

## 教育智能框架下计算机课程项目报告智能评估方法探索

孙清扬, 王天瑞, 盛 鹏, 王谦益, 章晓庆, 刘 江

(南方科技大学 计算机科学与工程系, 广东 深圳 518055)

**摘 要:** 针对现有计算机课程项目报告的评估标准模糊、评估过程主观性强并且耗时耗力, 从而导致课程项目报告评估质量与效率难以达到预期, 以及学生难以获得全面而有效的指导与反馈的问题, 提出教育智能框架下基于生成式人工智能技术的计算机课程项目报告智能评估方法, 通过人工智能导论课程的项目报告说明评估方法的有效性。

**关键词:** 课程项目报告; 智能评估; 生成式人工智能; 教育智能

### 0 引 言

在许多计算机及相关专业的课程教学中, 授课老师通常设置了课程项目模块并作为课程教学的重要环节, 旨在培养学生的实践技能、逻辑思维与学术写作能力, 直接关系到学生的学术发展和未来职业竞争力<sup>[1-4]</sup>。其中, 项目报告是课程项目的重要组成部分, 不仅可以评估学生对课程知识掌握程度, 还是一个促进他们技能全面发展的实践。因此, 课程项目报告撰写逐渐成为当代计算机系学生必须掌握基础且重要的技能。然而, 当前课程项目报告的评分标准尚不完善。同时, 授课老师精力有限, 面对大量的课程项目报告, 难以在短时间内容对它们进行详细地审阅与反馈。近来, 生成式人工智能技术 (Generative Artificial Intelligence, GAI) 的快速发展备受瞩目, 如大模型具有强大的知识推理与内容创造能力, 可以胜任多种任务, 包括教育智能领域<sup>[5-9]</sup>。

教育智能 (Education Intelligence, EI) 是人工智能与教育的深度结合, 指借助人工智能技术理解教育。在教育智能框架下, 探索如何利用大模型所具有强大的知识推理与生成能力, 实现课程项目报告的快速与精准评估, 有效地辅助授课老师在有限的时间内评估学生对课程知识的掌握程度和教学效果, 也使 学生获得及时且有针对性的指导与反馈, 更具有目的性且有效地完善与改进课程项目报告, 从而有助于学生进一步提升知识的掌握程度、逻辑思维与学术写作能力。

### 1 基于生成式人工智能的计算机课程项目智能评估方法

在传统的计算机课程项目报告评估过程中, 学生经过课程项目的实践以后, 将实验结果与自身对课程知识理解融合, 撰写成一份结构清晰、内容详实的课程项目报告。随后, 授课老师通常阅览学生所提交的项目报告, 并为报告给出评分和简短评阅作为反馈与指导。在这一过程中, 课程项目报告的评估标准存在一定模糊性, 不够普适与细化, 难以全面且准确地反映学生对课程知识的熟练度及报告在内容深度、逻辑结构、创新思维等多维度的优劣。与此同时, 在繁重的教学任务和科研压力下, 授课老师往往难以保证每份报告都能得到充分的阅读和细致的评估, 使得评估结果具有一定的主观性且缺乏有效的互动。以上原因在一定程

度上降低了课程项目报告评估工作的质量和效率。由于缺乏多维度的综合评价、具体的评语与修改意见，不利于学生全面了解自身的优势与不足，进而无法有针对性地调整学习策略和提升学习效果，同时增强他们的逻辑思维和学术写作能力。

为了解决以上问题，在教育智能框架下，基于生成式人工智能技术的计算机课程项目报告的智能评估方法如图1所示，旨在辅助教师快速量化分析课程项目报告，生成个性化的多维度反馈，从而提升教师的报告评估效率与质量，以及让学生能够在短时间内获得更加精准、个性化的评分与改进建议，这有助于学生定位自身不足并改进，从而实现对课程知识更好的掌握与综合素质提升。首先，学生在课程学习过程中，借助GAI辅助工具例如大模型去辅助课程项目实践及项目报告撰写，并在规定时间通过教学系统把项目报告提交给老师。其次，教师收到学生的课程项目报告后，对课程项目报告进行初步地评阅与打分，同时借助自身经验设计可量化的评分标准，并在相应的提示语帮助下利用GAI辅助工具对项目报告进行智能评估，得到评分和评阅意见。第三，教师结合评分和评阅意见对课程项目报告进行最终打分和撰写对应反馈。最后，学生收到老师对他们的课程项目报告的评分和有针对性的反馈，在GAI辅助工具引导下反思和改进自身劣势与不足，从而进一步弥补知识短板，提升学术写作能力和逻辑思维。

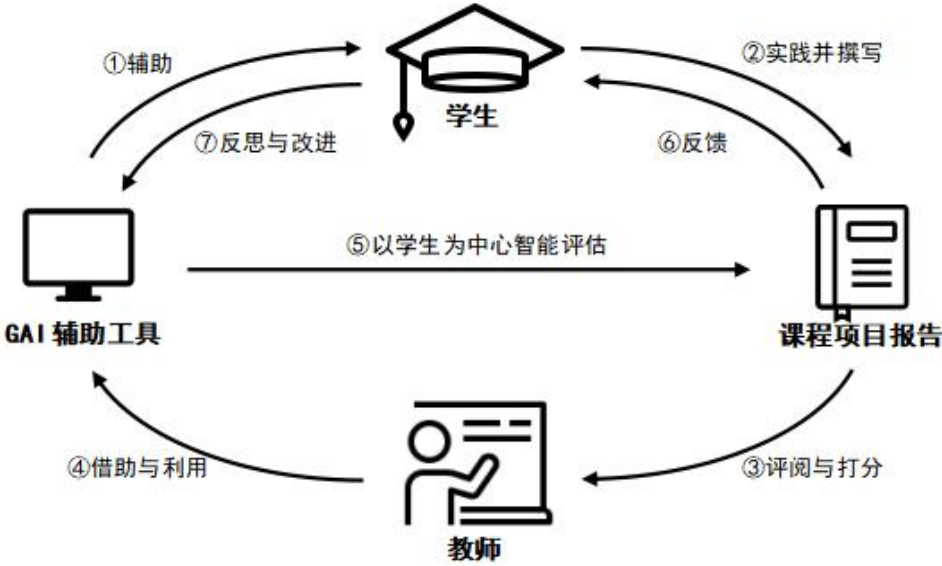


图1 教育智能框架下计算机课程项目报告的智能评估方法

## 2 计算机课程项目报告智能评估方法的实践

为了验证所提出的智能评估方法能够辅助教师高效准确的报告评阅与精准全面的评语撰写，进而有助于学生的知识短板弥补与写作能力提升，选取了来自人工智能导论课程项目报告进行测试。人工智能导论课程是面向南方科技大学全校学生开设人工智能基础入门课程，选修学生来自不同专业与年级，他们对人工智能知识的认掌握深浅不一，写作水平参差不齐。课程项目涉及多个领域，如甲骨文识别、自动驾驶、疾病监测等，这要求教师在批阅学生报告时具备丰富的跨学科知识与高度的专业素养，以准确评估不同水平的学生在不同技术前沿和应用领域的探索成果与创新深度。因此，高效评估学生提交的课程项目报告，并给出准确的评分和全面而有针对性的反馈是一项艰巨的任务，也能较好验证所提方法的有效性。

### 2.1 多维度课程项目报告评价标准及提示评分模版设计

为了全面而准确地评价课程项目报告，评估标准应从多角度综合考查课程项目报告的内

容质量、创新性与写作规范，这样可以确保报告评估的公正性和科学性，使评分过程更加标准化、可量化和自动化，从而提升评估的整体效率。根据教师的教学经验，课程项目报告的评估可以基于以下6个维度开展。

（1）结构完整性（20%）：评估课程项目报告的整体结构是否严谨，内容的布局是否合理。报告应包括清晰的引言、方法、结果、讨论和结论等部分，并确保内容的逻辑顺序和章节间的衔接顺畅。

（2）逻辑清晰度（20%）：考查学生在课程项目报告中所呈现的论证过程是否逻辑清晰、条理分明，重点关注论点的提出、论据的支持以及结论的得出是否循序渐进，是否准确地表达报告的核心思想。

（3）语言连贯性（20%）：评估报告的语言表达是否流畅，句子结构是否合理。报告中的用词是否准确，表述是否通顺无误，避免因语言问题导致的理解障碍。

（4）内容独特性和创新性（20%）：检查报告中的观点、方法和结论是否展现了独到的见解和创新性。评估学生是否在报告中提出了新颖的观点，是否能在已有研究的基础上提出独立的见解，并有效解决选题中的问题。

（5）参考文献规范性（10%）：考量学生在报告中引用文献的规范性，是否按照学术标准正确引用了相关文献。重点检查参考文献的选择是否科学合理，是否全面涵盖了研究所需的核心文献，且引用格式是否符合规范。

（6）课程知识掌握度（10%）：评估学生在报告中对所学课程知识的掌握和应用程度。检查学生是否能够将课程中学到的理论知识有效应用于实际问题的分析和解决，确保报告内容与课程教学目标的一致性。

选取人工智能导论课程项目报告共25份，教师根据所设计多维度课程项目评分标准对每份报告的各项指标逐项评分，并将各项评分加权计算出总分。所有课程项目报告的评分情况见表1。人工智能导论课程项目报告学生总得分分布如图2所示，课程项目报告总得分分布大致符合正态分布。

表1 基于多维度课程项目报告评价标准的人工智能导论课程项目整体评分情况

结构完整性	逻辑清晰度	语言连贯性	内容独特性	参考文献规	课程知识掌	总分	总分	总分
平均分	平均分	平均分	与创新性平	范性平均分	握度平均分	平均分	最高分	最低分
			均分					
8.08	7.92	8.18	8.18	6.36	8.88	8.00	9.10	6.30

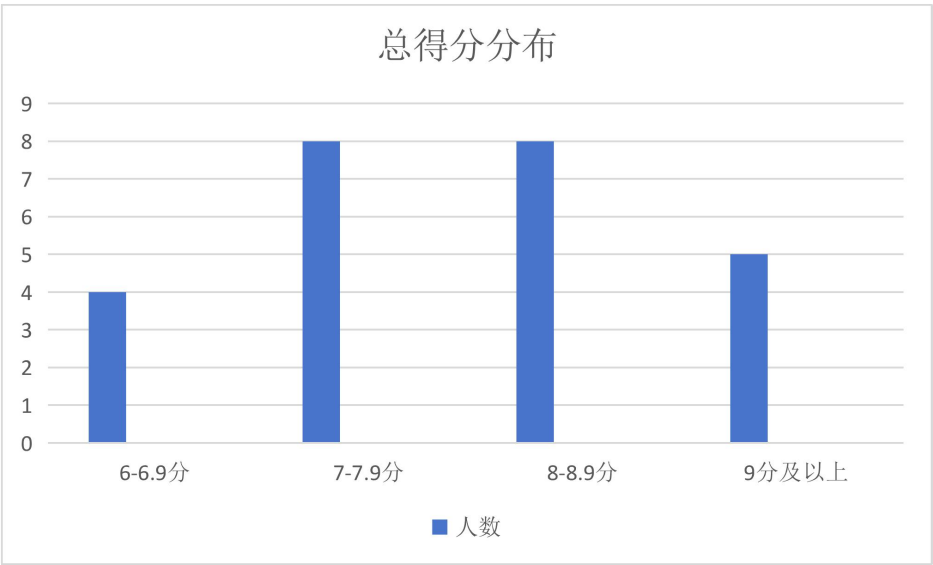


图2 人工智能导论课程项目报告学生总得分分布

在具体实践中，首先利用自然语言处理算法从课程项目报告中自动提取文本，再将文本输入到GAI辅助工具中进行深入分析和理解。模型将根据输入的报告文本反馈各维度的评分与详细评语，并提供具体的修改意见。为确保结论普适性和具有参考性，选择较为知名的中文大模型：文心一言、智谱；英文大模型：Gemini、Llama，这样的设计旨在比较不同模型间评分结果的差异，并更深入地分析各个模型的评分性能。为使GAI辅助工具准确理解评分任务，以多维度课程项目报告评价标准为基础，通过提示工程技术为模型设定特定角色，同时不断地优化和调整提示词，最终使用如下提示词：

“你是一位教授人工智能导论课程的大学教师，需要对学生提交的课程项目报告进行评估。请评估以下<报告文本>在描述<结构完整性>，<逻辑清晰度>，<语言连贯性>，<内容独特性和创新性>，<参考文献规范性>，<课程知识掌握度>方面的表现，并根据各指标<占比比例>进行打分与点评，打分范围0-10分。最终按照<评分模版>给出课程项目报告评分结果与修改意见”。评分模板如下：

“最终评分：<>（范围0-10分）

1. 结构完整性得分：<>，占比20%，原因如下：<>
2. 逻辑清晰度得分：<>，占比20%，原因如下：<>
3. 语言连贯性得分：<>，占比20%，原因如下：<>
4. 内容独特性和创新性得分：<>，占比20%，原因如下：<>
5. 参考文献规范性得分：<>，占比10%，原因如下：<>
6. 课程知识掌握度得分：<>，占比10%，原因如下：<>

修改意见：<>”

其中，<报告文本>为提取之后的项目报告，<>作为占位符，其内容可根据具体的课程项目报告数据集进行灵活的替换，以满足对应课程的需求，这套评分提示语模版保证了内容的格式化输出与评分结果的可解释性，有助于提高评分质量与反馈效率。通过利用GAI辅助工具实现报告的智能评估，教师可以高效地对项目报告实现全面的评估，并给予学生详细的反馈意见。学生可以从知识掌握度、内容创新性、写作水平等多个维度获得评分及评语，并根据教师有针对性的意见弥补知识短板，提高写作技能，实现逻辑思维的进步。根据评分模板，智能评估方法对输入项目报告的反馈结果如图3所示。



卷积神经网络在皮肤病相关的研究进展综述...

最终打分：8.5（范围0-10分）

1. 结构完整性得分：9，占比20%，原因如下：报告结构清晰，分为多个小节，包括CNN模型介绍、发展历程、算法及其改进、应用以及研究展望，每个部分都有明确的主题和内容，逻辑连贯，体现了良好的结构完整性。
2. 逻辑清晰度得分：9，占比20%，原因如下：...
3. 语言连贯性得分：8，占比20%，原因如下：...
4. 内容独特性和创新性得分：8，占比20%，原因如下：...
5. 参考文献规范性得分：9，占比10%，原因如下：...
6. 课程知识掌握度得分：9，占比10%，原因如下：...

修改意见：语言表达可以更加简洁，避免冗长的句子，并确保技术性术语的表达更加通俗易懂



图3 智能评估方法的反馈示意图

2.2 课程项目报告智能评估结果分析

表2为在相同评分提示模版下,不同GAI辅助工具对人工智能导论课程项目报告的评分结果,可知:①文心一言、智谱、Llama模型评分尺度均与教师相近,而Gemini模型在所有维度的评分普遍远低于标准,体现出与教师的评分风格的较大差异而。②文心一言模型对课程项目报告的结构完整性和逻辑清晰度评分较为精确,Llama模型在语言连贯性方面评估展现出更优的表现。③GAI辅助工具在内容创新性和课程知识掌握度这两项上的评分较为保守,然而教师充分认可学生在课程项目实践过程中所付出的辛勤努力,在这两项上都给予了学生较高的评价。④对于参考文献规范性,教师仅给出6.36分的平均分,而GAI辅助工具的评分均高于教师评分。学生在参考文献规范性这一评估项中得分较低,这体现出他们在学术写作过程中对于引用的规范性重视程度不够,具体表现为普遍存在缺少必要引用、引用格式不统一或不规范、以及引用文献数量较少(通常少于5篇)等问题。针对这一现象,教师在未来在教学中应当更加注重对学生引用规范性的训练和指导,确保学生在撰写学术论文时都能够充分引用相关文献,并且严格遵循统一的引用格式标准,从而提升学术写作的规范性和严谨性。⑤综合来看,文心一言模型综合表现最佳,在全部六个维度的评分表现较为接近教师评分标准,体现了GAI辅助工具在未来可以胜任辅助教师评估课程项目报告任务的潜能。

表2 不同GAI辅助工具基于多维度评分标准的结果对比

模型名称	模型版本	结构完整性平均分	逻辑清晰度平均分	语言连贯性平均分	内容独特性创新性平均分	参考文献规范性平均分	课程知识掌握度平均分
文心一言	ERNIE-Speed-128K	7.92	7.84	7.52	7.00	7.40	8.24
智谱	GLM-4	7.60	6.94	7.14	6.80	7.32	7.24
Gemini	Gemini-1.5-flash	6.80	6.16	6.76	4.92	7.32	6.36
Llama	llama-3.1-70b-versatile	8.52	8.44	8.10	7.76	8.38	8.38

此外,还计算了总分平均分、总分平均绝对误差与总分均方差以全面衡量不同模型的智能评分性能。平均打分作为模型对课程项目报告整体质量的直接量化,反映了模型评分倾向,平均绝对误差计算了模型评分与教师评分之间绝对差异的平均值,是误差大小的直接度量。均方差是模型评分与教师评分之间差异的平方和,量化了模型评分的波动性和准确性。

表3是使用GAI辅助工具对人工智能导论课程项目的打分结果,可知:①Gemini模型的评分偏低,所有报告的总分平均分为6.50分,其中仅有4份报告达到7分以上,与教师的评分尺度差异较大;文心一言、智谱、Llama模型的总分平均评分与教师评分风格较为接近。②在所有GAI辅助工具中,文心一言模型的评分表现最接近于教师的实际打分结果,其总分平均分与教师的评分仅有0.18分的差距,且总分平均绝对误差与均方误差在所有模型中均表现最好,体现出文心一言模型评分的准确性与稳定性。

表3 不同GAI辅助工具总评分结果对比

模型类别	模型名称	模型版本	总分平均分	总分平均绝对误差	总分均方误差
中文大模型	文心一言	ERNIE-Speed-128K	7.82	0.56	0.54
	智谱	GLM-4	7.39	0.76	0.89
英文大模型	Gemini	Gemini-1.5-flash	6.50	1.46	2.86
	Llama	llama-3.1-70b-versatile	8.24	0.59	0.58

通过课程项目报告智能评估方法的辅助,教师能在短时间内高效、全面地分析学生对课



程知识的掌握程度及报告在内容深度、逻辑结构、创新思维等多个维度上的表现,显著提升评估工作的质量和效率。对学生而言,这种智能评估方法不仅缩短了获得教师反馈的时间,更重要的是,这种智能评估方法能够提供多维度的具体评语与详尽的修改建议,这有助于他们全面、清晰地认识自己的优势与短板。在此基础上,学生可以更有针对性地调整学习策略,明确改进方向,从而有效提升学习效果,确保学习成果与社会实际需求紧密相连,实现知识的系统性掌握与实际应用能力的双重提升。

### 3 结 语

在教育智能框架下,课程项目报告的智能评估方法,利用GAI辅助工具提高课程项目报告的评估效率、可量化性、全面性,使得评分与反馈结果更具科学性、个性化。在人工智能导论课程项目报告上的实践证明,这种课程项目智能评估的方法,不仅具有辅助教师提高报告评阅的效率和准确性的潜力,还能提高学生的学术写作能力、逻辑思维以及对课程知识掌握程度。期待在未来,对GAI辅助工具进一步优化与深化,为学生和教师带来更为高效和智能的教学体验。

#### 参考文献:

- [1] 俞智慧,李富智,刘金华.基于渐近项目驱动的Python课程深度融合教学模式探索[J].计算机教育,2024(7):163-168,173.
- [2] 丁卉.开源AI平台助力人工智能类课程项目式自主协作教学模式研究[J].计算机教育,2024(6):120-124.
- [3] 牛瑞敏.基于深度学习的网页设计与制作项目式教学设计[J].计算机教育,2023(9):166-170.
- [4] 黎清万,钟嘉宝.AI视觉识别——人工智能项目式学习实践(PBL)设计[J].中国信息技术教育,2022(4):16-20.
- [5] 张红卓,周小宝,许玉焕,等.生成式人工智能赋能计算机程序设计类课程教学创新[J].计算机教育,2024(7):44-48.
- [6] 高洪皓,陈章进.人工智能赋能程序设计课程教学改革[J].计算机教育,2024(7):41-43,48.
- [7] 郭旦怀,吴若玲,卢罡,等.AIGC在大学计算机教育教学中的有效利用[J].计算机教育,2024(7):35-40.
- [8] 刘江,章晓庆.多媒体信息处理的主动学习教学框架[J].计算机教育,2023(6):150-155.
- [9] 刘江,章晓庆,巫晓,等.“ChatGPT+”驱动的课程教学创新设计——以南方科技大学多媒体信息处理课程为例[J].软件导刊,2024,23(2):172-176.

基金项目:广东省高等教育教学研究和改革项目“基于主动学习的多媒体信息处理课程教学的探索与实践——从‘多媒体’到‘真实世界融媒体’”(SJZLGC202202);南科大研究重点项目“用科研与应用给教学赋能——南科大交叉创新型本科生培养模式研究”(SUST23Z01)。

第一作者简介:孙清扬,女,硕士研究生在读,研究方向为教育智能,12332448@mail.sustech.edu.cn;刘江(通信作者),男,教授,研究方向为人工智能、精准医疗、眼脑联动,手术机器人和教育智能,liuj@sustech.edu.cn。