

韩国一家公益组织发起人李娟枝假扮男用户潜入“羞辱房”为调查人员收集视频、截图等信息

她力图揭开韩国“深度伪造性犯罪”的罪恶

39岁的李娟枝(Lee yeonjee)是韩国一家公益组织的发起人和CEO。她正在和蔓延全国的“深度伪造性犯罪”战斗。

有分析指出,把女性面部合成到不雅视频中并大肆传播的“深度伪造性犯罪”,已让韩国女性“陷入恐慌”,类似事件会让女性失去日常生活的安全感。

今年5月,首尔大学毕业生涉嫌利用深度伪造技术换脸合成色情照片和视频,在社交媒体“电报”上建立了近200个网络聊天室,共同参与制作、传播约1700份非法照片和视频。

3个月后,韩国仁荷大学再次发生深度伪造性犯罪的案例。超过1200名参与者在“电报”上合成深度伪造色情图像,并利用这些内容对女性进行敲诈和羞辱。8月下旬,韩国媒体报道称,“电报”上出现大量与全国各地学校相关的聊天群组,用户上传他们认识的女性照片,其他用户会将这些照片转换成深度伪造色情内容,还有针对特定受害者的“羞辱房”。

初步调查显示,受害者80%为女性,犯罪活动已经在初高中生等未成年、大学生、教师和护士等群体中蔓延,甚至连女性军人都是受害人。

这样的新闻极大地冲击了李娟枝,她看到有人在社交媒体平台X上发布消息调查“羞辱房”,找人帮忙揭露内部情况。李娟枝假装自己是男性用户,潜入其中,收集信息。她发现即便是在免费的房间里,非法深度伪造色情内容的传播也已经非常严重。

李娟枝将收集到的视频和截图集中汇集给一起工作的同伴,由她统一转交给调查人员。略显安慰的是,李娟枝听说,她们提交的材料帮助在部分案件中确认了涉案人员的身份。

“羞辱房”里的内容让李娟枝十分不适,甚至一度恶心到无法入睡,但她仍在继续行动。一方面,在线上开展请愿活动,要求强化惩戒标准;另一方面,在街头举行抗议活动。“政府必须采取坚定的立场来打击犯罪,执行‘零容忍’的惩治法律。”李娟枝说。

线上线下的抗议

9月6日晚上7点,首尔夜幕降临,淅淅沥沥下着小雨,李娟枝按照约定的时间来到了钟路区普信阁附近。过了一会儿,抗议活动正式开始,团体代表和个人率先上台宣读了各自的声明,随后,她们一行五六百人举着黑底白字的标语,走上街头。

当晚参与抗议的人包括现任教师、性教育组织、妇女团体和在校学生,她们高喊着“让我们过上没有恐惧和焦虑的生活”“监管推广深度伪造性犯罪的平台”“我们将推翻歧视女性的深度伪造”等口号。

抗议持续了约两个小时。李娟枝说,她和同伴们还计划9月21日在惠化地铁站举行第二场抗议,各地已经开始联系租借大巴车,届时会统一运送各地的抗议者到首尔来。

线下的街头抗议与韩国女性在线上社交媒体的发声遥相呼应。自深度伪造性犯罪的新闻报道引发韩国舆论重视以来,由于担心此次事件会被轻轻揭过,许多韩国女性在社交媒体上积极发声,使用“学校数字性犯罪”“公开深度伪造性犯罪者”“保护女性”等话题标签,推动“深度伪造性犯罪”的新闻得到全球关注,还有多名韩国女性通过中国互联网发出求助信息。

“目前来看,这次‘标签抗议’活动取得了不错的成效。9月1日,包括‘公开深度伪造性犯罪者’在内的标签在社交媒体平台X上实现了爆发,一共发布了60多万条帖子,全球实时趋势排名第二,也占据了韩国实时趋势的前两位。”李娟枝说。

线上线下联动造势外,李娟枝还开始调查研究涉及深度伪造性犯罪的处罚法案。其中,“以传播为目的”几个字尤其引起她的注意。有一篇分析文章认为,难以惩罚这些深度伪造性犯罪者的原因之一,在于很难证明他们有传播意图。

联想到此前韩国发生的“N号房”事件,除了少数关键罪犯,大部分人都没有受到惩罚。李娟枝意识到,正是现存法律体系存在漏洞,才让大规模深度伪造性犯罪成为可能。

在李娟枝潜入所谓的“羞辱房”期间,她发现尽管每个群组房间略有不同,但参与者通常需要提供一名女性的10张照片才能进入群组或合成一张照片,这为深度伪造提供了新的受害目标。部分群组房间会免费合成一次,接下来需要参与者提供“加密货币”或邀请新的男性进入房间,以继续使用服务。还有群组房间会张贴出“需提供300张女教师的照片”“高中生优惠”等告示。

“羞辱房”也会想办法确认其参与者的性别。有的群组房间会要求想入群的人将1韩元存入某个银行账户,通过银行账户存款名确认新入群人士是否为男性。

一些参与者会被收取费用以合成深度伪造图像或视频,还有人通过销售深度伪造内容获利,“电报”本身也会在用户较多的房间里产生广告收益。“他们就像犯罪组织一样分享利润。”李娟枝说,他们是由利益驱动的组织,不存在“被动观众”的概念。

由此,她产生了向国会请愿的想法。李娟枝要求将这些群组房间中的所有参与者都视为犯罪组织成员,并受到惩罚。因为韩国的法律只允许卧底调查涉及儿童及

青少年的案件,她同样希望请愿制度能够改进,允许卧底调查涉及成年人的性犯罪案件。

人工智能加持下“轻松的深度伪造”

位于深度伪造性犯罪丑闻中心的是即时通信应用程序“电报”。这是一款加密的消息应用程序,用户通常是匿名的,房间可以设置为“保密”模式,内容会被迅速删除,不留任何痕迹。原本“电报”这一设计是想要保护用户的言论隐私,结果却成了犯罪行为猖獗的主要场所。

用于制作深度造假内容的技术也不难获得。美国纽约州立大学布法罗分校计算机科学与工程系教授吕思伟(Siwei Lyu)对新京报记者解释称,“deepfake”一词本身就是深度学习(deep learning)和虚假媒体(fake media)的混合词,指的是使用依赖于深度神经网络模型的生成式人工智能模型创建的多媒体文本,其中包括文本、音频、图像与视频。生成式人工智能系统可以在大量无标记数据上进行训练,这些模型已足以生成令人信服的内容。因此,技术门槛被极大降低。

“有了人工智能技术,可以轻松、快速且廉价地生产深度伪造的视听材料,它本可以用于更有益的方面,例如电影、零售行业等。”荷兰蒂尔堡大学法律、技术与社会研究所副教授巴特·范德·斯鲁特(Bart van der Sloot)对新京报记者表示,不幸的是,深度伪造技术最常见的应用是色情内容,“尤其当年轻女孩成为受害者,这可能导致灾难性的结果。”

2019年一份调查深度伪造用途的报告发现,96%的线上深度伪造内容与色情有关,这一数字仍有增加的趋势。

以目前来看,打击深度伪造也存在不少障碍。斯鲁特发现,生成深度伪造色情内容时,受害者本人往往并不知情,而且内容也只有通过网络得到传播后,才会被外界发现并采取行动。网络上有如此多的深度伪造色情内容,执法部门也很难评估这些内容是否得到了当事人的同意,其中究竟是真人还是人工智能生成的人。即便找到犯罪者,确认受害者并删除材料,这时伤害往往已经造成,一切都太晚了。

李娟枝认为,事后追踪只是“亡羊补牢”,应该从源头加强平台层面监管,“如果犯罪分子隐藏在某个匿名平台之中,那么这个平台将成为犯罪的温床。”

针对具体措施,吕思伟认为,短期而言,应出台系列政策和法规,禁止或限制旨在产生有害影响的深度伪造内容。长期解决方案还是应回归技术本身,需要进一步开发用于自动检测、删除或跟踪深度伪造内容的技术。



深度伪造性犯罪引发韩国社会讨论后,当地时间9月2日,韩国警察厅国家搜查本部本部长禹钟寿在例行记者会上表示,警方以涉嫌“帮助深度伪造性犯罪”着手对社交媒体平台“电报”实施立案前调查。“电报”自己也发声承诺将积极打击平台上的有害内容。

始终未解决的“厌女症”

事件之外,外界开始追问为何韩国成为深度伪造的重灾区。

专注于身份盗窃保护的美国初创公司“Security Hero”在2023年发布的一份全球深度伪造报告显示,韩国是深度伪造色情内容的最主要目标,韩国歌手和女演员占此类深度伪造作品中人物的53%。

韩国的深度伪造性犯罪也日益猖獗,2021年156起,2022年160起,去年增至180起,而今年1月至7月,涉深度伪造的性犯罪案件已达297起。

观察人士认为,看似突然爆发的深度伪造丑闻,可能一直潜伏于表面之下,反映了长期存在的性别歧视问题。

在经合组织(OECD)国家中,韩国女性受教育程度最高,但劳动参与率却最低。韩国的性别工资差距在35个经合组织成员国中位列第一,差距为31.2%。

出于“只是觉得有趣”“用熟人的照片更真实”等原因,深度伪造变得像娱乐游戏一样蔓延。有分析称,无论是深度伪造性犯罪,还是韩国女性会在卫生间或更衣室遭遇微型隐藏摄像头偷拍,这些都是伴随技术发展而出现的最新形式的“厌女症”。

这与李娟枝的观察也不谋而合。

她最初关注的女性议题并不是深度伪造性犯罪,而是驾驶领域中的性别平等话题。

李娟枝发现,韩国女性在学习开车时常会遇到歧视问题。例如,当女性想要拿到一类驾照时(注:一种可以开高级车的高级驾照),她们会被劝说改换二类驾照。即便顺利拿到驾照,女性也不容易获得处理交通事故、汽车保险、汽车修理等方面的信息。女性司机往往会因为信息不足而对驾驶感到害怕,驾驶技术欠佳的女性也会被蔑称为“金女士”。

在韩国性别平等与家庭事务部的帮助下,2020年,李娟枝创建了名为“unniecar”的女性驾驶项目,提供讲座和书籍,旨在帮助初学驾驶的女性在没有“男性说教”的情况下,获得安全准确的信息。这几年来,李娟枝关注的女性社会议题外延也在不断扩大,强调保护女性权利。

李娟枝认为,从驾驶歧视,到深度伪造性犯罪。韩国社会中存在的“厌女症”一直没有得到有效解决。不仅如此,“厌女症”还被政治利用,进一步引发了男性群体的

怨恨,成为政客招揽选票的一种方式,“如今,有大量十几岁和二十几岁的年轻男性在物化和疏远女性,而非建立平等自然的两性关系。正是这样的环境让‘N号房’和深度伪造性犯罪事件得以出现。”

深度伪造丑闻曝光后,84个韩国女性权利组织签署了一份声明,直指歧视才是症结所在,“造成这种现象的根本原因是结构性的性别歧视,解决办法是实现性别平等。”

面对层出不穷的案件和社会团体的呼吁,韩国总统尹锡悦出面回应称,依靠匿名恶意利用深度伪造技术是犯罪行为,希望有关部门展开彻底调查,彻底根除这种数字技术犯罪。

韩国政府也有计划加强对深度伪造性犯罪行为的处罚。韩国执政党国民力量党政策委员会主席金相勋表示,国民力量党和韩国政府计划推进立法,将散布虚假视频等犯罪行为的最高刑罚从5年提高至7年。

不过,李娟枝对韩国政府的承诺仍有疑虑,担心行动会否落到实处。

这份疑虑也并非毫无依据。2022年竞选总统期间,尹锡悦一直否认韩国存在结构性性别歧视,将低生育率现象归咎于女权主义,声称要对虚假的性侵犯报告给予更严厉的惩罚。尹锡悦上台后,韩国政府还推动将“性别平等”一词从学校教科书上删除,取消了反对日常性别歧视项目的经费。

“长期以来,我没有看到政府出台解决性别歧视的政策,反而他们是在破坏性别平等行动取得的进展。”李娟枝表示,韩国出生率已经降到世界最低水平之一,如果不认真解决“厌女症”,性犯罪还将继续上升,如果性犯罪不受到严厉惩罚,未来罪犯利用新技术犯下更复杂的罪行,那时韩国社会将更加被动。

斯鲁特表示,从根本上应对深度伪造性犯罪,各环节联动缺一不可。政府需要将应对工作作为优先级,有明确的惩罚标准,设立标杆案件,若有必要可以专门建立一支应对深度伪造性犯罪的小队。学校和 workplace 也需要展开有关“非同意网络色情”的教育。网络提供服务者需要监控平台,阻止传播任何深度伪造色情内容,警告用户一旦通过他们的服务平台传播深度伪造色情内容,他们将被永久移除。

9月7日,李娟枝告诉记者,她向国会提出的请愿已获得所需的5万人同意,韩国政府有关部门有义务对请愿做出回应。

球已被踢到韩国政府一侧,“我很期待他们接下来的行动。”李娟枝说。