



边缘智能

赋能行业，智领未来

目录

IDC观点	03
第一章 AI拓展数字经济新疆域	04
1.1 数字经济为经济发展注入新动能	04
1.2 AI成为数字化转型核心驱动力	06
1.3 AI行业应用持续拓展	07
第二章 边缘智能拓展AI新边界	10
2.1 边缘计算融合AI走向边缘智能	10
2.2 边缘智能推动AI全面升级	14
第三章 边缘智能赋能千行百业	18
3.1 制造业	18
3.2 政府 (智慧城市)	22
3.3 电信行业	25
3.4 医疗健康与生命科学行业	29
3.5 零售行业	29
第四章 边缘智能走向未来智能	32
4.1 未来智能展望	32
4.2 IDC建议	34

IDC观点

AI在深化行业应用前景广阔但面临挑战

随着新一代信息技术与行业深度融合,数字化转型已经日趋成为企业发展的新动能。AI通过重构企业商业流程、用户体验、产品服务和商业模式,正成为企业数字化转型的核心。根据IDC预测,2024年,全球企业AI硬件、软件及服务方面的总支出将达1,106.6亿美元。但随着AI向行业渗透,传统云为中心的AI解决方案难以实现对海量的数据存储和传输要求的处理,也难以满足诸多对实时响应要求较高的应用场景的需求。

边缘智能助力AI释放应用潜能

边缘智能结合边缘计算和人工智能,在边缘侧进行实时、小数据的处理,开展AI模型的推理,实现智能检测,回传云端结果,在云端进行长周期、大数据的处理,开展模型的训练和数据分析,形成云-边-端协同的架构,有效的解决AI在行业应用面临的海量数据处理、实时响应、数据安全等问题,为AI在更多行业应用奠定基础。

边缘智能赋能各行各业

边缘智能在制造业、政府公共事业、通信、医疗等各类行业正在也将形成诸多应用场景,如在制造业,通过边缘智能开展智能质检,以实现降本增效。在智慧城市建设领域,通过边缘智能实现实现路况监测、路口交通控制、停车管理和违章抓拍等,提升城市交通效率。在通信行业,通过建设MEC边缘云,为运营商拓展了云VR、云游戏等创新业务空间。在医疗行业,通过边缘智能可应用在医学影像辅助诊断、基因检测、临床试验、医院内自助服务等场景中,提升医疗质量和效果。

01

第一章

AI拓展数字经济新疆域

1.1 数字经济为经济发展注入新动能



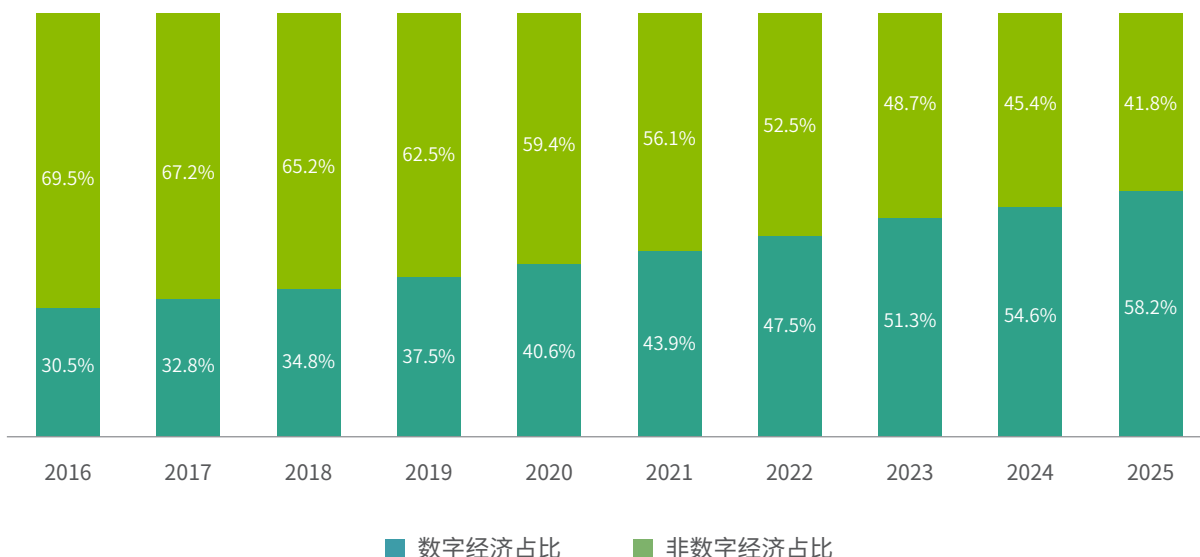
数字经济成为增长新动能

IDC认为,数字经济正在成为全球下一代经济发展的核心构件,全球经济加速履行“数字使命”。一方面新技术不断涌现,为发展带来诸多可能。伴随着云计算、大数据、物联网(IoT)、人工智能(AI)等为代表的新一代信息技术的创新和发展,越来越多的组织通过数字化技术实现增长。以数字化信息和技术为基础的数字经济逐渐成为新一轮增长点,正在助力各行各业重塑核心竞争力。

另一方面,数字经济作为新的发展动能,良好满足提升效率、改善增长质量的发展需求。一改传统“增加负债,拉动投资”的增长模式,数字经济深刻重塑了企业、用户以及社会之间的互动关系,提升供应链的协同效率、打破产品交付的地域限制、加速商业创新、为用户提供更加个性化的体验,应对经济体量增长带来的压力的同时盘活市场,产生持续的创新力。

IDC预计,2020年,全球由数字化产品和服务驱动的数字经济的占比将达到40.6%,而这个比例在2023年将超过50%,数字化产品和服务将成为主流,全球经济将全面进入数字经济时代。从2020年到2023年,全球数字经济将推动6.8万亿美元的IT支出。

图1 全球数字经济占比,2016-2025



来源: IDC, 2020



后疫情时代, 数字经济发展提速

2020年爆发的新冠肺炎疫情在客观上加速了数字经济的市场培育。疫情期间的防疫控制措施限制了面对面的交流, 大量数字化产品的设计、开发、测试、交付相关工作从线下转到线上, 协同效率逐步增加、客户体验逐步改善、市场认可度稳步提升。得益于数字经济在疫情期间显示出巨大优势, 预计疫情后, 非接触连接商业与服务将持续加速发展, 线上和线下持续深度融合, 助力中国及全球企业迈向下一个常态, 数字化技术如云、AI、5G等将成为推动这一趋势的引擎。

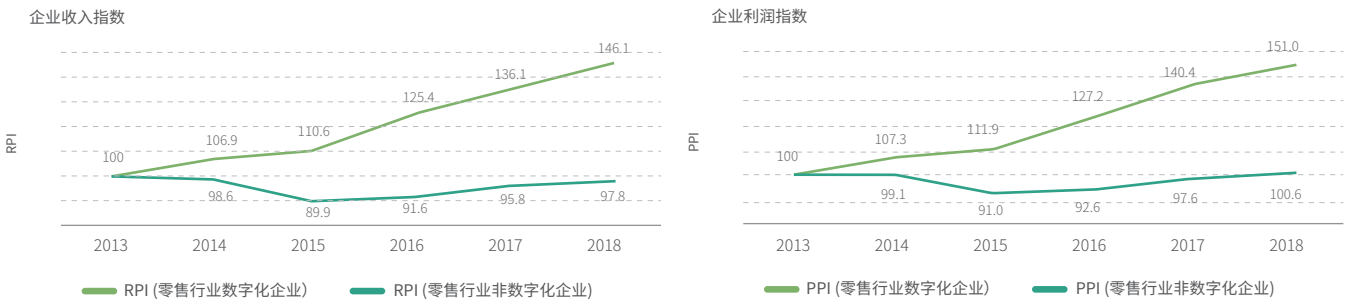
在中国, 政府出台新基建政策以进一步推进经济复苏。新型基础设施主要包括三个方面的内容: 信息基础设施、融合基础设施和创新基础设施, 其中多个领域与数字技术和数字经济密切相关, 包括5G基站与网络、AI、大数据中心和工业互联网等。**IDC 预计, 2020年新基建整体投资规模达2,750亿美元。**



企业开展数字化转型势在必行

时代的发展离不开行业以及企业的切实推动。在数字经济的浪潮中, 数字化转型已经成为企业保持其生命力, 实现长久发展的必然选择。数字化和非数字化企业的差距正在拉大。以零售行业为例, 从2013年到2018年, 实施数字化战略的零售企业的累计收入增长了约7,210亿美元, 复合年增长率为4.2%; 非数字零售企业下降了约370亿美元, 复合年增长率则为-1.5%。同期, 数字零售企业的累计利润增长了约2,270亿美元, 复合年增长率为4.9%。非数字零售企业的累计毛利润则下降约85亿美元, 复合年增长率为-1.3%。同样的鸿沟存在于制造业、银行等其他领域。**IDC预计, 截止2020年, 至少有55%的组织是数字化坚定的拥护者, 他们通过新的商业模式和数字化产品与服务来改变市场, 重塑未来。**

图2 零售行业数字化企业与非数字化企业收入及利润, 2013-2018



来源: IDC, 2020

1.2 AI成为数字化转型核心驱动力

前沿技术的发展和 innovation 是推进企业实现数字化转型的重要前提，新兴数字技术的兴起使颠覆和重塑市场格局成为可能，企业若想在新的竞争格局中脱颖而出，需能够逐步借助技术的力量发挥企业潜能。**IDC预测，到2024年，全球将有75%的企业将智能自动化嵌入技术和流程开发，并借助AI软件生成运营和体验洞察，引领创新。**

IDC认为，数字化转型将帮助企业实现4个重构：



重构商业流程



重构用户体验



重构产品和服务



重构商业模式

AI是帮助企业更高效释放数据能量的催化剂，可将数据转化为知识和行动，进而将行动转化为价值，从而为产品、客户和市场创建高价值的差异，并提供有意义的、增值的学习、预测和行动，从而改进商业流程、提升用户体验、促进产品和服务优化，进而完善企业决策机制，重构商业模式，创造企业新价值。



商业流程：AI驱动的企业架构和解决方案创新将促使AI在整个企业业务流程中被全面应用。通过对不同来源的类型丰富、结构多样、标准各异的数据进行有效实时的解析、汇集、分析，缩短业务响应时间；AI的应用实现业务流程执行自动化，满足合规有效的前提下提高业务流程准确性和预测准确率；AI环境下打通部门之间信息孤岛，以企业统一战略为目标，协作提升流程管理；利用自我学习、自动推理，具备解决问题的能力，进而提供更匹配业务的流程优化方案。



用户体验：AI技术使得企业能够通过大规模、多维度的数据分析准确了解客户的需求，并且实现更高频次的数据分析，实时感知客户需求的变化。对AI的日益依赖将使应用/系统开发人员重新思考用户触达方式，并突出AI在用户界面交互方面的增强作用，为客户提供更加个性化、场景化的体验，提升用户满意度。到目前为止，数字领域中的用户体验基本通过调动用户的听觉和视觉产生交互，未来，AI将会嵌入到用户界面的产品和设备中，借助计算机视觉，语音，自然语言，增强现实（AR）、虚拟现实（VR）和其他AI技术实现多维度人机交互。



产品和服务：借助AI技术，企业能够对来源广泛、内容多样的内外部数据进行多维度分析，从而更加全面地反映业务需求，避免单一来源数据造成视野局限；同时，多元化数据的交叉验证进一步提升分析的准确性、最大限度减少误差，为产品创新提供更好的数据支撑。未来，个性化及定制化服务将是企业产品和服务的核心竞争力所在。个性化推荐、定价甚至根据个人的特定需求量身定制产品将是未来发展方向，数据和先进的AI算法（例如强化学习）的出现使模型的自学习更精准，极大地促进了超个性化应用的发展。



商业模式:AI为企业提供重构商业模式的契机,梳理所有环节中“人”的因素以及技术能力,优化管理和技术架构,提升创新管理和风险管理效果。借助AI模型,挖掘市场参与者之间的内在关系,及时调整和创造适应市场供需能力,敏锐捕捉市场新机会,变被动服务为主动高效服务,利用具备高级算法的应用生成以数据为驱动的洞察,进而实现商业模式的创新,满足已有需求,创造新需求,创造新价值。

图3 借助AI释放激发企业潜能



来源:CIO视角:企业AI技术应用指南, 2019

1.3 AI行业应用持续拓展

从技术视角来看,单点AI技术不断演进。 2019年,图像、人体人脸识别等技术发展相对成熟,随着视频等非结构化信息数量的持续增加,未来视频分析技术会迎来进一步发展契机,计算机视觉仍是应用规模最大的技术领域;同时,智能语音、自然语言处理,机器学习发展也保持较高速发展。随着技术能力的成熟,企业逐步扩大对于AI能力的引入。**根据IDC预测,2024年,全球企业AI硬件、软件及服务方面的总支出将达1,106.6亿美元。**

图4 全球AI市场支出(十亿美元), 2019-2024



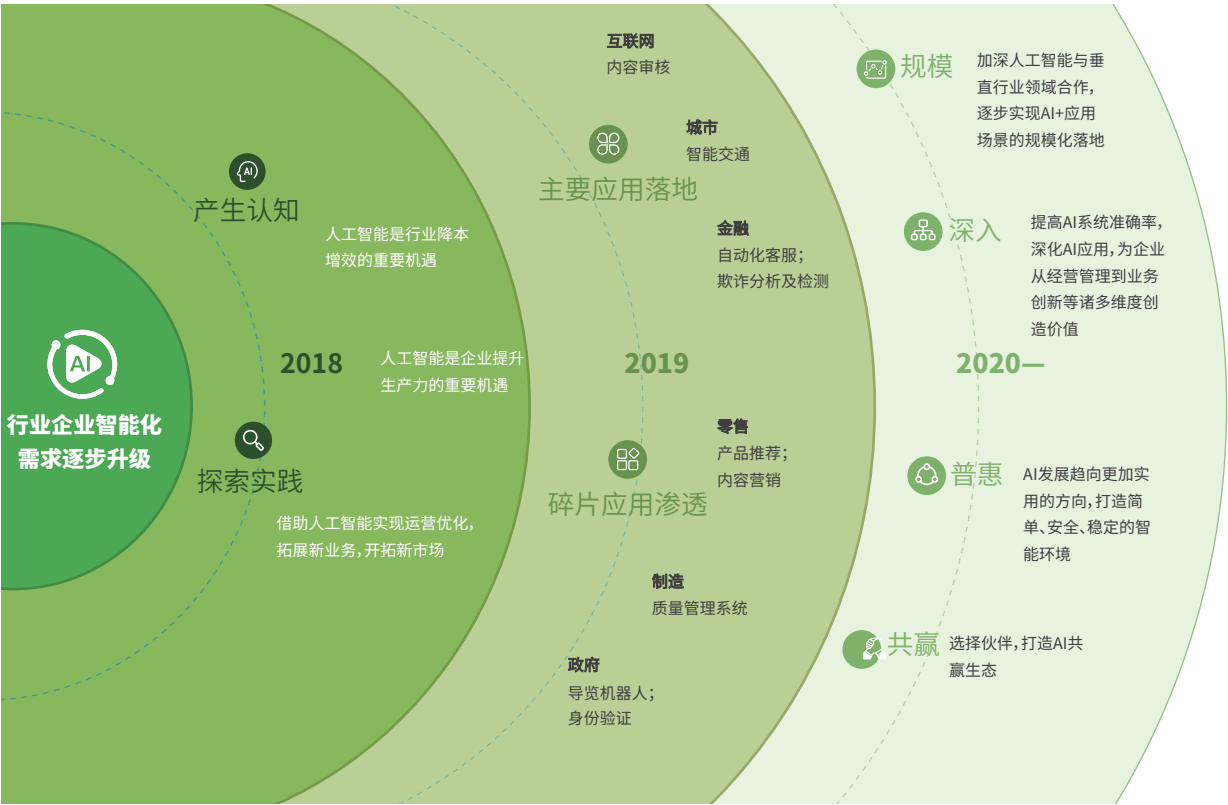
来源: IDC, 2020

企业将从对AI产生认知、规划通过AI赋能企业发展的愿景,到局部开展试点项目、累积AI能力,未来,其将加速AI在行业维度的渗透以及AI应用的成熟落地,如金融领域的反欺诈、政府以及信息服务领域的生物识别类身份验证、电商领域的智能客服等。**根据IDC数据,2019年中国企业级AI市场规模达到28.1亿美元,金融、政府、零售、制造、互联网及 IT 服务是AI应用规模较大的行业。**

- 以**金融**领域为例, AI技术有效整合海量历史交易数据、快速进行多元数据交叉比对,从而实时判断客户和交易风险等级,极大得提升了金融机构的风控效率;
- 以**电商**领域为例, 通过AI技术可实现大量基础运营工作的自动化,如上架和客服。此外, 结合结合大数据可实现精细化运营, 通过从海量异构数据进行实时分析, 提供个性化推荐和精准推送, 提升用户体验;
- 以**科学研究**领域为例, AI技术自动对海量实验和观测数据进行多维度交叉分析, 实现了远远超过人类的分析复杂度和分析速度, 为科研人员提供了有力的支持, 加速创新。

未来五到十年内, AI将继续向其他行业扩展。IDC认为, 未来AI有望在跨行业的视频结构化、业务流程自动化、金融领域的智能网点服务机器人、服务领域的营销互动、零售行业的供应链预测、电信行业的智能网络自动识别网络故障和潜在攻击以及医疗行业的辅助临床诊断决策、虚拟智能助理等方面实现广泛应用。在生产业务领域, 越来越多的行业也开始利用AI技术: 如能源行业的电力巡检, 交通领域的车辆识别等碎片化应用在中将开始有望实现广泛渗透。广泛的应用场景意味着企业对于智能化的需求将持续攀升, **实现规模化发展、深入化实施、普惠性服务以及打造合作生态共赢**将成为智能发展的新诉求。

图5 企业智能化需求逐步提升



来源: IDC, 2020

02

第二章

边缘智能拓展AI新边界

AI虽然未来有巨大的发展潜力,但与行业应用的深度融合才刚刚起步,成熟度不高,随着AI开始渗透到企业的核心业务,应用场景会越来越具体和细分,越发需要技术和应用上的不断创新和突破。

2.1 边缘计算融合AI走向边缘智能

AI向行业渗透面临新挑战

目前,AI更多地部署在云端的数据中心,在云端实现AI的训练与推理,充分利用云端强大的算力资源、提升处理效率;整合云端海量数据资源,有效改善AI模型的准确性和场景适配性。

伴随着数字化转型的兴起和数字经济的蓬勃,从产品的研发、生产、交付、使用,直至生命周期结束,每个环节都产生了大量的数据,数据成指数级增长以及连接的设备增加使几乎所有类型的业务都认识到,尽管进行了数据中心现代化和云迁移工作,但集中式计算模型并不是在所有情况下都有效,实践过程中,海量数据、低延时需求、网络受限以及高安全需求对企业带来多重挑战。



海量数据处理: IDC预测,到2025年,将有414亿台物联网设备,产生73,147EB数据。AI涉及的业务系统越来越多,并且逐步对接业务现场多样化的业务设备,与数据中心IT设备相比,业务现场的设备规模更大、应用场景更复杂,设备或传感器产生的数据将成倍增加,数据的类型也随之扩展,每秒收集一次读取的传感器(每次读取大约1KB数据)每天可以产生86MB的数据,那么数百甚至数千个传感器每天将产生千兆甚至TB级的数据流经网络。因此,完全在云端进行数据处理和AI分析往往超过系统和通信链路的处理极限,必须寻求新的应用方式分散云端的处理压力。



实时响应受阻: 不论是借助网络或增加端和服务器中间的跳数量,时延始终是限制响应速度提升的桎梏。在许多用例中,在云或数据中心往返的时间过长导致AI进一步发挥效应受阻。如自动驾驶汽车需要对传感器数据做出接近实时的响应来做出决策;在智能制造领域也是如此,在利用AI提高质量或产量的过程中,对时延的容忍度也较低。IDC估计,到2025年,全球产生的数据中大约有四分之一将是实时生成的。对于实时数据的分析和使用、缩短请求和响应之间所需的时间十分必要。



网络受限:物联网设备或其他相连的设备往往是可移动的,不受固定位置的束缚意味着设备有移出网络覆盖区域或具有有限带宽区域的可能性。例如,采矿设备将在地下数英里处工作,仍需要保持对遥测数据进行分析,以维持和安全运行;航运和物流公司需要在确保集装箱完整的前提下,全程追踪货物的运输情况,而运输环境则可能是在水、陆或空中切换。这意味着使设备在通讯带宽受限的条件下保持数据处理能力和采取行动的能力对于AI实际应用落地十分重要。



安全要求高:随着各国政府加强隐私保护以及数据安全,企业面临着合规带来的挑战,数据治理重要性逐渐凸显。对于企业而言,数据的汇聚和业务数据的增加使得安全变得更加重要,针对数据集中存储部位发起的攻击会引发大规模企业机密和客户隐私泄露,同时造成业务中断,带来营业收入和商誉损失;针对能源、化工、交通等事关国计民生的关键领域的攻击甚至会造成公众恐慌,安全成为智能化进程中部署和运行的重中之重,避免数据泄露或减少不必要的数据传输对企业来说具有深远价值。

上述挑战的出现意味着,在AI落地的过程中,企业有面临实际优化需求,在某些情况下,企业希望将处理移到离生成数据更近的地方,根据需求部署计算资源,更好地实时处理和存储数据,满足更快、更稳、更安全等需求。

边缘智能助力AI克服挑战

随着边缘计算的快速发展,计算能力快速向边缘迁移,AI也逐步从中心节点向更贴近数据源和业务现场的边缘侧拓展。但这并不是说边缘与云呈对峙状态,相反,两者可针对AI模型训练、应用开发、模型推理等任务形成分工合作局面,形成云-边-端协同的边缘智能架构:



边缘侧聚焦实时、小数据的处理,利用边缘网关和边缘服务器进行数据采集、预处理和存储、上传,运行训练好的AI模型进行推理,实现智能检测、控制等功能,并将结果回传云端。



云侧聚焦长周期、大数据的处理,在云端实现模型训练、数据共享及其它大数据分析等应用。

图6 云-边-端协同架构示意图



来源：IDC，2020

云-边-端协同的边缘智能架构，为云AI解决方案所面对的诸多挑战带来显著改善：



海量数据的获取和处理。边缘通过广泛的协议支持和连接支持，可以实现更广泛的多样化数据采集，同时将大量数据在边缘存储和处理，大大减轻网络传输和云端计算负荷。



实时响应。由于大部分计算和存储在本地完成，故相对于远端的云来讲能够显著降低网络传输延迟，满足实时控制等对延迟要求较高的边缘智能应用。利用边缘计算来定位数据创建点和数据使用点，可以将延迟从60-100ms +减少到1-5ms。



更好的网络适应性。边缘计算可以在边缘完成大部分计算，只将结果传输回云端，对传输网络带宽要求本身较低。同时，在网络不良或中断时，仍能运行全部或部分工作，降低了网络变化对应用的影响。

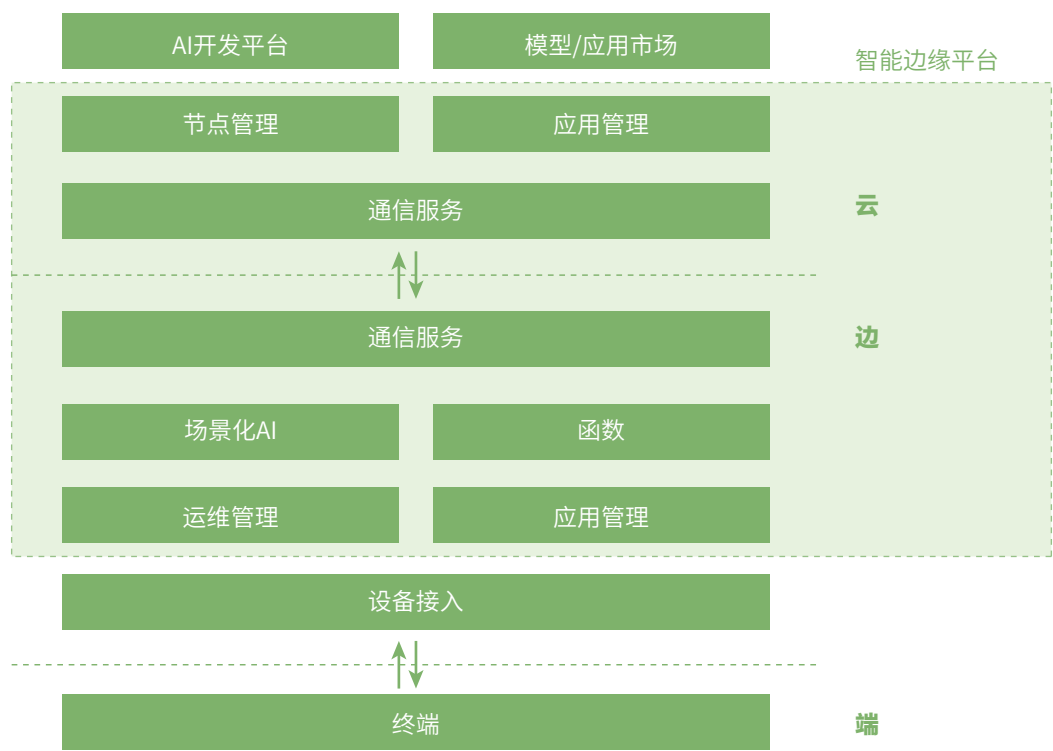


更高的数据安全性。边缘设备可以将敏感数据先在本地进行脱敏处理，之后再传递到云端，提高了数据安全性。边缘设备也可以将敏感数据仅在本地存储和处理，满足很多场景对数据本地存储的要求。

通过多方面改善，边缘智能将显著提升AI针对现场多样化的业务场景的适应性，从而更好地支撑业务运营、为客户创造更多的价值，为组织通过数据流获得洞察力，从而采取行动带来新机会。

为方便边云架构应用的开发部署，云计算服务商等厂商将一系列用于边缘节点管理、AI开发和部署的云原生的软件整合形成智能边缘平台，通过平台将云端AI应用、模型下发到边缘节点，并通过云端实现对边缘节点的统一远程监控、管理，为企业应用边缘智能提供完整的边缘协同一体化解决方案，同时也实现云端AI开发平台积累的大量AI模型和应用在边缘的复用。

图7 平台助力边缘快速智能化



来源：IDC, 2020

2.2 边缘智能推动AI全面升级

如果说AI的普及和逐步落地加深了对数据的“感知”，那么边缘智能将加速推进从数据中产生“认知”。边缘智能将更好助力企业实现数据智能自治、统一管理AI应用部署管理、汇聚生态资源，保障安全的前提下解决AI进一步行业落地的挑战。

图8 边缘智能开启AI新时代



来源：IDC，2020



数据智能自治

边缘智能将数据处理从中心节点迁移至更靠近数据源和业务现场的边缘侧，最大限度减少网络传输时延，提升业务响应速度，更好得适应边缘侧高实时性业务场景的需求，如工业设备控制；在本地处理海量数据，仅上传有价值的数据洞察和异常数据、减少数据传输的带宽消耗，如生产线高清质检视频本地处理；将数据处理限定在用户具有完全掌控能力的范围内，避免数据上云带来敏感数据外泄风险，如涉及企业技术机密的工艺参数；实现关键业务在一定范围内的自主处理，即使云边之间的网络中断，业务连续性业也不受影响。

同时，边缘智能通过标准化接口增强AI应用与底层硬件之间的协同，实现与业务场景相匹配的AI算力，并针对业务数据特点进行深度优化。

- ▶▶▶ **场景化AI能力**：边缘智能简化了上层AI应用调用底层硬件设备的AI算力，使得AI应用能够更好得适配不同业务场景中的硬件设备，实现AI能力在功耗、性能、成本等限制因素之间的最佳平衡。
- ▶▶▶ **行业数据大规模并行处理能力**：边缘智能使得AI应用高效调用设备算力资源，实现大规模并行计算，满足特点行业业务数据的快速处理需求，如工业设备参数矩阵运算、通信信号处理、视频编解码等。在视频安防、机器视觉质检、无人机巡检等视频AI应用高速发展的背景下，大规模并行计算能力的价值将持续增加。**IDC预测，2020年全球视频监控产生的数据约18.1PB，占同期物联网数据量的83.1%。**



云边一体的平台化部署

边缘智能提升了云边协同的效率，使得AI应用在云边之间平滑迁移，提升AI应用开发部署效率；以智能边缘平台为载体，汇集行业共性需求，简化AI行业应用开发；边缘智能统一云边之间IT架构，为边缘智能应用远程运维管理提供有力支撑。

- ▶▶▶ **云边联动，形成一体化AI平台**：边缘智能在边缘侧构建起云原生的运行环境，使得在云端已经稳定运行、经过充分验证的AI应用和算法，可以平滑迁移至部署在边缘侧的边缘智能设备，如基于NVIDIA的CUDA平台开发的AI应用，可以无需重新编程从云端下发到不同EGX服务器和安装了Jetson模块的边缘设备。不仅极大降低了开发测试的经济和时间成本，而且显著降低了业务风险；AI任务和业务数据处理可以在云边之间自由选择，在适合的位置处理适合的数据，充分发挥云端规模效益、全局视野的优势以及边缘侧实时响应、数据可控性强的优势；已有边缘智能应用的升级和新边缘智能应用开发可以实现“云端开发、边缘部署”，大幅提升开发测试效率、加速部署。

▶▶▶ **沉淀行业最佳实践:**随着边缘智能的逐步渗透,一方面,边缘承载的边缘智能行业应用日趋丰富、汇聚的行业共性需求和最佳实践也同步增长,通过汇总整合,形成软件开发工具包(Software Development Kit, SDK)等可复用的通用组件,加速行业应用开发;另一方面,行业化数据在边缘上快速积累,在合规的前提下,海量行业数据为AI模型提供了丰富的训练数据,通过AI模型预训练,可以有效缩短后续部署阶段的模型训练时间、加速行业应用的部署。

▶▶▶ **统一远程管理:**与中心化AI应用相比,边缘智能应用数量更多、部署更分散,大量应用部署在偏远地区和高危环境,现场维护成本高、效率低、甚至不可行,远程管理成为边缘智能的运维的首选方式。边缘智能统一了云端和边缘侧的应用部署环境,尤其是屏蔽了嵌入式设备的软硬件差异,使得在云端广泛使用、运维人员更加熟悉的运维工具软件和脚本可以在边缘侧稳定运行,显著提升运维效率。此外,云边协同可以使平台得以监控远程设备的资源和可用性,乃至在云上模拟终端设备的实时智能化实例,以便为物联网搜索和实时远程设备资源调配准备好硬件与系统基础。譬如,通过物联网搜索,电动汽车可以实时搜索充电桩和规划充电路线。IDC《中国边缘智能决策者调研》显示,92.2%的受访者使用嵌入式Linux做为边缘智能的操作系统。嵌入式设备支持成为统一远程管理的重点。



构建边缘生态

边缘智能通过构建统一的平台屏蔽不同软硬件之间的差异性,提升多厂商协同效率,推动商业模式创新。

▶▶▶ 实现多厂商分工协作和开放共赢

智能边缘平台统对上屏蔽底层硬件的差异性、提供统一的硬件资源接口;对下屏蔽上层业务应用多样性,提升上下游协同效率。

- 智能边缘平台厂商专注于平台通用ICT能力的提升,通过多样化的AI算力满足不同场景的AI需求,构建统一的平台实现算力资源的高效管理和调度,不断提升云边协同能力和效率,为AI应用开发调试提供更高效便捷的工具。
- 硬件设备厂商专注于硬件设备的制造,在不同业务场景中实现成本、功耗、性能之间的不同平衡;不断提升设备在多种严苛环境中的可靠性(如高温、高湿度、强电磁干扰等),适配更丰富的业务场景。
- AI应用开发者发挥行业专有知识,专注于业务逻辑的实现和流程的优化,通过AI不断解决用户的实际问题、为用户创造价值。
- 系统集成商专注于行业化解决方案的部署实施,根据用户需求整合硬件设备、算法、AI应用、云服务等多种资源,形成完整解决方案,快速实现行业落地。

商业模式创新

智能边缘平台云原生架构使得多应用有序共存、互不干扰，更好得支撑商业模式创新；云边高效协同使得智能边缘平台的运行更透明，更加有效保护各方利益。

智能边缘平台云原生架构的多租户隔离机制使得多个用户的业务应用彼此独立运行、互不干扰，为边缘智能共建共享提供了坚实的保障；云原生架构弹性扩容和云边高效协同使得多样化的定价策略成为可能，云端计费系统可以有效监测智能边缘平台的资源使用，实现按使用时长、数据处理量、调用次数等多种方式定价，不仅赋予用户更多的选择、加速边缘智能的普及，而且使得边缘智能建设和运营方的收入分成更加透明；智能边缘平台汇集海量行业化、场景化的数据资源，实现数据变现、创造新的收入来源。**IDC调研显示，33%的企业通过与生态系统合作伙伴扩展核心产品，将现有产品货币化；26%的企业将资产呈现给生态伙伴，以便共同创新。**



端到端安全

智能的边缘成为数据和业务应用的汇聚节点，安全成为边缘智能部署和运行的重中之重。

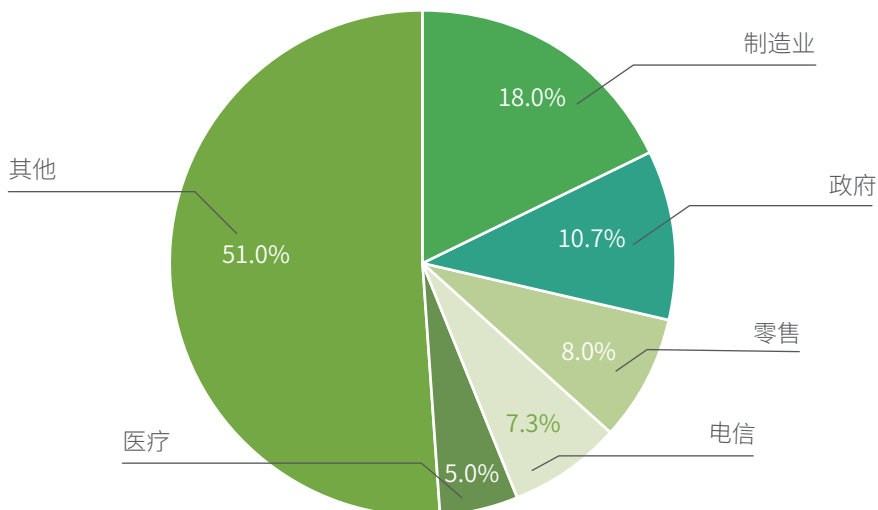
- ▶▶▶ **硬件安全**：智能化边缘数量大、部署分散，难以通过集中式管理保障设备安全；大量边缘智能部署在无人值守的设备上，攻击者更容易通过修改硬件电路或注入恶意代码的方式发起攻击，边缘智能自身的安全防护能力必不可少，尤其是对异常硬件组件和软件代码的主动识别能力。
- ▶▶▶ **数据安全**：智能化边缘位于用户掌控能力更强的边缘侧，处理的企业机密数据和客户隐私数据更多；高效云边协同在边缘侧被入侵时，一定程度上放大了网络入侵带来的危害。因此，边缘智能需要强化敏感数据在存储和传输过程中的安全性，通过数据加密、数字签名等机制防止数据被窃取或非法篡改；同时保证密钥自身的存储和使用安全，从根本上确保敏感数据的安全性。
- ▶▶▶ **AI模型安全**：边缘智能逐步向关键业务渗透的同时，AI模型也成为新的攻击目标，针对AI模型的非法篡改严重影响企业的业务连续性，甚至引发灾难性后果，诸如工业设备、医疗设备运行异常。边缘侧设备大量无人值守进一步加大了AI模型面临的威胁，AI模型安全日益成为边缘智能的默认功能。

边缘智能行业应用空间广泛

不同行业应用的差异化,决定了AI落地必然是场景化的,边缘智能是实现AI在行业差异化落地的关键。**IDC预测,到2024年,全球企业在边缘应用上的支出将达到2,500亿美金,其中与AI技术相关的支出将达到326亿美元,占整体边缘市场支出的13%。**

尤其是在制造业、政府、零售、电信、医疗等重点领域,敏捷联接、实时业务、数据优化、应用智能、安全与隐私保护已经成为核心需求,边缘智能将是实现行业数字化转型的关键。**IDC预测,到2024年,制造业、政府、零售、电信、医疗五大行业在边缘智能应用上的支出将达到159亿美金,占整体边缘智能市场支出的49%。**

图9 IDC全球边缘计算AI领域支出预测(分行业), 2024



来源: IDC, 2020

3.1 制造业

制造业智能化升级是大势所趋

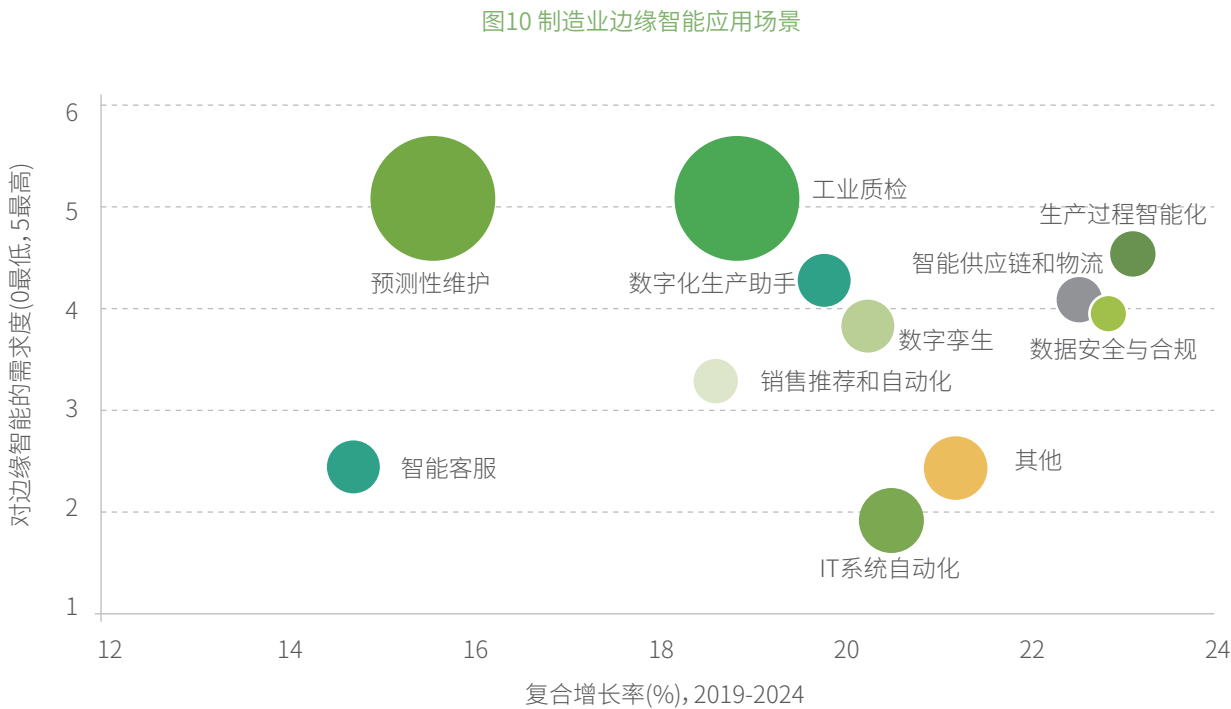
制造业是国民经济的基础性产业,为社会稳定和经济增长提供基础性保障。然而,在全球化竞争日趋激烈的大背景下,制造业正面诸如劳动力短缺、经验传承困难,品控要求趋严,交付周期缩短等诸多问题,给制造企业发展带来挑战。

面对这些挑战，制造业正不断加深推进新一代信息技术与制造业深度融合，加速数字化转型，提升制造业自动化和智能化水平，从而实现传承经验、提升品控、缩短交期，提升竞争力。

边缘智能在制造业应用前景广泛

制造业智能化发展已开展许久，边缘智能的应用将通过数据在边缘的计算和存储，实现降低延时、数据保护等效用，为助力AI在制造业的应用发挥重要作用。IDC《中国边缘智能决策者调研》显示，电子制造行业30.6%的被访者认为降低延时是边缘智能最大的价值，在所有的价值中位列第一，另有27.8%的被访者认为“保护公司数据安全”是其重要价值，13.9%的被访者认为边缘智能还可有效“保障数据主权”。

边缘智能在制造业的应用场景覆盖了生产、物流、经营管理、客服等诸多领域，并将保持快速增长。IDC预测，将在未来几年，边缘智能支出在生产过程智能化、智能供应链与物流、数据安全与合规等应用场景将保持20%以上的年均复合增长率。预测性维护和工业质检对边缘智能需求较高，是市场空间最大的边缘智能应用场景，也有望在未来几年保持14%以上的年均复合增长率。在智能制造向自动化、智能化升级的趋势下，未来边缘智能将持续帮助AI技术持续深化制造业应用场景，并在各个制造子行业中充分拓展。IDC预计，2020年，全球25%的制造业企业将引入边缘计算，提升决策速度、降低数据管理成本。



图中气泡大小代表市场规模

来源：IDC，2020

案例:边缘智能助力智能质检

需求背景

对电子、消费品、钢铁、服装等诸多领域的制造业企业而言,产品质量检测在生产过程中占有重要比重。而随着市场对产品质量要求的不断提高,以及设备升级、自动化水平提升,实现产品全检、提高质检水平成为生产企业的普遍诉求。

传统产品质检主要依靠人工目视检测和传统机器视觉等检测手段,但人工目视检测成本高、效率低,难以实现全检。传统机器视觉基于模式匹配技术,也很难满足复杂场景、复杂外观缺陷的检测需求,漏检、误检率高,两者均难以满足日益提升的产品质检需求。

在此背景下,通过AI技术显著提升机器视觉检测能力和适用性的智能质检系统,大幅减少人工目视检测,成为应对当前产品质检水平的重要手段。AI技术的应用带来显著的优势,但工业自身特点也对智能质检系统开发提出了一定挑战。例如智能质检系统对测量的周期和实时性要求高,对加速计算要求比较高。工业生产节奏加快,对系统开发时间的要求也在缩短,开发周期可能只有6-8周,但同时生产质检数据的积累速度远低于数字化产品的交付,AI训练数据的小样本问题突出,尤其是对于新产品的生产制造。此外,在丰富的实际生产现场环境中,不同生产设备和生成线在物理空间、散热条件、电磁干扰强度等方面千差万别,对AI设备的物理形态要求也各不相同,从轻边缘的嵌入式网关、工控机到重边缘的紧凑型服务器、标准服务器都可能存在,对系统的广泛适应性提出了更高要求。

关于NVIDIA EGX

NVIDIA EGX是面向多元异构AI计算的统一智能边缘平台,整合多种硬件、软件、云资源,形成完整的边缘生态体系、提升产业系统效率,加速AI行业应用的落地。

繁荣边缘生态赋能数据智能自治

EGX整合从边缘服务器高性能AI芯片T4到嵌入式低功耗AI芯片Jetson,为边缘智能应用和边缘数据处理提供统一的应用开发和数据交互接口,简化应用开发,满足边缘数据中心、生产作业现场、户外无人值守设备集群等多样化部署场景和业务需求。

众多EGX认证的标准服务器、紧凑型边缘服务器、工控机等成熟商用产品以及对主流容器(如kubernetes)、企业级操作系统(如Redhat Linux)、私有云(如VMware)、公有云(如微软Azure)支持进一步加速了基于EGX的边缘AI应用的落地实施。

丰富的云边协同开发工具加速边缘AI应用部署

EGX云原生的架构以及与主流容器技术的无缝集成使得云端经过充分验证的成熟AI应用可以无缝迁移至边缘侧,尤其是边缘数据中心,保护历史投资、加速边缘AI业务部署;云端开发、边缘部署的模式极大降低了边缘AI应用的开发测试周期和复杂度;丰富的远程管理功能使得海量边缘设备中的AI应用可以高效、及时得进行更新迭代,支撑边缘AI业务创新;ISAAC、Metropolis、Clara等行业SDK整合行业最佳实践、提供行业化算法库和开发测试工具(如基于机器视觉得质检应用的算法库、面向医学影像数据AI处理的工具包),加速行业专属边缘AI应用得开发。

端到端安全保证业务合规

EGX的多应用隔离管理机制有效避免不同应用之间的越权访问,保证边缘业务数据的安全性,针对边缘AI模型的安全校验从根本上保证了边缘AI业务的可靠性,避免算法模型被非法篡改带来的业务风险;EGX生态体系内基于可信计算技术的硬件,尤其是服务器,实现了对未知威胁和底层恶意软件(如BIOS病毒)的有效抵御,为边缘AI应用的安全可靠运行提供基础保障。

解决方案

天准科技是一家智能制造解决方案提供商，以机器视觉为核心技术，专注于工业领域，主要产品为工业视觉装备，包括精密测量仪器、智能检测装备、无人物流车（AGV）等，实现尺寸测量、产品缺陷检测、自动化生产装配、智能仓储物流等功能。

天准科技基于英伟达GPU及EGX平台，研发了多个型号智能边缘计算设备，并整合天准科技机器视觉算法、边缘计算设备、检测装备研发成光伏硅片分拣检测装备、平板显示检测装备、自动外观检测设备等一系列智能检测装备，将机器视觉技术应用于工业质检场景，实现对工业零部件和产品的实时在线检测，降低人员成本，提升检测效率。

以光伏半导体应用为例，因为硅片瑕疵形态多样，无明显的规律性，用经典的机器视觉检测方法容易造成误检、错检或漏检。通过引入深度学习AI算法，实现对单晶、多晶、黑硅等多种工艺硅片的尺寸、线痕、翘曲、崩缺、表面缺陷、隐裂、电性能等特性一站式高速全检和分级；对硅片中瑕疵特征进行高效、高识别率识别与提取，并将硅片品控过程中上料、全项检测、数据分析、下料分选四大工序集成到一个自动化设备中，满足新能源领域的硅片等类似产品的分选检测需求。通过英伟达GPU上运行TensorRT深度学习推理工具和CUDA并行计算加速工具，检测速度明显提升。

应用成效

- **降本增效**：天准借助机器视觉算法，将硅片分选检测设备产能提升到业界1.5倍以上，实现每小时8,500片的检测速度，大大提升工作效率，成本降低50%以上。
- **精度与速度并行**：GPU为AI任务处理提供了充沛的算力，满足生产线AI任务实时处理的需求。例如，在线质检会在一个节拍内（通常几秒钟）对生产线上的零部件或成品进行数百个参数的检查，GPU充沛的算力在保证质检精度的同时，避免质检作业延误整个生产线的运行。
- **提升数据安全**：数据在用户现场处理，避免设备运行和工艺流程数据在外网传输带来的数据失窃风险。
- **系统开发速度提升**：英伟达EGX丰富的生态体系提供了开源软件、工具链、基础算法、硬件产品、云平台支持，精简通用软硬件研发团队，专注于行业专有核心算法的开发和解决方案与生产线整合。借助EGX丰富的硬件产品组合，避免了系统方案硬件选型受限，满足诸多场景下标准化程度较低的产品的智能化制造需求。
- **高质量应用交付**：借助云原生架构实现AI应用从云到边的无缝迁移，开发人员可以在成熟算法和应用的基础上根据业务需求进行快速优化调整，而不必担忧云端开发测试环境与现场不同AI设备之间的差异性，不仅加速了应用开发，也为核心算法和应用的持续升级维护提供了稳定、统一的环境，避免硬件差异性带来的多分支、多版本维护。
- **模型持续迭代优化**：基于云边协同架构的智能质检系统，在业务初始阶段依托小样本训练的AI模型部署在边缘服务器，随着生产的进行，由技术人员在边缘进行在线判断和标注，持续动态迭代优化边缘的AI模型，同时基于云边协同将实际生产数据同步到云端，不断提升样本量，改进模型质量，再将改进后的模型更新到产线边缘服务器，进一步提升推理精度，从而使算法进一步匹配实际生产需求。

3.2 政府(智慧城市)

现代城市加速迈向智能生态

随着城市化进程的持续推进，现代城市发展也日益面临越来越多的压力，如交通拥堵、社区安全、环境恶化等，给城市管理者和城市居民带来巨大压力。城市管理者也一直在努力通过信息技术提升城市服务水平。然而，传统城市管理手段在面对产业升级、可持续发展、服务升级、高效应急处置等方面问题时，仍然乏力。

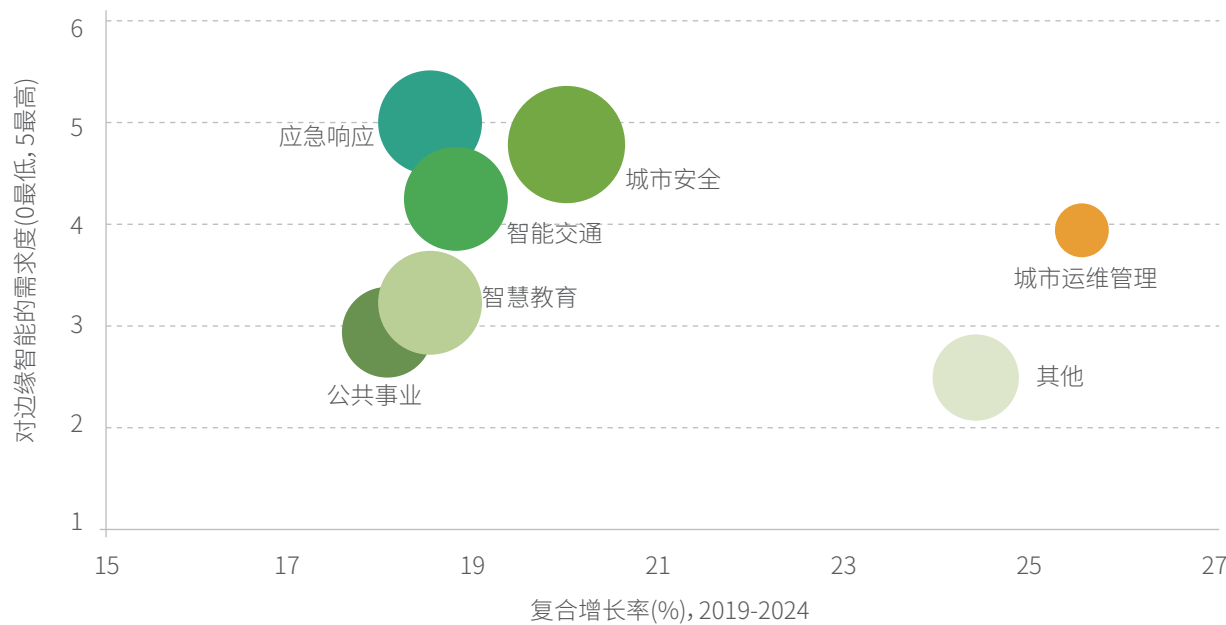
随着物联网、AI等新信息技术向城市融合，在政府、企业、社区、市民等各个实体中产生大量城市智慧治理、智慧民生、智慧交通等应用场景，为提升管城市精细化管理与智能化服务提供新出路，促进未来城市形成跨多行业的复杂智慧生态系统。

边缘智能助力智慧城市向精细化渗透

AI在智慧城市的应用已经渗透到城市管理、交通、民生等诸多方面，并形成了城市智慧大脑等中心化应用。但随着智慧城市向精细化和社区化下沉，中心化的智慧城市系统难以满足社区管理、局部交通管理、及时应急响应等精细化应用，依托边缘智能在社区等开展智能应用，成为助推智慧城市走向精细化的重要方向。

边缘智能在智慧城市的应用，涉及应急响应、城市安全、智慧交通、智慧教育、公共事业、城市运维管理等诸多方面。**IDC预测，边缘智能支出在城市运维管理场景将在未来几年保持25%的高增速。**智慧城市其他众多应用场景，如智慧交通、应急响应、城市安全等也对边缘智能需求较高，潜在市场空间都较大。

图11 智慧城市边缘智能应用场景



图中气泡大小代表市场规模来源：IDC, 2020

案例:边缘智能助力智慧交通

需求背景

交通行业是社会运行的动脉,支撑人员流动和物资交换,保障民生,促进产业协同。高效的交通尤其是城市交通方便民众日常生活,缩短通勤时间、改善出行体验;加速物流配送尤其是末端配送,提升物资供应效率;方便人员往来和交流协作,推动经济发展;为旅游、会展等行业提供更有力的支撑,改善游览和参观体验的同时,尽可能减少对本地居民日常生活的影响。

目前交通行业发展面临如下挑战:

- 道路交通拥堵:交通拥堵问题不仅严重影响民众的出行体验和生活质量,而且降低了人员交流效率,进而影响整体经济运行效率,尤其对以旅游、会展业为支柱的城市而言,更显著地制约了经济发展。
- 停车场资源紧张:停车场资源紧张不仅对居民日常生活和企业物流配送带来不便、影响民生和经济,而且一定程度上增加了违章停车甚至应急道路和消防通道的违规占用,带来安全隐患。
- 执法人员数量有限:城市机动车保有量的增长远远快于交通执法人员的扩编速度,单纯增加交通执法人员不仅应对效率低下,而且给公共财政带来沉重的负担,影响更加重要的民生保障和经济发展项目的建设。



解决方案

北京米文动力是一家智能边缘计算产品和技术提供商，研发软硬一体的智能边缘计算产品和解决方案，通过提供多种配置的硬件组合、统一的软件堆栈、垂直行业解决方案，满足物流运输、交通管理、车路协同、视频分析等多个领域在智能边缘端算力升级的需求。

米文动力基于英伟达EGX平台打造智能交通系统，以英伟达边缘智能芯片为基础，开发基于机器视觉和数据智能分析的边缘智能应用，此外，借助英伟达EGX中集成的Metropolis SDK，诸如Multimedia SDK、Deep Stream SDK、Tensor RT SDK等，形成统一的开发环境和丰富的功能模块，实现路况监测、路口交通控制、停车管理和违章抓拍等功能。

- **路况监测**:对摄像头拍摄的路况进行实时分析，获得路况数据，为城市交通运营管理提供数据支撑，包括通过识别车辆进行车流量统计；通过视频分析计算车路行驶速度、进而计算特定路段整体车流同行速度；识别疑似道路交通事故，触发告警。
- **路口交通控制**:综合分析路口各个方向的车流量和通行速度数据，对路口不同交通信号灯的时长进行动态控制，实现整体道路通信速度的优化。
- **违章抓拍**:对违章停车、违规占用应急车道、“闯红灯”、压线等违章行为进行实时识别和拍照取证，并且识别车牌号码，自动生成完整违章记录，包含违章类型、时间、地点、类型、车辆牌照、违章现场照片等，方便执法人员的后续处理。
- **停车管理**:通过摄像头对公共停车场和路侧停车位进行监测，实时更新车位占用信息，提高车位利用率；通过车牌识别进行计费管理，自动记录停车起止时间、生产账单数据，提升缴费效率，进一步加速车位周转速度。

应用成效

- **中心网络压力有效减轻**:高清视频数据在边缘侧处理大幅提升了处理效率，支撑路侧控制系统(如交通信号控制设备)进行实时决策；仅上传结构化数据(如车流量统计数据)和异常数据(如违章照片)有效节省了带宽，降低了市交通运营管理中心网络压力。
- **运维效率有效提升**:算法、应用、数据洞察在边缘侧的摄像头、交通控制设备与云端的城市交通管理系统之间无缝迁移，极大提升了运维效率，运维人员不再担忧云端算法、应用、数据格式在边缘海量异构设备上的兼容性，避免了复杂的适配工作。在某车路协同项目中，得益于EGX的云边高效协同，运维成本降低20%。
- **AI应用开发加速**:IDC《中国边缘智能决策者调研》显示，智慧城市领域66.7%的被访者认为“软硬件定制化开发导致项目周期长”是边缘智能实施过程中的最大问题，远高于其它挑战(初始投资高、投资前景不明朗等)。通过沉淀车牌识别等各类智慧交通相关基础开源算法以及各类通用和专用SDK整合，提速智慧交通行业应用开发。

3.3 电信行业

电信业在数字经济时代面临新机遇

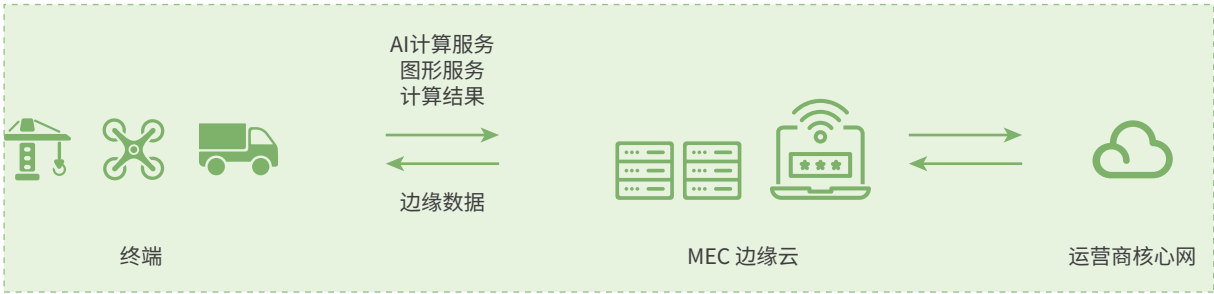
电信行业是数字经济时代的核心基础设施,通过广覆盖、可靠的数据传输通道和处理平台,为数字化产品的生产、传输、交付、使用提供基础性保障。宽带网络支撑不同企业、不同地域之间的大规模数据交换;移动网络不仅使得企业员工打破时间与空间的限制,进行实时无缝协同,而且保证消费者随时随地享受数字化产品;数据中心资源有力支撑了企业级的传统IT应用和云服务的部署、运行。

随着新一代信息技术发展,尤其5G和物联网的发展,电信行业不再局限于提供基本的电话和互联网服务,电信公司在面向消费者之外,具备了提供更多行业企业解决方案的空间,其中很多解决方案将应用AI技术。此外,电信公司多年积累的庞大的客户群、数据、通信基础设施也具备广阔的优化空间。

边缘智能助力电信业务优化拓展

5G时代,很多行业人工智能和创新应用,对网络延迟要求高,都需要边缘计算来满足应用需求。电信运营商通过建设具备GPU的MEC边缘数据中心,既可以基于资源共享为各类行业用户提供边缘计算服务,降低企业自建边缘计算中心的成本,实现运营商和企业双赢,同时又可以基于边缘的AI和图形图像计算能力,拓展VR游戏,云游戏,AR远程指导,高清视频直播等新消费业务形态,为运营商构建新的收入渠道。

图12 运营商MEC边缘云架构

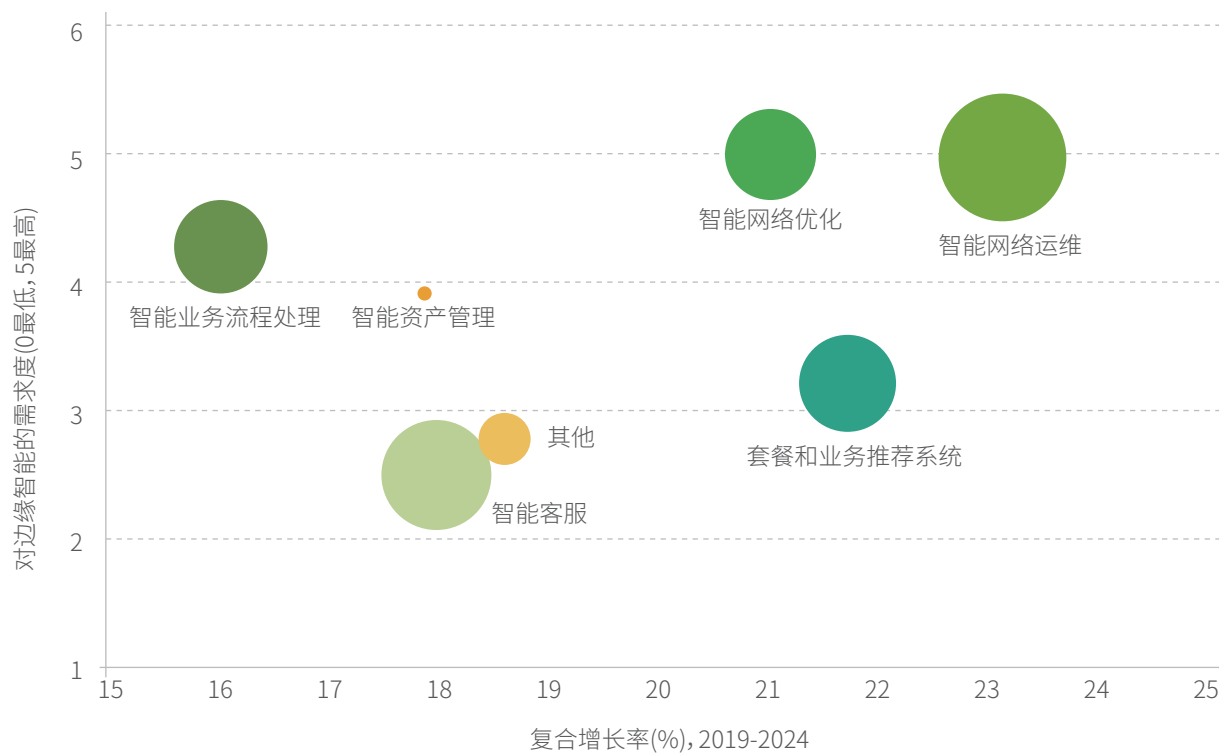


来源：IDC，2020



在消费者市场渗透总体趋缓的条件下，运营商也需要积极探索如何基于AI技术加速边缘智能的部署，通过在边缘应用部署智能网络运维、智能网络优化、套餐和业务推荐等应用，满足内部提升治理的需求，升级应用，加速业务创新。

图13 电信行业边缘智能应用场景



图中气泡大小代表市场规模

来源：IDC，2020

案例:边缘云助力中国移动业务拓展

需求背景

移动通信市场在传统网络用户渗透达到饱和和新兴技术加速拓展的背景下,经营面临挑战:

- 收入增长乏力:在主要发达国家和新兴市场经济国家,移动通信市场渗透率已超过或趋近100%,市场竞争从增量市场转向存量市场,通信服务营业收入增长持续放缓。
- 传统主营业务竞争激烈:互联网公司的跨界竞争严重冲击电信运营商的语音、短信、数据中心等业务。互联网公司的语音和即时通讯OTT业务迅速普及,运营商的语音、短信业务锐减,依赖于此类业务的彩铃、短信新闻等业务也同步快速下滑;互联网公司在云计算领域的生态和技术的领先优势,大幅分流了电信运营商主机托管、应用托管等传统数据中心业务。
- 资本支出高:一方面,技术迭代加速使得电信运营商的网络生命周期缩短、网络升级节奏加快;另一方面,介于电信行业的基础设施地位和消除数字鸿沟、推动地区间平衡发展的重要地位,各国政府对于电信运营商的网络覆盖、网络通信质量乃至建网速度的监管逐步强化,进一步增加了资本支出。



解决方案

中国移动中国内地最大的通信服务供应商之一，亦是全球网络和客户规模巨大、盈利能力领先、市值排名位居前列的世界级电信运营商。跟随5G建设整体发展趋势，中国移动基于英伟达EGX搭建边缘云，为企业边缘计算业务提供支撑。

- **边缘IaaS**：依托中国移动丰富、下层的边缘数据中心，构建边缘IaaS、适配不同部署环境（标准数据中心、普通机房、现场设备间等），使数据处理更加贴近用户侧，实现低时延和高可靠，支撑云游戏、云VR等消费者业务和智能制造等企业业务。
- **运维管理平台**：对规模庞大、部署分散的边缘云进行集中式统一管理，全景呈现边缘云及底层IT设备关键KPI，诸如工作负荷、网络流量、设备运行状况等。
- **细颗粒度AI算力虚拟化**：对不同性能的GPU进行统一的虚拟化管理，形成资源池；在单一GPU基础上虚拟出多个逻辑GPU，满足更小算力的场景。

应用成效

- **有效支撑业务创新**：MEC边缘云为中国移动在消费者服务和企业服务市场的业务创新提供基础设施保障，培养新的业务增长点，并帮助中国移动发挥云网融合、基础设施下沉的优势，与互联网厂商进行差异化竞争。GPU的大规模并行计算能力有效支撑VR图像渲染、云游戏等计算密集型任务的实时处理，为5G云VR、5G云游戏等业务提供坚实保障。
- **降低运营和开发成本**：MEC边缘云利用EGX，对不同算力的GPU使用标准的硬件和通信接口（PCIe接口），降低硬件部署和扩容的复杂度，同时将底层多样化的硬件抽象为统一的边缘智能资源池，并通过容器化方式加速边缘智能应用部署和迁移；基于EGX的虚拟化机制，细颗粒度的GPU虚拟化机制既可以整合大量GPU实现超强边缘智能算力，又可以将单个GPU虚拟为多个逻辑GPU，并且按需弹性扩容，拓展更丰富的应用场景，使移动边缘云灵活应对各类不同级别算力要求平滑升级，降低运营商CAPEX、加速边缘云部署。

3.4 医疗健康与生命科学行业

人工智能显著提升医疗技术水平和供给能力

医疗健康与生命科学行业是国民健康和社会进步的基础保障,不断进步的医疗技术为大众持续提供优质的医疗和保健服务,更有效的治愈各类疾病、攻克更多的疑难病症,从而改善大众身体素质、提升生活水平、减少疾病对社会发展的干扰与危害。

在数字经济时代,信息技术、医疗与生科科学技术以及创新的医疗保健商业模式是个人健康和整个社会健康水平进步的核心驱动力,信息技术支撑和促进医疗技术革新升级,显著提升疾病诊断和治疗水平,提高医疗资源的供给能力,有效应对医疗健康服务持续增长带给各国政府的挑战,促进民众整体健康水平的提升和社会进步。

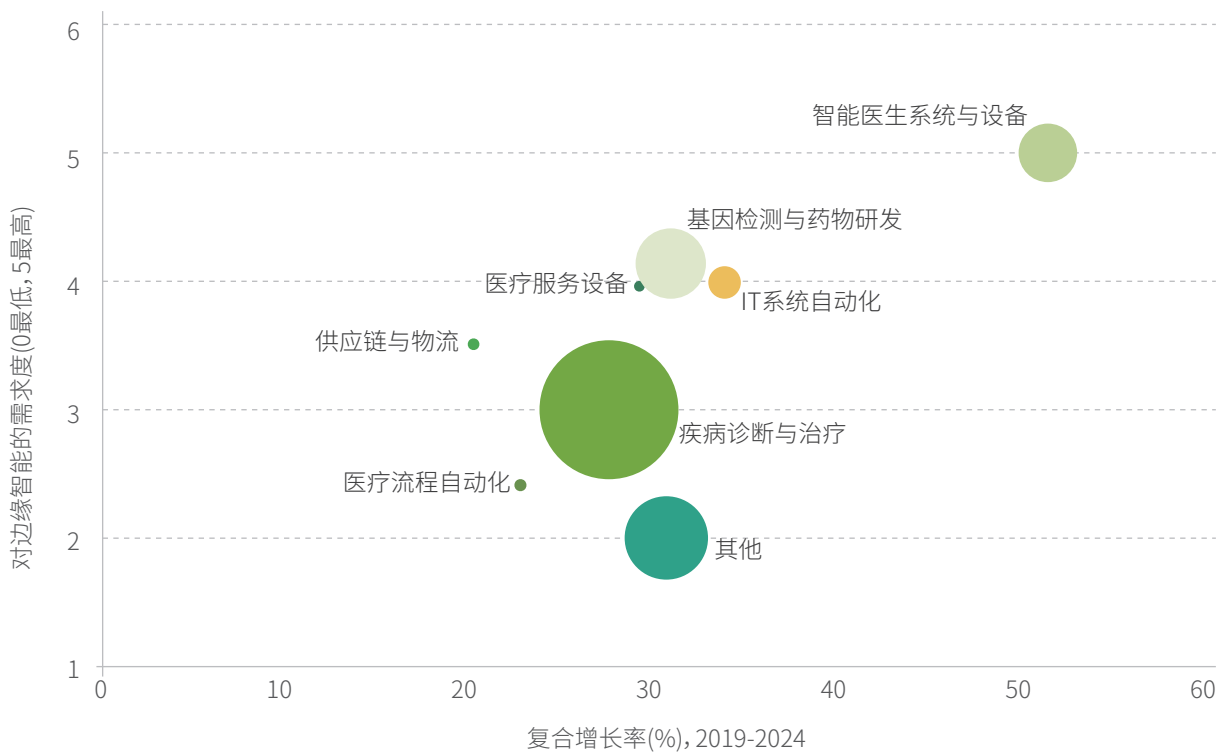
边缘智能助力医疗模式转型,优化医疗效率与效果

人工智能技术在医疗健康与生命科学行业已经广泛开展探索应用,并在医学影像辅助诊断、基因检测等一些领域深入发展,取得了良好的效果。

边缘智能的出现拓展了人工智能技术的应用深度和广度,在医疗健康与生命科学行业已经初步展示出良好的应用效果,并展示出巨大的应用潜力。边缘智能技术的主要应用领域及其未来发展趋势有如下几个方面:

- 边缘智能应用到医学影像辅助诊断中,在大型医疗机构的各类医学影像设备中部署边缘计算平台,支撑不同疾病AI辅助诊断软件集中部署与管理、联邦学习、模型开发等技术,提高辅助诊断技术水平与工作效率;在区域医疗的多机构之间构建云边一体化平台,支撑整个区域内的医学影像诊断,提高整体的医疗质量和效率。
- 在基因检测领域中,在特定区域、特定领域内甚至全国范围内部署云边一体化的边缘智能平台,支撑基因检测AI软件的开发、训练和部署,提高基因检测的速度;边缘智能技术尤其对于新一代基因测序技术具有显著的提升作用,云边一体化边缘智能技术能够支撑测序工作不受地点和实验室环境的制约,对于微生物、病毒和人类基因的测序能力和效率有巨大提升作用。
- 在医疗机构中,边缘智能能够部署于智能医生系统或者设备中,例如对于导诊机器人、康复机器人等进行统一的部署、升级与维护等,持续提高机器人的智能水平和工作质量;边缘智能技术也可部署在手术机器人设备上,提高手术机器人的智能化水平。中心云、边缘云通过对海量历史数据的深度学习和基于医生专家经验的算法设计,持续优化AI诊断模型,并下发到医疗设备和医生诊疗PC上,通过算法和AI模型形式固化专家经验、进行边缘推理,从而更高效得共享和传承专家经验。与此同时,采用联邦学习的模型训练方式,更好地保护隐私数据。
- 边缘智能也可以部署在慢病管理、临床试验、医院内自助服务等场景中,提升医疗质量和效果。

图14 医疗健康与生命科学边缘智能应用场景



图中气泡大小代表市场规模

来源：IDC, 2020

案例：边缘智能助力疾病快速诊断和流行性传染病防控

需求背景

基因测序已经成为当前病毒导致的流行性传染疾病防控的重要手段，通过对病原微生物的基因测序，可以在第一时间获取新发未知病毒的基因组序列信息，揭示病原相关特性、分析病毒的进化来源、研究病毒的致病病理机制等，为突发疫情的防控和后续研究提供帮助。2020年新冠肺炎疫情的防控过程中，各地疾控中心开展对病毒的基因测序，对疫情检测和防控发挥了重要作用。

当前基因测序设备主要基于二代测序技术，设备体积较大，主要部署在重点城市的实验室，一旦城市之外地区发生疫情，需要将病原体样本运送到具备条件的城市的实验室开展检测，存在两个问题：一是样本运输过程中可能存在病原体泄露，增加城市传播病毒的风险，二是样本运输和检测期间过程耗时较长，不利于病毒防控。

解决方案

某第三代基因测序公司,利用边缘智能技术,研发了可在病源地及时开展测序的移动测序解决方案。第三代测序通过检测DNA穿过纳米孔引起的电流变化,将电流信号转换为DNA序列信息。目前前端检测装置已实现便携化设计,可灵活携带至各地开展采样和检测,检测出的原始电信号还需通过AI转译成DNA序列信息。不过由于原始电信号数据量庞大,转译所需计算量大,传输到云端耗时较长。新方案使用安装英伟达计算卡的便携计算设备与检测装置配套,在本地及时开展转译,只将转译完成的DNA序列结果传输到远程云端。未来方案将进一步将计算能力与检测装置集成,使终端直接完成转译计算,或依托安装有GPU的运营商MEC基站,在基站边缘完成DNA序列信息转译。

方案中将电信号转译成DNA序列信息的算法的训练,主要在云端完成,基于英伟达EGX平台搭建训练环境,使用医疗Clara工具包开展算法的训练,并将训练完成的算法下发到便携计算设备,形成云-边-端协同的移动测序解决方案。

应用成效

- **检测及时高效:**可在疫情发生时,及时在发生地完成病原体的实时基因检测,缩短样本运输和数据传输时间,为及时了解病原体信息、制定防控策略、开展防控工作提供基础。同时,避免了病原体运送到城市实验室的潜在泄露风险。
- **助力疾控网络构建:**由于检测设备便携,可在条件有限的地区进行部署,布设并开展检测,如偏远地区或缺乏医疗资源的地市、县等,并及时将检测结果上传,从而助力在全国构建一体化的疾控检测网络体系。
- **快速开展模型训练:**企业基于EGX平台直接从英伟达NGC云下载容器镜像,由测序相关人员利用Clara的相关工具开展模型训练,节约安装部署开发环境的时间,实现分钟级别开箱即用,加速开发效率。同时,测序企业可减少IT运维和算法工程师的配备,可使企业更专注于基因检测专业工作,利用AI模型辅助判断、积累专业经验,加速科研效率。

3.5 零售行业

零售业加速向新零售转型

零售行业与全体国民息息相关，零售业的变革和进步，直接推动人们生活质量的提高，甚至引发生活方式的变化。随着时代发展，消费者诉求的变化和技术进步持续推动零售行业变革以适应新的市场趋势。2016年以来，在消费者数字原生化和电商持续冲击下，新一波“新零售”浪潮正在来袭。

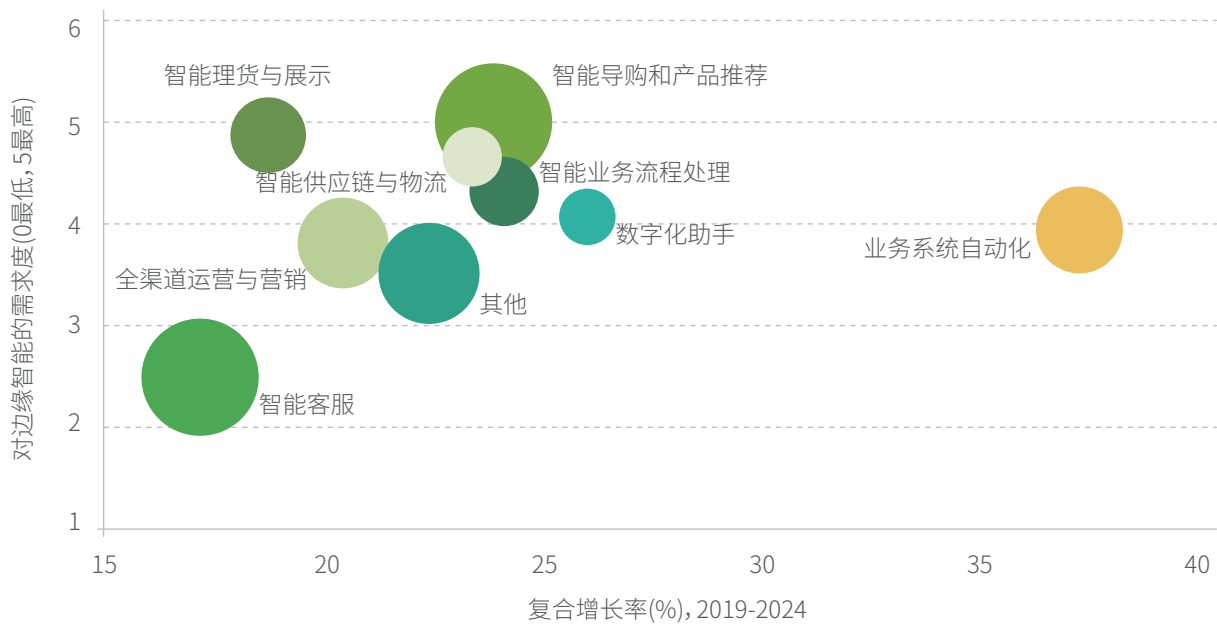
新一波“新零售”以数据为核心驱动力，以新兴技术作为实现手段，以消费者导向作为核心策略，以供应链和生态圈作为主要构成，对“人”、“货”、“场”的关系进行联接和重组。传统零售门店的数字化改造和数字化技术应用带来的新兴业态是新一波“新零售”的两种主要体现形式。

边缘智能助力体验式零售

新零售时代，零售商正通过自助售货、智能互动、全渠道运营等方式，让客户与线上线下的零售商店紧密联系，参与到搜索、评估、挑选商品、支付的全用户旅程，持续地创造无摩擦的、不间断的、沉浸式的“体验式零售”客户体验。AI在其中的客户服务、理货、全渠道营销、客户互动等方面扮演重要作用，催生无人零售等新兴零售业态。

AI在零售业的很多应用发生在线下门店，需要边缘智能的参与。IDC预测，智能理货与展示、智能导购和产品推荐、业务系统自动化等边缘智能典型应用场景，在未来几年将保持高速增长。其中自助结算为代表的业务系统自动化，在未来几年将保持35%以上的高速复合增长。

图15 零售行业边缘智能应用场景



图中气泡大小代表市场规模

来源：IDC，2020

案例:边缘智能助力自助结算

需求背景

为提高结算效率,降低人力成本,目前很多大型商超都已经引入了自助结算台、自助称重台等设施,提高结算效率和顾客体验。在超市出口,顾客可以在自助结算台自行对商品进行扫码,并直接通过APP等完成支付,减少在人工结算台的等待时间。在果蔬区,顾客可以使用自助蔬果秤,选择相应商品标签,完成称重并打印价签,减少排队等待的时间,优化消费者体验。根据2019年IDC对全球400余家零售商的调研显示,28%的被访零售商已经引入了自助结算解决方案。

但实际应用过程中,顾客在自助结算台结账时,可能存在个别商品错扫、漏扫现象,甚至出现通过偷换条码等方式,用廉价商品替代高价商品,减少支付费用,给商超造成经济损失。在果蔬称重时,用户需要在繁多的标签中努力寻找对应的商品,花费时间较长,且可能选错标签,既影响消费者体验,又可能给商超带来经济损失。据统计,零售商超货损可占到整个商超总收入的1.8%。



解决方案

码隆科技是一家专注商品识别解决方案的AI公司。码隆科技基于AI商品识别技术，推出了自助结算台资产保护解决方案和智能果蔬秤解决方案。自助结算台资产保护解决方案通过对自助结算过程的高清视频流进行实时AI分析，自动识别错扫、漏扫现象并及时通知店员，降低货损率，减少商超损失。智能称重解决方案则通过加装一台普通摄像头，自动识别秤上果蔬，供顾客确认，节约顾客寻找标签时间。

码隆科技的解决方案需要在现场独立部署和计算，依赖本地服务器和嵌入式处理器完成AI训练和推理。通过使用英伟达EGX平台的T4服务器、Jetson等计算硬件，借助EGX平台的Tensor RT模型框架、DeepStream视频流处理框架等软件支持，提升部署效率，加快数据处理速度，实现实时、准确响应，并借助Fleet Command对边缘的AI应用进行部署和管理的安全保障。英伟达Fleet Command混合云平台，作为EGX软件堆栈的一部分，可用于管理和扩展AI在边缘的部署。

应用成效

- **商超损失减少**:码隆科技自助结算台防损解决方案已经在超百家中大商超门店实现落地应用，可帮助挽回约占商超总收入0.4%的货损金额。同时，解决方案还能通过可视化数据分析，根据客流量和实际需求指导商超优化业务流程，减少人员需求，节约人员成本。智能果蔬秤可以将称重花费时间减少90%，降低50%以上的货损。
- **开发效率提升**:码隆科技通过使用TensorRT实现模型在GPU上约30%-50%的运行速度提升，节约了算法压缩优化的时间；通过DeepStream提升自行开发视频流处理工作管线的效率；并借助JetPack节省模型在Jetson平台部署的时间，进而在开发、部署、应用等多环节上实现效率的提升。
- **硬件适配简化**:借助EGX广泛的合作生态体系，诸多硬件、虚拟机系统等都实现了对GPU和EGX相关功能的优化和支持，使得基于EGX工具堆栈开发的应用可以实现与硬件设备无缝对接，进而节约应用适配硬件的时间，提升部署的效率。

04

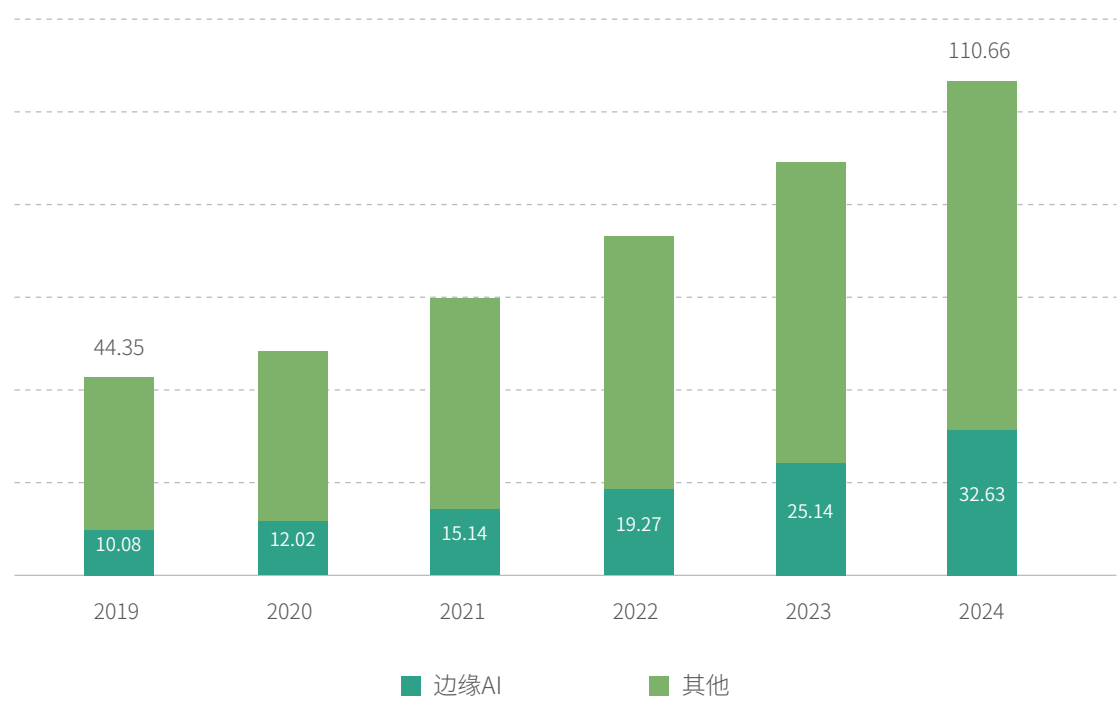
第四章 边缘智能走向未来智能

4.1 未来智能展望

边缘智能市场将持续拓展

随着边缘计算和AI在行业的持续深入应用和互相渗透,越来越多的行业用户将根据不同的业务需求把AI应用引入边缘,边缘智能应用领域日益丰富,应用场景百花齐放,不断释放巨大的市场空间。**IDC预测,边缘计算及边缘智能未来几年仍将快速发展,到2024年,全球AI市场支出将达到1,100亿美元,边缘智能将占到AI市场支出的29.5%。**

图16 IDC全球AI市场支出预测(十亿美元), 2019-2024



来源: IDC, 2020

边缘设备算力的提升,将加速业务向边缘迁移。**IDC预测,到2023年,超过50%的新部署的企业IT基础设施将会在边缘而不是企业数据中心,虽然现在这个比例不到10%;到2023年,70%的企业将在物联网的边缘处开展不同水平的数据处理;2024年,边缘的应用程序数量将增加800%。**

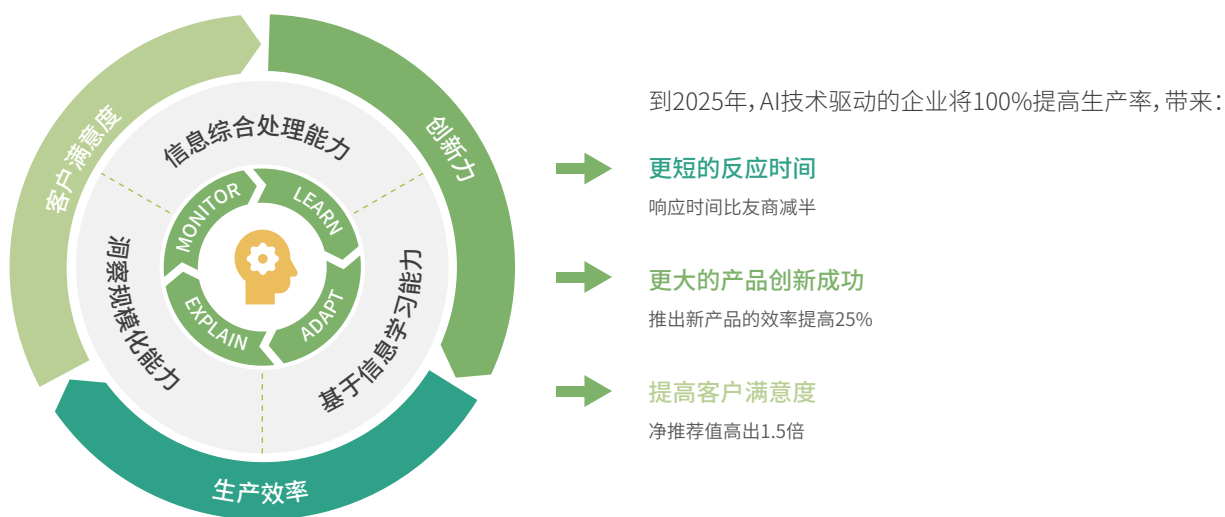
边缘智能开启“未来智能”新时代

未来智能旨在将人和计算机相连,使得情景智能可以渗透到运营、工作和经验中,产生规模化洞察力,实现产品和服务的经济和社会价值的大幅度提升。

企业拥抱未来智能的核心在于持续大规模学习能力,将未来智能与企业运营的方方面面深入结合,并不断改进优化,从而获得更大规模的洞察、创造更大价值,而边缘智能是持续大规模学习能力的关键:

- ▶▶▶ 具备更强的客户洞察能力,从而通过更强的同理心为客户提供个性化的体验与关怀:边缘智能推动未来智能向更加贴近客户的边缘侧渗透,更实时得感知客户需求,捕捉更丰富的行为细节,从而更高效得学习客户行为、更精确得营造同理心、获得更深入的客户洞察,为客户提供更加个性化、更具情感关怀的产品和服务。
- ▶▶▶ 具备更强的弹性扩展能力,从而实现借助高水平的业务流程自动化:边缘智能来在更靠近业务现场的边缘侧感知业务环境、将未来智能前置到更多的业务环节中,从而拓展业务场景学习的规模与实时性,并根据环境的细微变化进行自主实时优化,将业务流程自动化推向新的高度。
- ▶▶▶ 与人更紧密地结合,以增强智能的形式赋予企业员工“超能力”:边缘智能在边缘侧为员工提供“贴身”服务,以未来智能为依托,实时学习员工行为,为员工提供更加个性化、场景化的决策支持,辅助员工更高效得开展工作;发挥边缘智能大规模数据处理和一般流程性任务多路并行执行的优势,实现远超人工处理的效率与精度,使员工专注于复杂决策和创意性任务处理,实现员工能力的超级拓展。

图17构建未来智能能力



来源: IDC, 2020

伴随着企业边缘智能应用深度和广度的持续深化,情景智能渗透到企业方方面面,人工智能或许将逐步迈入“未来智能”时代。

4.2 IDC建议



掌握边缘智能发展趋势。边缘智能不是简单的嵌入新技术，边缘智能将驱动工作模式、工作流程的变革，企业应及时跟进行业发展态势、掌握边缘智能应用空间和发展趋势，充分认知边缘智能的重要价值，制定统一数字化战略，尽快布局边缘智能。



选择合适的场景启动应用。首先定义业务成果以找出可能的用例，根据边缘设备物理限制、数据类型、实时性、计算负载、网络带宽、安全性等方面的要求来定义项目成功标准，之后将它们映射到技术及应用的战略视图上，寻找到适合优先开展的场景，借鉴边缘智能实践案例，开展项目测试。



选择具有全栈技术体系的平台加速边缘智能应用。边缘多样化、碎片化的需求使得边缘智能的应用开发成本和周期受限，选择能够提供硬件、软件、服务全栈产品的厂商能够更大程度简化边缘智能开发和部署的难度，降低不同厂商产品带来的兼容性问题，实现平滑的云到边的迁移和部署，从而将精力和资源更专注在行业本身AI算法的构建和应用上来，快速迭代开发和部署边缘智能应用。



选择具有丰富开发者生态的合作伙伴。边缘侧行业化、场景化的业务需求涉及深入的行业专有知识和经验，通用边缘智能解决方案厂商往往并不具备，需要行业内开发者的协作。完善支持使得开发者更方便得使用通用AI技术和工具，加速边缘智能应用的开发、测试、部署、维护，而且逐步沉淀行业最佳实践，以行业SDK、行业算法库等形式为开发者提供更丰富的支持，进一步加速边缘智能行业落地。



全面布局端到端安全。边缘智能与中心化平台相比，数量大、部署分散，进一步增加了安全防护的难度。因此，边缘智能需要引入完整的主动防御机制，确保海量边缘设备能够有效识别攻击者的硬件篡改和恶意代码注入，进行安全告警；通过算法和算法安全机制保护边缘设备上AI算法和业务数据，避免企业技术机密和客户隐私数据外泄带来经济和商誉损失乃至监管机构处罚和客户法律诉讼风险。

关于 IDC

国际数据公司 (IDC) 是在信息技术、电信行业和消费科技领域, 全球领先的专业市场调查、咨询服务及会展活动提供商。IDC 帮助 IT 专业人士、业务主管和投资机构制定以事实为基础的技术采购决策和业务发展战略。IDC 在全球拥有超过 1100 名分析师, 他们针对 110 多个国家的技术和行业发展机遇和趋势, 提供全球化、区域性和本地化的专业意见。在 IDC 超过 50 年的发展历程中, 众多企业客户借助 IDC 的战略分析实现了其关键业务目标。IDC 是 IDG 旗下子公司, IDG 是全球领先的媒体出版, 会展服务及研究咨询公司。

IDC China

IDC 中国 (北京): 中国北京市东城区北三环东路36号环球贸易中心E座901室

邮编: 100013

+86.10.5889.1666

Twitter: @IDC

idc-community.com

www.idc.com

版权声明

凡是在广告、新闻发布稿或促销材料中使用IDC信息或提及IDC都需要预先获得IDC的书面许可。如需获取许可，请致信 gms@idc.com。翻译或本地化本文档需要IDC额外的许可。获取更多信息请访问www.idc.com，获取更多有关IDC GMS信息，请访问<https://www.idc.com/prodserv/custom-solutions>。

版权所有 2020 IDC。未经许可，不得复制。保留所有权利。