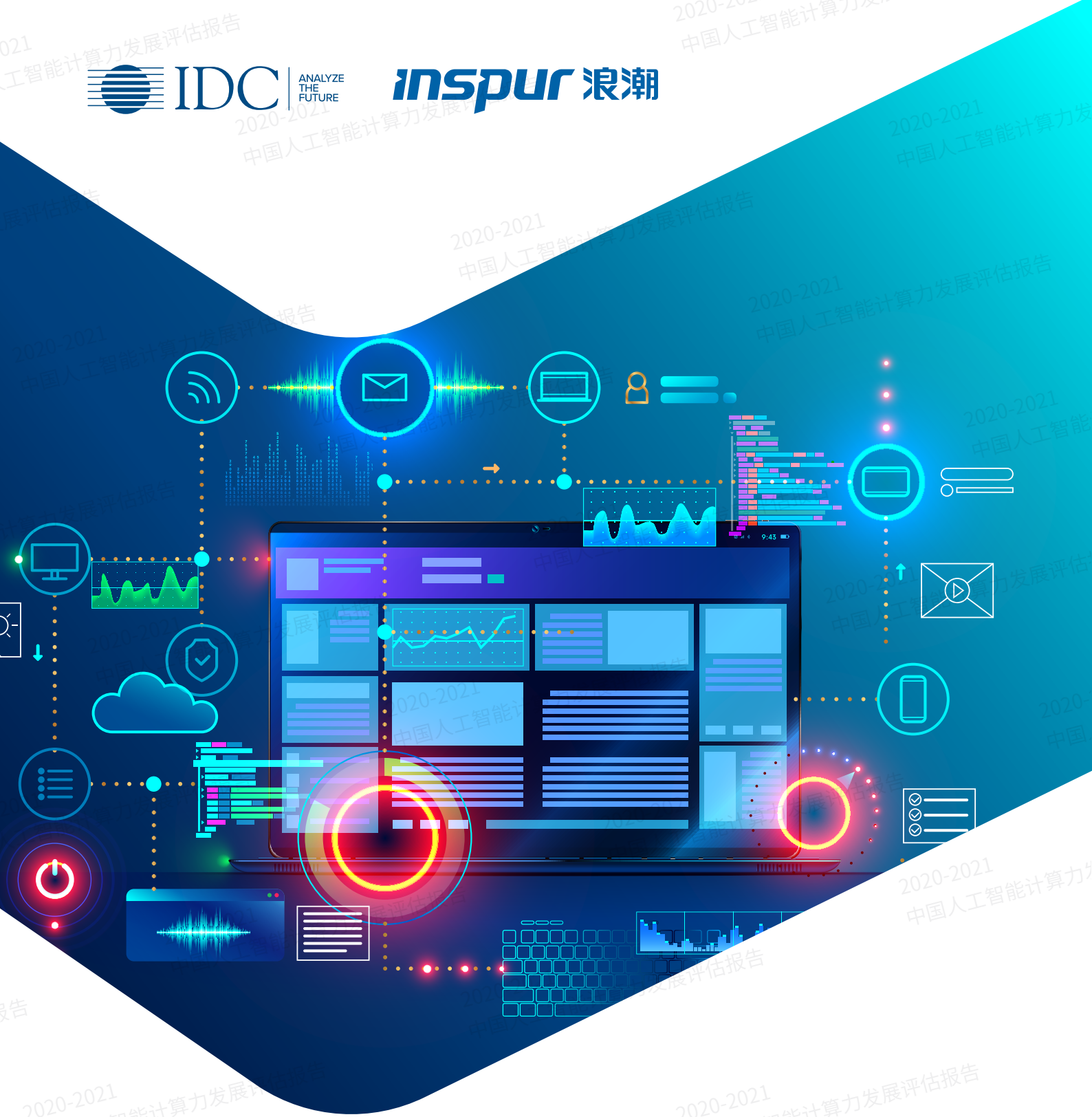
IDC ANALYZE THE FUTURE

inspur 浪潮



2020-2021
中国人工智能计算力发展评估报告

目录

一、IDC观点 3

二、前言 5

2.1 全球疫情的影响及AI抗击疫情的作用

2.2 新基建对中国数字经济的推动作用

2.3 人工智能高速发展对算力提出更高需求

三、中国人工智能算力基础架构发展现状及未来展望 9

3.1 人工智能算力基础架构的发展

3.2 人工智能应用的成熟度整体提升，场景开始落地

3.3 中国人工智能算力市场展望

四、中国人工智能算力发展评估与城市排名 20

4.1 人工智能行业应用发展评估

4.2 城市人工智能算力排行

五、全球人工智能算力发展评估 30

六、行动建议 32

6.1 对行业用户的建议

6.2 对技术供应商的建议

6.3 对产业发展的建议

IDC观点

- **新冠疫情对全球经济造成巨大影响，数字化转型的必要性已经凸显，新基建成中国数字经济化转型的主要驱动力。新基建**多个领域与ICT密切相关，2020年新基建整体投资规模预计将达到2,757.1亿美元，人工智能作为新基建的重要领域之一，将迎来快速增长。**IDC预计，2020年中国人工智能市场规模将达到62.7亿美元，2019-2024年的年复合增长率为30.4%。**

- **人工智能三要素中，算力成为驱动AI产业化和产业AI化发展的关键要素。**人工智能应用对算力最大的挑战依然来自于核心数据中心的模型训练，近年来，算法模型的复杂度呈现指数级增长趋势，正在不断逼近算力的上限。

- 人工智能芯片产业仍处在成长期，预计**2020年，中国的GPU服务器依然占据95%左右的市场份额，是数据中心人工智能加速方案的首选。**未来几年，随着推理工作负载在各个行业应用中不断增加，FPGA和ASIC等其他类型的加速芯片将在各个领域采用。**预计到2024年，其他类型加速芯片的市场份额将快速发展，人工智能芯片市场呈现多元化发展趋势。**

- **服务器是人工智能基础设施的核心，到2020年，中国人工智能基础设施市场规模将达到39.3亿美元，**其中服务器支出占比高达87%。人工智能服务器技术不断创新，包含多种互联方式和拓扑架构，以满足不同的应用场景需求，与此同时，人工智能基础设施正在向开放架构发展，以满足高效、灵活、可扩展的下一代人工智能数据中心的需求。

- **日益增长的业务实时性需求使边缘和端侧的计算能力变得越来越重要，IDC预测，到2023年，接近20%的用于处理人工智能工作负载的服务器将部署在边缘。**2020年是边缘计算广泛落地的元年，人工智能算力也会逐渐向边缘渗透，无论是更接近于端侧数据产生的轻边缘还是更接近核心数据中心的重边缘，都将迎来很大的发展契机。

- **产业AI化已经从早期的试点逐渐成为企业发展和生存的刚需。**过去一年，人工智能在行业的应用获得快速发展，通用型应用场景已经具有相当的成熟度，在业务需求的推动下，具有高行业属性的碎片化应用开始被广泛使用，并辐射到媒体娱乐、现代农业、智能家居、智慧电力等多个不同领域。**与2019年相比，疫情推动下的智慧医疗以及疫情常态下园区、办公楼宇、社区的生物识别类应用比预计发展得更快；5G、物联网和边缘计算在未来也将加速多个行业应用，尤其是制造和能源企业对于人工智能的采用。**

- **企业普遍希望采用具有公共基础设施意义的人工智能算力基础设施。**IDC调研发现，超过九成的企业正在使用或计划在三年内使用人工智能，其中74.5%的企业期望在未来可以采用具备公用设施意义的人工智能新型基础设施。**企业对于人工智能公共算力基础设施的TOP5需求分别是：用于人工智能训练的数据支撑、人工智能加速计算能力、配套的政策吸引、规模效应下的价格和成本因素，以及丰富的应用场景配置。**

- **2020年中国城市人工智能算力排行榜Top10排名变化不大，具有产业优势的人工智能应用的城市涌现。**2020年中国城市人工智能算力排行榜中排名前五的城市依次为北京、深圳、杭州、上海、重庆，排名6-10的城市为广州、合肥、苏州、西安、南京。**与2019年相比，深圳超过杭州位居第二，重庆进入第一梯队，西安超过南京位居第九。**除了TOP10城市之外，在自身产业优势及各种因素推动下，多个城市的人工智能应用取得了较大进展，例如东莞的智能制造、武汉的智慧医疗、合肥的智慧农业等，中国人工智能城市发展正遍地开花，未来将会出现越来越多结合城市特点的人工智能示范区，为产业发展树立标杆。

- **人工智能是全球IT产业发展最快的新兴技术应用之一，未来几年全球的人工智能发展潜力巨大。从2020年各个国家AI计算的发展水平来看，中国人工智能服务器占全球人工智能服务器市场的三分之一左右，是全球人工智能产业发展的中坚力量。**人工智能服务器的投资能够为人工智能应用的落地提供核心的算力保障，对于未来国家计算力指数的提高具有极大的推动作用。未来，将人工智能等新兴技术作为国家级的战略方针将有助于各国进一步提高计算力指数和经济发展的相互拉动力。



前言

2020年是人工智能产业发展“危”、”机”并存的一年，新冠疫情在短期内重挫全球经济，同时，未来几年疫情的不确定性也使企业开始重新思考如何更好地应对未来的挑战，越来越多的企业认识到数字化转型的重要性，人工智能、物联网等新兴技术已成为加速企业数字化创新的核心力量。

2.1全球疫情的影响及AI抗击疫情的作用

疫情的爆发使全球经济和IT宏观环境面临严峻压力和挑战。新冠疫情导致绝大部分的世界经济体采取了不同程度的封锁措施，对国际贸易、住宿与餐饮、旅游、批发零售、交通、制造以及中小企业等产生巨大的冲击，生产生活等都受到了很大影响，根据经济学人（TheEconomist）在2020年9月作出的最新预测，2020年全球GDP增速将下滑5.2%。受疫情的影响，中国乃至全球经济和ICT市场也受到巨大冲击。IDC预计，2020年全球ICT市场规模将同比下滑2.5%，随着企业适应新常态，2021年ICT支出将逐步恢复正常化。

短期来看，疫情对ICT市场的影响体现在三个方面：

削减采购预算：受延迟复工、供应链影响，中小规模企业无法按时生产和交付，不能现场拜访用户，进而导致销售收入下降，企业整体收入和利润都明显下滑，必然缩减支出，其中也包括IT支出。

推迟采购预算：由于延迟复工，一些企业以及政府与公共事业部门用户，推迟技术采购计划，采购计划的推迟短期内会对ICT市场造成不小的影响。

加速采购预算：对于医疗、教育、政府等公共事业，如涉及到疫情联防联控的公共部门，需要紧急公布及处理疫情，教育单位也需要通过远程教育的方式进行正常教学，这些机构迫切需要加速采购流程。互联网和云服务相关的需求也在短期内集中爆发，一些公司加速采购，如远程办公、协同办公、在线教育、电商、直播等业态在疫情期间呈现快速增长态势。

长期来看，一方面疫情防控的不确定性使未来几年全球经济增长也面临很大的不确定性，而“蝴蝶效应”还将会在多个方面在对未来全球的经济产生持续的影响。另一方面，疫情也使全球各行各业更深刻地认识到数字化转型的重要性和必要性，边缘计算、人工智能、云和数字服务、5G等技术在行业的应用都将加速并推动全球企业的数字化转型进程。

根据IDC对行业的调研，数字化转型领先的企业抗疫能力明显更强，工作资源转型领先的企业远程协同能力更好，整体工作效率更高。新冠肺炎疫情后，非接触连接商业与服务将加速兴起，线上和线下深度融合，“云+AI+5G/IoT”成为推动这一趋势的引擎，将加速推进中国及全球企业全方位数字化转型。

调整与加速实施数字化转型战略是企业疫后实现业务复苏、增长和创新的基石，其中算力是根本。

在新冠肺炎疫情肆虐全球期间，算力体现了巨大的社会价值，在算力的保障下，政府和企业不仅能及时获取疫情相关信息、开展远程协同工作、保证业务持续不停摆，还能有效打造数字化文化。

全球各国政府和企业将AI能力、大数据追踪判断能力、机器人服务应用到疫情防控的各个环节，这些数字化技术、数据智能的应用在疫情防控方面展现了显著的价值。在公共卫生防线、医疗诊治防线、重要场所防线和社区治理防线中，人工智能被应用在了常时统计报送、疫情监测分析、互联网诊疗、病毒基因检测、药物研发筛选、肺部影像分析、机器人、红外体温检测、健康码、智能语音操控电梯、智能外呼/问答机器人等多个应用场景，助力疫情防控并取得显著成效。

2.2 新基建对中国数字经济的推动作用

推动政府提出加快新基建的投资，这将促进中国宏观经济的进一步发展。2020年上半年，中央政治局、国务院、国家发改委等中央部门多次在各种会议上提出部署新基建项目，各级地方政府接下来也会陆续出台大量有关新基建的鼓励政策和项目规划，以推动中国经济增长和产业升级。

新基建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ICT密切相关，包括5G基站与网络、人工智能、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。

5G基站与网络：5G是面向行业应用的新一代移动通信网络，5G的场景化和确定性的网络特征，将进一步加速企业的数字化转型，从而推动企业对ICT产品和服务的需求。此外，5G采用云原生的基础架构，一方面会驱动运营商在网络建设方面更多采用通用IT硬件基础设施；另一方面，5G将加速数据中心下沉到向更加贴近用户的网络边缘，为用户带来更好的网络体验，为企业带来更高的网络性能。IDC预计2020年5G用户数将达到1.05亿，5G基站总数量将达到72万个，5G网络基础设施投资总额将超过170亿美元。

人工智能：IDC将人工智能定义为具备学习、推理和自我纠正能力的系统。系统通过自然语言、语音、图像、视频等方式与人类交互，从交互信息中抽取知识建立知识库，并采用机器学习方式建立预测模型，基于模型进行推理给出结果。数字经济正在进入以人工智能为核心驱动力的“智能经济”阶段，人工智能将驱动IT基础设施变革、人机交互方式变革以及未来企业的数字化变革，智能在未来将无处不在。IDC预计，2020年中国人工智能市场规模将达到约62.7亿美元，未来五年的年复合增长率为30.4%。从技术领域来看，计算机视觉仍是应用规模最大的领域，自然语言处理、对话式客服、语音分析以及机器学习市场也将保持高速增长。人工智能将会推动加速计算和异构计算的发展，IDC预计，到2023年，在人工智能等创新应用的驱动下，30%的企业将采用加速计算和异构计算等基础架构，以加快其业务及服务的速度、规模和智能化水平；到2023年，30%的核心企业数据中心和50%的主要边缘IT站点将采用机器学习和人工智能来改变运维方式，并提高能源以及资源的有效利用。随着人工智能在互联网、政府、金融、制造、电信、医疗、交通等行业应用的深入，人工智能也将在中国企业的数字化转型进程中扮演至关重要的角色。IDC预计，到2022年，凭借主动的超高速运营变革和市场响应，中国具备人工智能（AI）实力的企业对客户、竞争对手、监管机构和合作伙伴的响应将比同行快至少30%。

大数据中心：大数据中心作为大数据产业的一部分，更加聚焦于大数据硬件基础设施，也就是数据中心，将直接带动数据中心基础设施、软件和服务的建设。IDC预计，2020年中国数据中心市场投资规模约402亿美元，其中2020年数据中心IT基础设施市场规模约287亿美元，未来五年的年复合增长率约11.0%；2020年数据中心软件市场规模约49亿美元，未来五年的年复合增长率约14.0%；2020年数据中心服务市场规模约66亿美元，未来五年的年复合增长率约11.0%。

工业互联网：工业互联网是面向工业智能化发展的关键基础设施，是新一代信息技术和工业机理深度融合形成的IT应用模式。工业互联网新型体系架构关键要素包括网络、平台、安全、工业智能、应用等，通过与传统工业体系架构之间相互作用、深度融合，构成了新型工业体系架构。建设并完善这一体系，既要从顶层实现工业资源的统筹管理和调配，也对生产现场实时数据的采集分析与反馈控制提出了更高要求。随着智能生产、个性定制、远程运维等工业互联网应用场景的日益完善，工业边缘数据中心将迎来新的发展机遇。IDC预计，2020年中国工业互联网市场规模约800亿美元，未来三年的年复合增长率约 9%。

未来几年，新基建将成为关乎经济提振、产业转型升级的关键，不仅将刺激中国经济增长并带动 ICT市场商机，还将对企业数字化转型起到巨大推动作用。

IDC预计，新基建在未来几年直接带动的ICT商机将超过 1,400亿美元，而人工智能作为“新基建”的重要领域，被中央和地方政府高度重视，政府相继出台有关政策指导人工智能产业发展，产业化AI也将迎来新的发展契机。

2.3人工智能高速发展对算力提出更高需求

数据 - 数据量增长的步伐仍在加快，未来三年新产生的数据量预计将超过过去三十年的总和，到2024年，全球数据总量将从现在的44.9ZB，以26.0%的复合增长率达到142.6ZB。其中，中国在2017年已经成为全球数据量规模最大的国家，并将保持稳定的高增速，预计到2024年，中国新产生的数据量将达到36ZB，占全球总量的25.2%。数据为人工智能应用提供了丰富的土壤，新的应用场景在各行业不断涌现，其中，越来越多的场景对数据实时性提出了高要求，IDC预测，实时数据的占比将在未来五年不断增加，到2024年将达到数据产生总量的四分之一左右。实时数据的激增促进了边缘基础设施的发展，也使得边缘计算能力变得越来越重要，人工智能应用也将更加依赖于处在边缘的算力支撑。

数据为人工智能应用提供了丰富的土壤，同时也不断引发隐私和伦理上的争议，尤其当涉及到生物识别的使用时，例如智慧社区在实施过程中就频遇阻力，这类影响对于未涉及到个人隐私的应用则相对较小，例如制造业的智慧工厂、零售业的商品稽核等。目前，人工智能使用规范相关的政策法规依然处在早期的建立阶段，随着技术的进步和法规的完善，数据在未来将会发挥更大的价值。

算法 - 虽然数据的总量正在增长，但真正被有效利用的仍然不足1%。如何有效抓取高质量的数据并通过这些数据建立精准的模型，则取决于人工智能算法的开发能力。虽然近年来，新增的算法模型数量在逐年减少，但是单个模型的参数量和复杂程度都呈现指数级的增长趋势，例如Bert及其变种的算法模型，这些复杂的算法模型将极大地推动人工智能在行业应用的发展。到2020年，深度学习模型对算力的需求达到了每天百亿亿次的计算需求，特别是在自然语言处理等新兴认知智能应用领域对计算力的要求远超过图像识别、语音识别。

在算法模型创新的推动下，面向应用场景优化的各种模型不断出现，促进了中国机器学习开发平台市场的增长，其中使用自动化数据驱动方式的AutoML是最值得关注的前沿领域之一，也具有较为广泛的市场空间，尤其在还没有成熟AI软件产品的领域。在数据资源就绪度不足、数据科学家资源稀缺的现状下，市场需要更成熟的机器学习平台产品、更低门槛的AutoML产品、乃至更有效的落地方法，助力企业AI全面落地应用，提升业务价值。

2020年2月，微软发布了最新的智能感知计算模型Turing-NLG，参数量高达到175亿，是Google智能感知模型的50倍，使用125POPS AI计算能力完成单次训练就需要一天以上。随后，OpenAI又提出了GPT-3模型，参数量更达到1750亿，是微软Turing-NLG的十倍、GPT-2的100余倍，使得NLP模型参数体量达到了全新的高度。

算力-在人工智能三要素中，无论是数据还是算法，都离不开算力的支撑。IDC对用户的追踪调研发现，过去，用户对于人工智能的感知更多停留在数据层和应用层，随着非结构化数据的激增和算法框架的日益复杂，算力不足的问题逐渐显现出来，越来越多的用户认识到算力的重要性。尽管实时数据更多依赖于边缘和端侧的推理计算能力，但具有海量参数的模型训练仍完全依赖于核心数据中心的算力支撑，包括各类预训练模型，由于其庞大的参数体量，在给人工智能应用提供便利的同时，如果没有强大的算力支撑，很难发挥其价值，因此，人工智能的算法越来越依赖于算力的发展，如何布局算力是企业在未来很长时间内都需要关注的重点。



中国人工智能算力基础架构发展现状及未来展望

3.1 人工智能算力基础架构的发展

算力作为人工智能产业化的核心力量，其发展将对人工智能技术的进步和行业应用起到决定性的作用。本报告中IDC依然从算力基础架构层面，重点研究了人工智能芯片、服务器和边缘计算几个方面的变化，下文将就此进行逐一分析：

3.1.1 人工智能芯片呈现多样性发展趋势

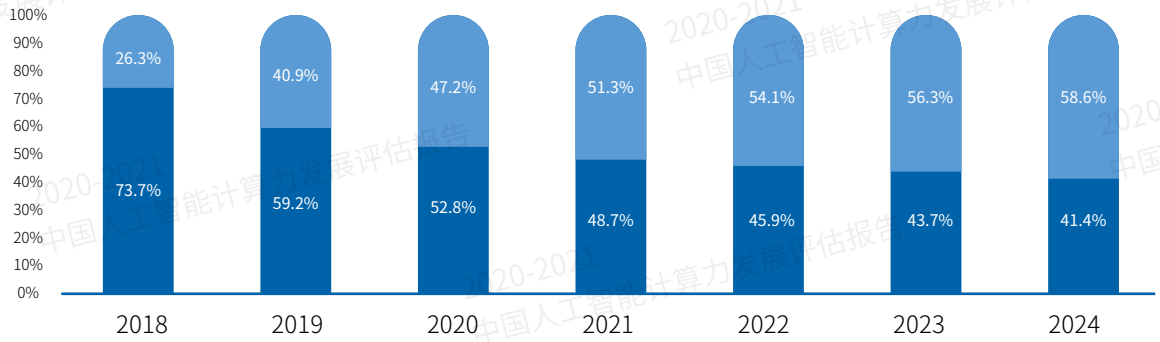
人工智能芯片仍处在高速发展的时期，面对不断增长的市场需求，各类专门针对人工智能应用的设计理念和创新架构将不断涌现。为了支持多样的 AI计算任务和性能要求，理想的AI芯片需要具备高度并行的处理能力、低内存延迟和创新的架构，以实现计算元件和内存之间灵活而丰富的连接，除此之外，还要考虑功耗和能效管理水平。在当前人工智能各领域的算法和应用还处在高速发展和快速迭代的阶段，针对特定领域而不针对特定应用的设计，将是AI芯片设计的一个指导原则，具有可重构能力的AI芯片可以在更多应用中广泛使用，并且可以通过重新配置适应新的AI算法、架构和任务。

目前全球人工智能产业在高速变化发展中，广泛的行业分布为人工智能的应用提供了市场前景，快速迭代的算法推动人工智能技术快速走向商用，算法的实现依赖于算力的基础——芯片，针对不同的计算任务，不同类型的芯片往往各具优势。

IDC的研究发现，预计2020年，中国的GPU服务器依然占据95%左右的市场份额，是数据中心人工智能加速方案的首选；而FPGA和ASIC等其他非GPU加速芯片将在各个行业和领域被越来越多的采用，尤其是部署在边缘和端侧的推理工作负载，以满足成本与功耗的需求。

IDC预测，用于推理工作负载的服务器市场份额将在不久的将来超过训练，并在预测期内保持这一趋势。现在的人工智能领域已经出现了多种类型的加速芯片，主要包括图形处理器（graphics processing unit，GPU）、现场可编程门阵列（field-programmable gate array，FPGA）、专用集成电路（application specific integrated circuits，ASIC）、神经拟态芯片等，未来，人工智能芯片市场会继续向灵活性和多样化的方向发展。

人工智能服务器推理和训练工作负载预测，2018-2024



数据来源：IDC，2020

3.1.2 人工智能服务器未来五年继续保持高速增长

中国人工智能基础设施市场规模在2020年达到39.3亿美元，同比增长26.8%，并将在2024年达到78.0亿美元，其中，2020年服务器市场规模占整体人工智能基础设施市场的87%以上，承担着最为重要的角色。人工智能服务器通常搭载GPU、FPGA、ASIC等加速芯片，利用CPU与加速芯片的组合可以满足高吞吐量互联的需求，为自然语言处理、计算机视觉、机器学习等人工智能应用场景提供强大的算力支持。

2020年上半年，从厂商销售额角度看，浪潮、华为、曙光排在前三位，占据了80%以上的市场份额，其中浪潮市场份额超过50%；从行业的角度看，互联网依然是最大的采购行业，占整体加速计算服务器市场近60%的份额，同比2019年上半年超100%的增长；政府行业和服务业分别依次位居第二位和第三位；从加速卡的角度看，2020年上半年的采购主要集中在搭载V100和T4的加速服务器上，其总和占整体市场的80%以上；从加速技术角度看，除了GPU以外，其它非GPU加速技术也在逐渐成长起来，包括FPGA，ASIC/NPU等，近年来发布的人工智能服务器和芯片种类越来越多，将支持更多的行业应用场景落地。

随着人工智能应用计算需求的指数级增长，在 GPU 之间以及CPU与GPU之间形成的芯片互联技术被更多地采用。虽然PCIe有着非常通用的标准化设计，但带宽有限将会产生瓶颈，人工智能芯片厂商NVIDIA、Intel、AMD等在芯片互联技术方面持续升级，以提供速度更快和扩展性更强的互联。

NVIDIA NVLink 是一种高速、直接的 GPU 到 GPU 互联技术，通过为多 GPU 系统配置提供更高的带宽、更多的连接和改进的可扩展性解决互连问题，NVIDIA 最新发布的 A100 GPU 单个最多支持 12 个第三代 NVLink 连接，总带宽为每秒 600 GB，是 PCIe 4.0 带宽的 10 倍。NVIDIA NVSwitch 将多个 NVLink 加以整合，在单个节点内以 NVLink 的较高速度实现多对多的 GPU 通信；AMD 的 Infinity Fabric 总线技术，目前支持 CPU-CPU 以及 GPU-GPU 之间的连接，下一代 Infinity Fabric 技术将最多支持 8 个 GPU 芯片的连接，并支持 CPU-GPU 间的连接，为加速计算提供动力；Intel 主推的 CXL 互联协议是基于 PCIe 5.0 之上的一个新的规范，CXL 在 CPU 和工作负载加速器（如 GPU、FPGA 和网络）之间创建了高速、低延迟的互连性，使设备之间实现内存一致性，允许资源共享，同时可以简化相关硬件的设计难度，降低整套系统的成本。在 PCIe 5.0 高带宽时代，人工智能通用 CPU 和加速芯片间的协同、互联以及内存共享不仅将成为主流趋势，还将推进人工智能服务器的性能扩展。

除了创新的互联方式之外，人工智能服务器另一个普遍采用的技术是多种拓扑架构的算法调优，以更好地适配多样的人工智能应用。例如浪潮部分型号的人工智能服务器在同一设备中可支持三种不同的拓扑架构，并可通过线缆接口快速切换，在不同的算法模型下实现更加优化的性能输出。随着人工智能应用的不断深化，如何最大限度地提高能效，确保人工智能产业的可持续发展也变得愈加重要。目前，越来越多新的技术被应用在IT基础架构中，包括模块化数据中心、液体冷却、风墙等，同时，采用超融合、异构计算（GPU、FPGA、ASIC）等新的产品和部署模式也能够极大提高IT系统的使用率和计算效率，从而降低服务器和数据中心的整体能耗。

在开源开放计算的大趋势下，人工智能基础设施也在向开放架构发展，各大主流服务器供应商和加速芯片供应商作为开放架构的发起者及参与者，积极为开源社区贡献自己的能力，以实现快速大规模地部署更加高效、灵活、可扩展的下一代人工智能数据中心。同时，在开放计算联盟中，加速计算的各大主流服务器供应商和加速芯片供应商与开源社区合作，在框架、库、编程模型、互联等方面贡献合作的方式，以实现更高的性能和可扩展性。

OCP加速器模块（OAM）是OCP基金会开放式加速器基础设施项目的一个子项目。OAM为机器学习、深度学习和加速计算解决方案中使用的硬件加速器提供通用规格设计。OAM还提供了通用于处理不同类型的加速器（如GPU、FPGA或ASICs）的产品和接口。基于该产品，系统运营商可以轻松地开发、部署和维护不同的加速器，从而节省时间和费用。包括脸书、微软、百度、谷歌、浪潮、阿里巴巴、腾讯在内的超大型运营商和计算平台提供商都参与了OCP加速器模块项目。

3.1.3 人工智能算力逐步向边缘渗透

在物联网技术的发展和企业业务需求的推动下，边缘计算将迎来快速落地的一年，例如在工业互联网中，对采集数据的准确性以及实时性要求会不断增加，对智能数据采集系统的需求也会越来越大。这些场景需要数据的采集系统能够形成多维、立体的图像和视频采集网络，使用机器视觉技术保证在实时监测的高峰期及时采集、处理及分析大量的图像视频数据，并全面整合与共享图像视频数据资源，实现整个系统智能高效地运行，确保后续的处理系统对图像和视频数据的处理、决策、分析以及信息服务能够全面、准确、及时，因此会需要更加先进的IT技术和基础设施来对数据进行采集和利用。**为了更加实时地分析和处理来自终端设备的海量数据并减轻网络传输的压力，人工智能算力会逐渐向边缘渗透，无论是更接近于端侧数据产生的轻边缘还是更接近核心数据中心的重边缘，都将迎来极大的发展契机。**

虽然边缘计算在发展过程中遇到了诸多阻力，例如边缘计算中心与不同的数据采集侧通讯协议的不统一；边缘应用场景碎片化，无法形成规模化的场景复制；用户的短期投资回报率低等问题，但边缘计算带来的响应速度提升，高可靠性，甚至是安全性使之成为必然的发展趋势。未来，随着物联网用例的增长、人工智能和机器学习解锁边缘数据，以及服务提供商增加对边缘计算领域的投资，都会驱动未来边缘计算市场快速发展。

物联网是边缘计算发展的重要推动力，根据IDC关于物联网的预测，全球物联网花费将保持两位数的增长，其中数据采集模块的支出占总物联网花费的比例为28.3%，而智能摄像系统又是数据采集模块中非常重要的一部分。随着机器视觉技术以及大量AI视频图像分析应用的兴起，越来越多的场景将会采用智能摄像系统，到2025年，全球视频监控摄像头的市场规模将达到440亿美元，年复合增长率为13.0%，机器视觉也将更贴近地服务于我们的生活。如智慧城市中，政府可能会加大环境保护及民生服务的智慧服务投入，而采用机器视觉技术和AI分析的智能视频图像采集系统除了能够快速获取大量视频图像信息之外，还能保证实体数据的安全，能够更好地服务于这些场景。这些智能的视频图像系统将会产生大量的数据，尤其是非结构化数据。IDC认为，未来五年非结构化数据在整体数据圈的占比将持续增加，对于人工智能的算力需求也会越来越大。

IDC预测，到2023年，将近20%用于处理AI工作负载的服务器将使用AI优化的处理器和辅助处理器，并部署在边缘侧；70%的企业将在物联网边缘运行不同级别的数据处理。与此同时，各组织将在物联网边缘基础设施上投入超过160亿美元；到2025年，将有50%的计算机视觉和语音识别模型将在边缘端（包括端点）运行。

3.1.4 AI云服务（AIaaS）是人工智能市场发展的重要驱动力

过去一年，在云服务厂商积极投入、多云混合云技术逐渐成熟等因素的推动下，中国AI云服务市场保持了高速的增长，2019年中国AI云服务市场规模达1.66亿美元，2018至2024年的年复合增长率预计将达到93.6%，AI能力已成为用户进行云服务选型时的重要考量因素。

在当前主流的人工智能用例中，采用AI公有云服务的应用场景主要包括：互联网行业的内容审核、推荐系统，制造业供应链预测分析，金融行业中小型企业的智能客服，零售行业智能营销、产品推荐、对话式客服，客流分析等等。从产品数量来看，当前领先的云服务厂商大都在公有云上提供了丰富的AI能力。从调用量以及营收来看，厂商在不同的细分技术领域会呈现一定的优势。百度智能云作为AIaaS的公有云代表，不管是智能音箱、家居等消费级产品市场，还是其他

企业级市场，百度智能云都建立了广泛的客户基础；其中，在自然语言处理领域，百度智能云是目前实现一定规模的商业化营收的唯一厂商。

考虑到部分行业的核心业务场景如金融行业的核心交易系统、医疗行业的核心医疗系统，对于安全、合规以及监管方面有特殊要求，因此在部署和使用AI时，更多采用私有化的方式，倾向于本地数据中心。浪潮AIStation作为私有化AI 算力调度的代表，可全面支持AI训练与AI推理两大场景，实现从模型开发、训练、部署、测试、发布、服务的全流程一站式高效交付。在AI训练环节，AIStation通过计算资源细粒度调度、训练数据缓存加速、分布式训练自动编排等策略，极大地提高用户的AI计算资源利用率，大幅缩短了模型开发周期。而在AI推理环节，则可实现推理服务资源敏捷调配，支持多源模型统一调度，使模型部署从几天缩短到几分钟，将有力地帮助企业轻松部署AI推理服务，从而大幅提升AI交付和生产效率，解决行业用户在私有化AI算力方面的需求。

随着AI开始渗透到企业的核心业务，应用场景会越来越分散，越发需要技术和应用上的不断创新和突破。云服务厂商有深厚的客户资源积累，强大且可持续的研发投入，以及云服务生态体系内大量的合作伙伴积累，将确保该类厂商在未来AI市场的竞争优势。AI云服务市场本身的竞争，2020年之后会更加激烈，尤其是第一梯队的厂商之间的竞争。未来的AI云服务竞争，不仅是技术的领先性、产品的丰富程度的竞争，更重要的是周边开发工具、合作伙伴生态等层面的系统性竞争。

3.1.5 AI基准测试逐步完善

相比2019年，人工智能性能基准测试已逐步完善，MLPerf开始引领通用型基础测试标准，具有行业属性的测试标准还需长期验证。MLPerf作为最具影响力的人工智能基准测试，可以衡量各类应用场景中的机器学习性能，凭借业内大量领先企业的参与（包括Intel、NVIDIA、Google、百度、浪潮、哈佛、斯坦福等），虽然从技术上讲仍不够完整，例如缺乏功耗测试，相信在未来几年会更趋成熟和稳定，最终为企业输出极具参考价值的性能基准测试结果。

与MLPerf不同，SPEC于2019年设立的SPEC Machine Learning技术委员会在一定程度上弥补了人工智能基准测试的单一性，该委员会由发起者浪潮担任首任主席，Intel担任秘书长，成员包括浪潮、Intel、AMD、ARM、HPE、IBM、Alibaba等12家企业，除了多个维度细颗粒度的性能测试之外，SPEC ML还关注其他参数，比如功耗，成本效益等，人工智能应用是未来的能耗大户，因此能效水平应该成为人工智能应用最重要的指标之一，不仅仅为企业的成本效益提供参考，也是人工智能应用未来可持续发展的重要因素。

2020年中关村智用研究院、百度、浪潮联合发起中国推出首个面向产业应用的AI开源评测基准—AI-Rank，通过建立开源、开放的AI基准测试体系，结合人工智能产业发展需求，对AI软硬件基础设施环境进行可定量、可复现、可对比的全方位测评，客观、全面、系统的反映人工智能产业的现状及其发展方向，助力全产业AI相关企业整体应用和技术成熟度提升。在具体的评测体方面，AI-Rank重视不同深度学习框架对AI训练任务的支持能力，并设置了多维度指标对框架能力进行综合评测。除此之外还对不同框架在各个领域所使用的深度学习任务支持能力进行评价。 AI-Rank指定了多个热门领域中的10+个有影响力的模型，并从框架性能、任务覆盖率操作系统支持几个维度对框架能力进行综合评价，旨在为产业研发和应用提供更为权威的、更为实用和贴近应用的真实数据，为AI在不同的产业提供权威的参考应用。

3.2 人工智能应用的成熟度整体提升，场景开始落地

过去一年，人工智能应用获得了快速发展，**从应用类型来看，计算机视觉目前仍然是中国最大的应用市场，也是过去一年增长最快的应用类型，2019年计算机视觉市场规模占整体市场的40.6%，**其次是对话式客服、自然语言处理和语音分析，这三者占比分别达到20.5%，17.8%和11.7%，随着自然语言处理算法模型的快速发展，预计未来语音语义市场规模也会进一步快速增长。第三个细分市场是机器学习开发平台，目前该细分市场的规模相对较小，占比不到10%，虽然落地的应用规模不大，但企业也开始加速探索，采用深度学习来提高预测的准确性。比如除了反欺诈，产品推荐这些已经落地的应用场景，企业在核心生产环境开始测试机器学习应用，比如说工业设备检测，配电站设备障碍评价等。

从行业角度来看，同2019年相比，电信和制造行业的应用场景更加丰富，市场潜力预计将有较大的提升。



电信运营商利用其在数据和算力上的优势，正在积极部署公有云上的人工智能服务，加上自身人工智能应用场景上的需求迅速增长，使得电信行业成为在人工智能投入上增长最快的行业之一。其中智能网络可以自动识别网络故障和潜在的网络攻击，也可以进行网络资源的优化，是电信行业潜力最大的应用场景；运营商也在着手智慧营业厅的建设，利用人工智能技术对消费者行为数据进行分析和指导。



制造业在过去几年对人工智能的投入增长相对缓慢，主要原因在于很多制造企业对投资回报率的把控严格，即使大部分制造企业认可人工智能所能带来的长期收益，但往往又妥协于企业的短期投资回报率。**未来，随着边缘计算的发展和物联网带来的数据累积的增加，人工智能得以更多地渗透在制造企业的生产和设计流程中，制造企业中的人工智能的应用场景也会更加丰富，更加成熟，**包括产品分拣、QC自动化、设备维护自动化、物流和供应链管理自动化等， IDC认为这些用例会促进越来越多的制造企业采用人工智能，因此，制造业未来具有较大的人工智能发展潜力，更多的制造企业将利用人工智能提高其自动化水平以及核心竞争力。

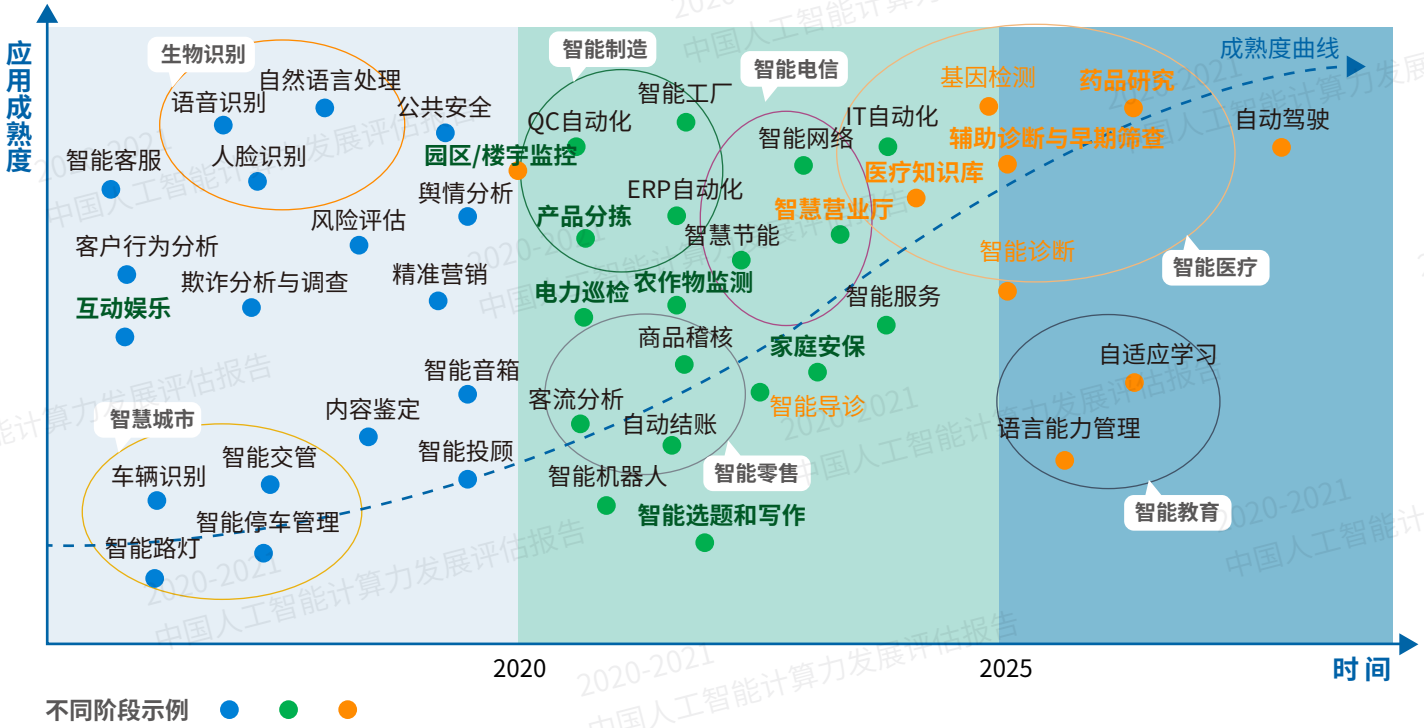


在全球疫情爆发后，在**医疗行业中，人工智能被应用于智能问诊、影响辅助诊断、药物研发及筛选、诊疗方案设计等环节，而智能导诊、辅助诊断与早期筛查、药品研究、医疗知识库和基因检测等多个应用场景也比预期发展更快。**预计新冠疫情的存在会成为一段时间甚至是未来几年的常态，疫情已成为医疗及其他相关领域人工智能应用落地的推动因素之一。疫情常态下园区、办公楼宇、社区的生物识别类应用比预计发展得更快，通过人工智能对人流量和活动轨迹进行识别，AI赋能的红外体温检测同时提高了工作效率，智能语音控制电梯等设备也有效降低了感染风险。

除了人工智能投入相对集中的行业之外，在业务需求的推动下，很多碎片化应用也开始被广泛使用，并辐射到媒体娱乐、现代农业、智能家居、智慧电力等多个不同领域，例如媒体娱乐中的智能选题、智能写作，现代农业中的产品分拣、农作物监测，智能家居中的家庭安保，以及智慧电力中的电力巡检等等。整体来看，包括生物识别、智能客服、精准营销在内的通用型应用场景已经具有相当的成熟度，人工智能的产业化已经从点到面，从通用应用场景渗透到更多行业特定场景，产业AI化已经从早期的试点逐渐成为企业发展和生存的刚需。

随着人工智能在行业的应用越来越广泛，头部厂商逐渐加大对人工智能生态建设的投入。作为中国领先的人工智能服务器供应商，浪潮凭借其在该领域的先发优势，逐渐加大对人工智能的投入以形成完整的人工智能产业链和生态，目前浪潮元脑Meta Brain 承载浪潮人工智能全栈能力，既包含浪潮全球领先的场景化人工智能基础设施、多样化的深度学习框架与工具，以及最新研发的人工智能 PaaS 平台和 AutoML Suite 等“有形”产品，同时也凝聚了浪潮多年积累的人工智能算法优化、系统优化服务等“无形”能力，以计算、资源和算法三大平台对外提供领先的AI计算力。

中国人工智能的应用场景发展 2020



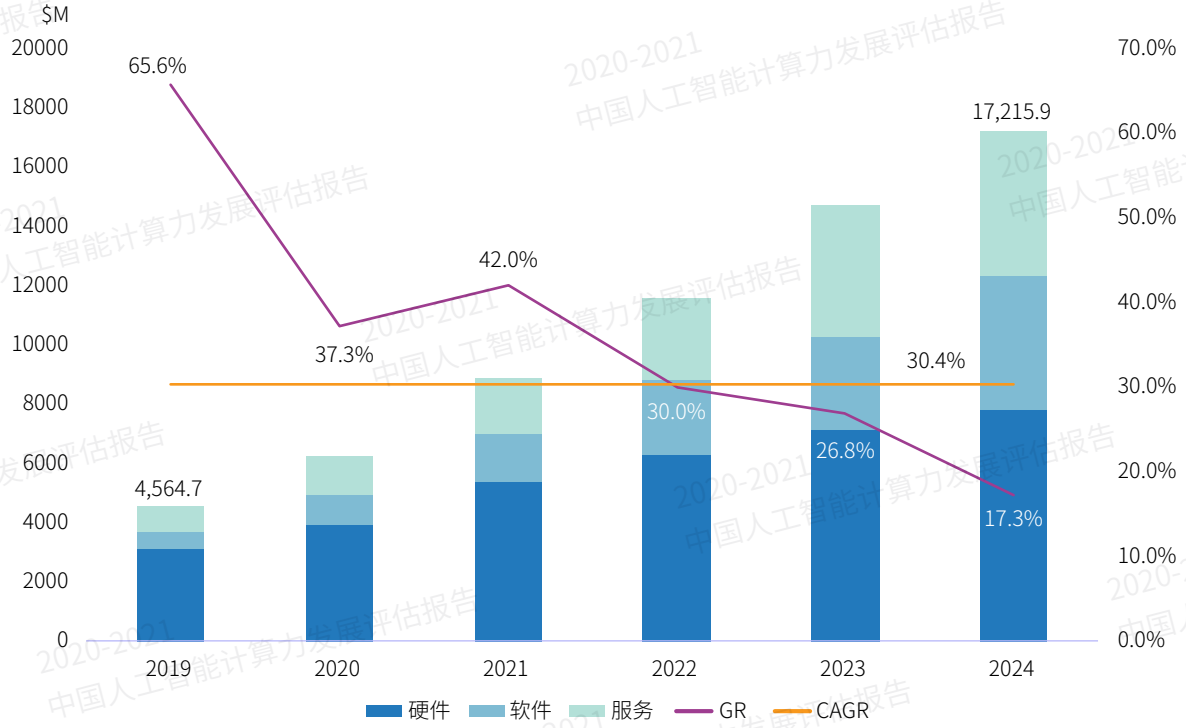
来源：IDC中国人工智能基础架构市场跟踪报告，2020H1

IDC针对重点行业人工智能的应用场景进行持续调研和梳理，如上图所示，纵轴是应用场景的市场成熟度，横轴是预测的应用场景得到广泛应用的时间，曲线则表示在该时间节点人工智能应用的平均市场成熟度。相比2019年，2020年人工智能应用成熟度有较显著的提升，整体来看，通用型应用场景由于被更大规模地采用并验证，具有较高的应用成熟度，而碎片化应用虽然已经逐渐被各行业用户所采用，但整体的应用成熟度仍相对较低，也是未来增长潜力较高的应用场景。

3.3 中国人工智能算力市场展望

IDC将人工智能市场技术类别分成硬件、软件和服务，在2019-2024年中国人工智能整体市场规模将保持30.4%的年复合增长率，预计到2024年将达到172.2亿美元的市场规模。虽然相比2019年的预测，因受到疫情影响整体市场规模比预期有一定的下滑，但仍将保持较高的增长率，同时也高于预测期内全球人工智能市场的平均增速。中国在全球人工智能市场的占比将从2020年的12.5%上升到2024年的15.6%，成为全球市场增长的重要驱动力。

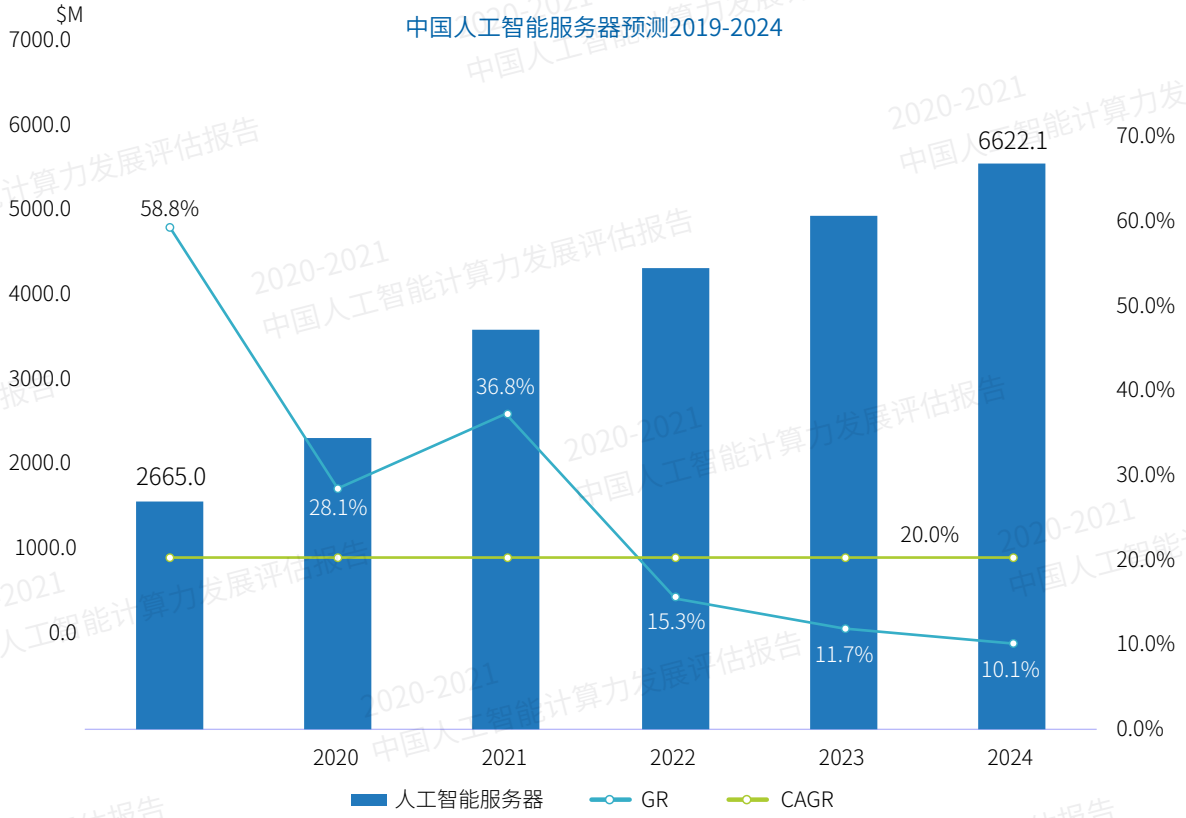
中国人工智能市场预测 2019-2024



数据来源：IDC, 2020

在整个人工智能市场中，算力扮演了极为重要的角色，过去，用户对于人工智能的感知更多停留在数据层和应用层，算力作为底层支撑似乎并不是关键的瓶颈，然而，随着人工智能算法突飞猛进的发展，越来越多的模型训练需要巨量的算力支撑才能快速有效地实施，而计算力受芯片工艺的物理限制，其在技术层面的提升速度已经落后于算法模型，某些模型已经逼近人工智能的算力极限，算力如果无法得到保障，未来算法模型本身的进一步创新也会造成影响。因此，算力已经成为未来人工智能应用取得突破的决定性因素。

IDC数据统计显示，2020年中国人工智能服务器市场规模将达到34.1亿美元，占人工智能基础架构硬件市场的87%，在包括软件、硬件、服务的整体人工智能市场占比也超过50%，数据直观地反映了算力在整个产业的重要性。



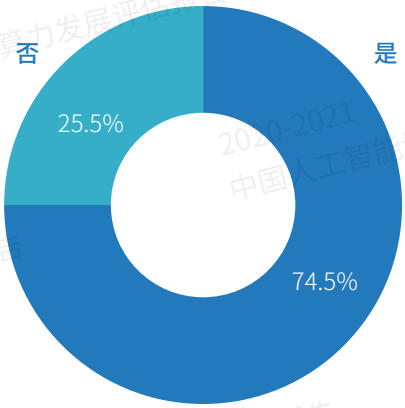
数据来源：IDC, 2020

为了更好地评估中国企业当前的人工智能算力发展现状，IDC于2020年6月至8月针对中国企业人工智能应用需求开展了专题调查和研究。

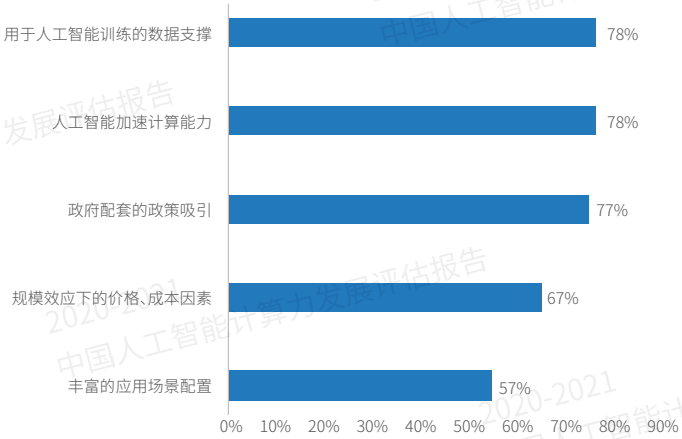
本次调研发现，企业对于人工智能带来的价值有了更深的认知，企业在人工智能应用上采取了很多积极的举措，最大的变化体现在新的应用场景尝试上，如前文所说，除了已经得到多个行业验证的通用场景之外，不同行业的用户也正根据行业的特性进行积极创新，尝试新的碎片化应用场景。**目前，有超过九成的企业正在使用或计划在未来三年内使用人工智能，其中，大部分企业采用了公有云、私有云加本地部署的混合架构来部署人工智能应用，而74.5%的企业期望在未来可以采用具备公用设施意义的人工智能新型基础设施，这一期望也是不同地区和不同规模的企业的共识。**无论是有能力自建私有云的大型企业还是严重依赖公有云服务的中小型企业，都希望可以在人工智能应用的部署方式上有更多、更灵活的选择。IDC调研显示，企业对于人工智能公共算力础设施的TOP5需求依次为：用于人工智能训练的数据支撑、人工智能加速计算能力、配套的政策吸引、规模效应下的价格和成本因素，以及丰富的应用场景配置。未来，随着更多规模化、普惠型的人工智能基础设施平台建成，整个中国人工智能产业将进入另一个发展的快车道——企业的刚需决定人工智能发展前进的速度，新型公共算力基础设施的建设则为人工智能发展拓宽了道路。

中国人工智能用户调研

贵单位是否有意采用人工智能算力中心?



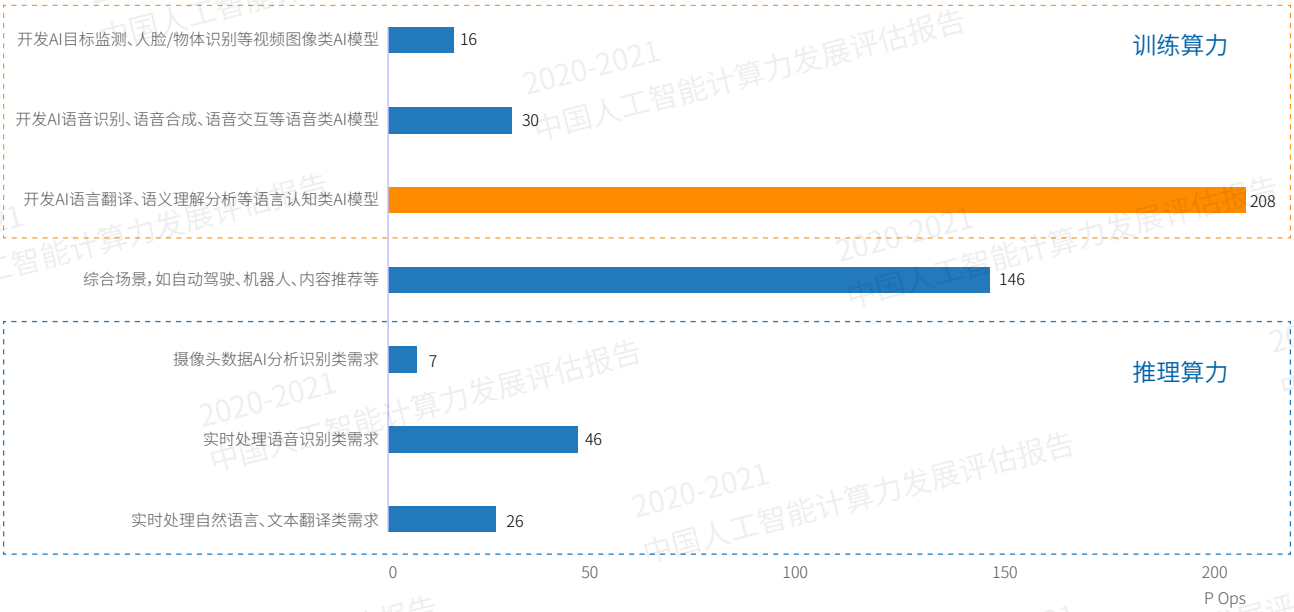
对人工智能算力中心的TOP5需求



数据来源:IDC中国人工智能用户市场调研,2020H1,N=300

本次调研中,IDC针对企业对于不同人工智能应用的模型,估算出了企业对于各项人工智能应用算力需求的平均值,未来三年,企业对于算力需求较为迫切的场景依然在语言认知类应用的训练上,对于该类AI模型训练的平均算力需求预计将达到208P OPS,远高于其他应用负载;对于自动驾驶、机器人、内容推荐等综合场景的算力需求也相对较高,平均算力值将达到146P OPS左右。

贵单位预计未来三年对人工智能应用的算力需求是?



数据来源:IDC中国人工智能用户市场调研,2020H1,N=300



中国人工智能算力发展评估与城市排名

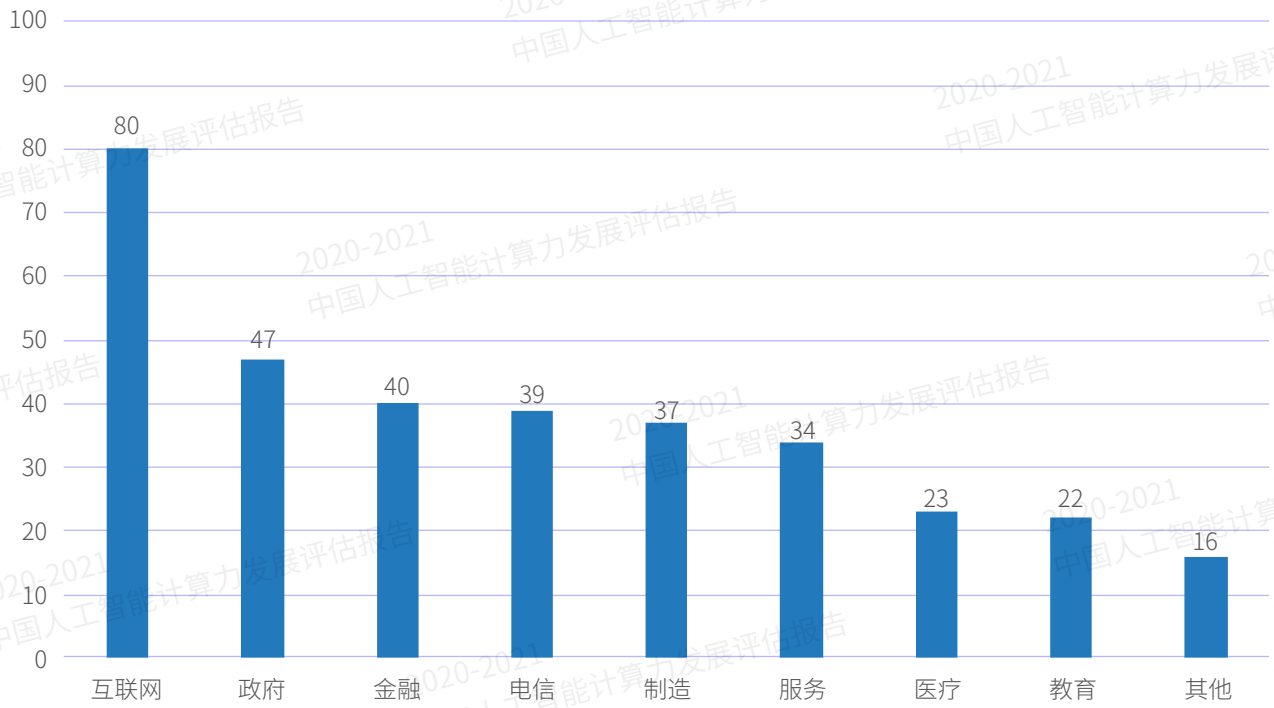
今年是IDC连续从行业和地域两个层面对中国人工智能算力进行发展评估的第三年，为了保持维度一致性，本次评估基本延续了过去两年的框架体系，分别从宏观层面、技术成熟度、劳动供给三个维度进行了评估。在宏观层面，通过考量某个行业或地区的经济整体发展水平以及在人工智能领域的支出占比，直观地反映出该行业或地区的人工智能的应用水平；在技术成熟度方面，重点考量不同行业对于人工智能的使用年限，应用的成熟度，以及第三平台技术的应用程度（第三平台解决方案，包括云、移动和大数据，推动了许多企业或组织的IT支出和新的人工智能项目）；在劳动供给层面，IDC重点考量了企业人工智能开发和运维的人员占比和技能水平，以及未来人才储备情况等因素。

4.1 行人工智能行业应用发展评估

人工智能在各个行业的应用程度逐渐加深，应用场景也越来越广泛，人工智能已经成为了企业寻求业务增长点、提升用户体验、保持核心竞争力的主要课题。IDC针对不同行业在人工智能方面的投入（包括人工智能投入占整体ICT投入占比）、行业应用场景成熟度、数据平台成熟度等维度进行了综合评估，基于持续的研究和最新的用户调研，**2020年，人工智能行业应用渗透度排名TOP5的行业依次为互联网、政府、金融、电信和制造，相比2019年，电信行业人工智能应用速度加快并超过制造业，位列第四。**

在2020年上半年，医疗行业在新冠肺炎疫情的影响下加速了人工智能应用的落地，在多方面取得了显著的成效，也促使医疗行业人工智能应用渗透度超过教育行业位列第七。2020年上半年，中国人工智能行业应用渗透度分布如下图所示：

中国人工智能行业应用成熟度，2020H1



资料来源：IDC, 2020

互联网——在中国，互联网行业是人工智能应用渗透度和投资最高的行业，互联网企业普遍重视人工智能技术的研究和价值，并加速其商业化落地。泛娱乐被公认为是互联网发展的重点领域，以IP价值挖掘为核心，电影、电视、动漫、游戏、音乐、网文等多领域逐渐打通，不仅行业的运作模式逐渐趋于成熟，对数字技术的接受和应用程度也处在领先地位。互联网企业在人工智能领域较早布局，同时随着互联网的深入发展和业务的多元化，其在自然语言处理、知识图谱、用户理解、计算机视觉、语音语义、深度学习等技术领域都有较大的投资和积累，除了将人工智能应用到自身的产品服务和运营职能等方面，还将人工智能通过云服务平台对外提供人工智能服务。

在企业自身应用层面，互联网企业将人工智能应用在身份验证、自动化客服、营销互动、精准营销、舆情管理、内容审核等多个应用场景。例如，人工智能作为百度的战略核心之一，运行于百度的服务和应用中，包括语音助理平台DuerOS、自动驾驶平台Apollo、百度云AI解决方案和云服务、百度搜索和百度Feed等；通过百度人工智能开放平台，百度向第三方开发者开放了百度人工智能功能并提供了百度云上的工具包。阿里巴巴已将各种人工智能技术广泛应用于数据经济体，以提升消费者体验和商业运营效率，包括基于深度学习和数据分析的个性化搜索结果和购物推荐、搜索功能中采用的语音识别和图像和视频分析技术以及智能客服，此外还将AI能力应用到其人工智能音箱天猫精灵产品中。

美团作为互联网生活服务平台的典型代表，拥有大量生活服务领域的场景和智能化升级的需求，以便深入的理解用户并提供更加个性化的服务，其中涉及外卖、餐饮、酒店、旅行等超过200个通过AI技术赋能的天然场景。美团通过图像视觉、语音技术、智能交互、自然语言处理、知识图谱、智能调度、运动控制、智能规划等技术，为用户提供更加智能化的体验，利用智能交互引擎为用户打造每个人都能享受的AI带来的红利，并利用AI享受更便利的生活。例如，美团以AI为关键技术引擎打造的智能物流网络辅助提升了配送的及时性和准确性；超脑调度系统基于海量数据和人工智能算法进行智能化自动分单，在消费者、骑手和商家三者之间实现最优匹配。丰富的线上线下生活服务场景，积累的丰富且海量的大数据反哺美团在大数据分析处理能力和模型训练，这也成为美团在AI能力上进行更深层次的技术迭代和升级的优势之一。

在疫情期间，美团利用人工智能技术为公司、客户和相关的政府部门提供帮助，例如通过对不同行业、不同区域以及不同人群的影响进行分析，能够以更加全面的视角对社会的经济复苏情况提供参考。未来，美团还将在人员、资金、技术等方面持续投入，发展无人配送，增加无人配送车等与骑手形成运力补充，进一步提升配送效率，改善全民互联网生活服务体验、赋能生活服务产业升级、助力社会治理。

医疗——相比2019年，2020年上半年，医疗行业的人工智能应用有了显著提升，虽然人工智能已经被中国的医疗机构和生命科学组织广泛接受，但是目前因为医疗人工智能的相关标准和规范还不完善，只有少量医疗人员参与人工智能开发和应用。在未来五年，随着中国政府对人工智能开发和应用相关法规的完善，人工智能应用将会快速拓展。

在医疗数字化转型的过程中，传统医疗向互联网医疗模式的转型趋势愈加明显，而人工智能、大数据等新技术的发展则使得疾病诊断和治疗的模式发生转变，并逐渐从单点开始逐步扩展到各个领域，带动了医疗信息化的全面升级。当前，医疗人工智能系统主要采用的技术包括计算机视觉、自然语言处理、机器学习等。医疗人工智能系统主要包括四种类型：医患交互式服务人工智能、医学影像辅助诊断和治疗的人工智能、临床决策支持人工智能和人机交互接口的人工智能，均已在医院中获得成功应用。

在抗击新冠疫情中，人工智能在疾病救治和疫情防控中发挥了巨大作用。对患者的肺部放射影像诊断需要医生检查大量的放射影像，耗费大量的精力和时间，医学影像 AI 辅助诊断系统的应用提高了诊断的效率；物联网技术与人工智能技术相结合用于监测人群体温，并对发烧者进行提示和监测。经过此次疫情防控，医疗数字化转型加速，全社会一体化的临床诊治与疾控及健康管理系统建立起来，人工智能作为医疗行业数字化转型的主要技术之一，连同大数据、机器人、物联网等技术将在健康管理中加速落地应用，并融入到医疗设备和器械的应用中，提高全社会疾病防控能力，并提高诊疗效率和护理水平，促进医疗健康服务模式的数字化转型。

IDC预计，到2023年，人机协作和人工智能驱动的接口将改变未来中国四分之一的医院工作方式及相应的医疗信息系统；到2024年，中国30%的医疗和生命科学组织的商业和临床决策将基于人工智能提供的洞察做出决策。

案例 – 联影医疗/联影智能

联影智能作为联影医疗技术集团公司之一，以AI赋能临床、赋能科研、赋能设备为使命，是全球少数能够提供贯穿成像、筛查、随访、诊断、治疗、评估全流程，覆盖全病种，全栈全谱智能解决方案的医疗AI企业。通过公司的uAI一站式解决方案，即一次扫描，AI辅助多病种筛查；一套诊疗工作流，AI全程优化；同一AI技术，多模态设备赋能；一大平台，AI助力科研与合作，为医疗行业的智能化提速增效（四个一）。

- 赋能设备：以AI赋能成像全流程，实现智能数据获取和处理以提升影像科工作效率，达到更全局的优化。
- 赋能临床：为医生的临床治疗提供贯穿成像、筛查、随访、诊断、治疗、评估的全栈式AI解决方案，助力医生精准诊疗。
- 赋能科研：打造医学影像AI共享平台与生态系统、覆盖从技术、应用的共享、跨界开放，加速临床、科研成果的落地与产业化。

2020年初，受新冠疫情的影响，云计算、大数据技术、融合人工智能等各种新应用在医疗健康领域蓬勃发展。尤其人工智能技术获得了医疗机构的青睐，医院积极推进并大规模部署人工智能应用，推动了AI在应用广度和深度上取得了较大进展。

这为以联影智能公司为代表的智能医疗技术服务提供商提供更广阔的市场空间，同时也对AI算力、存储等方面提出更高的要求。联影智能是联影集团在人工智能领域中的重要战略布局，在AI赋能医疗影像和放射科的成像、临床和科研等领域持续部署、重点突破。

电信——人工智能技术已经成为电信行业不可缺少的部分，运营商凭借庞大的用户基数获取海量的数据，其来源丰富、覆盖行业广泛、真实性高，形成了高价值的训练数据集，为人工智能在电信行业未来的发展奠定了基础。电信行业人工智能应用主要体现在两个方面，一方面，电信网络的构建及优化智能需要多项技术的融合，包括GPU加速、深度学习和分析技术等；另一方面，下一代智慧网络的打造，对云化网络的智能编排、调度、运营等也需要人工智能技术的支撑。另外，对运营商自身而言，越来越多的电信运营商着手于智慧营业厅的建设，捕捉消费者数据，比如：停留时间、行为习惯等，然后利用人工智能技术对获取的信息进行分析并及时作出反应，或者增加柜台客服人员，或者引导客户使用自助服务系统。电信行业的人工智能应用不仅优化和维护了基础设施、给客户创造了更好的体验感，还能提高业务营收、降低成本以及提高效率。目前云服务也是电信运营商的一个重要环节，电信运营商会选取x86架构为核心的服务器建设自己的公有云，并且大量引入GPU、FPGA等一系列加速芯片以适应未来的发展趋势，同时在自身的传统业务上，利用多样化架构实现更加优化的运行。

某电信运营商积极打造自身的人工智能平台，以网络智能化建设为重点，以开放合作共赢为目标，为各行各业用户提供基于人工智能的规模化的商业分析和商业策略。该人工智能平台涵盖深度学习平台和AI能力平台，提供从基础设施到核心能力的开放AIaaS服务，不仅汇聚了企业内部自研能力如语音语义、机器视觉、结构化数据分析、以及自然语言处理等，同时也积极与第三方人工智能技术服务商如商汤、讯飞等合作伙伴进行合作，开拓面向不同行业的顶尖AI应用能力，满足不同场景的智能化需求。目前该电信运营商着眼于自身的人工智能平台的发展，已经在交通、金融、工业、能源等领域，围绕数据采集、归类、分析和决策等，在图像、语音、文本等不同领域，为平台AI算法和模型的训练和优化进行广泛的实践和数据积累。

制造——人工智能是制造业迈向工业 4.0 和工业互联网时代的重要新兴技术应用。人工智能在该领域应用于工业自动化和智能化，实现非接触检测、提高加工精度、发现产品缺陷，进行自动分析和决策，是先进制造业的重要组成部分。在制造业中人工智能不断丰富和迭代自己的分析和决策能力，以适应不断变化的工业环境，提高生产效率或设备产品性能，具有自感知、自学习、自执行、自决策、自适应等特征。其本质是实现复杂工业技术、经验、知识的模型化和在线化，从而实现各类创新的工业智能应用。制造业在人工智能的主要应用场景包括：交互界面智能化、质量管理及推荐系统、维修及生产检测自动化、供应链管理自动化、产品分拣等。IDC预计，到2023年年底，中国50%的制造业供应链环节将采用人工智能，从而可以提高15%的生产率。

除此之外，业务的开展高度依赖线下的行业，如保险行业，传统保险业务的开展主要以代理人和客户面对面的方式进行，2020年初，由于突发的疫情，涉及营销、核保、理赔、客服以及风控等关键业务流程的系统在疫情的影响下，其人工智能应用的渗透度均获得不同程度的提升。

案例：某大型保险公司

某大型保险公司在疫情的影响下加速自身的数字化转型，一方面将公司的日常运营快速转成线上的形式，同时也加大对基础设施如人工智能服务器的投资，加强和通用算法领域的厂商合作，整合多方资源，在保单审核、智能客服、人脸识别、智能核保、客户信息录入等关键业务环节，以人工智能的图像识别、语音识别以及智能语音回呼等技术的融合应用助力保险业务的开展，在实现理赔的高效创新的同时提升客户在各个环节的体验，打造智慧保险。

4.2 城市人工智能算力排行

在2020年中国城市人工智能算力排行榜中，TOP5城市依次为北京、深圳、杭州、上海、重庆，排名6-10的城市为广州、合肥、苏州、西安、南京。与2019年相比，深圳超过杭州位居第二，重庆进入第一梯队，西安超过南京排名第9。除了TOP10城市之外，多个城市在自身产业优势及各种因素推动下，人工智能应用取得了较大进展，例如东莞的智能制造、武汉的智慧医疗、合肥的智慧农业等，中国人工智能城市发展正遍地开花，未来将会出现越来越多结合城市特点的人工智能示范区，为产业发展树立标杆。

2020中国AI计算力发展评估 – 城市排行

TOP 10 城市：

TIER 1

①北京

②深圳

③杭州

④上海

⑤重庆

TIER2

广州 合肥 苏州 西安 南京



资料来源：IDC, 2020

北京——作为全国政治中心，北京发布了多项针对人工智能发展的政策，比如，2019年年底，北京发布鼓励科学创新的相关政策——《2019人工智能关键技术源头创新专项申报指南》、《2019人工智能开源开放创新平台建设专项申报指南》（以下简称《指南》），指出重大人工智能创新项目最高可获得3,000万元支持。《指南》显示，创新项目主要聚焦于人工智能芯片、核心算法、框架、基础软件等关键领域。北京还布局建设了七个国家新一代人工智能开放创新平台，明确将人工智能作为北京科技创新和产业发展的重点领域，加大人才支持力度，聚焦创新和技术，支持科学家开展原创性探索研究，全市重大项目、重大工程向人工智能技术和应用开放，努力构建智慧城市。《北京市加快新型基础设施建设行动方案（2020-2022年）》中提到，到2022年，北京市将建设超300平方公里自动驾驶示范区；打造北京城市大脑，完善人工智能基础设施等。

除了政策的支持以外，北京拥有优质学术和人才资源，与人工智能相关的人才占全国总量的一半以上；相关人工智能企业数量过千，占全国三分之一以上。北京大学、清华大学、中科院自动化所、中科院计算所等全国过半数人工智能骨干研究单位都聚集在北京。

综上所述，北京在人工智能领域的人才、技术、企业、资金等方面的优势有助于其在人工智能领域取得领先地位，并将继续吸引新的企业和投资资金进入。

深圳——深圳是国家设立的五大经济特区之一，经过多年的发展，如今的深圳已是一线大城市。由于其靠近香港的优越地理位置，吸引了香港商人前来投资，收获资本，人才和经验。此外，深圳还有“世界工厂”之称，在政府的扶持和鼓励下，吸引了众多发达国家的产业落地，使得无数科技人才或者互联网人才大量涌入深圳；而华为，平安，腾讯，招商银行等知名企业也为深圳人工智能的发展打下了坚实的基础。平安科技和腾讯是人工智能技术的核心应用者和开发者之一，在这些企业的带动下，深圳市人工智能算力的投资位居全国第二位。

杭州——杭州市政府为了加快新一代人工智能创新发展，在2020年初也出台了若干鼓励性政策，比如：杭州市政府发布的《杭州市建设国家新一代人工智能创新发展试验区若干政策》指出，支持企事业单位开展关键核心技术研究，按项目研发投入给予补助，支持在杭州的企业、高校、科研机构建设开源开放、协同共享的人工智能科技创新服务平台，对实际投入500万元以上、服务产业取得良好成效的，按其软硬件投入的30%给予补助，最高不超过500万元；支持建设人工智能共性技术研发平台，按平台建设主体新购置研发设备总价的50%进行补助，单个平台补助最高不超过3,000万元等。7月初，杭州市科技局和浙江大学成立人工智能研究院，将针对技术创新、产业培育、融合应用三大主线进行研究与建设。

在阿里集团和浙江大学等科研院校的加持下，杭州在人工智能领域的起点颇高。杭州地区相关人工智能的专利申请和授权数已超过千件，位居国内人工智能的第一梯队。同时，杭州的学科建设历史悠久，创新研发底蕴深厚，早在1978年浙江大学就开始了人工智能方向的硕士培养和学术研究，1982年成立人工智能研究室并在1987年升级为研究所，是国内人工智能研究的起源地之一。



上海——上海是中国最大的经济中心城市，地处长三角地区，这里具有完备的产业体系、开放的市场环境，并且在智能制造和智能交通等领域构建了丰富的感知系统和应用场景，给人工智能应用落地的实验提供了优越的条件。

众所周知，数据是人工智能的要素之一，上海的人流，物流，车流，资金流，信息流都非常密集，每天产生和使用海量的数据。上海已经建成了法人库、实有人口库和空间地理库三大基础数据库。这为人工智能在上海的发展奠定了坚实的基础。



重庆——相比去年，重庆进入第一梯队，排在第五名，主要有以下原因：

首先，重庆市发布了人工智能“十大应用场景”，包括：重庆智慧城市智能中枢系统，智慧安防系统，智慧共享物流中心，礼嘉智慧公园，智慧广阳岛数字孪生平台，武隆全域智慧旅游平台，智慧诊疗伴侣平台，智慧法律公共服务平台，工业互联网安全服务及综合管理平台，西部自动驾驶开放测试基地。

其次，在教育部印发的《教育部关于公布2018年度普通高等学校本科专业备案和审批结果的通知》中提到，重庆大学的人工智能、机器人工程、数据科学与大数据技术、智能医学工程等4个专业全部获准设置。而人工智能为新设专业目录，重庆大学也成为全国首批35所设置人工智能本科专业的高校之一。此外，重庆理工大学开设了机器人工程专业、数据科学与大数据技术专业；重庆文理学院也开设了机器人工程专业等。诸多高校的加持为重庆市人工智能相关专业人才的培养奠定了基础。

最后，重庆拥有众多人工智能相关企业，其中重庆海云数据曾上榜“中国人工智能创新成长企业50强”，重庆市排名的提升也离不开这些企业的带动。

西安——西安在本次城市排名中上升至第9名，这一成绩首先离不开国家政策的支持。今年3月份，科技部在官网发布《科技部关于支持西安建设国家新一代人工智能创新发展试验区的函》，函中表示支持西安市建设“国家新一代人工智能创新发展试验区”。至此，获科技部支持建设的“国家新一代人工智能创新发展试验区”增至11个。同时，西安有清晰的人工智能规划、人工智能相关的人才，科研实力雄厚，人工智能企业超过100家，并且面向集成电路，大数据，机器人整机和关键核心部件产业，无人机系统产业等重要环节。另有海航集团、浪潮集团、吉利汽车等在人工智能领域有重点布局的大型企业已与西安市签约，在当地进行项目投资建设。

除以上城市外，还有些城市深耕特定的人工智能应用并取得了明显成果，成为城市的智能化新标签。如广东重要的制造中心和外贸口岸东莞市，在政府政策的支持下，引入人工智能技术，将人工智能技术和制造相融合，实现生产线自动化、智能化，从而进一步提高生产制造效率；武汉市在疫情期间，也利用人工智能技术进行高效的疫情防控，例如部分医院使用智能消毒机器人，可在指定地点喷洒雾状消毒剂，可节省大量人力，缓解疫情期间人员紧张的压力；合肥市率先开展了智慧农业的探索，使用人工智能技术，对果园进行管理和检测。利用人工智能图像识别技术，可准确识别每一棵果树，自动统计棵数，还可以利用人工智能技术对果树图片进行大量对比，及时发现长势不好的果树并做出分析和诊断，从而帮助果农及时止损、提高产量。中国人工智能发展的城市覆盖面正从核心的地区和城市向外扩展，未来将会出现越来越多具备城市特色的人工智能示范区，为产业发展树立标杆。



全球人工智能算力发展评估

放眼全球，在企业对人工智能基础设施的建设以及相关应用场景的大量试验之后，人工智能取得了迅速的发展，不仅产业生态更加完整，IT厂商的人工智能解决方案丰富多样，人工智能的行业应用也呈现百花齐放的局面。

全球人工智能市场发展蓬勃，IDC全球半年度人工智能IT支出报告显示，到2024年，全球整体人工智能市场规模将达到1,106.6亿美元，2019-2024年复合增长率为20.1%，其中硬件、软件和服务的年复合增长率分别为19.0%、22.5%和18.4%。未来五年中国人工智能市场的增速将超过全球人工智能市场，中国在全球人工智能市场的占比将从2020年的12.5%上升到2024年的15.6%，成为全球市场增长的重要驱动力。

从2020年各个国家AI计算的发展水平来看，中国和美国是人工智能服务器支出占整体服务器支出最高的国家之一，占比均超过10%。中国人工智能服务器在整体服务器占比中份额居高，中国人工智能服务器占整体服务器市场的16%左右，高于全球人工智能服务器市场在整体服务器中的平均占比。从另一方面来说，2020年中国人工智能服务器占全球人工智能服务器市场的三分之一左右，是全球人工智能产业发展的中坚力量。AI计算能力反映的是一个国家最前沿的计算能力，这也说明这些国家在战略层面对人工智能的重视，以及企业希望通过人工智能的发展契机提升核心竞争力的迫切愿景。人工智能服务器的投资能够为人工智能应用的落地提供核心的算力保障，对于未来国家计算力指数的提高具有极大的推动作用。随着人工智能在各行各业的加速渗透，AI计算能力将成为推动国家经济再次焕发活力的重要因素，尤其是对于计算能力和人工智能应用水平均衡发展的领跑者国家产生的经济效益最为显著。未来，将人工智能等新兴技术作为国家级的战略方针都将有助于各国进一步提高计算力指数和经济发展的相互拉动力。

从全球人工智能市场来看，各个参与主体积极对人工智能进行投资和部署。全球主流芯片厂商及IT厂商持续更新产品和解决方案、头部互联网企业加大对人工智能的投入和研发、大型行业用户利用人工智能赋能企业自身业务、人工智能领域的并购屡见不鲜。中国对人工智能的投资在近几年呈现快速增长的态势，以超大规模的互联网公司和大型行业用户为首，企业纷纷加大对人工智能的投资，来满足日益增长的数据量对于算力的需求，并寻求企业的数字化转型和中国整体产业变革和升级。

综上所述，人工智能依然是全球IT产业发展最快的新兴技术应用之一，未来几年全球的人工智能发展潜力巨大，各国政府均在战略层面对于人工智能高度重视，并出台相关政策指导人工智能发展并加快人工智能行业应用，中国作为全球人工智能发展的中坚力量，政府和企业也将持续加大对人工智能的投入和研发，随着更多传统行业与人工智能技术的不断融合，各类创新的应用场景也将加速落地。



行动建议

考虑到中国人工智能市场的现状，IDC针对行业用户和人工智能解决方案提供商分别提出了如下行动建议，希望对中国人工智能的发展和生态的成熟有所裨益。

6.1 对行业用户的建议

明确自身的里程碑和路标

人工智能已经从企业实验性采用转变为企业数字化转型的刚需。率先采用人工智能的企业已经从中获益，未来企业应该以更加长远的目光加大进行人工智能的投入。IDC建议致力于发展人工智能的企业，首先在企业的最高战略层制定规划和策略，明确自身在人工智能技术上的战略规划，结合自己的业务发展需要，明确自身智能化关键里程碑。与此同时建立自上而下的管理体系并设法进一步量化这种投资带来的收益。同时企业需根据本行业的特点和自身的实际情况，建立与人工智能应用和优化相关的KPI指标，使得运营人员可以更好地衡量并确认人工智能的战略投资是否合理，从而最大化人工智能的投资回报率。

在算力基础设施上提前布局

人工智能基础设施正在变得越来越重要，在明确人工智能在企业中的战略地位的同时应该积极在算力基础设施上提前布局，做好基础设施服务商的选型工作。除了性能之外，企业在部署人工智能基础设施时应该综合考量数据和应用的可用性、基础设施的利用率、成本优化等因素。企业可以考虑开放式硬件架构，提供模块化的“即插即用”体验，更好得保障技术基础架构的技术先进性和可成长性。从而提高人工智能基础设施的利用率，并实现更灵活的投资和部署方式。



注重AI算力产品的生态成熟性和技术领先性，真正把计算力转化成生产力

在选择人工智能算力基础架构相关产品时候，全方位进行评估和考虑，除了需要统筹不同范畴内的各种对象、接口、框架、功能、组件、算法等基础能力之外，还应重点关注服务商的生态完整度与成熟度，并且结合企业自身人工智能部署规划，关注服务商技术的先进性以及稳定性，真正的将计算力转换成生产力。

6.2 对技术供应商的建议



在硬件层面，人工智能基础架构服务商除了丰富和完善自身的产品之外，还应该关注AI计算平台的先进性，可根据客户应用场景和需求特点的不同，进行针对性优化方案，覆盖尖端推理边缘端 ，支持多样性发展的AI芯片，包括GPU/FPGA等，注重开放架构。



在软件层面，服务商应该注重打造自身的全栈能力，以帮助客户快速构建AI应用环境 ，真正做到能够从用户层，帮助致力于采用和部署人工智能的用户实现集高效部署与敏捷资源供应为一体的软硬件解决方案。



在生态层面，随着产业AI化的不断深入，技术提供商应加大产业AI化生态，在算力、算法和数据三要素层面实现有效连接 ，以增强产品和解决方案的互操作性和标准化，加速人工智能在行业落地，实现人工智能的规模化发展。

6.3 对产业发展的建议

IDC调研结果表明，企业在发展人工智能的过程中在数据、算力、优惠政策以及配套的服务等方面具有广泛的共性需求，具备公用设施意义的人工智能新型基础设施成为解决这些需求的不二之选。以政府机构为代表的社会服务主体，由于其在社会、经济、制度制定和管理等方面的综合影响力，是推动人工智能算力设施平台建设的关键力量。

在新型基础设施建设上，国家发展改革委明确了新型基础设施范围，将智能计算中心（以下简称“智算中心”）作为算力基础设施的重要代表纳入信息基础设施范畴，并且出台和制定加快新型基础设施建设和发展的意见，为建设“高效能计算基础设施”的建设做好顶层规划。

目前，国家信息中心联合浪潮发布了《智能计算中心规划建设指南》，明确了智算中心的概念、内涵、功能定位和技术架构，强调了智能计算中心对外提供算力服务、数据服务和算法服务的公共算力新型基础设施特征，在实施路径层面，强调政府主导投资、政府主导的政企合作以及“投-建-运”一体化的运营模式。

另外，IDC在本次研究中也针对企业在人工智能应用中普遍存在的需求和挑战进行了调研，其中缺乏模型训练所需的数据、算力基础架构存在不足、以及人工智能应用方案的成本过高等因素是绝大部分企业目前面临的主要挑战。因此，以政府为代表的社会服务主体，在推动人工智能公共算力基础设施的建设时至关重要，需要充分考虑企业目前面临的需求和痛点，与人工智能生态合作伙伴一起，携手关键的人工智能基础设施服务商，通力合作，共同构建加速产业发展的平台。

最后，在人工智能应用逐步落地的过程中，风险管控是整个产业可持续发展的基本保障，针对人工智能应用可能出现的风险，政府相关部门已经制定了一系列规范。未来，随着人工智能应用的进一步深化，政府也需要不断完善新一代人工智能治理规范，为人工智能产业化保驾护航。



关于浪潮

浪潮，全球领先的云数据中心基础架构产品及方案提供商，服务全球100多个国家和地区，以服务器、存储、云操作系统、信息安全技术为客户打造领先的云计算基础架构平台。

人工智能是浪潮战略重点业务之一，浪潮是全球第三大服务器供应商，也是全球领先的AI服务器提供商，在中国AI加速计算市场占有率超过50%，可提供从计算平台、管理套件、框架优化到应用加速的完整方案。

浪潮推动AI领域开放计算的发展，参与制定了OCP社区的OAM规范以及ODCC社区的GPU服务器规范，为不同的AI技术提供统一的技术规范。以元脑为技术平台，致力于为各行业用户提供软硬化一体的AI解决方案，形成从技术到应用的开放融合的AI技术生态。

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，全球领先的专业的市场调查、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 年的发展历史中，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

全文下载



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。