



“诸葛·深知” 为企业打开人工智能之门



“国家的创新驱动发展战略对科研院所提出了更高的要求，近几年，信息领域的发展也给中科院自动化所带来了前所未有的挑战。面对新的科技发展态势，我们始终坚持以智能技术作为研究所战略发展的主要方向，此次与戴尔中国共建联合实验室，将进一步加强研究所的计算平台优势，增强科研与产业的融合创新，助力研究所在人工智能领域的创新发展。”

中科院自动化所所长徐波

客户资料



| | |
|----|---------------|
| 机构 | 中国科学院自动化研究所 |
| 行业 | 科研 |
| 国家 | 中国 |
| 网址 | www.ia.cas.cn |

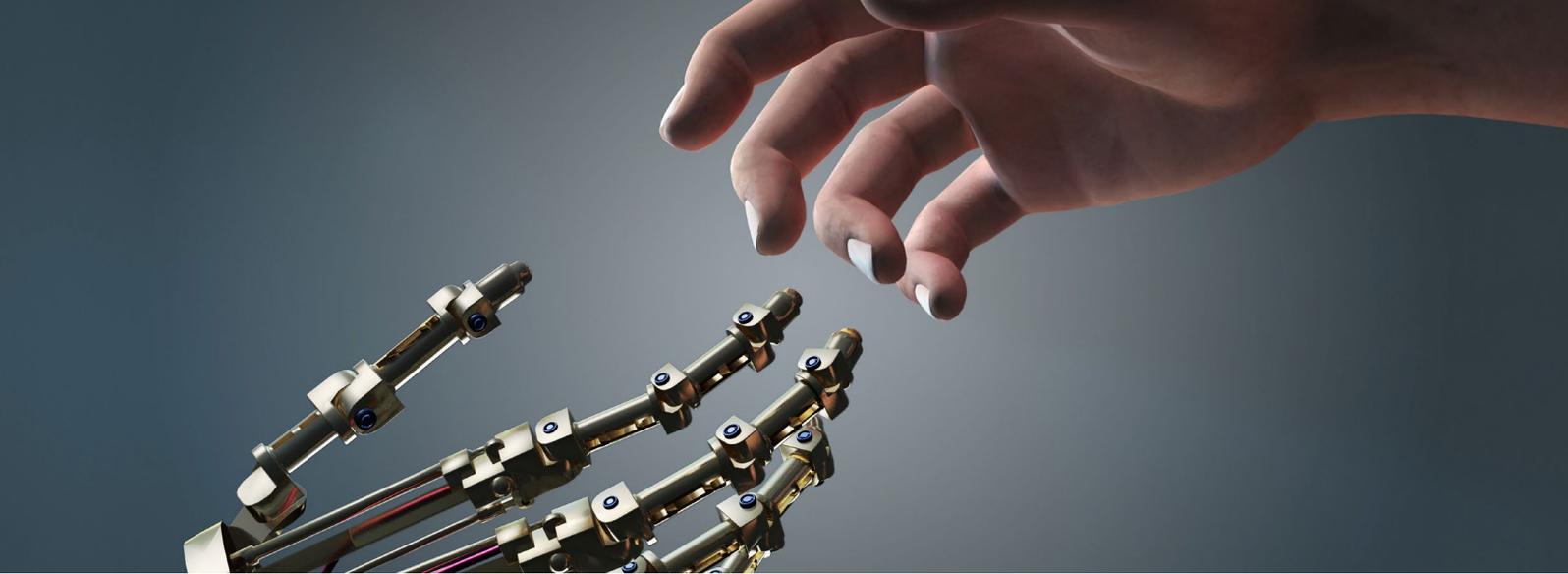
人工智能和深度学习新的发展机遇

向数字化转型和数字经济的迅猛发展使全球数据量爆增，为应对这一趋势，云计算、大数据等新兴技术应运而生，而物联网和“互联网+”的发展，又进一步推动了云计算和大数据技术的发展。同时，在从传统经济向数字经济转型的过程中，IT 技术成为这一转型中强大的助推器和核心引擎。正是大数据、物联网技术的崛起和高性能计算技术的进步，引发了人工智能的井喷式发展。

人工智能的提出可以追溯到上世纪 50 年代，但由于各种局限，一直没有得到快速发展。上世纪 80 年代机器学习出现，使人们找到了实现人工智能的方法。2010 年后，在高性能计算技术进步的推动下，深度学习取得了突破，成为实现机器学习的关键技术，并驱动了基于大数据的人工智能蓬勃发展。因此，毋庸置疑，大数据和高性能计算从根本上驱动了人工智能和深度学习的发展，同时，云计算加快了人工智能的发展速度。

据埃森哲的研究显示，到 2035 年，人工智能有望将企业盈利能力提升 38%，同时可能为中国经济增长率增加 1.6 个百分点。另据预测，到 2020 年，全球深度学习系统市场规模将超过 13 亿美元，2016 ~ 2020 年间的复合年增长率将达到 38.73%¹。

¹数据来源：Top 6 Vendors in the Deep Learning System Market from 2016 to 2020: Technavio.



由于中国市场数据量浩大，以及近年来在高性能计算方面的发展速度飞快，因此，对于基于大数据的人工智能/深度学习的发展，中国比其他国家更具有优势，有更多的发展机遇。尤其是中国政府近年来非常重视人工智能的发展，期望通过发展人工智能，进而带动智能制造业、智能农业、智能医疗、智能城市等的发展，最终实现传统产业转型升级的目标。因此，从2015年开始，国务院和相关部委频繁发布相关政策，推动人工智能的发展。比如《中国制造2025》、“互联网+”行动计划、“十三五规划”中都将人工智能列为着重发展的领域，国家发改委2016年9月还将人工智能纳入“互联网+”建设专项，并明确指出，针对我国人工智能基础支撑能力不足等问题，要建设深度学习技术及应用创新平台，支撑开展大规模密集型计算、海量数据训练、计算机视觉、生物特征识别、复杂环境感知、新型人机交互、机器翻译等技术的研发和工程化，提升我国深度学习算法综合运用能力。

未来5~10年，深度学习将向更多行业渗透，特别是数据密集型行业，如金融、证券、医药等，将颠覆传统行业数据分析和计算方式，形成基于大数据+深度学习的“智能+”。

在数据、计算、需求、政策的合力推动下，人工智能在中国的发展迎来了新一轮战略发展机遇。然而，对于很多正处于转型中的中国企业而言，在自身没有技术实力和优势的情况下，发展人工智能将会受到很大的限制。人工智能、深度学习技术的应用，将赋予科技创新企业更强的商业竞争优势。因此，一个帮助企业实现基于深度学习的大数据分析，快速步入人工智能时代的先进人工智能深度学习平台“诸葛·深知”诞生了。

“诸葛·深知”项目背景

“诸葛·深知”是中国科学院自动化研究所（下简称中科院自动化所）和戴尔联合开发的企业级深度学习应用与服务平台，为企业提供了深度学习工具包的统一接口，并提供了众多成熟算法模型，企业可轻松调用，帮助企业实现基于深度学习的大数据分析，步入人工智能时代。

早在2015年11月，戴尔就与中科院自动化所共建了“人工智能与先进计算联合实验室”，共同关注人工智能领域的新型计算架构的研究与应用，支持人工智能的创新发展研究，并且根据市场和行业发展的需要，对核心技术进行产品化和产业化。“诸葛·深知”平台正是双方从最初的平台基础设施搭建的合作到共创平台的深度参与的结果。

在人工智能领域，分布式机器学习开源平台的建立有着重要意义，它解决了数据持续增长和机器学习模型逐渐复杂化与计算能力受限的矛盾，大幅降低机构和個人进入分布式机器学习应用的门槛。而随着产业界和学术界越来越多的研究人员对平台的使用，将会贡献更多的数据，促进深度学习算法的完善。深度学习使得机器学习能够实现众多的应用，并拓展了人工智能的领域范围，使人工智能的进化节奏加快。

中科院自动化所是我国成立最早的国立自动化研究机构。近年来，该研究所把握信息与控制技术向智能化方向发展的趋势，聚焦类脑智能，部署了以脑科学与智能交叉融合为前沿、以类脑智能机器人与类脑智能信息处理为应用载体，进行类脑智能技术和产业的布局。作为以“智能立所”的国立科研机构，中科院自动化所在人工智能领域有着良好的理论基础和技术积累，凝聚了一大批优秀的专业人才队伍，面对当下智能科学与技术的发展趋势，结合行业应用的快速发展和快速迭代的行业特色，中科院自动化所迫切希望建立多渠道、多层次的协同创新平台，以强人工智能为目标，产业应用为引导，带动科研工作的创新发展。

深度学习算法的应用需要处理海量的数据集，以及大量的计算资源支持，其对于IT基础架构的需求将是企业级别，而非实验室级别，以满足系统部署与运维管理要求。当前深度学习框架、工具库众多，很多都是开源版本，适用场景各有不同，如何构建一个统一、高效的深度学习管理平台，满足多种深度学习框架及工具库的安装、配置、管理，以及多重框架的并行工作，这对多数企业和机构是一个巨大的挑战，尤其是对于大量技术力量薄弱的中小企业来说更是如此。而随着一系列深度学习框架及工具集的推出，操作人员相较以往则更容易在高性能计算集群上针对应用进行深度学习建模、训练、测试和调优，并且可以利用GPU技术实现成倍的浮点计算性能加速。



“诸葛·深知”定位及挑战

“诸葛·深知”为数据密集、价值型行业提供以深度学习为基础的大数据挖掘解决方案和计算平台，是企业级人工智能应用的入口。

设计目标

- 简单入门：针对目前市场上五花八门的深度学习工具包、复杂的算法，该平台将为企业提供深度学习工具包的统一接口，以及众多中科院的成熟算法模型，可轻松调用，帮助企业实现基于深度学习的大数据分析，使企业真正步入人工智能时代。
- 技术先进：平台包括了人工智能国家级科研队伍的最新技术成果，是基于算法优化的高性能计算平台，并有持续的研发作为保证。
- 安全可控：该平台提供全新的数据加密技术，保证客户数据安全，确保用于模型训练的客户数据全程加密。
- 定制服务：针对特定应用场景提供定制化的深度学习算法模型、训练及相关技术咨询培训，加速应用落地，满足中国企业的智能化需求。

面临的挑战

中科院自动化所在“诸葛·深知”的部署中面临的挑战主要来自三方面：一是对于脑科学乃至智能技术的研究来说，深度学习是一个重要的方法，但是深度学习的计算资源门槛非常高，进行大规模训练需要 GPU 计算群，而研究所原有的基础设施不能满足需求；二是深度学习、认知功能模拟以及内脑研究需要相应的技术架构支持，而研究所原有的架构无法支持深度学习；三是深度学习的基础是海量数据，而大量的数据对于存储以及网络性能也提出了更高的要求，这对于研究所来说也是一大挑战。

合作伙伴选择

基于深度学习应用与服务平台建设对 IT 基础架构的需求，中科院自动化所在全面了解目前各 IT 解决方案提供商的优势后，选择了与戴尔公司合作，双方共同成立了“人工智能与先进计算联合实验室”，并由此开启了合作搭建先进的通用平台基础设施，以及深度学习和机器人应用与服务平台的建设。

戴尔公司作为全球领先的 IT 厂商，秉承“在中国，为中国”戴尔中国 4.0 战略，不但能为中国企业级用户提供领先、开放、标准的企业级 IT 技术平台解决方案和高性能计算解决方案，同时，戴尔在全球拥有众多的成功案例和得以验证的参考架构，并构建了高性能计算、云计算等强大的生态圈。

为人工智能而打造的基础架构平台和应用平台

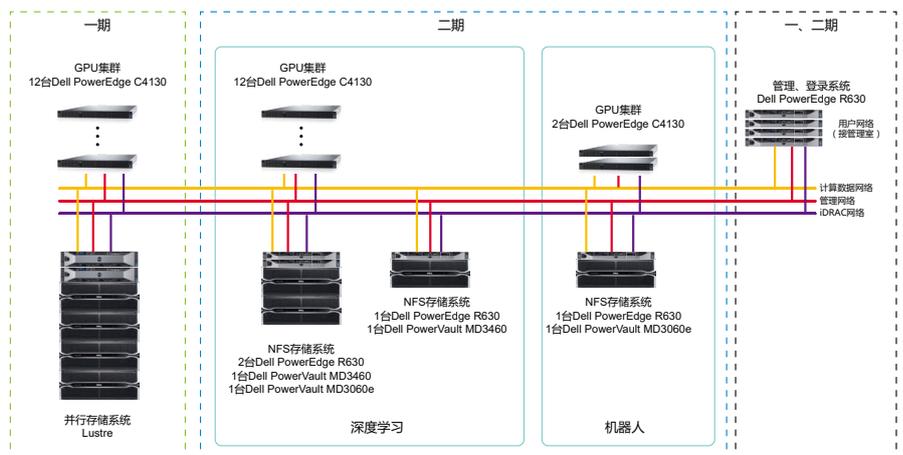
“诸葛·深知”企业级深度学习应用与服务平台项目分两期进行，第一期的目标是基础设施建设，即搭建先进的通用计算平台；第二期的目标是通过增强存储和计算能力、增强 GPU 等，满足深度学习和机器人平台的需求。

系统概况

“诸葛·深知”对于基础设施的需求较高，鉴于中科院自动化所用户数量多、种类多、应用多的实际情况，系统对于计算能力和存储能力的需求都非常大，如在影像处理方面，要求 800MB/s 单线程带宽访问，能够存储 32GB 的数据文件；在训练计算方面，要求 GPU 运算至少能支撑十几个应用并发，等等，在计算和存储方面都有较高的要求。

建设深度学习应用与服务平台有三个关键点需要考虑：

- 先进性和高性能，能满足深度学习领域对高性能计算和算法优化等先进性的需求；
- 稳定性和高可靠性，保证深度学习平台 IT 架构的性能稳定和安全可靠；
- 可扩展性和灵活选择，IT 平台具有可扩展性，能够满足未来应用的需求和提高 IT 投资回报率，同时，由于在深度学习领域有非常多的算法和框架，企业或机构的用户使用时可能有多种选择，如何让用户按需加载他们需要的文件或学习库是深度学习应用平台需要解决的一大问题。



企业级深度学习应用与服务平台第一、二期架构图



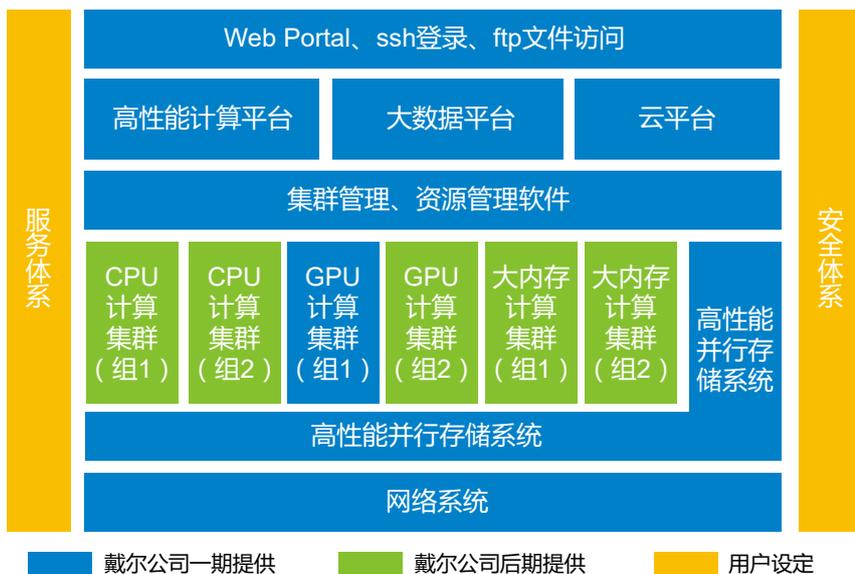
在平台建设过程中，中科院自动化所采用了戴尔及合作伙伴的先进技术和产品，包括戴尔高性能服务器、NVIDIA Tesla GPU、戴尔存储服务器和戴尔大容量存储设备、Intel Lustre 高性能并行存储、高性能 Mellanox 网络产品，以及业界领先的 BrightComputing HPC/Deep Learning 软件，并由戴尔中国的专业服务实施团队进行安装部署。系统主要组成包括：

- GPU 计算节点配置戴尔高性能服务器 PowerEdge C4130 ；
- 管理存储节点配置戴尔高性能的机架式服务器 PowerEdge R630 ；
- 存储节点配置戴尔存储设备 MD3060 和 MD3460 ；
- 网络部分配置戴尔领先的网络设备 56G InfiniBand 网络和万兆网作为计算网络，而千兆网组成管理和 iDRAC 网络 ；
- 戴尔 Storage for HPC with Lustre ；
- 集群管理软件 BCM ；
- 机柜。

IT 基础架构

在企业级深度学习应用与服务平台 IT 基础架构中，硬件系统由 CPU 计算集群、GPU 计算集群、大内存节点集群、高性能存储系统、网络系统等组成；软件系统由操作系统、集群管理和资源管理软件组成。可以提供高性能计算平台、大数据平台和云平台。用户的访问可以通过 Web Portal、SSH 登录和 ftp 文件访问。

此外，通用计算平台还包括可由用户设定的服务体系以及安全体系。



IT 基础架构图

一期通用计算平台

在一期通用计算平台中，采用 12 台戴尔 PowerEdge C4130 GPU 服务器组成了高性能计算集群，并采用容量为 960TB 的并行存储系统 Lustre，搭建了一个高性能计算平台。戴尔先进的技术和产品满足了中科院自动化所对于各类高性能计算应用的需求。

在高性能服务器方面，戴尔 PowerEdge C4130 服务器针对 HPC 作业和 VDI 环境，在 1U 的空间内提供了高密度的 PCI 插槽，并且支持最大 4 个 300W 的加速器、2 颗 Intel Xeon E5-2600 v4 CPU 和 1TB 内存，最大化了 IO 灵活能力。此外，PowerEdge C4130 还支持 InfiniBand FDR、NVIDIA GPU DIRECT 和 Intel Omni-Path 架构。

在存储方面，配置 Intel Lustre 的戴尔高性能计算并行存储通过单文件系统命名空间，实现容量和性能的轻松扩展，满足 PB 级文件存储的容量和性能要求。而通过动态存储扩展，又使每构建块峰值可以达到 18.5GB/s 的读取速率和 15.1GB/s 的写入速率。同时，使用 Intel Hadoop Adapter for Lustre (HAL) 大数据工作负载的解决方案设计将 HPC 和 Hadoop 的数据融合到一套存储系统中，以提升数据处理速度。

在网络方面，戴尔 Networking H 系列交换机基于 Intel Omni-Path 架构，缩短传输路径，让数据更靠近 CPU。而 Mellanox EDR InfiniBand 采用了每端口速率高达 100Gb/s 的交换机 IC 芯片 Switch-IB，专为高性能计算而设计，同时也是一个可实现全隔离和可扩展路由功能的交换机芯片，能够满足未来上千或更多节点的集群需要。

不仅如此，戴尔高性能计算解决方案的另一大优势是针对不同行业的工作负载进行设计并优化，使“HPC+ 数据分析解决方案”既可提供定制的高性能计算系统的灵活性，又具备预配置型系统的速度、简便性和可靠性，同时还配备了经过全面测试和验证的构建块，保障系统平台具备较高的可靠性。



“戴尔比以前的定位更加明确，即提供创新的端到端解决方案。在全新的“在中国，为中国”戴尔中国 4.0 战略下，戴尔将加强与中国研发机构的合作，推动前沿科技的发展。此次与中科院自动化研究所合作，共同建设“人工智能与先进计算联合实验室”，将计算资源与科研资源强强联合，推动人工智能跨越发展，抢占新一轮科技革命制高点，从而对人类的经济、社会产生深远影响。”

——戴尔大中华区总裁黄陈宏博士

二期应用平台

系统的二期主要针对深度学习和机器人应用而扩展和开发。在这一过程中，共使用了 12 台戴尔 PowerEdge C4130 GPU 服务器组成了集群，配合 2 个容量分别为 480TB 和 240TB 的 NFS 存储系统用于深度学习，另用 2 台 C4130 GPU 服务器和一个 240TB 的 NFS 存储系统用于机器人应用平台。

这部分同样使用了戴尔 GPU 计算、CPU 计算、胖节点等多种计算平台和 Infiniband、Omi-Path、40Gb、10Gb 等多种高速网络互联方案，以满足深度学习和机器人浮点计算强度以及网络传输需求。

在存储方面，戴尔提供的全套 NFS 解决方案的单个命名空间从最小 48TB 扩展至 480TB 的原始容量，同时以高可用性故障转移配置保障稳定性和可靠性，以优化的设计配置提供最高的数据吞吐量和峰值可达 6.07GB/2.23GB 的读写性能，满足深度学习和机器人平台对于性能的严苛需求。

而对于业界常用的人工智能 / 深度学习框架、工具集、函数库与硬件驱动，则是通过戴尔 Bright Cluster Manager (BCM) 集群管理软件来进行管理。Bright Cluster Manager 广泛支持当前使用的主流 Caffe\Torch\TensorFlow\Theano 等机器学习框架，MLPython\cuDNN\DIGITS\CaffeOnSpar 等机器学习工具，以及 CUDA drives\CUB(CUDA building blocks)\NCCL(library of standard collective communication routines) 等硬件驱动。

在深度学习 IT 基础架构中，NVIDIA GPU 通常是性能加速的关键部件。Bright Cluster Manager 集成了 NVIDIA 驱动、工具集及管理 API，可以通过 Bright Cluster Manager 部署 NVIDIA CUDA、DIGITS，运行 NVIDIA Docker 镜像，在各种深度学习框架下运行 GPU，支持深度学习应用跨越多台物理设备的 GPU 资源。在 GPU 运行过程中，Bright Cluster Manager 收集 NVIDIA GPU 卡运行状态的采样，自动进行健康检查，并通过图形化界面进行展示。

戴尔与中科院自动化所联合推出的“诸葛·深知”平台，充分利用了双方的技术优势，为中国企业和机构提供了深度学习工具包的统一接口，并能轻松调用众多中科院的成熟算法模型，帮助企业和机构实现基于深度学习的大数据分析，使他们能尽早从人工智能等先进技术中获益。

应用前景

对于中科院自动化所来说，与戴尔联手推出的“诸葛·深知”平台无疑将有力地推动他们在脑科学方面的科研项目开展。中科院自动化所 - 戴尔脑信息与智能技术先进计算研究中心致力于面向脑信息处理与智能技术领域的新型计算架构研究与应用，包括脑与认知海量数据分析、脑系统计算模拟、认知功能模拟、人工智能等方向的新型计算体系，现

在研究中心的研究团队联合国内外研究人员利用“诸葛·深知”平台已成功绘制出全新的人类脑图谱——脑网络组图谱，第一次建立了宏观尺度上的活体全脑连接图谱。不仅如此，中科院自动化所其他项目组也正在使用平台资源开展课题研究，加速成果转化。

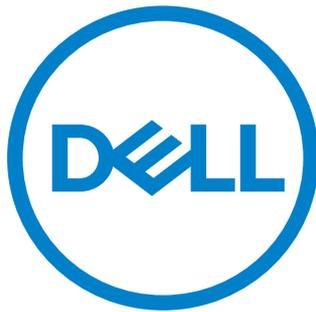
随着“诸葛·深知”等全新认知的人工智能平台的推出，为企业和机构打开了利用人工智能的方便之门。由于人工智能的采用，企业和机构明显改善了业务运营，不仅提高了工作效率，还扩大了市场覆盖范围。一些典型的行业包括公共交通、营销和媒体、法律服务、金融服务、医疗保健服务、石油和天然气乃至消费类电子产品等。对于数据丰富的行业，企业可充分利用基于大数据的深度学习 / 人工智能及认知技术所能带来的优势，提升行业的竞争力，加速向数字化转型。例如，在银行业，利用大数据，通过“诸葛·深知”平台进行实时防欺诈检测；在医疗保健行业，通过大数据和深度学习，进行流行性肺炎爆发预警和预测，利用医院大量的近千万级门诊数据和数以万计的确诊信息，加上近 3 年环保气象数据，通过“诸葛·深知”平台进行多元拟合模型 + 卷积神经网络分析等，从而做出疾病的预警；再如，通过“诸葛·深知”平台进行基因表达特征模型探索，对肿瘤基因表达特征应用分析，试图更准确地确定病情和提供更优的治疗方案。

“诸葛·深知”平台为各行业用户打开了基于大数据的人工智能 / 深度学习之门，通过这一平台，企业和机构将走上人工智能之路。我们相信未来随着“诸葛·深知”平台的不断发展和完善，将会有更多的行业和企业从中受益。



2016 年 10 月，戴尔（中国）有限公司荣获了中国计算机学会高性能计算专业委员会颁发的“人工智能技术实践创新奖”，以表彰戴尔在与中科院自动化所合作建立“人工智能与先进计算联合实验室”和“诸葛·深知”企业级深度学习应用与服务平台中所做出的卓越贡献。





如欲浏览戴尔的所有案例，请登录 [Dell.com/casestudies](https://www.dell.com/casestudies)

Dell、Dell 徽标、Dell ProSupport、EqualLogic、PowerEdge、PowerVault、Dell Precision、UltraSharp 和 Wyse 均为 Dell Inc. 的商标。本文中所提及的其他商标或产品名称是指拥有相应商标和名称的公司或其产品。Dell 对其它公司的商标和名称不拥有任何所有权。Dell 软件、解决方案和服务的可用性和条款因地区而异。本案例研究仅供参考。Dell 在此案例研究中不作任何明示或暗示的担保。© 2016 年 12 月，Dell Inc. 保留所有权利。