

国产AI大模型“百舸争流” 夯实算力底座正在进行时

■本报记者 许林艳

“国产AI大模型”成为近日的热门词汇。

随着多家上市公司官宣进入AI大模型领域,国产AI大模型的关注度不断提升。投资者互动平台上也“热闹非凡”,多家上市公司被询问是否会推出大模型产品、有没有为大模型产品提供技术支持等。大模型的热度由此可见一斑。

浙商证券表示,在ChatGPT背后,不断迭代的GPT系列使得大模型成为当下科技企业核心竞争力的重要体现,未来,大模型将成为AIGC时代的核心支撑。

国产AI大模型竞争开启

仅4月10日,就有多家上市公司发布在大模型领域的最新进展。昆仑万维宣布,公司与奇点智源合作自研、能够真正实现智能涌现的国产大语言模型——“天工”3.5发布在即,并将于4月17日启动邀请测试。据昆仑万维介绍,“天工”大模型已经非常接近ChatGPT的智能水平。消息发布后,4月11日,昆仑万维收涨9.5%,报45.76元/股。

同日,商汤科技对外发布了“日日新SenseNova”大模型体系,推出自然语言处理、内容生成、自动化数据标注、自定义模型训练等多种大模型及能力。软通动力也在互动平台上表示,在数据业务方面,公司正在开展多模态大模型的研究,并已将其应用于例如风险识别等以图、文、视为一体化的数据源的综合管理业务领域。

4月11日,阿里云正式发布自研大模型“通义千问”。此前,华为云人工智能领域首席科学家田奇也介绍了盘古大模型的进展及

应用,着重介绍了CV大模型和科学计算大模型的应用范例。还有多家公司透露了发布大模型产品的计划。国产AI大模型已呈“百舸争流”之势。

中信建投证券研究所所长武超表示,大模型在整个生态中的定位,类似于智能手机时代的操作系统或应用商店,这一领域最终能领跑的龙头公司有望达到万亿元级市值,但竞争也会非常激烈。

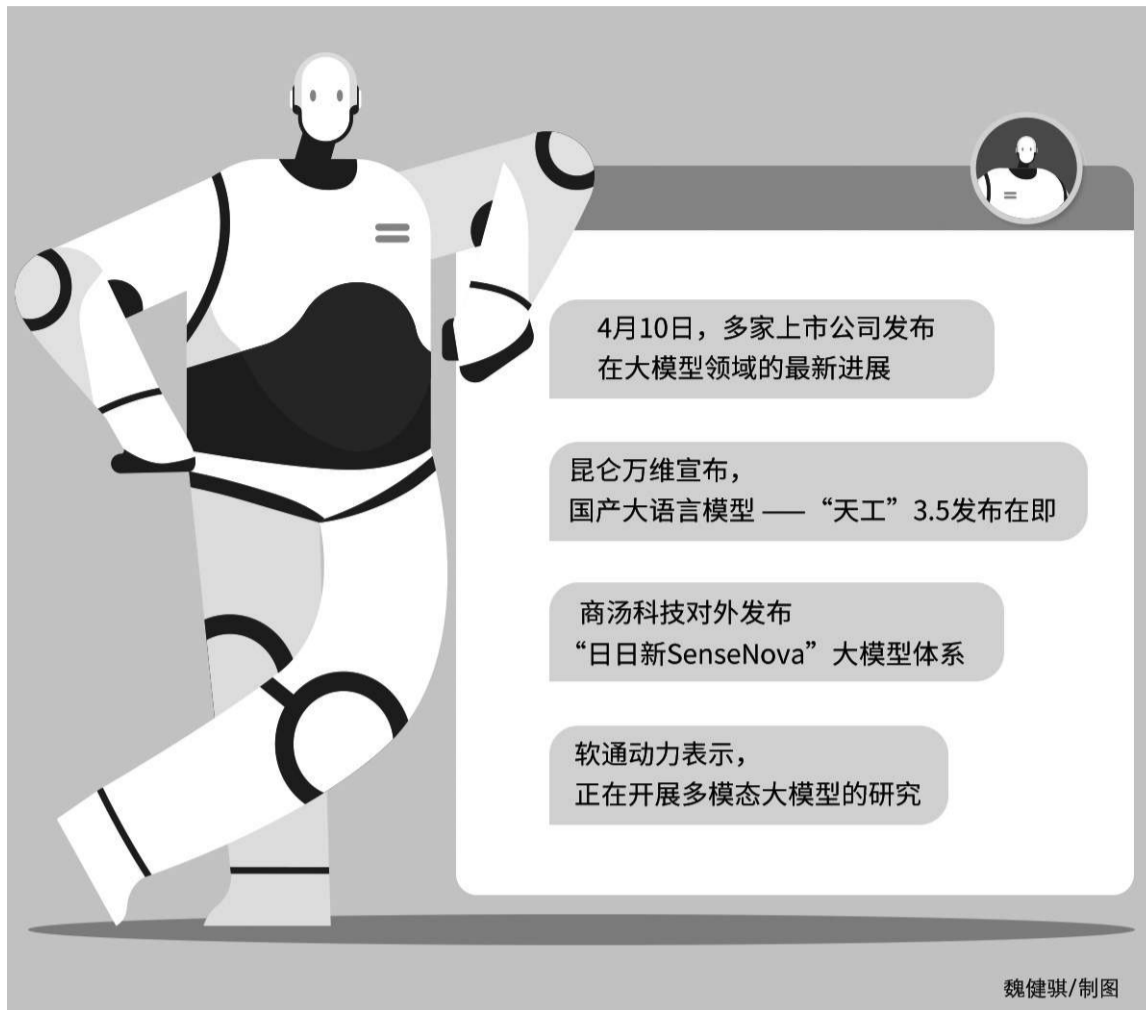
“我国AI产业偏重于视觉识别技术的研发和应用,近年来,视觉识别已经到了一个瓶颈期。近期,以自然语言处理为核心的GPT大模型崛起,给AI企业带来了新的发展思路。因此,许多公司开始向大模型方向布局。”深度科技研究院院长张孝荣在接受《证券日报》记者采访时表示。

“大模型与传统人工智能相比,能够更有力地促进传统行业的智能转型。国内大型互联网企业也基本认为大模型的开发是面向未来的一个重要机遇。”北京社科院研究员王鹏对《证券日报》记者表示,此外,以ChatGPT为代表的大模型推出以后,市场的关注度非常高,上市公司在这一领域布局新的业务条线,能够有效获取市场关注,有利于企业的未来发展。

算力芯片领域受关注

大模型发展的背后除了海量数据之外,还需要庞大的智能算力支撑。

国际数据公司(IDC)等发布的数据显示,2022年我国智能算力规模达到268亿亿次/秒,超过通用算力的规模;预计未来5年,我国智能算力规模的年复合增长率将达52.3%。



魏健骐/制图

智能算力的发展与算力芯片息息相关。“GPU芯片是加大AI算力的核心。目前适合构建大模型的算力芯片就是GPU芯片。”张孝荣说。

王鹏也认为,“大模型对算力的要求很高,传统芯片相对低端化,算力有限。高端化的芯片比如GPU,其算力速度相对更快,有利于大模型的训练。”

不过目前,我国GPU芯片的发展还面临着一些阻碍。据张孝荣

介绍,一是限于美国政府禁令,我国企业当前无法获得最先进的英伟达GPU;二是英伟达GPU价格昂贵,供应中国的单块芯片价格在1万美元至2万美元之间,而一个大模型需要数万块GPU,以此推算,算力投入至少需要几十亿元人民币,成本巨大。

“国家有关部门需加大政策扶持,进一步推动智能计算中心的建设运营,推进智能算力相关领域的发展,加强顶层设计和统

筹规划,尽快解决算力芯片领域的瓶颈问题。”中国电子商务专家服务中心副主任、知名互联网专家郭涛对《证券日报》记者表示。

上市公司正在积极推进算力芯片领域建设。4月11日,甬矽电子在互动平台表示,公司正在积极研发与算力芯片相关的封装技术及产业应用。此外,中兴通讯总裁徐子阳曾在业绩说明会上透露,公司将于今年年底推出支持ChatGPT的GPU服务器。

阿里大语言模型“通义千问”亮相 聚焦企业级市场提供普惠AI基础设施

■本报记者 许洁 见习记者 刘霄羽

4月11日,阿里巴巴集团董事会主席兼首席执行官、阿里云智能集团首席执行官张勇对外表示,让计算成为公共服务是阿里云始终不变的追求。

此前的4月7日,阿里云自研大模型“通义千问”开始邀请用户测试,引发市场关注。4月11日,阿里云智能首席技术官周靖人正式宣布推出大语言模型“通义千问”。

天使投资人、知名互联网专家郭涛在接受《证券日报》记者采访时表示,“通义千问”能够持续提高阿里生态圈的竞争力,在市场营销、客户服务、运营管理等领域不断为企业赋能,进一步提高效率、降低成本,助力企业高速发展。

赋能企业降本增效

张勇表示,阿里巴巴所有产

品未来将接入大模型,全面升级。“要建设全新的AI开放生态,需要从阿里内部开始,目前,钉钉、天猫精灵已经在测试接入大模型,钉钉能够实现自动写工作方案、生成应用等功能,天猫精灵接入‘通义千问’后显著提升了智能化水平。”

阿里云也将面向企业提供更普惠的AI基础设施和大模型能力。

据周靖人介绍,未来每一家企业在阿里云上既可以调用“通义千问”的全部能力,也可以结合企业自己的行业知识和应用场景,训练自己的企业大模型。比如,每家企业都可以有自己的智能客服、智能导购、智能语音助手、文案助手、AI设计师、自动驾驶模型等。

全球并购公会信用管理委员会专家安光勇对《证券日报》记者表示,“企业可以基于‘通义千问’开发智能对话系统,提供

更便捷、智能的客户服务、产品推荐、用户互动等应用,提升用户体验和黏性,拓展业务边界。‘通义千问’还可以应用于市场调研、舆情分析、用户情感分析等领域,为企业决策提供数据支持和智能分析。”

“‘通义千问’部分能力能够达到GPT3.5的水平,尤其是其代码生成能力,虽然还有不足,但至少能解决一些问题。”浙江大学国际联合商学院数字经济与金融创新研究中心联席主任、研究员盘和林对《证券日报》记者表示。

相比于ChatGPT,“通义千问”在面向企业客户时更有作为。郭涛认为,ChatGPT是通用型AI大模型,在企业级与消费市场均拥有广泛的应用场景,而“通义千问”是阿里巴巴整合商业资源和技术积累推出的产业版AI大模型,为企业自身生态体系或垂直行业赋能,更多聚焦企业级

市场。

让算力更普惠

事实上,阿里在云计算方面已进军To B领域。据了解,OPPO、复旦大学、南方电网等企业机构已经开始使用阿里云的算力和AI能力进行全新的AI探索。在阿里云的大模型魔搭社区上,超过100万开发者累计下载了超过1600万次各类模型,并直接调用阿里云的算力和一站式的AI大模型训练及推理平台。

在大模型技术层面,阿里同样具备深厚储备。国盛证券研报认为,阿里达摩院深耕NLP领域,在大模型技术路径上具备多年前技术积累。

早在2019年,阿里达摩院就启动了大模型研发,在超大模型、语言及多模态能力、低模训练、平台化服务、落地应用等多个方面,为中文大模型的发展做出一系列

探索工作。

AI大模型加速了智能时代的到来,张勇表示,所有企业在智能化时代都将站在同一条起跑线上,互联网公司和非互联网公司不存在先进或者落后的区别,传统企业有弯道超车的机会。“所有产品都值得用大模型重做一遍,未来将看到应用爆发、效率革命,AI将成为各行各业新型生产力。”

据介绍,未来十年,阿里云将全面拥抱智能化时代。张勇表示,智能化时代带来算力的爆发,同时也提出了更高的要求,为此,阿里云希望做好核心两件事:让算力更普惠,让AI更普及。

在张勇看来,云计算是AI大规模普及的最佳方式。“未来,我们希望在阿里云上训练一个模型的成本,能够降低到现在的十分之一,甚至是百分之一。即使是中小企业,也能通过云平台获得AI大模型的能力和服务。”

国产AI大模型竞相绽放 未来发展任重道远

■许洁

ChatGPT引发的大模型热潮依然汹涌。从上周末开始,各大企业AI大模型落地的消息接踵而至。

4月7日,阿里云自研大模型“通义千问”开始邀请用户测试。4月8日,京东集团副总裁戴晓晓称将在今年发布新一代大模型“ChaiJD”,同日,华为云介绍了“盘古”大模型的进展和应用。4月9日,基于360GPT大模型开发的人工智能产品矩阵“360智脑”率先落地搜索场景。更多的AI大模型正在飞奔而来的路上。腾讯、字节跳动、同花顺等企业的AI大模型产品近日将会陆续亮相。

还有越来越多的上市公司披露了大模型领域的最新进展。4月

10日,昆仑万维、商汤科技均公布了自家的大模型产品。

与OpenAI为ChatGPT引入插件,创造AI逻辑下的生态系统相似,4月11日,阿里巴巴集团董事会主席兼CEO、阿里云智能集团CEO张勇表示,阿里巴巴所有产品未来将接入“通义千问”大模型,并进行全面改造。张勇认为,面向AI时代,所有产品都值得用大模型重新升级。

可以预计,在中国规模庞大的互联网用户的支持下,国内企业在算法和算力上有望不断实现突破并取得领先。从算力到应用,已经接下了AI大模型发展的加速键。

但也要清醒地认识到,中国企业在AI大模型这条“直道冲锋”的赛道上,仍面临不少挑战。

首先,人才储备不足。AI大模型建设需要大量的数学、计算机科学和工程方面的人才。有业内人士称,全国做大模型的人才规模仅100人左右,在市场上属于“一将难求”的状态。虽然中国企业正在通过与国内外高校合作或自主培养等方式来解决此问题,但这需要一定的时间。

其次,算法和模型质量不高。在AI大模型领域,相关产品的算法、模型的质量和精度与国外相比还有差距。比如,笔者测试“通义千问”后发现,其已经具备一定基础知识与初步的逻辑思考能力,但在复杂理科计算方面仍有提升空间。

再次,商业模式有待探索。虽然AI大模型是技术的革新,可以为

企业带来很大的商业价值,但商业模式的探索和实践需要时间和大量的资源投入。中国企业也需要探索出适合自己的商业模式。

最后,安全问题亟待解决。3月29日,马斯克与上千名人工智能专家及AI行业高管签署联名信,呼吁暂停训练比OpenAI推出的GPT-4更强大的人工智能系统至少6个月,理由是这种系统对社会和人类构成潜在风险。事实上,保障AI大模型的安全性需要综合考虑多种因素,包括合理的数据管理、安全的算法和模型结构、可靠的系统、以及安全防护措施等,这些都离不开企业自律和政策监管。据悉,本周三,OpenAI、微软、谷歌将齐聚,讨论制定关于如何使用人工智能技术的标准。

国内政策也正在完善,4月11日,国家互联网信息办公室发布关于《生成式人工智能服务管理办法(征求意见稿)》公开征求意见的通知,违反办法规定的,由网信部门和有关主管部门按照《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律、行政法规的规定予以处罚。

诚如阿里云智能CTO周靖人所言,“通义千问”不是起点也不是终点。跟所有新科技一样,AI大模型未来一定会遇到波折和低谷。对于中国企业而言,不应追求速胜,而是应该坚持打“持久战”,花时间去上述问题一一填平,筑牢发展的根基,从而在这场不可缺席的较量中保持底气 and 信心。

生成式人工智能服务管理办法出台在即 有望推动AIGC产业规范发展

■本报记者 邢萌 田鹏

4月11日,国家互联网信息办公室就《生成式人工智能服务管理办法(征求意见稿)》(以下简称《征求意见稿》)公开征求意见。

《征求意见稿》重点强调了生成式人工智能产品训练数据及生成内容的真实性。其中,就提供生成式人工智能产品或服务作出五大要求,提出“利用生成式人工智能生成的内容应当真实准确,采取措施防止生成虚假信息”。同时,也对产品的训练数据作出五项要求,提出“能够保证数据的真实性、准确性、客观性、多样性”。

“《征求意见稿》的出台,意味着AI行业的标准化步伐正加快,AIGC产业可持续发展的政策红利期即将到来。”中关村智能科技发展促进会副会长兼首席数字经济学家陈晓华在接受《证券日报》记者采访时表示,从相关条例可以看出,我国极为重视本轮AI所引领的科技变革,有关部门在鼓励自主创新、推广应用、国际合作的同时,也致力于避免“先污染后治理”的困境。

强调生成内容真实性

随着ChatGPT火爆市场,大众对生成式人工智能技术的认知程度不断提升,激发市场需求的同时,也加速了相关企业布局。

不过,AIGC行业发展也面临着虚假信息的困扰。中国信息通信研究院和京东探索研究院发布的《人工智能生成内容(AIGC)白皮书(2022年)》显示,近年来,随着人工智能技术不断成熟,机器深度学习后生成的内容愈发逼真,能够达到“以假乱真”的效果。

在此背景下,《征求意见稿》着重强调了生成式人工智能产品训练数据及生成内容的真实性。

“这将引领‘可信’生成式人工智能服务的发展。”大成律师事务所高级合伙人肖飒对《证券日报》记者表示,对于生成式人工智能服务来说,“可信”是最难以达成的。核心问题指向服务生成内容可信程度的波动化,一时超常的准确表现,一时难以分辨的潜藏错误,是阻碍对该类服务赋予更多信任的主要原因。

此外,《征求意见稿》对于生成式人工智能产品的预训练、优化训练数据也进行了明确要求:一是要符合《中华人民共和国网络安全法》等法律法规的要求;二是不含有侵犯知识产权的内容;三是数据包含个人信息信息的,应当征得个人信息主体同意或者符合法律、行政法规规定的其他情形;四是能够保证数据的真实性、准确性、客观性、多样性;五是网信部门关于生成式人工智能服务的其他监管要求。

对此,肖飒认为,对服务背后的数据进行统一要求,最直接的意义便是对模型输入端进行优化,从而尽可能地保证输出端内容的纯净性、准确性。从长远来看,这是对数据相关产业链提出的新要求,是对整个行业数据收集、传输、交易、检验、维护与保护的全面规制,也符合现行立法对数据处理的各方面要求,能够最大程度地防止后续违法行为的发生。

《比较》杂志研究部主管陈伟对《证券日报》记者表示,总体来看,《征求意见稿》相关要求给生成式AI发展划定界限的同时,也为其留下了充分的发展空间,从而扬长避短,将技术发展的正向作用发挥到最大。

服务前应申报安全评估

在国内AIGC发展早期做好合规工作,可为行业后续稳健发展保驾护航。据《证券日报》记者不完全统计,截至目前,已有百度、华为、腾讯、阿里、京东、360等互联网企业布局生成式人工智能业务,并且部分企业已经推出相应产品。

“生成式人工智能快速发展,带来了诸多信息安全方面的风险。”平安证券研报表示,一方面,信息泄露风险,大量的敏感数据和AI交互之后,可能会被应用于迭代训练,造成敏感信息的泄露,尤其是国家、机构和个人信息,都面临着潜在风险;另一方面,AIGC被恶意应用的风险也在上升,可能被应用于恶意代码生成,发展智能化网络攻击武器,加大了网络安全防护难度。

深圳市信息服务业区块链协会会长郑定向对《证券日报》记者表示,生成式人工智能产品目前仅是一个机器程序,没有法律主体身份和行为能力,但如果被用来执行非法行为,例如攻击计算机系统、窃取数据或参与欺诈等,就会成为犯罪行为的工具或手段。基于此,为了确保人工智能的合法性和公正性,需要对提供生成式人工智能产品或服务进行必要的法律约束。

为此,《征求意见稿》强调生成式人工智能产品的安全性。《征求意见稿》提出,国家支持人工智能算法、框架等基础技术的自主创新、推广应用、国际合作,鼓励优先采用安全可信的软件、工具、计算和数据资源。但利用生成式人工智能产品向公众提供服务前,应申报安全评估。

郑定向表示,生成式人工智能技术和产品的发展归根结底是为了服务于人,科学、严密的法律规范是发展过程中必不可少的重要环节。

平安证券表示,新技术发展和监管合规是相辅相成的关系,离不开监管在机制体制建设上的完善,以及产业政策落地,一些潜在的风险将更多被认识和暴露,监管侧的监管措施也会更“对症”,《征求意见稿》只是第一步。