

智能体技术赋能数学分析教学改革的实践探索

——以WorkBuddy智能体平台为例

吴维新

新疆大学数学与系统科学学院, 新疆 乌鲁木齐

收稿日期: 2026年9月20日; 录用日期: 2026年10月17日; 发布日期: 2026年10月23日

摘要

数学分析是数学类专业最重要的基础课程之一, 其理论抽象、逻辑严密, 历来是教与学的难点。本文在系统梳理人工智能在数学教育中应用脉络的基础上, 以WorkBuddy智能体平台为例, 探讨智能体技术赋能数学分析教学改革的实践路径。文章首先给出该平台在数学符号处理与逻辑推理方面的能力分层: 它能稳定完成公式排版、定义复述、常规证明套路与逐句解说, 而在量词次序、构造性证明与一致性全局估计上则易出错; 继而以“一致连续性”教学为例, 呈现未经修饰的人机对话实录、智能体生成的分层习题与反馈原文, 并逐项作出专业评析; 再以一个含参变量反常积分的反面案例, 完整展示智能体的出错过程, 剖析其成因并提出教学警示。实践表明, 智能体能够在保持教师主导地位的前提下承担资源整合、分层辅导、评价辅助等程序性工作, 是推动数学分析课程数字化转型的一条可行路径, 但其输出必须经教师专业审核方可进入课堂。

关键词

数学分析, 智能体, 大语言模型, WorkBuddy, 一致连续性, 教学改革

The Practical Exploration of Empowering the Teaching Reform of Mathematical Analysis with Agent Technology

—A Case Study of the WorkBuddy Agent Platform

Weixin Wu

College of Mathematics and Systems Science, Xinjiang University, Urumqi Xinjiang

Received: September 20, 2026; accepted: October 17, 2026; published: October 23, 2026

文章引用: 吴维新. 智能体技术赋能数学分析教学改革的实践探索[J]. 教育进展, 2026, 16(10): 1639-1645.
DOI: 10.12677/ae.2026.16102270

Abstract

Mathematical Analysis is one of the most essential foundational courses for mathematics majors. Featuring abstract theories and rigorous logic, it has long posed challenges for both teaching and learning. On the basis of systematically sorting out the development of artificial intelligence applications in mathematics education, this paper takes the WorkBuddy agent platform as an example to explore practical approaches for empowering the teaching reform of Mathematical Analysis with agent technology. First, this paper establishes a capability hierarchy of the platform concerning mathematical symbol processing and logical reasoning: the platform can reliably perform formula typesetting, restate definitions, follow standard proof templates, and provide sentence-by-sentence explanations. Nevertheless, it is prone to errors when dealing with quantifier order, constructive proofs, and uniform global estimates. Taking the teaching of “uniform continuity” as a case, we present unedited human-agent conversation transcripts, tiered exercises generated by the agent, and original feedback texts, together with item-by-item professional evaluations. A counterexample involving improper integrals with parameters is then used to fully demonstrate the agent’s error-making process, analyze its root causes, and raise corresponding pedagogical caveats. Practice shows that, while preserving the leading role of instructors, agents can undertake procedural tasks such as resource integration, differentiated tutoring, and assessment assistance, offering a viable pathway to advance the digital transformation of the Mathematical Analysis course. Still, all outputs generated by the agent must undergo professional review by teachers before being used in classroom instruction.

Keywords

Mathematical Analysis, Agent, Large Language Model, WorkBuddy, Uniform Continuity, Teaching Reform

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

数学分析是数学类专业最重要的基础课程之一，是学生后续学习复变函数、实变函数、泛函分析等课程的基石，也是硕士研究生入学考试的必考课程[1]。多年来围绕该课程的教学改革从未间断，学者们在课程衔接、教学方法、能力培养等方面进行了大量探索[2][3]，但“难教、难学”的局面并未根本改观：一方面， $\varepsilon-\delta$ 、 $\varepsilon-N$ 等高度形式化的语言极为抽象，从常量思维到变量思维、从直观描述到严格论证的跨越始终是学习的最大障碍[4]；另一方面，高校扩招以来学生基础差异不断扩大，大班教学环境下因材施教难以落实，教师课后辅导与作业批改负担沉重[2]。传统改革多从内容重组与教法改进入手，受制于教师的时间与精力，难以在“个性化”这一关键环节上突破。

近年来，以 ChatGPT 为代表的生成式人工智能迅猛发展，正在深刻改变教育的形态与生态[5][6]；具备任务规划、工具调用与代码执行能力的新一代智能体(Agent)，其角色正从“辅助工具”转向“教学协作者”[7]。然而，既有研究对智能体在高等数学严格证明训练这一场景中的适用性与风险，尚缺少基于真实交互素材的实证考察。本文以 WorkBuddy 智能体平台为例，在呈现真实人机对话与生成文本的基础上，讨论智能体赋能数学分析教学的可行路径及其边界。

2. 人工智能在数学教育中的应用：述评与定位

2.1. 从计算机辅助教学到智能体的三个阶段

人工智能介入数学教育大致经历了三个阶段。第一阶段是计算机辅助教学(CAI)与计算机代数系统(CAS)时期, Maple、Matlab、GeoGebra 等软件在数值计算、符号运算与图形直观方面发挥作用[2][8], 但功能边界固定、需要专门的操作技能, 本质上仍是“被动工具”。第二阶段是智能导学系统(ITS)时期, 研究者开始关注学习者建模与适应性反馈[9]; 元分析显示其效应量已达 0.76, 与人类一对一辅导的 0.79 相当[10], 但知识表征依赖人工构建, 领域覆盖有限、开发成本高昂。第三阶段是当前的大语言模型与智能体时期: 模型通过自然语言理解意图、拆解任务并生成讲解与反馈[6][11][12], 智能体进一步获得调用工具、执行代码与多智能体协作的能力[7], 使个性化辅导的门槛与成本大幅下降。同时, 数学基准评测表明模型在需要多步符号推演与构造性论证的问题上准确率远低于常规问答任务[13][14], 存在结构性短板。

2.2. 本文的研究定位

与已有工作相比, 本文的特点有三。其一, 场景特定: 聚焦数学分析中以量词结构为核心的严格证明训练, 而非一般的解题训练或知识讲解。其二, 素材真实: 给出未经修饰的人机对话实录与智能体生成的习题、反馈原文, 并逐项作出专业评析, 而非转述或概括。其三, 视角双向: 既展示智能体的能力, 也给出一个出错的反面案例并剖析其成因, 避免对技术的单向赞美。需要说明的是, 本文所说的“智能体”, 特指以大语言模型为认知内核、具备工具调用与代码执行能力、可在多轮对话中自主规划任务的软件系统, 这使其区别于一般意义上的聊天机器人。

3. WorkBuddy 平台的技术特征与数学任务适配性

3.1. 技术架构与教学相关能力

WorkBuddy 是一款桌面端通用智能体平台, 以大语言模型为认知内核, 通过标准化协议接入外部服务, 整合了任务规划、工具调用、文件操作、代码执行与多模态生成等能力。与教学直接相关的有四项: 其一, 自然语言交互与内容生成, 可按提示生成教案初稿、习题变式与知识点讲解; 其二, 外部服务连接, 可在对话中直接调用文献检索与办公文档服务完成资料整理; 其三, 内置代码执行环境, 可现场编写运行程序、绘制函数图像与数据图表; 其四, 多智能体协作与周期化自动任务, 可将教研工作分解为子任务协同完成。与 Maple、GeoGebra 一类需预先制作脚本的被动工具[2][8]相比, 教师只需用自然语言表达需求, 即可由智能体理解意图、拆解任务并交付结果[6][7], 其角色由此从“辅助工具”转变为“教学协作者”。

3.2. 数学符号与逻辑推理的能力边界

在数学任务上, 平台的表现呈现明显的分层。能够稳定处理的是: 数学公式的规范排版(LaTeX/OMML); 定义与定理的标准复述; 常规证明套路, 如用 $\varepsilon-N$ 定义验证简单数列极限、用放缩法估计常见量; 以及对已有证明的逐句解说与启发式反馈。这一层能力已足以承担讲义整理、题解润色与答疑初答。容易出错的则是三类问题。其一, 量词次序的辨识: 一致连续与连续、一致收敛与逐点收敛的区别, 本质上是 $\forall\exists$ 与 $\exists\forall$ 的次序差异, 模型常在改写或举例中悄然交换次序而不自知。其二, 构造性证明中的参数选取: 需要构造具体的反例函数或给出 $\delta(\varepsilon)$ 的显式表达时, 模型倾向于给出形式合理、边界条件却失效的构造。其三, 一致性全局估计与跨段衔接: 长证明中模型可能遗忘前文已确立的条件, 或在局部正确的前提下

形成循环论证，对含参变量积分、函数项级数这类问题错误率尤高。

这一分层提示我们：智能体擅长的是“已有范式的复现与变式”，而不是“新结构的发现与自我验证”。数学分析教学的核心恰在于后者，这从根本上决定了智能体只能定位于协作者而非替代者——其产出必须经教师专业审核与再加工，教师始终是教学的主导者。

4. 教学环节中的应用设计

基于上述能力边界，本文在五个环节设计了智能体的介入方式。课前备课：由智能体汇总高中数学新课标的内容变化、梳理高中教材与数学分析在函数、不等式、数列上的重合与缺位[2]，并针对具体主题生成含概念引入、典型例题、易错辨析的教学设计初稿，供教师作专业修改。课堂教学：利用绘图能力现场演示函数性态(如讲解一致连续时对比演示 δ 随 ε 及基点位置变化的情形)，并把刘徽“割圆术”、《庄子》“一尺之棰，日取其半，万世不竭”等素材转化为问题情境；需要说明的是，杨小锋等的实践表明数学分析教学应以板书为主[2]，可视化材料只能用于呈现板书难以表达的动态过程。课后辅导：学生就课堂遗留问题与智能体对话，要求它针对表述漏洞追问而非直接给出完整证明；教师则让其按基础层(定义直接应用)、提高层(证明技巧与反例构造)、拓展层(联系后续课程)生成分层习题组[9][12]。学业评价：由智能体在教师批改前作初步筛查，按概念性、逻辑性、计算性错误归类，并统计多次作业的错误分布，形成班级共性与个体薄弱点的分析报告[15]。教学研究：借助文献检索与多智能体协作辅助选题论证、案例整理与文字润色[11]。在上述全部环节中，智能体的输出都只是“待审材料”，教师对进入课堂的内容负最终责任。

5. 应用案例：一致连续性的智能体辅助教学

“一致连续性”是数学分析公认的教学难点：学生容易接受“ δ 只依赖 ε 而不依赖点的位置”这一说法，却难以在量词层面区分连续性定义中 δ 对基点的依赖与一致连续性定义中 δ 对基点的独立，更难理解 x^2 在有限区间上一致连续而在 $[0, +\infty)$ 上不一致连续。该主题恰落在3.2节所说的智能体薄弱区，更能检验其高级能力，也更能暴露其局限。

5.1. 人机对话实录

以下是一段未经修饰的交互记录(S为学生，A为智能体；删除去无关寒暄外未作任何改动)。

S1：连续和一致连续到底有什么区别？定义我背下来了，但不知道它在说什么。

A1：区别在量词顺序。 f 在 I 上连续： $\forall x_0 \in I, \forall \varepsilon > 0, \exists \delta > 0, \forall x \in I, |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$ ，这里的 δ 在 x_0 给定之后才取，可依赖 x_0 与 ε 。 f 在 I 上一致连续： $\forall \varepsilon > 0, \exists \delta > 0, \forall x_1, x_2 \in I, |x_1 - x_2| < \delta \Rightarrow |f(x_1) - f(x_2)| < \varepsilon$ ，这里的 δ 在 x_1, x_2 之前取定，只能依赖 ε 。一句话：连续是“逐点地”给 δ ，一致连续是“统一地”给 δ 。

S2：能画出来看看吗？

A2：(执行绘图程序)对 $f(x) = x^2$ 与 $\varepsilon = 1$ ，在基点 $x_0 = 1$ 处可取 $\delta_1 = 0.4$ ，在基点 $x_0 = 10$ 处可取 $\delta_2 = 0.05$ 。可以看到，随着基点右移，同一个 ε 所需的 δ 越来越小并趋于0，因而在 $[0, +\infty)$ 上不存在通用的 δ 。(输出图像)

S3：所以要说明 x^2 在 R 上不一致连续，只要说清楚 δ 会趋于0就够了？

A3：不够，“趋于0”是描述而非论证。要写成反证：取 $\varepsilon_0 = 1$ ，对任意 $\delta > 0$ ，令 $x_1 = 1/\delta$ ， $x_2 = 1/\delta + \delta/2$ ，则 $|x_1 - x_2| = \delta/2 < \delta$ ，而 $|x_1^2 - x_2^2| = (x_1 + x_2)|x_1 - x_2| > 1 = \varepsilon_0$ ，故不一致连续。关键之处是： x_1 、 x_2 的选取依赖于 δ ，这正是定义中 δ 必须“先取”的后果。

S4：那闭区间上，为什么连续就能推出一致连续？

A4: 因为紧致性能把“逐点”变成“有限个”, 而有限个 δ 可以取最小值。具体地说, 对每个 $x \in [a, b]$ 存在 δ_x , 开区间族 $\{(x - \delta_x/2, x + \delta_x/2)\}$ 覆盖 $[a, b]$, 由有限覆盖定理取出有限子覆盖, 再取 δ 为这有限个 $\delta_x/2$ 的最小值即可。

5.2. 智能体生成的分层习题与反馈原文

【基础层】(1) 用定义证明 $f(x) = 3x + 1$ 在 R 上一致连续。(2) 证明 $f(x) = \sin x$ 在 R 上一致连续。

【提高层】(3) 设 f 在 $[0, +\infty)$ 上连续且 $\lim_{x \rightarrow +\infty} f(x) = A$ 存在, 证明 f 在 $[0, +\infty)$ 上一致连续。(4) 证明 $f(x) = x^2$ 在 $[0, 1]$ 上一致连续, 但在 $[0, +\infty)$ 上不一致连续。

【拓展层】(5) 设 f 在 R 上一致连续, 证明 $f(x)/x$ 在 R 上有界。(6) 举例说明: 一致连续的函数其导函数未必有界。

【对一份学生作答的反馈】学生作答(片段): “因为 $|l_2^x - 2_2^x| = |x_1 - x_2| \cdot |x_1 + x_2|$, 而 $|x_1 + x_2|$ 可以任意大, 所以找不到统一的 δ 。”智能体反馈: 方向正确, 抓住了“ $|x_1 + x_2|$ 无界”这一要害。但有两处需要改进: 第一, “可以任意大”是必要条件而不是论证, 需改写为反证形式——给定 ε_0 , 对任意 $\delta > 0$ 构造满足 $|x_1 - x_2| < \delta$ 而 $|l_2^x - 2_2^x| \geq \varepsilon_0$ 的具体两点; 第二, 结论须先限定区间, x^2 在 $[0, 1]$ 上一致连续, 在 $[0, +\infty)$ 上不一致连续, 答题时应先说明所论区间。

5.3. 作者对生成内容的专业评析

关于对话: A1 对两个定义的量词对比准确、表述凝练, 甚至优于部分教材的处理; A3 不仅给出了严格的反证构造, 还点明了“ x_1 、 x_2 的选取依赖于 δ ”这一学生最易忽略的要害, 属高质量的启发式反馈。但有两处需要修正。其一, A2 给出的 $\delta_2 = 0.05$ 未经核算: 对 $x_0 = 10$ 、 $\varepsilon = 1$, 由 $(10 + \delta)^2 - 100 < 1$ 解得 $\delta < 0.0499$, 取 0.05 恰好越界; 课堂演示无碍, 写入板书则会形成错误示范。其二, A4 的论证存在实质性跳跃: 从有限子覆盖“取半径的最小值”并不能保证 $|x_1 - x_2| < \delta$ 的两点落入同一区间, 正确做法应引入勒贝格数或改用 Bolzano-Weierstrass 定理的反证法。这是典型的“骨架完整、关键一步缺失”, 也是最需要教师拦截的一类错误。

关于习题: 基础层与提高层质量良好, (3)(4)为教材标准结论, 梯度合理; (6)有思维价值, 但宜补一句提示(如在 $[0, 1]$ 上考察 $f(x) = \sqrt{x}$)。拓展层(5)则有实质性缺陷: $x \rightarrow 0$ 时 $f(x)/x$ 未必有界, 取 $f(x) = \sqrt{|x|}$ 即为反例(它在 R 上一致连续, 而 $\sqrt{|x|}/x$ 在 $x \rightarrow 0^+$ 时无界); 应改为“证明存在常数 M 使 $|f(x)| \leq M(1 + |x|)$ ”, 或限定 $|x| \geq 1$ 。这类“结论大体正确、条件留有缺口”的错误最隐蔽, 也最需要教师审核。

关于反馈: 质量较高。它没有直接给出完整答案, 而是指出“必要条件而非论证”这一逻辑要害并要求补写反证构造, 符合“启发而非代劳”的原则, 且主动提示了区间限定; 经教师补充一句构造提示后, 已可直接用于课后辅导。

6. 局限性、风险与建议

6.1. 一个反面案例: 含参变量反常积分的一致收敛性

为检验智能体在更高难度内容上的表现, 笔者给出如下任务: 设 $F(y) = \int_0^{+\infty} ye^{-xy} dx$, $y \in [0, 1]$, 讨论该含参变量反常积分在 $[0, 1]$ 上的一致收敛性, 并求 $\lim_{y \rightarrow 0^+} F(y)$ 。智能体的回答(节录)为:

“对任意 $x \geq 0$ 与 $y \in [0, 1]$, 有 $|ye^{-xy}| \leq e^{-x}$ 。由于 $\int_0^{+\infty} e^{-x} dx$ 收敛, 由魏尔斯特拉斯判别法可知该积分在 $[0, 1]$ 上一致收敛。于是极限与积分可交换: $\lim_{y \rightarrow 0^+} F(y) = \int_0^{+\infty} \lim_{y \rightarrow 0^+} ye^{-xy} dx = 0$ 。”

事实核查表明这一回答是错误的。首先, 控制函数 $|ye^{-xy}| \leq e^{-x}$ 并不成立: 取 $y = 1/x$ ($x > 1$ 时 $y \in (0, 1)$), 左端为 e^{-1}/x , 右端为 e^{-x} ; 例如 $x = 2$ 时左端约为 0.184 , 右端约为 0.135 。其次, 直接计算即得 $F(y) = 1$

($y > 0$)而 $F(0)=0$ ，故 $\lim_{y \rightarrow 0^+} F(y)=1$ ，并非0；由于极限函数在 $y=0$ 处不连续，由一致收敛的极限函数连续性定理的逆否命题可知，该积分在 $[0,1]$ 上并不一致收敛。

6.2. 出错成因分析

上述错误可从三个层面解释。其一，控制函数的“样例外推”。模型并未对被积函数在参数域上的上确界作严格分析，而是选取了一个在端点 $y=1$ 处恰好取等、形式上也很“整洁”的 e^{-x} 作为优函数，这是以局部吻合代替全局验证。事实上 $\sup_{y \in [0,1]} ye^{-xy}$ 在 $y=\min(1, 1/x)$ 处取得，当 $x > 1$ 时为 $e^{-1/x}$ ，而 $\int_1^{+\infty} (e^x)^{-1} dx$ 发散，M判别法根本不适用。其二，缺少对结论的自我检验：若一致收敛成立则极限函数应当连续，而被积函数在 $y=0$ 时恒为零使 $F(0)=0$ 、 $y > 0$ 时 $F(y) \equiv 1$ ，间断一目了然。模型未做这一步检验，暴露出大语言模型“前向生成、缺少回验”的机制性弱点。其三，表达的完整性掩盖了推理的跳跃：回答中判别法名称、优函数收敛性、极限交换三步俱全，语气肯定、符号规范，读者若不逐字复核极易接受。容易造成“潘多拉魔盒”效应，在数学场景中正表现为这种“形式完备而实质错误”。

6.3. 教学警示与实施建议

这一案例对教学的警示有三。第一，教师审核必须落到“关键步骤复核”而非只看结论与排版；含参变量积分、函数项级数、一致连续与一致收敛这类涉及量词与一致性的内容是错误高发区，应列为必查项。第二，应训练学生用“必要条件检验”自查：一致收敛蕴含极限函数连续，若算出的极限函数不连续，即可立即判定非一致收敛，这一反查习惯在智能体时代尤为宝贵。第三，错误输出本身是优质的教学资源——把上述案例改编为课堂“找错”练习，学生对一致收敛本质的理解比被动阅读一个正确证明要深刻得多，这或许是智能体时代新增的一种课型。

此外仍需警惕三类风险：生成内容的可靠性风险；学生直接索要完整证明、跳过独立思考而致证明能力被侵蚀[3]；人工智能代笔边界模糊引发的诚信问题与使用条件差异带来的数字鸿沟[16]。相应地提出五条建议：其一，建立教师审核制度，凡进入课堂的智能体产出均须教师把关；其二，明确“启发而非代劳”的使用规则，作业要求附对话记录与反思笔记；其三，坚守板书主导的课堂模式，可视化材料仅作补充[2]；其四，把智能素养纳入教师培训，重点是能力边界的识别；其五，院校层面制定制度规范，明确使用边界与披露要求。

7. 结论与讨论

本文发现有三。第一，智能体的可靠能力集中在“范式复现”层面——公式排版、定义复述、常规证明套路与启发式反馈，而在“结构判断”层面——量词次序、构造性证明、一致性与全局估计——错误率显著上升。第二，经恰当设计，它能承担资源整合、分层辅导、评价辅助等程序性工作，把分层教学与个性化辅导转化为可操作安排。第三，其输出必须经教师审核，而错误输出经转化可成为培养批判性阅读能力的教学资源。

与已有研究相比：张志勇[8]强调可视化在“可见形式”与“抽象形式”之间的桥梁作用，本文案例印证了这一点，但补充了“生成数值须经核算”的限制条件；吴立宝等[15]构建了人工智能赋能课堂评价的框架，本文则给出证明题场景下“智能体评判只能作参考、评价权归教师”的边界；李澄锋等[12]论证了个性化学习在技术上成为可能，本文以分层习题组与对话式追问给出其落地形态；吴依曼等[7]提示了生成内容的可靠性风险，本文的反面案例把这一风险具体化为可逐步核查的数学错误。

本文的不足在于：案例来自单个平台、单门课程的小范围试用，未设对照班，也未作学习成效的量化测量；智能体迭代极快，本文记录的表现具有时效性。后续研究应在更大样本、更长周期上检验其长

期效果,并可围绕“学生对智能体错误的诊断能力”开展实证研究。

参考文献

- [1] 华东师范大学数学系. 数学分析(第四版 上册) [M]. 北京: 高等教育出版社, 2010.
- [2] 杨小锋, 黄冬霞, 刘迎洲. 数学分析的教与学[J]. 农业网络信息, 2015(5): 155-156.
- [3] 胡蓉, 冯小高. 谈数学分析的教与学[J]. 攀枝花学院学报, 2013, 30(1): 98-101.
- [4] 李莉, 谢琳. 关于学习极限概念认知障碍的研究与分析[J]. 数学教育学报, 2006, 15(1): 42-44.
- [5] 赵正杰, 叶阳. 生成式人工智能应用于高校思政教育策略探析——以 ChatGPT 为例[J]. 华声, 2025(1): 100-102.
- [6] 聂政威, 陈晔. ChatGPT/生成式人工智能时代的大学工程教育研究: 机遇、挑战与应对策略[J]. 广西教育学院学报, 2024, 39(6): 102-111.
- [7] 吴依曼, 简国锋, 曾子铠, 廖鑫桦, 陈泳妍. ChatGPT 在高等教育应用的潜能与风险及课程革新研究[J]. 电脑校园, 2025(21): 259-261.
- [8] 张志勇. 高中数学可视化教学: 原则、途径与策略——基于 GeoGebra 平台[J]. 数学通报, 2018, 57(7): 21-24, 28.
- [9] 刘凯, 胡静. 人工智能教育应用理论框架: 学习者与教育资源对称性假设——访智能导学系统专家胡祥恩教授[J]. 开放教育研究, 2018, 24(6): 4-11.
- [10] Van Lehn, K. (2011) The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, **46**, 197-221. <https://doi.org/10.1080/00461520.2011.611369>
- [11] 吴兰岸, 闫寒冰, 黄发良, 等. 大型语言模型在高等教育中的应用分析与现实挑战[J]. 现代教育技术, 2023, 33(8): 29-37.
- [12] 李澄锋, 王晶晶. ChatGPT/生成式人工智能促进研究生课程教学变革[J]. 中国高等教育评论, 2024, 19(1): 59-78.
- [13] Hendrycks, D., Burns, C., Kadavath, S., *et al.* (2021) Measuring Mathematical Problem Solving with the MATH Dataset. arXiv: 2103.03874.
- [14] Cobbe, K., Kosaraju, V., Bavarian, M., *et al.* (2021) Training Verifiers to Solve Math Word Problems. arXiv: 2110.14168.
- [15] 吴立宝, 曹雅楠, 曹一鸣. 人工智能赋能课堂教学评价改革与技术实现的框架构建[J]. 中国电化教育, 2021(5): 94-101.
- [16] 吴丰华, 韩文龙. 大型语言模型对高等教育的影响与中国应对[J]. 高等理科教育, 2024(1): 75-83.