

国产大模型掀起长文本竞赛



青岛财经日报/首页新闻讯

近日,国内多家企业相继宣布开放其AI大模型的长文本处理功能,开启国内技术竞赛新篇章。

AI大模型创业公司北京月之暗面科技有限公司(以下简称“月之暗面”)旗下Kimi产品200万字长文本处理功能的推出引起了广泛关注,阿里巴巴通义千问、百度文心一言以及360智脑随后跟进,展示了中国AI技术的迅速进步和巨大潜力。

Kimi支持200万字长文输入

Sora风头未过,Kimi又登场。

3月18日,月之暗面宣布其对话式AI助手产品Kimi智能助手已支持200万字的无损上下文输入。在去年10月发布时,Kimi可支持的无损上下文输入长度为20万字。

与此同时,月之暗面还为Kimi智能助手提供了更多数据源。月之暗面工程副总裁许欣然表示,当面对一个问题时,Kimi智能助手会尝试从不同的方向搜索并据此做出回答。在回答速度上也有提升。基于Infra层的优化,Kimi智能助手生成速度较去年10月份提升了三倍。

目前,Kimi智能助手已上线了网页版、安卓、iOS、小程序应用。据SimilarWeb数据显示,Kimi智能助手网页版于今年2月的访问量为291.9万人次,比上月增长104.99%。

提升上下文长度也存在一些技术路线,如基于检索增强的RAG(检索增强生成)、滑动窗口,但是这些方法在提升上下文长度的同时往往也会带来诸如智能水平下降、成本提升等负面效果。

至于月之暗面的解决办法,月之暗面创始人杨植麟此前曾表示主要在两方面,一是创新网络结构,二是进行工程优化。许欣然进一步表示,上下文长度的无损提升需要在数据、Infra、模型

训练、产品等层面协同合作才能实现,此次团队从模型预训练到对齐、推理环节均进行了原生的重新设计和开发。

超长文本输入能做什么

支持200万字超长无损输入,这意味着大模型能做什么?月之暗面表示,过去要10000小时才能成为专家的领域,现在只需要10分钟,Kimi就能接近任何一个新领域的初级专家水平。用户可以跟Kimi探讨专业领域的问题,让Kimi帮助自己练习专业技能,或者启发新的想法。

比如,用户可上传一份近百万字中医诊疗手册,Kimi可以针对用户问题,给出诊疗建议。

杨植麟表示,通往通用人工智能,无损的上下文将会是一个很关键的基础技术。历史上所有的模型架构演进,本质上都是在提升有效的、无损的上下文长度。上下文长度可能存在摩尔定律,但需要同时优化长度和无损压缩水平两个指标,才是有意义的规模化。

互联网大厂布局长文本领域

Kimi爆火后,国内互联网大厂也在长文本领域快速布局。

3月22日,阿里的通义千问官宣升级,向

所有人免费开放1000万字的长文档处理功能。用户可通过通义千问网站和APP快速读研报、分析财报、读科研论文、研判案情、读医疗报告、解读法律条文、分析考试成绩、总结深度文章。

360紧随其后,其宣布360智脑正式内测500万字长文本处理功能,即将入驻360AI浏览器。

另据报道,百度在下个月将对文心一言进行升级,届时也将开放长文本能力,文本范围在200万至500万字。目前,文心一言的文本最高上限在3.2万字。

业内人士表示,长文本能力目前在训练和效果上仍存在一些难点,比如如何避免遗漏细节。事实上,文本长度也可以通过RoPE(旋转式位置编码)等算法技术去拓展,但是此时会严重丢失细节,这就导致输出效果不准确。因此,核心标准是看输出结果的精准性和连贯性,并不是文本长度越长越好。

此外,算力挑战是长文本需面对的首要问题。华西证券认为,国内长文本带动大厂长文本能力迅速迭代,一方面将会缩小国内外大语言模型的代差,另一方面将会进一步加大对底层算力的需求,算力需求爆发引发国内异构算力的构建,会带来AI行业的又一次变革,AI模型、语料库、算力租赁、算力运维、国产算力等将会继续受益。

业界简报

国内首个智能机器人中试验证平台落户上海

青岛财经日报/首页新闻讯 为强化智能机器人应用赋能优势,加快推进新质生产力构建,近日,上海人工智能研究院与中国电子信息产业发展研究院(赛迪)完成签约,共同在上海建设国内首个智能机器人中试验证平台。

该智能机器人中试验证平台的建设,将推动我国实现三个“一流”:一流的垂类大模型测试床,解决垂类模型安全能力的评估难题,加速大模型落地进程;一流的人工智能大模型工业软件的自主可控测试平台,建立工业产品准入门槛;一流的基于区块链安全测试与性能测试平台,为数据强隐私要求场景下的数据安全提供“可见可控”的解决方案。

当前,人工智能大模型、人形机器人作为具有颠覆性和前沿性的产业,被视为新质生产力发展的代表,未来将展现出巨大的潜力和吸引力。智能机器人中试验证平台成立后,将成为推动技术成果从“实验室”迈向“应用场”的关键一环。同时,中试平台通过技术创新具备测评、调优等功能,为通用大模型、智能机器人排查风险点,为加速发展中的新兴产业系上“安全带”。

国内首个单池万卡液冷算力集群投运

青岛财经日报/首页新闻讯 中国电信近日宣布,天翼云上海临港国产万卡算力池正式启用,同时入驻首批用户。这是国内首个投入正式运营的国产单池万卡液冷算力集群,也是业内领先的全国产化云智一体公共智算中心。

此次启用的天翼云上海临港国产单池万卡液冷算力集群创新性地采用网络中置、算力分层的“魔方”型,实现了单一集群内万卡高速互联,满足万亿级参数大模型训练所需的多机多卡并行、高吞吐无损通信等需求。同时,为了实现绿色低碳的目标,全面采用融合液冷服务和IDC基础设施的新一代智算液冷DC舱,实现了数据中心的能效和智算集群的算效双提升,为“人工智能+”提供智能、弹性的绿色算力。

AMOLED显示驱动芯片今年需求将同比增19%

青岛财经日报/首页新闻讯 市场调研机构Omdia最新报告指出,在AMOLED显示驱动芯片市场,智能手机出货量仍占最大份额,2023年占据67%,2024年将占62%。得益于iPhone和华为手机的需求以及中国智能手机OEM厂商产品阵容的扩大,2023年下半年AMOLED智能手机显示驱动芯片的需求好于预期,推动2023年AMOLED智能手机驱动芯片的总需求同比增长8%。但由于电视、笔记本电脑和其他应用的需求下降,2023年AMOLED显示驱动芯片的总需求仅同比增长1%。

2024年,随着越来越多的LCD智能手机升级为AMOLED,AMOLED智能手机驱动芯片的出货量预计将同比增长11%。此外,由于iPad Pro采用OLED面板,预计2024年AMOLED平板电脑驱动芯片的出货量将达到9200万颗,同比增长224%。另外,由于智能手表、笔记本电脑和OLED电视的需求不断增长,Omdia预计2024年AMOLED显示驱动芯片的需求量将同比增长19%。

穿上即可行走的通用外骨骼面世

美国佐治亚理工学院机械工程师开发了一种控制机器人外骨骼的通用方法。无需专门训练,特别校准,对复杂算法进行调整后,用户穿上外骨骼就可以直接行走。新系统使用深度学习自动调整外骨骼为人类提供帮助的方式,现已证明这种外骨骼可顺滑地支持行走、站立以及爬楼梯或坡道等动作。

该领域之前的大多数研究都集中在单项活动上,例如在平地行走或爬楼梯,所涉及算法通常会尝试对环境进行分类。佐治亚理工学院团队将关注环境转移到关注人类,也就是关注肌肉和关节的活动。

与不佩戴设备的人相比,用户穿上新开发的髌部外骨骼能量消耗更少,他们自己的关节也不再那么“辛苦”。研究团队使用现有算法,并根据实验室收集的力量和动作的捕捉数据对其进行通用训练。不同性别和体型的受试者佩戴动力髌部外骨骼,在测力板上以不同的速度行走,爬上可调节高度的楼梯、上下坡道以及在这些动作之间进行转换。就像用于制作电影的动作捕捉工作室一样,每个动作都被记录并编目,以了解关节在各种活动中的表现和作用。

研究人员称,未来,机器人外骨骼将使士兵、航空行李搬运工或任何从事高强度体力工作的人群受益。 据《科技日报》张梦然/文

微软入局 AI个人电脑竞速升级

2024年被业界视为AI PC(AI个人电脑)元年,相关企业闻声而动。如今,微软也携自身的AI PC新品亮相了。

北京时间3月22日,微软宣布推出Surface Pro 10和Surface Laptop 6商用版,这也是微软首批专为商业用户打造的Surface系列AI PC,两款产品将于3月22日在微软官方商城开启预售。其中在中国市场,Surface Pro 10商用版售价为14088元,Surface Laptop 6商用版售价为12288元。

据不完全统计,目前包括联想、三星、戴尔、荣耀等厂商均已发布或亮相了自家的AI PC产品,苹果在发布新款MacBook Air M3机型时也特别介绍了产品的AI能力。可以说,今年是厂商们集中“秀肌肉”的节点。

如今,被视为“PC操作系统之王”的微软下场,又将为AI PC市场带来怎样新的想象空间?

微软AI个人电脑亮相

微软推出的Surface系列AI PC,重点落在了AI体验上。

硬件方面,此次微软推出的全新PC采用了最新的英特尔酷睿Ultra处理器,还拥有加快执行AI任务的神经网络处理单元(NPU)。

通过集成NPU,两款AI PC产品能够为用户提供包括Windows Studio Effects和实时字幕等在内的AI功能。

例如,微软Surface Pro 10商用版可以借助由AI支持的Windows Studio Effects始终保持发言人在画面的中心位置;Surface Laptop 6通过AI驱动的Windows Studio Effects,则能够确保用户在视频通话中展现出最佳状态。

此外,微软还特别强调为两款AI PC配备了“Copilot按键”。通过该按键,用户能够更快地调用Copilot并启用相关功能,包括寻找特定工作邮件并总结重点、进行会议总结纪要等文本处理,以及进行更换壁纸、清理存储空间、开启省电模式等功能操作。

不过相较于此前厂商们的产品而言,功能略有雷同。在今年1月的CES 2024上,联想集团一口气亮相了十余款AI PC。

据记者现场体验,联想的AI PC也同样能够进行上述类似AI功能。此外,联想发布的Yoga系列两款旗舰产品同样配备了Surface系列AI PC中力推的“Copilot按键”,从而便于更快地访问日常AI伴侣。

由此可见,在用户体验上,目前部分厂商的AI PC思路存在相似之处。不过从行业分析人士的角度看来,AI PC的功能还有很多想象空间。

“未来AI的发展场景非常多,不只是整理文档、AI字幕翻译、AI会议纪要等,所有对于图像处

理和文字处理和生成的方面,甚至于一些数据和编程的处理和生成都可能会有相关的应用场景产生。”IDC中国高级研究经理陈舒欽表示。

AI个人电脑元年之战

没有厂商愿意错过这一波AI PC的浪潮。

过去两年内,中国PC和智能手机市场一直处于负增长,如今市场则在回暖中。根据IDC预期,2024年PC市场整体向好,整体PC市场出货量同比2023年将增长3.8%。AI PC更是一枝独秀,2024年占比有望达到55%,到2027年市场渗透率有望达到85%。

在这样的趋势下,PC厂商已经先声夺人。今年1月,联想一口气亮相了10余款AI PC。其中Yoga系列两款旗舰产品Yoga Pro 9i和Yoga 9i二合一,搭载最新的英特尔酷睿Ultra处理器和联想AI芯片,并配备Yoga Creator Zone生成式AI软件。

其他厂商也同样紧随脚步。据不完全统计,目前市面上具备AI功能的PC产品包括三星Galaxy Book 4系列笔记本电脑、荣耀Magic-Book Pro 16、戴尔Inspiron灵越16 Plus、惠普HP Spectre x360商务本等多款产品,主打功能从双向AI降噪、AI文生图创作,到搭载AI引擎、AI软件等,不一而足。

不过,AI PC产业仍处于相对早期的阶段,行业内扎堆推出AI PC的同时,也在对AI PC这个新鲜事物加以定义。

此前,在接受媒体采访时,联想集团董事长兼CEO杨元庆指出,目前市面上PC的AI算力大约只能做到10 TOPS。“理想的情况是要到40个TOPS,这样才能算是AI第一代产品。现在叫第0代(AI PC产品)也可以,它能够在普通PC之上通过附加的方式实现部分AI功能。”

全新投资机会

在全新的AI PC产业生态里,不同类型的厂商各有分工。

其中,整机厂商是整合多样化异构算力的中间桥梁、AI生态的第一平台的整合者,还能从本地AI计算出发,为混合计算系统中的个人隐私保护和数据安全提供解决方案。

“联想处在直面客户的最佳位置,无论是做应用或是AI生态的厂商,都希望与我们进行合作。”杨元庆指出,“同时我们会更多地去做赋能,使这些厂商通过我们的AI PC,能够把应用做得更加丰富。”

尤其是在中国,整机厂商的重要性将更加凸显。据了解,微软目前在全球推行的Copilot服务暂时无法在中国本土落地。这也意味着在新一轮设备换代升级的周期内,中国PC设备本地化要求下整机厂商发挥的作用更加显著。

“从中国大陆地区来讲,一些微软的相关功能不能完整使用,所以中国PC厂商会进而发展自己的智能体,作为用户的助手,以此为切入点,从游戏、教育、工作、生活辅助等等方面进行全方位的AI加速。”陈舒欽认为。

与此同时,AI PC还将涉及多个产业。报告提示,AI应用繁荣将提振整体AI算力需求,无论是国际芯片厂商或是中国本土芯片厂商均有发展机会。模型厂商则迎来以端侧为起点的混合AI应用场景,构建于模型之上且能被智能体调用的AI应用预计将成为市场主流。

“在这场变革中,AI PC将带动PC行业开启全新成长周期,产业链有望迎来重新估值。”民生证券电子行业首席分析师方竞指出,在AI PC需求驱动下,PC产业链的上下游都会具有很大的投资机会。

供稿:《21世纪经济报道》杨清清/文