

# 超10款应用发布 百度迈入“AI原生时代”

■青岛财经日报/首页新闻记者 郭清鉴

“我们即将进入一个AI原生的时代。”昨日,百度创始人、董事长兼首席执行官李彦宏在2023百度世界大会上表示。

当天,李彦宏以《手把手教你做AI原生应用》为主题发表演讲,发布新款AI大模型,并带来新搜索、新地图等十余款AI原生应用。超千位企业家、创业者、高校专家、投资人、开发者等齐聚,与百度共同见证“AI原生时代”的到来。

据了解,百度AI大模型实现了基础模型的全面升级,在理解、生成、逻辑和记忆能力上都有显著提升。

## 丰富的原生应用是大模型价值所在

什么是AI大模型?简单来说,就是利用海量的数据和算力,训练出可以生成各种内容的超级人工智能。比如说,你想写一首诗,它就可以给你写一首诗;你想画一幅画,它就可以给你画一幅画。AI大模型的出现不仅改变着人们获取信息的方式,还将提升各行各业运转的效率。

李彦宏认为,AI原生应用的诞生得益于大模型的理解、生成、逻辑和记忆四大核心能力,百度的AI原生应用也是基于大模型来开发的。四大核心能力成就了生成式AI,成就了全新的想象和创新空间。

丰富的AI原生应用才是大模型的价值所在。李彦宏表示,百度新搜索具有极致满足、推荐激发和多轮交互三个特点。以极致

满足为例,当用户搜索问题时,新搜索通过对内容的理解,生成文字、图片、动态图表的多模态答案,让用户一步获取答案。生成式商业智能产品“百度GBI”可以通过自然语言交互,执行数据查询与分析任务,还支持专业知识注入,满足更复杂、专业的分析需求。

通过对海量文档、图片和视频的理解和再生成,百度网盘和文库拥有了创作能力。百度地图和智能办公平台“如流”通过理解、记忆等能力,变成了更贴心的出行助手和超级助理。插件则是一种特殊的AI原生应用,门槛最低,也最容易上手,能让开发者、创业者快速加入到生态中。此前,百度上线了灵境插件平台,目前已经有27万开发者申请入驻,覆盖法律、职场、学习等多个领域。

API是AI原生应用调用基础大模型的主要方式,企业和开发者可以在百度的千帆大模型平台上调用大模型API。目前,千帆大模型平台有42个主流大模型入驻,覆盖各行各业近500个场景。企业客户也可以登录百度智能云官网,在千帆大模型平台上申请测试百度大模型的API。

## 大模型将应用到各个领域

在百度世界大会现场,百度首席技术官王海峰针对百度AI大模型背后的关键技术和最新进展进行了解读。

百度集团执行副总裁、百度智能云事业

群总裁沈抖宣布全面升级“云智一体”战略,将其战略内涵全面升级为“云智一体,深入产业,生态繁荣,AI普惠”。此外,面向企业落地大模型的高效稳定的AI算力、大模型API调用、大模型二次开发、敏捷开发AI原生应用、便捷选购优质AI原生应用等五类需求,百度智能云“大模型超级工厂”分别给出了最佳服务方案;百度集团资深副总裁、百度移动生态事业群总经理何俊杰重点展示了新文库、百度电商慧播星等AI原生应用;在内容生态层面,通过百度APP“AI编辑器”等一系列应用为创作者生态赋能;在商业生态层面,通过AI Native商业产品“轻舸”等,助力企业效益提升。此外,何俊杰还公布了AI插件生态——“灵境矩阵”的最新成果,并面向全平台开发者、创业者发出邀请,共建大模型生态。

百度集团副总裁、小度科技CEO李莹表示,小度即将发布多款AI原生“新物种”,包括小度添添家庭机器人、小度青禾学习一体机以及全新智能音箱系列Tiantian Casa。

记者采访了解到,青岛早已启动布局人工智能赛道。2022年,青岛将人工智能产业列为24条重点发展的产业链之一,包括青岛市人工智能产业园等特色园区相继聚焦人工智能相关企业。目前,青岛吸引华为、阿里、中科曙光、创新奇智、以萨技术、极视角等一批头部企业来青布局,当前全市人工智能企业500余家,形成了人工智能与实体经济深度融合、双向突破的先导区建设新局面。

上半年市场规模达6.6亿美元

# 中国液冷服务器实现“爆发式增长”

在人工智能热潮带来的算力需求驱动下,今年上半年液冷服务器市场实现了“爆发式增长”。

国际数据公司(IDC)发布的最新报告显示,中国液冷服务器市场在2023年仍将保持快速增长。2023上半年中国液冷服务器市场规模达到66亿美元,同比增长283.3%,预计2023年全年将达到151亿美元。IDC预计,2022-2027年,中国液冷服务器市场年复合增长率将达到54.7%,2027年市场规模将达到89亿美元。

互联网厂商是中国液冷服务器市场的最大买家,政府、公共事业和教育占据次要市场份额,电信、金融、服务等行业的部署量也有稳健提升。未来更多传统行业如智能制造、医疗、能源等,将加大液冷相关投入。

## 互联网厂商是液冷服务器最大买家

数据显示,2023上半年,中国液冷服务器市场中,冷板式占到了90%,主要取决于传统数据中心对原有基础设施改造成本和难度的关注度,以及当前冷板式方案成熟度较高,商用基础较好。长远来看,浸没式在散热效率和单机柜功率、空间利用率等方面比冷板式具有显著优势,因此浸没式如何合理优化基础设施改造、建设成本,以及提高电子氟化液或其他冷却液的应用范围及可维护性,成为构建可持续发展绿色数据中心的必要条件。

从厂商销售额角度来看,2023上半年市场占比前三的厂商是浪潮信息、宁畅和超聚变,占据了七成左右的市场份额。越来越多的主流IT设备厂商均已公开表明将加大研发力度并加快液冷产品迭代速度,未来中国液冷服务器市场将持续保持高速增长。

从行业角度来看,互联网厂商是2023上半年中国液冷服务器市场的最大买家,上半年出货量中54%是由互联网企业购置,并有望在未来几年中持续加大液冷服务器的采购力度;政府、公共事业和教育占据次要市场份额;电信、金融、服务等行业的部署量也有稳健提升。

并且随着人工智能竞争加剧,“CPU+GPU”的异构计算服务器由于散热需求强烈,也开始逐步采用液冷制冷方式,并且有理由相信在生成式人工智能领域将加速构建可持续发展的绿色数据中心。

IDC中国服务器市场研究经理辛一认为,经过2022年的带动,互联网、超算中心对液冷制冷需求更加旺盛,2023年液冷服务器市场将持续大幅增长。同时伴随着更多成熟液冷项目落地,积攒更多经验,并将成功经验迁移至更多应用领域,以及人工智能大模型和5G的加持,未来更多传统行业如智能制造、医疗、能源等,将加大液冷相关投入。目前困扰中国液冷服务器市场发展的主要难题是如何平衡性能、能耗、成本之间的关系。尽管市场对数据中心制冷有旺盛的需求,但液冷技术从宏观上看仍处于发展早期阶段,产业生态建设仍不完善,存在多方竞争,无法形成凝聚力。

越来越多的行业标准推出带动生态链的有序发展。众多液冷生态链厂商及服务供应商在积极推动行业标准化,全力打造高水平液冷生态链,构筑开放生态,引领形成统一标



准,从传统意义的单纯项目供应,转变为液冷生态的主导者、设计者、构筑者,推进产业生态成熟。

## 液冷服务器成算力减碳“排头兵”

液冷服务器“爆发式增长”背后是AI浪潮下,算力需求持续攀升,进而带动大型数据中心等算力基础设施规模的扩大,而服务器作为大数据中心的基础配套设施,需求量随之提升。不过,伴随着数据中心蓬勃发展,其高耗能性带来了碳排放问题,这对绿色算力的发展要求带来了挑战。

国家能源局数据显示,2020年我国数据中心耗电量突破2000亿千瓦时,占全国总用电量的2.7%,创历史新高。IDC预计,2024年数据中心耗电量将占全社会耗电量的5%以上。数据中心节能减排势在必行。

对此,国家政策层对数据中心的建设提出了明确的指标要求。根据工信部发布的《新型数据中心发展三年行动计划(2021-2023年)》要求,新建大型及以上数据中心PUE降低到1.3以下(PUE值越接近于1,数据中心的绿色化程度越高),严寒和寒冷地区力争降低到1.25以下。“东数西算”工程中,要求内蒙古、贵州、甘肃、宁夏4处枢纽设立的数据中心集群PUE控制在1.2以内;京津冀、长三角、粤港澳大湾区、成渝4处枢纽设立的数据中心集群PUE控制在1.25以内。

为降低PUE值,液冷技术的发展成为必然,作为数据中心重要组成部分的液冷服务器也因此成为算力“降碳”的排头兵。有数据显示,液冷系统约比传统风冷系统节省电量30%至50%。此外,液冷服务器还可以排除海拔、地域和气温的差距,保证运行效率和性能,具备模块化能力。

## 成本是制约液冷服务器增长主因

在业内人士看来,成本仍然是制约液冷服

务器市场规模增长的主要原因。辛一告诉记者,目前国内数据中心大多以风冷为主,液冷服务器需要专业的液体冷却设备,因此改造成本较高,维护专业要求较高。

另一方面,辛一表示,没有统一标准也是一个重要因素。“由于目前液冷服务器市场处于早期发展阶段,国内相关液冷行业标准也在积极探索,今年以来有多个液冷技术标准发布,但是这些多以行业、团体标准为主,在液冷数据中心设计建设、机房运维、设备部署等环节依然参考、借鉴传统数据中心建设的经验,无法有效、快速部署,导致液冷数据中心的建设面临多种多样的问题,从而使用户对于采用液冷式制冷的信心不足。”辛一说。

浪潮信息服务器产品线总经理赵帅同样指出了这一问题,他告诉记者,当前液冷的价格仍然偏高,主要原因在于冷板的批量化生产目前还无法实现,而电脑里面造价比较高的快接头节点也没有统一标准。

“因此如果行业能够形成聚集效应,让大家都生产一样的东西,然后来降低它的整个的成本,可以促进产业的进一步蓬勃发展。”赵帅说。

在辛一看来,随着人工智能对算力芯片功耗提升,对液冷需求逐渐加强,并借助统一的技术实施规范标准,通过液冷项目的大规模部署,摊薄各项成本,液冷行业将进入螺旋上升的进程。

除此之外,他表示,长远来看,浸没式在散热效率和单机柜功率、空间利用率等方面比冷板式具有显著优势,因此浸没式如何合理优化基础设施改造、建设成本,以及提高电子氟化液或其他冷却液的应用范围及可维护性,成为构建可持续发展绿色数据中心的必要条件。

“不论是使用者、液冷生态链厂商以及行业专家,都希望尽快形成国家层面的技术规范标准,助力数据中心绿色节能、快速交付、统一管理,具备高度整合资源能力,推进液冷产业生态有序发展。”辛一说。

## ITMT 快报

### 小米启用新系统 将逐步接替MIUI

昨日,小米集团董事长兼CEO雷军宣布,小米全新的操作系统——小米澎湃OS(Xiaomi HyperOS),正式版已完成封包。小米14系列将是第一款搭载新系统的手机,已交付工厂,正式开始生产。从这一刻起,小米澎湃OS将逐步接替MIUI。

据了解,小米澎湃OS是小米“以人为中心,打造人车家全生态操作系统”。它基于深度进化的Android以及自研的Vela系统融合,彻底重写底层架构,为未来百亿设备、百亿连接做好了万物互联的公有底座。是小米迈向“人车家全生态”的关键跨越。据悉,小米澎湃OS可以为手机、小米汽车车机和物联网设备提供支持。

雷军回顾了MIUI的发展历程:13年前,2010年8月16日,小米第一款产品MIUI正式诞生。13年间,从几个工程师到四五千人的软件团队;从最初的“100个梦想赞助商”,到全球11.75亿累计用户;从第一版只有4个常用功能,到跨端互联互通的底层深度框架;与用户交朋友,做最易用、最人性化的系统体验,一直是MIUI不变的追求。

雷军表示,小米澎湃OS已准备了很久。“2014年,从IoT业务开始初具规模,我们就开始了探索式开发和验证。2017年,全新系统研发工作正式开启,以一套融合的系统框架统一支持全生态设备与应用,这就是我们的追求。”

综合

### 新型纳米电子设备 能效可提高100倍

美国西北大学工程师开发了一种新型纳米电子设备,可以迄今最节能的方式执行准确的机器学习分类任务。该设备的能源使用效率比当前技术提高100倍,可实时处理大量数据并执行人工智能(AI)任务,而无需将数据传输到云端进行分析。

该设备体积小、功耗超低且接收分析没有延迟,非常适合直接集成到可穿戴电子产品(如智能手表和健身追踪器)中,以进行实时数据处理和近乎即时的诊断。

为了测试该设备,研究人员用其对公开的心电图(ECG)数据集中的信息进行分类。结果表明,该设备不仅能够有效、正确地识别不规则心跳,还能从6种不同类别中确定心律失常亚型,准确率接近95%。

在机器学习工具可以分析新数据之前,这些工具必须首先准确可靠地将训练数据分类为不同的类别。对于当前的硅基技术来说,要对心电图等大型数据集的数据进行分类,需要100多个晶体管,每个晶体管都需要自己的能源设备来运行。但该纳米电子设备只需两台能源设备即可执行相同的机器学习分类,大幅降低了功耗。

团队先训练该设备来解释心电图数据,这项任务通常需要训练有素的医护人员花费大量时间。而该纳米电子设备能够从10000个心电图样本中准确识别每种心律失常类型。通过绕过将数据发送到云端的需要,该设备不仅为患者节省出关键时间,还保护了患者隐私。

未来这些纳米电子设备可整合到日常可穿戴设备中,根据每个用户的健康状况进行个性化定制,以实现实时应用。

据《科技日报》

### vivo 下月发布 自研大模型矩阵

vivo近日宣布,将于11月1日举行vivo开发者大会,本次大会将发布vivo自研AI大模型矩阵、操作系统OriginOS 4。

vivo自研AI大模型矩阵将在OriginOS 4系统中被首次应用,其中包括十亿、百亿、千亿三个不同参数量级的5款自研大模型,全面覆盖核心应用场景。

10月16日,C-Eval大模型评测榜单显示,vivo自研大模型在C-Eval全球中文榜单中排名第一。此前,vivo自研大模型已取得了CMMLU全球中文榜单第一的成绩。

分析人士认为,vivo自研AI大模型同时登顶两大榜单,展现出其在中文语言模型领域的领先实力,同时体现出业界对vivo的AI创新能力和发展潜力的肯定。

更重要的是,本次vivo打造的自研AI大模型矩阵将应用于新系统,证明vivo对大模型的探索已从技术研发阶段进阶至应用及产业布局阶段,对于整个行业而言具有十分积极的带动作用。

综合