单位代码 10006

****学 号 SY2303526



**智能生成大模型对于学术不端的影响**

|  |  |
| --- | --- |
| 院（系）名称 | 自动化科学及电气工程学院 |
| 学号 | SY2303526 |
| 学生姓名 | 杨和鹭 |
| 指导教师 | 石存 |

2023年11月

目录

[1. 文献综述 4](#_Toc150865091)

[2. 利用大模型撰写论文与学术剽窃 4](#_Toc150865092)

[2.1. ChatGPT不可靠的原因 4](#_Toc150865093)

[2.2. 怎样避免陷入学术剽窃 4](#_Toc150865094)

[3. 语言大模型会对论文的审查机制造成怎样的影响 5](#_Toc150865095)

[3.1. 语言大模型的滥用 5](#_Toc150865096)

[3.2. 监察机制应该做出改进 5](#_Toc150865097)

[4. 结论与建议 6](#_Toc150865098)

[5. 参考文献 7](#_Toc150865099)

智能生成大模型对于学术不端的影响

学 生：杨和鹭

指导老师：石存

摘 要

随着OpenAI发布语言聊天模型chatgpt以及百度推出语言大模型文心一言，各个领域的专业人士都对其给予了极高的评价。这些语言大模型的训练数据包括了几乎大部分的学术论文，因此在论文撰写上语言大模型都会给予使用者极大的便利，由此也引出了诸多问题，例如使用语言大模型撰写论文是否属于学术剽窃；使用语言模型会对论文的审核机制造成怎样的影响；应不应该准许语言模型参与论文的撰写。

本文针对这三个问题进行分析，结合实际例子分析语言模型对于论文撰写所产生的影响，论文审核过程应该做出怎样的改进，以及在实际科研中如何避免学术不端。

关键词：语言大模型；论文撰写；审核机制；学术不端

# 文献综述

语言大模型的出现为各个领域带来了许多的创新，但与其潜在的益处相对应

的是一系列的学术不端行为：语言大模型可以生成高质量的文本，但这也为沦为代写、抄袭行为提供了便利，因为文本很难被判断出是否是智能生成产生的。随着对抗生成网络的发展，一些研究者也会利用这些工具生成一些具有误导性的文本和图像，捏造实验数据与研究结果。文献[5]指出如果不及时制止，将会导致学术诚信的下降，损害科研可信度。文献[1]指出，为了缓解大模型生成对学术界的不良影响，需要发展更为精妙的抄袭检测工具、建立开放透明的实践标准等。文献[8]表明，在日本语言大模型在学术界也被广泛的应用，几乎有50%的大学生利用语言大模型进行论文撰写，有90%以上的学生利用其做作业，写报告等。应该认可语言大模型是一个强力的论文书写工具，但如何防止其用于学术造假，发挥其对于科研的积极作用，应该是我们应该探讨的重点。

# 利用大模型撰写论文与学术剽窃

## ChatGPT不可靠的原因

以ChatGPT为例，ChatGPT将大量的学术论文作为模型训练的数据集，其强大的自然语言处理能力可以给使用者提供实验的方法思路，跟踪一些研究现状，乃至直接撰写论文。文章[3]表明使用ChatGPT生成论文，这样做是否构成抄袭、剽窃也成为了学术界热议的话题，因为ChatGPT属于某物而非某人。

究其根源，ChatGPT的训练样本来自于互联网中的各个角落，包括各种论坛，各种百科等，因此如果滥用ChatGPT，这种行为本身和去互联网上找一篇知乎文章复制粘贴没有什么本质区别。

## 怎样避免陷入学术剽窃

这些信息来源良莠不齐，根本无法保证权威性与正确性，这也是为什么绝大多数的权威期刊（例如Science）明确禁止将ChatGPT列为合著者，且不允许在论文中使用ChatGPT所生产的文本。《细胞》和《柳叶刀》的限制性措施为：允许使用人工智能工具来提高文章的可读性，但不能取代作者完成关键任务，例如解释数据或得出科学结论，同时，作者必须声明他们是否以及如何使用人工智能工具。

可以看出，各大权威期刊为了防止语言模型带来的学术剽窃，对其使用范围进行了严格的限制。这种语言大模型毫无疑问会成为论文代写等的常用工具，对人工智能模型代写的审查与打击也肯定会愈加严格。在我们的日常科研中，可以将其应用在一些研究现状的跟踪，通过使用语言模型检索信息缩小自己的检索范围，避免将其应用到论文的核心部分，以免在今后论文的审查中由于使用了语言模型而导致论文撤稿，造成不必要的损失。

# 语言大模型会对论文的审查机制造成怎样的影响

## 语言大模型的滥用

随着利用语言大模型的代写事件频繁出现，对论文的审查手段、审查机制却没有很好的提升。

近日，一篇关于复杂数学方程新解的论文发表在了物理学领域期刊Physica Scripta上。法国图卢兹大学的计算机科学家Guillaume Cabanac发现论文中出现了 “重新生成响应（Regenerate response）”，这明显是利用ChatGPT生成的。最终这篇文章也是被期刊撤稿，给出的理由是违背了论文书写的道德以及无法判别论文其他部分是不是伪造的。

从论文书写方面来说，论造假的行为造假难度极低，打击难度极高。文献[5]表明，ChatGPT本身并不能创造出高水平的论文，但其作用在于可能在生成文本的过程中掺杂许多隐蔽、难以发现的错误，例如数学公式推导上，即使ChatGPT并不知道如何解题，但它最后都会给出一个“正确答案”。这些答案是来自于训练数据中类似答案的“缝合体”，利用其复杂且明显不正确的推导过程来欺骗用户。在论文撰写过程中，如果用户将这部分工作内容交给ChatGPT并撰写在论文当中，那么对于审稿人来说，如果不逐个符号进行排查的话很难发现其中的错误。对于撰写论文的人来说，想要瞒天过海学术造假，ChatGPT绝对是很好的助手[2]。

## 监察机制应该做出改进

对于论文审核来说，几乎很难通过一段文本本身反推其是不是出于语言大模型之手，文献[6]表明几乎所有的文本检测软件都不能确保检测结果完全正确，就连OpenAI自家的检测工具效果也不是很理想。因此想要筛除滥用语言大模型撰写的论文这种不端的行为，直接有效的方法就是耗费大量的人力物力，对论文进行仔细的审核，通过辨别论文本身是不是高质量的论文来防止语言大模型的滥用。文献[1]和文献[7]表明滥用语言模型撰写论文也能侧面反映出整个学术环境就大环境就有待完善，相关的论文写作规范，著作权版权意识等不够深入人心。因此相应的机构应该严入严出，推行相应的指标政策来约束这种行为。

# 结论与建议

语言大模型可以生成连贯、流畅的文本，帮助构建论文的框架和段落。也可

以帮助我们思考论文的结构、提供想法。同时我们也应该知道，语言大模型并不是专业领域专家，它的回答基于训练数据，可能不总是准确或完全符合学术标准。用语言大模型时，最好将其视为一个工具，而不是唯一的信息来源。大量的去查找文献，提高自己思维的宽度和广度才是重中之重，才能有自己的核心思想与技术路线。

因此我认为，编辑和润色文本以确保逻辑的一致性、语法的正确性等可以使用这些语言大模型工具，而在核心的Abstract、Method、Result这种论文的核心部分则要避免使用。同时在使用后要表明哪里用过大预言模型生成的文本，并且要关注所投期刊对于人工智能生成的文本是否有不同的限制。

# 参考文献

1. 王少.ChatGPT与学术不端治理:挑战与应对[J/OL].科技进步与对策:1-8[2023-11-13].
2. 白如江,秦明艳,张玉洁.科研数据学术不端影响因素研究[J].科技进步与对策,2023,40(20):110-121.
3. 关于学术不端行为的界定[J].校园心理,2023,21(05):380.
4. 杨励,程亦男.研究生学术不端预防处理体制机制构建研究[J].黑龙江教育(理论与实践),2023(09):38-42.
5. 张璐,郭晓亮,景勇等.ChatGPT对学术期刊的影响及应对策略研究[J].出版与印刷,2023(04):91-96.DOI:10.19619/j.issn.1007-1938.2023.00.043.
6. 沈锡宾,王立磊.人工智能生成学术期刊文本的检测研究[J].科技与出版,2023(08):56-62.DOI:10.16510/j.cnki.kjycb.2023.08.002.
7. 侯利阳,李兆轩.ChatGPT学术性使用中的法律挑战与制度因应[J].东北师大学报(哲学社会科学版),2023(04):29-39.DOI:10.16164/j.cnki.22-1062/c.2023.04.004.
8. 蓝茵茵,李婷.日本学术不端现象的成因与治理启示[J].教育评论,2023(06):163-168.

对于课程的建议：我认为在PPT展示的过程中小组的人应该全部上台参与展示，谁做了哪部分谁就讲哪部分。很遗憾在第三组的展示中给整个展示过程带来了一些不好的插曲，我认为我虽然有责任，但主要责任在于另外三个组员不积极参与PPT的制作，另外三个组员一共做了6页PPT，质量滥竽充数，而且PPT还无法播放，最后的PPT讲解也是没人想讲，我讲。并且我制作的前7页PPT是没有任何问题的，因此我感觉这件事就很令人郁闷，我认为我已经做到位了，每个人也应该为自己的所做负责任，结果事件发生后就都销声匿迹了，对此我只是想客观地阐述事实，希望能够得到正确的处理，对于课程本身没有其他意见，谢谢老师。