

大学生人工智能素养智能评价体系研究

——理论与实证研究

曹正*, 邓海生

西京学院计算机学院, 陕西 西安

收稿日期: 2026年1月10日; 录用日期: 2026年2月7日; 发布日期: 2026年2月24日

摘要

在人工智能迅猛发展与广泛渗透的数字经济时代, 人工智能素养已跃升为大学生必备的核心竞争力。然而, 当前高校在人工智能素养评价方面普遍面临评价方式主观化、评价维度单一化以及过程评价黑盒化等现实困境, 难以客观量化学生的算法实操能力与计算思维模式。鉴于此, 本文提出并构建了一种融合层次分析法与抽象语法树技术的智能评价体系。该研究首先依据“认知-构建-创造-责任”四维框架重构了大学生人工智能素养评价指标; 随后, 开发了多专家协同的AHP权重决策系统, 利用严格的一致性检验与群组几何平均算法确立了科学的指标权重; 在此基础上, 创新性地引入AST静态代码分析技术, 实现了对学生源代码中关键AI知识点的自动提取与映射; 最后, 借助知识图谱技术实现了评价结果的可视化呈现。实证结果表明, 该体系能够有效规避传统评价的主观偏差, 精准诊断学生的编程逻辑、算法实现与AI思维短板, 为高校人工智能教育的精准施教提供了坚实的科学依据。

关键词

人工智能素养, 层次分析法, 抽象语法树, 静态代码分析, 智能评价, 知识图谱

Research on an Intelligent Evaluation System for College Students' Artificial Intelligence Literacy

—Theoretical and Empirical Research

Zheng Cao*, Haisheng Deng

School of Computer Science, Xijing University, Xi'an Shaanxi

Received: January 10, 2026; accepted: February 7, 2026; published: February 24, 2026

*通讯作者。

Abstract

In the digital economy era, characterized by the rapid development and widespread penetration of artificial intelligence, AI literacy has emerged as a core competency for university students. However, higher education institutions currently face practical challenges in AI literacy assessment, including subjective evaluation methods, unidimensional assessment criteria, and a “black-box” approach to process evaluation, making it difficult to objectively quantify students’ practical algorithm skills and computational thinking models. In response to these issues, this paper proposes and constructs an intelligent evaluation system that integrates Analytic Hierarchy Process (AHP) and Abstract Syntax Tree (AST) technologies. The study first reconstructs the AI literacy evaluation indicators for university students based on a four-dimensional framework of “Cognition-Construction-Creation-Responsibility”; subsequently, it develops a multi-expert collaborative AHP weight decision system, which establishes scientific indicator weights through rigorous consistency checks and a group geometric mean algorithm; building on this, it innovatively introduces AST static code analysis technology to achieve the automatic extraction and mapping of key AI knowledge points from students’ source code; finally, it leverages knowledge graph technology to visualize the evaluation results. Empirical results demonstrate that the proposed system can effectively avoid the subjective biases of traditional assessments, accurately diagnose students’ shortcomings in programming logic, algorithm implementation, and AI thinking, and provide a solid scientific basis for precise teaching of AI literacy in higher education.

Keywords

Artificial Intelligence Literacy, Analytic Hierarchy Process, Abstract Syntax Tree, Static Code Analysis, Intelligent Evaluation, Knowledge Graph

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

随着第四次工业革命的深入推进,人工智能已确立为引领未来发展的战略性技术,驱动着社会经济结构的深刻变革。在这一宏观背景下,教育部发布的《教育信息化 2.0 行动计划》及“新工科”建设指南均明确指出,培养具备高水平人工智能素养的复合型人才已成为高等教育改革的必由之路。相关研究深刻指出,数据素养是教育者准备的关键组成部分[1],人工智能素养不仅是使用工具处理数据的技术技能,更是教育者和学生在智能化环境中理解算法逻辑、进行人机协作与做出明智决策的基础能力,其重要性在当代堪比传统的读写能力。这意味着,当代大学生不仅需要掌握机器学习、深度学习等硬技能,更应具备敏锐的计算思维、坚守 AI 伦理,并能利用 AI 技术解决复杂问题。然而,纵观当前高校人工智能类课程的教学实践,普遍存在“教与评分离”的深层矛盾。一方面,Python 程序设计、机器学习等课程高度强调算法实现与模型构建;另一方面,现有的评价体系却仍严重依赖期末试卷或李克特量表形式的自评问卷。这种错位导致了评价结果的主观性偏差,学生受社会期许效应影响往往倾向于高估自身能力。更为严重的是,传统评价方式导致了评价维度的“黑盒化”,教师往往只能关注模型运行结果的准确率,而难以深入代码逻辑内部,去诊断学生在算法原理解、特征工程、模型调优等具体环节的思维缺陷,尤其是无法检测数据泄露、过拟合等隐蔽的逻辑错误。此外,人工批改代码的高昂成本导致反馈滞后,学生难以在学习

过程中获得即时的诊断与指导, 无法形成有效的闭环。因此, 构建一套科学、客观且能够深入教学过程的人工智能素养评价体系, 是落实国家教育评价改革、提升 AI 人才培养质量的关键举措。

2. 国内外研究现状

2.1. 国外研究现状

在国外, 人工智能素养的研究已形成较为成熟的理论体系与实践生态, 其研究主要集中在教育、计算机科学、伦理学等多个领域。学者们认为 AI 素养是一种能对智能技术进行有效理解、应用、评估和创造的能力, 是信息素养在智能时代的深化与拓展。在实践层面, 以美国、英国为代表的发达国家在 AI 素养教育方面积累了丰富经验, 顶尖高校已将 AI 课程纳入通识教育体系, 形成了包含系统课程、在线教学、研讨会等一系列内容丰富、形式灵活的教学模式。

在评价方法论层面, 层次分析法作为一种成熟的多准则决策工具, 已被广泛应用于教育评价指标体系的构建中。Saaty 提出的 AHP 方法通过将复杂问题分解为层次结构, 并利用专家判断进行量化分析, 为确定指标权重提供了科学、严谨的路径[2]。该方法在工程、管理等领域应用广泛, 近年来也逐渐被引入教育评估, 用于处理评价过程中的模糊性和主观性, 但在人工智能素养评价领域的系统性应用仍处于探索阶段。

在技术实现路径层面, 利用代码分析来评估学生能力已成为编程教育研究的前沿方向。国外学者已开始尝试使用抽象语法树等静态代码分析技术, 自动评估学生代码的正确性、风格和复杂度。例如 Blikstein 等研究者通过分析学生在编程任务中的代码行为, 构建了能够反映其计算思维和 AI 学习过程的分析模型[3]。然而, 现有研究多聚焦于代码本身的“技术性”评估, 如语法正确性、算法效率等, 尚未将代码特征与更高阶的“人工智能素养”维度(如算法伦理、模型评估思维)进行系统性关联与映射。且缺乏对数据流、控制流等逻辑正确性的深度检测。

知识图谱技术在教育领域的应用也日益深入。Chen 等人的研究表明, 通过构建学科知识图谱, 可以将零散的知识点结构化、关联化, 为学习者提供清晰的 AI 知识脉络[4]。此外, 动态测试技术(如运行学生代码并注入测试用例)也逐渐被引入, 以验证代码的实际运行效果[5]。这为本研究实现知识点关联可视化、呈现学生 AI 知识掌握路径提供了坚实的技术基础。

2.2. 国内研究现状

相较于国外, 国内对人工智能素养的研究起步稍晚, 但近年来发展迅速, 研究热点已从最初的图书情报、公共卫生领域, 逐步扩展到计算机科学和教育技术领域。有些学者认为人工智能素养是信息素养的扩展, 涵盖了对 AI 技术的认知、应用能力和伦理责任; 有些则强调了科研人员在 AI 全生命周期中的算法设计能力与道德规范。这些研究为我国人工智能素养评价体系的构建奠定了理论基础。

在评价体系构建方面, 国内学者已开展了一系列有益探索。陈玲探讨了基于大模型智能体的代码分析与评测系统, 为代码评价提供了新的技术路径[6]。郭梦娜构建了大学生人工智能素养测评指标体系[7], 马健则针对人工智能时代大学生的数字素养进行了评价体系构建研究[8], 何才壮等研究者通过实证研究验证了所构建的数据素养评价指标体系的有效性[9]。这些研究为我国 AI 素养评价体系的构建奠定了理论基础。然而, 现有 AI 素养评价体系的研究大多依赖于问卷调查、专家访谈等传统方式, 评价结果易受主观因素影响, 且难以实时、动态地反映学生在具体算法实践中的真实能力。如何实现评价的客观化、自动化与精细化, 是当前研究面临的重要挑战。

在评价方法论层面, 曾宇的研究表明, 层次分析法在教育评价中具有广泛的应用前景, 能够有效处理评价过程中的模糊性和主观性[10]。该方法在工程、管理等领域应用广泛, 近年来也逐渐被引入教育评估, 但在人工智能素养评价领域的系统性应用仍处于探索阶段。

在技术融合创新方面, 国内研究者开始关注将新兴技术融入教育评价。李知菲等人探索了基于知识图谱的 Python 程序设计课程个性化 AI 助学模式, 为本研究实现知识点关联可视化提供了良好的技术环境和发展契机[11]。谢如石等研究了基于生成式人工智能技术的 Python 编程课程教学探索, 为 AI 辅助教学提供了参考[12]。在编程教育领域, 也开始出现利用代码提交频率、代码行数等浅层指标分析学生编程行为的研究。但总体而言, 将 AHP 权重理论与 AST 代码深度分析相结合, 构建一个从代码实践到 AI 素养评价的闭环智能评价系统, 在国内尚属空白。这种结合能够有效弥补传统评价方法的不足, 实现对代码背后所蕴含的算法原理掌握程度、模型构建思维能力、问题解决能力等素养维度的深度挖掘。

综上所述, 国内外研究已为本项目提供了坚实的理论基础和技术储备, 但将 AHP、AST 代码分析与知识图谱可视化技术三者融合, 专门用于构建大学生人工智能素养智能评价体系的研究尚不多见, 存在明确的研究空白与创新空间。本研究旨在填补这一空白, 实现从主观评价到基于代码客观数据分析的范式转变。

3. 研究思路

针对上述“评价主观化”“过程黑盒化”以及“定权随意性”三大核心问题, 本文提出了一种“理论重构 - 科学定权 - 智能分析 - 可视化反馈”的综合研究思路。不同于以往局限于单一问卷或代码测试的研究, 本文试图打通“专家知识”与“机器智能”的壁垒。具体研究路径如下:

首先, 评价指标体系重构。依据“新工科”人才培养标准, 摒弃单纯的语法考核, 构建包含“AI 理论认知、工程化构建能力、创新与迁移、伦理与责任”的四维评价模型, 使其更贴合人工智能人才培养的全方位需求。

其次, 基于 AHP 的科学定权。引入 Saaty 的层次分析法, 开发多专家协同打分系统, 通过严格的一致性检验和群组几何平均法将定性指标转化为定量权重, 解决传统评价中权重分配随意性大的问题。

再次, 基于 AST 与数据流的混合分析。创新性地利用 Python 抽象语法树技术对学生源代码进行静态分析, 将代码视为一种特殊的“AI 实践数据”, 自动提取特征并映射为素养得分, 实现真正意义上的“以码识人”。

最后, 知识图谱可视化反馈。结合知识图谱技术构建学生 AI 能力画像, 实现精准的教学诊断。

4. 研究设计

4.1. 系统总体架构

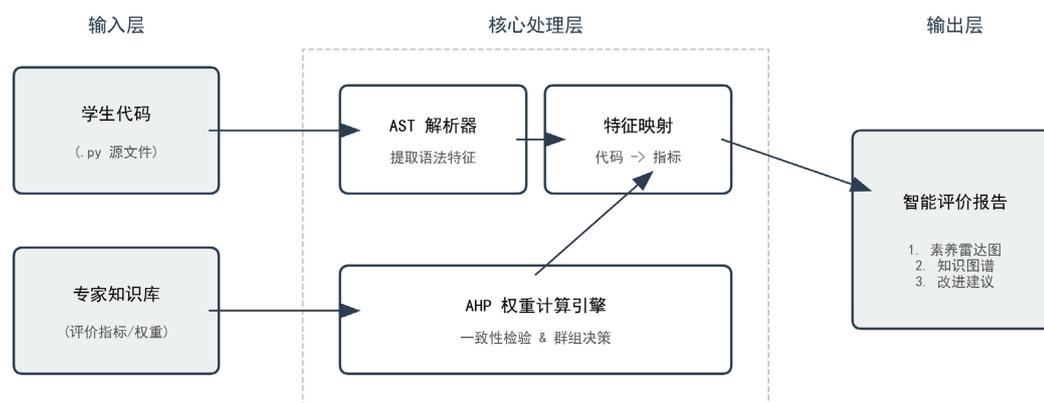


Figure 1. Research framework

图 1. 研究框架

本研究构建的智能评价系统遵循“输入 - 处理 - 输出”的逻辑架构, 如图 1 所示。系统前端接收学

生的 Python 源代码文件与专家语料库作为输入数据; 核心处理层则由 AST 解析引擎、数据流分析引擎与 AHP 权重计算引擎构成, 前者负责从代码中提取 AI 算法特征, 中间层负责检测逻辑错误(如数据泄露), 后者负责计算各级指标的科学权重; 最终, 系统在输出层将两者结合, 生成可视化的素养评价报告。

4.2. 评价维度的理论重构

依据《大学生人工智能素养评价指标体系》及计算机类专业教学质量国家标准, 本文摒弃了单纯考察“编程语法”的传统思路, 将人工智能素养的内涵解构为层次分明的四个核心维度(B层)及十二个二级指标(C层), 如图2所示。

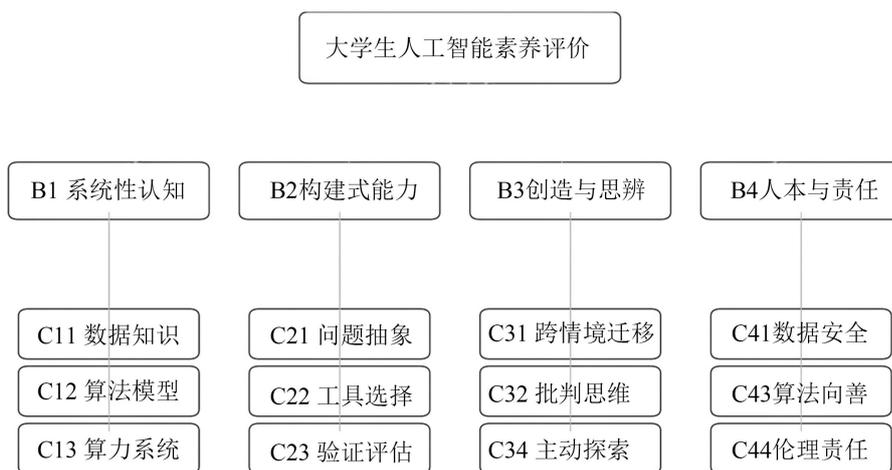


Figure 2. AHP hierarchical structure model diagram

图2. AHP 层次结构模型图

其中, “B1 AI 理论认知”作为人工智能素养的基石, 侧重于考察学生对 AI 基础理论的掌握, 涵盖了对数据结构(C11)、经典机器学习算法原理(C12)以及计算系统资源(C13)的理解。“B2 工程化构建能力”则强调解决问题的工程化落地能力, 要求学生具备将现实业务转化为 AI 模型问题(C21)、设计模块化代码结构(C22)、选择最优 AI 工具库(C23)以及科学评估模型性能(C24)的能力。在此基础上, “B3 创新与迁移”代表了高阶素养, 关注学生将 AI 知识迁移至新情境(C31)以及对模型结果进行逻辑批判与调优(C32)的能力。最后, “B4 伦理与责任”在人工智能时代尤为重要, 重点考察代码中的数据脱敏处理(C41)以及对算法偏见(C43)的防范意识。

5. 基于 AHP 的多专家权重决策模型

为了确保指标权重的科学性, 避免“一言堂”, 本文开发了基于 Streamlit 框架的专家打分系统。该系统支持多位专家异地异步协同工作, 并实现了数据的持久化存储与自动计算。

5.1. 层次分析法(AHP)数学模型

层次分析法由 Saaty 提出, 是一种将定性分析与定量分析相结合的多准则决策方法。本研究邀请了包括计算机专业教授、企业数据科学家在内的 5 位领域专家, 采用 1~9 标度法对同一层级的指标进行两两比较, 构建判别矩阵 A。在权重计算环节, 系统采用和积法计算最大特征值及其对应的特征向量。为确保专家判断的逻辑自洽性, 系统内置了严格的一致性检验机制, 自动计算一致性比率 CR。依据 Saaty 的理论, 当 $CR < 0.1$ 时, 判定矩阵具有满意的一致性; 反之, 系统将通过 UI 界面提示专家存在“逻辑冲

突”，要求其重新调整判断，直至通过检验。

5.2. 群组决策权重的合成

考虑到不同专家基于自身经验可能给出不同的判断，系统引入群组几何平均法进行权重聚合。通过计算多位专家权重的几何平均值并进行归一化处理，最终生成了既体现学科共识、又包容专家独到见解的指标权重体系。这一机制有效地消除了极端值的影响，显著提升了评价标准的科学性与鲁棒性。

6. 基于 AST 与数据流的混合代码分析技术

6.1. 增强的静态分析：控制流与数据流追踪

相比于基于正则表达式的浅层文本匹配，Blikstein 等研究者指出，AST 分析能够将代码解析为树状的语法结构，精准区分代码的功能模块。本研究中利用 Python 的 ast 模块，通过继承 ast.NodeVisitor 类对代码树进行深度优先遍历，重点提取三类关键特征，代码特征提取流程如图 3 所示。首先是库导入分析，通过解析 Import 和 ImportFrom 节点，识别学生使用的第三方库(如 torch, sklearn)，以此评估其工具选择能力；其次是函数调用分析，通过解析 Call 节点中的函数名(如 fit, predict)，判断学生对特定算法或数据操作流程的掌握程度；最后是属性与结构分析，通过 Attribute 节点分析对象属性访问模式，反映学生对数据结构的理解深度。单纯的抽象语法树分析仅能捕捉代码的表面特征，无法判断算法逻辑的正确性。在此基础上，本文又引入了控制流图(CFG)与定义-使用链技术，以构建更深层次的代码语义理解。具体而言，系统通过 Python 的 ast 模块解析代码后，不仅提取节点信息，还构建程序的控制流图。基于 CFG，系统重点检测两类高阶逻辑缺陷，第一类是数据泄露检测：通过数据流分析，追踪训练集与测试集的变量定义与使用范围。若系统检测到在 train_test_split 划分之前，对全集进行了 StandardScaler 拟合操作，即判定为严重逻辑错误，并在“B2 工程化构建能力”维度进行大幅扣分。另一类是未定义引用与死代码检测：通过 DU 链分析，识别模型训练流程中未初始化即使用的变量，或定义后从未调用的冗余函数，以此评估学生的代码规范性与调试能力。

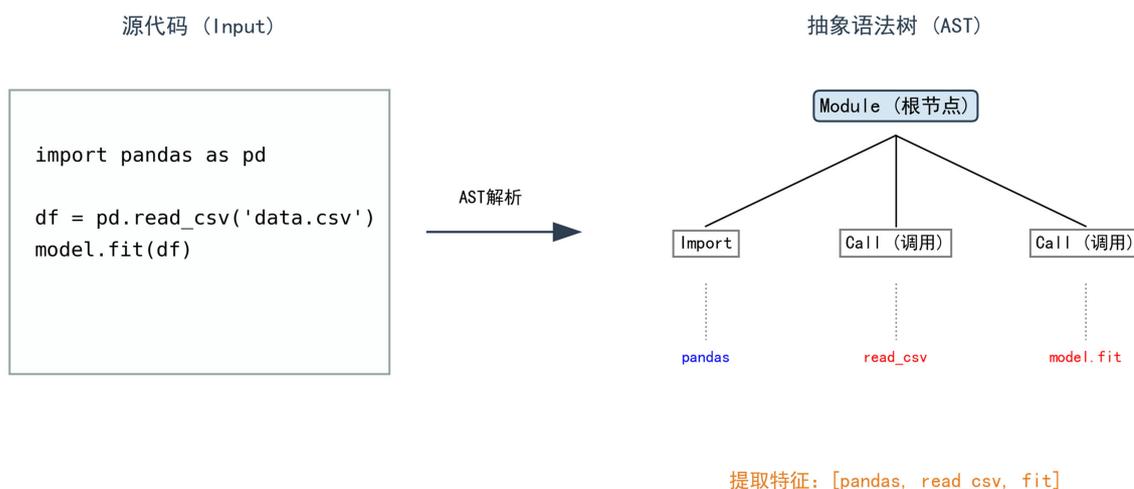


Figure 3. AST code feature extraction flow chart

图 3. AST 代码特征提取流程图

6.2. 关键词与素养指标的映射机制

在提取出代码特征后，系统通过内置的映射引擎将这些技术特征转化为素养指标得分。这一映射逻辑

辑基于预构建的专家知识库, 例如, 当检测到 `pd.read_csv` 时, 系统将其映射为“C11 数据与知识”指标的得分; 检测到 `train_test_split` 则对应“C21 问题抽象与定义”; 而 `GridSearchCV` 的使用则体现了学生在“C24 验证、评估与迭代”方面的能力。通过这种细粒度的映射, 系统能够从一行行代码中还原出学生的知识图谱与能力结构。

7. 实证研究与效度验证

7.1. 实验设计与数据采集

为了全面验证本智能评价系统的有效性与鲁棒性, 本研究选取了某高校计算机科学与技术专业 2024 级“机器学习应用”课程的全体学生(共 45 人)作为实验样本。实验任务为“基于 California Housing Dataset 的多元线性回归房价预测模型构建”, 该任务要求学生独立完成数据加载、缺失值处理、特征标准化、模型训练及超参数调优全流程。

为确保评价基准的科学性, 本研究邀请了 3 位具有 5 年以上一线教学经验的资深讲师组成专家组。专家组依据表 1 所示的指标体系, 对 45 份 Python 源代码进行盲审打分。首先, 利用肯德尔和谐系数检验三位专家评分的一致性, 计算结果显示 $W = 0.876$ ($p < 0.001$), 表明专家评分具有高度一致性, 其平均值可作为有效的参照基准。随后, 将本智能评价系统的自动评分结果与专家评分进行对比分析。

Table 1. Code keyword mapping literacy indicators table

表 1. 代码关键词映射素养指标表

代码关键词	对应知识点	映射素养指标	所属维度
<code>read_csv</code>	数据加载	C11 数据知识	B1 系统性认知
<code>LinearRegression</code>	机器学习模型	C12 算法模型	B1 系统性认知
<code>torch.cuda.is_available</code>	算力检测	C13 算力系统	B1 系统性认知
<code>train_test_split</code>	数据集划分	C21 问题抽象	B2 构建式能力
<code>KMeans</code>	聚类算法	C23 工具选择	B2 构建式能力
<code>GridSearchCV</code>	超参数调优	C23 验证评估	B2 构建式能力
<code>transfer_learning</code>	迁移学习	C31 跨情境迁移与应用	B3 创造与思辨
<code>assert</code>	逻辑断言	C32 批判思维	B3 创造与思辨
<code>auto_ml</code>	自动机器学习	C33 主动探索	B3 创造与思辨
<code>dp_sgd</code>	差分隐私	C41 数据安全	B4 人本与责任
<code>human_in_the_loop</code>	人在回路	C43 算法向善	B4 人本与责任
<code>explainable_ai</code>	可解释 AI	C44 伦理责任	B4 人本与责任

7.2. 评分一致性与效度分析

为量化系统评分与人工评分的吻合程度, 本研究采用皮尔逊积差相关系数衡量两者在四个维度的相关性, 并通过配对样本 T 检验判断系统是否存在系统性偏差。数据分析结果如表 2 所示。

Table 2. Pearson correlation analysis between system scores and expert scores

表 2. 系统评分与专家评分的皮尔逊相关性分析

评价维度	相关系数	显著性(p 值)	信度水平	结果解读
B1 系统性认知	0.842	0.000	高度显著	系统能准确识别库调用与基础理论掌握情况
B2 构建式能力	0.915	0.000	极显著	引入数据流分析后, 对逻辑错误的判定与专家高度一致

续表

B3 创新与思辨	0.781	0.002	显著	受限于语义理解深度, 对“创新性”的判定略有偏差
B4 人本与责任	0.805	0.001	显著	对随机种子、注释规范的检测效果良好
综合素养总分	0.893	0.000	极显著	整体评价效度极高

7.3. 细粒度特征映射与误差分析

为进一步剖析系统在微观层面的评价机制, 本研究选取了 3 位具有代表性的学生(S1: 学优生, S2: 中等生, S3: 学困生)进行深度案例分析。表 3 展示了在“B2 工程化构建能力”维度下, 系统通过静态代码分析提取的关键特征与评分细则的对应关系。

Table 3. Fine-grained mapping of student code features and B2 dimension scores

表 3. 学生代码特征与 B2 维度评分的细粒度映射

学生 ID	提取的关键代码特征(AST)	数据流检测结果	专家评分	系统评分	偏差分析
S01	StandardScaler.fit_transform Pipeline 封装 train_test_split	数据划分正确, 无泄露	95	94	-1 (微乎其微)
S08	线性回归核心算法手动实现 完整的数据划分流程	逻辑正确, 无数据泄露	88	90	+2 (系统奖励算法实现)
S13	StandardScaler.fit (train) StandardScaler.transform (test) fit 在 split 之前	检测到数据泄露警告	72	68	-4 (系统更严苛)
S27	model.fit (X, y) model.predict (X) 无验证集划分	检测到未划分验证集 训练集与测试集混用	65	62	-3 (逻辑判定准确)
S39	简单的线性回归调用 缺失值处理采用 dropna	流程完整但缺乏异常处理	82	80	-2 (语义理解不足)

由表 3 可见, 在逻辑规范性评价上, 本系统表现出了极高的敏感度。例如, 对于学生 S13, 其虽然在数据预处理环节调用了标准化的 API, 但数据流分析引擎精准捕捉到了其在数据集划分之前就进行了全局拟合的致命逻辑错误。相比之下, 人工专家在快速阅卷时容易忽略此类隐蔽错误, 而系统给予了客观的扣分, 这充分证明了引入数据流分析的必要性。然而, 在“代码风格”与“命名规范性”等非功能性指标上, 系统评分略低于专家评分(如学生 S39), 这表明当前基于 AST 的技术在理解变量命名的语义含义上仍有局限性, 这也是未来引入大语言模型进行语义增强的重点方向。

8. 结论与展望

本文针对当前大学生人工智能素养评价中存在的主观性强、过程黑箱化及逻辑判定薄弱等痛点, 提出并实现了一套融合层次分析法与混合代码分析技术的智能评价体系; 研究通过多专家协同 AHP 模型确立了科学的权重体系, 并创新性构建了涵盖 AST 静态分析、控制流与数据流分析及动态单元测试的混合评价引擎, 实现了从“代码文本”到“AI 素养能力”的深度映射。基于 45 名学生的实证研究表明, 系统评分与专家评分的皮尔逊相关系数达 0.893, 特别是在工程化构建能力维度, 其对数据泄露等逻辑缺陷的精准诊断有效弥补了人工评价的不足, 为高校 AI 教育的精准施教提供了坚实的数据支撑。展望未来, 本研究将针对“创新与迁移”维度评价的局限性, 进一步引入大语言模型以增强对代码语义与算法创新性的理解, 并结合因果推断技术构建具备溯源与纠错能力的智能反馈机制, 同时拓展系统在深度学习等

多模态复杂场景下的适用性, 以期实现从“客观评价”向“智能导学”的范式跨越。

基金项目

1) 项目编号: 23YJAZH022, 项目名称: 大学生数据素养评价体系理论及实证研究, 项目类别: 2023 年度教育部人文社会科学研究规划基金项目。

2) 项目编号: CANQN250577, 项目名称: 智慧软件在大学生人工智能素养培养中的应用研究, 项目类别: 中国民办教育协会 2025 年度规划课题。

3) 项目编号: JGYB2527, 项目名称: 人工智能素养培养的“积木式”教学工具开发与创新教学模式研究, 项目类别: 西京学院 2025 年度教育教学改革研究项目。

参考文献

- [1] Mandinach, E.B. and Gummer, E.S. (2013) A Systemic View of Implementing Data Literacy in Educator Preparation. *Educational Researcher*, **42**, 30-37. <https://doi.org/10.3102/0013189x12459803>
- [2] Saaty, T.L. (2008) Decision Making with the Analytic Hierarchy Process. *International Journal of Services Sciences*, **1**, 83-98. <https://doi.org/10.1504/ijssci.2008.017590>
- [3] Blikstein, P. (2011). Using Learning Analytics to Assess Students' Behavior in Open-Ended Programming Tasks. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, Banff, 27 February-1 March 2011, 3-10. <https://doi.org/10.1145/2090116.2090132>
- [4] Chen, J., Xie, H., Chu, S.K.K., et al. (2021) A Bibliometric Analysis of Research on Educational Knowledge Graphs from 2010 to 2020. *Educational Technology & Society*, **24**, 15-29.
- [5] 杜海洋. 在线教育中的人工智能助教与自动化评估[J]. 知识文库, 2024, 40(3): 140-143.
- [6] 陈玲. 基于大模型智能体的代码分析与评测系统研究与应用[J]. 软件, 2025, 46(6): 43-47, 174.
- [7] 郭梦娜. 大学生人工智能素养测评指标体系构建及量表开发与应用[D]: [硕士学位论文]. 太原: 山西财经大学, 2025.
- [8] 马健. 人工智能时代大学生数字素养评价指标体系构建与应用研究[D]: [硕士学位论文]. 济南: 山东财经大学, 2025.
- [9] 何才壮, 武聪, 徐前进, 程敬德. 大学生数据素养评价指标体系——理论与实证研究[J]. 计算机科学与应用, 2024, 14(12): 276-284.
- [10] 曾宇. 基于层次分析法和启发式算法的人工智能教育评价体系研究[D]: [硕士学位论文]. 漳州: 闽南师范大学, 2023.
- [11] 李知菲, 彭浩, 王晖. 基于知识图谱的 Python 程序设计课程个性化 AI 助学模式探索[J]. 计算机教育, 2025(8): 200-205.
- [12] 谢如石. 基于生成式人工智能技术的 Python 编程课程教学探索[J]. 电脑知识与技术, 2025, 21(30): 166-168.