AI检测AI：“矛”更利还是“盾”更坚

热点透视 redian toushi

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，笔者采访了有关专家。

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笔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。

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

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。

笔者了解到，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。

吴叶凡

攻克繁育难题　文昌鱼变身模式动物

不久前，笔者在厦门大学生命科学学院细胞应激生物学国家重点实验室的文昌鱼鱼房内看到，不同规格的水桶和鱼缸整齐排列，泵氧设备不停运转，文昌鱼正在这里健康生长。

文昌鱼作为无脊椎动物向脊椎动物进化过程中的重要过渡类群，是研究脊椎动物起源的理想模型。

长期以来，在实验室建立可持续繁育的文昌鱼种群，一直是世界各地实验室文昌鱼研究者努力的方向。如今，我国研究人员通过自主可控技术，调控水质、温度、光线等，实现了文昌鱼的人工繁育和精准生殖调控，为揭示更多关于脊椎动物起源和演化、基因表达调控等生命科学领域核心问题提供了稳定的研究样本。

“我们已搭建起具有国际领先水平的文昌鱼研究平台，基本实现文昌鱼实验动物化主体框架搭建。在此基础上，我们以文昌鱼为模型，在脊椎动物中枢神经等重要性状起源机制研究方面，取得可喜的成果。”厦门大学生命科学学院细胞应激生物学国家重点实验室教授李光在应邀前往芝加哥大学作学术报告时说，文昌鱼作为一种野生种群遗传多样性极高的海洋动物，要将其驯化成理想状态的实验室模式动物，仍需要许多基础性研究工作，也期待更多研究人员共同参与。

科研价值重要研究样本稀少

文昌鱼是一种两端尖细、身体透明、体长为3~5厘米的海洋生物。它们形似小鱼，在茫茫大海中并不显眼。但在生物学家眼中，它们却是一种极为特殊的海洋动物。

“文昌鱼并不是真正的鱼，它们没有脊椎骨，体内仅有一条脊索，没有明显的头部结构，是无脊椎动物向脊椎动物演化的过渡生物。”厦门大学生命科学学院教授王义权介绍，作为仍存活于地球的最原始脊索动物之一，文昌鱼也被称为反映演化进程的“活化石”。

“文昌鱼的身上保留着许多原始的宝贵信息，能够帮助我们理解脊椎动物的起源及其演化历程。”李光说，文昌鱼在演化链条中处于从无脊椎动物向脊椎动物过渡的关键阶段。在胚胎发育上，文昌鱼早期与无脊椎动物相似，但后期与脊椎动物相似；在基因结构上，文昌鱼基因组未发生大规模加倍，多数基因以单拷贝存在，是研究脊椎动物起源的理想动物模型，具有十分重要的教学和科研价值。

然而，栖息地环境的改变使文昌鱼种群数量呈下降趋势。目前，全世界仅存30余种文昌鱼种群，文昌鱼也被列为我国二级野生保护动物。因此，想要获得文昌鱼样本开展研究并非易事。

王义权说，尽管文昌鱼在我国厦门、青岛等沿海地区均有分布，也有自然保护区为其繁衍保驾护航，但获取胚胎材料受季节性繁殖限制等因素，为开展文昌鱼相关基础研究带来极大困难。

因此，要想破题，在实验室建立可长期繁育的文昌鱼种群成为关键。

开展技术攻关缩短繁育过程

为推动文昌鱼研究持续深入开展，厦门大学生命科学学院文昌鱼研究团队展开文昌鱼人工繁育技术攻关，逐步搭建以文昌鱼为动物模型的研究平台。

“中国的文昌鱼最早发现于厦门，这为我们开展文昌鱼研究提供了得天独厚的地缘条件。”王义权回忆，相关技术摸索始于二十多年前。

彼时，科研人员实地观察野生状态下文昌鱼的生长发育情况，几乎每月下海测量文昌鱼栖息地的水质、水温和盐度等环境因子，如此坚持数年。与此同时，他们同步在实验室建立稳定的文昌鱼及其饵料培养技术体系，率先在实验室成功繁殖出文昌鱼子二代。

要使文昌鱼成为可应用的实验室模式动物，仅仅实现其全人工繁殖还不够。李光介绍，开展研究需要常年获取文昌鱼新鲜胚胎材料，而多数文昌鱼一年仅产卵排精一次，且时间比较集中，采样窗口期短、难度大。

对此，团队在文昌鱼生殖调控方面继续开展研究，通过优化养殖温度、密度等条件，加速文昌鱼个体生长和产后修复。最终，团队攻克了文昌鱼产卵、产精诱导技术，实现文昌鱼一年多次、不受季节限制产卵。

“现在，我们可以根据实验需要，随时获取新鲜的文昌鱼胚胎材料。同时，我们还将文昌鱼的代时（从受精到性成熟）由原来的超过1年缩短为现在的3~6个月，将单一个体的产卵（精）频次由原来的1年1次，缩短为现在的半个月至2个月左右1次。”李光介绍。

此外，厦门大学生命科学学院将科研优势转化为教学优势，利用3D建模、虚拟仿真等技术，重建文昌鱼的结构、胚胎发育和形态特征等模型，开展文昌鱼成体解剖、野外采集、室内人工养殖及繁育等虚拟仿真实验教学。

跻身模式动物应用前景广阔

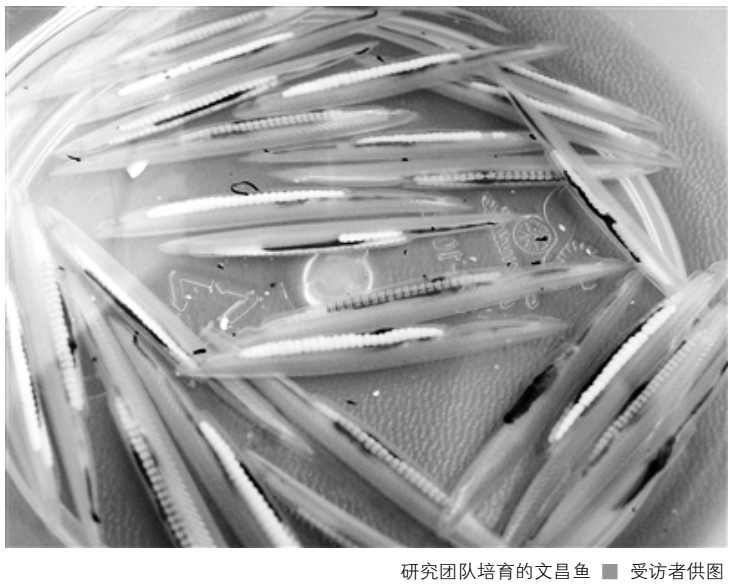
近年来，在建立稳定的实验室养殖系统基础上，厦门大学生命科学学院文昌鱼研究团队进一步建立了高效稳定的文昌鱼基因敲除技术，获得世界首个基因敲除突变体文昌鱼，在脊椎动物胚轴形成、中枢神经发生机制起源等重要基础科学问题研究方面取得系列成果。“现在，文昌鱼模式动物应用体系已基本建成。”李光说。

一直以来，模式动物开发都是推动生命科学进步的重要手段。从果蝇到小鼠再到斑马鱼等，这些模式动物不仅帮助科学家理解基因如何控制生物体的生长和发育，还为疾病研究和新药开发提供了关键线索。国际上有多个研究联盟和资源共享平台针对这些较为成熟的模式动物开展相关研究，但以文昌鱼作为模式动物的研究还有待发展。

研究人员认为，包括人类在内的脊椎动物拥有现生生物中最为复杂的中枢神经系统，揭示其如何起源一直是进化发育生物学研究领域的重要命题。文昌鱼进化地位独特，基因结构简单，作为研究脊椎动物复杂形状起源演化过程的模式动物，应用前景广阔。

在李光看来，目前他们搭建的文昌鱼研究平台拥有自主研发的文昌鱼无沙养殖系统，可养殖不同阶段的文昌鱼幼体和成体，而且大大简化了文昌鱼养殖流程。同时，该平台已囊括近百个文昌鱼突变体，为深入研究文昌鱼基因功能奠定了坚实基础。研究平台作为一种可复制、可借鉴模式，有望推动文昌鱼作为新型模式动物规模化发展，并能对其种群资源保护作出更大贡献。目前，厦门大学生命科学学院正在扩大和推广文昌鱼应用于更多学科领域，吸引更多实验室应用这一新型模式动物。

符晓波



研究团队培育的文昌鱼 ■ 受访者供图

创新杂谈 chuangxinzatan

不断深化职业教育产教融合

■ 谷业凯

前不久，欣旺达电子股份有限公司牵头联合84家企业发布了超1500个学生学徒和技培生岗位，合力培养新能源电池、智能制造等领域的高技能人才。作为广东省“产教评”技能生态链新能源产业的“链主”，这家企业通过与中高职技工院校、技术应用类本科院校、生态企业等联合，开展高技能人才前置培养、实训基地建设、职业技能等级自主评价等，积极培养高技能人才。

党的二十届三中全会《决定》提出“统筹推进教育科技人才体制机制一体改革”“加快构建职普融通、产教融合的职业教育体系”，为进一步推动职业教育发展、打造新型劳动者队伍指明了方向。职业教育是培育工匠人才的沃土，对于服务产业转型升级、培育发展新质生产力、助力推动经济社会高质量发展具有重要意义。

目前，我国已建成世界上规模最大的职业教育体系。乘着新时代职业教育的东风，大量中职、高职院校的毕业生成长为有技能、有学历、高素质的高技能人才，在现代制造业、战略性新兴产业等发展中发挥着重要作用。从“嫦娥”揽月、“羲和”逐日，到“蛟龙”深潜、“北斗”组网，再到最长的跨海大桥、世界领先的高速铁路……这些大国重器、超级工程的建设过程中，都印刻着高技能人才追求卓越的奋斗足迹。

人是科技创新最关键的因素。新一轮科技革命和产业变革加速演进，职业教育与产业结合的重要性更加凸显。培养好大量高技能人才，将为科技创新和高质量发展蓄势赋能。目前，我国职业教育与产业发展的“耦合”作用还不够强，产教融合“合而不融、融而不深”的问题仍较突出。从顺应科技发展趋势、满足产业转型升级的要求看，培养高技能人才，亟须形成产教融合、优势互补的协同育人模式，使之与产业结构、社会需求高度契合。

产教融合是职业教育的基本办学模式，也是实现产业链、创新链、人才链有机衔接的重要举措。深化产教融合，既要从学校端发力，开设更多符合市场需求的紧缺专业，形成紧密对接产业链、创新链的教学体系，提升办学质量；也要从企业端入手，通过开展校企合作办学、共同设立实习实训基地等方式，吸纳企业深度参与专业规划、课程设置、教材开发、教学实施等；也要做好政策引导，在全社会形成联动效应，探索建立招生、培养、评价、就业一体的全周期培养机制，不断壮大高技能人才队伍。

科技创新和产业转型升级的步伐不断加快，展示着经济社会发展的活力，也是职业教育发展的潜力和空间所在。进一步深化产教融合、健全协同育人模式，着力培养造就更多卓越工程师、大国工匠、高技能人才，提升整体人才素质，将不断壮大支撑中国制造、中国创造的重要力量。

川渝四地打造科学仪器共享“朋友圈”

“以前做检测需要把样本发到北京鉴定、测评，耗时很久，现在就近检测，基本当天就能拿到报告。”8月16日，在重庆市潼南区重庆檬泰生物科技有限公司，科技研发负责人刘雁成告诉笔者，该团队通过涪江流域科技创新走廊大型科学仪器共享平台，实现了邻近川渝地区科学仪器共享，极大加快了公司的科研进程和发展步伐。

刘雁成所在的公司是一家从事柠檬精深加工的高新技术企业。在日常生产中，该公司要对柠檬果胶分子分级鉴定和柠檬酯化度进行测定。这需要用到特殊液相色谱质谱联用仪，但公司目前还没有引进这项设备。在共享平台建立后，利用平台资源他们很快找到遂宁市的一家公司，当天就拿到了检测报告，检测周期比之前缩短一半以上。

为深入推动成渝地区双城经济圈建设，协同构建涪江流域科技创新走廊，重庆市潼南区与合川区、铜梁区及四川省遂宁市四地（以下简称“川渝四地”）协同建立涪江流域大型科学仪器设备共享平台，积极为科技企业及高校科研院所“架桥铺路”，打造科学仪器共享“朋友圈”。

潼南区科技局相关负责人介绍，涪江流域科技创新走廊大型科学仪器共享平台成立于2022年。平台为企业提供涵盖仪器共享、仪器研发、仪器首发、报告溯源、认证培训等的“一站式”服务，按照“四地协同、共建共享”的总体定位，促进跨行业、跨区域的协同创新。

通过“线上+线下、公益+市场、自营+中介”的共享模式，川渝四地逐渐形成跨区域、跨领域、多层次的开放共享服务体系，科学仪器利用率提升近30%。

此外，川渝四地设立仪器共享检测专项创新券，率先实现川渝四地创新券通用通兑，给予仪器检校费用60%的补助，有效缩短企业检测周期50%以上，降低检验检测成本超过20%。

截至目前，平台已集聚高校科研院所、医院、科技企业等成员单位380余家，实现1万台（套）大型科研仪器设备资源共享共用，科学仪器共享率达到60%以上，可开展指标检验检测1万余项。

雍黎

（上接A1版）在大同，网络服务以“服务进万家”为契机，今年上半年主动服务用户超过1.75万户，用户满意率超过90%。太原移动创新运用心级服务二维码，通过邀评送流量的方式，将触点用户转化为友好用户，进一步提升了客户满意度。

安全度汛，守护每一刻。面对极端天气等特殊情况，山西移动更是展现出了高度的责任感与使命感。筛查全省核心机楼、县局机房、传输骨干机房防内涝能力，加固杆路设备；省市县各级均与应急、水利、气象、水库等部门建立了点对点联络机制，实时掌握汛情和泄洪信息，严格落实7×24小时应急带班值班制度；绘制传输路由结构图和纤芯图，配置传输路由现场指导员；应急响应枕戈待旦，全部装备均检修保养完毕，通信保障人员在岗待命。

针对洪灾道路中断后运输困难的问题，今年山西移动还提前配置运输抢修类多旋翼无人机11架，在防汛关键期展现出高度的责任感与前瞻性。

匠心筑网，共绘美好。山西移动技术人员以匠心为引领，涉深涧、爬高塔、探洞穴、下矿井，用辛勤的汗水织就了一张技术优、覆盖广、网速快、体验好的精品网络。在吕梁市临县李家村，移动5G网络助力当地主播“跨界”直播卖货，为乡村振兴注入了新的活力；在方山生态文化旅游示范区，5G+VR/AR等新技术让游客在手机上即可“云”游方山，感受廉政教育的深刻意义。

山西移动正以5G为翼，以服务为心，深度赋能数字生活新篇章。

王蕾