

# 国内外人工智能工具在中文编校中 应用效果评价与建议\*

陈玮<sup>1)</sup> 叶飞<sup>2)†</sup>

1)南开大学《南开大学学报(自然科学版)》编辑部,300071,天津;

2)农业农村部环境保护科研监测所《农业环境科学学报》编辑部,300191,天津

**摘要** 编辑校对工作是编辑人员的核心工作之一,费时费力,而已有的智能编校工具效果不够理想。本研究通过在单句和段落中构建错例库检验人工智能工具的编校能力,对国内外6大知名人工智能工具:智谱 AI(ChatGLM)、通义千问(QianWen)、讯飞星火(iFLYTEK Spark)、文心一言(ERNIE Bot)、360智脑、ChatGPT进行了横向评价。评价结果表明:人工智能工具能够在逻辑、语言等方面辅助编校工作,但无法检查政治错误和参考文献错误,在数据检查方面能力各不相同;总体上人工智能工具在段落中的编校效果优于单句,国内人工智能工具中文编校效果优于 ChatGPT,尤其是通义千问和讯飞星火展现了优秀的编校能力。根据评价结果,本研究提出了编校工作中应用人工智能工具的建议。

**关键词** 人工智能;编校;ChatGPT;智谱 AI;通义千问;讯飞星火;文心一言;360智脑

**Evaluation and suggestions on the application effects of artificial intelligence tools in Chinese editing and proofreading work**//CHEN Wei, YE Fei

**Abstract** The editing and proofreading work is one of the core tasks for editors, requiring significant time and effort. However, existing intelligent editing tools often fall short of ideal effectiveness. This study examines the editing and proofreading capabilities of artificial intelligence (AI) tools by constructing a database of error cases both in a single sentence and in a paragraph. Six widely recognized AI tools both domestically and internationally are horizontally evaluated, including ChatGLM, QianWen, iFLYTEK Spark, ERNIE Bot, 360 Zhinao, and ChatGPT. The evaluation results indicate that AI tools can assist in editing and proofreading work, particularly in areas such as logic and language. However, they are unable to identify political errors and errors in referencing literature, and their capabilities vary in terms of data checking. Overall, most AI tools do better in a paragraph than in a single sentence, the domestic AI tools are better than the ChatGPT. Notably, QianWen and iFLYTEK Spark exhibit outstanding editing capabilities. Based on the evaluation results, this study proposes recommendations for the application of AI tools in editing and proofreading work.

**Keywords** artificial intelligence; editing and proofreading; ChatGPT; ChatGLM; QianWen; iFLYTEK Spark; ERNIE Bot;

360 Zhinao

**First-author's address** Editorial Department of Journal of Nankai University (Natural of Science), Nankai University, 300071, Tianjin, China

**DOI:**10.16811/j.cnki.1001-4314.2024.03.017

编辑校对(编校)是编辑的核心工作之一,一般包含编辑加工、三校一读等多个流程,非常烦琐而且耗时。国家很重视出版物的编校质量,《图书质量管理规定》《报纸期刊质量管理规定》分别对图书、期刊、报纸的编校质量提出了基本要求,其差错率分别不得超过万分之一、万分之二、万分之三,否则认定编校质量不合格。因此,广大编辑人员需要投入大量的时间和精力进行编校工作,董文杰等<sup>[1]</sup>通过调查问卷得出,80.00%的编辑在编校、排版等环节投入较多时间。

国内一些公司针对各出版单位编校工作任务重、要求高的难题,开发了一些智能编校工具,例如北大方正电子有限公司开发的智能辅助编校系统,北京黑马飞腾科技有限公司研发的黑马校对软件,善锋软件开发的参考文献自动校对系统,等等。这些工具在提升编校质量和效率方面发挥了作用,但是这些工具都是收费工具,而且存在算法、词库更新机制、产品架构等方面的局限性<sup>[2]</sup>,在应用过程中也出现了遗漏错误、应用范围有限等问题<sup>[3-4]</sup>。

自2022年底开始,人工智能技术取得了突飞猛进的发展,最显著的标志是2022年11月30日,OpenAI发布名为ChatGPT的模型研究预览版,由于其让人惊叹的性能迅速火爆,5天注册用户超100万,强烈预示着新一代人工智能时代的到来。人工智能技术在自然语言处理方面展现了强大的能力,因此有编辑尝试利用新一代人工智能工具辅助编校。例如,崔玉洁<sup>[5]</sup>和李侗桐等<sup>[6]</sup>利用ChatGPT对论文摘要进行润色,刘俏亮等<sup>[7]</sup>利用基于以ChatGPT为后台的工具Notion AI对论文全文进行润色,等等。可以看出已有的少量研究都是利用ChatGPT这一工具,而且测试的样本量较小,只研究了人工智能工具在语言润色方面的能力,而忽略了编校中更为重要的纠错能力,从结果看ChatG-

\* 中国高校科技期刊研究会专项基金课题(CUJS2023-C01)

† 通信作者

PT能够提高编辑的工作效率,但也存在提供错误信息、错用专业术语等缺陷。由于ChatGPT对注册用户的区域限制,我国大陆用户在利用ChatGPT方面存在很多困难,这极大地限制了ChatGPT的应用。因此,虽然人工智能技术突飞猛进,但我国广大编辑人员对于如何在编辑工作中方便地利用人工智能工具、利用何种人工智能工具存在较大的困惑。

已有的研究忽略了我国在人工智能方面的重大进展。在人工智能的浪潮下,我国科技巨头纷纷跟进,2023年,阿里巴巴、科大讯飞、百度等公司都开发出了人工智能工具,免费向公众开放。这些国产人工智能工具能否应用于中文编校工作中?相对于ChatGPT效果如何?相关研究未见报道。本研究归纳总结了编校工作中常见的错误,构建了包含政治错误、逻辑错误等7种类型的错例库,对国内外的6种主流的人工智能

工具进行了横向评价,根据评价结果提出了中文编校工作中人工智能工具应用建议。本研究结果有望指导编辑人员科学地利用人工智能工具辅助编校工作,提高编校工作效率和质量。

## 1 材料与方法

### 1.1 人工智能工具

我国人工智能技术高速发展,涌现了一批先进的人工智能大模型,为本研究的开展提供了丰富的工具。为全面比较人工智能工具在中文稿件编辑加工中的应用效果,选择目前比较知名的、被广泛认可且已经对公众免费开放的国内外人工智能工具。国内选择智谱AI、通义千问、讯飞星火、文心一言、360智脑等5款人工智能工具,国外选择ChatGPT,这些工具的详情见表1。分别登陆这些人工智能工具的官网进行注册、登录。

表1 研究所用人工智能工具详情

序号	工具名称	开发公司	网址	上线时间	是否免费
1	智谱AI	北京智谱华章	<a href="https://www.zhipuai.cn">https://www.zhipuai.cn</a>	2023年4月	是
2	通义千问	阿里巴巴	<a href="https://qianwen.aliyun.com">https://qianwen.aliyun.com</a>	2023年4月	是
3	讯飞星火	科大讯飞	<a href="https://xinghuo.xfyun.cn">https://xinghuo.xfyun.cn</a>	2023年5月	是
4	文心一言	百度	<a href="https://yiyan.baidu.com">https://yiyan.baidu.com</a>	2023年8月	3.5版免费,4.0版收费
5	360智脑	360	<a href="https://ai.360.com">https://ai.360.com</a>	2023年9月	是
6	ChatGPT	OpenAI(美国)	<a href="https://chat.openai.com">https://chat.openai.com</a>	2022年11月	3.5版免费,4.0版收费

### 1.2 错例库构建

为评价各个人工智能工具在编辑加工中的应用效果,本研究构建了丰富的错例库。错例库主要有4个来源:1)笔者多年编辑加工经验积攒的典型案列,并咨询了编辑同行;2)微信公众号“编辑课堂”(微信号:sanrenxing-peixun)中整理的各大报纸、期刊等错误实例;3)语文文字刊《咬文嚼字》每年发布的《年度十大语文差错》;4)已经发表的相关错例总结。<sup>[8-14]</sup>为了比较分析,将同一错例放置在单句和一个完整的段落中,待测。

为深入探讨人工智能工具对不同类型错例的纠错能力,参考《报纸期刊质量管理规定》中的分类原则,本研究将期刊编辑加工中常见的错误分为政治错误、逻辑错误、语言错误、名称错误、科技术语错误、数据错误、参考文献错误7大类。其中:政治错误是指文中出现的国别、领导人姓名、地区称号等错误,尤其是中文期刊中重点关注的涉港澳台问题;逻辑错误是指语句中的前后逻辑不通、指代错误等问题;语言性错误是指用词不当、动名词搭配不当等问题;对于名称错误本研究重点关注了人名错误、地名错误、机构名错误、国家标准错误;对于科技术语错误本研究重点关注了化学品名称错误、农业学科和数学学科等的科技术语错误;

数据错误是指文中出现的数值关系计算错误,本研究重点关注了倍数关系、趋势判断、常识性数据错误等几方面;对于参考文献著录错误本研究仅关注了作者、论文题目、期刊名称、出版年等项。错例库的详情见表2。

为保证不同人工智能工具的可比性,采用相同内容和相同数量的提示词,分别在6个人工智能工具中测试,如果人工智能正确指出错误之处,则得1分,否则得0分,分别统计各人工智能工具的总分和各错误类型中的得分,用以评估各人工智能工具在中文稿件编辑加工中的应用效果,得分越高则说明纠错能力越强,效果越好。

## 2 结果与分析

### 2.1 人工智能工具横向评价

分别统计各人工智能工具针对错例库的得分情况,见图1。

由图1可见,除了文心一言和智谱AI外,针对同一错误,其他人工智能工具均在段落中的纠错能力略强于单句,这说明人工智能工具更适用于段落的编校,这为人工智能工具辅助编辑工作提供了更多便利。其中,讯飞星火提升显著,在单句的测试中仅得27分,在段落的测试中得到42分,其他人工智能工具的差异均

表2 错例库详情

序号	问题类型	数量	错例示例(下划线为错误之处)
1	政治错误	10	1)中国、美国、德国、日本、 <u>台湾</u> 等国家。2)国立 <u>台湾</u> 大学植物科学研究所。3)香港特别行政区政府行政长官。4)所谓历史的原因,指的是老牌殖民者大英帝国武力战败了腐朽的清政府,以不平等条约强行霸占香港为 <u>殖民地</u> 。
2	逻辑错误	10	1)张宁能不能拿下这一局,是中国队战胜印尼队的关键。2)在本届世界杯决赛上,法国队将与意大利队争夺 <u>冠军</u> 。3)那年12月我应东京大学已故教授大泽吉博先生的邀请……
3	语言错误	10	1)节日之夜热闹非凡,到处都是灯火阑珊。2)作为一名普通的莘莘学子,他的创新精神让人敬佩。3)近两年,我市大力发展第三产业,吸收大量职工再就业,使一部分家庭的 <u>贫困</u> 和生活水平有较大幅度 <u>提高</u> 。
4	名称错误	10	1)三藩之乱, <u>吴三桂</u> 最终被平定。2)我在北京坐地铁,到 <u>崇文门</u> 站换乘5号线。3)她毕业于位于天津市的 <u>南开</u> 大学。4)根据国家标准《农业用腐殖酸钾》 <u>GB/T 33811-2017</u>
5	科技术语错误	10	1)实验用 <u>氢氧化那</u> 。2)水稻的生长过程中新的 <u>分蘖</u> 不断产生。3) <u>座果率</u> 是指果树生产中生理落果结束时树上的挂果数占开花时花的数目的比例。4)随机 <u>微分</u> 方程。
6	数据错误	10	1)A的含量为10mg,B的含量为20mg,B相对于A含量提高了 <u>50%</u> 。2)试验地沙质土壤,pH 7.96,含有机质 <u>230</u> mg/g,全氮 <u>21</u> mg/g,全磷 <u>17.2</u> mg/g,全钾 <u>21.8</u> mg/g。3)2020、2021、2022年的产量分别是100kg、120kg、150kg,产量总体呈 <u>下降</u> 趋势。4)试验于2017年5月15日水稻秧苗移栽开始,至2018年4月20日截止,共 <u>320</u> 天。
7	参考文献错误	10	错误的作者、论文题目、期刊名、出版年,等等。

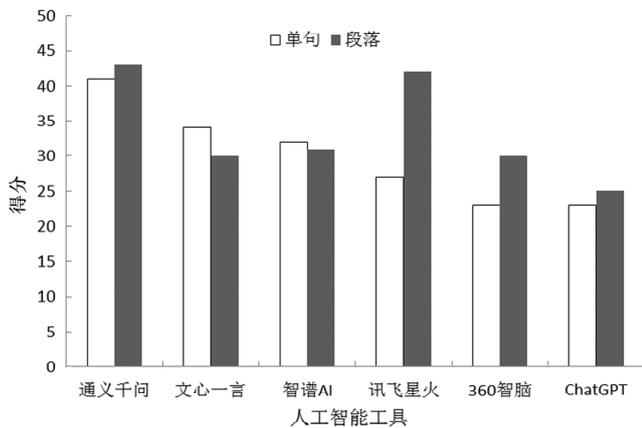


图1 各人工智能工具得分情况

较小。本次测试满分70分,人工智能工具得分最高的43分,最低的23分,这总体说明各个人工智能工具在中文稿件的编辑加工中能起到一定的辅助作用,但是

还不够完善。从各个人工智能工具在段落编校中得分排名看,大体分为3个梯队,通义千问和讯飞星火得分超过40,处于第一梯队,文心一言、智谱AI、360智脑得分分别为30、31、30分,处于第二梯队;ChatGPT得分25分,处于第三梯队。

从图1还可以看出,我国人工智能工具在中文稿件的编辑加工中效果都优于大名鼎鼎的ChatGPT,说明国产的人工智能工具更适用于中文信息处理。

2.2 人工智能工具对不同错误类型的处理能力评价

统计人工智能工具对不同错误类型的得分情况见图2。

由图2可见,总体上看,人工智能工具在段落中针对语言错误、逻辑错误的纠错能力强于单句中,其他类型的错误差异较小。这是因为较长的段落为人工智能工具提供了更多的信息和背景,因此对于依赖语境的

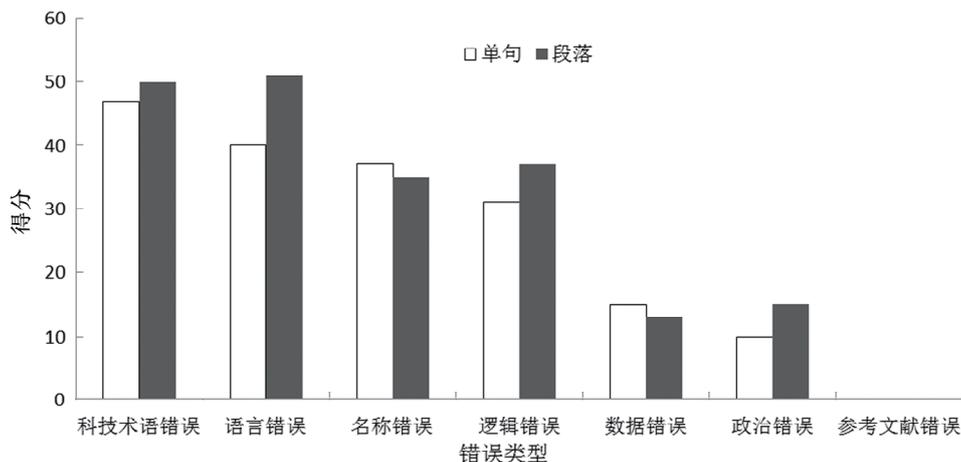


图2 人工智能工具对不同错误类型的处理情况

语言错误和逻辑错误有更好的识别能力,反之科技术语错误、名称错误、数据错误等较独立的错误受语料长度的影响很小。

从图2还可以看出,人工智能工具对不同的错误类型纠错能力差异很大,以段落数据为例,对语言错误、科技术语错误、逻辑错误、名称错误的纠错能力较强,得分分别为51、50、37、35分,对政治错误和数据错误的纠错能力较弱,得分分别为15、13分,对参考文献错误没有纠错能力。

出现差异较大现象的原因在于人工智能工具的学习资料来源于网络公开内容,因此对于科技术语、语言、名称等公开的、事实性的资料人工智能工具纠错能力较强,而参考文献信息属于非公开资料,需要购买专业数据库才能获取,因此人工智能工具对此类非公开的资料缺乏纠错能力。逻辑错误、数据错误的纠错需要对上下文有准确的理解,尤其是对汉语的表达习惯需要有较深的理解,这点人工智能工具尚欠缺。

### 2.3 详细评价

以段落数据为例,统计人工智能工具在不同错误类型得分的具体情况见表3。

表3 人工智能工具在各错误类型中的具体得分

问题类型	智谱 AI	通义千问	讯飞星火	文心一言	360 智脑	ChatGPT	平均分
政治错误	2	5	3	3	1	1	2.50
逻辑错误	6	8	8	5	6	4	6.17
语言错误	9	10	10	7	8	7	8.50
名称错误	6	7	8	6	7	1	5.83
科技术语错误	8	9	8	8	8	9	8.33
数据错误	0	4	5	1	0	3	2.17
参考文献错误	0	0	0	0	0	0	0

由表3可见,人工智能工具各有特点,对不同错误类型的纠错能力不同。在政治错误方面,人工智能工具整体表现较差,平均分只有2.50分,其中通义千问好于其他人工智能工具;在逻辑错误方面,通义千问和讯飞星火表现较好,均得到8分,其他人工智能工具较差;在语言错误方面,通义千问和讯飞星火表现很好,得到了所有错误类型中唯一的满分,智谱AI和360智脑得分也较高;名称错误中各人工智能工具的表现情况和语言错误很类似;科技术语错误方面,各人工智能工具表现均较好,平均分为8.33分,其中ChatGPT和通义千问得分最高;数据错误方面,通义千问和讯飞星火表现较好,智谱AI、文心一言、360智脑表现较差,得分几乎为0;各人工智能工具在参考文献错误方面的得分都为0。

总体来看,除了参考文献错误,通义千问和讯飞星火在各种错误类型纠错方面表现均衡,实力较强,尤其

擅长语言错误、名称错误、逻辑错误的纠错;文心一言也比较均衡,在语言错误、科技术语错误、逻辑错误方面表现较好,但是不擅长数据错误;ChatGPT在科技术语错误、语言错误、逻辑错误方面的纠错能力较强,在其他方面表现平平;实践中发现360智脑对于国家标准名称、编号错误方面的纠错能力较强。

## 3 启示与建议

### 3.1 国产人工智能工具能够在语言等方面辅助编校工作

本研究经过测试,发现国产人工智能工具在科技术语错误、语言错误、名称错误等方面的纠错能力较强,可以辅助编辑人员进行编校工作。而且人工智能模型时刻抓取网上公开的资料进行学习,相对于传统的利用词库进行纠错的编校辅助工具更加开放、适用范围更广。本研究所用的人工智能工具都是免费向公众开放的,注册方便,而且国产的5个人工智能工具都开发了App,可以在手机上随时调用,使用更加方便。

本研究发现,国产人工智能工具在语言错误、名称错误、逻辑错误方面的纠错能力要强于ChatGPT,显示国产人工智能工具的中文理解能力要强于ChatGPT,例如,对于错例库中历史人物错误(吴三桂错写为吴三佳,李白错写为杜甫),ChatGPT都无法识别,且对于逻辑错误这种需要一定的中文功底的错误,ChatGPT表现很差。可能的原因是相对于ChatGPT,国产智能工具的中文训练数据集更广泛,对中文的表达习惯积累的时间更长。因此,国产人工智能工具在中文编校中更有优势,有广泛的应用前景,能够辅助编辑人员提高编校效率和质量。

### 3.2 人工智能工具不能取代人工,编辑人员应着重加强政治和数据敏感性

从研究结果看国内外人工智能工具在政治错误方面的纠错能力较低,在测试时多次出现“系统检测到输入或生成内容可能包含不安全或敏感内容,请您避免输入易产生敏感内容的提示语”的提示,因此人工智能工具不适合于审查政治错误,编辑人员,尤其是社会科学方面的编辑人员应该增强政治敏锐性,加强学习,多查多看,谨慎处理稿件中涉港澳台、民族、历史等相关内容。

本研究设计的数据错误错例有难度区别,包含一些简单的倍数关系计算,也包含一些难度较大的问题,例如科技数据常识错误、规律总结、日期计算等。研究发现国内外人工智能工具对于简单的错例纠错能力较强,而对于难度大的问题普遍纠错能力低,这要求编辑人员应加强数据敏感性。例如,稿件中出现“试验地

沙质土壤, pH7.96, 含有机质 230 mg/g, 全氮 21 mg/g, 全磷 17.2 mg/g, 全钾 21.8 mg/g”的错误表达<sup>[13]</sup><sup>173</sup>, 这些数值的小数点都右移了1位, 扩大为原来的10倍了。发现这个错误要求编辑人员具有一定的专业基础知识和经验积累, 但人工智能工具显然还达不到这个水平。这提示我们面对人工智能的挑战, 编辑人员应着力加强学术判断能力, 深化对科技期刊出版流程的把控、编辑活动规律的认识、出版伦理和文化价值的坚守, 重塑编辑价值<sup>[15]</sup>。

### 3.3 人工智能工具应用建议

本研究根据评价结果提出如下5条建议:

1) 人工智能工具在中文编校方面展现了较强的能力, 因此建议广大编辑人员在编校工作中使用人工智能工具, 尤其是国产人工智能工具;

2) 各人工智能工具的提问均有字数限制, 无法将整篇论文交给人工智能工具编校, 但是人工智能工具对段落的纠错能力强于单句, 因此建议在使用中以段落为处理单元;

3) 从评价结果来看, 总体上通义千问和讯飞星火最优, 而且这2个人工智能工具比较全能, 因此编辑人员在编校工作中可以考虑优先使用这2个工具;

4) 人工智能工具各具特色, 可以结合使用;

5) 人工智能工具无法识别政治错误和参考文献错误, 不建议使用人工智能工具审查政治问题, 可以使用善锋软件等工具辅助编校参考文献。

## 4 结束语

本研究面对我国人工智能技术蓬勃发展的现状, 针对编辑人员编校工作量大和无法使用人工智能工具辅助编校的困惑, 通过构建错例库检验人工智能工具的纠错能力, 对国内外6大知名人工智能工具进行了横向评价。结果表明: 人工智能工具能够在逻辑、语言等方面辅助编校工作, 但无法检查政治错误和参考文献错误, 在数据检查方面能力各不相同; 总体上国内人工智能工具中文编校效果优于 ChatGPT, 尤其是通义千问和讯飞星火展现了较优秀的编校能力。根据评价结果, 本研究提出了编校工作中应用人工智能工具的建议, 这些建议对于提高编校工作效率和质量有重要

参考价值。

本研究的不足体现在错例库的学科构成较窄, 缺乏医学、经济学等一些重要学科的案例, 还需进一步完善, 另外本研究结果只适用于中文内容, 未涉及英文内容。这些需要在下一步的研究中深化。

## 5 参考文献

- [1] 董文杰, 李苑. 人工智能在科技期刊中的应用及启示[J]. 中国科技期刊研究, 2023, 34(11): 1401
- [2] 张晨. 智能编校工具辅助传统编务工作的探索与实践[J]. 科技传播, 2020, 12(20): 163
- [3] 张渊. 黑马校对软件应用刍议[J]. 出版广角, 2018(6): 58
- [4] 王静. 浅谈编辑校对工作中黑马校对软件的应用[J]. 电脑知识与技术, 2020, 16(28): 239
- [5] 崔玉洁. ChatGPT与人工编校相结合: 提高期刊编校效率和文章质量[J]. 编辑学报, 2023, 35(4): 429
- [6] 李桐桐, 高瑞婧, 田佳. ChatGPT在中文科技期刊摘要文字编辑中的实用性测试与分析[J]. 中国科技期刊研究, 2023, 34(8): 1014
- [7] 刘俏亮, 张洁, 刘东亮. 应用Notion AI辅助编校中文科技期刊论文[J]. 编辑学报, 2023, 35(5): 550
- [8] 常思敏, 孙玮. 农科论文中科技术语使用问题例析[J]. 编辑学报, 2002, 14(3): 180
- [9] 郝红梅, 胡成志, 丁丹丹. 农业科技期刊论文编校常见差错分析[J]. 农业图书情报学刊, 2011, 23(12): 180
- [10] 郝红梅, 胡成志, 丁丹丹, 等. 农业科技论文中常见错用字词辨析[J]. 农业图书情报学刊, 2011, 23(10): 179
- [11] 焦培娟, 赵爱群, 鞠善宏, 等. 农业科技论文中应避免名词术语的错用[J]. 编辑学报, 2007, 19(3): 185
- [12] 王贵州, 马永祥, 邝文国. 农业科技论文中常见误用专业术语辨析[J]. 中国科技期刊研究, 2010, 21(6): 880
- [13] 周林杰, 邝文国. 农业科技论文中常见数据错误的剖析[J]. 农业图书情报学刊, 2006(12): 172
- [14] 廖薇薇, 王明丰, 王维朗, 等. 医学学术论文中常见的政治性差错与防范策略[J]. 编辑学报, 2023, 35(2): 120
- [15] 王亚辉, 王晶. 人工智能之于科技期刊出版业态的变革及启示[J]. 中国传媒科技, 2023(1): 55

(2023-12-12收稿; 2024-03-13修回)