

强化全流程管理 引导技术向善

—《人工智能生成合成内容标识办法》解读

据新华社 近日,国家互联网信息办公室、工业和信息化部、公安部、国家广播电影电视总局联合发布《人工智能生成合成内容标识办法》。办法聚焦人工智能“生成合成内容标识”关键点,通过标识提醒用户辨别虚假信息,明确相关服务主体的标识责任义务,规范内容制作、传播各环节标识行为,将于2025年9月1日起施行。

专家认为,办法是我国推进人工智能领域安全治理、促进产业规范健康发展、引导技术向善的重要举措,标志着我国在生成式人工智能领域迈出了构建安全可信生态的关键一步。

近年来,生成式人工智能、深度合成等新技术快速发展,为生成合成文本、图片、音频、视频等信息提供了便利工具。

“然而,人工智能生成合成内容日益逼真,也催生虚假消息传播、身份信息冒充、恶意内容生成等新型安全风险,并削弱着公众对网络传播内容的信任根基。”中国工程院院士、浙江大学教授陈纯认为,面对人工智能安全治理这一世界性难题,四部门适时出台《人工

智能生成合成内容标识办法》,配套强制性国家标准《网络安全技术 人工智能生成合成内容标识方法》也同时发布,成为保障人工智能时代网络安全有序的关键手段。

国家网信办有关负责人表示,办法重点解决“哪些是生成的”“谁生成的”“从哪里生成的”等问题,推动由生成到传播各环节的全流程安全管理,力争打造可信赖的人工智能技术。

——针对服务提供者,办法明确应当对文本、音频、图片、视频、虚拟场景等生成合成内容添加显式标识,在提供生成合成内容下载、复制、导出等功能时,应当确保文件中含有满足要求的显式标识;应当在生成合成内容的文件元数据中添加隐式标识。

——针对互联网应用程序分发平台,办法提出在应用程序上架或者上线审核时,应当要求互联网应用程序服务提供者说明是否提供人工智能生成合成服务,并核验其生成合成内容标识相关材料。

——针对用户,办法要求使用网络

信息内容传播服务发布生成合成内容的,应当主动声明并使用服务提供者提供的标识功能进行标识。

此外,任何组织和个人不得恶意删除、篡改、伪造、隐匿办法规定的生成合成内容标识,不得为他人实施上述恶意行为提供工具或者服务,不得通过不正当标识手段损害他人合法权益。

“此前出台的《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》提出了标识有关要求,办法进一步细化了标识的具体实施规范。”北京航空航天大学法学院副教授赵精武认为,办法明确了生成合成内容制作传播各主体的责任义务,用户能够清晰识别人工智能生成的内容,减轻人工智能生成合成技术滥用危害,防范利用生成合成内容传播虚假信息、实施电信诈骗等风险行为。

一方面,清晰的内容标识有助于提升信息透明度,增强用户的知情权和选择权,培养公众对人工智能技术的理性认知。另一方面,对于服务提供者及传

播平台而言,内容标识制度既是责任更是机遇,在鼓励企业追求技术能力提升的同时,也将提升对产品社会影响的关注度,推动整个行业向更加规范、健康的方向发展。

良法善治,重在实施。专家认为,办法的落地需要各方协同配合,推动标识工作行稳致远。

公安部第三研究所副所长金波说,伴随标识管理与算法备案、安全评估等机制逐步实现有机衔接,生成合成内容标识合规或将成为相关部门开展人工智能监督检查、专项行动的重点关注领域。在此进程中,如何平衡好发展与安全、创新与责任,提升执法的专业化、精细化、智慧化水平,从而培育出安全、开放、公平的人工智能产业生态环境,还需要深入探究。

“此外,‘人的因素’应融入人工智能标识管理的全过程。”金波说,要着重提升公众对信息内容真实性、来源可追溯性的评估能力,积极培育公众的人工智能素养,确保人工智能技术成果普惠共享。

新华社记者 王思北

给商家打差评反惹官司,消费者的“差评权”咋维护?

据新华社 消费之后,在社交平台发布对商家“种草”或“避雷”的点评,是近年来兴起的消费新风潮。然而,部分商家认为负面评价损害了自身的名誉权,由此引发一系列官司。

消费者有没有权利在网上打差评?如果发生纠纷怎么维权?今年的“3·15”国际消费者权益日,重庆市第五中级人民法院披露了一起由差评引发的网络侵权责任案例,可以给消费者一些启发。

给商家打差评却坐上被告席

案件发生在重庆市綦江区。2023年8月,张某获得保险公司免费赠送的车辆保养服务,前往当地一家修理厂保养车辆。保养中,修理厂工人用力将机油尺把手抽出,导致机油尺刻度条脱落落在发动机油底壳里。工人告知张某,拆除油底壳取出机油尺刻度条需要收取费用。在张某未明确表示认可费用标准情况下,工人清洗了油底壳,安装了新的机油尺,对车辆进行了保养。

1天后,张某到修理厂取车,被告知需付320元。由于对费用存在异议,双方发生争执并报警。公安机关出警后组织双方进行协商,张某支付300元后开车离开。次日,张某在社交平台上发布了3条视频,描述了争议发生的过程,并对修理厂的行为作出批评。该修理厂遂以名誉受到侵害为由向法院提起诉讼。

“我将自己的真实经历发布在网络平台的个人账号里,并未歪曲事实,而且发布不久后我就已将这3条视频删除,客观上并没有对原告造成任何损失。”被告席上的张某十分委屈。

原告修理厂负责人则认为,张某发布的视频严重损害了修理厂的名誉和生意,其在前期试图与张某协商,未果,



理性对待“差评权” 新华社发 尚海春 作

不得已才起诉到法院。修理厂提出的诉讼请求包括要求张某停止名誉损害行为,公开道歉,赔偿经济损失2万元。

是合理批评还是恶意诽谤?

近年来,消费者与商家之间的差评纠纷屡见不鲜。记者梳理各地发布的典型案例,既有消费者的批评意见得到法院支持而赢得官司,也有人被法院认定为“恶意差评”而追究责任。

负责审理本案的重庆綦江区人民法院法官魏永康表示,评判的关键在于如何界定合理与恶意的边界。

消费者权益保护法明确,消费者享有对商品和服务以及保护消费者权益工作进行监督的权利。经营者应当听取消费者对其提供的商品或者服务的意见,接受消费者的监督。

“从张某的视频内容来看,基本属于客观描述及对该修理厂服务质量的个人感知,没有贬损和丑化的陈述,也没有出现明显恶意诽谤、侮辱性等词

语,不应认定存在主观恶意;其言论对修理厂产生的影响程度较低,不足以降低该修理厂社会评价的损害后果。”魏永康说。

法院最终判决,驳回该汽修厂全部诉讼请求。

无独有偶,中国消费者协会去年3月发布的2023年“全国消费维权十大典型司法案例”中,有一起消费者因差评纠纷起诉商家侵犯名誉权并胜诉的案例。

2022年9月,姚某某在网上预订了被告所经营的民宿。姚某某入住后结合自身体验对该民宿打出三星评价,后将评价改为一星,被告用“恶意差评”“缺乏心智”等言辞辱骂姚某某。因双方沟通未果,姚某某以被告侵犯其名誉权为由提起诉讼,被告又反诉要求姚某某赔偿26万元。

2023年12月,湖南省长沙市中级人民法院终审判决,被告向原告公开道歉,赔偿原告精神损害抚慰金1000元

并驳回被告全部反诉请求。

打差评要基于事实把握分寸尺度

消费者的“差评权”有法律撑腰,不意味着可以利用“差评权”为所欲为。

近年来,多地政法机关开展行动严打“职业差评师”,一些利用“恶意差评”牟利的不法分子被追究刑事责任。有遭遇过敲诈勒索的网店经营者表示,一旦短时间内涌入大量差评,轻则影响网店曝光率,重则会被平台和市场监管部门处罚,所以往往只能选择“花钱消灾”。

“对商品和服务进行评价是消费者的法定权利,但消费者应当基于事实在社交网络上发表对商品或服务的评价。”魏永康说。

今年2月,最高法发布6件依法惩治利用网络敲诈勒索犯罪典型案例。在其中一起案例的点评中,最高法有关负责人表示,消费者的评价和投诉对入驻电商平台商家的口碑及后续经营有着重要影响。合理差评和正当投诉有利于维护消费者的合法权益,提高商家的服务水平和竞争力。利用线上平台恶意“索赔”,不仅严重侵害了经营者的财产利益,也扰乱了正常的市场秩序。

“人民法院依法予以打击,有利于遏制恶意差评的蔓延,避免消费者被误导,维护企业合法权益,营造良好的营商环境。”这位负责人说。

网络不是法外之地。在消费后进行评价,是消费者行使监督权的正当行为。对于合理的批评意见,经营者应有“容人之量”,针对不足采取改进措施提升商品和服务质量,不能“玻璃心”甚至攻击报复消费者;消费者也不能滥用手中的权利,把握好批评的分寸尺度,依法理性维权,维护清朗网络空间。

新华社记者 周闻韬 白阳