

## 生成式人工智能训练数据陷入版权争端

## “投喂”大模型如何规范授权

◎ 实习记者 吴叶凡

近日,美国媒体《纽约时报》把OpenAI及其投资方微软公司告上法庭,指控二者未经授权就使用该媒体的数百万篇文章来训练人工智能大模型,要求被告销毁相关数据并对媒体损失负责。今年1月,OpenAI对此做出答辩,称其训练是合理使用,且它们已提供了退出的选择。

这一争端引发了公众对于大模型训练数据版权的关注。我国法律如何看待大模型训练数据的版权情况,如何对大模型使用数据进行有效治理?2月初,记者采访了相关专家。

## 训练数据面临较高法律风险

大模型的训练数据究竟是哪儿来的?  
去年,OpenAI首席执行官萨姆·奥尔特曼接受采访时表示,他们花费了大量的精力整合不同来源的数据,包括开源信息数据库、通过合作获得的数据以及互联网数据。但对于具体数据集的来源和细节,OpenAI尚未公开发布。

北京交通大学法学院副院长郑飞告诉记者,生成式人工智能的训练数据来源可以分为外界生产数据和自生产数据。其中,外界生产数据来源包括公共数据、数字图书馆、信息库、网络信息等,来源方式包括自行收集、公共下载、第三方购买、爬取、模拟生产等。自生产数据则来源于生成式人工智能应用时产生的相关数据。

郑飞指出,使用外界生产的数据通常面临着较高的法律风险。目前,OpenAI已经被多次指控侵犯著作权。早在去年9月,美国作家协会就组织包括电视剧《权力的游戏》原著作者在内的17位作家向法院提起诉讼,指控OpenAI在未经许可的情况下批量复制了他们受版权保护的作品。

外界生产的数据中,还有一类是开源数据。那么使用开源数据是否意味着可以规避法律风险?

“开源通常意味着免费,但免费并不意味着可以随意使用。”在郑飞看来,开源数据并不意味着可以完全规避风险。“许多开源数据虽然不存在著作权财产权保护问题,但会涉及署名权、修改权等人身权问题。以开源软件为例,所有的开源许可证均要求保留版权声明,在版权声明中列明开源软件的名称、作者或版权所有者的姓名或名称,以表明其身份。”郑飞说。

郑飞进一步解释,按照我国著作权法的相关条款,如果使用者在使用开源软件时不保留版权声明,不表明作者身份,违反许可证要求,就可能侵犯开源软件权利人的署名权。

北京智源人工智能研究院副院长兼总工程师林咏华也曾表示:“用于AI大模型训练的开源数据必须是合法地从公开或可公开获得的资源中收集的数据。”

## 侵权认定存在难点

针对生成式人工智能带来的法律风险,各国都在陆续出台、完善相关的政策法规。我国在去年7月公布的《生成式人工智能服务管理暂行办法》中,明确提到生成式人工智能服务提供者应当依法开展预训练、优化训练等训练数据处理活动,使用具有合法来源的数据和基础模型;对于其中涉及知识产权的,不得侵害他人依法享有的知识产权。

## 银发族网络“冲浪”越发便捷,但也遇到维权难等问题——

## 数字技术适老化要速度也要温度

◎ 实习记者 薛岩

“你看,这是朋友过生日的照片,我在照片上加上文字祝她生日快乐;这个视频是我用趣味运动会的照片做的……”84岁的薛冰玉现居北京。作为银发族中酷爱网上冲浪的一员,她经常借助手机App解锁新功能。

当前,互联网、智能手机等新技术、新载体向老年群体的渗透率正在不断提升。银发族对这些新技术的使用情况如何,怎样让新媒体技术真正惠及他们?近日,记者对部分老年人和相关专家进行了采访。

## 助力银发族无障碍“触网”

近年来,我国加快推动数字技术适老化改造。为进一步降低老年人便捷上网门槛,2023年12月,工业和信息化部出台《促进数字技术适老化高质量发展工作方案》。该方案立足老年人的实际需求,提出支持老年人常用的手机App及小程序积极参与改造,指导企业加快适老化关键技术科研攻关,简化语音助手、长辈模式等功能的开启方式。

数据显示,工业和信息化部已完成了1735家网站和App的适老化及无障碍改造。这为老年人进一步使用移动互联网技术提供了帮助。

除国家政策层面的大力支持外,全国

各地也积极通过多种形式助力银发一族无障碍“触网”。

“退休之后,我经常参加单位组织的手机培训课。购物、打车、查地图……你想学什么,他们就会教什么,很方便。”薛冰玉说。

记者调查发现,北京有多个社区以月为周期,组织开展老年人智能手机科普教学活动。这些活动邀请大学生志愿者为使用手机零基础的老年人提供“面对面、手把手”的帮扶指导。

苏州、厦门、衢州等城市制作“银龄数字课堂”相关视频。视频还原老年人使用手机时可能遇到问题的实际场景,并给出权威解答。

广东省要求省内老年大学将智慧养老课程纳入教学范围,同时对老年学员使用人工智能、虚拟现实等新技术提供体验机会,以此进一步弥合数字鸿沟。

## 老人数字权益更易受到伤害

尽管不少App已通过推出大字体、大图标、精简功能操作等方式进行适老化改造,但记者走访时,仍多次听到不少老年人抱怨自己糟糕的手机使用体验,如广告多、内容质量低、恶意安装程序、找不到老年模式切换入口等。

此外,老年人使用智能手机,个人权益难以保障是突出问题之一。

近年来,部分App违规收集个人信息,过度索取权限等行为时有发生。有报告显示,

七成App会在非必要的情况下获取用户隐私权限。老年用户由于缺乏信息辨别能力,很容易默许App提出的所有要求,对平台收集信息侵犯个人权利的行为也全然不知。

记者在采访中也发现,面对平台强制性收取个人信息的行为,老年人还未形成明显的权益保护意识。当下使用App,面对为何要注册、个人数据是否会被收集、被谁收集、作何用途这样的提问时,老年人往往有些招架不住,无所适从。

算法推荐服务的广泛应用也让不善于使用智能技术的老年人难以识别网络套路。相较之下,老年人的数字权益更易受到侵害。

“老年人由于缺乏个人信息保护意识,信息辨别能力较弱,更易陷入‘标题党’和‘养生保健党’等虚假信息构建的网络信息场。他们进而被诱导充值或是购买理财产品、保险产品,最终导致网络维权困难。”西南交通大学国际老龄科学研究院副院长、教授杨一帆说。

## 适老化改造无法一蹴而就

面对当前老年人“触网”中存在的问题,专家认为,数字适老化改造不能仅停留在对通用版本的修修补补,因为这种内嵌改造模式并不能彻底清除数字鸿沟。“还要增加精神性关照,给予更便捷的信息获取方式等。”杨一帆强调。数字适老化改造,既要追求速度,也要保证温度。

记者调查发现,近年来,很多互联网

现有侵权责任法,区块链、元宇宙也普遍面临新的侵权责任问题。“这一解决途径也存在问题。因为专门立法周期较长,难以适配日新月异的技术发展速度。”郑飞说。

二是创设单行监管条例。郑飞介绍,国家网信办自创设以来,就承担着监管职能,并不断根据技术发展动向及时进行单行条例立法。“特别是近年来在互联网算法、深度合成、推荐算法等方面都发布了不同程度的监管条例,为互联网行业合规经营设置了主要依据。”

其他方式还包括,在已有的侵权责任法律体系中添加关于人工智能责任的相关条款,为人工智能设计者和提供者添加特殊的条款来进行强调和补足;对现在已有的条款进行解释等。“就AI技术的发展水平及其当下立法技术成熟度而言,采取‘传统法律修正’模式是一种可行的方式。”郑飞说。

需要注意的是,对于新兴技术,法律治理的目的并不在于“禁止”,而是在于推动技术的合规发展、合法使用。北京大学法学院教授张平曾指出,我国目前的生成式人工智能技术创新还处在初级阶段,法律法规的制定应当给科技创新留有一定的发展空间,需要采取开放包容的规范原则。

大模型想要更“聪明”,就必须通过大量数据来强化学习。因此,如何平衡各方利益、寻求合作共赢是关键。正如郑飞所说:“人工智能训练数据合法性问题,本质上是个人利益与公共利益冲突的体现。如果缺少利益平衡原则,在利益分成时容易产生分歧。”

郑飞提出,版权集团或版权的集体管理组织可以通过集体授权的方式有效解决训练数据的权利许可问题,也可以通过知识共享许可协议搭建开源数据库,为大模型训练方提供权利许可便利。大模型训练方则可以为版权方提供更加优质、低费用的生成式人工智能接入服务,推动出版行业升级。

目前,大模型方正在与出版行业积极寻求合作。有消息称,OpenAI正在与数十家出版商洽谈内容授权协议。去年12月,OpenAI宣布与德国媒体巨头阿克塞尔·施普林格达成了“里程碑式”合作。根据协议,OpenAI将付费使用施普林格旗下出版物的内容,施普林格将提供其媒体品牌的内容,作为OpenAI大型语言模型的训练数据。

平台探索优化供给侧方式,针对银发人群需求推出个性化定制功能。比如,淘宝老年版在搜索页面新增了老年人感兴趣的“营养保健、居家养老”等品类推荐,页面底部还用橙色加粗字体凸出“按住说话搜一搜”语音功能。饿了么长辈模式增加“去买菜”“去买药”“逛超市”的醒目图标。为防止老年人过度上网,抖音老年版开通“时间管理”“时间锁”功能,对观看较久的老年用户进行提示和强制打断。

想要满足老年人日益强烈的融入智能技术新生活的需求,不仅需要他们自己积极参与和主动学习,还需要相关部门、企业和家庭等多方面的合力配合。

“数字技术适老化本身是一个逐步完善和均衡迭代的进程,其改造利在长远,不可一蹴而就。我们应该以积极的态度看待这个过程,让老年人跟上赛博时代的生活节奏。这也是全社会共同的责任。”杨一帆说。

今年1月,国务院办公厅印发《关于发展银发经济增进老年人福祉的意见》,其中第二十五条明确,依法规范智能化产品和服务中的个人信息收集、使用等活动,加强技术监测和监督检查,做好安全防护和风险控制。

杨一帆建议,数字技术适老化改造应该在广告投放与控制、信息和财产安全保障、防“忽悠”和防诈骗方面多做文章,建立起围绕老年群体的信息监管和保障体系,多措并举,撑起老年人权益的“保护伞”。

## 热点追踪

《关于加快应急机器人发展的指导意见》印发  
以科技筑起安全“防火墙”

◎ 本报记者 陆成宽

近日,应急管理部、工业和信息化部联合印发《关于加快应急机器人发展的指导意见》(以下简称《指导意见》),加快推动应急机器人技术发展与应用,推进应急管理体系和能力现代化。

为什么要加快推进应急机器人发展,《指导意见》对应急机器人的发展作了怎样的部署,怎样落实好《指导意见》,加强应急机器人的示范应用?就此,科技日报记者采访了相关专家。

## 促进应急机器人加速落地

应急机器人是在安全生产和防灾减灾救灾过程中,执行监测预警、搜索救援、通信指挥、后勤保障、生产作业等任务,能够实现半自主或全自主控制,部分替代或完全替代人类工作的智能机器人系统的总称。

“近年来,我国应急机器人处于快速启动和发展阶段,未来前景广阔。但总体而言还在应用初期,真正落地的应用并不是很多,尚未大规模普及应用。”中国科学院大学应急管理科学与工程学院院长李颖说,《指导意见》的发布必将有力地促进应急机器人的加速落地。

国家安全生产应急救援中心技术装备部主任吴三海也有类似看法。他说:“在安全生产应急救援中推广使用应急机器人很有必要。《指导意见》的出台,为我们安全生产应急救援领域的装备发展提供了重要指南,将有力促进高危行业领域应急救援机器人的广泛应用和发展。”

李颖从需求和供给两个层面分析了加快推进应急机器人发展的必要性。她认为,一方面,应急科技支撑能力有了新要求,急需加强应急机器人等智能装备的发展;另一方面,目前我国安全应急装备多集中在通用领域和低端市场,急需着力推动以应急机器人为代表的先进技术创新和成果转化,突破共性关键技术,增强产品供给,加快推广应用,打造应急机器人体系。

吴三海认为,在安全生产应急救援领域推广使用应急机器人是推动实施科技兴安的重要手段,是确保一线救援人员科学救援、高效救援、安全救援的有效途径。同时,应急机器人代表了应急管理装备现代化发展趋势,是衡量应急管理体系和能力现代化的重要标志。在安全生产应急救援领域推广使用应急机器人,可以有效提升整个行业装备专业化、精细化、科技化、智能化水平。

## 揭榜攻关破解应急救援技术瓶颈

《指导意见》主要任务中的第一项是加强应急机器人急需技术攻关。近年来,国家安全生产应急救援中心始终高度重视科技在应急救援中的支撑保障作用。

国家安全生产应急救援中心发布了6个攻关方向,具体的攻关项目包括可实现矿井水害事故水下环境勘察、被困人员搜索、定位的水下环境探测机器人,解决矿山(隧道)坍塌事故处理时无法对被困人员精准定位问题的生命探测仪等。

“我们通过揭榜攻关,调动社会优势力量投入资金协同攻关破解安全生产应急救援瓶颈技术、‘卡脖子’问题,促进安全生产应急救援领域装备科技化、智能化。”吴三海说。

李颖表示,我国是一个多种灾害频发的国家,应急机器人的发展要坚持系统观念全面发展。目前有一些重点领域急需优先发展。

“一是优先发展‘三断’(断网、断电、断路)等极端条件下提供紧急应急通信、灾情侦察、生命通道构建等最基础支持、具有自适应能力的应急机器人;二是优先发展灾害多发易发领域的应急机器人和高危、恶劣环境的应急机器人,如城市消防、抗洪抢险、地震救援、森林草原火灾救援、危化品救援、煤矿救援、电力救援等领域的应急机器人。”李颖说。

同时,李颖指出,以前应急机器人的发展重点主要集中在救援环节,现在需要加快监测预警机器人的发展,推动安全治理关口前移,防范化解重大安全风险,提升防灾减灾支撑能力。

## 加大应急机器人推广应用力度

近年来,国家安全生产应急救援中心通过开展国家安全生产应急救援队伍装备项目建设,针对各类事故特点,为国家安全生产应急救援队伍配备了各类救援机器人装备。

《指导意见》提出,到2025年,研发一批先进应急机器人,大幅提升科学化、专业化、精细化和智能化水平;建设一批重点场景应急机器人实战测试和示范应用基地,逐步完善发展生态体系;应急机器人配备力度持续增强,装备体系基本构建,实战应用及支撑水平全面提升。

对此,吴三海表示,下一步,国家安全生产应急救援中心将认真贯彻落实《指导意见》有关要求,一方面,针对水下环境探测机器人等揭榜攻关成果,开展联合测试、演练等,加大力度进行推广应用。例如通过评审验收的攻关成果集结成册、开展试点应用、支持申报有关科技奖励等。

与此同时,应急管理部 and 财政部联合印发了《安全生产应急救援力量建设总体方案》,中央财政将连续5年安排补助资金,购置包含应急机器人等先进智能化装备在内的国家重要特殊救援装备。

“我们将综合考虑各地实际情况、事故灾害类型和国家安全生产应急救援队伍能力,为国家安全生产应急救援队伍配备一批矿山救援、危化救援、灾情侦察、应急医疗机器人等,进一步提升我国应对重特大、复杂生产安全事故灾难的救援能力,为社会稳定和经济安全发展提供坚实的保障。”吴三海说。



在2023世界设计之都大会上,开展馆的生成式人工智能展区受到众多观众、业内人士及媒体的关注。视觉中国供图



图为2023世界机器人大会现场展出的应急救援机器人。新华社记者 张晨霖摄