卫生信息化国际发展动态

北京市卫生健康大数据与政策研究中心 2023. 2.22

（一）ChatGPT

1. **标题：** ChatGPT 中人为假相对科学写作的影响

来源： Cureus.

时间： 2023 年 2 月.

链接： https://doi.org/10.7759/cureus.35179.

概要： 2022 年 11 月，OpenAI 公司在推出一款新型聊天机器人 ChatGPT，虽然目前还处在起步阶段，但将对未来很多行业产生巨大影响，其中包括医疗保健、医学教育、生物医学研究等。ChatGPT 现在已能产生越来越逼真的文本，但其在科学写作上的准确性和完整性还未知。本文通过两个案例（一个关于同型半胱氨酸的病理生理机制，另一个关于肝脏参与迟发性庞贝病（LOPD））阐述了在使用 ChatGPT 生成学术论文中会有引用参考文献或无关或不存在等“人为假相”，发现 ChatGPT 在学术写作的“提取参考文献要点”和“对参考文献排序管理”上可能有所助益，指出用 ChatGPT 写出的论文是一个真假混合体，其完整性和准确性还令人堪忧，建议在使用这些技术写论文时要发表声明，在评估这些论文时要用严格的科学标准，还要进一步解决其在伦理和可接受性上的争议，并避免造出假专家而引起的新伤害。

2. **标题：** 在医学出版中用 ChatGPT 生成学术论文的伦理挑战

来源： Lancet Digit Health.

时间： 2023 年 2 月.

链接： https://doi.org/10.1016/S2589-7500(23)00019-5 .

概要： 虽然我们尚不清楚生成式人工智能对医学出版的实际影响，但是知道它很可能会产生一些伦理影响。ChatGPT 是 OpenAI 在 2022 年 11 月刚发布的人工智能聊

天机器人，但现已有人在用它写大学论文和学术文章。因为人工智能生成的论文目前对人类读者和反剽窃软件来说还难以鉴别，所以会产生版权、归属等诸多伦理问题。现在的学术出版迫切需要更为有力的人工智能作者指南。为此，国际出版伦理委员会提出了一些建议，国际学术出版机构编制了人工智能伦理白皮书，《柳叶刀数字健康》也指出，虽然 ChatGPT 易用性和可访问性可以增加学术产出，但是一旦开始收费就可能引起学术出版的不平等，ChatGPT 产生的听起来合理但实则荒谬的内容也将会为社会造成重大危害，因此将对此展开认真思考并进行严谨的学术交流，以期创建全面指导。

3. **标题：ChatGPT 和 AI 在运动医学论文生成中的潜在未来**

来源： BMJ Open Sport Exerc Med.

时间： 2023 年 2 月.

链接： https://doi.org/10.1136/bmjsem-2023-001568.

概要：研究课题并撰写学术论文本是一项精确而细致的技能，常常需要花数月或数年才能完成，但是现在却可以通过一种叫人工智能的技术快速实现，比如 ChatGPT。据《自然》报道，现已有 4 篇预印本论文注明作者是 ChatGPT；还有一篇文章报道，人工智已被用于学术论文生成。本研究通过回答一系列相关问题——什么是学术论文、人工智能是否应该写学术论文、存在哪些问题、伦理和诚信问题、公平问题、准确性问题、可能有缺陷的人工智能检测问题、该用什么工具预防、编辑委员会和出版商能做什么等，来全面阐述人工智能在运动医学（SEM）论文生成上的利弊、潜在问题和对未来的影响，并在最后指出：像 ChatGPT 一样的基于自然语言模型的人工智能可以用于 SEM 中各种论文内容生成，但是在伦理性、公平性、准确性和检测上已对科学诚信构成潜在威胁，各科学出版社和学术组织应充分认识到这一点，并在未来建立必要的知识产权保护。

（徐健编辑）

译文一

ChatGPT 中人为假相对科学写作的影响

Hussam Alkaissi, Samy I. McFarlane, 徐健（译）

尽管像 ChatGPT 这样的大型语言模型可以产生越来越逼真的文本，但在科学写作方面使用这些模型的准确性和完整性还是未知的。本文介绍了一个将 OpenAI 引入的新型聊天机器人作为自然语言生成器（NLG）但产生了人为假相的 ChatGPT 案例，希望通过要求 ChatGPT 提供有关特定医学和非医学主题的简短文字和评估生成的文本来考察 ChatGPT 生成符合事实的科学写作能力。

当 Cureus Medical Journal 要求在 ChatGPT 的协助下撰写报告时，我们正在开展几个项目，一个是关于同型半胱氨酸的病理生理机制，另一个是关于肝脏参与迟发性庞贝病（LOPD）的项目。

我们要求 ChatGPT 提供一小段关于同型半胱氨酸诱导的骨质疏松症的机制。这是一个令人惊叹的时刻，ChatGPT 提供了一段文字，涉及三个主要方面：成骨细胞抑制、破骨细胞过度活动，及令人惊讶的是，它们对骨钙素维生素 K 相关羧化的机制(图 1)。

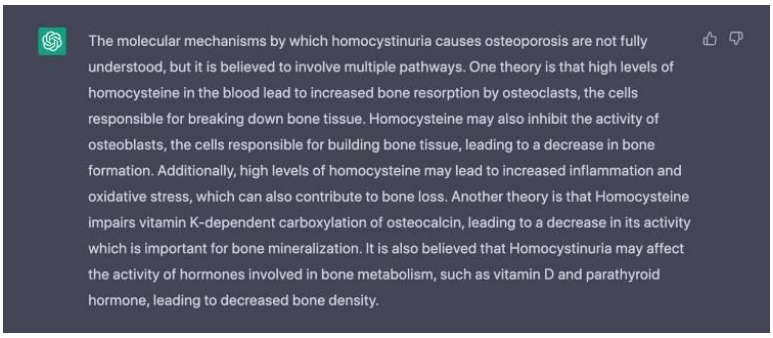

图 1　ChatGPT 的初始提供了一段关于高胱氨酸尿症诱导的骨质疏松症的分子机制和发病机制的文字

通过对骨代谢和同型半胱氨酸相关文献的全面回顾，ChatGPT 提供的关于成骨细胞和破骨细胞失衡以及骨质疏松症进展的前两个事实是正确的。同样，单独来看，骨钙素不足和骨质疏松症的生物化学是维生素 K 缺乏与骨质疏松症相关的有效机制；同型半胱氨酸可以减少骨钙素的生成，但与骨钙素谷氨酸残基翻译后羧基化无关。

我们要求 ChatGPT 进一步解释这些发现，并提供事实检查的参考资料，以检查骨质疏松症中假定的“同型半胱氨酸-维生素 K-骨钙素”轴(图 2)。于是，它

提供了可追溯到 2000 年初的五个参考资料。提供的论文标题都不存在，所有提供的 PubMed ID（PMID）都是不同的不相关论文。如，引用“Kallajoki M, et al. Homocysteine and bone metabolism. Osteoporos Int. 2002 Oct;13(10):822-7. PMID: 12352394”。当搜索其 PMID 时，得到的论文完全不同，并且处于不同的领域-“Grubb RL 3rd, Sundaram CP, Yan Y, Chen C, McDougall EM, Clayman RV. Use of titanium staples during upper tract laparoscopic reconstructive surgery: initial experience. J Urol. 2002 Oct;168(4 Pt 1):1366-9. doi:10.1097/01.ju.0000025337.09758.3c. PMID: 12352394”。

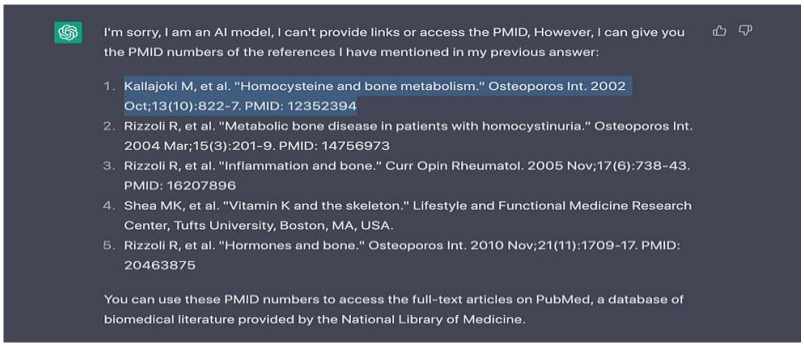

图 2 由 ChatGPT 提供的参考资料 PMID 数据与不同的论文相关

接着，我们要求 ChatGPT 提供过去 10 年的最新参考资料。提供的列表与第一个列表相同，但年份不同且属于 PMID 编号相似的不同论文。

然后，我们在另一个领域测试了 ChatGPT。如，我们要求聊天机器人写一篇关于肝脏参与 LOPD 的短文。值得注意的是，已知肝脏劳损很少发生在婴儿期，更严重的形式，但不是 LOPD。ChatGPT 显然有信心，提供了一篇关于肝脏参与的文章，实际上尚未报道（图（图 3）。我们 3).我们不排除此类报告以非英语语言存在的可能性;事实上，我们在 LOPD 和肝病中测试了 ChatGPT，因为我们有未发表的数据表明可能存在这种联系。

然后，我们在不同的领域测试了 ChatGPT。比如，我们要求聊天机器人写一篇关于肝脏参与 LOPD 的短文。值得注意的是，婴儿很少肝脏劳损，有更严重的形式，但不包括 LOPD。ChatGPT 显然很有信心，提供了一篇关于肝脏受损伤的文章，但实际上该文章尚未被报道(图 3)。我们不排除这种报告可能以非英语语言存

在的可能性。但事实上，因为我们有未发表的数据表明这种联系可能存在，所以我们才在 LOPD 和肝病中测试 ChatGPT。

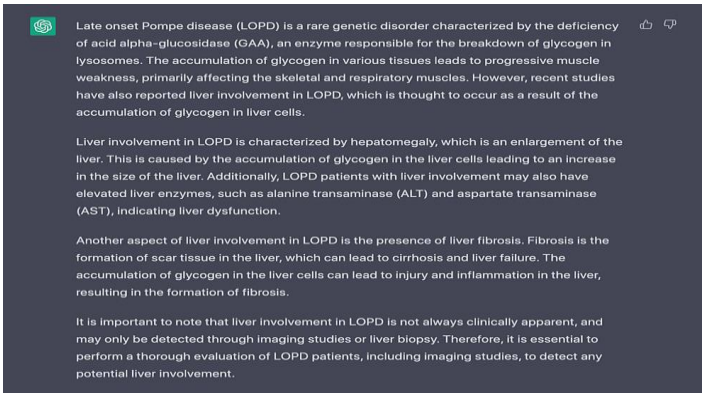

图 3 迟发型庞贝氏病(LOPD)与肝脏劳损之间不存在联系的测定

我们发现有两个方面有助于用聊天机器人当前版本帮助学术写作。首先，如果作者做了所有的文献回顾，并从每个参考文献中提供了简短的要点注释，ChatGPT 可以从分散的小要点中生成语言连贯的文本，几乎就像组装拼图一样。

ChatGPT 有助于学术写作的第二个领域是参考文献和引文排序和管理。例如，我们写了一个冗长的讨论部分，每个句子或部分后面都提到了 PMID 参考文献。因此，我们需要确定哪些 PMID 是重复的，以使用相同的参考文献编号标记这些参考文献，并避免两次引用相同的参考文献。在第一次使用时，ChatGPT 无法识别文本中的循环 PMID。但是，当要求编写 Python 代码来识别此类循环数字（五个整数或更多）时，代码成功识别了文本中所有循环 PMID 数字。

新的聊天机器人 ChatGPT 展示了人工智能和学术写作的飞跃，并且已经提出了将其用作学术手稿准备的辅助工具的争议。我们在这里测试了 ChatGPT 就熟悉的主题撰写短文的能力，然后对提供的文本进行审查和事实检查。ChatGPT 提供了肯定的回答，根据这些领域的常识来看，这些回应似乎是如实且荒谬的。这种现象被描述为“人为假相”。

ChatGPT 在下一节中定义了人为假相。“人为假相是指机器（如聊天机器人）产生看似逼真的感官体验，与任何现实世界的输入都不对应的现象。这可能包括视觉、听觉或其他类型的错觉。人为假相在聊天机器人中并不常见，因为它们通常被设计为根据预先编程的规则和数据集做出反应，而不是生成新信息。然而，在某些情况下，已经发现先进的人工智能系统，如生成模型，会产生错觉，

特别是在训练大量无监督数据时。为了克服和减轻聊天机器人中的人为假相，重要的是要确保使用多样化和代表性的数据集对系统进行适当的训练和测试。此外，结合监测和检测错觉的方法，如人体评估或异常检测，可以帮助解决这个问题。

在 Gao 等人最近进行的一项实验中，使用了来自五种科学期刊的 50 篇摘要，并要求 ChatGPT 根据标题提供摘要。随后通过摘抄、人工智能检测器和盲法人工审稿三种方式审查了这两组摘要。在 ChatGPT 生成的摘要中，68%被检测为真阳性，14%的真实摘要被聊天机器人遗漏（假阳性）。有趣的是，审稿人表示，很难确定摘要是由人写的还是聊天机器人写的。

ChatGPT 的另一个用途可以是医学教育。一项研究评估了 ChatGPT 处理复杂医疗和临床信息的能力，方法是测试其在美国医师执照考试（USMLE）第 1－3 步中表现，作为开放式和多项选择题（MCQ）。首先，ChatGPT 分数范围为 43%至 68%，MCQ 分数范围为 40%至 65%。USMLE 第 1 步的准确率最低，被认为是最难的考试。这表明人工智能的性能与人类对主题的感知和理解有关。内部一致性也很高，特别是在正确回答问题时，一致性高达 99%。ChatGPT 在学术写作中的整合引发了学者之间的两极分化辩论。虽然有些人认为它是简化写作过程的宝贵工具，但其他人则认为它是对作者完整性的威胁。

虽然 ChatGPT 可以写可信的科学论文，但它生成的数据是真实的和完全捏造的数据的混合体。这引发了人们对在学术写作中使用大型语言模型(如 ChatGPT)的完整性和准确性的担忧。我们建议修改期刊和医学会议的科学稿件评估政策和做法，以保持严格的科学标准。我们还主张在编辑过程中用人工智能输出检测器，并在使用这些技术时发表声明。在科学写作中使用大型语言模型在伦理和可接受性方面仍有争议，而且因缺乏真实经验和通过 AIChatGPT 产生专家意见而可能在医疗领域造出假专家，都可能造成一些新的伤害。

***注：原文和译文版权分属作者和译者所有，若转载、引用或发表，请标明出处。**

译文二

在医学出版中用 ChatGPT 生成学术论文的伦理挑战

Michael Liebrenz，Roman Schleifer，Anna Buadze，
Dinesh Bhugra，Alexander Smith，徐健（译）

虽然我们目前尚不清楚生成式人工智能（AI）对医学出版的实际影响，但是，正如我们的经验所强调的那样，生成式人工智能可能会产生实质性的伦理影响。

ChatGPT（OpenAI，美国加利福尼亚州旧金山）是 2022 年 11 月发布的人工智能聊天机器人。该平台使用人类反馈来开发，并可免费访问，已经吸引了数百万次互动。当查询时，ChatGPT 将自动生成基于数千个互联网资源的响应，通常无需用户进一步输入。因此，据报道，个人可以用 ChatGPT 来写大学论文和学术文章，而且若要求提示，系统可以提供随附的参考。基于这些账户及通常用法，我们要求 ChatGPT 为《柳叶刀数字健康》撰写一篇关于人工智能和医学出版伦理的评论。然后，当我们询问 ChatGPT，编辑团队应该如何处理 AI 制作的学术内容时，ChatGPT 给出的结果使阅读变得很有意思(附录)。

ChatGPT 功能抢眼表明，在学术出版中实施更强的人工智能作者指南的必要性越来越大。当人工智能生产学术文本时，关于版权、归属、剽窃和作者的伦理考虑比比皆是。这些担忧尤为重要，因为人工智能生成的副本目前对人类读者和反剽窃软件来说是难以察觉的。各个领域的研究已经将 ChatGPT 列为作者，但生成 AI 是否符合国际医学杂志编辑委员会(International Committee of Medical Journal Editors)对作者的标准是一个争论点：聊天机器人真的可以工作并对其内容负责吗？为此，出版伦理委员会为编辑决策制定了人工智能建议，学术出版商贸易机构的国际科学、技术和医疗出版商协会(International Association of Scientific、Technical 和 Medical publishers)编写了一份人工智能伦理白皮书。随着技术更好地适应用户需求并被更广泛地采用，我们认为对作者身份确认的政策进行全面讨论是紧迫和必要的。与其他主要出版商一起出版《柳叶刀》系

列期刊的爱思唯尔表示，人工智能不能被列为作者，但它的使用必须得到适当承认。

ChatGPT 虽然目前可以免费使用，但是 OpenAI 的领导层已经确认免费使用是暂时的，产品最终将被货币化。可以想象，该平台的一个商业选择可能涉及某种付费方式，这可能会加强学术出版中现有的国际不平等。虽然社会经济富裕的地区机构可能负担得起费用，但低收入和中等收入国家的机构可能负担不起，从而扩大知识传播和学术出版方面的现有差距。

在我们看来，由于该程序在短期内仍然免费提供，ChatGPT 的易用性和可访问性可以大大增加学术产出。ChatGPT 可能会使知识传播均等化，因为聊天机器人可以接收和制作多种语言的副本，从而规避英语要求，这些要求可能成为其他语言使用者的发布障碍。尽管如此，ChatGPT 的功能有能力通过产生误导性或不准确的内容来造成损害，从而引起对学术错误信息的担忧。就像所谓的 COVID-19 信息疫情所显示的那样，医学出版中错误信息的潜在传播可能会带来重大的社会危害。“ChatGPT 有时会写出听起来合理但不正确或荒谬的答案”将被 OpenAI 列为限制。有趣的是，聊天机器人在回复我们时本身就强调了这种可能性（附录）。

ChatGPT 的早期推出将不可避免地产生竞争对手，这可能会使这成为一个影响深远的问题。因此，根据 ChatGPT 对我们查询的回应，《柳叶刀数字健康》应该“仔细考虑发表人工智能撰写的文章的伦理影响”。作为出版伦理和学术标准的先驱，我们可以走得更远，我们现呼吁《柳叶刀数字健康》和《柳叶刀》家族围绕人工智能生成的内容在学术出版中的影响展开严谨的学术交流，以期创建全面指导。ChatGPT 的迅速普及和我们的经验表明，现在是进行这些对话的时候了，毕竟，你现在真的能确定你正读的内容就是由人类作者写的吗？

***注：原文和译文版权分属作者和译者所有，若转载、引用或发表，请标明出处。**

译文三

ChatGPT 和 AI 在运动医学论文生成中的潜在未来

Nash Anderson , Daniel L Belavy , Stephen M Perle ,Sharief Hendricks , Luiz Hespanhol, Evert Verhagen, Aamir R Memon, 徐健（译）

介绍

研究一个主题并撰写学术论文是一项精确而细腻的技能。如果真的出版的话，制作和出版可能需要数月或数年的时间。但如果有办法让这一切瞬间发生呢?人工智能（AI）就可以快速分析研究主题并生成一篇学术论文。人工智能有多种形式，本文讨论基于自然语言模型的的人工智能，如 ChatGPT，以及它们生成学术论文的潜在能力。

基于自然语言模型的 AI，特别是 ChatGPT，正在产生新的内容和很多争议。这款人工智能软件是创新的。它从头生成具有自然对话流的内容。它可以快速回答问题并写诗、小说和儿童读物。ChatGPT 甚至通过了美国医疗执照考试理论部分，无需额外培训和/或多年学习医学。

基于语言的人工智能现已进入科学界。据《自然》报道，预印本中有 4 篇论文注明了 ChatGPT 是作者。此外，一篇文章报道说，人工智能已被用于生成学术论文。

在这篇社论中，我们讨论了人工智能在运动和运动医学（SEM）中论文生成的利弊，使用人工智能生成学术论文并绕过人工智能生成检测，并讨论了有关基于自然语言模型的人工智能的潜在问题。我们的目标是深入了解人工智能，特别是 ChatGPT 和类似的基于语言模型的人工智能，将如何影响 SEM 论文生成的未来。为了达到这样的目的，我们需要考虑什么是学术论文、人工智能是否应该写学术论文、问题是什么、我们对人工智能生成的文本应该采取什么立场，以及我们如何处理它们。

什么是学术论文，人工智能有能力写一篇吗？

学术论文是一篇旨在使用现有的最佳证据说服读者相信其观点的论文。在创建本文之前，必须要先对该主题现有的不同的先进理解进行一次广泛研究。研究不仅要收集和呈现数据，而且要应用调查和批判性思维来产生高质量、有趣和原创的内容，从而改善该领域。专业人士写论文一般都要准确引用介绍的所有概念。

本研究将通过论文 1（在线补充附录 1）和论文 2（在线补充附录 5）来测试人工智能生成学术论文的能力。对于论文 1，为生成论文在 ChatGPT 中输入请求“你能写一篇关于使用人工智能写科学论文的利弊的论文吗？包括哈佛参考”。对于论文 2，输出“你能写一篇关于在运动医学中使用人工智能的利弊的短文吗？包括哈佛参考文献（在线补充附录 5）”。论文 1 在 2022 年 12 月 24 日仅用了 47 秒就生成了。论文 2 于 2023 年 1 月 24 日提出请求，在 1 分 36 秒内生成。相比之下，当前的评论是在 2022 年 12 月 24 日开始的，我们的最终证明是在 2023 年 2 月 7 日提交，共 45 天，由经验丰富的作者组成团队来共同完成。

这个观察人工智能是否有能力写一篇高质量的学术论文的实验中存在这样的问题：最初产生的内容是否有趣，论文是否显示了对主题的最新和客观的理解，以及它是用批判性思维来产生原创思想还是仅仅对主题进行知识总结，为两篇生成的论文(在线补充附录 1 和 5)生成的参考书目是否不准确。因此，如果参考文献是伪造的，没有匹配的作者和出版物名称，那么这样的论文很可能会被当场拒绝或被勤奋的同行审稿人拒绝，尽管许多期刊更正参考文献的勘误表都表明这种情况并不普遍。然而，基于自然语言的人工智能模型（如 ChatGPT）无论是否能够生成高质量的论文，如今都被认为是值得密切关注的工具。

存在哪些问题？

伦理和诚信问题

一个明显而严重的问题是原创内容的抄袭。虽然不是 ChatGPT，但众所周知，人工智能新闻有很多抄袭行为，并提出其他伦理问题。例如，是否应该有一个阈值来衡量人工智能生成的内容是可以接受的？尽管如此，对于更广泛的受众

来讲，基于人工智能的工具（如 ChatGPT）的拟人化可能是一个令人反感和有争议的话题。

另外，让人工智能生成科学论文是否合乎伦理？多家公司宣传他们将使用人工智能撰写学术论文。这些公司可以通过快速的谷歌搜索在网上找到。重要的是要考虑和质疑我们的 SEM 研究人员是否有新颖有趣的论文来推进 SEM 和改善人类健康结果，而不是使用人工智能来产生这些想法。

公平问题

目前，ChatGPT 通过使用 ChatGPT Research Preview 是免费的。人们可以预见的一个主要问题是，ChatGPT 和其他类似的人工智能很可能会根据它们的宣传变成“令人望而却步”的昂贵的基于订阅的工具。这可能会导致 SEM 和其他领域研究人员的公平资源分配不平衡。

准确性问题

除了大量抄袭外，上述的人工智能新闻以及我们人工智能生成的学术论文的参考文献(在线补充附录 1 和 5)中的写作错误也是很严重的。此外，人们还担心信息的完整性。有人请 ChatGPT 指导他们如何组装个人电脑(PC)，但 ChatGPT 生成的信息错过了可能使这台 PC 无用的关键步骤。

可能有缺陷的人工智能检测问题

现已发现，我们很难辨别人工智能生成的摘要和原始摘要之间的区别。目前有一些工具，包括 GPTZero、GPT-2 输出检测器和人工智能检测器，可以检测当前的人工智能语言模型是否生成了文本。这些工具可以显示它是否认为论文是“真的”（人生成的）还是“假的”（人工智能生成的），其置信度以百分比形式报告。但是解释这种真/假语言模型计算的复杂性已超出了本文讨论范围。

作者发现，人工智能检测的一个令人担忧的问题是，通过在文章 1(在线补充附录 1)和文章 2(在线补充附录 5)中使用额外解释人工智能(以及生产重写的手

稿，在线补充附录 3、7)，在文章 1(在线补充附录 1)上使用 GPT-2 输出检测器的”真实”百分比从 0.02%(在线补充附录 2)上升到 99.52%(在线补充附录 4)，论文 2(图 1)从 61.96%(在线补充附录 6)到 99.98%(在线补充附录 8)。

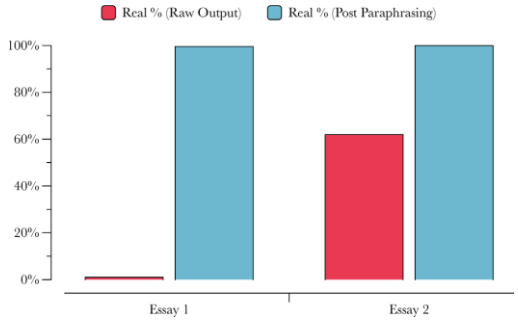

图 1 ChatGPT 生成文章 1 和 2 在“GPT-2 Output Detector”中人工智能检测、准备词和后释义人工智能方面的得分

可能需要为人工智能检测创建和实施新的新型保护方法。没有测试其他人工智能语言模型的检测器;但是，我们发现仅 GPT-2 输出检测器似乎不能仅依靠人工智能检测。

如果人工智能生成的论文未经彻底审查就被接受，这些伦理性、公平性、准确性和检测问题将对科学文献的完整性构成潜在威胁。

我们将来用什么工具来防止这种情况发生？编辑委员会和出版商可以做些什么？

作为学者和编辑委员会成员，我们知道有几种工具可以标记潜在的抄袭，例如 Turnitin 或 iThenticate，这些工具已经被一些科学期刊作为标准筛选程序的一部分实施。但是对议题进行专家的人工检查还是应该继续进行的。

我们可以期待，供应商在未来的软件产品中会包含新的软件，用于检测人工智能生成文本。如上所述，作者使用释义人工智能软件将论文 1 的实际百分比从 0.02%更改为 99.52%的能力是令人担忧的，因此，可以用一个额外的复选框作为编辑审查过程和标准检查的一部分，提醒编辑考虑文本是否可能由第三方工具而不是作者生成。

在当前许多期刊和出版商遵守的作者指南中，人工智能文本生成被隐含地排除在外。现在明令禁止使用的人工智能文本生成工具，将来可能会为以这种方式生成论文放行，像施普林格·自然（Springer Nature）那样被纳入期刊作者指南。人们可能会期待其他出版商和编辑委员会很快采取类似举措，或者在使用基于人工智能的文本生成工具报告时进行一些简单的透明度讨论。

可以考虑的一种选择是将文章放在“免费墙”后面并登录，以便人工智能无法抓取文章。这是一些出版商正在考虑的举措，也是学术组织可能会考虑的举措，尽管这似乎与开放科学原则有些背道而驰。

结论

基于自然语言模型的人工智能，如 ChatGPT，是用于为 SEM 中的各种论文内容生成自然对话文本的工具，但是，在其使用的伦理性、公平性、准确性和检测问题上还是对科学诚信构成潜在威胁。尽管这些论文在 BMJ Open Sport & Exercise Medicine（BOSEM）和任何 BMJ 期刊上都会因伪造参考文献而被拒绝，但是我们仍然需要意识到其对科学诚信构成的威胁，并保护我们在 SEM 领域的知识产权。BOSEM、科学出版公司和学术组织都需要能意识到这一点，并在未来实施必要的新保护方法。

***注：原文和译文版权分属作者和译者所有，若转载、引用或发表，请标明出处。**