

基于大语言模型的课程思政教学效果多维度评价指标体系构建

王枝宁¹ 陈少华² 屈海东¹ 杨晓鹏^{1*}

1.韩山师范学院 数学与统计学院 广东 潮州 521041

2.潮州市教育局 教育研究与教师发展中心 广东 潮州 521041

【摘要】：课程思政教学效果具有价值目标隐性、生成周期较长、表现载体分散等特征，传统评价容易陷入重投入轻产出、重静态轻发展、重主观印象轻过程证据的困境。大语言模型在长文本理解、语义关联和证据归纳方面的能力，为整合教案、课堂互动、学生作业与反思日志等多源数据提供了新条件。本文采用政策文本分析、文献研究与逻辑推演方法，以“教师融入度—学生接受度—育人达成度”为主线，构建由3个一级指标、13个二级指标及相应观测点组成的课程思政教学效果评价框架，提出“通用指标+学科特色指标”的模块化结构，并设计数据采集、语义编码、证据融合、人工复核和反馈改进的实施流程。研究认为，大语言模型应定位为评价证据的处理工具，而非价值判断主体；只有建立基于量规的受控分析、来源可追溯的证据链和教师最终裁量机制，才能兼顾评价效率、育人导向与结果可信度。本研究可为高校课程思政评价从经验判断走向多源证据支持提供可操作的框架。

【关键词】：大语言模型；课程思政；教学效果；多维度评价；指标体系

DOI:10.12417/2705-1358.26.15.078

1 引言

《高等学校课程思政建设指导纲要》提出，人才培养效果是课程思政建设评价的首要标准，并要求健全多维度考核评价体系^[1]。现有研究已运用扎根理论、德尔菲法、层次分析法和CIPP模型细化指标^[2-6]，但仍较多依赖问卷和期末材料，课堂讨论、作业、项目报告等过程证据利用不足；通用量表也难以呈现科学精神、工程伦理等学科差异。大语言模型能够按照量规对长文本进行检索、分类和比对，为整合多源证据提供了条件^[7-8]，同时存在事实幻觉、价值偏差和隐私风险^[9-10]。因此，本文以证据链而非“自动判定”为中心，构建人机协同的多维评价框架：模型负责证据发现与辅助分析，教师和专家负责价值判断与结果确认。

2 指标体系构建的逻辑依据与基本原则

课程思政是“教学设计—学生接受—价值转化”的连续过程。教师融入度对应教学输入，考察目标、内容和方法是否把价值引领嵌入专业知识；学生接受度对应学习过程，考察理解、

互动、情感认同和意义建构；育人达成度对应学习输出，考察政治文化认同、职业伦理、科学精神、社会责任与实践行为的增值。三者既避免只看教师“讲了什么”，也避免脱离教学过程测量学生态度。

指标设计遵循五项原则：一是价值导向，以国家政策、课程标准和专业伦理规范为依据；二是专业融合，评价知识逻辑与价值逻辑的内在联系，而非思政内容出现次数；三是发展性，关注前后变化，不以一次结果给学生贴标签；四是可观测性，以课堂行为、作品和实践记录支撑判断；五是审慎智能化，坚持最少数据、证据可追溯、人工复核和异议可申诉。

三个维度在评价中各有侧重，又不能彼此割裂。教师准备充分，并不必然带来学生的真实接受；学生在课堂上积极回应，也未必已经形成稳定的价值判断。因此，评价时既要沿着教学活动追查效果产生的原因，也要允许不同环节出现不一致。比如，学生接受度较高而育人达成度变化不明显，可能说明课程引起了兴趣，却缺少把认知转化为专业行动的实践任务。保留这种差异，比用总分掩盖问题更有诊断意义。

作者简介：王枝宁，男，汉族，讲师，主要从事不确定性问题的数学理论和人工智能在课程思政教学中的应用研究。

陈少华，女，汉族，中小学高级教师，主要从事思想政治教育与课程思政研究。

屈海东，男，汉族，副教授，主要从事数值计算、机器学习及课程思政教学研究。

*通讯作者：杨晓鹏，男，汉族，教授，主要从事模糊数学优化、人工智能及课程思政教学研究。

本文为广东省教育评估协会2025年度课题《基于大语言模型的课程思政教学效果多维度评价指标体系构建》(编号：BDPG25109)、2024年度广东省中小学德育课题《大思政课背景下基于双融双真理念的小学道德与法治情境教学实践研究》(编号：2024ZXDY190)、潮州市哲学社会科学“十四五”规划2024年度课题《以〈习近平新时代中国特色社会主义思想学生读本〉教学推进潮州新时代好少年培育的实践探究》(编号：2024-C-02)、2024年度广东省本科高校教学质量与教学改革工程项目《校企联合实验室：数据科学创新创业实验室》(编号：粤教高函〔2024〕30号)、2025年度广东省本科高校教学质量与教学改革工程建设项目《面向新工科的AIGC赋能<深度学习>智慧金课建设实践研究》和《专项人才培养计划：数学筑基，AI跃迁：智能时代跨学科创新人才培养计划》(编号：粤教高函〔2026〕4号)、广东省数学会2025年度高等教育教学改革项目《数智融合赋能一流本科课程建设——以<深度学习>为例》、韩山师范学院2025年智慧课程建设项目《深度学习》(编号：E25178)的阶段性研究成果。

3 多维度评价指标体系的结构设计

3.1 指标生成路径

本研究对政策文本和已有量表进行概念提取，按“投入—过程—产出”合并交叉项目，删除只能反映材料数量而不能说明效果的指标，再把抽象品质转化为行为化观测点，形成3个一级指标和13个二级指标。本研究提出的框架在正式应用时需经德尔菲咨询检验专家权威性和协调程度，并用层次分析法或组合赋权法确定权重。

专家遴选宜兼顾课程思政、教育评价、学科教学和人工智能应用背景，并吸收一线专业课教师参与。咨询时不只询问“指标是否重要”，还应请专家判断指标含义是否重叠、证据是否能够取得、表述是否适合不同专业。对意见分歧较大的项目，应回到课程任务和学生作品中试评，而不能单凭平均分决定保留或删除。这样形成的指标才可能从文字上的完整走向课堂中的可用。

3.2 一级、二级指标及观测要点

一级指标	二级指标	主要观测点与证据来源
教师融入度	目标契合度	课程目标与立德树人、专业培养目标一致；教案、课程标准可核查
教师融入度	内容融合度	思政元素与知识逻辑、学科史和行业问题自然关联；课件、案例、课堂实录可核查
教师融入度	方法适切度	采用讨论、案例、项目或实践等促进价值判断的方法；课堂观察与任务设计可核查
教师融入度	资源规范度	资源来源可靠、表述准确、更新及时，不含价值偏差；资源库及引用记录可核查
教师融入度	评价反思度	设置过程评价并依据反馈改进教学；评价方案、教学反思与修订记录可核查
学生接受度	认知理解	能准确说明专业知识中的价值意涵，避免口号化复述；测验、作业与访谈可核查
学生接受度	情感认同	对家国责任、职业使命等形成有理由的认同；反思日志和访谈可核查
学生接受度	互动参与	主动提问、讨论、协作并回应价值冲突；课堂实录与平台记录可核查
学生接受度	迁移判断	能在新专业情境中识别伦理问题并权衡方案；案例分析与项目报告可核查
育人达成度	政治与文化认同	能把个人发展、专业学习与国家社会需要联系起来；综合作品与访谈可核查
育人达成度	科学精神与职业伦理	表现出求真、诚信、规范、安全和责任意识；实验记录、代码、报告可核查
育人达成度	社会责任与协作	关注公共利益、公平包容并履行团队责任；项目过程与同伴评价可核查
育人达成度	实践行为增值	在实践、竞赛或志愿服务中将价值认知转化为行动；前后测与成长档案可核查

评价采用四级行为锚定量规，主要为“未呈现或证据相反”、“零散呈现且依赖提示”、“稳定表现并能说明理由”、“可迁移到新情境并形成行动”。每一级必须关联原始证据；证据冲突或不足时，模型输出“不确定”，不得强行评分。

3.3 “通用指标+学科特色指标”的模块结构

不同专业共享目标契合、内容融合、认知理解与社会责任等通用指标，同时在行为锚点中嵌入学科特色：以数学与统计学院为例，数学类专业应突出严谨求真、理性思辨和模型责任；统计类专业应突出数据诚信、抽样公平和证据意识；数据科学类专业应突出隐私保护、算法偏差和技术向善。高校应建立指标版本库，兼顾横向可比与学科真实性。

以概率统计课程为例，教师介绍抽样方法时若只附加一个先进人物故事，很难说明内容已经融合；若进一步组织学生比较不同抽样方案对弱势群体代表性的影响，并要求其在报告中说明误差、适用范围和可能后果，则可以同时观察内容融合度、迁移判断和社会责任。高等数学课程也可围绕模型假设与现实边界设计讨论，使“严谨求真”落实在推导、检验和修正的学习过程，而不是停留在课件末尾的一句概括。

4 大语言模型支持指标落地的实施机制

评价语料由规范依据、课程材料和学习证据组成。前者包括政策、培养方案、课程目标和专业伦理规范；后两者包括教案、课件、课堂文本、作业、项目报告和反思日志。模型任务分为“证据检索—指标匹配—等级建议—理由生成”。系统先从可信知识库检索指标定义与行为锚点，再要求模型摘录原文、注明位置、给出建议等级和不确定性。报告以“指标表现+证据充分度+人工说明”呈现，重点说明哪些知识点融合薄弱、哪些活动促进了有效讨论以及何处需要支持。教师据此调整案例、问题链和任务，下一周期采集同类证据比较变化。模型、量规、数据范围和人工改判均写入日志，形成“评价—反馈—改进—再评价”闭环。

5 质量检验与应用边界

体系使用前应检验内容效度、结构效度、内部一致性及评价者一致性，并按学科和任务类型考察稳健性。模型可识别公开表达中的价值线索，但不能读取学生内心或作政治性标签；可提出诊断建议，但不能替代教师承担育人责任；任何自动化结果均须允许复核、纠正和申诉。

试评可先选择两至三门材料较完整的课程，由两名教师独立使用量规评价同一批作品，再比较人工判断与模型建议的差异。若某一指标频繁出现“无证据”或评价者理解不一，应优先修订指标定义和教学任务，而不是立即调整模型参数。经过小范围试评后再扩大样本，有助于把测量问题、教学问题与技术问题区分开来。

6 结语

本文构建的“教师融入度—学生接受度—育人达成度”框架，以多源证据连接教学设计、学习过程和育人增值，并以通

用与学科特色模块兼顾可比性和适配性。后续应通过专家咨 识库、人工复核和数据治理约束下真正服务教学改进。
询、权重测算和课程试点持续检验,使大语言模型在量规、知

参考文献:

- [1] 教育部.教育部关于印发《高等学校课程思政建设指导纲要》的通知[EB/OL].(2020-06-05)[2026-09-07].http://www.moe.gov.cn/srbsite/A08/s7056/202006/t20200603_462437.html.
- [2] 高一波.地方高校课程思政教学评价指标体系的构建与探究[J].实验室研究与探索,2022,41(11):242-248.
- [3] 孙跃东,曹海艳,袁馨怡.理工科课程思政教学评价指标体系构建研究[J].江苏大学学报(社会科学版),2021,23(6):77-88+112.
- [4] 马孟伟,王茜,金莉,等.护理本科专业课程思政教学评价指标体系的构建[J].护理学杂志,2022,37(1):6-10.
- [5] 黄玲芳,王宗华,孔德辉,等.基于 CIPP 模型军校护理本科专业课程思政教学评价指标体系研究[J].军事护理,2023,40(8):105-108.
- [6] 王丽,潘虹.课程思政教学评价研究:评价原则、关键问题及评价框架[J].天津农学院学报,2025,32(2):38-43.
- [7] 穆肃,陈孝然,周德青.生成式人工智能赋能教学设计分析:需求、方法和发展[J].开放教育研究,2025,31(1):61-72.
- [8] 周光玲,黄义灵.大语言模型赋能高校思政课教学探究[J].学校党建与思想教育,2025(13):70-73.
- [9] 蒋贵友,殷文轩.变革抑或危机:大语言模型赋能大学教学及其限度——基于斯坦福大学的案例考察[J].电化教育研究,2025,46(1):122-128.
- [10] 李焕宏,薛澜.加强生成式人工智能在高等教育领域中的风险管理:基本框架与关键举措[J].高等教育研究,2024,45(2):31-38.
- [11] 官玉平,徐艳,顾冯霞,等.“双一流”背景下高校课程思政教学评价指标体系构建研究——以成都中医药大学为例[J].成都中医药大学学报(教育科学版),2025,27(1):120-123.
- [12] 陈世月,董晗钰,李茂婷,等.高校课程思政教学评价体系的构建[J].医学教育管理,2023,9(6):693-699.