

原创文被判定为“人工智能生成”，长相精致通不过真人检测

向AI证明“我不是AI”频频困扰用户

紫牛头条
在这里遇见不同

最佳深度媒体 | 未经授权,不得转载摘编



毕业季到来,很多人都在忙着写论文,在社交媒体网站上经常看到有人吐槽,辛辛苦苦写的文章,竟被判定为AIGC(即人工智能生成内容)。事实上,AI的这种误判不仅针对文章,有人想以自己为原型,用AI制作数字人,竟然无法通过“真人审核”。网友对此表示,一些人正在因为表现得“太优秀”或“太标准”,而在AI算法面前失去了作为人的“入场券”。

扬子晚报/紫牛新闻记者 宋世锋 张冰晶

该片段AIGC疑似度综合值:

智能降AIGC

综合疑似度

65.36%

中度疑似AIGC

该片段AIGC生成概率比较高,有可能是AIGC和人类共同完成。

► 免费检测提供“智能降AIGC”增值服务



网友上传真人素材被判定无法通过真人审核

辛苦原创的文章被判定为“人工智能生成”

临近毕业,在社交媒体网站上,吐槽论文进行AIGC率检测时数据离谱的帖子频频可见。

小达(化名)近日曾在网上发帖,询问论文AIGC率检测20%左右够不够。因为现在检测平台众多,他不知道自己的论文在不同的平台会有什么结果。

他告诉扬子晚报记者,他所在的学校规定论文的AIGC率不能超过30%。起初,他检测的AIGC率达到60%多,后来用了检测平台提供的降低AIGC率的服务,花了数百元钱,降到16%左右。

小熊(化名)把市面上免费的AIGC检测网站都试了一遍,发现结果都比较偏高,绞尽脑汁降低AIGC率以后,还有20%左右。后来她用学校提供的AIGC检测系统,数值低了很多。她的不少同学也遇到过这些问题,在免费网站测出30%—40%多,在学校提供的系统中测出来1%多。

记者试用了AIGC率检测,把AI大模型流行前写过的一篇报道提交到一个检测平台,检测出AIGC总体疑似度达18.12%,数段内容被标上橙色,判定AIGC生成疑似度在60%—70%。

活生生的真人被系统当成“数字人”

受到AIGC检测困扰的不只有论文作者们。

年轻女孩小林(化名)为尝试制作专属AI数字人,将自己真人拍摄的原片视频上传至某专业剪辑平台,却接连收到平台“未通过真人审核”的提示。她多次更换账号重新上传,结果均以审核失败告终,这一乌龙事件让女孩倍感疑惑与无奈。

小林告诉记者,她出于对AI数字人技术的好奇,打算用自己真实拍摄的视频素材,借助平台功能生成个性化虚拟形象。视频全程由本人出镜拍摄,画面、动作均为真实记录,无任何AI合成、换脸等后期处理。

可提交审核后,平台系统却始终判定其视频并非真人拍摄,无法进入后续的数字人制作流程。反复尝试无果后,小林不禁提出疑问:“明明是活生生的真人,难道被系统认成数字人了?”

此事一经分享,迅速引发讨论。不少网友表示感同身受,称如今AI生成技术愈发逼真,虚拟形象与真人差距不断缩小,导致平台审核算法出现“识别混乱”。

业内人士分析,当前一些内容平台为防范AI虚假视频、侵权换脸等问题,设置了严格的真人

审核机制,但算法在判定时易过度敏感,将面部特征精致的真人素材,误判为AI合成的虚拟内容,进而出现此类反向乌龙。

文章改到句子不通顺才能过检

论文查重有非常明确的结果,是否抄袭一目了然。AIGC率检测如同黑箱,系统到底如何判定,没有人知道。各个检测平台都没有具体说明如何检测AIGC率。小熊说:“我就是这种感觉,查重的话,起码能标注和哪个文献有相同的地方、相同处有多少字。我感觉AIGC率挺虚无缥缈的,不知道平台基于什么算法。”

为了通过AIGC率检测,已经有人总结了一些经验。一位AIGC率从85%降到10%的用户介绍方法说,文章要多用“的、了、到、过、会”等废话字,并把“首先、其次、最后”这类AI常用的词语删了,可以改用“一是、二是、三是;一方面、另一方面”等代替。要把短句合成成长句,经常出现的词换成同义词,改变句子的前后结构,如“效率提升了30%”改成“提升了30%的效率”等。读起来句型要乱一点,不要全是一个节奏,语气不要太书面。

另外,每个AIGC率免费检测网站都会提供增值服务,如果

检测出来的数值比较高,就向用户推出降低AIGC率的服务。记者在检测报告页面中也看到,系统提供智能降AIGC服务,但需要购买旗舰版才能获得。

毕业生如果论文被判定AIGC率过高,可能有利影响。小熊迫于无奈,只能花钱使用检测平台的服务降低AIGC率。小熊说,降低AIGC率之后,她的论文在学校提供的系统中AIGC率检测出来是3.7%，“标红的内容都在致谢部分,有时候手写的也会标注为AIGC,所以感觉检测不准。”

小熊发现,所谓降低AIGC率,“其实一般是把语句改得不通顺一点,句子结构更换一下。而且检测平台降低AIGC并不是人工修改,也是在使用AI进行处理。”经历了降低AIGC率的小达也表示,降AIGC率对文章的表述有影响,“有些句子都不通顺了,真的无语。”

如何判定?AI对人类进行“反向图灵测试”

技术专家表示,目前的AI检测工具并非真的能“读懂”AIGC内容,它们本质上是基于统计学的特征分析。AI是通过学习人类最优秀、最标准的数据训练出来的,AI写作非常稳定,而且逻辑

严密,没有语法错误。如果一篇文章文笔极佳、逻辑无懈可击,AI算法就会因其缺乏人类常见的“困惑度”而判定为AIGC。对数字人的判断也是这样,如果真人长相标准,表情举止到位,AI算法会认为这种高度的对称性和稳定性超出了自然人类的随机误差,从而判定为AIGC。

有些人正在因为表现得“太优秀”或“太标准”,而在算法面前丧失了作为人的“入场券”。而且AI判定不像人类的民事法庭,并非“谁主张谁举证”,被判定为“疑似AI生成”后,人类必须自证清白。

人类评判AI是否具有人类智能时,有一个著名的图灵测试,核心是探讨机器是否能够模拟人类智能,使人类无法区分机器与真人的回答。AI是否真正通过图灵测试尚有争议,而AIGC率检测已经成了AI对人类的“反向图灵测试”。

现在的论文检测和数字人判定是AI在测试人类是否“够乱、够笨、够不标准”,只有通过了这种“反向图灵测试”,人类在AI眼中才有人的身份。

人类面对AI的检测,如果为了“证明自己是人”而不得不故意把文章写“烂一点”、让相貌更“糙一点”,故意展示瑕疵来证明自己,这标志着我们进入了一个“以不完美定义人类”的荒诞时代。

紫牛热点

路面坑洼导致路过车辆爆胎,谁之责

法院:管理单位未完全尽责,承担八成责任

扬子晚报讯(记者 郭一鹏)在日常驾驶中,不可避免地会遇到坑洼不平的路面,有时会导致爆胎或轮毂损坏。车主如果有此遭遇,应该由谁来承担赔偿责任呢?扬子晚报/紫牛新闻记者注意到,云南省威信县人民法院近期审理了一起关于此类事件的财产损害赔偿纠纷,一审判决道路管理方承担八成责任。

中国裁判文书网上发布了关于此案的一审民事判决书。2025年8月22日22时5分许,陈

某驾驶小型轿车,沿GXX国道由某地方向往县城方向行驶,至某大道与GXX国道交叉口260米处,其所驾车行驶过程中遇路面坑洼,造成了车辆左方前后轮受损的交通事故,事故经威信县某局某大队认定,出具了《道路交通事故证明》。之后,陈某更换受损的轮胎及钢圈,产生费用1500元。另查明,案涉路段属城市道路,被告威信县某局负责该路段的管理工作。事故发生时,坑洼处未设置锥形桶、警示

灯等警示标志。

陈某认为,事故成因系路段路面坑洼,自己已尽到合理安全驾驶义务,威信县某局作为市政道路管理维护主体,道路出现安全隐患时,未及时维护并设置警示标志,存在重大过错,应承担事故全部责任。对此,威信县某局辩称,案涉路段已经长期反复进行修复,不存在管理不到位。同时,为建设某高速路段,经常有大车、浇灌车等大型车辆通过,对案涉路段造成长期的损

坏,不能要求他们作为一个完美的管理者,实时进行兼顾。此外,在该路段出事的第一时间该局已经对该路段进行了修复。威信县某局认为,陈某作为一个成年人,在阴雨天气驾驶车辆时,对驾驶车辆以及驾驶路段的评估应负有责任。

法院经审理认为,威信县某局作为案涉事故发生路段的管理维护义务人,在道路出现坑洼的情况下,未及时发现并设置警示标志,足以认定被告未完全尽

到对案涉路段的管理或维护义务,对案涉事故的发生存在过错,应对原告的损失承担主要的赔偿责任。另一方面,事故发生路段地面湿滑,且是夜间行车,陈某未尽到谨慎注意义务,对损害的发生存在一定过错。综上,法院酌情确定威信县某局对本次事故造成的损失承担80%的责任,陈某对本次事故造成的损失承担20%的责任。最终,法院判决威信县某局赔偿陈某车辆维修费1200元。