

格言录

法贵止奸，不在过酷。

——(北魏)郭祚

出版机构应否禁止作者运用生成式人工智能

□ 李泊毅

此产生以下问题：是否所有作者运用生成式人工智能辅助写作的方式都不可被接受？如何有效规避作者不当利用生成式人工智能写作的风险？

二、对生成式人工智能辅助写作的担忧

总反对作者使用生成式人工智能辅助写作的理由，可大致将其归纳为如下三类：第一类是对质量的担忧。例如，认为生成式人工智能、大语言模型存在“黑箱”“不可解释”的问题，甚至会一本正经地胡说八道，编造不存在的内容（出现“幻觉”），人工智能生成合成内容的质量得不到保障，这种担忧部分是因为不了解生成式人工智能所致。第二类是对独创性和著作权的担忧。在著作权法中，对作品的定义是应当具备独创性，所谓的独创性大体上是指作品由作者独立完成并能体现出作者独特的智力判断和选择。人工智能生成合成作品是否具备独创性，以及体现的是哪个主体的独创性，都存在争议。因此，作者使用人工智能生成合成内容可能导致著作权争议。第三类是对创作伦理的担忧。例如，认为作者提交完全由人工智能生成合成的内容而不声明，属于不诚信行为，在学术领域甚至会被定义为学术不端。一种更深刻的担忧是，生成式人工智能将会使人类失去创作的积极性，甚至丧失写作能力，例如一些本应该具备写作能力的群体，如高等院校的研究生，将因为有捷径可走而缺乏写作训练。

一、问题的提出

笔者在参加某次出版专业技术人员继续教育时，授课老师就编辑对作者使用生成式人工智能辅助写作的态度进行了一次现场调查，结果是现场200余名编辑中的绝大多数人都对作者使用生成式人工智能应用辅助写作持不欢迎的态度。在生成式人工智能特别是大语言模型迅速发展但被广泛运用的当下，从上述调查中可以看出，编辑群体对作者使用生成式人工智能辅助写作充满担忧，这表明生成式人工智能的运用对出版工作者的心理或出版工作产生了一定的冲击，由

对生成式人工智能的担忧并不罕见，例如国家对生成式人工智能的态度是“坚持发展和安全并重，促进创新和依法治理相结合的原则，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管”（《生成式人工智能服务管理暂行办法》第三条），言下之意，既充分认识到潜在的风险，也肯定其积极意义，所以有“安全”“治理”“审慎”“监管”的表述，但也强调“发展”“促进创新”“鼓励”“包容”。

国家的态度表明，生成式人工智能本身是具有积极意义的技术手段，人类对相应原理下产出结果的运

其最终是发挥积极作用还是产生消极影响，关键在于使用者的利用方式。所以，应当引导使用者合理利用生成式人工智能，而非不问缘由不分青红皂白地一律禁止使用。

三、可以被允许的运用方式

从规范运用的立场出发，我们审视上述对生成式人工智能辅助写作的担忧，就可以发现许多担忧可能被过分放大了，以下详述。

(一) 质量问题

生成式人工智能、大语言模型确实存在“黑箱”“不可解释”的问题，以及存在会产生“幻觉”的问题——生成看似合理但实际错误的内容。但“黑箱”“不可解释”并不会导致人工智能生成合成内容质量的降低，相反，“黑箱”“不可解释”恰恰是生成式人工智能、大语言模型存在的原因。生成式人工智能的本质是一个函数（模型），但这个函数如大语言模型可能存在上亿个参数，如上文提及的，我国DeepSeek的参数量为6710亿个，如此巨大的参数量，使得模型的内部状态和决策过程极为复杂，想要逐一分析每个参数对输出结果的影响几乎是不可能的，这使得模型的决策过程很难被解释。而且，大语言模型通常采用深度神经网络架构，包含众多层次，如Transformer架构的模型有数十至数百个解码器层，每一层都对输入信息进行复杂的变换和提取，这些变换通过大量的矩阵运算和非线性激活函数实现，这使得各个神经元（简单函数）之间的相互作用错综复杂，难以直接追踪和理解信息是如何在不同层次之间传递和处理的。输出依赖于输入的上下文和模型内部的动态计算路径，相同的输入在不同情境下可能激活不同的参数组合，即模型的输出通常是基于概率分布的，具有一定的不确定性，导致结果难以稳定复现。这种决策过程难以被人类直观理解的情况，被称为算法的“黑箱”或“不可解释性”。本质上，“黑箱”或“不可解释性”恰恰是大语言模型强大能力的代价。某种原理难以被人类直观理解，并不影响人类对相应原理下产出结果的运

用，例如大部分人并不理解无线网络通信技术（Wi-Fi）的原理，但这丝毫不影响人们利用该技术接入无线网络。

当模型遇到超出训练数据范围的输入时，由于缺乏足够的相关数据来准确生成合理的输出，就会基于已有的不完整模式进行推测，从而产生“幻觉”。所以，“黑箱”或“不可解释性”与文本质量并不直接关联。人类作者也会编造不真实的内容，但我们绝不可能因为部分人存在编造行为，就概括性地排斥写作，我们反对的只可能是编造行为。同理，单纯以“幻觉”为由反对作者使用生成式人工智能应用辅助写作，反对理由成立的前提似乎只能是作者不加甄别地采纳上述内容，而不是使用生成式人工智能本身。

作者应当对自己作品的质量负责，编辑应当对作品的质量进行把关，作品质量不高的原因是作者不负责任、编辑没把好关，而非使用了某种工具。因此，“黑箱”“不可解释”“幻觉”问题似乎都不足以成为断然拒绝作者使用生成式人工智能辅助写作的理由。

(二) 独创性和著作权问题

人工智能生成合成内容是否可以被授予著作权，是存在争议的问题。目前比较主流的观点是，作品的作者必须是人类，只有人类创作者的贡献才能获得著作权，人工智能生成合成内容不能单独获得著作权保护。若某人使用人工智能辅助创作或作品中包含人工智能生成的内容，并不影响著作权保护资格，但申请人在申请著作权时须注明哪些部分由人工智能生成，并承认这些部分不受著作权保护。在大多数国家和地区，人工智能生成合成内容能否获得著作权保护，关键在于其是否含有自然人创作者的独创性智力投入，如果人工智能只是作为创作工具帮助创作者完成作品，且作品能够体现创作者的个性和审美选择，那么该作品具有获得著作权的资格，著作权归属于创作者。当然，不同国家甚至同一国家的不同机构对何谓“作者的独创性智力投入”会有不同理解，例如，美国版权局发布的《版权与人工智能报告》（Copyright and Artificial Intelligence Generated Content），因此被简称为AIGC。以大规模文本数据为基础，训练而来的生成式人工智能模型被称作大语言模型，大语言模型中的“大”是形容其巨大的参数量，例如，我国DeepSeek的参数量为6710亿个，GPT-4（OpenAI）据推測有超过5000亿个参数。

gence，以下简称《报告》）在可版权性这一部分（Part 2: Copyrightability）同样认为，完全由人工智能生成的内容因缺乏足够的人类控制（sufficient human control）不受版权（英美法系国家将著作权称为版权）保护，但利用人工智能辅助创作可获得版权，比如，人类通过修改、编排或输入原创性表达（如上传手绘草图并指认AI修改）主导创作过程，则人类贡献部分可享有版权。《报告》列举了3种典型的运用人工智能进行创作的场景，即仅输入提示词、输入原创性表达、对人工智能生成的内容进行二次创作（如编辑、修改），如上所述，对于后两者，《报告》都认为可以获得版权的保护——在输入原创性表达这一场景下，输出内容中可感知的人类表达部分受保护；在对人工智能生成的内容进行二次创作这一场景下，衍生作品的人类贡献部分可受保护，但保护范围不延伸至AI生成的原始内容。但对于仅输入提示词，《报告》则认为人类对创作过程的控制程度不足，向生成式人工智能提供提示词等同于提供“创意”，“创意”本身是不受保护的（unprotectable idea），然而，对于这一点，不同的法院存在不同的法律见解。

综上，关于独创性和著作权问题，我们可以归纳出最大公约数，即利用人工智能辅助写作，并非绝对不受著作权的保护，也就是说某些利用人工智能辅助写作的方式，是能够体现出作者的独创性的。尽管对于何谓“作者的独创性智力投入”有不同理解，但我们可以发现最没有争议的部分——对人工智能生成合成的内容进行编排、修改，输入原创性表达的部分具有独创性。所以，从独创性和著作权的角度审视，我们反对的应该只是完全由人工智能生成合成的内容以及独创性不足的内容。（如仅修改个别标点符号）

(三) 创作伦理问题

在创作伦理方面，对生成式人工智能运用的担忧，首先集中在诚信方面，如前所述，如果作者运用生成式人工智能的方式对创作过程的控制程度不足，那么会被认为对内容的独创性智力投入不够，无法获得版权，此时如果作者将这样的内容以自己为作者提交发表，在生成式人工智能所有方未明确将产生内容的著作权授予使用者的情况下，作者

和出版机构都面临著作权争议的风险，即便生成式人工智能所有方明确将著作权授予使用者（如ChatGPT），作者的行为仍存在抄袭与剽窃、未尊重他人劳动成果（未清晰界定贡献）的伦理问题，如果这种情况发生在学术出版的语境下，则涉嫌学术不端。因此，为避免诚信问题，作者应当注明哪些部分由人工智能生成，自己在创作过程中扮演了什么角色。

一种观点是生成式人工智能将会使人类失去创作的积极性，甚至丧失写作能力。如果人们过度依赖生成式人工智能，可以预见，部分人掌握语法、组织结构、整合信息的能力都可能被削弱，这就如同计算机削弱了人类的书写能力。但从历史上看，技术的进步未曾成为过人类创造力的桎梏，以人文社科领域为例，人类的创造力突出地表现为理解与同情，这恰恰是人工智能所不擅长的，因为人工智能掌握的是文本规律，而非真正理解文本后的意涵，无理解自然不存在基于理解的同情。换言之，人工智能无法取代人类核心的创造能力。如果人工智能的表达出现同质化，有创造性的人自然将追求与众不同的表达；如果表达不再成为问题，人们的思想或许更加深刻。

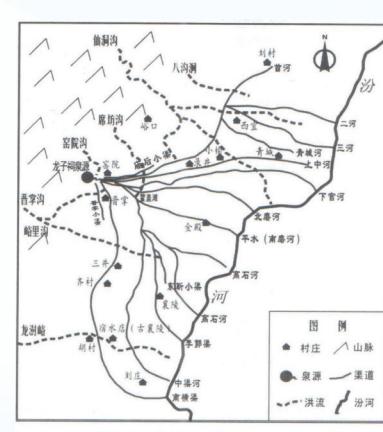
四、结论

制度伦理学上，好的制度让人做道德上对的事情，同时也是对人有利的事情。如果禁止作者使用生成式人工智能辅助写作，无疑将使作者陷入悖论，即如果选择对自己有利的事将被谴责，一旦违反规范的成本微低，人们就容易越界。从规范有使用价值但存在一定风险物品（如酒）的实践来看，最有效率的管理方式既不是完全放开，也不是绝对禁止，而是有条件地允许使用。借用王敏远教授的话，试图阻止现代科技在出版中的运用是徒劳的；不规范现代科技在出版中的应用是危险的。所以，与其玩猫捉老鼠的游戏，不如规范作者的使用方式，即请作者标注使用生成式人工智能的部分，并表明自己在其中扮演了何种角色，仅拒绝那些独创性智力投入不足的使用方式，这也与现行的规范保持了一致。如此，编辑人员恰可以将自己的有限精力投入到作者声明其具有原创性的部分。

（作者单位：《人民法院报》杂志社）

从龙子祠有关水利诉讼碑刻看“均利以息讼”与“以德化人”

□ 张奇平



息讼者，其惟今户部尚书范中奉公乎。……未几，多讼水利于公者。公既察前弊，复有言于公者曰：□□□常流之水自下而上，则上无邀决之患，若通利渠、丽泽渠是也。□源恭流之水自上而下，则无走泄之患，若麻柳洞、望州谷是也。诚能如此，则利均□□□讼水息矣。公瞿然，斯言得之矣。于是改纪上官河□□，大署文榜，溉田之令。自下而上，令行禁止，信赏必罚，民咸称便。……桐乡之于朱邑……，虽然昔者孔子适卫，欢其既庶且富，而必以教终之，尽使民知教则悌行于家，忠信著于乡，礼义廉耻之风震然成俗，……虽使之争而有不肖之者矣。夫子尝曰：听讼，吾犹人也，必也使无讼乎。……”

从以上碑文内容，有下面几点关于诉讼的启示。

当时，作为临汾重要的灌溉水渠“上官河”，取自龙子祠泉水，是一条人工开挖的灌溉河，牵扯到36村，田地两万多亩，上游之地以势之便，专利而多灌溉，下游之地，因地势之下，而不能保证按时灌溉浇地，往往误了农时，致粮食欠收，因之而起的争水而斗的事件时有发生，斗者百千人，有人因此而死者，数起斗死人命案，此种情况累积有数十代，成为

“习以为常”的事情。到了元至正二十六年范国英当政时，事情有了转机，通过官府订立新的渠规，颁布溉田之令，令行禁止，信赏必罚，民咸称便，从根本上杜绝了这一因水而争的诉讼之累讼，所确立的“自下而上、依次行浇”的渠规深入人心，沿用600多年，而造福当地百姓。

一是群体利益之争，均利以息讼。如灌溉水利诉讼，多是村庄之间的群体利益之争，牵扯到多数人群体的利益，均利就显得更为必要。所以，在处理这件数十年代的水利之讼时，当时的官员就确立了“均利息讼”的正确处理诉讼的指导原则，为积讼数十年的累讼找到一个好的处理办法。

二是注重田野调查，均利强弱。为处理好“多讼水利于公者”，时为负责案件的当事官员晋宁路总管范国英看现场听取意见。有人提出，“常流之水自下而上，则上无邀决之患”，“无源恭流之水自上而下，则无走泄之患”，这些话的意思是，有源泉的常流之河水，用水浇地时，自下而上，是从下边的田地开始，则上流没有决口的祸患。没有源泉的季节性洪水的涧河水，用水浇地时，从上而下，从上边的田地开始，就没有水走泄的忧虑。这

也符合不浪费水，提高灌溉效率的用水规律，让人容易接受。所以被到达现场的官员范国英采纳，“瞿然，斯言得之矣”，重要的是，“均利强弱”通过让处于下游的相对上游的弱势的用水者，先用水，以强其弱势，达到“均利”的效果，并通过官府行文成为定制，“大署文榜，溉田之令。自下而上，令行禁止，信赏必罚”，这种提高灌溉效率、均利强弱的用水渠规，破解了多年“水利不均”的顽疾，斗水事件得到了根治，达到了息讼的目的。从现有文献记载看，自此，上官河因争水而讼的大案子就很少了。

三是追求道德教化，必以教终之。从上述碑文看，最后讲到了“桐乡之于朱邑”“孔子适卫”“听讼，吾犹人也，必也使无讼乎”三个典故。

朱邑（公元前61年），西汉官员。朱邑初任桐乡啬夫（县以下基层政权组织乡的长官，掌管一乡的诉讼和赋税等），后任大司农，为朝廷重臣，他秉公办事、不贪钱财，以仁义之心广施于民，深受吏民的爱戴和尊敬。是地方官吏为政者的榜样。据《汉书·循吏传》记载，朱邑廉洁公平，不苛刻乡民，以守护百姓的利益为自己的职责。秉公执法，以德化人，从来就不因一己之怨而对老百姓动用刑罚。朱邑在主管乡里诉讼时，审理定

案时未以功之以刑而动之以情，注力调解劝导，感化顽劣；同时教民稼穑，倡导子弟读书。《汉书》记载其“廉平不苛，以爱利为行，未尝笞辱人，存问耆老孤寡，遇之有恩，所部吏民爱敬焉。”“为人淳厚，笃于故旧，然性公正，不可交以私。天子器之，朝廷敬焉。”“身为列卿，居处俭节”。苏轼的《浣溪沙·忆旧》词曰：仲卿终不避桐乡。讲的就是汉臣朱邑。明代诗人胡伊就在他的《七绝·题大司农朱公祠》中写道：“古木森森覆墓亭，大农祠在古桐城。后来为政应千辈，能使斯民有情。”碑文这里表达了地方官员为政追慕先贤的情操，重农宣教、清正廉洁、所居民富、所去见思的施政追求。“孔子适卫”是孔子周游列国期间的重要事件，核心内容为与卫将军文子关于礼治与刑罚的治国思想辩论，强调道德教化的作用，即碑文所述的“必以教终之，尽使民知教则悌行于家，忠信著于乡，礼义廉耻之风震然成俗。”“子曰，听讼，吾犹人也，必也使无讼乎。”意思是，历朝历代都有类似的规定。朱邑在主管乡里诉讼时，审理定

（作者单位：临汾铁路运输法院）