

生成式人工智能视域下外语教育的风险与应对策略

赵晨薇^{1,2} 付裕淇¹

1.四川文理学院 外国语学院 四川 达州 635000

2.多语种翻译与国际传播研究中心 四川 成都 611844

【摘要】：随着生成式人工智能在教育领域的广泛应用，其对外语教育的影响日益深化。AI在提升教学效率与个性化学习方面展现出巨大潜力，但也引发内容真实性、文化偏见、创造力下降及教育公平性等挑战。本文在梳理相关研究的基础上，系统分析了生成式人工智能在外语教育中的风险机制及教育影响，并提出相应对策，包括构建人机共审机制、优化语料结构与文化平衡、强化创造力导向教学设计及建立公平性算法评估体系。研究旨在为外语教育的智能化转型提供理论支持与实践启示。

【关键词】：生成式人工智能；外语教育；内容真实性；文化偏见；创造力

DOI:10.12417/2705-1358.25.23.006

引言

2025年《教育部等九部门关于加快推进教育数字化的意见（教办〔2025〕3号）》指出，应以教育数字化为突破口，开辟教育发展新赛道，全面支撑教育强国建设。文件强调要“加快建设人工智能教育大模型与构建高质量自主可控的教育多模态语料库”，并要求“将人工智能技术融入教育教学全要素全过程”，以推动课程、教材与教学体系的智能化升级^[1]。在此政策背景下，生成式人工智能迅速融入外语教育场景，被广泛应用于写作辅助、语音评测与个性化学习等领域，为教学创新与学习支持提供了全新机遇。然而，AI技术在赋能教育的同时，也带来了内容真实性与安全性、文化偏见、创造力下降及教育公平性等风险。本文旨在从以上四个方面系统探讨生成式人工智能赋能外语教育的潜在风险机制与应对策略，以期教育数字化转型提供理论支持与实践启示。

1 生成式人工智能赋能外语教育的主要风险

1.1 内容真实性与安全性问题

在生成式人工智能广泛融入外语教育的过程中，内容生成的准确性问题逐渐成为学界与教学实践关注的焦点。虽然语言模型在表达流畅性与上下文衔接方面表现优越，但其生成逻辑主要基于语料的统计概率关系，而非真实世界的知识验证机制。这种“语义拟合优先、事实验证缺位”的模式，导致AI在外语学习任务中容易出现“语言幻觉”现象，即语言表面正确、语法连贯，却存在语义失真或语用偏误。

这一问题在外语学习初期尤其具有隐蔽性与破坏性。一方面，外语初级或中级学习者由于缺乏足够的判断能力，往往会将生成式人工智能生成的错误信息视为可靠输入并加以模仿，从而在词汇搭配、句法结构和语用策略等方面形成系统性偏误。已有研究表明，学生在使用AI辅助写作时，往往直接采纳其生成的表达方式，尽管这些语句在形式上符合目标语规则，但其语义逻辑常存在误导性断裂，削弱了语篇的连贯性与逻辑一致性^[2]。长期依赖“看似正确”的语言输出，容易使学习者在潜移默化中固化错误表达，造成语法结构的过度泛化与交际表达的程式化倾向^[3]。这不仅影响语言输出的真实性与自然度，也可能干扰语言迁移与交际策略的形成。

另一方面，教师在教学过程中亦难以及时识别学生语言偏误的生成源头。由于AI输出具有不可解释性，语言反馈的溯源路径被中断，教师往往难以判断错误究竟源于学生内部语言知识体系的缺陷，还是外部系统输入的偏差。这一现象在作文批改与口语纠错环节尤为明显。已有研究指出，教师普遍低估了AI建议对学生写作策略与表达方式的潜在影响，容易在教学中错失关键纠错节点，从而导致语言学习中的偏误累积^[4]。

1.2 刻板印象与文化偏见

生成式人工智能在外语教育中的应用虽为教学创新提供了契机，但其潜藏的文化偏见与刻板印象问题也日益受到关注。其根源在于主流语言模型在训练阶段普遍采用开放网络文本作为预训练语料，而这些语料在地域分布、意识形态、性别角色建构等方面往往存在显著的不均衡与主流文化导向倾向。

作者简介：赵晨薇（1993-），女，河南洛阳人，博士，副研究员，研究方向：教育心理与外语教学。

作者简介：付裕淇（1992-），女，河北石家庄人，博士，讲师，研究方向：外语教学。

基金项目：本文系四川省高等学校人文社科重点研究基地多语种翻译与国际传播研究中心资助课题“生成式人工智能赋能多语种教学的认知支持机制研究”（项目编号：RCMTIC2511）研究成果。

这种偏差一旦被嵌入模型参数,便可能在生成过程中表现为价值判断的倾斜、文化形象的扁平化及交际范式的单一化,对外语学习者的文化认知与语言使用产生潜移默化的影响。

从外语教学的视角来看,文化偏见不仅仅是语言风格差异,更深层地影响学习者对语言背后语境、身份与文化意涵的理解与运用。当AI生成系统在例句、翻译或写作建议中强化特定性别、职业、族裔或国家形象时,学习者可能在语言编码与语义映射中形成刻板化角色图式,削弱其在多元语境下的语言迁移与表达灵活性。已有研究表明,在AI辅助写作训练中,学生倾向于复用系统生成的模板句,而这些模板大多源于欧美主流文化语境,导致非英语母语学习者在跨文化交际中出现语用错配与策略干扰^[5]。这种“隐性文化输入”进一步固化了学习者的语言思维模式,使其难以发展批判性语言意识与跨文化理解能力。此外,AI生成的语料在语用层面也存在区域性局限。如AI口语平台生成的对话集中于“标准美式”语用模式,忽视了不同文化语境中的礼貌层级、交际目的与语域差异,导致部分学习者在真实交际中出现语气不当、表达过于直接或语用层级错位等现象,从而影响语言得体性与交际效果^[6]。

造成此类文化偏差输出的根本原因在于,现有AI系统普遍缺乏“文化多样性建模能力”与“语用自适应机制”。主流模型的训练体系主要依赖北美、英国等地区的媒体与出版物,其语料在价值取向、文化符号与语言风格上高度单一,缺乏对非西方文化、少数群体语言实践及真实跨文化交际情境的覆盖。同时,外语学习的核心目标不仅是实现语法正确与词汇丰富,更在于培养学习者具备“语用适切—语义可控—文化得体”的综合表达能力。然而,当AI输出长期受制于单一文化范式时,学习者的文化敏感性与语用灵活性将受到削弱,其跨文化理解与多元表达能力难以充分发展。这种隐性的文化单向性与语用偏差,实质上限制了生成式人工智能在促进文化理解、增强语用意识及提升跨文化交际能力方面的教育价值。

1.3 创造力与多样性下降

生成式人工智能在提供语言生成辅助的同时,其内在的生成逻辑也可能对学习者的语言表达的创造力与个体风格建构造成潜在抑制。当前主流语言模型多采用基于概率的最优词序排列机制,即倾向输出语料中出现频率最高、共现概率最大的表达方式。这种“趋于平均”的生成策略虽然能提升文本的连贯性与语法正确性,却往往以牺牲语言多样性与表达个性为代价,导致语义构造趋同、风格特征弱化。

在外语教育场景中,学习者通常处于从模仿性输入向自主性表达过渡的关键阶段。若学生在写作与口语表达中过度依赖AI辅助,其语言产出将逐渐模板化、标准化,缺乏在句式、语用策略与语篇组织上的主动探索,从而削弱语言创新能力与表

达独创性。研究显示,在AI辅助写作中,尽管学生的语法准确度有所提高,但其“语言风格丰富性”与“表达独创性”得分显著低于未使用AI的对照组^[7]。学习者往往直接采纳系统推荐句式,并依赖AI生成的逻辑框架进行段落重组,而非通过独立推理实现语篇构建。类似现象也出现在AI语音交互任务中。相关研究发现,学生在多轮对话中倾向反复使用AI生成的高频句式应对不同语境,导致语言表达程式化,缺乏灵活调整与语用适应性^[8]。这种“策略固化”现象在长期学习中不仅限制了表达自由,也削弱了语用创造力与交际灵活性。

此外,AI辅助平台往往以“高分模板与高效句型”为导向,鼓励学习者在有限时间内追求输出效率而忽视表达质量。这种“结果导向偏移”强化了语言产出的标准化,使学习者逐渐忽略表达的多样性与语义创新空间。从教学法视角看,任务型教学法强调以真实交际任务为驱动,促使学习者在语境、目的与听众间进行动态的语言选择与策略建构。然而,当AI过度介入教学过程,承担信息整合与语篇规划职能时,学生的语言决策权被削弱,其认知参与思维激活程度显著下降。这种现象在开放式写作与讨论型口语任务中尤为突出,导致语言创造力与批判性思维能力逐步弱化。

1.4 智能鸿沟与教育公平性

随着生成式人工智能在外语教育中的深入应用,技术可达性与资源可用性的差异正逐渐演化为新的教育壁垒。智能系统在赋能教学的同时,也引发了“智能鸿沟”问题,即技术的获取、使用与收益能力在不同地区和群体间分布不均。这种鸿沟不仅体现在硬件设备的拥有差异上,更延伸至教学平台接入机会、AI模型定制程度、教师培训质量以及系统评估算法对不同学习背景的适应能力等多个层面,最终导致学习成效、评估结果与发展路径上的结构性不平等。

在资源接入方面,发达地区学校更容易获得具备大模型调用能力的智能平台,并拥有能够理解与运用AI技术的教师队伍,从而实现教学的智能化融合。相较之下,经济欠发达或偏远地区受限于网络条件、设备配置与经费支持,往往只能使用功能受限的本地模型或公开API版本,导致教学体验与反馈精准度显著下降。调查显示,约半数来自三线城市的高中教师因设备不足难以稳定开展AI辅助教学,学生亦反映平台任务推荐与实际水平存在明显错配^[9]。

在评估机制层面,当前多数AI语言评估模型的训练语料主要来源于主流母语群体、标准口音语音及中高水平语言输入。这种语料结构的单一性导致系统在处理带有区域性口音、方言干扰或非典型表达策略的学习者时,常出现误判、偏分或反馈失效等问题,尤其在口语评分与写作生成任务中表现突出。研究表明,AI写作评分系统对中低水平学习者的语篇结构

容忍度显著低于人工评分,促使学生为“迎合系统”而过度调整语言策略,从而限制其真实表达与思维发展的空间^[10]。这种算法性偏差在无形中扩大了不同语言基础学习者之间的发展差距,削弱了教学评估的公平性与教育指导功能。

2 应对策略与优化路径

2.1 内容真实性与数据安全保障

针对生成式人工智能在外语教学内容真实性方面的潜在风险,应建立以“人机共审”为核心的内容质量保障体系。在教学平台中引入“双通道反馈机制”,即AI生成内容后同步开放教师复核与标注通道,并在学生端设置“人工审查通过”标识,以确保输出内容的准确性与教学权威性。同时,通过知识图谱与数据库比对构建“事实追踪模块”,自动验证生成语句中的事件与数据,并提示潜在偏误;在教学环节中引导学生开展“可信度标注训练”,对AI生成内容进行自我评估与标注,以提升其语言辨识力与元认知水平。除内容真实性外,数据安全亦是关键议题。学习过程中生成的个性化数据若被用于模型再训练,可能引发隐私泄露与算法歧视。为此,应建立“知情授权—用途限定—可撤回访问”机制,在数据采集阶段明确使用范围与目的,遵循“最小调用原则”,并提供“一键删除”与数据撤回功能。教育机构可通过“本地部署模型+私有数据池”构建封闭系统,从源头上降低数据风险,保障AI教育的安全与伦理可控。

2.2 文化多样性与语料平衡机制

为缓解生成式人工智能输出中的文化偏差,应从语料、平台与教学三方面入手。在模型训练中构建多语料融合体系,系统引入不同地域与文化类型的教育语料,防止单一文化主导。在平台设计上增设“文化语境提示”与“表达多样性切换”功能,帮助学习者理解表达的文化来源与语用差异。在教学环节中,教师可设置“文化偏差辨析任务”,引导学生识别并改写AI输出中的刻板表达,培养跨文化批判意识与语用敏感性。同时,可开发“多文化角色模拟模块”,通过情境互动提升学生的文化理解与表达灵活性,降低文化单一化带来的隐性偏差风险。

参考文献:

- [1] 教育部等九部门.关于加快推进教育数字化的意见(教办[2025]3号)[EB/OL]. 教育部, 2025-04-15.
- [2] Wang Y, Xie Q. Diagnosing EFL undergraduates' discourse competence in academic writing[J]. *Assessing Writing*, 2022, 53. 100641.
- [3] Chen C, Gong Y. The role of AI-assisted learning in academic writing: A mixed-methods study on Chinese as a second language students[J]. *Education Sciences*, 2025, 15(2): 141.
- [4] Han J, Li M. Exploring ChatGPT-supported teacher feedback in the EFL context[J]. *System*, 2024, 126. 103502.

2.3 创造力导向的教学设计

为防止生成式人工智能在外语学习中引发表达趋同与创造力弱化,应从教学流程、平台功能与评价体系三方面入手。教学中可采用“AI辅助—学习者再创作”模式,鼓励学生在AI生成初稿基础上进行改写与风格化表达,以强化语言掌控力与创造意识。在平台设计上引入“创造力引导模式”,提供多种语体与风格版本供学生比较选择,促进语言创新与表达灵活性。在评价体系中,将“语言多样性”和“创造性使用”纳入评分标准,弱化对语法准确性的单一导向,鼓励学生发展个性化与创新性表达,构建以创造性学习为核心的外语教学生态。

2.4 教育公平性与教师赋能

为缩小生成式人工智能应用中的“智能鸿沟”并促进教育公平,应从“基础设施—平台—算法—教师”四个方面协同推进。基础设施层面,政府与企业应支持边远地区部署轻量化AI教学系统,开发可离线运行的学习终端,保障技术普惠可达。平台层面,构建“分层自适应任务库”,依据学习者水平与设备条件动态调整任务难度与反馈机制,避免学习错配。算法层面,扩充“多口音、多水平”样本,引入包容性评估模型并设立“人工复核与申诉机制”,防止算法偏差。教师层面,常态化实施“AI公平性培训”,提升教师识别偏差与引导学生理性使用AI的能力,实现教育数字化的公平与包容。

3 结论

生成式人工智能的迅速发展,为外语教育的智能化与个性化转型提供了前所未有的机遇。然而,技术进步的同时也伴随内容真实性、文化偏见、创造力下降与教育公平性等多重挑战。如何在创新与规范之间取得平衡,已成为教育数字化进程中的核心议题。未来的外语教育应以“人机协同、伦理优先、包容共享”为原则,通过建立内容审核机制、强化算法治理、完善教师培训与学习者素养培养,实现生成式人工智能的可持续应用。唯有在制度保障、技术优化与教育理念三者的协同作用下,才能真正实现智能技术赋能教育的初心,促进语言学习的公平性、创造性与人本发展。

- [5] Wang C. Exploring students' generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers[J]. *Technology, Knowledge and Learning*, 2024: 1–22.
- [6] Erdogan N, Christina K. Integrating AI in language learning: Boosting pragmatic competence for young English learners[J]. *LatIA*, 2025(3): 115.
- [7] Lee G. AI-based writing feedback in ESP: Leveraging ChatGPT for writing business cover letters among Korean university students[C]//*Proceedings of the International CALL Research Conference*. 2024: 131–136.
- [8] Han J, Yoo H, Myung J, Kim M, Lee T, Ahn S, Oh A. RECIPE4U: Student-ChatGPT interaction dataset in EFL writing education[J]. *ArXiv*, 2024..
- [9] Huang M. Student engagement and speaking performance in AI-assisted learning environments: A mixed-methods study from Chinese middle schools[J]. *Education and Information Technologies*, 2024: 1–23.
- [10] Zhang M, Johnson M, Ruan C. Investigating sampling impacts on an LLM-based AI scoring approach: Prediction accuracy and fairness[J]. *Journal of Measurement and Evaluation in Education and Psychology*, 2024, 15(Special Issue): 348–360.