

## BDM 编辑部张月红 | ChatGPT 的感悟：既生瑜，何生亮？

原创BDM 编辑部张月红生物设计与制造 BDM2023-02-24 08:31发表于浙江

[https://mp.weixin.qq.com/s/Zn2Sv4\\_1vacPpRKvGkRSXA](https://mp.weixin.qq.com/s/Zn2Sv4_1vacPpRKvGkRSXA)

生物设计与制造 BDM 198 篇原创内容 公众号

### 编者按

前天，国内一家不错的医学英文学术期刊编辑约我从期刊编辑视角谈：

### 如何应对 ChatGPT 的洪流？

我真的不敢对 ChatGPT 说三道四啊！因为我也试着问它“开放科学当下的策略是什么”，它周密的 7 个答案让我对它的强大毋庸置疑，但最后我还是“很违心地拒绝”用它的 7 个答案，诚信地在一篇文章中写出了自己的粗浅理解（笑）。故而我直接向这位编辑推荐了《科学》主编在今年 1 月 26 日发表的“ChatGPT 很好玩，但不能当作者”（ChatGPT is fun, but not an author, 26 Jan. 2023, <https://www.science.org/doi/10.1126/science.adg7879>）的这篇社论，文中他强调说：

*“《科学》系列期刊正在更新版权许可和编辑政策，明确由 ChatGPT(或任何其他人工智能工具)生成的文本不能用于科学作品中，其图像或图形也不能是此类工具的产品。总之，人工智能的程序不能成为作者。”*

*违反这些政策将构成学术不端行为，与修改图像或抄袭的不端行为没有区别。当然，在研究论文中有许多由 AI 生成的合法数据集（而不是论文的文本）不包括在这次政策的更改中。”*

说实话，我心里还是很赞许这位约稿编辑的前卫性和敏感度，毕竟有人说 ChatGPT 或许是第四次科技革命的前兆，直觉也告诉我们，仅在信息业，ChatGPT 是继 WWW 互联网和 iPhone 之后的第三次信息革命，拭目以待吧？！

今天我看到最新一期的《科学》又一次聚焦“随着科学家们探究人工智能撰写文本，期刊也在制定政策 (As scientists explore AI-written text, journals hammer out policies, <https://www.science.org/content/article/scientists-explore-ai-written-text-journals-hammer-policies>)”，这或许是当下学术圈和出版业最热议的话题，读后感觉对 ChatGPT 横空出世的来龙去脉及逻辑关系叙述详尽，不由地想做一些解译和提亮 (highlight)，以飨我们的期刊作者和编辑，也算是对这位期刊编辑的热邀做一点空间回应吧。但是也不得不说出心里话，那就是“既生瑜，何生亮？”个中体会，大家去评说吧.....

在此，我根据自己的理解，特意加了几个小标题：

## 要求作者公开透明已是共识

文章开场很直接：许多期刊出台新政策要求作者公开文本生成工具的使用情况，并禁止将 ChatGPT 等大型语言模型列为合著者，以强调人类作者对确保文本准确性的责任。

如国际光学与光子学学会 (SPIE) 出版总监帕特里克·弗兰岑 (Patrick Franzen) 说：“自 2022 年 11 月以来，我们一直在讨论这个问题。”

ChatGPT 可以写出极为流畅且看似消息灵通的报告、文章和科学手稿。出于对内容的伦理性和准确性担忧，弗兰岑和其他期刊的管理人员正在积极地忙于保护学术文献，以防可能出现的全部或部分由计算机程序编写的手稿泛滥。

尽管有些出版商尚未出台应对政策，但多数已积极表态要坚决禁止人工智能生成的文本，正如 SPIE 的举措，要求作者必须披露使用自动化工具的实情。的确当下，编辑和同行审稿专家们几乎没有其他选项，因为眼下尚缺乏有效的执行工具。截至目前，还没有软件能够在一定时间内从始而终地检测出合成的文本。

这段时间，在线工具 ChatGPT 免费给公众使用，科学家们蜂拥而至尝试它（ChatGPT 创建总部位于美国 OpenAI 公司，该公司已经开始限制用户的使用）。许多人报告说，它有着前所未有的、不可思议的

能力，可以创造听起来似是而非的文本，充斥着看似事实的细节。ChatGPT 和它的兄弟，包括本月早些时候为部分用户推出的谷歌的 Bard，以及在 2022 年 11 月短暂向公众开放 Meta 的 Galactica——一种大型语言模型的人工智能算法，均在互联网上提取大量文本的样本上进行训练而为。该软件可以识别单词之间的模式和关系，从而允许模型对问题和提示做出相关的回答。

某些情况下，生成的文本与人们所写的文本难以区分。例如，根据 2022 年 12 月的预印本，阅读由 ChatGPT 生成的医学期刊摘要的研究人员未能识别出其中三分之一是由机器编写的。人工智能的开发人员可能有望创造出更强大的版本，包括专门针对科学文献进行训练的版本，这一“前景”在学术出版业引起了轩然大波。

目前为止，科学家们报告说，他们正在研究 ChatGPT，以探索它的功能，还有一些人已将 ChatGPT 列为手稿的合著者。出版界的专家们担心这种有限的使用，可能会导致大量的手稿会包含人工智能编写的大量文本。

而期刊管理者关心的一个问题是准确性。计算机科学家已经发现，如果软件没有接触到足够的训练数据来生成正确的回答，它通常会编造一个答案。如谷歌子公司开发的信息搜索聊天机器人 Sparrow 对 2022 年的一些预印本分析研究中发现，高达 20% 的回答包含错误。

人工智能文本也可能会偏向于算法训练内容中所包含的既定科学思想和假设。期刊编辑更担心伦理道德问题，他们认为使用文本生成器的作者有时会像自己写的那样呈现结果，这被称为“剽窃”。

现在，如《自然》和 Nature-Springer 旗下所有期刊 (*Nature* and all Springer Nature journals)，美国医学学会所有杂志，以及全球出版伦理委员会 COPE 和世界医学编辑协会 (World Association of Medical Editors) 等这些国际品牌出版团体纷纷对此出台实践政策。也如上述《科学》(*Science*)系列期刊一月宣布了强硬的立场：全面禁止生成文本。但颇为有趣的是《科学》主编也不忘在他的这篇强硬的社论中“留有余地”，即“期刊或许未来会放宽这一政策，这取决于科学界对文本生成器可能接受的使用方式来决定。”社论的结尾，他也不忘幽默地表示“放松标准要比收紧标准容易得多”。

### **新技术会来监控 ChatGPT**

当下，一些出版官员仍在研究细节，比如何时要求期刊工作人员、编辑或审稿人检查或核实作者披露的生成文本？泰勒与弗朗西斯出版集团 (Taylor & Francis Publishing Group) 出版伦理与诚信的主管萨宾娜·阿拉姆(Sabina Alam)表示，即将出台的规定可能会要求公开这类文本，编辑可能会要求作者具体说明他们手稿的哪些部分是用 AI 工具写的。她说，如将其结果剪切粘贴到某个学术视角或观点在文章中“就不合适，因为这不是你的视角。”她说，随着出版业处理这类

稿件的经验越来越丰富，这项政策可能会不断演变。“我们认为这是一个分阶段的方法，现在还时尚早。”

期刊管理人员还表示，希望使用更多的技术来监控这项新技术 - ChatGPT，即可标示合成文本的自动探测器。但高级研究科学家多梅尼克·罗萨蒂 (Domenic Rosati) 说，这并不容易。Ai 是一家开发软件以协助科学出版商的公司。“如果因为文章的流畅性或缺乏真实性能够说出 (某些文本显然来自) 机器，我们恐早已过时了，尤其是在科学领域。”

所以，这个领域正在热切期待探测器不断改进。仅 OpenAI 也承认了一些局限性。但是相关的检测工具仍在开发中。更好的解决方案可能不久即将出现。OpenAI 在 2022 年 12 月表示，他们正在研究为生成的文本添加水印的方法。它将对其模型进行编程，从而创建一个搜索引擎可以探测到的秘密代码。上个月，斯坦福大学的一个团队发布了一份描述 DetectGPT 的预印本，它与其他探测器不同，DetectGPT 不需要训练。该算法通过创建多个随机变体来检查文本，并查询文本生成器来对它喜欢的版本进行排名。该团队研究的生成器 (由 OpenAI 开发，与 ChatGPT 类似) 更喜欢原始文本而不是修改版本，这在人类编写的文本和人工智能生成的文本中是一致的，这使得 DetectGPT 能够预测样本来自特定机器的可能性。但是，领导斯坦福大学团队的博士生 Eric Mitchell 说，在期刊编辑完全依靠 DetectGPT 的结果对

稿件做出决定之前，DetectGPT 还需要更多的技术发展。TurnItIn 公司销售了一款广泛使用的抄袭检测器，该公司上周表示，计划最早在 4 月份推出一款合成文本检测器。TurnItIn 表示，该工具经过学术写作训练，可以识别由 ChatGPT 生成的 97% 文本，误报率为百分之一。

一口气读到此，我不由地想到《三国演义》的一句经典“既生瑜，何生亮？”不是吗？！既要如此查探 ChatGPT，何必让 ChatGPT 诞生呢？也不由地倒吸一口冷气：人类啊，你的期望是否有极限？谁又能控制了你的智慧和探索？

但 BDM 在 2023 年 2 月 20 日第一次学科编辑会议上明确共识：对投稿中使用 ChatGPT 生成文本的态度与《科学》的政策保持一致。



——BDM 编辑部张月红