基于大语言模型的人机协同课堂教学质量评价新范式

曾勋勋 陈 飞 阮 奇

摘 要：为了解决人工督导缺乏过程性量化评价指标以及大模型评价的“幻觉”问题，提出基于大语言模型的人机协同课堂教学质量评价新范式。该范式基于哈蒂“可见的学习”理论，首次提出“可见的督导”新模式，涵盖课堂教学六个维度，构建16项可量化指标与3项难量化指标。针对可量化指标，通过引入AI技术，对课堂教学中的视频等多模态教学数据进行智能计算；针对难量化指标，结合布鲁姆认知分类法，提出“认知路径评价法”进行等级评定。基于该范式引导大语言模型生成课堂教学质量评价报告，可为课堂教学提供更加全面、精准和个性化的评价结果。

关键词：教学督导；大语言模型；人机协同；课堂教学质量；评价指标

一、引言

课堂是学校教育的主阵地，课堂教学质量直接关系着国家教育发展和人才培养水平，而课堂教学评价是驱动教学质量提升的重要环节。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》，要求强化过程性评价，充分利用信息技术提高教育评价的科学性、专业性、客观性。2025年8月，国务院印发《关于深入实施“人工智能+”行动的意见》，明确要求将人工智能深度融入教育教学全要素、全过程。在此背景下，将人工智能深度融入课堂教学质量评价，探索人机协同督导，成为赋能评价范式革新、全面提升课堂教学评价效率与质量的关键路径。

课堂教学质量评价是指通过对课堂教学系统化地观察、测量和分析，对教师课堂教学活动的有效性、学生学习成效以及教学目标的达成程度进行质量评估与判断的过程[1]。随着教育信息化、数字化与智能化的发展，传统课堂教学质量评价方式已经难以满足现代教育的需求[2]。课堂教学质量评价是教育管理的重要环节[3]，现有的方法通常可分为定性评价和定量评价。定性评价是一种以非数值化方式评估课堂教学质量的方法，主要通过观察、访谈、问卷等手段，研究课堂教学过程中的重要特征和关键因素。通常采用督导专家听课、学生反馈和同行教师评议等多角度评价。定性方法容易受个人喜好、人情世故、个人教学经验等因素影响。层次分析法[4]和机器学习法[5]是常见的定量评价方法。层次分析法将课堂教学分解为教学目标、教学内容、教学方法等多层次指标，通过一致性验证和权重计算综合评分。机器学习法利用课堂教师讲授和学生学习产生的数据信息构建机器学习模型，建立教学数据库，对教学质量进行预测和评价。近年来，随着深度学习、计算机视觉、大语言模型等技术的快速发展，利用AI辅助课堂教学质量评价可以有效减少主观偏差，但是存在数据依赖、成本较高的问题[6]。

课堂教学是一个高度动态且复杂的过程[7]，现有的课堂教学质量评价方法在实际运作中存

曾勋勋，福州大学数学与统计学院副教授；陈 飞，福州大学计算机与大数据学院教授；阮 奇，福州大学化工学院教授。

中国知网 https://www.cnki.net

在着如下问题：（1）人工督导缺乏过程性量化评价指标。课堂教学过程存在着多变量因素的影响，信息量庞大且瞬息万变。督导专家长时间处理高维度的信息，容易产生主观偏差。（2）大模型课堂评价存在幻觉现象[8]。利用生成式人工智能大模型在输出内容时，因数据有限会产生一些虚构或错误的信息，从而出现评价与实际情况不一致、不准确的情形。

针对上述问题，本文基于哈蒂“可见的学习”理论，提出“可见的督导”，构建16项可量化指标与3项难量化指标。结合督导专家的经验和AI技术，建立基于大语言模型的评价范式，实现课堂教学质量精准评价。

二、可见的督导

哈蒂的“可见的学习”理论[9]对数亿学生进行实证研究，发现“教师是影响学生学业成就的关键因素之一”。该理论强调教学应成为“可见”的过程。一方面，教学对学生应是透明的，学生要清晰了解学习目标与学习内容。另一方面，教师应通过评估“看见”学生的学习情况，并据此调整教学策略。围绕课堂教学质量评价，将哈蒂的“可见的学习”理论应用于评价课堂，提出“可见的督导”新模式，课堂教学评价聚焦“可见”的两端：教师的教（教学策略是否清晰、是否提供高效反馈）与学生的学（学习态度是否积极、进步是否可衡量）。将教师的“可见”和学生的“可见”融合，构建督导的“可见”（图1）。

依据教育部《关于一流本科课程建设的实施意见》、全国高校青年教师教学竞赛“课堂教学”评分细则，将督导的六个“可见”对应六个评价维度：教学态度、教学目标、教学策略、

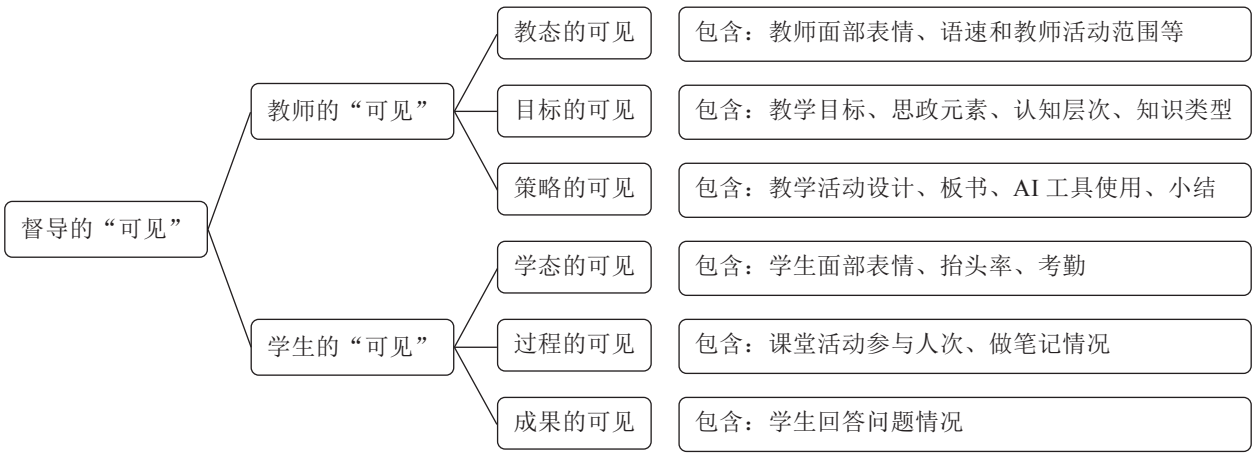


图1　课堂教学质量评价维度

学习态度、学习过程和学习成果。基于AI技术、督导专家经验以及相关研究[3]，对六个维度展开分析如下。

（1）教学态度。利用计算机视觉技术识别教师面部微笑等表情特征，并实时检测教师在课堂上的移动轨迹，计算其活动范围。利用语音识别技术能够准确检测教师的语言清晰度和语速，分析出教师语言表达能力和上课节奏。通过教态的“可见”，评估教师的课堂感染力与亲和力。

（2）教学目标。利用深度学习方法对教师的课件和语音文本进行分析识别，判断是否包含教学目标和思政元素。引入机器学习方法构建语义相似性度量指标，用于评价教师授课的语音文本与课件内容、授课计划内容的匹配程度。另外，还需要评估授课内容的广度、深度和挑战度，以及教师是否引导学生思考、联想和是否给予创新的启迪。通过目标的“可见”，评估教学的科学性、专业性以及对学生成长的帮助。

中国知网 https://www.cnki.net

（3）教学策略。通过分析教师的教学活动，判断是否启发和引导学生达到教学目标。通过统计教师提问中是否包含“你有什么解决方案”“还有其他方法吗”“今天学到的方法还可以应用于其他什么地方”等高阶开放性问题，体现教师引导学生深度思考、培养批判性思维的意图和能力。板书时间占比可用于客观反映教师对传统板书教学方式的运用情况。教师在课堂上是否使用 AI 工具，反映了教师应用人工智能技术赋能课堂的教学策略。通过策略的“可见”，评估教师设计的教学活动的合理性，对教学目标达成的支撑情况。

（4）学习态度。利用计算机视觉技术分析学生面部表情，统计注视讲台区域的抬头率，可客观观测学生注意力集中程度；同时，对缺席、迟到等考勤数据的记录，也为评价学生学习态度提供了重要依据。通过对学态的“可见”，评估学生的学习心理。

（5）学习过程。利用课程平台数据和计算机视觉技术统计全班学生参与课堂活动的次数、学生做笔记的情况。通过对过程的“可见”，评估学生课堂参与度与学习过程。

（6）学习成果。利用课程平台数据统计学生回答问题的情况。通过对成果的“可见”，客观衡量学生的学习成效。

三、评价指标

基于大语言模型的生成能力，在上述课堂教学质量评价的六个维度下设计了对应的 19 个评价指标，包括 16 个可量化指标和 3 个难量化指标，支持课堂教学质量的“可见”评价（图 2）。

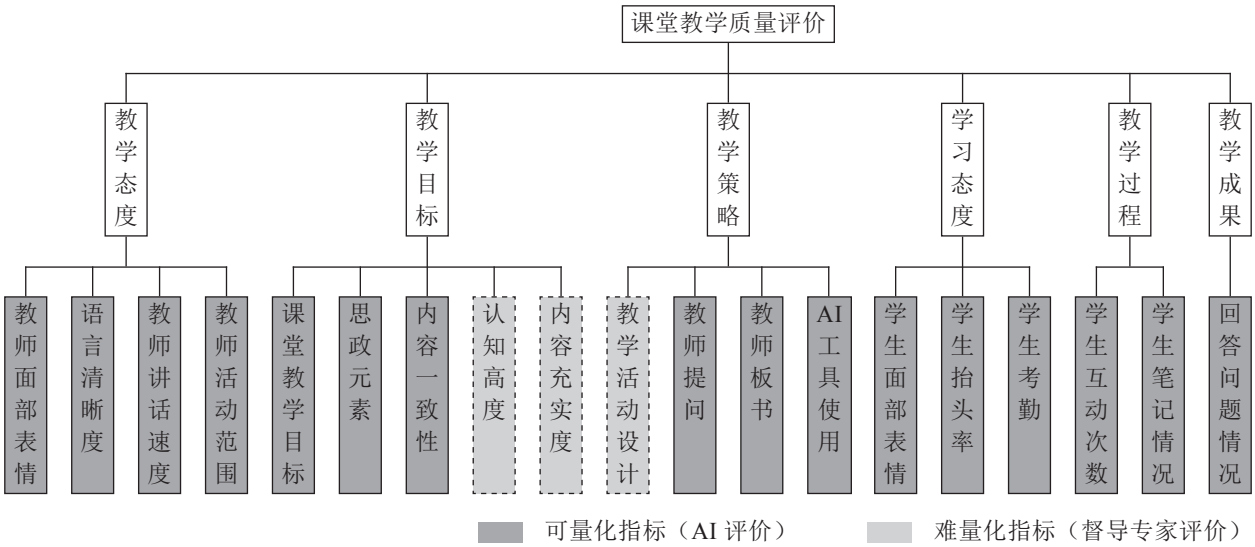


图 2　课堂教学质量评价指标

1．可量化指标

利用人工智能、大数据、物联网等技术，实现课堂教学行为的自动采集、多维度智能分析，为课堂教学评价提供客观、精准的数据支撑与决策依据。智慧课堂借助摄像头、麦克风、智能平板等终端设备以及一体化大数据教学平台，可实时采集课堂教学过程中师生行为视频、课堂语音、课件及板书文本等多模态数据。通过高清摄像头捕捉师生人脸图像，并基于深度学习与计算机视觉算法，精准识别教师面部表情、教师移动轨迹、学生抬头情况及视线方向等细粒度行为特征。利用麦克风阵列录制课堂音频，结合语音识别与自然语言处理技术，对教师语言清晰度、语速和师生互动频次等进行量化分析。在此基础上，系统能够关联学生状态与教师行为[10]，动态评估课堂活跃程度、讲授内容与课件 / 板书的匹配度等。

2．难量化指标

针对课堂教学质量评价中 3 个难量化的指

中国知网 https://www.cnki.net

标——“认知高度”“内容充实度”和“教学活动设计”，结合布鲁姆认知分类法，提出可操作性强的“认知路径评价法”。布鲁姆理论将认知能力分为六大层次，即从低阶的记忆、理解、应用，到高阶的分析、评价和创造，将知识分为四种类型：事实性知识、概念性知识、程序性知识和元认知知识。认知路径评价法将教学活动按大类分为五种，即“教师讲授”“小组讨论”“平台练习”“生生互评”“AI 助学”。以知识类型为横坐标、教学活动为纵坐标、认知层次为竖坐标构建课堂教学三维坐标系，画出一堂课中每个教学内容的坐标点（知识类型、教学活动、认知层次），连接这些坐标点绘制“三维活动路径”。通过评估教学目标、教学内容和教学活动确定知识类型、认知层次。一堂课的成效，取决于其在认知层次上达到的高度，以及教学活动是否为学生的攀升提供了足够支撑[11]。爬升的认知层次越高，教学活动的设计越丰富，学生的思维认知水平也就越高。

通过分析一堂课所涵盖的知识类型数量，可以评估其知识的“广度”与“深度”，进而判断教学内容的充实程度。教师所设计的教学活动，应有效启发并引导学生从低阶思维（如记忆、理解、应用）逐步迈向高阶思维（如分析、评价、创造），实现认知层次的递进与支撑。“有效教学活动”指在课堂中能引导学生思维有序攀升，并高效达成教学目标的教学设计。课堂中有效教学活动的数量，是衡量教学策略多样性的直观体现，也是评估教师启发与引导能力的核心指标。提出的认知路径评价法是将一堂课视为一条引导学生从起点（已知）走向终点（新知）的“认知路径”。评价的重点在于这条路径的“高度”“广度”“有效长度”三个关键属性。该评价法不仅可以有效服务督导专家的课堂评判，也有助于指导教师优化其课堂教学设计。

以高等数学课程里定积分的几何应用节段为例，授课教师重构了教学内容，在计算图形面积和体积的知识中加入了实际应用案例——求解加油站储油罐的油容量。督导专家采用认知路径评价法对该教学节段进行评价时，将课堂中的教学内容及其在课堂教学三维坐标系中对应的坐标分类为截面积可求立体定义（事实性知识、教师讲授、记忆），体积计算公式（概念性知识、平台练习、理解），微元法（程序性知识、教师讲授、理解），储油罐储油的截面积（程序性知识、小组讨论、应用），储油实际情况（元认知知识、生生互评、分析），储油罐的油容量（程序性知识、小组讨论、评价），复杂情况求解（元认知知识、AI 助学、创造）。从而分析出课堂教学包含了四种知识类型和五种有效教学活动，实现了难量化指标的精确评价。在课堂教学三维坐标系中绘制出的“三维活动路径”，也体现了教师在该节课引导学生思维从低阶向高阶跃升的一个过程。

四、评价模型

基于多模态数据，融合机器学习、自然语言处理与计算机视觉等人工智能技术，构建基于大语言模型的人机协同课堂教学质量评价模型（图 3），该模型工作过程主要包含八个步骤。

（1）数据采集。通过教务系统收集教师该节课的授课计划，利用教室摄像头、麦克风、课程平台等自动采集师生语音、表情、肢体动作、互动文本等多维度数据。

（2）课堂教学长文本生成和可量化指标提取。利用自然语言处理、计算机视觉、语音识别等 AI 技术由机器自动将教师语音、PPT 文本和课堂板书转化为课堂教学长文本，同时自动计算如教师语速、提问频次、学生抬头率、回答问题等 16 个可量化的教学指标。

（3）难量化指标评定。督导专家结合布鲁姆认知分类法，按照提出的认知路径评价法，构建三维教学路径，对 3 个难量化指标进行等级评定。

（4）生成语义评价词。将 16 个可量化指标和 3 个难量化指标通过预先设计好的二维映射表进行转化（详见本文评价词设计），得到 19 个语义评价词。

中国知网 https://www.cnki.net

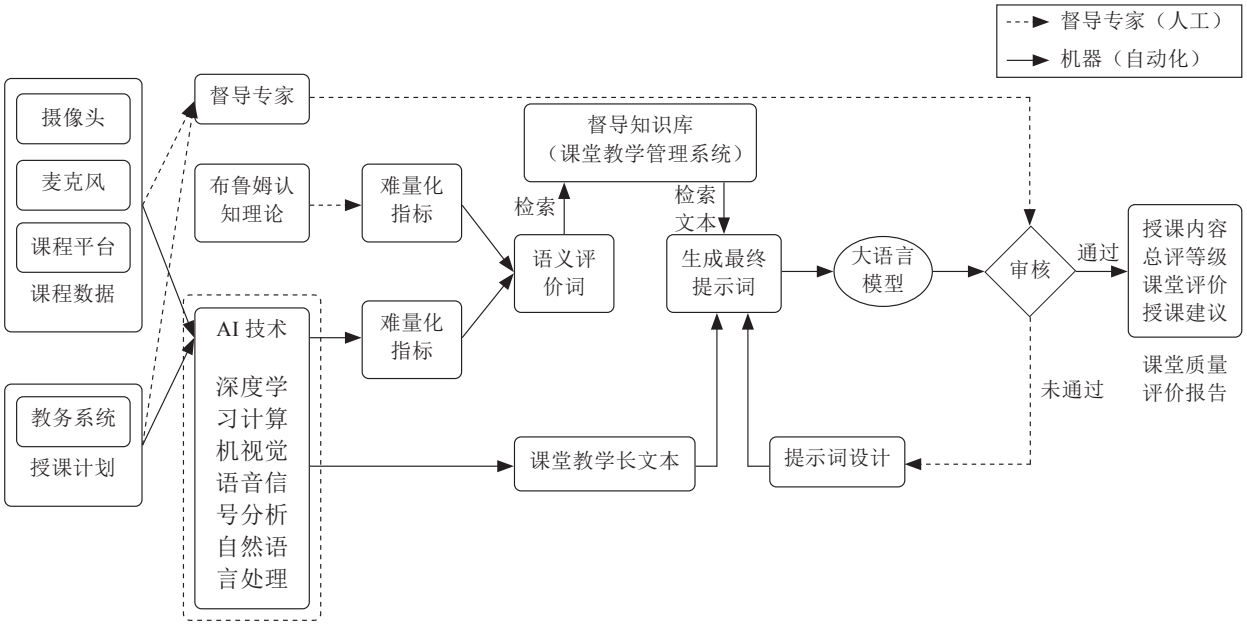


图 3　基于大语言模型的人机协同课堂教学质量评价模型

（5）生成检索文本。基于预先构建好的督导知识库，该知识库汇总了福州大学督导简报、课堂教学管理系统里近三年的督导听课记录，引入检索增强生成技术[12]，利用语义评价词从督导数据库中检索相关案例和信息，生成检索文本。

（6）生成大语言模型最终提示词。将预先设计好的提示词、课堂教学长文本、检索文本合并生成大语言模型的最终提示词。

（7）生成评价报告。大模型通过分析最终提示词，识别核心需求与上下文，将碎片信息整合成结构化提示，生成课堂教学质量评价报告。

（8）督导专家审核报告。督导专家结合经验评价报告的合理性，如果审核通过，形成最终的评价报告；如果不合格，则返回修改提示词设计，重新生成新的评价报告，直至审核通过。

人机协同课堂教学质量评价的核心优势在于实现了督导专家与机器的优势互补。机器凭借精准的数据分析与 AI 技术，能对课堂教学数据 16 项可量化指标进行不间断全覆盖监测，极大提升了督导的广度、效率与客观性。而督导专家则能从烦琐重复的监督任务中解放出来，将精力专注于课堂教学的高阶性等 3 项难量化指标。图 3 中实线箭头和虚线箭头分别表示机器自动化过程和督导专家人工流程。这种协同将机器的“精准高效”与人类的“经验智慧”深度融合，最终形成数据驱动、人机协同的新范式，使督导工作更加科学、精准且高效。

五、评价报告

1．提示词设计

在实际课堂听课评价中，督导专家除评定等级外，还需撰写详细的评价报告。课堂教学评价报告通常包括四个核心部分。一是课堂授课内容概述，简要说明教学主题、重难点等；二是五级制（优秀、良好、中等、及格、不及格）的总评等级；三是课堂评价，从教学态度、教学目标、教学策略、学习态度、学习过程和学习成果进行客观分析，指出亮点与不足；四是给任课教师的具体建议，基于课堂评价提出具有针对性、可操作的改进方向，以促进教师教学能力提升。

利用大语言模型生成高质量、可信且有用的评价，需要系统性的步骤和精心设计的提示词。为了有效评估课堂教学，提出一个结构化框架，让模型基于六个评价维度进行推理和生成。例如，在课堂教学质量评价提示词设计模

中国知网　https://www.cnki.net

板中，教学内容的提示词是汇总教师这节课的授课内容（不超过 200 字）；教学态度的提示词为分析教师面部表情、语速和活动范围；教学目标的提示词为上课是否有教学目标、思政元素，并分析课程的广度、深度、挑战度；教学策略的提示词为是否有分析、启发、引导、板书、课堂小结、AI 工具使用情况；学习态度的提示词为分析学生面部表情、抬头率、出勤率；学习过程的提示词为分析学生课堂互动比例、做笔记情况；学习成果的提示词为分析学生回答问题情况；最后，给任课老师的建议不超过 100 字。若仅将大语言模型直接应用于课堂教学质量评价，会有一些局限性。由于其依赖训练时所获取的静态知识，生成内容缺乏时效性，并容易出现“大模型幻觉”问题。通过将可量化的评价指标转化为语义明确的评价词，结合预先设计好的提示词生成最终提示词，可以有效约束和塑造大模型的生成过程，从而精准引导其输出符合预期结构和内容的评价结果。

2．评价词设计

在课堂教学评价中，可量化的观测指标（如“教师语速偏快”“授课内容”等）通常源于具体的课堂表现证据。为生成贴合实际的专业评价，评价词的设计应围绕每一个评价维度，紧密结合观测指标，系统整合督导评价的专业术语，避免空泛褒扬或无依据的判断，并始终保持中立、客观的语态。通过这种结构化的评价语言设计，大语言模型能够有效模拟督导专家的分析研判过程，最终输出推理清晰、证据扎实、实用可靠的高质量课堂教学评价报告。

在对可量化指标进行评价时，需明确划分等级。以“优、良、中、差”四级为例，若指标取值范围为［0, 1］，则可按以下区间划分：“优”对应［0.9, 1］，“良”对应［0.8, 0.89］，“中”对应［0.6, 0.79］，“差”对应［0, 0.59］。若取值为集合形式 {0,1}，则定义为“0”表示“偏少”，“1”表示“偏多”等。提出的认知路径评价法包含三个维度，即认知层次依据布鲁姆认知分类法划分为 6 级（由低到高），知识类型划分为 4 类，有效教学活动上限为 5 种。表 1 针对每个评价指标提供了一组具体示例。通过构建此类结构化的评价词条，并将其转化为符合督导专家用语习惯的专业表述，再融入大语

表 1　评价词设计与示例

序号	评价指标		取值范围	评价词示例
1	可量化指标（AI评）	教师面部表情是否合理	{0，1}	教师授课表情较严肃/面带微笑
2		教师讲话清晰度	{0，1}	教师发音含糊/标准
3		教师讲话速度是否合理	{0，1}	教师语速偏慢/偏快
4		教师活动范围是否合理	{0，1}	教师课堂活动范围偏小/偏大
5		是否有明确的教学目标	0	教学目标对学生不可见
6		是否包含思政元素	[0，1]	课程思政元素偏少/偏多
7		授课与教学目标一致性	[0，1]	教师授课内容与教学目标有偏差/一致
8		教师提问次数是否合理	[0，1]	教师提问次数较少/善于提问
9		教师板书是否合理	[0，1]	教师板书偏少/合理
10		AI工具使用	{0，1}	教师缺少/善于使用AI教学工具
11		学生面部表情是否合理	{0，1}	学生面部表情迷茫/专注
12		学生抬头率是否合理	[0，1]	学生低头听课/课堂关注度高
13		学生考勤是否合理	0	存在旷课、迟到现象
14		学生互动比例	[0，1]	学生互动频率偏少/课堂活跃
15		学生笔记情况	0	较少学生做笔记
16		回答问题情况	[0，1]	课堂沉闷/学生积极回答问题
17	难量化指标（专家评）	认知层次	{1,…,6}	注重低价思维/高阶思维训练
18		知识类型	{1,…,4}	内容缺乏/注重前沿性和时代性
19		教学活动设计	{1,…,5}	缺乏教学设计/精心设计教学活动

中国知网 https://www.cnki.net

言模型的生成流程中，可显著提升模型输出与课堂实际状况的吻合度，增强评价报告的专业性和可信度。

六、评价结果与分析

1．评价结果

以高等数学课程为研究对象，系统采集授课过程中的音频、视频、平台数据和教务处系统数据，采用主流深度学习算法（包括目标检测、图像分类、语音识别等）进行智能处理，从中提取出 16 项可量化的课堂教学评价指标。在此基础上，结合督导专家对 3 个难量化指标进行评价，以 DeepSeek 为大语言模型，实现了课堂教学评价报告的自动生成。以隐函数的导数一堂课作为评价示例，使用评价词的课堂总评等级为“中等”。从评价结果来看，依托大模型强大的生成能力，报告内容在总体上与实际情况较为吻合，有效整合了教师课堂授课的核心内容。课堂评价围绕教学态度、教学目标、教学策略、学习态度、学习过程和学习成果展开，在评价词的引导辅助下，评价表述更加贴近实际教学表现，针对性更强。报告中对“教学态度”的描述是：教师表情严肃，语速偏快，活动范围有限，基本以讲台为中心，虽讲解清晰，但缺乏亲和力与感染力，可能影响学生的听课情绪。其中，“表情严肃”“语速偏快”等是现场观察者容易捕捉且可信的评价。此外，报告还能够针对课堂评价中识别出的薄弱环节，为任课教师提供具体可行的改进建议——建议适当放慢语速，增加板书与互动环节，融入课程思政元素，可尝试使用多媒体或 AI 工具辅助教学，增强学生参与感与思维深度。

为了验证评价词引导生成的课堂教学质量报告的效果，对比未使用评价词生成的课堂评价。针对同一堂课（隐函数的导数），未使用评价词的课堂总评等级为“良好”。可以看出，默认状态下仅依赖课堂讲授语音和文本的大语言模型生成结果偏于正面，评价内容较为空泛，针对性不足，尤其会出现与课堂实际表现不符的概括性表述。例如，在描述教师表现时，其“语速适中”等评价与“语速偏快”的实际情况存在出入；所提及的“思政元素”（如“节约用电”）也属自行添加；将学生的“被动”听课与“响应较少”的互动状态，描述为“主动”和“积极回应”。从评价示例中可以看出，大语言模型根据常见评价模板生成“幻觉”内容，对评价内容进行了“美化”或“理想化”，将一种可能的结果陈述为既定事实，因此给出的总评等级都较高。

2．评价分析

为评估人机协同课堂教学质量评价（后文称“人机协同督导”）的表现，邀请 5 位经验丰富的督导专家填写调查问卷，基于效率、一致性、个性化、准确性与解释性五大指标，对人工督导、机器督导及人机协同督导三种模式以［0, 1］区间量化进行评价（表 2）。

（1）效率：指在单位时间内完成课堂教学质量评价报告的能力，完成单节课教学评价报告所需的平均时间越短，效率越高。

（2）一致性：指对同一节课堂进行评价时，评价分数的一致程度，反映执行评价标准的稳定性。

（3）个性化：指对不同课堂所给出的评语在内容上的区分度，避免评语趋于模板化。

（4）准确性：指督导结论与课堂实际教学水平之间的吻合程度，体现评估结果的可靠性与有效性。

（5）解释性：指出课堂中真实存在的问题，并提供具有针对性、可操作的改进建议。

从表 2 可以看出，在效率和一致性上，机器督导展现出其固有优势，其效率超出人工督导 0.87，比人机协同督导高 0.03，其一致性高出人工督导 0.12，高出人机协同督导 0.03，这是因为机器督导能毫无偏差地重复执行督导标准化流程，而人工督导在这些方面受限于生理、情绪及专注度，表现较弱。而在个性化、准确性、解释性方面，人机协同督导优势明显，其个性化高出人工督导 0.16，高出机器督导 0.05；

中国知网 https://www.cnki.net

表 2　五位专家对三种督导模式的评价

	指标	专家1	专家2	专家3	专家4	专家5	平均
人工督导	效率	0.10	0.15	0.02	0.05	0.11	0.09
	一致性	0.91	0.77	0.89	0.87	0.79	0.85
	个性化	0.79	0.61	0.58	0.69	0.72	0.68
	准确性	0.93	0.86	0.90	0.89	0.84	0.88
	解释性	0.85	0.78	0.91	0.79	0.83	0.83
机器督导	效率	0.99	0.98	0.95	0.99	0.89	0.96
	一致性	0.99	0.95	0.97	0.96	0.98	0.97
	个性化	0.73	0.84	0.92	0.79	0.67	0.79
	准确性	0.83	0.73	0.84	0.88	0.76	0.81
	解释性	0.41	0.35	0.48	0.52	0.37	0.43
人机督导	效率	0.95	0.91	0.96	0.85	0.98	0.93
	一致性	0.96	0.92	0.91	0.99	0.94	0.94
	个性化	0.84	0.75	0.89	0.83	0.91	0.84
	准确性	0.92	0.94	0.95	0.91	0.96	0.94
	解释性	0.83	0.89	0.92	0.97	0.93	0.91

准确性高出人工督导 0.06，高出机器督导 0.13；解释性高出人工督导 0.08，高出机器督导 0.48。这是因为机器督导在解释性上存在明显短板，其决策过程如同“黑箱”，难以让人理解和信任。人工督导在处理大量数据时难以实现个性化，但在准确性上，凭借情境理解和常识判断，能在复杂、非结构化课堂评价中表现优异。人机协同督导模式成功融合了人与机器的长处，在五个维度上都达到了或接近最佳的水平。

七、结论

围绕课堂教学质量评价，提出基于大语言模型的人机协同课堂教学质量评价新范式。在指标设计方面，基于哈蒂“可见的学习”理论，提出“可见的督导”新模式，从教学态度、教学目标、教学策略、学习态度、学习过程和学习成果六个维度，构建 16 项可量化指标与 3 项难量化指标。在指标量化方面，针对可量化指标，利用 AI 技术，由机器实时从课堂教学多模态数据中计算获得；针对难量化指标，结合布鲁姆认知分类法，提出“认知路径评价法”，由督导专家给出评价等级。在评价模型方面，提出人机协同督导的新范式，将量化指标转化为评价词，结合督导知识库，引入检索增强生成技术，结合提示词，融入大语言模型，生成精确的教学质量评价报告。这种人机协同课堂教学质量评价新范式，大幅降低了教学督导的工作量，解决了人工督导缺乏过程性量化指标和大语言模型的“幻觉”问题，在一定程度上实现了教学评价的公平性、客观性。

参考文献：

[1] 薛辉，徐文彬 . 近 20 年来我国教学评价研究的知识图谱 [J]. 高等理科教育，2018（4）：1-8，14.

[2] 宋宇，许昌良，穆欣欣 . 生成式人工智能赋能的新型课堂教学评价与优化研究 [J]. 现代教育技术，2024，34（12）：27-36.

[3] 燕姣云，安俊丽，孙国红 . 课堂教学质量评价指标体系重构 [J]. 中国大学教学，2023（12）：74-78，91.

[4] 邱文教，赵光，雷威 . 基于层次分析法的高校探究式课堂教学评价指标体系构建 [J]. 高等工程教育研究，2016（6）：138-143.

[5] 傅翠霞，罗亦泳 . 基于 RVM 机器学习方法的高校理论与实验教学质量评价模型 [J]. 现代电子技术，2019，42（13）：181-186.

[6] 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.

[7] 王连照，张雪 . 数智化时代的课堂教学：特征、

中国知网　https://www.cnki.net

实施与展望 [J]. 教育科学探索，2025，43（2）：99-106.

[8] 刘泽垣，王鹏江，宋晓斌，等 . 大语言模型的幻觉问题研究综述 [J]. 软件学报，2025，36（3）：1152-1185.

[9] 约翰 • 哈蒂 . 可见的学习：对 800 多项关于学业成就的元分析的综合报告 [M]. 北京：教育科学出版社，2015：3.

[10] 季诚钧，何菊芳，卢双坡 . 高校教师课堂教学行为分析 [J]. 中国大学教学，2010（5）：17-20.

[11] 孙建彬，蒋平，于海跃 . 基于布鲁姆认知模型的线上教学目标和活动设计 [J]. 教育教学论坛，2025（5）：5-8.

[12] 杜修平，王崟羽 . 检索增强生成赋能智能导学系统构建研究 [J]. 中国电化教育，2025（5）：117-127.

[福建省本科高校教育教学研究重大项目（项目编号：FBJY20240257）]

[责任编辑：黄　强]

（上接第 36 页）

参考文献：

[1] 关于进一步弘扬科学家精神加强作风和学风建设的意见 [M]. 北京：人民出版社，2019：4-6，14.

[2] 习近平著作选读：第一卷 [M]. 北京：人民出版社，2023：500.

[3] 刘萱，马健铨，李正风 . 科技传播中的科技伦理价值传递与治理原则 [J]. 中国科技论坛，2025（1）：162-168.

[4] 万长松，程磊 . 新时代中国特色科学家精神的传承与发展 [J]. 河南师范大学学报（哲学社会科学版），2022（5）：1-7.

[5] 中办国办印发《关于进一步加强青年科技人才培养和使用的若干措施》[N]. 人民日报，2023-08-28（1）.

[6] 习近平春节前夕赴贵州看望慰问各族干部群众 向全国各族人民致以美好的新春祝福 祝各族人民幸福吉祥 祝伟大祖国繁荣富强 [N]. 人民日报，2021-02-06（1）.

[7] 曲建武，陈晓军 . 教育家精神融入教学全过程的三维探究 [J]. 中国大学教学，2025（4）：53-58.

[8] 习近平 . 思政课是落实立德树人根本任务的关键课程 [M]. 北京：人民出版社，2020：20.

[9] 在教师节到来之际 习近平向全国广大教师和教育工作者致以节日祝贺和诚挚慰问 强调不忘立德树人初心 牢记为党育人为国育才使命 不忘作出新的更大贡献 [N]. 人民日报，2020-09-10（1）.

[国家社会科学基金重大项目（项目编号：21ZDA020）；安徽省新时代育人质量工程项目（项目编号：2024jyjxggyjY072）]

[责任编辑：周晓燕]

中国知网 https://www.cnki.net