

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之间是否存在哪些差异?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?带着这些问题,记者采访了有关专家。

AI创作套路化明显

“虽然大模型在不断发展迭代,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生鲍光胜说。

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向记者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,它每次都套用相同的套路——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,“零样本分类法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为弱化甚至移除。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。

记者了解到,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。

例如,西湖大学文本智能实验室团队在DetectGPT基础上研发的Fast-DetectGPT模型,可提升AI检测准确性,缩短检测时间。“Fast-DetectGPT与其他零样本分类器原理一致。其中一个创新点在于,我们提出通过条件概率曲率指标进行检测。”鲍光胜说,“与DetectGPT相比,Fast-DetectGPT在速度上提升340倍,在检测准确率上相对提升约75%。”

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。

(科技日报)

七部门发文
做好汽车以旧换新有关工作

新华社北京8月16日电 商务部等7部门16日对外发布关于进一步做好汽车以旧换新有关工作的通知。

根据通知,对符合《汽车以旧换新补贴实施细则》规定,报废旧车并购买新车的个人消费者,补贴标准由购买新能源乘用车补1万元、购买燃油乘用车补7000元,分别提高至2万元和1.5万元。对《汽车以旧换新补贴实施细则》印发之日2024年4月24日至2025年1月10日前提交的符合条件的补贴申请,均按本通知明确的标准予以补贴。其中,对已按此前标准发放的补贴申请,各地按本通知明确的标准补齐差额。

参加申请补贴的报废汽车所有人和新购置汽车所有人应为同一个人消费者,其所报废的国三及以下排放标准燃油乘用车或2018年4月30日前注册登记的新能源乘用车,应当于2024年7月25日前登记在申请人名下;自本通知发布之日起,在补贴申请审核期间,其所新购置的汽车应登记在申请人名下。

通知要求,优化汽车报废更新审核、拨付监管流程,按程序将补贴资金拨付至申请人提供的银行账户。各地要加快制定汽车置换更新实施方案,合理确定补贴标准、补贴条件和实施方式。

登录全国汽车流通信息管理系统网站或“汽车以旧换新”小程序,即可按要求申请汽车报废更新补贴、查询进度。

截至今年6月底
全国登记在册个体工商户
达1.25亿户

新华社北京8月16日电 截至今年6月底,全国登记在册个体工商户达到1.25亿户,占经营主体总量的66.9%。

这是记者16日从国新办举行的“推动高质量发展”系列主题新闻发布会上获悉的。

市场监管总局副局长蒲淳介绍,市场监管总局正在研究制定《个体工商户促进发展和规范登记管理规定》,明确市场监管部门促进个体工商户发展的重点举措,解决“个转企”登记、经营权继承、市场退出等重点难点问题,更好支持个体工商户长远发展。今年将建成全国“个体工商户名录”,让个体工商户的“家底”更加清晰、帮扶措施更加精准有效。

近年来,市场监管部门建立完善经营主体发展质量评价指标体系,部署促进民营企业发展22条举措,推动企业注销等“一件事”高效办理,精准有效支持个体工商户发展。2023年全国登记在册经营主体达到1.84亿户,企业活跃度同比提高0.69%,退出企业平均寿命延长0.64年。

按照国务院统一部署,市场监管总局持续深化“高效办成一件事”改革,企业和群众办事环节更少、周期更短、流程更优、成本更低。

市场监管总局综合规划司负责人王国伟介绍,企业信息变更环节由7个压减到1个,提交材料减少13份,审批时间减少6个工作日;企业注销办理环节由10个压减到4个,普通注销办理时间由原来最快70天缩短为目前最快46天,其中包含法定公告期45天;开办餐饮店由37个工作日压减到15个工作日,提交材料减少14份。

当前,越来越多的经营主体关注自身的信用状况。市场监管总局通过信用监管赋能经营主体高质量发展,以多种方式向各部门共享企业信用信息。

国家企业信用信息公示系统上线以来,日均访问量突破1.31亿人次,日均查询量超过1600万人次。广大消费者和市场交易对象可以通过系统快捷、全面地掌握经营主体相关信息及其经营状态,自主判断经营主体的信用情况。

据介绍,市场监管总局探索向平台企业开放共享信用监管数据,支持浙江、北京市海淀区试点在互联网平台企业开放共享信用监管数据,已在深圳、杭州等地开展向微信、支付宝核验个体工商户信息的试点工作,为强化与平台企业信用共治提供更多有益探索。

我国变速抽水蓄能机组
自主研制取得突破

新华社成都8月16日电 国家能源局能源领域重大技术装备项目——300MW级变速抽水蓄能机组发电电动机1:1转子16日顺利通过飞逸试验,转子动态稳定性、可靠性等性能得到充分验证,标志着我国变速抽水蓄能机组自主研制取得突破。

变速抽水蓄能机组发电电动机转子具有动态稳定性、可靠性要求高等特点,是机组最核心的部件,也是科技攻关最难啃的“硬骨头”。为此,承担自主研制任务的东方电气集团创新采用端部整体护环固定结构,应用新型转子绕组接线方法,开发超大直径非磁性金属护环,攻克了高转速大尺寸绕线式转子结构的设计开发、工艺制造、装配安装等难题,确保了整个转子系统的稳定性和可靠性。

记者了解到,新能源产业的迅速发展对电力系统灵活性提出了更高要求。变速抽水蓄能机组运行转速在一定范围内可调,具备运行灵活、稳定可靠、反应迅速等特点,适用于新型电力系统构建需要。

抽水蓄能作为安全、稳定、经济的储能方式,在用电低谷将电能转化为水的势能,在用电高峰时将水的势能转化为电能,相当于一个巨型“充电宝”,可以作为电网的稳压器、清洁能源的存储器。

广东阳江
千帆竞发迎开渔

8月16日,渔船驶离广东省阳江市闸坡国家级中心渔港(无人机照片)。

当日,随着伏季休渔期结束,停泊在广东省阳江市闸坡国家级中心渔港的渔船纷纷出港,开始进行捕捞作业。

新华社发



江苏:数字影像上“云”让跨院就医更便捷

患者跨院复查无需拎着片子到处奔波,医生可随时调取患者不同阶段的影像资料进行比对,医疗机构的传统物理胶片耗材明显减少……近年来,江苏省持续推进省卫生健康云影像平台建设,实现省内影像检查结果互认,患者省时、省心、省钱的同时,医生的诊疗效率和准确性也随之提升。

今年暑期,家住江苏连云港的3岁女童欢欢(化名)在当地医院接受CT检查时,影像结果提示存在骨质破坏,疑似骨肿瘤或骨纤维结构不良,这让她的父母十分紧张,一家人随即赶到南京市儿童医院进行咨询。

欢欢的父母本以为跨市就诊需要再给孩子做一次CT检查,却发现医生可在电脑中直接调阅欢欢在连云港拍摄的影像资料,局部图像还能放大缩小,比传统物理胶片更加清晰。医生看过图像后,很快判断这是儿童发育中的正常现象,无需过多干

预,这让欢欢的父母松了一口气,他们表示,“这项改革很便民,不仅可以让我们尽快知道结果,孩子也能少受辐射。”

南京市儿童医院放射科主任杨明介绍,婴幼儿和儿童是辐射损伤的敏感人群。通过查询云影像平台,医生可以调阅患儿近年来在全省公立医疗机构和部分民营医疗机构做过的影像检查数据,帮助他们避免不必要的重复影像检查。“我院还在诊断报告上添加了一个二维码,患者即便在省外就医,也可以扫码查看我院影像检查的图像及报告,或将二维码发给专家进行远程会诊。”杨明说。

记者在采访时了解到,江苏省卫生健康云影像平台的建设不仅可以方便群众跨院就医,还能有效提升医疗机构质控水平和医生诊断的准确性。

在江苏省人民医院的放射科读

片室,放射科主任吴飞云正和同事进行医学影像分析。吴飞云介绍,他们所看的病例是一名年近七旬的女性患者,在云影像平台输入信息获得授权之后,可以查询该名患者近年来在省内乡镇卫生院、县级医院和省人民医院所做的所有影像检查。与此同时,患者的检验报告、用药记录、医嘱记录等信息均能在该平台显示,方便医生迅速知晓患者的整体情况。

“这名病人患有中枢神经系统淋巴瘤,通过与她此前影像资料的对比,我们可以分析病灶几年来的变化情况。”吴飞云说,随着医学影像技术的发展,一次检查会生成上千张图像,传统胶片只能容纳关键角度图像,影像数字化则可以保证检查结果完整呈现,帮助医生做出更加精准的判断。

据江苏省卫生健康信息中心副主任唐凯介绍,近年来,为打通群众就医“堵点”,江苏在全省公立医疗机

构和106家民营医疗机构完成江苏省卫生健康云影像平台建设,接入平台的医疗机构给患者做完影像检查、出具影像报告后,数据资料可在30分钟内上传至平台。截至8月15日,云影像平台实现报告存储约11901万份,医生通过工作站调阅影像数据共计约1183万次,按照云影像平台全面接入以来的运行情况测算,每年可节约影像检查费用22.4亿元。

与此同时,江苏还积极推动影像数据的智慧化应用,利用人工智能技术开展对包括肺结节在内的肺部疾病的早期筛查,帮助预防和控制疾病传播,目前已对全省60家社区卫生服务中心的37951张正位胸片进行了人工智能阅片。

唐凯说,医生看诊的病例越多,经验越丰富,云影像平台同样如此,数据越多、应用场景越丰富,就越“智慧”。(据新华社南京8月16日电)

关于对2024年苏州市环卫行业道路机械化清扫职业技能竞赛路段实施临时交通管制的通告

2024年苏州市环卫行业道路机械化清扫职业技能竞赛将在我市举行,为确保比赛的顺利进行,根据《中华人民共和国道路交通安全法》及其实施条例等相关规定,决定对科教新城南渡路(荆石路-复游城路)、江申大道(荆石路-复游城路)、南星路(荆石路-复游城路)、及复游城路(南渡路-南星路)、复游城路(南星路-文昌路段

北半幅)实施禁止车辆、行人通行的临时交通管制措施,管制时间为2024年8月25日至2024年9月15日。请过往车辆驾驶人提前规划出行安排和通行线路,服从现场执勤人员的指挥疏导。

太仓市城市管理局
太仓市科教新城管理委员会
太仓市公安局交通警察大队
2024年8月17日



拍卖公告(202416期)

我受委托,将于2024年8月26日12时前对下列标的在互联网的太仓得金网(www.tcdj.cn.com)进行网上公开拍卖。

- 新毛垃圾中转站内报废垃圾压缩机一台。
- 浏河镇新塘镇132号原清宇特种塑料公司内报废珍珠棉生产线等设备一批。
- 显示屏、工控主机、投影机、电脑、打印机、空调等报废办公教学设备;彩超、血球计数仪等报废医疗设备;发电设备、摩托车、桌椅等报废物资一批。
- 空调、电子屏、音响、电视机、电脑、桌椅、沙发等报废物资一批。
- 城厢镇兴业北路150号4幢第1-3层约305.1m²;第3层约300m²两年期承租权(分两个标的)。
- 浮桥镇荷池路10号荷池花园商业

- 街1楼109、114号;2楼201-204、222-223,共8间商铺租赁权(分4个标的)。
 - 太仓市区70处公共自行车亭棚橱窗广告位壹年期使用权。
- 竞买须知:凡有意参加竞买者,请于拍卖截止日10点前,交纳参拍标的保证金,并持有效证件及报名资料至我公司办理网上竞拍及出价手续。
- 苏州太仓产权交易拍卖有限公司
地址:太仓市公正路11号联合大厦六楼
咨询电话:53588642
18915795588
详情可扫描二维码查看或咨询。



东盛弄施工通告

因娄东街道东盛弄改造工程施工需要,为保障工程的顺利进行及施工期间的道路交通安全与畅通,根据《中华人民共和国道路交通安全法》及其实施条例等相关规定,对娄东街道东盛弄实行半幅封闭施工半幅双向通行的交通管理措施,施工期限:2024年8月23日至2024年10月30日。请过往车辆、行人经过该路段时按照交通标志指示信息通行,并服从现场执勤人员的指挥和疏导。

太仓市人民政府娄东街道办事处
太仓市公安局交通警察大队
2024年8月14日