

今日聚焦 ■ 人工智能加速演进,打造智能经济新业态

词元高质量出海,大幕拉开

算电协同助中国大模型持续提升全球竞争力

没有超大平台入口,却抓住这波风口,技术早布局厚积薄发 沪产大模型领跑词元出海关键局

大模型推理,产出的是代码、分析、翻译和创意,是一种更深层的嵌入。一旦开发者对模型形成习惯,其切换成本会随时间积累而越筑越高

本报记者 李晔

本报记者 查睿

越来越多海外开发者及“龙虾”(OpenClaw)用户,对中国大模型愈发青睐,从而引发中国“词元(Token)出海”现象。

一名新加坡开发者,曾用美国模型公司 Anthropic 的旗舰模型 Claude Opus 4.6 编程,发现其“好用但贵”,单日词元消耗数千个,每日成本动辄上百美元。这名开发者转而拥抱中国大模型公司 MiniMax 的 M2.5 开源模型,性能相当,词元消耗所产生的费用仅为美国头部模型的 1/17。此非个案,而是风潮——碾压级的性价比,催生高频调用。

数据显示,近一个月来,在中美大模型 API 调用量的较量中,中国模型更胜一筹,调用量第一、第二和第四均为中国模型,分别为 MiniMax M2.5、阶跃星辰 Step 3.5 Flash,以及 DeepSeek V3.2。同时,平台前十模型的词元总消耗量的 60% 以上来自中国模型。

这意味着,全球开发者更喜欢调用中国大模型,愿意为中国词元付费。大模型推理所需算力、电力均在中国完成,却创新实现“词元出海”这一数字服务贸易形式。

其中,我国低成本电力无疑是核心优势之一。利用乌兰察布、宁夏等地可再生能源丰富及气候优势,我国实行“东数西算”,绿色电价较美国电价低 50% 到 70%,从而降低了数据中心运营成本。字节、阿里等大厂在中西部地区布局数据中心。

“词元出海”的更硬核辅助,在于中国大模型的技术实力,以及硬件厂商的高质量算力。

全球大模型聚合平台 OpenRouter 最新数据显示,3月16日至22日,中国 AI 大模型词元(Token)调用量连续第三周超越美国,中国词元正成为全球开发者的主流选择。其中,上海大模型 MiniMax 与阶跃 Step 多次登顶调用量榜单。

国家统计局数据显示,截至今年3月,我国日均词元调用量超过 140 万亿,相比 2024 年初增长超 1000 倍,比年初也增长 40% 以上。摩根大通预测,中国 AI 推理词元消耗量将在 2030 年增长至 3900 千万亿。

中国词元反超美国,也折射出全球 AI 竞争格局的转变。过去,全球模型竞争的核心是智力上限,追求万亿级以上参数规模和榜单跑分。如今,在巨额算力成本的压力下,AI 大模型越发务实,更关注在易用性与成本上解决用户问题,创造商业价值。

“这背后的核心逻辑不是价格博弈模型。”

MiniMax 多次登顶 OpenRouter 调用量排行,创始人闫俊杰在 2026 上海全球投资促进大会开幕式上介绍,MiniMax 致力于 AI 普惠,早在“龙虾”大火前就定下目标:复杂智能体运行 1 小时,成本 1 美元。

国内芯片企业同样励志。据记者了解,国内某头部模型厂商与国内自主芯片企业的合作算力已达数千卡规模。国产 GPU 在易用性与成本上的综合优势,对于日均词元调用量上亿乃至十几亿的头部大模型而言,是其商业竞争力的关键因子。

这一波“词元出海”,寒武纪、昆仑

奔,而是‘模型能效’与‘场景适配’的胜出。“阶跃星辰首席技术官朱亦博告诉记者,OpenClaw 等 AI 智能体执行的是词元高消耗与长周期任务,尽可能降低模型使用门槛才能赢得 AI 竞争。在算力进口受限的背景下,国产大模型通过底层架构创新,推理成本被大幅压缩,用更少词元可完成同类任务。前期的技术积累,迎来了当下的厚积薄发。

“上海很幸运,抓住了这一波风口。”采访过程中,多位业内人士都表达过类似观点。上海是全国最早布局 AI 产业的城市之一,虽然没有超大平台入口,但 OpenClaw 的崛起改变了这一局面。

快速反应,让上海大模型抢到第一张入场券。MiniMax 率先推出“编程套餐”模式,帮助开发者解决词元包月使用的痛点。阶跃星辰自研 StepClaw 率先接入微信,用户可在聊天界面直接向模型发送指令。

一同进化的还有上海大模型的能

力迭代。3月18日,MiniMax M2.7 首次展示“模型自我进化”路径,让模型深度参与自身训练与优化流程,在内部评测集上实现约 30% 的效果提升。上海扎实的算力基础,是中国词元出海的底气。目前,上海智算规模已突破 12 万 PFLOPS(每秒千万亿次浮点运算),上海市级智算调度平台纳管算力近 2 万 PFLOPS。上海正在升级一批新型算力设施,预计到 2027 年智算总规模力争达到 20 万 PFLOPS。

前不久,总投资 400 亿元的阿里飞天云智能华东算力中心宣布落户上海,将成为华东地区规模最大的智算枢纽之一。记者获悉,该算力中心将重点部署上海平头哥“真武”芯片,构建从底层芯片、智算平台到上层应用的全栈自主算力底座。

“未来 3 到 5 年,全球 AI 算力都非常紧缺。”阿里巴巴首席执行官吴泳铭表示,平头哥“真武”芯片出货量已达数十万片,未来将为上海词元出海提供更充足的算力保障。

能转化为有效吞吐。可预期,指在大规模集群、复杂业务负载和功率约束下,算力性能、时延和稳定性能被准确评估并持续兑现。如天数智芯有业内的独家 IX-SIMU 全栈软件仿真系统,可在集群部署前仿真模拟实际运行效果,减少“上线后发现不行”的风险。可持续,则指芯片不只跑今天的算法,还能持续向“后”兼容。

大模型推理,产出的是代码、分析、翻译和创意,是一种更深层的嵌入。一旦开发者对模型形成习惯,其切换成本会随时间积累而越筑越高。此番词元高质量出海,大幕刚刚拉开。

实现“生物级反应速度”,为具身智能奠定感知物理基础。其三是国产化具身“大脑”实现技术性突破。通俗来说,现在的机器人需要背巨大的“机箱”来装载设备,又重又耗电,且续航短、散热难。全栈国产化的 T200 和 T80 打破这一“不可能三角”,在极小的体积内实现超高性能,扫清具身智能规模化落地工程障碍。

第四项成果是全球首创“德沃夏克架构”具身专用计算体系。传统机器人的“大脑”负责思考决策,“小脑”负责运动控制,“感知系统”负责接收信息,互相之间沟通慢。德沃夏克架构首次将芯片层面把这三项功能合在一起,更关键的是,它让机器人拥有“进化”能力:见过的人、走过的路、做过的动作,都能实时储存并进化。

大会期间,人形机器人 AI 软硬件生态协同系列标准工作,“全球具身智能生态共建计划”正式启动。“具身智能研究院(筹)”同时揭牌。灵境智源将把技术能力与生态资源向全行业开放,助力上海乃至全国构建具身智能的创新策源地。

实现“生物级反应速度”,为具身智能奠定感知物理基础。其三是国产化具身“大脑”实现技术性突破。通俗来说,现在的机器人需要背巨大的“机箱”来装载设备,又重又耗电,且续航短、散热难。全栈国产化的 T200 和 T80 打破这一“不可能三角”,在极小的体积内实现超高性能,扫清具身智能规模化落地工程障碍。

第四项成果是全球首创“德沃夏克架构”具身专用计算体系。传统机器人的“大脑”负责思考决策,“小脑”负责运动控制,“感知系统”负责接收信息,互相之间沟通慢。德沃夏克架构首次将芯片层面把这三项功能合在一起,更关键的是,它让机器人拥有“进化”能力:见过的人、走过的路、做过的动作,都能实时储存并进化。

上汽集团量产线打造人形机器人“练兵场”

首位“员工”上岗,更多“实习生”排队转正中

本报记者 吴丹璐

记者昨天从上汽集团获悉,其首位人形机器人员工——“能仔 1 号”,正式在别克至境 E7 电池量产线上岗。这是上汽集团首个投入量产线应用的人形机器人案例,也是国内汽车行业首批应用案例。

目前,“能仔 1 号”在别克至境 E7 电池量产线承担电芯抓取、上料等核心工序。在电池包制造环节,传统人工上料存在带电操作风险与技能门槛,而传统工业机器人则面临占地大、柔性不足等局限。“能仔 1 号”具备视觉感知、双臂协同、力控抓取等拟人特性,可自主完成电芯来料识别、抓取路径智能规划,无需依赖固定编程与来料的精确定位。其导航引导轨迹累计定位精度±0.1 毫米,单件作业效率约 2 秒,占地面积不足传统自动工位的 15%,可全面匹配量产节拍要求与安全标准。

“能仔 1 号”由上汽通用与智元机器人联合研发,后者是上汽金控投资的企业之一。早在 2023 年,上汽集团就围绕旗下整车企业的生产需求,启动人形机器人产业调研与测试;同年,秉持“投早投新”策略,上汽金控开始布局人形机器人赛道,投资相关

工业场景验证的企业。

2024 年,在上汽金控牵头下,上汽通用与智元机器人实现首次对接,三方共同完成上百个工位、数十个场景的摸排验证,历经近千次测试与持续优化选定电芯产线作为首个场景,并最终实现量产落地。

打通汽车制造场景,对于人形机器人产业发展意义重大。一方面,汽车制造被公认为是人形机器人的最佳“练兵场”;另一方面,智能化汽车的许多核心技术,如感知芯片、毫米波雷达等,都与机器人存在技术复用性。小鹏汽车创始人何小鹏曾公开表示:“车企现有的技术储备 70% 可以直接迁移给机器人。”

总体来看,“产投联动”是上汽集团布局具身智能赛道的关键策略。当前,上汽金控对人形机器人的累计投资已超过 8 亿元,涵盖机器人本体、具身智能、触觉传感器、行星滚柱丝杠等多个核心领域。除智元机器人外,还有更多“实习生”排队转正中,如已投资企业银河通用的具身智能机器人也已在上汽延锋推进测试。

与此同时,上汽集团旗下整车与零部件企业正同步开展双足人形机器人在生产和物流等环节的应用试点,为适配更复杂地形和生产环境做好技术储备。

海洽会举办“以商会友·沪联全球”中外商协会对话对接“走出去引进来”全链需求

本报讯(记者 胡幸阳)昨天,“以商会友·沪联全球”中外商协会对话和企业对接专场在上海东方枢纽国际商务合作区举办。本次活动是首届海外投资与综合服务展洽会的重要组成部分,对接企业“走出去、引进来”全链条需求,为中外商协会、企业、产业园区搭建共话合作、共谋发展的务实平台,吸引来自全球 16 个国家和地区的 30 余家海内外商协会、百余家企业与会。当前,民营企业在上海企业出海格局中扮演着越来越重要的角色,已成为上海企业出海的“主力军”。2025 年,在全市对外投资备案项目中中方投

20 号胶期权和国际铜期权下月上市并向 QFII 开放 提升大宗商品市场价格影响力

本报(记者 张杨)上海期货交易所有全资子公司上海国际能源交易中心宣布,将于 4 月 22 日(含 4 月 21 日晚连续交易时段)正式挂牌上市 20 号胶期权和国际铜期权,两个期权品种作为境内特定品种,上市时间同步向合格境外投资者(QFII)开放。20 号胶是轮胎产业核心原料,铜是新能源、高端制造等领域的关键材料,均是支撑实体经济发展的重大大宗商品。我国是全球最大的 20 号胶消费国和铜消费国,相关产业链企业深度融入全球市场,对精细化、多元化的风险管理工具需求日益迫切。2018 年以来,上期能源上市了 20

号胶期货、国际铜期货和原油期货等 6 个期货和期权品种,市场运行平稳,交易、结算、行权和对外开放等各环节得到了市场检验。据统计,2025 年原油期货日均成交同比增长 23.61%,日均持仓同比增长 32.71%。市场普遍认为,上市 20 号胶期权和国际铜期权并对外开放,不仅能丰富境内外产业链企业的风险管理工具,助力企业对冲价格波动风险、提升风险管理水平,还能与对应期货品种相辅相成,完善价格发现体系,进一步发挥期货市场资源配置作用,提升我国在全球相关大宗商品市场的价格影响力,助力上海国际金融中心建设。

越剧双甲子推出百余场精品演出 上海越剧院龙华中路新址年末启用

本报(记者 诸葛漪)上海越剧院昨天发布 2026 全年演出计划。今年适逢《戏剧振兴三年行动计划(2026—2028 年)》开局之年、越剧诞生 120 周年、越剧宗师王文娟诞辰 100 周年,上海越剧院推出百余场精品演出及系列纪念活动。今年年末,上海越剧院将迎来龙华中路新址正式启用。新剧场以现代化设计理念融合传统戏曲美学,配备专业设施与沉浸式观演空间,为越剧艺术创作演出提供全新平台。结合新址开幕系列演出,剧院正与知名 IP 及著名艺术家携手合作,拟为新剧场打造沉浸式环境体验,形成独具特色的剧目社区氛围,开展常态化驻场演出。今年,《孔雀东南飞》《玉蜻蜓》《珍珠塔》等经典之作将相继登台亮相,在演出中让宗师的艺之花在新时代表

越剧双甲子推出百余场精品演出

舞台上焕发新生。剧院还将精心策划举办纪念越剧宗师王文娟诞辰 100 周年专场及系列演出,复排优秀传统剧目《追鱼》,让“王派”艺术薪火相传,生不追。新剧《华山奇缘》以《沉香太子全传》为基础,根据京剧传统戏《宝莲灯》和越剧传统戏《劈山救母》改编,对传统表演方法与现代舞台科技的结合进行有益尝试,计划于今年夏季首演。2025 年,在上海市委宣传部指导下,东方卫视举办《越剧青春》,上海越剧院第十代青年演员参加,获得广泛好评与关注。今年将由第十代青年演员传承演出的《桃李梅》人物众多、个性鲜明,呈现青春靓丽的风采。同样由第十代青年演员主演的环境式越剧《红楼梦》将在长江剧场驻场演出,打破传统剧场观演边界。上海越剧院还将启动全国规模化巡演,持续夯实越剧艺术代际传承、青年受众培育的坚实基础。

企业级智能体系统从“模速空间”起步

让 AI 参与经营决策

本报(记者 舒抒)昨天,“模速空间”生态企业特赞科技正式发布企业级智能体系统 GEA (Generative Enterprise Agent),目前该系统已在全球超 180 家企业部署,其中包括 60 余家世界 500 强企业。

此次发布的企业级智能体系统 GEA,不同于传统生成式 AI 依赖模型提示词驱动任务执行,而是以企业运营过程中产生的“上下文”作为推理基础,让 AI 理解品牌规范、历史决策逻辑、用户数据与业务流程,从而参与企业跨部门、跨场景的经营过程。

特赞创始人、CEO 范凌表示,企业级 AI 的发展重点并不在于生成什么、回答什么,而在于其能否真正理解企业独有的业务语境,进入复杂的经营流程,并且在有约束的条件下进行判断。“随着企业软件从流程编排系统向推理系统转变,AI 的竞争焦点正从模型的构建能力转向模型的上下文理解能力。”

具体技术层面,GEA 的技术基础来自特赞自研的 4 层企业级智能体架构,包括“意图层”“编排层”“智能

体技能层”以及“上下文层”。其中,“意图层”负责将企业的经营目标转化为结构化的任务路径。“编排层”由特赞自研的分散推理模型驱动,承担任务拆解、多路径推理与模型调度功能,可在超过 30 种基础模型之间动态选择最优的执行路径。“智能体技能层”提供超过 400 个可调用的技能模块,能够将真实的企业工作流沉淀为可重复采用的执行能力。

特赞发布的个性化智能体“GEA Claw”可调用企业的上下文系统,用户研究资产与技能模块,在企业真实的业务流程中持续推进策略生成、内容生产和决策优化,让企业具备可 7x24 小时运行的智能体执行网络。

基于 GEA 架构,特赞还形成了 4 类企业级智能体产品体系,覆盖企业关键的经营流程,包括洞察研究智能体、内容增长智能体、设计创作智能体以及产品创新智能体,使智能体从单一的工具升级为跨部门的协同系统,最终奔向企业级智能体的发展愿景——成为连接组织知识、业务流程与执行能力的新一代 AI 基础设施。

具身智脑自研成果在“大零号湾”发布

破解机器人“缺智”

本报(记者 黄勇娣)昨天,“人形机器人 AI 软硬件生态融合工作组”工作推进会暨首届具身智脑技术生态大会在上海大零号湾科创大厦举行。灵境智源集中发布 4 项自主研发的具身智脑核心技术成果,覆盖计算架构、感知处理、运动控制、智脑底座全技术环节,助力破解国产机器人“缺智”困境。

上海人工智能研究院智脑创新中心、灵境智源创始人孙博介绍,这 4 项成果是全栈国产化具身智脑灵境方案的核心技术支撑。

首先是国产化具身“小脑”性能革命性升级。灵境智源推出行业最强算力小脑 T40,搭载首发全链路软硬件协同视觉控制模型,通过单阶段端到端训练融合视觉感知与运动控制,结合深度拟人模仿学习,让机器人在“小脑”上即实现“目视行走、跨沟避障”。

其二是在行业内首发“硅基皮层”实时多模态融合系统。这相当于给机器人构建了一个“感知中枢神经”。它能把眼睛看到的、身体感受到的触觉、力度、平衡感在硬件层面瞬间融合,延迟低于 1 毫秒,让机器人



上海欢乐谷 国潮文化节

上海欢乐谷 2026 国潮文化节正在火热进行中。本次国潮文化节融合限定剧本游、热门 IP 联动、国风演艺、潮玩互动等多元玩法,打造沉浸式国潮体验,邀请市民游客化身“见习宝藏探员”,在欢乐中解锁国潮新乐趣,实现遛娃、潮玩两不误。欢乐谷还打造了丰富的亲子游玩内容,让家庭游客共度温馨时光。萌宠乐园里,浣熊、羊驼等 10 余种萌宠可与游客近距离互动;童趣音乐会、萌偶见面会、欢乐海狮秀等童趣演出全天不间断上演;海洋奇幻馆、饼干警署特训营、金银岛等 30 余个亲子游乐项目可畅玩到底,让孩子们尽情释放童真。活动将持续至 4 月 30 日。 本报记者 孟雨涵 摄