



北京市经济和信息化局



北京市中小企业服务中心

2023 AIGC产业链 北京专精特新企业图谱



北京股权交易中心
Beijing Equity Trading Center

新京报



贝壳财经

更多免费行业报告

并购家

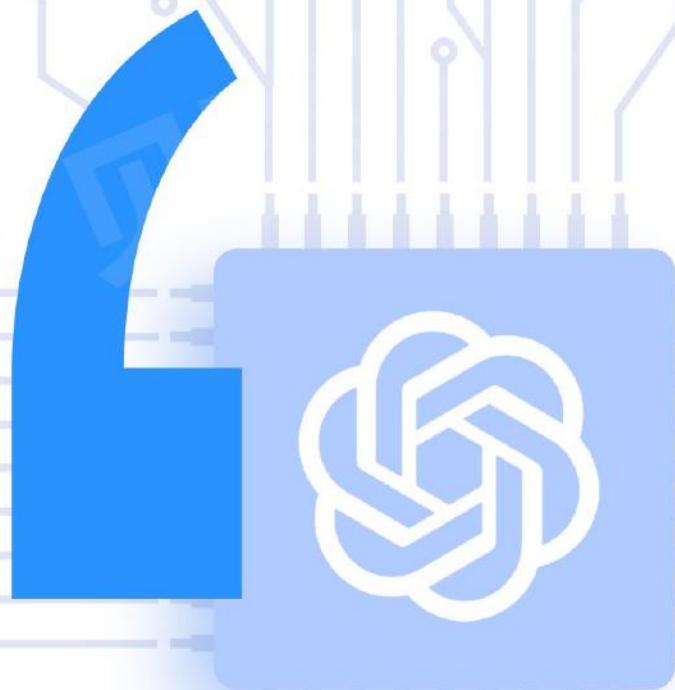
www.ipoipo.cn

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

01 AIGC行业概况

| ChatGPT成为行业发展里程碑



历史上用户数量增长最快的
消费者应用程序

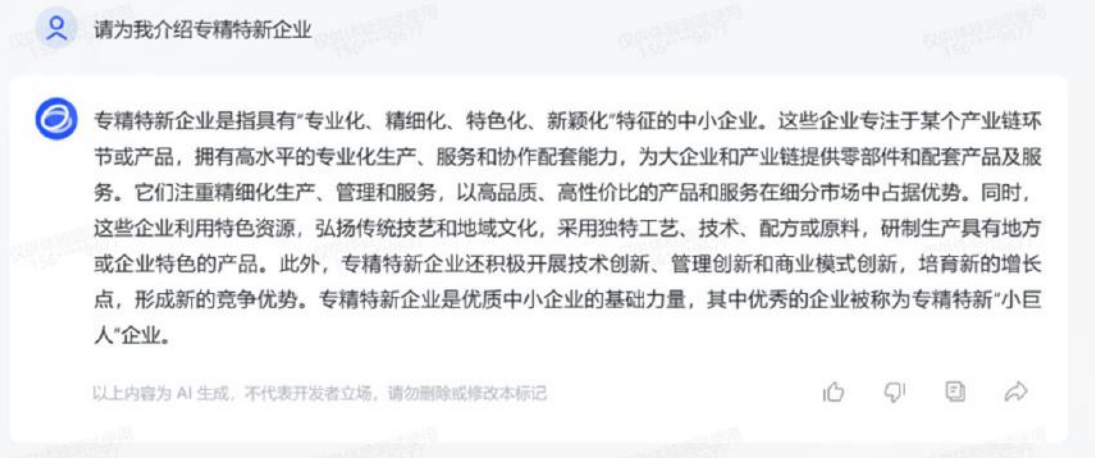
2022年11月30日，OpenAI推出对话式通用人工智能工具——ChatGPT，凭借其出色的语言理解、生成和知识推理能力在上线5天后活跃用户数就达到了100万，在2个月内活跃用户数达1个亿，成为历史上用户数量增长最快的消费者应用程序，受到各国政府、企业界、学术界的广泛关注，推动AIGC行业迈出了开拓性的一步。

| 什么是AIGC?

AIGC (AI-Generated Content)，生成式人工智能，指利用人工智能技术生成的数字内容，包括文本、图像、音视频、代码等形式，属于人工智能技术催生的新兴行业。AIGC是基于自然语言处理、生成任务算法、预训练大模型、多模态技术融合的前沿人工智能技术，通过数据训练寻找规律，利用泛化能力输出使用者需要的内容。



图：AI根据指令生成文字内容



图：AI根据指令生成图片内容



图：AI根据指令生成图片内容



| 中国AIGC行业市场规模

根据量子位智库测算，到2030年中国AIGC行业市场规模将达1.15万亿元。

11491亿元 2030



170亿元
2023

2024

2025

2026

2027

2028

2029

| AIGC行业广受政策支持



2023.05

北京市经济和信息化局发布《北京市通用人工智能产业创新伙伴计划》

汇聚产业链上下游合作伙伴，构建政产学研用深度融合的协同联动产业体系，赋能千行百业数智化转型，助力全球数字经济标杆城市建设

2023.05

北京市人民政府办公厅印发《北京市促进通用人工智能创新发展的若干措施》

着力发挥本市算力资源优势，实施算力伙伴计划，为创新主体提供多元化优质普惠算力，保障人工智能技术创新和产品研发算力需求

2023.06

北京市委、市政府印发《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》

提出培育人工智能生成内容产业发展，发展人工智能生成语音、图像和自然语言等内容，丰富合成数据供给

2023.07

国家7部门印发《生成式人工智能服务管理暂行办法》

国家**首个**面向生成式人工智能行业的规范性文件政策

2023.07

北京市海淀区人民政府印发《关于加快中关村科学城人工智能大模型创新发展的若干措施》

提出强化布局大模型软硬件技术体系；建设公共算力，降低大模型研发成本；探索数据利益共享机制；推动大模型赋能实体经济，加快产业重塑

| 开发ChatGPT具体步骤拆解

01.数据收集

收集大量自然语言文本数据，如书籍、网络信息等。这些数据将用于训练ChatGPT的语言模型

02.数据清洗

对收集到的数据进行清洗，去除掉无用的信息

03.数据标注

对数据进行标注，以便训练模型时能够使用。例如，可以为每个单词标注词性、命名实体等

04.模型开发

设计ChatGPT的模型架构，包括编码器、解码器等。ChatGPT使用的是Transformer模型，是一种基于自注意力机制的深度神经网络模型

05.模型预训练

在大量文本数据上进行模型预训练，但没有特定的任务目标，目的是让模型学会理解语言的结构、语义和一些基本的知识。训练过程中需要使用大量的算力资源，如搭载GPU的AI服务器。训练时间长达几天甚至几周

06.模型微调

微调的目的是使预训练模型适应特定任务，例如情感分析、机器翻译等。在微调期间，模型将一小部分标记好的训练数据（例如，带有情感标签的文本、源语言和目標语言的对应文本等）作为输入，并利用这些数据优化模型参数

07.数据评估

对训练好的模型进行评估，检查其性能是否达到预期。可以使用一些指标，如准确率、召回率等来评估模型的性能

08.模型交互

将训练好的模型部署到服务器上，使用户可以与模型进行交互

| AIGC重点发展行业

自然语言处理和 计算机视觉

用于文本生成、机器翻译、图像识别与生成、视频分析等方面。

自动驾驶

用于车辆感知、路径规划和驾驶决策等方面，提升自动驾驶车辆的安全性和行驶效率。

生物医药

通过模拟和预测药物分子的结构和性质，提高药物研发的效率和成功率。

金融行业

基于市场数据和经济指标，生成投资组合策略，优化风险和回报的平衡，提供金融投资意见。

AIGC产业链 北京市专精特新企业情况

| AIGC产业链分层

应用层

包括各种使用AIGC技术的应用程序和产品，例如，聊天机器人、智能客服、智能证件照等，是AIGC行业的商业化落地途径



数据层

AIGC产业链的基础层级，包括用于训练和微调模型的各种文本、图片和音视频数据集，为模型提供深度学习的“养料”



模型层

AIGC产业链的核心层级，例如用于自然语言处理和生成的Transformer模型，通过学习数据层中的数据知识，具备了理解和生成自然语言的能力，是实现AIGC功能的关键



算力层

为AIGC产业链提供底层硬件算力支撑，包括用于训练和部署模型的AI芯片、AI服务器等计算设备，使模型能快速训练和优化

| 北京AI企业数量居全国各省市第一

2200家

全国人工智能企业

占比 **28%**
618家

北京人工智能企业

数据来源：中国新一代人工智能发展战略研究院发布的
《2023中国新一代人工智能科技产业发展报告》

更多免费行业报告

并购家

www.ipoiipo.cn

| 北京有57家专精特新企业处在AIGC产业链

算力层

数据层

模型层

应用层

14家

11家

10家

22家

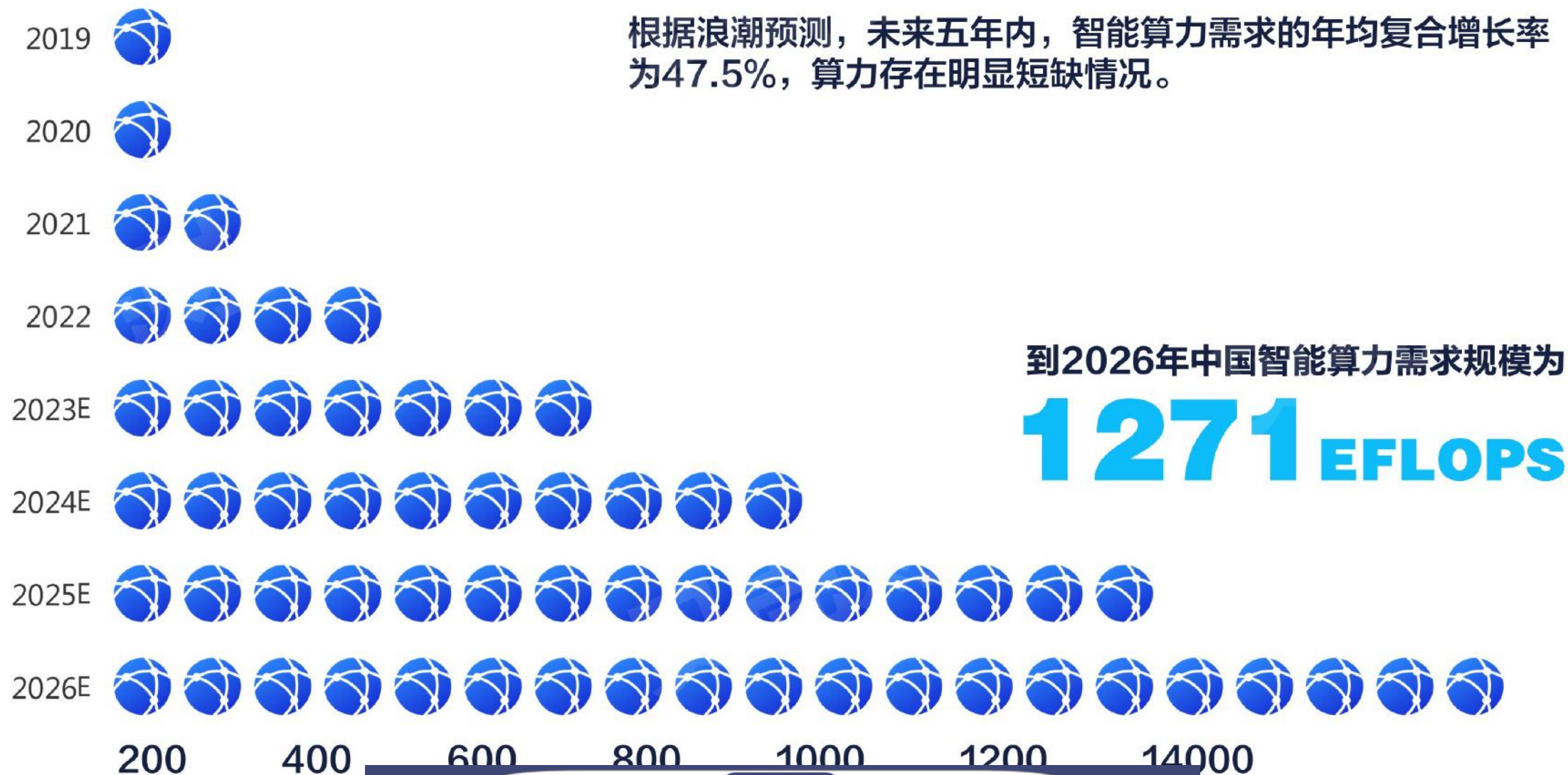
数据来源：北京市经济和信息化局

| 算力层：价值占比最高

算力占据AIGC产业链约65%的成本，数据及模型算法占据剩下35%的成本。



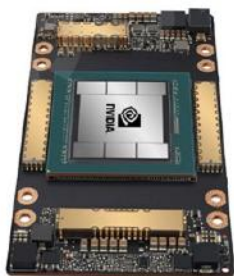
| 算力层：算力存在明显短缺情况



| 算力层：AI芯片领域被“卡”

AI芯片是算力硬件中的核心，价值占据算力成本的50%以上，目前以GPU为主。其中高算力GPU市场基本被美国英伟达公司垄断，而美国政府禁止英伟达向中国出口其尖端人工智能芯片，涉及A100、H100等型号，国产GPU在性能和良率方面存在较大的差距，因此我国AI芯片领域存在严重的“卡脖子”问题。

被卡



X 8个

组成



组成



8个英伟达A100
GPU作为核心部件

多个英伟达
DGX A100 AI
服务器作为核心
部件

AI数据中心机柜

| 算力层：北京专精特新企业情况

在AIGC行业的算力层，有14家北京市专精特新企业深耕领域，具有较强竞争实力，他们平均每家拥有发明专利16个，平均深耕行业8年以上，平均营业收入超1.8亿元。

14家企业

16个专利

平均营收

1.8亿元

| 算力层：有代表性的北京专精特新企业



摩尔线程

由前英伟达全球副总裁张建中创立，国内为数不多能够覆盖GPU研发设计、生产制造、市场销售、服务支持等完整流程的创业公司，并于2022年发布首颗国产全功能GPU，能够为AI大模型提供强大计算加速能力。

截至目前，摩尔线程估值超150亿元，迈入独角兽行列。



中科驭数

国内DPU芯片领域唯一拥有自研核心架构能力的企业，被评为中国信科“潜在独角兽”。

DPU是继CPU、GPU之后，数据中心场景中的第三颗重要的算力芯片。

中科驭数在芯片领域有近二十年的技术积累，至今已进行三代DPU芯片的研发迭代。



清微智能

核心团队来自清华大学，公司专注于可重构计算（CGRA）芯片的研发和创新，为AI训练、推理，NLP，自动驾驶，生物计算等通用计算场景提供高性能算力支持。

公司曾获国家技术发明奖、中国专利金奖、科学技术奖技术发明一等奖等多个奖项。

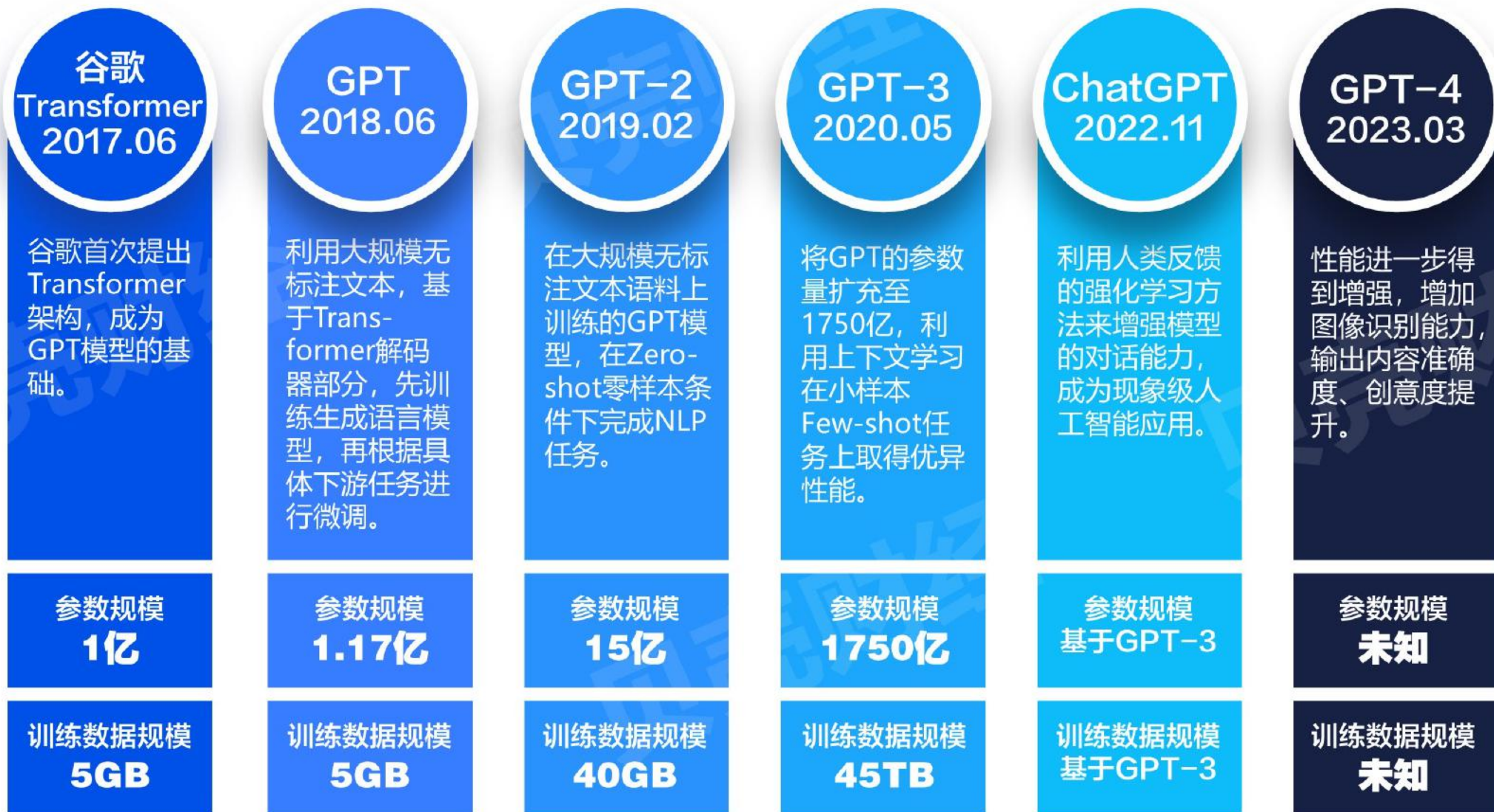


算能科技

国内领先的通用算力提供商，专注AI、RISC-V CPU等算力产品的研发和推广应用，为智算中心、智能制造、AIGC等应用场景提供算力产品及整体解决方案。

自2016年以来，旗下品牌算丰 SOPHON 系列产品已完成多次迭代，每代产品相较于前代产品均实现能耗比倍数级提升。

| 数据层：数据是AIGC行业的“能源”



| 数据层：数据集环节存在“堵点”

斯坦福大学吴恩达教授提出
二八定律：80%的数据
+20%的模型=更好的AI

当前，主流数据集多以英文为主，中文数据集紧缺。根据360分析，中文语料数据集占比低于5%。此外，中文预料数据集质量差，数据集种类和内容缺乏多元化，专用领域数据集基本空白，在政府、医学、金融等公共领域的开源数据稀缺。对比之下，美国拥有大量的文本、目标检测数据集。因此，中国AIGC产业链的数据集环节存在“堵点”。

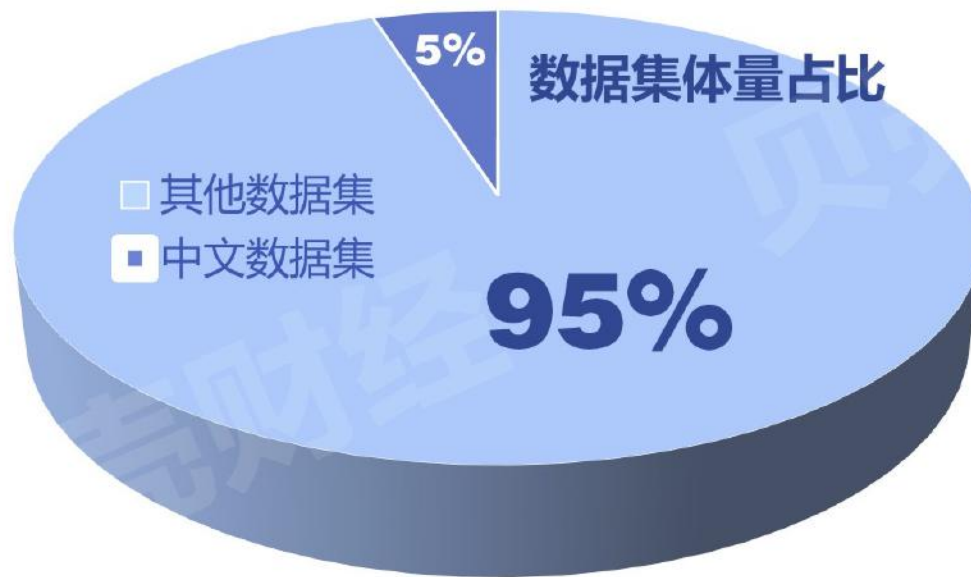


80%的数据

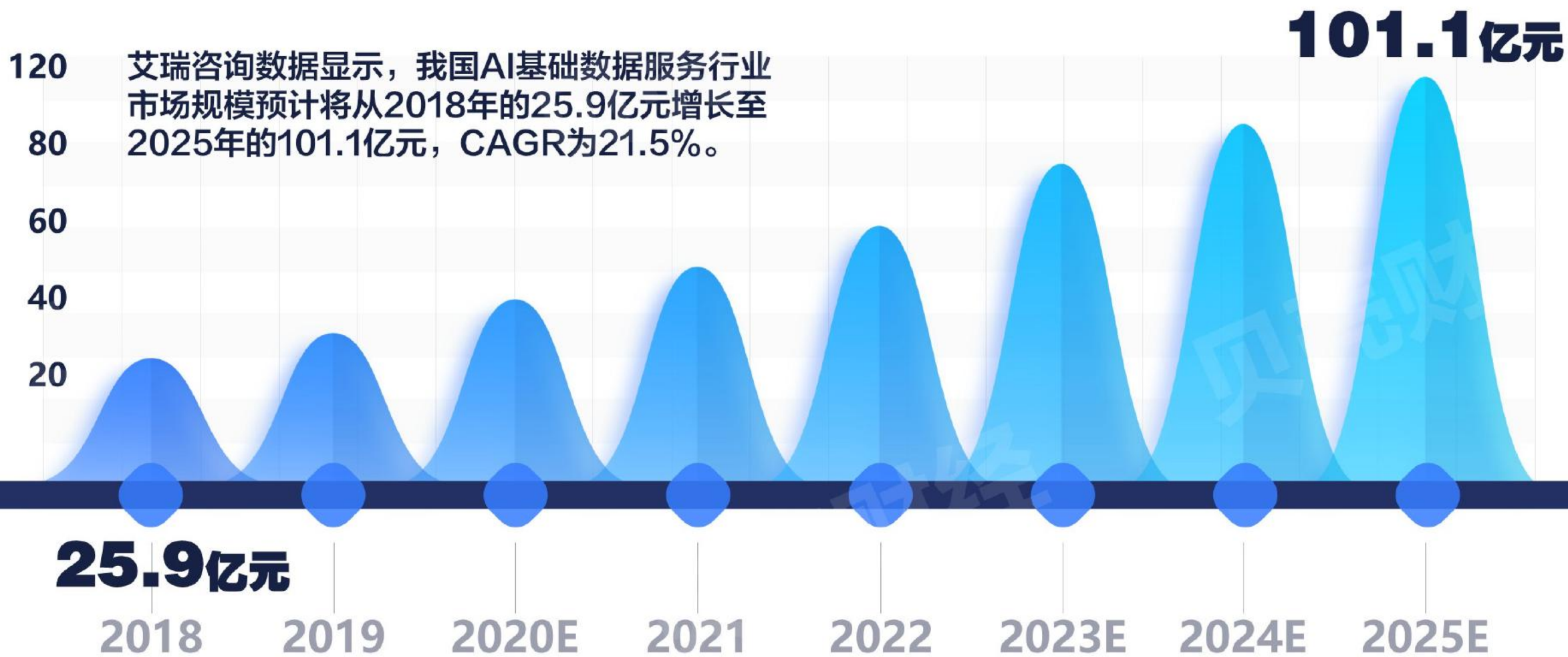


更好的AI

20%的模型



| 数据层：我国AI数据服务市场规模持续扩容



| 数据层：北京专精特新企业情况

在AIGC行业的数据服务领域，有11家北京市专精特新企业处于行业领先地位，平均研发强度超20%（研发投入/营业收入），营收增长率超30%，平均每家企业软件著作权数量超50个。

平均研发强度

20%

平均营收增长率

60%

平均软著数量

50↑

| 数据层：有代表性的北京专精特新企业



海天瑞声

我国最早从事AI训练数据解决方案提供商之一，向全行业提供多语言、跨领域、跨模态的人工智能数据及相关数据服务，覆盖全球近200个主要语种及方言，于2021年8月在科创板上市



数据堂

现拥有全球最大规模的AI版权数据库，建立了以AI版权数据、AI数据生产工厂、AI数据评测系统为主导产品的服务体系，覆盖了智能语音、生物认证、无人驾驶、智能安防等领域3.5PB规模的版权数据库



晴数智慧

在传统数据采集标注基础上，在数据生成、清洗、分类、微调、测试等环节，提供专业的DaaS（Data as a Service）解决方案，打造高质量数据集解决方案公司



星尘数据

拥有行业内自动化水平最高的标注平台Rosetta平台，目前平台年处理数据量数亿，自动化水平达到60%以上，数据质量达到99.9%，达到国际领先水平

| 模型层：数量上，北京全国领先

据不完全统计，全国范围已发布124个大模型，整体呈现“百模大战”状态，北京市相关企业和科研单位已累计发布54个大模型，数量在全国各省市中排名第一，占全国总数的44%。



| 模型层：北京专精特新企业情况

在AIGC行业的大模型开发领域，有10家北京市专精特新企业，他们过去一年的平均营业收入超7600万元，平均研发投入约为5800万元，平均超60%的员工为研发人员。

10家企业

研发员工

60%

平均营收

7600万元

| 模型层：有代表性的北京专精特新企业



智谱华章

在国内首家推出开源大模型 ChatGLM-6B，截至目前在全球的下载量已超过百万

同时研发双语千亿级超大规模预训练模型 GLM-130B，构建高精度通用知识图谱，基于此模型打造了对话式产品 ChatGLM，部分能力超过 ChatGPT



澜舟科技

公司于2023年初发布孟子MChat可控大模型，支持BERT（分析类）、GPT（生成类）和T5（翻译类）等不同的模型架构，能够覆盖阅读理解、文本生成、机器翻译、对话等不同场景，并可根据行业和需求做出快速调整，在面向垂直领域、专业赛道等应用场景时有显著优势



第四范式

拥有输入输出多模态、知识库、Copilot、思维链等核心技术，可实现私有化部署，降低企业算力成本，并可实现信息输出时定位原始出处，做到“有据可查”，随着用户的反馈“知错能改”。为企业使用大模型提供数据安全、内容可信、成本可控的三大保障

| AIGC共分为六类应用

AI大模型的出现对各行各业都带来深远影响，由AI生产的内容将逐渐成为众多行业中不可或缺的一部分。基于AI大模型的模态分类，AIGC应用可被划分六类。

其中，文字生成应用发展时间较长、易于落地，而跨模态生成赛道的发展潜力最高。



| 应用层：北京专精特新企业情况

在AIGC行业的应用层，有22家北京市专精特新企业积极探索AI大模型在各行各业的应用，他们平均营业收入超1亿元，平均研发投入超3000万元。

22家企业

1亿元营收

研发投入
3000万元

| 应用层：有代表性的北京专精特新企业

AI X coder

硅心科技

专注于AIGC编程领域的创业公司，由北京大学软件研究所主持孵化，国际最早开展基于深度学习的代码生成与代码理解研究的科研团队，开发了国内首款基于代码大模型的智能编程系统

百炼智能
BAILIAN.AI

百炼智能

国内B2B智能营销的先行者，提供数据驱动、深耕场景的B2B智能营销SaaS产品及数字化解决方案，致力于通过AI技术帮助企业提升从市场洞察到销售线索开发、筛选、联系、转化的营销全流程效率。

2023年3月公司推出全面覆盖营销人员实际工作场景的营销大模型“爱迪生”

感易智能
Senseddeal.ai

感易智能

在数据智能领域的算法、产品及系统架构研发经验平均超过10年。

公司围绕多模态深度语义理解、图计算和大数据等相关技术创新，研发出金融信息认知大模型，依托AI技术为金融机构提供智能服务

AIGC产业链 北京市专精特新企业图谱

AIGC产业链北京市专精特新企业图谱



北京股权交易中心
Beijing Equity Trading Center

新京报



贝壳财经

更多免费行业报告

并购家

www.ipoiipo.cn