

几秒钟就合成一篇假新闻 互联网上眼见不一定为实 莫要被 AI 假信息忽悠了

不久前,两封疑似AI生成的“重磅”信件引发网友大量转载,其中一封署名英伟达CEO黄仁勋的内部信表达对DeepSeek的看法,乍看信件文采斐然、自带人文情怀,幸好辟谣及时来到,明确信件完全是虚假捏造。

1月,美国加利福尼亚州洛杉矶多地连续发生山火,“在加州山火中幸存的教堂”图片、“好莱坞巨型标志牌被加州山火吞没”视频等在社交媒体广泛流传,均被检测出是人工智能生成或深度伪造的内容。

从灾难虚假图片到仿冒名人事件,AI发展这把双刃剑,让“眼见为实”的固有信念遭受前所未有的挑战。

现象 AI崛起降低造假门槛

“假新闻并非新生事物,但AI大模型的崛起大幅降低了造假门槛。”浙江大学百人计划研究员、博导王永威表示,深度伪造技术(Deepfake)能轻松合成逼真的图片、视频等,实现换脸、视频无缝合成及语音精准模拟,内容令人难辨真伪。

如今AI模拟真实世界的的能力显著提升,现在只需简短对话和操作,几秒钟内便能生成一篇图文并茂的假新闻。浙江绍兴警方曾抓获一个利用ChatGPT制作虚假新闻的团伙,其中有的成员仅有初中学历,此前从未

接触过电脑技术,凭借易上手的AI大模型,就能批量生成AI假新闻。暴力冲突、灾难现场等社会热点议题,是AI假新闻的重灾区。

算法把肉眼难辨的假新闻推送到人们面前,大多数人易被情绪左右,不假思索地点赞、转发,不经意间成为谣言传播的帮凶。验证新闻的真实性,公检法、媒体等需调动大量资源,被造谣者会陷入无尽的自证陷阱,而等到水落石出,假新闻早已“吃饱”流量。

AI伪造不仅骗取公众的情感和流量,还被用于

诈骗。

蚂蚁集团AI核身安全负责人刘健表示,AI模拟物理和生物信息的能力越来越强大,定制化诈骗让人防不胜防。如2024年香港一家公司财务被AI伪造的“高管”通过视频会议诈骗2亿港元。还有犯罪分子用AI换脸伪造事主的家人进行电信诈骗等案例。

“AI伪造给社会带来严峻的挑战,带来信任体系的崩塌。”刘健分析,大量虚假信息会让人难以分辨真伪,逐渐麻木,甚至可能侵蚀人类社会的温情与关怀。对网民来说,未来的数字世界里还能相信什么?

应对 用AI技术打败AI造假

面对来势汹汹的AI伪造,一些科技企业率先筑起防线。宇视科技有关负责人陶沁沁介绍,AI伪造有从数字领域渗透至物理世界的趋势,犯罪分子利用真人照片AI生成的3D结构模型制作人脸头套用于非法解锁和诈骗。对此,宇视科技开发了3D结构光技术,确保只有本人活体才能通过识别系统,其摄像头能识别真人面部轮廓变化,结合多种验证手段精准判断。

蚂蚁集团的相关安全团

队采取了三大策略应对AI深度伪造挑战:“不变应万变”,通过AI防伪大模型识别真实世界中的共性;“交互逻辑”,通过随机实时交互增加造假难度,如高风险场景中的AI实时视频电话验证;“持续监测”则实时监控异常行为,及时阻止造假。为此,蚂蚁集团研发推出以安全防伪大模型为基础的新一代核验平台。此外,其还通过举办挑战赛推进技术进步,吸引了全球2200多名参赛者。

浙江大学王永威团队提出一种无须训练即可检测AI合成文本的方法,通过单词预测的概率分布判断文章是否由AI生成,准确率超过美国商用文本检测模型“GPTZero”近6个百分点。在AI合成图像检测方面,该团队研发的6个专家模型,能综合多个维度检测特征,对当前近20种生成模型鉴伪,准确率达92%,处于世界领先水平。目前部分技术已在高校论文鉴伪、内容审核等领域应用落地,未来还将向更多领域拓展。

建议 每位用户都要擦亮眼

守护AI安全,不仅是技术问题。“AI攻防就像‘矛和盾’,是一个持续博弈的循环过程。”刘健认为,需要构建一个超越技术层面的,涵盖监管、法律、制度及教育等多维度的综合性社会防护网。

刘健建议,从源头管理、信息传播链路监控以及防控机制的建立这三个维度构建法制框架,做到有法可依、有据可查。源头服务商需强化内容把控,利用水印等技术手段实现内容可追溯;社交平台承担内容核实与筛查责

任,阻断虚假信息流通渠道;权威机构与国际组织携手建立信息联盟,确保鉴定与举证透明高效。

目前,我国2023年实施的《互联网信息服务深度合成管理规定》,明确要求深度合成服务提供者履行安全评估及配合检查的责任。微信等平台也积极响应,开展打击AI仿冒名人营销等行动。

公众的辨别能力同样重要。王永威的团队正筹备相关实验室,科普AI伪造技术和案例,提升公众认知与防

范意识。刘健将肉眼识别AI伪造比作“找茬游戏”,他指出“比如生物学上的不合理之处,地震小男孩有6个手指这种细节,或是光影效果不符合物理学原理。”公众可利用这些细微线索辨别虚假信息,增强鉴别能力。

AI技术是一把双刃剑。陶沁沁呼吁,AI安全守护需政府、企业、公众等多方面的共同努力。在这场没有硝烟的战争中,每一方都是不可或缺的战士,共同编织着一张守护数字时代安全的密网。 据中新网

自2月8日起面包糕点 为何不再使用这种添加剂?

根据国家卫生健康委、市场监管总局发布的《食品安全国家标准 食品添加剂使用标准》(GB 2760-2024),自2025年2月8日起,脱氢乙酸钠(脱氢乙酸及其钠盐)不再用于面包、糕点、果蔬汁(浆)等7类食品,在腌渍蔬菜中的最大使用量也由1克/千克调整为0.3克/千克。

为何对这种食品添加剂的使用进行调整?之前吃了添加脱氢乙酸钠的食品,是否有危害?怎样科学看待食品添加剂?记者采访食品安全专家予以解读。

为何调整使用脱氢乙酸钠?

专家告诉记者,脱氢乙酸钠是一种常见的食品添加剂,对酵母菌、霉菌、腐败菌具有良好的抑制作用,过去几十年里被多国许可,广泛应用于食品中,以延长保质期。

那么,新国标为何对脱氢乙酸钠的使用范围和使用量作出调整?

科信食品与健康信息交流中心副主任阮光锋介绍,对一种食品添加剂进行重新评估,一般有两种原因:一是在安全性上有新的证据发现,需要重新评估;二是食品消费结构发生变化,当一种食品的消费量由少变多时,要考虑其中某种食品添加剂累积之后会不会超过安全限值。

此前,有动物实验研究显示,多次大量食用脱氢乙酸钠,可能造成动物取食减少、体重下降、凝血能力下降、肝肾组织变化等问题。

中国农业大学食品科学与营养工程学院教授范志红认为,这类动物实验并不能证明人类少量吃这种添加剂也会导致同样危害,但往往会让人们对该物质有更严格的限量。比如,烘焙产品近年来消费量明显增多,其中的食品添加剂含量就需要重新考量。相对来说,腌渍蔬菜消费量并不多,所以只是降低了最大限量。

专家表示,食品安全追求“尽可能降低风险”。随着最新科研发现以及国内食品消费结构改变,食品添加剂使用也会相应作出调整,防患于未然。

国家食品安全风险评估中心标准三室主任张俭波表示,修订脱氢乙酸钠的使用规定,是根据食品安全风险评估结果和行业实际使用情况调研作出的决策。

食用后健康风险有多大?

脱氢乙酸钠已经在食品行业使用了较长时间。不少人关心:此前吃过这种食品防腐剂,会造成身体伤害的风险有多大?

食品安全专家表示,风险大小跟摄入量密切相关,即要看消费者实际一次吃进的量,以及一年当中吃的频次。

范志红表示,脱氢乙酸钠在人体内能够被代谢掉。相关实验数据显示,脱氢乙酸钠在毒理学上不属于高毒成分,动物实验发现其有害作用,是在“长期”“反复”“大量”食用之后的结果。人们日常摄入量一般不足实验中发现有害量的十分之一,无须过于担心健康风险。

阮光锋表示,脱氢乙酸钠并非被“禁用”,它依然是允许使用的食品添加剂,可以用在腌渍蔬菜、发酵豆制品等食品中。这说明风险评估结果显示,合理使用脱氢乙酸钠依然是安全的。

阮光锋还介绍,美、日、韩等多国也仍允许在部分食品中使用脱氢乙酸钠,但各有不同规定。例如,美国允许用于切块或去皮南瓜和草莓,最大使用量不超过65毫克/千克;日本、韩国允许用于黄油、奶酪、人造黄油等食品中,最大使用量不超过0.5克/千克。

食品添加剂还能不能吃?

出于对食品安全的考虑,还有部分消费者担心防腐剂等食品添加剂“添加”了风险。

对此,中国工程院院士、中国食品科学技术学会理事长孙宝国表示,食品添加剂的使用历史悠久,例如古代卤水点豆腐时使用的卤水,其主要成分氯化镁,便是一种添加剂。现代人的生活更是离不开食品添加剂。

孙宝国介绍,我国相关标准规定了23类2300多种食品添加剂,包括防腐剂、着色剂、膨松剂、甜味剂等,按照标准使用,不超过最大使用量和最高残留量,不会引发食品安全问题。过去发生的一些食品安全事件,往往源于违规使用“非法添加物”或滥用食品添加剂。

孙宝国表示,为保障公众健康和食品产业发展,我国对食品添加剂一直实行严格的准入和淘汰机制,开展持续、动态的跟踪评价,不断调整其使用范围和使用量,并对部分物质禁止使用,应当科学、理性看待标准的调整。新华社记者董瑞丰