

2023年Q3 AIGC产业观察

Q3 2023 AIGC INDUSTRY OBSERVATION

创业邦研究中心

2023.11

目录

01

—

季度洞察

Quarterly Insight

02

—

政策变革

Policy Evolution

03

—

技术发展

Technological development

04

—

市场动向

Market trends

05

—

投融资分析

Investment analysis

06

—

总结

Summary

季度洞察

- ◆研究背景与主要观点
- ◆数据速览



研究背景与主要结论

AIGC产业历经上半年“百模大战”的军备竞赛，应用层生态的“百花齐放”，第三季度以来，站在了从“玩具”到“工具”快速迭代的关键时期，主战场大模型市场格局发生深刻变化，行业关注焦点也普遍转移到了困扰人工智能发展的“终极命题”——应用与商业化落地。与上一波深度学习的浪潮不同，本次AIGC研发范式的变革从根本上大幅提升数据生产速率，同时降低使用者和开发者的使用门槛，从需求侧推动“智慧飞轮”，有望对千行百业产生深远影响。

基于此，本报告主要围绕产品体验和价值创造两大生成式AI第二阶段的特征，全面扫描第三季度AIGC产业的政策、技术、市场、资本发展趋势。一方面帮助期待应用大模型技术的企业、期待投资大模型企业的投资者建立客观认知；另一方面，研判目前市面上大模型的产品化、生态构建等能力，洞察应用场景机遇，展望行业落地的潜在方向。

主要发现如下：

由于大模型带来的革命类似但有别于互联网，因此，传统的互联网思维不足以覆盖AIGC的发展路径。

大模型落地面临挑战与机遇，行业共同关注放大模型能力的有效途径。当大模型参数量普遍达到一定水平，在边际效用递减的规律下，将有更多的方式来提升模型的能力和效用，比如微调、提示工程、搜索增强生成、AI Agent等技术手段。AI Agent是挖掘大模型潜力的关键环节，但不是终点，也不可独立运作。开源模型迅猛发展，产品向终端延伸，结合更多AI应用技术，有利于推动应用场景多元化发展。

由于政策面向C端设置准入门槛，标准体系覆盖多个行业，涵盖模型行业能力、模型工程化性能、模型算力网络、模型安全可靠等方面，数据、算法、模型、安全因素重要性凸显，意味着“百模大战”回归理性，行业格局迈入整合阶段。

由于整体经济环境仍处于下行期，行业示范效应尚不明显，大模型应用落地路径处于探索阶段，投资相对谨慎。

进入第三季度以来，国内AIGC创投市场CVC融资事件占比显著提升，下一阶段，国资有望发挥更大作用。

Q3中国AIGC行业融资数据速览

<div>融资</div>	<div>35 起</div> <div>融资事件</div> <div>39.61亿 人民币</div> <div>融资金额</div>	<p>2023年Q3国内AIGC行业发生融资事件35起，已披露融资金额39.61亿人民币，涉及公司33家，涉及投资机构数量51家。</p> <p>已披露融资规模以1000万~3000万区间居多，发生9起；</p> <p>早期事件（种子轮~天使轮）数21，占比63.64%；</p> <p>北京仍为最活跃地区，融资事件为11起，以模型层居多，上海超越北京成为国内AIGC融资最吸金的地区，总融资金额22.08亿元。</p> <p>2023年Q3国内AIGC行业发生融资事件中，通用大模型（6起）、工具平台（6起）两个细分赛道相对活跃。在应用层中，元宇宙/数字人（5起）和营销（5起）是融资事件最频繁的细分领域。</p>
<div>IPO</div>	<div>1 家</div> <div>AIGC企业完成上市</div>	<p>2023年Q3有1家国内AIGC企业完成上市——第四范式（决策类人工智能公司）。</p>
<div>并购</div>	<div>1 起</div> <div>AIGC并购事件</div>	<p>2023年Q3国内AIGC行业发生1起并购事件——美团收购光年之外，融资额20.65亿元。</p>

政策变革

- ◆政策准入
- ◆政策支持



国产大模型面向C端建立准入机制，应用发展迎来新拐点（1/2）

- 政策方面，8月15日，我国首份生成式人工智能监管性文件正式实施；8月31日，首批通过备案的人工智能大模型名单出炉，意味着其生成式AI产品可以正式面向公众开放注册、提供服务，此前的大模型产品多采用邀请注册制，仅开放给企业合作伙伴使用。
- 面向C端的对话式应用产品吹响了号角，带动B端产业大模型陆续落地变现，“百模大战”的风已经吹向了“百模应用之争”。

大模型企业面向C端的应用产品

注：按企业首字母排序

企业/单位名称	大模型名称	应用产品名称	产品定位	应用进展
阿里巴巴	通义千问	通义听悟、通义灵码、通义星尘等	大模型AI助手	6月1日，阿里云宣布通义大模型进展，聚焦音视频内容的AI新品“通义听悟”正式上线，听悟融合了十多项AI功能，可以全面提升知识从音视频向图文形态转化的效率。2023年10月31日云栖大会上，正式对外发布“通义灵码”。
百川智能	Baichuan大模型	Baichuan系列大模型产品	大模型	截止8月31日，Baichuan-7B/13B两款开源模型在Huggingface下载量已突破三百万。
百度	文心一言	文心一言	“你的智能助理”	8月31日登顶苹果APP Store免费应用下载榜首。
抖音	云雀大模型	豆包	AI机器人	“豆包”提供聊天机器人、写作助手以及英语学习助手等功能，可以回答各种问题并进行对话，帮助人们获取信息。
华为	盘古大模型	大模型在矿山、政务、汽车、气象、医学、数字人、研发等领域的创新服务	大模型产业化应用	盘古大模型3.0分为L0基础大模型、L1行业大模型、L2场景模型三层架构，将重塑千行百业。

国产大模型面向C端建立准入机制，应用发展迎来新拐点（2/2）

大模型企业面向C端的应用产品

注：按企业首字母排序

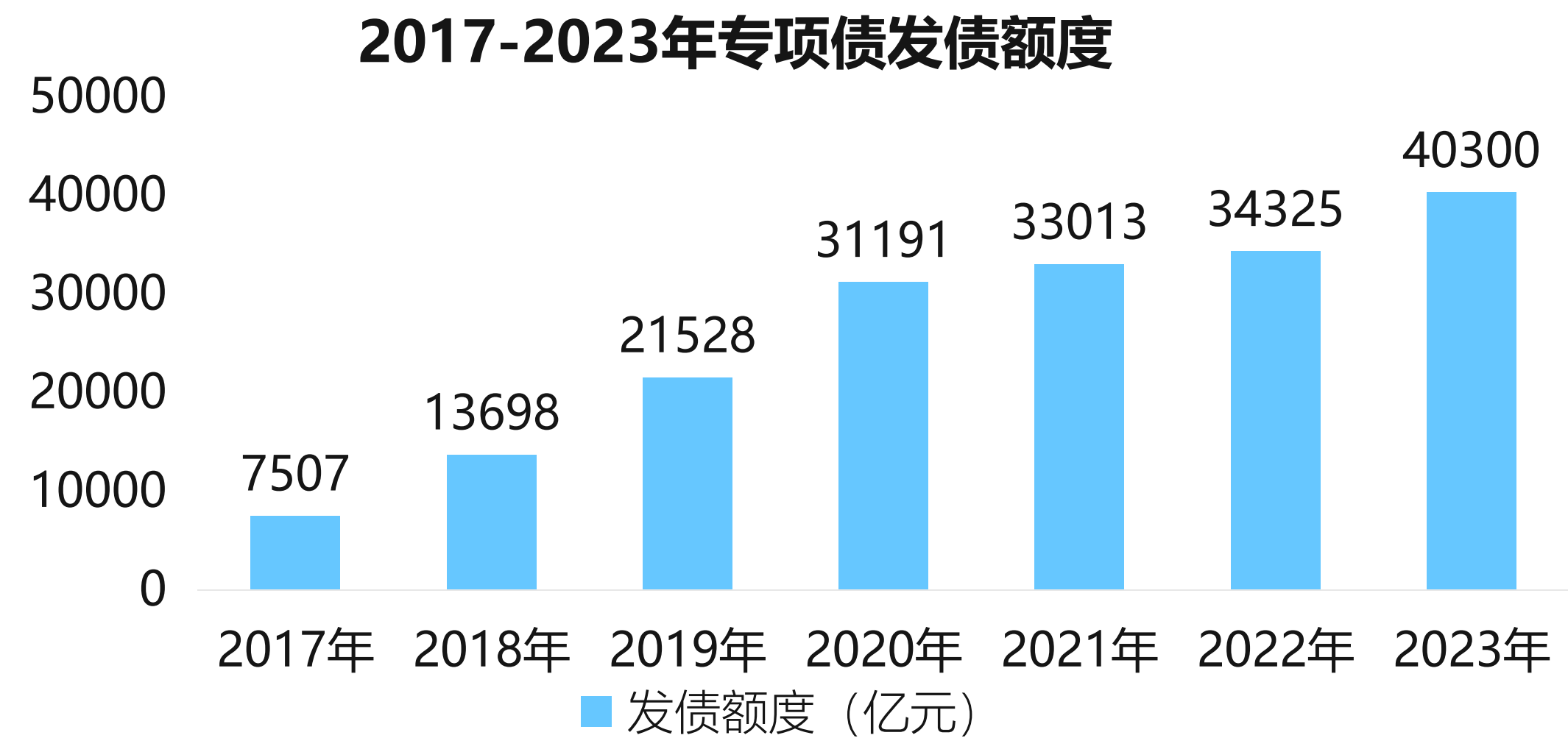
企业名称	大模型名称	应用产品名称	应用产品定位	应用进展
科大讯飞	讯飞星火认知大模型	智能编程助手 iFlyCode 1.0、星火教师助手等	大模型在垂直领域的应用产品	于9月5日凌晨面向全民开放，随即“讯飞星火”APP登录各大应用市场，用户可以直接注册使用。
MiniMax	ABAB大模型	MiniMax开放平台	面向B端用户的大语言模型技术平台	MiniMax开放平台已接入超百家付费客户，提供办公协作、互动娱乐、客服、搜索、教育等十余个行业场景。
商汤科技	日日新SenseNova大模型	商量SenseChat等	AI聊天机器人	于8月31日正式开放，用户通过官网注册使用，可体验文本编辑、数理计算、编程等功能。
上海人工智能实验室	书生通用大模型	书生通用大模型	涵括视觉、语言和三维等在内的大模型体系	在WAIC 2023上，上海AI实验室联合香港中文大学、上海市测绘院发布全球首个城市级NeRF实景三维大模型书生·天际。
智谱AI	ChatGLM	智谱清言	生成式AI助手	截止10月，智谱AI的开源模型全球下载量超过一千万。
中科院	紫东太初	紫东太初大模型开放平台	跨模态通用人工智能平台	自9月1日面向公众开放服务，已在文本创作、医疗、物流、政务等多个行业实现应用落地。

中央专项债支持，地方引导基金为AI产业发展注入活水

创业邦睿兽分析的数据显示，中国AIGC投融资事件占人工智能投融资事件总数的比重、中国AIGC金额总数占人工智能投融资获投金额的比重呈上升趋势，尤其是 2023年H1尤为显著，2023年上半年 AIGC事件数占AI赛道的比重为 23.9%，远高于前几年。但就上半年数据，AIGC行业融资事件和融资金额涨幅不明显，主要与经济下行的大环境有关。

因此，为增强民间资本投资信心，2023年专项债投向新增两大领域，分别是新能源和新基建，其中新基建包括数据中心、人工智能基础设施。**今年以来，新型基础设施建设步伐加快，上半年新基建投资同比增长16.2%。步入下半年，广西、贵州、河南、福建等多地就推进新基建作出新的部署，进一步释放出新基建提速的信号。**

Q3人工智能地方引导基金除上海相对活跃外，其他还未充分释放。



来源：公开数据

地方政府AI引导基金进展情况

地区	Q3产业基金动向
北京	<ul style="list-style-type: none">9月18日，北京市科委、中关村管委会网站公布《投资引导基金关于公开遴选北京市人工智能产业投资基金管理机构的公告》。北京市科学技术委员会、中关村科技园区管理委员会、引导基金现公开遴选基金管理机构。基金目标总规模为100亿元，其中，引导基金出资不超过30亿元。基金可分两期实施，首期目标规模40亿元。基金围绕北京市在人工智能领域的总体布局开展直接股权投资，重点投向人工智能芯片、训练数据及相关软件等底层技术领域，大模型算法创新、具身智能、可信AI等关键领域，以及大模型等人工智能技术产品开发和垂直行业创新应用等相关领域。9月22日举办的《北京市促进未来产业创新发展实施方案》新闻发布会上，门头沟区重点发展人工智能产业，设立10亿元的人工智能产业引导基金。
上海	<ul style="list-style-type: none">7月8日，2023世界人工智能大会上，上海市经济和信息化委员会副主任张英介绍了上海即将出台的大模型创新发展政策要点，将围绕“创新能力、创新要素、创新应用、创新环境”4大方向，重点打造“3项计划+5大平台”。签约的32个全新人工智能重大产业项目中，涵盖了算力、数据、大模型、硬件、软件等不同领域。项目投资总额288亿元。当天，“上海人工智能产业投资基金元宇宙智能终端子基金暨上海人工智能开源生态产业集群”正式发起成立，以助力加快打造人工智能“上海高地”。
成都	<ul style="list-style-type: none">8月，成都高新策源投资集团有限公司拟发起成都高新区人工智能产业基金的遴选，总规模不超过50亿元。

来源：创业邦研究中心整理

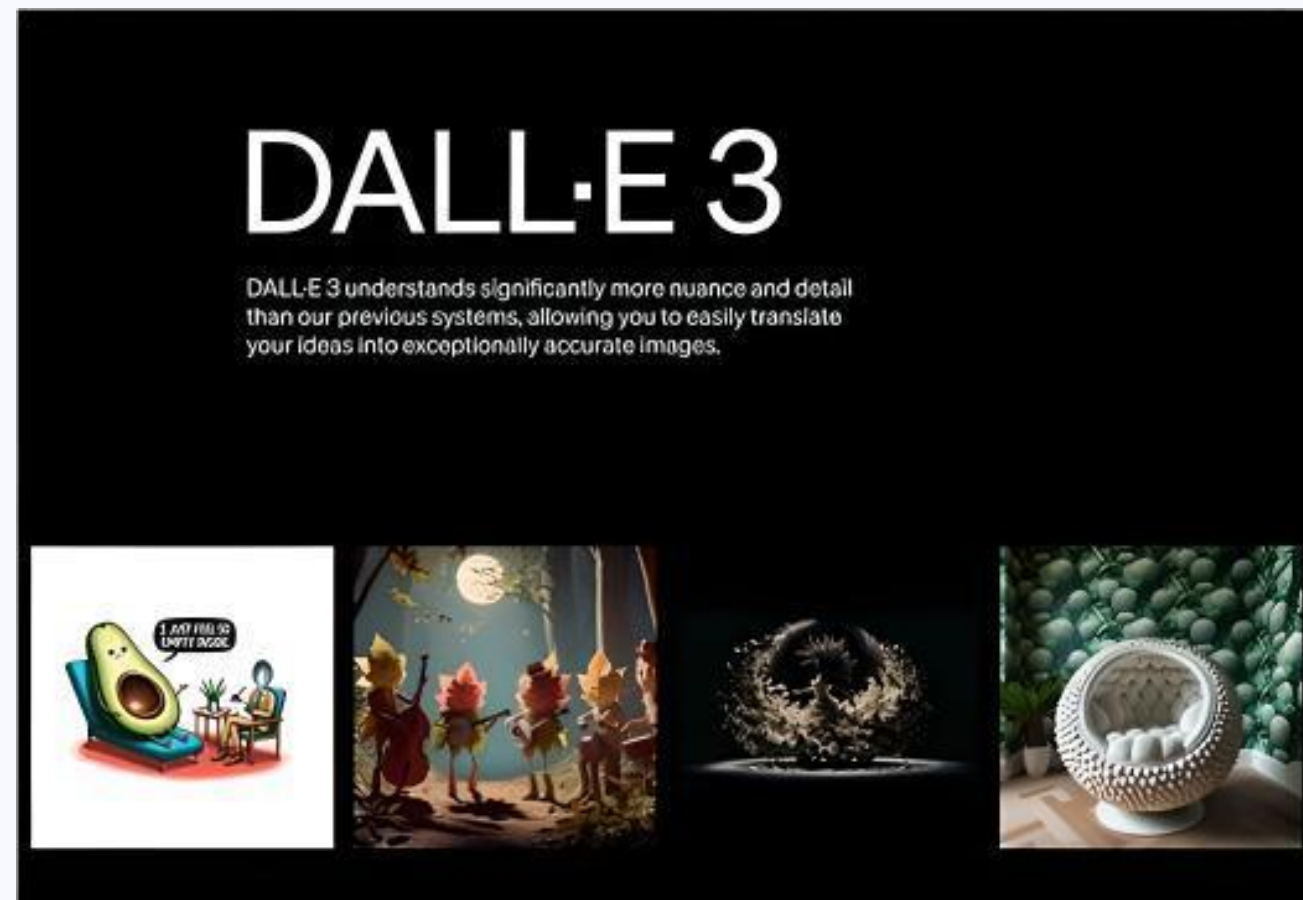
技术发展

- ◆ 技术迭代
- ◆ 技术趋势



技术迭代一：多模态大模型DALL·E 3带来产业冲击

背景：多模态大模型（MLLM）是将文本、图像、音频和视频等多模信息结合起来训练的模型。相比LLM（大语言模型），MLLM更符合人类感知世界的方式。多模态输入的支持，使用户可以用更灵活的方式与智能助手进行交互，继而利用强大的大模型作为大脑来执行多模态任务。



9月21日，OpenAI宣布生成式AI艺术平台DALL-E的第三个版本——DALL·E 3已经进入研究预览阶段，并将于10月初面向ChatGPT Plus和企业用户上线。另外ChatGPT-4V的推出为其增加了视觉智能，使其不再局限于文字处理，还能识别、理解、分析和预测图片信息。

影响：DALL·E 3的出现不仅再次冲击艺术界的绘画产业，也同时给电商、设计等行业产生了实质性的影响。但同时，安全和版权保护等社会问题也尤为值得关注。

对标产品：昆仑万维天工大模型Skywork-MM、MiniMax多模态大模型矩阵等。

技术迭代：

- 1) DALL·E 3能更好地捕捉细微差异的语义描述，实现了提示词的完美遵循，还能高效避免混淆详细请求中的元素，在画面呈现方面有了明显进步。
- 2) 文生图模型与 ChatGPT 的结合，极大地弱化了提示工程的约束。



技术迭代二：长文本技术增强产品用户体验

背景：在LLM中，“上下文长度”是指大语言模型在生成预测时考虑的输入文本的长度。对大模型而言，具备更长文本的建模能力意味着模型可以观察到更长的上下文，可以避免因观察窗口限制导致重要信息的丢失。大模型应用效果通常取决于两个核心指标，一是模型参数量，决定了大模型的“计算”能力；二是上下文长度，决定了大模型的“内存”能力，即长文本可以通过提供更多上下文和细节信息，来辅助模型判断语义，进一步减少歧义，并且基于所提供事实基础上的归纳、推理也更加准确。

7月18号，斯坦福大学发布了新型注意力算法FlashAttention-2，更好的并行化和工作分区，比标准Attention提速5-9倍，极大加速了现有模型的训练和推理。9月21号，香港中文大学联合MIT推出微调方法LongLoRA。两项研究成果能在资源受限的情况下，极大扩展大模型的上下文。

影响：在诸如金融、法律和科研等特定行业或场景，需要对长篇幅文档进行分析、归纳抽取、多文档信息对比以及关键信息总结。因此，长文本建模能力是大模型目前行业应用能顺利落地的前提条件。

对标产品：月之暗面Kimi Chat等。

国内外关注度较高的模型上下文可接受长度表		
公司/机构/团队	模型/产品名称	上下文Tokens
OpenAI	GPT-3.5 GPT-4	4K-16K 8K-32K
Anthropic	Claude Claude2	100K 100K
Meta	LLaMA LLaMA2 Llama 2 Long	2k 4k 32K
IDEAS NCBR、Google DeepMind等	Long LLaMA	256k
Moonshot(月之暗面)	Kimi Chat	400K
港中文贾佳亚团队、MIT	LongAlpaca	32K-100K

技术展望：

国内外对于文本长度的探索还远没有达到“临界点”，在通往未来Agent和AI原生应用的道路上，长文本依然扮演着重要的角色。Agent任务运行需要依靠历史信息进行新的规划和决策，AI原生应用需要依靠上下文本来保持连贯、个性化的用户体验，这也是月之暗面、OpenAI等一众大模型公司在当下聚焦长文本技术的原因所在。

技术迭代三： Llama2掀起大模型市场新格局

背景： LLaMA(LargeLanguage Model MetaAI) 是Meta发布的一款开源大型语言模型。该模型仅使用公开数据集进行训练，确保了其与开源协议的兼容性和可复现性， LLaMA已成为AI社区中最受欢迎的开源模型之一。然而， 由于其开源协议的限制， LLaMA仅限于学术研究使用， 不能进行免费商业用途。

Llama 2 was trained on **40% more data** than Llama 1, and has double the context length.

Llama 2		
MODEL SIZE (PARAMETERS)	PRETRAINED	FINE-TUNED FOR CHAT USE CASES
7B	Model architecture:	Data collection for helpfulness and safety:
13B	Pretraining Tokens: 2 Trillion	Supervised fine-tuning: Over 100,000
70B	Context Length: 4096	Human Preferences: Over 1,000,000

7月19日， Meta在官网发布了开源大语言模型—Llama 2。


影响： 对上游， Meta与微软云服务Azure合作， 向全球开发者首发基于Llama 2模型的云服务， 与高通合作， 打破市场上英伟达、AMD处理器对AI产业的垄断。对下游， 通过“模型基座+迁移学习+微调”的垂类AI模型开发范式， 使得更多的企业和个人开发者可以快速加入到AIGC热潮中， 也极大地加速了行业AI应用开发效率， 弱化和颠覆OpenAI等公司在新兴的生成式人工智能软件市场中建立的早期主导地位。Llama2推动了国内大模型的开源进程和商业化变革。

对标产品： 百川智能部分模型、智谱AI GLM-130B开源双语预训练语言模型等。

- 技术迭代：**
- (1) 相比于 Llama 1， Llama 2的预训练语料库增加40%， 提升到2万亿Tokens;
 - (2) 9月Llama2 long已达32,768个token;
 - (3) 采用了分组查询注意力机制， 对文本语义的理解更强;
 - (4) Llama 2 70B 在MMLU和GSM8K上接近 GPT-3.5。

技术迭代四：AI Agent深入挖掘大模型潜力

背景：Agent（智能体）指在一定的环境中体现出自治性、反应性、社会性、预动性、思辨性（慎思性）、认知性等一种或多种智能特征的软件或硬件实体。Agent = 大模型+记忆+主动规划+工具使用。AI Agent能够理解、规划、执行、自我调整，解决更复杂的问题。相比LLM，AI Agent能独立思考、调用工具去逐步完成给定目标的能力；和RPA的区别在于能够处理未知环境信息。



The image shows a presentation slide for Amazon Bedrock, titled "Choice of foundation models". It lists several providers and their models: AI21labs (JURASSIC-2), cohere (COMMAND + EMBED), stability.ai (STABLE DIFFUSION XL 1.0), ANTHROPIC (CLAUDE 2), and amazon (AMAZON TITAN). Each model is accompanied by a brief description of its capabilities. A person is standing on the stage in front of the slide.

7月26日，亚马逊推出Amazon Bedrock Agents，可以自动分解企业AI应用开发任务；8月初，AI独角兽Inflection在开发私人AI助理等（另，10月底GPT-4重磅更新，推出了整合了画图、插件、代码等所有工具的All Tools功能。）

影响：研发侧，由于AI Agent需要主动感知感知环境信息，多模态大模型成为新的行业热点之一，同时Agent能自主调用工具，其改变了以往的软件研发方式和生态应用方式；应用侧，为企业和个人提供更加个性化、可定制的AI伙伴。

对标产品：联汇科技OmBot欧姆智能体、实在智能TARS-RPA-Agent等。

AI Agent与其他技术选型方案发展及优劣势比较

选型	优势	劣势
Fine-tuning 大模型微调	使用特定的下游任务知识对基础模型进行指令微调，改变参数权重以适应领域，增加大模型的领域知识理解力和领域专业能力。	高质量的指令微调语料构建难度大；微调资源消耗巨大，且难支持频繁更新；微调效果难符合预期。
Prompt Engineering 提示工程	给通用大模型传递指令、示例、背景信息，控制大模型的输出，可定制化生成答案，且一定程度上提升LLM逻辑推理能力。	给通用大模型传递指令、示例、背景信息，控制大模型的输出，可定制化生成答案，且一定程度上提升LLM逻辑推理能力。
RAG 搜索增强生成	结合传统搜索框架、构建领域知识库检索，利用LLM对提问的理解能力和对答案的二次加工能力（概括、分析、推理），输出可控结果。	对知识解析、知识存储、搜索匹配的技术和策略要求高。
Agent 智能体	针对API、插件等工具使用对基础模型进行指令微调，使LLM能根据提问自主思考决策调用工具，增加LLM对垂直领域任务和工具的理解。	高质量的工具指令微调数据构建难度大；微调资源消耗巨大；大模型决策的可控性和准确性低，生成内容有合规风险。

技术趋势一：拥抱开源精神，国产模型的崛起已成燎原之势

- 国产模型已成为大语言模型阵营中的中坚力量。虽然国产模型起步较晚，并且在GPU高端芯片中遭受国外穷追围堵，但在国家大力扶持和头部厂商的推动下，国产模型的崛起已成燎原之势。
- 同时，在智源研究院等诸多顶尖的人工智能科研院所的带领下，国内部分头部大模型创业企业、有云端业务基础的互联网大厂积极推动开源生态体系的构建。

国内AI大模型发展进程（截至2023年Q3）



国内AI大模型发展进程，追溯了2023年Q1~Q3以来语言模型的发展。开源模型由蓝色边框示意，其余为闭源模型。

由图可知，2~3月大模型市场整体还处于萌芽阶段，4~7月是国产大模型集中爆发的阶段。7月，应用层医疗（京医千询）、教育（子曰）、旅游（问道）等产业级大模型以及赤兔等企业服务大模型密集发布。8月~9月大模型发布数量有所减少，但在产品功能上呈现出差异化特征，比如，MiniMax ABAB 大模型擅长数据增强和数据压缩，云雀大模型擅长多媒体内容生成和内容理解等。通过拓展纵深，不断挖掘大模型应用潜力，渗透千行百业。

来源：创业邦整理

技术趋势二：大模型产品向终端延伸，推动应用场景多元化发展

- 大模型开源、多模态、Agent等技术趋势下将产生全新的、个性化的、人性化的人机交互体验。
- 未来大模型部署在手机、PC、汽车、人形机器人等终端，能够缓解云端AI在成本、能耗、性能、隐私、安全和个性化等方面存在的问题，也将拓宽自动驾驶、智慧教育、智慧家居等场景的多元化应用，但如何在端侧轻量部署，实现软硬件深度融合仍是难点问题。

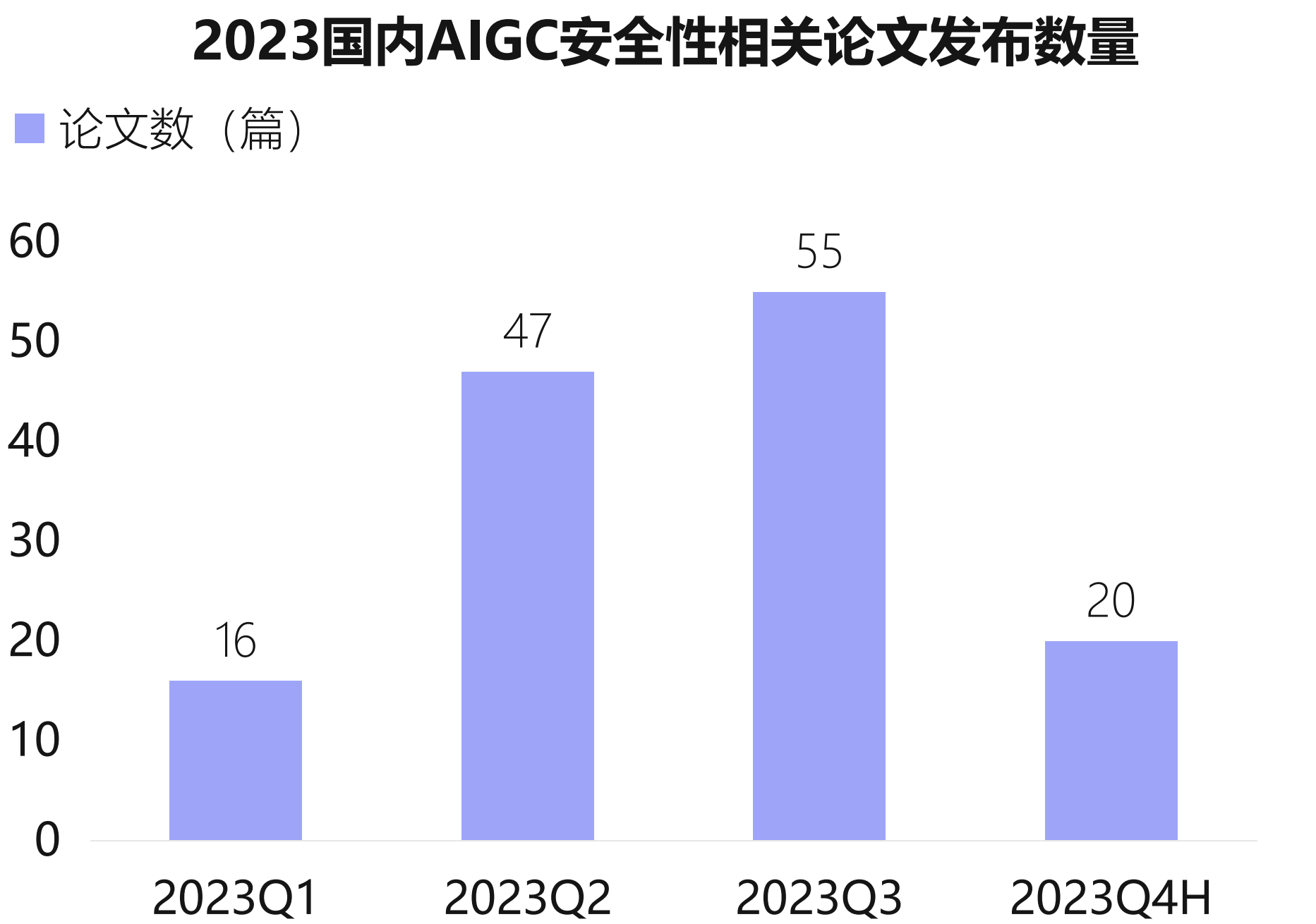
企业名称	AIGC终端厂商Q3动态
小米	8月，小米官宣“轻量化、本地部署是小米大模型技术主力突破方向”。小米自研的13亿参数端侧模型已经在手机端跑通，而且部分场景效果可以媲美60亿模型在云端的运算效果。另外，在数据层面，小米自己挖掘整理的训练数据占比达到了80%，其中小米自有的产品和业务数据量达到3TB。
三星	9月4日，三星电子家用电器部门副总裁Miyoung Yu在IFA2023展会上宣布,三星电子明年将在所有家用电器中引入基于Bixby的生成式AI功能。
华为	8月4日，华为开发者大会2023在中国·松山湖举行，宣布随着AI大模型技术的不断成熟，HarmonyOS 4的小艺智慧能力取得了跨越式的成长，将具备AI大模型能力。华为小艺融合了盘古自然语言大模型、盘古视觉大模型和盘古多模态大模型，最高版本高达1000亿参数，于今年9月开启众测，首批支持机型为Mate 60系列手机。
VIVO	8月，vivo的自研大规模预训练语言模型“vivo_Agent_LM_7B”现身C-Eval。
高通	9月，高通已在终端侧实现支持超过10亿参数的生成式AI模型，比如Stable Diffusion，并计划未来在终端侧支持参数高达数百亿的模型。10月24日，在骁龙峰会期间，高通宣布推出全新旗舰移动平台——第三代骁龙®8，它是一款集终端侧智能、顶级性能和能效于一体的强大产品。而且这个芯片还是首个专为生成式AI而精心打造的移动平台。

资料来源：创业邦研究中心整理

技术趋势三：安全性成为国内关注的焦点问题

由于政策的准入要求，以及政企客户落地应用的必然要求，安全性成为国内关注的焦点问题。业内对安全性的担忧主要集中在以下方面：1.合法合规；2.数据安全；3.算法安全：具有不可控性，容易出现幻觉；难以解释，参数规模大。安全问题的明确和解决有利于降低大模型风险。

区域	AIGC安全性议题Q3动态	Q3技术迭代
国内	国内行业发出可信AI的倡议。在世界人工智能大会期间（7月6日-8日），中国信通院、上海人工智能实验室、武汉大学、蚂蚁集团等多家单位共同发起的《AIGC可信倡议》。	7月6日-8日，清华大学与蚂蚁集团联合发布AI安全检测平台“蚁鉴2.0”，运用智能博弈对抗技术； 9月7日，在2023腾讯全球数字生态大会上腾讯混元大语言模型正式亮相，并宣布通过腾讯云对外开放。不依赖外挂，在预训练阶段通过“探真”算法进行事实修正，让混元大模型的幻觉相比主流开源大模型降低了30%至50%。
国外	7月21日，OpenAI、微软、谷歌等美国头部AI公司向社会做出公开承诺以负责任的方式发展AI。	9月，DeepMind发布全球首个数字水印SynthID，提升AI图片生成安全。



数据来源：以“AIGC”、“安全”为检索词在知网搜索的国内论文结果。从今年年初ChatGPT爆红以来，国内相关论文发布数量日益提升，在Q3达到最高点，可见AIGC安全性已成为学术关注的焦点问题。

资料来源：创业邦研究中心整理

技术趋势四：企业私有化部署大模型综合成本持续降低

大模型的应用落地成本，主要包含数据成本、模型成本和应用开发成本三个维度，其中模型成本包含授权成本、算力成本等。Q3阶段，随着Llama2带动了国内模型的商用免费化，且MaaS作为一种新型商业模式逐渐被市场接受，授权成本过高的壁垒正在消失。尤其通过QLoRA微调和GPTQ量化，中小企业也可以上手千亿级模型，很大程度上降低了模型的算力成本。企业私有化部署综合成本持续降低有利于增加大模型对B端市场的渗透。

大模型应用落地成本构成：



大模型应用落地成本预估：

API调用（小微企业）

综合成本：10万以内

方案：价格低廉，通过直接调用或调教第三方API，同时采用Prompt提示词做查询优化。

群体：想快速低成本搭建产品的小微企业

高效微调（中小企业）

综合成本：10~100万

方案：私有化部署，有助于提高服务安全性，一般采用高质量训练集，配合LangChain等技术栈。

群体：注重数据隐私和垂类数据精度的中小企业

预训练（中大型企业）

综合成本：100万起步

方案：一般通过RLHF训练策略进行垂直化训练，增强模型泛化能力。

群体：想构建品牌力的中大型企业

附：主流厂商第三方API接口成本

厂商	模型	参数规模（B）	描述	上下文长度（K）	价格（¥ / 1K tokens）		免费额度	有效期
					输入	输出		
OpenAI	GPT-4	万亿级	广泛的通用知识和领域专业知识，可以用自然语言遵循复杂的指令，准确地解决难题	8 ~ 32	0.21 ~ 0.43	0.43 ~ 0.87	-	-
	GPT 3.5	175	gpt-3.5-turbo是该系列的旗舰，并对dialog进行了优化	4 ~ 16	0.01 ~ 0.02	0.01 ~ 0.03	5美元	3个月
智谱AI	ChatGLM	130	按量计费，简单灵活，提供满足不同业务规模的产品方案	8 ~ 32	0.002~0.01		18人民币	-
科大讯飞	讯飞星火v3	-	套餐制，拥有跨领域的知识和语言理解能力，完成问答对话和文学创作等任务	32	0.024 ~ 0.03		20万Tokens	一年
百度	文心一言 ernie-bot	-	包含ernie-bot和ernie-bot-turbo	-	0.008 ~ 0.12		-	-
阿里巴巴	通义千问	千亿级	DashScope灵积模型服务建立在MaaS的理念下，围绕AI各领域模型，通过标准化的API提供多种模型服务	-	0.012 ~ 0.14		30万Tokens	180天
百川智能	Baichuan	53	融合了意图理解、信息检索以及强化学习技术，结合有监督微调与人类意图对齐，在知识问答、文本创作领域表现突出	-	0.01 ~ 0.02		-	-
腾讯	混元	千亿级	分高级和标准两种方案	-	0.01 ~ 0.1		-	-

* GPT接口按2023年11月14日汇率转换，除OpenAI外，其他API接口Token收费均已包含输入和输出





市场动向

- ◆ 市场潜力
- ◆ 市场格局
- ◆ 细分领域分析
- ◆ 典型案例

国内AI应用市场“金矿”待挖掘，龙头企业落地成果初现

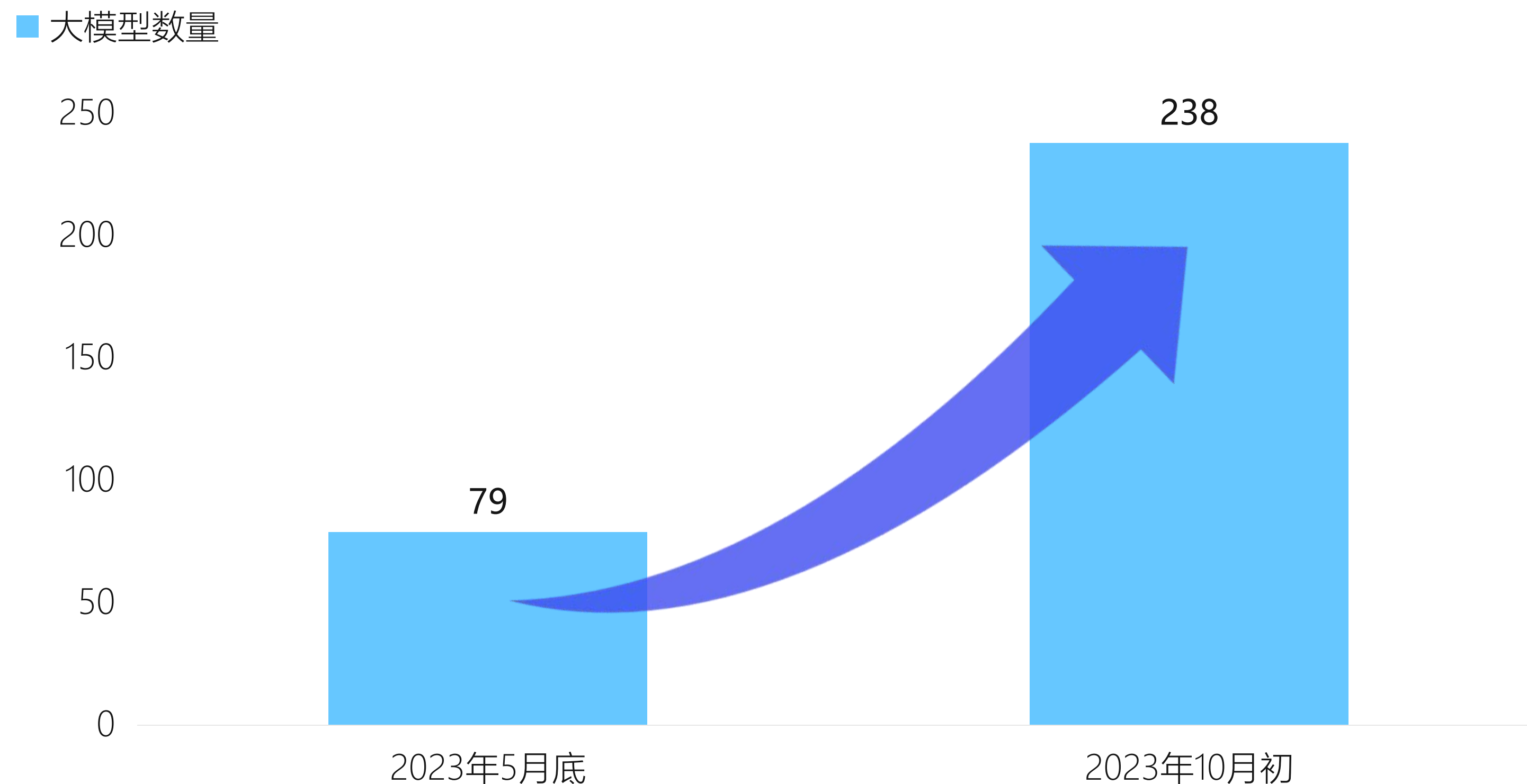
- **AIGC在企业的渗透率低，AIGC市场应用期望值高。**2023年5月，据市场调查机构 Gartner调查数据显示，在中国有6%的用户已经成功部署了与生成式AI相关的技术，有26%的用户正在积极试点生成式AI技术，还有4%左右的中国用户表示将在未来0-6个月内部署生成式AI相关的应用或技术。到2026年，超过80%的企业将使用生成式AI API，或部署生成式AI的应用程序，而2023年这一比例还不到5%。医疗保健、生命科学、法律、金融服务等部门，对生成式AI的需求正在不断增加。
- **因此，国内龙头企业亟需打造标杆案例，探路大模型边界。**Q3阶段，阿里巴巴妙鸭相机在C端图像生成场景，智谱AI在B端的制造业、金融、医疗、政务、教育等行业场景等等，国内龙头企业有一定落地案例初现。但总体看来，国内AIGC产业在落地应用和商业化发展方面仍处于探索阶段。

企业名称	标杆案例和产品特色	应用成果
	<ul style="list-style-type: none">• 妙鸭相机，一款基于AI人脸识别的美图应用。通过AI学习消费者上传的照片来构建人脸模型，然后把人脸模型套用其他模型合成照片，生成无数风格的写真。	妙鸭相机6月30日邀请制内测、7月17日正式上线，第三方数据平台显示，截至8月16日，妙鸭相机在苹果应用商店的下载量已经超过了533万次。
	<ul style="list-style-type: none">• 2023年6月，发布全面升级的ChatGLM2模型矩阵，多样尺寸，丰富场景，模型能力登顶C-Eval榜单。• 2023年8月，正式上线首款生成式AI助手—智谱清言。	为了满足不同应用场景的需求，智谱AI提供了三款 ChatGLM API服务，对标 OpenAI全模型产品线。

大模型数量激增，市场热情高涨

- 大模型是AIGC行业主战场，在Q3阶段，市场热情高涨。据2023年5月底中信所发布的《中国人工智能大模型地图研究报告》，目前中国10亿参数规模以上的大模型已发布79个，主要集中在北京和广东，其中北京38个大模型，广东20个大模型。中国大模型中，已有半数以上大模型开源，高校/科研机构是开源主力。根据北京市经信局公布的数据，截至2023年10月初，国内公开的AI大模型数量为238个，近五成集中在北京。
- 这说明：1.在政策鼓励和外界对AIGC行业市场前景的普遍认同下，企业踊跃参与大模型布局；2.大模型存在同质化问题，主要由于基座模型入局门槛过高，导致大部分中小企业基于开源的基座模型在垂直行业二次开发或者作为工具平台推动产业链的完善；3.市场集中度较低，产品竞争较为激烈。

2023年5月底~10月初国内10亿参数以上大模型数量变化



国产芯片与大模型适配能力日益提升，向量数据库降低用户门槛

算力：国产芯片与大模型适配能力日益提升

2023年7月6-8日，2023世界人工智能大会上，各类人工智能芯片公司、算力提供商针对大模型展示了相应的方案。

企业名称	AIGC算力企业Q3动态
瀚博半导体	发布了第二代GPU SG100，并推出南禺系列GPU加速卡VG1600、VG1800、VG14，以及LLM大模型AI加速卡VA1L、AIGC大模型一体机、VA12高性能生成式AI加速卡等6款新品。
燧原科技	发布了燧原曜图文生图MaaS平台服务产品。该产品以燧原科技“邃思”系列芯片为算力支撑，由首都在线提供计算服务，燧原曜图MaaS平台服务产品为用户提供面向AIGC时代的高效易用、安全可靠、企业级的文生图服务。
登临科技	展示了最新一代创新通用GPU产品Goldwasser II系列以及基于开源大语言模型可交互界面。
天数智芯	展示了在大模型训练、推理所取得的显著进展，包括图片识别/以图搜图、3D建模、大模型推理等。

数据：向量数据库进一步降低用户使用门槛

定义：向量数据库是专门用来存储和查询向量的数据库，其存储的向量来自于对文本、语音、图像、视频等的向量化。与传统数据库相比，向量数据库可以处理更多非结构化数据。在机器学习和深度学习中，数据通常以向量形式表示。

优势：基于大模型和向量数据库，可以整合企业数据进行企业知识的向量化并实现高效的知识搜索，并通过持续学习和用户反馈，为企业构建一个智能的知识库。

国内代表企业：



资料来源：创业邦研究中心整理

头部企业动态：模型能力升级，产品功能加码

企业名称	AIGC头部企业Q3动态
百度	9月5日，2023百度云智大会宣布，千帆大模型平台上月活企业数已近万家，覆盖金融、制造、能源、政务、交通等行业的400多个场景。
腾讯	9月7日，2023腾讯全球数字生态大会上，腾讯混元大模型正式亮相。这是一款由腾讯公司自主研发的通用大语言模型，其模型参数达到千亿级别。
阿里巴巴	7月6-8日，2023年世界人工智能大会上，阿里云正式推出AI绘画新品“通义万相”。 9月25日，阿里云开源通义千问140亿参数模型Qwen-14B及其对话模型Qwen-14B-Chat，免费可商用。
华为	7月7日，华为开发者大会2023（Cloud）上，华为常务董事、华为云CEO张平安正式发布华为云盘古大模型3.0。
字节跳动	8月，字节跳动AI产品Grace新版本更名为“豆包”并启动测试，网页版、移动端均已上线，支持手机号、抖音账号、Apple ID登录。目前拥有文生文的功能。
科大讯飞	8月15日，科大讯飞正式发布了讯飞星火大模型V2.0。发布代码能力和多模态能力升级，同时发布并升级搭载讯飞星火认知大模型V2.0能力的多项应用和产品。
智谱AI	8月31日，生成式AI工具“智谱清言”正式上线。 9月，增持聆心智能股份，双方在股权层面上达成深度合作，业务上强强联合。
百川智能	7月11日，王小川旗下百川智能今日发布Baichuan-13B大模型，号称“130亿参数开源可商用”。 9月25日，百川智能发布Baichuan2-53B，开放API全面进军To B领域。
MiniMax	7月6-8日，2023年世界人工智能大会上，MiniMax表示，在把模型能力做到企业应用的地步、输出安全性和服务的稳定性三方面做了大量的工作。目前自研模型以每周更新的速度进行迭代，这对整套大模型系统有着非常严格的算法要求和工程标准。为了提供更加稳定可靠的模型服务能力，MiniMax推出MaaS互信推理专区方案，保证了企业数据隐私及模型的安全和互信。

资料来源：创业邦研究中心整理

大模型创业标杆案例：智谱AI-今年累计融资超25亿元

- **成立时间**：2019-06-11，北京
- **融资轮次**：B++轮，社保基金中关村自主创新基金（君联资本为基金管理人）、美团、蚂蚁、阿里、腾讯等
- **行业定位**：致力于打造新一代认知智能大模型，专注于做大模型的中国创新。
- **主营业务**：公司于2020年底开始研发GLM预训练架构，2021年训练完成百亿参数模型GLM-10B，同年利用MoE架构成功训练出收敛的万亿稀疏模型，2022年合作研发了中英双语千亿级超大规模预训练模型GLM-130B，于2023年3月14号推出ChatGLM对话模型并开源ChatGLM-6B模型，并基于GLM系列模型打造大模型平台及AIGC产品矩阵。
- **服务对象**：公司践行Model as a Service (MaaS) 的市场理念，目前合作伙伴涵盖了不同的行业领域，包括制造业、金融、医疗、政务、教育等。



智谱AI观点：大模型的原生应用时代正在到来，改变原有的软件开发方式和生态应用方式

【过去挑战】 预训练大模型是AIGC时代的基座，在训练千亿大模型时有三个挑战：第一，训练成本高昂；第二，人力投入极大；第三，训练过程不稳定。大模型落地应用的主要挑战来自：外界普遍对AI 助手技术的认知参差不齐，需要通过磨合达成共识，因此智谱AI通过公开交流表达行业观点，通过技术开源，帮助客户企业理解大模型背后的价值、难度、成本；同合作伙伴建立生态合作机制，结合其所处的行业和能力需求，有针对性地开展技术培训、联合创新等工作。

【过去突破】 1.“中国创新”：和清华大学联合研究，提出了自主研发GML算法框架，融合了GPT和BERT两种训练框架解决训练模型问题。2022年8月，智谱AI开源1300亿参数规模的双语预训练模型GLM-130B，收到70余个国家1000余个研究机构的使用需求。2.商业化：在不损失任何模型精度和推理速度前提下，不仅能让大模型运行成本降低75%，还能适配国产化硬件（智谱AI发起GLM适配计划，目前已经与30多家国产硬件厂商进行适配。）3.技术迭代：公司全自研的第三代基座大模型ChatGLM3，多尺寸多场景适配，拥有更强大的性能，提升跨模态能力、AI Agent（较ChatGLM二代提升1000%）、代码生成以及强化搜索能力。全模型布局正逐渐成型。此次升级后，智谱 AI 是目前国内唯一一个有对标 OpenAI 全模型产品线的公司。

【市场策略和商业化探索思路】 公司践行MaaS的市场理念，负责模型推理训练、算力、适配等高复杂事项，行业和企业端则可以无压力地部署应用，全面开源，共创共建。通过提升核心技术推动商业化落地，大模型技术比上一代 AI 技术有更强的通用性和泛化能力，因此公司能尽量避免定制化。

【Agent趋势的理解】 AI Agent是大模型技术体系中非常重要的关键环节，但并非完全独立运作，本质上还是挖掘大模型本身的潜力。其优势在于，将理解、规划、执行以及自我调整等模型的基本能力、原子化能力结合起来，解决更复杂的问题，拓展模型应用范围。

【行业预判】 大模型的原生应用时代正在到来，对企业业务模式、管理流程、技术架构产生深刻影响：1.大模型改变传统软件开发范式。开发者从借助产品的平台能力转化成用户价值，到从模型本身特性出发开发新应用，展现大模型原生能力，思维范式与过去相反。2.大模型改变传统软件开发方式和生态应用方式。早期建立的平台，由软件开发者定义和实现用户功能、修复漏洞，现在大模型真正赋予软件生态智慧，将更自动、更智能地打通应用产品背后的数据、逻辑、流程。

【未来规划】 继续追赶世界顶尖水平，智谱AI的技术迭代周期约3-6个月，做更适合中国企业的原生基座大模型；面向企业端，即在应用场景给客户带来更大的价值——不仅提质、降本、增效，还有创新，做大市场蛋糕，与合作伙伴一同高速发展。



Agent模式应用案例：360智脑大模型

- **成立时间：**2023-03-29，北京
- **行业定位：**人工智能大模型
- **主营业务：**360智脑是360自研认知型通用大模型。依托360多年人工智能技术积累及搜索、浏览器等大模型场景先发优势，360智脑已拥有千亿参数规模，具备生成创作、多轮对话、逻辑推理等十大核心能力、数百项细分功能。
- **服务对象：**面向个人，360智脑全面接入360全端应用，打造个人AI助理，提供智能创作、智能办公、智能陪伴等百余种生活及工作场景下的服务；面向企业级用户，360将大模型拉下神坛，构建企业级应用九层架构，从普及通用大模型，到深度定制垂直大模型，再到支持大模型驱动业务连接和协作，三步走循序渐进推进模型落地，赋能百行千业数转智改，让大模型成为辅助企业和员工的“副驾驶”。
- **落地情况：**360智脑及其解决方案已在金融、医疗、教育、税务、企服等近20个行业落地应用，赋能企业级市场提高生产力和生产效率，服务产业数字化。

360智脑观点：大模型并非万能，在大多数商业化场景落地过程中需要扬长避短

大模型的“长”

- 内容生产、内容理解、逻辑推理
- 极大降低使用者和开发者的门槛。但目前产品落地开发和使用的过程仍然复杂，比如提示词学习、二次开发等。

扬长

- 公司根据不同用户的开发能力提供不同的产品解决方案。
- 采用Agent模式，针对API、插件等工具使用对基础模型进行指令微调，使其能根据提问自主思考决策调用的工具，增加大模型对领域内任务和工具的理解。
- 进一步降低开发者门槛，搭建软件、低代码工具等应用平台。

大模型的“短”

- 通用大模型经常会遇到幻觉、遗忘、安全性问题，垂直领域难落地。
- 大模型应用在垂直领域缺乏行业know-how。

避短

- RAG搜索增强生成，结合传统搜索框架、构建领域知识库检索，利用大模型对提问的理解能力和对答案的二次加工能力，输出可控结果。
- 先在企业内部落地测试，采用“副驾驶”模式。
- 业务倾向于提供通用大模型基座，垂直行业落地选择通过生态合作伙伴共同解决。公司设立GPT产业联盟，与入驻企业建立包含技术、产业、资金、政府等等合作。其中，与垂直行业内的SaaS公司的合作模式对大模型落地最为切实可行。

【过去挑战】大模型需要达到足够参数量来提升性能。

【过去突破】360智脑及其解决方案已在金融、医疗、教育、税务、企服等近20个行业落地应用，赋能企业级市场提高生产力和生产效率，服务产业数字化，其中360智脑数字人业务已服务上千家企业。

【第三季度市场洞察】模型精度需求相对较低、容错率较高，能够人机结合的赛道和场景更有潜力。

【第三季度需求洞察】中小微企业希望大模型产品“开箱即用”，更关注其对业务带来的降本作用；很多大企业更关注增效作用；政府客户首要考虑安全性，其次是技术能力。

【第三季度趋势洞察】1.行业更关注技术应用，包括Agent能力；2.模型算法合规性很重要。近期会经过备案的开源视觉大模型，能安全解决垂直领域的绘画需求；已推出的新产品LoRA360，能够解决通用大模型在文生图场景会遇到的生成内容不可控、质量较低等问题。

合纵连横，大模型+垂直模型直击场景痛点是第一步

大模型与垂直领域模型的结合可以优势互补，拓展应用价值，提高模型的泛化能力、计算效率、满足特定需求以及促进知识迁移等方面的性能。这种结合方式可以为不同领域和不同场景提供更全面、更准确、更高效的解决方案，推动人工智能技术的进一步发展。



行业应用案例：乐言科技-AI SaaS+全链路数智化解决方案

- **成立时间：**2016-04-17，上海
- **融资轮次：**D轮，中金资本、上海人工智能产业投资基金等
- **行业定位：**大模型、场景应用：客服机器人、RPA机器人、语音外呼机器人、智能客户管理系统。
- **主营业务：**人工智能整体解决方案提供商，为电商行业提供AI SaaS+电商全链路数智化解决方案。旗下拥有全平台智能客服机器人、智能客户管理系统、智能工单RPA、智能语音外呼、海外电商服务运营平台等产品，覆盖咨询服务、新客转化、老客复购的完整客户生命周期。
- **服务对象：**服务6万+电商商家，触达终端买家超数十亿人次，服务覆盖国内100多个城市以及东南亚地区。
- **落地情况：**乐言科技自主研发的电商行业大模型已正式上线商用。乐言电商GPT大模型是主要聚焦To B领域，深耕电商行业和泛零售行业。



乐言科技观点：通用大模型+垂类SaaS合作打造标杆用例，将带领行业向前

【上半年趋势洞察】“百模大战”，大模型军备竞赛；应用“百花齐放”，商业化初见端倪。

【第三季度趋势洞察】技术方面，GPT-4 上线代码编辑器，标志着在某些工作内容中，人类和 AI 协同办公的Copilot 模式已进入到下一阶段，即AI 能够完成绝大部分的工作的Agent模式；OpenAI升级了DALL·E 3，也推动自然语言+绘图的人机交互迈向新的里程碑，即AI不但理解多模态信息，还能将实际业务需求与其结合，扮演“操作系统”的角色。行业环境方面，监管越来越严格和清晰。

【第三季度需求洞察】C 端场景更加多样化，相对繁荣；B端，在电商领域，越来越多客户关心产品落地，需求旺盛，但鲜少有标杆案例出现。小微企业更关心降本提效，包括节省人力成本，优化业务流程等等方面，但缺乏资金投入；中型企业需要在市场营销、经营决策等的业务环节带来增益，但相对缺乏对具体场景的明确诉求，难度在于整合碎片化的需求；相对前瞻性的大型企业希望运用大模型提升行业认知度和竞争壁垒，希望乐言科技提供数据和行业 know-how，合作训练行业大模型，或在其数据安全合规的前提下，有效地利用大模型找到合适的落地场景；也有自研能力的大型企业，本身拥有AI 技术团队，选择内部自研创新和探索，较少寻求第三方合作。

【行业预判】当前中小企业由于资金、数据、技术储备还不足以支撑大模型的落地，且普遍对技术发展还未形成客观认知。行业亟需头部通用大模型企业与乐言科技这样的垂直领域的SaaS服务商共同打造标杆案例，在技术、数据、算力、业务场景、市场运营等全方位试炼与试错。第三、四季度公司大模型业务也将致力于在生态合作模式以及大模型在垂直领域的落地方式的探索。

【大模型成本】主要来自几个方面：1.数据标注和模型训练；2.算力投入；3.贴合场景做模型调试和运营；4.数据安全相关的投入。

【大模型行业格局】分为两类玩家：1.打造通用大模型和生态体系的超级大玩家，目前看寡头效益比较明显。2.围绕行业场景做应用落地的玩家，to c和to b领域都有很多活跃的玩家。

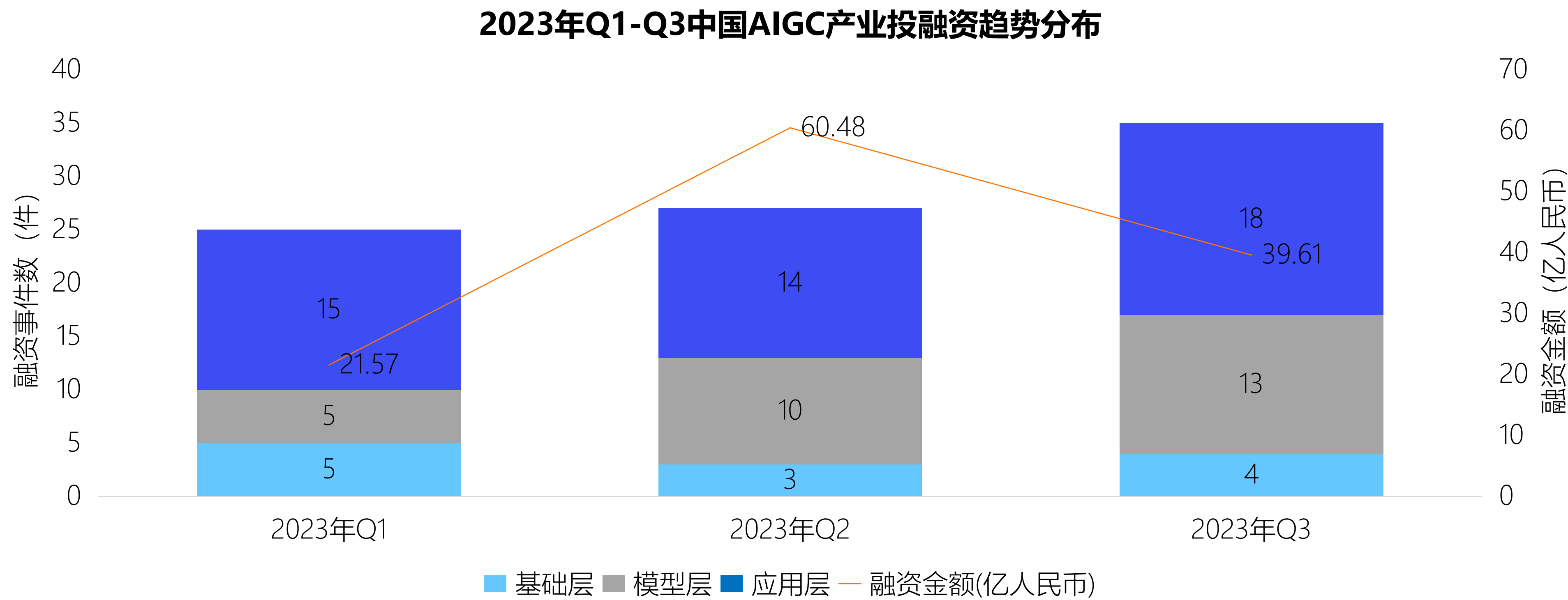
投融资分析

- ◆Q3投融资整体趋势
- ◆Q3融资规模分布
- ◆Q3热点赛道、热点城市、活跃机构
- ◆Q3大额融资事件



投融资趋势：融资事件日益活跃，金额冲高回落

- 据睿兽分析数据，中国AIGC行业2023年第三季度发生融资事件35件，已披露融资金额39.61亿人民币，涉及公司33家，涉及投资机构数量51家。其中，智谱AI（AI知识智能技术开发商）和兔展（AIGC内容引擎与营销云核心平台）在第三季度获得两轮融资。
- 2023年AIGC行业融资事件不断上升，说明从年初AIGC赛道火热以来，市场普遍看好其应用前景，在AIGC行业确定性的趋势下，投资活动日益频繁。
- Q3投融资金额经历冲高回落，融资金额均值从Q2的2.24亿元环比下降至1.13亿元。主要原因在于：应用层和工具平台玩家增多，共建行业生态，稀释融资额均值。整体经济环境仍处于下行期，行业示范效应尚不明显，大模型应用落地路径处于探索阶段，投资相对谨慎。

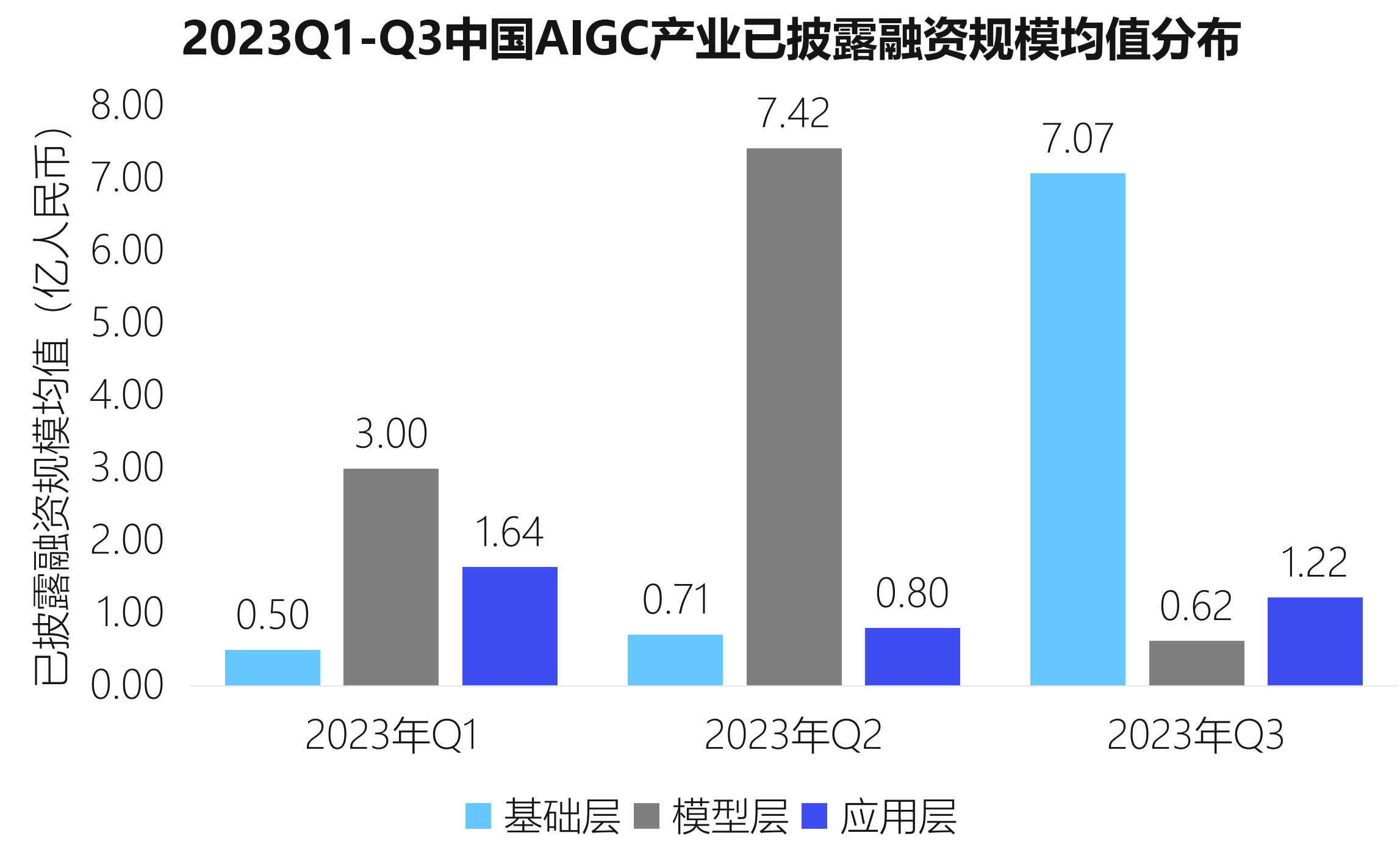
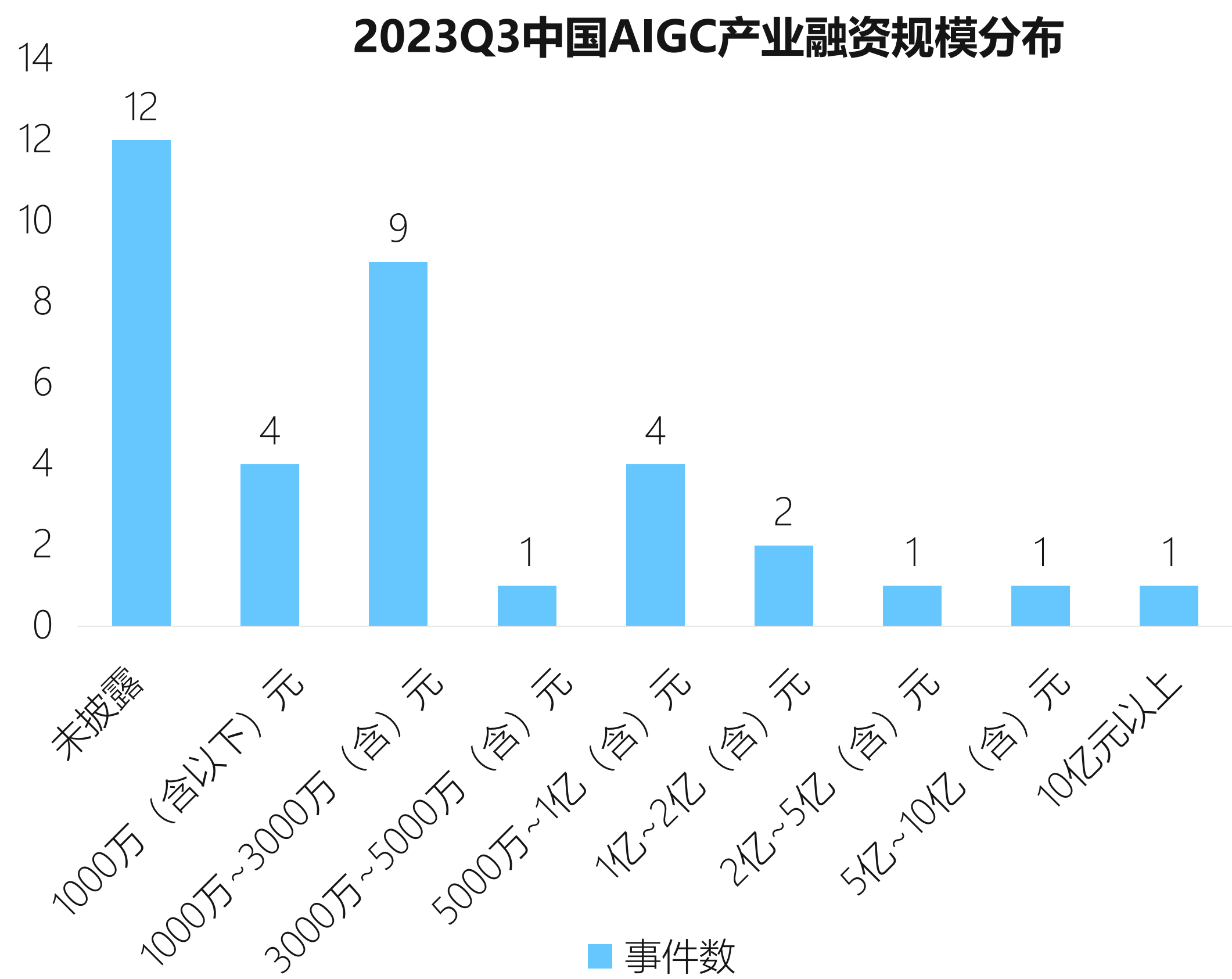


注：融资事件为种子轮至Pre-IPO轮次发生的事件，不含IPO

资料来源：睿兽分析，创业邦研究中心

融资规模：基础层均值显著提高，模型层下降

- 据睿兽分析数据，2023年Q3发生的融资事件中，已披露融资规模以1000万~3000万区间居多。基础层融资规模均值最高，达到7.07亿元。主要由于燧原科技（人工智能领域云端算力平台）融资20.00亿人民币。
- 2023年Q1~Q3，由于产业对人工智能的竞逐，市场对算力和数据具有刚性需求，大额融资逐渐注入算力、数据两大AIGC基础层领域；模型层融资金额均值从Q2的7.42亿元环比下降至Q3的0.62亿元。其中，工具平台已披露融资规模均值为0.3，通用大模型均值为0.93。Q3大模型技术迭代过快以及通用大模型和应用层之间、和垂直大模型之间的边界不清晰的情况凸显，具有底层技术创新能力和商业化能力的创业企业稀缺，投资策略难以形成共识。

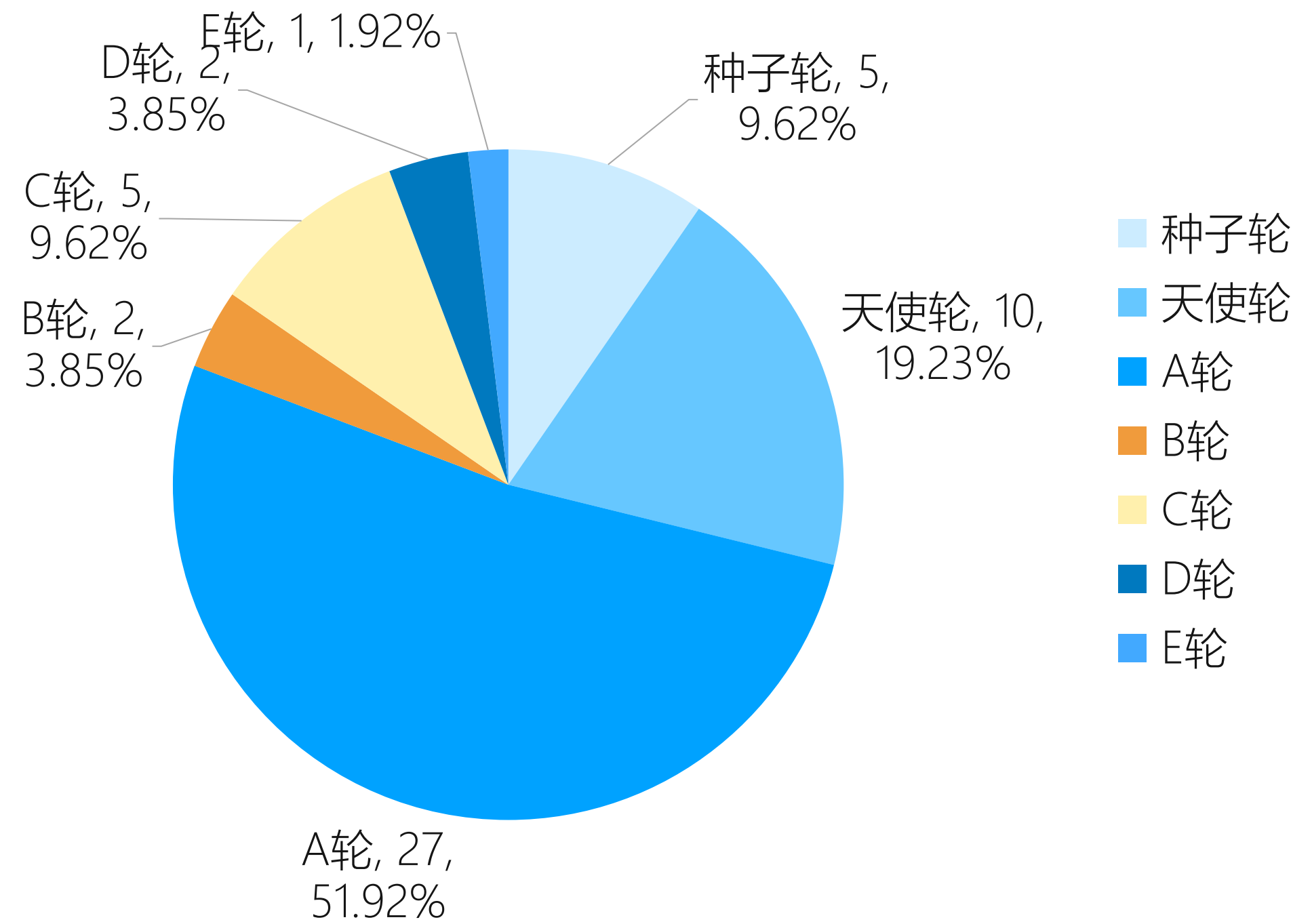


资料来源：睿兽分析，创业邦研究中心

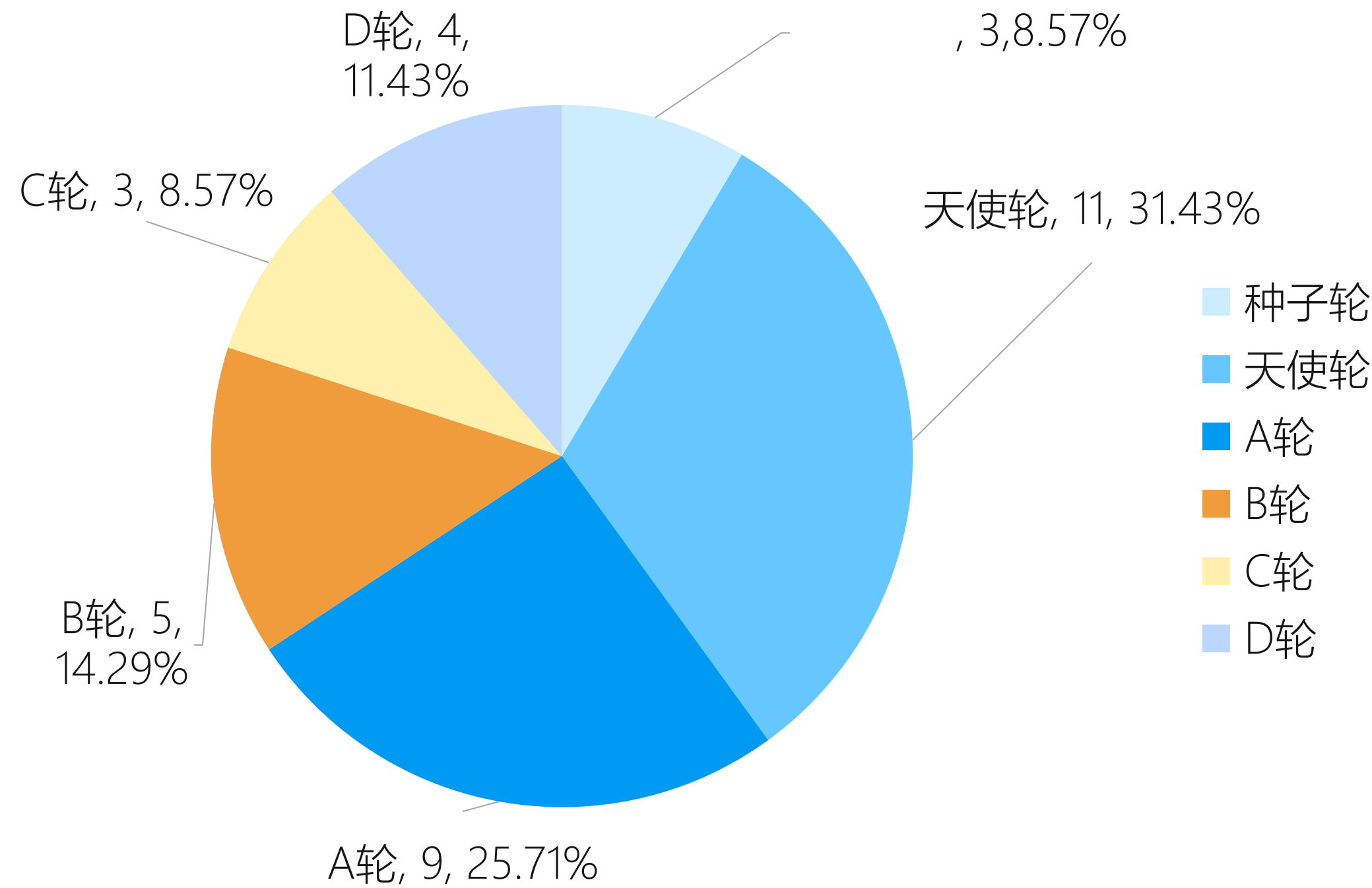
融资轮次：行业转型促进成长期和后期投资增长

- 据睿兽分析数据，2023年上半年，从交易数量看，国内AIGC产业融资事件大多发生在A轮（事件数27，占比51.92%）、天使轮（事件数10，占比19.23%）、种子轮（事件数5，占比9.62%），早期事件数（A轮、天使轮、种子轮）占比达到80.77%；而相比较而言，Q3成长期和后期（B轮及以上）占比大幅提升，从H1的19.23%上升至34.29%，早期事件数占比从H1的80.77%下降至65.71%，说明在行业转型的需求下，成长期和后期的创业企业融入AIGC的浪潮中来。

2023年H1AIGC产业融资轮次交易数量分布



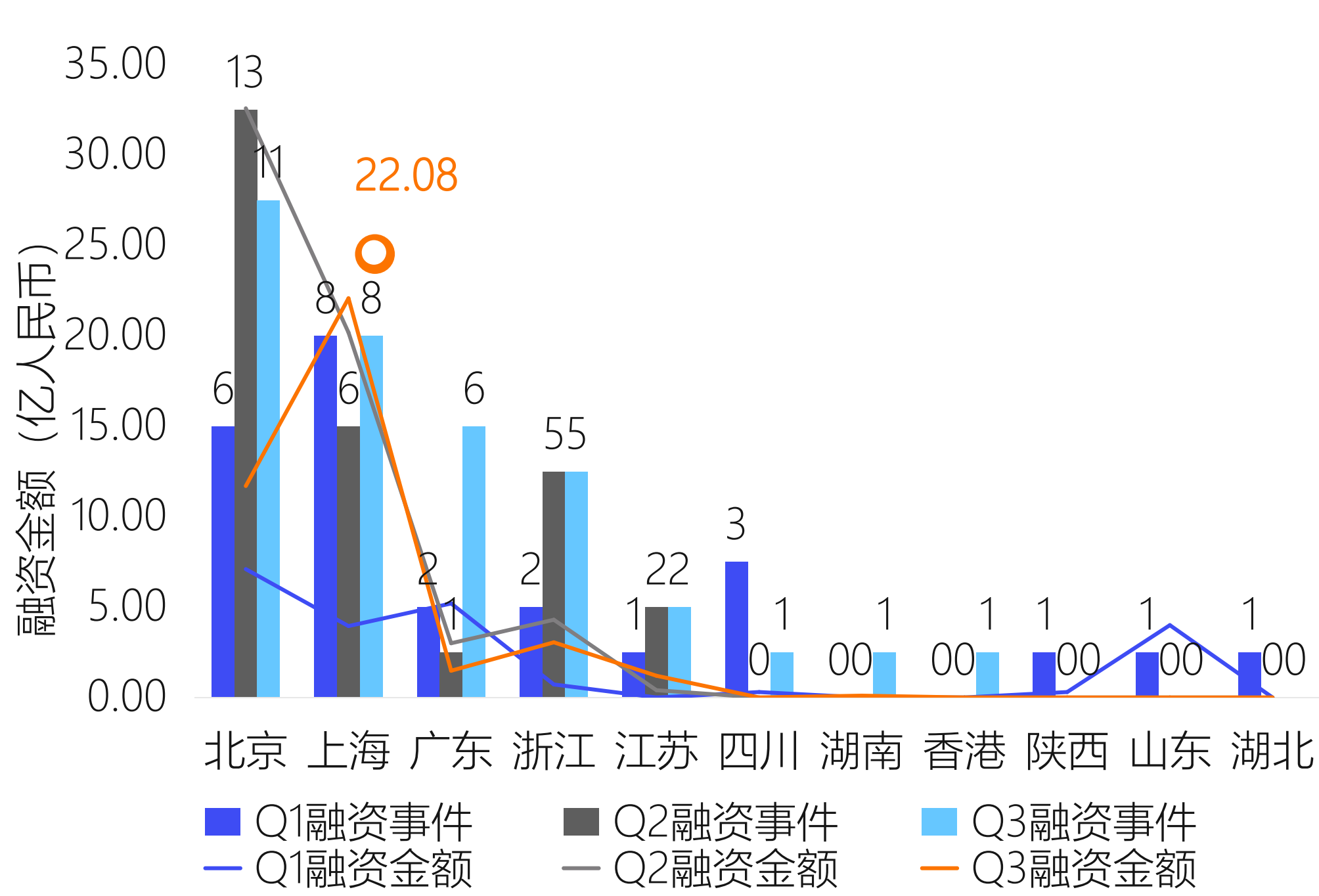
2023年Q3AIGC产业融资轮次交易数量分布



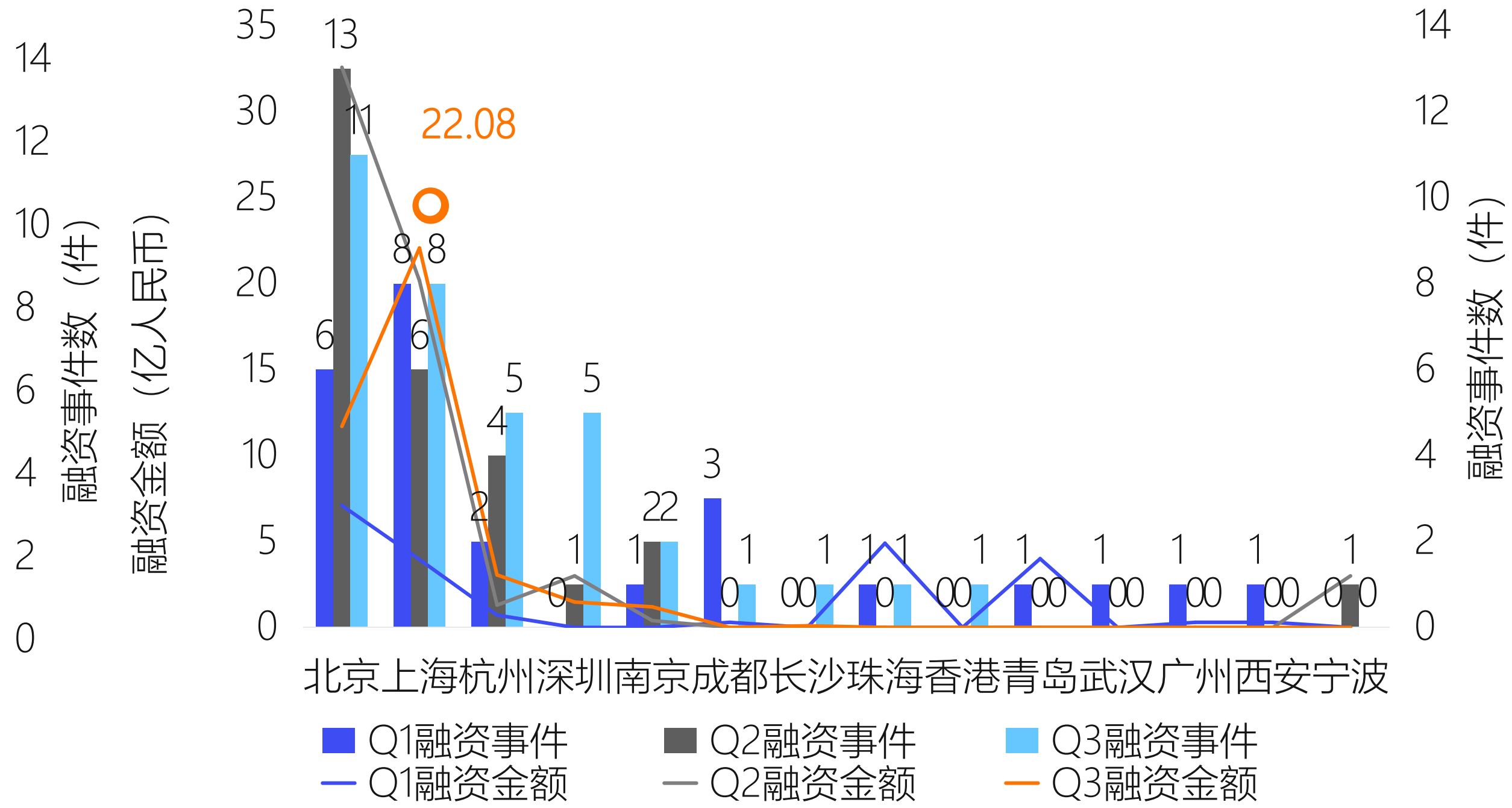
热点区域： 国资领投， 上海超越北京成最吸金地区

- 从热点地区分布来看，相比2023年上半年，进入第三季度，北京仍为最活跃地区，融资事件为11件，以模型层居多，包括智谱AI（AI知识智能技术开发商）、爱诗科技（视觉多模态算法开发商）、生数科技（多模态生成式大模型与应用产品开发商）等，应用层的深势科技（人工智能和分子模拟算法平台）融资金额最高，超过7亿元人民币。
- 上海超越北京成为国内AIGC融资最吸金的地区，融资总额达22.08亿元。主要由于燧原科技（人工智能领域云端算力平台）融资20.00亿人民币，由上海国际集团旗下子公司及产业基金，包括国际资管、国鑫创投、国方创新、金浦投资旗下上海金融科技基金、国和投资联合领投。上海集聚全国最多的智能芯片创新企业，初步形成了智能芯片产业集群，包括沐曦、燧原科技、瀚博半导体、天数智芯等。据上海市经信委披露，2022年~2023年8月，已有17块AI芯片点亮，其中推理芯片7块、训练芯片4块、车载智能芯片3块。

2023Q1-Q3中国AIGC产业投资区域分布



2023Q1-Q3中国AIGC产业投资城市分布

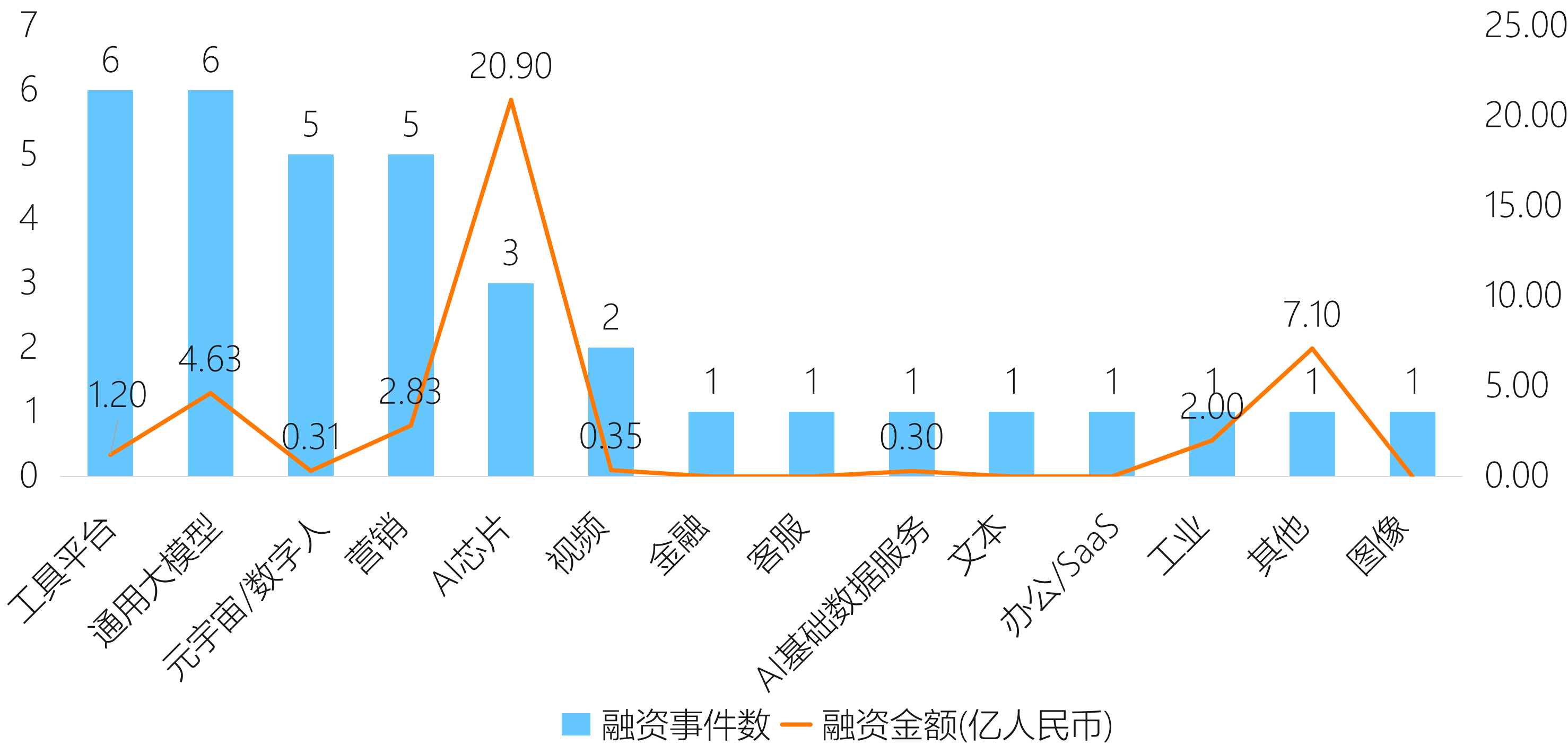


资料来源：睿兽分析，创业邦研究中心

热点赛道：工具平台和通用大模型依然最受追捧

- 据睿兽分析数据，2023年Q3AIGC行业发生的融资事件中，国内AIGC产业在通用大模型（6次）、工具平台（6次）两个细分赛道相对为活跃。由此看出，技术仍为大模型行业最主要的投资要素之一。
- 梳理Q3国内5家获得融资的通用大模型企业画像，可见开源、自适应智能体、多模态大模型已成为行业发展的重要趋势。
- 在应用层中，元宇宙/数字人和营销是融资事件最频繁的细分领域。虚拟数字人不仅能够模拟人类的外貌和语音，还有望通过情感识别技术理解人类情感，实现更真实的互动，在营销场景的商业化路径较为清晰。知识图谱能够协助管理项目、进行数据分析、提供专业建议等等，使得数字员工作为虚拟助手具备更高级的能力。

2023年Q3AIGC产业热点赛道分布



公司名称	公司画像
OpenCSG	AI大模型 开源 公司
无限光年	生成式AI和AGI研发商，核心技术自我进化的通用大模型以及领域自适应 智能体 的研发
智子引擎	多模态大模型 初创公司
生数科技	多模态生成式大模型 与应用产品开发商
智谱AI	AI 认知智能 技术开发商

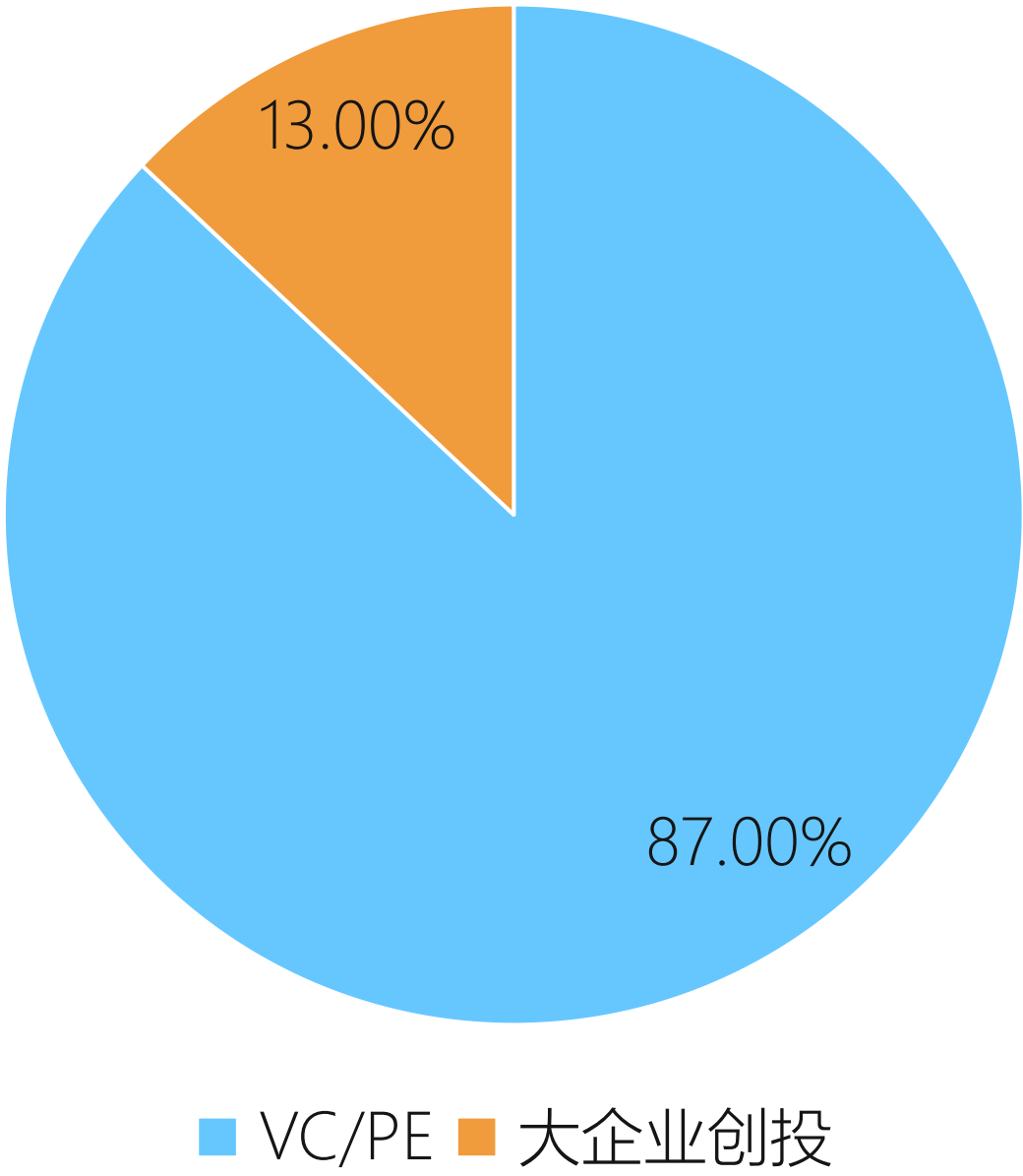
▲Q3国内5家获得融资的通用大模型企业画像

活跃机构：原美元基金与互联网大厂CVC扮演最重要的角色

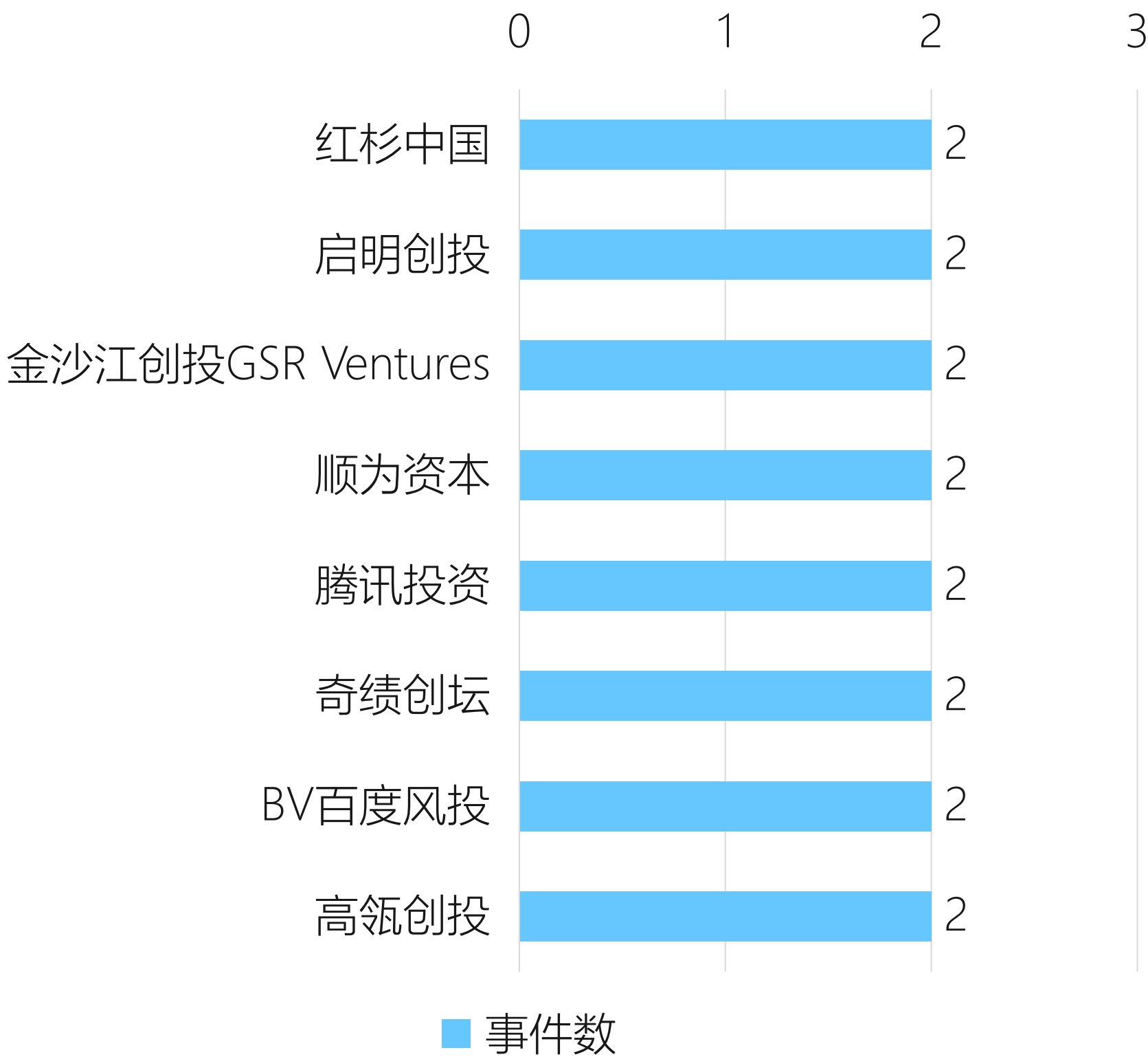
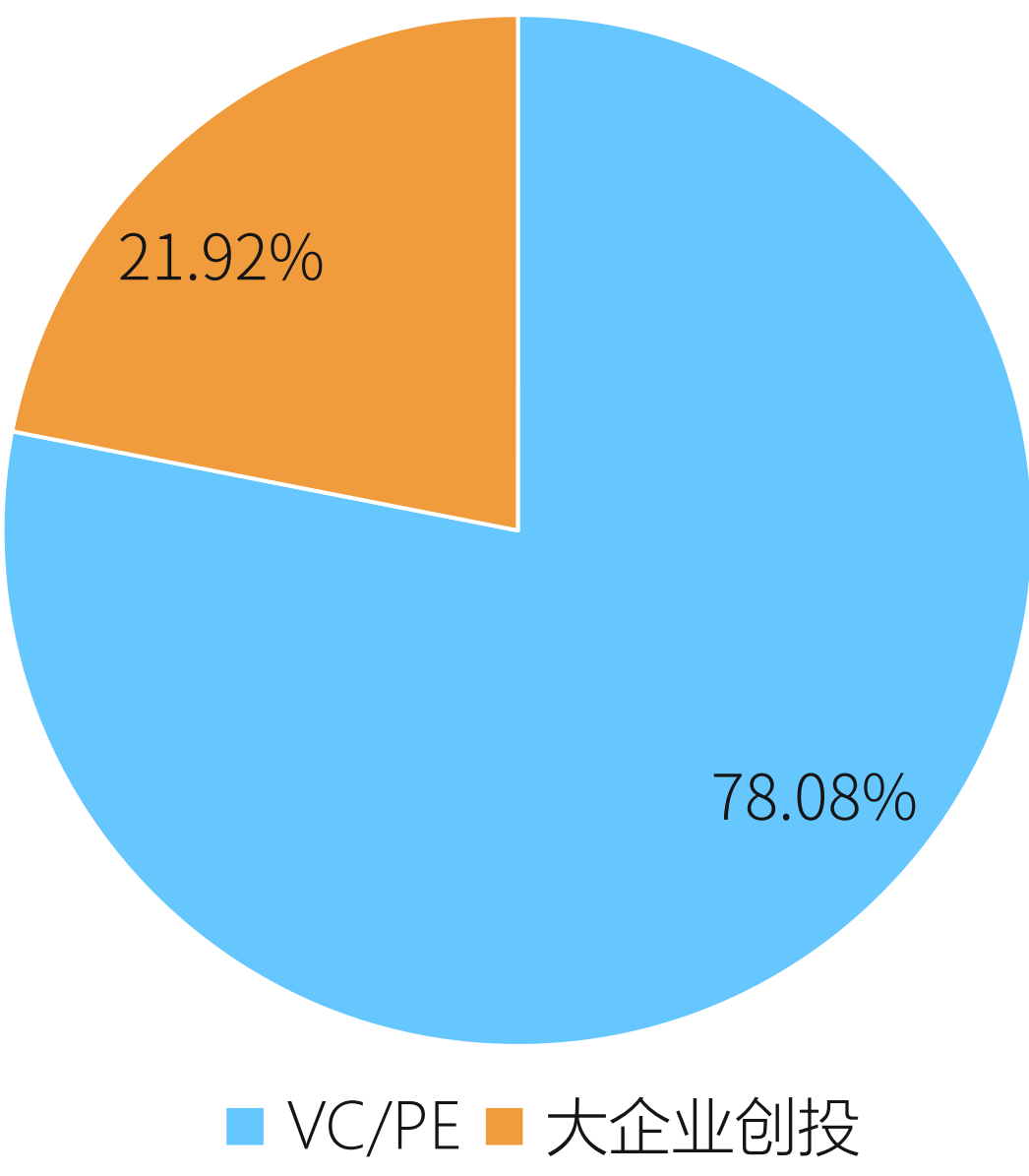
- 据睿兽分析数据，第三季度，投资机构类型以VC/PE为主，占比78.08%，大企业创投占比21.92%，相较上半年，Q3CVC明显在AIGC市场发挥了更大的作用，CVC参与的事件数占比从13.00%提升至21.92%。CVC以技术、资金、应用场景、市场渠道等多种形式合作，数据要素流动频繁，推动行业生态日趋完善。驻扎在人工智能领域的产业资本，一直以来主要集中在始终在进行前沿技术探索的互联网大厂、部分人工智能公司以及未来业务可以与之结合的游戏、机器人、汽车等应用型公司。
- 红杉中国、启明创投、金沙江创投、顺为资本等机构投融资相对活跃。

2023年Q3AIGC产业活跃投资机构

2023年H1AIGC产业投资机构类型分布



2023年Q3AIGC产业投资机构类型分布



■ 事件数

资料来源：睿兽分析，创业邦研究中心

2023年Q3国内超1亿元大额融资事件

公司	简介	国家和地区	融资轮次	融资金额	融资时间	投资方
燧原科技	人工智能领域云端算力平台	中国上海	D轮	20.00亿人民币	2023-09-28	疆亘资本、上海国际资管、国鑫创投、国方创新(王磊)、金浦投资、上海国和投资、腾讯投资(姚磊文)、美图(吴欣鸿)、武岳峰科创、允泰资本、弘卓资本(何燕青)、红点中国、广发乾和(何宽华)、达泰资本、浦东投资控股
深势科技	人工智能和分子模拟算法平台	中国北京	B+轮	过7.00亿人民币	2023-08-18	哈勃投资、众源资本、和玉资本、正心谷资本
智谱AI	AI认知智能技术开发商	中国北京	B+轮	数亿人民币	2023-07-19	美团战略投资部
Style3D	柔性快反供应链管理解决方案提供商	中国浙江	B轮	2亿人民币	2023-08-23	鼎晖投资、高瓴创投、BV百度风投
兔展	AIGC内容引擎与营销云核心平台	中国广东	D+轮	过亿人民币	2023-08-23	南山战新投

2023年Q3海外超1亿元大额融资事件

公司	简介	国家和地区	融资轮次	融资金额	融资时间	投资方
Anthropic	人工智能安全服务商	美国加利福尼亚	C4轮	13.00亿美元	2023-09-25	Amazon亚马逊
Hugging Face	AI聊天机器人研发商	美国纽约	D轮	2.35亿美元	2023-08-23	Salesforce Ventures、Google、Sound Ventures、Amazon亚马逊、高通创投、英特尔资本、IBM Ventures、NVIDIA英伟达、AMD Ventures
Writer	AI智能营销平台研发商	美国加利福尼亚	B轮	1.00亿美元	2023-09-18	ICONIQ Growth、Insight Partners、Balderton Capital、Gradient Ventures、WndrCo、Aspect Ventures、Accenture Ventures、Vanguard
Anthropic	人工智能安全服务商	美国加利福尼亚	C3轮	1.00亿美元	2023-08-13	
Inworld AI	生成式AI创建NPC平台研发商	美国加利福尼亚	A+轮	5000.00万美元	2023-08-02	Lightspeed Venture Partners光速创投、First Spark Ventures、LG Technology Ventures、M12、Samsung NEXT、Stanford University斯坦福大学

总结

◆行业洞察

◆投资洞察



行业洞察

- 由于大模型带来的革命类似但有别于互联网，因此，传统的互联网思维不足以覆盖AIGC的发展路径。
- **技术趋势：**当大模型参数量普遍达到一定水平，在边际效用递减的规律下，将有更多的方式来提升模型的能力和效用，比如微调、提示工程、搜索增强生成、AI Agent等技术手段。AI Agent是挖掘大模型潜力的关键环节，但不是终点，也不可独立运作。产品向终端延伸，结合更多AI应用技术推动应用场景多元化发展。
- **政策趋势：**由于政策面向C端设置准入门槛，标准体系覆盖多个行业，涵盖模型行业能力、模型工程化性能、模型算力网络、模型安全可靠等方面，数据、算法、模型、安全因素重要性凸显，意味着“百模大战”回归理性，行业格局迈入整合阶段。
- **市场趋势：**模型层企业市场集中度低，竞争激烈，应用层生态“百花齐放”。
- **产业趋势：**应用层，各个行业垂类应用更加清晰，调用大模型接口或自研做一定的训练。模型层，大模型原生时代正在到来。大模型根本上革新研发范式，因此应用方式和思路也需要相应改变。但目前以“大模型+垂直领域模型”打造行业标杆案例还处于探索阶段。相比于上一次深度学习计算机视觉的浪潮，AIGC大模型对底层数据生产的效率起到了决定性的影响，从应用层驱动智慧飞轮也变得更加容易，因此各行业企业在应用大模型技术的过程中需理解大模型的长处，找到应用场景的痛点，并与垂直领域应用进行深入磨合，构建知识工程和创新体系，发挥大模型的价值。
- **挑战与机遇：**大模型目前逻辑思维能力不强；知识边界模糊，会出现幻觉等问题；需要更便捷、更安全的大模型软硬件协同解决方案；训练、推理成本依然很高。这些也是大模型商业化路径探索的关键议题与机遇。

投资洞察

■ 重要的投资因素

- **技术成熟度：**与支撑应用落地存在一定距离，大模型技术成熟度还有待提升，目前仍然处于追赶国外领先产品的阶段。
- **创新潜力：**C端活跃度相对较高，B端缺乏行业标杆案例。未来一段时间内，大模型预计与各行各业的应用场景不断磨合，直至找到具有可行性的落地路径。
- **商业模式：**MaaS作为大模型理想的商业化模式，其有效实现依赖大模型进一步的通用性和泛化能力。当前阶段，除大模型服务外，还存在大模型+算力、大模型+应用等“打包”售卖的形式。

■ 趋势洞察

- 2023年AIGC行业融资事件不断上升，说明从年初AIGC赛道火热以来，市场普遍看好其应用前景，投资日益频繁。Q3投融资金额经历冲高回落，主要原因在于应用层和工具平台玩家增多，共建行业生态，稀释融资额均值；整体经济环境仍处于下行期，行业示范效应尚不明显，大模型应用落地路径处于探索阶段，投资相对谨慎；大模型技术迭代过快，通用大模型和应用层之间、和垂直大模型之间的边界不清晰的情况凸显，具有底层技术创新能力和商业化能力的创业企业稀缺，投资策略难以形成共识。
- 由于大模型带来的革命类似但有别于互联网，开发和应用的思维也需发生转变，因此，跨周期思维定式带来的旧的投资逻辑不再适用，具有“模型基因”、尽其所长的企业更有潜力价值，也是需要行业共同探索的方向。但目前外界对“AI助手”技术的认知参差不齐。
- 相比VE/PE，许多互联网大厂具备实践经验、大数据、资金、场景，与垂直领域的模型提供商合作将有效推动行业发展，找到落地路径，因此进入第三季度以来，国内AIGC创投市场CVC融资事件占比显著提升。下一阶段，AIGC创投市场国资有望发挥更大作用。

关于我们

创业邦

创业邦是领先的国际创新生态服务平台，为高成长企业、金融机构、产业大公司、政府园区提供全方位的媒体资讯、数字会展、数据研究、创新咨询、教育培训、资本对接等服务。



立即扫码关注



睿兽分析

睿兽分析是创业邦旗下横跨一二级市场的综合性创新数据平台，致力于通过即时、有效、可触达的行业一手数据，为大企业、地方政府、金融机构、投资机构等经济主体，提供强有力的创新驱动与投资决策依据。



立即扫码关注

