

从CV“四小龙”到大模型“四龙五虎” AI混战进行时

本报记者 曲忠芳 北京报道

时代的洪流滚滚向前，牵引前进的车轮也在不断更替迭代。在科技行业，业界时常会将

一些新兴的、有潜力的、四家典型的独角兽公司称为“四小龙”，体现了对这些“明星”公司的重点关注和期待。在人工智能（AI）领域，以计算机视觉识别

（CV）起家的“四小龙”还未远去，大模型“四小龙”“五虎”已然闪亮登场。

《中国经营报》记者注意到，CV“四小龙”指代的是商汤、旷

视、云从、依图四家计算机视觉识别技术公司；当下关于大模型的“四小龙”“五虎”的指代对象，业界尚未达成完全的一致，但大多指向智谱AI、月之暗面、百川智

能、MiniMax四家。除此之外，面壁智能、零一万物等明星公司也有望争取“五虎”席位。我们总结前后两代“小龙”的特点，不外乎出身或履历光鲜的创始团队、受

到资本机构的青睐与追捧，在市场中已获得大型客户的关注等。而不论是CV“四小龙”，还是大模型“四小龙”，AI技术的商业化价值仍在验证的过程中。

亮丽的创始团队 对风险投资的依赖

CV“四小龙”均有着光鲜亮丽的创始团队，并受到资本界的热捧。

从2006年到2009年，人工智能领域实现多项技术突破，包括ImageNet数据库建立、深度学习模型发展、自然语言处理（NLP）、机器学习算法优化等，推动了图像识别、语音识别等领域的快速发展。在此背景下，CV“四小龙”在2010年前后纷纷成立，获得资本的热捧，迅速完成多轮融资，成长为耀眼的AI明星企业。

CV“四小龙”均有着光鲜亮丽的创始团队：商汤的创始人汤晓鸥是顶尖的人工智能科学家，香港中文大学信息工程教授，旷视科技的实控人印奇出身于清华大学“姚期智实验室”；依图云科技的周曦出身于中国科学院重庆研究院，云从科技正是由中国科学院的人脸识别团队孵化而成；依图的创始人朱珉和琳晨曦同样是杰出的人工智能科学家，拥有出色的学术及工作履历背景。

大模型明星公司被冠以“四小龙”名称时，“大力出奇效”的大模型及生成式AI技术已于2022年年底被OpenAI所验证，近两年迅速在全球科技界走红，引领了新一轮的技术变革。大模型明星公司同样普遍拥有出色的创始人或创始团队背景。创立于2019年的智谱AI初始团队来源于清华大学计算机系知识工程（KEG）实验室，首席科学家唐杰是清华大学计算机系教授，CEO张鹏是清华计算机系博士；百川智能由互联网行业“老兵”王小川于2023年4月创立，王小川入局大模型一度带动了国内的AI创业热潮；月之暗面创立于2023年3月，创始人杨植麟毕业于清华大学和卡耐基梅隆大学，其大模型产品Kimi率先因“长文本”出圈，成为新锐大模型的



在2024世界人工智能大会上，多种搭载了AI大模型的智能终端纷纷亮相。视觉中国/图

典型代表；MiniMax创立于2021年12月，创始人闫俊杰毕业于中国科学院自动化研究所，曾在商汤担任副总裁；面壁智能的核心团队来自清华大学的自然语言处理实验室成员。

AI明星公司的另外一个显著特征便是受到资本界的热捧。经记者统计，CV“四小龙”在IPO或冲击IPO之前几乎都完成了10轮左右的融资，投资方涵盖海外资本、互联网巨头、国资基金等。截至目前，大模型“四小龙”同样也完成了数轮融资，智谱AI继在2023年累计获得超25亿元人民币融资后，又在今年6月完成了新一轮约4亿美元的融资，最新估值约30亿美元；月之暗面在今年2月完成新一轮10亿美元融资后，投后估值达到25亿美元；百川智能继去年10月完成3亿美元的A1轮融资后，据媒体今年4月报道，该公司正在寻求新一轮融资。零一万物作为创新工场“塔尖孵化”

项目，仅一轮融资后估值便超过了10亿美元，由创新工场董事长兼CEO李开复亲自带队。MiniMax自成立以来至今年6月也完成了多轮融资，估值超过了12亿美元。

CV“四小龙”凭借人脸识别、图像识别在政企市场迅速攻城拔寨，软硬件产品及解决方案在政务、智慧城市、商业、零售、制造、文旅、交通等多个行业领域得到广泛应用。在推进商业化的同时，CV“四小龙”纷纷发起IPO，试图在二级市场募资“输血”。旷视最早于2019年8月就向港交所递交上市申请，然而适值国际环境突变，港股闯关未果后转向A股科创板，目前仍在注册环节中。商汤于2021年12月30日在港交所上市，云从则于2022年5月27日登陆科创板。早于旷视、云从向科创板提出IPO申请的依图，则最早“离场”，2021年7月主动撤回了IPO申请。

AI商业化的故事由大模型接棒

当前大模型、生成式AI的技术演进仍在加速向前，而关于国产大模型“四小龙”的故事才刚刚开始。

从CV“四小龙”提交的招股书来看，几家企业均处于长期亏损状态，研发投入居高不下。而从已上市的商汤、云从两家的财报数据来看，这种亏损状态目前仍未扭转。

商汤2023年财报显示，公司全年营收34.06亿元，同比减少10.6%，归属母公司股东的净亏损为64.40亿元，同比扩大6.5%。大装置和大模型服务带动生成式AI收入实现11.8亿元，同比增长近200%，在总收入中的占比达到34.8%，传统AI、智能汽车两大业务板块收入则分别贡献占比为53.9%、11.3%。

而据云从科技财报，该公司去年总营收实现6.28亿元，同比增长19.33%，归母净亏损6.43亿元，同比收窄约26.0%。

值得注意的是，商汤、云从在过去几年的销售毛利率保持在50%上下的水平，但销售净利率却一直在一100%以上，这表明企业在销售产品或服务时能够覆盖直接成本并获得毛利，但包括诸如管理费用、销售费用、研发费用等在内的运营成本保持着过高的水平。

需要指出的是，在大模型时代，尽管属于CV“四小龙”的星光已“黯淡”不少，但这些AI企业尚未远去，目前都在向大模型领域转型发力，尤其是商汤旗下“日日新”大模型在前不久闭幕的世界人工智能大会（WAIC）期间已升级至5.5版本，在智慧城市、智慧医疗等多个行业领域中实现落地。无论是商汤，还是云从，目前营收来源均以To B业务为主。

To B市场的合作、大模型赋能行业的落地，这一商业化逻辑也被大模型明星公司所沿用，大多数通用大模型公司选择了向B端市场拓展业务的路线。

智谱AI方面向记者介绍，该公司提供MaaS（模型即服务）的商业化服务模式，包括轻量级API接入、云端私有化部署、本地私有化、软硬结合等方式，通过“AI+”来推动各行各业生产范式革新。月之暗面的商业模式则更倾向于To C端，其负责人介绍，Kimi采取向开发者、企业开放API接口的标准模式，另据记者观察，Kimi平台近来也在试水用户打赏、搜索等新模式。

知名经济学家、工信部信息通信经济专家委员会委员盘和林认为，大模型仅靠在To B、To G等方面打转或许是难以真正打开市场的。

不同于国产大模型，ChatGPT这些海外AI企业的落地路径是先在C端取得了成功，再去和B端企业谈应用的融合。如果大模型公司只采用一些简单的AI产品或模型先在B端落地变现，可能较难取得成功，因此才会有“你方唱罢我登场”、明星AI企业轮换的现象。要想持续地占据市场主导地位，需要企业更多的耐心与持续的投入。

谦询智库联合创始人龚斌指出，当通用泛化能力不足的时候，往往只能做To B的生意。如果按照OpenAI提出的通用人工智能五级能力评估体系来看，当AGI（通用人工智能）达到第二阶段（Level 2）时，或许就具备了在消

费级（To C）市场大规模流行的所谓“爆款”AI原生应用的条件。相比于前一波的计算机视觉明星公司，这一波大模型公司已推出了面向C端的应用，接近PMF（产品市场契合点），年收入有望做到几亿美元，而且团队规模都不算大，跟以往的纯烧钱模式有了很大区别。龚斌指出：“AI作为一种新的通用目的技术，其发展会遵循基本的创新扩散S曲线规律，必须等到突破临界阈值，包括技术成熟度即AI模型本身的能力阈值、市场接受度即AI应用跨越鸿沟，尤其是2C应用达到引爆点或者说大众市场流行的拐点，才能解决商业化的窘境。”

需要说明的是，就在前不久，根据彭博社报道，OpenAI为跟踪旗下AI系统在超越人类方面的进展，将AI发展进程划分成了五个层级：第一级是聊天机器人，能使用对话语言的AI；第二级是推理者，能解决人类级别问题的AI；第三级是智能体，能采取行动的AI系统；第四级是创新者，能辅助发明的AI；第五级是组织者，能够完成组织工作的AI。在OpenAI看来，该公司的AI目前正处于第一级，即将达到第二级“推理者”的水平。

毫无疑问，当前大模型、生成式AI的技术演进仍在加速向前，CV“四小龙”的镜鉴不远，而关于国产大模型“四小龙”的故事才刚刚开始，旧人、新人以及综合性科技大厂的AI混战也刚刚开始，谁能持续稳居“第一梯队”，谁能最终找到商业化的金钥匙，显然还有待时间的检验。

简单算术题难倒主流大模型 大模型偏科严重

本报记者 李玉洋 上海报道

近日，音乐综艺《2024歌手》两名选手“13.8%和13.11%的得票率谁高”的话题吸引关注。该话题不仅暴露了一些网友堪忧的数学水平，也让大模型集体“扑街”。因为媒体测试了12个国内外主流大模型，包括ChatGPT-4o、月之暗面kimi、智谱清言、阶跃星辰跃问、百川智能百小应等模型均答错了。

为什么这些大模型会在小学数学水平都能应付的比较数字大小问题上出错了？对此，月之暗面回应称：“其实我们人类对大模型的能力探索，无论是大模型能做到

什么，还是大模型做不到什么，都还处于非常早期的阶段。我们非常期待用户在使用中能够发现和报告更多的边界案例（Corner Case），不管是最近的‘9.9和9.11哪个大’、13.8和13.11哪个大’，还是之前的‘strawberry有几个r’，这些边界案例的发现，有助于我们增加对大模型能力边界的了解。”

“大模型有token（文本处理的基本单位）化和自回归的特性，有几类简单问题对大模型来讲都很困难，比如问strawberry里有几个r，还有9.11和9.9谁大。这种看似简单，对大模型都算难题。”MiniMax方面对《中国经营报》记者如

此表示。

结合今年高考期间，有不少人用大模型去做高考题，包括GPT-4在内7个大模型在高考测试中语文和英语考试水平普遍不错，但数学全不及格，最高分也只有75分。为什么数学对于大语言模型（LLM）而言这么难呢？

一些行业人士将大模型们数学不好的原因归结于LLM的架构问题，大语言模型往往是通过预测下一个词的监督学习方式训练。对此，做智能客服这类大模型垂类应用的容联云大模型产品负责人唐兴才表示：“现在生成逻辑确实是这样的，大模型目前最擅长

的还是语义理解。”

2024年被称为大模型应用落地元年，大模型如果连“9.9和9.11哪个大”这些简单的数学问题都出错，那么怎么去应用落地？“还是看应用场景，得在模型上去做迭代，引入新算法；或者之前有通过Zero-shot-CoT增强推理能力。”唐兴才表示。

AI算法专家、资深人工智能从业者黄颂则表示，LLM虽然数学不太好，但应用落地“不影响，用它所长”，数学一定是和计算相关的，这不是LLM使用的Transformer架构所擅长解决的问题，“现在比较好的解法是调用外部工具”。

于Tokenization的方法，错误地认为小数点后的数字具有不同的权重，导致它们认为11大于9。”

而通义实验室产品经理王晓明也表示，大模型是基于Transformer架构实现，本质是做next token prediction，而非直接进行算术计算，因此在处理比大小等简单数学题目时，依赖于预测模型的成功率。

此外，在处理类似“9.11比9.9大”的场景时，大模型通常会通过分词器（tokenizer）进行处理。分词器在解析这类表达时，可能会把数字辨认为日期或版本号进行比较，最终导致回答错误，这种处理方式是由分词器的特定算法和机制决定的。

如何让大模型数学变好

“问大模型关于简单的数字、字符串等问题，都是比较容易出现错误的。整个行业都在探索解决办法。”MiniMax方面表示。

根据第一财经的报道，新浪微研新技术研发负责人张俊林指出，早期LLM的Tokenizer一般不会对数字进行特殊处理，经常把连续的若干数字切在一起形成一个Token，比如“13579”，可能被切成3个Token，“13”是一个，“57”是一个，“9”是一个，哪些数字被切在一起组成Token，这取决于数据集合理的统计情况，在这种不确定的情况下，LLM要想做多位数字计算，是非常困难的。

那么，大模型如何撕掉算术差等生的标签？在思维能力上，更核心的可能还是训练语料的问题。大语言模型主要通过互联网上的文本数据进行训练，而这些数据中数学问题和解决方案相对较少，导致模型在数学推理和问题解决技能上的训练机会有限。

王晓明表示，提高大模型数学能力的核心在于提供高质量的数据支持，特别是在数学计算和逻辑推理方面。例如，通义千问针对这类场景，有针对性地加入高质量数据进行训练，使得面对此类问题时保持较高准确率。

此外，黄颂也表示，调用外部工具也是现在克服大模型数学差的比较好的解法。他曾以“从出生到现在一共活了多少天？”这个算术问题询问ChatGPT、Gemini、Claude 3 Sonnet、llama 3、文心一言、智谱GLM等10

个大模型，答对的4个模型几乎无一例外使用了外部工具帮助（python脚本）。

唐兴才还特别提到，有的论文在尝试引入新的算法解决大模型数据计算差的问题。例如来自复旦大学、上海人工智能实验室的5名研究者在今年6月发表了一篇技术报告，展示了一将大模型与蒙特卡洛搜索树（MCTS）算法相结合，从而使LLaMa-3 8B奥数水平拉到比肩GPT-4的高度。

该技术报告指出，为了解决LLMs在准确性和可靠性方面的挑战，特别是在策略和数学推理中，MCTS利用系统性探索和启发式自我精炼机制来改善LLMs中的决策框架。该算法通过迭代过程的选择、自我精炼、自我评估和反向传播构建蒙特卡洛搜索树，并使用改进的上置信界（UCB）公式来优化探索-利用平衡。

在被业界称作大模型应用落地元年的2024年，复杂推理能力关乎可靠性和准确性，是大模型在金融、工业等场景落地需要的关键能力。“这可能得在模型上去做迭代，引入新算法。”唐兴才表示。

同时，月之暗面也表示：“要彻底解决问题，又不能仅仅依赖于逐一修复每个案例，原因在于这些情况就像自动驾驶会遇到的场景一样是很难穷尽的，我们更加要做的是不断提升底层基础模型的智能水平，让大模型变得更加强大和全面，能够在各种复杂和极端情况下依然表现出色。”

Transformer架构的数学差综合征

AI大模型聪明得令人难以置信，但同时也会蠢得令人震惊。这种矛盾充分体现在“9.11和9.9哪个大”的回答上。

在回答该问题时，比如大模型全球第一梯队的ChatGPT-4o认为小数点后面的数字“11大于9”，因此9.11大。而智谱清言的回答虽然提到9.11的十分位是1，而9.9的十分位是9，但仍得出结论“9.11整体大于9.9”。

诸如此类“一本正经地胡说八道”的回答，还出现在月之暗面Kimi、字节豆包、商汤商量等大模型产品上。

“理解和推理没问题，展开和演算可能不太行。”黄颂表示，生成式的大语言模型自诞生以来做算

术就非常不靠谱，经常把一些简单的计算弄错。

大模型之所以在算术上犯错，有人指出是因为上下文语境不清楚，提示词（prompt）也不够明晰。“提问的方法也很重要，你可以试试：9.11和9.9这两个日期哪个大？”密度首席技术官刘益东表示，在这个提问下，大模型明确指出“在数学语境下，9.9是更大的数值”，所以“问题本身就可能会有歧义，语义缺失也可能带来误导”。

“不是所有的大模型都能做好数学理解题。”市场研究机构Omdia AI首席分析师苏廉节表示，大模型是按照被输入的问题做推测，这些被输入的问题就是大家熟悉的提示词，好的提示工

程师能调度大模型提供对的答案，像“9.11和9.9哪个大”这种问题的局限就是在于问题本身不是一个好的提示词，没办法让大模型完全掌握问题的语境。

事实上，如何让LLM给出“9.11和9.9哪个大”的正确答案，提示词很重要。比如把ChatGPT-4o的人设确定为数学家，或者先提问“哪个更大”再给出具体数字，LLM就会给出对的答案。

然而，大模型为什么会把“9.11和9.9哪个大”算错，这却是一个值得思考的问题。对于包括GPT-4o和Claude 3.5 Sonnet等大模型在该问题上出错，腾讯元宝给出的一个解释是：“Tokenization误解：一些AI模型在处理小数时，由