

GDC 2025 全球开发者先锋大会

GLOBAL DEVELOPER CONFERENCE

中国·上海 | 2025 2.21-2.23

本报记者 邵阳

算算,深度求索公司 DeepSeek 发布大招,已经过去30天了。

“DeepSeek-R1 性能对标 OpenAI o1 正式版+开源模型权重+网页版和 App 免费可用+API 每百万 tokens 输入只要 1 块钱+论文训练技术全部公开”,1 月 20 日的这套“组合拳”下来,全球大模型产业巨震。

上海稀宇科技有限公司,更为人们熟知的名字是 MiniMax,副总裁刘华有个判断:大模型仍处于快速

发展阶段,“保持着每隔几个月就会取得重要突破的迭代频率”。

仔细想想,甚至还有点“保守”。北京时间 2 月 18 日,大洋彼岸埃隆·马斯克旗下的人工智能公司 xAI 发布新一代 AI 模型 Grok3,号称“地球上最聪明的 AI”。

这 30 天,你可以说是“你方唱罢我登场”,你也可以说是“大风起兮云飞扬”。但有一点谁都承认,大模

型人的腰杆硬了。前几年,中国企业每每推出模型,都要和 GPT3.5、GPT4 比比;到昨天,Grok3 的“超越对象”成了 DeepSeek——这本身就是一种成功。

再过两天,以“模型全球,无限可能”为核心主题的 2025 全球开发者先锋大会,就要在世界人工智能大会的发源地——徐汇西岸举行了。本报记者试图用尽可能通俗的语言,梳理这一个月,国内外,特别是上海的大模型同行们,都在做什么。

国内外大模型企业“你方唱罢我登场”

DeepSeek 爆火 30 天 上海的同行人都在忙啥

作为大模型独角兽的两家申城 AI 企业阶跃星辰和 MiniMax,都在各自产品中“上新”了 DeepSeek。当然,将这两家“单拎”出来讲,也是考虑到它俩的主攻方向是多模态,并未和 DeepSeek 构成强竞争关系。

若是没有 Grok3“搅局”,18 日“大模型日报”的“头版头条”,本来是属于阶跃星辰的——一家被权威科技商业化智库《麻省理工科技评论》“点名”值得关注的中国 AI 企业。当天,阶跃星辰联合吉利汽车集团,开源了两款多模态大模型:全球范围内参数量最大的开源视频生成模型 Step-Video-T2V,还有行业内首款产品级开源语音交互大模型 Step-Audio。

和引爆科技界的 DeepSeek R1 一样,

在拥有自研大模型的上市公司里,总部位于上海徐汇的商汤科技主要通过 MaaS(模型即服务)模式来提供 DeepSeek 模型服务。

记者梳理发现,近 30 天里,商汤的“官宣”吸睛有限:1 月 20 日,日日新融合大模型交互版(SenseNova-5o)正式对外提供实时音视频对话服务,限时免费使用;26 日,宣布商汤日日新融合大模型交互版(SenseNova-5o)在实时音视频中文测评基准 SuperCLUE-Live 首测榜单中以总分 68.59 排名国内第一。2 月 8 日,商汤大装置上架 DeepSeek 系列模型,企业客户和开发者可享受 3 个月内 1000 万 tokens 免费使用权益。

2025 年,AI Agent 有望成为新一年最重要的产品形态,引领 AI 从传统的“工具”

阶跃星辰、MiniMax 各自“上新”DeepSeek

阶跃星辰的 Step-Video-T2V 视频生成模型,采用最为宽松的 MIT 许可协议,支持免费商用、任意修改和衍生开发。阶跃星辰创始人、CEO 姜大昕博士告诉记者,这次发布也意味着阶跃星辰成为大模型开源世界的又一股中国力量。

这两款大模型也在 AI 社区引发了关注,“抱抱脸”(hugging face)官方转发了中国区负责人给予的高度评价:“简直就是下一个 DeepSeek。”

此外,阶跃星辰在春节前也没闲着——1 月 21 日宣布多模态理解大模型 Step-1o

Vision 上线,语音模型 Step-1o Audio 能力升级;22 日又发布视频生成模型 Step-Video V2 版本;24 日,推出“跃问 AI 创意板”,在跃问最新版本上线,支持 0 代码基础、3 步开发应用。

相比之下,连续 2 个月保持全球 AI 视频产品榜首的 MiniMax,“官宣动作”少了些——仅在 1 月 20 日宣布全新升级 T2A-01 系列语音模型,全球同步上线海螺语音产品。不过,2025 年以来,MiniMax 是十分“活跃”的,曾在十天内连发四个 AI 模型,包括基础语言大模型 MiniMax-Text-01 和视觉多模态大模型 MiniMax-VL-01,以及视频模型 S2V-01、语音模型 T2A-01。而两个 MiniMax-01 系列模型,更是公司创建以来首度开源。

商汤科技 发布新应用开发框架

角色向更具互动性与协作性的“伙伴”角色转变。在本周的走访中,记者获悉,商汤科技将在 2025 全球开发者先锋大会上发布面向开发者的一站式开源 Agent 应用开发框架——LazyLLM。

“LazyLLM 利用轻量网关实现分布式应用一键部署,当智能体开发完成后,应用者能够一键将其部署到网页、企业微信、钉钉等平台。”商汤大装置产品总监刘叶枫介绍。举例来说,假设要开发一款多模态客服问答机器人,传统方法可能需要开发者花费一周时间进行模型微调、串联各个功能,最终才能形成一个完整的 Agent 应

用。而通过 LazyLLM,就可以通过低代码拖拉拽快速实现各种功能,无论是查询快速、智能客服,还是转接人工服务,通过大模型的意图识别和分割,都能够将客户指向不同的意图识别流程,让开发更加高效。

另据透露,商汤科技还将在本次大会上发布“代码小浣熊”2.0 版本,为开发者和企业用户带来更加全面和强大的 AI 支持。

“前两年,一些新的 AI 技术让大家感觉需要花很多的教育成本。然而,随着 DeepSeek 爆火,更多人尝试使用,对于整个 AI 生态和产业发展来说,无疑是个利好消息。”商汤科技产品经理贾安亚表示,“我们看到了一些开源模型在能力上的亮点,也希望保持产品的开放性、博众家之长,去接入更多更好的模型,更好服务 B 端和 C 端用户。”

他山之石

马斯克的 Grok3 “最聪明”也要靠“有钱”

Grok3,在“诞生”前就被描述为“地球上最聪明的 AI 模型”。在北京时间昨天的发布会上,Grok3 的表现可以说“名副其实”,在数学推理、科学逻辑推理和代码写作等方面的能力,在多项基准测试中均取得了比 DeepSeek-V3、GPT-4o 和 Gemini-2 pro 更优的成绩。

在被誉为大模型竞技场的全球知名 AI 模型评测平台 Chatbot Arena 中,Grok3 版本的得分达到 1402 分,超过了包括 GPT-4o、DeepSeek-R1 等在内的所有其他模型。

在发布会的演示过程中,马斯克及其团队成员让 Grok3 寻找去火星的时间窗口,演示了其信息收集和推理能力,同时还创作一款融合《宝石迷阵》和《俄罗斯方块》的游戏,展示了 Grok3 的创意能力。

“算力军备竞赛”中,DeepSeek 走出了一条“四两拨千斤”的突破之路。而 Grok3 还是信奉“大力出奇迹”,由 10 万个以上的英伟达 H100 GPU 驱动,提供了超过 2 亿 GPU 小时的计算资源。当然,它不是开源模型,而且还要付费,典型的“好且贵”。

马斯克的 xAI,毫无疑问成为国际 AI 大模型的强有力竞争者。意欲逐鹿中原的,除了渐露锋芒的国内大模型企业外,OpenAI、谷歌、Anthropic 等亦是支支劲旅。事实上,2025 年前两个月,全球大模型玩家仿佛集体“冲业绩”——

OpenAI 于近日宣布,公司将在未来几个月内推出新一代人工智能模型 GPT-5 和 GPT-4.5。而在本月初,谷歌高调更新了 Gemini 2.0“全家桶”,包括通用模型 Gemini 2.0 Flash、谷歌最强的模型 Gemini 2.0 Pro,以及“性价比最高”的模型 Gemini 2.0 Flash-Lite。

相关链接

2025 全球开发者先锋大会

2 月 21 日至 23 日,大会将在徐汇的多个场地举行,包括西岸大剧院、穹顶艺术中心及西岸艺术中心的 A 馆和 B 馆,还有漕河泾会议中心。大会以“模型全球,无限可能”为核心主题,致力于推动 AI 产业的实际应用和加速其商业化步伐;聚焦于大模型、算力、语料、工具和软件平台等关键技术,涵盖硬件开发、云计算、大数据、物联网、AI、机器人、区块链及元宇宙等多个领域的开发者群体;定位为“社区的社区”,广泛吸引国际开发者社区的参与。

业内说法

DeepSeek 的横空出世,让更多普通人关注起大模型产业。行业创新的重点是否要转移到应用?“价格战”打不打、怎么打?对于 AGI 来说,多模态究竟有多重要?这些问号,都等待着被拉直。

毫无疑问的是,在大模型“撞墙论”(发展遇到瓶颈)不绝于耳的当下,行业任何时候都更为迫切地渴求创新。

不过,MiniMax 副总裁刘华显然是对“撞墙论”嗤之以鼻的一派,“未来两到三年,除增加训练和推理算力的 Scaling Law(尺度定律)以外,Transformer 架构变化的算法创新、合成数据、强化学习、多模态融合等,都将成为模型能力带来几倍量级提升的技术空间。”

他是有绝对底气说这话的。今年 1 月的那

AI 会越来越“渗透”进生活

次开源,MiniMax 首次在一个 4560 亿参数的超大规模商用模型上,引入了有别于传统 Transformer 架构的“线性注意力”机制,以极低的算力成本,试图为困扰着整个大模型行业的难题提供一个新的解题思路。

近年来,大模型研发框架敏捷迭代,这就要求研发人员对技术变革抱有极致的热情,而不能习惯于从过去的经验中寻找答案。刘华一针见血地指出,大量聘用年轻研究员并充分授权,打造科学的研发组织体系,是组建高效大模型研发团队的唯一路径。

最“新鲜”的佐证是,DeepSeek 团队昨天发布一篇新论文,内容先放一边,在合著名单里,论文的第一位作者是在实习期间完成的这项研究。而就 MiniMax 来说,研发团队也非常年轻,是“90 后”的天地,大部分人都 35 岁以下,好多研究生在博士阶段就来实习,一毕业就签约了。

对普通大众而言,在 DeepSeek 的推动

下,大模型正朝着越来越智能且免费的方向发展,这也加快了大模型“飞入寻常百姓家”的脚步。“高水平模型的持续开源,能让大众免费体验优秀大模型的技术能力,这将推动 AI 的渗透率,还能吸收更多开发者的力量。”刘华告诉记者。

“一款模型做得好,不代表其他模型就没有了机会。很多友商觉得 DeepSeek 很好,自己也主动接入、愿意用,这是一个非常开放的心态,我觉得这挺好。”刘华说。



徐汇滨江汇集了众多 AI 企业 本报记者 陈梦泽 摄