

低成本、上线快、易调试,可在特定任务上反超通用大模型

轻量化小模型兴起,中小企业也能搭上AI“快车”

本报记者 陶稳

近两年,人工智能快速发展迭代,大语言模型如雨后春笋般涌现,文本生成、文生图、语音处理、代码处理、视频处理等生成式人工智能能走进人们的生活。同时,一些企业开始发力研发可在电脑、手机端训练的轻量化小模型。

“我们需要高铁、飞机、游轮等大型交通工具,也需要私家轿车、公交车,以及摩托车、自行车等小型交通工具。因为在不同场景下,不同人群有不同需求。”在青岛自然语义公司联合创始人、首席架构师孙燕群看来,满足特定市场的需求,是专而精的小模型兴起的重要原因及其存在的价值。

今年3月,自然语义研发的Euler模型通过中央网信办生成式人工智能服务备案。不同于大模型动辄千亿级的参数量,Euler的参数量只有2.5B(25亿),是典型的轻量化小模型。相较于大模型,小模型有何特点?应用前景如何?记者对此进行了采访。

低成本、易调试的端侧小模型兴起

关于小模型,目前并没有明确定义。孙燕群表示,在行业内,参数量低于100B的模型就算比较小的模型。在实际应用时,要想在笔记本电脑端实现微调,模型参数量一般在3S左右。

相较于大模型,小模型在算力消耗、使用成本方面更具优势。具体来说,一是训练和推理所需的硬件资源较少,使得成本较低;二是使用更便捷,可在手机、电脑、物联网设备等计算场景中实时运行;三是结构简单,开发者能快速定位问题,易于调试。

记者了解到,随着大模型的比拼日益激

阅读提示

随着人工智能快速发展迭代,一些企业开始押注小模型。相较大模型,低成本、上线快、易调试的小模型,以更高的性价比为中小企业和个人用户提供了打开人工智能大门的钥匙。

烈,一些企业开始押注小模型。2024年8月,微软和英伟达就相继发布过小型语言模型。国内不少企业也开始研发在“断网、弱网”环境下,让各种智能终端具备自主思考能力的端侧小模型。例如,在今年3月举办的中关村论坛年会期间,北京的面壁智能公司就发布了应用于汽车智能座舱的纯端侧超级智能助手。今年1月,广东佛山移动牵头联合40家单位成立佛山市AI小模型产业联合体,致力于通过人工智能小模型提供个性化服务,助力企业完成智能化升级。

“小模型让我们实现了与科技巨头们的错位竞争。”提到小模型的具体应用场景,孙燕群举例说,Euler通过备案后,已经开始面向中小企业和个人用户提供服务。如与山东某市级档案馆合作,上线了档案模型;与青岛当地一家仪器仪表设备公司合作,开发了内网技术相关的模型。

退而求其次后的“主动作为”

“客观来说,大模型的能力强于小模型,一个参数量2.5B的模型无论如何都不可能超过200B模型的算力。”孙燕群直言,这也是大公司都在做千亿级模型的原因。模型小,神经网络参数就少,容纳的知识量不够,“体现在文本生成上,就是容易出现上下文不连贯的情况”。

自然语义最初的目标并不是小模型。2019年底,该公司考虑要做大模型,但由于发展方向不明朗,同时在购买高算力GPU上面面临资金和货源压力,因此未能实施。

“没想到,仅过了两年时间,国外的大模型就发布了。”孙燕群回忆说,后来随着越来越多的企业涌人大模型赛道,作为基础硬件的高算力GPU价格水涨船高。“那段时间,听说哪里有GPU,我们就坐飞机去买,常常是提前交了定金也不一定能买到。”最终,公司用10多台设备组成一个算力集群,能支撑做出7B参数的模型。

为了在现有设备基础上尽可能提高参数,自然语义公司工程师经过头脑风暴,想出了新的分词方式,以及将向量计算从实数空间转换到复空间等各种办法。“经过模型训练,这些技术都达到了比较好的效果,能让一个3B参数的模型,达到了150B参数模型60%左右的能力。”孙燕群表示。

中国信息通信研究院人工智能研究所副总工程师王蕴韬告诉记者,小模型之所以能在“瘦身”后仍保持可观性能,得益于一系列成熟的模型压缩与高效架构技术,包括剪枝、量化、知识蒸馏、设计先天高效的网络架构等。

让“小身材”跑出“大能量”

“小模型发展大有可为。”在王蕴韬看来,

未来面向特定应用场景的小模型,潜力将会进一步释放。以AI终端为代表的应用形态及产品服务,将成为小模型释放大能力的主战场。

关于小模型的应用前景,王蕴韬进一步解释,一是为离线办公、文档摘要、私密对话等场景铺平道路;二是随着处理器架构和神经处理单元技术的应用,手机、车载和物联网设备将成为小模型的天然舞台;三是在垂直领域与“专精特新”行业,如金融、医疗、法律、教育等已出现6B及以下参数的定制模型,成本低、上线快,可在特定任务上反超通用大模型。

“市场关心的是能否解决实际问题,不关心背后模型细节,能够与场景深度结合,拥有行业知识,尤其是可信的小模型至关重要。”王蕴韬说。

从用户端来看,在实际应用中,小模型的性价比优势也十分明显。北京某互联网公司算法工程师张先生向记者表示,在现有技术条件下,想要在本地部署大模型存在一定难度。“专业GPU芯片价格太高,无法应用到低价格的终端上,如手机、机器人等配备的芯片就无法撑起大模型。另外,这些终端所配备的电池,往往也支撑不了高性能芯片的耗电等。”张先生表示,这些硬件性能有限的终端,更适合小模型施展。

王蕴韬分析认为,未来将是大小模型混合的系统范式。“云端大模型负责通用推理,端侧小模型承担即时响应与私域数据处理。”他还强调,小模型并非“大模型的低配版”,而是面向资源受限环境与专用任务的高性价比解法。通过配合端云混合部署和行业数据精调,企业完全可以让“小身材”跑出“大能量”,在AI商业落地的下一程中获得确定性收益。

6月CPI同比上涨0.1%



图为7月9日,消费者在山东省枣庄市一家超市选购调味品。
新华社发(孙中喆 摄)

本报北京7月9日电(记者时斓娜)国家统计局今天发布的居民消费价格指数(CPI)数据显示,6月,扩内需、促消费政策持续显效,CPI同比上涨0.1%,环比下降0.1%;扣除食品和能源价格的核心CPI同比继续回升,上涨0.7%,创近14个月以来新高。

从同比看,6月,CPI连续下降4个月后转涨。CPI由降转涨主要受工业消费品价格有所回升影响,工业消费品价格同比涨幅由上月的1%收窄至0.5%。其中,油价变动影响能源价格降幅比上月收窄1个百分点;受国际大宗商品价格变动影响,金饰品和铂金饰品价格同比分别上涨39.2%和15.9%;促消费政策效应持续显现,文娱耐用消费品、家用纺织品和家用器具价格同比分别上涨2%、2%和1%;汽车价格降幅逐步收窄,燃油小汽车和新能源小汽车价格本月同比分别下降3.4%和2.5%,降幅分别为近28个月和26个月最小。

从环比看,6月,CPI环比下降0.1%,降幅比上月收窄0.1个百分点。受高温、降雨较常年同期偏多影响,食品价格环比下降0.4%,降幅小于季节性水平0.5个百分点。受国际油价变动影响,汽油价格环比由上月下降3.8%转为上涨0.4%,带动能源价格由上月下降1.7%转为上涨0.1%。毕业季房屋租赁需求有所增加,房租价格上涨0.1%。

重庆上半年出入境外国旅客创新高

7月8日,在重庆江北国际机场,重庆出入境边防检查总站民警解答入境旅客疑问。

随着免签政策不断优化、“ChinaTravel”持续火热,重庆江北机场口岸出入境人员数量呈持续增长态势。2025年上半年,重庆出入境边防检查总站累计查验出入境人员114万余人次,与2024年同期相比增加35%,其中上半年出入境外国旅客33万余人次,创历史新高。

新华社记者 黄伟 摄

前5个月我国共查处违法广告案件14315件

本报北京7月9日电(记者杨召奎)记者从市场监管总局获悉,今年以来,全国市场监管部门持续加大对违法广告的打击力度,防范经营主体借违法广告加剧“内卷式”竞争,切实维护公平有序的广告市场秩序。截至今年5月底,全国市场监管部门共查处违法广告案件14315件,罚没款1.01亿元。

今天,市场监管总局选取一批已办结的违法广告典型案例予以公布。其中包括:四川省成都市成华区市场监管局查处成都成华韩后医疗美容医院有限公司违法广告案。经查,成都成华韩后医疗美容医院有限公司在以直播电商形式推销“半岛超声炮”“欧洲之星面部全模式”“星耀超光子”等医疗美容项目过程中,宣称的治疗功效与实际情况不符,并使用“单次我们的续航能力是9-12个月”“能够解决肌肤20多种问题”等宣传用语为功效作保证。今年1月,成都市成华区市场监管局依据《中华人民共和国广告法》有关规定,对当事人作出罚款40.98万元的行政处罚。

其他案例包括:河南省郑州市市场监管局查处市野(郑州)网络科技有限公司违法广告案;安徽省阜阳市颍东区市场监管局查处安徽全康药业有限公司违法广告案;浙江省杭州市拱墅区市场监管局查处杭州迅优贸易有限公司违法广告案等。



我国锂矿找矿取得又一重大突破

湖南探获4.9亿吨超大型锂矿

本报讯(记者王鑫 方大丰)我国锂矿找矿取得又一重大突破。记者7月8日从湖南省自然资源厅获悉,湖南省地质院矿调所在郴州市临武县鸡脚山矿区通天庙矿段探获超大型蚀变花岗岩型锂矿床,提交锂矿石量4.9亿吨,其中的氧化锂资源量达131万吨。同时伴生铷、钨、锡、铌、钽等多种战略矿产,均达中型以上规模。

据勘探项目负责人介绍,此次探获的锂矿素有“白色石油”之称,锂是动力电

池、储能电池等产品的核心材料。随着全球新能源产业的发展,锂资源需求量急剧增加。我国锂资源一度面临对外依存度较高、进口量占比大的局面。今年初,自然资源部中国地质调查局宣布,我国锂矿储量已从全球占比6%升至16.5%,从世界第六跃至第二,重塑了全球锂资源格局。

据勘探项目负责人介绍,此次探获的锂矿,矿体规模大、延伸稳定,矿石属较易选类

型。探获的氧化锂资源量相当于13个大型锂矿。项目克服矿段地形复杂、钻探施工条件等难题,通过探索建立花岗岩型锂多金属矿找矿模型,采用索道和大型无人机吊装钻探设备技术,采用先进轻便型液压钻探工艺和设备,在短短一年内完成钻孔227孔,钻进尺9.2万米,极大提高了施工效率,实现了绿色勘查。“这样的勘查速度在以前是不可想象的。”该项目负责人说。

推动车企践行60天账期承诺 这个窗口为中小企业护航

新华社记者 唐诗凝 周圆

工业和信息化部9日在“全国违约拖欠中小企业款项投诉平台”开通“重点车企践行账期承诺线上问题(建议)反映窗口”,受理中小企业有关问题建议。业内人士认为,这一举措为中小企业提供了便捷的投诉与建议渠道,有利于推动汽车企业切实履行账期承诺,保障中小企业合法权益,促进汽车产业可持续发展。

产业链供应链是汽车产业的“筋骨血脉”,是产业转型升级的关键一环。近期,一汽、东风、广汽、赛力斯等17家重点车企作出“供应商支付账期不超过60天”的公开承诺,如何进一步落实兑现备受社会关注。

记者了解到,此次开通的线上反映窗口主要受理4类问题建议,包括重点车企未践行60天支付期限承诺,在采购合同中约定的付款期限超过60天;重点车企设定不合理的支付期限起算时间、无正当理由拖延出具检验或验收合格证明等方式变相延长支付期限,以及以收到第三方货款作为支付条件或按照第三方支付比例支付款项;重点车企强制或变相强制中小企业接受商业汇票、应收账款电子凭证等非现金支付方式;重点车企落实《保障中小企业款项支付条例》不力的其他问题。

中汽中心中国汽车战略与政策研究中心主任王铁表示,这4类问题涵盖了从合同条款设计、付款流程漏洞、验收规则模糊、支付工

具滥用等各个环节,有利于精准识别和解决中小企业面临的实际问题,确保账期承诺落到实处。

“60天账期承诺和开通问题反映窗口,是针对供应链资金承压问题采用‘点穴’式手段对症下药,是推动汽车行业践行社会责任、改善产业生态的重要一环。”王铁说。

中国汽车工业协会常务副会长兼秘书长付炳锋告诉记者,现金流如同企业的“血液”,按期回款尤为重要,影响到企业日常经营、项目投资和技术创新等方方面面,现金流畅通了,企业发展才能有活力。

“我们了解到,车企作出‘供应商支付账期不超过60天’的公开承诺以来,已在积极推

进管理流程改善、支付方式优化等,加快相关

问题的解决。但规范供应商支付账期是一项系统工程,仍需各方协同努力。”付炳锋说。

王铁说,为实现汽车行业健康可持续发展,要引导行业建立长期互信的整零关系,推动产业从“零和博弈”转向“共生共赢”。“建议建立动态监测长效机制,针对可能存在的市场增长乏力、盈利能力下降、竞争恶化等风险及时识别研判,并对症施策。”

工业和信息化部有关负责人表示,下一步,还将指导行业机构研究制定汽车行业结算支付规范,推行合同范本,进一步规范汽车企业供应商货款支付流程,推动构建“整车一零部件”协作共赢发展生态,促进汽车产业健康可持续发展。

(新华社北京7月9日电)

宁夏10部门合力推进美丽城市建设

本报讯(记者马学礼 李静楠)近日,宁夏回族自治区生态环境厅、住建厅、工信厅、财政厅等10部门联合印发《宁夏回族自治区推进美丽城市建设的工作方案》(以下简称《方案》),明确了提升城市绿色低碳发展水平、城市生态环境质量、城市生态宜居品质、城市环境健康安全保障能力和城市数智治理效能等5方面19项任务,并根据城市特点和工作基础,选定银川市、固原市开展美丽城市先行区建设试点。

根据《方案》,宁夏将采取“一城一策”差异化建设方式,打造绿色低碳、环境优美、生态宜居、安全健康、智慧高效的美丽城市。为提升城市绿色低碳发展水平,宁夏将加快绿色制造体系建设,推动建设绿色工厂、绿色园区,力争到2027年建成绿色工厂180家,绿色园区创建率达到50%以上。同时,严格执行建筑节能降碳强制性标准,加快既有建筑节能改造,推动新建建筑中绿色建筑比例提高到100%。

《方案》要求,推动城市移动源结构优化和高效监管,深入实施清洁柴油车(机)行动,推进火电、钢铁、煤炭、焦化、铝工业等行业大宗货物清洁运输,推动淘汰国三及以下排放标准柴油货车和国一及以下非道路移动机械。同时,统筹推进固体废物综合治理。严控危险废物填埋处置,加强危险废物分级分类精细化管理和全过程信息化监管。