

## AI 检测 AI

## “矛”更利还是“盾”更坚

近年来,人工智能(AI)技术推动生产力快速发展,但同时也因技术滥用导致各种问题。

为监督 AI 技术使用,如今市面上不乏各类用于检测 AI 生成内容(AIGC)的工具,如普林斯顿大学学生开发的 GPTZero、斯坦福大学研究团队推出的 DetectGPT 等。我国一些研究团队也陆续发布各类检测工具,如西湖大学文本智能实验室研发的 Fast-DetectGPT。

人类的创作与 AIGC 之间存在哪些差异? AI 检测工具如何根据差异进行识别? AI 检测工具如何应对越来越聪明的大模型?

## AI 创作套路化明显

“虽然大模型在不断迭代,但到目前为止,AIGC 与人类的创作在用词用语、逻辑语法等方面依旧存在明显区别。”Fast-DetectGPT 研发者之一、西湖大学文本智能实验室博士鲍光胜说。

在用词用语上,AIGC 有相对固定的偏好。“不难发现,一些词语会反复在语段中出现。”鲍光胜举例说,有研究发现,大模型应用于英语学术论文写作时,“delve”(深入研究)一词的使用频率大大提高,这是因为大模型习惯用这个词对语句进行润色修改。

在逻辑语法上,AIGC 惯常使用的一些语法搭配方式,在人类创作中可能并不常见。“受模型建模的影响,AIGC 有相对固定的行文逻辑和表述模式,且这些模式会不断地被重复。人类在行文上则更为灵活,没有固定套路。”鲍光胜说。

北京大学信息管理系师生比较了 AI 生成与学者撰写的中文论文摘要。研究结果同样显示,AI 生成的摘要具有较高同质性和较强写作逻辑性,并惯用归纳总结等学术话语体系;学者撰写的摘要则具有显著个性化差异,使用凸显实际含义的搭配较多,并常用与国家政策密切相关的词语。

哈尔滨工业大学一名研究生讲述了他使用大模型的实际感受:“当我给大模型提供一些材料让它扩写,它每次都用了相同的套路——把给定的材料拆解开,分为若干点论述。总体来说感觉它写得比较‘僵’。”

AIGC 相对套路化的创作,可能会影响人类的用语习惯。“随着越来越多人用 AI 创作或润色文字,人类会受到潜移默化的影响,这或将影响整个社会对语言的使用。”鲍光胜说。

## 三种路径识别文本

如何准确识别 AI 生成内容? 鲍光胜介绍,目前主要有三种技术路径进行检测,分别是模型训练分类器法(也被称为监督分类器法)、零样本分类器法、文本水印法。“三种检测方法本质上都是利用 AI 检测 AI,且各有优劣。”鲍光胜说。

模型训练分类器法,首先要收集大量人类创作内容与 AIGC,然后以此为基础训练一个能区分两类内容的分类器。“这是目前被广泛使用的一种方法,但缺点较为明显。”鲍光胜解释,用于训练分类器的数据有限,很难覆盖所有类型和语言的文本。分类器在训练数据覆盖的文本领域或语言上检测准确率较高,反之准确率则较低。而且,模型训练往往需要较高成本,数据规模越大,训练成本越高。

相比之下,零样本分类器法不需要对机器进行训练,也无需收集数据。它利用已训练好的大模型,抽取语言模型生成文本的特征,据此来区别人类与机器。“似然函数是零样本检测法中比较常用的基准之一,它可以简单理解为一段文本在某个模型的建模分布中出现的概率。概率是一种特征,不同的概率体现了人类创作内容与 AIGC 的差异。”鲍光胜进一步解释,“零样本分类通过综合考虑多种函数特征来区分人类创作内容与 AIGC。”

如今,很多大语言模型几乎覆盖了互联网上的全部数据。因此,相比于模型训练分类器,零样本分类器在不同领域、不同语言的文本上表现较为一致。

不过,零样本分类器也存在明显缺点。一方面,现有零样本分类器依赖生成文本的源语言模型进行检测,这意味着如果是未知源模

型生成的文本,分类器就无法准确检测。另一方面,为提高检测准确率,零样本分类器往往需要多次调用模型,这增加了模型的使用成本和计算时间。

“文本水印法则是一类‘主动方法’。区别于前两类方法,它不是检测已生成的文本,而是在 AI 生成文本时加入水印。人类虽然看不出这些水印,但却能通过技术手段检测出来。”鲍光胜说,文本水印法的准确率较高,但缺点在于水印可能被人削弱甚至移除。此外,对于无法访问模型内部结构的大语言模型,技术人员可能无法在生成内容时成功加入水印。

## 检测技术需不断改进

“未来,我们要不断更新、完善现有技术,力争实现快速、准确、低成本检测,在大模型这把‘矛’越来越锋利的同时,让检测技术这面‘盾’更为坚固。”鲍光胜说。

据了解,为提升检测准确性,目前市面上的商用 AI 检测软件大多融合了多种技术手段。国内外研究团队也在进一步完善相关技术。

对 AI 检测 AI 的前景,有两种截然不同的观点。一种观点认为,未来 AIGC 将会与人类创作极为相似,以至于检测工具无法判别。还有一种观点认为,随着技术发展,检测技术或将赶超大模型技术,实现对 AIGC 的有效识别。

“目前,无论是 AI 生成的文字、图片还是视频,都在技术上可识别的范畴之内。相较于文字,图片和视频甚至可以直接被专业人士肉眼识别。期待未来通过大模型技术的不断进步,推动检测技术发展。”鲍光胜说。

(吴叶凡)

谷神星一号完成  
“海上一箭6星”发射

据星河动力航天公司消息,日前,在太原卫星发射中心海上发射团队的组织指挥下,该公司在山东附近海域发射谷神星一号海射型遥三运载火箭,顺利将6颗卫星送入535公里高度太阳同步轨道,任务取得圆满成功。

本次发射的云遥一号15至17星,用户为天津云遥宇航科技有限公司,3颗卫星技术状态一致,配置全球导航卫星系统掩星载荷及长波红外相机,主要用于提供气象观测服务。吉天星A-03星作为“吉天星A”光学遥感卫星星座的首发星,配置高光谱遥感载荷,主要用于高分辨率光学遥感技术验证,由苏州吉天星舟空间技术有限公司独立研制与运营。苏星一号01星研制单位为上海埃依斯航天科技有限公司,用户为西北工业大学太仓长三角研究院,卫星配置光学相机载荷,主要用于光学遥感成像技术验证。天辅高分二号卫星研制单位为湖南航天卫星科技有限公司,用户为厦门天卫科技有限公司,卫星配置高光谱相机载荷,主要用于提供高光谱遥感数据服务。(付毅飞 张强 李宸)

新飞行机器人  
源于甲虫翅膀启发

《自然》日前发表了一项研究,分析了犀金龟如何展开和收缩后翅,表明这一过程无须肌肉活动。该发现有助于改进飞行微型机器人的设计。

在所有飞行昆虫中,甲虫的翅膀机制最为复杂,其包括两组翅膀:一对硬化的前翅,称为鞘翅,以及一组精细的膜质后翅。虽然对甲虫折纸式的翅膀折叠已有大量研究,但人们仍不甚了解它们是如何展开和收缩后翅的。过往研究认为胸部肌肉驱动了甲虫的后翅基部运动,但缺乏支持这一理论的实验证据。

瑞士洛桑联邦理工学院团队结合高速摄像机和动态相似的飞行机器人,来探索这一研究空白。团队观察到,犀金龟在展开和收缩翅膀时使用了被动机制。展开是一个二阶段过程,在这一过程中,甲虫的后翅会以类似弹簧的方式部分弹出,然后拍打的动作会将后翅上升至飞行位置。他们还发现,甲虫会用鞘翅来被动地降低后翅至静止位置。

受到这些观察结果的启发,团队制作了一个模仿甲虫翅膀被动展开与收缩的微型机器人。该机器人成功地起飞并维持了飞行。研究表明,将甲虫后翅的被动过程转化到机器人的扑翼设计,有助于改进需在有限或杂乱空间中运作的小型机器人的能力。(张梦然)

## 胡昌喜:深耕供应链 开拓“研产销一体”新蓝海

2013年成立的深圳市海利轩科技有限公司,正逢其时,以“助力中国制造出海,引领全球优质生活”为企业使命,业务范围覆盖美国、英国、德国、法国、日本等国家,自成立以来业绩每年以100%的增速高速发展。作为海利轩科技总经理,胡昌喜堪称跨境电商先行者,近年来他抢抓行业发展机遇,成功探索出独具特色的“跨境电商研产销一体”模式,为企业完成了产品交易端到生产端的整合。

跨境贸易的短板在于要看供应端的“脸色”。2022年TEMU全托管模式横空出世,席卷全球流量,这种模式下产品和供应链竞争力就显得更加重要。此时深耕跨境贸易近十年的胡昌喜及时抓住窗口期,创新开拓出“研产销一体”新模式,于当年成立喜利健公司,重点投入产品研发和工艺研发,并通过自

主研发生产来提升产品竞争力。

在胡昌喜的领导下,研发生产工厂建成投产后,2023年工厂年产值4000万,目前,工厂月产能十万件,2024年产值预计达8000万人民币。他每年推出二十余款拥有自主知识产权的新产品,基于B2C市场的及时数据反馈,快速立项研发、打板,及时精准解决客户需求,以精细研发赋予产品持久生命力,确保市场占有率遥遥领先。

在关注研发的同时,胡昌喜认为应该用供应链思维来对研发端进行思考,叠加贸易领域和生产领域的优势,深入优化供应链整合,聚焦核心物料供应链。一方面,他精准控制质量水平,降低成本,提高周转效率,减少资金占用。另一方面,他还广泛布局市场渠道,提高渠道销售能力,将自主研发产品供给内部B2C渠道销售的同时,也开放给更多

跨境电商B2C客户。

在研发过程中,胡昌喜提出产品应该“更聪明”,让“智能”成为产品的生命力。为此他带领团队在设计之初,就对市场进行了深入的调研,通过对用户的生活习惯、行为习惯跟踪调研,分析用户的需求点在哪里,同时深入医疗领域,了解康复和保健患者的特点,让设计团队不仅能够深度挖掘市场需求,还能更加深入地理解如何让产品更具有“超前”的特征。

近年来,胡昌喜已带领团队申请商标10多项,发明专利60多项,覆盖护具类、控制器类等,这些专利也先后转化为公司参与市场竞争的无可替代的技术优势。海利轩科技凭借广泛的行业知名度和影响力,先后斩获2021年阿里巴巴国际站SKA、AliExpress最佳合作护肤仪器品牌、2023年Gold Plus Sup-

plier Assessment Certificate等行业荣誉,他个人也荣获pingpong中国最有远见跨境卖家、出海者联盟最佳分享者等荣誉称号。

面对如此荣誉,胡昌喜认为民营企业应该担负起更多的社会责任,让自己实战所得能够影响更多的青年人。为此,胡昌喜与南京信息工程大学签署了校企合作协议,作为学校的优质就业基地为学校的应届生提供实习岗位。同时,他还作为学校客座教授贡献专家智慧,定时定期开展专题讲座,将自己多年积累的跨境电商实战运营经验做系统性分享,让学生从纸上谈兵跨入到实战案例。胡昌喜表示,跨境贸易的起点是货物的买与卖,随着中国在世界供应链体系中的身份不断变化,他期待为行业培育更多的年轻人,为行业创新与发展贡献重要力量。(刘智明)