

# 全模态大模型：连接“数据孤岛”的桥梁

## ——“解读人工智能前沿技术趋势”系列报道之二

◎本报记者 沈唯

11月26日，在湖北武汉举行的2024东湖国际人工智能高峰论坛上，“紫东太初”大模型3.0全球首发。

2021年发布的“紫东太初”大模型1.0是全球首个图文音三模态大模型，由中国科学院自动化研究所联合武汉人工智能研究院研发。2023年，“紫东太初”大模型2.0发布，其在原有三模态基础上，进一步融入视频、信号、3D点云等多种模态数据，具备全模态理解能力、生成能力和关联能力。

日前在2024年世界科技与发展论坛期间发布的《2024年人工智能十大前沿技术趋势展望》中，“全模态大模型：打破数据壁垒”被列入其中。与多模态大模型相比，全模态大模型有何不同之处？它将为相关行业带来哪些变化？科技日报记者就这些问题采访了相关专家。

### 延伸升级多模态

全模态大模型是一个综合性概念，指的是一种能够处理、理解和生成多种模态数据的人工智能模型。中国科学院自动化研究所紫东太初大模型研究中心常务副主任、武汉人工智能研究院院长王金桥解释，多种模态包括但不限于文本、图像、音频、视频、传感器数据，以及结构化和非结构化数据等。

“全模态大模型的核心目标是通过统一架构，完成多模态数据的感知、理解、生成和推理任务，提供通用解决方案，而无需针对特定模态单独开发模型。”北京科技大学智能科学与技术学院副教授王耀祖介绍，全模态大模型致力于实现模态间的无缝交互与信息融合，构建一个统一的智能体系。

相比于全模态大模型，多模态大模型这一概念似乎更为人所知。“从单模态到多模态，再到全模态是人工智能大模型技术发展的必然趋势。”王金桥认为，全模态大模型在多模态大模型的基础上，进一步融合了更多种类的模态数据，提升了模型的认知、理解和创作能力。因此，全模态大模型可以被视为多模态大模型的延伸和升级。

## 地球科学领域“元古大模型”上线

科技日报（记者吴纯 实习生邹文慧）12月7日，“元古大模型”在中国地质大学（武汉）正式发布。该模型综合运用图生文、文生图、文生文的生成式预训练模型架构，成功验证了在全国产化算力与算法基础上，利用多模态大模型解决地球科学领域问题的技术路线。

“元古大模型”由中国地质大学（武汉）地球科学学院教授宋海军团队与科大讯飞合作完成，旨在利用国产化算力

基础设施和多模态大模型集成能力，构建地球科学研究、教育教学和社会服务一体化智能化系统。

宋海军介绍，“元古大模型”的命名灵感源自地质时代的“元古宙”，是地球历史上存在时间最长的一个宙。团队对62万余张古生物化石图片、3000篇地学文献中有关地质实体及其属性的45万余条信息进行数据清洗与标注，构成模型基石。

目前，“元古大模型”已上线古生

## 音乐大模型市场“群雄逐鹿”

◎本报记者 崔爽

多样的音乐创作玩法、不断优化的模型效果，让音乐大模型成为今年垂直大模型赛道的一大亮点。

今年3月，音乐生成模型Suno V3面世，被一些业界人士称为音乐行业的“ChatGPT时刻”。最近，Suno发布V4版本，允许用户生成音质更高、更好听、时长达4分钟的音乐。这意味着，AI可以创作一首完整的歌曲。

不仅是Suno，国内AI音乐领域的资深“玩家”们也纷纷入局。在Suno V3发布后，国内大模型厂商纷纷上线类似产品，形成“群雄逐鹿”的格局。

### 产品层出不穷

记者梳理发现，目前国内AI音乐大模型企业主要有三类：以字节跳动豆包旗下海绵音乐为代表的大型科技公司，以昆仑万维旗下的天工SkyMusic为代表的新兴大模型厂商，以及以趣丸科技旗下的天谱乐为代表的垂直赛道公司。

海绵音乐主打利用AI技术帮助用户

生成个性化音乐作品。用户只需输入一句灵感或上传一张图片，即可生成专属音乐作品。相比Suno，海绵音乐生成的中文歌曲吐字清晰度和演唱流畅性更高，曲风更符合国人喜好。

天工SkyMusic基于昆仑万维“天工3.0”超级大模型打造，能够生成多种风格的音乐作品。它采用音乐音频领域类Sora模型架构，支持生成80秒44100赫兹采样率双声道立体声歌曲。

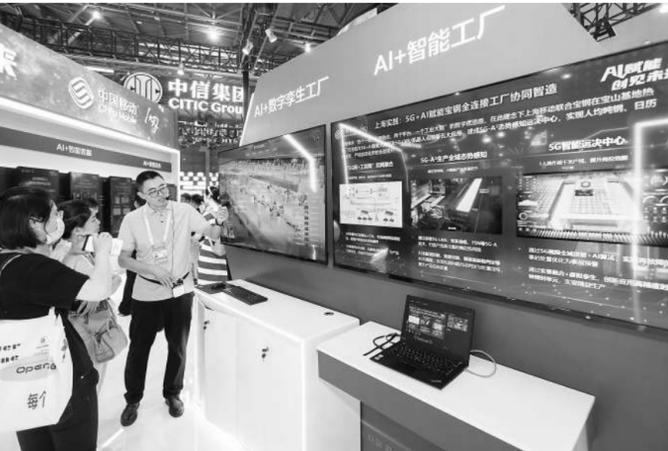
天谱乐于今年7月发布，是一个多模态配乐大模型。它集成图片理解算法、旋律生成算法、视频理解算法、配置识别算法等技术。值得注意的是，天谱乐不仅支持文生音乐、音频生音乐，还具备图片、视频生成音乐的功能。

此外，腾讯音乐和网易云音乐也推出了相关产品。

### 落地场景广泛

业内普遍认为，与全球巨头同场竞技，AI应用是国内人工智能企业的优势所在。国内市场需求大、可落地的商业化场景多，为国产大模型进化提供了坚实基础。

在趣丸科技副总裁贾朔看来，“降



2024世界人工智能大会暨人工智能全球治理高级别会议上，人们在了解AI赋能智慧工厂。

的数据类型，让分析更准确、决策更智能、控制更精准。

王金桥认为，在医疗领域，全模态大模型可以更高效处理各种模态的海量病历数据，进行临床症状描述并检验检查结果，为医生提供更加全面、精准的诊断建议；在交通领域，它将进一步助力智能交通系统建设，提升交通运输效率和安全性，有效缓解交通拥堵；在教育领域，它能为每个学生量身定制学习方案，提高学习效果和满意度……

“凭借超大规模参数和复杂计算结构，全模态大模型在多个行业中展现出巨大应用潜力，未来能够助力许多行业实现智能化、高效化和个性化发展。”王金桥说。

### 开发仍面临难题

受访专家认为，全模态大模型有利于进一步拓展人工智能的应用场景，但这并不意味着全模态大模型将完全替代多模态大模型。王金桥说，在某些特定应用场景下，多模态大模型已经能满足需求，这种情况就无需引入全模态大模型。

王耀祖也认为，多模态大模型通常针对特定模态组合进行优化，适用于特

定领域应用。全模态大模型虽然通用性强，但可能在某些领域表现不如专门设计的多模态大模型。同时，全模态大模型对计算资源的需求远高于多模态大模型。对于许多中小企业或资源受限的场景而言，多模态大模型仍是更为实际的选择。

“全模态大模型的开发与应用，本身还面临着一些挑战，涉及技术、数据、资源、行业接受度等多个方面。”王金桥说，全模态大模型需要大量高质量数据进行训练，收集、整理和标注这些数据可能耗费大量资源。数据的质量和多样性对大模型的性能有很大影响，而某些领域的数据可能难以获得或获取成本过高。

王耀祖还提到，大模型容易受到数据偏差的影响，可能生成带有偏见或有害的内容。此外，大模型在处理隐私信息时也存在一定风险。因此，开发全模态大模型时，还需要进一步研究、监测大模型中的潜在偏见，并且在训练过程中引入隐私保护机制，以保护用户数据安全。

“未来，我们需要优化数据采集与处理、简化模型结构、加强行业合作与定制化开发、建立伦理和监管框架，逐步克服难题，推动全模态大模型在更多领域应用。”王金桥说。

研究成果向社会服务转化。

古生物学是地质学的重要分支学科。技术方法的突破是带动科学创新的关键，未来古生物学的发展需要以技术发展为支撑。“我们将逐步上线地学智能搜索引擎和古生物科普数字人，为地球科学专业的教学、实践及科研活动提供服务支持。”宋海军表示，未来“元古大模型”将不断丰富数据类型、优化模型能力、开发创新应用，满足更加多样化的应用需求。

提供服务。对此，Suno进行回击，指责唱片公司滥用版权诉讼，企图借此维护自己在AI音乐市场的主导地位。如何在法律框架内实现高质量音乐音频数据的运营和商业化，是AI音乐公司未来必须思考的问题。此外，音乐大模型生成的音乐作品是否享有版权保护，其版权归属如何界定，尚无明确的法律规定，这给音乐大模型的应用带来了不确定性。

贾朔认为，可以利用技术手段规避一些纷争。比如在产品侧，当用户在使用AI创作时，可利用相关技术判断其是否有模仿现有艺术家作品的主观意图，并视情况做出弹窗提示。

同时，音乐的创意与情感是否可以被大模型实现也有待验证。如何提升大模型的创意能力，使其能创作出更具个性和创新性的音乐作品，是当前音乐大模型需要突破的方向之一。

此外，AI音乐“演唱者”的声音距离真人也有一定距离，解决这一问题需要相应的技术手段。昆仑万维董事长方汉说，天工SkyMusic通过自研模型技术，实现乐器、人声、旋律、音量、音符的一体化端到端生成，在人声合成、情感表达等方面具有一定优势。

## 首款依托鸿蒙系统的焦炉巡检机器人发布

科技日报（记者赵向南）12

月9日，山西戴德测控技术有限公司发布其研制的焦炉巡检机器人，这是国内首款采用边缘计算模型进行巡检识别的焦炉巡检机器人，也是国内首个依托鸿蒙系统研发的焦炉巡检机器人。

“传统焦炉巡检存在规范化程度不高、不能及时发现排查安全隐患等问题。”山西戴德测控技术有限公司总经理蔡岳松介绍，焦炉巡检机器人不仅能解决这些问题，还能高效完成焦炉废气盘、地下室等关键部位的巡检任务，如检查废气盘连接杆是否脱落、搬把是否到位等。

据介绍，这款机器人在巡检趟次、巡检里程、巡检速度、算法识别

率方面均有显著提升，已实现无故障累计巡检7000趟次、无故障巡检里程1800千米，巡检速度提升2.5倍，算法识别率高达98%。

这款机器人采用离线巡检技术，即使在无网络的环境下也能正常工作，确保巡检数据完整、准确。同时，它还引入单炉单机、多机单平台集控、可视化报告系统等技术，进一步提升了巡检效率和管理水平。

这款机器人采用国产化整机架构，国产化器件使用率75%以上，核心控制单元和机器人核心系统完全国产化。“凭借鸿蒙系统强大的兼容性和稳定性，这款机器人为焦炉巡检提供了更为安全和高效的解决方案。”蔡岳松说。

## 无人零售智能柜行业加速规范化发展

◎本报记者 罗云鹏

“无人零售智能柜内商品可选择种类比较多，基本能满足上班族日常临时需求。”在深圳前海一家公司工作的职员吴潇瑛日前接受科技日报记者采访时说，“这种即拿即走、自动结算的即时购物店，我和同事更愿意选择无人零售智能柜。”

在深圳，无人零售智能柜已在办公场所、产业园区、医院、学校、交通枢纽、物流中转站等场景中随处可见。

丰e足食是一家从事无人零售智能柜业务的运营商，线下终端规模达14万台。该公司首席技术官蔡亮介绍，无人零售智能柜主要使用射频识别、物联网、视觉识别、人工智能和移动支付等技术，由前端智能柜与后端智能运营两部分组成。前端智能柜主要面向使用者，包含刷脸屏、摄像头等部分；后端智能运营侧主要借助人工智能算法实现销售预测，动态调整商品结构和数量，同时根据各类细分消费场景需求进行决策。

由于具有使用便利等优势，无人零售智能柜行业前景广阔。相

关预测显示，预计到2025年，无人零售市场规模将接近2万亿元，覆盖消费群体或达2.5亿人。记者梳理发现，近年来，丰e足食、友宝在线等众多企业纷纷入局这一行业，但其中绝大部分仍是中小微运营商，运营体量多则千台、少则百台，市场竞争比较激烈。

业内人士指出，该行业在投放智能柜标准、智能识别技术、用户隐私安全保护等方面仍需完善，行业的规范化、标准化建设已提上议程，后续各类标准规范有望快速落地。

记者了解到，今年，多项与无人零售智能柜相关的团体标准出台。这些标准涵盖智能柜识别技术要求、用户交易视频隐私保护及智能零售动态柜技术规范等方面，旨在统一智能柜的投放标准，为消费者提供专业、安全的无人零售服务。

据介绍，丰e足食已率先实现个人信息与隐私数据的全生命周期管理。“加强用户隐私安全保护，制度管理非常重要。我们设立相关岗位，制定严密的信息安全管理规定，并对全体员工定期开展专业培训和考核。”蔡亮说，随着公众和企业信息保护意识的增强以及相关标准的出台，这一行业规范化程度将日益提高。

## AI可“拦”住电动自行车进楼入梯

◎通讯员 王锐娟

本报记者 王禹涵

“这是电动自行车安全监管监测平台中的AI识别功能。当平台监测到电动自行车进入电梯时，电梯门不会关闭。电梯不但会停止运行，还会发声提醒乘客不要让电动自行车进入电梯，直到其离开后电梯才会恢复正常。”12月6日，陕西省宝鸡市麟游县九成宫镇南社区党支部书记王妙向居民介绍

小区电动自行车充电桩300个，解决居民充电的后顾之忧。“麟游县消防救援大队参谋张柏林介绍，通过这些举措，小区电动自行车进楼入梯、飞线充电现象大幅下降，有效解决了电动自行车进楼难查、充电桩等问题。



工作人员正查看电动自行车安全监管监测平台的后台数据。受访者供图