

科技巨头打响“AI军备竞赛”

随着科技巨头们2024年财报的出炉,一场史无前例的AI军备竞赛也正式打响。

“战火”最猛烈的地方是海外。

2024年,微软、亚马逊、谷歌、Meta等4家美国科技巨头的资本支出分别达到756亿美元、777亿美元、525亿美元和373亿美元,均创下各自的历史新高。

2025年,这些公司资本支出将继续增长。微软预计投入800亿美元,亚马逊预计投入1000亿美元,谷歌预计投入750亿美元,Meta预计投入超600亿美元。这些巨额支出的流向也十分一致——都将用于AI基础设施建设。

毫无疑问,这次由ChatGPT掀起的新一轮AI浪潮,已经成为未来十年甚至更长远的变革力量。

即便今年DeepSeek的出现加速了技术平权,但在前沿智能的探索上,规模定律依然有效,头部企业想要抢占未来,仍要依赖“大力出奇效”。

在此背景下,国内的科技巨头们也不甘示弱。尽管在资本支出的体量上仍与海外企业有差距,但这已是中国企业AI投资的历史顶峰。

阿里腾讯争相加码

目前,在资本支出方面给出最明确信号的,是阿里巴巴(以下简称“阿里”)和腾讯。

2024年,阿里和腾讯的资本开支分别为725.13亿元、767.6亿元,同比分别增长197.04%和221.27%。

尤其是在2024年第四季度,阿里和腾讯不约而同地加大了投资力度。期内,阿里的资本开支为317.75亿元,环比增长81.66%,同比增长258.76%;腾讯为365.78亿元,环比增长114%,同比增长386%。

仅2024年第四季度的资本支出,阿里和腾讯均超过了自身2023年全年的资本开支总额,创下了单季支出的历史新高。

在财报业绩沟通会上,腾讯总裁刘炽平表示,从2024年下半年开始,腾讯就已经发觉AI的发展有加速迹象,因此也加大了对AI芯片的投入。

数月前,腾讯重组了AI团队,并开始增加AI相关的资本开支,以及加大了对原生AI产品的研发和营销力度。

腾讯董事会主席兼首席执行官马化腾认为,



AI应用大爆发的机会已经到来。“我相信各行各业都会受益于AI的普及,未来每个行业也会因为AI发生翻天覆地的变化”。

阿里认为这波AI机会可能是几十年一遇的行业变革。阿里首席执行官吴泳铭表示,从集团层面来说,AI战略的首要目标是追求通用人工智能(AGI)。

“AGI的实现成为我们的第一目标,可能远远超过现在可见的任何一个应用场景。现在的应用场景都只是AI能力提升过程中产生的机会。因此智能能力的提升才是这一轮AI技术生产力变革的核心主轴。”吴泳铭说。

尽管阿里强调底层、腾讯看重应用,但在AI技术的核心价值上,他们已形成了共识。因此这两家科技巨头均大幅提高了未来的资本支出预算。

其中,阿里计划未来三年将投入至少3800亿元用于建设云计算和AI基础设施。据吴泳铭透露,这笔投资在年度周期内的支出安排会相对平均,因此2025年,阿里的资本开支将超过1200亿元。

腾讯表示,2025年会进一步增加资本支出,预计占收入的低两位数百分比(通常为10%至

20%)。2024年,腾讯营收6602.57亿元,同比增长8%,按照这一营收增长率以及低两位数百分比的中值15%计算,腾讯2025年的资本支出或达到1070亿元。

如何支配超千亿元资金

中金公司在一份研报中指出,参考海外经验,近年来美国科技七巨头资本开支持续增加,支撑市场表现走强的同时,对其他行业以及总量层面的投资增速回升也具有积极意义。

因此,当国内互联网大厂纷纷表示加大对AI投入时,也有望带动AI算力产业链需求增长,并推动市场对科技资产重新定价。

对于未来一年如何支配这超千亿元资金,阿里和腾讯均透露了一些方向。

吴泳铭表示,阿里将加大三个AI领域的投入,分别是AI和云计算的基础设施建设、AI基础模型平台以及AI原生应用、现有业务的AI转型升级。

腾讯首席战略官詹姆斯·米歇尔则列出了资本支出的四个用途:第一个用途是购买GPU以支持广告技术,同时小部分用于游戏业务;第二

个用途是使用GPU进行大语言模型训练;第三个用途与云业务相关,腾讯会购买GPU服务器再出租给客户;最后则是用于面向消费者的推理业务。

然而,科技巨头们规划的巨额资本开支,也引发了投资者对其盈利状况的担忧。

吴泳铭坦言,未来三年是阿里历史上最大规模的资本支出建设周期,这些大规模投入的折旧摊销必然会对利润率产生一定的影响。

但他同时表示,从现在看到的需求以及对行业未来的预判上来看,这些投入相信很快会被内外部客户需求所消化。

詹姆斯·米歇尔更是直言资本支出的增加不一定会导致利润率的下降。他认为,腾讯的资本支出用于具体业务时,不仅不会带来压力,反而会产生不错的利润率。他举例称,2024年四季度,腾讯广告业务在AI技术的推动下,收入同比增长了60%。

目前在投资回报上最具不确定性的,是大模型训练。实际上,在之前很长一段时间内,外界都认为每一次大语言模型的更新都需要数量级更多的GPU,但随着DeepSeek的出现,现在整个行业都能利用现有的GPU实现更高的训练效率,因此对于GPU的需求已经不像之前那么巨大。

刘炽平指出:“总体而言,中国科技公司的资本支出占收入的比例低于一些西方同行。但我们认为,这一现象的原因是中國公司对效率的重视和GPU的高效利用,这并不一定会影响技术的最终效果。”

除此之外,阿里和腾讯还有一个共同点,即他们都拥有云计算业务,这也是现阶段最为明确的一个商业模式。

吴泳铭指出,现在各家公司大模型的差异化在逐渐收窄,开源模型和闭源模型的差异也越来越不明显,这样的现状反而利好于云计算公司,因为无论是开源模型还是闭源模型,大部分模型都要托管在云计算网络上。

对此,詹姆斯·米歇尔表示,在腾讯的投资组合中,云计算业务可能不是回报率最高的,但它确实能产生正回报,足以覆盖GPU的成本,进而抵消随之而来的折旧。

当下,AI大模型的进化浪潮仍在奔涌向前,这场技术革命远未抵达终局。在这个充满不确定性的竞技场中,阿里与腾讯的千亿押注,是在为下一个十年埋下伏笔。他们深知:放缓脚步就意味着失去先机,唯有持续加码投入,才有可能在未来的AI时代掌握话语权。

供稿:《21世纪经济报道》作者:白杨

大模型为科学研究装上“加速器”

人工智能驱动的科学(AI for Science)是加速我国科学研究范式变革和能力提升,提高我国科技创新能力,抢占全球科技制高点的重要支撑。中国工程院院士、之江实验室主任、阿里云创始人王坚认为,人工智能不是一次工具的革命,而是一次科学革命的工具或者是科学革命性的工具。

监测青藏高原水资源,对珊瑚礁进行生态分析,实现自主天文观测……在AI for Science浪潮下,中国科学院多个研究所正通过接入阿里通义千问QwQ-32B推理大模型,开启从“上天”到“入地”,从“观星”到“察水”的科研智能化远征。

精准预测太阳耀斑爆发

耀斑是太阳最剧烈的活动现象之一,每隔约11年就会大爆发。为了破解耀斑爆发之谜,近年来,学者们从数据驱动角度出发,开始用深度学习、机器学习等技术开展研究。然而,随着观测数据的不断积累和数据特征维度的不断增加,对算法规模的要求越来越高,研究人员亟须对海量多模态数据进行有效处理。

大模型的出现带来新解法。在中国科学院国家天文台,科研人员利用QwQ-32B构建的太阳物理大模型“金乌”,对太阳耀斑活动进行精准预测。

国家天文台人工智能小组核心成员李瑀昉表示,在阿里通义千问系列模型的基础上,团队通过监督学习、强化学习,训练模型理解、回答太阳物理问题,以及认识、分析太阳图像的能力。其中,太阳耀斑爆发预报任务以SDO(太阳动力学天文台)卫星公开数据、怀柔基地35厘米磁场望远镜数据、“夸父一号”(ASO-S)全日面矢量磁像仪数据为主要数据源进行训练和测试,预测准确率达到了领域内领先水平。

“天文大模型‘星语3.0’接入国家天文台兴隆观测站望远镜阵列Mini‘司天’,可自主控制观测、分析数据并推荐后续计划。”李瑀昉介绍,“全新升级的‘星语’大模型正在向智能体方向发展,将吸收更多细分领域知识,融合现有科研模型、算法,进一步提升科研效率。”

李瑀昉表示,科研对数据安全要求极高,因此,部分科研场景会把模型部署在本地。QwQ-32B可提供相对更低的部署成本,并能满足相关研究对模型能力的要求。

探寻青藏高原“水密码”

青藏高原是“世界屋脊”“亚洲水塔”,也是全球对气候变化响应最敏感的区域之一。第二次



青藏高原发现,青藏高原固态水正在快速融化,液态水呈增加趋势。气候变化导致的不确定性,会造成青藏高原潜在的水资源和能源保供风险。

面对这一挑战,中国科学院青藏高原研究所联合阿里云打造了首个专注于气候变化适应领域的“水-能-粮”多模态推理大模型——“洛书”。

“水-能-粮”耦合是指水资源、能源系统和粮食生产之间复杂的相互依存和影响关系,研究“水-能-粮”耦合对制定更有韧性的应对策略至关重要。

中国科学院青藏高原研究所助理研究员夏萃慧介绍,“洛书”以研究所自研的科学模型“思源”为底层,依托青藏高原的时空数据训练而成。其输出结果包含两部分:一部分是直接支撑水电生产预测的径流量,另一部分是精准描绘水文过程的高维数据。“但人类无法直接理解和使用这些数据,‘思源’接入通义千问后,可实现自然语言的查询和输出,实现高维数据可视化呈现,一线人员可据此直接进行决策。”夏萃慧说。

更值得一提的是,“洛书”结合“思源”和QwQ-32B,可直接对数据进行推理分析,得出结论。“比如,在枯水期要观测什么、未来要做哪些工作来适应气候变化等问题,都可以通过推理大模型得到解释,从而辅助决策。”夏萃慧说,在这项研究中,团队还借助阿里云提供的AI计算资源、数据存储和深度学习平台,高效处理海量数据和复杂的计算任务,实现对模型的快速实验和迭代,大幅提升科研效率。

未来,“洛书”还将接入Qwen-VL以实现高效识别图像数据,并与智能体、具身智能观测和空天地一体化动态数据中心相协同,共同为青藏高原生态保护与可持续发展提供科技支撑。

据《科技日报》作者:崔爽

► 科技前沿

科研人员研制出硬币大小生物电池

近日,中国科学院深圳先进技术研究院(简称“深圳先进院”)定量合成生物学全国重点实验室钟超团队联合深圳先进院集成所神经工程中心刘志远团队、深圳大学王任衡团队在《先进材料》上发表最新研究成果。他们利用3D打印活体水凝胶技术,成功研制出直径仅20毫米的微型便携式微生物燃料电池,创新性地整合了生物电刺激装置,通过刺激神经元实现了对电生理和血压的精准调节,在疾病治疗方面有较大应用潜力。该研究推动了便携式生物器件的发展,拓展了活体能源材料的研究前沿。

在该研究中,研究团队基于希瓦氏菌开发了一种创新的3D打印活体水凝胶材料,这种生物材料具有独特的弹性,使其能够通过3D打印技术形成包括精细的蜘蛛网状、叶片状等涵盖一维到三维的复杂结构。为确保微生物在装置中保持活性,研究人员首先将处于溶液状态的微生物封装在藻酸盐水凝胶中,并通过添加纳米纤维素、氧化石墨烯对体系进行优化,显著提升了材料的机械强度和导电性能。

设计并制作生物电池需要融合多学科知识,涉及合成生物学、材料科学和生物医学等领域。在深圳先进院,跨学科合作具有天然优势,

钟超团队基于突破性材料,与深圳先进院集成所神经工程中心研究员刘志远团队合作,共同开发微型生物电池系统。

受传统锂电池制造技术启发,研究团队采用阴阳极分离的优化设计:以活体水凝胶作为“阳极”,含铁氰化钾的藻酸盐水凝胶作为“阴极”,通过3D打印技术制备出高性能电极结构,最终成功构建了直径仅20毫米的微型生物电池系统。

实验结果表明,这块微型电池可稳定输出450毫伏电压,并实现长达10次完整的“自充电-放电”的供能循环。进一步的性能测试表明,生物电池具有出色的循环稳定性,能量损失极低,同时完全避免了传统电池中钴、锂等稀缺金属以及有毒电解质的使用,在环保方面具有显著优势。

尽管生物电池的技术概念新颖,但存在功率和输出电压波动(受细菌活性影响)两大局限,使得生物电池难以胜任需要持续稳定供电的场景。

基于这些特性,团队瞄准了瞬时神经刺激这一精准医疗领域——通过集成电容器系统实现电能的精准调控,成功开发出适用于神经调控的生物电池应用方案。

据《深圳商报》作者:陈小慧 朱诗颖

Meta发布开源人工智能软件

近日,Meta发布了最新的开源人工智能软件Llama 4的首批大模型版本Llama 4 Scout和Llama 4 Maverick。这也是该公司迄今为止最强大的两款AI大型语言模型。

不过Meta表示,更强大的一款被命名为Llama 4 Behemoth的大模型仍在训练中,该模型将在Meta的混合专家模型架构中充当“新模型的老师”。

这也是Meta基于Llama软件首次采用的MoE模型架构。在MoE模型中,单个token仅激活总参数的一部分。Meta表示,混合专家模型架构在训练和推理时计算效率更高,在固定训练FLOPs预算下,相比密集模型提供更高的质量。

以Llama 4 Maverick模型为例,该模型拥有170亿激活参数和4000亿总参数。Meta采用交替的密集层和混合专家层来提高推理效率。在

这种方式下,尽管所有参数都存储在内存中,但在服务这些模型时,只有总参数的一部分被激活。

Meta最新大模型的发布也意味着科技巨头在生成式人工智能大潮中的投资竞赛进一步升级。此前有报道称,Meta推迟了大模型最新版本的发布,因为在开发过程中,Llama 4在技术基准方面没有达到Meta的期望,特别是在推理和数学任务方面。

“我们的目标是打造世界领先的人工智能,将其开源,并使其普遍可用,让世界上的每个人都能从中受益。”Meta首席执行官马克·扎克伯格表示,“我认为开源人工智能软件将构建出领先的模型,而随着Llama 4的出现,这种情况开始出现了。”

据《第一财经日报》作者:钱童心