

中国人工智能基础层行业发展 研究报告

2021年



海量行研报告免费读



- **人工智能基础层定义**：算力、算法、数据是人工智能产业发展的三大要素。基于此，艾瑞定义人工智能基础层是支撑AI应用模型开发及落地的必要资源，主要包括**智能计算集群、智能模型敏捷开发工具、数据基础服务与治理平台**三大模块。
- **人工智能基础层价值**：①AI基础层可多环节提效AI技术价值的释放，解决需求方人工智能生产力稀缺问题；②依托AI基础层资源，AI企业可有效应对下游客户的长尾应用需求，将其高频应用转化为新主营业务，寻找业务增长突破点；③基础层工具属性标志着AI产业社会化分工的出现，AI产业正逐步进入各产业深度参与、双向共建的效率化生产阶段。



- **人工智能基础层市场规模**：据艾瑞测算，2020年中国人工智能核心产业规模超过1500亿元；其中**人工智能基础层市场规模为497亿元**，为AI产业总规模的33%，AI芯片的高增长是产业规模增长的主要拉动力。2021-2025年，中国人工智能基础层市场规模CAGR为38%，整体产业规模发展速度较快、空间较为广阔，于2025年将达到2475亿元。



- **智能计算集群**：提供支撑AI模型开发、训练或推理的算力资源，包括系统级AI芯片和异构智能计算服务器，以及下游的人工智能计算中心等
- **智能模型敏捷开发工具**：包括开源算法框架、提供AI技术能力调用的AI开放平台和AI应用模型效率化生产平台。API规模经济与效率化生产平台杠杆增效共拓产业广度与深度。
- **数据基础服务与治理平台**：实现AI应用所需的数据资源生产与治理，提供AI基础数据服务及面向AI的数据治理平台



- **人工智能基础层资源向集约型发展**：伴随各行业智能化转型的迫切需求，艾瑞认为人工智能基础层发展总体向好。在人工智能由技术落地应用阶段向效率化生产阶段转变的背景下，人工智能基础层的各模块工具有望走向**集约型的生产模式**。
- **人工智能基础层资源自主可控展望**：基础层全栈的自主可控建设目前还处在萌芽阶段。未来或将有更多企业自研开源框架，国产的操作系统与数据库等软件配套设施将稳步崛起，算力模块的智能服务器国产化率也将逐步提升。

| | |
|---------------|---|
| 人工智能基础层概念界定 | 1 |
| 人工智能基础层需求篇 | 2 |
| 人工智能基础层供给篇 | 3 |
| 典型人工智能基础层企业案例 | 4 |
| 人工智能基础层发展洞察 | 5 |

人工智能基础层定义

支撑各类人工智能应用开发与运行的资源和平台

算力、算法、数据是人工智能产业发展的三大要素。据此，人工智能基础层主要包括智能计算集群、智能模型敏捷开发工具、数据基础服务与治理平台三个模块。**智能计算集群**提供支撑AI模型开发、训练或推理的算力资源，包括系统级AI芯片和异构智能计算服务器，以及下游的人工智能计算中心等；**智能模型敏捷开发工具**模块主要实现AI应用模型的生产，包括开源算法框架，提供语音、图像等AI技术能力调用的AI开放平台和AI应用模型效率化生产平台；**数据基础服务与治理平台**模块则实现AI应用所需的数据资源生产与治理，提供AI基础数据服务及面向AI的数据治理平台。AI基础层企业通过提供AI算力、开发工具或数据资源助力人工智能应用在各行业领域、各应用场景落地，支撑人工智能产业健康稳定发展。

人工智能基础层涵盖AI算力、算法平台与数据资源模块



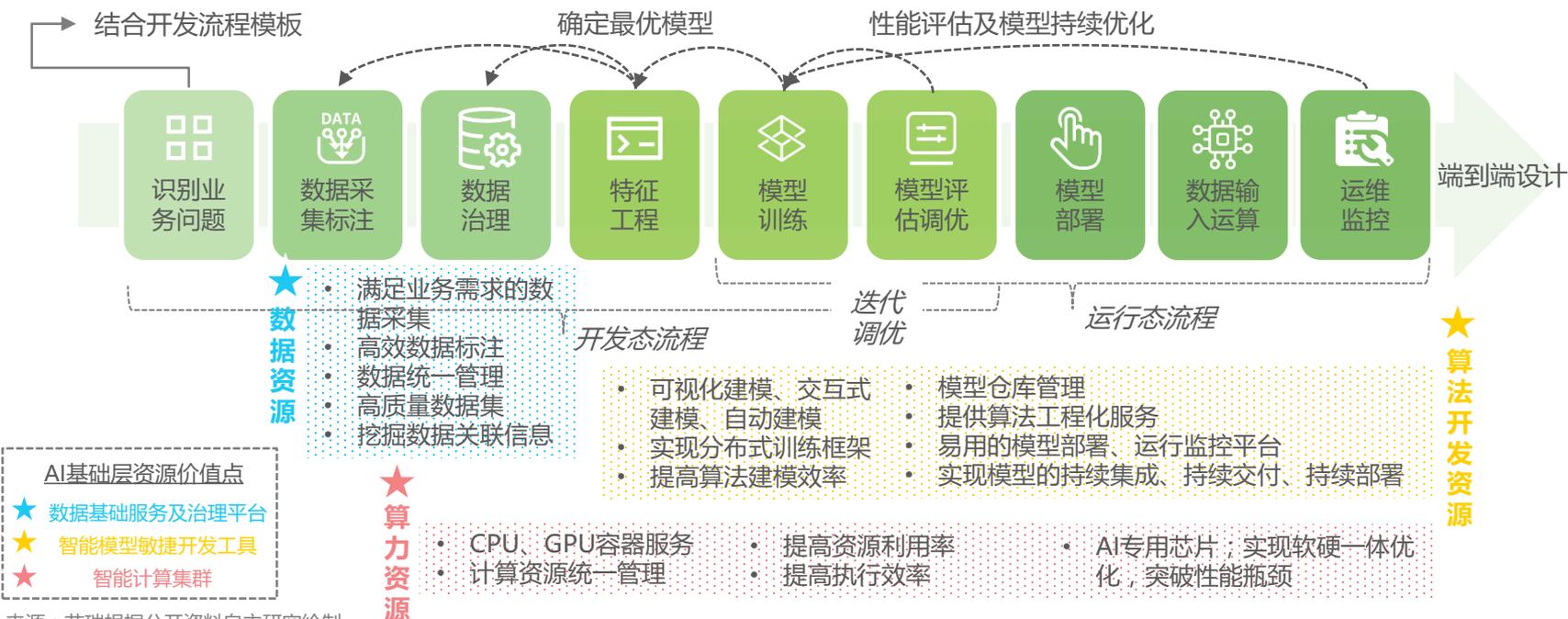
来源：艾瑞根据公开资料自主研究绘制。

人工智能基础层价值

AI基础层是支撑AI应用模型开发及落地的必要资源

开发一项人工智能模型并上线应用大致需经历从业务理解、数据采集及处理、模型训练与测试到运维监控等一系列流程。过程中需要大量的AI算力、高质量数据源、AI应用算法研发及AI技术人员的支持，但大部分中小企业用户并不具备在“算力、数据、算法”三维度从0到1部署的能力，而财力雄厚的大型企业亦需高性价比的AI开发部署方案。依靠AI基础层资源，需求企业可降低资源浪费情况、规避试错成本、提高部署应用速度。作为支撑AI模型开发及落地的必要资源，AI基础层可在多环节提效AI技术价值的释放；其工具属性也标志着AI产业社会化分工的出现，AI产业正逐步进入低技术门槛、低部署成本、各产业深度参与双向共建的效率化生产阶段。

AI应用开发上线流程与AI基础层资源价值点

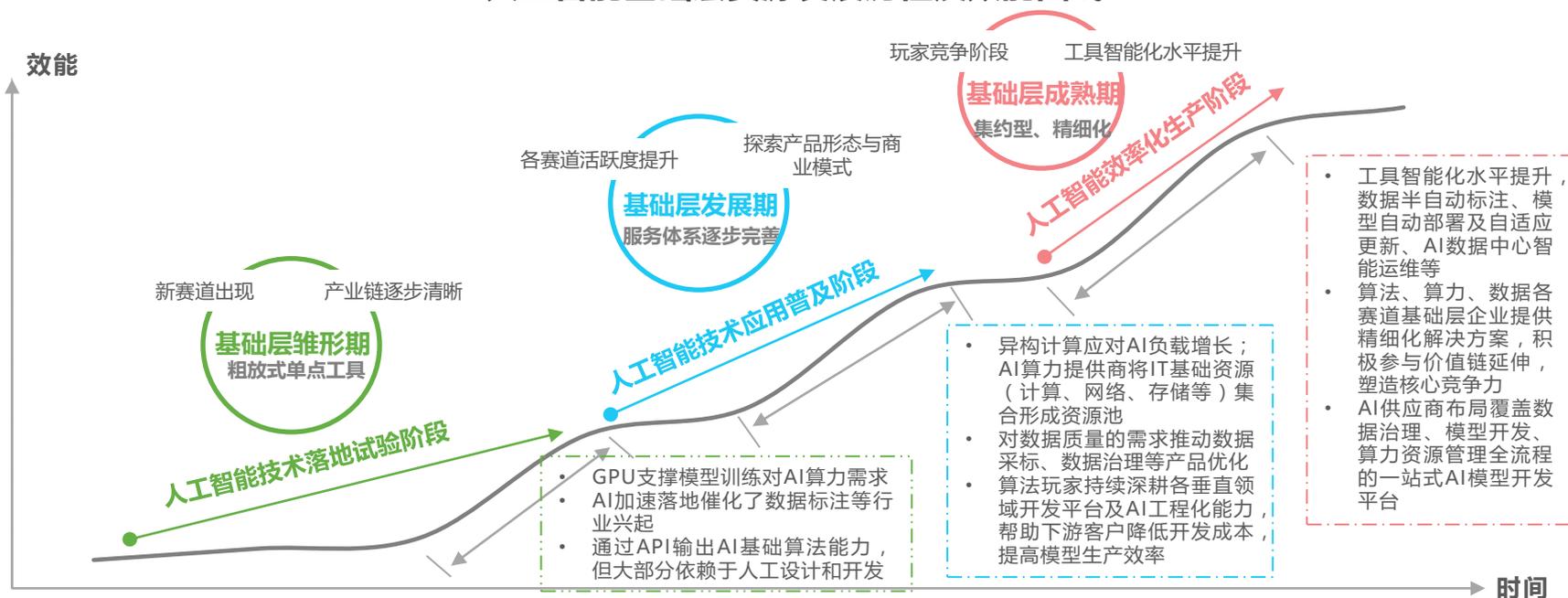


人工智能基础层进阶之路

粗放式单点工具向集约型、精细化资源演进

智能化转型趋势下，企业部署AI项目的需求正经历着变化，对数据质量、模型生产周期、模型自学习水平、模型可解释性、云边端多样部署方式、人力成本及资金投入、投资回报率等的要求都逐步走高。在上述需求特点及自动机器学习、AI芯片硬件架构等技术发展的共同推动下，AI基础层资源的整体效能水平也在不断进化，以有效降低需求企业的AI开发成本。大致涵盖相互交融的三个阶段：雏形期，算法/算力/数据各模块多为粗放式的单点工具，新兴产品及赛道逐步出现；快速发展期，各赛道活跃度显著提升，参与者积极探索产品形态与商业模式，基础层服务体系逐步完善、资源价值凸显；最后则向成熟阶段过渡，各赛道内企业竞争加剧，逐步跑出头部企业。同时各赛道间企业生态合作增多，一站式工具平台出现。

人工智能基础层资源发展历程及效能曲线



人工智能基础层相关政策指引

相关政策以规划、鼓励、建设为倾向，政策红利持续释放

从2017年到2021年初，国务院、国家发改委、国家工信部、科技部以及各省市、地方政府都发布了有关人工智能基础层的政策，在算力、数据、算法开发多层面皆有其侧重点。在算力层面，强调超算中心、AI芯片等基础硬件设施的源头战略地位，倡导为后续的AI应用开发做好硬件底层铺垫；在数据层面，突出实时数据的开发、治理、共享以及数据安全与隐私建设，支持为人工智能的研发及应用做好高效高质、合规合法的数据准备；在AI开发平台层面，提倡开源开放、互助共享的理念，引导具备领先AI资源与技术能力的企业、高校或机构构建开源社区或开放平台，释放优势互补的协同效应。

人工智能基础层相关政策解读

| | 人工智能芯片 | 超算中心 |
|-----------|--|---|
| 算力 | <p>政策扫描总结：有关政策以规划和建设为引导，围绕超算中心、数据中心、AI芯片等展开基础硬件的强化与升级</p> <p>2021.1.13 国家工信部 《工业互联网创新发展行动计划（2021-2023年）》</p> <ul style="list-style-type: none">重点任务中指出要加强工业互联网基础支撑技术攻关，支持工业5G芯片模组、边缘计算专用芯片与操作系统、工业人工智能芯片及工业视觉传感器等基础软硬件的研发突破。 | <p>2019.11.13 国务院 《国家创新驱动发展战略纲要》</p> <ul style="list-style-type: none">战略任务中提到，要“强化原始创新，增强源头供给”，建设一批支撑高水平创新的基础设施和平台，建设超算中心和云计算平台等数字化基础设施。 |
| 数据 | <p>政策扫描总结：有关政策以建设为倾向，覆盖海量实时大数据采集、数据治理、数据共享、数据安全与隐私等方面</p> <p>面向AI的数据治理</p> <p>2020.8.7 国家发改委、科技部等 《国家新一代人工智能标准体系建设指南》</p> <ul style="list-style-type: none">建设内容中，提出支撑技术与产品标准，涵盖大数据标准，核心内容为规范AI研发与应用等过程涉及到的数据存储、处理等相关技术要素，包括数据治理、数据共享开放等。 | <p>人工智能基础数据服务</p> <p>2021.2.9 北京市人民政府 《2021年市政府工作报告重点任务清单》</p> <ul style="list-style-type: none">任务指出要大力发展数字经济，构筑高质量发展新优势。数据基础设施建设中提到，要强化数据安全管理和个人隐私保护，探索建立数据分类分级保护体系。 |
| 算法 | <p>政策扫描总结：政策以鼓励及建设为导向，重点落脚于开源开放的理念，以实现先进AI资源与技术能力的共享</p> <p>人工智能技术开放平台</p> <p>2019.8.1 科技部 《国家新一代人工智能开放创新平台建设工作指引》</p> <ul style="list-style-type: none">建设原则中，提出以企业为主体，鼓励人工智能细分领域领军企业搭建开源、开放平台，面向公众开放AI技术研发资源，向社会输出AI技术服务能力，推动AI技术的行业应用。注：截止到目前，新一代人工智能开放创新平台已增加至15个，包含百度、阿里云、腾讯、科大讯飞、商汤等企业。 | <p>2017.7.8 国务院 《新一代人工智能发展规划》</p> <ul style="list-style-type: none">总体要求中，以开源开放作为基本原则之一，促进产学研用各创新主体共创共享。总体部署中，提出“构建开放协同的人工智能科技创新体系”，以针对性地解决原创性理论基础薄弱、重大产品和系统缺失等重点难点问题。 |

来源：艾瑞研究院根据公开资料自主研究绘制。

人工智能基础层概念界定

1

人工智能基础层需求篇

2

人工智能基础层供给篇

3

典型人工智能基础层企业案例

4

人工智能基础层发展洞察

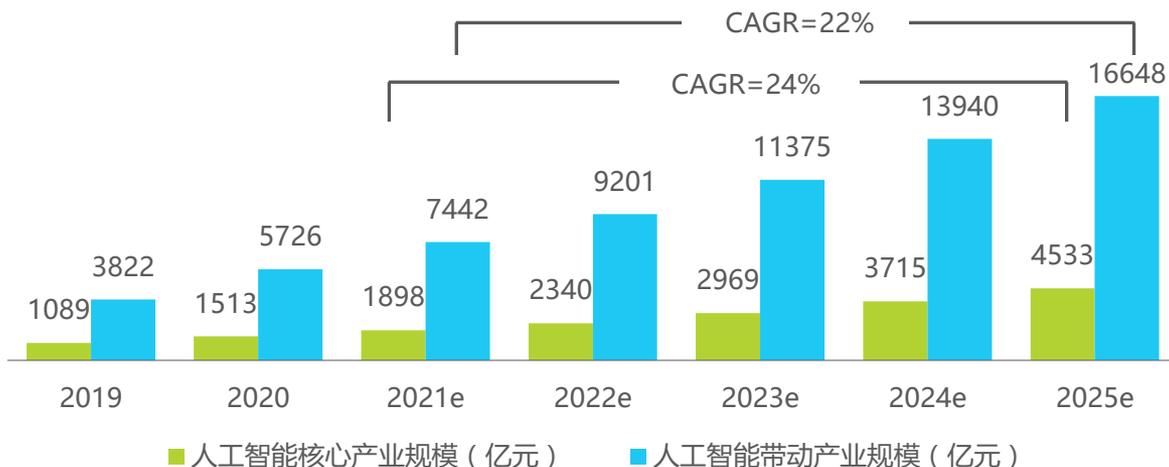
5

人工智能产业的社会经济价值

智能化转型趋势下，人工智能产业及其带动规模维持高增速

AI技术作为智能化转型的重要工具，是指通过机器学习、计算机视觉、智能语音和自然语言处理等技术，处理企业的结构化与非结构化数据等，以提升业务流程效率、进行产品与服务创新并对整体业务作出分析与反馈等。现阶段，各行业企业在改善价值链、降本增效的内在需求驱动和人工智能被列入“新基建”的外在因素影响下，产生了多样化的智能化转型升级需求。据艾瑞测算，2020年中国人工智能核心产业规模超过1500亿元，至2025年预计超过4500亿元，2021-2025年人工智能核心产品CAGR为24%；2020年人工智能带动相关产业规模超5700亿元，至2025年将突破16000亿元，2021-2025年人工智能带动相关产业CAGR为22%。在新产业、新业态、新商业模式经济建设的大背景下，企业对AI的需求逐渐升温，人工智能产值的成长速度令人瞩目；且基于AI的“赋能”特性，已逐步展现出从单向的产品供应向各产业深度双向共建的特征发展，带动相关产业发展，回馈社会经济。

2019-2025年中国人工智能产业及带动相关产业规模



注释：核心产业规模包括计算机视觉、智能语音、人机交互、机器学习、知识图谱与NLP、AI芯片等；带动产业规模包括AI核心产品所带动的工程服务、大数据平台与应用建设、计算机通信产品整机销售收入、甲方企业产值与效益提升的规模总和。

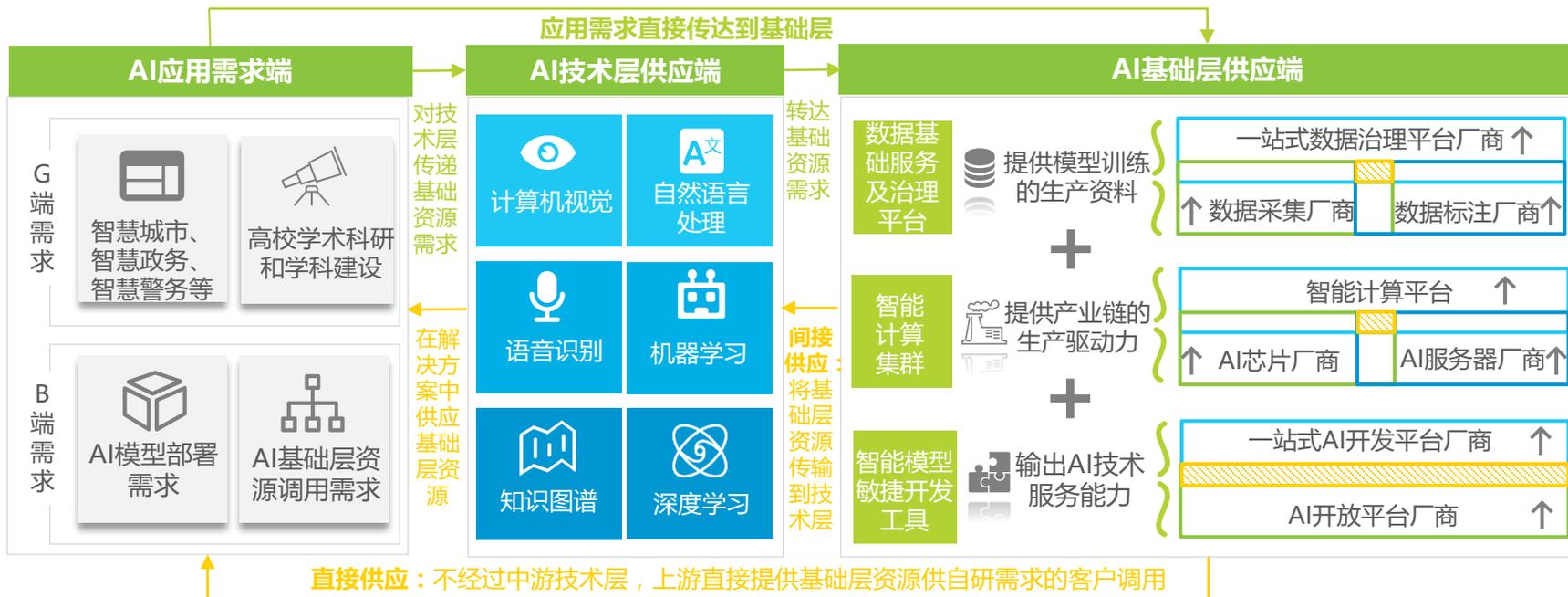
来源：艾瑞长期政府及企业服务数据监测，结合行业专家访谈，根据数据测算模型，自主研究绘制。

基础层初步成型是AI产业链成熟的标志 iResearch 艾瑞咨询

基础层资源促进AI产业链各环节价值传导顺畅、分工明确

现阶段，已初步成型的AI基础层资源可有效缓解下游行业用户逐渐增长的、从感知到认知多类型的AI应用模型开发及部署需求。细看基础层内部，一方面，数据资源、算力资源和算法开发资源三者之间的分工更为明确和有序。数据基础服务及治理平台企业为AI产业链供应数据生产资料；智能计算集群产出高质效的生产力；智能模型敏捷开发工具则负责模型开发及模型训练等，输出AI技术服务能力，提高AI应用模型在各行业的渗透速率与价值空间。另一方面，基础层厂商的数量保持增长、厂商业务范围持续扩大，可提供专业定制化或一站式的基础资源服务。由此，基础层完成AI工业化生产准备，通过直接供应和间接供应的形式，将基础层资源传送到下游的AI应用需求端，产业链向顺畅的资源输送及价值传导方向演进。

AI产业链价值传导机制示意图



注释：绿色箭头代表需求传导链，黄色箭头代表供应传导链，需求端与供给端的传导链并不完全独立，可同时并存。AI企业可具备AI技术层供应和AI基础层供应双重角色。
来源：艾瑞研究院自主绘制。

AI基础层解决人工智能生产力稀缺问题 iResearch 艾瑞咨询

基础层资源缓解甲方在对待人工智能投资上的“矛盾”

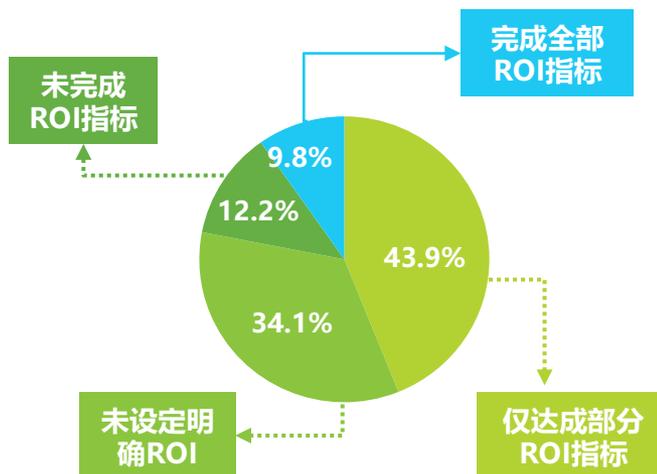
根据艾瑞2020年执行的CTO调研，2019年超过51%的样本企业AI相关研发费用占总研发费用比重在10%以上，2020年65.9%的企业AI研发占比达到10%以上。一方面是甲方企业不断增长的对智能化转型的强劲需求，一方面则是在AI应用开发与部署过程中企业普遍面临的数据质量(49%)、技术人才(51%)等基础资源配置难点。且目前只有少数企业可以完成AI项目实施前设定的全部投资回报率（ROI）标准，因此甲方企业在投资AI项目时相对审慎。AI基础层资源则可有效缓解甲方利用AI技术重塑自身业务时的投资矛盾，提升模型生产效率，降低部署成本：数据资源集群具备数据采标与数据治理能力，且一站式的数据平台可对实时数据进行统一管理，提高数据利用率；高效的AI算力集群与调度系统可满足模型训练与推理需求，降低总拥有成本（TCO，Total Cost of Ownership）；基于算法开发平台演化出的语音识别、计算机视觉、机器学习等专业的AI模型生产平台，可提供高效、一站式的AI模型生产服务。

甲方企业部署AI基础层资源面临难点

| | |
|--|---|
| 数据资源  | <ul style="list-style-type: none">● 数量多、标准不统一：数据量激增，半结构化与非结构化数据难以标准化，数据价值有待发掘，数据开发缺少统一标准● 标注成本高：海量数据标注耗费大量时间与人力● 风险隐患：数据存在违规违法泄露风险，数据共享隐患多 |
| 算力资源  | <ul style="list-style-type: none">● 算力需求强劲但利用率较低：各业务实时性算力需求攀升，业务与算力适配难度较大● 算力资源部署难、调度难：硬件采购成本高以及交付周期不可控，计算资源调度门槛高● 算力集群操作系统复杂：各类异构算力的操作系统自动化水平低，交互友好性低 |
| 算法开发资源  | <ul style="list-style-type: none">● 产品二次开发：由于企业业务特性不同，长尾化的AI开发需求难以满足。产品二次开发普遍存在需要定制化的AI能力● 沟通不畅：开发涉及多团队沟通、沟通对接冗余 |

来源：艾瑞研究院自主绘制。

2020年中国甲方企业AI项目投资回报率实现情况



来源：艾瑞咨询《2020年中国人工智能产业研究报告》大中型企业CTO/CIO调研，N=41，2020年9月。

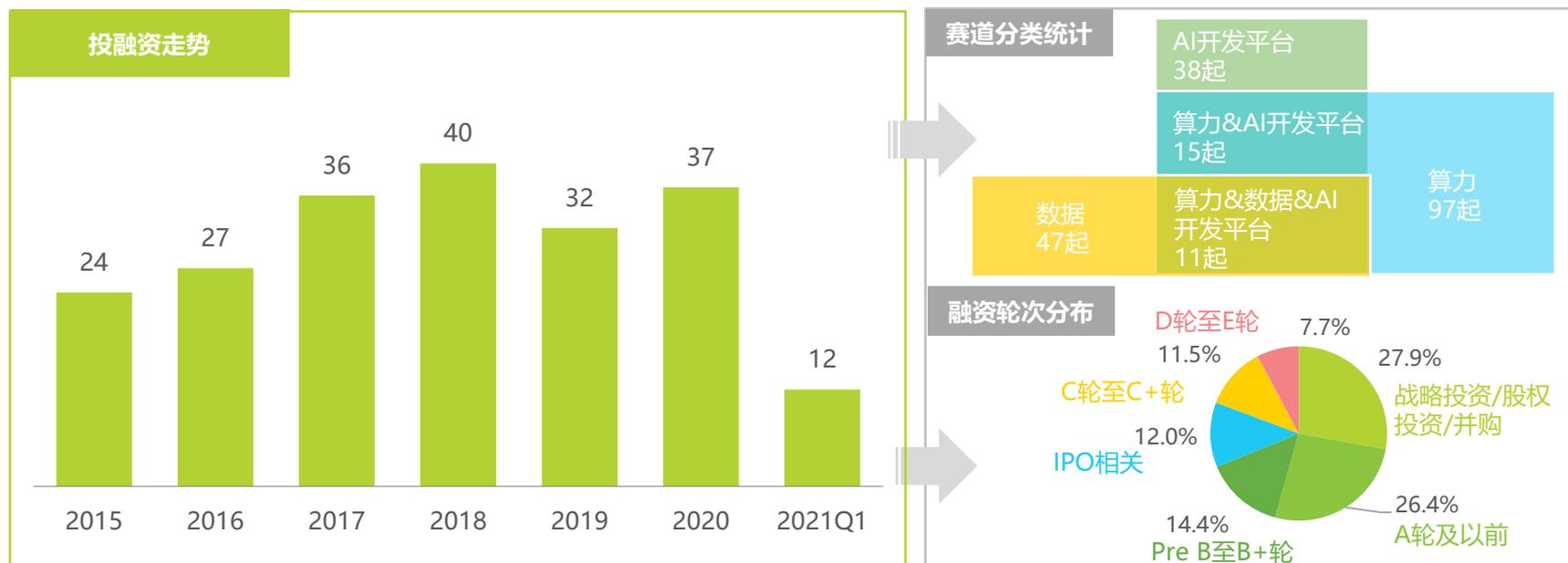
| | |
|---------------|---|
| 人工智能基础层概念界定 | 1 |
| 人工智能基础层需求篇 | 2 |
| 人工智能基础层供给篇 | 3 |
| 典型人工智能基础层企业案例 | 4 |
| 人工智能基础层发展洞察 | 5 |

人工智能基础层资本情况

融资热度较高，资本市场广度与深度拓宽

据艾瑞不完全统计，2015年至2021年Q1，中国人工智能基础层融资事件共计208起。各赛道中，算力资源赛道目前形成以AI芯片和智能服务器为主、高性能计算中心为辅的情况，而数据资源赛道的融资集中分布于数据资源定制服务与数据集产品。同时，综合类融资事件围绕着AI芯片、高性能计算平台、AI模型生产平台与数据资源服务而展开。从融资轮次分布情况看，人工智能基础层资本市场呈现多轮次的局面。融资多聚集在战略投资、股权投资、并购与A轮及以前，IPO与Pre-B轮后融资较为分散。这意味着该市场存在新鲜度、活跃度较高的企业同时，不乏产业成熟度较高的企业。艾瑞认为，随着多类型的玩家入局与产品丰富度提高，该资本市场的广度与深度将进一步拓宽。

2015-2021Q1中国人工智能基础层融资事件分布情况



来源：艾瑞研究院根据公开投融资数据自主研究绘制。

人工智能基础层产业图谱

2021年人工智能基础层产业图谱



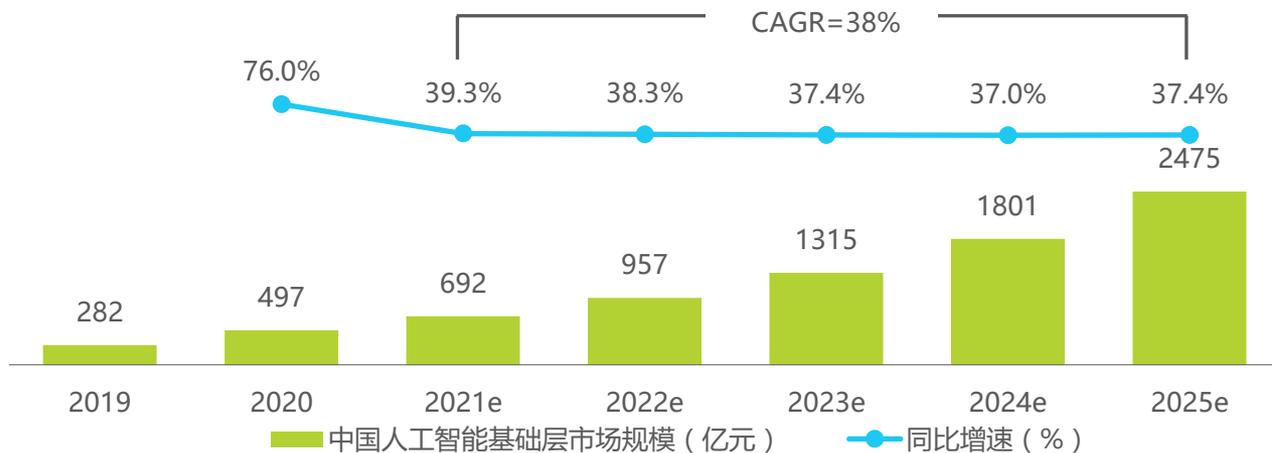
注释：以企业主营业务为主，展示部分行业代表性企业。图谱中所展示的公司logo顺序及大小并无实际意义，不涉及排名。
来源：艾瑞根据公开资料研究绘制。

人工智能基础层市场规模

AI企业业务突破、智能化转型趋势等多因素驱动产业规模增长

AI企业突破业务增长瓶颈的需求是人工智能基础层发展的驱动力之一。当前人工智能核心产业规模保持线性增长态势，且增速趋于平稳与常态化。为寻求产值增长突破点，AI企业发力探索开拓市场的有效手段。依托人工智能基础层资源建设，AI企业可有效应对下游客户的长尾应用需求，再将高频应用转化为新主营业务。此外“新基建”、半导体自主可控等相关政策扶持、传统行业智能化转型等因素也都在助推人工智能基础层资源的发展。据艾瑞测算，2021-2025年，人工智能基础层市场规模CAGR为38%，整体产业规模发展速度较快、空间较为广阔，总体呈现持续增长的走势。2020年，中国人工智能基础层市场规模为497亿元，为人工智能产业总规模的33%，市场规模较去年同比增长76%，AI应用模型效率化生产平台创收增长、AI芯片市场规模随着云端训练需求出现较高增长等是同比增速的主要拉动力；2021-2024年同比增速趋于平缓下降，市场开始恢复稳步增长态势。到2025年，中国人工智能基础层市场规模将达到2475亿元，云端推理与端侧推理芯片市场持续走高使得人工智能基础层整体市场同比增速稍有抬升。

2019-2025年人工智能基础层市场规模



注释：此处的人工智能基础层规模由AI芯片（算力）、AI基础数据服务（数据）、面向AI的数据治理（数据）、AI技术开放平台（算法）与AI应用模型效率化生产平台（算法）构成；算力模块的计算核心为AI芯片，服务器其他硬件构成与AI算力不存在直接相关性，故不予统计计算。
来源：艾瑞研究院自主研究绘制。

智能计算集群

智能计算集群产业链及图谱

企业自建数据中心需求增长，AI服务器占比逐渐增高

智能计算集群的产业链结构为：1) 上游：材料与基础设施；2) 中游：智能计算集群供应商；3) 下游：各行业算力需求企业。中游的三类服务商中，系统级AI芯片厂商为服务器厂商提供异构计算所用的芯片或板卡；异构智能计算服务器厂商则为各类型数据中心/超算中心提供服务器。值得注意的是，智能云、互联网等领域企业的自建数据中心需求增长，且其中AI服务器比例逐渐增高。

智能数据集群产业链及图谱

下游：各行业算力需求企业



注释：图谱中所展示的公司logo顺序及大小并无实际意义，不涉及排名。

来源：艾瑞研究院结合公开资料自主研究绘制。

超算/数据中心的存量与增量判断

从算力需求与节能减排规定看存量替换与增量增长

现有的超算/数据中心以建设单位为标准，可分为两类：1) 以国家或地方为建设单位的G端超算中心，服务对象主要为国家牵头的重点科研单位、高校研究院等，此类超算中心是解决国家安全、科学进步、经济发展与国防建设等重大挑战性问题的重要手段，近两年受国家与地方的高度重视与扶持，建设与升级超算中心的趋势愈加明朗。但由于数据网络安全与计算精度要求高，建设周期较长，此类超算中心的数量在中短期内增长缓慢，长期来看则会成为替换存量与增量增长的贡献主力之一。2) 以智能云厂商或IDC服务商为建设供应主体的B端超算/数据中心，为互联网公司、其他类型的企业或事业单位提供主机托管、资源出租、增值或应用服务，是存量与增量市场变化主要推力。从市场变化趋势来看：1) 存量市场：日渐增加的AI计算负载需要处理力更强、能耗承受度更大的数据中心，同时，一系列有关控制数据中心PUE值的节能审查规定相继出台，一味盲目扩建、新建数据中心已难合时宜，促进老旧数据中心绿色化改造的减量替代方案因此诞生。微型、中小型数据中心会逐渐被改造为集约型的大型数据中心，符合节能减排相关标准、机柜数量与异构组合增多的集约型超算/数据中心将在存量市场中占据主流。2) 增量市场：考虑到边缘计算可分担AI计算任务、兼具低延时优势，管理边缘计算中心则需要布局相应的大型云端数据中心，故增量市场会被异构的边缘计算数据中心与云端超算或大型数据中心扩充。



注释：此面积图是根据研究资料总结所得的历史趋势判断进行的定性绘制。

来源：艾瑞研究院结合国防科技大学学报《中国超算技术赶超发展模式探析》、专家访谈、桌研公开资料自主研究绘制。

AI服务器综合能力维度

资源配置相互协作，共同决定AI服务器综合能力强弱

吞吐量、响应时间、并发数与业务成功率等技术评价指标往往是一台AI服务器在各种资源运行协作下产生的结果，要评价AI服务器的综合能力，需追溯到最根源的资源配置上，查看软硬件搭配的综合情况。艾瑞在此处将AI服务器的综合能力维度分成三部分：板卡芯片组配置、硬盘-网络带宽-操作系统、机箱结构设计。板卡芯片组是AI服务器运行效率的核心，主频高、核心数多的CPU搭配小面积、高频率、算力强的加速芯片，并结合互联互通的总线与充足的内存，才能保证更高的吞吐量、更多的并发数与更低的延迟。硬盘-网络带宽-操作系统是AI服务器的运作基础，硬盘存储空间与网络带宽的大小影响着AI应用与用户数据存储的上限与线上接入数据的快慢，操作系统则影响着数据与访问的安全性、执行的效率与稳定性。由于AI服务器配置多个GPU或其他的系统级芯片，在助跑AI模型时会散发较多的热量，因此AI服务器物理环境的安全稳定、机箱的通风散热设计亦是需要考虑的因素。

AI服务器综合能力解析

计算核心：板卡芯片组配置

CPU

很大程度上决定服务器的工作速度与效率，影响吞吐量。CPU的主频越高、核心数越多，服务器的运转速度就越快

GPU or FPGA or ASIC

AI芯片是算力强弱的核心。架构与制程、理论与实际算力效果、功耗、存储宽带接口、与物理主机的适配性等共同影响着AI芯片性能的高低

缓存、内存与总线

缓存越大，越能提高CPU对内部数据的命中率；内存的容量越大、频率越快、延迟越低，越能提高CPU调用数据的效率；总线涉及到芯片、硬盘、设备、网卡之间的通信效率

来源：艾瑞研究院结合公开资料自主研究绘制。



辅助：机箱结构设计

AI服务器为多卡配置，需对系统的结构、散热、拓扑做针对性设计，保证性能稳定

运作基础：硬盘-网络带宽-操作系统

硬盘

服务器硬盘种类的选择关联到CPU的占用情况与数据传输速度的快慢，目前一般采用SATA和SCSI接口类型的硬盘

网络带宽

部署服务器的前期需做好带宽预估工作，带宽影响单位时间内数据量的传输效率与业务响应时间

操作系统

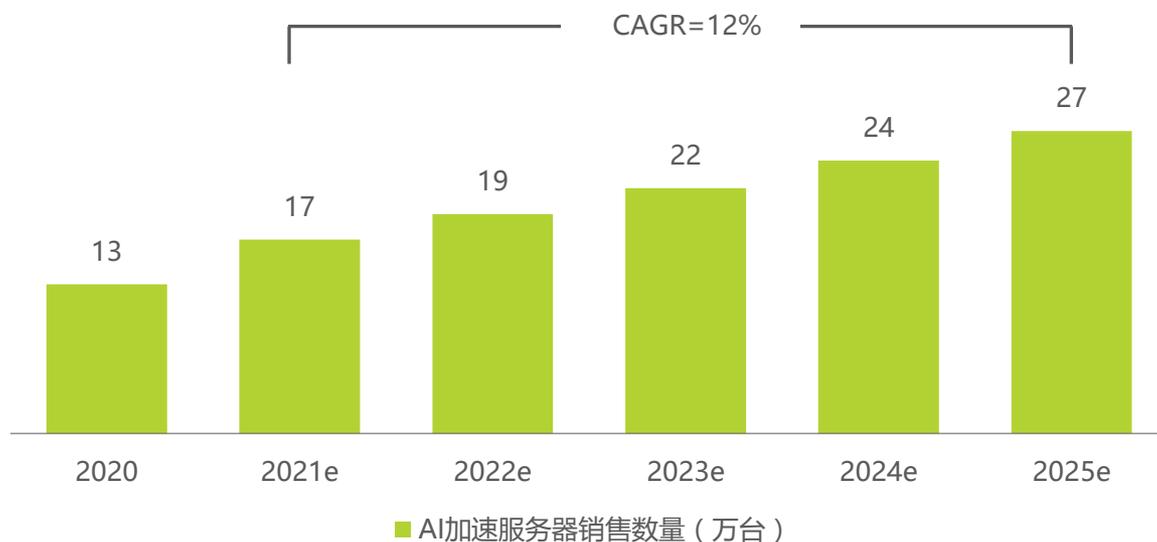
当前服务器的主流操作系统为Unix、Linux、Windows Server和NetWare，各操作系统的特性与应用场景有差异，客户需根据自身行业与业务性质选择

AI加速服务器的出货情况

销售数量持续走高，市场处于成长期

尽管现阶段各类人工智能应用模型层出不穷，但是人工智能应用层的各种理念远远走在底层算力架构之前，现有的计算资源难以满足人工智能的算力需求，人工智能底层算力架构的设计与优化已是大势所趋。CPU+GPU/FPGA/ASIC类的异构AI加速服务器能够为机器学习、深度学习、计算机视觉、NLP、知识图谱、智能语音等AI技术加速，支持更大规模的深度神经网络模型的同时，还能提高训练精度，因而成为AI算力突破的新业务增长点。据艾瑞统计与预测，2020年我国AI加速服务器销售数量为13万台，到2025年，我国AI加速服务器销售数量将达到27万台，2021-2025年的相关CAGR=12%，我国AI加速服务器出货市场整体呈现上扬趋势，AI加速服务器在服务器市场中的渗透率将会继续走高，并持续为服务器市场注入较为强劲的增长活力。

2020-2025年中国AI加速服务器销售数量



注释：AI加速服务器是指采用异构形式部署、搭载异构计算加速卡的服务器。
来源：艾瑞研究院结合桌研情况、专家访谈自主研究绘制。

端-边-云的算力协同

端-边-云实现AI算力泛在，加速AI模型训练与推理

在人工智能与5G等技术的冲击下，设备端产生大量实时数据，若直接上传到云端处理，会对云端的带宽、算力、存储空间等造成巨大压力，同时也存在延时长、数据传输安全性等问题。因此，为缓解云端的工作负载，云计算在云与端之间新增了若干个边缘计算节点，从而衍生出端-边-云的资源、数据与算力协同。在算力协同的业务模式下，靠近云端的云计算中心承担更多的模型训练任务，贴近端侧的各设备主要进行模型推理，而二者之间的边缘侧则负责通用模型的转移学习，帮助云端分散通用模型训练任务、处理实时计算的同时，也解决了终端算力不足、计算功耗大的难题。未来，边缘计算的发展会催生出更适宜边缘计算场景的算力集群异构设计，其异构化程度将会高于传统的数据中心，异构设计的突破将会进一步提高端-边-云的整体计算效能，进而加速AI模型的训练与推理。

端-边-云的算力协同运作流程



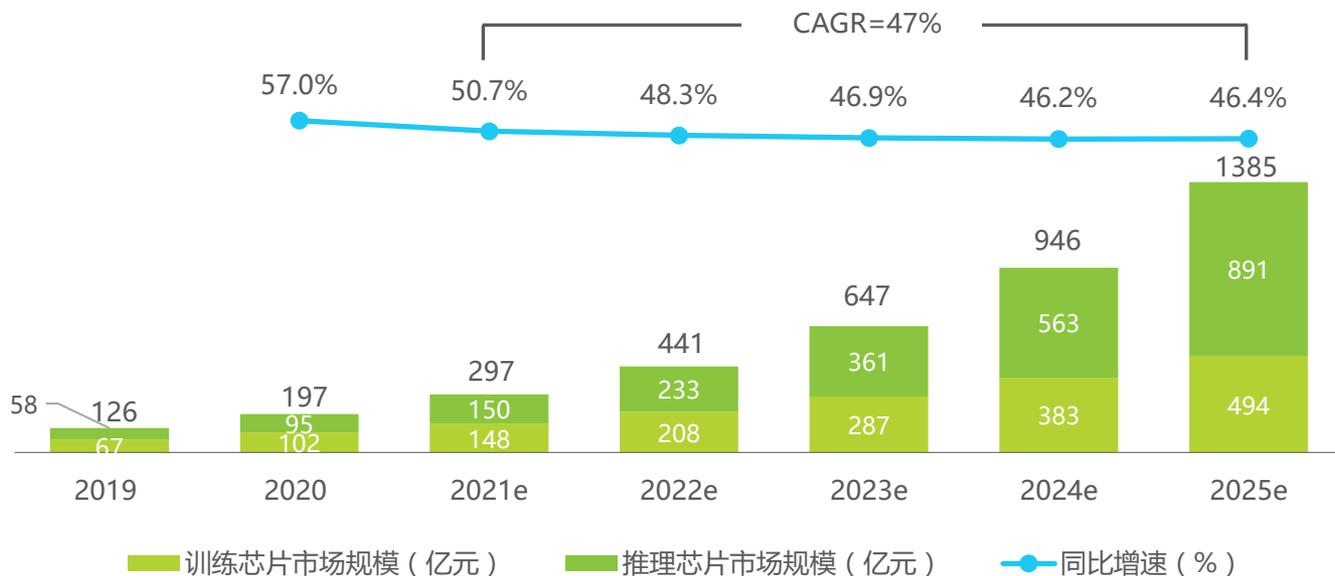
来源：艾瑞研究院结合公开资料自主研究绘制。

AI芯片市场规模

当前以训练需求为主，推理需求将成为未来市场主要增长动力

AI芯片是人工智能产业的关键硬件，也是AI加速服务器中用于AI训练与推理的核心计算硬件，被广泛应用于人工智能、云计算、数据中心、边缘计算、移动终端等领域。当前，我国的AI芯片行业仍处于起步期，市场空间有待探索与开拓。据艾瑞统计与预测，2020年我国AI芯片市场规模为197亿元，到2025年，我国AI芯片市场规模将达到1385亿元，2021-2025年的相关CAGR=47%，市场整体增速较快。从AI芯片的计算功能来看，一开始，因AI应用模型首先要在云端经过训练、调优与测试，计算的数据量与执行的任务量数以万计，故云端训练需求是AI芯片市场的主流需求。而在后期，训练好的AI应用模型转移到端侧，结合实时数据进行推理运算、释放AI功能，推理需求逐渐取代训练需求，带动推理芯片市场崛起。2025年，云端推理与端侧推理成为市场规模增长的主要拉动力，提升了逐渐下滑的AI芯片市场规模同比增速。

2019-2025年中国AI芯片市场规模



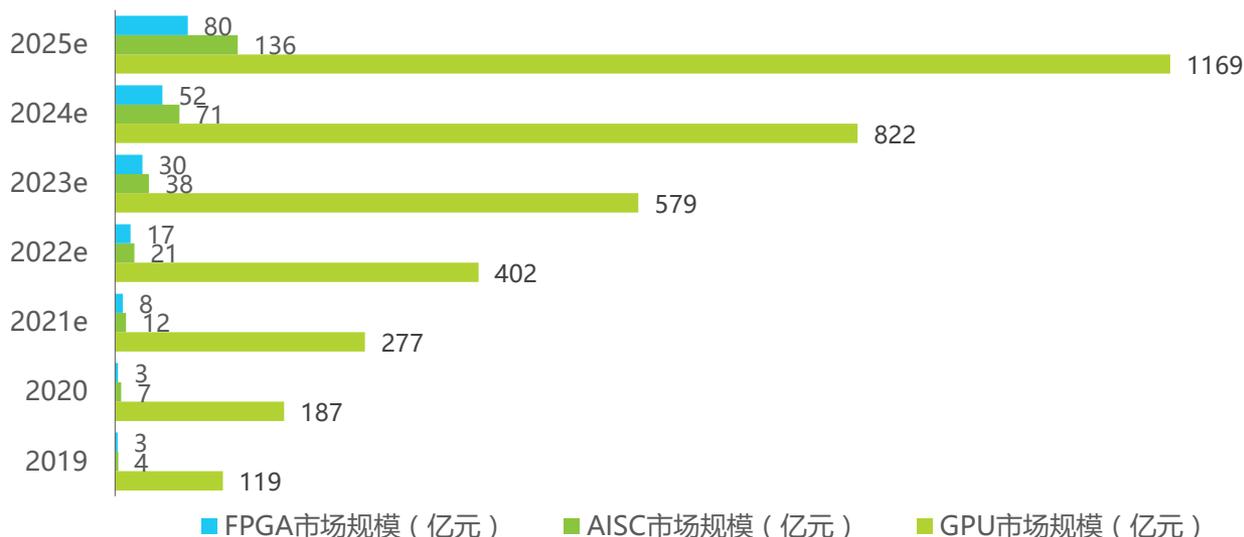
来源：艾瑞研究院通过统计分析市场主流AI芯片供应商数据、结合桌研行业趋势与专家访谈自主研究绘制。

以技术架构分类看AI芯片市场规模

GPU、ASIC、FPGA各司其职，协同开拓细分架构市场

从AI芯片的技术架构看，运用得最为广泛的加速计算芯片为GPU，这主要是因为GPU具备高密度与高并行的计算优势，可用于以深度学习为主的AI算法模型训练。但是，在以下典型的两种情况下，GPU的优势难以得到充分发挥，AI计算需要更好的替代方案：1) 在细分长尾的计算场景中，通常需要设计可实现特定功能的计算模型，相比更为通用、更适用于图像类运算的GPU而言，ASIC芯片因其专有性更适合此类计算。2) 在执行推理任务时，数据量是实时而小批量的，使用GPU显得有些“大材小用”，而FPGA同时拥有流水线并行与数据并行的功能，更适用于此类情况，而且时延更低。因此，AI芯片市场同时需要GPU、ASIC与FPGA等多类芯片来针对不同的应用模型进行计算。2019-2025年，GPU的市场规模保持着主要地位，2025年起市场规模将达到1169亿元。ASIC与FPGA的市场规模还处在起步阶段，增长态势偏稳定，存在较大的拓展空间，有望替代GPU执行更为复杂或更有针对性的计算任务。

2019-2025年中国AI芯片细分架构市场规模

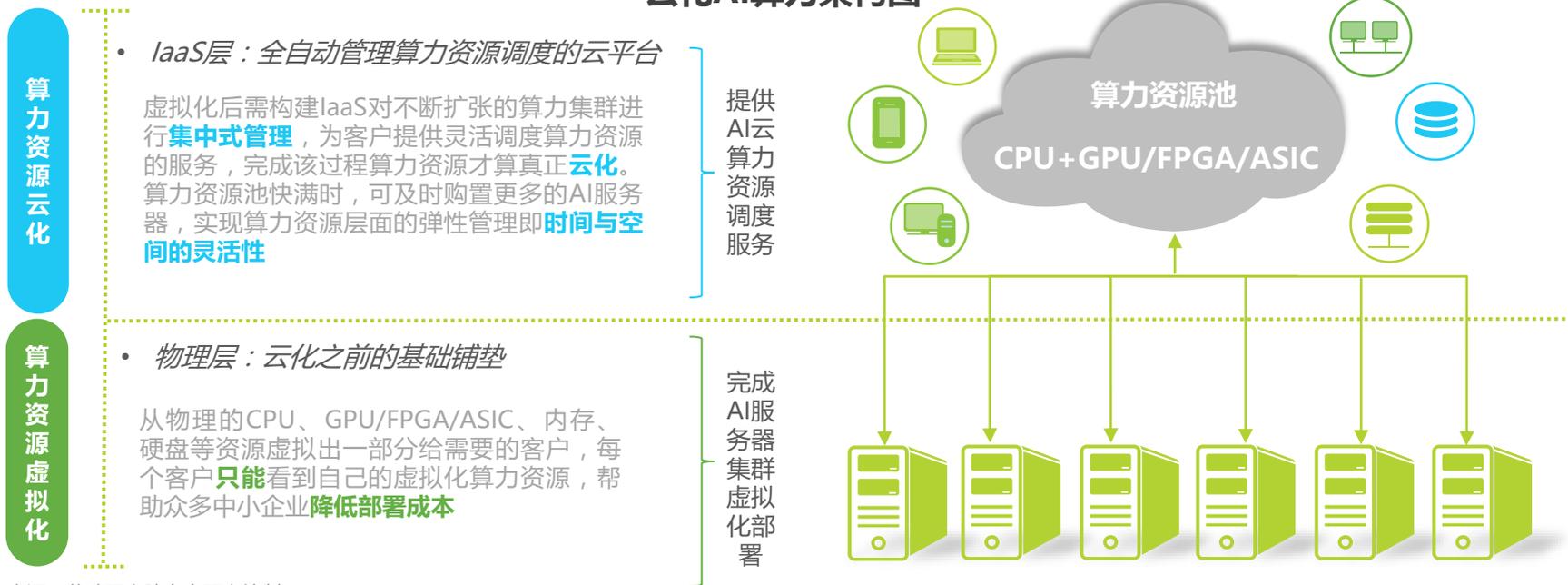


来源：艾瑞研究院通过统计分析市场主流AI芯片供应商数据、结合桌研行业趋势与专家访谈自主研究绘制。

开放共享虚拟AI算力资源，实现AI模型海量训练与推理

AI是一种高资源消耗、强计算的技术，AI算力的强弱直接关联到AI模型训练的精度与实时推理的结果。若企业独立部署AI算力，不仅需要建设或租用机房这类重资产与网络宽带资源，还需要购置物理机、内存、硬盘等硬件设备，而且购置设备存在采购周期不确定、硬件资源过度铺张、专业管理团队缺乏等问题。所以，独立部署AI算力资源是一项耗时耗力的工作。将AI算力云化是一种高效能、低成本的有力解决方案。具备先天性业务优势的云服务商搭建数据中心，先将AI服务器算力资源虚拟化，开放给AI模型开发者，做到按需分配，如给短视频业务的开发者优先配备CPU+GPU方案，而后对算力资源的调度工作进行统一管理。由此，“物美价廉”的算力有序注入各行各业的AI模型中，减轻了井喷式数据爆发所带来的模型训练负担，并能及时根据用户使用情况弹性扩充或缩减虚拟算力资源空间，达到方便、灵活、降本增效的效果。

云化AI算力架构图



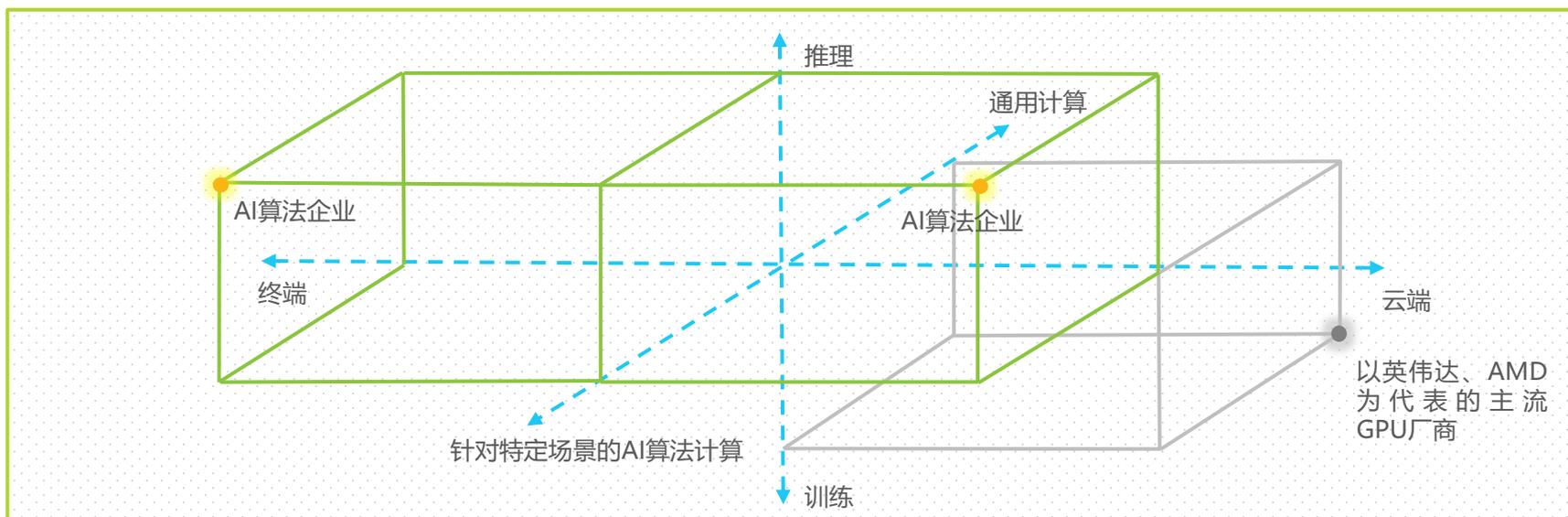
来源：艾瑞研究院自主研究绘制。

AI算法企业的芯片之路

芯片架构与业务场景、算法的融合

现阶段仍为人工智能产业发展的起步期，CPU与GPU搭配的传统计算模式仍然是AI应用模型计算的首选。但是，随着人工智能应用场景的不断细分与算法的加速演进，这些芯片与场景、算法脱节的问题日渐凸显，针对特定场景而定制开发的ASIC芯片架构成为AI芯片市场的新发力点。以算法起家的AI企业以此为机会，凭借自身在算法技术领域的优势以及对业务场景的前瞻性理解，开始了算法芯片级封装的尝试，并取得了不错的成绩。因计算机视觉的业务广泛性强、技术成熟度高，多数AI企业都投入到与之相关的AI芯片研发中，应用于安防、自动驾驶、短视频等高频业务场景中。就AI算法企业的主打市场而言，云端训练芯片市场相对集中，市场份额基本被英伟达这类芯片生态圈强大的早期进入者所垄断，进入壁垒较高，故AI算法企业的切入点一般会向云端推理、终端推理等较为分散、应用场景各异的市场倾斜，这也侧面反映出未来的推理芯片市场存在一定的扩张趋势。

AI算法企业与传统芯片厂商芯片布局三维象限



来源：艾瑞研究院自主研究绘制。

智能模型敏捷开发工具

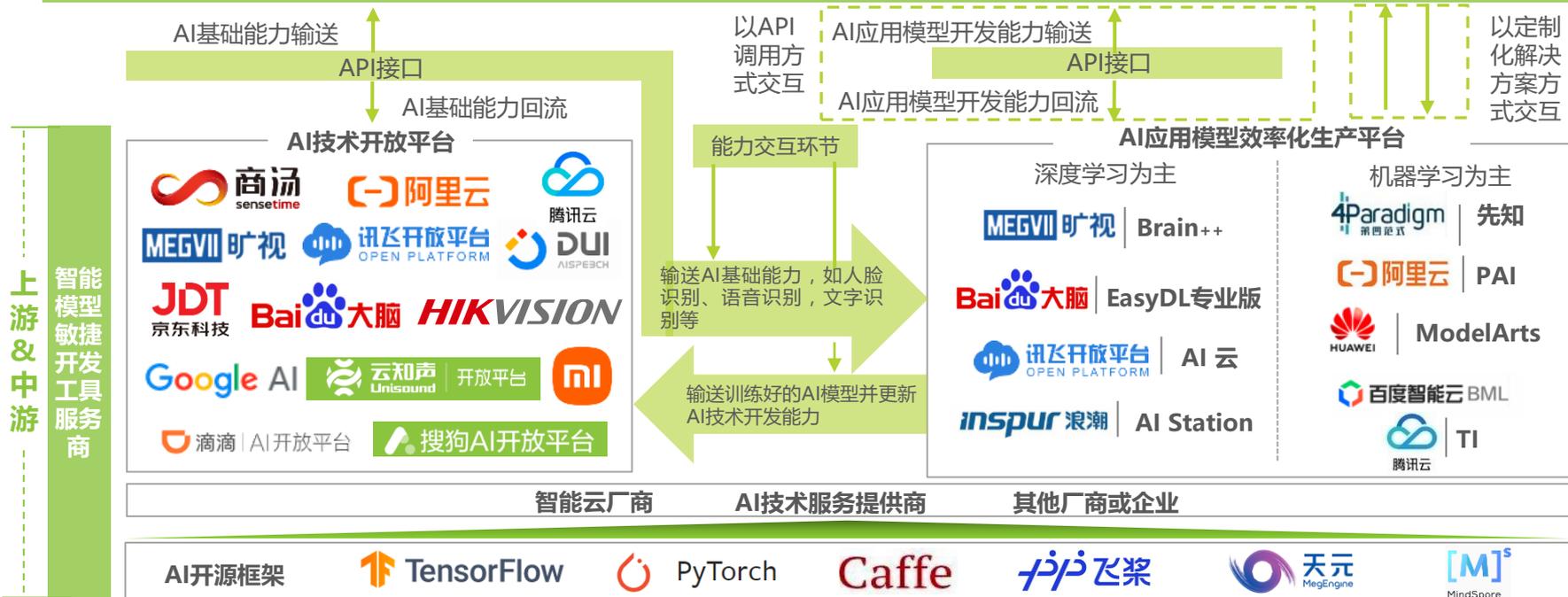
智能模型敏捷开发工具产业链及图谱

AI能力开放互通，技术与业务协同前进

智能模型敏捷开发工具的上游与中游重合，AI技术服务商、智能云厂商与其他厂商或企业通过自有或外部的开源框架开发出AI应用模型，将模型放入AI技术开放平台或AI应用模型效率化生产平台中。从能力输送的方向看，一方面，AI技术平台与AI应用模型效率化生产平台均开放API接口，将对应能力输送给下游，并吸收下游更新的产品与技术，从横向与纵向拓展业务的广度与深度；另一方面，AI技术开放平台与AI应用模型效率化生产平台之间也会进行能力的互换，共同促进AI技术的发展。

智能模型敏捷开发工具产业链及图谱

下游：各行业AI应用开发企业、个人开发者或企业使用者



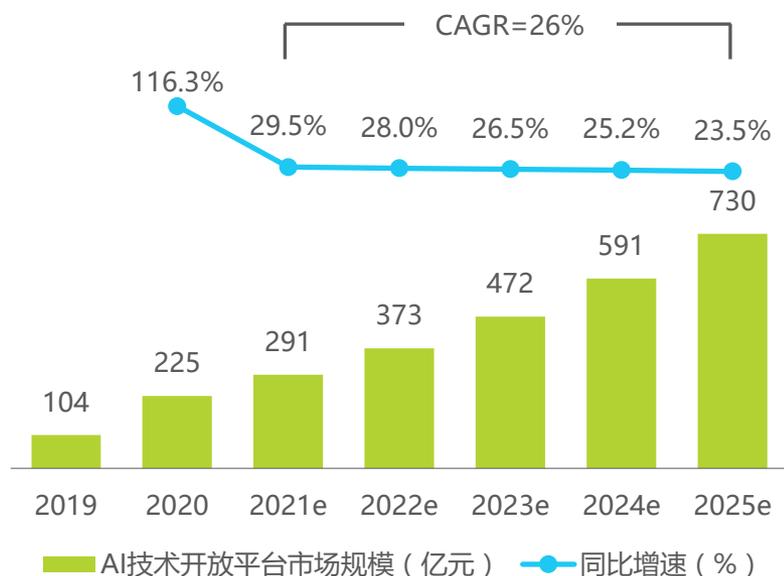
注释：图谱中所展示的公司logo顺序及大小并无实际意义，不涉及排名；
来源：艾瑞咨询《2020年中国人工智能API经济白皮书》；艾瑞研究院自主研究绘制。

AI技术开放平台市场规模

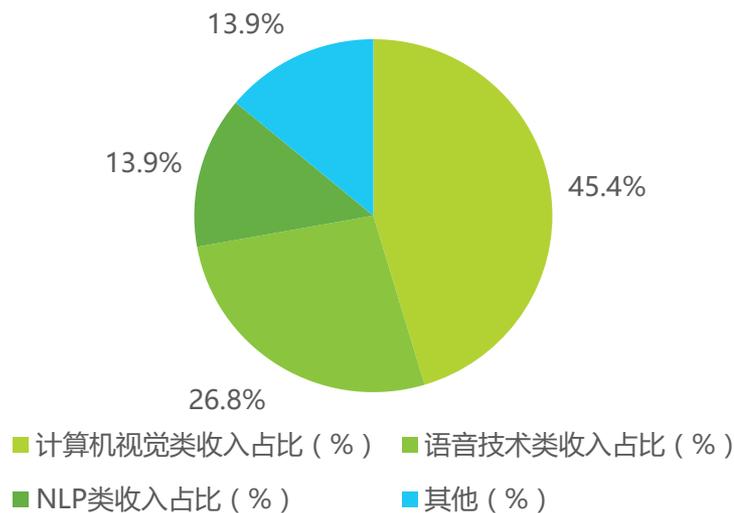
产业受API经济带动，主要收入贡献来源为计算机视觉类

随着数据量与AI算力的提升，可落地的场景与算法的交互变得愈加频繁，二者结合开发出的AI应用模型就需要更大量地通过API调用AI技术开放平台的AI技术能力。据艾瑞统计与预测，2020年我国AI技术开放平台市场规模为225亿元，到2025年，相应规模可达到730亿元，2021-2025年的相关CAGR=26%。受API经济兴起的影响，2020年市场规模同比增速走高较快，2021年下滑后恢复平稳态势。按AI技术能力划分，计算机视觉类与语音技术类收入占比达72.2%，是收入的主要贡献来源。人脸识别、人体识别、OCR文字识别、图像识别等构成了计算机视觉类业务的主要技术能力，且计算机视觉类的技术价格相较于其他技术而言更高，应用领域也更为广泛。现阶段的市场集中度相对分散，未来，能持续投入成本、研发出强劲算法的厂商有望占领更多的市场份额，市场集中度亦会因此提升。

2019-2025年中国AI技术开放平台市场规模



2020年中国AI技术开放平台市场规模结构 分类：按AI技术能力



注释：规模尚未统计企业或集团内部之间的API调用结算收入。
来源：艾瑞研究院结合市场主要供应商的数据统计、行业专家访谈与公开资料自主研究绘制。

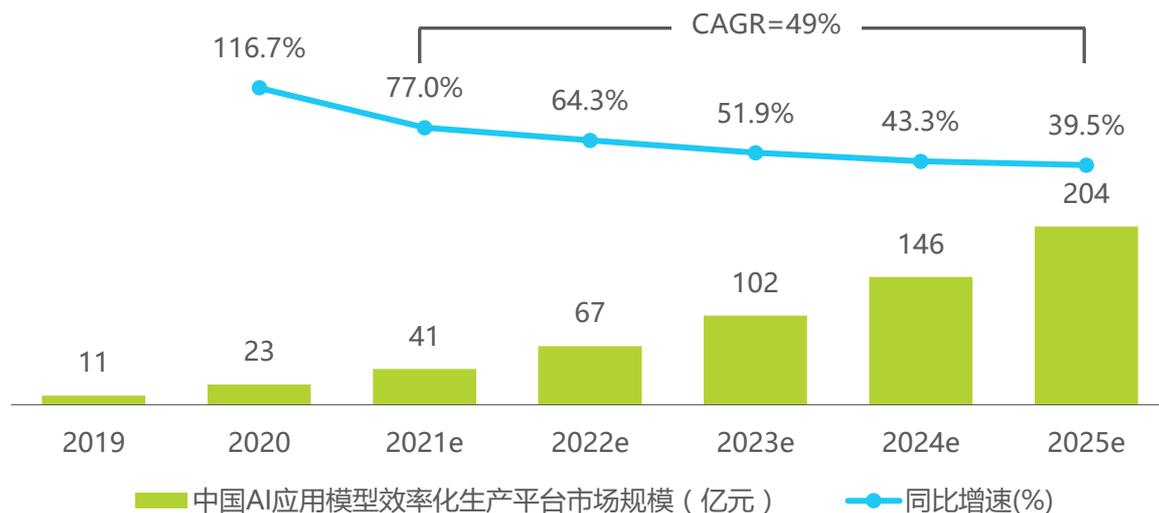
来源：艾瑞研究院根据专家访谈与公开资料自主研究绘制。

AI应用模型效率化生产平台市场规模

集成式的模型开发工具包，产业恰逢伊始，前景有待开拓

AI应用模型效率化生产平台是全栈式的、可实现流水线开发的AI应用模型生产工具。假若每次开发模型都需要算法工程师单独完成从生产到上线的全流程搭建，就会导致很多时间的耗损与AI模型开发成本的浪费。集成了数据、算法与算力的相应开发工具的模型开发工具包——AI应用模型效率化生产平台应运而生。据艾瑞统计与预测，2020年我国AI应用模型效率化生产平台市场规模为23亿元；到2025年，相应规模可达到204亿元，2021-2025年的相关CAGR=49%。2020年，AI应用模型效率化生产平台相关业务拓展相对较快、产品恰逢创收伊始阶段，故同比增速增幅较快。与此同时，因参与技术门槛偏高，具备能力的厂商较少，市场尚未形成稳定状态，市场集中度偏高。

2019-2025年中国AI应用模型效率化生产平台市场规模



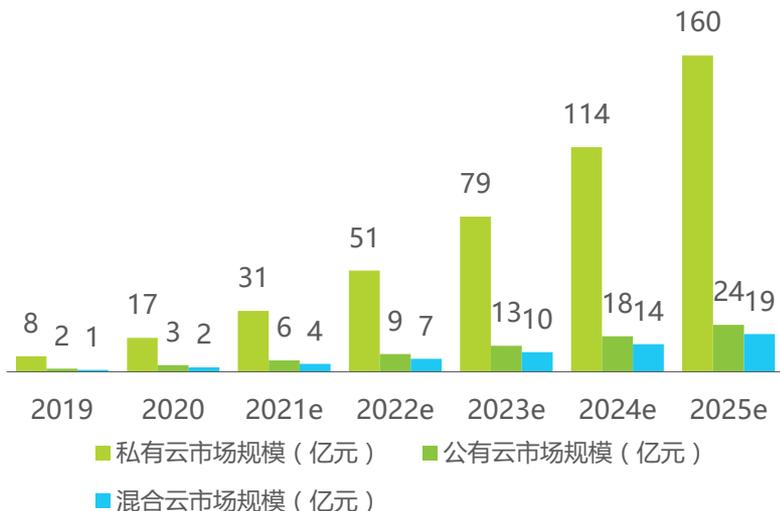
来源：艾瑞研究院统计市场主流供应商相关数据、结合专家访谈与公开资料自主研究绘制。

以部署方式分类的市场规模细分结构

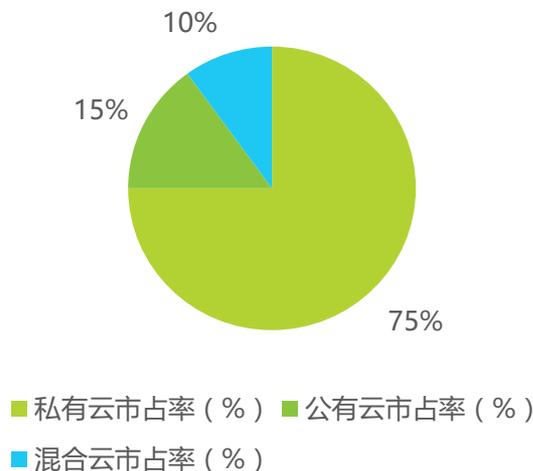
以私有云部署为主，私有云市场仍是一片蓝海

AI应用模型效率化生产平台有三种部署方式，分别是公有云、私有云与混合云。从部署的市场规模量级与市占率来看，当前与未来均以私有云为最主要的部署方式。2020年，私有云部署方式市占率高达75%，市场规模达到17亿元；公有云与混合云部署方式市占率分别为15%与10%，二者的市场规模仅有5亿元。2025年，私有云部署方式市场规模将达160亿元。当前，该产业尚处萌芽阶段，参与者少且潜在的下游客户市场空间巨大，下游客户往往出于数据安全与隐私的考虑，偏好私有云部署，私有云市场仍是一片蓝海。在公有云部署方式下，虽然客户数量偏多，但单价偏低，其收入贡献程度有限。另外，公有云市场基本已经被云计算业务成熟度高的公有云厂商所垄断。在混合云部署方式下，客户的私有敏感数据会存储到私有云中，非敏感数据则存储到公有云中，但使用该种部署方式的客户偏少。

2019-2025年中国AI应用模型效率化生产平台部署模式市场规模



2020年中国AI应用模型效率化生产平台部署模式市场占有情况



来源：艾瑞研究院统计市场主流供应商相关数据、结合专家访谈与公开资料自主研究绘制。

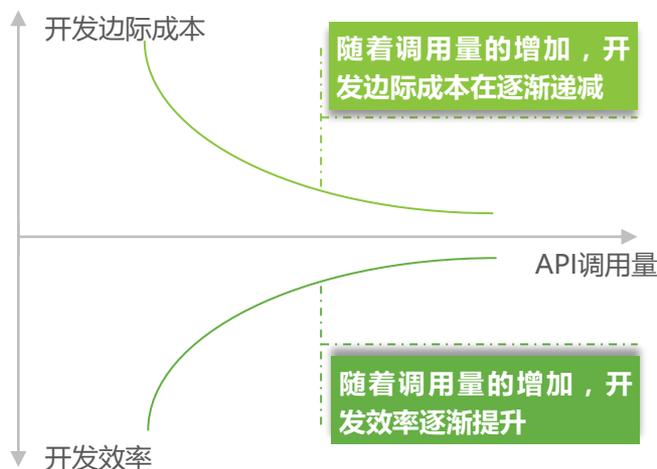
来源：艾瑞研究院统计市场主流供应商相关数据、结合专家访谈与公开资料自主研究绘制。

智能模型敏捷开发工具商业价值分析

API规模经济+AI应用模型效率化生产平台的杠杆增效

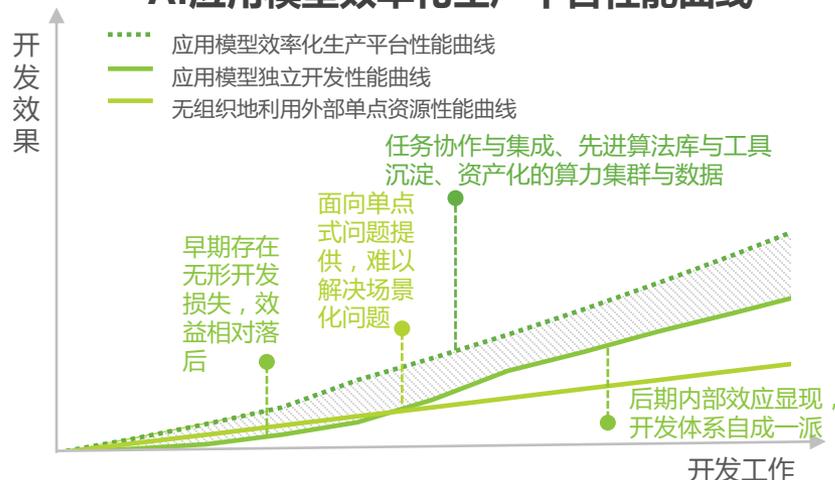
智能模型敏捷开发工具的出现与驱动AI业务的外因以及企业自身的内因紧密相关。从外因看，规模化多场景的业务不断衍生出长尾需求，原有的应用需及时更新；从内因看，囿于开发企业有限的经营成本与AI技术人才，其资源主要投放到现阶段的主营业务，现有人员难以推动业务的智能化改造。针对此局面，可有效解决AI应用模型设计与开发过程中通用或特有问题的智能模型敏捷开发工具逐步成为备选方案。AI开放平台与AI应用模型效率化生产平台作为其中的代表性工具，不仅能减少由0到1的开发成本，而且可降低人工智能市场的参与门槛，提升开发效果。AI开放平台属于API资源的一种，其可帮助技术领先企业开放AI能力与先进资源，从而延伸价值链、构造产业生态，形成规模经济与长尾经济，且利用开发者的创新应用来反哺开放平台；同时，其亦可在减轻基础设施建设投入的条件下协助开发者打造自身产品或服务，节省开发时间。从总体上看，AI应用模型效率化生产平台可提供较为前沿的技术、经济合理的模型生产经验以及为实现敏捷开发而打包的数据、算力与算法资源。具体而言，其采用自动机器学习技术，很大程度上降低了机器学习的编程工作量、节约了AI开发时间、减轻了对专业数据科学家与算法工程师的依赖，让缺乏机器学习经验的开发者用上AI，加快开发效率。

API调用量与边际成本及开发效果的关系



来源：艾瑞咨询《2020年中国人工智能API经济白皮书》；艾瑞研究院自主研究绘制。

AI应用模型效率化生产平台性能曲线



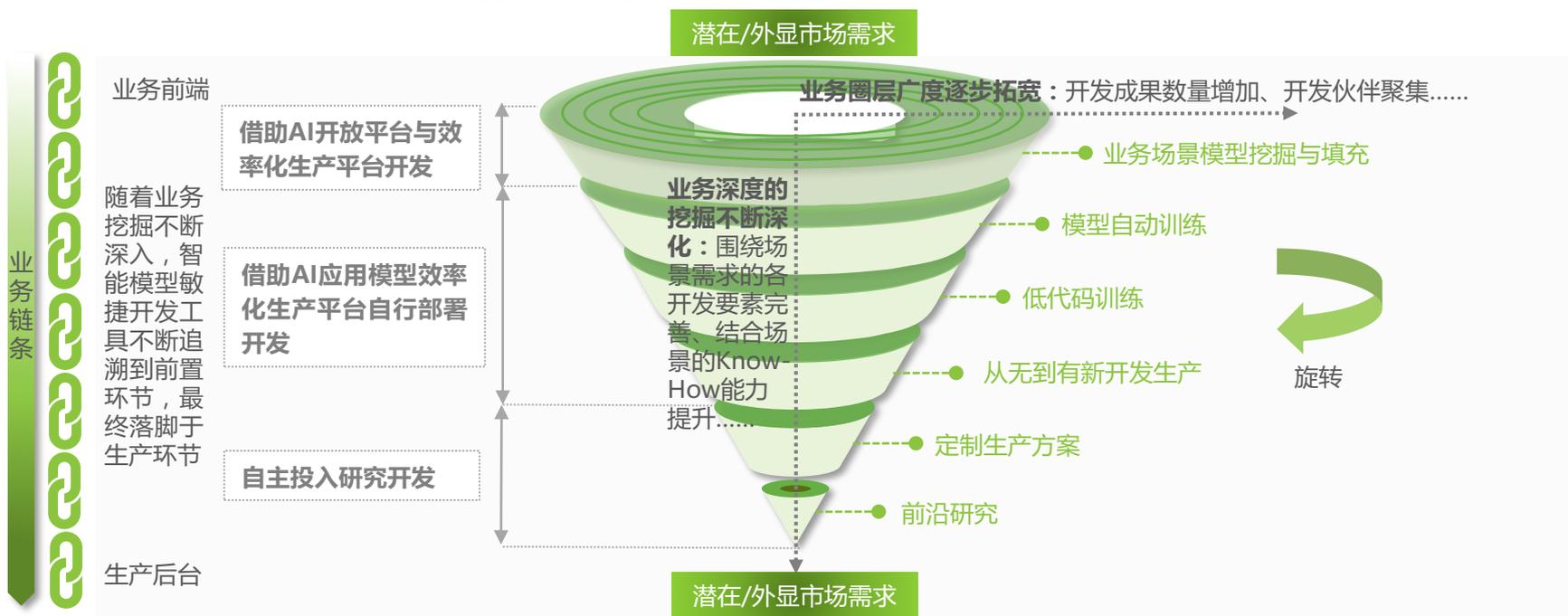
来源：艾瑞咨询《中国人工智能应用模型效率化生产平台白皮书》；艾瑞研究院自主研究绘制。

智能模型敏捷开发工具增长模型

API与定制方案共拓产业广度与深度，AI柔性生产贴近需求

在人工智能产业发展的过程中，智能模型敏捷开发工具可持续拓宽与深挖AI业务的广度和深度。从广度讲，AI开放平台形成平台效应，调用平台API的开发者聚集创新，针对不同业务场景的开发成果数量逐渐增多，提高了技术产品的利用率，打造出轻量化的输出模式、降低单位开发成本，并且构建出动态更新的服务池；与此同时，一站式AI应用模型效率化生产平台逐步填充因场景多元化而衍生出的长尾业务模型，丰富模型供应市场的种类与数量。从深度讲，二者均从业务前端发掘潜在或外显的市场需求，针对刚需应用与高价值环节延伸出多条增量建设与运行需求业务线，瞄准市场风口的同时，敏捷、经济地消化个性化或碎片化需求，根据需求柔性匹配生产。

智能模型敏捷开发工具助力AI业务增长螺旋



来源：艾瑞研究院自主研究绘制。

AI开放平台综合能力维度

服务可用、高性价比、性能稳定、功能丰富是主要关注点

AI开放平台作为API的载体，其受众主要为AI开发者。众多开发者在选择AI开放平台时最为关心的维度包括：服务是否可用、服务的性价比高低、平台接口功能是否丰富、接口性能好坏、服务管理质量高低、平台设计是否简易、AI能力编排强弱、应对场景长尾化的能力强弱。在上述方面更具竞争力AI开放平台更受开发者青睐。

AI开放平台综合能力维度评价罗盘

- 人工智能的细分场景和细分应用极多，**通用性强**的平台可将适配度高的资源应用于各种方案，故打造能同时供应CV、NLP、智能语音等AI技术的通用AI开发平台是应对长尾细分业务的必要准备
- 在AI基础服务之上进行服务编排与组织，可针对性形成语音转文本、智能推荐、客服质检等带有业务色彩的AI服务，而这一点较为考验平台提供商的AI编排能力
- AI技术厂商需警惕落入“**聪明综合征**”的陷阱，避免追求算法的时髦与浮华，**清晰有效**的设计才是开发者所需要的
- 平台资源调用的体验贯穿整个服务周期，服务商需配备好相应的团队应对服务周期各个环节的不同需求



- 一方面，平台资源的部署最终都落脚于各行各业的具体业务场景中，需考虑**资源与业务的耦合度**；另一方面，后台带宽、存储、计算等硬件设施影响平台运行时间的长短，需做好**底层设施准备**
- 在开发者中，绝大多数为**中小企业**，其开发预算有限，所以在挑选AI开放平台时会着重考虑性价比，因此平台供应商应合理选择收费模式、科学定价
- 优秀平台的API接口应做到**快速调用与查询**不同技术能力与海量数据，避免开发者重复寻找接口、造成时间的消耗与浪费
- 具备稳定性能的AI开放平台才能够连续释放资源与能力，这就要求平台接口的性能指标要达到较高的标准，配备强大的**基础计算资源**将带来接口稳定性优势

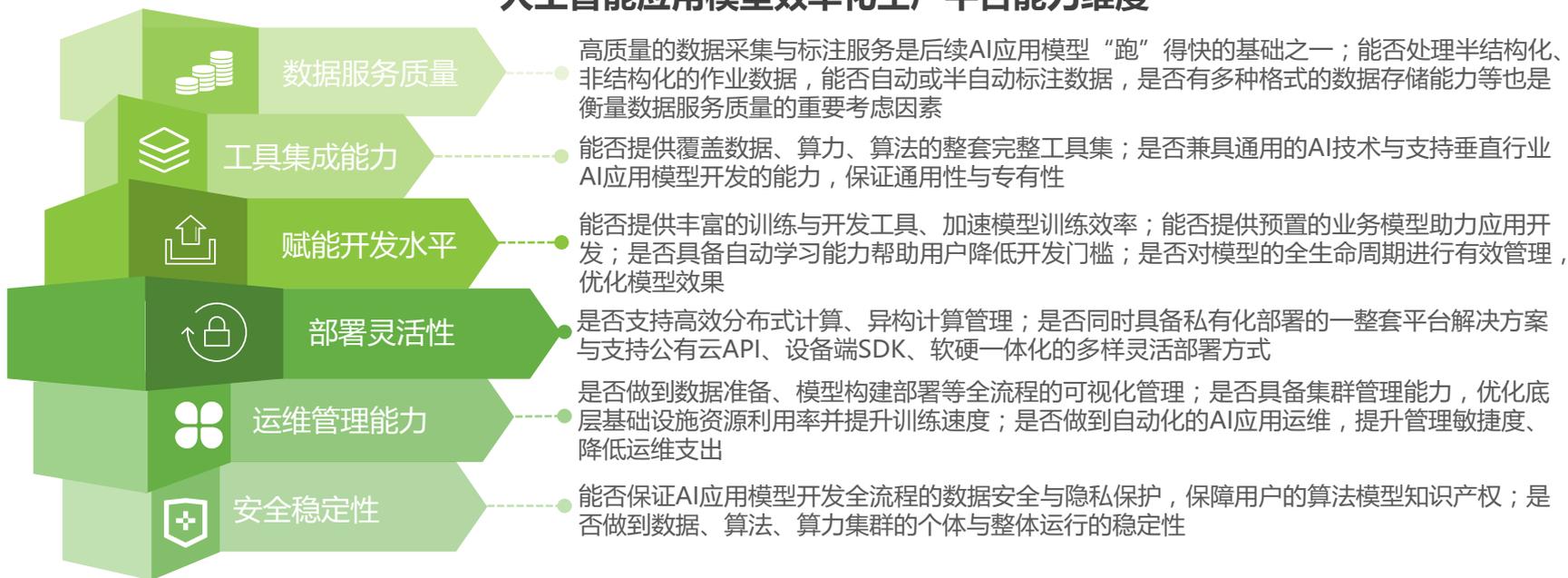
来源：艾瑞咨询《2020年中国人工智能API经济白皮书》，艾瑞研究院结合公开资料自主研究绘制。

AI应用模型效率化生产平台能力维度

六大衡量要点，衡量侧重点因受众而异

AI应用模型效率化生产平台的受众群体通常以规避资源浪费、降低试错成本为出发点，并结合自身业务特点、应用模型开发规划与发展战略来衡量AI应用模型效率化生产平台的能力。衡量的维度包括数据服务质量、工具集成能力、赋能开发水平、部署灵活性、运维管理能力与安全稳定性六大要素，且每一要素在不同受众面前所获的期待程度也各具差异。对于以企业为代表的受众而言，如何借助AI应用模型提高其自身业务价值才是关键。但并非每家企业都具备足够强大的实力让AI与业务深度融合，所以需要AI厂商提供灵活高效、成本合理、可以切实解决业务问题的AI应用模型效率化生产平台。具体而言，技术水平高的企业主要关注引擎算法模型；而预算有限、技术能力欠缺的中小企业则比较看重项目开发流程的易用性及运营管理难度。

人工智能应用模型效率化生产平台能力维度



来源：艾瑞研究院自主研究绘制。

数据基础服务与治理平台

AI数据基础服务与数据治理产业图谱

中游服务商提供可用数据助力AI落地

AI基础数据服务与数据治理产业链结构为：1) 上游：数据源与数据产能；2) 中游：数据产品开发工具与管理服务；3) 下游：人工智能应用。处于中游的两类服务商中，AI基础数据服务商使用数据采集与标注工具处理图片、语音、文本等非结构化数据，面向AI的数据治理平台服务商则负责使用数据治理的各组件管治多源异构数据，使其形成数据资产，从而提高数据质量。二者处理后的数据可直接为下游的AI训练所用，使AI应用落地能够省时省力。

AI数据基础服务与数据治理产业图谱



注释：图谱中所展示的公司logo顺序及大小并无实际意义，不涉及排名。
来源：艾瑞研究院自主研究绘制。

人工智能数据基础服务定义

以AI训练与调优为目的提供的数据采集、标注与质检等服务

人工智能基础数据服务是指为各业务场景中的AI算法训练与调优而提供的数据库设计、数据采集、数据清洗、数据标注与数据质检服务。整个基础数据服务流程围绕着客户需求而展开，最终产出产品以数据集与数据资源定制服务为主，为AI模型训练提供可靠、可用的数据。数据采集、数据标注与数据质检是较为重要的三个环节。数据采集是数据挖掘的基础，提供多源的一手数据和二手数据；数据标注对数据进行归类与标记，为待标注数据增加标签，生产满足机器学习训练要求的机器可读数据编码。数据质检为数据的客观性和准确性设置检验标准，从而为AI算法的性能提供保障。AI基础数据服务商可着重在以上三个环节建立壁垒，以巩固行业地位。

人工智能基础数据服务流程与主要产品



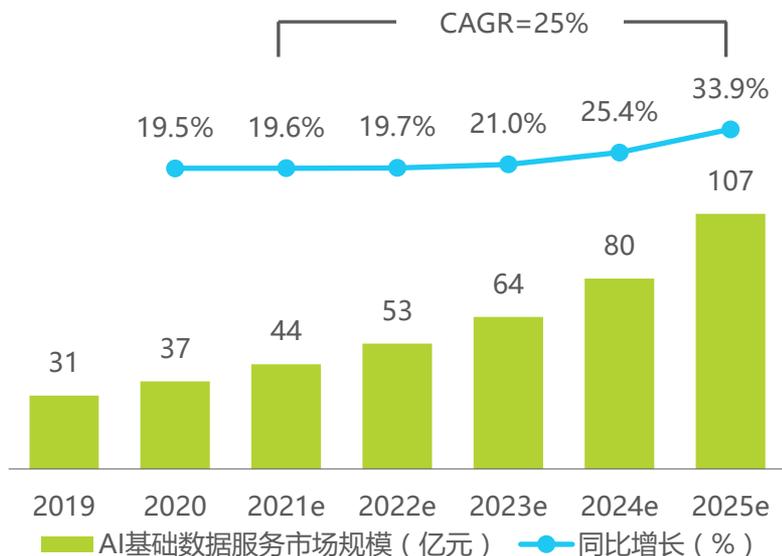
来源：艾瑞研究院根据公开资料自主研究绘制。

AI基础数据服务市场规模

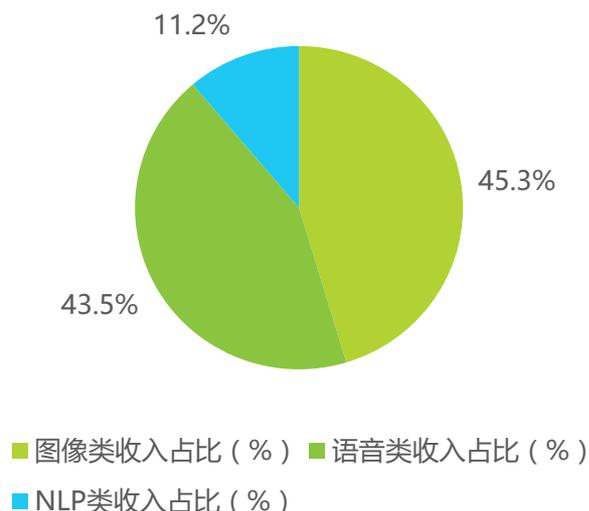
行业规模稳步向前，图像、语音类内容继续向新兴场景开拓

高质量的数据是提高AI应用模型训练速度与精度的必要准备之一，而行之有效的AI基础数据服务又为提高数据质量奠定了坚实的基础。因而，提供通用化、精细化、场景化的AI基础数据服务才能满足日渐增长的AI应用模型训练需求。据艾瑞统计与预测，2020年我国AI基础数据服务市场规模（含数据采集与标注）为37亿元，到2025年，相应规模可达到107亿元，2021-2025年的相关CAGR=25%，整体增速呈现稳步提升的趋势。从市场细分收入结构来看，图像类与语音类收入占总收入规模的88.8%，是业务的主要构成部分；图像类与语音类收入基本持平，图像类业务以智能驾驶与安防为主，而语音类业务以中英大语种、中国本土方言以及外国小语种为主。目前，行业中也相应地分成了以图像类或语音类为主的供应商阵营，各类供应商将会继续立足于主营业务，深挖现有应用场景的业务细分需求，从而带动未来收入的增长。

2019-2025年中国AI基础数据服务市场规模



2020年中国AI基础数据服务市场规模细分结构 分类：按业务内容



来源：艾瑞研究院通过统计国内主流的数据采标服务商数据，结合行业趋势分析而得。

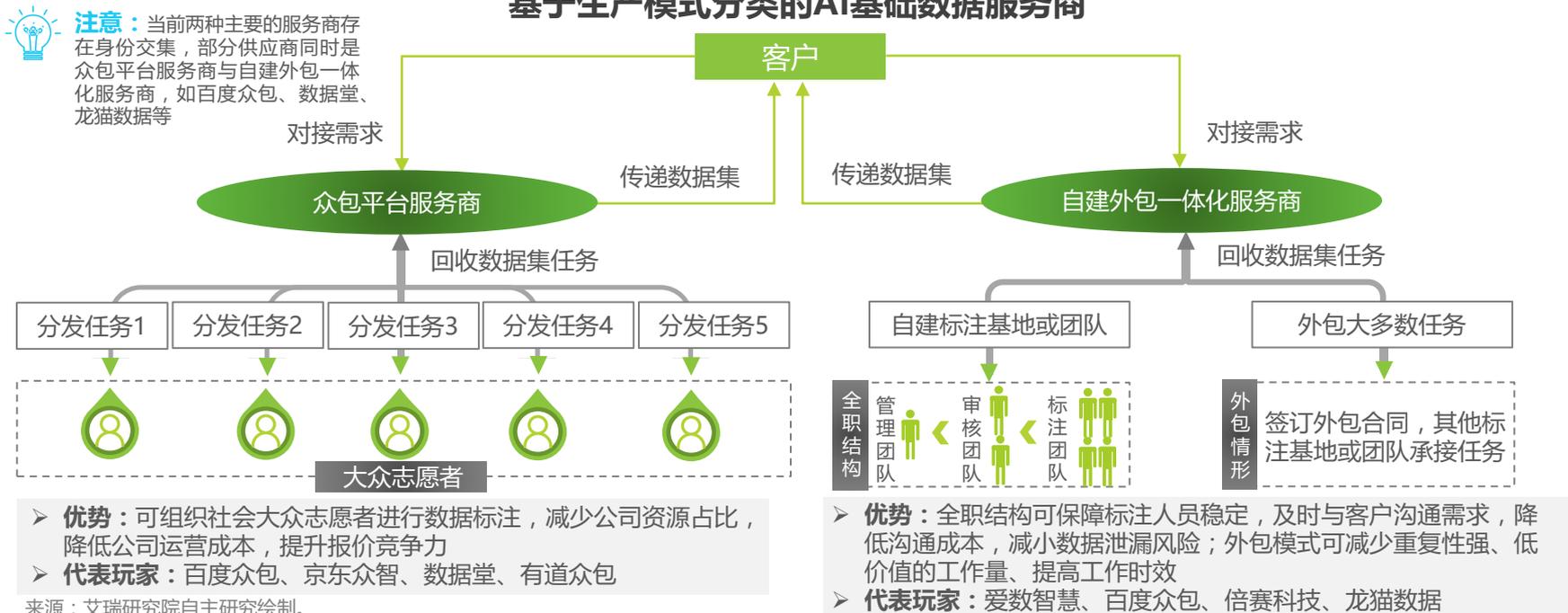
来源：艾瑞研究院结合专家访谈、供应商业务结构自主研究绘制。

从生产模式看AI数据基础服务商

两种身份存在交叉，众包与外包辅助专业开发

AI基础数据服务属于TO B业务，现市场上有两种主流的服务商：众包平台服务商与自建外包一体化服务商。众包平台服务商是连接客户与大众志愿者的纽带，其先对接客户的数据标注需求，而后根据需求分发任务包给大众志愿者，形成“需求客户——众包平台服务商——大众志愿者”的众包结构。自建外包一体化服务商既自己提供全职的采标人员执行任务，又将部分任务外包给其他标注基地或团队，兼具标注团队管理的专业性与繁琐任务外包的灵活性。值得注意的是，两种身份的服务商并非完全割裂的关系，而是存在交叉的关系，也就是说，部分企业同时拥有两种身份。在众包与外包模式下，服务商有更多的余力专注于数据集产品的开发与项目执行，而自建模式下的服务商则可以培养专业的执行团队，深化对数据集产品的理解。随着市场需求的变化，众包形式因专业性与稳定性等缺陷，会导致业务的支撑力度减弱。

基于生产模式分类的AI基础数据服务商



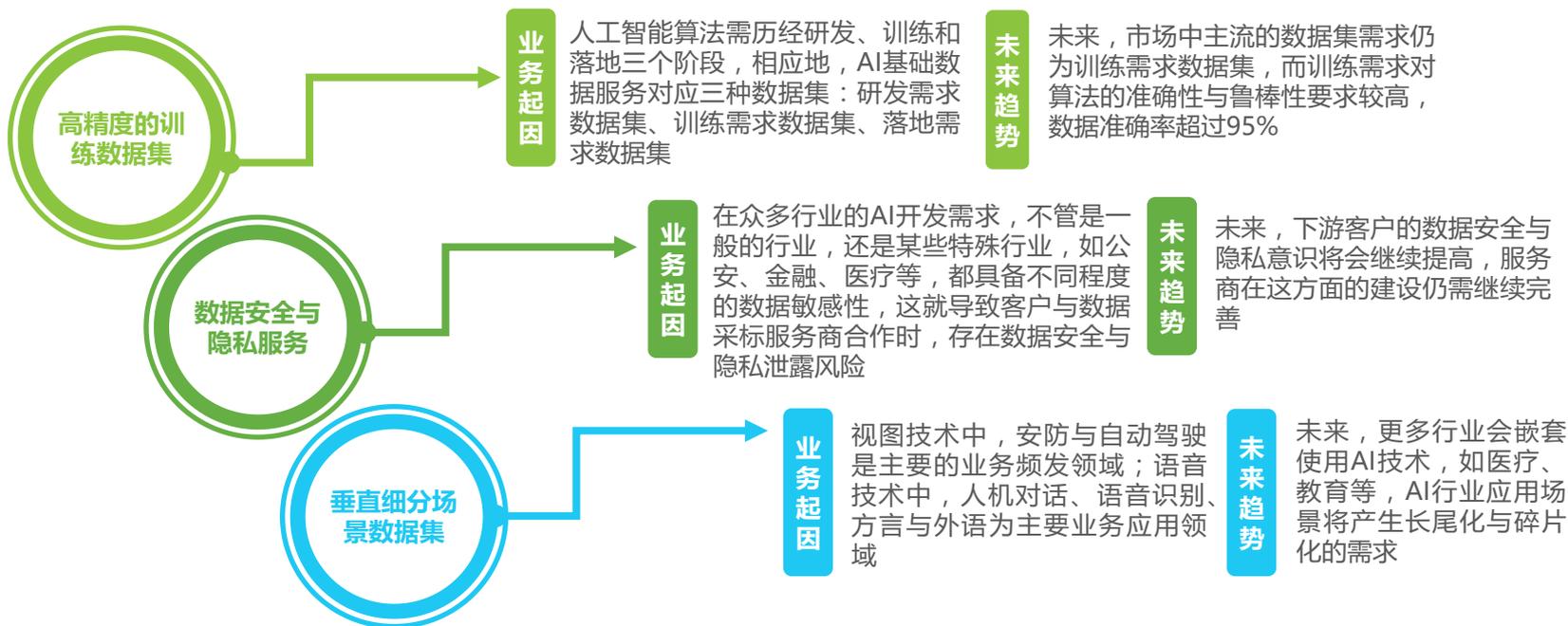
来源：艾瑞研究院自主研究绘制。

AI基础数据服务的业务增长点展望

数据精度抬升、垂直场景细分、安全隐私意识提高

目前以及未来3-5年内，需求方在选择AI基础数据服务时，主要会考虑三个因素：1) 服务商是否有高精度的训练数据集以及数据精准度提升能力；2) 服务商是否有垂直细分场景数据集以及相应的开发能力；3) 服务商是否具备数据安全隐私资质与对应的私有化部署能力。高精度的训练数据集与垂直细分场景对服务商的标注能力提出更高的要求，需要投入更多的成本，拥有专业的标注团队、能承担新需求所引发业务成本、紧跟AI需求变化的服务商才能在标注门槛不断上涨的发展趋势中存活。此外，为数据的安全与隐私提供敏感性保护服务，防止客户数据与隐私泄露，也是服务商提升自身业务竞争力的必要选项。

未来AI基础数据服务主要业务增长点



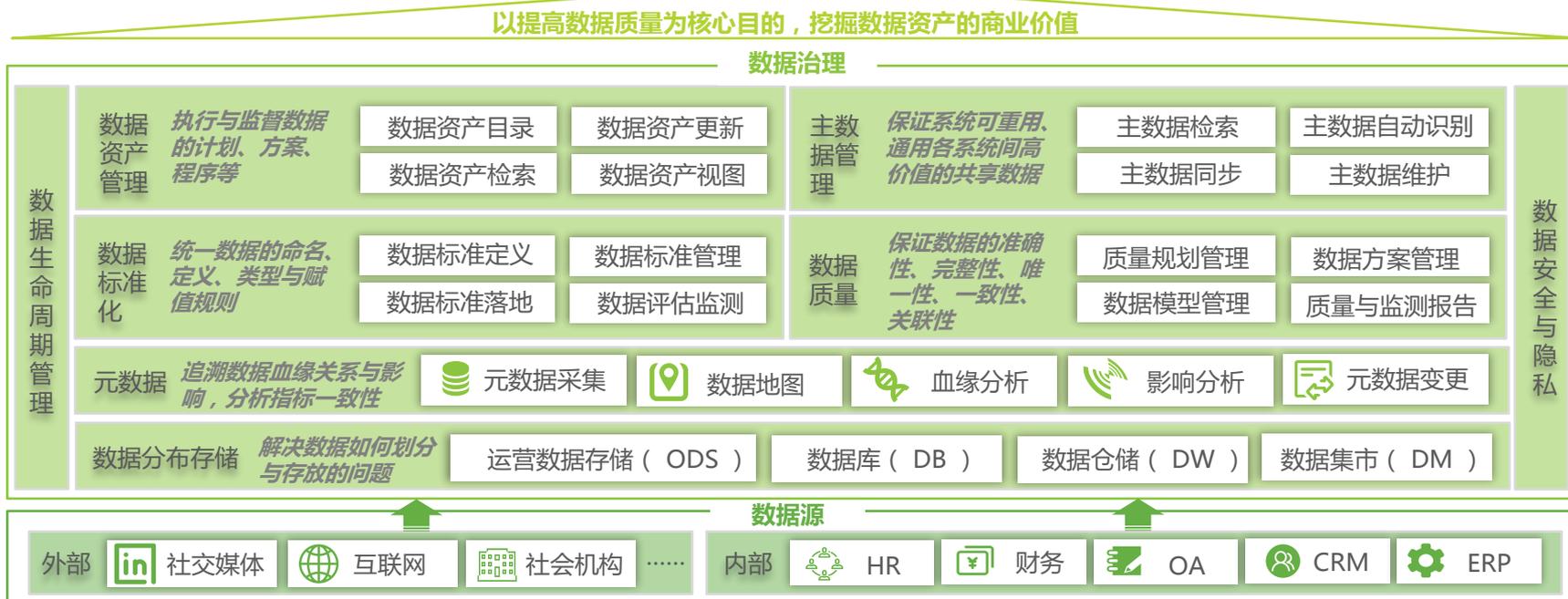
来源：艾瑞研究院自主研究绘制。

传统数据治理介绍

提高数据质量，挖掘数据资产的商业价值

由于企业的数据来源多样，企业内外部系统以及内部各系统的数据之间的互通性欠缺，常常导致信息孤岛、数据口径不统一与数据错漏等问题，数据治理需求就随之产生。传统的数据治理是指使用各个相关的数据技术组件、持续性地对数据进行规整与管理。数据治理技术组件的工作流程一般为：数据分布存储-元数据-数据标准化与数据质量-数据资产管理与主数据管理。因为数据生命周期管理覆盖数据的生产、加工与销毁，而数据安全须确保每个环节的数据存储、流转和使用得到加密与保护，故二者贯穿数据治理的全流程。数据治理的意义在于持续不断地提高数据质量，最大程度地挖掘隐藏在数据资产背后的商业价值，为企业管理者改善决策与规划提供依据。

传统数据治理示意图



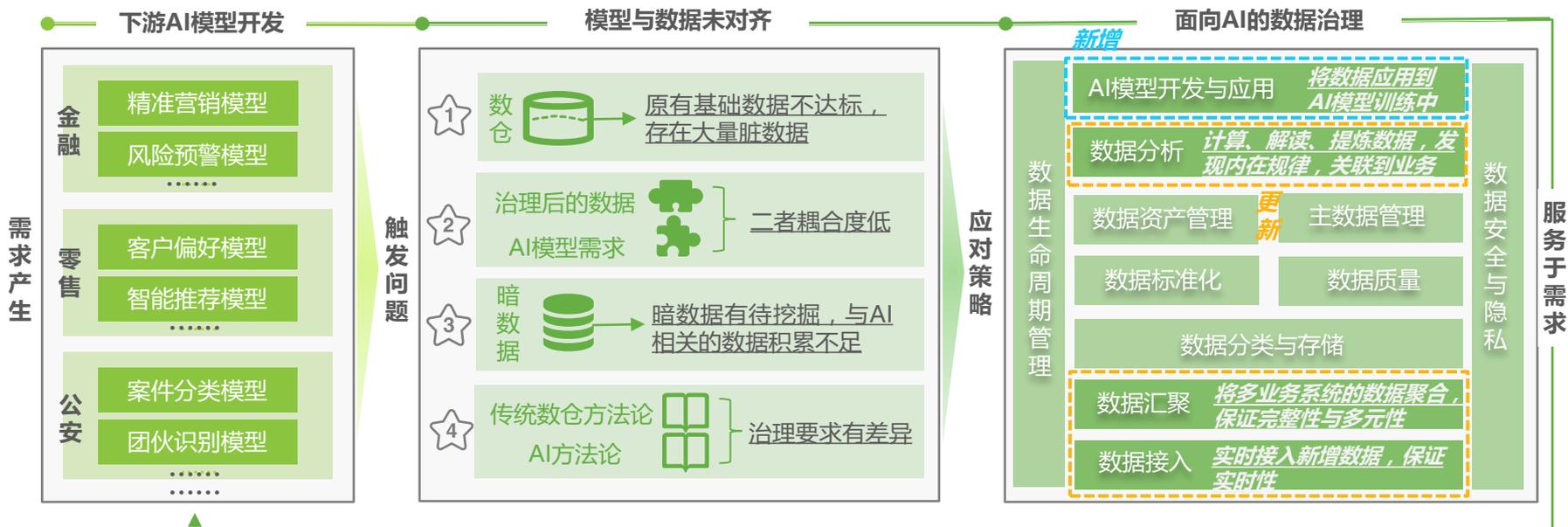
来源：艾瑞研究院自主研究绘制。

面向人工智能的数据治理定义

产生于业务，围绕于AI，追溯于治理

在大数据时代背景下，金融、零售、公安、工业等不同行业的业务场景衍生出诸多应用，多元的AI模型开发需求因此产生。AI模型开发的原材料是数据，但在挖掘模型数据时，往往面临模型与数据无法拉齐的问题，所以需要溯源到前置环节，从一开始就把数据治理的工作做好，面向AI的数据治理这一概念也就由此出现。面向AI的数据治理是指，以具体业务产生的AI模型开发与训练为目的，使用各个数据组件与人工智能技术，对数据进行针对性与持续性的诊治与管理。相比于传统的数据治理，其更新了数据接入、数据汇聚、数据分析的功能，并新增了AI模型开发与应用组件，以应对海量实时数据迸发、模型需及时对接数据等情形。面向AI的数据治理的特点在于，其对接企业现有的数据、积累新的AI数据而非重新进行AI数据的数据库建设，而且提供针对实时数据的处理办法、优先解决业务落地的困难，并持续挖掘具体业务的数据资产价值。

面向AI的数据治理的触发逻辑



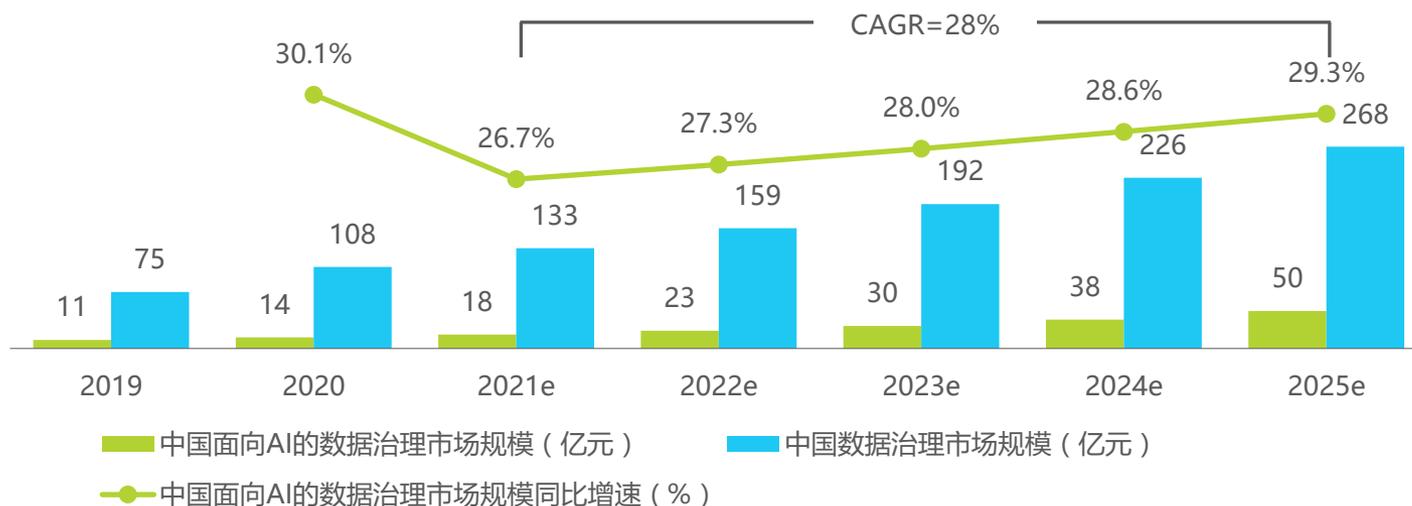
来源：艾瑞研究院结合专家访谈与公开资料自主研究绘制。

面向AI的数据治理市场规模

存在数据中台带来的业务冲击，后期恢复稳步爬坡态势

在数字化转型与智能化转型的大趋势下，数据治理工作一般伴随着数据中台的搭建以及AI应用模型的开发而展开。数据治理属于数据中台的构成组件，数据治理的工作与服务则属于数据中台建设、运营与维护中不可缺少的环节。与此同时，以AI应用模型所需的数据标准去治理数据，结合AI技术提升数据的可用度与模型的训练效率，才能够更好更快地为人工智能业务前端服务，改善供应商的业务流程与消费者的消费体验。据艾瑞统计与预测，2020年我国面向AI的数据治理市场规模为14亿元，到2025年，相应规模可达到50亿元，2021-2025年的相关CAGR=28%。2018年，数据中台概念兴起，其规模在2020年处于爆发点，而数据治理作为数据中台的组件，也于同期迎来增长爆发点，从而带动面向AI的数据治理。2020年后，数据中台市场规模增速开始降温，数据治理也随之回落，面向AI的应用模型开发业务在该过程中的带动作用有限，故2021年的业务同比增速出现拐点。后期，面向AI的应用模型开发业务的带动效应逐步凸显，规模增速呈现稳步爬坡态势。

2019-2025年中国数据治理与面向AI的数据治理市场规模



来源：艾瑞研究院通过统计分析数据中台、数据治理、面向AI的应用模型业务的相关数据与信息，结合桌研行业趋势与专家访谈自主研究绘制。

面向人工智能的数据治理的内在AI技术

数据治理类脑系统=业务场景模型+NLP+KG+ML

面向人工智能的数据治理的先后使用NLP——知识图谱——机器学习三项认知智能技术，这三项技术作用于数据治理的多数组件，影响着数据处理能力的强弱，进而影响着AI应用系统智能化程度的高低。读时模式数据完成接入与汇聚后，NLP便开始一系列提高数据可读性、可用性与准确率的工作，将数据转化为人类可理解的文本。知识图谱是紧随其后的一项技术，主要负责构建具备逻辑推理关系的框架、模型，进行知识挖掘与符号推理，但当前NLP大数据处理的准确率有待提升，所以知识图谱的应用还比较受限。机器学习比较特殊，其运用于除数据接入与汇聚外的所有组件，发挥自动化处理的效果，提高数据处理效率。此外，其也是技术环节的最后一环，首先提取知识图谱网络中的特征，然后不断调优算法、训练模型，完成决策与行为动作。至此，面向人工智能的数据治理的类脑信息系统建设基本告一段落。

AI技术与数据治理相辅相成



机器学习 (ML) ——学习“教材”，认知智能加载，“考试”合格后步入社会

所涉及数据治理组件

- 除数据接入与汇聚外的所有组件，较为重要的是AI模型应用环节

处理

特征选择 -----> 训练模型 -----> 测试与评估模型
选择前置环节 处理好的数据 作为变量命名 选择最佳算法后，用训练数据集做训练 对比真实数据与预测数据，判定模型好坏

知识图谱 (KG) ——汇编“教材”，梳理知识提纲

所涉及数据治理组件

- 数据分析
- AI模型开发

处理

知识建模 -----> 知识获取 -----> 知识融合 -----> 知识计算
将传统关系型数据库的ER关系转化为KG的 Schema 导入Schema，采用知识抽取技术处理读时数据，整合到现有的KG中 通过实体消歧、指代消解等知识融合技术对KG的质量进行管理 通过图计算进行知识发现、知识推理和知识挖掘等工作

NLP (主要含NLU与NLG) ——寻找可用的“教材”，让“教材”变得可理解

所涉及数据治理组件

- 数据分类与存储
- 数据资产管理
- 数据标准化
- 主数据管理
- 数据质量
- 数据安全与生命周期

处理

NLU -----> NLG

进行文本表示、实体挖掘 关系抽取、属性提炼等操作，构造词林 进行机器翻译、系统问答，提供结构化的金融数据，生成人类可理解的文本

来源：艾瑞研究院结合专家访谈与公开资料自主研究绘制。

面向人工智能的数据治理供应商分类

三类厂商各具优势，未来行业集中度提升

面向人工智能的数据治理供应商主要有云服务厂商、数据智能厂商与数据解决方案厂商，三类厂商的数据治理业务产生的原因各不相同，也各具优势。云服务厂商因自身集团业务系统繁多、数据体量庞大，存在自发的数据治理需求，其首先向成熟的数据咨询公司学习方法论，内化为自有的数据治理知识与软硬件工具集合并累积实操经验，之后形成通用的数据治理产品对外销售，具备天然的先发优势与坚实的业务基础。数据智能厂商与数据解决方案类厂商则更多出于服务下游客户的大数据需求的原因而产生，前者的数据治理软件、AI算法及技术较为成熟，后者围绕客户的核心需求展开方案设计与产品汇聚，具备系统集成优势。目前，三类厂商是客户的主要选择，行业集中度相对靠拢。未来，行业集中度将随着各类厂商业务的开拓与优势的发挥而继续提升，资金、研发实力与业务经验较为薄弱的中小企业将面临较为残酷的竞争局面。

面向人工智能的数据治理的主要供应商分类

- **特点：**通过与数据咨询公司或AI公司长期合作从而累积方法论，如IBM方法论，在集团内形成通用的数据治理组件，进而打入内部产品线对外出售
 - **主要收费模式：**云服务租用模式
 - **代表玩家：**华为云、阿里云、百度云
- **特点：**把大数据技术软件或AI应用开发作为主业，产品开发过程中，与AI算法和技术结合密切
 - **主要收费模式：**License模式，产品的物权具有永久性，厂商后期收取维保费
 - **代表玩家：**第四范式、明略科技、星环科技
- **特点：**作为数据治理产品的集成商，比较专注于某个行业，如金融、电信，行业属性较强，同时参与软件的设计与研发
 - **主要收费模式：**外包模式，与项目实施方共同抽成
 - **代表玩家：**东方国信、宇信科技、中软国际



来源：艾瑞研究院结合专家访谈与公开资料自主研究绘制。

面向AI业务的数据积累与治理模型

锚准方向，双轮驱动为智能化转型速度与质量赋能

在大数据应用的驱动下，具有相当数据规模的企业多条业务条线往往会产生大量的结构化与非结构化数据，愈加需要企业内部的数据及时融通，但企业不可能完全抛弃现有的数据库系统、更换一套完全符合面向AI业务开发的数据治理系统；另一方面，智能化转型浪潮推动着企业的AI应用开发需求增长，但数据开发缺少统一标准、数据与业务场景割裂，让面向AI的数据治理的工作面临两难的局面。对此，艾瑞认为面向AI的数据治理并非完全舍弃已有的数据治理结构，而是在原有的基础上，进行数据治理结构的改造，让治理工作更多为AI开发而服务，从而完成AI业务数据的积累。面向AI的数据治理工作完成后，才能驱动AI应用模型开发业务高效、高质运行，而业务模型的开发反过来为面向AI的数据治理工作提供指导，业务系统与数据系统像两个锚准工作方向的齿轮，共同滚动。此外，符合业务场景需求的AI应用模型与ML\KG\NLP等AI技术加速促进两个齿轮的转动，使企业的业务系统运转效率向高质高效发展，也为企业带来更可观的智能化转型业务发展速度与业务服务质量。

面向AI业务的数据积累与治理模型简图



来源：艾瑞研究院自主研究绘制。

| | |
|---------------|---|
| 人工智能基础层概念界定 | 1 |
| 人工智能基础层需求篇 | 2 |
| 人工智能基础层供给篇 | 3 |
| 典型人工智能基础层企业案例 | 4 |
| 人工智能基础层发展洞察 | 5 |

SenseCore AI大装置推动人工智能通用型技术研发

商汤科技是领先的人工智能平台公司，是科技部指定的“智能视觉”国家新一代智能开放创新平台。为满足深度学习网络模型不断攀升的算力需求、多行业长尾细分的应用需求以及摆脱人力密集型的开发模式，商汤构建了新型人工智能基础设施——SenseCore商汤AI大装置。其通过整合强大的算力基础和领先的算法能力，致力于拆解与碰撞海量数据、深入挖掘数据的潜在价值、打破认知和应用的边界。在该装置的架构组成中，算力层以商汤自主建立的新型人工智能计算中心（AIDC）为基础，整合AI芯片及AI传感器，集训练与推理于一体，可支持海量数据与算法模型的计算；平台层融合了商汤数据平台、深度学习训练框架、深度学习推理部署引擎和模型生产平台，打造创新的人工智能通用算法开发平台，实现从数据存储、标注到模型训练、生产、部署、测试的全链路、批量化过程；算法层则基于平台层生产的创新算法模型，向企业开发者提供算法工具，为企业赋能，助力其实现智能化升级；同时商汤通过与香港中文大学商汤联合实验室（MMLab）打造算法开源计划（OpenMMLab），与社区开发者共同构建创新生态。

SenseCore商汤AI大装置驱动因素与架构组成

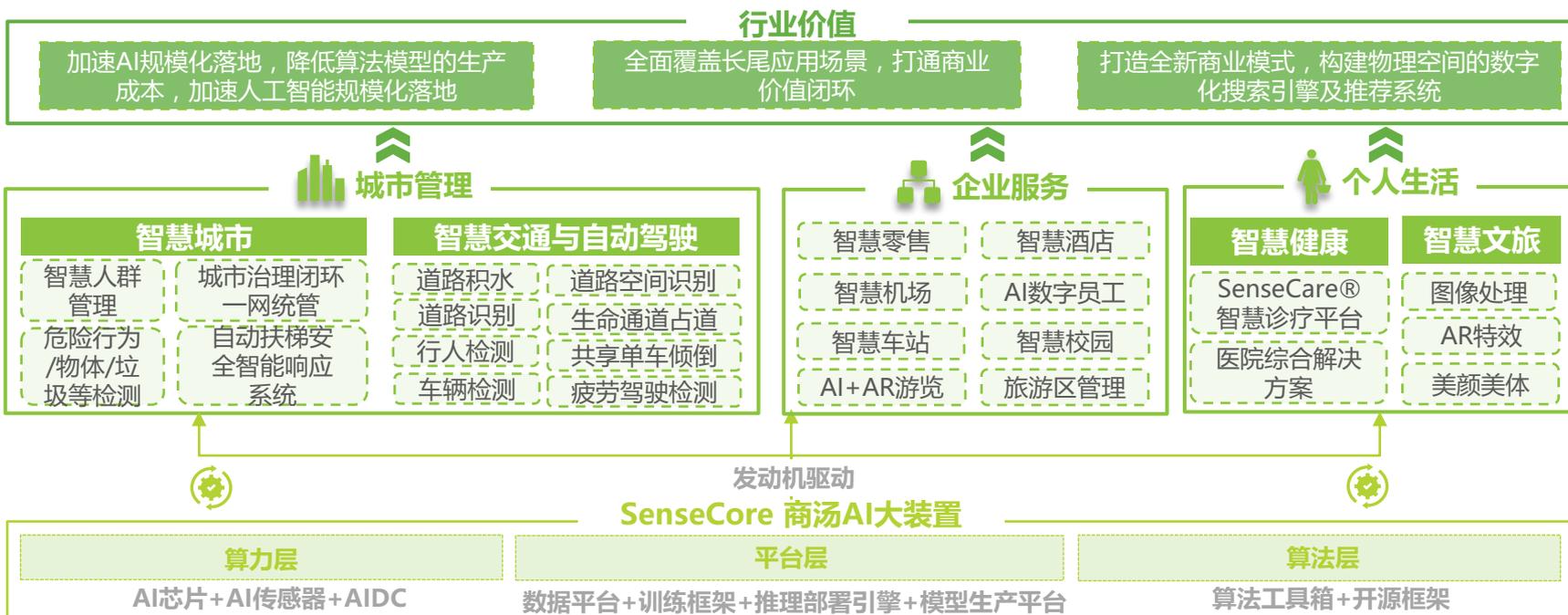


来源：艾瑞研究院自主研究绘制。

实现高效率、低成本和规模化的AI创新与赋能

SenseCore商汤AI大装置是商汤打造的新型人工智能基础设施，作为三位一体的业务智能化发动机，其构建了一整套端到端的架构体系，打通算力、平台和算法之间的连接与协同，能够大幅降低人工智能生产要素的成本，从而实现高效率、低成本、规模化的AI创新和赋能。商汤AI大装置的推出将实现人工智能以自动化、自适应的方式进行生产和落地，解决城市管理、企业服务和个人生活中的长尾应用问题，打通商业价值闭环，进而构建物理空间的数字化搜索引擎和推荐系统。通过SenseCore商汤AI大装置，商汤科技致力提高人工智能的生产效率的本质，推动人工智能进入工业化发展阶段，向着“AI赋能百业”的目标砥砺前行。

SenseCore商汤AI大装置致力“AI赋能百业”



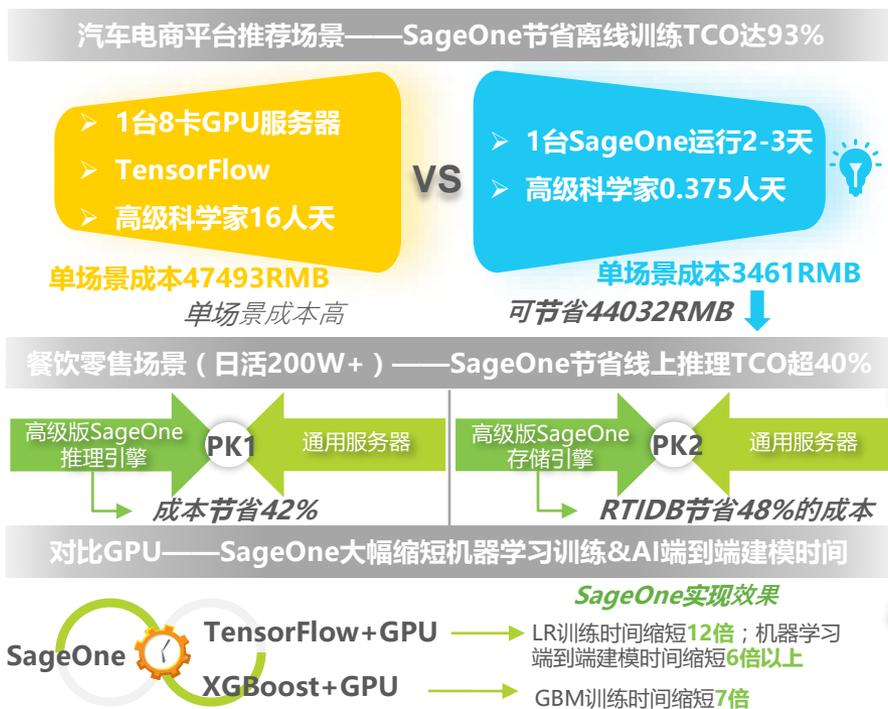
来源：艾瑞研究院自主绘制。

第四范式

提供覆盖软硬一体算力、数据治理及模型开发的AI基础资源

第四范式成立于2014年，是人工智能平台与技术服务提供商。依托国际领先的机器学习技术与丰富的行业实践经验，第四范式可解决企业在智能化变革中面临的AI应用门槛较高、落地价值受阻、算力投入激增等实际难题。其SageOne软件定义算力平台、Sage AIOS 2.0人工智能平台底层操作系统、HyperCycle ML/CV等应用开发工具为下游客户提供了覆盖算力、数据治理与模型开发部署等各模块的AI基础资源。第四范式通过助力下游客户轻松落地AI应用并降低AI总体拥有成本，全面支撑企业智能化能力构建，目前已在金融、零售、制造、医疗、能源、互联网等领域成功落地上万个AI应用。

第四范式SageOne-软件定义算力平台



来源：艾瑞根据公开资料整理绘制。

第四范式Sage AIOS 2.0企业级AI操作系统

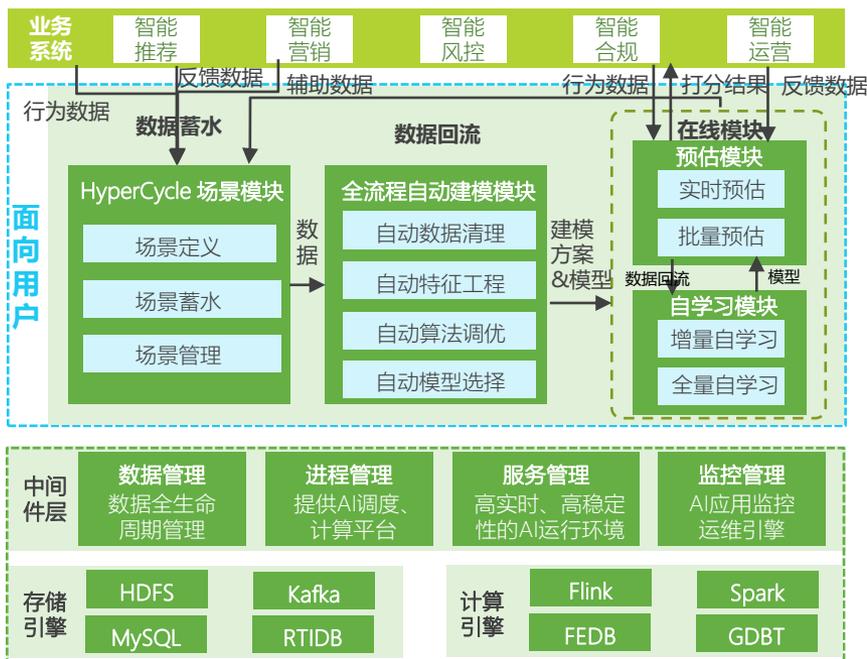


来源：艾瑞根据公开资料整理绘制。

自动决策类机器学习平台赋能企业加速实现智能化转型

AI技术在各领域多业务环节的落地应用可有效驱动营收增长、降本增效、优化业务模式，为企业带来显著收益。现阶段甲方企业对AI应用模型开发的诉求日益增长，但同时面临着AI技术门槛高、规模化落地效率低；模型效果不佳、AI难以提升业务表现；AI整体成本投入昂贵等痛点。第四范式自动决策类机器学习平台基于HyperCycle ML技术架构，可自动完成从数据引入到模型应用的AI闭环全流程，帮助企业降低应用门槛，赋能业务专家参与AI建设；实现高维、实时、闭环自学习的模型效果；从AI全流程出发提升算力应用效率，助力企业端到端解决AI落地难题、加速抢占市场机会、维持领先优势。

HyperCycle ML技术架构实现AI全流程



HCML平台助力金融客户加速实现AI规模化落地

某大型国有银行2019年落地100+业务场景



- 依托HCML快速实现AI应用场景的落地，加速业务AI转型
- 基于AutoML技术的模型效果媲美人类数据科学家模型效果，为业务带来真正的价值
- 帮助客户将上万名数据分析人员转化为AI数据科学家，实现人才梯队培养

某股份制银行落地手机银行APP组合推荐项目



来源：艾瑞根据公开资料自主研究绘制。

来源：艾瑞根据公开资料自主研究绘制。

三大业务基石共同赋能下游五大行业，战略布局升级

爱数智慧是一家多模态数据服务商，为从事语音识别、语音合成、自然语言理解等人工智能领域企业和机构提供数据服务，迄今已与上百家伙伴达成合作。为更好服务下游企业，爱数智慧布局了MagicHub.io开源社区、对话式AI训练数据集与Annotator®5.0智能化标注平台三大业务基石。MagicHub.io开源社区提供多元、海量的开源数据集，让AI开发者迅速适配训练数据、达到开箱即用的效果；AI训练数据集累计时长超15万小时，覆盖60多种语种，包含ASR、TTS等训练数据集，且采集设备、对话场景多样，保障数据的广覆盖、多维度、强时效、高精度与合规性；Annotator®5.0智能化标注平台分为企业私有化部署版和个人SaaS版，在企业端部署时具备简单快捷、易运维、AI辅助高效标注等特点，预计可降低近50%的综合成本，提升100%的数据标注工作效率。三大业务基石相互辅助，共同为下游五大行业提供相应的数据解决方案，赋能下游各行业的场景应用。

爱数智慧战略布局——三大业务基石覆盖五大行业

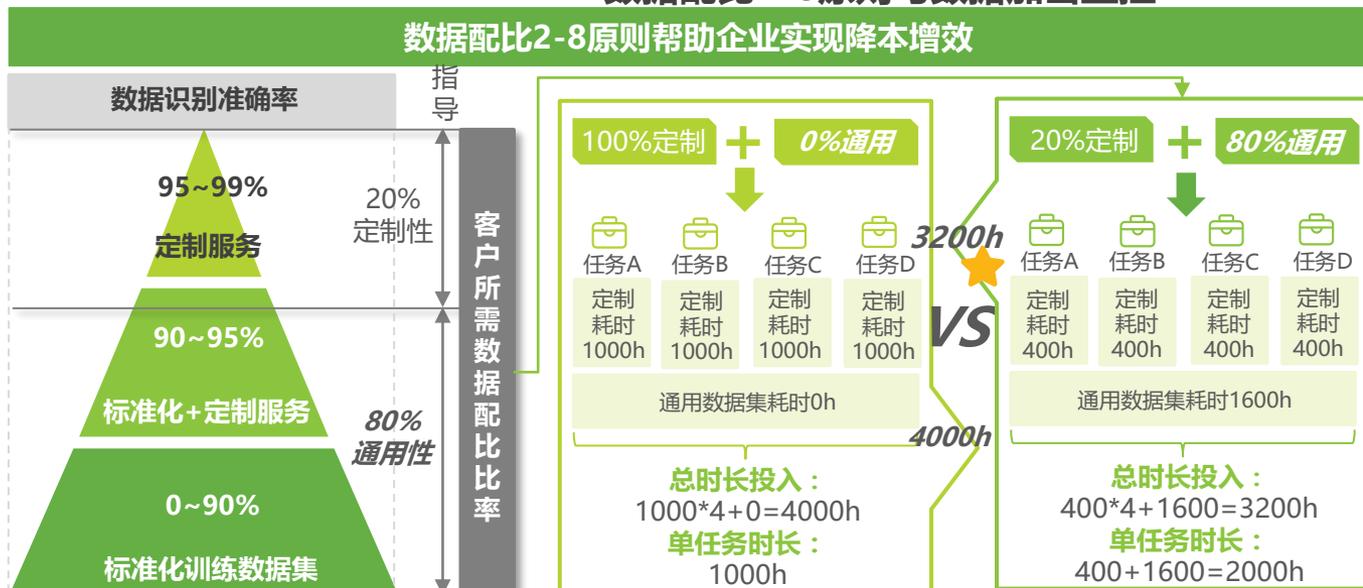


来源：艾瑞研究院自主绘制。

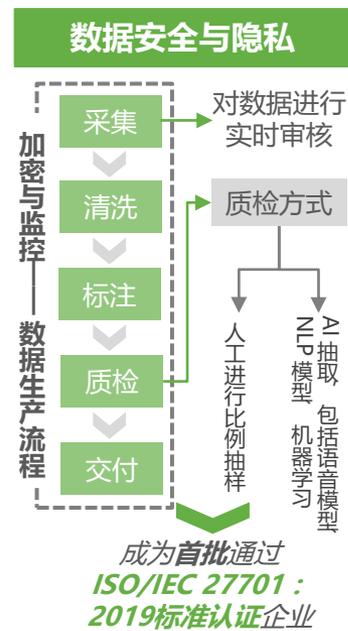
两大助力：数据配比2-8原则与ISO/IEC 27701认证

为帮助下游企业在AI应用模型开发环节中降本增效，爱数智慧提出数据配比2-8原则，即数据配比需考虑80%数据的通用性与20%数据的定制性。数据识别率包括三个等级：第一等级为0-90%，第二等级为90%-95%、第三等级为95%-99%。第一等级的大多数场景只需使用标准化训练数据集即可满足开发要求；第二等级的场景可搭配使用标准化训练数据集与定制服务，第三等级才真正需要定制的数据服务。在实际任务开发中，若客户基于数据通用性的考虑、搭配标准化训练数据集，就可满足其大部分的业务需求，节省完全定制所消耗开发成本，提升开发效率。此外，爱数智慧加密与监控数据的整个生产流程，这为其通过ISO/IEC27701：2019标准认证创造了必要条件之一。在采集环节，爱数智慧对数据进行实时审核，确保数据来源的可靠性；在质检环节，爱数智慧采取人工与AI协作的方式对数据进行检验，NLP模型、机器学习等AI技术提升了质检效率的同时，也保证了审核的准确性。

数据配比2-8原则与数据加密监控



总时长投入降低，开发成本降低；单时长提升，任务处理效果更好



来源：艾瑞研究院自主绘制。

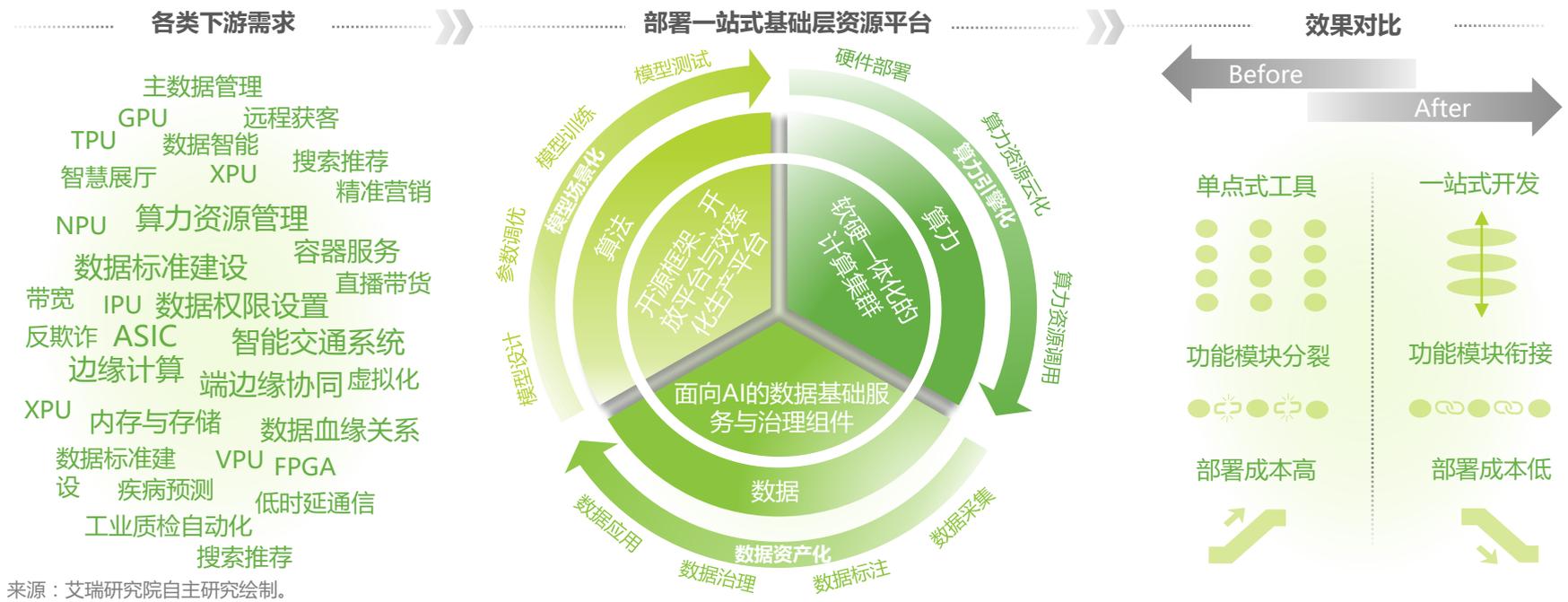
| | |
|---------------|---|
| 人工智能基础层概念界定 | 1 |
| 人工智能基础层需求篇 | 2 |
| 人工智能基础层供给篇 | 3 |
| 典型人工智能基础层企业案例 | 4 |
| 人工智能基础层发展洞察 | 5 |

一站式基础层资源平台

泛在需求下AI模型生产模式的变迁与资源集成

在人工智能由技术落地应用阶段向效率化生产阶段转变的背景下，艾瑞认为人工智能基础层的各模块工具有望走向集约型的生产模式。该模式主要能赋予开发企业以下价值：1) 开发方式改进：从客户需求分析到解决方案部署形成独立的闭环，构建端到端的工作流。在强大算力的支持下，完成数据采集、数据标注、数据治理、数据应用、模型设计、参数调优、模型训练、模型测试、模型推理的全栈式流水线生产。2) 管理效率提升：将数据、算法与算力委托给专业的服务商，实现一站式托管，打通三者之间的衔接壁垒，提高交互友好性，让开发者专注于业务。3) 部署成本降低：集成数据、算法、算力的各个软件与硬件，企业可在一个平台内按需选择自己所缺失的模块组件并自由搭配，有效避免因采购不同供应商的产品或服务而带来的隐性成本损失与显性成本损失。

一站式基础层资源平台生产模式展望



来源：艾瑞研究院自主研究绘制。

基础层全栈自主可控展望

自主可控稳步向前，内外兼修

信创产业涉及到核心技术问题，受到国家的大力扶持。比如，2020年12月，财政部、发改委、工信部等部门就联合发布了《关于促进集成电路产业和软件产业高质量发展企业所得税政策的公告》，文中明确指出：国家鼓励的集成电路线宽小于130纳米(含)，且经营期在10年以上的集成电路生产企业或项目，第一年至第二年免征企业所得税，可见国家对国产芯片的重视。在信创产业稳步推广的趋势下，人工智能基础层的各模块也在逐个突破“卡脖子”的关键点，朝着全栈国产化的方向迈进。算法模块相对其他两个模块而言，因开源框架协助，算法开发相对容易，但依然面临开源框架商用版限制的潜在风险，同时，使用开源框架难以友好对接到AI企业的业务逻辑，基于这两点，部分企业已开始自研开源框架并取得一定成效；数据模块的各类操作系统与应用软件在较大程度上仍以国外企业为主导，而国产的操作系统与数据库等软件配套设施正在稳步崛起，已存在相应的产品与服务可供客户选择；算力模块的智能服务器的国产化率逐步提升，AI芯片虽然仍以英伟达的GPU为主导，但国内部分企业开始自研AI芯片，产生了一批针对通用GPU、ASIC与FPGA的先行玩家。总的来说，基础层全栈的自主可控建设还处在萌芽阶段，未来将在“可用”的建设要求上打好根基，向“好用”的状态演变，并且从以政府政策引导为主的局面向以企业产品自由竞争的局面转变。

人工智能基础层所面临自主控制风险与应对策略

外部策略：以双循环为主，加强外部交互，学习国内外先进企业、前沿开发者的技术能力，在“前人”的基础上加深理解，寻求突破与创新



注意：此处的内外策略同时指代国内策略与企业内部的策略；加强自主研发能力并非重新开始与闭门造车

来源：艾瑞研究院结合公开资料自主研究绘制。



行业咨询

- 市场进入 为企业提供市场进入机会扫描，可行性分析及路径规划
- 竞争策略 为企业提供竞争策略制定，帮助企业构建长期竞争壁垒



投资研究

- IPO行业顾问 为企业提供上市招股书编撰及相关工作流程中的行业顾问服务
- 募 投 为企业提供融资、上市中的募投报告撰写及咨询服务
- 商业尽职调查 为投资机构提供拟投标的所在行业的基本面研究、标的项目的机会收益风险等方面的深度调查
- 投后战略咨询 为投资机构提供投后项目的跟踪评估，包括盈利能力、风险情况、行业竞对表现、未来战略等方向。协助投资机构为投后项目公司的长期经营增长提供咨询服务

关于艾瑞



艾瑞咨询是中国新经济与产业数字化洞察研究咨询服务领域的领导品牌，为客户提供专业的行业分析、数据洞察、市场研究、战略咨询及数字化解决方案，助力客户提升认知水平、盈利能力和综合竞争力。

自2002年成立至今，累计发布超过3000份行业研究报告，在互联网、新经济领域的研究覆盖能力处于行业领先水平。

如今，艾瑞咨询一直致力于通过科技与数据手段，并结合外部数据、客户反馈数据、内部运营数据等全域数据的收集与分析，提升客户的商业决策效率。并通过系统的数字产业、产业数据化研究及全面的供应商选择，帮助客户制定数字化战略以及落地数字化解决方案，提升客户运营效率。

未来，艾瑞咨询将持续深耕商业决策服务领域，致力于成为解决商业决策问题的顶级服务机构。

联系我们 Contact Us

 400 - 026 - 2099

 ask@iresearch.com.cn



企 业 微 信



微 信 公 众 号

法律声明

版权声明

本报告为艾瑞咨询制作，其版权归属艾瑞咨询，没有经过艾瑞咨询的书面许可，任何组织和个人不得以任何形式复制、传播或输出中华人民共和国境外。任何未经授权使用本报告的相关商业行为都将违反《中华人民共和国著作权法》和其他法律法规以及有关国际公约的规定。

免责条款

本报告中行业数据及相关市场预测主要为公司研究员采用桌面研究、行业访谈、市场调查及其他研究方法，部分文字和数据采集于公开信息，并且结合艾瑞监测产品数据，通过艾瑞统计预测模型估算获得；企业数据主要为访谈获得，艾瑞咨询对该等信息的准确性、完整性或可靠性作尽最大努力的追求，但不作任何保证。在任何情况下，本报告中的信息或所表述的观点均不构成任何建议。

本报告中发布的调研数据采用样本调研方法，其数据结果受到样本的影响。由于调研方法及样本的限制，调查资料收集范围的限制，该数据仅代表调研时间和人群的基本状况，仅服务于当前的调研目的，为市场和客户提供基本参考。受研究方法和数据获取资源的限制，本报告只提供给用户作为市场参考资料，本公司对该报告的数据和观点不承担法律责任。

为商业决策赋能

EMPOWER BUSINESS DECISIONS



海量行研报告免费读