

未来大模型：更多能 更轻量 更亲民

热点透视

rediantoushi

去年起,全球掀起以大模型为代表的生成式人工智能行业浪潮,人工智能迈向全面应用新阶段。国外以 OpenAI、微软为代表,谷歌、Meta 等巨头一个不落,国内头部科技企业则悉数入场。

近日, 国家数据局局长刘烈宏表示,中国 10 亿参数规模以上的大模型数量已超 100 个, 行业大模型深度赋能电子信息、医疗、交通等领域,形成上百种应用模式,赋能千行百业。

大模型技术日新月异,产业化脚步追星赶月。在这个热闹夹杂争议的行业,不同的技术路线和选择正在延展开来。

多模态正成标配

360 集团董事长周鸿祎在关于 2024 大模型发展的十大趋势判断中明确表示,多模态将成为大模型标配。未来,大模型不仅能听会说,还能看懂图片和视频,更能识别理解。

中国科学院自动化研究所研究员刘静在《GPT-4 对多模态大模型在多模态理解、生成、交互上的启发》一文中谈到,模态是指表达或感知事物的方式,例如人类的听觉、视觉、触觉等。在人工智能领域,多模态是指利用多种模态的信息来进行感知和理解。多模态技术可以让机器同时接收并处理不同模态信息,如文本、图像、音频等,从而提高机器感知和理解能力。

刘静进一步解释,相比传统的单模态大模型,多模态大模型更加符合人类的多渠道认知方式。它能将不同模态信息相互补充,提高信息的完整性和准确性,从而更好地应对复杂环境、场景和任务。如在语音识别中,多模态大模型可以结合语音和文本信息,让机器更准确地识别和理解语音内容。在图像识别中,图像和文本信息的结合可以让机器更深入地理解图像内容。

刘静介绍,落实到应用层面,多模态能使机器更好地理解人类的意图和需求,提供精准、个性化服务。例如,在目前智能技术已经深入落地的客服领域,多模态大模型可以根据用户的文本和语音信息,综合理解用户问题并提供解决方案。在智能家居领域,多模态大模型通过与智能家居设备的交互,可以根据用户需求智能调整家居环境。在医疗领域,多模态大模型则可以融合医学图像、病历文本等信息,辅助医生进行疾病诊断、

制定治疗方案。

结合国内大模型产业布局,多模态大模型是近年主流厂商普遍选择的发力方向,如“紫东太初”2.0 全模态大模型可实现文本、图片、语音、视频、3D 点云、传感信号等不同模态的统一表征和学习。

落地应用越来越“轻”

为提升大模型性能、增进用户体验,大模型正以肉眼可见的速度越来越“大”。开源模型的参数从几百亿增加到几千亿,闭源模型也在沿着尺度定律路径不断升级,在千亿参数的基础上向着万亿参数攀登。但与此同时,人工智能产业算力吃紧,难以满足大模型参数规模的指数级增长。

为加速大模型落地应用,人工智能领域正尝试对大模型进行轻量化,通过打造更小型、更高效、成本更低的模型吸引行业客户,让模型在更广泛的场景有更高应用价值。

大模型轻量化通过降低模型的计算复杂度和内存占用,实现模型性能与资源消耗的平衡。这不仅有助于提高计算效率,降低运行时资源消耗,还可以减轻计算系统的硬件和软件压力,提高系统的可靠性和稳定性。此外,轻量化还有助于提高模型部署的灵活性和可扩展性,为模型在各种场景下的应用提供更多可能性。

轻量化主要通过模型压缩来实现。模型压缩是指在不降低模型性能的前提下,减小模型的计算复杂度和内存占用。模型压缩的方法有很多,剪枝、低秩分解等是业界常用的方法。具体来看,剪枝是指删除一些不必要或冗余的参数,低秩分解是指将高维数据或高维模型分解为低维数据或低维模型。这些方法都是通过降低模型的复杂度,减少参数和计算量,来达到让模型“更轻”的目的。

轻量化不仅有助于提高计算效率,降低运行时资源消耗,还可以减轻计算系统的硬件和软件压力,提高系统的可靠性和稳定性。

对于通用大模型服务垂直领域、行业大模型落地应用来说,轻量化更是关键步骤——通过合理的数据准备、模型选择、超参数设置和训练验证过程,可以使模型更好地适应特定领域,提高模型性能。

联想集团副总裁、联想研究院技术战略与创新平台总经理王茜萋曾表示,用好大模型主要有三个要素,第一是场景化,第二是隐私化,第三是轻量化。轻量化不仅是将个人大模型进行压缩并部署到用户设备上,还包括企业内部轻量化的,甚至云端的轻量化。轻量化意味着使用成本的降低。



多模态将成为大模型标配,未来,大模型不仅能听会说,还能看懂图片和视频,更能识别理解。

端侧成厂商新“角力场”

去年以来,大模型正逐步走向“智能终端侧”, 国内外一些厂商纷纷宣布加速推进大模型在移动终端的部署。端侧大模型,正成为行业热词。

所谓端侧, 是指模型服务不部署在云端,而直接存储在终端内部的芯片中,利用芯片算力生成结果。这样的大模型服务不需要联网,数据也不需要被上传到云端。

相比于云端大模型, 端侧大模型的优势主要在于:无需云端处理信息,安全隐私性更好;不消耗云成本,高频使用下的成本更可控;弱网和无网环境下也可以使用,用户无需担心云端服务器宕机,交互体验更稳定。

中国工程院院士邬贺铨表示,通过模型压缩和定制人工智能芯片提升算力,将具有百亿参数大模型的推理能力嵌入手机,可以提供低成本、低时延、高安全的智能助手服务。

相比 ChatGPT、Midjourney 等人工智能应用依赖云端服务器提供服务,端侧大模型主打在本地实现智能化。甚至有厂商提出,让每个人在手机上都拥有“个人大模型”。

目前,在中国手机市场份额排名前五的企业中,除了苹果公司,其余均已发布自有端侧大模型产品。手机厂商提出的包括通话记录自动生成、人工智能图像处理等应用场景,日渐成为消费者的日常。

端侧大模型同样也是个人电脑产品发布时的高频词。在相关宣传中,端侧大模型不仅能够提升日常使用及办公效率,还是保护隐私和数据安全的最优解。

艾媒咨询 CEO 兼首席分析师张毅表示, 人工智能可能成为今年度的手机新卖点, 尤其是手机的社交价值功能呈现方面,将有更多故事可讲。对于手机厂商而言,大模型是公认的未来发展赛道和方向。尽管现在人工智能在智能手机终端的应用仍较为初级,但未来的普及和推广仍值得期待。

不过,现有技术条件下,要做好端侧大模型并不容易。目前条件下,手机性能远远不足以支撑大模型运行。对于大模型来说,参数量是模型能力的直观体现,如今云端大模型的参数量均在千亿级别,而手机端最高参数规模则不过百亿。

联想集团董事长兼 CEO 杨元庆认为,未来十年是人工智能的十年,人工智能将改造所有业务。将来人人都会有自己的智能体,后者通过交互学习越来越懂用户,最终成为每个人的“人工智能双胞胎”。

但他也不讳言市场的不确定性。在他看来,内置个人智能体需要更高配置,如果这些对用户不是刚需,可能就不需要这么高配置。

国际数据公司认为,未来人工智能终端将在创作与创意、游戏和虚拟世界、语音合成与转换、视觉和图像处理、聊天机器人等十大领域广泛落地发展。

崔爽



丰翼方舟 40 物流无人机在晋中左权上空亮相。■ 张文军摄

晋中左权：着力推动“空中送餐”项目应用

近日,在山西省晋中市左权县泽城村写生基地,一款型号为丰翼方舟 40 的物流无人机进入了最后调试及空运审批阶段。

左权县正着力推动“空中送餐”项目在中国北方国际写生基地应用。届时,从基地出餐到用户取餐,最快只需 10 分钟。

“将来在山上作画,外卖也能打飞的,再也不用担心吃饭问题了。”前来泽城村写生的学生郑景文听说“空中送餐”项目后高兴地告诉笔者。

据了解,承担送餐任务的无人机采用八轴设计,单次可携带 10 千克货物,平均飞行速度达每小时 50 千米,充满电最长可飞行 20 千米。

工程师李鑫介绍,目前“空中送餐”项目以泽城驿站为起点,配送东安山村、西安村两个配送点。无人机配送具有速度快、过程可控的优点,是即时物流运输的得力工具。

“本项目可以极大缩短送餐时间,省时省力。”李鑫介绍,物流无人机拥有高精度的自动降落及避障系统,能够平稳、准确、自动降落,降落精准度达厘米级。通过无人机信息回传系统,可以实时查看无人机飞行数据,监控快件安全。

“低空经济对于促进地区经济发展、提升交通运输效率、推动科技创新等具有重要意义。”山西省晋中市左权县相关负责人表示。

韩荣

荒山披锦绣 葡香引燕来

(上接 A1 版)多年来,如燕入治理荒山荒坡,造林绿化,营造生态养生环境,持续造林 4500 余亩,累计种植树种超过 30 种。与此同时,酒庄还不断探索防治病虫害和应对极端天气的方法,并在与大自然的紧密合作中为动物及鸟类提供了更理想的栖息地,极大地丰富了酒庄的生物多样性,吸引包括金腰燕、环颈雉、青羊等数十种野生动物在此繁衍栖息,以天然肥料反哺土地。适中的土地养分有利于葡萄藤向下扎根,更深的根系提高了葡萄的品质。

酒庄所处的大峪坪村自古就有“滴水如油”的说法,为了解决灌溉难题,如燕人遍访周边群山,终于在酒庄以南约 12 公里处的尖山深处,找到了清冽的山泉,其矿物质含量达到国家天然饮用水标准。筑坝、饮水、蓄水、滴灌,这项翻山越岭的“南泉北引”工程堪称浩大,仅滴灌网的总长度就超过了 450 公里。

从 2010 年至今,如燕酒庄根据实际情况在太谷区委、政府及乡镇有关部门的大力支持下进行基础设施建设。现有基础设施包括太邢线通往区域主路 7 公里、区域内干线 30 公里、大型蓄水池 4 个(单个容量 1000 立方米)、灌溉蓄水库 1 个(容量 40000 立方米)。铺设引水管线共 16 公里,引入区域内的自流山泉水每天 200 吨左右,年引入总量约 73000 立方米,这些基础设施不仅支撑了企业的发展,也为当地村民带来便利。

选种育苗抓管理

品种的选择是酿酒的关键,2010 年,公司高层开始参观考察国内外各类知名酒庄,引进赤霞珠、品丽珠、美乐、霞多丽等法国优质葡萄品种,2013 年春季完成所有葡萄种植,种植范围 1 千余亩,共 33.5 万株,同年荣获“县级龙头企业”及“繁荣县域文化先进会员单位”等荣誉称号。

“田地管理是葡萄酒酿造的第一车间,采用天然有机肥是我们的种植理念。”山西如燕酒业有限公司总经理孙建伟介绍,“酒庄每年使用 1500 立方米有机肥料改良 300 余亩土地,并采用隔年施肥的方式来平衡土地养分。专业的葡萄种植师与种植团队共同践行对水、空气和土壤的保护,为可持续发展创造条件。”

人工采收是高品质葡萄酒的基础工序,温和的采摘手法可以最大限度减少对果实的损伤,同时让葡萄在田间完成第一次筛选。为了精选出品质上乘的山地葡萄,如燕比其他酒庄多付出了近两倍的时间和人力。“我们与 100 多名农民工签订了雇佣合同,帮助我们修剪、采摘、筛选、搬运,虽然严苛的筛选标准会限制葡萄的产量和增加额外的劳动量,但我们相信,正是这一切的精益求精,才会让如燕葡萄酒成为独具个性的甄选佳酿,拥有世界一流品质。”孙建伟说。

2019 年,如燕酒庄引进国外先进的前处理整套设备,严格按照国际 OV 酿酒标

准,以品种、产地、年份三大品质路线生产葡萄酒,于同年 10 月投产。

“我们引进了法国布赫前处理整套设备、意大利 GAI 自动灌装线,国内先进冷热自动控温不锈钢罐等设备,在地下恒温恒湿的窖窖中选用法国橡木桶陈酿 12-18 个月,秉承‘与大自然的和谐共存中酿造美酒’的品牌理念,确保葡萄酒品质上乘。”酒庄葡萄管理师王恢介绍。

先天优质的种植基因、如燕酒庄十余年的种植经验加上从国外引进的优质酿酒葡萄品种以及严格的管理等,让如燕酒庄产的如燕系列葡萄酒在中国国际精品葡萄酒挑战赛、亚洲葡萄酒质量大赛上屡屡得奖。

反哺社会勇担当

矗立在山峦之间的如燕酒庄远望似欧洲城堡建筑,近看似有战争时代之碉堡风貌。进入酒庄内部,迎面而来的是向社会免费开放的葡萄酒文化展厅,这是一个集视频展示、光影互动、雕塑造型和产品展示于一体的文化艺术展厅。通过引入先进的科技手段,展厅将葡萄酒制作过程生动呈现给游客,让其仿佛置身于葡萄园和酿酒工坊的场景中。

在这个富有教育性和趣味性的展厅里,游客不仅可以近距离地观察到葡萄的种植和采摘过程,还能通过光影互动亲自参与红酒的酿造过程。专业导游会向游客详细介绍

每个环节的重要性,并解答他们对葡萄酒知识的疑问。此外,展厅还设置了品酒区域,供游客品尝正宗的山西风土红酒,让他们能够亲身感受到葡萄酒的独特风味和品质。

如燕酒庄自开放以来,始终坚持合理规划、全力保护、注重文化传承、有序开发和形成特色的工作思路,酒文化展馆自建立以来,接待游客达 10 万人次。为了保护旅游资源,在园区内设立了各类专职人员队伍,实行全天候保洁、巡视,确保环境整洁。孙建伟说:“我们希望带给游客一次难忘的体验,让游客领略中国红酒文化的魅力,感受匠人精神的伟大,并品鉴到原汁原味的山西太谷红酒。无论是专业爱好者还是旅行者,这里都是一个不可错过的旅游胜地。”

目前,酒庄已与西北农林科技大学葡萄酒学院、山西农业大学食品学院、晋中学院、晋中职业技术学院等高校院所建立校企合作关系,建立山西农业大学食品科学与工程学院大学生创业实践和就业实习基地、山西农业大学食品科学与工程学院产教融合创新基地、西北农林科技大学葡萄酒学院理事会单位等,每年接受数百名大学生参与葡萄采收与酿酒榨季生产顶岗实习工作。

凭借着对土地的热爱、对家乡建设的热情,如燕人成为了凤翼山的“拓荒者”。用燕子筑巢殷殷而不舍的精神,让数千亩荒山荒坡变成了有机葡萄园,在打造百年酒庄的道路上不断前行。

创新杂谈

chuangxinzatan

4 月 24 日是中国航天日。今天上午,在酒泉卫星发射中心召开的神舟十八号载人飞行任务新闻发布会上, 中国载人航天工程新闻发言人、中国载人航天工程办公室相关负责同志表示,将抓紧研究推动国外航天员以及太空游客参与空间站飞行。

推动国外航天员参与空间站飞行,是我国载人航天事业发展的既定规划,也是太空领域科学研究国际合作的必然趋势,在此前的相关新闻发布会上也有提及。研究推动太空游客参与空间站飞行,于中国空间站而言则是“新鲜事”。太空游客,顾名思义就是让普通人经过必要的培训之后,搭乘航天器进入空间站。历史上,“和平号”空间站和如今唯一在轨运行的国际空间站,都曾多次接纳过太空游客的到访。我国提出推动太空游客参与空间站飞行,表明中国空间站全面建成进入常态化运营阶段,有能力接待普通人的到访,更体现出了我国在航天事业发展中的包容和开放。

人类对宇宙始终是向往的。在过去的数十年间,除去科学研究目的,世界多国都曾开展过亚轨道和地球轨道太空商业旅游项目,满足普通人进入太空的需求。包括我国民营商业航天公司中科宇航也曾表示,待技术成熟后将开始提供亚轨道载人旅游服务。人类当前对宇宙的认知是浅显的,对太空的探索是无止境的,这也意味着探索宇宙将会耗费大量的人力物力财力。国际空间站上到访的太空游客,都曾花费巨额资金,而这些资金也可以作为有益补充,充实太空研究和空间站运行维护。因此,循序渐进、逐步开展合理商业活动,如太空游客参与空间站飞行,从航天事业长远发展可持续发展角度看,是符合发展规律的举动。

当然,研究推动太空游客参与空间站飞行,能够预想最终落地成功探访中国空间站的普通人不会很多,但对全社会而言,这同样是个积极信号。普通人有机会进入太空、进入中国空间站,本就是一次播撒太空梦想的种子过程。普通人有希望作为游客参与空间站飞行,是一个可实现的目标。神舟十六号乘组成员桂海潮是一名教授、博导,并非职业航天员,他成功进入空间站就曾引发轰动,并产生了积极影响。当普通人有进入太空的机会,就会吸引大量对太空探索心怀梦想的人群为之奋斗,为科学事业发展注入新鲜血液,而这或是经济社会发展所需要的实干者。

从更长远的太空发展规划看,我国还将持续推进载人探月工程等多项太空探索项目,人类在宇宙中显然会留下更多身影。或许在不久的将来,人类就能够在地球之外尝试延续生命,而这显然需要更多普通人参与进来。

刘星星：“隔空诊脉”的“大机医生”

(上接 A1 版)作为劳模创新工作室负责人,刘星星积极发挥传、帮、带作用,带领身边职工进行故障溯源,把出故障的配件原理、结构、动作过程、养护维修方法编制成系统;编制“运、检、修”闭环流程和“电气组标准化一次作业流程”,满足 13 种车型调度的百余台大机年检调试的需求。团队累计完成职工创新成果 200 余项,其中 50 项成果获局级合理化建议改进奖,风动齿轮箱报警装置、集中润滑自动控制系统等 30 余项成果在全局范围推广应用,5 人获得局级技术能手。

“现在每年大修、维修线路总里程超一万公里。大机不断升级换代,我们的检修技术也要与时俱进。”为此,他们成立大机现场课堂,举办大机检修人员专项培训班,传授大机结构原理、故障维修、检修规程等理论知识,利用施工保养间隙,手把手指导故障排查、应急处置、精捣调试等多专业实操,为施工现场输送近 300 名检修人员,显著提升了全局大机检修技术水平。

近几年,大秦线日均运量 110 万吨左右。为确保施工、运输两不误,太原局集团公司把每一次集中修任务都精确计算到分钟。

2022 年国庆节期间,刘星星遇到了在大秦线上最棘手的问题——桥梁偏心和落道问题集中整治。大秦线采用的是重轨重枕,桥梁大部分在弯道处,拨偏心量超过以往单次最大拨道量,而且留给大机作业的时间仅有 1 个小时。

“大拨道量、多次作业,钢轨回弹、人员操作、石砟密度等影响因素非常多。”刘星星说,任何一个环节的微小变化都会带来严重后果。

为了攻克这一难题,太原局集团公司召开专题会议,太原工务机械段成立攻关课题组,刘星星作为中坚力量,调试好车辆的的同时,指导车组人员控制回弹量确保拨道精度,最终圆满完成集中整治任务。

“根据不同线路、不同操作手的情况,他按需调整车辆作业状态,基本做到一线一策、一车一策。”该段副段长郝燕青对刘星星精湛技术能力、保障作业安全的精神满是赞赏。

星光不问赶路人,时光不负有心人。刘星星记不清有多少个夜晚是在线路上度过的,但他知道只要不停下前进的脚步,就能战胜困难,实现自己的梦想。

“变”赢市场 “通”创未来

(上接 A1 版)在精密制造的世界里,正捷不断迎接挑战。据了解,正捷主要从事精密零件代加工业务。营业范围主要包括:机械零件、零部件加工;机械设备的研发;通用设备制造;光通信设备制造;互联网设备制造;通信设备制造;网络设备制造等。公司于 2020 年通过了 ISO9001 质量体系认证和太原不锈钢园区环境管理体系认证。

车间内,立式加工中心水花飞溅,工作人员拿着一块凹凸不平、有很多不规则小孔的铁板告诉记者:“正在生产的就是这款制作苹果手机的底板,订单有 2 万单,正在稳步生产陆续发货,现在我们看到的生产线上的这一批将在 3 天后统一运往富士康。”

正捷作为一个小微企业在 2023 年创下了 600 万元营业收入,在继续扩大生产能力,今年又购置了 12 台新设备,立志突破 1000 万元营收大关。“变则通,通则久。”李瑞祥说。

抓创新不问“出身”,这句话极大地鼓舞着正捷人。面对市场,正捷坚持做强主业、延伸多元化发展、建设最具竞争力的企业;面对未来,正捷将更加努力,加大创新,进一步为地方和国家作出更大贡献。

当普通人有希望到访中国空间站

■ 光明网评论员