AI检测AI：“矛”更利还是“盾”更坚

AI世界

◎本报记者 吴叶凡

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？带着这些问题，记者采访了有关专家。

2024世界人工智能大会暨人工智能全球治理高级别会议上，观众在参观由人工智能生成的图片。 视觉中国供图

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。

2024年上半年 我国智能手机销量同比增长4%

科技日报讯（记者**罗云鹏**）记者从日前举行的“Counterpoint行业对话：中国手机市场上升周期洞察”主题沙龙获悉，中国智能手机市场回暖，今年上半年中国智能手机销量同比增长4%，预计全年销量可回升至2.7亿部以上。

Counterpoint是一家市场调研机构。该机构中国区研究负责人齐英楠介绍，2024年上半年，中国智能手机市场前6名分别为vivo、荣耀、苹果、华为、OPPO及小米。

根据该机构数据，vivo与荣耀手机销量同比增长5%和9.6%，均高于行业平均水平。“只有用户满意，vivo才有能力、有可能持续让员工、合作伙伴和股东满意。”vivo副总裁、vivo中国区总裁程刚说。

苹果手机销量同比下降13%。通过降价促销，苹果在3月和6月两个时间段保住了16%以上的市场份额。

齐英楠介绍，在中国智能手机市场，OLED（有机发光二极管）屏幕占比已超七成，半数以上智能手机采用5000万像素主摄像头，国产品牌正推动大容量电池普及。

值得一提的是，人工智能、折叠屏等已成为中国智能手机厂商竞逐的新赛道。荣耀、vivo、OPPO、小米等均推出手机端侧自研大模型，vivo、OPPO等厂商还在开发语音和图像处理大模型。

Counterpoint预计，2025年中国智能手机增长动能将主要来自产品力的提升。齐英楠说，基于对领先厂商的观察可以发现，伴随用户消费观念和需求的升级，高端化和人工智能等前沿技术已经在实践中展示出对品牌增长的强劲推动力。在此背景下，手机厂商纷纷在旗舰产品矩阵、产品形态功能创新、品牌建设等方面加大投入，以把握新的增长机遇。比如，不少手机厂商都在影像赛道上发力。影像看似是拍照，实则涉及算法、模型等多方面，其更新迭代看似是性能的提升，其实也是在满足手机用户的消费需求。

与会专家认为，随着生成式人工智能应用场景的普及，消费者有望认可人工智能手机概念，有助于提升中高端手机用户换机意愿。

第三届世界元宇宙大会将在武汉举行

科技日报讯（记者**操秀英**）8月14日，记者从中国仿真学会获悉，第三届世界元宇宙大会将于11月22日—24日在湖北省武汉市举行。

本届大会采用“1+2+N+X”模式，即举办1场开幕式、2场全体会议、N场主题论坛、X场特色活动。其中，2场全体会议为大会主旨报告和主题报告，N场主题论坛分别是元宇宙关键技术论坛、生成式人工智能论坛、工业元宇宙论坛、汽车元宇宙论坛、文旅元宇宙论坛、教育元宇宙论坛、医疗元宇宙论坛、空间计算技术论坛等，X场特色活动包括元宇宙投融资路演活动、元宇宙圆桌论坛、科技成果展览及体验等。

近年来，相关部门出台一系列政策推动我国元宇宙产业发展。2023年，工业和信息化部、教育部、文化和旅游部、国务院国资委、国家广播电视总局等五部门联合印发《元宇宙产业创新发展三年行动计划（2023—2025年）》，提出强化人工智能、区块链、云计算、虚拟现实等新一代信息技术在元宇宙中的集成突破，推动智能生成算法、分布式身份认证、数据资产流通等元宇宙关键技术在国家重大科技项目中的布局。

在位于江苏苏州昆山市的昆山元宇宙产业园元宇宙展示厅里，参观者体验VR游戏。 新华社记者 杨磊摄

盘古大模型将药物设计效率提升33%

人工智能加速数字医疗时代到来

◎本报记者 崔爽

“医药卫生改革发展进入新时期，以数字健康为动力推进健康中国建设迈入新阶段。充分运用数字技术和人工智能将提升人民健康水平，推动中国乃至全球数字医疗时代加快到来。”近日，在海南博鳌举行的华为云医药健康AI高峰论坛上，海南省卫生健康委员会副主任张毓辉说。

人工智能正在为医药健康产业注入新活力。华为云副总裁黄瑾认为，国家政策持续支持生物医药，医药健康企业纷纷增加科研创新投入，但我国在创新药研发等领域与国际先进水平相比仍有差距，人工智能快速发展为国内医药健康领域提供了赶超新契机。

华为云中国区副总裁、首席营销官张鹏说，近年来大众对健康愈发关注，线上平台迅速延伸至医药行业，不断压缩药企和线下零售药店空间。在此背景下，企业亟须重构自身能力，数智化或将成为破局关键。

记者了解到，华为云致力于赋能生命科学、药械企业、医疗健康三大场景，加速医药健康产业智能化跃迁。例如，在药物研发领域，针对药企在药物发现、临床试验、基因分析等环节面临的难题，华为云人工智能可助力药企实现突破；在药械领域，华为云能帮助药企构建研、产、供、销、服全链路的数字化能力；在医疗健康领域，华为云可助力医疗服务平台构建全栈自主的智慧医疗场景能力，打造诊前、诊中、诊后的人工智能应用，提升医院及区域医疗服务水平。

华为云大数据与人工智能领域总裁尤鹏介绍，华为云盘古大模型正在深耕医疗健康行业，解最难的题、做最难的事，用人工智能守护健康。盘古药物分子大模型全新升级，新增靶点口袋发现、分子对接、分子属性预测、自定义属性建模等十大人工智能制药核心场景，将药物设计的效率提升33%，实现早研阶段的全流程加速。

长安链推出全场景技术平台

科技日报讯（记者**崔爽**）记者8月14日从国家区块链技术创新中心获悉，我国首个自主可控的区块链软硬件技术体系——长安链正式推出全场景技术平台ChainBridge“链桥”。该平台支持所有异构、同构的区块链完成协作，可满足跨领域、跨地域、跨行业、跨层级的任意类型区块链之间数据互联互通需求。这对于破解目前国内区块链“各自为链”带来的孤岛难题具有重要意义，有助于加速聚链成网，织密国家级区块链网络，实现全国数据要素可信、高效流通。

作为前沿信息技术，区块链既能利用非对称加密和冗余分布存储实现信息不可篡改，又可利用链式数据结构实现数据信息可追溯，在数字经济各领域有很高应用价值。随着区块链技术发展，全球各地区、各行业正在加快形成区域链、行业链。然而，由于企业间存在技术差异，“各自为链”带来的孤岛现象日益突出。安全、易用、功能更强的区块链协作技术成为扩展区块链应用边界的关键。

长安链研发负责人介绍，同构区块链是指基于相同区块链产品构建的区块链网络，异构区块链则是基于不同区块链产品构建的区块链网络。区块链协作技术的关键，就在于让异构区块链便捷地实现数据交互。ChainBridge“链桥”团队创造性地将区块链之间的协作功能标准化、规范化，可在不干扰业务链与智能合约的情况下，实现多个异构链之间协作、互通，大大降低区块链间相互协作的开发与运维成本。开发者不需要对原有链系统进行任何改造，只需使用通用网关就能实现需求。

为确保安全性，在ChainBridge“链桥”平台上，区块链之间协作产生的每笔交易和数据都在“桥接系统”中完成链上存证，确保可追溯不可篡改。ChainBridge“链桥”还支持非链系统，数据库等传统中心化系统也能通过协议对接。

作为长安链技术体系的一部分，ChainBridge“链桥”已原生集成到系统合约中。“如果是用长安链作为底层技术开发的应用型区块链，可直接通过已开发好的平台满足区块链之间的协作需求，用户不必支付任何开发成本。”长安链负责人补充。

2024政法智能化建设技术装备及成果展上，观众参观“区块链+大模型”展台。 视觉中国供图

图说智能

智能警务机器人“上岗”

科技日报讯（记者**王禹涵**）记者8月13日从西安市公安局高新分局获悉，1名移动式智能机器人和1名AI智能机器人近日在白沙路暖心警务会客厅正式“上岗”。移动式智能机器人具有自主巡逻、全景监控、智能识别、宣传防范、一键报警等功能；AI智能机器人具备迎宾服务和导览讲解等功能，可对公安行政服务的2246个事项提供准确引导。图为AI智能机器人。 李聪迎摄