

# 人工智能时代高等教育评价体系的 范式转型与实践重构

吴佳一<sup>1</sup> 李拓宇<sup>2</sup>

(1. 浙江大学 公共管理学院, 浙江杭州 310058; 2. 浙江大学 中国科教发展研究院, 杭州 310058)

**[摘要]** 生成式人工智能深度改变了学生学习和评价的方式, 高等教育评价体系与学习实践之间开始出现系统性失齐。本研究以比格斯提出的建设性对齐理论为分析视角, 指出评价体系在评价目标、评价任务、评价过程与评价标准的失齐表现, 并以广度、深度与时间三个维度分析评价体系失齐现象的深层逻辑。基于此, 研究提出从目标重校、主体重构、方式重塑到功能重定方面推进评价体系范式转型, 形成目标层、运行层与保障层联动的实践框架, 并提出框架在不同学科条件下的适用边界、价值取向与治理逻辑。文章认为, 人工智能时代评价改革的关键不是调整考核方式, 而是形成持续校准评价与学习匹配的制度能力。

**[关键词]** 人工智能; 学习评价; 建设性对齐; 人智协同素养; 评价体系重构

**[中图分类号]** G649.2

**[文献标识码]** A

**[文章编号]** 1007-2179(2026)05-0080-09

## 一、问题提出

以 ChatGPT、DeepSeek 为代表的生成式人工智能正以超预期的速度嵌入高等教育。学生借助大语言模型可在极短时间内完成结构完整的学术论文、生成可运行的程序或构建系统性的研究方案, 人类智能与机器智能之间的边界日益模糊(王佑镁等, 2023)。这一变革不仅改变了知识获取与加工的方式, 也从根本上动摇了高校长期存在的学习评价范式。

现行大学评价体系的有效性, 建立在评价任务由学生独立完成、任务成果能够反映学生真实认知水平、统一标准能够保障评价公平等假设之上。当生成式人工智能全面渗透学习过程后, 上述假设

开始失效, 评价体系与教学实践之间出现裂隙。面对这一挑战, 部分高校诉诸人工智能检测工具与考试监控技术, 试图维护既有评价秩序; 另有高校采取有限的开放策略, 允许学生在部分课程评价中使用人工智能工具。然而, 无论是技术监督还是局部调适, 都只是在既有评价框架内作出调整, 未能触及问题核心, 即评价体系的底层逻辑与学习实践脱节。本研究把这种脱节称为失齐, 即评价体系所预设的学习图景与学生实际的学习方式不一致。失齐不只带来评价效度的损失, 还会产生逆向引导效应: 当高校继续按照前人工智能时代的假设评价学生时, 学生学习的努力方向会被引向人工智能擅长的低阶领域。

基于此, 本研究围绕“评价体系为什么必须整

**[收稿日期]** 2026-05-19

**[修回日期]** 2026-09-01

**[DOI编码]** 10.13966/j.cnki.kfjyyj.2026.05.009

**[基金项目]** 中国工程院科技战略咨询项目学部重大项目“国家战略急需人才培养机制与路径研究”(2025-XBZD-07); 国家自然科学基金项目“中国原创研究分类评价及长周期资助模式探究”(L25240108)。

**[作者简介]** 吴佳一, 博士研究生, 浙江大学公共管理学院, 研究方向: 人工智能赋能高等教育(12322074@zju.edu.cn); 李拓宇, 副研究员, 博士生导师, 浙江大学中国科教战略研究院副院长, 研究方向: 科教管理。

**[引用信息]** 吴佳一, 李拓宇(2026). 人工智能时代高等教育评价体系的范式转型与实践重构[J]. 开放教育研究, 32(5): 80-88.

体范式转型”与“应如何推动范式转型并落到可操作实践层面”问题,以建设性对齐理论为分析视角,概述传统评价体系在目标、任务、过程和标准四个层面的系统性失齐,阐明评价体系范式转型的内在逻辑,构建评价体系的实践框架,并反思讨论重构路径的理论边界与实施条件。

## 二、分析视角

### (一)建设性对齐的核心主张

比格斯提出的建设性对齐强调教育实践中预期学习成果、教学活动与评价任务三者的一体化,且共同指向同一目标(Biggs, 1996; Biggs & Tang, 2011)。该理论揭示了评价在教—学—评一体化中所处的杠杆位置:学生依据评价任务反推学习应达到的水平,并据此选择深层或表层的学习策略。当三者一致时,评价引导学生走向深度学习;反之,无论教学评价设计和实施多么完善,教学目标也难以实现。这一主张与逆向设计、教—学—评一致性理论相互印证。前者主张单元设计从想要达到的学习结果出发,评价证据的确定先于教学活动设计(格兰特·威金斯等,2017);后者强调所教即所学、所教即所评、所学即所评三重匹配,强调评价应嵌入学生完成学习任务的过程之中(崔允漷,雷浩,2015)。三者的共同点在于,评价不是教学结束后的测量环节,而是目标得以落地的证据设计。

本文以建设性对齐理论作为评价体系的分析视角,主要基于以下两点:一是它将评价理解为关系性的存在,评价的好坏不取决于测量技术本身是否精确,而取决于是否与目标、教学一致。当前考试与论文的评价形式并未失灵,失灵的是它们与学习实际发生方式之间的对应关系。二是它提供了可以逐项检查的诊断标准,包括评价目标指向的能力是否仍然值得培养,任务完成能否取得可信的能力证据,学习过程能否可视化,标准在不同课程之间是否自洽等。评价体系的挑战就不再是笼统的不适应,而被具体化为目标、任务、过程与标准的失齐。

### (二)分析视角拓展

建设性对齐原本是课程层面的教学设计理论,其成立依赖三个假定:分析单元是课程,制度环境作为外部条件;被评价主体是单一的学生个体,其

能力可由任务成果推断;对齐是一次达成的稳定状态。当人工智能介入后,三个假定逐一失效,对齐关系有待重新梳理。

第一,分析单元由一门课程扩展为课程、制度与生态的嵌套结构。课程理论把制度环境处理为课程之外的既定条件,教师在课程内部完成三者匹配即可。人工智能的使用规则、过程性证据能否进入学分认定、学术诚信的边界如何界定等都由校级层面规定,不是任课教师所能调整的。学生同期修读的其他课程是否沿用旧有范式,也直接决定一门课程内部的对齐能否成立。对齐因此不再是课程内部的匹配,而是跨层级的嵌套匹配。与之对应,失齐可能发生在课程内部,也可能发生在层级之间。

第二,对齐要素由三项增补为四项,即在目标、教学、评价之外,把完成任务的主体一并纳入。课程理论默认接受评价的是学生个体,主体是谁无需专门交代,其能力可由任务成果推断。人工智能介入后,完成任务的可能是学生个体,可能是人机协同体,也可能是协同成果中归属于人的那一部分,主体因此成为评价设计必须先行说明的变量。相应的,失齐可能出现在评价的方式,也可能出现在由任务成果推断能力的逻辑本身。

第三,对齐由一次达成的稳定状态改为周期性重建的动态关系。模型能力以月为单位更替,人才培养方案与考核制度数年修订一次,一次达成的匹配会随技术演进自行失效。

## 三、失齐表现与逻辑

失齐并非某一环节的技术性失误,而是评价体系与学习实践之间的结构性错位。从建设性对齐的视角看,生成式人工智能同时冲击了对齐关系的多个要素:改变了最终的学习成果和完成任务的主体,也改变了教学活动的组织方式,却没有改变评价任务的设计范式。一致性因而在评价目标、评价任务、评价过程与评价标准层面发生断裂。

### (一)失齐表现

#### 1. 目标—能力失齐

现行评价体系的目标设定仍指向知识的记忆、理解与标准化再现。当生成式人工智能能够在极短时间内调取海量知识并生成结构完整的文本时,知识再现已不再是区分学生能力水平的有效指标

(戚佳等, 2024), 更需关注的是评价的导向作用。评价内容决定着学生的学习投入方向和程度。评价继续锚定在知识再现层面, 意味着把学生的学习精力导向人工智能胜任的低阶领域, 而批判性思维、创造性问题解决等真正需要在大学阶段养成的能力反而被冷落(胡小勇等, 2025), 从而把学生学习投入引向与能力发展背离的方向。

## 2. 任务—行为失齐

传统评价任务的设计, 建立在学生凭借自身认知资源独立完成任务、任务成果直接反映其知识结构及思维品质的前提之上。生成式人工智能的介入使这一前提难以成立。在当前的学习实践中, 从选题构思到文献梳理再到文本生成, 个体的智力贡献与人工智能的生成内容彼此交织, 边界日益模糊(陈默等, 2025)。教师面对高质量的课程论文, 很难判断哪些部分体现了学生的真实理解。在未来职业实践中高效运用人工智能解决复杂问题本身已是核心竞争力, 隔绝人工智能的评价情境已不是当前学生真实的学习情境。而不加限制地允许人工智能介入, 又可能使评价丧失基本的区分度。这一两难反映出评价设计范式滞后于学习行为模式的变迁。

## 3. 过程—结果失齐

当前的大学评价仍以结果性评价为主, 聚焦学生在特定时间节点提交的最终成果, 对学习过程的认知发展轨迹关注不足。一方面, 学生借助人工智能生成高质量的文本或代码, 其成果与真实学习水平之间可能存在显著偏离(王佑镁等, 2025); 另一方面, 学生与人工智能的交互策略、提问的层次与质量、对人工智能输出的批判性评估行为, 恰恰是判断其高阶能力的关键依据(李艳燕等, 2026), 但这些信息在以结果为中心的评价范式中往往被忽略。评价由此陷入失衡, 能反映学生能力的证据没有被采集, 而被采集的证据又难以支撑判断学生的真实能力。

## 4. 标准—生态失齐

学术诚信方面, 传统评价以独立性与原创性为核心原则, 人工智能时代的知识生产方式正在重新定义这两条原则的内涵。不同学科、不同课程对人工智能使用的规定差异较大, 某一做法在某门课程中受到鼓励, 在另一门课程中却可能被判为违规,

学生因此长期处于规则不确定的状态(Cotton et al., 2024; 吴龙凯等, 2025)。公平性方面, 学生在人工智能工具的获取渠道与使用素养上差异显著, 这些差异最终会以评价结果的形式固化为系统性的不公平(汪益民等, 2026)。教师素养方面, 多数大学教师的评价能力形成于前人工智能时代, 设计情境化评价任务、解读学习分析数据等能力缺失, 使评价改革即便在理念上获得认同, 也难以在课堂层面落地(吴砥等, 2025)。

四个层面的失齐并非彼此孤立, 而是相互强化。目标的滞后使任务设计失去方向引领, 任务与行为脱节使过程性信息难以被纳入评价, 标准与生态的矛盾又从制度层面限制了问题的解决空间。

## (二) 深层逻辑

以上论述表明, 评价改革必须是覆盖评价目标、评价主体、评价方式与评价功能的整体性范式转型, 而不是考核方式的局部更新。

就广度而言, 失齐不在某门课程内, 而在课程与制度、生态之间。教师即便完全依照建设性对齐的原理设计一门课程, 失齐问题仍无法解决。过程性证据难以进入学分认定与成绩记载的既有通道, 学术诚信规范由学校制定而非教师所能界定, 学生修读的课程仍沿用旧有范式。技术监督与局部开放两种常见方式之所以无效, 原因正在于二者都只是限于单一级别内的调整。

就深度而言, 失齐并不发生在评价方式层面, 而在评价赖以成立的推断逻辑。学习评价长期依赖一条推断链, 即由任务成果推断认知过程, 再由认知过程推断个体能力。这一推断链的前提是完成任务的主体单一且责任分工明确。人工智能介入后, 认知过程分布于人与智能系统之间(Hutchins, 1995), 成果不再仅属于学生, 推断链的第一环随之断裂。不先行明确评价的成果是学生完成还是人机协同完成, 评价效度就无从谈起。

就时间性而言, 失齐不是一次性事件, 而会随技术迭代反复出现。评价任务之所以能测出学生的真实水平, 正是因为当下的人工智能还无法完成。当模型迭代更新, 测试任务很可能就不再具有区分度。人才培养方案与考核制度往往数年修订一次, 而人工智能模型以月为单位迭代, 两种周期的不匹配会使改革的有效性随技术更新而递减。

表1列出了上述三种常见的改革思路的不足——在某一课程内推进改革、更新考核形式、一次性完成方案设计。因此,评价体系改革必须在广度上贯通课程与制度,在深度上触及评价逻辑本身,在时间上具备持续调整的能力。

表1 失齐逻辑与改革要求

| 失齐逻辑 | 失齐表现                 | 改革要求                    |
|------|----------------------|-------------------------|
| 广度   | 在课程、制度与生态之间,而非某门课程内部 | 改革需在课程、制度与生态层面协同推进      |
| 深度   | 在于评价赖以成立的推断逻辑,而非评价方式 | 更新考核形式无法恢复效度,需先界定任务由谁完成 |
| 时间性  | 反复出现,而非一次性事件         | 改革目标由一次性方案转为持续调整的制度能力   |

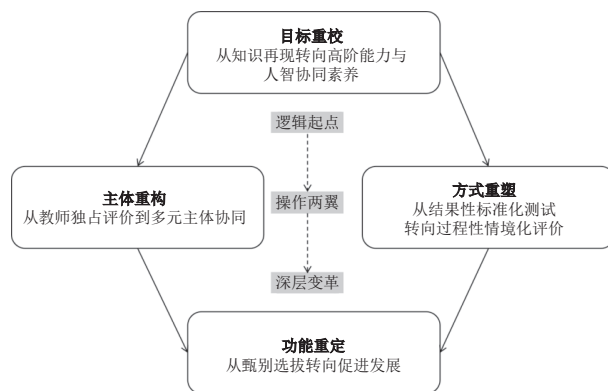
#### 四、范式转型

重建对齐并非回到旧有对齐状态,而是重新建立四者之间的协调关系。依据前述分析框架,“目标重校”重建预期学习成果,“主体重构”重建被评价的主体,“方式重塑”重建评价任务,“功能重定”重建评价所服务的目的。四者之间依次递进、相互制约(见图1)。目标重校是逻辑起点,为其他转型提供方向;主体重构与方式重塑是操作两翼,二者相互支撑;功能重定体现最深层的价值变革,反过来又重新界定目标的价值取向。

(一)目标重校:从知识再现转向高阶能力与人智协同素养

目标重校回应的是主体变化后的失齐。既然评价的推断链以完成任务的主体为前提,那么当主体由学生个体扩展为人与人工智能的协同体时,预期学习成果就必须随之扩展,否则评价体系将持续处于失齐状态。

目标重校的核心在于将评价重点从人工智能可替代的低阶能力转向人工智能难以替代的高阶能力。重校后的评价目标包含四个能力维度:学科知识的深度理解与跨情境迁移、高阶思维能力、人智协同素养、自主学习与元认知能力。其中,人智协同素养并非指在原有能力清单上追加新技能,而是主体结构改变后新增的评价内容,也是人与智能系统协同时所承担的那部分认知功能,即判断哪些任务可以由系统完成、如何评估系统的输出、如何



据以作出决策。它指向判断而非操作,由此与人工智能素养、数字素养相区分:后两者指向工具运用与信息环境中的行动,与评价性判断力相通(Tai et al., 2018)。

从知识建构的视角看,人智协同学习已不再单纯依赖个体认知资源,而是要求学习者与智能系统形成协同认知结构(岳彦龙等, 2025; 黄涛等, 2025)。目标重校需要确保新目标与教学活动对齐:如果教学鼓励学生运用人工智能,但评价仍关注脱离人工智能情境的知识记忆,便会导致目标与教学错位(余亮等, 2025b)。

(二)主体重构:从教师单独掌握评价权到多元主体协同

主体重构指向对齐关系中新纳入的主体要素。当完成学习任务的主体由学生个体变为人机协同体时,判断学生在其中承担哪些认知功能,需要过程数据;判断这部分功能完成得如何,需要专业眼光;判断学生是否意识到自己在协同中的作用,则需要学生本人的反思。

在评价领域,主体重构的关键在于依据不同主体的能力优势合理分工。人工智能系统适合承担过程数据的自动采集、标准化指标的即时评估与学习行为模式的识别分析,但其作用应定位在辅助参考而非最终判定。教师专注于高阶能力的质性判断与提出个性化发展建议,发挥人类在把握认知发展深层特征与教育价值取向方面的作用。学生通过结构化的自评与互评参与评价过程,在反思与对话中发展元认知能力(兰国帅等, 2026; 祝智庭等, 2023; 余亮等, 2025a)。三类主体围绕同一条证据链各司其职:人工智能提供过程证据,教师作出能

力判断,学生解释其在协同中的决策依据。三者的输出完整刻画一次学习表现。

(三)方式重塑:从标准化测试转向过程性、情境化与多模态评价

评价方式转型的核心是取证方式的改变:一是从结果性评价为主转向过程性与结果性评价的有机整合。这种评价借助学习分析技术与人工智能辅助工具,将评价嵌入学习过程本身,即通过持续采集和分析学习行为数据实现动态评估,既能提供更丰富的学生能力发展信息,又有助于缓解学习成果在人工智能介入后的测试效度下降问题(Sadler, 1989; Black & Wiliam, 1998)。二是从去情境化的标准化测试转向情境化的真实任务评价。这种方式把评价任务嵌入真实或高度仿真的问题情境中,允许或要求学生使用人工智能工具,评价的重心从能否获得正确答案转向如何在人智协同中展现判断力、创造力与策略性(Wiggins, 1993)。三是从单一文本成果评价转向多模态评价。这种评价利用人机对话记录、操作行为轨迹、文本修改历程与协作交互数据等多源异构信息,构建多维能力画像,突破仅依赖文本成果推断能力的局限(廖宏建等, 2024)。

(四)功能重定:从甄别选拔转向促进发展

评价功能的重新界定是四重转型中最深层的变革。传统评价的核心功能在于通过统一标准对学生分类与排序,为资源分配和机会配置提供依据。在人工智能时代,评价的首要功能转向促进学生的个性化发展。其实质在于重新定义评价信息的使用方式。评价结果的运用从结果性的等级标定转向诊断性的发展分析,即借助人工智能的数据处理能力为学生提供能力发展诊断与个性化改进建议;评价反馈从期末的延迟反馈转向学习过程中的即时反馈,使评价信息发挥学习引导作用(Boud & Falchikov, 2006);评价标准则在确保基本能力的前提下,允许学生形成差异化的发展路径。

已有元分析研究表明,生成式人工智能对学生发展的影响具有多维性和复杂性(胡钦太等, 2025),简单的甄别性评价已无法捕捉这种复杂的发展特征。不过,功能重定的实质不是用发展性功能替代甄别性功能,而是在不同场景中实现两者的合理平衡。日常教学过程评价以促进发展为主,学业认证

与选拔场景保持必要的甄别功能,同时探索把发展性评价信息融入甄别性判断。

## 五、实践框架

依据建设性对齐,有效的评价需要目标、评价任务与主体三者相互匹配,并需要外部条件保证匹配的持续性。本研究据此构建了涵盖三个层面的实践框架(见图2):目标层承载预期学习成果,回答评什么;运行层承载评价任务与主体分工,回答怎么评;保障层承载匹配得以持续的制度条件,回答如何使评价与学习的匹配。

(一)目标层

目标层的重构是框架的逻辑起点。它包括学科知识的深度理解与迁移、高阶思维能力、人智协同素养与自主学习和元认知能力,遵循两条逻辑:一是人工智能替代性与互补性逻辑,即知识记忆与标准化文本生成属于人工智能的突出能力,深度理解与创造性问题解决属于人类认知的优势能力,人智协同素养处于两者交叉地带;二是能力层级递进逻辑,即深度理解构成高阶思维的基本要素,高阶思维是人智协同素养的前提,自主学习与元认知能力为前三者的持续发展提供内在驱动力(Boud & Falchikov, 2006)。评价实践应根据学科性质与教学目标调整各维度权重。各维度核心指标见表2。

(二)运行层

运行层将目标维度转化为可观测的评价活动,包含任务设计与主体分工两个方面:任务设计规定评价活动的形态,主体分工统筹考虑评价活动由谁承担。

任务设计面临的新困难在于既要把学生的真实认知贡献与人工智能的工具性辅助区分开,又不宜退守为完全隔绝人工智能的封闭式测试。据此,本研究提出真实情境嵌入、人智协同开放与过程可追溯三条设计原则,其含义与实施要点见表3。各形态在人工智能介入程度、时间跨度与认知要求上形成梯度差异,课程评价方案根据教学目标与学科特征组合使用(Mollick & Mollick, 2023)。

本研究基于教师、人工智能系统与学生三类主体的能力特征,提出评价各环节的主体分工(见表4)。分工遵循两条原则:人工智能的评价输出仅作为教师专业判断的参考而非替代,教师对结论

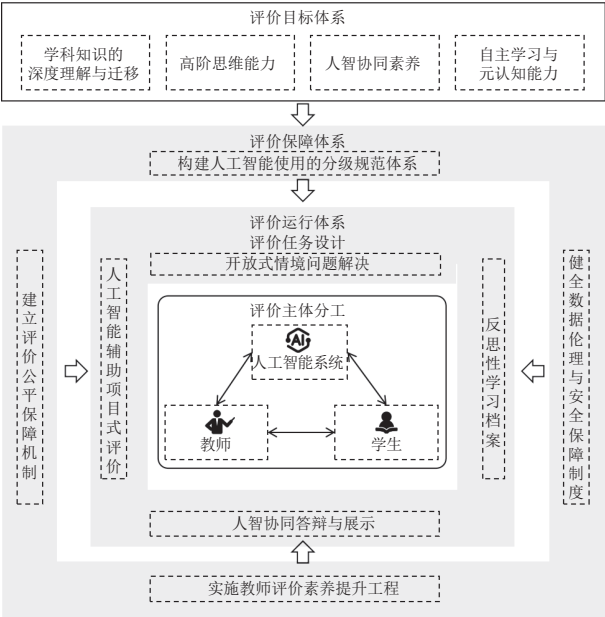


图 2 三层联动实践框架

有最终裁决权(刘邦奇, 2025); 学生通过结构化的自评工具与规范化的互评流程参与评价。

(三)保障层

保障层为框架的运行提供制度基础与条件支

撑。它包含四个方面:一是规则判定,构建人工智能使用的分级规范体系,按课程性质与任务类型建立禁止使用、允许辅助使用与要求使用三级规则,并配套实施情况的透明声明制度(Chan, 2023),为运行层中任务由谁完成提供制度依据;二是公平保障,确保学生能够平等获取评价所需的工具与资源;三是教师评价素养的提升,聚焦情境化任务设计与人智协同评价的协调;四是数据伦理与安全制度,过程数据的采集与使用遵循知情同意和最小必要原则,评价算法保持必要的透明度。

保障层还承担周期性重建匹配关系的功能。院校需要建立评价方案与教育目标匹配度的定期审查制度,将技术能力、学生使用方式与社会期待的变化及时纳入常规监测,并据此修订目标、任务形态与人工智能使用规则。

六、讨论和结论

(一)实践框架的适用边界与实施条件

该实践框架的适用性受学科性质、制度环境与技术条件的约束。学科差异主要体现在两个维度:一是人工智能对该学科核心能力的可替代程度,

表 2 四个能力维度与核心指标

| 评价维度         | 核心指标                                                 |
|--------------|------------------------------------------------------|
| 学科知识的深度理解与迁移 | ①概念关联能力;②跨情境迁移能力;③知识整合应用能力                           |
| 高阶思维能力       | ①批判性分析能力;②创造性方案生成能力;③复杂问题结构化拆解能力;④跨学科综合推理能力          |
| 人智协同素养       | ①有效运用人工智能工具的能力;②批判性审视人工智能输出的能力;③人智协同决策能力;④人工智能伦理认知能力 |
| 自主学习与元认知能力   | ①学习策略选择与调整能力;②学习过程自我监控能力;③反思与自我改进能力                  |

表 3 四类评价任务形态: 特征、适用条件与实施要点

| 任务形态        | 核心特征                              | 人工智能使用规则                                      | 考查重点                                                |
|-------------|-----------------------------------|-----------------------------------------------|-----------------------------------------------------|
| 开放式情境问题解决   | 设置多种可能路径的真实问题,无唯一答案               | 允许使用人工智能辅助信息检索与方案比较,但需明确标注人工智能使用方式与个人判断依据     | ①问题分析的深度与多角度;<br>②方案的创新性与可行性;<br>③批判性筛选与整合人工智能信息的能力 |
| 人工智能辅助项目式评价 | 围绕真实问题设计跨周期项目,学生自主决定人工智能使用方式与时机   | 要求人工智能作为项目工具之一,维护人工智能使用日志并说明人机如何分工            | ①完成项目的策略选择与调整能力;<br>②人智协同的合理性与有效性;<br>③作品的原创性贡献     |
| 人智协同答辩与展示   | 学生展示成果并回应现场提问,重点考查对作品的深度理解与即时思辨能力 | 作品准备阶段可使用人工智能,答辩环节不使用人工智能,考查学生脱离人工智能辅助的独立思考能力 | ①深度理解作品的核心论点;<br>②即时回应质疑的思辨能力;<br>③个人贡献与人工智能贡献的区分度  |
| 反思性学习档案     | 学生持续记录学习关键节点的决策、困惑、突破与反思          | 允许使用人工智能辅助整理和结构化反思内容,但核心判断由学生完成               | ①元认知监控与调节能力;<br>②自我认知的准确性与发展性;<br>③学习策略的选择与优化过程     |

表4 评价主体的角色分工

| 评价环节     | 教师职责                                             | 人工智能系统职责                                     | 学生职责                              |
|----------|--------------------------------------------------|----------------------------------------------|-----------------------------------|
| 评价方案设计   | ①确定评价目标与维度权重;<br>②设计评价任务与评分标准;<br>③设定人工智能使用规则    | ①提供学情数据分析优化方案;<br>②生成评价任务的备选方案供教师参考          | ①了解评价方案与标准;<br>②参与人工智能使用规则的讨论     |
| 过程数据采集   | ①观察学生课堂表现与协作行为;<br>②记录答辩的关键表现                    | ①自动采集人机对话记录、操作行为轨迹与文本修改历程;<br>②生成学习行为数据可视化报告 | ①维护学习日志与反思档案;<br>②记录人工智能使用情况与决策理由 |
| 形成性评估与反馈 | ①基于人工智能提供的学情报告进行判断;<br>②为学生提供个性化发展建议             | ①生成学习进度报告与能力发展画像;<br>②识别学习困难模式并触发预警          | 开展自我评估并对照人工智能反馈和教师建议进行反思          |
| 终结性评价    | ①对高阶能力进行质性判断;<br>②综合多源信息给出最终评定;<br>③审核校准人工智能评价结果 | ①提供量化评估数据;<br>②生成过程性表现的综合分析报告                | ①完成自评报告;<br>②参与同伴互评               |
| 评价结果应用   | ①基于评价数据优化后续教学设计;<br>②为学生制定个性化发展指引                | ①生成个性化学习资源推荐;<br>②更新学生能力发展模型                 | ①基于反馈调整学习策略;<br>②制定个人发展计划         |

二是该学科学习过程的可数据化程度。两者组合形成四种情形(见图3)。

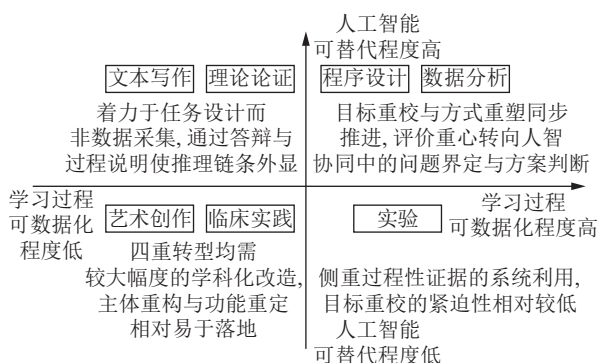


图3 学科条件与实践框架着力点的对应关系

程序设计与数据分析类课程被人工智能替代的程度高,学习过程又容易留下痕迹,对它的评价要从最终代码转向问题界定与方案取舍,学生提交需求分析、备选方案的比较与选择理由,并把版本记录、人机对话记录作为过程证据。教师可就关键代码追问其取舍依据以及约束条件改变后的调整思路。

文本写作与理论论证类课程被人工智能替代的程度同样高,但学生的思维过程难以由行为数据刻画。对它的评价,教师可要求学生把期末论文拆成选题说明、提纲与文献取舍、初稿、修改说明环节分段提交,并增加现场答辩,请学生说明论点从何而来、哪些材料被舍弃及舍弃的理由。

实验类课程较难被人工智能替代,但学习过程数据完整,关键点在于把过程记录真正用于评价。

教师可将实验记录、操作步骤、异常处理与数据处理过程纳入评分系统,重点关注学生在实验失败或数据异常时的判断与调整。人工智能可用于数据处理与文献查阅,但实验方案的确定与异常判断须由学生完成并说明。

艺术创作与临床实践类课程的评价,教师可引入导师评议与同行评议,把重心放在创作过程或诊疗决策的推演上,通过创作手记、病例讨论、口头复盘等方式取证;人工智能主要用于资料准备,作品或方案的核心判断需由学生完成。

## (二)评价体系重构的教育价值取向

评价体系重构的深层困难不在于技术,而在于教育价值取向,即高等教育评价究竟服务于何种目的。人工智能的介入使过程数据的持续采集、学习行为的即时分析与个性化反馈不再是难题,但技术可行性不等于教育应用的合理性。

第一,过程性评价的大规模实施意味着学习行为被持续记录,学习场景变为被数据监测的透明场域,这虽能提升诊断精度,但也可能侵蚀学习的心理安全感。学生意识到学习过程被持续记录后,可能转向安全而可预期的学习策略,回避风险高但认知收益大的深层探索。第二,数据驱动的能力画像存在系统性盲区,学术志趣的形成、知识立场的确立与伦理敏感性很难被还原为行为指标,评价若过度依赖可量化的数据,这些发展维度就可能被边缘化。第三,评价也是师生之间关于学习意义与发展方向的持续协商,其教育性内嵌于人与人之间的理解与回应,无法被技术系统完整记录。评价体系的

智能化不应以教育关系的技术化为代价。

### (三) 评价体系变革的长期逻辑

失齐会随技术迭代反复出现,试图一劳永逸地解决问题并不现实。这意味着改革思路需要从关注方案设计转向关注能力建设;前者指设计尽可能完善的方案并长期有效,后者指培育高校在评价制度层面的组织学习能力,使其具备感知变化、诊断失调与快速调整的制度韧性,具体包括评价方案与教育目标匹配程度的定期审查、教师评价创新经验的结构化交流,以及将学生评价体验反馈纳入制度改进的机制。

对高校管理者而言,建立评价改革的持续迭代机制,比投入资源开发短期即过时的评价方案更具长期价值。评价改革的终点不应止于形成一份完善的制度文本,而应指向院校自我校准和持续回应变化的能力。

最后,本研究的局限性在于:失齐表现与深层逻辑的分析主要通过概念辨析与逻辑推演展开,尚缺乏经验研究的直接支持;实践框架与适用边界虽已讨论学科条件与制度环境,但这些条件的实际作用方式仍有待检验。未来研究可在以下方向推进:一是在具体课程中采用设计性研究方法,检验本研究提出的任务设计原则与目标体系;二是开展识别跨层级失齐与观察指标的经验研究;三是探索人工智能辅助评价的技术架构与质量保障机制,特别是多模态数据的采集、分析与应用;四是考察不同学科与制度环境下评价改革的差异化路径。

### [参考文献]

- [1] Biggs, J. (1996). Enhancing teaching through constructive alignment[J]. *Higher Education*, 32(3): 347-364.
- [2] Biggs, J., & Tang, C. (2011). *Teaching for quality learning at University* (4th ed.). [M]. McGraw-Hill Press: 83.
- [3] Black, P., & Wiliam, D. (1998). Assessment and classroom learning[J]. *Assessment in Education: Principles, Policy & Practice*, 5(1): 7-74.
- [4] Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning[J]. *Assessment & Evaluation in Higher Education*, 31(4): 399-413.
- [5] Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning[J]. *International Journal of Educational Technology in Higher Education*, 20(1): 38.
- [6] 陈默, 杨玉辉, 杨清元, 陈娟娟, 陈文智 (2025). 智能体赋能高等教育变革: 基于 DeepSeek-R1 的范式重构与“浙大先生”实践探索[J]. *现代教育技术*, 35(5): 111-118.
- [7] Cotton, D., Cotton, P., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT[J]. *Innovations in Education and Teaching International*, 61(2): 228-239.
- [8] 崔允灏, 雷浩 (2015). 教—学—评一致性三因素理论模型的建构[J]. *华东师范大学学报(教育科学版)*, 33(4): 15-22.
- [9] 格兰特·威金斯, 杰伊·麦克泰格 (2017). 追求理解的教学设计(第二版)[M]. 闫寒冰, 宋雪莲, 赖平, 译. 上海: 华东师范大学出版社: 14-25.
- [10] 胡钦太, 梁心贤, 刘颜帆, 王姝莉 (2025). 生成式人工智能如何影响学生发展——基于 31 项实验与准实验研究的元分析[J]. *现代远程教育研究*, 37(2): 83-91.
- [11] 胡小勇, 刘雪旋, 陈丽诗, 刘颜帆, 刘晓红 (2025). 人工智能融入基础教育: 助内卷抑或提新质?[J]. *开放教育研究*, 31(2): 45-54.
- [12] 黄涛, 张振梅, 刘三女牙 (2025). 以共存求共生: 人智协同共育如何可能[J]. *教育研究*, 46(1): 147-159.
- [13] Hutchins, E. (1995). *Cognition in the Wild*[M]. Cambridge, MA: MIT Press: 287.
- [14] 兰国帅, 蒋顷烁, 郑明扬, 肖琪, 宋帆, 张一春 (2026). 人智协同赋能大规模因材施教的生态构建与路径创新[J]. *开放教育研究*, 32(2): 55-65.
- [15] 李艳燕, 李红慧, 张香玲, Natthawin Cho, 包昊翌 (2026). 人智协同视域下的虚拟科学探究学习: 内涵逻辑与实践路径[J]. *现代教育技术*, 36(5): 38-48.
- [16] 廖宏建, 王慧敏 (2024). 从信息到生态: 融入 AIGC 的反馈素养与教学意涵[J]. *开放教育研究*, 30(6): 55-65.
- [17] 刘邦奇 (2025). 智能教育协同创新的原理与路径[J]. *开放教育研究*, 31(3): 92-101.
- [18] Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts[R/OL]. The Wharton School Research Paper, Available at SSRN: <https://ssrn.com/abstract=4391243>.
- [19] 戚佳, 徐艳茹, 刘继安, 薛凯 (2024). 生成式人工智能工具使用对高校学生批判性思维与自主学习能力的影 响[J]. *电化教育研究*, 45(12): 67-74.
- [20] Sadler, D. R. (1989). Formative assessment and the design of instructional systems[J]. *Instructional Science*, 18(2): 119-144.
- [21] Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work[J]. *Higher Education*, 76(3): 467-481.
- [22] 汪益民, 欧阳彬 (2026). “α 世代”的群体画像、教育困境与重构路径[J]. *开放教育研究*, 32(2): 36-43.
- [23] 王佑镁, 王旦, 梁炜怡, 等 (2023). “阿拉丁神灯”还是“潘多拉魔盒”: ChatGPT 教育应用的潜能与风险[J]. *现代远程教育研究*, 35(2): 48-56.
- [24] 王佑镁, 王旦, 王海洁, 郭府宁, 毛聪聪, 柳晨晨 (2025). 生成式人工智能教育应用的“去技能化”危机与应对——基于反转型逆向思维分析框架[J]. *开放教育研究*, 31(4): 97-108.
- [25] Wiggins, G. (1993). *Assessing Student Performance: Explor-*

ing the Purpose and Limits of Testing[M]. Jossey-Bass Press: 193.

[26] 吴砥, 郭庆, 李佳平(2025). 智能技术进步何以赋能教师发展[J]. 教育研究, 46(5): 43-54.

[27] 吴龙凯, 程浩, 黄敬御, 王正鑫(2025). 技术伦理视角下人机协同教育评价的运行机制与实践策略[J]. 中国电化教育, (1): 8-16.

[28] 岳彦龙, 罗江华, 李宏展(2025). 从支持到协同再到共生: 知识建构的演进脉络、形态重塑与未来图景——基于技术发展视角[J]. 开放教育研究, 31(5): 70-78.

[29] 余亮, 邓双洁, 张馨月(2025a). 人工智能技术赋能教育的演进脉络、内在逻辑和发展趋势[J]. 电化教育研究, 46(6): 13-20+28.

[30] 余亮, 张馨月, 邓双洁(2025b). 混合式教学中的人智协同学习: 本质特征与实践样态[J]. 西南大学学报(社会科学版), (5): 189-201.

[31] 祝智庭, 戴岭, 赵晓伟(2023). “近未来”人机协同教育发展新思路[J]. 开放教育研究, 29(5): 4-13.

(编辑: 李学书)

## Paradigm Shift and Practical Reconstruction of Higher Education Assessment Systems in the Age of Artificial Intelligence

WU Jiayi<sup>1</sup> & LI Tuoyu<sup>2</sup>

(1. School of Public Affairs, Zhejiang University, Hangzhou 310058, China; 2. China Institute of Science and Education Development, Zhejiang University, Hangzhou 310058, China)

**Abstract:** Generative artificial intelligence has produced a systemic misalignment between assessment systems and the learning practices in higher education. Taking Biggs' theory of constructive alignment as its analytical lens and extending it to the assessment system, this study first analyzes how the misalignment manifests itself across four aspects of assessment, namely objectives, tasks, processes and standards, and analyses its underlying logic along three dimensions: scope, depth and temporality. On this basis, the study proposes a fourfold paradigm shift for the assessment system, comprising the recalibration of assessment objectives, the restructuring of assessment agents, the redesign of assessment methods, and the redefinition of assessment functions. It also develops a three-tiered, interconnected practice framework comprising an objective layer, an operational layer, and a safeguard layer. It discusses the framework's boundaries of applicability under different disciplinary conditions, as well as the value orientations and governance logic involved in the reconstruction. The study suggests that the focus of assessment reform should shift towards building the institutional capacity to keep assessment and learning continuously aligned.

**Key words:** artificial intelligence; learning assessment; constructive alignment; human-AI collaborative literacy; assessment system reconstruction