

中国云原生AI开发平台白皮书

©2021.11 iResearch Inc.



行业背景：近年来，国内人工智能技术成熟度持续提升、服务种类不断丰富，在企业经营管理各环节的价值已得到市场的初步验证。然而，当前国内甲方企业在进行人工智能开发和应用时仍然面临着技术**人才储备不足、AI应用部署存在困难、投入产出比不达预期**等问题，亟需能够帮助企业解决这一问题的高效AI开发和应用工具。



产品&关键技术：云原生AI开发平台融合了成熟的人工智能开发框架以及云原生工具灵活调用云资源、高效部署云应用的能力，一方面帮助企业开发者**提高算法模型的开发效率**，另一方面**提升交付、部署、运维环节的效率并降低TCO**。横向对比甲方企业可能采用的诸多获取人工智能能力的平台和方式之后，我们认为云原生AI开发平台在AI开发应用全生命周期视角下具备一定的综合优势。



应用场景：云原生AI开发平台在诸多人工智能密集应用的下游场景和行业具备通用性，包括**互联网、金融、自动驾驶、政务、制造、营销**等。本报告挑选了部分应用场景，梳理了上述场景下企业进行AI开发和应用过程中面临的实际需求和难点，展示了典型云原生AI开发产品的服务架构以及对企业管理的价值。



发展趋势：AI开发平台还将朝着易用性、专业化、综合性、产用协同等方向发展，我们认为在这一过程中，AI开发平台的**产品广度将进一步提升**，并有望集成DevOps、AIOps等运维方法和工具，全方位融入企业的数字化经营体系。同时，AI开发平台的服务业态还将向**软硬一体化**方向演进，深度融合技术交流社区等平台，形成**学用一体化**的技术传播与升级环境。

行业背景与产品综述

1

云原生 AI 开发场景与实践

2

云原生AI开发平台发展展望

3

AI应用背景

- 在企业数字化转型的客观需求以及政策对发展前沿IT科技的支持下，我国数字经济高速发展，为人工智能发展创造了积极的经济环境。近年来，国内人工智能技术成熟度持续提升、服务种类不断丰富，在企业经营管理各环节的价值已得到市场的初步验证，伴随云计算的普及和云原生技术的发展，云服务将有望帮助企业解决现阶段开展人工智能应用存在的难点，提升人工智能的效用。



人工智能发展环境 (1/2)

政策引导AI算法的协同开发与AI应用的产业化落地

过去数年间，国务院、国家发改委、工信部、科技部以及各省市、地方政府陆续发布了有关推进人工智能算法开发以及应用落地的政策。在算法开发层面，政策明确倡导开源开放、互助共享的理念，支持具备人工智能资源与技术优势的企业、高校构建促进AI能力开源开放的平台，释放优势互补的协同效应，缩小我国人工智能技术与领先国家的差距；在人工智能应用层面，政策鼓励人工智能等数字化能力在企业层面加大应用力度、在区域层面实现项目落地，通过人工智能等前沿IT技术驱动工业化和信息化深度融合。整体来看，国内政策对于人工智能始终保持积极态度，人工智能产业化和规范化的发展前景会更加明朗。

近年人工智能应用相关政策解读

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》

- 指出**推动人工智能等科技创新重大项目在京落地**，推动各方科技力量优化配置和资源共享，支持包括**人工智能在内的新型研发机构**发展，健全创新创业服务体系。

2021.1.13 国家工信部

《工业互联网创新发展行动计划（2021-2023年）》

- 重点任务中指出鼓励大型企业**加大人工智能等数字化技术应用力度**，全面提升研发设计、工艺仿真、生产制造、设备管理、产品检测等**智能化水平**，实现全流程动态优化和精准决策。

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》

- 建设内容中，提出支撑技术与产品标准，其中“**关键通用技术**”和“**关键领域技术**”涉及机器学习、知识图谱、自然语言处理、计算机视觉和智能语音等重要AI算法和技术的应用。

2019.11.13 国务院（修订）

《国家创新驱动发展战略纲要》

- 发展新一代信息技术，加强**类人智能**、自然交互与虚拟现实研究，推动宽带移动互联网、云计算、物联网、**大数据**等技术研发和综合应用，加快工业化和信息化深度融合。

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》

- 建设原则指出要以企业为主体，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**。

2017.7.8 国务院

《新一代人工智能发展规划》

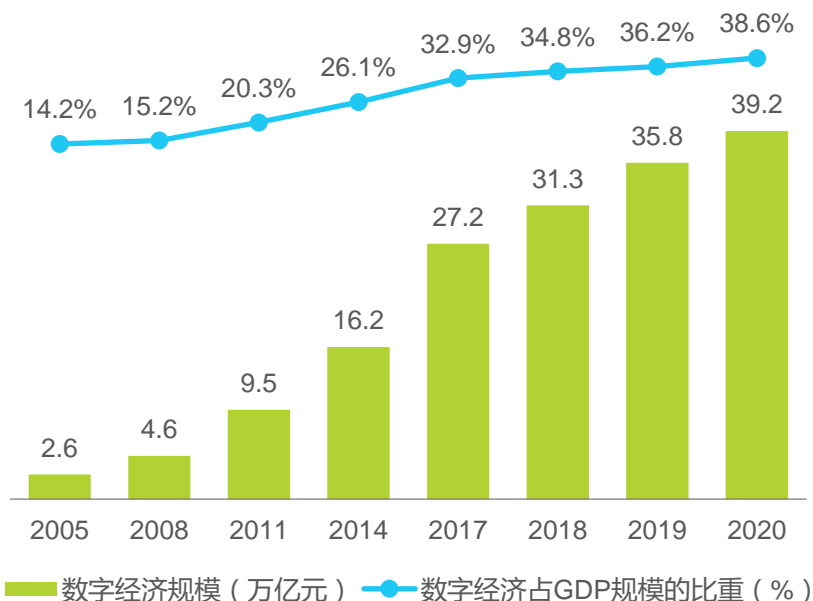
- 指出要将人工智能提升至国家战略高度，以**开源开放**作为基本原则之一，促进**产学研用**各创新主体**共创共享**，**构建开放协同的人工智能科技创新体系**。

人工智能发展环境 (2/2)

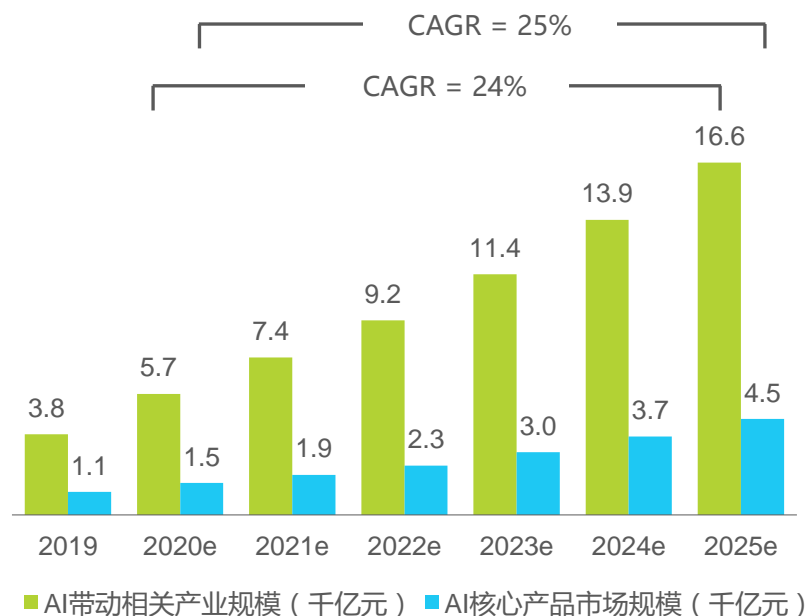
人工智能相关产业规模高速增长，市场需求更加明确

随着数字产业化和产业数字化的不断深化，我国数字经济高速发展，在宏观经济中的重要性持续提升。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，人工智能产业的成长速度令人瞩目，并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发展特征，带动相关产业发展，回馈社会经济。在这样的发展环境下，企业对人工智能的需求逐渐升温，人工智能在企业端的应用成熟度也渐入佳境。

2005-2020年中国数字经济规模及占GDP比重



2019-2025年中国AI产业及带动相关产业规模



来源：中国信通院（2020），艾瑞咨询研究院根据公开资料研究及绘制。

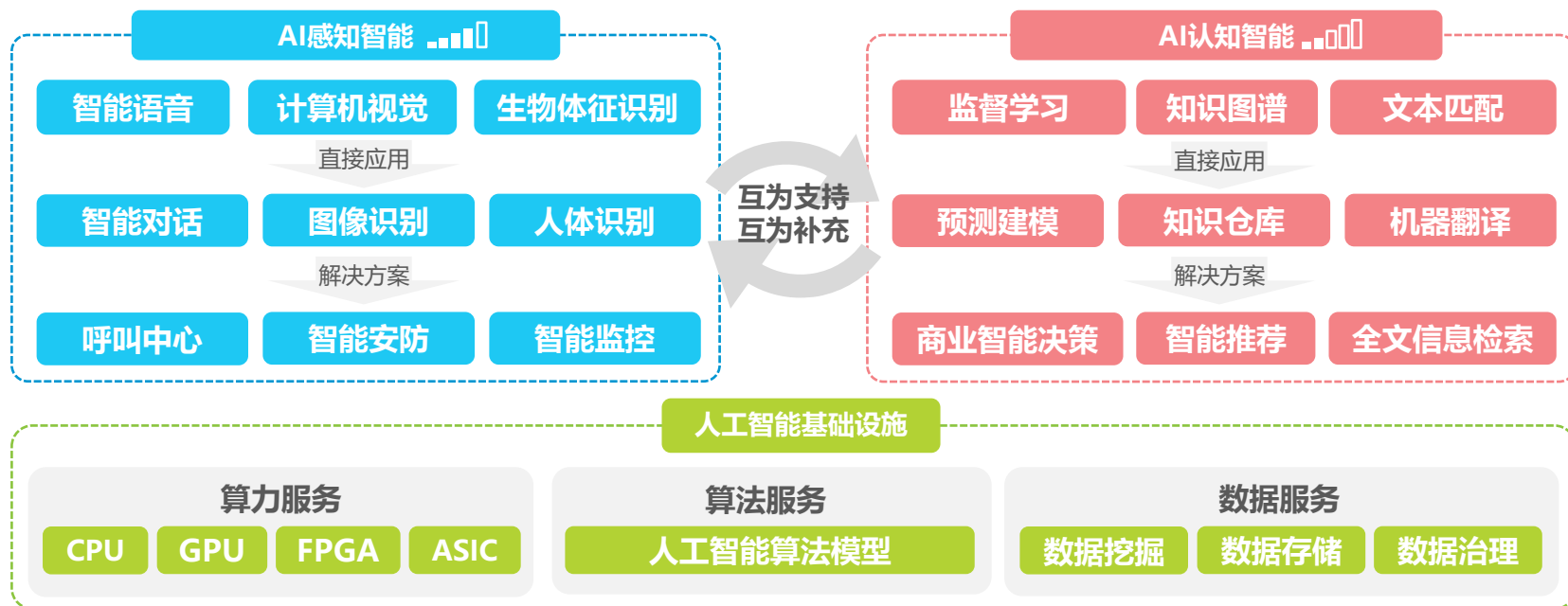
来源：艾瑞咨询《2020年中国人工智能产业研究报告》。

人工智能应用现状 (1/2)

感知智能相对成熟，认知智能加速发展

一般认为，人工智能可分为感知智能和认知智能两大类。感知智能以智能语音、计算机视觉和部分生物体征识别（如体态识别）为核心底层技术，对应的上层应用为智能对话、图像识别和人体识别等，分别直接应用于客服机器人、智能安防、智能监控等解决方案；认知智能以机器学习/深度学习、知识图谱和自然语言处理为核心底层技术，对应的上层应用为预测建模、知识仓库、机器翻译等，分别直接应用于商业决策、智能推荐、全文信息检索等解决方案，二者是互为支持和补充的关系，同一个AI解决方案中往往包含多种技术。目前我国感知智能的算法研发和应用落地相对成熟，而认知智能正在加速发展过程中，未来将有着更广阔的应用空间。

当前AI感知智能&认知智能算法及应用



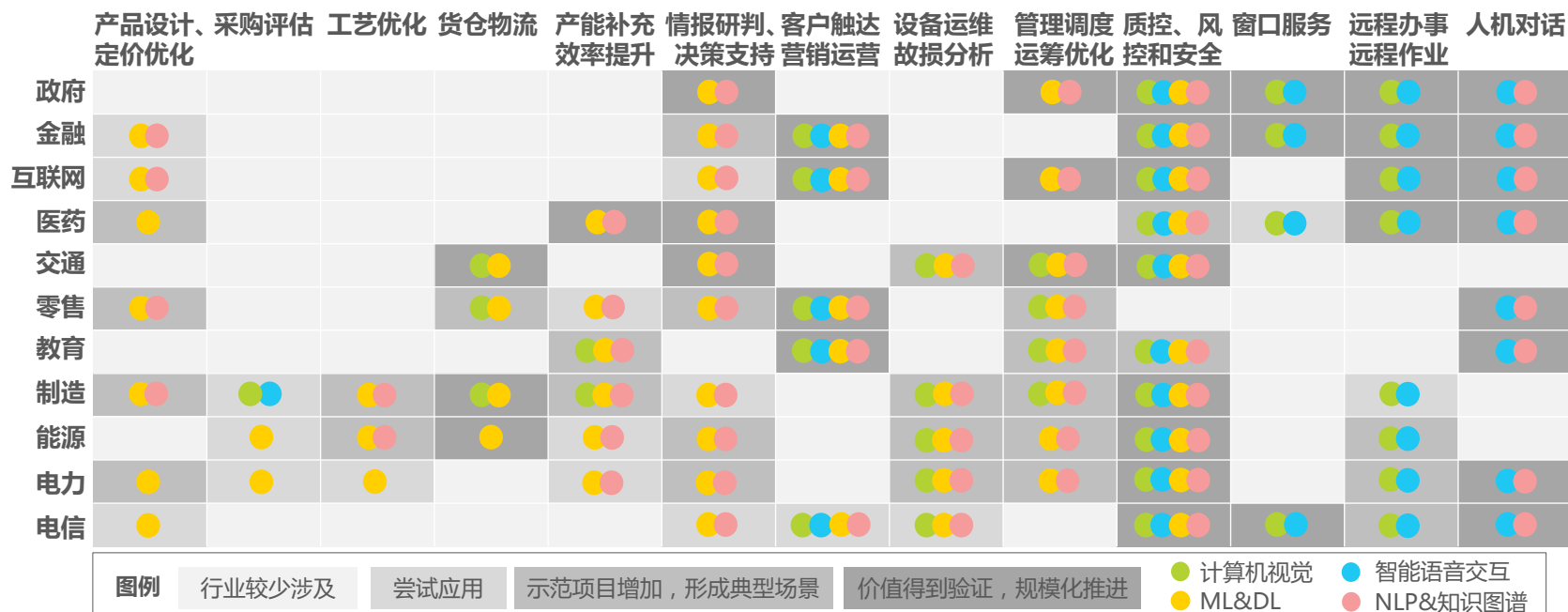
来源：艾瑞咨询研究院自主研究及绘制。

人工智能应用现状 (2/2)

营销、客服、质检、安防等应用的市场价值已获得验证

具体到实际应用中看，人工智能已广泛渗入各行业经营管理的诸多环节，为企业带来更高效的自动化流程、更精准的情报分析以及更智能的运营管理。当前国内人工智能应用成熟度较高的领域包括：①在金融、互联网等需要密集与客户沟通交流的行业中用于智能呼叫、客服、销售等环节，帮助企业提升触达能力和服务质量，从而提升用户粘性；②赋能公安、交通和企业内部的安防监控以及制造业企业的物流配送、产品质检等环节，代替人眼进行大规模监测并提升监测精准度、从而提高调度和管理效率。整体来看，感知智能的诸多应用对于企业的应用价值已得到了市场验证，而涉及认知分析、智能决策的认知领域的应用成熟度也正在逐步提高，已在部分行业展开试水，应用渗透有望加速。

人工智能应用于我国不同行业企业经营管理活动的主要环节



来源：艾瑞咨询研究院自主研究及绘制。

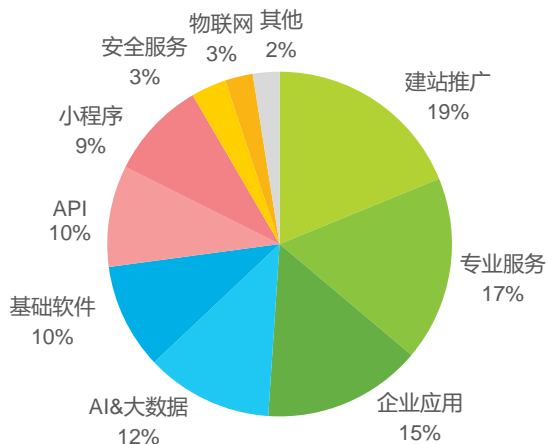
人工智能与云服务 (1/3)

云服务平台成为企业获取和应用AI能力的重要渠道

当前各类前沿信息技术彼此融合促进，界限正逐渐模糊，无论是在技术开发、实施还是应用阶段，包括云计算、大数据、人工智能等在内的多项技术都互为依托和补充。云计算在我国经历了十余年的发展，目前基础云服务已经进入成熟阶段，各类上层应用以云服务平台为技术基座和分发渠道，逐步构建起云上的IT服务生态闭环。具体到人工智能领域，云计算为人工智能研发和部署提供计算集群、存储阵列等基础设施，通过大数据和AI算法PaaS提升企业进行AI开发的节奏和效能，而信息安全、敏捷开发等应用则间接提升了企业采用云上AI能力的稳定性。对于企业的AI开发和应用工作而言，云平台已成为他们加强AI能力的重要助力。

云上人工智能服务品类及应用现状

2021中国云服务产品市场份额



企业AI能力使用情况调研

93%的企业使用云上的AI能力



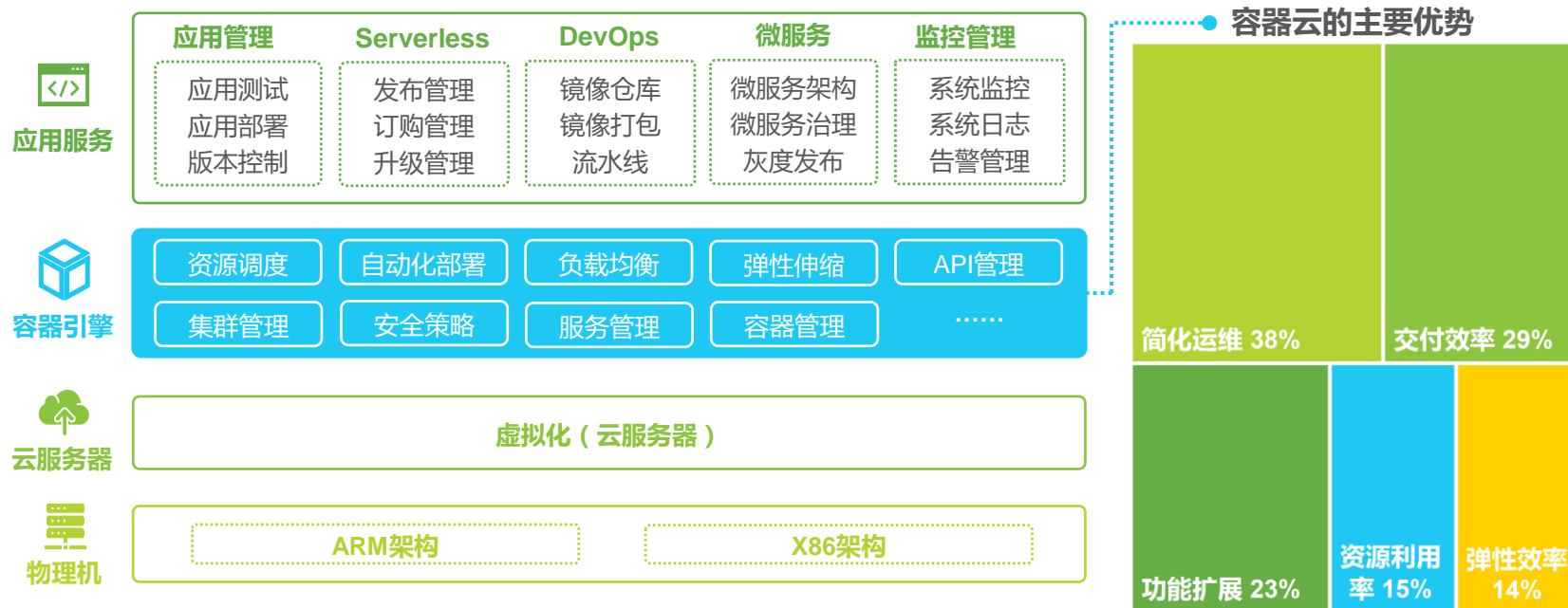
来源：1. 艾瑞咨询《2021年中国基础云服务行业发展洞察》；2. 德勤（2020），艾瑞咨询研究院根据公开资料研究及绘制。

人工智能与云服务 (2/3)

云原生成为业界认可的云计算技术发展方向

以Docker和K8s为代表的容器和容器编排技术是云原生应用的典型代表，容器对基础资源的调用相较虚拟机更加轻量、敏捷、高效，能够直接部署于物理机上作为资源调度器，但在当前的企业用云实践中，无论对于公有化还是私有化部署模式，容器引擎普遍架构于虚拟机之上，对虚拟化的IT基础设施实行弹性资源调度、流程自动化以及集群管理。整体来看，云原生架构具备弹性扩容、敏捷分发、高效易用、兼容适配等主要优势，在云计算成为企业数字化转型标配的今天，云原生带来了更加灵活的用云模式，能够帮助用户降低用云成本、提高云服务可用性和云端服务的质量，其价值已获得了产业界的普遍认可，云原生也被认为是云计算未来的技术发展方向，诸多云端服务也被业界证实能够与云原生架构充分融合并带来使用性能的提升。

容器云计算平台一般系统架构及容器云的主要优势



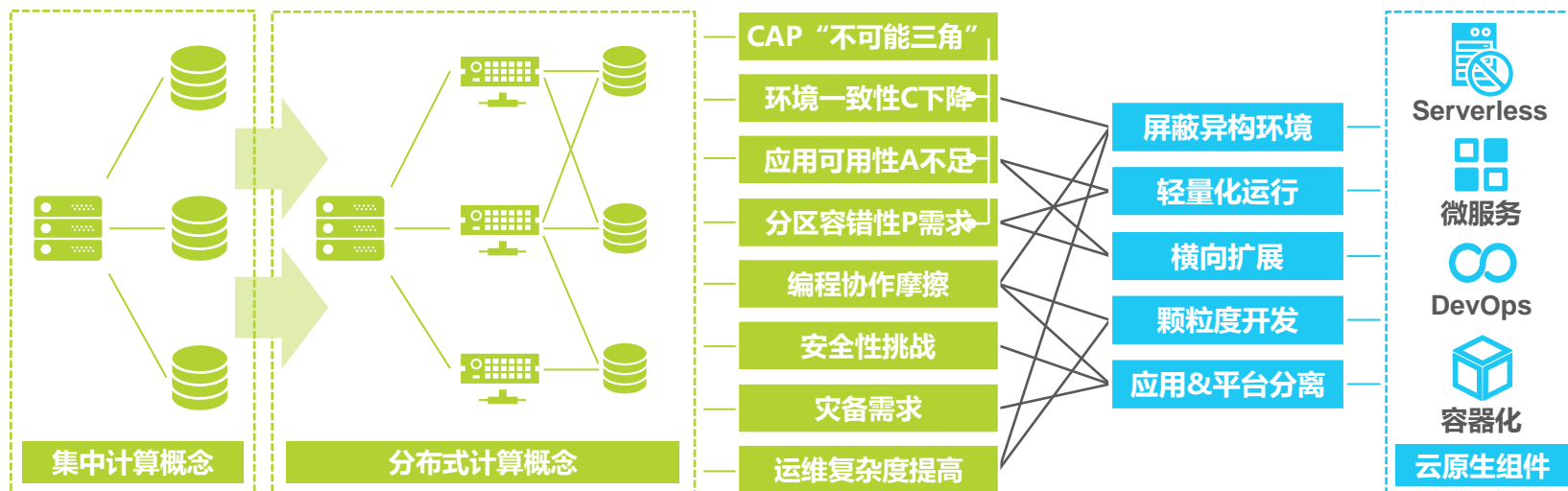
来源：中国信通院（2021），艾瑞咨询研究院根据公开资料研究及绘制。

人工智能与云服务 (3/3)

云原生融合分布式计算性能更具优势，应用前景广阔

伴随着数据量的大爆发和数据应用的丰富，传统的集中式计算架构已无法支撑超大规模的数据处理需求，现今AI应用主要部署在以分布式为基础的云平台之上，AI模型开发以及应用事实上也是以分布式计算为基础。分布式计算通过将数据负载分配到不同的终端进行统筹处理，以异步、并行、多线程的方式提高计算效率，同时，分布式系统的复杂化带来了环境一致性下降、可用性不足、容错能力降低等问题，对于复杂的模型训练和超大规模的应用部署尤其突出。作为云计算未来的总体发展趋势，云原生自诞生之始就以轻量的模块组合以及分布统筹为核心理念，其性能优势可以帮助使用者应对分布式计算架构带来的诸多问题，从而为云原生与人工智能的融合提供了广阔的应用前景。

分布式计算系统架构带来的一般性问题&云原生潜在性能优势



伴随大数据的爆发、AI应用在各场景不断渗透和云计算的普及，基于分布式计算架构进行人工智能模型训练和应用部署既存在客观需求，也具备了技术条件

相较于传统模式，分布式计算带来了计算能力整体的提升，但系统复杂度随之增加，带来包括CAP难题在内的一系列负面影响，而云原生理念不仅与分布式概念天然契合，其应用架构和性能恰在一定程度上为分布式带来的问题提供了解决方案，二者相辅相成

来源：艾瑞咨询研究院自主研究及绘制。

企业AI应用困境

- ▶ 尽管人工智能技术成熟度正在不断提高，但目前国内甲方企业进行人工智能应用仍然面临着技术人才储备不足、AI应用部署存在困难、投入产出比不达预期等问题。我们认为，企业通过配备适宜于AI开发的高性能软硬件基础服务，有望能够利用底层技术的复用和IT资源的灵活配置优化AI开发和部署流程，提升AI的价值创造能力。

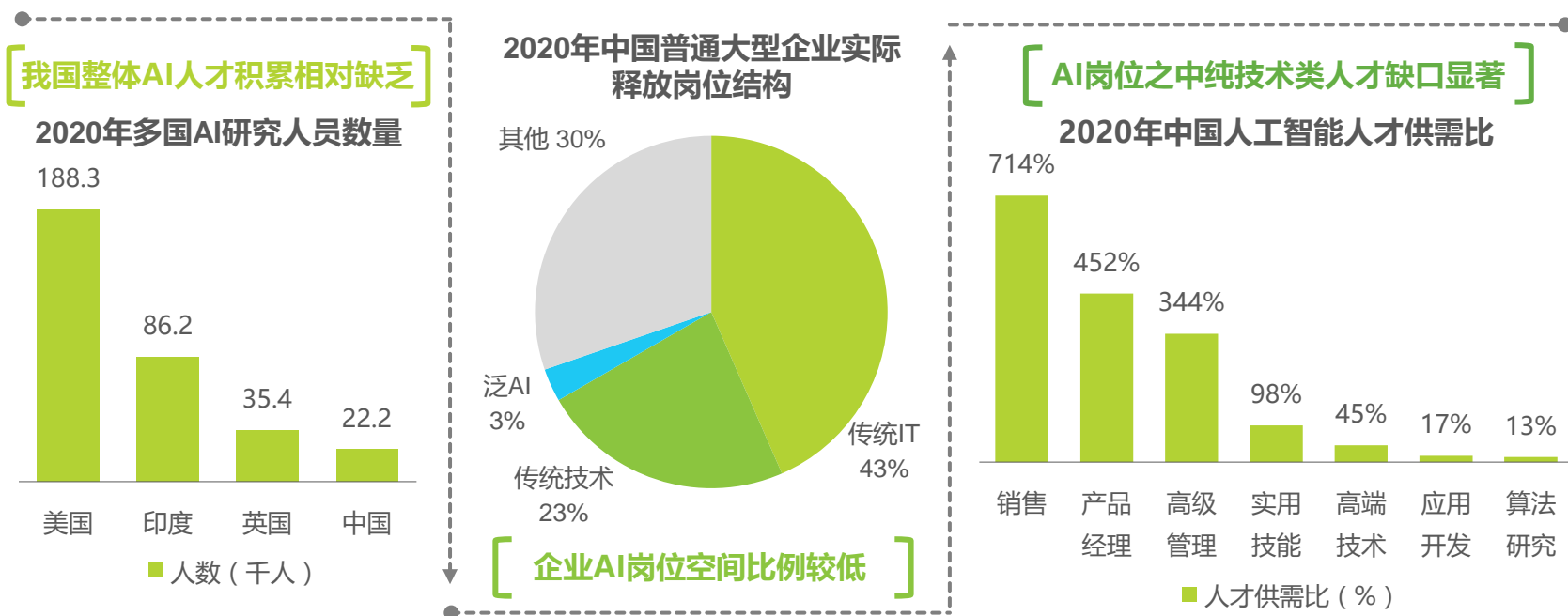


AI人才仍短缺

人才短缺限制企业进行AI开发和落地的步伐

作为前沿IT技术的代表，人工智能产业近年来高速发展，带动了市场对AI人才的集中需求。与许多发达国家相比，我国的AI人才总数仍处于短缺状态，而在企业微观层面上，AI人才市场表现出人才相对集中于互联网科技公司，且技术人才缺口更加显著等问题。人才短缺导致企业的AI需求无法得到及时满足，而人才的培养也非一日之功，长期来看这依赖于IT教育的转型和结构调整，而在短期则更需要产业端通过复用成熟能力、降低应用难度、促进技术交流等方式来提升业内人员的整体素质。

我国人工智能人才处于短缺状态



来源：1.Element AI (2020)，艾瑞咨询研究院根据公开资料研究及绘制；2.艾瑞咨询《2020年中国人工智能产业研究报告》；3.工信部人才交流中心，N=2224，包括人工智能公司和互联网、软件及传统公司，艾瑞咨询研究院根据公开资料研究及绘制。

生产环境对AI基础设施、算法及数据质量要求更高

企业在实际应用人工智能的过程中主要面临基础设施、AI算法及数据等方面的阻碍。在基础设施方面，企业的IT基础设施在海量数据参与运算和采取分布式架构的条件下可能面临算力不足以及不兼容问题，从而降低企业AI应用的可用性；在算法领域，以“大模型”为代表的人工智能算法模型体现出模型参数的数量不断增加的趋势，高度复杂化的模型融入应用程序后可能带来应用延迟的增加，而在交付和部署后对模型的修改也会给用户带来服务中断等问题；此外，由于训练/测试环境和生产环境存在差异，实际应用中的数据存在较多噪音、与模型匹配度较低，也会降低AI应用的质量。总体来看，尤其是中小企业在应用和部署AI应用过程中所遇到的障碍更需要高质量的软硬件平台提供支持，降低开发者在基础配置和运维方面的消耗的精力，并帮助开发者提升模型优化能力。

人工智能应用在生产环境中面临的应用和部署障碍



来源：艾瑞咨询研究院自主研究及绘制。

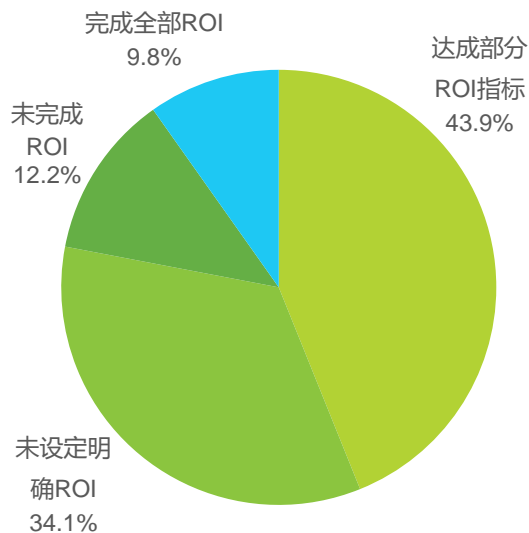
投入-产出比不足

企业应用人工智能的回报尚不及预期

据统计，目前我国企业，尤其是非科技类甲方企业应用人工智能的效果还不尽如人意，许多企业表示人工智能落地后并未达到预期的投入产出比。在成本和支出方面，企业进行AI开发所需的人才和IT资源价格较高并处于相对稀缺状态，引入AI给企业业务更迭和内部管理带来的隐性成本也可能成为降低企业应用AI效果的因素；在价值回报方面，尽管人工智能应用在许多领域和场景已经得到验证，但对于不同的企业和具体工作环境，其效能可能并不稳定。因此，成本和效能两方面原因共同提高了AI应用对于甲方企业的门槛，导致企业不能充分享受人工智能带来的红利。

企业应用人工智能回报不及预期的现状及原因分析

2020年中国甲方企业AI应用ROI



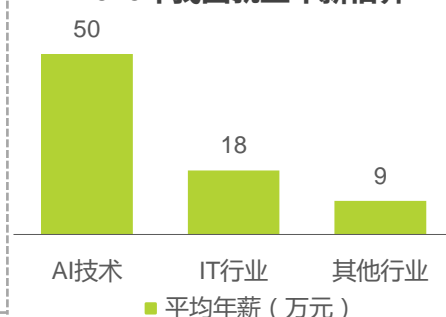
成本和支出

- 企业应用人工智能的投入的资金成本包含人力和IT资源支出，前者主要为AI和数据专家的工作报酬，后者包含自建或购置云服务资源的成本，二者均位于较高水平
- 在实际工作中，模型产出周期长、完成度低等问题给企业带来管理和运维方面的额外投入，构成了企业进行AI投入的隐性成本

效能和价值

- 由于经验的缺乏和模型的不足，企业应用AI于生产环境中面临模型精度不足这一直接问题
- 在生产环境中，负载的动态变化、场景的快速迭代都将给模型的架构和设计带来考验

2020年我国就业年薪估算



2019年部分AI模型训练成本 (美国)

AI模型	硬件	碳排放lb	云计算成本
Transformer _{big}	P100x8	192	\$635
ELMo	P100x3	262	\$953
BERT _{base}	V100x64	1438	\$8161
NAS	P100x8	626155	\$2072348
GPT-2	TPUv3x32	-	27955

来源：1. 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；2. 专家访谈，国家统计局，艾瑞咨询研究院根据公开资料研究及绘制；3. 《Energy and Policy Considerations for Deep Learning in NLP》，Emma Strubell等，2019年6月。

云原生AI开发平台

- 云原生AI开发平台以云原生容器服务为基础架构调用云资源，配合大数据计算、人工智能计算以及分布式计算框架，服务于企业的数据分析、模型训练以及AI应用部署需求。借助容器、微服务、无服务器等云原生优势，企业和开发者得以在IT成本优化的条件下实现AI算法高效训练、应用敏捷开发、程序灵活部署和全生命周期管理。

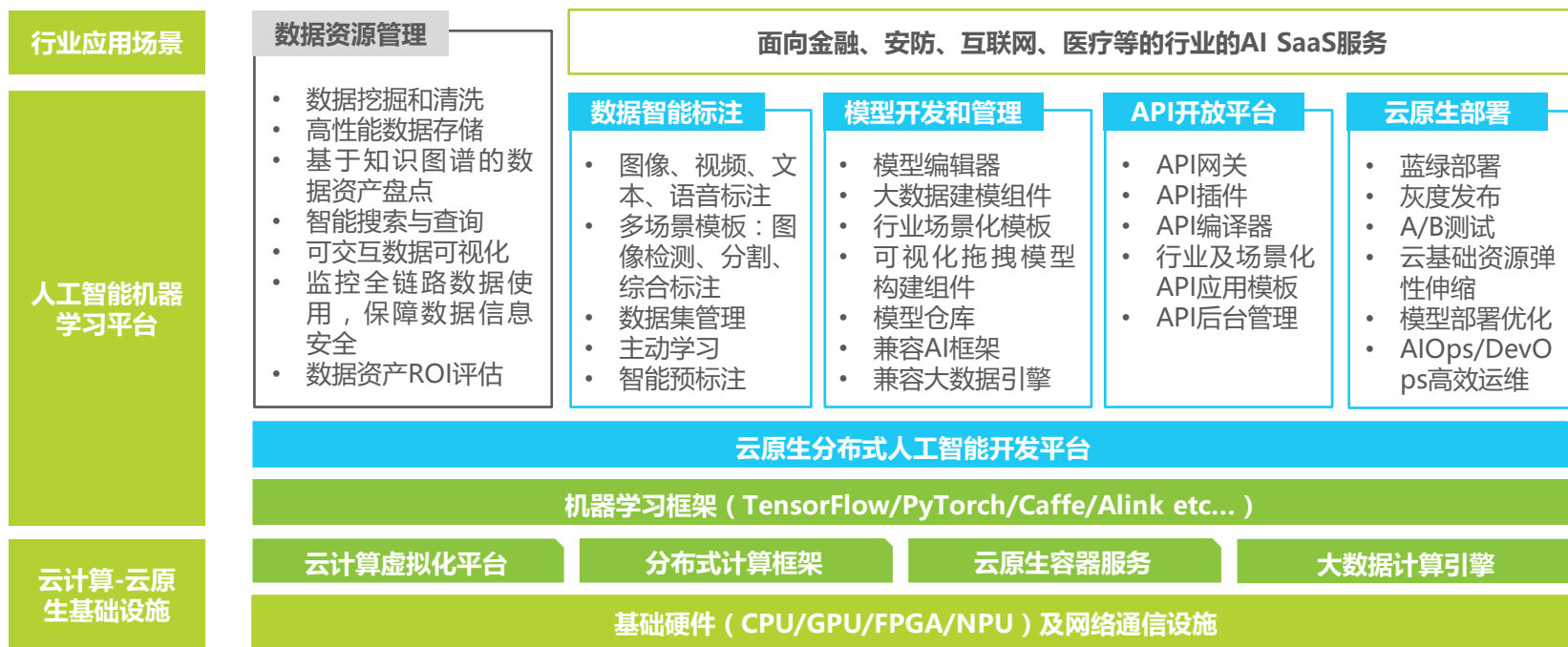


云原生AI开发平台架构

以云原生的敏捷高效赋能人工智能应用开发与部署

云原生AI开发平台以云计算为基础，因为考虑到信息安全和数据隐秘性，该类项目在实践中通常以私有化部署和专有化部署的云服务器为基础，通过容器组件进行IT资源的调用，以微服务架构指导应用设计和开发，并配置分布式、大数据和人工智能计算框架作为底层计算平台。云原生AI开发平台内置数据智能标注、智能模型开发、API开放平台管理以及云原生应用部署等功能模块，辅以包括数据挖掘、网络安全等在内的数据资源管理系统，帮助企业敏捷、高效、安全地利用数据进行人工智能应用开发，并在应用部署过程中实现成本优化和灵活的版本控制。

云原生AI开发平台产品和服务架构



来源：艾瑞咨询研究院自主研究及绘制。

相关应用图谱

云原生AI开发平台相关应用图谱

云原生AI开发平台



智能模型效率化生产



云原生应用部署



AI&大数据计算框架



容器引擎



容器编排



云计算基础设施



注释：本图谱企业排名不分先后。
来源：艾瑞咨询研究院自主研究及绘制。

核心技术优势

资源灵活运用，模型敏捷开发，应用高效部署

我们着眼于企业在云上进行AI开发的各个主要环节，可以发现这一过程也符合著名的“2-8法则”，在AI应用中模型设计和算法精度才是决定其应用价值的核心关键，但企业和开发者实际上将大部分的时间和精力投入了平台搭建、系统调试、团队磨合、监控运维等“非核心”的工作中，云原生与AI开发平台的结合帮助开发者减少对基础IT资源的关注，并通过底层技术复用、开发流程可视化等方式提高开发效率，因而在IT基础设施调用、模型编写和测试优化、模型和应用封装、应用交付及运维等领域更具备成本效率优势。

企业应用云原生AI开发平台的优势与价值



80/20

人工智能应用开发也遵循普遍的“2-8法则”，通常认为最有价值的部分是模型设计及算法精度，但实际工作中大量的成本被用于平台搭建、系统调试等，云原生组件在这一领域帮助企业 and 开发者节省大量的时间和精力

来源：艾瑞咨询研究院自主研究及绘制。

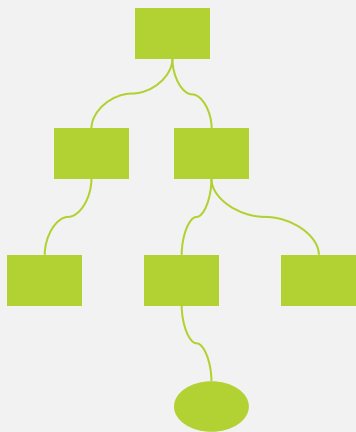
核心技术优势：模型敏捷开发

多种开发模式，赋能人工智能低门槛、高效能开发

云原生AI开发平台配备多种人工智能模型开发模式，其中较为典型的是可视化建模和程式化建模。前者利用JavaScript脚本等组件对算法进行封装，使得用户能够通过拖拽等图型界面进行模型开发，帮助对编程语言不熟悉的甲方企业的开发者乃至业务部门进行定制化开发；另一方面，程式化建模一般基于Tensorflow、PyTorch、Caffe等AI开发框架，利用Python等通用编程语言进行模型开发，由于上述开源框架普遍具备完备的功能，能够赋予开发者更多的开发选项和进行编程优化的空间。云原生组件如容器和微服务框架也能够从底层架构方面对编程、测试等过程进行支持，进一步提升开发者的开发效率。

可视化&程式化建模为开发者提供多种AI开发方式

可视化建模视图



- 利用JavaScript脚本等组件对算法和计算过程进行封装，让使用者能够通过图型界面进行人工智能算法的开发
- **降低了AI开发门槛，提升了模型编写效率**
- 尤其适合甲方企业内部、对AI开发框架熟悉度较低的开发者乃至业务人员

- 通过软件开发者熟悉的Python等通用编程语言进行人工智能模型的开发，融合Tensorflow、PyTorch、Caffe等人工智能开发框架等
- **能够调用人工智能开发框架成熟的底层能力，拥有相对完备的功能和更高的自主性**
- 更适合有编程以及人工智能开发经验的熟练开发者

程式化建模视图

```
Worker:  
  replicas: 3  
  restartPolicy: OnFailure  
  template:  
    metadata:  
      annotations:  
        sidecar.istio.io/inject:  
"false"  
    spec:  
      containers:  
      - name: tensorflow  
        image: gcr.io/your-image  
        command:  
        - python  
        - -m  
        - trainer.task  
        - --batch_size=32  
        - --training_steps=1000  
.....
```

核心技术优势：云原生部署

云原生发布、部署、运维组件提供便捷的AI应用管理渠道

除人工智能算法编写开发之外，云原生AI开发平台利用丰富的云原生组件赋能企业更好地对AI应用进行发布、部署、运维等方面的高效管理。除前文所提及的蓝绿部署、灰度发布和DevOps运维等功能之外，云原生环境中常用的API接口和网关、便捷移植和扩容、边缘侧部署功能均对人工智能的规模应用形成支撑，帮助企业提升用户的使用体验，并降低企业的IT运营成本。

云原生在AI应用部署及运维管理方面的优势



经横向对比，云原生AI开发平台更具成本效用优势

甲方企业获得AI能力有多种渠道和方式，其中，从零开始进行自主研究能够最大程度贴合企业自身需求，但考虑到一般甲方企业的AI人才有限，此方法无论是从IT成本角度还是从最终应用水平角度上看都不合算，如果直接购买产品化的软件，虽免去了开发流程，但未必能够完全满足企业自身的特殊需求。目前，通过AI开发平台进行自主定制化开发是企业常规的选择，在此基础上考虑各类基础资源服务平台，则基于云原生架构的AI开发平台在IT成本优化、开发和部署效率、运维效果等方面具备较好的表现。尤其对于中小企业而言，云原生AI开发能够帮助企业在更大程度上弥补由资金和人才短缺带来的技术水平限制，利于企业平衡AI成本收益，打造符合企业需求的AI产品。

甲方企业AI开发模式效果对比

AI开发模式		IT成本优化	客制化	开发/部署效率	运维效果	AI性能
完全自主研究		☆	★★★★	★	★	★
购买商业化软件		★★★★	★	-	-	★★★★
基于AI开发平台进行研发	开源平台	★	★★★☆	★★	★	★★
	纯公有化商业平台	★★	★★★☆	★★★☆	★★★☆	★★★☆
	纯私有化商业平台	★	★★★☆	★★	★★	★★★☆
	云原生商业平台	★★★☆	★★★☆	★★★★	★★★★	★★★☆

注释：1.本对比不代表具体产品性能及特点，仅代表行业一般水平；2.图中开源平台的部署模式考虑为私有化/非云部署。
来源：艾瑞咨询研究院自主研究及绘制。

行业背景与产品综述

1

云原生 AI 开发场景与实践

2

云原生AI开发平台发展展望

3

互联网娱乐场景的AI需求分析

AI提升用户体验，并帮助企业实现精准营销

互联网娱乐行业的成长植根于各类前沿IT技术，其产品服务和商业模式的演进与IT技术的发展紧密相关。近年来随着人工智能技术的进步，互联网娱乐产品的形式和内容也在不断迎来创新，互联网企业运用各类智能组件为用户带来了更强的沉浸感、交互感和趣味性，同时也利用用户画像和智能营销系统充实客户群体、提高用户粘性。总的来说，人工智能在互联网娱乐领域的应用非常广泛，同时，互娱企业在应用人工智能的同时也面临着设备、用户体验、推荐模型精度等方面的不足，仍有较大的改进空间。

人工智能在互联网娱乐场景的应用需求和难点

互联网娱乐主要场景中的AI应用需求

影视+直播领域	音乐领域	游戏领域	通用
智能特效	智能音效	防沉迷系统	用户画像
美颜/滤镜	真人K歌	反作弊系统	关联交友
低时延交互	风格识别	低时延交互	个性化推荐
高并发访问	平滑切换	高并发访问	语音交互
画质重塑	智能创作	AR/VR	内容分发
自动标签	听歌识曲	动作捕捉	负载均衡
音质优化	外语转译	操作评估	智能运维
视频摘要	音乐鉴权	机器人玩家	智能客服

互联网娱乐场景中的AI应用难点

设备 包括互联网厂商的服务器、网络基础设施、用户使用的终端在内设备本身的性能限制了包括AI在内的应用的性能，对于游戏、超高清视频应用尤其如此。	体验 无论是直播互动、音视频增强还是机器人玩家，在互娱场景，AI应用始终需要关注的是用户体验，目前许多领域的AI应用仿真度、实时性等并不能让用户完全满意，还有很大提升空间。	精度 互联网娱乐行业对流量逻辑有着较强依赖，利用AI进行精准用户画像和智能营销已成为企业重要的获客手段，然而目前这一领域的算法精度还有待提高。
-----------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------

来源：艾瑞咨询研究院自主研究及绘制。

互联网娱乐场景应用案例：南瓜电影

智能推荐助力打造私人定制影院，提升差异化服务体验

南瓜电影是专注于影视精品化运营的垂直类视频产品，通过移动互联网、OTT等客户端提供精品化、差异化的内容运营服务。随着互联网娱乐行业对用户时间和关注度的持续争夺，通过智能算法对用户进行精准画像并提供内容推荐成为行业趋势，但在实践中存在推荐精度不足、算法参数复杂等问题。阿里云智能推荐为南瓜电影APP提供了“南瓜为你推荐”、“根据您看过的《XX》推荐”、“类似影视”板块的推荐服务，结合电影语义分析、内容关联以及用户行为分析，融合多目标优化，包括停留时长、点击率、消费等指标，可根据不同用户的习惯智能生成用户画像、结合用户长短期兴趣喜好进行个性化推荐。

南瓜电影×阿里云：长/短视频智能推荐服务

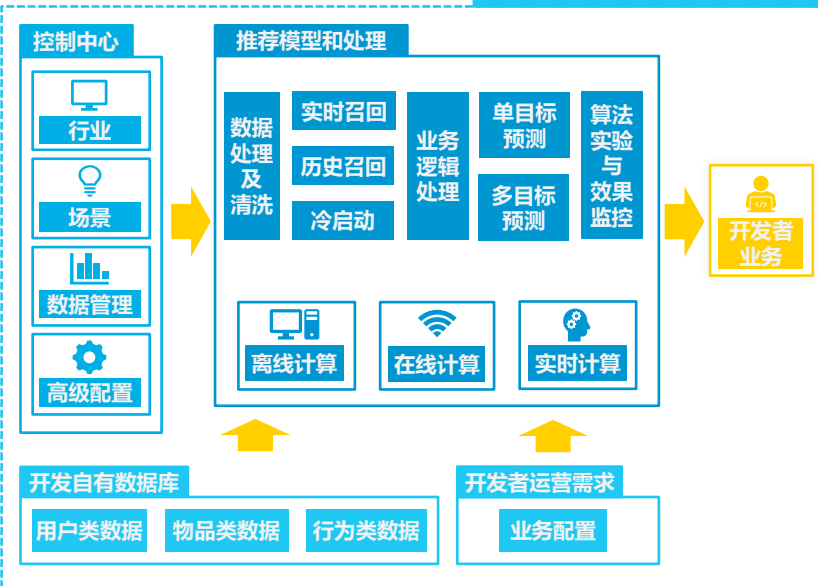
业务难点

- **长视频推荐需兼顾多目标**：电影、连续剧作推荐具备时长、集数等概念，需要兼顾点击率、停留时长、播放数量等目标和参数
- **推荐精准度要求高**：需要结合电影的素材、主题、核心内容片段内容等包含语义理解、深度解读等特征来丰富物料画像
- **曝光过滤设置**：有时用户缺少完整的时间段来完成一次长视频观看，平台需考虑用户的重复触达需求，设置合理的曝光过滤时间
- **长短视频的多样性结合**：短视频板块将电影中的精彩画面集中呈现，为用户带来独特的观影体验，推荐需要根据用户的需求、习惯，融合不同视频类型的特点、目标，多方位激发用户兴趣，增强用户粘性，增加用户停留时长

解决方案&产品价值

- **丰富的电影标签**：结合电影题材介绍、影评等内容，通过语义分析和自动打标来丰富电影标签，并通过专业的运营、剪辑团队，将这些标签作为物料的重要特征输入到推荐系统中，提高推荐的精准度
- **时间因素分析**：经试验分析，计算出将曝光过滤时间设为3天较合理。对于用户群体而言，可将内容适当重复触达，达到提升推荐效果的作用
- **打造短视频专区**：南瓜电影为用户打造了电影片段（短视频）的推荐专区，既能为用户的碎片时间提供别致的观影体验，又能将电影精彩片段快捷呈现，引导一次完整的观影行为
- **用户习惯全方位打通**：通过长短视频的智能算法接入，将用户的兴趣全方位打通，根据用户习惯，喜欢观看的电影题材，实现进一步的精准推荐。同时，兼顾推荐电影的多样性，让用户拥有新发现，促进观看时长的增加

阿里云智能推荐产品架构



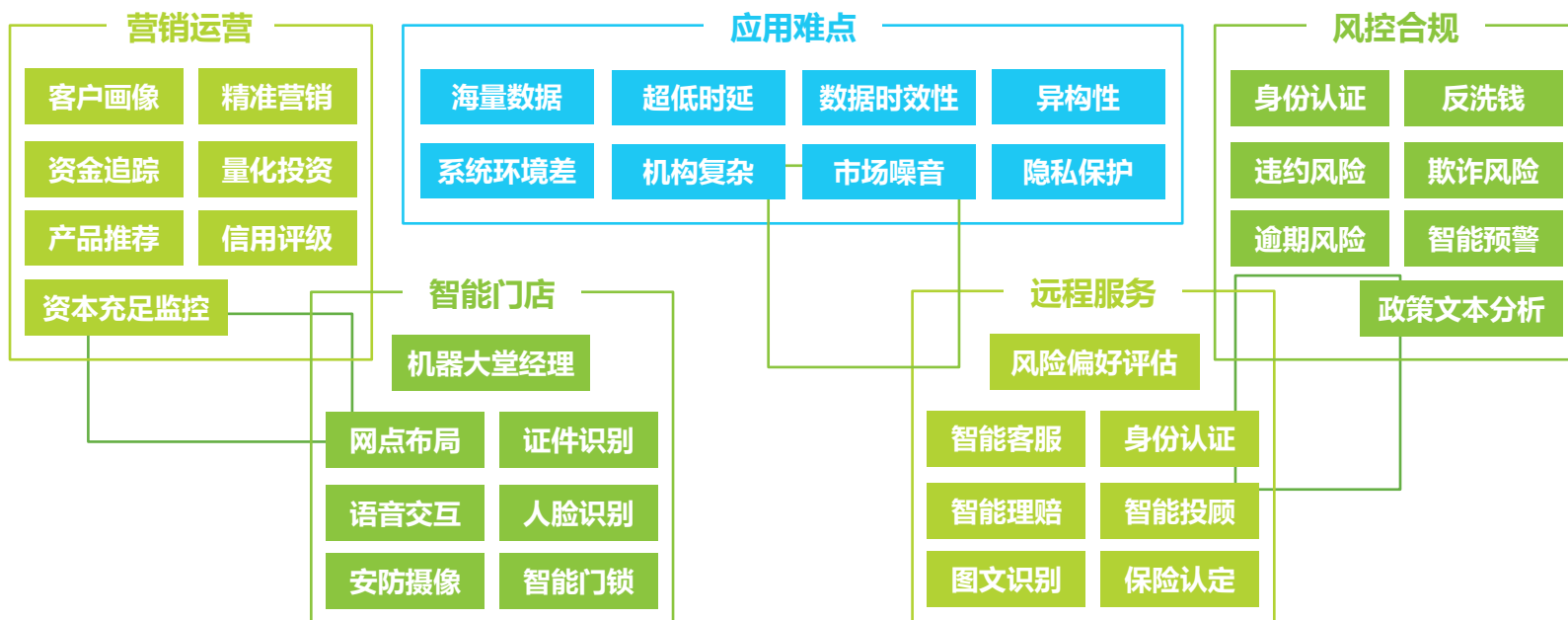
来源：艾瑞咨询研究院自主研究及绘制。

金融场景的AI需求分析

AI贯穿金融机构运营多个流程，提升机构经营质量

金融行业的运行无处不与数字打交道，这一特性决定了金融业务与大数据和人工智能应用有着天然的契合性，加之金融机构普遍资金充足、信息化水平相对较高，因此对人工智能的投入度较高，与IT企业的合作开发也较多。具体来看，人工智能在金融机构的营销决策、智能门店建设、在线服务、风控合规等领域都有较多的应用，与低代码、RPA等效率工具配合，能够替代金融机构员工完成大量重复性高且复杂的工作，提升机构运营的效率，并避免人工操作带来的疏漏。另一方面，金融机构客户群体较大，需要实时处理海量数据，且金融服务对交易数据实时性和同步要求超过其他任何行业，更需要人工智能在满足客户和机构需求的同时不能够对系统性能造成显著影响。

金融业务场景放中的AI开发需求与难点



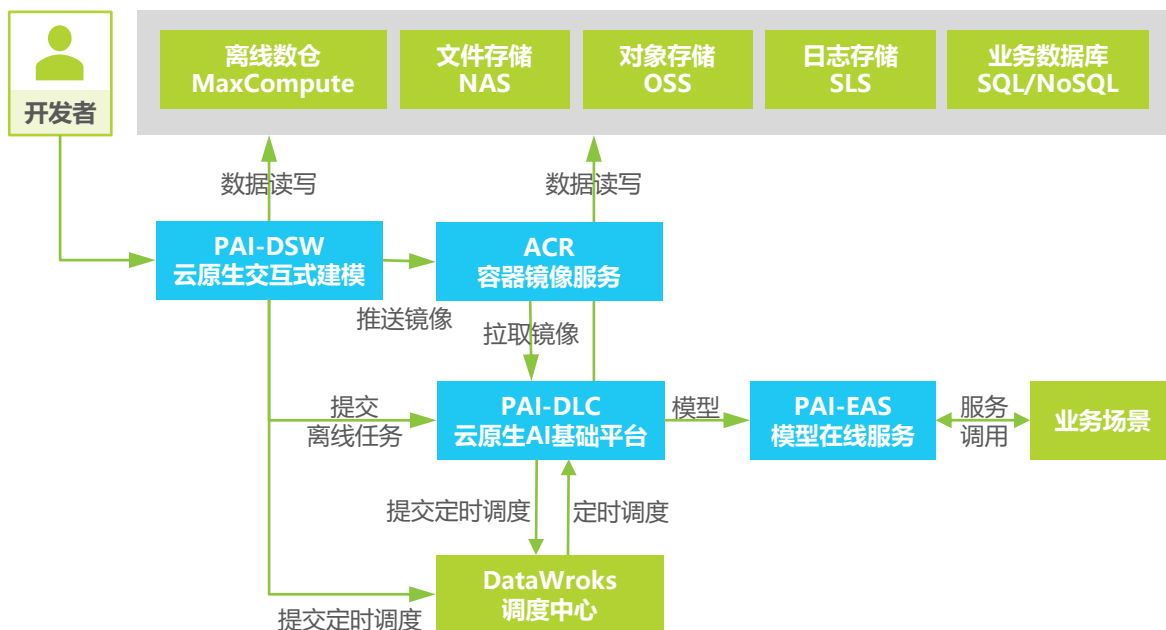
来源：艾瑞咨询研究院自主研究及绘制。

金融场景解决方案：金融量化科学

集成科学计算引擎Mars，高效利用计算资源赋能业务

量化金融是大数据时代金融行业的发展方向之一，金融量化科学计算大量依赖NumPy、pandas等库来处理数据，这些工具运用符合直觉的处理方式来分析数据和提炼有用信息，然而，金融量化数据具备数据计算量大、性能要求高、开发资源不够集中等特点。阿里云自研的科学计算引擎Mars与NumPy和pandas API兼容，能够利用分布式和GPU硬件提升计算框架的性能，并与PAI-DSW深度集成，使得用户能够在PAI-DSW中直接使用Mars。用户通过阿里云PAI进行金融量化计算，能够充分实现计算环境容器化、计算资源弹性化、离线任务工程化，帮助解决环境管理困难、资源分配复杂、策略回测难度大等量化计算的关键问题。

阿里云：金融量化科学计算方案



场景及方案介绍：

金融量化数据在大规模计算上常常遇到性能瓶颈。PAI平台金融量化科学技术实践为客户带来基于PAI-DSW、Mars和PAI-EFLOPS进行大规模科学计算的加速方案

解决问题：

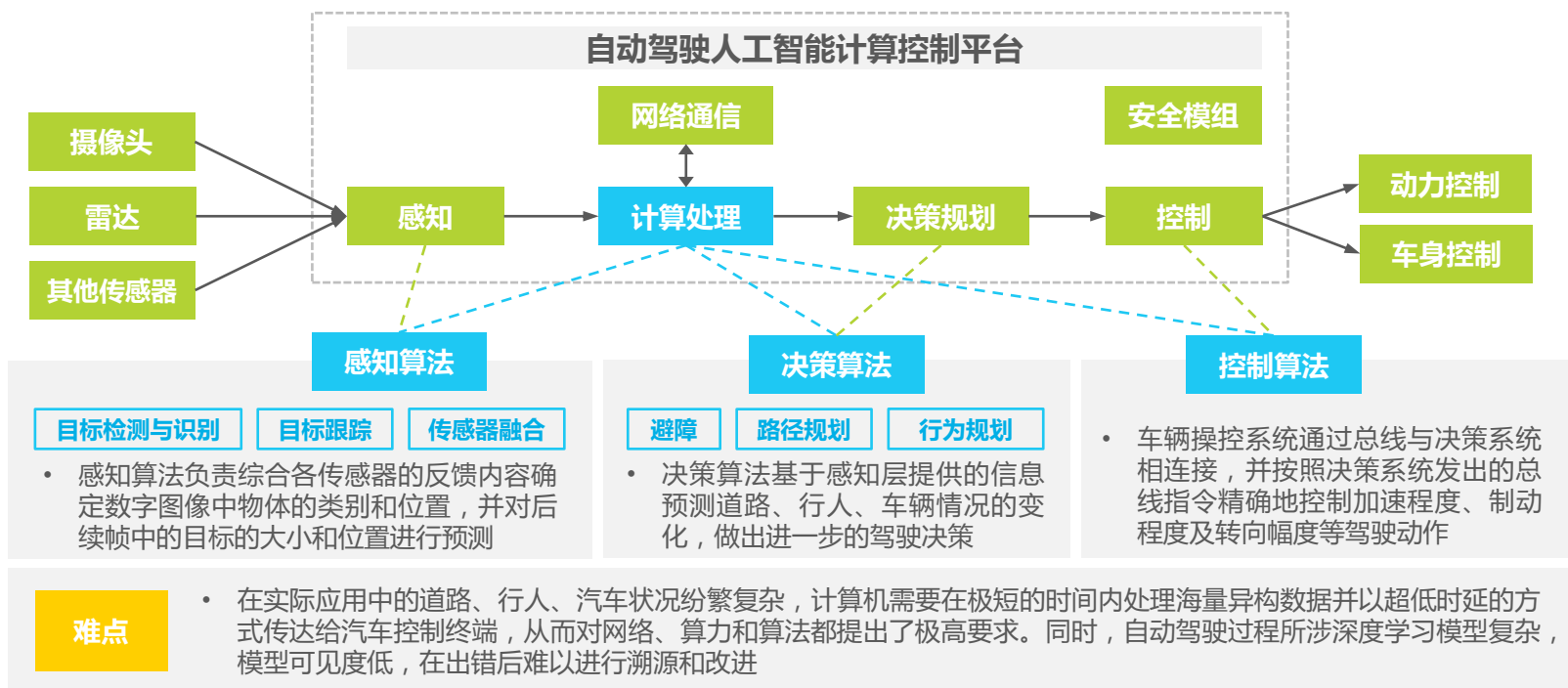
- 环境管理困难
本地化自建策略研发平台容易导致环境不一致
- 资源分配复杂
任务资源分配不均导致资源抢占
- 策略回测麻烦
数据量大、算法版本不一致导致策略测试困难

自动驾驶场景的AI需求分析

AI为实现自动驾驶的核心技术，准度和性能要求极高

在智能化、网联化、电动化、共享化的背景下，自动驾驶成为智能网联汽车行业发展的重点，也成为了下一代汽车企业转型升级的技术高地。自动驾驶对算力、实时性、安全性等系统性能的要求极高，且自动驾驶的工作环境复杂多变又涉及公共交通安全，要求计算机在极短时间做出正确的判断，因此需要以强大的算力、算法和数据传输能力作为支撑。在自动驾驶实现过程中，人工智能算法承接传感器接收的实时信息并进行数字化处理和分析，经过决策算法对周围环境进行的判断和预测，并对驾驶操作进行调整。

自动驾驶场景中的人工智能应用需求及难点



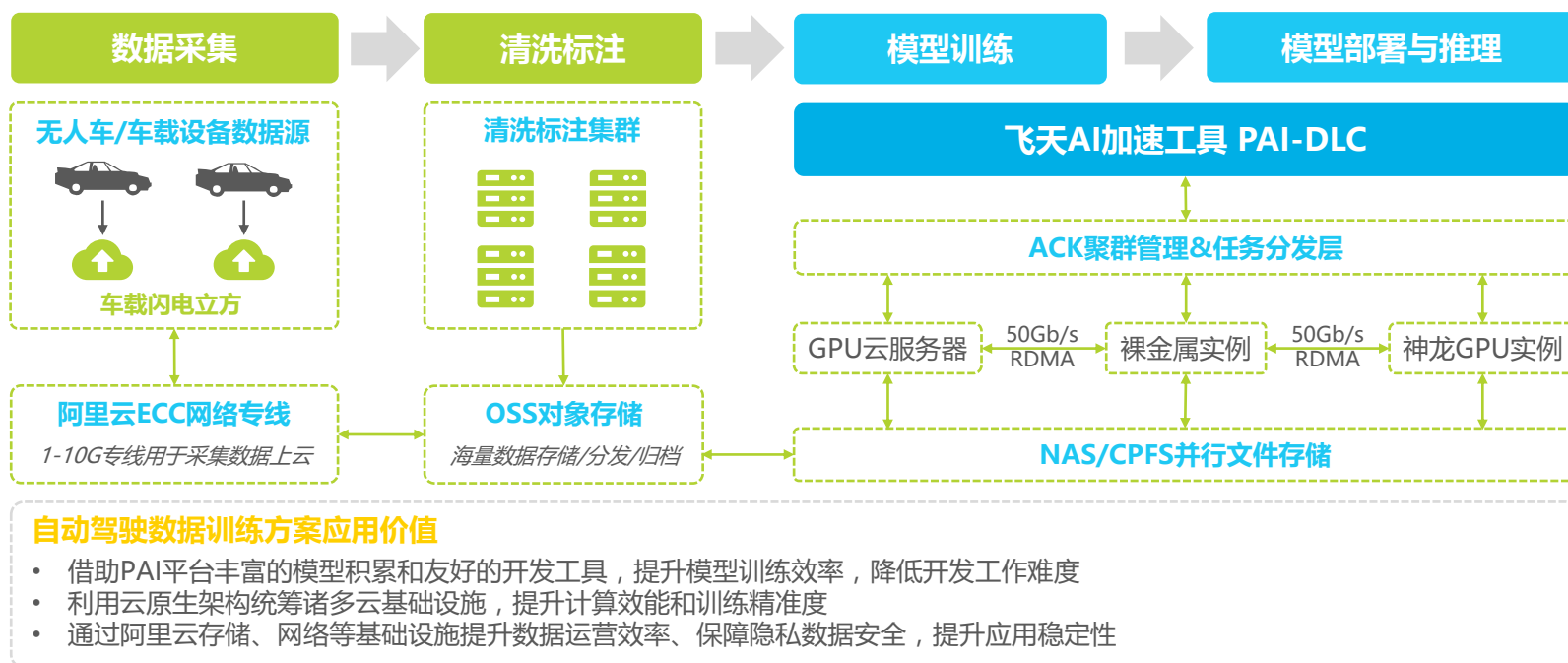
来源：《2018-2019年中国自动驾驶产业发展蓝皮书》，艾瑞咨询研究院根据公开资料研究及绘制。

自动驾驶场景解决方案：数据训练

融合AI开发、存储及专网服务，提供高性能数据训练服务

为了应对复杂的自动驾驶实际环境，一方面需要后台的自动驾驶算法尽可能做到精确，另一方面也需要人工智能应用通过高效的部署模式来提高运行效果，才能够最大程度保障行车安全。阿里云自动驾驶解决方案以飞天AI加速工具PAI-DLC为核心，一方面借助网络专线、对象存储等功能收集行车数据进行清洗标注和模型训练，另一方面利用云原生工具进行应用部署和任务分发，更好地调用GPU、裸金属服务器、神龙等底层算力设施的能力。整体来看，该解决方案面向自动驾驶场景中的诸多难点，全方位融合了AI开发、存储及专网服务，打造一体化的数据服务能力。

阿里云：自动驾驶数据云上训练解决方案



来源：艾瑞咨询研究院自主研究及绘制。

制造业场景AI需求分析

通过高效自动化和智能决策，帮助企业实现降本增效

以人工智能应用赋能制造业发展是我国进行制造业产业升级、建设工业互联网平台的关键。当前人工智能主要通过三种模式融入工业互联网：第一，将AI模型嵌入工业互联网平台层，提升平台层数据分析能力；第二，提供工业AI软件系统，并通过云端部署形成工业互联网SaaS层应用；第三，提供工业互联网框架下含软件和边缘侧硬件的完整系统。具体到制造企业的生产流程中，上述应用广泛融入企业从产品设计、供应链管理到生产制造和销售运营的各个环节，能够通过提升自动化和决策智能化水平为企业带来降本增效价值。

制造业场景中的人工智能应用需求及难点



效能

人工智能帮助制造业企业提升自动化程度和数字化程度，更好地利用数据实现企业经营的降本增效

应用难点

- **IT基础差**：企业尤其是工厂往往地处较偏远位置，网络、算力等基础能力较差；部分经过早期数字化升级的工厂采用的数据标准、通信协议不统一，造成数据孤岛问题，需进行改造
- **AI成本较高**：制造业人工智能应用需要较多的传感器、机器人硬件设备作为基础，后期的算法开发、运维成本也相对较高
- **工控要求高**：对于部分高精制造类应用，要求IT系统提供超低时延、超高精度的识别和控制性能

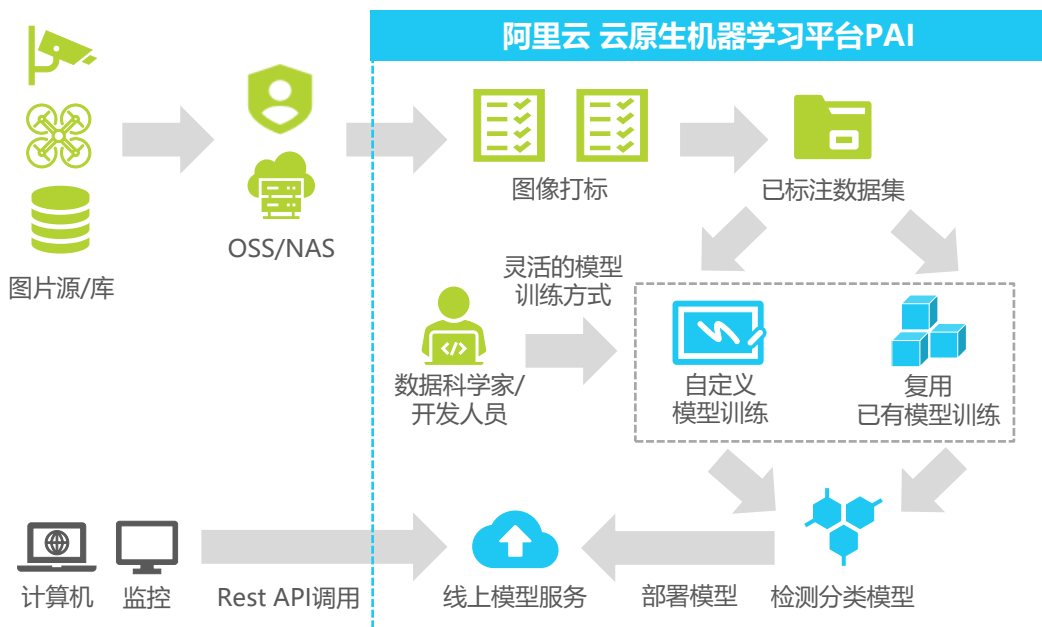
来源：艾瑞咨询研究院自主研究及绘制。

制造场景解决方案：线路安全检测

智能监控高压线路状况，降低生产安全隐患

在制造场景的诸多人工智能应用中，电源线路检查属于最基本的需求之一，不仅关系到最核心的人身和财产安全性问题，在生产场景中的部署密度也较高。一般而言，生产场景中的电源线路可能面临的安全隐患包括线路自然老化、交通工具擦挂、异物附着、烟火点燃、其他施工机械接触等，上述隐患通过人工检查不仅工作繁杂，在实践中更难以面面俱到。阿里云高压线路安全检测方案依托云原生机器学习平台PAI，以计算机视觉为基础，配合丰富的图像数据库和精准模型算法，再辅以分布式部署的线上模型推断服务，帮助制造企业以更高效、高精度和低成本的方式进行线路检查，保护企业的经营安全。

阿里云：高压线路安全生产检测方案



人工智能线路安全检查难点

- 线路较长，途径生产环境多变
- 高压线路可能面临自然老化、交通工具擦挂、异物附着、烟火点燃、其他施工机械等安全隐患
- 高压线路如果发生意外，对人身和生产安全的危害极大，安全检测能力须达到极高要求

解决方案综述

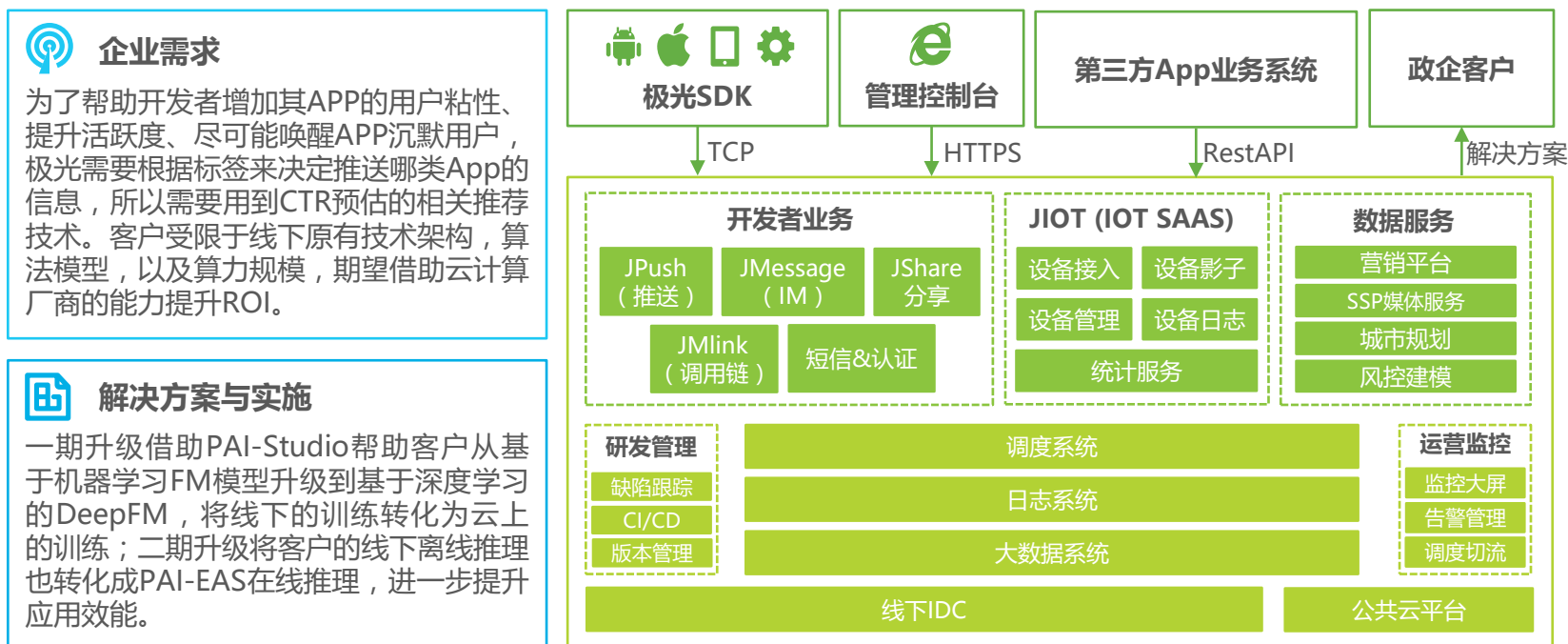
- 该解决方案结合物联网设备采集和系统图库进行图像打标，依托阿里云PAI丰富的模型训练框架进行图像识别和决策判断，并利用云原生架构进行灵活的边缘侧应用部署，提高AI应用的准确性和速度

其他场景应用案例：极光

算法与IT基础能力共同提升，显著优化关键业务ROI

极光 (Aurora Mobile) 成立于2012年，专注于为开发者提供稳定高效的消息推送、一键认证以及流量变现等服务，助力开发者的运营、增长与变现，在市场洞察、金融风控与商业地理服务等场景中都有应用。企业希望优化基础资源能力和算法能力以提升消息推送效率，最终提升用户粘性和活跃度。通过与阿里云PAI的合作，企业将FM模型转化成DeepFM模型并将线下的Angel模型转化成PAI-TF，实现了算法模型指标的提升；同时，企业利用线上的模型训练环境大幅增加了模型训练的迭代次数，摆脱了此前线下IT基础设施的限制，最终实现了AUC、上线后点击率等相关业务指标的显著提高。

极光×阿里云：通过PAI平台实现推送效果提升



来源：艾瑞咨询研究院自主研究及绘制。

行业背景与产品综述

1

云原生 AI 开发场景与实践

2

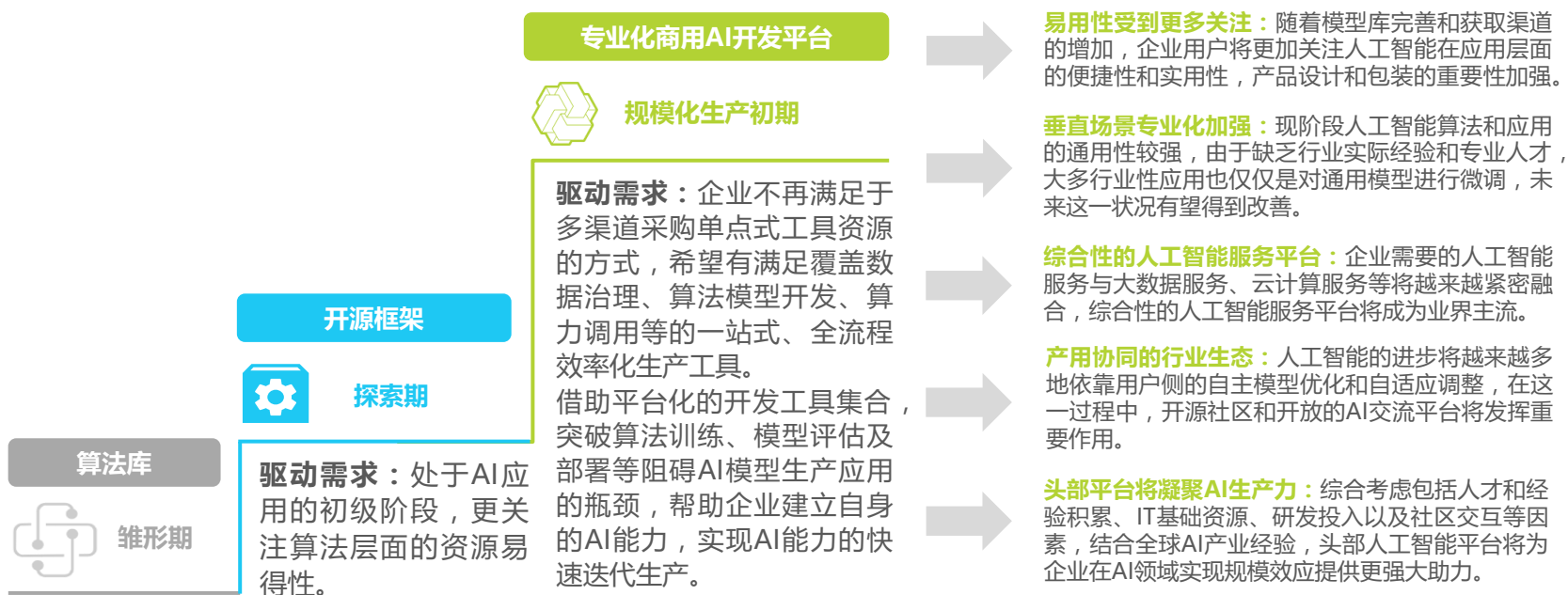
云原生AI开发平台发展展望

3

产品易用性、综合性不断提升，帮助企业形成AI规模经济

企业进行人工智能应用开发的核心资源大致经历了三个发展阶段：在雏形期，应用开发主要依靠开源模型或开发者自发分享的代码进行；随着AI技术发展和需求激增，开源框架陆续出现，助力开发人员快速生成及训练模型，开源框架也成为了众多AI开发平台的基础；后续AI产业化进程加速，为满足智能化转型需求方对多层次资源的多样性需求，专业化的商用的AI开发平台出现，致力于为甲方提供一站式的AI模型生产服务。为了进一步满足企业运用人工智能过程中全方位的需求，AI开发平台还将朝着易用性、专业化、综合性、产用协同等方向发展，在这一过程中，在底层技术、IT资源、人才和经验等方面具备优势的头部平台产品将能够更好地协助企业形成人工智能应用的规模经济。

人工智能开发的核心资源演化以及未来发展趋势



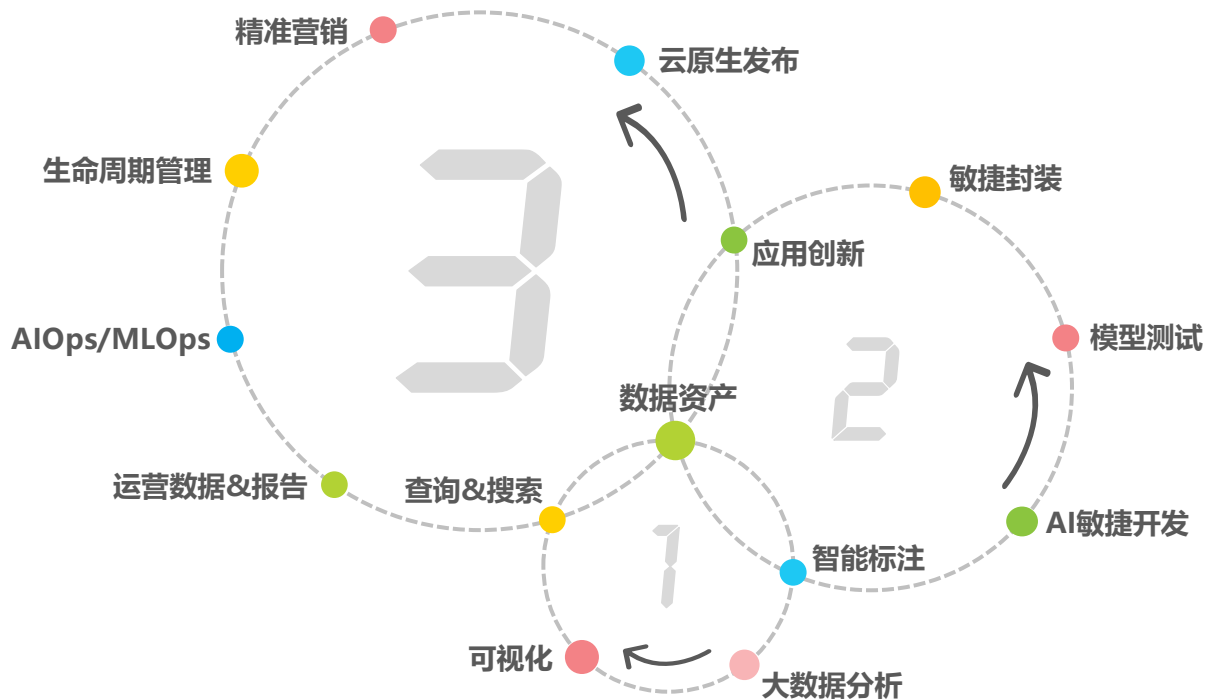
来源：艾瑞咨询研究院自主研究及绘制。

产品广度 (1/2)

以数据资产为核心，数据智能/模型开发/数据应用紧密协同

现今大数据产业在独立的领域如数据库、智能搜索、数据可视化及人工智能领域均实现了一定突破，应用价值得到了普遍的认可。未来大数据相关产品将以数据资产为中心，实现进一步整合和协同，最终形成商业数据智能、AI模型开发、数据智能应用三大飞轮，分别为企业的数字化运营提供精准决策、生产工具、经营管理方面的支持。同时，三个飞轮所代表的业务体系和流程将彼此融合、紧密关联，对应着一类链路更加复杂的工具，IT厂商或将以现有的云原生AI开发平台为基础进行服务扩充和集成，而云原生组件则有望在这一高度集成的服务链中发挥服务编排作用。

以数据为核心，融合AI与大数据分析的企业数据应用飞轮



3 数据智能应用飞轮

—— 以数据为核心的企业运营，形成产品创造数据、数据打磨产品循环

2 AI模型开发飞轮

—— 数据价值挖掘，提升自动化水平，实现业务创新

1 商业数据智能飞轮

—— 经营数据分析、赋能企业精准决策

来源：艾瑞咨询研究院自主研究及绘制。

DevOps和AIOps有望进一步提升AI开发运维效率

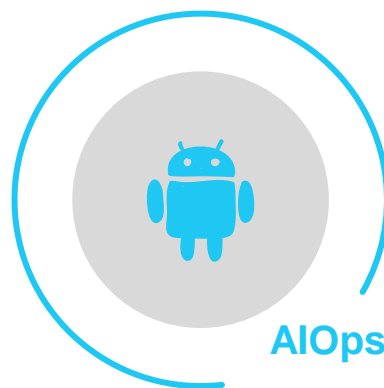
头部云计算和人工智能厂商在不断对包括人工智能在内的数字化产品进行丰富和升级，DevOps和AIOps/MLOps均有望与当前的云原生开发平台进行深度集成，进一步提升企业进行人工智能开发和应用效率。其中，DevOps是云原生服务于软件开发的应用流程，主要用于软件的协作编程开发，目前和AI算法模型的编写流程并不完全相通，但仍能用于模型封装成软件应用之后的开发流程。另一方面，AIOps/MLOps原本主要用于软件上线部署后的运维，不仅能够帮助甲方企业进行复杂AI应用的监测，从而提升客户体验，也能利用自动化流程来进行模型效果的测试。目前AIOps不仅是国内IT运维产业的发展新方向，也有望成为云原生AI开发平台优先的集成方向。

DevOps与AIOps与云原生AI开发平台的协同



DevOps

- 云原生的主要概念之一
- 适用于软件编程开发
- 通过流水线组件提供便捷的协同开发环境
- 通过容器环境，让开发者介入运维流程，打破开发-运维部门的壁垒，提升运维流程



AIOps/ML Ops

- 新一代IT运维工具
- 以AI算法代替人工进行系统故障检查、根因分析、事前预警
- 需要大量系统数据沉淀以及AI算法的加持

AI开发平台与IT基础设施进一步融合，提升计算能力

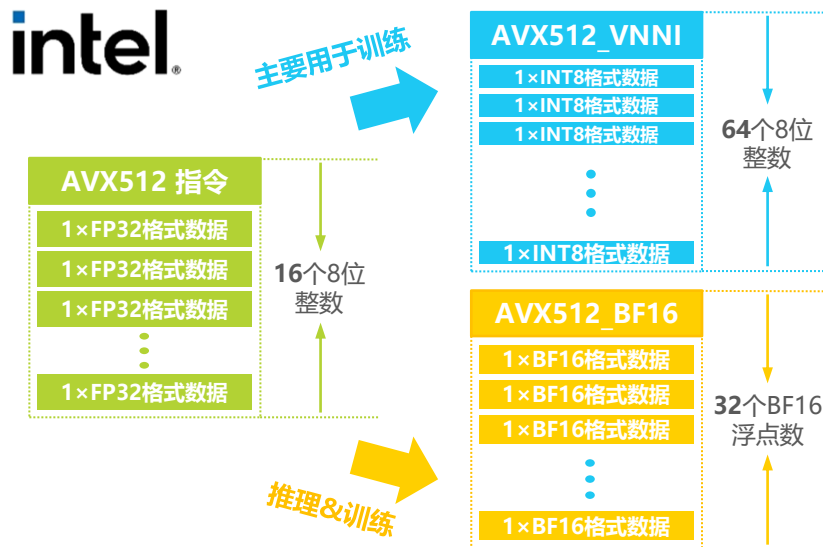
随着数字经济大发展，数据量爆发，应用类型丰富，原有的IT基础设施无法满足AI应用需求。近年来，适宜于大规模数据集群计算的CPU、GPU、DPU等计算芯片不断推陈出新，内存、虚拟化、网络等IT基础层技术也都在向着大规模并发、超高速传输的方向演进，并配合主流的人工智能框架进行优化升级。未来人工智能开发与上述IT基础设施深度融合将成为大趋势，这一趋势或将从AI技术和硬件厂商的合作起步并逐渐深化，例如，Intel与阿里云达成合作，阿里云AI开发平台PAI利用Intel的至强处理器以及内置的深度学习加速（DL Boost）功能进行AI应用优化，帮助用户实现AI应用的降本增效。

面向人工智能的IT基础设施不断进步

人工智能应用对整个IT基础环境提出了较高的要求，以CPU、GPU、DPU等计算芯片为代表，内存、虚拟化、网络传输技术的发展都在对人工智能发展形成支撑。

Intel至强处理器内置DL Boost功能

大多数深度学习应用采用 32 位浮点精度数（FP32），Intel的DL Boost采用8位整数（INT8）或16位浮点数（BF16），通过降低内存的负载更高效地利用计算资源、提升AI应用的运行表现。



来源：艾瑞咨询研究院自主研究及绘制。

来源：Intel，艾瑞咨询研究院根据公开资料研究及绘制。

AI用户与开发者角色交融，加速产品创新与市场拓展

以开源社区为代表，全球范围内有着众多软件技术交流社区与平台，在这些组织之中，软件的开发与使用者对产品的架构、功能、缺陷进行讨论，为开发者创造了解用户真实需求和实际使用难点的环境，帮助开发者进行产品打磨，也为用户提供了深入产品底层、基于自身需求对产品进行改良的平台。随着软件市场活力不断提升和开发者人群技术的增长，国内市场上出现了越来越多的技术交流社区。天池大数据众智平台作为阿里云旗下技术学习、交流、竞赛社区，为大数据和AI开发者提供了丰富的学习资源、实践工具和交流环境，帮助开发者加速获取技能和经验。

开源社区对AI开发&应用的价值



来源：艾瑞咨询研究院自主研究及绘制。

阿里云天池平台简介



合作产品



PAI



视觉智能
开放平台



九章算法

合作机构

政府机构

中国气象局、中国地震局、多地政府

企业

Intel、OPPO、美年大健康、新浪微博等

科研&医院

中国计算机学会、中国科学院国家天文台等

来源：阿里云天池平台，艾瑞咨询研究院根据公开资料研究及绘制。

艾瑞新经济产业研究解决方案



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。

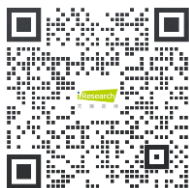
如今，艾瑞咨询一直致力于通过科技与数据手段，并结合外部数据、客户反馈数据、内部运营数据等全域数据的收集与分析，提升客户的商业决策效率。并通过系统的数字产业、产业数据化研究及全面的供应商选择，帮助客户制定数字化战略以及落地数字化解决方案，提升客户运营效率。

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

iResearch

艾 瑞 咨 询