紫光集团 | H3C
核心企业 | 数字化解决方案领导者

新华三人工智能发展报告白皮书

2020年4月
AI研究院&2029战略研究院

新华三集团

北京总部
北京市朝阳区广顺南大街8号院 利星行中心1号楼
邮编:100102

杭州总部
杭州市滨江区长河路466号
邮编:310052

www.h3c.com

CN-203X30-20200421-BR-SD-V1.0

序

REPORT

纵观人类的发展历史，每一次重大变革，都会使一些组织或行业产出成指数级增长。改良蒸汽机促进了工业时代的到来，而计算机的发明则引领了信息时代的到来，两个时代的技术革命都使生产力实现了革命性提高。

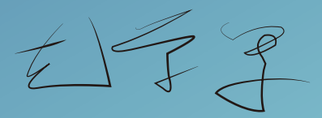
在国家层面，早就提出：加快发展新一代人工智能是我们赢得全球科技竞争主动权的重要战略抓手，是推动我国科技跨越发展、产业优化升级、生产力整体跃升的重要战略资源。2020年3月4日，在强调加快新型基础设施建设进度的中央会议上，人工智能更是作为重点领域被再次提及和关注。

在企业数字化转型过程中，AI技术将承载更多重任。打通从感知智能到认知智能的道路上，需要融合传统机器学习、知识图谱、自然语言处理、语音识别、图像识别、深度学习等多种AI技术。紫光旗下新华三集团通过打造包含了AI中台与数据中台的智能计算平台，促进认知智能的发展，并与ICT产业协同演进，让AI与工业、医疗、教育、政务、安防等行业深度融合，帮助传统行业客户走向AI化。

如今我国在人工智能技术应用方面，已经走在了世界前列。新华三在ICT领域中深耕多年，面对新兴的人工智能技术，新华三将借助ICT产业与人工智能协同发展契机，全面推动人工智能的跨层级布局，进一步提高企业的综合实力，与云、网、端协同共进，构建灵活高效的数字化IT体系，让“智能+”深入百行百业。

未来世界科技产业将以人工智能为基础进行转型与升级，与各行业高效协同、生态融合，在这个过程中，新华三将全力加码行业AI发展，引领产业智能迈向新阶段。

新华三集团联席总裁、首席技术官 尤学军



目录

CONTENTS

引言
INTRODUCTION

随着人工智能技术的普及，人类社会正在从信息化时代步入智能化时代。我们在生活中已经能切身感受到人工智能带来的便利，从虚拟语音助手到自动驾驶汽车，很多场景中能够找到人工智能的身影。国家层面也极其重视这项变革性的技术，围绕各个领域的智能化制定发展战略。人工智能作为新一轮产业变革的核心驱动力之一，对社会和经济将产生深远影响。人工智能与行业场景深度结合，会产生显著的效益：行业场景拥有第一手数据资源，拥有丰富的场景需求，人工智能可以助力传统行业实现跨越式升级，同时人工智能技术本身也得以持续进化。目前安防、金融等行业的人工智能变革已经取得了较好的成果。

人工智能技术如火如荼的发展得益于信息与通信技术（ICT）的有力支撑。当前的人工智能技术是以海量数据驱动的学习算法为主，需要强大的算力来支撑。近几年高性能GPU服务器、计算集群、大数据技术以及高性能的网络和存储等基础设施为人工智能的迅猛发展提供了得天独厚的条件。紫光旗下新华三集团作为数字化解决方案领导者，拥有计算、存储、网络、安全等全方位的数字化基础设施整体能力，可提供包括人工智能在内的一站式数字化解决方案。基于新华三在人工智能领域多年的探索整理本报告，旨在阐明新华三对人工智能技术的理解与认知，分享新华三在人工智能应用方面的实践。

新一轮科技变革来临
智能化浪潮势不可挡

每次科技变革都推动人类社会跨越式发展

在人类近代发展历史上，经历了三次重大科技革命。每一次科技革命都带来了人类社会的巨大变革。科技作为第一生产力，已经成为人类社会发展进步的关键要素。前两次科技革命，分别使人类社会迈进“机械时代”、“电力时代”。始于上世纪中期的第三次科技革命，主要起源于美国、前苏联和欧洲各国，以原子能、电子计算机、空间技术和生物工程的发明和应用为主要标志，涉及信息、新能源、新材料等诸多领域的一场信息技术革命，人类社会从此进入“信息化时代”。第三次科技革命极大地推动了人类社会政治、经济、文化领域的变革，同时也很大程度上改变了人类的生活方式和思维方式。当下，我们正在进入以人工智能、物联网、5G通信、机器人、新能源、新型工业材料等前沿技术为代表的第四科技革命，可以预见这次科技革命在规模、影响力方面都将远远超过前几次，并且会改变当前国家竞争格局。因此美、日、韩以及欧洲各国政府竞相出台政策加大对前沿技术的投入，以抢占这轮科技革命的领先优势，而且人工智能作为这轮科技革命中的头雁技术，更是备受关注。

顶层政策支持，国家战略地位

2018年5月，习近平总书记在两院院士大会强调：“新一轮科技革命和产业变革正在重构全球创新版图、重塑全球经济结构，科学技术从来没有像今天这样深刻影响着国家前途命运，从来没有像今天这样深刻影响着人们生活福祉”，“现在，我们迎来了世界新一轮科技革命和产业变革同我国转变发展方式的历史性交汇期，既面临着千载难逢的历史机遇，又面临着差距拉大的严峻挑战”。

连续三年相继出台了很多人工智能相关的政策。2017年7月，国务院印发了《新一代人工智能发展规划》，明确了我国发展人工智能的战略目标，并进行了总体部署，设立了“三步走”目标：

- 到2020年，人工智能技术和应用与世界先进水平同步，人工智能产业成为新的重要经济增长点，人工智能核心产业规模超过1500亿元，带动相关产业规模超过1万亿元；
- 到2025年，人工智能基础理论实现重大突破，部分技术与应用达到世界领先水平，人工智能成为带动我国产业升级和经济转型的主要动力，核心产业规模超过4000亿元，带动相关产业规模超过5万亿元；
- 到2030年，人工智能理论、技术与应用总体达到世界领先水平，成为世界主要人工智能创新中心，核心产业规模超过1万亿元，带动相关产业规模超过10万亿元。

2017年12月份，工业和信息化部印发了《促进新一代人工智能产业发展三年行动计划（2018-2020年）》，从推动产业发展角度出发，以三年为期限明确了多项任务的具体指标，对《新一代人工智能发展规划》相关任务进行了细化和落实，以信息技术与制造技术深度融合为主线，推动新一代人工智能技术的产业化与集成应用。同时，大力鼓励和支持传统产业向智能化升级，陆续出台《智能制造发展规划（2016-2020）》、《产业结构调整指导目录（2019年）》等重要文件，为产业升级提供了有力的政策保障。

时间	政策
2017年3月	《2017年政府工作报告》
2017年7月	《国务院关于印发新一代人工智能发展规划的通知》
2017年10月	十九大报告
2017年12月	《促进新一代人工智能产业发展三年行动计划（2018-2020年）》
2018年3月	《2018年政府工作报告》
2018年4月	《高等学校人工智能创新行动计划》
2018年11月	《新一代人工智能产业创新重点任务揭榜工作方案》
2019年3月	《2019年政府工作报告》
2019年3月	《关于促进人工智能和实体经济深度融合的指导意见》
2019年6月	《新一代人工智能治理原则——发展负责任的人工智能》

表1 近三年我国政府出台的人工智能相关政策

近期，人工智能又被纳入新型基础设施建设，成为“新基建”七大方向之一，属于信息化领域的通用基础技术。概括来讲，“人工智能新基建”是指围绕提供基础智慧能力的一系列芯片、设备、算法、软件框架、平台等的统称。推动“人工智能新基建”有助于加速传统产业智能化升级，反过来也促使人工智能技术的升级进化。

人口老龄化趋势加重，智能化升级迫在眉睫

我国上世纪末进入老龄化社会，从2000年到2018年，60岁及以上老年人口从1.26亿增加到2.49亿，老年人口占比从10.2%上升到17.9%，提升幅度是世界平均水平的2倍多。而且未来较长一段时期内，老龄化的趋势还将持续下去。相应地，随着人口老龄化带来的劳动力资源短缺以及劳动力成本的增加，将会对我国经济和社会发展产生一定的阻力。2019年底，国务院正式印发的《国家积极应对人口老龄化中长期规划》明确指出，充分发挥科技创新引领带动作用，把技术创新作为积极应对人口老龄化的第一动力和战略支撑。

利用人工智能、机器人等作为劳动力替代及增强技术来应对劳动人口减少的挑战，产业智能化升级，用科技手段从根本上对冲人口老龄化对经济发展所带来的不利影响是必然选择。

基础条件已渐趋成熟，人工智能应用将进入爆发阶段

近几年来，随着数字化基础设施的不断完善，再加上以深度学习为代表的算法上的突破，人工智能技术日渐成熟，已经在安防、金融、客服、工业制造等领域，取代了大量重复性高、繁琐枯燥或者大量使用人工并不经济的工作，不仅降低成本，而且生产效率提升也十分显著。人工智能技术在京津冀、长三角、珠三角地区已经初步带来产业规模效益。据IDC报告显示，预计到2023年中国人工智能市场规模将达到979亿美元（包含软件、硬件、服务等），2018-2023年复合增长率为28.4%，我国人工智能关联产业将进入了快速发展阶段。

随着5G商用落地，高带宽、低延迟、大接入的特性将会进一步拓宽人工智能应用场景的边界，未来3-5年，为人工智能技术在产业智能化的爆发奠定坚实的基础。

PART 1

人工智能商业化加速
将深刻改变人类社会

人工智能从学术研究走向商业应用

2016年，谷歌AlphaGo以4:1的成绩战胜了人类顶尖围棋选手李世石，让人工智能走进了大众的视野。人工智能如今已不再停留在学术研究阶段，开始大规模的应用到商业环境中。

人工智能发展历史

人工智能最早可追溯到上世纪的四五十年代，被誉为“人工智能之父”的艾伦·图灵，在其论文《计算机器与智能》中，提出了非常著名的图灵测试，即被测试的机器是否能够表现出与人类等价或无法区分的智能。

人工智能概念正式提出是在1956年，在美国达特茅斯学院举办的夏季学术研讨会上，约翰·麦卡锡、马文·闵斯基、克劳德·香农等学者参与讨论“让机器像人一样认知、思考和学习”，这次会议上首次使用了“人工智能”这一术语。因此，业内也一般都认为1956年是人工智能元年。

在过去的六十多年里，人工智能发展跌宕起伏，经历了三次大的浪潮：

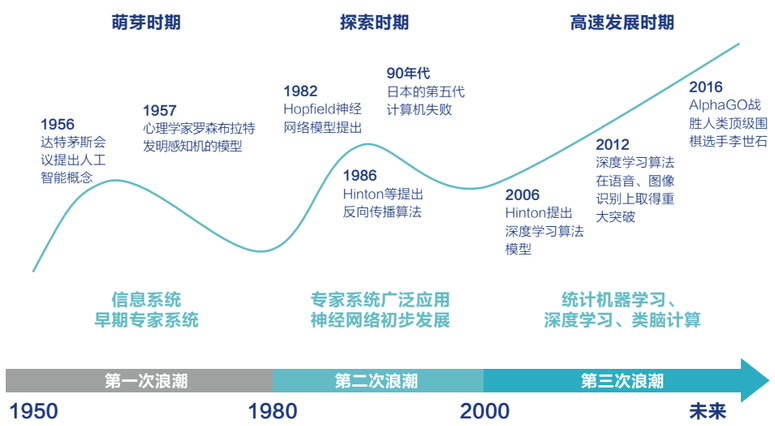


图1 人工智能发展的三次浪潮

第一次浪潮（20世纪50～80年代）：人工智能的起步阶段，期间提出了人工智能的概念，取得了一些突破性的研究成果，如机器定理证明、跳棋程序、LISP编程语言、首个聊天机器人等，但当时的算法理论、计算机的性能等因素，无法支持人工智能应用的推广。

第二次浪潮（20世纪80～90年代）：这阶段主要以专家系统和日本的第五代计算机为代表。专家系统促使人工智能从理论研究走向实际应用，并在医疗、气象、地质等领域取得成功。但随着人工智能应用范围的扩大，专家系统的缺点也逐渐显现：应用领域狭窄、推理方法单一、缺乏常识性知识等，人工智能的发展又进入了停滞状态。在这阶段也出现了神经网络算法，但是由于当时计算机的性能限制，最终也没有较好的落地效果。

第三次浪潮（2000年～现在）：随着信息技术蓬勃发展，为人工智能的发展提供了基础条件。这阶段人工智能的理论算法也在不断的沉淀，以统计机器学习为代表的算法，在互联网、工业等诸多领域取得了较好的应用效果。2006年，多伦多大学Hinton教授提出了深度学习的概念，对多层神经网络模型的一些问题给出了解决方案。标志性事件是在2012年，Hinton课题组参加ImageNet图像识别大赛，以大幅领先对手的成绩取得了冠军，使深度学习引起了学术界和工业界的轰动。近几年，以深度学习为代表的人工智能算法，在图像分类和识别、语音识别、自然语言处理等领域取得了巨大的进步。究其原因，一方面计算机的性能得到了极大的提升，新型人工智能芯片、云计算技术都为大规模神经网络计算提供了基础平台；另一方面是互联网、大数据技术的发展，积累了大量的数据资源。算法、算力和数据三者的结合，直接促成了这次浪潮，将人工智能再次推向繁荣期。

根据人工智能的研究领域、周边技术和涉及的产业，可以将人工智能的技术体系分为三个层次，如图2所示，具体包括：基础层、技术层和应用层。

图2 人工智能技术体系层级

这三个层级的技术，彼此依赖，相互促进。

应用层：人工智能技术与行业深度结合，针对具体的场景来实现智能化的方案，目前主要的应用行业领域包括安防、金融、医疗、交通、教育、制造、互联网、电力等，未来将会拓展到更多的领域。当前，人工智能产品种类也比较多，比如机器人方面，包括家用机器人（扫地、陪伴、教育等用途）、工业机器人等；再如自动驾驶汽车，其中就使用到了大量的人工智能技术，包括通过计算机视觉技术来识别车道线、交通标志、信号灯等，进一步利用人工智能算法进行决策分析，做出正确的动作指令。未来将会有更多的人工智能产品进入生产生活当中。

技术层：产业界和学术界都比较关注的层面。底层包括各种机器学习/深度学习的开源框架等。以学术界为代表，对人工智能的底层理论算法的研究，包括近年来比较主流的深度神经网络算法、传统机器学习算法，正是因为这些基础理论取得突破，才使得当下人工智能技术在产业化方面取得突飞猛进的发展。应用算法层主要的研究领域包括计算机视觉、语音识别、自然语言处理、决策规划等，涉及感知、认知、决策不同的智能方向。在每个研究领域中，又有很多细分技术研究领域，比如计算机视觉领域，包括图像识别、目标跟踪、视频理解、行为分析、图像超分、多维特征识别等等。技术层是人工智能中最为令人关注的，也是最具挑战的，其优劣直接决定了行业应用落地的成效。

基础层：作为人工智能产业的底座支撑，包括硬件、软件和数据的技术支持。硬件主要是为人工智能应用提供强大的算力支撑，包括计算资源如GPU、FPGA、ASIC等加速芯片，网络资源，存储资源，以及各种传感器件；系统平台包括操作系统、云计算平台、大数据平台等；数据资源是人工智能技术（尤其是深度学习）获得长足发展不可或缺的组成部分，犹如为发动机提供充足的“燃料”。

场景化是人工智能商业落地的关键

人工智能技术只有在实践中解决了具体的问题，才能产生价值。因此合适的商业场景是人工智能技术落地的关键。当前人工智能技术主要是以深度学习方法为主，通过大规模数据驱动的机制，挖掘数据中蕴含的潜在规律。这种方法，机器并没有真正的推理和思考的能力，并没有人类所具有的高阶智能，一般只能解决特定领域内的问题。目前取得较好成效的主要在单任务、单领域的视觉感知方面上，有些已经做到了非常极致，甚至超越人类，比如图像识别技术在安防、交通流量监测、闸机身份验证等特定场景中，可以代替人工完成这些重复性的工作，取得了很好的效果。但在认知方面目前效果不尽人意，还达不到像视觉感知领域的效果。随着谷歌BERT等算法的突破，对于自然语言语义的理解和认知方面，也渐有起色。

由于目前人工智能算法机制对数据集的重度依赖，需要有足够的数据，而数据都是在行业场景中积累产生的，比如医疗影像数据、金融交易数据等。因此，将人工智能技术与行业场景结合才能发挥人工智能的价值。并且只有在场景历练通过不断的反馈机制，使数据形成闭环，才能持续不断迭代优化和提升算法精准度。

人工智能将对人类社会产生深远影响

人工智能带来生产效率提升

人工智能对企业变革影响巨大，在未来15年内，人工智能和自动化技术将取代40-50%岗位，同时也带来效率的提升。

例如，在工业制造领域，AI技术将深度赋能工业机器，将会带来生产效率和质量的极大提升。采用AI视觉检测替代工人来识别工件缺陷，带来的益处：

- 识别精度，基于图像数字化，可以达到微米级的精度；
- 无情绪影响，可以长时间保持稳定工作；
- 检测速度，毫秒级就能完成检测任务。

人工智能改变人们的生活方式

随着人工智能技术的普及，人们的居住、健康、出行、教育、娱乐等多方面的生活方式都将从中受益。

智能家居将会是人工智能技术应用的一个重要突破口。未来，智慧家居助理会统筹管理所有智能家居设备，使其协同工作，根据不同的活动场景，为人们营造更加舒适和安全的居住环境。人们不再是通过双手去操作使用各种电器，而是通过更加自然的方式与智慧家居助理交流，轻松地让各种电器完成任务。

医疗也将是人工智能大展身手的领域。AI技术的推广，可以很大程度缓解当下的医疗资源紧缺、医护人员工作强度大等问题，使更多的民众受益。另外，通过健康穿戴设备，监测人们的生理数据，对人们的日常健康状况进行检测管理，做到疾病的提前预防。

人工智能改善人类的生存环境

人工智能在粮食保障、能源利用、气象预测、环境污染、自然资源保护等领域上应用，可有效改善人类生存环境，促进人与自然和谐共生。

农业是人类赖以生存的基础，为人类提供每天所需的食物。据《2019年全球粮食危机报告》显示，全球仍有1亿多人处于重度饥饿状态。自然灾害和气候变化是导致粮食不安全的部分关键因素。人工智能在一定程度上可以改善农业所面临的问题。例如2019年底在全球较大范围内发生的非洲蝗虫自然灾害，造成部分地区粮食大幅减产。有些机构组织开始着手研究如何利用人工智能技术结合卫星遥感地理信息，对类似的自然灾害进行预警，减少农业损失。另外，利用人工智能技术对小地域范围内实时、精准的气象预测，可以指导农业实施过程，在什么时间适合进行播种、施肥、灌溉、采摘等。人工智能还可以用于筛选优良种子，达到粮食增产的目的。

人工智能面临的挑战

正因为人工智能技术能够对人类社会产生巨大效益，国家政策、资本等方面也大力支持，企业积极布局人工智能战略，增加研发投入、加快商业落地。人工智能产业一片向好的景象。但在繁荣的背后，人工智能也面临诸多挑战。据《IDC中国人工智能软件及应用市场半年度研究报告，2019H1》显示，面临的挑战主要有缺乏人工智能技术人员、缺乏高质量数据集、应用场景、成本等多个方面。

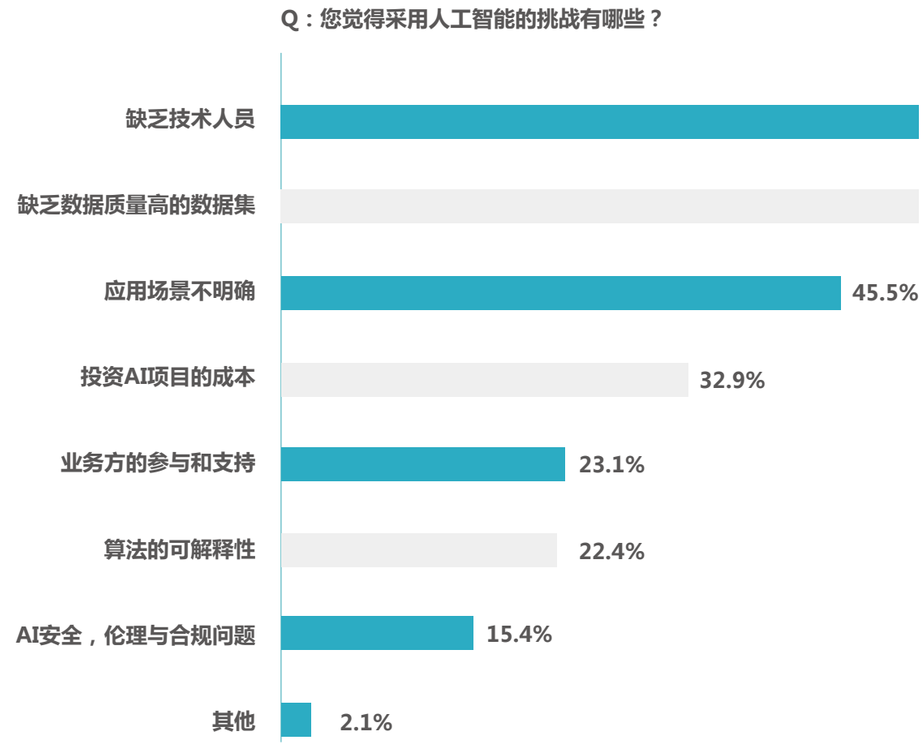


图3《IDC中国人工智能软件及应用市场半年度研究报告，2019H1》市场调研

面对这些挑战，我们应该理性对待，寻找合适的解决方法，打造有利于人工智能健康发展的良好环境。

场景化落地面临的挑战

目前，人工智能商业落地效果比较好的是安防、金融等行业领域，在其他领域的部分场景中，落地效果并不是太理想。究其原因，一方面是安防、金融等落地效果好的领域，都是有良好的数字化基础的，多年来积累了大量有价值的数据，利用人工智能技术来挖掘数据价值自然是水到渠成。另一方面，是对当前人工智能算法所能解决问题的边界没有厘清，与用户期望的有偏差，用户期待的效果，可能当前AI算法还达不到成熟标准，而AI算法能解决问题的场景，还有待进一步挖掘。

对此，建议各行业领域的企业，在实施人工智能应用落地过程中，优先完成数字化改造，积累行业数据，然后再实施合理的智能化业务。

技术方面的挑战

在人工智能技术层面上，也面临一定程度的风险，主要表现在数据和算法上。

当前算法严重依赖有标注的数据

数据在人工智能商业化落地中有着不可替代的作用，目前人工智能算法以有监督的深度学习为主，即需要标注数据对学习结果进行反馈，在大量数据训练下，算法才能取得预期的效果。算法从大量数据中进行学习，挖掘数据中蕴含的规律。数据决定了人工智能模型精度的上限，而算法则是不断逼近这个上限。

高质量数据需求导致数据成本高昂

为了提高数据的质量，原始数据需要经过数据采集、清洗、信息抽取、标注等处理环节。得益于大数据技术的快速发展，当前采集、存储海量数据已经不再是难事。在时间和成本上，数据标注成了制约环节。目前数据标注主要是人工标记为主，机器自动化标注为辅助。但是人工标注数据的效率并不能完全满足算法的需求，研究提升机器自动化标注的精度，是提高效率的重要思路，也是数据标注的一个重要趋势。

数据噪声、数据污染会带来人工智能安全问题

人工智能训练模型时用到的训练数据，如果数据本身有较大的噪声，或者数据受到人为破坏，都可能会导致模型决策出现错误。由于一些客观因素，训练数据中不可避免含有噪声，如果算法模型处理的不得当，可能会导致模型漏洞，模型不够健壮，给黑客有了可乘之机。另外，也存在黑客故意在训练数据中植入恶意数据样本，引起数据分布的改变，导致训练出来的模型决策出现偏差，进而按照黑客的意图来执行。从数据源角度进行攻击，会产生严重的后果。例如在无人驾驶车辆上，会诱使车辆违反交通规则导致事故。

当前深度学习算法有一定局限性

深度学习算法通过构建大规模多层次的神经网络模型，从大量数据中学习经验规则，从而达到拟合复杂的函数来解决实际问题。深度学习模型的学习能力强，效果也非常好，但在实际应用过程中依然面临资源消耗、可解释性、安全等方面的挑战。

模型计算量大，对硬件要求高

深度学习训练的时候需要处理大量的数据，模型单元也会做大量的计算，所以会耗费大量的存储和计算资源，成本高昂。即使是在模型推理阶段，计算量相对较小，但在边缘、端侧部署深度学习模型，仍然需要对模型经过压缩、剪枝等出来，来进一步降低计算量。目前国内很多企业在研究端侧的AI芯片，提升边缘侧的计算能力，相信未来计算力的问题会得到解决。

模型复杂，存在不可解释性

人工智能模型的可解释性，是指人类能够理解机器做出决策原因的程度。由于深度神经网络模型异常复杂，参数量巨大，导致模型成为“黑箱”，我们很难获知模型预测结果的准确原因，也不知道模型会在什么时候或条件下会出错。这就导致了在一些如医疗、无人驾驶等关键场合中，使用深度学习都比较谨慎。当然在学术界，也在积极研究可解释性的人工智能，包括如何改善用户理解、信任与管理人工智能系统。

模型鲁棒性弱，易受对抗攻击

深度神经网络非常容易受到对抗样本的攻击的。一些图像或语音的对抗样本，仅有很轻微的扰动，以至于人类无法察觉这种扰动。但对于模型却很容易觉察并放大这个扰动，进而处理后输出错误的结果。这个问题对于在一些关键场合下危害非常大。对抗与攻击也是深度学习研究领域的一个热点，已经有很多防范攻击的方法来降低风险。

社会规范方面的挑战

人工智能技术是一把双刃剑，一方面能推动社会进步和经济发展，另一方面也会带来法律、隐私保护、伦理等的风险。人工智能技术的运作效率极高，如果被不法分子利用了，发起网络攻击或者窃取机密信息，将会产生巨大的危害。另外，深度学习依赖于数据，在数据采集过程中，不可避免的会收集到用户的一些隐私数据，涉及个人的生活习惯、健康等数据，如果这些数据不加以监管被乱用，势必会造成隐私侵犯。针对这方面风险，国家也在研究应对措施。在《新一代人工智能发展规划》中明确指出，到2025年，我国初步建立人工智能法律法规、伦理规范和政策体系，形成人工智能安全评估和管控能力。在2019年6月，《新一代人工智能治理原则——发展负责任的人工智能》发布，提出了人工智能治理的框架和行动指南。

相信随着技术上的进步，法律、社会规范的出台，人工智能将会朝着安全可靠、公平、保护隐私等正向发展，促进人类福祉。

PART 2

人工智能产业化落地
ICT技术是关键支撑

算力仍然是人工智能的核心支撑

从人工智能概念的提出，半个世纪以来，其发展一直都比较缓慢。究其原因，主要是以前的人工智能应用效果并不理想，不能得到大规模推广和应用。决定着人工智能应用成效的关键因素很多，算力是其中之一，算力对人工智能的发展起到支撑作用。

算力突破推动算法创新，促成第三次AI浪潮

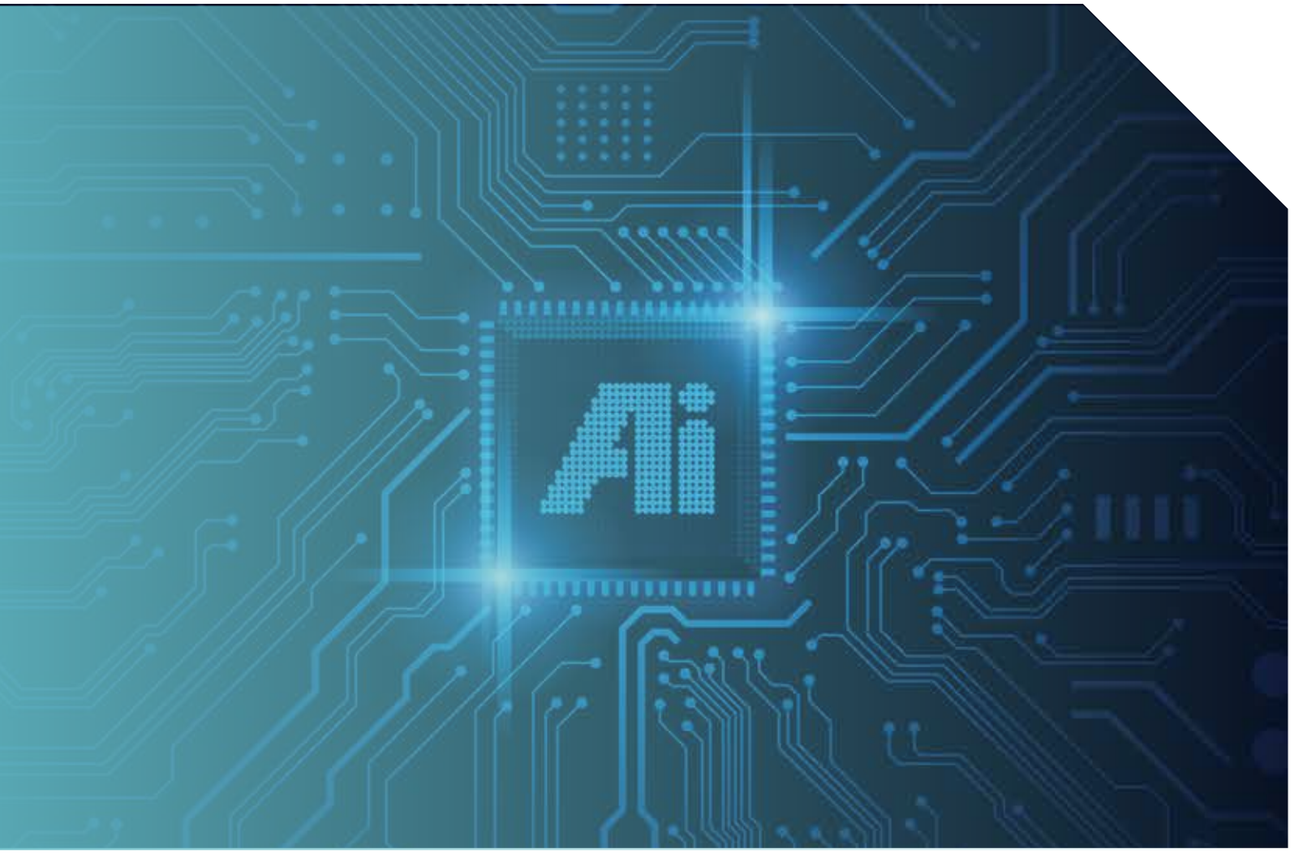
在2012年，Hinton课题组参加ImageNet图像识别大赛，其AlexNet模型以大幅领先对手的成绩取得了当年的冠军，使得深度学习算法一时间轰动整个学术界和工业界。

深度学习算法本质上也是神经网络，早在上世纪80年代就已经诞生。AlexNet模型使用了比以前更加深层的网络，参数量高达千万级，使用了大规模的图像样本进行训练，当然也有一些细节上的算法创新。当时支撑AlexNet模型的实现，是基于两块英伟达GTX 580的GPU，完成了当时CPU难以短时间完成的任务。从此，业内普遍认同了两方面的事实：一方面是神经网络的模型规模增大有助于提升识别效果；另一方面，GPU卡可以提供非常高效的算力，用来支撑大规模神经网络模型的训练。

近几年，业内各厂家意识到算力的重要性，分别推出多种加速卡如GPU、谷歌的TPU等，用于加速人工智能计算，直接推动了人工智能算法飞跃式的创新。从2012年到2018年期间，以计算机视觉为主的感知类智能取得了突飞猛进的发展，有些领域如多维特征识别等，其识别率远远超越了人类水平。在2018年末，谷歌发布的BERT模型，在11项不同的NLP测试取得最佳成绩，直接推动了NLP认知类智能的突破。在这惊人成绩的背后，是强大算力提供的支撑。跟据作者描述，BERT-Large模型是在33亿词量的数据集上训练的，拥有3亿多的参数。试想一下，如果没有能支撑这么大计算量的算力资源，也许很难验证算法的效果，算法创新也就更加不易。

另外，数据的爆发式增长，对算力的依赖也十分强烈。根据IDC报告显示，“数据总量正在以指数形式增长。从2003年的5EB，到2013年的4.4ZB，在2020年将达到44ZB”。面对海量的数据，使用人工智能算法挖掘其中的价值，也必须有强大的算力支撑才能实现，这也直接关系到人工智能应用的创新和发展。

异构计算集群实现算力横向扩展，支撑大规模AI运算



当前这种以深度学习训练算法为主的时期，对算力和数据的需求是惊人的。OpenAI对近年来的模型训练和算力需求做过一个分析总结，自2012年以来，最大规模的AI训练运行中使用的计算量呈指数增长，且翻倍时间为3.4个月，远快于芯片工艺的摩尔定律。

为了支撑巨大的算力需求，一种行之有效的方法就是采用异构计算集群。在人工智能领域中，异构计算是指联合了通用的CPU和面向AI运算加速的GPU/FPGA/ASIC等不同计算体系结构处理器的计算系统。另外，单颗芯片的计算能力是有限的，且随着摩尔定律失效，仅从芯片角度来提升算力相对来说比较困难。业界一般采用计算集群的方式来扩展算力，通过把成千上万颗计算芯片，整合在一个系统中，为人工智能模型的训练和推理应用提供支持。目前，鉴于GPU的通用性、性能和生态等因素，面向人工智能的异构计算集群，仍然以CPU+GPU的方式为主流，但在一些特定应用场景中，CPU+FPGA/ASIC的方式也有一定的优势。

另外，异构计算集群实现算力的扩展，不单是硬件设备上堆砌。由于人工智能特有的计算模式，设计面向人工智能计算的集群需要区别传统通用计算集群，如在进行模型训练的时候，集群计算节点间需要大量且频繁的周期性数据同步等，都是需要考虑的因素。为了提升性能，需要考虑系统软件和计算框架层面上的优化，如何合理的调度AI任务来最大化地利用计算资源。同时也还需要考虑高性能的网络和存储，来保障集群整体性能。

AI芯片“井喷”式发展，满足多样化的算力需求

提升算力的另一条途径，就是从芯片层面去实现。相对于传统程序，AI计算有着明显的特征，导致传统处理器无法满足：当前很大一部分AI应用，处理的是视频、语音、图像等非结构化数据，计算量巨大且多数为矩阵运算，非常适合并行处理；另外，深度学习模型参数量非常多，对存储单元访问的带宽和时延直接决定了其计算的性能。

为此，一方面可以通过不断的改进优化现有计算体系芯片的计算能力，从早期的CPU，到专用于并行加速计算的GPU，以及在特定场景应用的FPGA和ASIC芯片，都是在朝着适应AI计算模式的方向优化，加速AI运算过程。这种方式是目前AI计算加速的主流方式。另一方面可以采用新型计算架构，如类脑芯片、量子计算等，从根本上颠覆现有计算模式。2019年8月，清华大学类脑计算研究中心研制的Tianjic芯片登上了《自然》杂志，展示了类脑芯片的潜力，是未来AI芯片的一个重要方向。

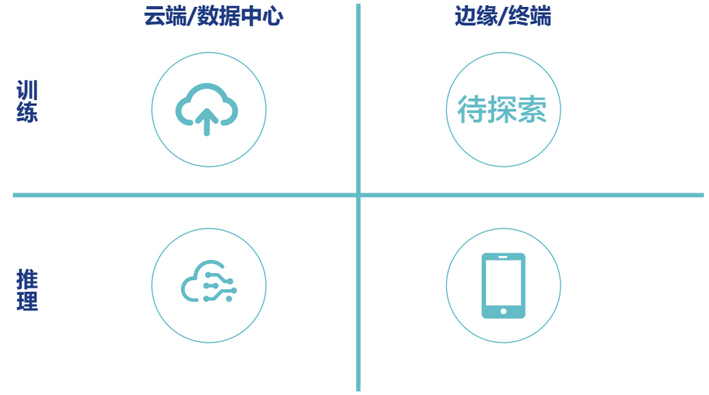


图4 AI加速芯片及应用场景

不同的计算场景对算力的需求特点是有差异的：

- 在云端/数据中心的训练场景中，更多的关注算力的性能、精度、扩展性、通用性、可编程、能耗效率等；
- 在云端/数据中心的推理场景中，对算力考量的侧重于吞吐率、延时、扩展性、能耗效率等；
- 在边缘端的推理场景中，考虑更多的是延时、能效、成本等。

随着市场的强劲需求和国家政策的引导，国内研发AI芯片呈“井喷”趋势，众多厂家加入到了造芯行列当中。针对不同的人工智能应用场景，各个厂家都在打造各具特色的芯片。尤其是随着物联网的普及，端侧应用场景更加繁杂，AI芯片百家争鸣的态势，有助于解决AI多样化的算力需求。

大规模AI训练场景，对网络和存储提出挑战

数据、算法、算力是人们常说的AI发展三要素，必然在AI中起着至关重要的作用。那么除了这些，是否还有其他因素关系其发展呢？我们试想，AI是一辆火车，数据、算法、算力、好比其燃料、发动机，有更多，更好质量的燃料，才能让火车跑的更远，更先进的发动机才能使火车跑的更快。不过，在实际火车运营中，仅仅这些是不够的。火车要在铁轨上运行，也就是有了更好的路，火车才能四通八达、通畅无阻。AI面对实际应用也是如此，其爆棚的数据量和超高的算力要求都不是一台普通的服务器能够完成的，需要大规模的集群，集群中服务器、存储设备间的互联网络就是AI中的“路”，而这些当前的“路”是不能满足大规模AI训练场景需求的。除了“路”之外，火车是用来运输货物或者人，那车厢本身的存储容量以及装卸车的速度也是火车运营的重要指标。对应到AI应用中就是存储容量及数据读写访问技术。

大规模AI训练场景对网络之“路”要求很高，有多方面原因。首先，AI相关业务通常包含大量的图像、视频等非结构化数据，数据量上有一个指数级的增长，需要保证这些海量非结构化数据顺畅、快速通过才能使AI系统平稳运行。其次，AI运算相比以往运算更加复杂，一次智能化业务背后要几百个模型计算，每次计算并非一台服务器能完成的，需要庞大算力和复杂的异构计算，背后实现往往是通过大规模集群并行处理的，那么集群中的服务器快速通信就成为完成一次计算任务的关键要素之一。第三，AI业务很多需要实时学习，算法在框架层和应用层需要保持高精度一致。这些要求都是现存以太网所不具备的，其中千分之一的网络丢包对AI的影响都是巨大的。这个如同以前的马车走土路，压过一块小石头，或许就是有个小颠簸，不会发生什么大问题，但是如果铁轨上有一块小石头，可能就会造成火车的出轨，后果不堪设想。

当前铺设的这条网络“路”主要技术有TCP/IP及以太网，这是最常用的网络传输技术，其优点是应用范围广，成本低，兼容性好，缺点也很大，网络利用率低，传输速率不稳定等。InfiniBand是一个用于高性能计算的网络标准，服务器间、服务器与存储设备间、存储设备之间均可以使用其进行传输。它的优点就是传输性能好，可惜在大规模应用中支持不好，而且需要特定网卡和交换机的支持，成本相对高昂。还有诸如Intel提出的Omni-Path等技术，都是为了优化网络性能，不过均存在各种兼容、成本等问题。

要满足AI的大规模训练需求，我们需要一种综合的网络解决方案，既能广泛大规模使用，价格低廉、成本可控，又能够完成高性能AI计算的需求。这首先要保证网络达到90%以上的带宽有效利用率的同时，网络中无丢包，并确保低时延。通过RoCEv2、Lossless无损网络流控技术综合方案可以实现上述需求。RoCEv2即RoCE（RDMA over Converged Ethernet，基于以太网的远程直接内存访问）的第二个版本，较第一个版本支持跨IP子网的通信能力。该技术主要解决两大问题：

- 通过远程直接的内存访问绕过操作系统内的多次内存拷贝，远程节点的CPU无需介入，降低CPU负载，数据直达对端应用buffer。测试显示数据从CPU到网卡出口时间通过RoCEv2技术可以有效提升8倍，RoCEv2在提高网络吞吐量的同时极大的降低了数据包传输延时。如图5所示，传统TCP/IP与RDMA方式的数据移动对比。

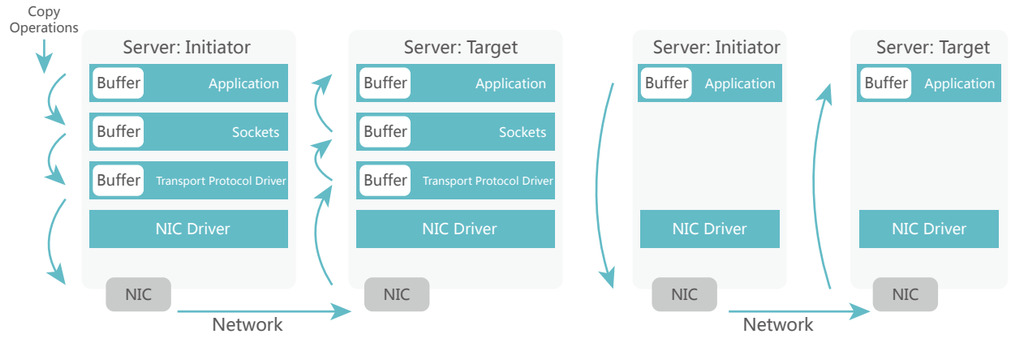


图5 传统TCP/IP与RDMA方式数据移动对比

- RoCEv2是RDMA在以太网上传输的实现，部署时仅两端点需要采用专用的网卡硬件，中途路径采用原有以太网线路及设备即可，相较InfiniBand等技术大大降低了成本。

RoCEv2解决了成本、延时、吞吐等问题，这样还是不够的，上面提到面对大规模AI计算，网络中是不能出现丢包。这就需要Lossless无损网络流控技术来保证。如图6所示，无损网络解决方案部署参考。

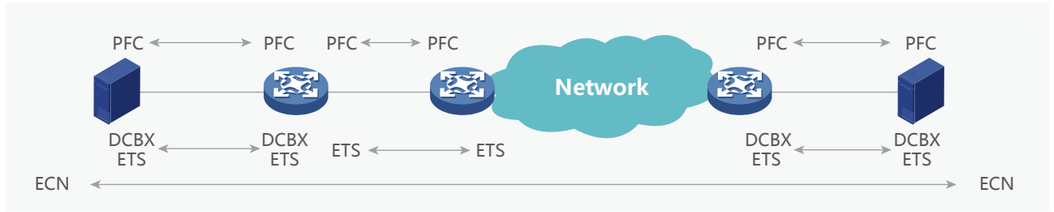


图6 无损网络部署参考

其实现包括如下几个方面：

- 需要支持PFC流控能力，当某一优先级报文发送速率超过接收速率时，通过向上一跳发Pause帧通知上一跳设备暂停发送本优先级报文，实现不丢包机制；
- 开启快速ECN能力，向服务端快速进行通告反压，保证流量将要出现丢包时，快速通知发送端进行降速；
- 用户可选择开通ETS将网络中的流量优先级分成不同的优先级组，为每组分配一定带宽，如果一个组未消耗完为其分配的带宽其他组可以使用这些未使用的带宽，达到资源的合理分配及充分使用；
- 交换机与服务器网卡之间，通过开启LLDP协议的DCBX TLV，其报文中携带ETS/PFC配置状态，实现全网的DCBX能力通告和协商，保证网络无丢包。

综上通过RoCEv2、Lossless无损网络流控技术给网络带来极大的能力提升，为AI的发展提供了一条宽大、平坦、通畅的“路”。

面对AI大规模训练场景除了对“路”提出新的挑战，对存储及其访问也提出了新的挑战。随着NVMe（Non-Volatile Memory Express）和SSD技术的发展，对存储数据的读写访问目前已经有了飞速提高。然而，AI训练模型的精准度对数据集是强依赖的，数据样本越大，训练出来的模型越精准。因此AI对存储带来了一系列新的挑战：

- 通常的一个训练模型需要千万甚至上亿的文件数量，面对这样的海量数据访问，传统分布式文件存储架构（如HDFS，MooseFS等）就显得相形见绌了；
- 很多的训练模型都依赖于图片、音视频片段，为了进行更有效的特征分析，即便是大文件也会被切片成小文件。有些特征文件小到几十、几百字节，也有很多都在几KB到几MB之间。而传统分布式存储是针对大文件设计的，集群容量是其首要考虑的问题，面对AI训练场景，80%以上是小文件，首要解决的是文件系统支持海量小文件的问题；
- 业务部门数据组织存储的不确定性，导致系统管理员不知道数据怎么存储的，很可能将大量文件放在同一个目录节点上，这样在AI进行训练时，会同时读取一批数据，数据所在目录的元数据节点成为“热点”被大量访问，从而导致训练性能出现问题。

这几个问题就如同过去的绿皮车时代，车次少，乘客少，停车时间还长，那么上下车就没什么特别要求，大家慢慢上，慢慢下，反正时间很充裕。而现代高铁时代，车次多，有的地方甚至十五分钟左右一班车，车厢长了，乘客还都满员，每站停车时间几分钟，有些甚至1分钟，这样就要求有合理的上下车次序和分流等手段进行优化。

针对AI对存储访问的特殊应用需求，同样需要针对性的进行优化。如将单点MDS（Metadata server，元数据服务器）进行横向扩展，形成MDS集群。MDS集群可以缓解CPU、内存压力，同时存储更多的元数据信息，并提高海量文件并发访问性能。这点像火车乘车进站以前的一个两个检票口，现在扩充到十个左右，减轻一两个检票口的压力，同时能够一起进出更多的乘客。针对小文件，可进行小文件内联、聚合，客户端读缓存等优化手段。这点可以理解为，老人小孩的，一家人一起提前检票进站。而“热点”访问问题，可采用目录镜像扩展或增加虚拟子目录的方式。同样映射到坐火车场景，可以理解为乘车时点餐服务。以前是大家都到餐车排队购买，现在是将二维码都贴到每个座位上，自己使用手机扫码就可以点餐，到时乘务员会按照座位把餐送来。

综上，我们可以看到，真正的AI时代，不仅仅是其三要素数据、算法、算力技术发展就能满足的，同时对AI的运行环境也提出了更多挑战。当前是把AI效能发挥最大的一系列技术共同发展的时代，而非仅AI技术本身，相关技术要合力前行。无论是网络还是存储技术应走到更前面，在全球产业智能化转型中充当开路者的重要角色，为AI提供更顺畅的运行环境。

云边端协同，满足多样化的AI应用场景

多样化的应用场景对云端AI提出挑战

云计算的核心依靠云端超强的计算能力来完成计算要求很高的任务。进入云计算时代，由于云计算在成本、效益、规模、自动化和集中性等方面给企业带来的好处，大量人工智能服务完全部署在云上或者在很大程度上依赖于云。与此同时，随着物联网等技术的不断发展、数据的不断增加，如何在数据从生成到决策再到执行的整个过程中，保持尽可能小的延迟，就显得尤为关键。在一个只有“云”的世界中，数据可能要传输几千甚至上万公里，较大的延迟是在所难免的。

对于一些时延敏感的人工智能应用场景，如自动驾驶汽车，对实时性要求极高，纯粹依靠云端的能力是难以满足的。另外，一些数据敏感的场景中，将数据上传到云端进行智能计算，也会面临一定程度的风险。云端服务在这些人工智能场景中的应用效果大打折扣，而边缘计算则可以有效解决这一问题。

智能下沉是对云端AI能力的延伸

边缘计算作为云计算的延伸拓展，是一种分布式处理和存储的体系结构，它更接近数据的源头。它是将计算任务从数据中心迁移到靠近数据源的边缘设备上，因此它更擅长处理实时性、安全性要求较高的计算任务。基于边缘计算的方式，大大降低了网络延迟，处理数据更加快速，支持企业更快更好的做出决策。

图7 边缘计算模型

在人工智能应用场景中，将一些重量级的AI训练任务，或者对时延不敏感的任务，放置在云上进行，而将一些轻量级、或者对时延敏感、或者对数据安全有要求的AI计算任务，下沉到边缘设备或者终端设备中执行，通过边缘、终端和云端协同来实现快速决策、实时响应。在万物智联时代，只有云、边、端紧密协同工作，才能更好地满足各种AI应用场景的需求，从而最大化AI的价值。

云边端协同将进一步拓展AI应用边界

云边端协同工作将成为人工智能应用部署的重要方式，可以满足云端AI短板，即时延或数据安全等方面，为支持更多有严苛要求的AI应用场景铺平道路，提升应用效果。

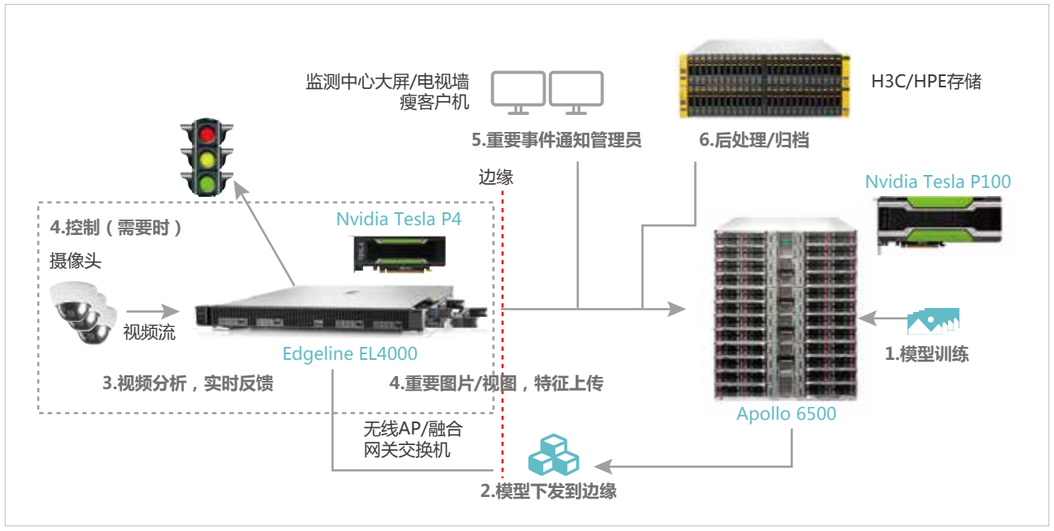


图8 云边协同的智能安防应用

在智慧安防场景中，传统方式下需要将大量摄像终端采集到的视频数据，通过网络直接传输至云端或服务器进行存储和处理，不仅加重了网络的负载，也难以满足业务低时延快速响应的需求。通过增加边缘计算节点，将摄像采集终端采集的数据汇聚到边缘节点，从而有效降低网络传输压力和业务端到端时延。此外，智慧安防与人工智能相结合，在边缘计算节点上搭载AI人工智能视频分析模块，面向智能安防、智慧安防、轨迹跟踪、多维特征识别等AI典型业务场景，以低时延、大带宽、快速响应等特性弥补当前基于云端AI的视频分析中产生的时延大、用户体验较差的问题，实现本地分析、快速处理、实时响应。

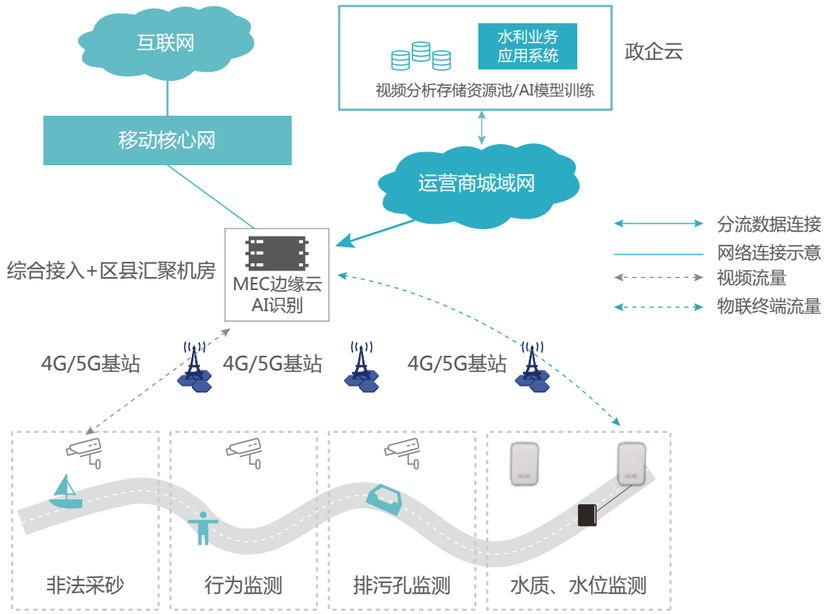


图9 云边协同的智能水利应用

在智慧水利场景中，5G、智慧安防、边缘云和AI分析紧密结合，可以智能的识别出水利业务中的异常场景（河道漂浮物、钓鱼、游泳、非法采砂等），做到无人值守，实时告警。采用边缘计算（MEC）的网络结构在河道附近部署无线摄像头，在运营商本地机房部署MEC平台。实时性要求高的业务部署在边缘云，其他业务部署在中心云，实现云边协同。视频流经MEC分流后，将流量进行本地化分流，在本地完成AI智能分析，实施将告警信息上送中心云。在本地进行业务流量的分流和处理，不仅提高了响应速度，而且减轻对运营商核心网络的数据传输压力。

在智能家庭场景中，边缘计算节点通过各种异构接口就近汇聚、存储和处理边缘节点上的各类异构数据，执行AI任务，对敏感数据就地处理，不出本地，有力地保护数据隐私，同时将处理后的非敏感数据统一上传到云平台。用户不仅仅可以通过网络连接边缘计算节点，对家庭终端进行智能控制，还可以通过访问云端，对过往非敏感数据进行访问。

在智慧交通场景中，汽车作为边缘计算节点，通过集成的采集装置采集实时数据，并与路侧边缘节点进行交互。边缘计算节点进行视频的就地处理和识别，将识别的车辆和位置信息通过5G等通信手段回传到云计算中心。云计算中心通过大数据和人工智能算法，为边缘节点、交通信号系统和车辆下发合理的调度指令，从而提高交通系统的运行效率，最大限度的减少道路拥堵。

人工智能应用普及，安全备受关注

人工智能技术发展迅速，目前在各行各业的应用已经日益普及，但人工智能系统和技术自身的安全风险也越来越成为不能回避和不可忽视的风险，甚至在某些场景下还会带来很大的问题。

图10 人工智能系统自身面临的安全风险

人工智能应用导致用户隐私泄露风险

目前人工智能在智能手机、办公设备、智能家居上的应用越来越多，很多人家里都有了智能音箱，另外不少电视、冰箱、电饭煲、空调、窗帘等都具备了人工智能的功能，人们使用语音或者手势就可以指挥它们帮人们完成查询天气预报、查找信息，甚至烧饭做菜，调节室内环境等。由于这些智能设备为了随时响应主人的召唤，需要实时在线，加上其日益强大和不断升级的语音、图像和视频的感知、认知能力，有可能对主人家里每个人的一举一动了如指掌，用户在享受了人工智能带来的便捷服务的同时也带来了自己和家庭隐私泄露的隐患。

人工智能平台和模型泄密风险

人工智能平台和模型泄密风险主要有：模型窃取攻击和用户数据窃取攻击。

机器学习模型窃取攻击

指的是攻击者基于反复查询并分析人工智能系统的输入、输出参数和其它外部信息，从而推测和猜测出系统的模型参数、训练参数和训练数据等信息。目前很多云服务商提供了AI即服务（AIaaS），由AI服务商负责

模型训练、识别等服务，对公众开放，用户可使用开放接口进行各种人工智能识别等操作。但通过反复调用AIaaS的识别接口，有经验的攻击者就可能通过多次返回的信息从而还原出AI模型的各种参数等关键特性，从而把AI模型窃取到。或者即使不能完全窃取到原模型，也可以通过窃取到的信息构建机器学习的对抗样本或模型，从而对人工智能系统进行下一步更深层次的攻击。

用户数据隐私窃取攻击

在用户提供训练数据的情况下，攻击者可能通过反复查询训练好的机器学习模型，获取到用户的隐私数据。

人工智能系统输入数据真实性风险

当前的人工智能模型和算法非常依赖于输入数据的真实性、完整性和全面性。从攻击者视角，恶意的数据注入是进行对抗样本攻击的重要手段。数据真实性风险主要体现在训练数据真实性和判断数据真实性两个方面。

训练数据真实性问题

攻击者在训练数据中掺入的恶意数据，可能会大大影响机器学习模型训练的有效性，降低人工智能模型的推理能力。例如，研究者发现，只需要在训练样本中掺杂少量的恶意样本（药饵攻击），就能很大程度感染AI模型的准确率。通过加入药饵数据，在人工智能健康数据库应用中，攻击者可以使模型对超过一半的患者的用药量建议阐述超过四分之三的变化量。

判断数据真实性问题

在机器模型的判断阶段，对被判断数据样本加入少量噪音，即可能大幅改变判断结果的准确性，甚至出现风马牛不相及的结果。比如著名人工智能科学家Ian Goodfellow曾发布论文，通过图像生动阐述了基于判读数据投毒的对抗样本攻击概念，一张原本是熊猫的图片，在加入了少量干扰白噪声后，人眼看还是熊猫，但机器学习模型直接将其识别为长臂猿，且可信度高达99.3%。

人工智能基础设施安全风险

人工智能的基础软硬件的安全风险

包括TPU等AI专用芯片，GPU，CPU，FPGA，还有大到AI计算服务器集群，小到我们的智能手机、终端，都可能存在软硬件设计缺陷、安全漏洞、后门。例如处理器硬件的安全风险，可能很多人并不陌生，如2018年全球最大处理器生产商英特尔爆出的Meltdown漏洞，该漏洞被认为是史上最严重的处理器漏洞之一，本质上是英特尔处理器的预测执行技术设计缺陷，但由于预测执行读取的数据防护不当，破坏了位于用户和操作系统之间的基本隔离，从而可能允许恶意代码访问主机任意内存，进而窃取其他应用程序以及操作系统内核的敏感信息。这个漏洞“熔化”了由硬件来实现的安全边界。允许低权限用户级别的应用程序“越界”访问系统级的内存，从而造成数据泄露。而且漏洞修复会不可避免地造成处理器性能的降低。另外，研究人员发现，在芯片制造过程中也可植入后门，或者硬件木马。攻击者只需要通过短时间在处理器上运行一系列看上去非常安全的命令，就能够地触发处理器的某个隐藏逻辑，从而获得操作系统的高级权限。而更加让人担心的是，这种非常微小的硬件后门基本无法通过任何硬件检测和安全分析手段检测出来，并且可能只需要芯片工厂中的某位普通员工就能完成此项任务。至于软件设计、编码过程中由于不小心、不遵守设计和编程规范等，无心埋入的软件Bug，甚至别有用心的软件后门的植入，一直都是软件开发和应用全生命周期中需要解决的重大课题，在人工智能软件系统中也不例外。而且由于人工智能系统的黑盒性和不可解释性，使得软件后门更难以被检测。

人工智能算法和人工智能框架的设计缺陷、安全漏洞

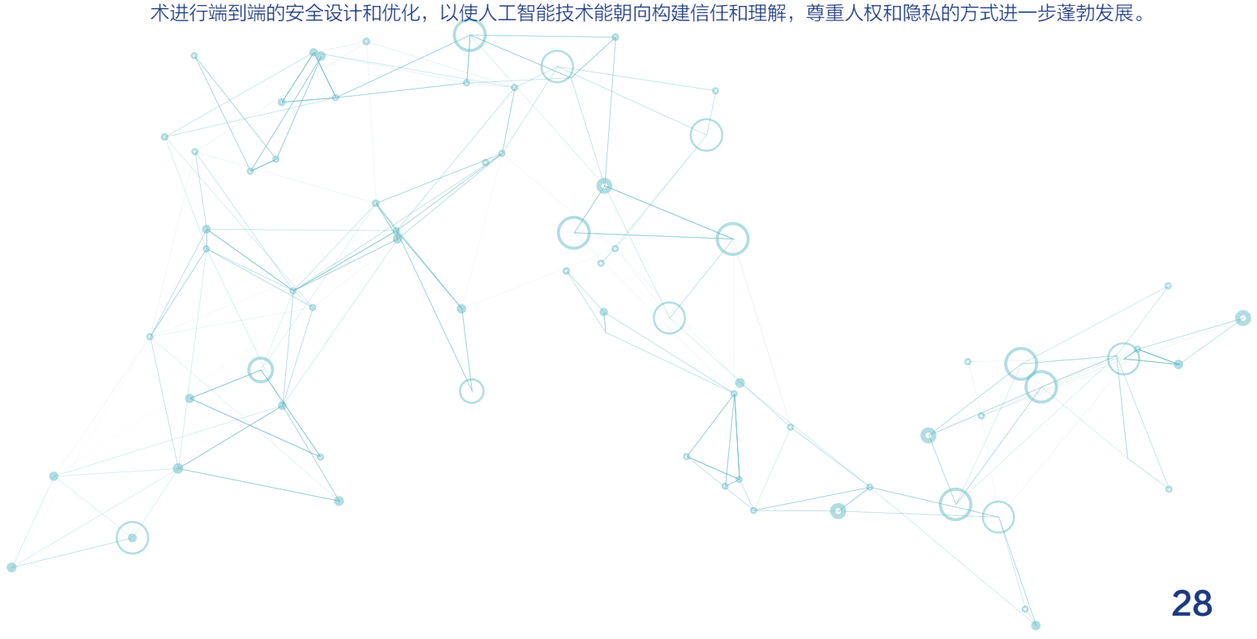
腾讯安全平台部预研团队曾发现某著名人工智能系统框架存在自身安全风险，可被黑客利用，生成恶意模型文件，对使用该框架和平台的人工智能研究者进行攻击，受害者自身的人工智能应用可能被窃取或恶意篡改、破坏。该漏洞危害面较大，一方面攻击成本低，不需要太高深的人工智能技术能力，普通攻击者即可实施攻击；另一方面迷惑性强，使用该平台的大部分人工智能研究者可能毫无防备；同时因为利用了该框架自身的跨平台机制，其在PC端和移动端版本均会受到影响。

人工智能架构和流程的安全风险

人工智能架构、操作模式和运作流程设计的不合理。比较典型的例子有，去年某著名快递企业的快递柜，被人发现使用用户的照片就可以轻松通过其多维特征识别系统的安全验证，从而取走物品；目前还有一些企业的无接触考勤系统也未能基于三维特征来进行识别，也存在类似问题，这种由于各种原因导致的架构或工作流程设计缺陷使得人工智能系统的安全性存在漏洞，容易被不法分子利用。

另外，AI模型的可检测性、可验证性、可解释性普遍不足，在目前AI应用优势领域的语音、图像、棋类竞技类场景，可解释性差可能问题不大，因为结果一般是可以快速取得并且显而易见的，只要AI系统识别的结果是好的，人们可以忍受它继续以黑盒形式存在。但对于有些场景，不可解释性则会带来一些法律上或者业务逻辑上的关键风险。例如在银行给用户发放贷款前的AI评估系统中，如果AI模型无法给出做出相应判断的依据和来龙去脉，那就无法获得用户的充分信任，如果连其深层次的判断原理和规则都无法得知，该系统也就很难说是一个安全的系统。

综上可见，人工智能技术是一把双刃剑，用好了可以造福人类，而如果用不好，甚至被恶意利用，也会给个人、企业、社会甚至国家的安全带来危害。未来我们需要更多地从基础技术到顶层设计上，从AI应用的全流程上考虑，对人工智能系统和技术进行端到端的安全设计和优化，以使人工智能技术能朝向构建信任和理解，尊重人权和隐私的方式进一步蓬勃发展。



PART 3

ICT产业与人工智能协同发展，共创智慧生态

新华三在ICT领域中深耕多年，面对新兴的人工智能技术，从内部产品赋能，到外部智慧能力输出，新华三一直在不断探索和实践。

在新华三内部，从无接触考勤到生产制造环节，都已经广泛应用到了人工智能技术，同时，在自研的产品里也集成了大量的人工智能技术，提升产品智能化程度。在对外智慧能力输出，除了能提供人工智能计算的软、硬件基础设施，也拥有完善的人工智能行业解决方案。

AI基础设施

智能计算平台

目前，新华三拥有丰富的人工智能基础产品，从底层的GPU服务器、网络、存储等硬件，到上层的公共科学计算AIOS软件平台，可以为客户提供软硬件一体化的高性能人工智能解决方案，帮助客户快速实现人工智能应用的落地。

如图11所示，搭配新华三R系列GPU服务器与RDMA网络设备，可获得卓越的计算性能，可以满足训练、推理等多种人工智能计算场景的需求。

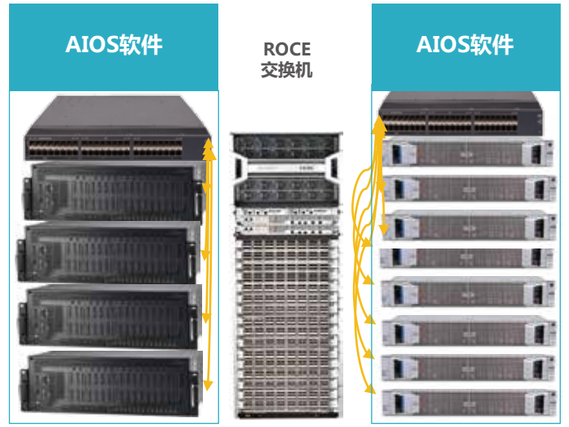


图11 新华三高性能人工智能平台方案

同时，新华三提供的AIOS开发平台，从基础硬件的部署和软件安装，到交互式开发环境的一键启动，从模型的深度训练和调优，到多机多卡GPU作业灵活调度，提供了十分简洁的使用方式，实现了资源的整合、弹性扩容缩容和合理调度。同时，AIOS软件平台提供丰富的可自定义的软件和镜像以及二次开发的API接口，既可以独立部署，也可以方便的集成到用户原有的SaaS平台。

图12 一站式AI开发平台

AIOS平台为用户提供了强大的全流程可视化管理平台，大大加速了用户在开发人工智能解决方案时的速度。

- **统一的集群管理：**负责整个系统计算资源的集中管理、统一分配与作业调度，包括GPU资源池的集中管理与分配、多租户方式隔离计算资源、以作业方式动态分配计算资源以及计算资源回收等。
- **统一的监测运维：**实时监测管理集群资源使用情况和集群状态，包括作业状态、GPU使用率、集群健康度等，并分析每一类的资源占用情况，提供触发预警机制。
- **统一的开发环境：**提供一站式的交互开发操作界面，帮助用户完成模型脚本在线编辑、模型训练、模型验证以及模型推理等核心功能，并结合硬件资源可视化、作业调度器，最大化提高系统硬件资源的利用率。

场景及成效

天津大学智能与计算学部综合平台建设与服务项目，是学部教学科研的基础设施，旨在全面支撑人工智能、大数据处理分析、高性能计算等相关的教学实验、科学研究工作。主要功能包括：

- 提供与业界接轨的实验环境，为人工智能、高性能计算、大数据等方面的专业实践课程提供实验环境，为各类专业比赛提供实训环境。
- 为学部各类科研项目提供的数据存储和计算分析服务，有能力支持国家科技重大专项、国家自然基金、以及省部级以上的各类项目。
- 为校企结合的前沿科学研究和工程项目提供支持，促进产学研转化，充分发挥平台的示范辐射效应。

目标是建成国内一流的人工智能、计算机科学人才培养的综合实验平台，取得高水平的产学研成果。

- 该方案整体上是以HPC高性能计算集群、AIOS集群（AI深度学习推理、演算的科研计算）、虚拟化集群等为主要组成部分，以云平台统一纳管的技术方案。能随需调整并可持续扩展，能为日常运行提供可靠、安全的保障，保证数据及业务安全。
- 高性能计算集群、GPU集群、虚拟化集群和云平台通过高速万兆光纤网络互联，HPC集群支持批处理模式的科研计算，GPU集群支持AI推理、演算的科研计算，云平台的虚拟机用于前期的程序调试，以及小规模实验教学。整体集群支持多样化的计算需求，优化资源利用。

新华三以此方案，顺利地帮助客户实现平台落地，并展开相应的科研工作。

智能网络方案

人工智能与网络应用相结合的探索早已有之，在网络应用的众多子领域中，智能运维就是备受关注的研究方向之一。智能运维系统就是通过传统运维系统的智能化升级，更好地管理海量AI基础设施，保障AI应用的稳定运行。传统运维模式解决故障需花费IT人员40%时间，且大于85%的故障是投诉后才发现。由于网络问题源头难以识别，跨域问题定界困难，问题根因定位复杂，导致运维人员90%时间都在定位问题。人工智能技术是协助解决数字化转型中企业IT运维管理所面临挑战的一个利器。2017年Gartner提出AIOps（Artificial Intelligence for IT Operations）的概念，也即智能运维，并预计到2022年，所有大型企业中，40%的企业将会结合人工智能技术来支持和部分取代现在的监测、服务台以及自动化流程和任务。

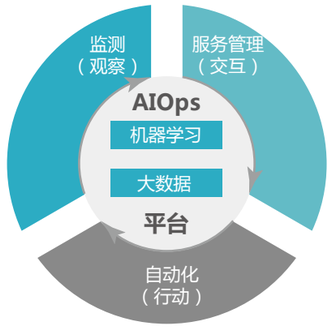


图13 Gartner给出的智能运维含义

智能运维的目标是基于已有的运维数据（日志、监测信息、应用信息等），借助AI技术，通过主动发现、动态数据分析和预测，增强ICT的技术能力，实现所维护产品的更高质量、合理成本及高效服务。

先知网络架构（SeerNetwork Architecture，简称SNA）是新华三公司推出的下一代智能网络架构，用户只需在单个界面，即可完成网络设计、仿真、部署、保障等全部运维管理的过程，帮助用户拉通数据中心、园区、广域网等不同业务场景的网络解决方案，实现网络建设全生命周期的AI赋能与端到端的业务保障。SNA主要体现在融合和智能两大方面：

融合

- 场景融合，数据中心、园区、广域网跨场景统一编排，统一入口、统一策略、统一管理，用户体验升级；
- 技术融合，设计、编排、控制、保障形成全生命周期的网络闭环管理，降低管理复杂度。

智能

- Telemetry，更精细、更实时的网络洞察能力；
- SeerAnalyzer（先知分析器），大规模、多维度、细粒度的数据采集，深度的网络分析；
- 数据中心、园区、广域网，端到端的业务可视与保障能力。

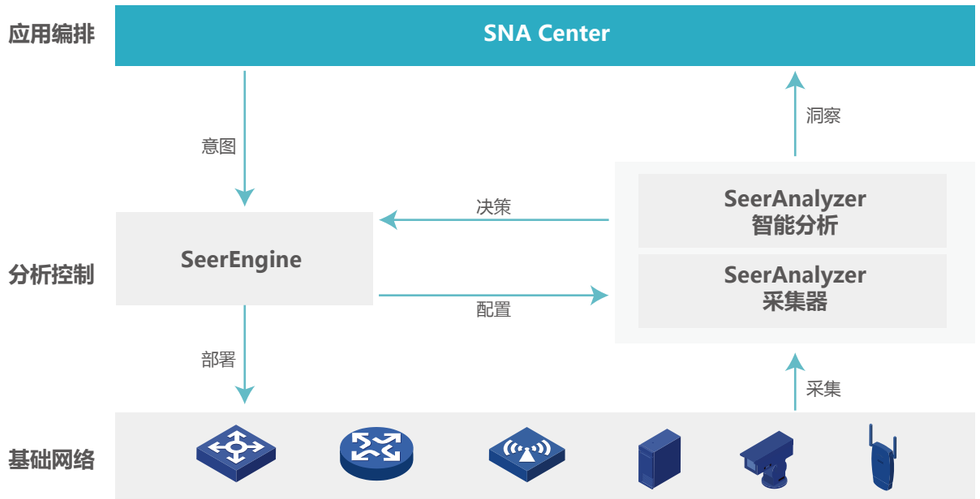


图14 新华三SNA先知网络架构

SNA先知网络架构，由以下部分组成：

SNA Center

先知网络中心（SNA Center）是先知网络架构的核心组件，它具有对网络进行全局统一管理、控制和智能分析、业务编排的能力。SNA Center拥有全局视角，能够解决不同网络管理域之间的协同问题，从而简化运维、降低运营成本；同时SNA Center可以实时感知网络状态，基于用户意图或网络状态分析可以实现业务自动化部署和风险预测，让网络更加智能，能够更加简洁、智慧、高效的为业务服务。SNA Center与SeerEngine控制器和SeerAnalyzer分析器配合来实现上述功能。

SNA Center通过北向API接口，实现网络服务能力的承接。通过南向API，实现与控制器的对接。SNA Center的主要工作是创建模型及数据集训练调优，因此是一套软硬件结合的算力系统。

支撑SNA Center的关键技术包括新华三AI平台软件AIOS、高效计算集群以及低时延高带宽无损网络等。

SeerEngine控制器

实现数据中心解决方案（AD-DC），广域网解决方案（AD-WAN），园区解决方案（AD-Campus）等多个场景的控制器平台统一，提升网络部署的自动化能力。

- SeerEngine-DC是面向数据中心方案的控制器，可管理基于Fabric架构的数据中心网络，提供Underlay设备自动化上线、可视化运维及灵活的业务部署。提供了统一的控制平台、高可靠的管理平台和良好的可扩展平台。
- SeerEngine-WAN是面向广域网方案的控制器，负责广域网领域的业务自动化、网络调优、控制策略下发等功能。
- SeerEngine-Campus是SNA框架下开发的园区网络智能控制系统，实现网络与业务的完美结合。支持基于VXLAN的不同规模的园区网络，提供包括园区网络的自动化上线、业务端到端自动化部署、接入用户名址绑定、有线无线一体化管理等一系列功能。整体架构基于SNA架构思想之上，是园区网络运维的有力助手。

SeerAnalyzer分析器

SeerAnalyzer作为SNA的智能分析引擎，借助大数据和人工智能技术完成网络数据的采集、存储、分析及应用功能。

SeerAnalyze的主要功能包括数据采集、数据分析（关联分析、根因分析、应用分析、流量分析）、精细可视、业务保障（通过经验、根因分析、形成决策的依据、自动部署联动）等。同时也提供了丰富的应用和数据API接口，方便与第三方的网络应用对接集成。

SeerAnalyzer已经用于健康度评估、故障感知及其分析预测等方面的实践，比如在光模块故障以及链路流量的智能预测上：

- 使用分布式采集器实时采集全网设备接口带宽利用率及光模块状态数据，经过数据清洗和标准化后统一存入数据仓库；
- 利用机器学习算法对接口带宽利用率及光模块状态历史数据进行持续学习，预测未来一段时间的流量趋势及潜在的光模块故障，为人工干预提供依据。

企业大脑

方案介绍

企业数字化转型需要企业将现有的业务流程进行数字化改造、智能化改造。新华三在ICT领域长期深耕，深刻理解数字化转型的意义，不断在企业内部进行探索和实践，形成了完善的企业大脑解决方案，主要包含企业智能运营中心、远程工作接入和现场工作环境。

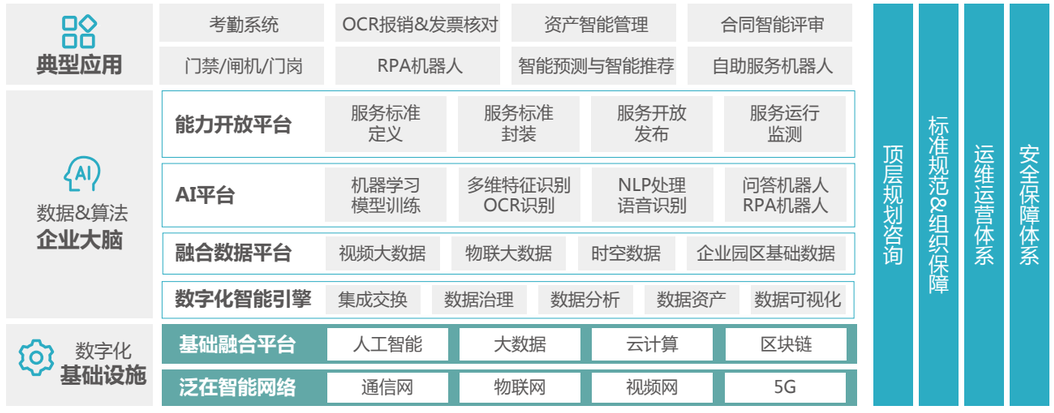


图15 企业大脑方案

智能运营中心——AI中台和数据中台

AI中台

以H3C AIOS为基础建设的AI中台，可以实现OCR识别、多维特征识别、语音识别、智能预测模型、NLP、RPA机器人等多种基础能力集。

数据中台

以H3C DataEngine为基础进行数据中台建设，通过数据中台对数据进行加工后，快速响应业务及运营，通过数据中台为AI中台提供数据基础。

- 构建不同业务场景的主题集，统一各应用基础数据及业务查询，提升准确性及接入效率；
- 通过智能算法及数据支撑，实现备件及库存的精准预测，优化库存结构，降低业务库存成本；
- 通过中台实现历史数据归档，节约成本，提升业务系统运行效率；
- 对业务数据通过数据中台及NLP技术提供智能的数据问答场景。

远程工作

云桌面

基于新华三自研的桌面传输协议VDP开发的VDI服务端、软件客户端以及多种硬件终端，支持大规模远程部署虚拟云桌面，为每个用户提供专属虚拟机，提供远程工作环境，支持立体声、低延时、高音质音频；保障音视频同步流畅播放；支持云端定时/手动备份，故障快速恢复；支持终端出现异常快速还原。

多种安全机制

支持桌面显性水印&盲水印，可通过多种方式显示，如果信息外泄可溯源找到泄密源头；支持独立控制外设通道，灵活实现信息安全。

支持云桌面补丁及时更新，引擎及病毒库及时更新，防止黑名单软件运行；各类操作有迹可循，违规事项可溯源；掌控虚拟终端敏感数据分布、及时阻断信息外泄。

支持准入控制，确保从接入请求到进入企业内网接入安全性。支持通讯安全多重安全通讯机制，为端到端的传输提供安全保障；支持多类VPN通道，支持多种安全认证方式；支持桌面连接协议加密。

云端安全通过SSMS服务器安全监测系统，从被动防御转为主动防御，为用户提供云桌面主机安全保障：支持主动防御，安全监测系统，提供云桌面主机安全保障；支持资产清点、病毒查杀、风险发现、安全基线、入侵检测、安全日志六大功能。

现场工作

入园管控防范潜在风险，风险不入园

人员入园时自动识别是否佩戴口罩，单画面可同时识别5人。多维特征分析+多维特征识别+人员轨迹分析，还原园区活动轨迹，根据账号、MAC地址和姓名查找人员轨迹。做到事事可溯源。车辆入园时自动进行车牌识别分析，白名单自动放行，平台统一汇总数据，确保可溯源。

无接触考勤

避免交叉感染，预防疫情传播。

热门路径重点部署

采用大数据，对人员路径数据采集并实时分析，对热门路径重点消毒及部署人员巡查；实时呈现人员迁徙路径云图变化，对迁徙路径进行跟踪；对热门路线动态展示，辅助决策热门路径周期性重点消毒及园区安保力量布置。

人员聚集管控无线大数据实时分析人员密度，及时发现、及时预警

基于无线接入大数据，对无线接入密度分析，分析人流密度变化，及时发现人员聚集情况，超过阈值及时告警。

入园告警针对重点人员布控，及时告警入园风险

针对重点区域事前录入信息，实时布控；一旦发现风险人员，立即联动客户端、APP生成告警。

事后追溯，轨迹回放

提供事后追溯功能，可以回放人员行动轨迹，快速查找密切接触者。基于轨迹查询人员活动地点、逗留时间及密切接触者。

智能办公——移动办公

为实现员工灵活的移动端办公需求，提供了常用场景的应用入口，包括移动考勤、内部通知公告、移动邮箱、流程审批、通讯录查询等，方便员工随时随地接入。手机移动端部分办公APP如下，左图为办公功能，右图为考勤功能：

图16 移动端办公APP界面

办公业务智能化

在新华三企业内部，已经在多种办公场景的应用中，实践了人工智能赋能。

OCR发票识别

基于AI中台提供的基础能力，构建的OCR票据识别应用，在企业内部广泛使用。目前可以支持增值税发票、住宿发票等10多种票据信息的自动识别和提取，极大地提升了办公效率。

自动问答机器人

基于自然语言处理和知识图谱技术，构建了多种不同用途的问答机器人，节约人力成本的同时，也提高客户体验。目前主要服务于HR及行政业务咨询、市场业务及流程自助咨询、IT自助问答等场景，未来还将拓展到运维服务等更多场景中。

智能检索

利用自然语言处理及推荐技术，实现合同智能评审，自动检测出合同各条目风险指数，以及标明需要人工修正条目。还有智能简历筛选，可以自动根据招聘岗位要求，匹配出最适合的简历。

电子标签

智能资产管理，通过智能电子标签提高库存的准确性和效率以及信息透明度，为企业提供资产评估、决策提供更为可靠的依据，避免企事业单位在资产管理环节上可能造成的隐患。

智能制造，通过智能电子标签实现设备电子化管理，对生产物流的优化实现生产过程的优化与产品质量的可追溯。

适用场景及成效

企业研发办公

为研发体系员工提供了VPN移动开发、测试和办公解决方案，保障了研发办公在特殊时期下的全面开展，同时支持管理员设定数据访问权限，并具有屏幕显性水印和盲水印两种模式，确保敏感数据安全不外泄。

使用家庭电脑安装新华三VDI客户端软件，通过云桌面终端统一访问实验环境时，如同多人在一个物理实验室上班。可将共用文件传输至云盘以供进行下载使用，配合云桌面的安全策略，全方位防止信息泄露。并通过自研的VDP协议对音视频进行重新编码压缩，保障在更低带宽下享受更好使用体验。

其他人员场景

引入RPA机器人技术辅助解决重复性的业务操作及日常工作，通过RPA技术实现IBPMS平台。引入NLP技术实现智能知识检索，智能简历检索，智能合同评审，智能客服机器人等场景。

应用成效

四大成效：稳定的远程办公环境，可靠的远程办公保障，事前风险预防，事后及时追溯。

总体上，方案支持上万人的接入能力，超过1000台桌面云服务器、100台存储设备运营能力，超过PB级的数据管理能力，支持图片、自然语言、OCR等算法，且具备开放的兼容性与扩展性。

通过云桌面终端统一访问实验环境时，如同多人在一个物理实验室上班。针对不同专业需求，定制化选择服务器性能，并采用VDI架构为员工分配足够性能，方便员工进行自主学习及编译代码，保障云桌面完美支持专业软件运行，真正实现研发无忧。移动办公门户为业务的办理，会议及培训，信息的发布与查看提供了更易用，更便捷的使用体验。通过终端准入控制，拒绝非法身份、隔离不合规身份，基于身份动态授权，保障远程办公安全接入，提供8000名研发远程加密接入服务，实现开发、测试远程协同。

支撑部门通过RPA机器人实施，每年可以节约5000+人天；通过OCR技术实现，解决发票录入，等效4000+人天工作量。通过NLP技术引入到智能简历与智能合同评审领域，预计可节约搜索简历过程3000人天、合同评审800+人天。还支撑了销售体系大规模的线上培训，为员工线上培训、合作伙伴知识赋能。

工业互联网

方案介绍

工业互联网作为新一轮科技革命和产业变革的重要驱动力量，是新一代网络信息技术与制造业深度融合的产物，借助大数据和人工智能等新技术必将推动形成全新的工业生产制造和服务体系。

新华三在ICT领域耕耘多年，也较早参与了工业互联网的开发和落地，结合机器视觉的优势，新华三发布了《工业品外观智能检测方案》，包括：边缘计算平台、IaaS层、工业PaaS层、视觉业务应用层。

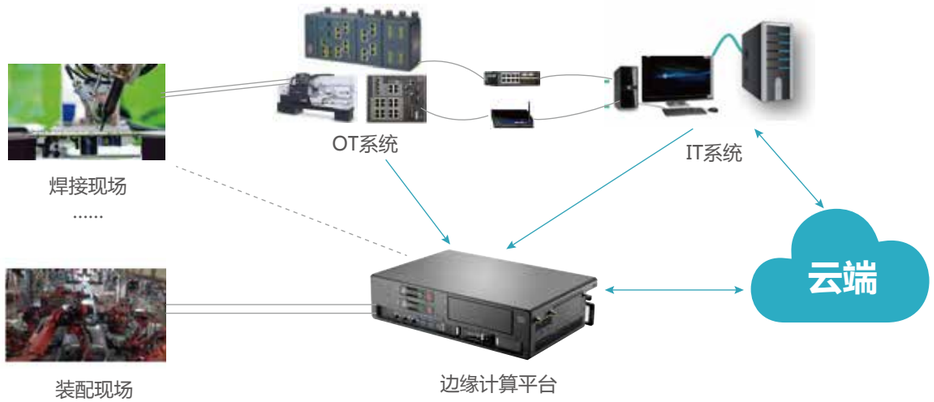


图17 工业品外观智能检测方案

边缘计算平台

就近部署在工业现场，进行产品外观产品的识别和处理；是AI算法的主要承载体；还可以通过云边协同，方便从云端更新最新模型等。

IaaS层

作为基础设施主要提供联接、计算和存储等能力，通过引入虚拟化和Docker技术，逻辑隔离各类业务应用，使应用具有更大的弹性，也更方便与工业云平台对接。

工业PaaS层

作为抽象和能力提供层对边缘前端和各类数据进行清理和管理，还提供丰富的算法库和模型。

视觉业务应用层

包含丰富的组件，通过图形方式，用户可以方便地以"拖拉"方式快速构建客户的应用。本方案采用的边缘智能计算架构，引入的AI技术，针对复杂的工业现场进行了技术专门开发，具有准确高、稳定、可扩展和方便部署等优点。

适用场景

传统产品外观检测主要依赖“人眼+简单工具”，实现对产品外观的瑕疵进行识别，从中选出其外观有缺陷的产品，效率低下，漏检率高。随着AI技术快速发展，其中机器视觉具有准确率高，稳定和快速等特点，逐步应用在产品质量检测等环节，新华三的《工业品外观智能检测方案》通过视觉智能检测技术可以快速应用于工业品外观检测。

本方案既适合刚性产品外观检测，也适合部分形变较小的产品外观检测等场景。刚性外观类产品质量检测有：钢材、主承外观等；部分容易产生变形的产品质量检测有：纸张、标签等。

钢材缺陷检测

钢材的种类繁多、形态各异，给质量检测带来了很多难度，当前主要采取肉眼和工具探伤等手段，存在劳动强度大、检测准确率不稳定等问题。在视觉检测方面，新华三联合某钢铁公司共同攻关，凭借对钢材材质和AI的理解，不断丰富《工业品外观智能检测方案》。通过对钢材表面的缺陷、位置等信息进行学习并识别，优化后的算法可有效筛选出在常规、高反光或高亮度等场景下很难识别的缺陷，具有识别速度快、准确率高、通用性强，可以解决各种不同形状、大小、材质的建材在生产过程中遇到的视觉难题。

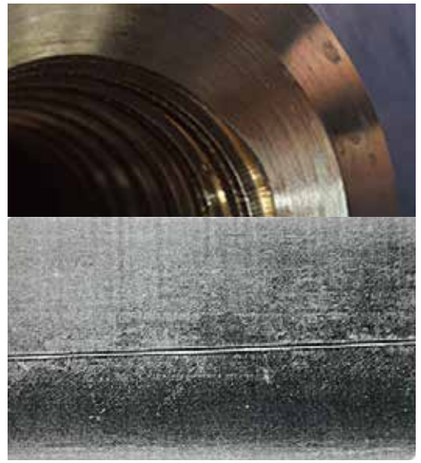


图18 钢材检测

钢材的生产流程和工艺复杂，常会出现结疤、裂纹、划痕、色斑等缺陷，人工检测工作量和难度都很大。新华三检测平台只需将有缺陷的表面或横截面的标注，经过迭代训练，把相应模型部署到指定工序中就可以替代人工检测了，还可以根据阈值对产品残缺程度的评分，灵活控制误检率和漏检率，极大地降低了人力物力成本。其速度相当于一个普通工人的4~6倍，还可以7*24小时工作，释放更多劳动力。

生产车间组装

作用于机器人，可以辅助机械手臂拥有3D视觉能力，靠视觉导引、定位，成为夹取物件的要件。除了视觉定位，通过视频扫描、分析还可以轻松判断出装配的效果。

电子焊接制造

在焊线技术中，因为芯片维度的缩小，需要较强大的影像放大功能。在此环境中，高质量的成像镜头系统必须满足特殊的最佳化需求。本方案具有操作简便、可靠度及视觉算法的高准确度，从而很好地解决了芯片焊接过程中的诸多问题。

产品自动化分拣

自动化分拣是工业生产、特别是产品批量生产中的必需环节之一。工业生产中需要根据产品特性及其生产/出厂质量要求进行分拣，它可以代替人工进行货物的分类、搬运和装卸工作，提高生产和工作效率，从而实现自动化、智能化、无人化。

药品质量检测

在医疗领域中，传统的药品包装、药瓶、标签等基本上通过人眼完成，通过本方案，可以轻松完成对图像信息的采集、存储、管理、处理及传输等功能，把不符合的筛选出来，进而完成药品生产环节的质量检测。

图19 药品质量检测

应用成效

从实践效果来看，本方案的产品检测率高达100%，超越人工检测的精度，完全能代替人工检测和识别部分，从而达到节省人力的目标。

商业价值

本方案引入边缘计算和AI技术，将AI应用延伸到工厂现场侧，AI的训练可以采用成本低廉的公有云；通过对传统工业的智能化改造，开创新的商业共赢模式，同时还可以推广到其他领域。

经济效益

以CPU贴膜检测项目为例，商业价值估算如下：原人工检测：20万+/年，通过本方案一次性设备投资10万，云服务按需付费，可以节省50%的检测投资；同时还提升产品检测效率和正确率，节省产品缺陷类引起的负资产等。

社会价值

本方案完全可以替代人工检测，进而减少了由于人工检测引发的眼睛疲劳职业病，解放员工“眼睛”，减少和预防职业病。

社区安防

方案介绍

智慧社区建设中，安防作为社区安全的第一道屏障，是全社区安全守护者，十分重要。当前智慧社区安防建设正在从“传统”向“技防”转型，为社区百姓提供安全与便利，从而推动和谐社区建设。新华三结合人工智能技术和强大ICT基础能力打造的新一代智慧社区安防解决方案。

新华三智慧社区安防方案整体结构如图20所示：

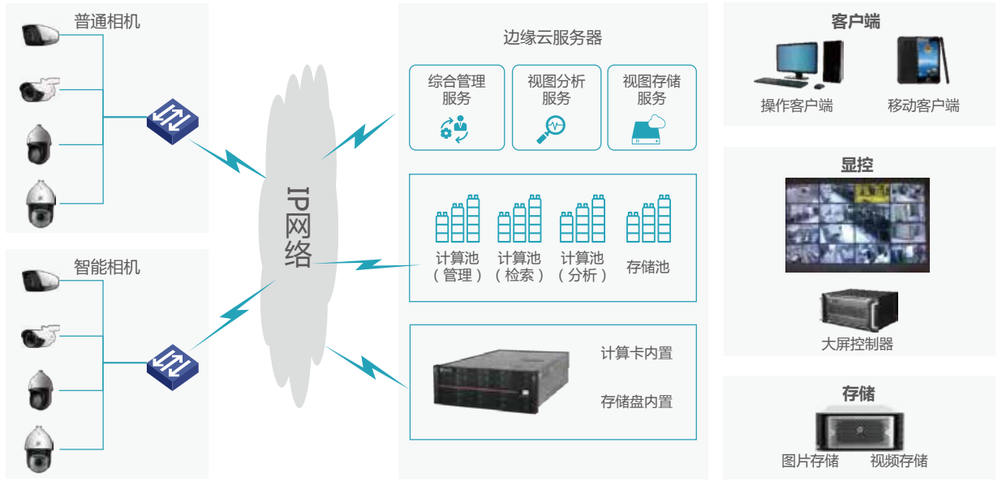


图20 新华三智慧社区安防方案整体结构

智能增点，优化布局

构建“设置科学、布控高效、感知精准、主动预警”的前端智能感知布局。以小区主干道和小区出入口为第二级防控圈，重点关注人脸轨迹、车辆轨迹、摩托车轨迹等线索，做到有迹可循，有据可查。以各单元楼出入口为一级防控圈，通过人脸相机和人脸门禁部署，避免惯偷惯犯各楼间游窜，黑名单报警。以外围商业区和周围车站为第三级防控圈，重点关注全局治安态势，以及打架斗殴和人员聚集等事件，同时是人员控制抓捕圈。

另外，通过部署软件定义摄像头，将富余的算力共享给周边已部署的普通摄像头，通过1台主设备带动周边N台普通设备（1拖N方案），完成利旧改造，推进传统小区前端高清化和智能化改造。

全面云化，逐级联网

制定云化架构标准，多云协同，资源共享高效处理海量数据，为警务大数据的研判赋能。

- **资源高共享：**物理分散、逻辑统一，省厅统一管理，市区两位一体，真正实现资源解耦与标准协议协同；
- **计算高效率：**资源灵活调度，应对并行计算；任务弹性调度，潮汐缓急有别。整体资源利用率提升30%；
- **存储高性能：**分布式云存储架构实现视图及数据存储的高性能、高可用、易运维。

严格遵循《治安防控系统技术规范指南》统一标准规范，实现小区社区汇聚采集、数据逐级级联。

应用驱动，主导实战

深化实有人口、档案管理、治安防控等应用功能，做到社区治安管理易用、好用。多类数据研判分析模型，治安态势智能预警。多功能移动警务助力基层民警工作效能。

- 可通过智能数据采集系统实现小区实有人口、实有车辆、实有房屋、实有安防设施、实有装备力量、实有警情事件的采集；
- 具备智能安防、人脸门禁、车辆卡口等前端采集设备，并可通过物联网关利旧接入小区前端设备；
- 系统通过人员、车辆、房屋等信息的关联，实现全息档案一键检索；
- 通过疑似迁入迁出实现流口管理；对重点人员进行大数据模型分析预警；通过系统平台掌握社区安全、警情、重点人员等治安态势等。

	视图应用	一标六实管理		治安防控分析	全息档案管理	数据可视化	监督考核
地市（区县）	视频查阅	实有人口	实有力量装备	重点对象管理	一人一档	数据融合展示	工作轨迹
	人车布控	实有房屋	实有安防设施	异常研判分析	一车一档	治安态势展示	积分排行
	人车轨迹	实有单位	实有警情事件		一屋一档	巡防路线展示	

	接触警中心	事件处理	信息综合查询	社区定义管理	社区可视化
社区警务室	告警处理	事件查询	模糊搜索	网格小区管理	数据展示
	分派策略	事件审核	全息查询	门牌楼管理	警情展示

图21 实战应用平台

互联网+，便民利民

移动公众号提升居民参与度，社区隐患主动上报，落实社会治安群防群治。

适用场景

新华三智慧社区安防解决方案适用于家庭监测、社区监测、全息管理场景：

- 家庭监测：防盗报警、防火防气报警、紧急求助；
- 社区监测：无线关爱、社区安防监测、社区卡扣监测、人脸门禁；
- 全息管理：人、车、物登记管理。

应用成效

在重庆两江新区康庄美地公租房试点联合创新项目中，新华三联合紫光华智与两江新区分局，以视频云为核心，以社区网格管理为基础，实现了智安社区平台，立体预防和控制，视频大数据平台和数据精准服务等多级智能应用的部署，有效地改善了社区治理和防控。

政务服务

方案介绍

新华三智慧政务方案，通过营造一个数据资源流动、融合的基础环境，打造面向未来的“智慧政务体系”。

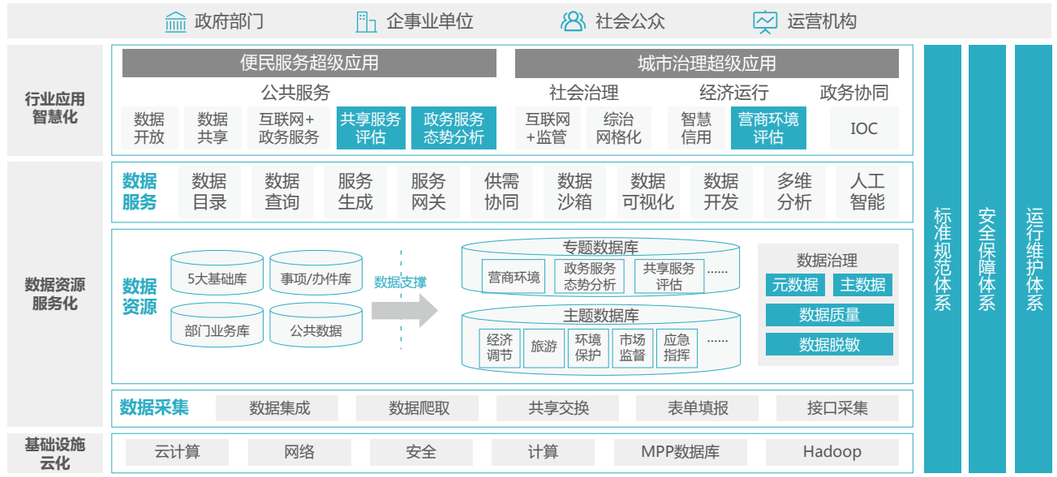


图22 智慧政务方案整体架构

方案整体架构中，主要包括基础设施云化、数据资源服务化、行业应用智慧化模块组成，同时还配套有标准规范、安全保障、运维保障等体系。

- **基础设施云化：**在政务云基础架构建设初具成效后，后续重点在业务上云、数据共享、应用服务创新等方向上；
- **数据资源服务化：**汇聚公共数据、行业数据、互联网数据等非政务数据，接入政务数据，通过数据管理和数据治理，打造业务及基础数据等融合数据产品，承载政务重点领域智慧应用；
- **行业应用智慧化：**通过数据共享/开放服务，对外提供公共服务，形成服务于民生、综合治理、营商环境评估等各方面的智慧类应用，提升政务效率。

下面重点介绍几个典型的政务智能化应用案例。

智慧市政

市政服务效率提升是AI在城市管理中的重要应用场景，针对如垃圾桶、井盖、路灯为治理对象，实现垃圾清运、井盖管理、照明控制智慧化，同时依托路灯杆合并建设应急呼叫系统、信息发布系统。

图23 智慧市政场景

垃圾桶管理

垃圾桶内置的传感器会检测桶内垃圾高度。工作人员可以设置垃圾桶垃圾高度的阈值，当垃圾堆满到达设定高度时，会通知监测中心垃圾箱已满的消息事件，然后工作人员通过电脑或手机上的Web工具能实时查看到管辖区域内所有垃圾桶的状态，从而通过该系统能帮清洁车规划好回收垃圾的最优路径。

井盖管理

随着社区化进程的进一步加快，市政公用设施建设发展迅速。市政、电力、通信等部门有大量市政设备、资产需要管理。其中窨井盖成为了不可忽视的一项。大量在外井盖由于缺乏有效的实时监测管理手段，给不法分子提供了可乘之机，移动、偷盗井盖等违法行为时有发生，同时，破损、损坏、丢失的井盖也因无法及时获知而得不到及时修复，这样不仅影响了相关设备的正常工作，造成巨大的直接或间接经济损失，而且丢失井盖的井口也会对道路上的车辆、行人造成极大的危害，对社会安定、安全造成了极大负面影响。

路灯管理

社区照明包括道路照明（路灯）和景观照明（夜景），不仅是重要的社区基础设施，更是体现社区建设水平和社区形象的重要标志。其中路灯是重要载体，实现物联网控制和智能化管理势在必行。

随着信息化建设的深入，传统路灯照明已经表现出诸多不利因素：

- 手动、光控、钟控易受天气和季节影响，造成资源浪费；
- 无法远程控制路灯开关；
- 人工巡检调度能力差，费时费力；
- 设备故障难定位；
- 无法监测路灯运行状态，依赖人工巡视和市民投诉，不能实时、准确、全面监测路灯。

为了解决以上问题，推进社区智能化进展，新华三推出了智慧路灯的概念。可以对传统路灯进行改造或者新建智慧路灯。

图24 路灯管理解决方案

路灯远程控制：通过对单个、多个路灯进行开关控制，亮度控制（分级调光、传感器按需调光、自动调光），解决目前路灯粗放管理造成的电力资源浪费。

应急广播：在地质灾害、堵车等紧急状况发生时，通过灯杆上的扬声器进行应急疏导。

信息展示：在重大活动、节日时，通过灯杆上的屏幕展示城市风貌、节庆画面。

智慧环保

智慧环保以环境治理和生态保护为核心目标，是AI提升城市管理效率的典型应用，方案中会针对空气、水质、土壤、噪声、污染源（重点排污企业）、水面漂浮物、地面垃圾为监测对象。实现环保态势实时感知、环保事件实时上报、环保治理线上调度。

图25 智慧环保场景

水质监测

智能城市水质综合解决方案以水质在线生物毒性预警技术为基础，设计遵从“一条流程、多层把关、多参数选择、多方式实现”的整体原则，针对水源地、进厂水、出厂水、管网水、终端水等各个关口，进行实时针对性的可选多参数检测。实现了从“水源地”到“水龙头”的全流程监测。

运用4G/NB-IoT/5G/mMTC通讯技术与数据专家分析系统相结合，实现水质数据监测、预警，和突发性水污染事故处理系统，形成饮用水水质监测、预警和应急的综合技术体系，及时发现水体污染情况和扩散趋势。

空气监测

空气监测以气体测量模块为基础，结合空气质量自动监测站以及其他测量方式，将监测数据汇总至环保局的空气监测网，针对环境管理部门的需求，建立大气环境质量在线监测系统平台，可实时监测空气环境质量，实现在线数据查询及统计报表、在线数据自动报警、电子地图与环保信息综合发布等。

噪声监测

城市噪声对于居民的干扰和危害日益严重，已经成为城市环境的一大公害，形势不容乐观，噪音治理势在必行。目前城市噪声主要有：工业噪音、建筑噪音、交通噪音、生活噪音等。一些违规噪音的排放缺乏监管，如建筑工地、"三厅"和餐饮企业超国家法律规定的时限（每日22:00-次日6:00）排放噪声，使得噪声监管方面存在量大、分散、随机和取证不容易等难点。为了使噪声环境能得到有效的监测，需要建立稳定、可靠的声环境在线监管系统。

噪声监测综合解决方案是将噪声污染源的状态利用传感技术、通信技术、GIS技术、信息业务处理系统有机结合而构成的新型环境噪声监测和管理系统。系统将以GIS技术和信息业务处理系统为核心，利用在线监测终端设备，把相关噪声数据传送到服务器，实现对噪声跟踪监测，并根据远程检测终端返回的数据，安排执法工作人员对出现的噪声参数超标等不正常情况及时做出相应的处理，并根据噪声的历史数据，为管理部门决策提供相关支持。

适用场景

新华三智慧城管解决方案适用于市政、环保、交通等城市管理场景：

- **市政：**智能垃圾桶、智能井盖、智慧路灯杆、智慧消防、智慧管网；
- **环保：**空气检测、水系监测、噪音监测、土壤监测、排放监测、围棋监测；
- **交通：**智能红绿灯、AR导航、车联网、无线充电、辅助驾驶、自动驾驶；

应用成效

在成都锦里智慧垃圾桶项目中基于新华三智能垃圾桶解决方案实现了垃圾溢满告警和垃圾清运路线智能规划，保障了景区的干净卫生、减轻了环卫工人的压力。

医院管理

方案介绍

智慧医院解决方案，是新华三结合人工智能技术和强大ICT基础能力打造的新一代智慧医疗解决方案。本方案的建设目标是，打造就诊便捷化、自助化，医疗过程高效化、透明化，管理标准化、数据化的智慧医疗解决方案。

该方案整体框架如图26所示：

图26 新华三智慧医院解决方案1+1+N整体框架

1+1+N整体方案，表示一体化运维体系+统一安全管理+N个智慧应用。

智慧医疗应用系统

以新华三人工智能、共享数据、数据分析、物联网和云计算等平台为基础建设的智慧医疗应用系统，为医院和患者提供智能的医疗辅助诊断、智慧门诊、智慧病区、智慧后勤、智慧医联体等多种服务，提高医疗效率和准确率，提升患者体验。

智慧医院信息系统

基于H3C DataEngine作为数据中台，通过数据中台对数据进行加工，快速响应智慧医疗业务及运营需求，并通过数据中台为AI中台提供数据基础。基础数据的互联互通，院内业务之间、医院之间、医院与患者以及医院与各级机构的数据共享；数据不仅共享出去，还可以拿进来，比如患者的生活习惯、职业特点、运动数据、甚至包括真系亲属的信息等，真正实现让“信息多跑路、患者少走路”。

智慧医院数据分析决策平台

智慧医院数据分析决策平台基于人工智能、基础数据处理和分析结果，对管理者提供可视化的运营管理、决策支持辅助；对医护人员提供全面的病患信息和病程进展、危重病人的基于历史数据的多基础疾病关联分析、输出高效率的辅助诊疗建议；对患者提供更人性化、精细化的精准医疗诊疗服务、个性化治疗方案。人工智能新技术大幅提升医生、护士的劳动强度和患者的就医体验。

一体化运维和信息安全

从运维和安全顶层设计上就通盘考虑，从医疗计算、网络、存储系统各种基础设施防护、网络防护、终端安全、准入控制上确保信息安全和系统运行的可靠性。

适用场景

新华三智慧医院解决方案适用于各类智慧医院建设、数字化和智能化改造场景。

应用成效

新华三智慧医院解决方案已经在全国700多家三级甲等医院，包括顶级100强医院的80多家落地应用。以下是无锡市人民医院智慧医院项目情况介绍。

项目背景

无锡市人民医院现有3.0T核磁共振、炫速双源CT等大型先进医疗影像设备40余台，日均接近3000的影像检查量和的诊断处理需求。

解决方案

采用新华三人工智能智慧医疗报告辅助系统，通过图像处理引擎实时计算处理，帮助医生高效完成患者的影像诊断，降低了医生工作强度。

客户价值

提高医生工作效率，增加医院收入，减少患者等待和诊疗时间，提升就诊体验。

智慧校园

方案介绍

新华三智慧校园方案建设的核心理念是“以人为本，以数据为中心，以流程为基础，以服务为目标，全面服务学校教、管、研”。

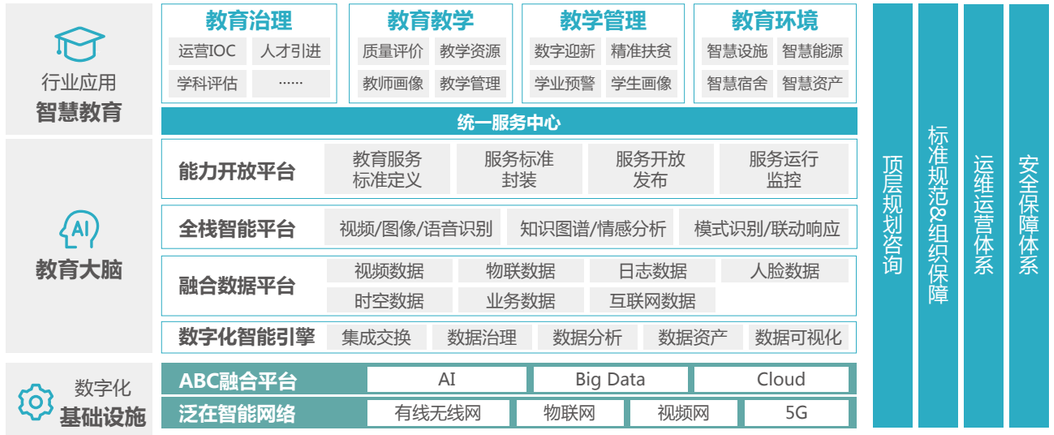


图27 新华三智慧校园整体方案1+1+N整体框架

新华三智慧校园方案的整体框架为“2+4+4+N”的建设体系，其中：

- **“2”**指的是数字化基础设施，包含校园的泛在智能网络和ABC融合平台。校园泛在智能网络实现有线无线网、物联网、视频网、5G网络的构建；ABC融合平台实现AI、大数据、云计算的计算能力构建。
- **“4”**指的是教育数字大脑，分别包含数字化智能引擎、融合数据平台、全栈智能平台、能力开放平台，实现数据采集、治理、标准化、开放共享。
- **“4”**指的是智慧校园顶层规划咨询、标准规范体系、运维运营体系和安全保障体系。
- **“N”**指的是服务于教育行业教学、科研、管理、服务等各个方面的智慧应用。

以“N”里面的智慧教学系统为例，其核心功能架构如图28所示。智慧课堂包含教师端和学生端两大组成部分，分别可以支持PC、平板、智能手机接入。含盖课前、课中、课后的互动，以及各类基本信息的管理。同时结合学校业务系统、以及互联网数据的支撑，经过逐层数据清洗后，提供“招、知、学、管、评、就”一体化的智慧教学人工智能应用：

- **“招”**即招生分析，可辅助学校全面分析生源数据，深入挖掘生源质量；
- **“知”**即专业认知，明确专业特点和树立目标意识；
- **“学”**即智慧课堂，基于互联网+的实时课堂互动设计，贯穿课前、课中、课后各个环节，随时随地构建学习氛围；
- **“管”**即教学管理，通过智慧课堂全面拉通各方面学习过程数据，实现教育过程管理“可视化”，增强了管理状态的可监测性；
- **“评”**即全面的评价体系，科学的评价分析，辅助提升教学的针对性，全面优化教学的整体水平；
- **“就”**即精准就业，辅助正确制定培养方向，全面提升学生综合素质，提供精准就业服务。

图28 “招、知、学、管、评、就”人才培养体系

智慧招生

生源结构分析，全面呈现新生生源地、性别、政治面貌、民族、输出高中等各方面基本信息，以及计划招生人数、实际招生人数、实际报人数、录取率、报到率等关键指标，辅助招生处及时掌握最新招生动态和关键指标完成情况。

生源质量分析，通过对生源地、招生类别、招生批次、招生专业、高收费专业等多维度群体进行立体建模，从而深入挖掘不同新生群体的志愿满足率、最低过线分、最高过线分、平均分差、基础成绩等关键指标，辅助招生处及时掌握多维度的、准确的生源质量数据，并能为学校长远的招生规划提供数据支撑。

学业发展分析，全面跟踪各类生源入学后的延期毕业情况、留级情况、挂科情况、四六级考试情况、获得奖学金情况、研究生录取情况等信息，以及关联学生最后的毕业流向、就业方向。在学校掌握各方面生源信息后，可进一步通过此模块关联分析各生源类别的发展趋势，为学校进一步优化招生计划、培养方案、就业指导等相关工作提供核心的数据支撑。

智慧培养

通过爬虫获取招聘网站中各类岗位的社会需求信息，结合学校专业设置情况形成专业岗位库，通过人工智能技术来分析专业所对应岗位的地域分布，薪资分析，行业趋势等，再根据学生个人特点和意愿，个性化定制培养方案，为学生的个性化成长提供智慧指导。

采集学校历年来毕业的优秀学生数据，展示其成长轨迹、获奖情况等各方面信息，以榜样的力量渲染学习氛围，引导学生正确制定学习计划，提升学生的积极性。

智慧教学

为促进现代信息技术与教育的深度融合，新华三将构建全新的教室学习环境来提高学习质量和效率。以高度现代化软件、硬件的结合，为智慧教学提供实施场所，促进课堂教学交互开展，实现教学全过程数据统一管理。

智慧教室用最便捷有效的手段将人工智能、智慧课堂系统、直播录播等先进的技术，结智能互联黑板、智能课桌、智能讲台等智能硬件，深度融入到教学场景中，为教学过程提供智能化的信息支持，为学生和教师提供全新的教学体验。

对智慧课堂考勤、课堂交互、作业考试等相关数据进行采集分析，及时了解教学情况，辅助课程设置和教学改进。

与精品在线开放课程运营平台进行无缝对接，实现资源、用户、数据等的全面同步，实现推广线上线下混合教学模式，支撑时时可学、处处可学的校园泛在学习环境建设。

智慧管理

为了提升学校的管理水平和办学效率，可以采用“管理化+服务化”的思路，从教学、生活等多方面入手，借助于大数据和人工智能技术，实现学校各类资源的整合和配置优化。

教学方面，基于院系、专业、课程、班级等多个维度生成学情周报，帮助学校及时掌握学情综合概况。综合分析学生复习情况、习题情况、以及课堂表现，提供学生整体学业预警。综合分析学生各门课程的课堂表现，全面绘制学生画像。帮助学校了解教学质量、教学评价等各方面情况。

生活方面，统筹建设校园一卡通系统、智能楼宇系统、后勤保障系统等，并实现数据共享，业务互联，辅助进行人工智能分析。

智慧评价

智慧教学体系的质量评价系统，通过教学质量评价、督导评价等模块实现教学评价数据的全面采集，从而构建更加科学、更加多元化的教学评价体系。

教学质量评价系统，除通过传统问卷方式等进行评价数据采集外，还可通过学生上课的姿势识别、上课专注度、互动频率等信息进行人工智能算法分析，从而让教师、学生以及管理者得到及时的反馈，从而尽快地改善教学过程。

督导管理系统贯穿从督导工作的安排管理、督导开展过程中的评价、督导结束后的综合分析、以及最后的结果公示等各环节，能够辅助学校全面的、发展性的开展督导工作，更快速地采集、处理和公开督导数据。推动教学质量的优化提升。

智慧就业

基于学生在校的学习情况，从专业技能、沟通、文档读写、获得证书、实习经历等方面，构建学生的能力模型。

通过分词解析技术对互联网招聘数据的深度挖掘分析，全面分析学校某专业所对应的社会岗位以及各岗位所需技能和技能熟练度等信息，构建社会需求模型。通过两个模型进行智能匹配分析，计算学生能力与社会需求的匹配度，针对学生的就业提供针对性改进意见，提供精准就业服务，全面保障学生就业，使学生学有所用，为社会输出高质量专业人才。

适用场景

新华三智慧教育解决方案，主要服务高校教育。

从人才培养的社会需求入手，构建专业技能知识图谱，指导相关专业培养方案建设与课程设置。根据学生画像，个性化定制培养方案，辅助教师提前规划整体教学思路。

将网络课堂与实体课堂深度打通，支持以线上线下混合为核心的教育教学资源建设，推进教学模式改革，建设教育资源共享平台；收集教学全过程数据进行科学的评价分析，辅助提升教学的针对性，全面优化教学的整体水平。

根据学生的能力模型与就业意向，个性化推荐就业方向，提供精准就业服务，向社会输送高质量专业人才。

应用成效

新华三智慧校园方案，已经在众多高校取得显著成效。

实践案例：清华大学校园云及毕业生画像

服务校园信息化

- 新华三全融合校园云解决方案完整落地清华，纳管2000+虚拟机，服务5万师生，实现教学、科研、管理IT资源实现分钟级实时分配，为每人提供极速100GB云网盘，也是首个支持IPv6的校园云。
- 新华三无线解决方案应用于清华大学156栋建筑，覆盖教学科研及室外区域，9000余台AP精细部署，并率先实现用户IPv6接入的安全管控。
- 新华三大数据解决方案助力清华大学开展机器数据分析、毕业生画像等应用，毕业生画像在两天时间内访问量达到89620次，为毕业学生四年大学生活带来美好回忆。
- 云融合大数据服务，通过H3C DataEngine作为校园应用创新引擎，服务校园科研创新应用与未来智慧园区建设。

服务科研高性能计算

为清华生命科学院搭建高性能计算与存储系统，应用于海量冷冻电镜数据中冷冻固定、显微、图像处理等数据的收集、处理，实现实验结果的快速、清晰呈现。

PART 4

把握机遇积极布局
迎接智能化时代到来

经过学术界、工业界、投资界、政府机构等多方共同努力，人工智能产业在近几年取得了引入瞩目的成果。未来随着人工智能应用大规模的推广，我们将进入一个全新的智能化时代。面对未来的挑战，下面的一些趋势值得我们关注：

ICT基础设施在人工智能产业中地位将愈加重要

海量人工智能应用需要稳固的ICT基础设施支撑，也促使硬件设备的设计向着满足AI应用需求的方向倾斜，比如更加可靠的网络、更快速的存储以及更高效的计算设备。为了支撑巨大的计算需求，AI芯片在未来几年将持续成为热点，通过架构上的不断优化来满足计算性能，尤其是端侧AI芯片需求量将会持续增加。另外，类脑芯片也将会是一个热点方向，可以有效解决对大数据和大算力依赖的问题。随着物联网设备增多，在边缘端进行人工智能分析将会成为趋势，未来边缘计算设备的需求量也将会激增。

认知智能时代到来，智能持续得到提升，应用场景将更为丰富

随着以谷歌的BERT及其衍生方法为代表的算法在自然语言处理领域上取得重大突破，语义理解准确率显著提升，同时知识图谱技术赋予机器具备知识推理能力，人工智能技术正在从“感知智能”迈向“认知智能”新阶段，也将会渗透到行业场景中更多关键业务环节中。语音、语义、视觉及情感等多模态融合成为AI领域新的研究方向。未来较长一段时间里，仍然以人机协同为主，机器处理其比较擅长的事情，而人类则可以处理机器所不能处理的高级事务。这种人机协同的模式，将会创造出更多的应用场景。

人工智能与5G、IoT、VR/AR技术相结合，将赋予更大的想象空间

随着5G、IoT和VR/AR技术的普及，将会产生更多新的人工智能应用场景。5G通信技术提供了智能时代万物互联的基础，IoT提供了洞察万物的数据基础，而VR/AR带来了全新的交互体验。几种技术交互融合，产生的效应超乎想象，将彻底改变人们的生活方式。

智能化时代即将到来，让我们翘首以待！

附录：缩略词表

缩略词	全称	中文名称
ABC	AI BigData Cloud	人工智能/大数据/云计算
AD-Campus	Application Driven Campus	应用驱动的园区解决方案
AD-DC	Application Driven Data Center	应用驱动的数据中心解决方案
AD-WAN	Application Driven Wide Area Network	应用驱动的广域网解决方案
AI	Artificial Intelligence	人工智能
AIaaS	Artificial Intelligence as a Service	人工智能即服务
AIOS	Artificial Intelligence Operating System	人工智能操作系统
AIOps	Artificial Intelligence for IT Operations	智能运维系统
API	Application Programming Interface	应用程序接口
APP	Application	应用程序
AR	Augmented Reality	增强现实
ASIC	Application Specific Integrated Circuit	专用集成电路
BERT	Bidirectional Encoder Representation from Transformers	来自变换器的双向编码器表征
CT	Computed Tomography	X射线计算机断层成像
DCBX	Data Center Bridging eXchange	数据中心桥接交换
ECN	Explicit Congestion Notification	显式拥塞通知
ETS	Enhanced Transmission Selection	增强流量选择
FPGA	Field Programmable Gate Array	现场可编程逻辑阵列
GIS	Geographic Information System	地理信息系统
GPU	Graphics Processing Unit	图形处理单元
HDFS	Hadoop Distributed File System	Hadoop分布式文件系统
HPC	High Performance Computing	高性能计算
ICT	Information and Communications Technology	信息与通信技术
IoT	Internet of Things	物联网
LLDP	Link Layer Discovery Protocol	链路层发现协议
MAC	Media Access Control	媒体存取控制
MDS	Meta Data Server	元数据服务器
MEC	Mobile Edge Computing	移动边缘计算
mMTC	massive Machine Type of Communication	海量机器类型通信
NB-IoT	Narrow Band Internet of Things	窄带物联网
NLP	Natural Language Processing	自然语言处理
NVMe	Non-Volatile Memory Express	非易失性内存主机控制器接口
OCR	Optical Character Recognition	光学字符识别
PCIE	Peripheral Component Interconnect Express	高速串行计算机扩展总线标准
PFC	Priority-based Flow Control	基于优先级的流量控制
RDMA	Remote Direct Memory Access	远程直接数据存取
RPA	Robotic Process Automation	机器人流程自动化
RoCE	RDMA over Converged Ethernet	融合以太网上的RDMA
RoCEv2	RDMA over Converged Ethernet version 2	RoCE的第二个版本
SaaS	Software as a Service	软件即服务
SNA	Seer Network Architecture	先知网络架构
SSMS	SQL Server Management Studio	SQL数据库管理工具
TLV	Type Lenght Value	类型、长度和值
TPU	Tensor Processing Unit	张量处理器
VDI	Virtual Desktop Infrastructure	虚拟桌面架构
VDP	Virtual Desktop Protocol	虚拟桌面协议
VPN	Virtual Private Network	虚拟专用网络
VR	Virtual Reality	虚拟现实
VXLAN	Virtual eXtensible Local Area Network	可扩展虚拟局域网络