
使用AI虚构参考文献的作者将被arXiv封禁一年

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/39939.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

使用AI虚构参考文献的作者将被arXiv封禁一年

。据《自然》报道，物理科学知识库arXiv近日宣布，若投稿被发现包含由人工智能（AI）工具虚构的参考文献，其作者将被禁止在该平台发表稿件，禁期一年。该禁令同样适用于其他有确凿迹象表明使用了生成式AI且未经仔细核查的稿件的作者。

美国俄勒冈州立大学计算机科学家、arXiv计算机科学部主席Thomas Dietterich称，一年处罚期满后，除非该作者的研究成果已被“声誉良好的同行评审期刊”接收，否则他们仍无法在arXiv上发布任何手稿。

arXiv的新举措是预印本服务器应对日益泛滥的AI“垃圾内容”，即使用生成式AI制作的低质量或无意义内容的最新且影响最广的例子之一。该举措在社交媒体上引发研究人员大量评论，正面和负面的评价都有。一些平台如arXiv一样，对不遵守其准则的作者实施封禁，而其他平台则直接排除了某些引发对生成式AI使用担忧的投稿类别。

尽管科学家们日益频繁地使用大语言模型来完成文献综述等多种合理的任务，arXiv的公告仍赢得了许多研究人员的支持。德国AI初创公司sci2sci的联合创始人Valeri Kremnev在社交媒体上发文称这是“很棒的举措，完全支持”，并希望相关举措不止于AI幻觉，而是要更全面的打击垃圾内容。

不过，并非所有人都确信此类举措是正确的。利用AI帮助研究人员进行同行评审的平台Reviewer3的创始人Natalie Khalil认为，arXiv的举措治标不治本，“如果一名研究人员被arXiv封禁，他们依然会做研究，只是换个地方发罢了。”

对此，Dietterich回应道，各平台需要通力合作，剔除大语言模型生成的错误参考文献和其他有问题的内容输出。“一名不负责任的研究人员可以在别处发表不负责任的研究，这并不能成为允许他们在arXiv上公布这些内容的理由。”

在接受《自然》杂志采访时，Dietterich表示，尽管arXiv之前已对多种违反其行为准则的行为进行了处罚，但该服务器直到最近才针对不当使用生成式AI制定了标准化的应对措施。

Dietterich说，现在公布制封禁举措是为了震慑采取此类不当行为的作者。在他看来，现在的研究人员过于信任大语言模型，没有花足够时间去分析模型生成的结果。他还指出，部分此类AI生成内容源自论文工厂。

Dietterich指出，AI垃圾内容在arXiv的计算机科学版块最为普遍。该版块的论文量约占该预印本服务器所有投稿的一半。“这个领域的作者是大语言模型技术的早期使用者，也是早期的滥用者。”

其他预印本服务器同样充斥着AI垃圾。比如，开放科学中心去年10月关闭了其预印本集合平台OSF Preprints，原因是低质量投稿大量涌入，其中许多是AI生成的垃圾内容。

心理学预印本服务器PsyArXiv的科学咨询委员会主席、爱尔兰梅努斯大学的Dermot Lynott表示，PsyArXiv对未披露大量使用生成式AI的作者实施永久封禁。

社会学预印本平台SocArXiv同样会永久封禁提交AI生成的垃圾的作者。“我们接收的任何论文都会被搜集起来投喂给下一代大语言模型，我们不想助长这种由废话和虚构内容构成的恶性循环。” SocArXiv主任、美国马里兰大学的Philip Cohen说。

生物和医学预印本平台bioRxiv和medRxiv的联合创始人、其运营商openRxiv的负责人Richard Sever表示，他们未曾考虑对投稿包含虚构参考文献的作者实施封禁。相反，他们正在探索开发帮助识别此类参考文献的技术。

Dietterich认为，生成式AI最终可能不再“胡言乱语”，这将使其产出的论文与真正研究论文更加难以区分。“届时，我们可能需要一些能让作者证明他们确实开展了论文中实验的方法了。”

作者：许悦 来源：中国科学报

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发