

性化特征和能力倾向等得到了全面呈现,这既有利于排除因考官个人倾向性等因素对考生成绩产生影响,也有利于解决传统面试场景下不同面试小组间的横向可比性问题。因此,ChatGPT 可以被视为个性化选拔人才模式下解决公平性难题的适切方案。

4. 服务于学生个性化成长。在 ChatGPT 支持下,过程性评价从过去的对整个学业过程进行评价,具体化为对考试过程进行评价,考试组织成本变化不大,但评价方式发生了巨大变化。这一变化可以抽象化为:考试过程作为整个学业过程的浓缩样本,不仅是对学习结果的客观反映,还是对学习过程、思维过程的客观反映。这一评价方式的转变,有利于激励学生加强思辨能力、培养创新思维,服务于学生的个性化成长,而不是片面追求考试分数。这样就自然解除了考生对于个性化成长与结果评价之间内在冲突的担忧,从而打开阻碍学生个性化成长的枷锁。(摘编自:徐光木、熊旭辉、张屹、魏晴晴《ChatGPT 助推教育考试数字化转型:机遇、应用及挑战》,《中国考试》2023 年第 5 期)

认知智能大模型在教育考试数字化转型中的应用:

1. 试题命制。随着认知智能大模型技术的日渐成熟,亟须开发试题命制环节的专用模型或定制化命题助手,以满足未来试题命制过程中的个性化、自动化、多样性需求。

个性化命题是指针对学生的特点和能力设计不同的考试题目或任务,以更好地评估他们的知识、技能水平。认知智能大模型可以根据学生的实际水平和需求生成个性化试题,还可以全面评估学生的知识和能力水平,并根据学生特点和学习进度调整命题的难度和深度,使考试更贴近学生的实际情况。自动化命题是指应用计算机技术自动生成考试题目并自动评分,具有提高效率、保证公平和降低错误率等优点。认知智能大模型通过学习历年试题和评分标准自动生成新题目,既能减轻命题人员的工作负担,又能提高命题效率和一致性。多样性命题是指在考试中采用不同形式的题目测试学生的知识和技能,并评估他们的综合能力和适应能力。多样性命题不仅是题型多样,还强调对学生多种能力的考查。认知智能大模型可以生成多样性试题,且涵盖不同的知识点和问题解决方式,鼓励学生提高创新思维和综合能力,使考试更具有启发性和挑战性。

需要注意的是,应用认知智能大模型技术进行试题命制时,必须结合教育领域专家的指导和人工干预,以确保命题的合理性、公平性和教育价值。

2. 交互式智能语言测试。将对话机器人应用于口语考试领域,可以打造具有直接测试功能的机器人考官,满足口语考试评测全面性要求。通过将认知智能大模型与口语评测技术、学科知识图谱相结合,机器人考官能够通过直接问答方式对考生进行口语测试。

在评测效果与考试组织效率方面,机器人考官的优

势表现为 3 点:首先,可以增加口语测试过程的互动性和及时反馈;其次,可以扩充评价维度,通过多模态技术采集考生答题过程中的神态动作,辅助评价考生的口语运用与表达能力;最后,可以应用认知智能大模型的逻辑推理能力,结合考生作答情况设置与其能力水平一致的问题,从而实现个性化口语评测。

3. 智能化评阅卷。认知智能大模型给智能评分技术带来新的发展机遇,主要有以下 3 个明显特征:第一,更全面的题型支持。新一代智能评分技术从对部分学科、部分主观题型的智能评分拓展为可覆盖全学科全题型的智能评分。随着认知智能大模型在语言理解、数学能力以及多模态方面不断取得进步,其对数学、物理、化学、生物等学科的手写文本、公式识别和自动评测能力也逐步在新一代智能评分中得到应用。第二,更深度的评价能力。智能评分系统能够更好地理解学生的语言表达和思维逻辑,从而提供更为准确、全面的反馈。新一代智能评分技术有能力从整卷出发对学生的作答进行全面评估。第三,更多维度的评卷质量检测。即通过人机独立评分方式对教师评卷结果进行质量检测。新一代智能评分技术从基于评卷数据的统计分析扩展到支持多模态教师评卷质量监测,即通过分析教师的行为动作、评卷轨迹、评卷分数等多模态数据,分时段对教师评卷质量进行精准预测,从而保证教师高质量、高效率地完成评卷工作。

4. 拓展基于考试数据的教育评价。认知智能大模型通过强大的数据分析和预测能力,可以从覆盖考前、考中、考后全流程数据中提取有价值的信息,进而为学生、教师和教育决策者提供有针对性的评估和指导。具体表现在 4 个方面:一是预测考试结果,并通过自动学习学生的学习模式、强项和弱项等信息,给学生提供有针对性的预测和建议。二是评估考试难度,从而评估考试质量、调整教学、改进教学大纲。三是提供个性化评估和反馈,有助于学生全面了解自己的学习状况、发现自身知识缺陷,并在教师指导下制订个性化的学习计划。四是改进和干预教学,为教师提供判断和决策支持。

5. 智能化考试管理与服务。基于认知智能大模型在语言理解、逻辑推理多模态交互方面的能力,可将其与教育考试管理系统相结合,作为教育考试主管部门的智慧助手,提升考务工作效率,赋能教育考试服务的全流程。以监考为例,基于认知智能大模型的智能化监考系统除实时记录、反馈考试画面外,还能根据已有考试的海量监控数据做机器训练和测试,并调节、制订准确度较高的异常情况监考数据组。同时,智能化监考系统还可以在网考、机考的考试端增加虚拟电子考官,并根据不同等级的异常情况由电子考官给出告警和提示。在考生服务方面,可用于考试咨询等话务服务。

(摘编自:汪张龙《认知智能大模型加速教育考试数字化转型》,《中国考试》2023 年第 8 期)

江苏教育考试科研月报

2023 年第 11 期(总第 107 期)

江苏省教育考试院编印

科研动态

聚焦新招考 共谋新发展

——中国教育发展战略学会教育考试专业委员会 2023 学术年会在苏州召开



11 月 16—17 日,中国教育发展战略学会教育考试专业委员会(简称“考专委”)第一届理事会第四次会议暨第四届学术论坛在苏州召开。会议回顾总结了 2023 年考专委工作并研究部署下一年的重点工作。教育部高校学生司副司长宁小华,中国教育发展战略学会副会长、考专委理事长姜钢出席会议并讲话;江苏省教育厅副厅长袁靖宇,西交利物浦大学党委书记钞秋玲致辞;江苏省政协常委,教卫体委员会分党组书记、主任,江苏省教育厅原厅长葛道凯作主题报告。常务副理事长刘敏主持会议。

会议指出,2023 年是我国建设教育强国的关键之年,也是深化教育考试改革的奋进之年。在这一年中,考专委始终坚持以习近平新时代中国特色社会主义思想为指导,坚持学术研究和服务并重的原则,紧密围绕考试招生工作领域中的热点、难点和痛点问题,组织召开了第五批高考综合改革省份研讨会、学术研究第二期课题中期评审会;积极开展“高考护航行动”,连续四年成功开展全国普通高校招生咨询指导公益活动;完善考试招生宣传媒体协作机制,加强招考战线全国宣传平台建设;联合北京师范大学珠海校区为多省招考系统干部开展专题培训。考专委按照既定目标,全面推进各项工作,较好完成了预期的工作任务。

会议强调,2024 年全国高考综合改革实施 10 周年,考专委要紧紧围绕教育强国建设这一新时代教育发展的重大战略主题,凝聚发展合力,发挥平台优势,持续加强学术研究和会员服务,更好服务于新时代教育考试招生工作,促进考试招生工作高质量发展。

会议要求,要聚焦招考改革,继续加强学术年会及专题会议的组织,开展更多学术研究和研讨交流活动,深化学术交流;要统筹全国资源,将公益咨询指导活动常态化、深入化开展下去;要优化培训模式,分批次为常务理事单位,高校、中学、基层会员单位等进行培训,切实提升队伍素质;要严格规范程序,抓实抓细,确保换届选举工作顺利运行。

考专委第四届学术论坛同期召开。学术论坛聚焦新时代背景下我国考试招生发展中的热点、难点,在考试招生高质量发展的内在要求下,以高考综合改革推进过程中的问题为导向,围绕新高考背景下的选考命题、投档模式、心理服务体系、学业水平考试系统研发以及人工智能在教育考试中的应用等领域进行深入探讨,旨在进一步凝聚改革共识,推动我国考试招生事业高质量发展。论坛上,来自教育部教育考试院、湖南师范大学,以及北京、四川、河南、江苏教育考试机构的 6 位报告人作了精彩的主题报告。

32 个常务理事单位,部分高校、中学、基层会员单位等 79 家单位,160 余人参加本次会议。

【编者按】11月17日,中国教育发展战略学会教育考试专业委员会第四届学术论坛在苏州召开。本次论坛上,江苏省教育考试院作《人工智能技术在教育考试中的应用研究》课题汇报,从课题研究背景、目标与研究过程、进展情况、存在的问题与改进方向出发,全方位、多层次讲解了人工智能技术在教育考试中应用的意义和研究进展。

人工智能等新技术在教育考试中的应用研究

(中国教育发展战略学会教育考试专业委员会第二期<2022年度>教育考试专项立项课题,课题编号: jyks202201)

课题承担单位:江苏省教育考试院

课题负责人:钱建良

一、研究背景

(一) 考试评卷中的通用需求

1. 提高效率。网上评卷需要大量评卷教师在电脑屏幕前付出巨大的工作量,需要进一步提高工作效率。
2. 质量与公平。利用人工智能技术,在现有人工评卷的基础上进一步提高评分质量,保障公平。
3. 成本与组织形式。人工成本越来越高,疫情等突发事件下的组织难度越来越大。
4. 统计与分析。繁重的评卷任务使得想通过评卷环节的分析来提升考生答题水平和评卷质量定量分析难以施行。

(二) 高利害考试评卷中的痛点

1. 定标阶段。极不常见的正确解法有可能未被纳入定标范围;难定位典型错误、了解不同的错误变形。
2. 评卷阶段。同一题目解法多,个别评卷教师有可能出现失误。

(三) AI 在考试行业的发展阶段

第一阶段是考试加计算智能。利用穷举和匹配搜索等方法实现海量学习资源的存储与传递,构建智能化考试信息管理系统。

第二阶段是考试加感知智能。主要是通过数学建模和基于大数据的深度学习等方法对人类感知能力进行模拟,以实现考生认证、行为分析、口语测评和智能评分等功能。

第三阶段是考试加认知智能阶段。模拟人类的推理、联想、知识组织能力,使得机器人一样能够理解、会主动思考并采取合理行动,具备一定的概念、意识和观念,能够实现真正的自适应学习与评判和考生行为预测。

本课题研究着力初探第三阶段的应用成效,即将具备逻辑推理能力和数值运算能力的认知智能与考试评卷相结合。

二、研究目标与研究过程

(一) 研究目标

通过“人工智能+考试评卷”实现公平性和高效性。公平性是通过人工智能技术完善优化评分标准制定、优化评分标准执行来实现;高效性是通过机器评卷减少评卷人员的投入、缩短评卷工作时间来实现。

(二) AI 在考试评卷中的主体流程

1. 识别阶段。基于手写体的中文、英文及公式的识别,力求做到准确高效。
2. 评分细则制定阶段。通过统计出常见解并进行解法聚类,来辅助完善评分细则的制定。
3. 评分阶段。有两种模式,一是通过学习专家评分来进行数据训练,二是无需数据训练,直接输入评分细则。
4. 复核阶段。如果有人机评分不一致的进入专家复核。

(三) 技术难点

1. 文本识别。对科学符号的识别,如数学符号、上下标等;排版的识别、手写字符识别与印刷体的排除;相似字符 AI 的判定。
2. 机器评分。对考生的回答进行语义理解并给分;对等价答案、复杂回答的给分;对复杂答案的给分需要与人保持一致。
3. 效率问题。对百万到千万级别的题目回答数据在一定时效内完成批阅。

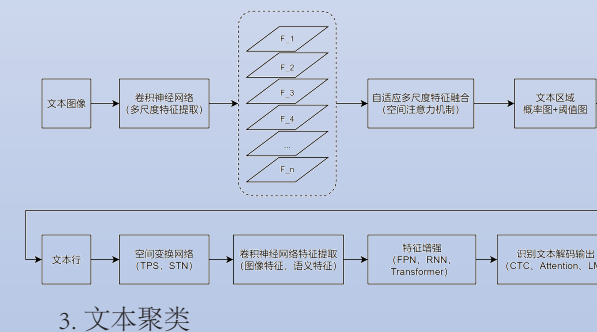
(四) 核心技术

1. 图文转写(手写字符识别)

主要的技术难点包括:图像版面结构复杂;答题内容存在大量删除、涂改、句中字符插入,且规范不统一;文本行字符存在倾斜、弯曲、交叉、粘连的情况;考生作答内容涵盖中文、英文、公式等特殊符号,等。
2. 算法流程

文本图像通过卷积神经网络(多尺度特征提取)、自

适应多尺度特征融合(空间注意力机制)、形成文本区域概率图和阈值图,在此基础上,再通过空间变换网络,卷积神经网络特征提取(图像特征和语义特征)以及特征增强等算法,识别文本解码输出,以解决之前提到的图片识别中的难题。

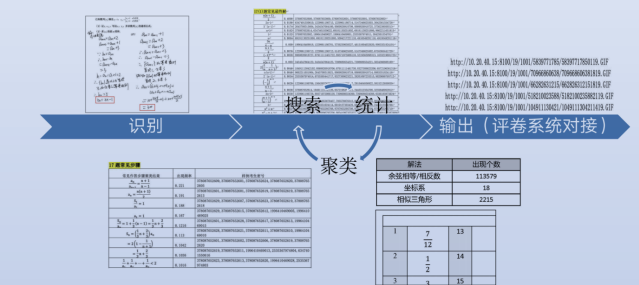


三、研究进展

高考评卷数学科目应用流程示例:

基于 ChatGPT 的教育考试数字化应用场景:

1. 海量题库的自动化生成。AI 人工智能的算法有很多,如决策树、随机森林算法、逻辑回归、线性回归等。借助一定的算法,ChatGPT 把海量的互联网信息和背后团队的知识贡献融为一体,使得瞬间自动生成海量题库成为可能。第一,在 ChatGPT 助手协助下,可在很大程度上减轻命题教师的非智力劳动工作量,节省大量的时间成本与货币成本,方便教师将主要精力用于创新性劳动;第二,题库容量被无限拓展;第三,除了能够自动生成试题,ChatGPT 同样能够对生成参考答案。
2. 基于 ChatGPT 的虚拟考场。首先,ChatGPT 凭借海量题库的自动化生成,使在线考试、虚拟考场的公正性获得进一步的保证,从而强化前者之于传统纸笔考试的优势地位;其次,目前虚拟考场已经实现了对人的表情、动作等的 AI 识别,如果再辅之以 ChatGPT 条件下试题自动随机生成等功能,其优势将得到进一步凸显;最后,随着 AI 技术的不断发展,传统纸笔考试和在线考



主要成果: 数学计算题、证明题

- (1) 可提高抽样效率;
- (2) 辅助完善评分细则;
- (3) 利用分类数据可降低评卷老师的误差以提升评卷质量。

四、存在的问题与改进方向

1. 存在的问题

- (1) 课题研究的考试类型偏少,只有高考;
- (2) 课题研究的考试形式单一,只有纸笔考试;
- (3) 缺少安全保密、自主控制的算力环境。

2. 改进方向

- (1) 扩展更多考试、学科等落地应用;
- (2) 与评卷系统无缝链接、与其他采集类型探索兼容;
- (3) 探索多种本地化部署及算力环境整合形式。

观点分享

试、虚拟考场将经历此消彼长的过程,前者将在越来越多的考试领域中被后者取代,如学校的平时测验和期末考试,地方性的学业水平考试,全国性的英语水平考试、高等教育自学考试等。在 ChatGPT 条件下,根据课程目标等自动随机生成符合要求的试题,并作出初步评判,使虚拟考场有能力成为 ChatGPT 等 AI 技术服务于教育考试数字化的重要增长点。

3. 服务于学校多元选拔人才。面试作为一种选拔个性化人才的方式,在世界范围内得到广泛应用,深受学生和学校的欢迎。以 ChatGPT 为代表的 AI 技术能够凭借其客观性、智能化等特征,摒除人情、技术等因素的困扰,形成一对一、个性化的考试选拔场景。在这种场景下,考官所担任的不再是提问者和倾听者的角色,而是对考生与 ChatGPT 之间的对话记录进行评阅的角色。每名考生因为回答上一道问题的内容和方式不同,ChatGPT 结合自己理解自动生成的下一道问题又有所不同。随着考生与 ChatGPT 对话的不断深入,考生的个