

前言

2023年，大模型与生成式AI的崛起无疑成为了技术领域的焦点。在这一年里，大模型和生成式人工智能（AI）的讨论持续“破圈”，各类商用大模型和开源大模型的发布和更新将技术创新推向了新的高峰，相关产品不断发布。

然而，AI领域的巨大变革外，其他数字技术也在助力实体经济的道路上稳步迈进。5G、数据资产、云计算、数字孪生等技术正融入工业互联网的各个环节，产品设计、资源勘探、生产制造、设备管理、安全生产各个环节之间也日益融合。5G工厂、双跨平台等工业互联网建设成果不断累积和爆发。

在此背景下，InfoQ研究中心从人工智能新纪元和以数强实稳根基两大篇章，从数据定量和公开资料、专家访谈定性的角度，深入回顾2023年技术市场的重大进展和演进，并以此为基础，探究2024年技术发展的未来方向。

我们期望，在这个VUCA时代，这份报告能够帮助各位开发者、科技领导者以及各类对技术领域感兴趣的读者朋友，把握未来技术发展的脉络。InfoQ研究中心也将继续以专业、客观和持续的报告产出，同大家一起见证技术发展的每一个重要时刻。

——InfoQ研究中心

1

桌面研究

通过对行业公开信息进行桌面研究，资料包括但不限于专业机构研究报告、技术领域书籍、技术社区网站、技术企业官网及专家公开演讲内容等；

2

技术模型

通过技术专利数量、技术诞生时间、技术舆论指数指标，进行技术发展阶段类型模型设计，作为细分技术发展阶段展现；

3

专家访谈

InfoQ 研究中心针对本次研究定向邀请了多位技术企业管理层、垂直技术领域专家、技术从业者、产业从业者等各方面专家；

研究数据说明——数据来源

本次研究通过收集并整理160+细分技术领域的技术专利数量、技术诞生时间、技术舆论指数、技术融资事件作为2023年重要技术发展回顾与技术发展特征解读的基础。

01

技术专利数量

技术专利数据来自国家知识产权局专利数据库。使用对应技术领域关键词进行检索，截止时间2023年11月30日。

02

技术诞生时间

考虑到数据可得性与统一性，技术诞生时间使用知网论文库进行相关技术领域论文最早收录年份统计。

03

技术舆论指数

技术舆论指数数据来源为搜索引擎和主要技术媒体，其中包括百度、InfoQ中文站和CSDN社区等网站，具体数据通过对应技术领域关键词搜索涉及的文章数量获得，指数通过加权获得。

04

技术融资事件

技术融资时间数据来源为IT桔子融资事件库和36氪项目库，分别按照不同细分技术领域进行融资时间、融资金额、投资机构的统计。

CONTENT

目录

01 | 2023年技术市场回顾

02 | 2024年技术趋势展望

01

2023年技术市场回顾

2023年技术领域热点新闻：AI、云、稳定与安全仍然是重点

2023年3月，新一代serverless事件中间件EventMesh顺利通过Apache孵化器项目管理委员会的毕业评估，晋升为Apache基金会顶级开源项目。这意味着以事件驱动为代表的serverless进入新发展阶段，微服务serverless，边缘serverless，AI serverless，通用serverless 成为发展新趋势。

Mojo将Python的特性，同C、C++和CUDA的系统编程功能结合了起来，并与其它Python速度增强方案区分开来。在运行速度上比原始Python快上3.5万倍。

2023年Q3：金融、医疗、交通、营销、政务、法律、气象等行业均发布了国产大模型产品：蚂蚁集团AntFinGLM金融大模型、度小满开源大模型轩辕-70B等；百度的医疗AI大模型灵医大模型、京东健康的京医千询医疗大模型等。

2023年内，B站、微信、阿里云、滴滴等多款产品和云服务发生崩溃事件，引发对于基础设施稳定性的广泛讨论。

谷歌发布的OSV-Scanner会基于OSV数据库评估项目的依赖项，显示与项目相关的所有漏洞。

新一代Serverless事件中间件EventMesh正式毕业

全新编程语言Mojo发布比Python快35000倍

开源大模型发展迅猛，应用场景得到进一步探索

年内多款产品崩溃服务稳定性引发关注

谷歌发布开源漏洞扫描工具OSV-Scanner

OpenAI官宣首个ChatGPT IOS 应用

OpenAI 公司表示，新款ChatGPT 应用将向用户免费开放，不设广告且支持语音输入，但发布初期仅面向美国用户。

多家云服务商对Ipv4进行收费，IT行业遭受冲击

AWS、Google Cloud Platform等云服务商纷纷宣布对IPv4收费，云服务商自己掏钱购买 IPv4 地址空间的时代即将结束。

Spring Boot 应用的新命令行工具：Just

Just无需任何配置即可自动加载有变动的源码、构建文件或 Docker 编排文件，提升 Spring Boot 应用构建时的 Java 开发体验，同时该项目也支持生成（原生）应用及（原生）Docker 镜像。

微软Bing Chat全面开放同时官宣多项重大升级

新版Bing Chat 将允许第三方参与 Bing Chat AI 的结果，同时Bing Chat 除了能做普通的文字交流问答、编程学习外，还能帮助用户“生成图片”、视频。此外，微软还在 Bing Chat 中添加了一项“历史记录”的功能。

GitHub 发布 Copilot X

Copilot X允许程序员以ChatGPT的风格，通过语音聊天来编写和调试代码。

- GitHub Copilot Chat：可实现与AI对话完成编码；
- Copilot for Pull Requests：由AI协助程序员拉取请求；
- Copilot for Docs：智能文档编写工具；
- Copilot for CLI：命令工具；
- Copilot Voice：语音扩展功能，说话就完成程序编写。

数据来源：InfoQ研究中心根据InfoQ官网文章阅读量等公开资料整理

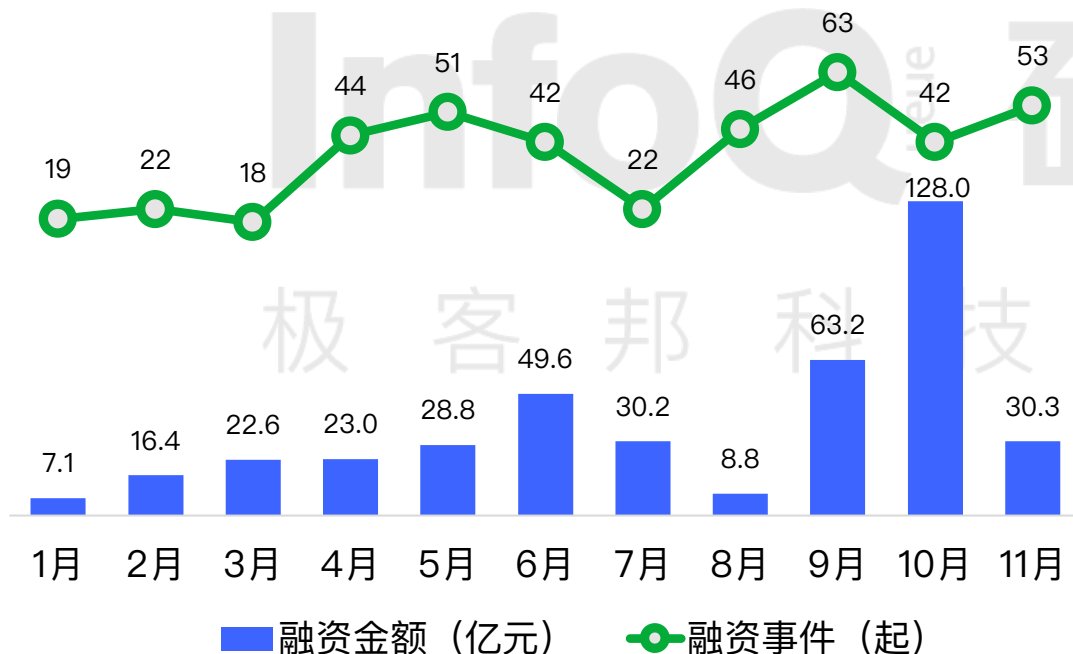


人工智能新纪元

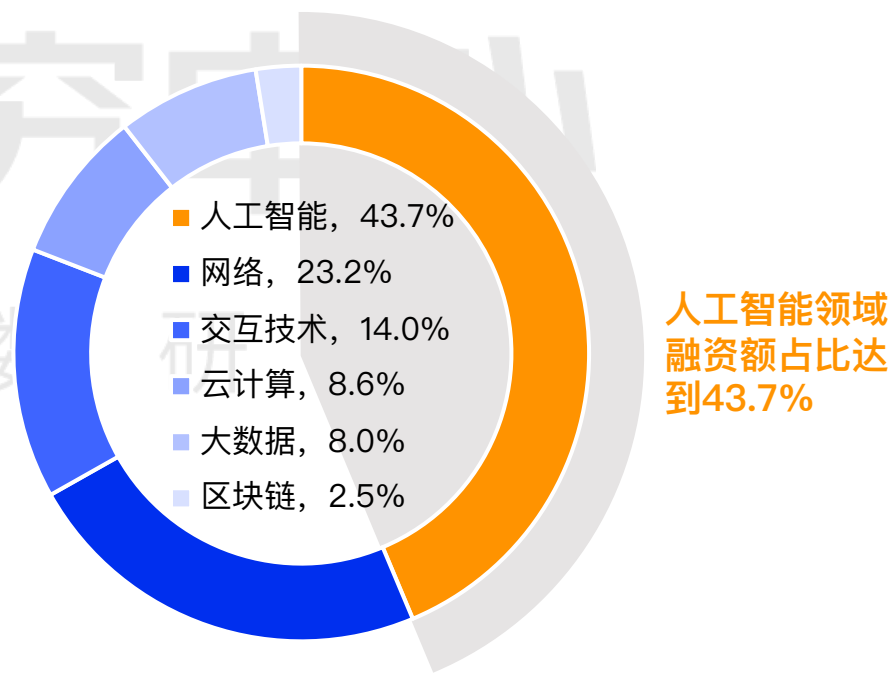
投融资：2023年技术领域共融资409.5亿元，人工智能遥遥领先

- 2023年技术领域融资事件共432起，共融资409.5亿人民币。
- 从细分领域来看，人工智能领域吸纳资金能力凸显。2023年1月-11月，仅人工智能领域融资额就占到整个技术领域的43.67%。

2023年1月-11月融资金额与事件数量统计



2023年1月-11月技术领域融资金金融统计

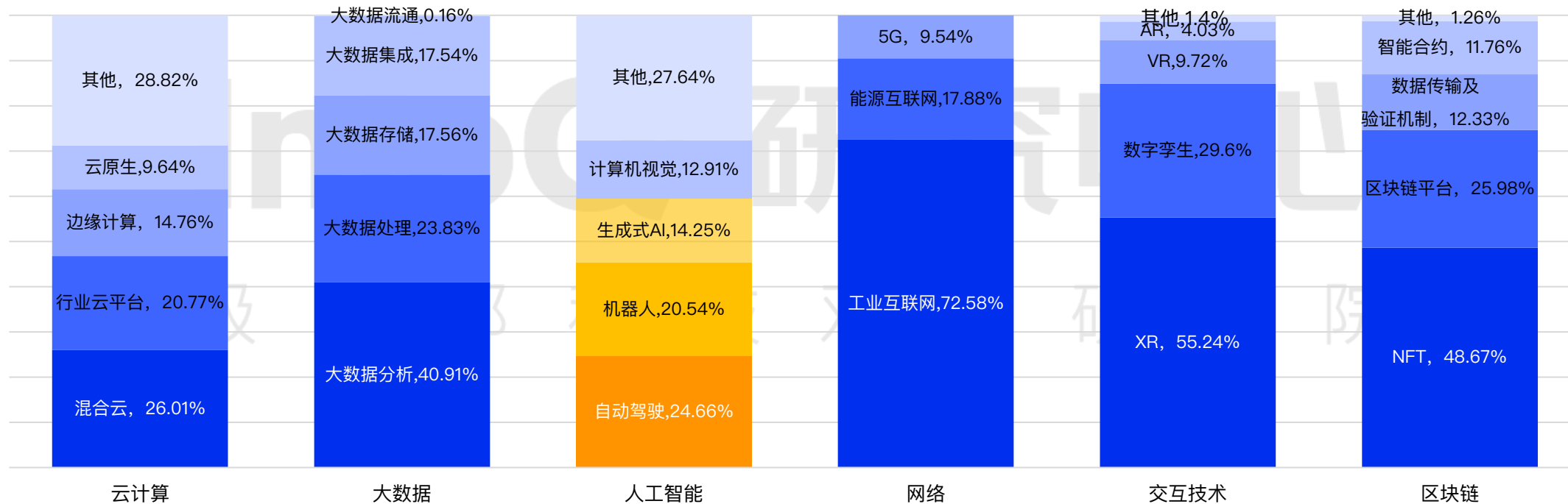


数据来源：InfoQ研究中心根据投融资新闻、资讯等公开资料整理

投融资：自动驾驶、机器人和生成式AI，占比超AI领域50%

- 人工智能领域的资金也主要流向了自动驾驶、机器人和生成式AI。这三个领域融资额占到整个人工智能领域的59.45%。

2023年1月-11月细分领域融资金额分布百分比



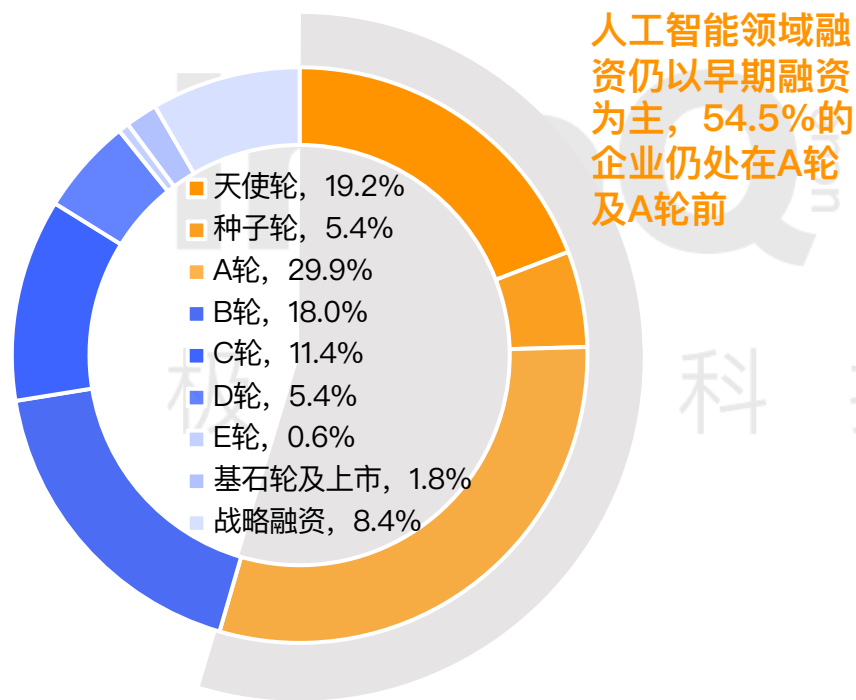
数据来源：InfoQ研究中心根据投融资新闻、资讯等公开资料整理

右侧数据表中，融资金额统一按照2023年前11个月的平均汇率（7.0366）进行换算，并经过四舍五入取整数处理。

投融资：人工智能仍处早期融资

- 人工智能领域在融资方面表现亮眼，不仅总融资额庞大，而且大额融资事件也频频发生。据统计，人工智能领域过亿融资事件达到了52件。

2023年人工智能领域融资事件轮次分布情况



2023年人工智能领域融资金额最多的中国公司 TOP10

公司简称	成立时间	总部	AI方向	最新融资轮次	年融资总额
智谱AI	2019/6	北京	大模型	B轮	25亿人民币
百川智能	2023/3	北京	大模型	A轮	21亿人民币
光年之外	2018/7	北京	大模型	天使轮	21亿人民币
燧原科技	2018/3	上海	AI芯片	D轮	20亿人民币
Minimax	2021/11	上海	大模型	A轮	18亿人民币
速腾聚创	2014/8	深圳	自动驾驶	基石轮	12亿人民币
滴滴沃芽科技	2019/3	上海	自动驾驶	战略融资	10亿人民币
达闼科技	2015/8	上海	机器人	C轮	10亿人民币
纽劭科技	2017/1	上海	自动驾驶	B轮	10亿人民币
小马智行	2016/12	北京	自动驾驶	D+轮	7亿人民币

数据来源：InfoQ研究中心根据投融资新闻、资讯等公开资料整理。

右侧数据表中，融资金额统一按照2023年前11个月的平均汇率（7.0366）进行换算，并经过四舍五入取整数处理。

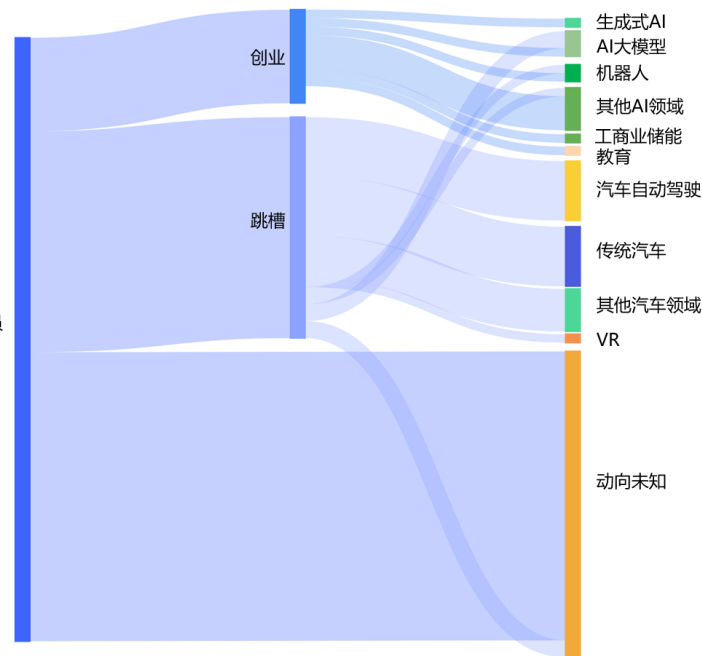
开发者：科技领导者大变动，AI成为科技领导者创业新走向

- 根据InfoQ研究中心不完全统计，2023年中国至少有70位科技领导者出现了职位变动。11位选择创业的科技领导者中，有8位的创业项目与AI相关，例如大模型、生成式AI、机器人等。

2023互联网离职人员代表图

	<p>贾扬清（原阿里技术副总裁）</p> <ul style="list-style-type: none">• 去向：创业• 创业领域：To B方向的AI应用层搭建
	<p>顾晓韬（原华为云大模型核心成员）</p> <ul style="list-style-type: none">• 去向：跳槽• 跳槽企业及领域：智谱AI —AI领域
	<p>吴新宙（原小鹏汽车自动驾驶副总裁）</p> <ul style="list-style-type: none">• 去向：跳槽• 跳槽企业及领域：英伟达全球副总裁—自动驾驶

2023年科技领导者流向图



数据来源：InfoQ研究中心根据文章、资讯等公开资料总结整理

政策关注：大模型引发民众热议，下半年受到政策持续关注

- 根据InfoQ研究中心统计和计算，在160+细分技术领域，通用大模型舆论指数排名第3，基于生成式AI的对话机器人（ChatBot）排名第15，行业大模型排名第21。
- 从7月开始，伴随着民众的广泛讨论，大模型相关政策频繁发布。

序号	发布时间	文件名	主要内容
1	2023年7月	《生成式人工智能服务管理暂行办法》	明确了生成式人工智能“提供者”内容生产、数据保护、隐私安全等方面的法定责任及法律依据，确立了人工智能产品的安全评估规定及管理办法。
2	2023年7月	《上海市推动人工智能大模型创新发展的若干措施》	推动上海大模型创新发展，营造通用人工智能创新生态，加快打造世界级人工智能产业集群。
3	2023年8月	《电子信息制造业2023—2024年稳增长行动方案》	鼓励加大数据基础设施和人工智能基础设施建设，满足人工智能、大模型应用需求。
4	2023年10月	《人形机器人创新发展指导意见》	到2025年，人形机器人创新体系初步建立，“大脑、小脑、肢体”等一批关键技术取得突破，确保核心部件安全有效供给，开发基于人工智能大模型的人形机器人“大脑”。
5	2023年11月	《北京市人工智能行业大模型创新应用白皮书（2023年）》	从大模型全球发展态势、国内外行业应用概述、北京应用情况和发展建议等方面进行了系统分析和阐述，旨在进一步推动大模型应用落地，展示北京市大模型应用成果，促进大模型价值传播和供需对接。
6	2023年11月	广东省《关于加快建设通用人工智能产业创新引领地的实施意见》	22条举措打造国家通用人工智能产业创新引领地。
7	2023年12月	《关于加快推进视听电子产业高质量发展的指导意见》	支持骨干企业做大做强，支持人工智能企业研发视听应用大模型。
8	2023年12月	《“数据要素x”三年行动计划（2024—2026年）（征求意见稿）》	以科学数据支持大模型开发，深入挖掘包含科技文献在内的各类科学数据，通过细粒度的知识抽取，构建科学知识资源底座，建设高质量语料稿库和基础科学数据集，支持开展通用人工智能大模型和垂直领域人工智能大模型训练。

数据来源：InfoQ研究中心根据政府官网等公开资料整理。

基础设施：算力是大模型的「发动机」，云算力成为AI应用的首选方式

- 大模型训练中，算力资源是「发动机」样的存在，支持大模型处理超大规模的训练数据和实现复杂的模型架构。但GPU等硬件设备的需求规模大，自建算力成本高。在这种背景下，云算力就成为了企业构建AI算力的一种主要方式。根据用户调研，有32.80%的企业主要使用算力云服务构建AI算力。

大模型训练成本计算公式

$$\text{训练成本} = \frac{\text{AI模型训练所需的浮点运算次数}}{\text{AI算力集群单位时间有效浮点计算数}} \times \text{AI算力集群单位时间价格}$$

其中：

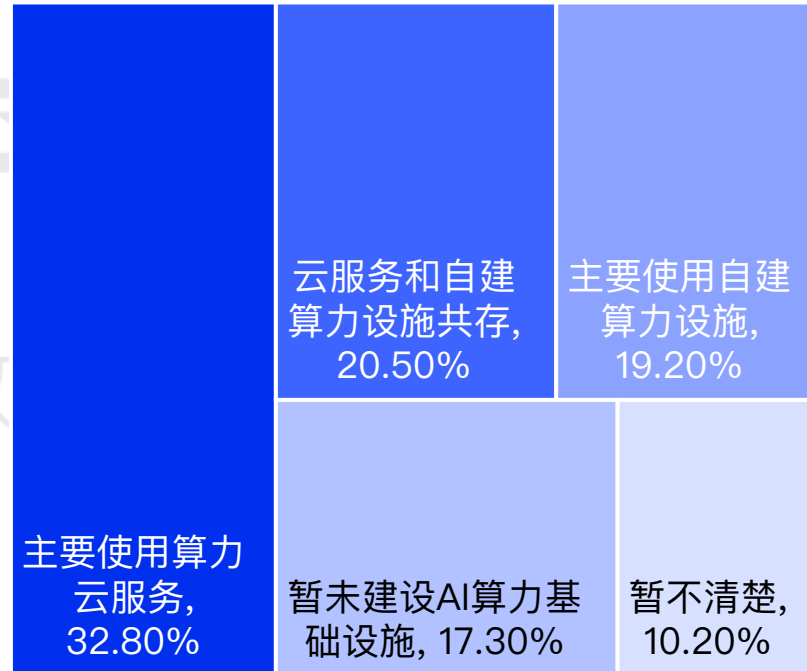
AI模型训练所需的浮点运算次数 $\approx 6 \times$ 模型参数量 \times Token总训练数

AI算力集群单位时间有效浮点计算数 = GPU数量 \times GPU峰值FLOPs \times GPU利用率

典型模型训练过程细节数据推测

模型名称	模型参数	训练Token数	训练时长	GPU使用情况
GPT-3 175B	1750亿	3000亿	14.8天	10000张V100
LlaMA-65B	650亿	14000亿	21天	2048张A100
GLM-130B	1300亿	4000亿	57天	768张A100

用户调研：企业AI算力构建方式



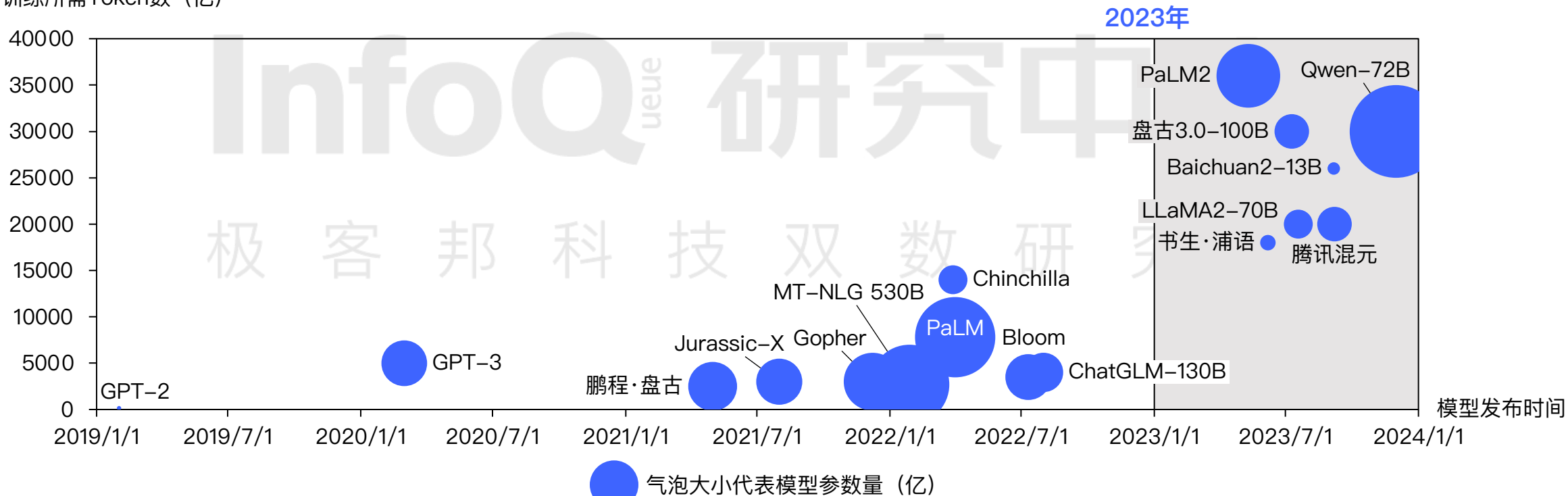
数据来源：左侧数据为InfoQ研究中心根据OneFlow、国金证券及陈巍等公开资料整理得来，训练数据根据《Carbon Emissions and Large Neural Network Training》、《LLaMA: Open and Efficient Foundation Language Models》、《GLM-130B: An Open Bilingual Pre-Trained Model》
 右侧数据来自InfoQ在2023年11-12月进行AI应用开发者调研，N=1217，Q：请问您/公司的AI算力是哪种类型？

基础设施：数据是「燃料」，模型训练数据量消耗迅速提升

- 数据大模型训练过程中不可或缺的「燃料」，其规模和质量也是影响大模型性能的核心要素。根据公开资料整理，2023年发布的模型，对模型训练的Token数量有明显增长。2023年以前，已知训练消耗Token数最多的模型为ChinChilla，训练过程中共计消耗了1.4万亿Token。而2023年5月发布的PaLM2-340B，训练过程中使用了3.6万亿的token。

2019-2023年发布大模型的模型参数与训练Token数

模型训练所需Token数（亿）



数据来源：InfoQ研究中心根据各大模型发布的论文、报告、文章、宣传资料总结整理

生态：从大模型井喷式爆发进入更多行业领域

行业大模型

游戏电商 	社交文娱 	教育 	汽车 	金融 	医疗 	其他
-----------------	-----------------	---------------	---------------	---------------	---------------	---------------

通用大模型

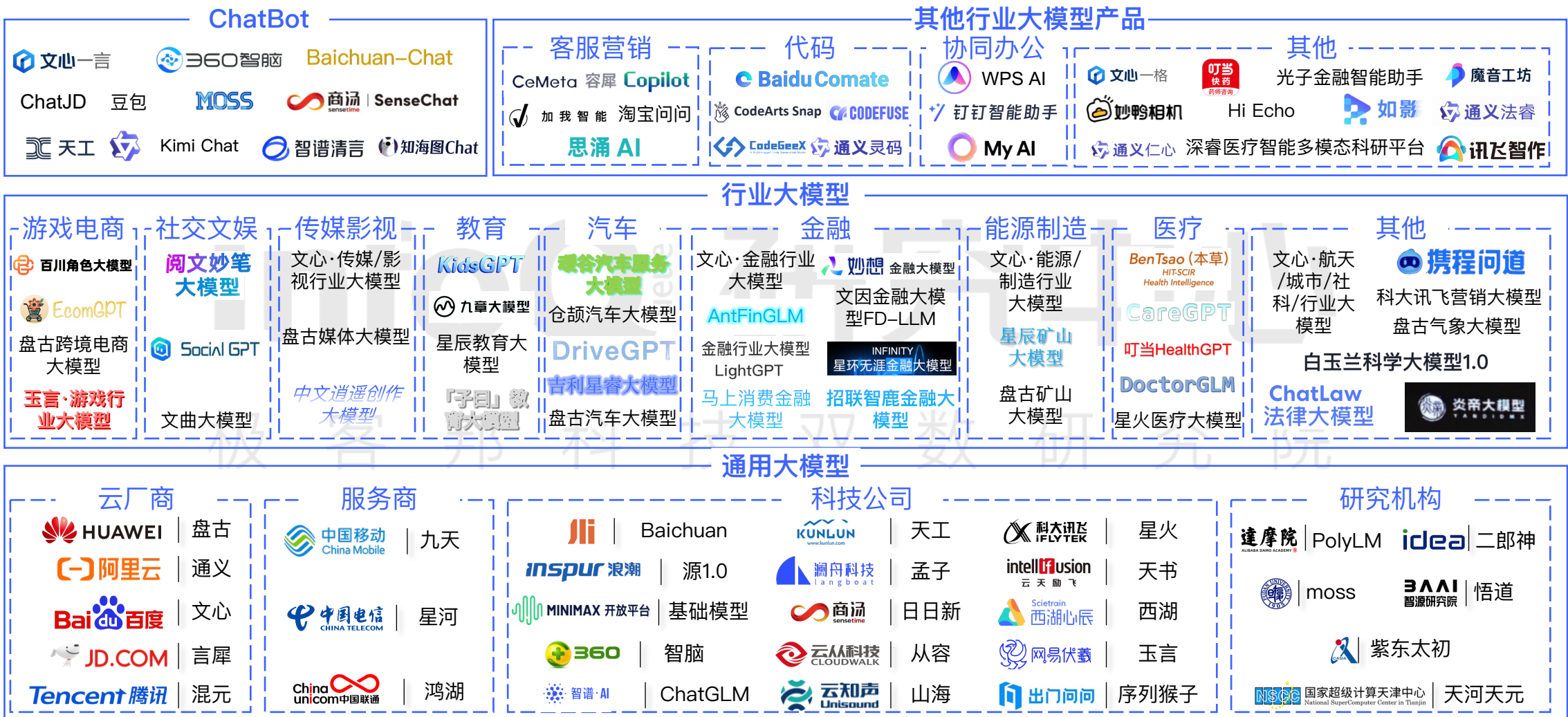
服务商 	科技公司 	研究机构
----------------	-----------------	-----------------

芯片

半定制芯片 (FPGA) 	类脑芯片 	通用芯片 (GPU) 	全定制芯片 (ASIC)
-------------------------	-----------------	-----------------------	-------------------------

数据来源：InfoQ研究中心根据厂商官网、媒体资讯等公开资料整理

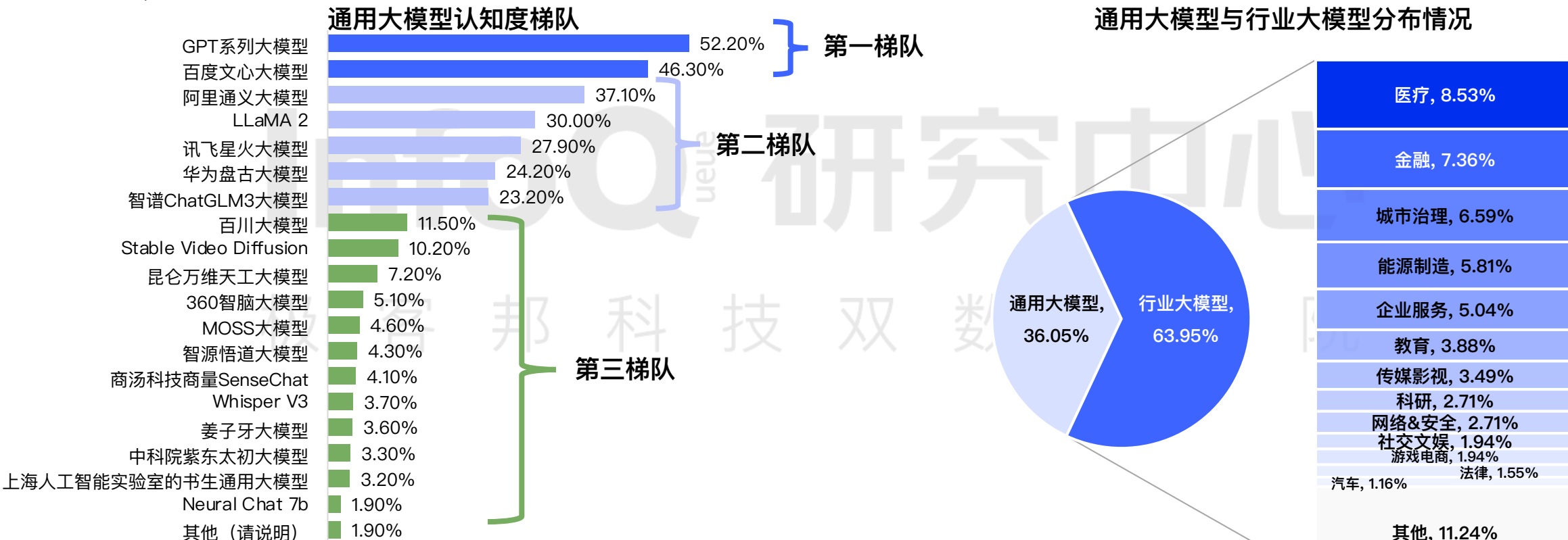
生态：模型层产品百花齐放，并逐渐开始产品探索



数据来源：InfoQ研究中心根据厂商官网、媒体资讯等公开资料整理
极客邦科技双数研究院

产品：通用大模型认知度梯队初现

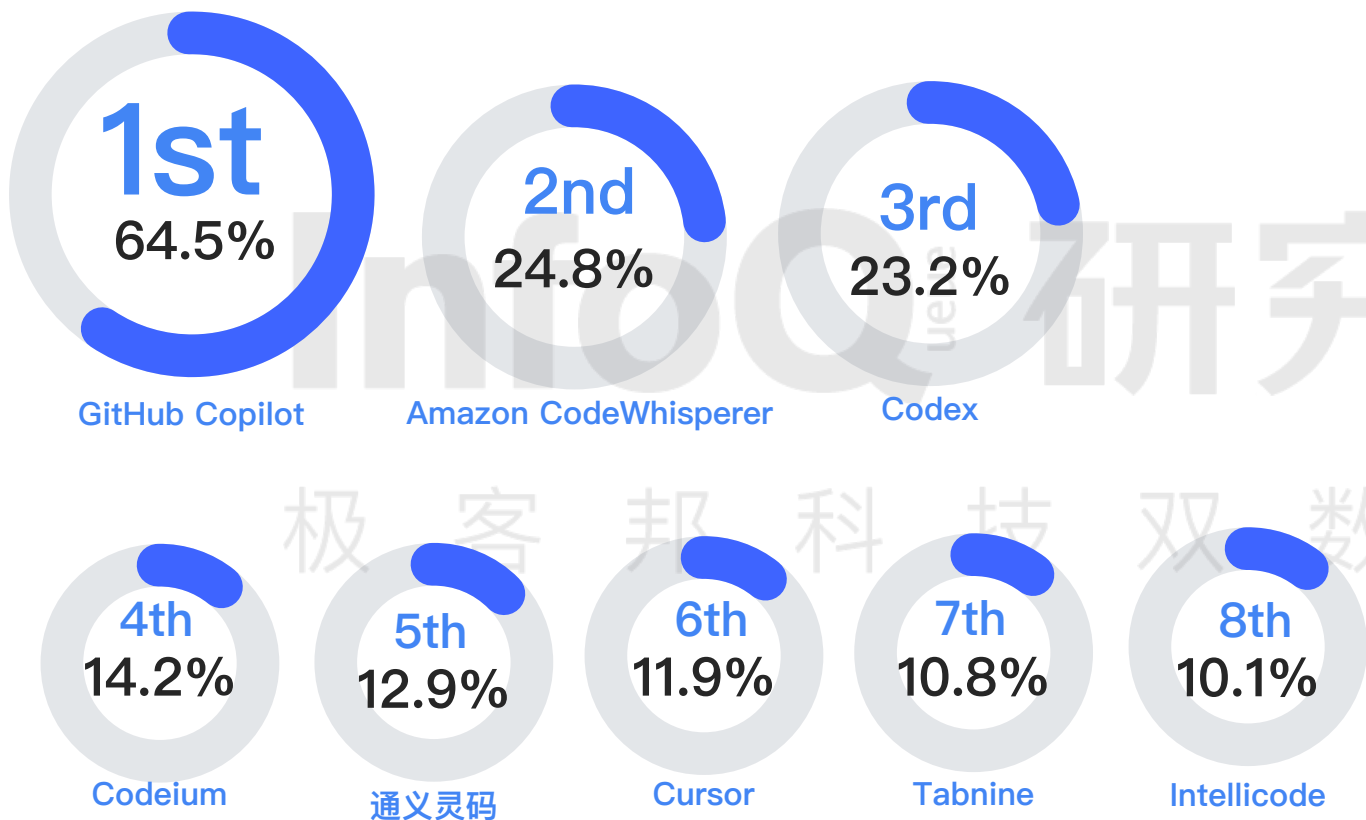
- 2023年，AI应用开发者对通用大模型的认识度梯队已经逐渐形成，根据InfoQ在2023年11-12月发起的AI应用开发者调研，GPT系列大模型以52.20%的认知度排名第一。同时，根据InfoQ对市面上大模型的不完全统计，目前中国公布的行业大模型数量已经达到165个，占到统计的258个模型的63.95%。



数据来源：认知度数据来自InfoQ在2023年11-12月进行AI应用开发者调研，N=1217
模型层数据由InfoQ根据公开资料、企业官网发布的258个大模型统计生成；其中通用大模型包括大语言模型和多模态模型；

产品：开发者智能编码助手使用习惯养成

- 离开发者最近的智能编码助手，也是大模型产品中目前探索较深，产品形态较完善的产品形态。



根据InfoQ发起的生成式人工智能开发者调研显示，智能编码助手产品的开发者使用率超过10%的产品有8款。

其中，GitHub Copilot 以显著优势位居第一，占比64.5%；Amazon CodeWhisperer 以24.8%位列第二；Codex 以23.2%的份额位于第三位。

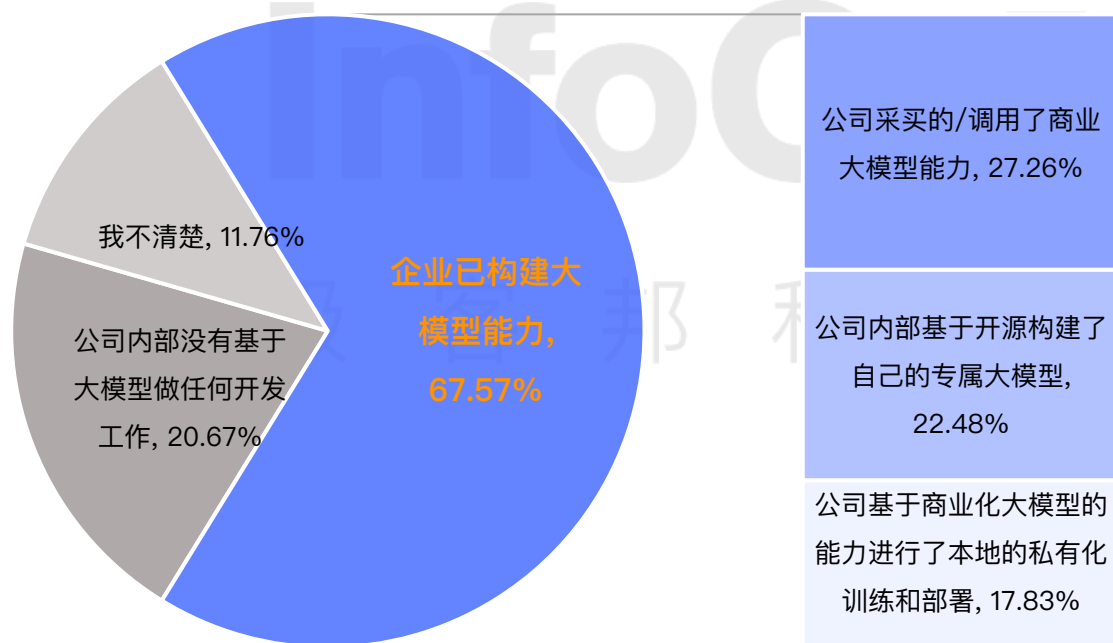
8款产品中唯一一款国内企业研发的产品为阿里旗下的通义灵码，使用率占比12.9%排名第五。

数据来源：InfoQ在2023年11-12月进行AI应用开发者调研，N=1217

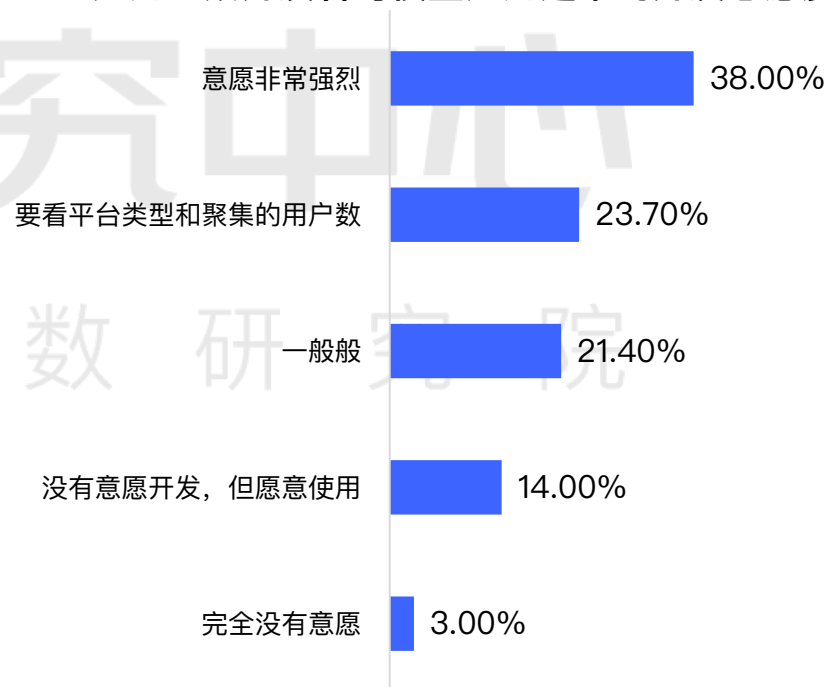
产品：大多企业已开始构建大模型能力，应用超市受开发者关注

- 67.57%的AI应用开发者所处的企业已经开始构建大模型能力，主要构建方式分为采买/调用商业大模型、基于开源构建和基于商业大模型的本地化部署。
- 此外，对于类似GPT Store这类的模型应用超市，38%的开发者的开发意愿非常强烈，但23.7%的AI应用开发者表示要看平台类型和聚集的用户数。这表明模型应用超市本身基于的模型以及聚集的用户数，对于能否吸引开发者进行应用开发至关重要。

AI应用生成企业大模型能力构建方式



AI应用生成开发者对模型应用超市的开发意愿度



数据来源：InfoQ在2023年11-12月进行AI应用开发者调研，N=1217

终端：成大模型产品流量入口，手机和电脑厂商探索AI终端新卖点



小艺小艺

2023年8月4日，华为宣布语音助手小艺已融入盘古大模型的部分能力。9月1日，华为小艺大模型版本现已开启众测推送。

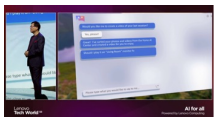


小爱同学大模型版本邀测

2023年8月14日，小米宣布小爱同学将升级大模型能力，并已经开启了大模型版本的小爱同学的邀请测试。



2023年10月11日，OPPO宣布基于AndesGPT大模型打造的OPPO 新小布 1.0 Beta版尝鲜体验正式开启。



2023年10月24日，联想宣布并展示了联想首款AI PC。2023年12月15日，联想ThinkPad X1 Carbon AI、联想小新Pro 16 AI 酷睿版两款AI PC产品正式发售。



2023年11月10日，三星宣布推出Galaxy AI，用Galaxy旗舰手机承载端侧大模型能力。



2023年11月13日，vivo X100系列正式发布，并融合vivo 蓝心大模型能力，支持10亿和70亿参数的大模型本地部署。



2023年11月30日，魅族宣布其AI助理Aicy已经实现搭载Flyme AI大模型，并发售搭载Flyme AI大模型的AR眼镜产品MYVU探索版。

2023年下半年，手机和电脑厂商开始探索将AI大模型能力搭载在终端上，期望各类终端能够成为大模型产品流量的入口。

大模型推进智能生活助理



- 总结摘要+App调用+对话式交互，三步养成自己的生活助理。
- 在通话、阅读等环境下使用内置大模型智能总结要点和代办，通过对话式交互完善生活决策，并智能调用App，帮助使用者高效快捷生活。

多模态造就图像处理新玩法



- 多模态大模型能力，让手机成为用户专属的、个人定制的照相馆。
- 照片的美化、修复、抠图、扩图、清除等，让用户快捷造就「人生照片」。

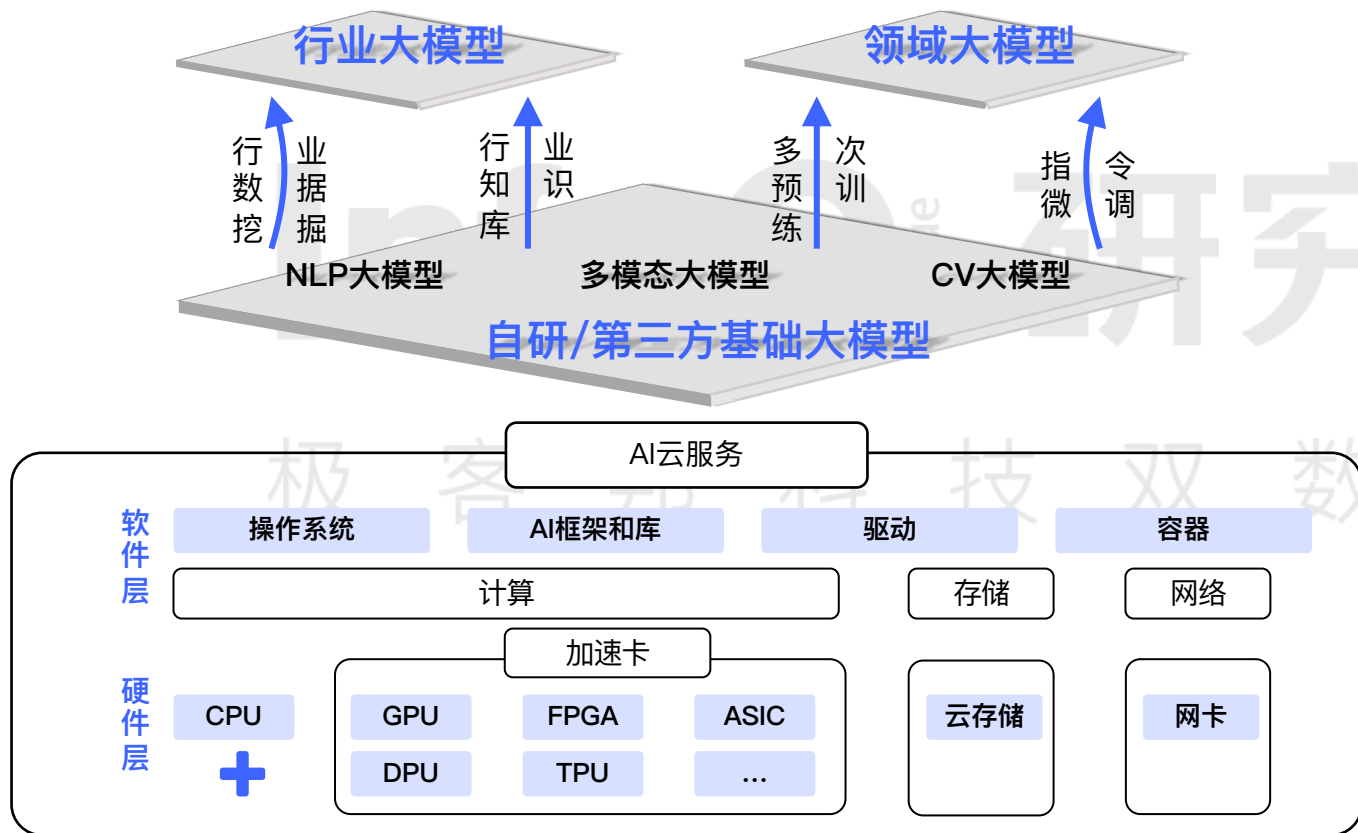
数据来源：InfoQ研究中心根据公司官网、文章、资讯等公开资料总结整理



以数强实稳根基

云计算——AI浪潮下，云厂商集中探索MaaS模式

- 大模型带来的技术浪潮，为云厂商指引了一条探索的路线。事实上，云厂商在2020年之后，便陆续开始以机器学习平台的形式，加大云计算资源与AI的联系。2023年，更多的云厂商开始探索MaaS模式。



- 云计算为AI提供能够便捷获取的底层计算与存储资源
- 云计算优化AI开发过程，帮助企业更便捷地开发和管理AI应用
- 云计算立足基础模型，优化微调过程，帮助企业实现大模型的快速落地

一站式企业级大模型平台，提供先进的生成式AI生产及应用全流程开发工具链

百度智能云千帆大模型平台

火山方舟

开放大模型生态，驱动智能化未来



期望大模型成为云计算新增量

数据来源：InfoQ研究中心于2023年9月发布的《互联网行业再进化——云上AI时代》

云计算——成本优化推动2023年云服务降价潮

- 与云厂商的降价潮相对应地是，企业对于云等IT资源投入的保守倾向。在一项用户调研中，认为公司的IT资源投入仍将保持稳定提升的企业比例下降了8.71%，但认为IT资源投入会保持不变或者会逐步减少投入的企业比例却分别增加了5.72%和5.47%。
- 这当中，尤其是对于本身利润就不高的工业制造企业来说，作为企业IT资源投入最多的云计算资源，也将受到直接的影响。

云服务年中出现降价潮

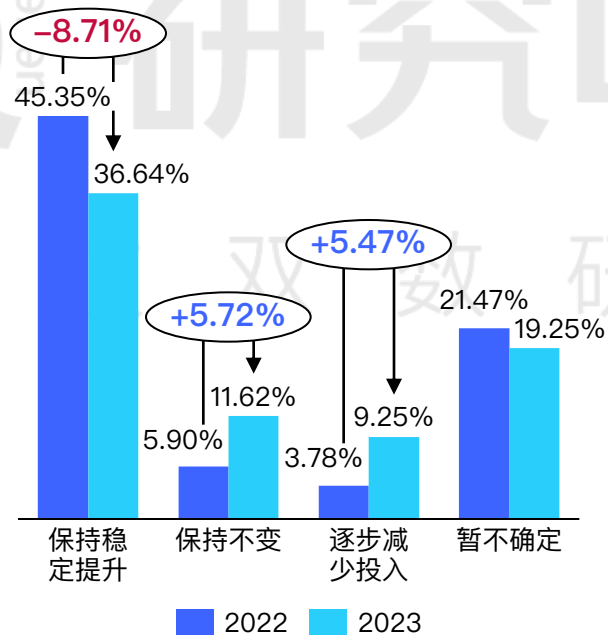
阿里云 4月26日，阿里云在其合作伙伴大会上，宣布核心产品价格全线下调15%至50%，存储产品最高降幅达50%，也被称为是“阿里云史上最大规模降价”。

腾讯云 5月16日，腾讯云宣布对多款核心云产品降价，部分产品线最高降幅达40%，降价政策将在6月1日正式生效。

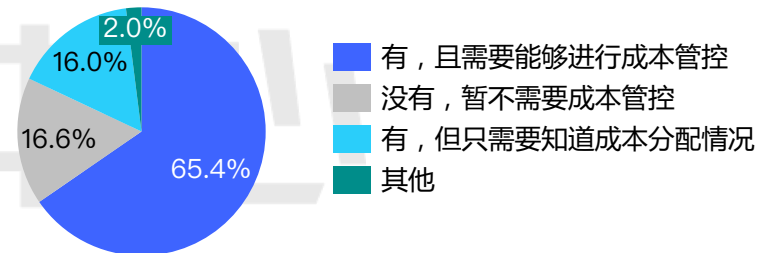
移动云 5月16日，移动云宣布其通用入门型云主机、通用网络优化型云主机均降价60%，云安全中心、云硬盘备份降价50%。

京东云 5月23日，京东云宣布“全网比价”，比价活动涉及京东云全系核心产品，包含：计算、存储、网络、数据库、中间件、安全、视频服务、云电脑。

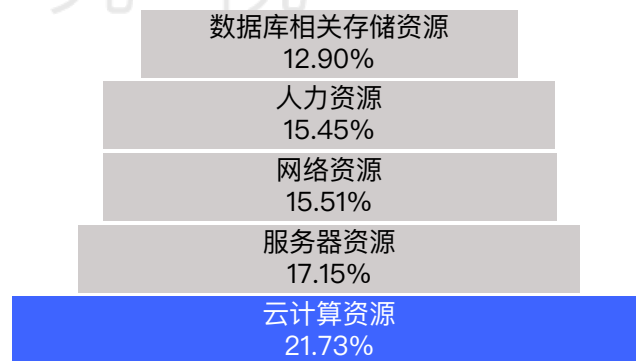
企业IT投入倾向分布



企业对FinOps的态度分布



企业IT投入金额排名占比



数据来源：信通院发起的《中国FinOps现状调查报告（2023）》、新闻报告、公司官网等公开渠道

云计算——六大环节云化加速工业行业互联进度

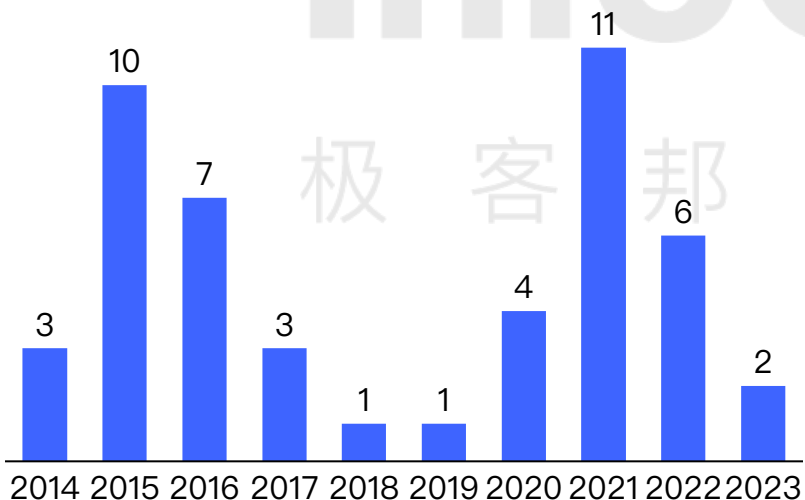
- 工业互联网双跨平台的底层本质是行业云平台，以云计算为底层，支撑研发设计、设备管理、生产管理、用户服务、经营管理和供应链管理的云化。



大数据——数据资产入表在即，数据商业模式新开端年

- 2023年8月21日，财政部正式发布《企业数据资源相关会计处理暂行规定》，并规定自2024年1月1日起施行。
- 2023年10月25日，国家数据局正式揭牌。
- 2023年12月15日，国家数据局发布《“数据要素x”三年行动计划（2024—2026年）（征求意见稿）》，计划提出到2026年底，数据要素应用场景广度和深度大幅拓展，数据产业年均增速超过20%，数据交易规模增长1倍。
- 以上种种都是数据资产，在评估、定价、交易环节流程标准化的新进展，国家、数据交易中心、企业等主体目前正在深入探索数据商业模式。

2014-2023年数据交易中心/所的成立情况

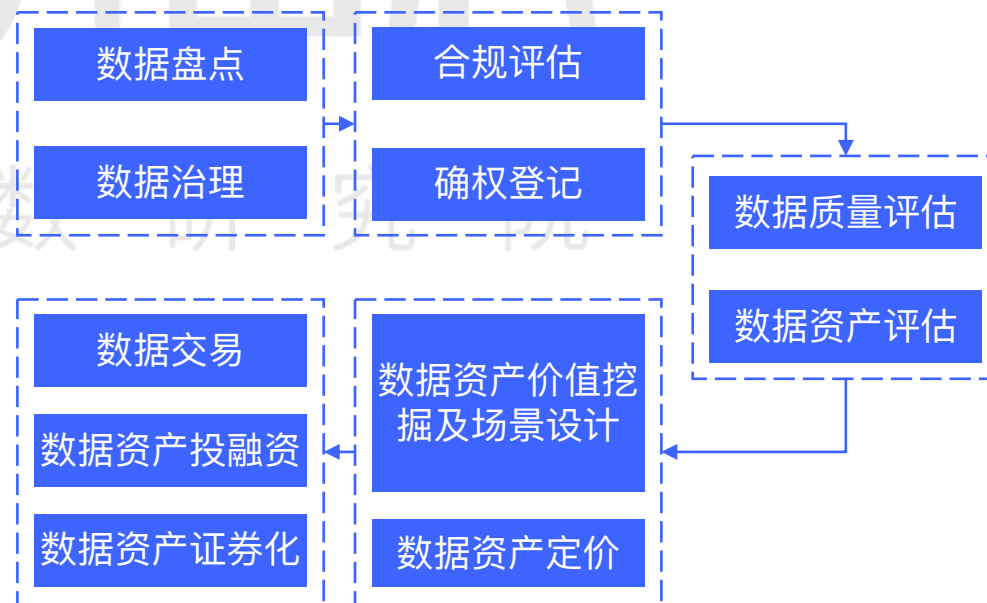


数据来源：InfoQ研究中心根据数据交易网等公开资料整理总结

工业互联网涉及的数据资产



数据资产入表及定价交易流程示意图

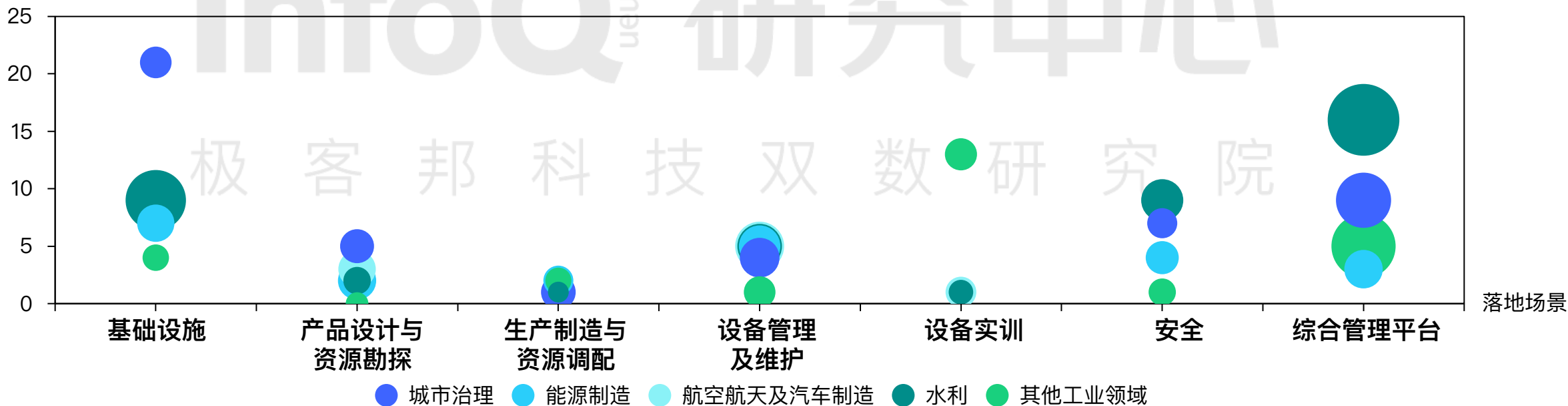


人工智能——数字孪生助力工业生产全周期提效

- 数字孪生作为工业互联网的重要技术载体之一，目前不同行业已经在产品设计与工艺规划、生产制造与资源调配、设备管理及维护、设备实训、安全等环节和场景进行了落地尝试。根据InfoQ研究中心统计的150个过百万的数字孪生中标项目，水利和城市治理在整体中标项目数量上位列前二，水利和能源制造行业基于自身安全性的高要求，数字孪生安全项目数量多，项目平均中标金额高。
- 水利行业数字孪生项目数量爆发的背后，是“构建以数字孪生流域为核心的智慧水利体系”被写进《数字中国建设整体布局规划》。
- 从中标厂商来看，数字孪生目前以设计院/规划院、运营商为主，地域属性明显。千万以上的项目通常以联合体形式投标。

中标项目数（个）

2023年1-11月数字孪生中标项目分布情况



数据来源：InfoQ研究中心根据招标公告等公开资料总结整理

气泡大小大表平均中标金额（万元）

人工智能——行业大模型研发探索赋能工业数字化进程

- 2023年下半年，“模型热”正在延伸到工业领域，工业大模型频繁发布。但工业企业逐渐发现大模型不是“万能”的，对于大模型的能力边界也有了较为清晰的认知。目前落地成本、场景数据和人才，仍然困扰着想要尝试落地大模型的工业企业。

2023年发布的工业大模型（不完全统计）

模型名称	发布机构	机构类型	发布时间
智工·工业大模型	中工互联	工业软件企业	2023年6月4日
COSMO-GPT	海尔卡奥斯	制造企业	2023年9月
Alno-15B	创新奇智	技术服务商	2023年9月1日
羚羊工业大模型 V1.5	羚羊工业互联网	互联网	2023年9月
IndustryGPT V1.0	思谋科技	技术服务商	2023年11月6日
星辰矿山大模型	中国电信	运营商	2023年11月14日

工业大模型落地三大难点

成本：企业利润水平较低，在新技术投入上谨慎。大模型落地的算力成本和部署成本都存在较大负担。

数据：工业制造场景众多，但碎片化现象明显，企业间的数据基础差距大，造成有效数据难以形成规模。

人力：既懂工业领域Know-how，又对大模型较为了解的人才较少，对大模型实际落地造成了阻碍。

数据来源：InfoQ研究中心根据公司官网、新闻资讯等公开资料整理

工业互联网——双跨平台发展路径日渐清晰

- 2023年，我国已培育50家双跨平台，其中主要平台建设/运营方分为4类：制造企业、工业软件企业、运营商和互联网企业。

双跨平台		典型平台	优势	典型应用
<p>双跨平台是跨行业跨区域工业互联网平台的简称，是国家从现存300+工业互联网平台中筛选出来的、代表工业互联网发展水平第一梯队的50家代表平台。</p> <p>双跨平台在产业链内，注重研发、设计、生产、仓储、物流、服务等重点环节的数据流通与协同。同时，双跨平台普遍会搭建供需对接平台，以增强产业链供应链韧性。</p>	泛制造	<p>制造企业</p> <p>海尔卡奥斯 中石油昆仑数智 徐工汉云等</p>	<p>依托自身制造业工业场景理解力，逐渐对外辐射服务能力</p>	<p>海尔卡奥斯依托长期家电制造的工业基因，将自身对于工厂、生产、设备、仓储管理的理解，对家电、纺织、日化等相近行业辐射。</p>
	工业软件	<p>东方国信Cloudiip 宝信软件宝联登xIn^3Plat 金蝶星域 用友精智等</p>	<p>跟随客户需求，依托软件云化，拓展以采购、营销为代表的产业链管理能力</p>	<p>用友凭借长期服务工业企业的经验，围绕ERP、MES等工业软件的协同，助力工业企业的高效采购、营销和设备管理</p>
	技术服务商	<p>运营商</p> <p>移动OnePower 电信天翼云 联通格物unilink 华为云FusionPlant</p>	<p>依托长期通讯和终端服务能力，打造5G+工业互联网能力</p>	<p>移动致力于5G+工业互联网场景的探索，通过5G+智能终端实现远程设备操作、智能仓储管理、数字化巡检。</p>
	互联网	<p>阿里supET 腾讯WeMake 百度智能云开物</p>	<p>依托云、边缘计算、AI、物联网大数据等技术服务能力优势</p>	<p>腾讯借助自身在云、AI、大数据分析、音视频交互等方面的技术积累，助力工业企业建设数字孪生工厂，视频远程操作、工业视觉检测、设备数据分析下的预测性维护等应用</p>

数据来源：InfoQ研究中心根据企业官网、双跨平台名单等公开资料整理总结

2024年技术趋势展望

技术模型：多模态生成整体迈入早期推广，数字孪生进入准成熟

2024年中国技术成熟度评估曲线

InfoQ 极客传媒



新增：6G

- 前沿技术中，新增了**6G**，根据6G推进组负责人王志勤的专访，6G目前在技术储备期，预计在2025年发布我国的6G标准化，并在2030年实现6G的商用。

阶段性变化：数字孪生与多模态生态

- 在2022年技术成熟度曲线中，**数字孪生**处于早期推广技术，2023年进入准成熟阶段。
- 多模态生成整体迈进早期推广技术，文本生成图像率先迈进准成熟技术。

数据来源：InfoQ研究中心根据技术专利数量、技术诞生时间、技术舆论指数、技术融资事件等数据计算与专家访谈得出。

1、人工智能浪潮引领基础设施新必然——算力军备竞赛继续

解决算力问题 在开发者中已经形成共识

4th

算力问题是开发者认为未来需要解决的问题排名

算力问题
36%

开发者认为算力
是未来需要解决的问题占比

中国算力发展趋势



智能算力规模持续扩大

- 工业和信息化部、中央网信办等6部门联合印发《算力基础设施高质量发展行动计划》，提出到2025年算力规模超过300EFLOPS，智能算力占比达到35%。



异构算力占比扩大

- 虽然2024年英伟达算力供应链有望阶段性改善，但是英伟达+国产算力的异构模式也将成为市场发展的弹性方案。



算力效能的进一步提升

- CPU、GPU、DPU三大芯片的协同有望得到进一步关注，从而加快验证DPU在智算中心的场景应用，整体提升智算中心海量数据的处理能力。

数据来源：InfoQ在2023年11-12月进行AI应用开发者调研，N=1217

2、人工智能浪潮引领基础设施新必然——数据库升级

驱动因素



大规模数据交互广泛应用

数据存储形式
更为边缘化

数据冷热交替
特征更为显著

要求更大的数据库**规模弹性**
和更多元的**负载处理能力**

云上数据库的**Severless**
和**可观测能力**重要性提升

数据库升级点



多模态能力成为标配

语音、图片、视频等数据
形态在大模型中日益成为
标准功能配置

向量数据处理能力成为未
来数据库能力标配



TensorDB



腾讯云向量数据库



Chatbot和AI Agent成为主流

C端大规模交互要求更为
高效的数据提取能力

技术门槛更高

要求技术使用门槛更低；
既能检索又能提示的能力
匹配更大范围应用研发

RAG将随大模型应用生态
拓展被广泛应用

Graph RAG

Anything

3、大模型激战——国内场景争夺战打响，应用试炼拉开序幕



24% 的受访开发者希望AI应用缺乏好的应用场景的问题未来能够得到解决

数据来源：InfoQ在2023年11-12月进行AI应用开发者调研，N=1217

01 多模态

- 除了文本，图片、视频等多模态大模型的应用探索如何？
- 多模态之间的转换场景探索如何？

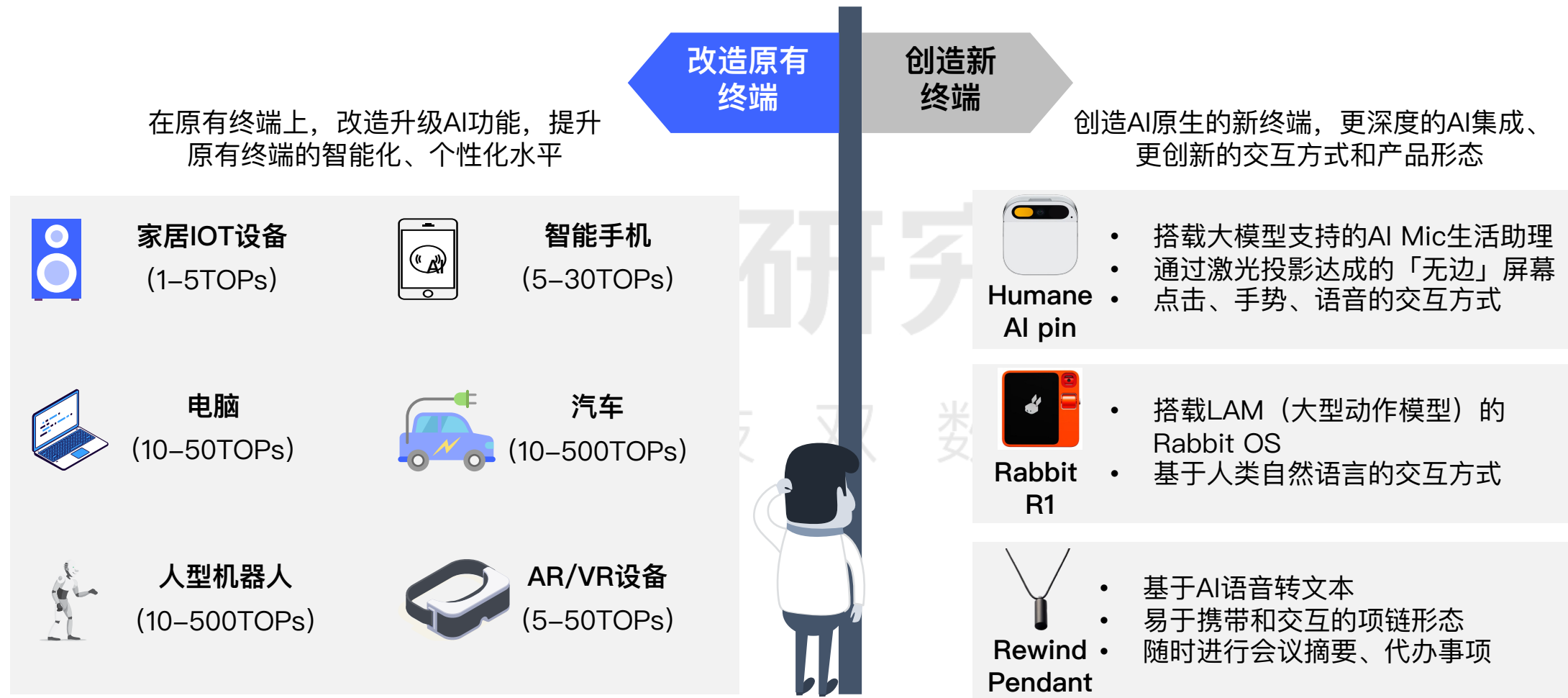
02 产品形态

- 除了对话机器人形态之外，to C端还会产生什么样的「杀手级」应用？
- To B端，大模型如何同企业各类信息系统融合？

03 应用场景

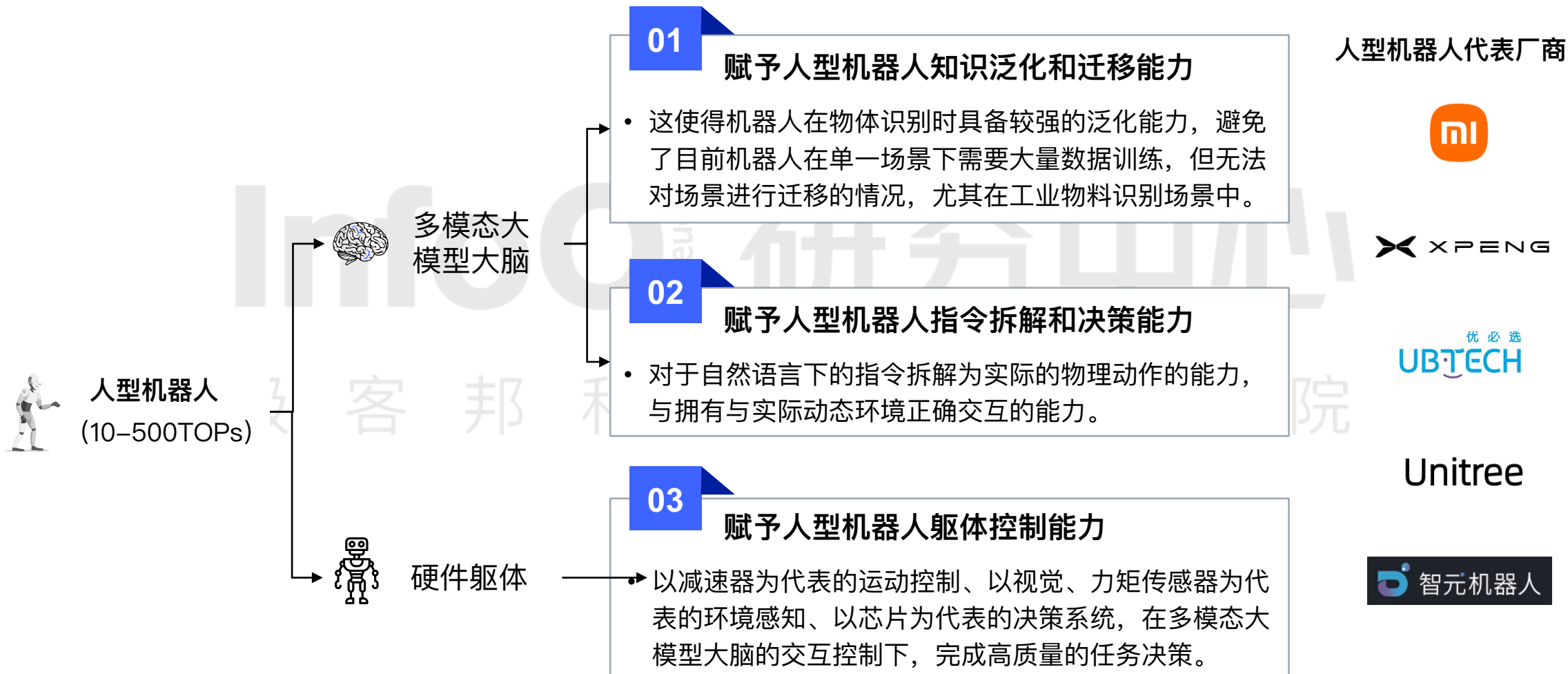
- 如何从以协同办公、客服、营销为代表的**非核心业务场景**，向以生产制造、风控、交易、销售为代表的**核心业务场景**深入？

4、端侧AI序曲——优先改造原有终端，AI原生终端崭露头角

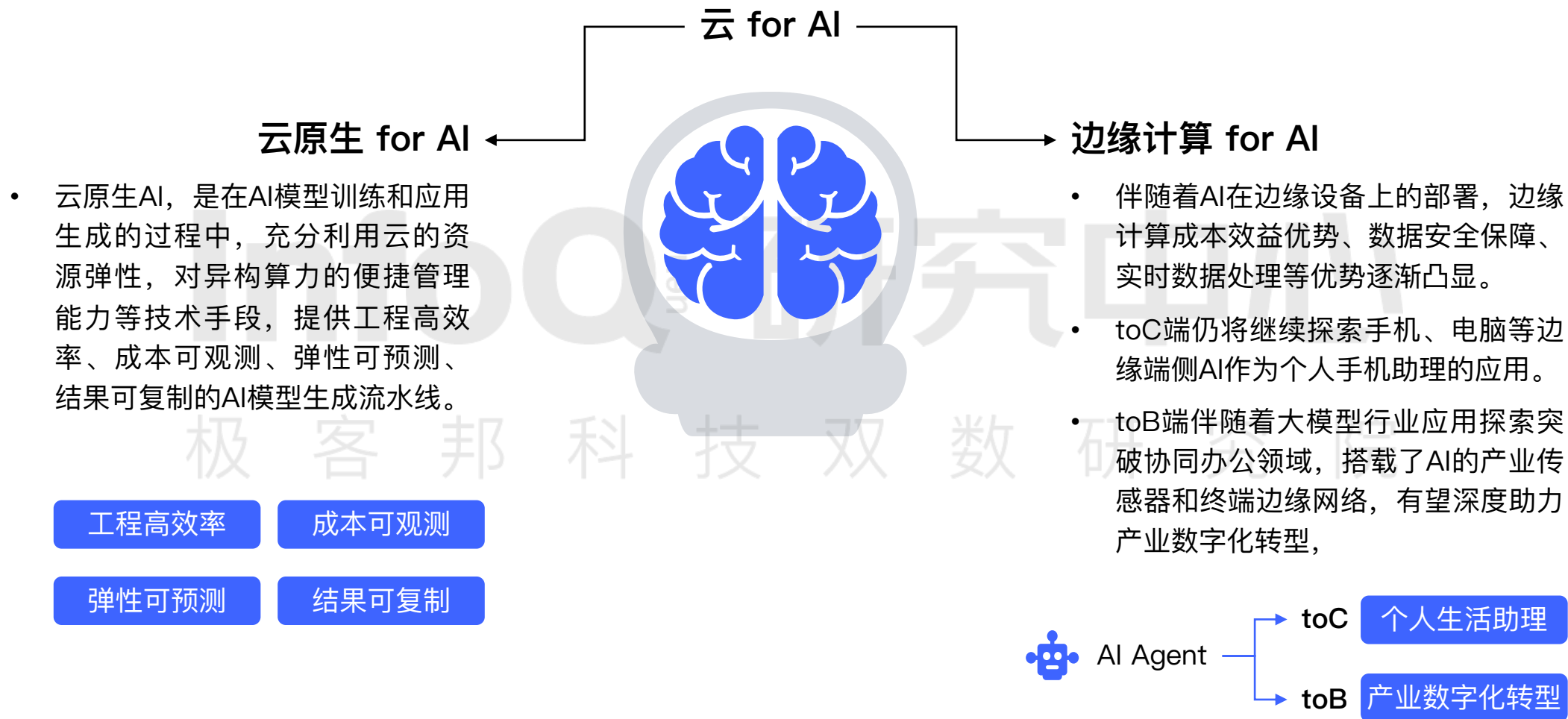


数据来源， 2023年12月 InfoQ 编辑部发起的《开发者调研》， N=1217

5、端侧AI智能大脑探索——多模态大模型与硬件躯体的协同进化



6、云 for AI 新阶段——支撑数字化 Ready for AI



7、云成本优化——云计算成本控制解决方案日益成熟



Serverless满足企业对成本的关注

- 企业对云服务成本的关注仍将保持高涨，而**Serverless**通过其按需缩放和自动扩展的特性，为企业提供了灵活性和成本效益，满足了对成本管控的需求。
- 但**Serverless**的灵活性使得预测成本变得复杂，因为其计费受到多重因素影响包括计算、存储、IO请求量、容灾等因素呈现的随机性和非平稳性。这些都会导致serverless成本估算更加复杂。因此，企业在使用时，需要特别关注这一方面。各大云厂商针对这一难题，也在不断推陈出新，如基于FinOps的用户成本中心，I/O Optimized, One Zone Service等方案层出不穷。总体来讲，云成本可观测将会成为通用serverless领域重要发展分支。

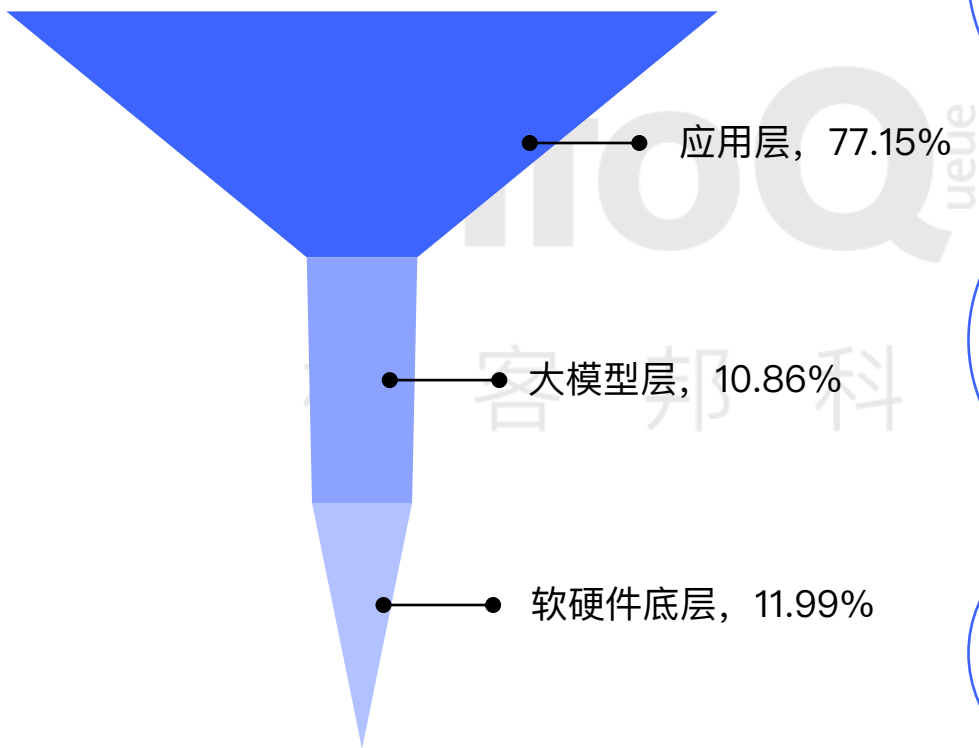


可观测性构建满足企业对稳定性的关注

- 云原生时代，东西向流量逐渐成为企业数据通路的主要载体。大量业务流量在容器、虚拟机内部进行直接交互。通过链路追踪和运行时监控等，企业能够实时分析并准确定位故障发生的位置。
- 但对于Nginx、DNS、MySQL、Redis、MQ等基础软件监控盲点往往无能为力。通过原生eBPF的应用零侵扰采集能力，则可以弥补传统虚拟网络监控的可观测盲区，为端到端的应用全景分析，调用链路追踪与回溯提供新思路。尤其对于异常敏感的网络丢包，抖动与长尾时延优化带来新的可能。
- 因此，构建**可观测性**系统在保障云服务稳定性和进行高效根因分析方面具有关键性作用。

8、AI开发者版图形成——大模型层短期聚能效应明显

AI开发者人才层级分布



数据来源：来自InfoQ在2023年11-12月进行AI应用开发者调研，N=1217。



应用层人才占比最大，核心是场景理解力

- 随着AI的成熟，以及AI应用在各个行业领域的探索与普及，应用层人才的需求逐渐旺盛。
- 对于应用层人才，核心是能否理解使用场景，并依托使用场景生成能够实际解决问题的AI应用。



大模型层核心人才吸引力强，开发者聚集化倾向明显

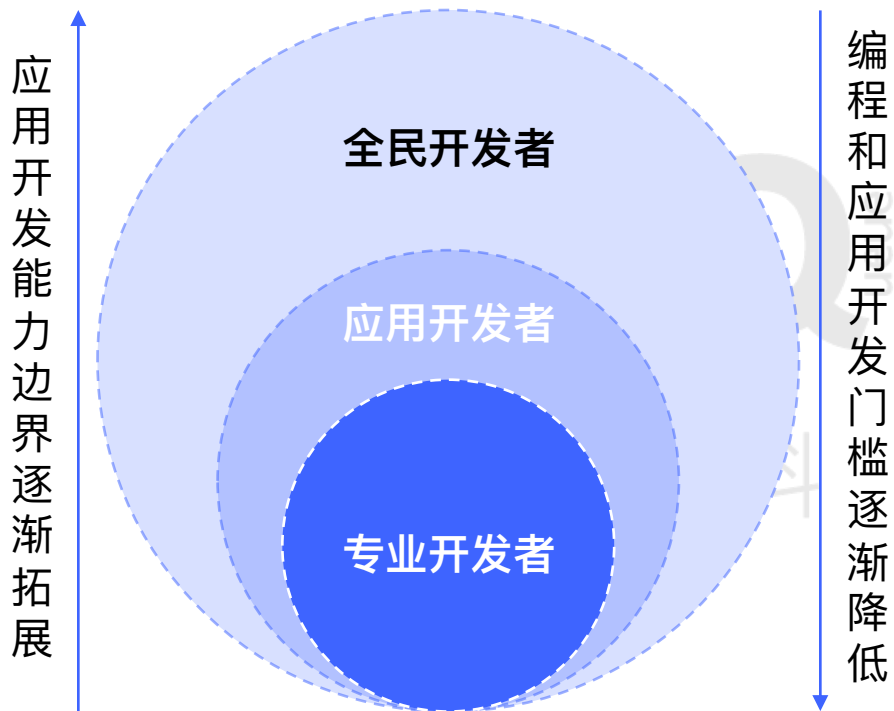
- 大模型层本身开发的门槛高，需要的资金和人才资源高，因此最终能从「百模大战」中生存下来的公司就是少数。
- 最终，拥有核心人才的团队，会吸引大量人才，并形成人才聚集。



软硬件层壁垒仍然高位，人才培养难度大

- 虽然底层软硬件对于大模型和应用层至关重要，但其专业性壁垒高，人才培养难度大，短期内人才仍不会存在大量爆发。

9、开发者概念泛化——编程与应用开发能力新天地



使用自然语言/图片进行程序开发

对于原本受限于编程能力而无法自主开发应用的人群，在此次AI浪潮中，可以通过AI提供的编程能力，参与程序开发。

- 自然语言程序开发
- 拍照生成程序开发

使用AI生成应用工具提升数据分析效率

对于工作中涉及数据分析与决策的应用开发者，在此次AI浪潮中，可以充分利用AI工具，提升数据分析效率。

- 监测指标智能推荐
- 自动化数据分析

使用AI提升编程效率和创造力

对于原本就编写代码的专业开发者，在此次AI浪潮中，可以充分利用AI工具提升编码生成和开发效率。

- AI自动化代码生成
- 自动化测试用例生成
- 代码优化智能分析
- 自动化代码审查

10、数字化转型企业类型下钻——转型成果向中小企业普及

- 以工业互联网为代表的数字化转型探索，在经历多年的应用探索后，2024年的重点是数字化转型成果向中小企业普及，夯实中小企业上云基础，培育中小企业创新动能。同时，普及过程中，双跨平台为中小企业打造产业链生态服务体系，带动中小企业转型发展。

核心主题：数字化转型成果向中小企业普及

- 双跨平台及工业互联网试点园区，会有针对性的为中小企业开发轻量化的应用，扩大工业互联网应用供给，为中小企业打造一批“工具箱”。
- 同时中小企业在研发设计、设备管理、生产管理、用户服务、经营管理和供应链管理的云化基础进一步夯实。

工业大模型应用产生典型试点

工业大模型会选择典型场景进行试点，但探索仍然以企业自主选择为主。伴随着试点进程的继续，会陆续产生一批有场景、有代表、有路径的工业大模型应用产生。

依托双跨平台打造产业服务体系

以泛制造企业和技术服务商打造的双跨平台作为载体，在支撑重点行业和中小企业的技术与场景融合探索、产业链金融与需求匹配、诊断咨询、人才培养等产业服务体系进一步完善。

应用

服务

数字孪生应用转向全局

- 数字孪生应用依托技术标准，底层数据打通，开始从单环节应用，向产品设计与资源勘探——生产制造与资源调配——设备管理与维护——设备实训——安全生产。

数据要素市场服务体系试点搭建

- 数据要素进入报表中的定价方式、确权、登记、流通交易、融资等市场服务体系逐渐出现大量试点。
- 下半年有望从政策段总结典型应用，推动试点服务体系搭建。

报告专家团 (按姓氏首字母排序)

感谢各位专家为报告编制工作提供的宝贵建议!

蔡 超 汇量科技 Mobvista 技术 VP 兼首席架构师

董 晓 聪 作业帮 基础架构 负责人

冯 嘉 分布式基础软件领域资深专家, 现任华为云中间件首席专家, 部门负责人

郭 炜 白鲸开源 CEO, Apache Software Foundation Member, TGO 鲲鹏会 (北京) 负责人

胡 宗 星 九章云极 DataCanvas 高级产品总监

蒋 林 泉 阿里云CIO&aliyun.com负责人

李 洁 北京阿哇科技 创始人

李 飞 数势科技AI负责人

刘 道 儒 星汉未来CEO

刘 东 网易杭州研究院 人工智能专家、AI 算法团队负责人

刘 松 PingCAP 副总裁

芦宇峰 (肉山) 小电科技 董事兼VP, TGO鲲鹏会 (杭州) 董事

孙百冬 中国石油昆仑数智 数字化咨询中心副主任, 油气数字化转型资深专家

王一鹏 虎嗅科技&医疗组组长

吴 晟 Tetrade 创始工程师, Apache SkyWalking创始人

宜 博 大模型农场 LLMFarm 创始人, TGO 鲲鹏会 (北京) 董事

张文举 人工智能领域专家

Fabarta技术团队

极客邦科技双数研究院

InfoQ LEUE 研究中心

InfoQ 研究中心隶属于极客邦科技双数研究院，秉承客观、深度的内容原则，追求研究扎实、观点鲜明、生态互动的目标，聚焦创新技术与科技行业，围绕数字经济观察、数字人才发展进行研究。

InfoQ 研究中心主要聚焦在前沿科技领域、数字化产业应用和数字人才三方面，旨在加速创新技术的孵化、落地与传播，服务相关产业与更广阔的市场、投资机构，C-level 人士、架构师/高阶工程师等行业观察者，为全行业架设沟通与理解的桥梁，跨越从认知到决策的信息鸿沟。

内容咨询: researchcenter@geekbang.com

商务合作: hezuo@geekbang.com

- 极客邦科技，以“推动数字人才全面发展”为己任，致力于为技术从业者提供全面的、高质量的资讯、课程、会议、培训等服务。极客邦科技的核心是独特的专家网络和优质内容生产体系，为企业、个人提供其成功所必需的技能 and 思想。
- 极客邦科技自 2007 年开展业务至今，已建设线上全球软件开发知识与创新社区 InfoQ，发起并成立技术领导者社区 TGO 鲲鹏会，连续多年举办业界知名技术峰会（如 QCon、ArchSummit 等），自主研发数字人才在线学习产品极客时间 App，以及企业级一站式数字技术学习 SaaS 平台，在技术人群、科技驱动型企业、数字化产业当中具有广泛的影响力。
- 2022年成立双数研究院，专注于数字经济观察与数字人才发展研究，原创发布了数字人才粮仓模型，以此核心整合极客邦科技专业的优质资源，通过 KaaS模式助力数字人才系统化学习进阶，以及企业数字人才体系搭建。
- 公司业务遍布中国大陆主要城市、港澳台地区，以及美国硅谷等。十余年间已经为全球千万技术人，数万家企业提供服务。



促进数字技术领域知识与创新的传播



科技领导者同侪学习社区



数字人才的移动知识库



一站式数字技术学习 SaaS 平台

洞察技术创新趋势，推动数字化商业升级



InfoQ 公众号



InfoQ 视频号

内容咨询：researchcenter@geekbang.com

商务合作：hezuo@geekbang.com
