-具体指标层”的三级递进结构, 将抽象的数学核心素养转化为可实施、可反馈、可优化的评价体系, 既保障了评价的专业性与科学性, 又为生成式AI的深度参与奠定了内容基础。

3.3. 评价方式

本模型基于人机协同教育理念, 融合学生自评、教师评价以及生成式人工智能辅助评价三大主体, 构建了一个多维度、动态化、发展性的初中数学课堂评价方式。首先, 学生作为学习的主体, 通过自评积极参与到评价过程中, 不仅能够反思自身的解题思路、策略选择与表达规范, 还能在与标准答案或AI反馈的对比中识别认知偏差与思维盲区。

其次, 教师作为专业引导者, 在评价体系中发挥着不可替代的核心作用。凭借对数学学科本质、学生认知发展规律及课堂情境的深刻把握, 教师能够对AI初步生成的诊断结果进行价值判断与人文调适,

尤其在评估高阶思维品质(如创新性解法、模型迁移能力)和情感态度(如解题毅力、合作意识)等方面, 确保评价的专业性、公平性与教育性。

生成式人工智能则作为智能协作者, 依托其强大的多模态理解能力(包括对自然语言、数学符号、几何图形的联合解析)、逻辑推理引擎与大数据分析技术, 对学生的数学作答进行结构化解析。它不仅能自动识别计算错误、符号误用、步骤缺失等表层问题, 更能深入追踪解题逻辑链, 判断推理是否严密、建模是否合理、结论是否可靠。此外, 生成式 AI 还可通过比对学生历史作答数据, 动态识别个体认知模式, 并结合课程标准自动生成个性化提示, 如“请检查方程建立是否遗漏约束条件”或“你的证明缺少对全等三角形判定依据的说明”。尤为关键的是, 该系统具备持续学习与优化能力, 能够不断吸收教师修正意见与学生修改行为, 迭代更新其评价模型, 从而不断提升诊断的准确性与反馈的适切性。

这一多元协同的评价方式, 不仅为学生提供了即时、具体、多层次的数学学习反馈, 更有效激发其元认知意识与自主改进意愿。学生在接收来自 AI 的初步诊断与教师的深度点评后, 需结合自身理解进行反思性修改——例如重梳推理过程、优化解法路径或完善表达规范。在此过程中, 生成式 AI 依托自适应反馈机制, 推动形成“作答 - AI 初评 - 教师复核 - 学生修改 - 再评估 - 提升”的高效闭环, 使数学评价真正从“终结性判断”转变为“发展性支持”, 助力学生在持续迭代中实现数学思维的深化与核心素养的进阶。

4. 评价模型的实现路径与应用案例

4.1. 实现路径

本模型的评价流程系统地融合了教师、生成式人工智能(AI)及学生三大核心主体, 通过精心设计的三个关键阶段——准备阶段、评价阶段、总结反思阶段——确保整个评价过程具备全面性、精准性与发展性。每个阶段均清晰界定各主体的角色任务与互动关系, 为初中数学课堂的人机协同评价提供了科学、可操作的实施路径。

4.1.1. 准备阶段

在正式开展数学任务评价之前, 准备阶段是保障后续评价有效性的前提。该阶段的核心目标是明确评价导向、统一评价标准、激活学生认知准备。教师需根据教学目标与学生学情, 设计具有真实情境的数学任务(如建模问题、几何证明或开放探究题), 并制定与数学核心素养对应的评价量表, 明确各维度的具体要求。同时, 教师向生成式 AI 输入任务描述、评分细则及班级整体认知特征等指令。生成式 AI 基于其语义理解与教育知识库, 对任务难度、常见错误类型及潜在思维障碍进行预判, 并为教师提供优化建议。学生则在此阶段认真阅读任务要求, 理解评价标准, 并完成初始解题作答, 为后续评价奠定内容基础。

4.1.2. 评价阶段

评价阶段细分为初评、复评、反馈与修改、终评四个递进环节, 体现“诊断 - 分析 - 改进 - 确认”的闭环逻辑。

在初评环节, 生成式 AI 首先自动接收学生提交的多模态作答(包括文字、公式、手绘图形等), 利用符号识别、逻辑推理与自然语言处理技术, 对解题过程进行初步解析, 识别计算错误、逻辑漏洞、表达不清等问题, 并生成结构化初评报告。

进入复评环节, 生成式 AI 将初评结果与师生反馈数据进行融合, 结合预设权重, 通过算法模型进行综合再评估, 输出包含分项得分、错误归因、改进建议与拓展提示的精细化评价报告。

在反馈与修改环节, 学生与生成式 AI 展开多轮交互: 通过点击提示、追问澄清或尝试新解法, 学生

在 AI 的启发式引导下主动修正错误、优化策略、完善表达。教师则适时介入，对共性难点进行集中讲解，或对个别学生提供情感支持与高阶思维引导，确保反馈被有效内化。

在终评环节，学生将修改后的终稿重新提交至系统。生成式 AI 对比初稿与终稿，从进步幅度、错误修正率、思维深度提升等维度进行定量与定性综合分析，生成学习成长报告。教师据此进行最终评定，并纳入过程性档案，作为后续教学调整的重要依据。

4.1.3. 总结反思阶段

作为评价流程的收尾环节，总结反思阶段聚焦于元认知提升与系统优化。学生需回顾整个解题与修改过程，撰写简要反思日志，强化自我监控能力。教师则对本轮评价全流程进行复盘：一方面分析生成式 AI 在错误识别、反馈有效性等方面的表现，评估其与教学目标的契合度；另一方面收集学生对反馈内容、交互体验的意见，识别流程中的不足。基于此，教师可动态调整任务设计、量表权重或人机分工策略，持续优化评价体系的科学性与育人效能。通过三阶段、多主体、闭环式的评价流程，不仅实现了对数学学习过程的全程追踪与精准干预，更将评价转化为促进学生思维发展与教师专业成长的双向赋能机制。

4.2. 应用案例

为验证本模型在真实教学场景中的可行性与有效性，本文选取八年级下册《一次函数的应用》单元中的典型任务——“手机套餐资费比较”作为实施案例，完整呈现评价流程的运行机制。

(1) 准备阶段

教师围绕“数学建模”与“数据分析”两大核心素养，设计如下真实问题情境：某运营商提供 A 套餐(月租 30 元，通话 0.2 元/分钟)与 B 套餐(无月租，0.5 元/分钟)，要求学生建立数学模型，分析不同通话时长下的最优选择。同时，教师制定结构化评价量表，明确“变量定义”“函数关系”“图像/表格辅助”“结论解释”等二级维度的具体指标，并将任务描述与评分标准输入生成式 AI 系统。AI 基于语义解析预判学生易忽略定义域及图像表达，建议教师在任务说明中强调“可借助图像辅助分析”。学生据此理解任务要求，完成初稿作答并提交至平台。

(2) 评价阶段

在初评环节，生成式 AI 对学生的多模态作答进行自动解析。以某生初稿为例，其正确建立函数模型 $y_1 = 0.2x + 30$ 与 $y_2 = 0.5x$ ，并通过解不等式得出“当 $x > 100$ 时选 A 套餐”的结论。然而，AI 识别出三项缺失：① 未限定自变量 $x \geq 0$ ；② 未说明 $x = 100$ 时费用相等；③ 缺乏图像或表格支持。系统随即生成初步反馈：“你的计算正确！但请补充定义域、交点意义，并尝试画图让结论更直观。”与此同时，学生开展自评，对照量表发现“图像辅助”维度未达标；教师抽样审阅后确认该问题具有普遍性。

进入复评与反馈修改环节，AI 融合教师标注的共性难点，优化反馈策略，在后续交互中主动引导图像绘制。该生追问“如何作图？”AI 即时响应：“横轴为通话时间(分钟)，纵轴为费用(元)，绘制两条直线，交点为(100, 50)。”学生据此补画函数图像，并完善结论表述：“当 $0 \leq x < 100$ 时，B 套餐更经济； $x = 100$ 时费用相同； $x > 100$ 时 A 套餐更优。”教师在后台监控修改过程，给予正向激励。

在终评环节，AI 对比初稿与终稿，量化识别其在“结果解释”与“表征多样性”维度的显著进步，生成成长报告。教师结合该报告，给予最终评分(92/100)并撰写发展性评语：“能主动完善模型表达，具备良好的反思意识！”

(3) 总结反思阶段

学生在学习日志中反思：“原来光算对还不够，还要说清楚、画明白。AI 提醒我漏了图像，现在我真正理解了数形结合的价值。”教师则基于全班数据复盘：一方面肯定 AI 在诊断“隐性缺失”(如图像

缺位、定义域忽略)方面的优势; 另一方面发现学生对实际问题中变量范围的敏感度不足, 计划在后续教学中增设专项训练。同时, 教师建议优化 AI 提示词库, 以更好识别和鼓励创新性解法(如使用表格对比)。

本案例表明, 该评价模型通过“准备-评价-反思”三阶段闭环, 有效实现了对学生数学建模过程的全程追踪、精准干预与发展性支持。生成式 AI 不仅提升了评价效率, 更成为促进学生高阶思维发展的智能支架, 而教师则在人机协同中持续发挥专业引领与人文关怀的核心作用。

5. 技术局限性与应对策略及实施挑战与伦理考量

在推进生成式人工智能(Generative AI)深度融入初中数学教育评价体系的过程中, 对其技术局限性、实施障碍及伦理风险进行客观审视, 是确保模型科学、公平、可持续落地的关键前提[8]。本研究立足“以人为本”的教育价值导向, 从技术瓶颈、现实挑战与伦理治理三个维度出发, 构建“问题识别-策略应对-制度保障”的闭环逻辑框架, 旨在为“师-机-生”三元协同的智慧评价生态提供理论支撑与实践路径。

5.1. 技术局限性的多维表征与优化策略

当前生成式 AI 在数学作答解析、逻辑推理与反馈生成等核心环节仍面临显著技术瓶颈, 主要体现为以下三方面:

其一, 多模态融合识别精度不足。数学作答常涉及符号表达、手绘图形与文本说明的混合形态, 而现有 AI 系统在处理非结构化输入时存在识别盲区。例如, 学生手绘函数图像的坐标轴偏移、几何证明中辅助线标注不清等问题, 易导致 AI 对关键要素定位失准; 同时, 符号逻辑链中的隐含前提或跳跃步骤难以被完整捕捉, 从而在复杂推理路径追踪中产生“黑箱式”误判。

其二, 逻辑推理可解释性薄弱。尽管 AI 能够识别答案错误或步骤缺失, 但其反馈往往缺乏对思维断裂点的精准溯源。例如, 在代数推导或几何证明中, AI 难以区分“概念混淆”“规则误用”或“逻辑跳跃”等不同性质的错误类型, 削弱了教师对学生认知过程进行针对性干预的能力。

其三, 反馈生成趋于模板化。受训练数据分布与生成机制限制, AI 倾向于输出标准化、程式化的改进建议, 难以识别并鼓励学生在解题过程中展现的非常规思路或创造性策略, 存在抑制高阶思维发展的潜在风险。

针对上述问题, 本研究提出四项系统性优化策略:

- (1) 设计“定义域-函数式-图像-结论”四步结构化答题模板, 引导学生规范作答, 降低 AI 对非结构化输入的解析难度;
- (2) 开发教师“一键修正”功能模块, 支持教师对 AI 初评结果进行快速标注、补充说明或覆盖调整, 构建“AI 初评 + 教师调适”的双重校验机制;
- (3) 引入学生反馈评分机制, 允许学生对 AI 建议的适切性、启发性进行动态评分, 形成“评价-反馈-优化”的闭环迭代;
- (4) 融合符号计算引擎(如 SymPy)与高精度图像识别算法(如基于 CNN 的几何图元检测), 提升数学图形绘制的坐标精度与多模态语义对齐能力。

5.2. 实施推广中的现实挑战与协同应对机制

模型的大规模应用需直面教育场域中的结构性不平等与资源约束。主要挑战包括: 教师 AI 素养参差不齐、区域间算力与网络基础设施差异显著、城乡学校数字鸿沟持续存在等。

为此, 本研究构建三级协同支持体系: 首先, 建立“基础操作-评价设计-伦理规范”三阶教师培

训体系，通过工作坊、微认证与校本研修相结合的方式，系统提升教师对 AI 评价工具的操作能力、任务设计能力及伦理判断力；其次，依托区域教育云平台实现算力资源池化共享，采用“轻量化 AI 部署”方案(如边缘计算 + 云端协同架构)，适配不同学校的硬件条件，降低技术准入门槛；最后，建立动态资源调配与技术支持响应机制，优先保障偏远地区学校的模型更新、故障响应与专业指导服务，推动教育数字化成果的普惠共享。

5.3. 伦理风险防控与“技术向善”制度保障

在数据隐私保护方面，严格遵循“最小必要”原则，仅采集与评价直接相关的作答数据；所有学生信息实行匿名化处理，并采用端到端加密存储；制定《学生数据使用规范》，明确教师、AI 系统与平台运营方的数据访问权限边界，并通过家长知情同意书强化程序合法性。

在算法公平性保障方面，引入多维度公平性评估框架(如 Demographic Parity、Equalized Odds 等指标)，定期检测 AI 评价结果是否因性别、地域、学业水平等因素产生系统性偏差；同时，利用区块链技术对评价全流程(包括原始作答、AI 评分、教师修正、学生反馈)进行不可篡改存证，实现操作可追溯、责任可认定，增强评价过程的透明度与公信力。

6. 结语

本研究结合生成式 AI 技术的发展与初中数学评价改革的需求，构建了一个基于生成式 AI 的人机协同评价模型。该模型整合教师的专业判断与 AI 的数据处理能力，从评价理念、维度、方式、技术和流程五个方面提出了一种融合数据驱动与社会交互的新型评价框架。通过引入数学核心素养导向的三级评价维度及“准备 - 评价 - 反思”的闭环流程，实现了对数学学习过程的结构化刻画与智能化支持。

尽管尚未进行大规模实证检验，但通过一次函数建模任务的应用案例，展示了该模型在真实教学中的可操作性和潜在价值。提出的评语生成与反馈机制旨在为教师提供智能辅助，推动评价从“结果判定”转向“过程引导”，促进精准化和个性化的数学教育评价。

随着 AIGC 工具逐渐融入教育生态，教师需主动适应技术变革，在减轻评价负担、优化反馈策略等方面提升专业素养，并审慎应对 AI 带来的挑战。未来教育应向“师 - 机 - 生”三元协同结构发展，促进人机优势互补与认知协作。本研究提出的理论框架为迈向“人类智能”与“人工智能”双脑协同的教育新思维奠定了基础。

基金项目

本文系黄冈师范学院 2025 年研究生工作站课题“生成式 AI 赋能初中数学课堂评价研究”的研究成果，课题编号：5032025014。

参考文献

- [1] 宛平, 顾小清. 生成式人工智能支持的人机协同评价: 实践模式与解释案例[J]. 现代远距离教育, 2024(2): 33-41.
- [2] 孙丹, 朱城聪, 许作栋, 等. 基于生成式人工智能的大学生编程学习行为分析研究[J]. 电化教育研究, 2024, 45(3): 113-120.
- [3] 刘邦奇, 聂小林, 王士进, 等. 生成式人工智能与未来教育形态重塑: 技术框架、能力特征及应用趋势[J]. 电化教育研究, 2024, 45(1): 13-20.
- [4] 吴砾, 郭庆, 吴龙凯, 等. 智能技术赋能教育评价改革[J]. 开放教育研究, 2023, 29(4): 4-10.
- [5] 彭绍东. AIGC 时代基于双向赋能的人工智能教育创新框架[J]. 教育文化论坛, 2023, 15(4): 12-26.
- [6] 刘明, 吴忠明, 廖剑, 等. 大语言模型的教育应用: 原理、现状与挑战——从轻量级 BERT 到对话式 ChatGPT [J].

- 现代教育技术, 2023, 33(8): 19-28.
- [7] 方海光, 孔新梅, 李海芸, 等. 人工智能时代的人机协同教育理论研究[J]. 现代教育技术, 2022, 32(7): 5-13.
- [8] 卢宇, 余京蕾, 陈鹏鹤, 等. 生成式人工智能的教育应用与展望——以 ChatGPT 系统为例[J]. 中国远程教育, 2023, 43(4): 24-31+51.