2026年1月13日 星期二
www.jfdaily.com
解放日报

部分爆款文章、热点视频并非出自真人之手，记者调查发现——

AI造假产业链是这样运作的

“用大模型7分钟写了9篇短篇小说，都签约了”“下班后借助大模型搞虚拟资料，半个月自动接了750单”……打开社交平台，诸如此类的宣传语随处可见。你有没有想过，网络上的爆款文章、热点视频，有相当一部分并非出自真人之手。更重要的是，由AI批量生产的“工业制品”中，有相当一部分违背事实，通过造谣的方式博取关注度，获得收益。

记者历时半个多月调查发现，一条涵盖AI内容生成、教程售卖、流量裂变、商业变现的完整造假产业链已悄然形成，不仅会误导网民，更是对正常的网络传播秩序和商业环境造成冲击。

目前，不少社交平台对明显有AI生成痕迹的内容进行强制标注。网络截图

一键“二创”，AI裂变

从各地网信部门和公安部门披露的案例看，使用AI大模型生成文字、视频造谣，已经不再稀奇。

具体来看，整条产业链涉及不同的盈利方式，最常见的一种是造假者利用自媒体矩阵，发布大量AI造假内容，为自己增加关注度，既能通过平台的流量分成机制或植入广告获利，又能利用关注度推销自己的服务。同时，大量企业或个人在推销“AI教程”。他们深谙平台流量和广告分配机制，继而开发出对应的“AI生成课程”以及相关服务，通过销售课程和生成服务，也能牟利。

这些盈利方式各有特点又互相交织，核心都是“AI造假”。在实际生活中，一条谣言的诞生和传播，往往穿插着不同的利益链条。

通过“眼球效应”能赚多少钱？记者通过社交平台购买了一套声称“10天受益3000元”的“视频号AI生成励志语录保姆级教程”。下单后，记者收到对方发送的网盘链接，其中介绍了如何将他人发布的心灵鸡汤去除水印、用AI生成图片或视频进行二创的方法，对方表示只要用对工具，每天可以生成数百条。至于“10天受益3000元”，靠的是平台的植入广告分成机制、直播带货和微店销售，主要来源是广告分成。

知情人士更是为记者算了一笔细账：以某社交平台的植入广告分配制度为例，一个拥有1000粉丝、单篇访问量1000的视频账号，如果支持植入广告，根据账号定位、内容是否为原创、植入广告数量及位置等，单篇广告收入从十几元到上千元不等。“不要小看十几元，因为对那些靠吸引眼球牟利的企业来说，旗下有数百上千个账号。所以同一主题，用AI写出不同角度、不同细节的文章，都能被平台认定为‘原创’，继而获得相对较高的广告收入。综合起来看，一条谣言可能获得数千元收入。”该人士提醒。

爆款有“教程”，造假有“选题”

在多个社交平台和知识付费平台上，“AI爆款写作秘籍”“零基础玩转AI文案”等教程随处可见，价格从十几元到数千元不等。

使用AI生成的文字到底什么样？记者联系到一名推销AI副业的“导师”，他毫不讳言地表示：“靠AI赚钱，选题是关键。”为证明自己“经验丰富”，他向记者介绍了几个核心“选题方向”。

“卖惨”是他给的一个建议。他表示，用AI生成卖惨类话题很简单，常见的“骑手卖惨”是典型话题，“用AI生成骑手遭遇恶劣天气送餐、被客户误解、家庭困难仍坚守岗位等内容，搭配煽情文案，极易引发同情和转发”。

“根据热点事件生成‘心灵鸡汤’类型的分析稿件也有市场。”该“导师”说，在这类稿件中，可以用真实事件作为主体，然后用AI生成“治愈”“励志”“煽情”“批判”等不同风格的“心灵鸡汤”，“就是对真实事件进行解读、分析、延伸。”

该“导师”称，如果记者购买付费服务，他还能传授如何修改AI生成内容的痕迹、如何判断敏感词规避平台审核等。可见，在AI造假牟利上，已经形成了“教程售卖—工具推荐—内容生成—商业接单”的完整灰色链条。

“用魔法打败魔法”

AI生成的虚假内容在网络空间肆意传播，真的处于无人监管的真空地带？答案显然是否定的。

去年9月1日，《人工智能生成合成内容标识办法》正式施行。办法明确规定，所有AI生成合成的文本、图片、视频等内容，必须添加显式标识，确保用户能够清晰辨别内容来源。

记者实测验证发现，主流大模型在生成内容时已初步落实该规定：生成文稿时，结尾会自动标注“本内容包含AI生成部分”；生成图片或视频时，内容角落也会添加半透明的“AI生成”水印。同时，将这类已标注AI生成的内容上传至多个社交平台时，发布界面会跳出弹窗提示：“经检测，该内容含AI生成元素，将按规定展示标识后发布”，强制完成合规标注。

不过，造假者也在总结“去标识”技巧：文本类内容直接删减末尾的AI标注语句；图片类通过裁剪、拉伸或轻微调色抹去水印，视频类利用剪辑工具剪掉带标识的片段；更进一步，通过调整格式或二次转码的方式，清除文件属性中的AI标记……经过以上处理，部分AI生成内容就能堂而皇之地以“原创”身份发布在社交平台，普通用户难以辨别。

面对这种规避监管的行为，治理之路该如何推进？AI行业从业者尤骑给出的答案是进一步发挥“用魔法打败魔法”的力量，即以AI技术对抗AI造假。越来越多的大模型开发者开发出了鉴别工具，能精准捕捉AI生成内容的隐性特征。国内不少科技企业和科研机构也研发出专门的AI生成内容检测工具。这些工具通过深度分析文本的语义模式、句式规整度、逻辑连贯性，以及图片的光影合理性、人物肢体细节，构建多维度鉴别模型。虽然识别准确率不能达到100%，但有很强的参考价值。

平台能否接入“鉴假”大模型？

在提升技术鉴别能力的同时，也需要平台作为“把关人”，承接鉴别能力的落地。

这个话题，恐怕是“老生常谈”。只不过这一次可以用上技术力量——由平台接入成熟的AI生成内容检测大模型，建立“发布前检测—疑似内容预警—违规内容拦截”的常态化审核流程。

具体而言，对用户上传的文本、图片、视频等内容进行实时检测，直接拦截存在AI造假且未标注的内容，不允许其上线；对多次发布违规造假内容的账号，采取限流、禁言直至永久封号的梯度处罚措施。尤其对于“AI副业”培训中高频出现的热点议题，更要强化定向审核，及时清理虚假内容，遏制不良导向传播。

治理AI造假不能仅靠监管和技术，公众的辨别能力同样重要。在信息爆炸时代，面对动辄10万+的网络爆款、看似治愈的心灵鸡汤，公众也要保持理性判断，不盲目相信、不随意转发。如果仅凭情绪共鸣就轻信传播内容，很可能被AI生成的虚假内容迷惑，陷入更深的信息茧房。 本报记者 任翀

时评

火爆App背后深切焦虑应被看见 独居群体身后保障托举应有更多

顾杰 周昱帆

这几天，一款名为“死了么”的App火了。

但实际上，除了名字之外，它的功能属实没什么新意：用户每天签到报平安，如果连续两天未签到，系统会自动发邮件通知紧急联系人。

市场上类似功能的产品已经推出多年。最简单的有手机自带的安全功能，一键即可呼叫紧急联系人；升级版的有智能手表等可穿戴设备，能够检测出摔倒、车祸等特殊情况。于是，一些商家便出来求关注，“看看我们吧，我们的产品明明更早也更好”。

但显然，“死了么”App这把确实赢在名字上。网友们说，这款App就像是一群西装革履的人中，突然多了一个“穿寿衣”的，“很难不注意到”。

不得不承认，它以一个似乎“很晦气”的名字，精准命中当代最现实的一个问题：在日益原子化的社会里，很多人在物理层面上与其他人的实时连接越来越少。这些独居群体有一种普遍的焦虑：“我会不会悄无声息地死去？”

但无论是已经花了8元下载并用上这款应用的年轻人，还是开发者本人，恐怕都不敢拍胸脯说，仅靠这一个App就能满足独居者实际的安全需求。

这款App的逻辑在于将使用者的安全责任部分分摊给“紧急联系人”，紧急联系人可以是家人，也可以是朋友。但现实情况是，很多在大城市独自打拼的独居青年，家人亲属根本不在身边，可能也没什么可以依赖的朋友。哪怕写了联系人，联系人也未必能第一时间赶到身边。

追捧这款软件的人，真正底层的焦虑或许可以概括为同一个问题：当我独自一人时，甚至是没有“紧急联系人”时，我的安全还可以托付给谁？

换个视角，这就变成了另一个正变得越来越迫切的议题：我们的社会如何为规模庞大的独居群体编织一张可以信赖和依靠的社会支持网络？

这个问题背后，有很多待填补的空白。它不仅是一个简单的技术问题，更是一个关乎社会分工的问题。其核心就在于，要构建一个多方力量相互协同且能流畅运转的机制。

记者想起之前采访针对独居老人研发的“一键报警器”等智能产品，底层逻辑和“死了么”App有点类似，但真正的难点往往不在于发出警报的技术本身，而更多在技术之外。

比如，警报响起之后由谁来处理？如果警报需要发送给后台，谁来组成后台？如何保证后台的运营维护能随时在线？如果发生漏报、误报，产生的风险谁来承担？一旦出现紧急情况，又该如何和物业、社区、居委会、急救、医疗等不同主体对接？这背后涉及一系列流程响应、人力调配和法律权责的问题，远非一个应用程序所能囊括。

从这个层面上看，技术或许只是其中“微不足道”的一环，在这条“安全链条”的前端和后端，有很多问题有待梳理。

在前端，可能涉及社区独居群体的日常管理和识别，如果平时就打不开居民的家门，又怎么指望社区能在第一时间发现意外？在后端，可能牵涉独居者出现意外时，由谁在手术单上签字的问题，这就又可能涉及意定监护等一系列法律和政策问题。

说这些，不是意在强调这件事难做，而是想说，所谓的“独居安全”，不单单是个人的事，更是全社会的事。这也不是简单的线上与线下孰优孰劣的争论，更不必陷入一味排斥技术或盲目迷信技术的极端。归根结底，这是一个需要社会多方力量共同参与、形成合力的系统工程。

唯有当技术的进步、社会的支持、政策的保障等多种力量无缝衔接、流畅协作，独居者的安全，才有了真正的依托。

建言 投稿 报料 求助
扫码参与互动