

人工智能驱动下的高校混合教学评价体系变革研究

闫锦熙

西安外国语大学教育学院, 陕西 西安

收稿日期: 2026年5月21日; 录用日期: 2026年6月30日; 发布日期: 2026年7月8日

摘要

在人工智能上升为国家新质生产力的背景下, 高校混合教学评价体系面临“唯分数”惯性、数据利用不足、教师负荷过重和维度单一等现实困境。本文以技术可供性与政策支持为分析框架, 梳理了2018~2026年国家层面从“数据驱动”到“AI赋能评价改革试点”的三阶段政策演进, 并指出当前仍存在的四大短板: AI评价主体地位未确立、伦理规范缺失、复合脑能力指标空白、教师AI素养不足。针对上述问题, 论文提出“政策优化 + AI-ESS系统架构”双轮驱动的改进路径: 政策层面通过立法确权、伦理沙盒、财政常模与标准协同为改革护航; 技术层面构建“全域感知 - 智能决策 - 人机共生”五层AI-ESS系统, 实现评价的过程 - 结果耦合、差异个性化、三元协同、动态生成及复合脑能力多维评价。研究认为, 人工智能不会取代评价, 而将重塑“师 - 生 - AI”评价共同体, 未来需在认知数字孪生、量子隐私计算与生成式评价伦理等方向持续探索, 以技术增强理性、以人文守护价值。

关键词

人工智能, 混合教学, 评价体系, AI-ESS, 政策优化, 个性化评价

Research on the Reform of the Evaluation System for Blended Teaching in Universities Driven by Artificial Intelligence

Jinxi Yan

College of Education, Xi'an International Studies University, Xi'an Shaanxi

Received: May 21, 2026; accepted: June 30, 2026; published: July 8, 2026

Abstract

Against the backdrop of artificial intelligence rising to become a national new quality productivity, the mixed teaching evaluation system in universities is facing practical challenges such as the inertia of “score only”, insufficient data utilization, heavy teacher load, and single dimension. This article takes technology availability and policy support as the analytical framework, and summarizes the three-stage policy evolution from “data-driven” to “AI-enabled evaluation reform pilot” at the national level from 2018 to 2026. It also points out four major shortcomings that still exist: the unclear status of AI evaluation subject, the lack of ethical norms, the blank of composite brain ability indicators, and the insufficient AI literacy of teachers. In response to the above issues, the paper proposes a dual wheel driven improvement path of “policy optimization + AI-ESS system architecture”: at the policy level, reform is escorted through legislative confirmation, ethical sandbox, fiscal norm and standard coordination; at the technical level, a five layer AI-ESS system consisting of “global perception, intelligent decision-making, human-machine symbiosis” will be constructed to achieve process result coupling, personalized differentiation, tripartite collaboration, dynamic generation, and multi-dimensional evaluation of composite brain capabilities. Research suggests that artificial intelligence will not replace evaluation, but will reshape the “teacher-student-AI” evaluation community. In the future, continuous exploration is needed in areas such as cognitive digital twins, quantum privacy computing, and generative evaluation ethics, to enhance rationality with technology and protect values with humanities.

Keywords

Artificial Intelligence, Blended Learning, Evaluation System, AI-ESS, Policy Optimization, Personalized Evaluation

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. 问题提出

混合式教学是通过线上线下相结合的方式开展教学的一种教学模式，常见的混合式教学包括以下四种混合形式：翻转课堂、流动式学习中心、整体流动、弹性混合，自 2020 年线上教学的大规模推广使用为契机，如今已发展成为一种常规教学模式[1]，线上教学时代到来使得混合式教学迅速蓬勃发展，进入加速改革的阶段。2024 年政府工作报告中首次将“开展‘人工智能+’行动”写入年度重点任务并明确提出“深化大数据、人工智能等研发应用，开展‘人工智能+’行动，打造具有国际竞争力的数字产业集群”[2]，意味着人工智能已从“前沿技术”上升为“国家新质生产力”的核心引擎，报告进一步要求深入实施科教兴国战略，人工智能必须深度融入教育主战场，成为撬动“教育强国”建设的战略支点。

与国家战略的高频推进形成鲜明对比的是，混合教学评价体系的现实困境愈发凸显：结果导向的“唯分数”惯性依然强大，过程数据难以纳入正式评价；“标准化”考试难以适配“大规模个性化”培养诉求；教师作为单一评价主体，面对海量线上、线下多模态数据，徒增负荷，评价滞后；评价维度仍停留在“知识-技能”二维，无法覆盖 AI 时代所需的“复合脑能力”（批判性思维、提示工程素养、伦理责任等）等问题[3]。在人工智能上升为国家新质生存力的当下，教育如何与人工智能融合从而适应未来的发展趋势，是急需思考并解决的问题[4]。混合教学评价如何借助技术可供性与政策窗口期，完成从理念、方

法到制度的系统变革，来回应国家和教育的需求，是本文研究的核心问题。

2. 演变动力——技术可供性与政策支持

2.1. 技术可供性

人工智能技术的发展为教育评价体系的变革提供了坚实的技术基础。人工智能技术在数据处理、自然语言处理、机器学习、深度学习等方面具有显著优势，能够实现对学习过程的实时监测、动态分析和个性化反馈。生成式人工智能是指能够基于算法和规则及预训练数据库自主生成内容的人工智能技术，在我国教育研究领域，更常见的是采用 AIGC 的表述[5]，生成式人工智能指具有文本、图片、音频、视频等内容生成能力的相关技术及基于此的应用与服务，其在与人类的交互中已表现出与以往人工智能截然不同的自主交互能力[6]。生成式人工智能(Generative AI)，以下简称 GAI 的技术簇群——以大规模预训练模型为核心，辅以多模态对齐、强化学习人类反馈(RLHF)与提示工程(Prompt Engineering)——正在重塑教育评价的技术基底，其数据可供性、算法可供性、交互可供性等特性给教育评价体系的变革提供了更多的可能性与实施性，从技术上为混合式教学评价体系提供了坚实的基础。

2.2. 政策支持

人工智能作为一项新兴事物的发展离不开国家政策的引导与规范，2020 年以来，国家与地方层面连续发布的多份权威文件，从战略定位、财政投入、标准研制、队伍能力建设四个维度，为人工智能驱动下的混合教学评价体系提供了系统性、持续性的政策支持，本文将按照四个维度对政策进行大致划分。

其中《政府工作报告》2024 首次将“开展‘人工智能+’行动”纳入年度重点任务，明确提出“深化大数据、人工智能等研发应用”，在战略定位上为 AI 全面融入教育评价奠定了国家层面最高政策依据[2]。2025 年中共中央、国务院发布的《教育强国建设规划纲要（2024—2035 年）》在专设促进人工智能助力教育变革章节中，要求建立基于大数据和人工智能支持的教育评价和科学决策制度，并将师生数字素养与 AI 伦理纳入综合素质评价[7]，这提供了一个鲜明的政策信号，混合教学评价将不再是局部试点，而是国家教育治理体系现代化的必答题。教育部办公厅在认证一流本科课程时指出，各高校要以一流课程建设为抓手，主动拥抱教育数字化，推动人工智能与专业教育深度融合，加强教师数字素养与 AI 教学能力培训，培育一批 AI+课程示范案例，形成可复制、可推广的教学改革模式[8]。高校教师的 AI 素养直接关系到混合式教学评价体系改革的实施与实践效果，在教师队伍建设上，2023~2025 年，教育部高校教师网络培训中心实施“高校教师数字素养与 AI 教学评价能力提升专项”，面向全体本科高校专任教师，每年开设 60 学时线上课程，涵盖学习分析仪表盘解读、生成式 AI 评价任务设计、算法公平性自检三大模块，完成培训可计入高校教师继续教育学时[9]。

综合上述政策文本可见，国家层面通过“目标 - 资源 - 规则 - 能力”四维耦合的制度设计，已为人工智能驱动下的混合教学评价体系变革提供了可持续的激励结构，混合教学评价体系不再局限于技术试验，而被纳入国家教育治理现代化的总体进程，其变革动力已由潜在需求转化为制度化的现实推力。

3. 特征演变：人工智能驱动下的评价体系五大转向

3.1. 从“结果评价”到“过程 - 结果耦合评价”

传统混合教学评价以终结性结果为核心，通过期末考试、课程论文等进行评价，其本质是对学习结果的抽样检测，无法反映学生的学习过程与成长轨迹。过程 - 结果耦合评价(Process-Product Coupled

Assessment, PPCA)是指借助学习分析(LA)与多模态感知技术,将学习历程中的微观行为数据(如注视序列、停顿时长、协作话语密度等)与最终学习产出进行动态关联,以揭示学习发生的内在机制,其理论根基来自 Scriven 提出的形成性-终结性评价统一体,经 Black 和 Wiliam 发展为嵌入式评估概念,强调评价即学习本身[10]。生成式 AI 的多模态数据采集与分析能力,使“过程数据的量化评价成为可能”,推动评价体系从结果导向转向过程-结果耦合导向——即以过程数据反映学习规律,以结果数据验证学习成效,二者协同构成完整的评价闭环。

3.1.1. 过程评价的核心维度与技术实现

首先是行为维度。过程评价中通过学生的学习参与度与时间管理能力来判断学生的行为,核心指标包括线上学习时长(与课程平均时长的对比)、视频回放次数(反映知识点掌握难度)、作业提交及时性(是否在截止日期前提交)、论坛互动频率(发帖/回帖次数、被点赞数),根据这些指标设立合格标准,当所得学习指数低于 60 则引发教师干预。

其次是认知维度。即评价学生的知识建构过程与问题解决能力,核心指标包括思维导图修改次数(反映知识结构优化过程)、虚拟实验操作步骤(是否符合规范、是否存在重复操作)、AI 对话 prompt 设计质量(是否能清晰表达需求、是否能通过多轮对话优化结果)。通过学生提交的代码记录,生成式 AI 可以分析学生并形成学生的认知发展报告,指出学生在学习相关知识点时的掌握轨迹。

最后是情感维度。对学生的学情绪状态与学习动机进行评价,核心指标包括课堂发言情感倾向(通过语音情感分析识别积极/消极/中性)、弹幕评论情感极性(通过 NLP 分析满意/困惑/不满)、学习日志中的情绪表达(如这个知识点太难了反映困惑情绪)。通过检测学生在学习平台的弹幕、评论、留言等中的情感表达,对学生表示困惑、疑难的知识点进行重点标记,对教师给予提醒,在教学中重点讲解。

3.1.2. 过程与结果耦合的评价闭环设计

过程评价与结果评价并非孤立存在,而是能够通过过程数据预测结果、结果数据验证过程形成耦合闭环。过程数据对结果的预测:生成式 AI 基于历史过程数据与结果数据的相关性,构建学习成效预测模型,通过实时过程数据预测学生的结果表,并提前干预通过分析学生的历史数据(过程数据包括线上学习时长、实验操作正确率,结果数据为期末考试成绩),发现不同类型学生期末考试通过率的不同,基于此模型,AI 可在学期中期对过程数据不达标者发送“预警通知”,并推荐个性化学习资源(如实验操作视频、知识点微课)。结果数据对过程的验证:通过期末考试、课程论文等结果数据,验证过程评价的有效性,并优化过程评价指标,在过程性评价中相关指标得分较高的学生如果期末成绩也较高,则说明过程性指标具有有效性,相反则需要改进过程性指标。

3.2. 从“标准化评价”到“差异化-个性化评价”

传统混合教学评价采用“标准化”模式——统一的评价内容(如同一套试卷)、统一的评分标准(如选择题每题 2 分)、统一的权重分配(如期末考试占比 70%),其优势是“效率高、易操作”,但弊端是“忽视学生的个体差异,无法满足个性化培养需求”。生成式 AI 的自适应算法与个性化反馈能力,推动评价体系从“标准化”转向“差异化-个性化”:差异化指根据学生的学习基础调整评价内容难度,个性化指根据学生的能力短板提供定制化反馈与改进建议。“差异化-个性化评价”(Differential & Personalized Assessment, DPA)以学习者画像-自适应任务-动态阈值三元结构替代传统常模参照,其核心是让每个学生与自身的“最近发展区”对话[11]。AI 的自适应测试与生成式任务技术使大规模个性化成为可能[12]。

3.2.1. 差异化评价:基于学习基础的动态难度调整

生成式 AI 能够通过“学前诊断-动态调整-分层评价”的流程,实现评价内容的差异化。在学前诊

断阶段,在课程开始前,AI 通过适应性测试来评估学生的初始能力水平,根据学生的答题情况(如正确率、答题时间)将其分为“基础薄弱组”(正确率 <60%)、“中等水平组”(正确率 60%~80%)、“优秀水平组”(正确率 >80%)。动态调整阶段则根据学生的实时学习数据,调整后续评价内容的难度,“基础薄弱组”的评价任务以“知识记忆与基础应用”为主,“优秀水平组”则以知识拓展与创新应用为主。若“基础薄弱组”学生在某次评价中正确率能够超过 80%,AI 将自动将其调整至“中等水平组”,提升后续评价难度。最后是分层评价阶段,需要针对不同水平组制定差异化的评分标准,这种标准设计避免了基础薄弱学生因标准过高而失去信心,优秀学生因标准过低而缺乏挑战的问题。

3.2.2. 个性化评价：基于能力短板的定制化反馈

生成式 AI 通过“能力诊断 - 定制反馈 - 改进建议”的流程,为学生提供个性化评价服务。第一是能力短板诊断:基于多模态学习数据,识别学生的具体能力薄弱点。例如,针对“学术英语写作”混合课程,AI 通过分析学生的论文草稿,识别“语法错误(如时态混淆)、逻辑结构问题(如段落衔接不自然)、学术规范问题(如引用格式错误)”等短板,并生成“能力雷达图”,直观展示学生在“语法、逻辑、规范”等维度的表现。第二是定制化反馈:针对能力短板提供具体、可操作的反馈,而非笼统的“成绩差”评价,给予学生特殊差异化的建议,让学生能够实际操作、切实可行的建议,而不是笼统概括的评价。最后也同样重要的是给予个性化改进建议:结合学生的学习习惯与能力短板,推荐定制化的学习资源与任务。若学生在某个知识模块上存在短板,AI 推荐设计包含“合适数量的练习”;若学生的“时间管理能力”不足,AI 可以生成个性化学习时间表,给学生提供适合的时间表作为参考,一定要切合实际,让学生更容易完成任务实现目标。

3.3. 从“教师主评”到“师 - 生 - AI 三元协同评价”

在传统混合教学评价中教师是唯一的评价主体,需要承担多达“作业批改、试卷评分、成绩录入”等全部工作,通常会面临“负荷过载、反馈滞后、视角单一”的问题。生成式 AI 的评价能力与学生的元认知发展需求,推动评价主体从“教师单一主导”转向“师 - 生 - AI 三元协同”:AI 可以承担“基础性、重复性评价任务”(如客观题批改、数据统计),学生需要参与“自评与互评”(培养元认知能力),教师则负责“高阶性、创新性评价”(如复杂能力诊断、评价结果解读),三者分工协作,提升评价效率与公平性。三元协同评价(Triadic Co-Assessment, TCA)将教师的专业判断、学生的主体反思与 AI 的数据洞察置于协商式框架中,体现了第四代评价理论所倡导的“共同建构”:精神[13]。

3.3.1. 三元主体的角色定位与职责分工

Table 1. List of responsibilities of different evaluation subjects
表 1. 不同评价主体职责表

评价主体	核心职责	技术支撑/能力要求	评价内容示例
AI	基础性评价、数据统计、初步诊断	生成式 AI (NLP、CV、机器学习)、多模态数据融合算法	客观题批改(选择题、判断题)、作业查重、学习行为数据统计、能力短板初步识别
学生	自评(反思学习过程与成果)、互评(同伴间的能力评价)	元认知能力(自我反思、自我调节)、批判性思维(客观评价同伴成果)	学习日志撰写(反思“本周学习的优点与不足”)、小组作业互评(评价同伴的“观点贡献度”“任务完成质量”)
老师	高阶性评价、评价结果审核、个性化指导	AI 素养(解读学习分析报告)、专业能力(判断创新性与学术规范性)	课程论文主观题评分、AI 评价结果审核(纠正算法偏见)、针对学生短板的一对一指导

根据“师 - 生 - AI 三元协同评价”的概念,将不同主体的角色职责体现于上方表 1。

3.3.2. 三元协同评价流程

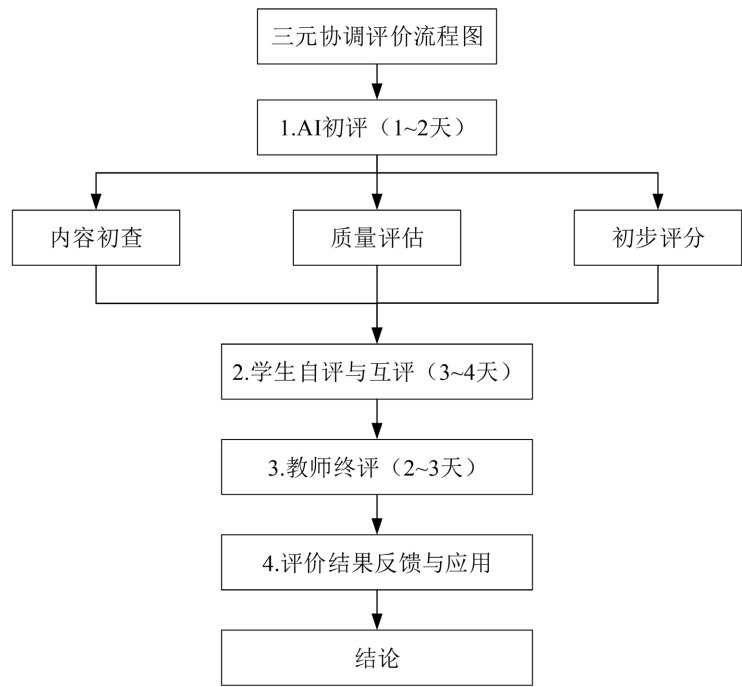


Figure 1. Ternary collaborative evaluation process diagram
图 1. 三元协调评价流程图

如图 1，三元协调评价流程依次推进：首先开展为期 1 至 2 天的 AI 初评，同步完成内容初查、质量评估与初步评分；接着进入 3 至 4 天的学生自评与互评阶段；之后是耗时 2 至 3 天的教师终评；完成终评后进行评价结果反馈与应用；最终得出结论构成完整的评价流程体系。

3.3.3. 协同评价的优势与挑战

三元协同评价的优势显著：能够提升评价效率，AI 承担大部分基础性评价工作，教师的评价时间大大缩短；增强评价公平性，学生互评与教师审核可纠正 AI 的算法偏见。培养学生能力，自评与互评过程提升学生的元认知能力与批判性思维。其面临的挑战主要包括：学生互评的客观性不足，部分学生因“人情关系”给出过高或过低分数，需通过 AI “互评一致性分析” (如计算小组内互评分数的标准差，超过 10 分则提醒教师干预)加以规范；教师的 AI 素养要求提升，教师需具备“解读 AI 分析报告”“纠正算法偏见”的能力，需通过持续培训强化。

3.4. 从“静态评价”到“动态生成性评价”

传统混合教学评价是“静态”的——评价时间固定(如期中/期末考试)、评价内容固定(如开学初确定的试卷范围)、评价结果固定(一旦评分完成则不再修改)，无法反映学生的动态成长过程。在学习中经常出现这种情况，学生在学期中期掌握了某一薄弱知识点，但因期末考试内容固定，无法通过评价体现这一进步，学生在期末考试前临时记忆知识点，但学期结束后很快遗忘，静态评价也无法识别这一问题。但是生成式 AI 的实时数据采集与动态分析能力，推动评价体系从“静态”转向“动态生成性”——即评价伴随学习全过程，内容随学习进度动态调整，结果随学生成长实时更新，最终生成反映学生完整成长轨迹的评价报告。动态生成性评价借鉴 Feuerstein 的中介学习理论，将评价视为“潜能激发 - 实时调节 -

再评估”的螺旋上升过程；AI 通过持续生成新任务、新反馈，使评价始终“走在发展前面”[14]。

3.5. 从“单一维度评价”到“复合脑能力评价”

复合脑能力指 AI 时代人类区别于机器的综合性高阶心智素养，包含批判性思维、提示工程素养、AI 伦理责任、跨学科整合能力、元认知与协作创新五大可观测维度，是混合教学评价从知识导向转向素养导向的核心指标。理论来源为脑科学、建构主义学习理论、布鲁姆认知目标分类理论与未来素养框架。AI 评价与经典理论深度对话则指对泰勒目标评价模式，突破固定目标，走向动态生成；对形成性评价，实现规模化落地；对第四代评价，支撑多元协同共建；对动态评价，以技术实现潜能诊断与全程干预，推动传统评价理论创新升级。传统混合教学评价聚焦知识记忆 - 技能掌握单一的二维维度，评价内容以课本知识点背诵(如名词解释)、基础技能应用(如公式计算)为主，无法覆盖 AI 时代对人才的“复合脑能力”需求，然而 AI 时代的核心竞争力不再是 AI 可替代的知识记忆与简单技能，而是 AI 难以替代的复杂认知能力——即复合脑能力，特指 AI 时代人类区别于机器的综合性心智品质，包括批判性思维、提示工程素养、AI 伦理责任、跨学科整合能力、创新思维五大维度[15]。生成式 AI 的复杂能力诊断算法，使复合脑能力的量化评价成为可能，推动评价体系从“单一维度”转向“复合脑能力多维评价”，评价重心由“知识 - 技能”二维扩展到“知识 - 技能 - 思维 - 伦理”四维[16]。

3.6. AI 评价的局限性与批判性反思

在肯定技术价值的同时，需以批判性视角避免技术乐观主义。其一，AI 难以有效评价高阶思维、创造性与情感共鸣，对不可量化的质性能力存在本质短板。其二，过度依赖 AI 反馈会削弱学生内在动机与自我反思能力，元认知发展空间被挤压。其三，易催生“应试 AI”新型教育异化，师生围绕算法指标优化行为，背离育人本质，必须坚持人机协同、以人为本；AI 承担基础性、重复性评价任务，教师主导价值判断、发展指导与人文关怀，以教育逻辑引领技术应用。

4. 政策演进与四大短板

4.1. 政策演进梳理

4.1.1. 探索期(2018~2020)《教育信息化 2.0 行动计划》，提出“数据驱动教学评价”

2018 年教育部印发《教育信息化 2.0 行动计划》，首次在国务院部委层面提出“数据驱动教学评价”，该文件将大数据技术定位为教育评价变革的“新基建”，强调依托国家数字教育资源公共服务体系，实现对学生学习行为数据的伴随式采集与可视化分析[17]。然而这一时期的政策文本仍停留在宏观号召层面：一方面，尚未明确数据采集、存储、使用的权责边界；另一方面，对评价结果如何与学籍管理、升学考试衔接缺乏细则[17]。因此，高校在落地时多以“示范性项目”形态存在，尚未形成规模效应。

4.1.2. 深化期(2021~2023)《深化新时代教育评价改革总体方案》，明确“破除唯分数、唯升学”

2020 年 10 月，中共中央、国务院印发《深化新时代教育评价改革总体方案》，以“破五唯”为切入口，将教育评价改革上升为国家战略[18]。该方案首次提出“强化过程评价、探索增值评价、健全综合评价”，并强调“利用信息技术改进结果评价”[18]。2021~2023 年间，教育部配套出台《高等学校课程思政建设指导纲要》《本科教育教学审核评估实施方案（2021—2025 年）》等文件，将“学习分析证据”纳入审核评估指标体系，标志着政策从价值宣示走向操作化[19]。但是由于 AI 算法在教育场景中的合法性、正当性尚未确立，高校在采集与分析课堂行为数据时仍面临“合规焦虑”[20]。

4.1.3、融合期(2024~至今)《教育数字化战略行动 3.0》，提出“AI 赋能教育评价改革试点”

教育部提出专设 AI 赋能教育评价改革试点板块，提出到 2027 年形成 30 个左右评价改革示范区[2]。该阶段政策的最大特征是将 AI 评价工具纳入政府采购目录，并通过财政专项(每门混合式课程 30~50 万元)支持高校建设“AI 评价中台”[21]。教育部等五部门发布《“人工智能+教育”行动计划》，对教育算法实施“分级分类 + 全流程”监管[22]，称之为“教育数字化战略行动 3.0”，至此政策链条完成“价值 - 规则 - 资源 - 能力”的四维耦合，AI 评价正式嵌入国家教育治理现代化进程。

4.2. 四大短板

4.2.1. 制度：AI 评价主体地位未确立

现行教育评价相关法规(如《深化新时代教育评价改革总体方案》)仅将“教师、学生、家长、社会机构”为合法评价主体并未提及 AI，导致 AI 生成的评价结果(如过程性评分、能力诊断报告)无法作为“学生评奖评优、升学就业”的正式依据，仅能作为“参考信息”。AI 评价结果的法律效力、责任归属、申诉渠道等制度设计缺位，使高校陷入“想用不敢用”的困境[20]。

4.2.2. 伦理：隐私缺规范

AI 评价需要通过采集面部微表情、语音声纹、眼动轨迹等高敏感度数据来进行工作，在实践中部分高校以“教学改进”为由扩大数据采集范围，但未获得学生明示同意，容易引发诉讼风险[20]。

4.2.3. 标准：无“复合脑能力”指标体系

《深化新时代教育评价改革总体方案》虽然提出了要进行“完善德智体美劳过程性评价”，但对 AI 时代所需的“批判性思维、提示工程素养、伦理责任”等复合脑能力缺乏操作性定义[18]，并且目前由于缺乏全国统一标准，各高校自建指标的信度、效度差异显著[23]。

4.2.4. 能力：教师队伍能力不足

技术介入教育的深度与广度不断拓展，技术工具的更迭既释放了教育的生产力，又加速了教育本质的异化，导致教师群体的技术焦虑不减反增[24]。教师对于 AI 的学习和使用未进行系统完整的培训，缺乏规范的一练习以及实践指南。

5. AI 教育评价的伦理困境与治理框架

人工智能驱动的混合教学评价在提升效率与精准度的同时，也带来数据权力、算法透明、过度管控与评价公平等深层伦理挑战。从教育哲学与社会学视角看，评价是对人的发展的价值判断，而非单纯的数据量化；AI 评价重构了教育权力结构，若缺乏治理约束，易走向技术规训与数字监控。

一是数据隐私与权力失衡，多模态感知采集高敏感信息，学生处于被动授权状态；二是算法黑箱与问责缺失，评价逻辑不可解释、偏差难以追责；三是全程监控侵蚀学习自主性，易引发表演性学习；四是算法偏见加剧教育不公，训练数据偏差可能形成系统性歧视。

为此，构建可操作治理框架：坚持伦理前置，落实以人为本、最小必要、公平无害；实行数据分级分类治理，强化知情同意与脱敏去标识；推行算法可解释与备案审查，保障评价可追溯、可校验；建立人机双审申诉机制，保障学生救济权利；运用伦理沙盒，在可控范围内试点创新，实现创新与风险的动态平衡。

5.1. 政策优化

为切实减轻高校教师教学负担，政策层面需多管齐下。教育主管部门应制定科学合理的教学工作量核

算政策,明确不同类型教学任务的量化标准,避免过度强调单一指标,同时应该加大对高校教学资源的投入政策倾斜,如划拨专项资金用于引进先进教学设备、建设在线教学平台等,为教师教学提供硬件支持。在职称评定政策方面,降低对科研成果的过度依赖,提高教学质量、教学改革成果在职称评定中的权重,引导教师将更多精力投入到教学工作中。为了提升教师 AI 智能素养,应该出台教师培训政策,定期组织教学能力提升培训,帮助教师掌握先进教学方法与技术,提升教学效率,减少不必要的教学时间损耗。

5.2. AI-ESS 系统架构

AI-ESS (Artificial Intelligence-Enhanced Support System)系统架构以“全域感知-智能决策-人机共生”为核心理念,围绕高校教学全生命周期展开设计,力求在数据采集、处理、决策、交互与持续优化五个维度上,为教师提供一站式、可解释、可定制的智能支持,从而切实缓解教学负担并提升教学质量。

5.2.1. 数据采集层：多源异构、细粒度、全场景

设立学习管理系统(LMS)接口,实时捕获学生在 MOOC、SPOC、翻转课堂中的点击量、停留时长、重播片段、测验轨迹与论坛发帖情绪,形成毫秒级学习事件日志;构建智能教室感知网,部署 4K 云台摄像机、阵列麦克风、毫米波雷达与电子墨水板,采集教师授课语速、板书轨迹、走动热力图,以及学生抬头率、表情微变化、举手频次等行为信号;记录教学过程日志:对接教务系统课表、PPT 翻页记录、课堂即时反馈(弹幕、抢答、云班课),形成教师教学行为数字孪生流;注意隐私与合规性,在边缘节点完成人脸模糊化、声纹去标识化,确保学生个人信息“可用不可见”。

5.2.2. 数据处理与分析层：深度洞察、教育语义、可解释

利用机器学习引擎描述学生画像,结合聚类将学生划分为“探究型”“依赖型”“焦虑型”等学习风格簇;可解释输出对每条推荐策略生成 SHAP 值分解图,教师可直观看到“语速放慢 10%”对“中等生理解度”的边际贡献,并支持自然语言追问式交互。

5.2.3. 智能辅助教学层：生成式备课、自适应作业、全天候辅导

利用智能备课助手,依据课程大纲与知识图谱,自动生成含三维目标、情境案例、互动环节的 PPT 模板,并提供与专业契合的开放教育资源链接;通过检索增强生成、技术,实时抓取最新论文,为教师推荐可嵌入课堂的“前沿微案例”,实现“课程-科研”同频共振;智能作业系统:客观题自动批改覆盖选择、填空、编程题,支持多语(Python/Java/C++)代码沙箱运行与抄袭检测 MOSS 算法;主观题采用“两段式”流程:先由大模型给出要点级评分与改进建议,再由教师一键确认或微调提升批改效率,依托规范、透明的学业评价体系,能够引导学生从功利化的“分数导向”转向深度参与的“成长导向”,进而提升整体学习投入水平[25]。

5.2.4. 用户交互层：沉浸式、可定制、跨终端

设置教师驾驶舱,一页式仪表盘集成“班级画像、教学热力图、风险预警、资源推荐”四大模块,支持语音唤醒与触控拖拽;“一键生成课堂叙事”功能,可将实时数据自动转化为教学反思报告,并同步到超星学习通、雨课堂等常用平台[26]。搭建学生端小程序,利用“学习体检”卡片展示本周学习曲线、情绪热力、知识漏洞雷达图,支持一键设定“番茄钟”专注模式;个性化推送“微课+习题”组合包,依据学生注意力曲线动态调整学习路径。注意无障碍与多模态,适配屏幕阅读器提供语音导航、高对比度模式;支持微信、钉钉、飞书、企业微信小程序,实现“消息即服务”。

5.2.5. 持续优化层：闭环反馈、群体智慧、生态共建

评价即服务,允许第三方企业、研究机构调用 AI-ESS 能力,打造教育评价 App Store,形成可持续

的开源生态。

通过上述五大层级、十余种智能工具与多维度保障机制, AI-ESS 系统不仅能将教师的重复性、低认知负荷工作自动化, 更重要的是以“数据 + 算法 + 解释”的方式, 为教师提供循证决策支持, 最终实现“技术减负”与“教学提质”的双重目标。

6. 结论与展望

6.1. 核心结论

人工智能不会取代评价, 但会重构评价共同体的边界。AI-ESS 的实践表明, 教师、学生与 AI 之间已形成“认知分工”, AI 负责大规模数据计算与实时反馈, 教师负责价值判断与情感支持, 学生则在持续对话中提升元认知水平[27]; 政策演进的三阶段模型揭示, 技术嵌入教育制度的关键在于“法律确权 - 伦理沙盒 - 财政常模”三位一体, 缺一不可; 五层系统架构验证了“可解释性”与“高并发”并非零和博弈, 通过微服务与边缘计算可在毫秒级延迟下提供像素级解释, 为 AI 教育应用树立了工程范式。

6.2. 未来研究方向

第一是认知数字孪生, 构建每位学生的“认知数字孪生体”, 实现“先仿真后教学”的精准干预; 第二是 AI 伦理建构, 培养学生全面多维的伦理素养, 亟须发展“教育生成式 AI 伦理学”这一新兴交叉学科。

AI 驱动的混合教学评价体系无论在任何发展阶段都最终必须回归教育本质, 以技术增强理性, 以人文守护价值。

参考文献

- [1] 姜凤春, 赵子博, 郭新伟. 高校混合教学本质的生态学审视[J]. 湖南师范大学教育科学学报, 2023, 22(5): 105-111.
- [2] 政府工作报告——2025 年 3 月 5 日在第十四届全国人民代表大会第三次会议上[EB/OL]. 2024-03-05.
https://www.gov.cn/yaowen/liebiao/202503/content_7013163.htm, 2025-08-12.
- [3] 刘建达. 人工智能时代的语言测评: 机遇与挑战[J]. 现代外语, 2024, 47(6): 859-869.
- [4] 教育部, 中央网信办, 国家发展改革委, 等. 教育部等九部门关于加快推进教育数字化的意见[EB/OL].
https://www.gov.cn/zhengce/zhengceku/202504/content_7019045.htm, 2024-07-18.
- [5] 易凯谕, 韩锡斌. 从混合教学到人智协同教学: 生成式人工智能技术变革下的教学新形态[J]. 中国远程教育, 2025, 45(4): 85-98.
- [6] 张树华, 张夏添. 从工具到共生: 生成式人工智能技术驱动下知识生产与“新文科”建构[J]. 郑州大学学报(哲学社会科学版), 2025, 58(4): 21-28+168-169.
- [7] 中共中央 国务院印发《教育强国建设规划纲要(2024—2035 年)》[EB/OL].
https://www.gov.cn/zhengce/202501/content_6999914.htm, 2025-01-19.
- [8] 教育部办公厅. 教育部办公厅关于开展第三批国家级一流本科课程认定工作的通知[EB/OL]. 2023-11-27.
http://www.moe.gov.cn/srcsite/A08/s7056/202312/t20231214_1094582.html, 2025-12-03.
- [9] 教育部高等教育司关于开展人工智能赋能高等教育人才培养系列师资培训的通知[EB/OL].
<https://statics.scnu.edu.cn/pics/em/2024/1209/1733707615317077.pdf>, 2024-12-05.
- [10] Black, P. and Wiliam, D. (2018) Classroom Assessment and Pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25, 551-575. <https://doi.org/10.1080/0969594x.2018.1441807>
- [11] 戚业国, 杜瑛. 教育价值的多元与教育评价范式的转变[J]. 华东师范大学学报(教育科学版), 2011, 29(2): 11-18.
- [12] Shute, V. and Rahimi, S. (2021) Stealth Assessment. In: Lukin, R., Ed., *AI and Learning*, Springer, 55-72.
- [13] Guba, E. and Lincoln, Y. (1989) *Fourth Generation Evaluation*. SAGE.
- [14] Feuerstein, R. (1979) *The Dynamic Assessment of Retarded Performers*. University Park Press.

-
- [15] World Economic Forum (2023) Future of Jobs Report 2023. World Economic Forum.
- [16] 朱旭东, 赵志群, 白滨. 职业教育能力本位评价研究[M]. 北京: 教育科学出版社, 2020.
- [17] 中华人民共和国教育部. 教育部关于印发《教育信息化 2.0 行动计划》的通知[EB/OL]. http://www.moe.gov.cn/srcsite/A16/s3342/201804/t20180425_334188.html, 2018-04-13.
- [18] 中共中央 国务院印发《深化新时代教育评价改革总体方案》[EB/OL]. https://www.gov.cn/zhengce/2020-10/13/content_5551032.htm, 2020-10-13.
- [19] 教育部关于印发《普通高等学校本科教育教学审核评估实施方案(2021—2025 年)》的通知[EB/OL]. https://www.gov.cn/zhengce/zhengceku/2021-02/07/content_5585584.htm, 2021-02-05.
- [20] 姚荣, 王倩. 生成式人工智能教育应用的风险及其法律规制[J]. 清华大学教育研究, 2025, 46(6): 131-142.
- [21] 教育部关于一流本科课程建设的实施意见[EB/OL]. http://www.moe.gov.cn/srcsite/A08/s7056/201910/t20191031_406269.html, 2026-06-27.
- [22] 教育部等五部门关于印发《“人工智能+教育”行动计划》的通知[EB/OL]. http://www.moe.gov.cn/srcsite/A16/s3342/202604/t20260410_1433240.html, 2026-04-08.
- [23] 教育部. 深化新时代教育评价改革总体方案配套政策汇编[M]. 北京: 人民教育出版社, 2022.
- [24] 马晓榕, 李明, 陈午阳. 生成式人工智能时代教师发展的平衡之道[J]. 高等建筑教育, 2026, 35(3): 9-16.
- [25] Tomlinson, M. (2014) Exploring the Impacts of Policy Changes on Student Approaches and Attitudes to Learning in Contemporary Higher Education: Implications for Student Learning Engagement. *Higher Education Academy*, 12-15.
- [26] Aiken, A., Joyner, D., et al. (2024) MOSS: A System for Detecting Software Similarity. <http://theory.stanford.edu/~aiken/moss/>
- [27] Luckin, R. and Cukurova, M. (2023) Designing Educational Technologies in the Age of AI: A Learning Sciences Perspective. *Learning, Media and Technology*, 48, 1-17.