Advances in Education 教育进展, 2026, 16(8), 2113-2123
Published Online August 2026 in Hans. https://www.hanspub.org/journal/ae
https://doi.org/10.12677/ae.2026.1681858

Hans汉斯

从问答AI到智能体AI：高校辅导员工作重构

李滟婷

四川外国语大学国际金融与贸易学院，重庆

收稿日期：2026年7月21日；录用日期：2026年8月21日；发布日期：2026年8月31日

摘　要

目的：分析生成式人工智能由问答型向智能体型应用演进后，高校辅导员工作流程、育人关系与责任结构的变化，提出治理边界。方法：采用概念分析、功能比较和规范分析，以教育政策、智能体研究及典型学生工作任务为材料，从任务自主性、工具调用、后果外溢和责任归属四个维度展开。结果：智能体AI可能使技术由“生成答案”进入“组织任务并触发行动”，推动辅导员工作由内容生产转向任务编排、由事件响应转向过程支持、由结果审核转向全链条监督；风险也由内容错误扩展至错误执行、数据越界、标签固化、关系替代和责任悬空。结论：高校应按任务风险实行“可用－受限－禁止”分级，在关键节点设置人工批准、谈话核实、会商研判、异议复核和责任追溯。智能体AI宜限于低风险、可逆、可核验的辅助任务，涉及学生权益、危机处置和价值判断的事项必须由具名人员负责。

关键词

智能体AI，高校辅导员，人机协同，数字思政，教育治理

From Question-Answering AI to Agentic AI: Reconfiguring the Work of University Counselors

Yanting Li

College of Finance and Economics, Sichuan International Studies University, Chongqing

Received: July 21, 2026; accepted: August 21, 2026; published: August 31, 2026

Abstract

Objective: To examine how the shift from question-answering AI to agentic AI changes university counselors' workflows, educational relationships, and accountability, and to define governance boundaries. Methods: Conceptual, functional-comparative, and normative analyses were applied to

文章引用: 李滟婷. 从问答 AI 到智能体 AI: 高校辅导员工作重构[J]. 教育进展, 2026, 16(8): 2113-2123.
DOI: 10.12677/ae.2026.1681858

educational policies, research on AI agents, and typical student-affairs tasks across task autonomy, tool use, consequence spillover, and responsibility allocation. Results: Agentic AI may move from generating answers to organizing tasks and triggering actions. Counselors may therefore shift from content production to task orchestration, from event response to process support, and from output review to lifecycle supervision. Risks expand to erroneous execution, excessive data access, persistent labeling, relationship substitution, and accountability gaps. Conclusion: Universities should classify uses as permitted, restricted, or prohibited according to task risk and require human approval, direct verification, collective review, appeal, and traceable accountability. Agentic AI should be limited to low-risk, reversible, and verifiable assistance; decisions involving student rights, crisis intervention, or value judgment must remain with identifiable human professionals.

Keywords

Agentic AI, University Counselors, Human-AI Collaboration, Digital Ideological and Political Education, Educational Governance

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

《教育强国建设规划纲要（2024—2035 年）》提出以人工智能助力教育变革，教育部等九部门随后要求推进教育智能化并健全安全治理机制[1] [2]。《普通高等学校辅导员队伍建设规定》将思想理论教育、日常事务管理、心理健康教育、职业规划和危机应对纳入辅导员职责，这些任务同时涉及价值引导、关系信任、学生权益与组织责任[3]。教育部 2025 年高校辅导员人工智能专题培训也将技术理解、实践创新和伦理判断并列为能力要求[4]。因此，人工智能能否进入辅导员工作，不能只看技术能力，还要考察授权范围、核实主体和错误后果。

现有研究已从数字思政、精准思政和辅导员智能体建设等角度讨论人工智能的应用价值与风险。相关成果梳理了智能体的功能类型与部署问题，分析了生成式人工智能对思想政治教育时空结构、内容生产和主体关系的影响，并指出“智能思政”不能被简化为数据处理或媒介升级[5]-[8]。但实践中，问答机器人、知识库助手、工作流自动化和工具调用型智能体常被统称为“AI 辅导员”。不同系统的行动能力和风险后果差异明显，仅以“生成后人工审核”作为治理逻辑，难以覆盖流程中的权限、执行与责任问题。

本文关注的不是模型参数变化，而是人工智能从“给出回答”转向“组织步骤、调用工具并影响后续行动”。文章依次区分问答 AI 与智能体 AI 的边界，分析辅导员工作重构，并结合典型场景提出任务分级与风险控制框架。其核心观点是：智能体越接近流程和行动层，人工监督越应前移。

2. 研究设计

2.1. 研究方法与材料

本文属于理论与规范研究，不进行文献计量，也不据此判断高校智能体应用的普及程度。材料包括三类：教育数字化、辅导员队伍建设和个人信息保护等政策规范；智能体技术研究；生成式人工智能、高等教育与思想政治教育领域的相关成果。

分析分三步进行：首先，通过概念分析区分问答 AI 与智能体 AI；其次，按照“任务输入 – 系统处理

－外部行动－后果承担”分解辅导员工作，识别技术介入后的控制点；最后，依据数据敏感性、后果严重度、可逆性和专业判断需求，对典型任务进行风险分级。该框架旨在形成适用于不同平台和校本系统的判断原则。

2.2. 核心概念与技术栈

问答 AI 主要通过自然语言生成文本、解释信息或提出建议，通常不直接改变外部系统状态，其责任控制点主要位于输出之后，即由使用者核验并决定是否采纳。

智能体 AI 通常在大语言模型基础上增加目标分解、步骤规划、工具调用、记忆或状态更新和反馈修正等能力。ReAct 框架将推理与行动交替组织，相关综述将规划、记忆、工具使用和环境交互视为典型组成；生成式智能体与 AgentBench 等研究表明，智能体可以执行持续、多轮任务，但长程推理、指令遵循和稳定决策仍有明显局限[9]-[12]。本文所称智能体 AI，是在人工授权和制度约束下参与有限数字任务的系统，不将拟人化界面或连续对话等同于智能体。

从技术栈看，智能体风险并非只来自基础模型。大语言模型可能产生幻觉、价值偏差与指令误解；规划器可能错误拆解目标，或在长链任务中发生目标漂移；记忆库可能将临时困难固化为长期标签，并在跨场景调用中放大偏差；工具集则可能因权限过宽而误读、误写或误发。对应辅导员工作，上述风险可分别表现为政策解释失真、帮扶对象筛选偏差、学生画像固化，以及就业提醒、资助材料或危机处置流程中的越权执行。因此，控制点应覆盖模型输出、规划路径、记忆写入和工具调用，而非只审核最终文本[9]-[12]。

2.3. 理论基础与判断标准

公平、问责与透明度(fairness, accountability and transparency in machine learning, FAT/ML)为本文提供规范性分析视角。公平要求识别训练数据、标签和代理变量是否使特定学生群体承受系统性不利；问责要求能够确定数据授权、模型配置、人工采纳和纠错主体；透明度要求向辅导员和学生说明 AI 参与环节、数据来源、推荐依据及能力边界。FAT/ML 研究同时强调，公平并非模型内部可以脱离情境独立实现的指标，而应将算法、组织程序和受影响主体作为社会技术系统共同考察[13] [14]。

NIST 人工智能风险管理框架以“治理－映射－测量－管理”四项功能组织全生命周期风险活动，为本文的场景识别、风险评分、控制措施和持续审计提供操作化参照[15]。据此，本文将 FAT 原则落实为三个判断：是否对特定群体形成不利差异，是否存在具名责任与申诉路径，是否能够解释数据、规则及 AI 介入程度。

区分两类应用可考察四项指标：是否自主安排多步骤任务，是否调用数据库或业务系统，输出是否触发提醒、分类、推荐或状态变更，以及错误是否会影响学生权益、资源分配或危机处置。技术形态、工作介入和责任控制的同步迁移见图 1，两类应用的差异及人工控制要求见表 1。

3. 辅导员工作重构

3.1. 任务编排

问答 AI 多用于起草通知、生成班会提纲、修改简历或解释政策，任务顺序仍由辅导员掌握。智能体则可能串联多个环节。例如，在就业帮扶中，系统可读取脱敏台账、整理未落实去向学生和岗位信息、生成联系清单并设置回访提醒。该例仅说明功能可能，并不表示相关流程已普遍部署。此时，辅导员需要决定数据能否读取、分类规则是否合理、提醒对象是否准确以及流程何时停止。

智能体提高重复性事务效率，也使“任务设计”成为专业劳动。辅导员需把工作要求转化为可审计

规则，明确输入范围、禁用数据、人工审核节点、异常中止条件和保存期限。其劳动并未消失，而是由手工处理转向流程设计与质量控制。

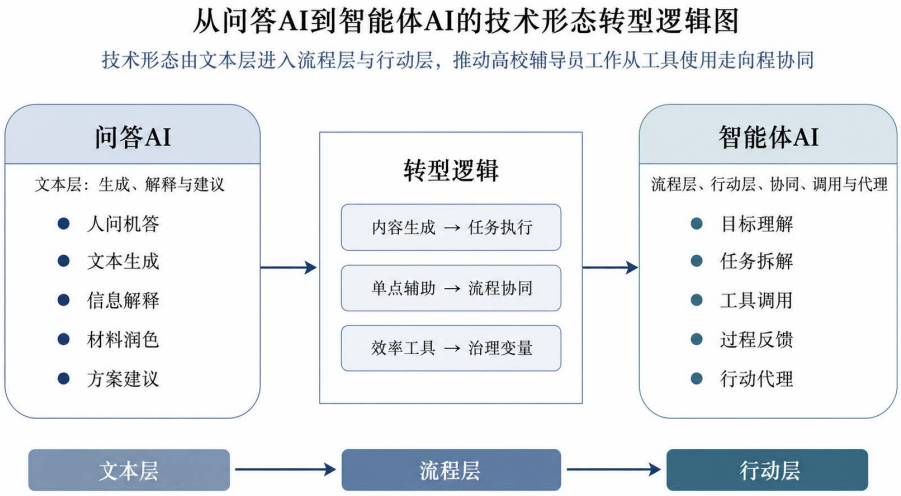

Figure 1. Logic of the technological transition from question-answering AI to agentic AI
图 1. 从问答 AI 到智能体 AI 的技术形态转型逻辑

Table 1. Differences between question-answering AI and agentic AI in university counseling work
表 1. 问答 AI 与智能体 AI 在高校辅导员工作中的差异

比较维度	问答 AI	智能体 AI	主要人工控制要求
任务对象	单次问题、文本或方案	目标导向的多步骤任务	先明确任务范围和终止条件
系统能力	生成、解释、改写、建议	规划、调用工具、更新状态、反馈修正	限制工具权限并设置关键节点批准
介入位置	文本层和建议层	流程层，部分进入行动层	监督从末端审核前移至过程控制
典型用途	通知初稿、谈话提纲、政策解释	材料核验、流程提醒、信息归并、反馈跟踪	核对数据来源、对象和执行结果
主要风险	幻觉、偏差、表达失当、代写	错误执行、数据越权、用途漂移、自动化偏信	保留日志、撤回机制和人工复核
责任结构	使用者对最终文本负责	授权、配置、审核、采纳与处置形成责任链	具名人员作出最终判断并承担责任

3.2. 过程支持

传统学生工作常从成绩下降、连续缺勤、就业受阻或材料缺失等事件开始。智能体可持续整理不同时间点的信息，将多次通知未读、材料退回或主动求助记录汇总为待核实线索，但这些信号并不自动等于学业困难或心理风险。

过程支持的价值在于减少遗漏，而非持续监测学生。相同数据在不同专业、年级和个人经历中含义不同，脱离谈话便可能将暂时状态固化为“高风险”“低参与”等身份标签。因此，智能体可以提示“需要了解什么”，不能直接判断“学生是什么样的人”。

3.3. 数据辅助研判

辅导员经验来自长期接触、组织观察和谈心谈话，能够识别数据难以呈现的关系变化与生活处境；智能体则擅长整理大量材料、追踪节点和发现重复模式。较稳妥的分工是系统负责整理与提示，辅导员通过直接沟通核对事实、理解原因并决定是否会商。

对“精准思政”而言，精准不应等同于细密标签化。既有研究既肯定人工智能识别需求、匹配资源的作用，也警惕算法偏见、价值偏差和主体弱化[6]-[8]。数据只能提高问题被看见的概率，不能替代对具体学生的理解。

3.4. 人本主导的三元互动

在可能的使用情境中，学生可先向 AI 咨询政策、学习或就业问题，再向辅导员确认；辅导员也可借助系统准备谈话、整理资源和记录安排，技术由此进入师生互动。UNESCO 相关文件强调以人为中心、保护人的能动性，并要求教育者理解技术局限、伦理风险和数据责任[16] [17]。因此，AI 可以扩展服务入口，但不应被塑造成替代真实关系的“数字辅导员”。

在情绪困扰、价值冲突和重大选择面前，学生还需要倾听、理解和现实支持。将陪伴与关怀外包给拟人化系统，可能削弱学生与真实支持网络的连接。技术应减少机械事务，让辅导员有更多时间开展面对面的教育与陪伴。

3.5. 全链条责任

问答 AI 的错误通常可在文本发布前修改，智能体错误却可能沿流程传导：字段读取错误会造成遗漏，不当匹配会形成持续偏向，自动提醒也可能在敏感情境中产生压力。因此，应明确谁批准数据接入、谁设定规则、谁检查中间结果、谁采纳建议以及谁处理申诉。

人机协同责任并非让机器分担责任，而是把人的责任落实到各节点。开发者和平台管理者负责安全与维护，学校负责制度、采购和数据治理，业务部门负责场景规则与流程监督，辅导员或专业人员负责具体判断和学生沟通。无法说明责任主体与复核路径的应用，不宜进入真实学生工作。

4. 典型场景边界

高等教育中的人工智能任务差异明显。相关研究显示，技术已用于内容生成、反馈、推荐、预测、画像和评价，同时伴随隐私、透明度、受益原则和教育者参与不足等问题[18]-[21]。因此，不能笼统判断某一场景“可用”或“禁用”，而应区分信息准备、辅助研判和权益决定。五类学生工作的适用边界见表 2。

Table 2. Scenario boundaries for counselors’ use of agentic AI
表 2. 高校辅导员使用智能体 AI 的场景边界

场景	可用任务(低风险)	受限任务及必要条件	禁止由 AI 独立完成的事项
思想引领	公开材料检索；班会提纲初稿；匿名反馈归纳	议题推荐和内容适配须核验来源、政治表述及适用对象	政治立场评价；对个体思想状态自动定性；未经审核发布内容
学业支持	课程政策解释；学习资源整理；节点提醒	基于脱敏数据形成待核实线索，须由辅导员谈话并与教学信息交叉核验	自动认定学业困难；据标签限制机会；向无关人员披露学习记录
心理健康	普及教育材料；匿名共性问题统计；预约流程提醒	个体线索仅限校内合规平台和专业流程，必须由专业人员复核	心理诊断、危机等级判定、治疗建议和转介决定

续表

就业指导	岗位信息归类；面试模拟；投递进度提醒	岗位匹配须解释依据，允许学生更正资料和拒绝推荐	虚构经历；自动排除学生；替代学生作职业选择
事务管理	材料缺项提示；政策问答；办理进度和截止日期提醒	资助、评优、处分材料可作形式核验，但不得绕过组织程序	资助资格、评奖结果、处分结论及其他权益决定

4.1. 低风险任务

通知初稿、公开政策检索、材料排版、缺项提示和一般流程提醒，通常不需推断学生个体状态，结果也易核验，适合优先试点。但仍须检查政策时效、事实来源和表达方式，防止旧规则、虚构链接或不当表述被直接发送。

4.2. 受限任务

学业线索整理、岗位匹配和个性化推荐涉及个体数据与分类判断，使用时应满足四项条件：数据来源合法明确；仅读取必要字段；推荐依据可理解；学生可以更正信息、拒绝推荐或申请人工复核。输出应表述为“待核实线索”或“可能需要支持”，不得将概率结果包装为结论。

4.3. 禁止自动化任务

心理危机等级、资助资格、评奖评优、纪律处分和重大舆情处置等事项后果重大，通常需要专业资质、集体程序和充分听取当事人意见。智能体可检查材料、汇总信息或提醒节点，但不能独立决定或执行。尤其不能将学生文字直接转化为心理诊断或危机等级。

5. 风险升级

5.1. 错误传导

大语言模型可能产生事实幻觉、错误引用和不稳定输出。问答场景中，这些问题主要形成错误文本；当系统可以调用工具时，错误还可能改变任务路径。例如，误解政策条件后，系统可能进一步筛选错误对象、生成错误清单并安排提醒。AgentBench 等研究指出，长程推理、决策和指令遵循仍是智能体可靠应用的障碍[12]，一次成功演示不能代表真实业务中的稳定能力。

控制错误传导不能只依靠末端审核。数据读取后应核对字段和范围，分类后抽查样本，发送或提交前由具名人员批准，执行后检查对象和结果。不可逆或难以补救的操作应取消自动执行权限，仅保留建议与草稿功能。

5.2. 数据与标签风险

学生工作资料可能包含身份、家庭经济、健康、处分、谈话和社会关系等个人信息，其中部分属于敏感个人信息。《个人信息保护法》要求目的明确、最小必要并加强敏感信息保护。对于向境内公众提供生成式人工智能服务的场景，《生成式人工智能服务管理暂行办法》还要求保护用户输入信息和使用记录[22] [23]；高校内部未向公众提供的系统虽不直接适用该办法，仍须遵守个人信息保护法。智能体连接更多文档和业务系统后，风险不仅是泄露，还包括将提醒数据用于评价、将临时帮扶标签长期保留等用途漂移。

数据治理应覆盖收集、输入、调用、输出、保存和删除全过程。学校不能以“内部使用”为由默认所

有数据均可进入模型，也不应使用公共平台处理可识别材料。校内平台仍需设置角色权限、字段隔离、访问日志、保存期限和删除机制，并为风险标记设定失效与复核规则。

5.3. 偏信、关系与责任风险

结构化输出、风险分值和整齐报告容易造成客观、专业的外观。研究表明，使用者可能因自动化偏信而忽视相反信息或系统错误[24]。在工作压力下，辅导员可能把系统结论当作事实，将谈话核实变成确认算法判断；学校也可能把回应、安抚和陪伴交给拟人化系统，以技术可用性替代真实支持。

责任悬空常发生在系统“只是建议”但实际上主导行动的情形。平台、学校和使用者可能相互转移责任。为此，应记录相关输入、模型输出、人工修改和最终批准人，并向学生提供可理解的说明和异议渠道。相应控制点见图 2。

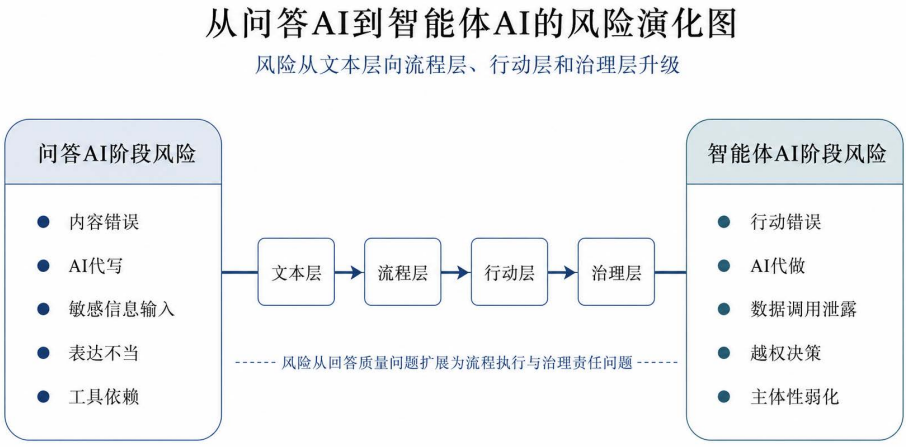

Figure 2. Risk evolution from question-answering AI to agentic AI
图 2. 从问答 AI 到智能体 AI 的风险演化

6. 应用治理框架

6.1. 建立任务分级清单

任务清单应以具体动词而非产品名称分类(见表 3)。例如，“生成公开活动通知初稿”可列为可用，“根据个体数据生成学业风险线索”列为受限，“自动认定资助资格”则应禁止。每项任务还应注明可读取数据、输出用途、审核人和保留期限，并在工具权限变化后及时调整风险等级。

Table 3. Risk assessment checklist for agentic AI tasks
表 3. 智能体 AI 任务风险评估清单

评估维度	0 分(低风险)	1 分(中风险)	2 分(高风险)
数据敏感性	公开或匿名信息	一般个人信息	敏感个人信息
行动权限	仅生成草稿或只读	提醒、推荐或分类	写入、提交或自动执行
后果严重度	易核验且不影响权益	错误需人工纠正	影响权益、危机处置或价值判断
可逆性	可立即撤回	可追溯并部分恢复	难以撤回或后果不可逆
专业判断需求	规则明确，无需个案判断	需辅导员情境判断	需专业资质或集体程序

五个维度合计 0~3 分为“可用”、4~6 分为“受限”、7~10 分为“禁止 AI 独立完成”。若任务涉及心理危机、资助资格、纪律处分或政治立场评价，无论总分高低，均不得由 AI 独立决定。评分结果应由业务部门复核，并在数据范围、接口权限或模型版本变化后重新评估。

6.2. 设置人类控制闭环

高风险控制不能只依赖最后一次确认。建议建立“AI 线索整理–辅导员谈话核实–学院会商研判–分类帮扶实施–反馈复评”的闭环。智能体负责汇总和提醒，辅导员核实事实与学生意愿，学院或专业力量处理需集体判断的事项，帮扶由明确责任人实施并记录。若信息、标签或建议有误，应及时纠正、撤回并追踪影响。闭环机制见图 3。

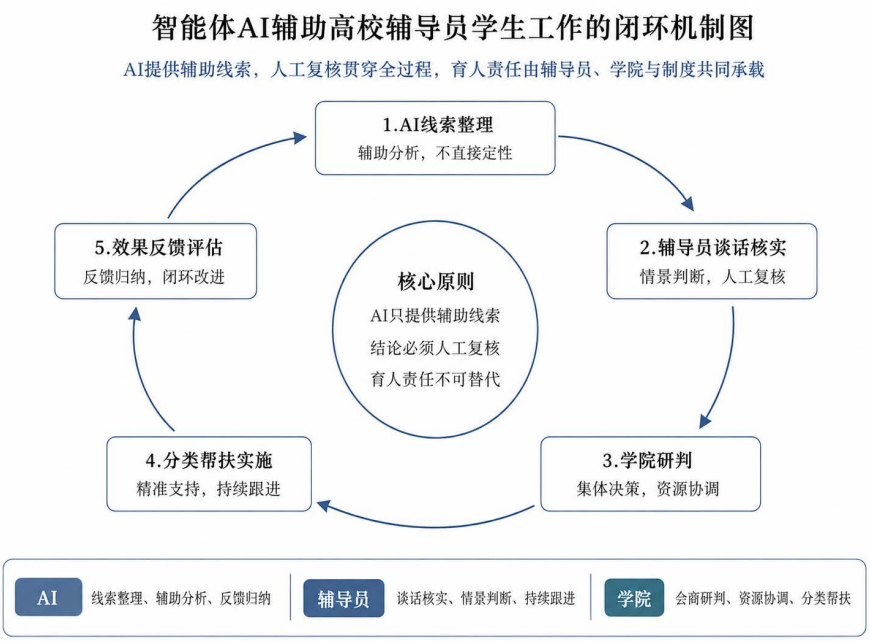

Figure 3. Closed-Loop mechanism for agent-assisted university counseling work
图 3. 智能体 AI 辅助高校辅导员学生工作的闭环机制

闭环机制落地至少需要三类资源：具备权限控制与日志审计能力的校内平台，心理、资助、就业等专业会商力量，以及用于人工复核和申诉处理的时间配置。潜在挑战包括跨部门数据口径不一致、人工复核负担增加、供应商更新引发权限漂移，以及学生对标签化和 AI 介入的信任顾虑。高校宜从公开、低敏、可撤回的任务小范围试点，同时记录错误率、复核工时和申诉数量，达到预设阈值后再扩展应用范围。

6.3. 数据全生命周期管理

学校应优先采用经过安全评估、能够实施权限控制和日志审计的校内环境。数据进入系统前，应区分公开信息、一般个人信息和敏感个人信息，对必要数据去标识化，避免将姓名、联系方式、健康与处分材料直接输入公共模型。任务完成后，应删除临时文件和中间结果，不得长期保存推断性标签。

知识库还应避免“只增不减”。奖助政策、学籍规定和就业信息具有时效性，应标注发布部门、有效期和更新责任人；政策依据无法确认时，应转交人工，而非生成语气确定的答案。

6.4. 审计、异议与责任追溯

对影响学生机会、评价和权益的智能辅助，应保存数据来源、任务指令、工具调用、系统输出、人工修改和最终批准等必要日志，以便争议发生时还原过程。学生应知晓 AI 参与环节，有权更正基础信息、对推荐或分类提出异议并获得人工复核。

责任制度还应覆盖采购与外包，明确数据是否用于训练、服务终止后的删除方式、安全事件报告和系统升级后的权限变化。无法提供技术审计和数据退出机制的产品，不宜处理真实学生资料。

为使责任追溯可执行，可建立“一任务一责任单”，记录任务发起人、数据授权人、系统与模型版本、任务规则、工具调用、人工修改、最终批准人、执行结果、保存期限和申诉处理。业务部门负责定期抽查具体任务，信息化部门保障技术日志完整，学院承担个案决策责任，学校层面则负责制度、采购和供应商责任。发生争议时，应能够沿责任单还原“谁授权、谁配置、谁审核、谁决定、谁处置”的完整链条。

6.5. 提升专业化 AI 素养

辅导员培训不能止于提示词技巧，还应包括识别智能体边界、拆解可审核任务、判断数据必要性、发现自动化偏差、坚守心理与权益场景的专业边界，以及向学生解释技术参与方式。教育部专题培训和 UNESCO 教师 AI 能力框架均将伦理判断、人本理念和专业发展置于核心位置[4] [17]。案例复盘、权限演练和错误情境模拟，可帮助辅导员掌握何时不用 AI、何时中止流程。

学生评价也不能只检查最终文本，还应关注信息来源、修改过程、口头说明和真实实践。学校应明确 AI 使用边界并要求学生对提交内容负责，但不宜将 AI 文本检测结果作为认定违规的单一证据，现有工具仍存在误判和漏判[25]。规范使用的目标是保护学习、表达和选择的主体性。

7. 讨论

智能体化并不意味着系统越自主越先进。学生工作中的高价值应用更适合采用“受限智能体”：读取经过筛选的信息、完成可核验的流程准备，但不能自行扩大目标或跨越权益决定。将其限制在可逆环节，更符合教育场景对信任、程序正当和责任清晰的要求。

从 FAT/ML 视角看，“受限智能体”并非简单降低自动化程度，而是把公平、问责和透明度嵌入流程：通过比较不同学生群体的错误率与资源获得差异识别不公平，以具名批准、日志和申诉落实问责，以说明数据来源、任务规则和 AI 参与程度保障透明。NIST AI 风险管理框架的全生命周期思路进一步表明，治理应随模型、数据和权限变化持续更新，而非一次性验收[13]-[15]。

问答 AI 与智能体 AI 只是分析风险的理想类型，现实系统往往处于两者之间。知识库工具在获得邮件发送、日程写入或业务审批接口后，可能产生行动后果；被称为“智能体”的产品也可能只是多轮对话界面。因此，学校应按实际权限和任务链评估风险，而非依据营销名称。

本文主要依据政策、技术研究和典型任务开展规范分析，尚未比较不同高校的部署效果，也未以访谈或案例追踪检验边界的可执行性。后续可在学业支持、就业帮扶和事务提醒等低至中风险场景中考察错误率、复核成本、学生接受度与申诉情况；心理健康、资助和处分等高风险场景则应优先研究治理流程与伦理审查。

8. 结论

从问答 AI 到智能体 AI，变化的核心是技术与现实行动之间的距离。问答 AI 主要改变内容制作和获取方式，智能体 AI 则可能进入数据读取、任务排序、对象分类、提醒发送和结果跟踪等流程，辅导员也

由答案使用者扩展为任务设计者、数据守门人、过程监督者和最终责任人。

高校推进相关应用应坚持三项原则：按任务风险而非产品名称划定边界；将人工控制前移至数据调用、分类推荐和行动执行之前；涉及学生权益、危机处置和价值判断的事项，必须保留直接沟通、组织程序和具名责任。只有当智能体承担低风险、可逆、可核验的辅助任务，且学生拥有知情、纠错和申诉渠道时，技术效率才能转化为育人支持。

基金项目

四川外国语大学校级科研项目2024年度专项项目“社交媒体中大学生心理问题的心理机制、预警模型及干预方案：自我调节的视角”(项目编号：SISU202427)。

参考文献

[1] 中共中央、国务院印发《教育强国建设规划纲要（2024—2035年）》加快建设中国特色社会主义教育强国[EB/OL]. 2025-01-19. https://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/202501/t20250119_1176166.html, 2026-07-17.

[2] 教育部等九部门关于加快推进教育数字化的意见[EB/OL]. 2025-04-16. https://www.moe.gov.cn/srcsite/A01/s7048/202504/t20250416_1187476.html, 2026-07-17.

[3] 普通高等学校辅导员队伍建设规定: 中华人民共和国教育部令第43号[EB/OL]. 2017-09-21. https://www.moe.gov.cn/srcsite/A02/s5911/moe_621/201709/t20170929_315781.html, 2026-07-17.

[4] 教育部举办全国高校辅导员人工智能专题培训班[EB/OL]. 2025-05-20. https://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/moe_1485/202505/t20250520_1191356.html, 2026-07-17.

[5] 马成瑶, 岳爱武. 高校辅导员智能体建设: 现状、审思与发展[J]. 思想理论教育, 2025(8): 93-99.

[6] 陈哲. 生成式人工智能视域下大学生思想政治教育的建构与调适[J]. 自然辩证法通讯, 2024, 46(12): 95-100.

[7] 冯琳, 倪国良. 基于生成式人工智能的思想政治教育数字化转型[J]. 思想教育研究, 2024(2): 46-53.

[8] 颜佳华, 李睿昊. 网络思政、虚拟思政、数字思政、数据思政、智能思政与智慧思政概念及其关系辨析[J]. 湘潭大学学报(哲学社会科学版), 2024, 48(3): 103-110.

[9] Yao, S., Zhao, J., Yu, D., *et al*. (2023) ReAct: Synergizing Reasoning and Acting in Language Models. *The Eleventh International Conference on Learning Representations*, Kigali, 1-5 May 2023, 1-33.

[10] Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., *et al*. (2024) A Survey on Large Language Model Based Autonomous Agents. *Frontiers of Computer Science*, **18**, Article ID: 186345. https://doi.org/10.1007/s11704-024-40231-1

[11] Park, J.S., O'Brien, J., Cai, C.J., Morris, M.R., Liang, P. and Bernstein, M.S. (2023) Generative Agents: Interactive Simulacra of Human Behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, San Francisco, 29 October-1 November 2023, 1-22. https://doi.org/10.1145/3586183.3606763

[12] Liu, X., Yu, H., Zhang, H., *et al*. (2024) AgentBench: Evaluating LLMs as Agents. *The Twelfth International Conference on Learning Representations*, Vienna, 7-11 May 2024, 1-58.

[13] Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S. and Vertesi, J. (2019) Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Atlanta, 29-31 January 2019, 59-68. https://doi.org/10.1145/3287560.3287598

[14] Kroll, J.A., Huey, J., Barocas, S., *et al*. (2017) Accountable Algorithms. *University of Pennsylvania Law Review*, **165**, 633-705.

[15] National Institute of Standards and Technology (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST. https://doi.org/10.6028/NIST.AI.100-1

[16] UNESCO (2023) Guidance for Generative AI in Education and Research. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

[17] UNESCO (2024) AI Competency Framework for Teachers. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104

[18] Zawacki-Richter, O., Marín, V.I., Bond, M. and Gouverneur, F. (2019) Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators? *International Journal of Educational Technology in Higher Education*, **16**, Article No. 39. https://doi.org/10.1186/s41239-019-0171-0

[19] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., *et al*. (2023) ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences*, **103**, Article ID: 102274. https://doi.org/10.1016/j.lindif.2023.102274

[20] Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., *et al*. (2024) Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review. *British Journal of Educational Technology*, **55**, 90-112. https://doi.org/10.1111/bjet.13370

[21] Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R., *et al*. (2023) What If the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education. *Smart Learning Environments*, **10**, Article No. 15. https://doi.org/10.1186/s40561-023-00237-x

[22] 中华人民共和国个人信息保护法[EB/OL]. 2021-08-20. https://www.cac.gov.cn/2021-08/20/c_1631050028355286.htm, 2026-07-17.

[23] 生成式人工智能服务管理暂行办法[EB/OL]. 2023-07-13. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm, 2026-07-17.

[24] Goddard, K., Roudsari, A. and Wyatt, J.C. (2012) Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators. *Journal of the American Medical Informatics Association*, **19**, 121-127. https://doi.org/10.1136/amiajnl-2011-000089

[25] Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., *et al*. (2023) Testing of Detection Tools for AI-Generated Text. *International Journal for Educational Integrity*, **19**, Article No. 26. https://doi.org/10.1007/s40979-023-00146-z