

■一家之言

人工智能大模型开源的版权之问

□辜凌云

人工智能大模型开源作为理论与实践相结合的新型范式,有效促进了算法的创新和优化,推动着人工智能技术的迭代升级。在此环境下,人工智能大模型的研发依托开源运动,不仅实现技术共享、创新和发展,还促进跨学科、跨领域的知识交流,为技术创新提供强有力支持,拓展了人工智能向前发展的边界。

然而,随着人工智能大模型开源的不断深入,开源许可证中的版权问题逐渐浮现并成为一个显著的发展瓶颈。一般而言,人工智能大模型开源中的版权问题主要表现为人工智能大模型学习语料的获取困境、人工智能大模型开源的版权许可政策以及人工智能大模型生成内容的版权属性判断问题。本文重点关注的是人工智能大模型学习语料获取的版权困境、开源许可证中的版权政策逻辑及带来的市场变化,以明晰大模型开源后的商业利益考量与商业选择策略,推动人工智能技术的广泛应用和负责任地发展。

获取学习语料易侵权

在开源运动的加持下,人工智能大模型的研发与应用获得了前所未有的推动力。作为以数据驱动的人工智能大模型,其研发与迭代升级无法脱离高质量的数据训练。这直接导致了人工智能大模型在获取学习语料时,需要在尽可能短的时间内多获取高质量数据源,尤其是包含人类智慧精华的作品数据。

然而,根据我国《生成式人工智能服务管理暂行办法》和《著作权法》规定,人工智能服务提供者应使用具有合法来源的数据及不得侵害他人依法享有的知识产权。实践中,以美国《纽约时报》诉OpenAI为例,原告《纽约时报》指控被告OpenAI公司擅自使用大量文章数据训练其人工智能大模型,侵犯了原告的版权。尽管该案暂未判决,

但人工智能服务提供者在未获取版权人授权的情况下使用作品数据,很容易构成侵权。

人工智能大模型学习语料的获取能否适用《著作权法》中的合理使用制度同样存在争议。具体而言,人工智能大模型的研发涉及对数据的收集获取、微调等训练过程,一旦在此过程中涉及收集使用与作品相关的数据,相应行为便存在侵权风险。与此同时,基于前端未获授权作品训练的人工智能大模型一经开源,大量代码研发者可以在原有训练基础上继续利用作品,以此迭代更新后续大模型,这种做法极易造成数据泄露等风险的扩大,影响技术的创新发展。

开源版权许可政策有差别

在人工智能大模型开源之后,相关主体往往在开源许可证中设置相应的许可政策以确保开源的目标得以顺利实现。一般而言,许可证中的许可政策是指对某种软件的使用、复制、修改和分发等方面的控制策略,它们定义了用户可以如何使用软件及其源代码。其中,开源人工智能大模型的版权许可政策集中体现为不同类型许可证在版权方面的管理宽松程度。

开源许可证一般分为两大类:宽松许可证(如MIT许可证或BSD许可证),以其对用户自由度的极大扩展而著称。这些许可证通常仅施加最低限度的限制,允许用户在保留版权声明等基本条款的前提下,自由地使用、修改及重新分发软件。相对地,严格许可证(如GPL许可证)除了提供自由的使用权,还要求用户在修改后的软件或衍生作品中,需以相同的许可证发布其源代码,这被称为“强制性开源”。

然而,人工智能大模型的开源并非完全将大模型归属于公共领域,实际上仍是基于商业利益考量作出的市场选择。以Llama许可证为例,Meta公司在



资料图片

将人工智能大模型开源的同时,通过Llama许可证设置相应的开源要求。其中,Llama3.1许可证不仅要求对“使用Llama材料或Llama材料的任何输出或结果”标注来源,同时还明确了严格的知识产权诉讼条款。这种严格且带着商业利益的许可证规定同样见于苹果所开源的7B模型之中,其中apple-ascl许可证更是明确“未经Apple事先书面许可,不得使用相关的知识产权标记”,并规定免责条款以规避风险。

以上限制均意味着人工智能大模型的开源并非纯粹地进入公有领域,而是基于商业利益衡量后的一场市场化行动。

版权授权难题何解

进入智能开源时代,人工智能大模型的开源运动为学术研究与企业实践带来了前所未有的机遇。然而,随之而来的是版权问题的复杂性和挑战性。对这一难题的化解,不仅需要我们从综合考虑到开源社区的技术开放和商业化的双重需求,而且需要用更加灵活的措施以适应版权制度的发展。

一是研发主体应尽量在事前获取版权许可,巧用技术手段规避侵权风险。

人工智能大模型研发主体需要尽量在事前获取相应的训练数据版权许可或者通过集体管理组织获得版权授权。此外,利用技术手段对数据集进行版权检测,识别并排除可能侵犯版权的内容。同时,为了减少版权风险,可以通过技术手段对所使用的数据进行清洗和脱敏,移除版权保护的内容,确保在不侵犯版权的前提下使用数据。

二是创新合理使用制度,促进人工智能技术发展。一方面,通过创新《著作权法》合理使用制度,在条件成熟之时,针对人工智能大模型的数据训练提出灵活的适用性条款,从而适应技术的发展,保护创作者的合法权益。另一方面,通过法律解释对特定行为进行个案认定,更好地理解人工智能技术的特点和需求,从而制定出更加合理的法律规则,促进人工智能技术的发展。

三是提高风险防范意识,实现开源知识共享与商业使用的利益平衡。针对人工智能大模型开源的版权许可问题,开源可以促进技术的开放和共享,促进创新,但对开发者来说,开源也可能涉及放弃部分商业化的机会。因此,在选择开源的同时,需要在保持技术开放性和保持商业利益之间找到一个平衡点。对于人工智能大模型使用者而言,在选择相应大模型之前,应谨慎阅读与开源相关的开源许可证规定,尤其是其中的版权许可内容,通过事前明晰条款内容,以避免后续侵权风险的发生。

科技浪潮汹涌,人工智能大模型的开源已是不可逆转的趋势,这不仅加速了技术的发展,也极大地促进了学术界和产业界的合作。人工智能大模型开源中的版权之问作为科技与法律互动协调的问题,理应在技术商业化与知识共享之间寻求问题的解决之道,为大模型的健康发展找到合适的路径,以期最终实现人工智能技术的广泛应用和社会价值的最大化。

(作者单位:北京大学法学院)

■海外速览

国际奥委会强力阻止奥运盗版

众所周知,国际奥林匹克委员会(以下简称国际奥委会)非常重视其对知识产权的保护。在过去的一年时间里,国际奥委会已经向谷歌发送了数千份《数字千年版权法》(DMCA)删除通知。这些通知显示,洪流网站已成为过去的问题,直播网站才是目前的主要关注点。

使用奥运五环的图像,甚至只使用“Olympics”一词都可能涉及侵权问题,尤其是在商业环境中使用时。然而,这其中最有价值的是国际奥委会的广播权。由于事关数十亿美元的利益,国际奥委会及其许可合作伙伴正在竭尽全力防止人们在未经许可的情况下收看他们的赛事。

但是时至今日,奥运盗版仍然是一个棘手的问题,国际奥委会目前将直播门户网站视为当前的主要威胁。不久前,国际奥委会与其他委员会一起在巴黎司法法院获得了网站屏蔽令。该命令要求大型法国互联网服务提供商阻止对几个盗版流媒体网站和网络电视(IPTV)提供者的访问。

虽然屏蔽令目前有效,但它目前仅限于25个域名,并且仅适用于法国。不过,国际奥委会并没有就此止步,它还邀请了反盗版合作伙伴Friend MTS发出更传统的DMCA删除请求。

这些DMCA工作大多是在公众视线之外进行的。在谷歌的透明度报告和Lumen数据库的帮助下,人们仍然可以发现这些删除行动的痕迹。

公开数据显示,国际奥委会已经发送了大量删除通知,共同针对了数千个统一资源定位器(URL)。这些通知要求谷歌从其搜索引擎中删除这些链接,以使公众更难找到盗版的奥运会直播流媒体。国际奥委会表示:“请注意,国际奥委会是奥运会和奥林匹克财产的所有权利的所有者,特别是为2024年巴黎夏季奥运会制作的视听内容的权利……未经国际奥委会事先明确书面批准,不得通过互联网或任何其他互动媒体、电子媒体传输或传播此类奥运内容,而根据我们的记录,国际奥委会尚未授予你们相关权利。”

据了解,截至目前,国际奥委会已要求谷歌从其搜索引擎中删除了5907个URL。这份名单上主要是盗版流媒体网站。

(中国保护知识产权网)

浙江省杭州市上城区持续深化“护苗”行动,探索多元活动形式与内容——

暑期“护苗”有新意 绿色守护有心意

□杭州市上城区“扫黄打非”办公室

前不久,一场内容丰富的“护苗·绿书签”市集暨盛夏“绿书签”主题游园活动在杭州图书馆举办,该活动吸引了上百名少年儿童参与。活动由杭州市上城区新闻出版局、上城区四季青街道联合主办,旨在以“绿书签”主题活动为载体,倡导尊重和版权保护的理念。

“护苗·绿书签”市集设置了多个精彩活动区,包括“绿书签”主题画展、NPC剧情体验、“绿书签”趣味拼图、诗词接龙、绘本演出等。这些活动趣味十足,富有创意,寓教于乐中向广大青少年传播“护苗”知识以及《未成年人保护法》等相关内容。

此外,杭州图书馆还特设了“修补时光 惜书爱书”环节,引导青少年读者树立爱书护书意识。活动中,孩子们了解图书出版、装订知识,学习图书修补方法,亲手让旧书换新颜,成就满满。

上城区“扫黄打非”办公室相关负责人介绍,暑假期间,为了给广大青少年营造更加健康、阳光、绿色的成长环境,全区积极响应“护苗”行动号召,开展了一系列丰富多彩的暑期“护苗·绿书签”活动,通过常规动作扎实执行与创新活动有效开展,取得了显著的宣传成效。

活动丰富,形式多样有特色

为解决儿童“假期无处去,家长看护



“护苗·绿书签”市集暨盛夏“绿书签”主题游园活动人气爆棚。

难”问题,笕桥街道组织开展了暑期夏令营。当“绿书签”牵手夏令营,一场场“护苗”阅读分享会、红色故事小课堂火热开展,让孩子们在学中玩,玩中学。

采荷街道开展了一系列寓教于乐的“护苗”活动,如“扫黄打非”主题剪纸、好书共读会、“绿书签”衍纸艺术创作、“护苗”环保袋DIY等,不仅让孩子们暑期生活更加丰富多彩,也让他们在轻松愉快的氛围中,潜移默化地受到教育。

围绕绿色阅读、文明上网、尊重知识产权、抵制盗版等内容,南星街道开展了“书香润童心·护苗共成长”活动,老师向孩子们宣传“护苗”“绿书签”知识,并教孩子制作手工树。大家纷纷表示要多读书、读好书,文明上网、拒绝非法出版物及有害信息。南星街道“扫黄打非”办公室相关负责人说:“希望通过活动培养孩子的绿色阅读习惯,知道版权保护的重要性,也呼吁家长要重视对孩子阅读能力的培养,了解‘护苗·绿书签’行动,为孩子的健康成长保驾护航。”

彭埠街道为引导未成年人科学合理使用网络,提高网络安全意识和自我保护能力,通过案例讲解、互动问答等方式,向参加活动的孩子及家长讲解沉迷网络的危害及如何防范网络沉迷,现场还发放了《致家长的一封信》、“绿书签”等宣传材料,倡导绿色阅读、文明上网,呼吁家长加强陪伴、以身作则,



上城区望江街道举办“绘”出童年“画”出梦想”活动,孩子们用奇思妙想作画。



上城区笕桥街道开展了“绿书签—衍纸艺术创作”活动,丰富青少年的暑期文化生活。

护助未成年人健康成长。

有效引导,动手动脑长本领

闻弄口街道依托濮家社区“护苗”教育服务基地,开展以“闻梦启航·时代少年向阳夏令营——护苗绿书签”为主题的团扇制作、阅读分享、绘画、展览等系列宣传教育活动。孩子们纷纷表示:“绿书签”不仅仅是书签,还可以时刻提醒自己远离非法出版物。通过系列活动,孩子们了解非法出版物的危害和“绿书签”行动的意义,明白要拒绝有害内容和不良信息,争做绿色阅读的倡导者和践行者。此外,活动还充分发挥“小手拉大手”作用,鼓励孩子们向父母、朋友宣传“护苗·绿书签”行动,积极营造良好文化氛围。

凯旋街道依托景华社区新时代文明实践站,开展了“护苗手绘帆布袋 童心童聚绘清风”活动。志愿者用通俗易懂的语言向学生介绍什么是“绿书签”、如何辨别正版图书和盗版图书,引导学生热爱读书、快乐阅读。活动中,同学们用画笔在帆布袋上绘制出文明阅读、抵制盗版等和“护苗”行动相关的图案。一个个平平无奇的帆布袋,摇身一变

成了“护苗”小书包。通过活动,孩子们不仅丰富了暑期文化生活,提高了动手能力、创作能力,还增强了辨别和抵制不良信息的能力,提高了自我保护意识。

彭埠街道为了让孩子和家长更直观感受正版图书和盗版图书的区别,工作人员将盗版书带到暑期课堂,向大家讲解通过书籍外观、装订、纸张、色彩、价格等鉴别书籍真伪的方法,阅读盗版书籍的危害,倡导大家通过正规书店、网站购买书籍,自觉抵制盗版书籍和不良信息,筑牢自我保护安全防线。

丁兰街道联合上城区图书馆开展“书香润夏·阅暖上城”活动,倡导拒绝盗版、绿色阅读,让文化惠及民生。社区工作人员向家长和孩子开展“护苗”宣传,引导未成年人提高自我保护意识,为他们健康成长营造绿色的阅读环境,倡导积极向上的社会新风尚,居民连连点赞。

加强监管,清明市场共守护

笕桥街道利用街道、社区的微信公众号、微信群转发关于“绿书签”“文明上网”等内容的宣传视频,向大家传播绿色阅读及文明上网的理念,呼吁自觉

抵制非法出版物和有害信息。同时,对学校周边打印店、文具店等文化经营场所展开“扫黄打非”专项检查,严厉打击各类侵权盗版、违规印刷等行为;向商家宣传文明、诚信、守法的经营理念,督促商家规范、守法经营,以实际行动营造风清气正的文化环境。

闻弄口街道在暑期开展联合执法行动,针对辖区文具店、文印店、网吧等经营场所进行排查,重点检查是否销售含有淫秽色情、暴力恐怖、封建迷信等内容的非法出版物。同时,在检查过程中宣传文明经营、守法经营理念,提高广大群众的自觉性。

值得一提的是,采荷街道有一支“小荷苗”观察团。“小荷苗”都是来自社区的少年儿童,担任文化环境小小守护者,以小主人翁的姿态对校园周边文化市场进行巡查。

下一步,上城区将持续深化“护苗”活动,探索多元活动形式与内容,吸引更多青少年儿童参与,让他们成为文化传播的参与者和推动者。同时,上城区将加强学校、家庭及社会各界的联动,携手共建一个健康、和谐、绿色的文化环境,为青少年的健康成长保驾护航。