几分钟生成一篇论文，传统查重工具无法识别——

AI代写论文现象如何科学治理

深瞳工作室出品
采 写：实习记者 吴叶凡
本报记者 付丽丽
策 划：刘 恕 李 坤

“太不像话了！学生用人工智能生成的期末论文糊弄我。”近日，上海某高校教师在社交媒体上“吐槽”自己遇到的新难题——一些想偷懒的学生开始用人工智能技术完成论文。

以ChatGPT为代表的生成式人工智能技术（AIGC）横空出世，似乎为人们写论文提供了新帮手。从提供选题到文稿润色、从统计分析到图表制作……其功能之强大，几乎覆盖了学术论文写作过程的方方面面。

面对ChatGPT等工具的潜在风险，争议随之而来。不少人质疑，人工智能到底能不能用于辅助学术论文写作。有人认为，它只是提高科研效率的工具。有人则对此持审慎态度，认为容易引发大规模的学术诚信问题。

人工智能技术在论文写作中的应用程度如何？技术应用的边界在哪里？如何对这一技术进行有效治理？科技日报记者对此进行了深入采访。

AI生成的文本“非常水”

有多少人尝试过用人工智能技术写论文？去年《自然》杂志对全球博士后的一项调查发现，约有三分之一的受访者使用人工智能聊天机器人来优化文本、生成或编辑代码、整理文献。

当记者尝试在社交媒体上搜索“AI”“论文”“写作”等关键词，五花八门的AI论文写作指导教程映入眼帘。其中大部分宣称能够教会用户在几分钟内通过几个简单的步骤，生成一篇几万字的“优质”论文。这些教程的浏览量最高已达数百万。

AI真的能生成一篇完整的“优质”论文吗？记者按照教程开始了尝试：“请提出与民族志纪录片有关的论文选题。”几乎无需等待，几个看起来很“靠谱”的选题就出现在对话框里。

“请就某一选题生成写作大纲。”几秒后，7个像模像样的章节全部生成完毕。“请就提纲中某项内容，详细描述2000字。”重复几次操作后，一篇几万字的“论文”很快就完成了。但记者浏览后发现，其生成的段落中，存在大部分重复且言之无物的内容。

除了说“车轱辘”话，某985高校人工智能专业硕士研究生温睿还发现了此类论文的行文特点：“一般是先写一句话，然后进行分条论述。当老师看到这样套路化的内容就会猜测，这类文章很大程度上是人工智能写的。”

文章开头那位教师的经历印证了温睿的发现。“这样的论文看似条理清晰、层次丰富，但实际上每个层面的内容都很少，而且非常空洞。我马上就怀疑是AI生成的。”该老师说。

不少期刊编辑、审稿人也发现了同样的问题。

某人文社科期刊审稿人徐彬向记者透露，用AI写论文的关键在于提示词。如果提示词选用的不恰当，就极有可能得到一篇套路化的文章。他目前已经收到过五六篇“一眼就能看出来”用AI写的稿子。

“这些文章的共同特点就是非常水。虽然它生成的语言连贯性不错，但是缺乏深度，创新性也不强。”对此，徐彬略显无奈，“综述类文章是使用AI的重灾区，但目前期刊还缺乏相关的评价标准和处理机制。”

伪造数据集更具隐蔽性

在清华大学人工智能国际治理研究院副院长梁正看来，论文核心评价标准包括作者发挥的创造性、对论文的贡献程度。一篇大部分由AI生成且隐瞒使用情况的文章，既没有作者智力的贡献，也不符合科研诚信的要求，属于学术造假。

AIGC造成的学术造假还发生在数据领域。记者在采访过程中，多位业内专家提到了伪造数据集问题。相比直接的文本生成，这一方式更具有隐蔽性。

GPT-4的ADA功能是一种结合了计算机编程语言Python的模型，可以执行统计分析和创建数据可视化。梁正向记者讲述了一则真实的案例：国外某机构研究人员先是要求GPT-4 ADA创建一个关于圆锥角膜患者的数据集，后又要求它编造临床数据，用以支持深板层角膜移植术比穿透性角膜移植术效果更好的结论。但真实的临床数据证明，两种手术效果并无明显差别。

“针对某个问题，提出方法来解决，并通过实验来证明方法的可行性——这是专业论文的常用模式。人工智能不能做实验，哪怕它给的实验数据再理想，也都是虚假的。”温睿认为，虚假的数据背离了科学研究的真正意义。

除了数据处理，更多人使用AIGC来解释概念。温睿发现AIGC生成的概念简洁明了，查重率也非常低。但当记者询问这些概念是否正确时，温睿显得有些迟疑：“我也没有把握，通常默认它是对的。”

为了验证AIGC给出答案的准确性，记者就一些新兴概念提问，但它给出的答案往往和真正概念毫不沾边。当记者让AI生成5篇某领域的重点参考文献，它又胡编乱造了5个不存在的作者和不存在的文献。

在人工智能领域，描述AI“一本正经地胡说八道”的专业名词是“AI幻觉”。哈尔滨工业大学（深圳）特聘校长助理、教授张民解释，AI幻觉是指AI会生成貌似合理连贯，但与输入问题意图不一致、与现实或已知数据不符合或无法验证的内容。这多是由于AI对知识的记忆不足、理解能力不够、训练方式固有的弊端及模型本身技术的局限性所导致。

“如果不警惕AI幻觉，很有可能损害科学研究的真实性和客观性。”梁正表示，AI生成的错误信息一旦被广泛传播，不仅会造成“学术垃圾”泛滥，还将影响学术生态的良性发展。

一场你追我逃的“猫鼠游戏”

一项新技术的出现，对于社会的发展往往是把双刃剑。虽然人工智能技术存在种种隐患，但其在图文创作、数据处理等方面的强大能力已被大多数人认可。“归根结底，我们认为AI将增加人类的智慧，而非取代人类。其使用应在人类监督之下，并将道德因素考虑在内。”施普林格·自然集团发言人说。

推动AI向善发展，需要借助行之有效的技术手段。值得注意的是，AI生成的论文并不能被查重工具检测出来。因此，国内外都在探索研发专门针对AIGC的检测工具。

从原理看，AIGC检测技术是在“用AI打败AI”。同方知网数字出版技术股份有限公司副总经理柯春晓介绍：“人类的创作往往是随机且富有灵感的，而接受过大量文本训练的AI已经形成了生产文本的‘固有’范式，倾向于使用‘一致’的结构和规则，因此具有更高的可预测性。”AIGC检测的核心就是依托海量的文本和数据样本，识别出人类和AIGC工具在平均句子长度、词汇多样性和文本长度等方面的不同点，从而揪出AI论文“枪手”。

一些期刊出版机构通过检测工具发现了AIGC代写论文的痕迹。“从去年7月底到现在，我们发现涉嫌AI写作的论文数据每个月都在上升，大约有六七十篇的文章疑似使用AI的程度超过了50%。”《中华医学杂志》社有限责任公司新媒体部主任沈锡宾介绍。

沈锡宾向记者展示了检测过程：一篇论文经过检测系统后，会显示疑似AI生成占全文比重，相关疑似段落也会被标红。但记者注意到，和传统的查重报告单明确标注重复痕迹不同，AIGC检测报告单只是指出某些文本AIGC的“置信度”，并不能回答为什么是这个值。

“这使得报告单往往只起到参考和警示作用。”柯春晓说。

目前，人工智能大模型正在以“周”为单位进行迭代升级。如何适应不断升级的技术，是摆在AIGC检测工具面前的一道必答题。

作为使用者的人类本身也在不断“进化”。“类似人们逃避查重的方式，如果人们了解到AI检测的方式，也可以重新组织相关内容，对AI生成的文本进行人工润色。这样很可能就检测不出来了。”沈锡宾说。

作弊与反作弊的过程，实质上是场“猫鼠游戏”。只要技术不断升级，两者间的博弈就不会停止。目前，AIGC检测技术仍处在萌芽期。如何对AI生成的虚假图片、虚假数据进行识别仍是难点。因此，人们引入智能检测技术的同时，也要建立人工审查机制。

“审稿人要当好‘守门人’，发挥同行评议的作用，仔细甄别判断论文的数据是否和认知存在偏差。出版机构也可以要求作者提供原始数据，多管齐下，确保科研诚信。”沈锡宾说。

技术向善要他律更要自律

加强技术治理的同时，各方都在翘首以盼，期待达成某些共识以及相关政策尽快出台。“教育、科研、出版各方都很关注AIGC使用的边界，期待对合理使用AIGC形成一个共识性规范。”知网技术专家呼吁。

其实，早在去年初，中国科学技术信息研究所（以下简称中信所）就牵头爱思唯尔、施普林格·自然、约翰威立等国际知名出版集团和科研信息分析机构，在广泛调研并梳理业内相关研究和探索工作的基础上，完成了中英文版的《学术出版中AIGC使用边界指南》（以下简称《指南》），并于去年9月20日在国内外同步发布。

去年12月21日，科技部发布的《负责任研究行为规范指引（2023）》（以下简称《指引》）更是受到了业内的广泛关注。

《指引》和《指南》就如何负责任地使用AIGC，解答了令科研工作者、期刊编辑、审稿人困惑的一些问题。

首先是披露问题。《指引》提出，使用生成式人工智能生成的内容应明确标注并说明其生成过程，确保真实准确和尊重他人知识产权。《指南》中更是提供了声明的模板，供科研人员参考。

对于一些人想用AIGC投机取巧的行为，《指引》明确提出，不得使用AIGC直接生成申报材料；《指南》规定，AIGC不应该用来产生研究假设、直接撰写整篇论文文本、解释数据、得出研究结论。研究人员使用的数据必须是研究人员进行实验并收集所得，如使用AIGC提供的统计分析结果需进行验证。

随着AIGC的使用边界不断清晰，越来越多的出版机构达成共识，制定了使用规范。施普林格·自然集团发言人介绍说，他们目前已经明确了有关作者身份和图像方面的规定。例如，人工智能不能担任作者，真正作者如使用大语言模型须加以透明描述，AI生成的图像通常不能用于发表等。

“《科学》杂志在去年1月份发布的政策是禁止使用任何AIGC工具。而11月16日他们更新了投稿规则、放宽了限制，表示只要进行了适当披露，使用工具是可以接受的。”中信所博士郑雯雯说道。

“《指引》覆盖较为全面，对AIGC的使用总体呈现出平衡包容、敏捷治理的态度，而非一味禁止。这也说明治理的目的并不是阻止科研工作者使用新一代人工智能技术，而是让科研工作者能够负责任地去使用。”梁正提到，在政策制定的行为框架之下，还要关注学科差异问题。“使用AIGC可能因学科的不同而有所差异，其伦理问题也要根据学科特点细化。”

例如，在自然科学领域；AIGC的强大功能更多体现在数据处理领域，如果失范使用，往往难以发现。而对于人文社科领域，直接使用AIGC生成内容的痕迹非常容易被发现，尤其是在高水平的研究当中，优劣之分更为明显。

“因此，对于更加注重文字表达、数据资料支持的学科，比如企业管理、理工科、医学等，需要防范产生虚假的数据集或论证材料。”梁正说，“对AIGC使用的披露程度、疑似度的数据指标等，都需要学术共同体进一步探索，来推动形成广泛共识。”

此外，尽管国家出台了相应的规则，但从外部监督到行业自治还需要一个过程。AIGC的使用涉及包含研究人员、出版机构、相关行业组织、政府等方方面面。如何厘清各方关系，各司其职是关键。“简单说，就是出了问题，谁来查？有没有能力查？”郑雯雯强调。

记者了解到，中华医学会杂志社在今年1月9日公布了其对于AIGC技术使用的有关规定。其中不仅涉及了作者要遵守的细则，还提出了查处方式——经编辑部研判的违反AIGC使用的情形，将直接退稿或撤稿；情节严重者，将列入作者学术失信名单。

“我们下一步的目标是把存在问题的文章作一个归纳总结，进一步摸清AIGC使用的规律，为科学治理积累经验。”沈锡宾说。

“尽管新兴技术有着潜在风险，但也有着无可比拟的优势，不宜一味封堵，而是要做好引导、合理合规地使用新技术。”郑雯雯表示，归根到底，科学研究的主体是人。如果心中的那杆“秤”倾斜了，即使再完善的监管政策、再高端的检测技术，也难以抵挡学术不端的侵袭。

梁正也强调，作为科研诚信的第一责任人，科研人员一定要保持严谨的学术态度，关注研究领域的真问题，坚守学术研究的基本原则，如原创性和透明性；明确认识到ChatGPT等工具的潜在风险，避免使用不当而造成学术不端。

“科研诚信和伦理是科研的生命线，科研人员一定要存敬畏、有底线。一旦在这方面有瑕疵，职业生涯或将葬送。”梁正提醒。

（文中温睿、徐彬均为化名）

本版图片由视觉中国提供