

# “普慧”算力 开启新计算时代

2023年7月

[kpmg.com/cn](http://kpmg.com/cn)

# 摘要

历史总在重演，科技永远向前。纵观过往三次科技革命，无不是由某种基础性的通用技术（GPTs, General Purpose Technologies）推动，站在ChatGPT爆发的当下，通用人工智能极有可能成为新一代的GPTs。从GPT（生成式预训练模型）到GPTs，第四次科技革命即将踏浪而来，人工智能也将迎来第四次发展浪潮。双浪叠加背景下，人类的数字化大航海征途再启新篇。算力作为新型生产力是驱动经济社会发展的底层动力，其供给规模受限与需求指数级增长之间的矛盾日渐突出，传统计算范式正在失效，计算模式从中心式演进到分布式，再到云计算阶段，即将进入融合智能计算阶段，标志着新计算时代的开启。大国和大企业主导的算力竞争将愈演愈烈，资金、技术、人才都不可或缺，而最根本的是要准确把握发力方向，方能决胜未来。

未来算力发展将具备两大核心特征：“数字经济的基础设施”和“通用人工智能的核心动力”。与之相对应，算力将在两个关键维度上加速发展：普适（Inclusive）与智慧（Intelligent）。“普适”意味着算力将成为人人可得（Affordable）、人人可用（Available）、人人适用（Adaptable）的“3A”型基础资源，是发展数字经济的坚实基础。“智慧”意味着算力将具备自适应（Self-adapting）、自学习（Self-learning）、自进化（Self-evolving）为代表的“3S”智能，是实现通用人工智能广泛应用、推动数字经济迈向更高阶段的必然要求。

放眼全球，算力或将成为影响国家综合实力和国际话语权的关键要素之一，已是全球主要经济体的战略必争之地。美、中、欧、日作为全球算力产业规模的领跑者，均在持续强化相关布局，各国算力投资或补贴计划金额均在千亿元以上。面对波诡云谲的国际形势，中国始终依靠技术创新来夯实算力基础，已初步建立起产业链条持续完善、产业创新不断活跃、市场规模迅速增长、国际竞争力稳步提升的算力产业生态。经初步测算，到2025年，中国算力核心产业规模将不低于4.4万亿元，成为与新能源汽车比肩的超万亿级高潜赛道，若将视野扩大到算力关联产业，规模可达24万亿元。

立足中国，算力产业的发展蓝图是由千行百业的算力需求落地所绘制的。实体经济呼唤新计算时代的新型算力，基于对代表行业的调研评估，可大致将各行业的算力需求潜力分为四类，包括以ICT和制造业为代表的“普惠双驱型”，以汽车行业为代表的“智慧拉动型”，以金融业为代表的“普适促进型”，以及以医疗和教育产业为代表的“发展酝酿型”。不难看出，尽管各行业的算力需求各有千秋，但落脚点均在“普惠”。“普惠”算力包括三大关键要素：算力设施（Infrastructure）、算力应用（Application）、算力服务（Service）。在三大关键要素的作用下，算力将经历从产生到落地的完整闭环，彻底转化为新型生产力。

预测未来最好的方式就是创造未来。正如“普惠”算力的叙事离不开数字经济和通用人工智能这两个宏大主题，“普惠”算力的产业实践也将扎根于数字经济、人工智能大模型等提供的发展土壤。

在数字经济方面，数字经济最核心的资源是数据要素，“普惠”算力作为数字经济的基础设施，将充分盘活数据资产，助力中国充分释放“数据红利”，中国数字经济体量有望在2030年突破百万亿元。在通用人工智能方面，AI大小模型协同进化是未来发展方向，“普惠”算力作为通用人工智能的核心动力，既是科技公司搭建AI模型的“卖铲人”，又是检验算力公司创新能力、制造能力的“试金石”，更是推动中国迈向科技强国的“加速器”。驱动“普惠”算力前行的双轮上，一边是安全，一边是发展，只有统筹安全与发展，以政府、企业、科研机构等多方力量，坚持走高质量发展道路，落实各项具体举措，才能真正以“普惠”算力开启新计算时代。



# 目录

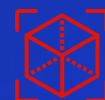
01



顺势：  
发展背景篇

05

02



明义：  
概念界定篇

13

03



有为：  
大国战略篇

21

04



识局：  
发展评估篇

29

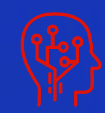
05



应变：  
需求洞察篇

41

06



筹谋：  
趋势展望篇

69

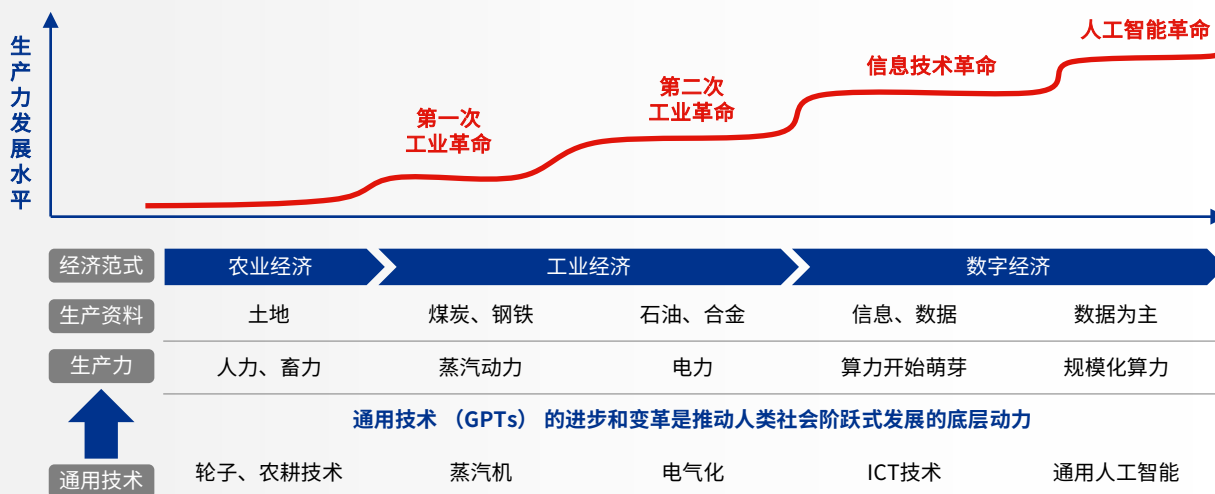
# 01 顺势： 发展背景篇



## 从GPT到GPTs，通用人工智能点燃第四次科技革命

人类历次科技革命的爆发标志都是新一代通用技术（GPTs, General Purpose Technologies）引发生产力跨越式发展，进而使生产资料、经济范式等发生颠覆式变革。

图 1 | GPTs推动人类历次科技革命



资料来源：公开资料，毕马威分析



区别于SPTs（专用技术, Special Purpose Technologies），GPTs代表着对人类经济社会产生巨大、深远和广泛影响的革命性技术，典型的技术包括轮子、农耕技术、蒸汽机、水煤电、计算机、互联网等。

在农业经济时代，轮子、农耕技术等普及，提高了人们迁徙和获取食物的效率，人类可以在大规模开发土地资源的基础上持续繁衍，为经济发展提供源源不断的底层动力。但由于纯体力劳动效率较低、土地资源有限（尤其是耕地属于不可再生资源），农业经济增长相对缓慢且存在明显的“天花板”。

在工业经济时代，蒸汽机和电气化分别引发了第一次工业革命和第二次工业革命，这两项GPTs从根本上改变了体力劳动的性质，代之以机械化大生产，在蒸汽动力、电力等新型生产力驱动下，人类获取矿产等自然资源的效率大幅提升，以大规模流水线生产为特征的现代工厂加速兴起，推动人类社会实现了跨越式发展。

在数字经济时代，信息通信技术（ICT, Information and Communications Technology）作为推动信息技术革命的GPTs，则从根本上改变了某些类型的脑力劳动，人类获取信息、处理信息的效率得到极大提升，以算力为生产力，信息、数据为生产资料的数字经济模式开始萌芽。

2022年11月底，OpenAI推出的ChatGPT（聊天生成预训练转换器, Chat Generative Pre-trained Transformer）引发全球热议，人工智能技术有望成为新一代的GPTs——通用人工智能（Artificial General Intelligence, 简称AGI）。AGI将推动数字经济更加蓬勃发展，使人类社会发展迈向更高阶段，从GPT（生成式预训练模型）到GPTs（通用技术），人工智能有望作为新的底层通用技术，点燃第四次科技革命！

# 人工智能发展并非一帆风顺，正迎来第四次浪潮

AGI也称强人工智能（Strong AI），是指与意识、感性、知识、自觉等人类特征相联结，能够执行人类智力行为的机器智能。相对应地，应用型人工智能（Applied AI）也称弱人工智能（Weak AI），是指能实现专用或特定技能的机器智能，如人脸识别、机器翻译等。

迄今为止，人类所接触的各种人工智能系统，都属于应用型AI，这类应用能在某些单项智能上超越人类，例如下围棋等。但如果要在更广泛意义上达到甚至超越人类智能，必须实现AGI。不过，人工智能过往的发展并非一帆风顺，面向AGI的未来也将不会一蹴而就。

1956年，达特茅斯（Dartmouth）会议召开，标志着AI元年开启，在近70年的发展历程中，AI产业经历了“三起两落”：



## 一起一落（1950s-1970s）

达特茅斯会议之后，大量AI程序和创新研究涌现，包括搜索式推理、自然语言等，相关研究获得了大量经费支持。但好景不长，到了70年代，AI发展遭遇瓶颈，原因是大部分AI程序都不具备解决复杂问题的能力，AI研究者们也遇到了无法克服的基础性障碍并受到了哲学家等的批判，研究机构随之开始大幅削减对AI课题的资金投入，经历过1974年经费削减的研究者们以“AI之冬”一词概括了AI产业的第一次低谷。



## 再起再落（1980s-2000s）

80年代，名为“专家系统”的AI程序问世，它能够依据一组从专门知识中推演出的逻辑规则在某一特定领域回答或解决问题，大大增强了AI的实用性，大部分公司开始采用“专家系统”并内设AI部门。1980-1985年，各类公司对AI的投入总计达十亿美元以上，AI再一次大获成功。然而，“专家系统”的实用性只局限于特定领域，持续升级难度高且维护成本居高不下，严重阻碍了其推广应用。1987年，AI硬件市场需求大幅下跌，Symbolics等厂商生产的LISP机<sup>1</sup>不仅价格昂贵，性能也不敌Apple和IBM生产的台式机，开始迅速被市场淘汰，标志着AI的第二次寒冬来临。

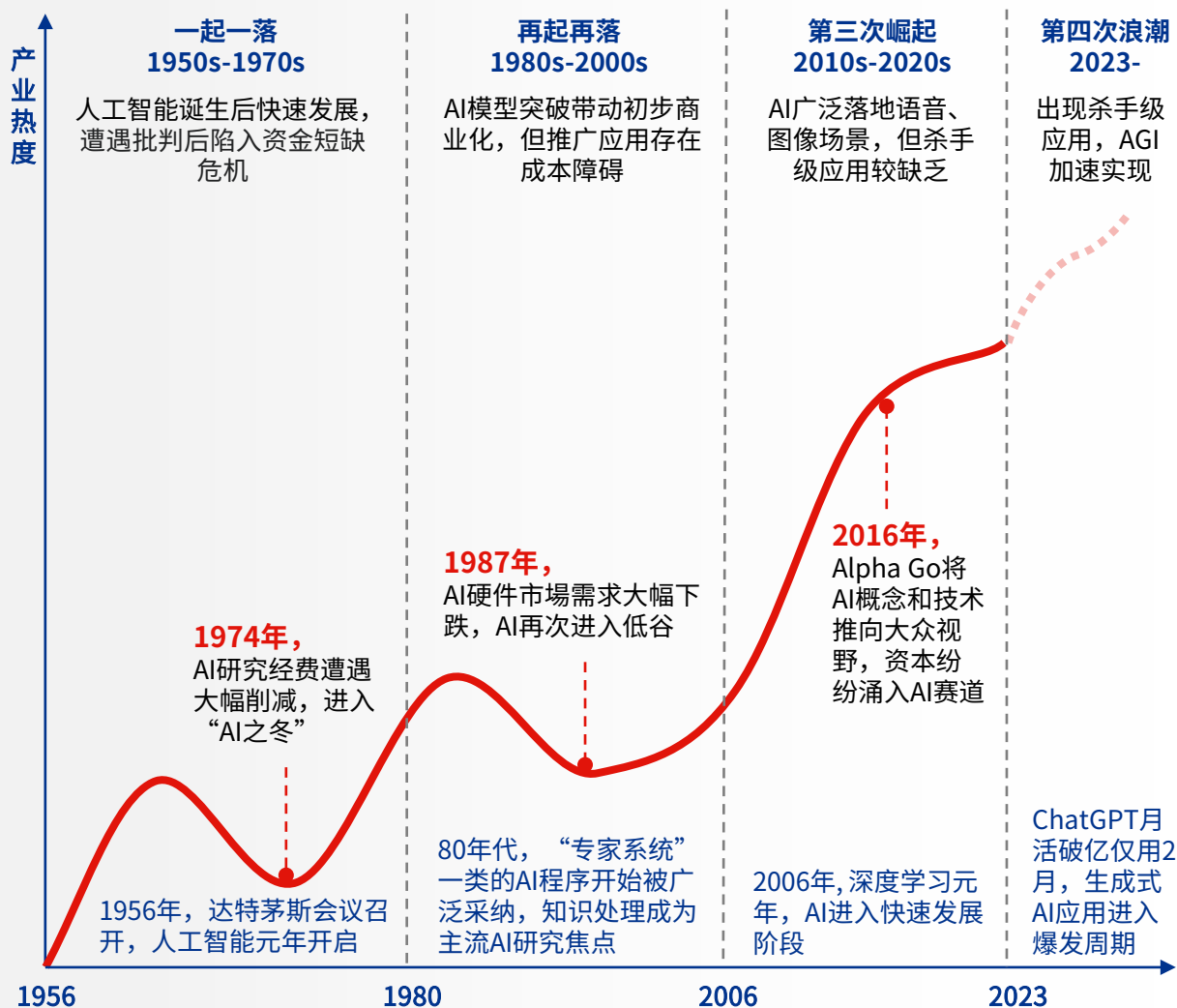
<sup>1</sup> LISP机是一种直接以LISP语言的系统函数为机器指令的通用计算机，是首先进入市场并广泛应用的人工智能机。



## 第三次崛起（2010s-2020s）

21世纪以来，互联网的飞速发展推动人类进入大数据时代，数据、算法、算力三要素齐头并进，以深度学习为核心的机器学习算法取得了持续突破。2006年，加拿大多伦多大学教授Geoffrey Hinton及其团队在《Science》发表的一篇文章中，首次提出了深度学习的概念，并详细给出了“梯度消失”问题的解决方案——通过无监督的学习方法逐层训练算法，再使用有监督的反向传播算法进行调优。该解决方案的提出，开启了深度学习元年。深度学习元年之后，AI进入快速发展阶段。到2016年，谷歌DeepMind研发的AlphaGo在围棋人机大战中击败人类棋手李世石，人工智能开始被大众熟知，各界对AI的热情被充分调动，语音识别、图像和视频处理等AI应用开始逐渐渗透到各行各业。

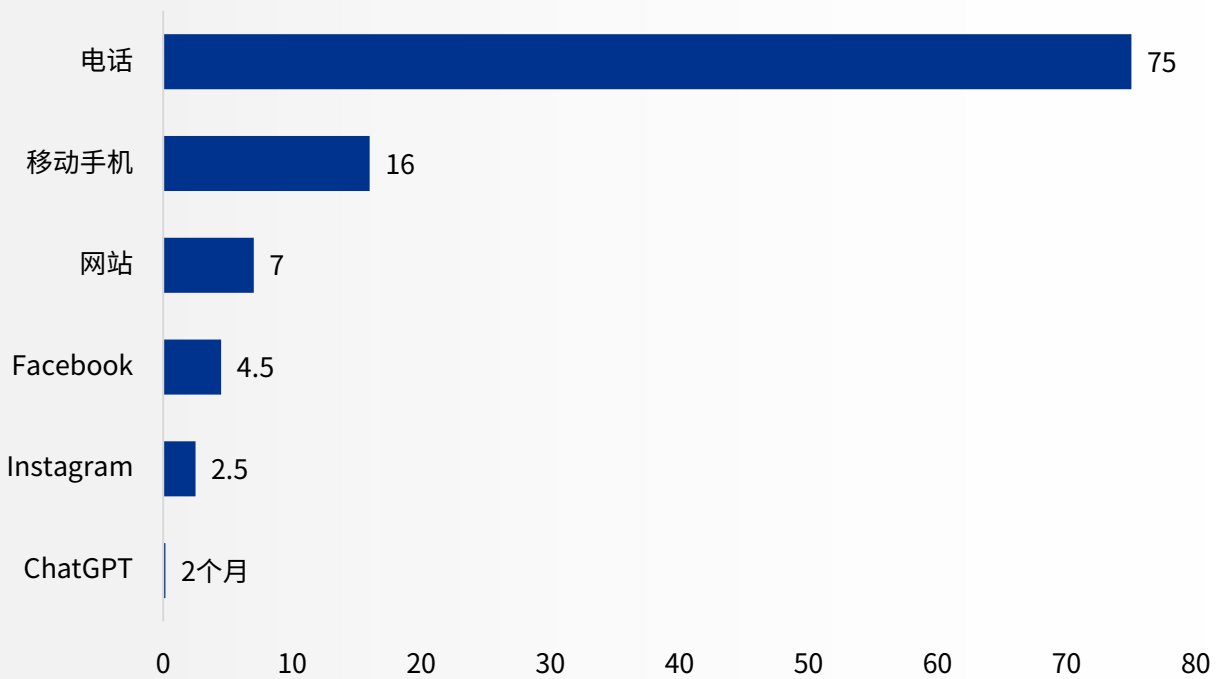
图2 | 人工智能经历“三起两落”迎来第四浪



资料来源：公开资料，毕马威分析

ChatGPT推出后迅速成为现象级应用，仅用两个月就实现了月活用户数量破亿，成为了历史上达到这一指标最快的产品，有望推动AIGC（人工智能生成内容，AI Generated Content）等进入爆发周期，加速实现AGI。

图3 | 各类典型产品达到全球1亿用户所用时间（单位：年）



资料来源：World of engineering，毕马威分析

根据一项由OpenAI和宾夕法尼亚大学等机构学者进行的预测<sup>2</sup>，以GPT（生成式预训练模型，Generative Pre-trained Transformer）为代表的大型语言模型（LLM, Large Language Model）将给美国劳动力市场带来重大影响，呈现出了GPTs的特征。从影响范围来看，大约80%的美国劳动力至少有10%的工作任务将受LLM使用的影响，大约19%的美国劳动力可能会感受到至少50%的任务受到影响。从影响程度来看，通过获得LLM支持，美国大约15%的工作任务可以在相同质量水平下更快地完成，当在LLM上面搭建软件和工具时，这一份额将增加到所有任务的47%-56%。

当前，实现AGI还面临着技术、数据、计算能力、知识表示和人类智能理解等多方面的挑战，ChatGPT作为AGI发展的重要里程碑，其核心技术壁垒是数据、算法、算力等要素资源的精巧组合，证明了沿着“大数据、大模型、大算力”路径发展AGI的可行性，预计随着各国科技巨头加紧布局相关技术，AI产业经历“三起两落”后，将迎来第四次浪潮。

<sup>2</sup> Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock, GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models, March 27, 2023 (arxiv.org)

# 双浪驱动新计算时代来临，算力紧缺亟待解决

第四次科技革命与AGI双浪叠加的时代背景下，新一轮科技创新周期将肇始于底层算力创新，繁荣于AIGC等上层应用爆发，大数据、物联网、人工智能、区块链等数字技术趋向于组合式进化，将释放出巨大的乘数效应。算力是多技术、多领域的交汇点，更是一切创新成果从首次落地到大规模扩散的元动力，算力需求指数式增长的同时，传统计算架构正失去竞争力，一个打破思维定式、充分释放算力“底层动力”价值的新计算时代即将开启。

具体来说，科技创新周期一般会经历“技术创新——模式创新——应用创新”三阶段。以个人电脑和移动互联网的普及为例，两者均是围绕ICT硬件创新展开，进而带来门户网站、平台化等商业模式创新，最后是各类软件应用的不断涌现。对应到产业层面，竞争焦点也会从硬科技实力逐步过渡到产业生态构建等软实力。

随着PC、智能硬件渗透率增长见顶，互联网流量红利消退，科技创新即将开启新一轮循环，各类前沿技术走向深度融合，由硬科技实力主导的技术创新再次回到聚光灯下，单一硬件层面的革新已不足以帮助企业构建起核心竞争力，而是要掌握更核心的凝聚力和更底层的驱动力——算力。

狭义的算力可以用一台计算机理论上具备的每秒浮点运算次数（FLOPS）来衡量<sup>3</sup>。例如，世界上第一台通用计算机“ENIAC”每秒只能进行5,000次计算，目前全球排名第一的超级计算机“Frontier”每秒能进行百亿亿次运算（1.194Exaflop）。计算机性能一路狂飙，是信息技术大规模普及的必然结果，过程中，算力对经济社会发展的作用越来越重要，算力内涵得以逐渐摆脱物理载体限制，扩充成了数字经济时代集信息计算力、数据存储力、网络运载力于一体的新型生产力<sup>4</sup>。

算力内涵扩充的逻辑升华点在于，从“技术工具”进阶为“底层动力”，背后离不开计算模式的不断演进：

## 新计算时代



### 1960-1980

#### 集中式阶段：

在由封闭网络和巨型应用主导的集中式计算架构下，以IBM大型计算机为代表，算力被集中式地分配使用。



### 1980-2000

#### 分布式阶段：

随着小型计算机性能每18个月翻一番，最终可与大型计算机相媲美，且价格和耗电量大幅降低，PC互联网时代来临。在由开放网络和小客户端应用主导的分布式计算架构下，算力进入分布式阶段。



### 2000-2020

#### 云计算阶段：

随着移动通信技术快速发展，计算机、手机、传感设备等能随时随地接入网络，人类进入移动互联网时代，共享计算资源、提高算力利用率等需求催生出云计算模式，算力进入以云计算为中心的集中式共享阶段。



### 2020 以来

#### 融合智能计算阶段：

从万物互联到万物智联，数据爆炸、场景爆炸、大模型爆炸的发展趋势下，计算密度越来越高，计算节点分布越来越广，算力需求不仅体现为量的井喷，更呈现出“万物融合，智能无界”的显著特征，即“万物皆数据”“无数不计算”“无算不智能”。

<sup>3</sup> 算力的英文如何翻译？中国计算机学会通讯，2022年9期，[https://dl.ccf.org.cn/article/articleDetail.html?type=xhtx\\_thesis&\\_ack=1&id=613366915296665](https://dl.ccf.org.cn/article/articleDetail.html?type=xhtx_thesis&_ack=1&id=613366915296665)

<sup>4</sup> [http://www.news.cn/fortune/2022-08/10/c\\_1128902885.htm](http://www.news.cn/fortune/2022-08/10/c_1128902885.htm)

计算模式在前三阶段的演进主要由摩尔定律和冯诺依曼计算架构两大底层理论支撑，前者持续引导着芯片制程工艺改进，后者实现了计算机硬件设计和软件设计的分离，大幅提升了计算灵活性，形成了以“CPU（Central Processing Unit，中央处理器）”为核心的通用计算体系。然而，摩尔定律正不断逼近物理极限，依靠先进制程提升计算性能的方式越来越难以为继，愈发多样化、个性化的计算需求倒逼着通用计算架构向通专并行、提高专用性、实现定制化等方向创新，例如有CPU+GPU、CPU+ASIC、CPU+FPGA等创新模式。

一个明显的迹象是，产业界正加速推动芯片和计算架构创新，CPU正不断丧失主导地位。例如，谷歌自2016年以来就不断创新专为机器学习定制的专用芯片TPU（Tensor Processing Unit，张量处理器），并利用TPU进行了大量的人工智能训练工作。英伟达则抓住AI大模型爆发契机大力推广“GPU+加速计算”方案。此外，也有观点认为TPU、GPU都并非通用人工智能的最优解，指出量子计算具有原理上远超经典计算的强大并行计算能力，IBM在2023年宣布了将与东京大学和芝加哥大学合作建造由10万个量子比特（量子信息处理的基本单位）驱动的量子计算机，有望推进量子计算在新药物研发、探索暗物质、破译密码等方面的应用<sup>5</sup>。

总而言之，在融合智能计算阶段，算力作为“底层动力”对经济社会发展的影响更为广泛且深刻，算力正由终端计算等需求驱动的“被动式”发展，转向促进AI大模型训练、实现通用人工智能、超越经典计算等代表的“主动式”发展。新硬件、新架构竞相涌现，现有芯片、操作系统、应用软件等都可能被推翻重来，新计算时代呼之欲出。

无论被动还是主动，都指向同一个结论，算力供给增速明显难以满足指数式爆发的需求，储备算力成为各行各业的必要举动。而算力具有高投入、高风险、强外部性和高垄断特征，意味着围绕算力的竞争将会是大国和大企业参与为主的竞争。面向下一轮竞争，硬核的技术实力、雄厚的资金储备、高水平的人才队伍等都是参与下一轮竞争的“加分项”，而唯有准确掌握发力方向，才是决胜未来的“必选项”。

<sup>5</sup> IBM计划10年内建成10万量子比特的量子计算机，与两所大学合作，澎湃新闻，2023年5月31日，[https://www.thepaper.cn/newsDetail\\_forward\\_23294150](https://www.thepaper.cn/newsDetail_forward_23294150)





# 02 明义： 概念界定篇



## “普慧” 是未来算力的核心特征

未来算力发展趋势将具备两大特征：“数字经济的基础设施”和“通用人工智能的核心动力”。因此，算力将在两个关键维度上加速发展：普适（Inclusive）与智慧（Intelligent）。“普适”意味着算力将成为人人可得（Affordable）、人人可用（Available）、人人适用（Adaptable）的“3A”型基础资源，是发展数字经济的基础。“智慧”意味着算力将具备自适应（Self-adapting）、自学习（Self-learning）、自进化（Self-evolving）为代表的“3S”智能，是实现通用人工智能广泛应用、推动数字经济迈向更高阶段的必然要求。

图 4 | “普慧” 算力概念示意图

算力是数字经济的基础设施

未来算力发展两大趋势特征：

Computing for the future Digital Economy and AGI

算力是通用人工智能的核心动力

资料来源：毕马威分析





提出“普惠”算力（Inclusive and Intelligent computing）不仅是面向未来的重大选择，更是立足当下的深刻思考。

当前，算力发展面临着巨大的挑战，体现为传统计算架构缺乏并行管理架构、效率低、能耗高，已然无法适应爆发式增长的算力需求和愈发复杂的计算任务。CPU、GPU、ASIC、FPGA、NPU、DPU等“XPU”芯片的出现使得算力日趋多元化<sup>6</sup>，传统x86架构之外，ARM、RISC-V、MIPS等多种架构也正在被越来越多的芯片公司所采纳，异构计算随之加速崛起。但是，业界对于未来计算架构的发展方向一直存在争论。在实际商业环境中，大部分企业选择根据业务场景、数据类型、支出成本选择合适的计算架构，如何就新型计算架构的技术路线、标准体系达成行业共识成为必答题。针对现实挑战，“普惠”算力强调立足实际使用需求，安全合理地使用数据，以高效的算法，实现可靠的数字化、智能化效果。

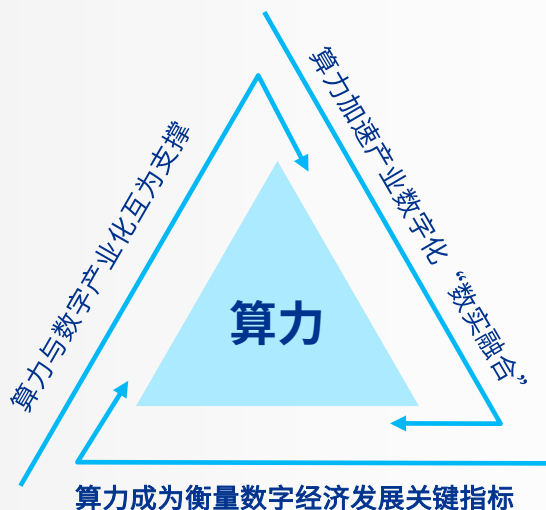
---

<sup>6</sup> CPU（Central Processing Unit，中央处理器）、GPU（Graphics Processing Unit，图形处理器）、ASIC（Application Specific Integrated Circuit，特定用途集成电路）、FPGA（Field Programmable Gate Array，现场可编程门阵列）、NPU（Neural network Processing Unit，神经网络处理器）、DPU（Deep learning Processing Unit，深度学习处理器）。

# 算力是数字经济的基础设施

当今世界正经历百年未有之大变局，数字经济正成为重组全球要素资源、重塑全球经济结构、改变全球竞争格局的关键力量，算力“地基”夯实与否，关系到数字经济这座“大厦”能否巍然屹立。

图 5 | 算力全方位支撑数字经济



资料来源：毕马威分析

**算力成为衡量数字经济发展关键指标。**全球范围内来看，算力对数字经济规模乃至GDP总量的带动作用愈发明显，信通院研究得出，2016-2021年，全球算力规模平均每年增长34%，数字经济规模和GDP每年分别增长8%和4%。作为全球数字经济主导力量之一<sup>7</sup>，中国自2012年以来数字经济增速已连续11年显著高于GDP增速，2022年数字经济规模更是首次突破了50万亿元。各级政府正积极推动数字经济转向深化应用、规范发展、普惠共享的新阶段，其中，普惠共享强调数字经济发展成果更广泛、更公平地惠及全体人民，要求在大规模扩张算力规模的基础上，让算力成为人人用得起、用得好的基础性资源，加强算力智能调度和灵活部署，使算力与算法、数据、应用资源相互协同，向经济社会和产业发展各领域广泛深入渗透，促进数字技术、应用场景和商业模式融合创新，实现全要素生产率提升。

<sup>7</sup> 根据中国信息通信研究院的统计，中国数字经济规模仅次于美国，2021年美国数字经济规模为15.3万亿美元，中国为7.1万亿美元。



**算力与数字产业化互为支撑。**一方面，算力底层融合了集成电路、服务器、数据中心、云计算、人工智能等数字化技术。例如，云计算将算力资源池化，推动算力成为覆盖端、边、云、网全架构场景的泛在能力，人工智能促进算力升级，算力与多模态感知技术融合实现智能化应用。另一方面，随着数字产业化由规模化发展转向高质量发展，大到5G通信、卫星互联网，小到出行线路规划、外卖订单系统优化、影视特效制作，都离不开算力支撑。从超级计算机、数据中心、云计算中心，到智能手机、智能电视等各类智能终端，算力专用性、可拓展性不断增强，有望与各类信息化技术相结合，充分实现性能功耗的平衡，广泛融入各类科研机构和大中小科技企业的产品业务之中，全面推动数字经济发展。

**算力加速产业数字化“数实融合”。**随着实体经济企业借助算力不断将产品、服务和业务流程转化为数据，数据要素深入渗透实体经济肌理，产业数字化转型的关键任务从“数字化”转变为“数智化”。在“数智化”阶段，决定数据要素价值的不再是如何生产数据，而是如何深入挖掘数据价值，不仅要持续消耗巨量算力，还对算力的质量、效率提出了更高要求。质量方面，数字化场景越复杂、数据颗粒度越精细、决策精确度要求越严格，就越需要合理分配算力资源，借助智慧管理系统、智能调度等智能化技术，全方位统筹协调算力服务的部署位置、实时状态、负载信息、业务需求，算力将以更加智慧的方式支撑产业数字化；效率方面，智能终端和传感器的广泛部署使得边缘算力需求更加显著，通过云-边-端协同计算架构将算力下沉到数据源头或关键价值交付点，算力将成为实时流动、快捷可取的资源，高效融入实体经济的方方面面。



# 算力是通用人工智能的核心动力

对AGI发展来说，数据、算力、算法三大核心要素中，算力向下扎根于数据，向上支撑着算法，将是驱动AGI发展的核心动力。

深度学习出现之前，用于AI训练的算力增长大约每20个月翻一番，基本符合摩尔定律；深度学习出现之后，用于AI训练的算力大约每6个月翻一番；2012年后，全球头部AI模型训练算力需求更是加速到每3-4个月翻一番，即平均每年算力增长幅度达到惊人的10倍；目前大模型发展如火如荼，训练算力需求有望扩张到原来的10到100倍<sup>8</sup>，算力需求的指数级增长曲线将更加陡峭。

然而，这也意味着发展AGI需要巨大的算力成本投入，以构建GPT-3为例，OpenAI数据显示<sup>9</sup>，满足GPT-3算力需求至少要上万颗英伟达GPU A100，一次模型训练总算力消耗约3,640PF-days（即每秒一千万亿次计算，运行3,640天），成本超过1,200万美元，这还不包括模型推理成本和后续升级所需的训练成本。

合理且高效地为通用人工智能注入源源不断的核心动力，已逐渐成为产业界共识，算力发展将满足：

## 大规模



在大数据、大模型、大算力的相互作用下，全球人均算力有望实现G-T-P三连跳，即从GFLOPS级别跃升到TFLOPS再到PFLOPS<sup>10</sup>，乃至更高级别。罗兰贝格测算得出<sup>11</sup>，当前全球人均算力<sup>12</sup>约在100-2,500GFLOPS的量级，而当AGI目标基本达成时，智能场景会实现所有行业的全覆盖和大部分行业的高渗透，全球人均算力将超过29TFLOPS。正如工业经济模式下，人均用电量和人民生活质量成正比，在数字经济时代，人均算力也将会是衡量人民智能化生活水平的重要指标，算力需同电力一样成为规模化发展的公共资源。

## 低成本



放眼全球，除了几大科技巨头外，几乎没有企业承担得起大模型训练的高昂成本，算力不实现普惠化，任何关于AI和科技革命的畅想都将是空中楼阁。目前偏大规模的算力需求以B端和G端为主，算力降本需求十分迫切。算力降本并非减少投入，而是利用有限资金获取更多元的算力，同时优化算力资源配置和利用效率。如何实现算力投入产出比最优，给算力供给侧带来挑战的同时，也将催生大量产业机遇。

## 低能耗



数据中心作为大规模算力聚合、调度、释放的重要节点，是各国算力发展的重中之重，但也存在着严重的碳排放问题，随着“碳中和”目标成为全球共识，大部分国家都提出了较为严格的数据中心节能减排要求。电能利用效率（PUE）、水资源利用效率（WUE）、碳利用效率（CUE）等指标已成为各国建设绿色数据中心的重要抓手，例如，中国工信部明确提出“到2023年，新建大型及以上数据中心PUE降低到1.3以下，严寒和寒冷地区力争降低到1.25以下”<sup>13</sup>，未来，只有绿色低碳算力才符合人类社会可持续发展的愿景要求。

<sup>8</sup> ChatGPT到底需要多少算力，华尔街见闻，2023年2月15日，<https://wallstreetcn.com/articles/3681960>

<sup>9</sup> Tom B. Brown, Benjamin Mann, Language Models are Few-Shot Learners, May 28, 2020 (openai.com)

<sup>10</sup> 一个GFLOPS (giga FLOPS) 等于每秒十亿 ( $=10^9$ ) 次的浮点运算，一个TFLOPS (tera FLOPS) 等于每秒一万亿 ( $=10^{12}$ ) 次的浮点运算，一个PFLOPS (peta FLOPS) 等于每秒一千万亿 ( $=10^{15}$ ) 次的浮点运算

<sup>11</sup> 泛在算力：智能社会的基石，华为，2020年2月

<sup>12</sup> 测算范围覆盖美国、中国、印度、日本等27个主要国家

<sup>13</sup> 《新型数据中心发展三年行动计划（2021-2023年）》，工业和信息化部，2021年7月4日，[http://www.gov.cn/zhengce/zhengceku/2021-07/14/content\\_5624964.htm](http://www.gov.cn/zhengce/zhengceku/2021-07/14/content_5624964.htm)

# 普适、智慧相融合，算力将无处不在

“普适”强调以自然交互的方式提供算力，算力将成为人人可得（Affordable）、人人可用（Available）、人人适用（Adaptable）的“3A”型基础资源；“智慧”强调以认知驱动的方式提供算力，算力将具备自适应（Self-adapting）、自学习（Self-learning）、自进化（Self-evolving）为代表的“3S”智能。对于构建“人-机-物”三元融合系统来说，普适和智慧缺一不可，算力终将无处不在。

图6 | 未来算力将在普适和智慧两个维度上加速发展



资料来源：毕马威分析

## 普适



普适计算的目标在于让人们即取即用算力而不必见，要求计算设备感知环境变化，并根据环境变化自动调整算力供给方式，人只需少量干预甚至不用干预算力管理、调度、控制、提供、运维的过程。从技术落地角度来讲，这意味着计算需要“以人为中心”，人们可以通过可穿戴设备、移动设备、小型计算设备等“随时、随地、按需”地获取算力资源，算力表现出大规模、低成本、通用化、共享化等特征。20世纪末，以IBM为代表的IT巨头公司已经意识到普适计算将取代主机计算和桌面计算，成为下一代主流计算模式，纷纷布局普适计算，随着小型计算设备制造、小型计算设备操作系统、移动通信等相关技术日渐成熟，普适计算逐步走向落地。



“普适”一词来自于普适计算，1991年由施乐公司PALOATO研究中心的首席科学家Mark Weiser在The Computer for the 21st Century一文中提出<sup>14</sup>，Mark Weiser认为最深刻的技术是那些已经消失的技术，计算需要融入到日常生活用品中，使人们能够自然地和计算机进行交互。

<sup>14</sup> Mark Weiser, The Computer for the 21st Century, pp. 94 – 104, Sep 1991. (Scientific American)

## 智慧



当前，算力虽然融入了人们的生活空间，人却很难自如地融入物理信息系统（CPS, Cyber Physical System）。具体体现在，人们在使用算力时常常会问“我要如何使用相关硬件和应用”才能满足实际需求，尤其对于企业类客户来说，在解决场景化需求时，问题可能更为棘手。

让人们能按需使用算力而不必问，关键就在于强化算力的“智慧”特征，使算力更自主地理解需求、强化学习、深度思考，即算力应具备自适应、自学习、自进化为代表的“3S”智能。例如，可以让机器模拟人脑神经元的思考过程，推动算力由“感知人”转向“认知人”；通过多模态学习让机器掌握更全面的信息，增强算力认知的广度；通过协同训练算法让机器学会多角度思考，增强算力认知的深度。

## “普适”和“智慧”有机融合

从PC互联网、移动互联网、产业互联网再到元宇宙，实际上是不断构建“人-机-物”三元融合系统（HCPS, Human-Cyber-Physical system）<sup>15</sup>的过程。在HCPS中，物理空间分别与信息空间、社会空间源源不断地进行着信息交互，信息空间与社会空间则进行着计算属性和认知属性的智能融合，因此，算力的“普适”和“智慧”必然会走向融合，在这一过程中，“绿色”“安全”“可信”等全新特征也将融入“普慧”算力的内涵之中，驱动“人-机-物”更加紧密融合，算力最终将无处不在。



<sup>15</sup> “人”是指人通过各种方式联系而构成的社会空间；“机”是由计算机和软件构成的信息空间，包括计算、存储等资源，以及相关的软件服务；“物”指的是物理空间中的各种设施、设备。

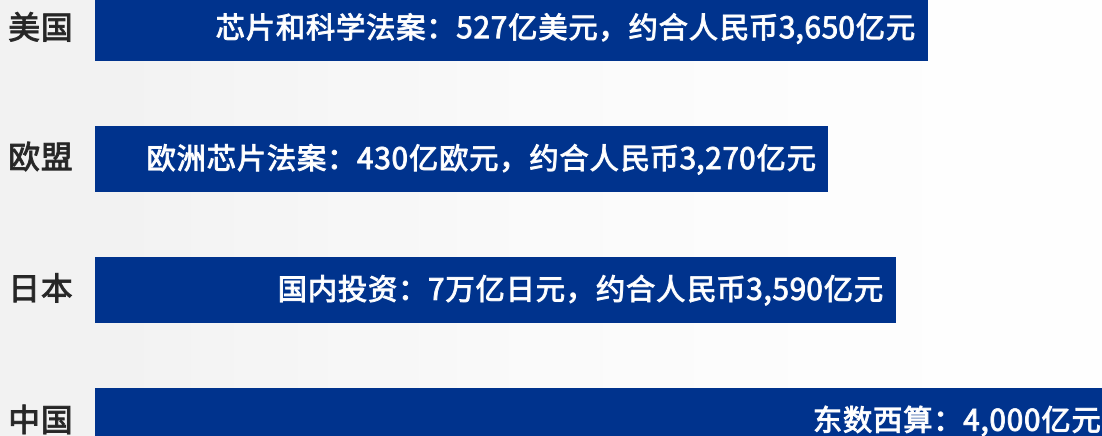
# 03 有为： 大国战略篇



## 算力已成为各国科技战略布局重点

新一轮科技革命深入发展，国际力量对比深刻调整。算力作为数字经济的基础设施，重要性愈发凸显，算力自主可控和高效利用已成为大国科技竞赛的关键制高点。

图 7 | 各国算力投资或补贴计划



资料来源：公开资料，毕马威分析



## 美国： 强化布局多种前沿算力，力保AI等关键科技领域的 领导地位

美国在云计算、数据中心等领域的战略布局较为领先，在2011年和2012年就分别通过《联邦云计算战略》《大数据研究和发展计划》把云计算和大数据上升到了国家战略高度。近年来，美国持续加强布局量子计算、超级计算、类脑计算等前沿算力，很大程度上是为了保障本国在人工智能等关键领域的领导地位。

2019年2月，美国前总统特朗普签订《维护美国人工智能领导地位的行政命令》，明确提出加强对人工智能研发运用的数据和计算资源支撑。同年11月，美国白宫科技政策办公室（OSTP）发布《国家战略性计算计划（更新版）：引领未来计算》，提出了开拓数字和非数字计算的新领域（如量子计算、神经形态计算等），开发、扩展和推进计算基础架构和生态系统的目标，并从计算机硬件、软件和整体基础设施，以及开发创新的、实际的应用程序和机会等方面，给美国政府提出了详细建议。紧接着，美国白宫在2020年11月相继发布《关于利用云计算资源推进联邦资助的人工智能研发的倡议》《引领未来先进计算生态系统：战略计划》回应特朗普的行政命令和OSTP提出的目标及建议，由此，美国确立了对未来先进计算系统的设想，目标是维持其在科学工程、经济竞争和国家安全方面的领先优势，并提出了以政府、学术界、非营利组织、产业部门共同参与为主的举国方案。

在人工智能方面，美国政府于2021年1月将《2020国家人工智能倡议法案》加以修订后纳入《2021财年国防授权法案》，人工智能倡议法案正式升级为法律，标志着美国政府最终完成人工智能战略部署及政策设计。此后，美国政府陆续成立了国家人工智能行动办公室、人工智能咨询委员会、人工智能机构间委员会等职能机构，建立起了系统化的管理协调机制。

在半导体方面，美国总统拜登在2022年8月正式签署《芯片和科学法案》，有意通过政府扶持本土半导体产业，以增强美国半导体产业竞争力，计划为美国半导体研发、制造以及劳动力发展提供527亿美元，其中390亿美元将用于半导体制造业的激励措施，110亿美元用于补贴半导体研发，20亿美元用于汽车和国防系统芯片，5亿美元用于建设可靠的半导体供应链补贴，2亿美元用于培养半导体行业人才。



## 欧洲：

### 发展可持续半导体、高性能计算、安全云，助力数字化转型

欧洲算力产业布局基本围绕“数字化转型”这一核心，自2020年初，欧盟陆续发布《塑造欧洲的数字未来》等一系列文件，全方位多层次地设计了欧洲各领域的数字化转型方案。2021年3月欧盟委员会发布《2030数字罗盘》，较为系统地指出了欧盟到2030年要实现的量化目标，其中，数字基础设施、企业数字化两方面的目标需要算力相关产业带来强力支撑：

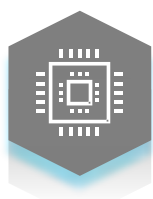
- 构建安全、高性能和可持续的数字基础设施。到2030年，欧盟生产的尖端、可持续半导体产业的产量至少占全球总产值的20%（产能效率将是目前的10倍）；到2025年，生产出第一台具有量子加速功能的欧洲量子计算机，到2030年，欧洲处于量子领域前沿。
- 致力于企业数字化转型。到2030年，75%的欧盟企业应使用云计算服务、大数据和人工智能。

针对半导体产业，欧盟着力提高半导体产业链生态的弹性，从而减少外部依赖。2022年2月欧盟委员会发布《欧洲芯片法案》，再次重申半导体全球份额目标的同时，制定了详细的半导体生态建设规划，预计撬动总投资金额超430亿欧元。

针对量子计算机研制，欧盟一直都十分注重对高性能计算、量子计算等前沿技术的研发投入。早在2018年1月欧盟就通过“高性能计算共同计划”，提出在2022-2023年开发具有准百亿亿次性能（E级计算）的高性能计算系统，并在同年10月成立了欧洲高性能计算联合执行体（EuroHPC JU），该组织专门负责部署和运行世界级的高性能计算和数据基础设施。2020年欧盟委员会出台针对EuroHPC JU的新章程，将量子计算基础设施补充进入重点工作内容，拟投资80亿欧元以支持E级计算和量子计算等前沿计算研究。

针对企业数字化转型，欧盟强调构建安全、高效、可互操作的云服务，持续推进欧洲云计划GAIA-X，为欧洲提供安全可信的数据基础设施，尤其是在工业领域，欧盟相继推出《欧洲工业战略》《新欧洲工业战略》，持续推动工业数据、边缘和云联盟等重点项目。





## 日本： 重振半导体产业，坚持以应用、绿色为导向发展算力

日本近年来频繁强调振兴半导体产业，2021年6月，日本经济产业省发布《半导体和数字产业发展战略》，提出将半导体和数字产业定位于基础性产业，作为国家重要事业，紧接着又在2021年底提出了“三步走”实施方案，可概括为“恢复半导体产能、推动下一代半导体发展、为未来技术奠基”，而以半导体产业为核心，日本算力产业布局呈现出两大特征：

**应用导向。**从市场层面来看，日本主要面向具有高增长潜力的新兴应用需求，部署尖端半导体的研发设计、产能扩充方案，重点应用领域包括边缘计算、物联网、自动驾驶等，还强调根据应用场景和服务类型构建有针对性的云服务体系。从社会层面看，日本大力促进算力的灵活泛在应用，推动日本数字社会向“以应用为中心”转变，早在2016年日本就提出促进云计算服务关联技术的灵活应用，日本“六五计划”<sup>16</sup>提出“使任何人都能随时随地方便、安全地使用大数据和人工智能”和“最大程度地利用数据和人工智能等技术”，同样体现了以应用为导向发展算力的鲜明特征。

**绿色导向。**为实现碳中和、实现经济社会早日脱碳，日本强调数字化和绿色化协调发展。在发展半导体器件方面，要求兼顾性能提升和功耗降低，尤其是针对关键零部件功率半导体，指出必须从巩固本国产业竞争力的角度，研发新型创新材料（SiC、GaN、Ga<sub>2</sub>O<sub>3</sub>）。在建设数据中心方面，则大力推进节能措施和数据中心节能绩效评价标准，引导政府采购和民间投资向绿色化方向发展等，并计划通过“绿色创新基金”（2万亿日元）等推动“下一代数字基础设施建设”等重点项目。

2021年日本经济产业省邀请台积电在熊本县设厂并提供了6,000亿日元的资助，2022年日本丰田汽车株式会社等八家公司投资73亿日元成立半导体公司Rapidus，日本经济产业省提供了700亿日元的补贴。2022年底日本政府表示，将投入约7万亿日元来扩大日本国内投资，主要包括为日本国内尖端半导体生产提供投资支援等<sup>17</sup>。

<sup>16</sup> 即“第六期科学技术与创新基本计划”，是指导日本2021-2025年科学技术与创新发展的纲领性规划。科技创新基本计划是日本政府依据《科技基本法》制定的科技振兴计划，自1996年以来，该计划根据世界科技形势、日本科技发展状况的变化，每5年更新一次。

<sup>17</sup> 日本政府将投7万亿日元促企业国内投资，日经中文网，2022年12月9日，<https://dev.cn.nikkei.com/politics/economy/economic-policy/2022-12-09-10-10-41.html>



## 中国：

# 算力融入数字中国蓝图，“东数西算”织就全国一体化算力网络

科技自立自强是中国发展的战略支撑，自2020年4月算力被纳入“新基建”范畴以来，国家依靠自主创新持续夯实算力基础。2023年2月国务院印发《数字中国建设整体布局规划》，提出要夯实数字中国建设基础，系统优化算力基础设施布局，促进东西部算力高效互补和协同联动，引导通用数据中心、超算中心、智能计算中心、边缘数据中心等合理梯次布局。

新型数据中心<sup>18</sup>是中国建设算力基础设施的重要抓手，《新型数据中心发展三年行动计划（2021-2023年）》明确提出，“用3年时间，基本形成布局合理、技术先进、绿色低碳、算力规模与数字经济增长相适应的新型数据中心发展格局”。在此过程中，对算力产业实际影响最大的莫过于“东数西算”工程，该工程通过构建数据中心、云计算、大数据一体化新型算力网络体系，将东部算力需求有序引导到西部，对于优化算力基础设施布局具有重要意义。

随着“东数西算”工程正式全面启动<sup>19</sup>，中国有望建成国家级的算力网络，将分散在通用数据中心、超算中心、智能计算中心、边缘数据中心的算力资源整合起来，真正将算力转变为水、电、燃气一样的基础性资源，将算力以最简单高效、低成本的方式输送到算力需求侧。根据国家发展改革委高技术司公布的信息，截至2022年8月，8个国家算力枢纽均进入深化实施阶段，新开工数据中心项目达60余个，新建数据中心规模超110万标准机架，项目总投资超4,000亿元<sup>20</sup>。

继东数西算之后，超算互联网有望成为建设数字中国算力基础设施的另一大重点工程，2023年4月科技部启动国家超算互联网部署工作，提出用互联网思维运营超算，将全国众多超算中心通过算力网络连接起来，构建一体化算力服务平台，计划到2025年底，国家超算互联网将可形成技术先进、模式创新、服务优质、生态完善的总体布局。

<sup>18</sup> 新型数据中心是指以支撑经济社会数字转型、智能升级、融合创新为导向，以5G、工业互联网、云计算、人工智能等应用需求为牵引，汇聚多元数据资源、运用绿色低碳技术、具备安全可靠能力、提供高效算力服务、赋能千行百业应用，与网络、云计算融合发展的新型基础设施。

<sup>19</sup> 正式启动！“东数西算”工程全面实施，中国政府网，2022年2月17日，[http://www.gov.cn/xinwen/2022-02/17/content\\_5674322.html](http://www.gov.cn/xinwen/2022-02/17/content_5674322.html)

<sup>20</sup> 国家发展改革委高技术司牵头“东数西算”工程进展情况（2022年8月），发改委，2022年9月23日，[https://www.ndrc.gov.cn/fzggw/jgsj/gjss/sjdt/202209/t20220923\\_1336061.html](https://www.ndrc.gov.cn/fzggw/jgsj/gjss/sjdt/202209/t20220923_1336061.html)

# 全球算力产业竞速，中国有望迈入十亿级赛道

算力产业链涵盖由基础硬件、基础软件、基础设施构成的上游产业，由第三方数据中心、云计算服务、网络运营服务、IT外包服务、系统集成服务构成的中游产业，由互联网、制造业、金融、能源等各行业企业用户构成的下游产业。

图 8 | 算力产业链构成

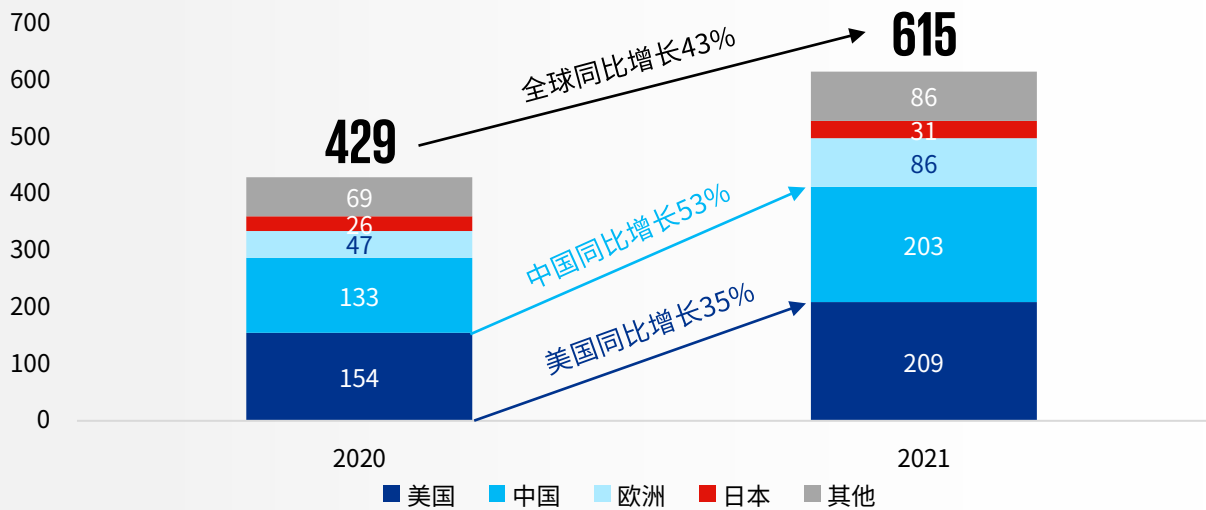


资料来源：公开资料，毕马威分析

算力下游各类业务场景中智能硬件、工业设备、智能汽车等多类型终端大量接入网络，对计算实时性、可靠性、吞吐能力、能耗等的要求愈发复杂，算力需求规模不断扩张的同时，呈现出明显的专用化、边缘化、智能化等特点，以通用计算为主的传统计算范式已然无法满足需求，将驱动算力供给侧不断提高对需求的适配性，构建由通用算力、智能算力、高性能算力、边缘算力等构成的多元算力供给体系。

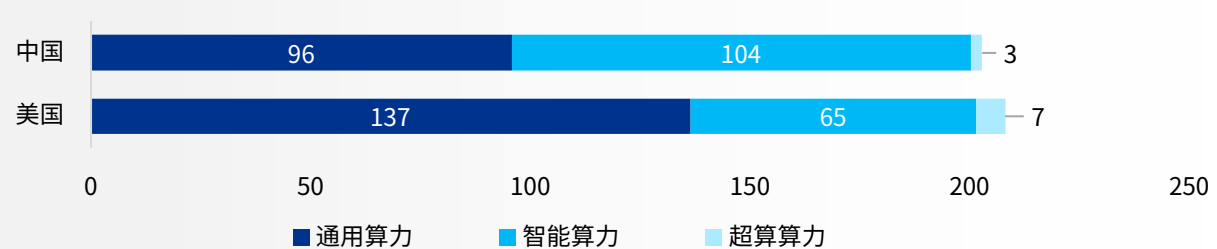
根据信通院测算数据，美、中、欧、日基本稳居全球算力产业规模前四名，美、中两国处在领先地位且中国算力规模增速明显领跑。

图 9 | 2020-2021全球算力规模分布情况（单位：EFLOPS）



此外，对比中美两国细分算力规模来看，中国的智能算力规模已经超过美国60%，主要得益于人工智能模型训练算力需求快速增长背景下，中国不断加快以智能计算中心为代表的智能算力基础设施建设。公开资料显示，截至2023年2月，中国投入运营和在建的人工智能计算中心已达23个<sup>21</sup>，结合IDC预测来看，2021-2026年中国智能算力规模年复合增长率有望达到52.3%，预计未来3-5年中国算力发展仍主要由智能算力拉动，且有望继续保持全球领先地位。

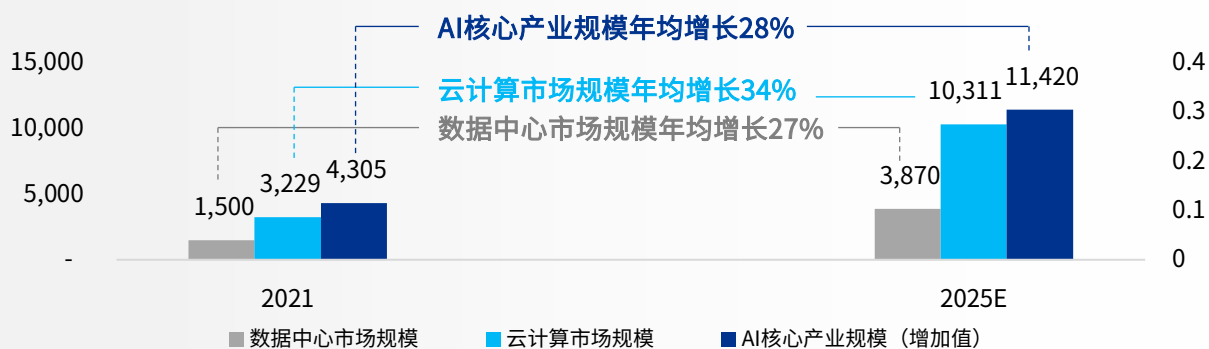
图 10 | 2021中美细分算力规模对比（单位：EFLOPS）<sup>22</sup>



当前，中国已初步建立起了产业链条持续完善、产业创新较活跃、市场规模迅速增长、国际竞争力稳步提升的算力产业生态。信通院数据显示，2021年中国算力核心产业规模已超1.5万亿元，关联产业规模超过8万亿元，主要涉及数据中心市场规模1,500亿元，云计算市场规模3,229亿元，人工智能核心产业规模4,305亿元。

<sup>21</sup> 行业洞察 | 2023年中国人工智能计算中心分布概览，至顶网，2023年2月15日，<https://insights.zhiding.cn/2023/0215/3147551.shtml>

<sup>22</sup> 根据信通院测算口径，基础算力主要是基于CPU的存量服务器算力水平，智能算力主要是基于GPU、FPGA、ASIC或其他加速器支撑的高并行、高密度计算能力的异构算力水平，超算算力主要是基于国际知名排行榜TOP500、中国高性能计算机性能TOP100统计的超算算力规模。

图 11 | 中国数据中心<sup>23</sup>、云计算、人工智能市场规模预测（单位：亿元）

数据来源：基于信通院公开历史数据预测得出，毕马威分析

预计“十四五”期间，大数据中心投资将以每年超过20%的速度增长，累计带动各方面投资将超过3万亿元<sup>24</sup>。经初步测算，到2025年，数据中心、云计算、人工智能市场规模总计将突破2.5万亿元，算力核心产业规模将不低于4.4万亿元，关联产业规模可达24万亿元。

横向对比来看，能够突破万亿元乃至十万亿元大关的高潜赛道屈指可数，算力产业景气度尤为亮眼。

能源革命作为数字革命之外另一大社会变革主流趋势，正强力带动风电、光电、氢能、储能、新能源汽车等风口，业界对于这些风口的规模预期均在万亿元以上。以新能源汽车为例，结合中国汽车工业协会统计数据，2022年中国新能源汽车销量为688.7万辆，占汽车新车销售总量比例已达25.6%，预计到2025年该比例有望达到30%，假设新能源汽车未来三年增速维持在50%以上，2025年中国新能源汽车销量将不低于2,300万辆，市场规模将跃升至万亿级。

此外，就数字经济各细分赛道来说，《数字中国发展报告（2022年）》数据显示，当前中国数字经济规模为50.2万亿元，其中，10万亿元以上的赛道只有电子信息制造业和软件业。也就是说，算力关联产业有望崛起为数字经济蓝图中规模突破十万亿元的第三大超级赛道。

图 12 | 预计2025年中国算力关联产业规模可达24万亿元



数据来源：基于信通院公开历史数据预测得出，毕马威分析

<sup>23</sup> 数据中心市场规模主要依据数据中心牌照企业的收入数据统计计算，不包含云计算业务收入

<sup>24</sup> 高技术产业延续增长势头主要指标亮眼，发改委，2022年5月6日，[https://www.ndrc.gov.cn/wsdwhfz/202205/t20220506\\_1324276.html](https://www.ndrc.gov.cn/wsdwhfz/202205/t20220506_1324276.html)



# 04 识局： 发展评估篇



本报告通过构建“普惠”算力行业发展评估框架，结合定量分析和定性调研，研判各行业“普惠”算力需求及差异。评估结果有利于算力供给侧积极应对算力需求变化，及时调整市场战略，推动算力产业沿着“普惠”算力的方向持续发展。

## 设施、应用、服务是算力的三大关键要素

“普惠”算力包括三大关键要素：



在三大关键要素的作用下，算力将经历从产生到落地的完整闭环，彻底转化为新型生产力。



图 13 | “普惠”算力三大关键要素



资料来源：公开资料，毕马威分析

## 算力设施

算力设施是算力的基础，主要包括芯片、服务器、数据中心等底层硬件，以及操作系统、数据库等底层软件，如果说底层硬件决定了算力的物理极限，底层软件则决定着如何释放算力。随着算力硬件设计越来越复杂，底层软件越来越难以驾驭，要兼顾硬件的高性能和软件的灵活性，亟需实现软硬件协同创新和优化。符合“普惠”算力要求的算力设施将具备以下特征：



### 多元异构

大数据、物联网、人工智能、元宇宙等迫切要求释放多元算力，以CPU、GPU、FPGA、ASIC为代表的定制化芯片不断涌现，通用算力和专用算力出现明显分流。从源头开始，“普惠”算力就包含着多元化基因，硬件系统随之将朝着超异构、超融合和存算通信等多维一体的方向发展，以实现最优效率和性能比。



### 软硬件协同<sup>25</sup>

“人-机-物”融合共生关系不断深化，海量、多源、异质、异构的资源，使得资源管理难度呈指数级增加，软硬件协同设计工具链将成为高效管理各类硬件资源、支持多类型资源和场景有效隔离的重要创新手段，从而提高算力可扩展性，实现算力跨平台部署、多场景兼容。



### 绿色集约

面向双碳目标，平衡算力提升和能耗降低成为关键命题，通用数据中心、智能计算中心、高性能计算中心作为规模化算力的重要载体，在保证业务性能的前提下，尽可能提升服务器等的CPU利用率，实施集约化、精细化、智能化管理，充分利用可再生能源是必然趋势。



### 云边端一体化

数据中心之外，算力载体已涵盖“云端数据中心”“边缘数据中心”“智能手机”“智能传感器”等多种形式，算力部署呈现出云边端三级架构。通过云边端一体化，即云端负责统一管理和大规模集中式计算，边缘进行数据敏捷接入和实时计算，终端实现泛在感知和本地智能，能有效推动“普惠”算力真正满足各类场景需求。

<sup>25</sup> 泛在操作系统实践与展望研究报告，北京大学、腾讯研究院，2022年8月



## 算力应用

算力应用实现了算力的“功能设计”，主要包括各类应用软件，能够赋予算力解决问题的能力，随着计算机视觉、智能语音、自然语言处理等不断嵌入应用软件之中，算法已逐渐成为应用软件的灵魂，在各类智能应用软件中，“普惠”算力的功能呈现出了明显的升级趋势：

### 从感知型算力到认知型算力

基于各类智能传感器，算力不仅能执行计算和存储任务，还具备了视觉、听觉、触觉等感知能力。随着知识图谱、图计算等技术不断成熟，算力将具备更多思维逻辑和认知能力，从而更好地理解人类指令，可以预见，认知型算力将大幅提升ChatGPT回答的准确率。

### 从应用型算力到通用型算力

以ChatGPT为代表的生成式AI更加成熟，意味着向AGI更近一步，AGI赋能应用软件成为可能，相应地，算力也能从面向特定领域、特定问题的应用型算力，转变为通用型算力。

## 算力服务

算力服务代表着算力的提供方式，是决定“普惠”算力最终能否转化为新型生产力的临门一脚，针对算力需求激增、硬件性能提升受限等复杂情况，产业界的当务之急是设计出合理的算力服务模式，降低算力服务门槛，实现多方利益最大化。“普惠”算力有望成为水电一样的可交易资源，提供方式能部分参考水电，但远比水电复杂，目前主要包括三大类型：

图 14 | 2022年中国三类算力服务规模现状

算力提供方式	算力提供方	服务器数量 (万台)	机架数量 (万个)
新IT服务	算力设备等IT服务商	合计约200	
	数据中心服务商		
	地方大数据产业园		
云计算服务	阿里云	200-300	
	腾讯云	100-200	
	字节跳动	约100	
	百度	60-80	
	华为云	约100	
算力网络服务	中国移动	71	46.7
	中国电信	100-200	51.3
	中国联通	约50	36.3

资料来源：财经十一人，毕马威分析

## 新IT服务

传统算力设备制造商等IT服务企业习惯简单销售硬件，但随着各类实体经济产业数字化转型逐渐深入，IT服务企业对内整合产品、技术等资源，对外提供一站式解决方案成为主流趋势。以“端（智能物联网终端）—边（边缘计算）—云（云计算）—网（5G）—智（行业智能）”为代表的新IT技术架构逐渐成为市场共识，在新IT架构下，算力可以实现跨层级、跨领域、跨生态赋能。从落地角度来看，IT服务模式的目标客户比较适合定位为中大型企业类用户，主要原因有二，一是此类用户资金预算和时间预算都比较充足，对前期成本投入大、回报周期长、建设流程复杂等情况接受力较强；二是此类用户往往在自身产业链生态中具有一定影响力，可以借助新IT架构对外共享“普惠”算力等资源，带动产业生态协同发展。

## 云计算服务

云计算服务主要通过分布式计算、虚拟资源管理、云原生等技术，实现服务器、存储、网络等基础设施资源的池化、抽象化和标准化，在整合规模化算力的基础上，成功将算力转化成虚拟商品，算力需求方可以结合自身资金实力和IT基础建设情况，按需采购算力，能极大节约算力使用成本。云计算服务模式的潜在客群以中小型企业为主，主要原因就是此类用户成本控制较为严格，云计算具备按需服务、灵活计费、泛在接入等特点，能有效推动“普惠”算力低成本普及。

## 算力网络服务

算力网络<sup>26</sup>的目标是建成水网、电网一样的网络，使算力“实时接入、即取即用”，该建设目标和“普惠”算力发展需求最为接近，有利于推动“普惠”算力渗透到广大个人用户群体中，但目前算力网络的落地还亟需解决“广覆盖”“深应用”两大问题。在广覆盖方面，数据中心等基础设施的布局应当保证用户能平等获取算力，必然要依靠国家机器强力推进，中国“东数西算”工程就是典型示例。在深应用方面，算力一旦成为公共服务必然要面对大量差异化需求，因此算力网络建设过程中，要兼顾通用算力、边缘算力、智能算力、高性能算力等多元算力，合理规划一体化大数据中心、人工智能计算中心等基础设施布局，而现阶段，国内外对于算力网络建设的技术标准、建设思路仍在探索之中。

<sup>26</sup> 算力网络指网络和计算的高度协同，将计算单元和计算能力嵌入网络，实现“云、管、边、端”的高效协同，提高计算资源利用率。

# “普慧”算力行业发展象限图

基于以上分析，我们将算力拆分为“普”和“慧”两大核心维度，并以算力设施、算力应用、算力服务三大关键要素为一级指标，进一步细分成13项二级指标，并根据专家打分、文献分析等设置了各指标的相对权重。







图 15 | “普慧”算力行业发展评估框架

维度	一级指标	二级指标
普	算力设施	硬件设施（芯片、服务器、存储器等）
		软件设施（操作系统软件、中间件等）
	算力应用	管理类应用软件（ERP/SCM/CRM）
		通用类软件（办公、数据分析等）
	算力服务	IT架构实施服务
		云计算IaaS服务
企业级网络部署服务		
慧	算力设施	高性能硬件设施（AI服务器、高性能服务器等）
		定制化软件设施（专用操作系统软件等）
	算力应用	智能应用软件
		定制化应用软件
	算力服务	一站式IT实施和托管服务
		云边协同服务

资料来源：毕马威分析

在具体评估过程中，各级指标的基础分数主要以2023-2026年IT投入金额的预计增速为关键参考，数据范围覆盖ICT、制造、汽车、医疗、教育、金融六大行业，最后通过对各级指标的加权计算得出了各行业在“普”和“慧”两大维度上的算力需求潜力综合评分结果。

图 16 | 2023年各行业IT投入规模及增速预测

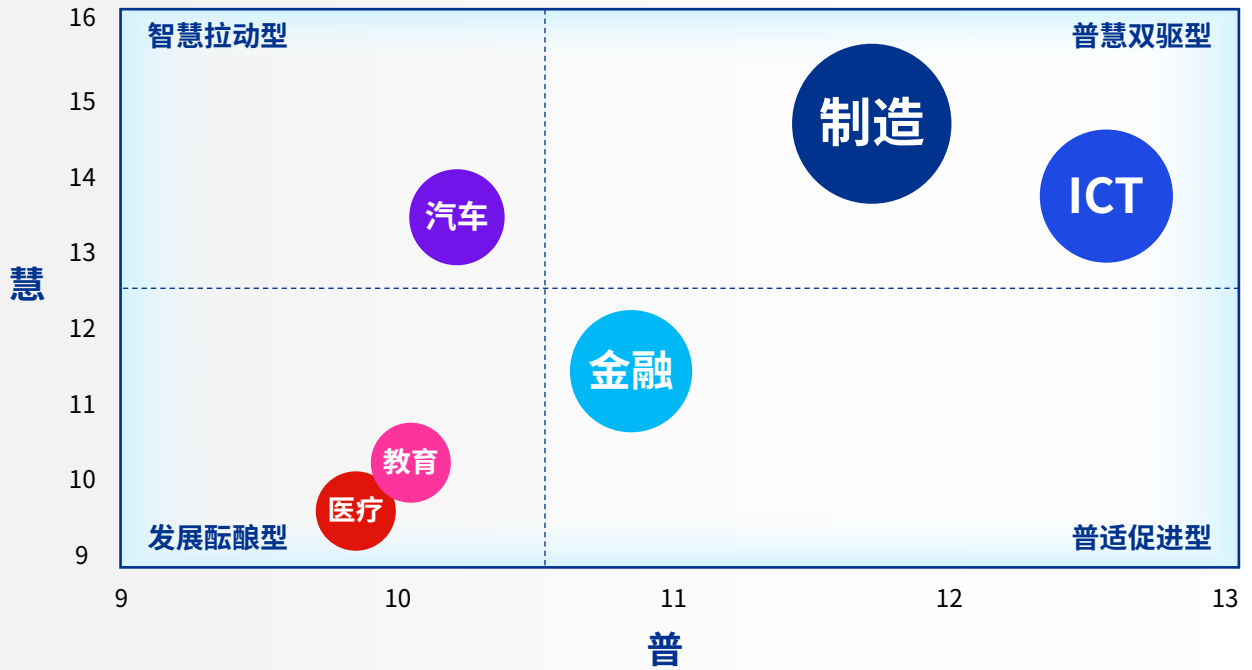
2023年 IT投入预测金额		2023-2026 CAGR		2023-2026 CAGR
5,000 亿元以上	制造 	10%	ICT 	12%
3,000-5,000 亿元	金融 	9%		
1,000-3,000 亿元	汽车 	10%		
1,000 亿元以下	教育 	9%	医疗 	10%

数据来源：Gartner，毕马威分析

\* 气泡大小代表各行业2023年IT投入预测金额的相对规模

通过在象限图横轴反映各行业在“普”维度上的评分，在纵轴反映各行业在“慧”维度上的评分，可将各行业对“普惠”算力的需求分为四类，分别是“普惠双驱型”“智慧拉动型”“发展酝酿型”“普适促进型”。

图 17 | “普惠”算力行业发展象限图



\* 气泡大小代表各行业2023年IT投入预测金额的相对规模，一定程度上能反映当前各行业“普惠”算力需求规模对比情况  
数据来源：Gartner，行业调研，毕马威分析



## 普惠双驱型：“普惠”算力需求高地



ICT行业汇集海量数据资源，促进数据要素价值释放过程中亟需“普惠”算力

ICT包含5G、物联网、云计算、大数据、人工智能等新一代信息技术，涉及通信业、电子信息产业、互联网等众多关联行业。在数字化转型浪潮中，海量优质的数据资源在ICT行业产生、汇集，又随着ICT赋能传统行业数字化转型的步伐走向流通、共享。

随着“数据二十条”落地实施，中国将持续充分发挥海量数据规模和丰富应用场景的优势，激活数据要素潜能，做强做优做大数字经济，围绕释放数据要素价值，ICT行业将产生大量“普惠”算力需求。

在普惠方面，科技巨头不断加码算力进行大模型训练的同时，大量中小型企业因资金和技术限制被拦在门槛之外，但在大模型走向场景化、实用化的过程中，将会形成“大模型+小模型”的产业生态，即大型企业负责搭建底层通用大模型，中小型企业在大模型的基础上搭建面向特定任务场景的小模型，相较于大模型而言，小模型训练的资金投入更低，因此，服务此类企业小模型训练需求的算力增量仍然可观。

在智慧方面，自深度学习、机器视觉等技术不断普及以来，ICT企业一直是各类“智能+”“AI+”创新应用的主要推动者，具备较强的智能化解决方案能力。随着To C市场的流量天花板逐渐显现，ICT企业纷纷转战To B市场，面向制造业等传统行业进行数字化转型赋能，业务类型从传统的门户网站、视频、社交转向智能制造、智慧交通、智慧城市等。To C类业务一般仅对存储和访问要求较高且需求较为同质化，相比之下，To B类业务对网络带宽、计算时延、灾备容量等指标的需求差异性更为显著，要求在算法及软件层面融合更多垂直行业洞察（know-how），使得行业算力需求从通用型转向专用型。





## 制造业数字化场景丰富，“普惠”算力需求分梯队、分阶段、分批次爆发

“十四五”时期是中国建设制造强国，实现经济高质量发展的重要阶段，智能制造将是主攻方向。中国制造业具有产业体系完善、产业链条长、产业集群多等特点，相应地，制造业数字化场景十分丰富，在数字化转型方面有望充分实现规模经济和范围经济。伴随着系统性产业变革的发生，“5G+工业互联网”“百万企业上云”“两化融合”等将带动智能制造场景全面拓展，制造业正成为“普惠”算力最大的潜在市场。

制造业算力需求表现为“普惠双驱”，但在“慧”维度上的评分明显高于“普”，原因是中国智能制造发展具有“龙头先行，中小微跟进”、“先小步慢跑，再大步快跑”、“先单点应用，再逐步推广”的特点，使得算力需求分梯队、分阶段、分批次爆发，不同主体在不同阶段、不同场景中对算力特征的要求各不相同。

具体来说，龙头类企业的数字化转型成熟度更高，算力需求规模较大且侧重强调特定场景下算力的智能化升级；多数中小微企业尚处在数字化转型初期，算力需求基本以降本增效、可快速落地为主要目的，多面向数据通联互通、可视化分析等基础性场景。可以说，强化算力的智慧特征关乎数字化转型的深度，强化算力的普适特征关乎数字化转型的广度。

在智慧方面，边缘智能是大势所趋，通过在边缘计算平台上进行智能推理加速和多节点智能训练算法的联动，能实现本地轻量级、低延时的数据分析和优化决策，将推动算力在高可靠、高安全等场景中深化应用。此外，中国工业软件还需加速突破，在利用深度学习等技术使工业知识与工艺实现软件化的过程中，云原生、工业互联网、数字孪生等技术会不断向具体场景渗透，也会推高对专用化、定制化算力的需求。

在普适方面，在国家不断强化中小企业数字化赋能专项行动、助推中小企业专精特新发展的背景下，中小型企业算力需求有望持续增长，而此类企业往往不会选择自建算力平台，相关需求主要由云计算等ICT企业和制造业龙头类企业承接，前者主要面向跨行业的中小制造企业提供IaaS、PaaS、SaaS等多类型算力服务，后者主要面向自身产业生态中的中小制造企业共享算力资源，有利于提升垂直产业链的整体数字化水平。



## 智慧拉动型：“普慧”算力以智取胜



### 汽车成下一代移动智能计算终端，软件定义汽车的发展趋势要求算力实现智能升级

中国汽车产销总量已连续14年居全球第一<sup>27</sup>，汽车消费在加快构建以国内大循环为主体、国内国际双循环相互促进的新发展格局中发挥着重要作用。当前汽车呈现出“电动化、网联化、智能化”三大趋势，有望成为继PC、智能手机之后的下一代移动智能计算终端，预计汽车智能化和汽车产业数字化转型会是“普慧”算力需求爆发的两大主阵地，相关算力需求呈现出较明显的“智慧拉动”特征。

在汽车智能化转型方面，算力需求会随着自动驾驶等级的提升呈现阶段式爆发，L2、L3、L4、L5对应的车载芯片算力起点分别是2TOPS<sup>28</sup>、24TOPS、320TOPS、4,000TOPS，大算力芯片已经成为自动驾驶的必备基础。而随着“软件定义汽车”时代的到来，单纯堆砌硬件已无法满足车载操作系统、智能座舱等软件系统不断升级的需求，强化算力算法实时协同正在成为各路造车势力的比拼重点，“人-车-路-网-云”五维一体化使算力不再局限于单车智能，亟需算力实现跨区、跨级的“连接+感知+计算”。预计随着自动驾驶等级和智能汽车渗透率的不断提升，将持续带来芯片需求和边缘计算中心等底层设施建设需求。

在汽车产业数字化转型方面，汽车消费者的个性化、多元化需求倒逼上游车企转向柔性生产，研发设计、生产制造、营销、服务等环节的不确定性和复杂性大幅提升，车企生产数字化对算力的场景适配能力提出了更高要求，上升到产业链层面来看，制造端、供应端和消费端的数据和信息亟需横向拉通，算力有望成为产业生态级共建共享的底层资源，需要更智慧的算力实现多节点、多链条、多组织的精准赋能。

<sup>27</sup> 我国汽车产销总量连续14年居全球第一，中国政府网，[http://www.gov.cn/xinwen/2023-01/12/content\\_5736536.htm](http://www.gov.cn/xinwen/2023-01/12/content_5736536.htm)

<sup>28</sup> Tera Operations Per Second，1TOPS代表处理器每秒钟可进行一万亿次（ $10^{12}$ ）操作



## 普适促进型：“普惠”算力持续进化



### 金融回归服务实体经济本质，“场景+金融”要求接入广泛且灵活的安全算力

服务实体经济，是金融的宗旨。近年来，金融与科技走向深度融合，带来了大量新应用、新模式、新业态，实现了金融与实体经济共荣共生，金融科技已成为各大金融机构重点投入方向。据最新年报数据，2022年6家国有银行在金融科技方面的投入均超百亿元，其中工商银行投入金额最高，约为262.24亿元，此外，多家股份制银行的投入同比增长均超20%。预计随着数据元素注入金融服务全流程，金融业的交易量、数据量和每秒峰值会呈现指数级增长，计算需求将持续迸发且愈发复杂多样，面向金融场景的算力服务正迎来创新窗口期。

金融的算力需求具有明显的“普适促进”特征，原因是金融机构正积极将金融服务嵌入到个人和企业用户的多元场景之中，为之提供全时段、全方位的敏捷响应服务，数字金融服务加速转向“场景+金融”模式，倒逼着算力供给侧实现算力的多节点部署，提供无处不在的接入能力。此外，由于金融发展一直是转型提速与监管趋严并重，在IT架构、网络基础、数据存储和管理等方面存在着严格的行业标准，因此金融机构数字化转型的规划、建设和运维均会大量采购外部专业性服务，“金融数据安全”“用户隐私保护”等拉高了算力服务（尤其是安全类算力服务）的需求潜力。

值得注意的是，以CPU为通用服务器的传统计算方案，已难以满足所有场景需求，部分特定场景需借助专用芯片，获取更适配、更灵活的算力支持，例如基于FPGA（Field Programmable Gate Array，现场可编程门阵列）芯片的硬件加速方案，正试点落地券商极速行情系统、经纪订单系统等细分场景，为金融量化交易带来高性能算力支撑。预计随着“场景+金融”模式走向成熟，金融算力需求将转变为“普惠双驱型”。





## 发展酝酿型：“普惠”算力后劲十足



### 教育、医疗的算力需求来自普罗大众，“普惠”算力降本后有望迎来爆发

教育、医疗关乎民生福祉，在“中国教育现代化2035”“健康中国2030”等战略引领下，近年来中国一直积极推动教育、医疗信息化工作，保障每个公民都能够平等地享有教育、医疗等领域的资源，有望带来巨大的“普惠”算力需求，并能真正让算力实现人人可得、人人可用、人人适用。

由于算力使用门槛较高、公共型算力设施建设尚不完善等问题，目前教育、医疗领域的很多使用者仍未能真正感受到算力的有效服务，“普惠”算力需求潜力还在酝酿阶段，预计未来一段时间内两大行业的算力落地需以降低成本、提高普及率为主要途径。

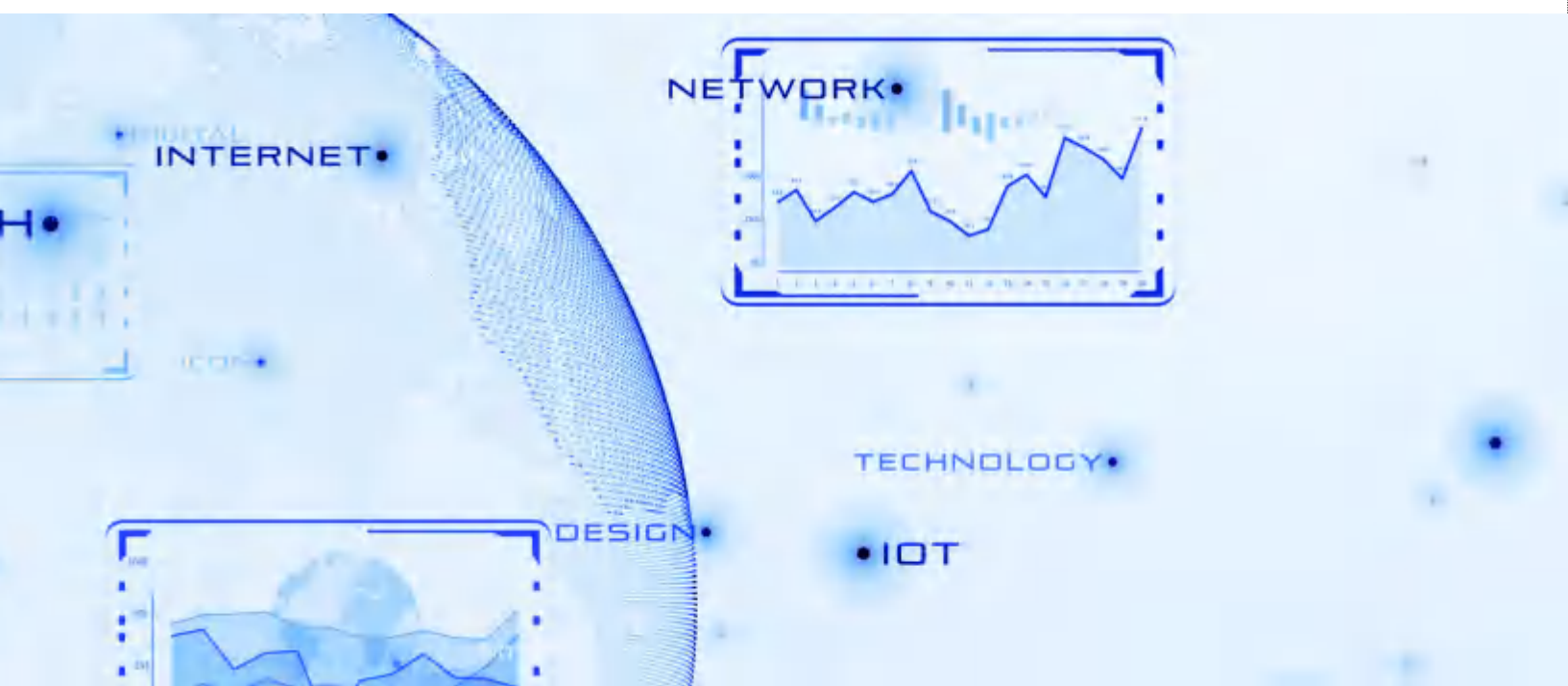
以高校信息化为例，目前大部分中国高校缺乏自建算力，且许多高校老师、信息中心员工等最终用户对算力来源等情况尚未形成清晰认知。中国教育和科研计算机网2022年的调研数据显示<sup>29</sup>，接受调研的1,998所高校中，91.6%的高校未建设校级算力平台；从校外算力资源使用情况来看，已使用校外算力资源的高校仅占6.8%，未使用校外算力资源的高校占比高达40.8%，剩下52.4%的高校受访人员则表示说不清。而在这些高校提出的算力服务优化建议中，“性价比高”“资源共享”“云计算”“免费测试”等关键词在一定程度上表明算力亟需降本。

在医疗信息化方面，目前中国优质医疗资源总量不足、城乡之间分布不均衡的问题仍较明显，根据《第51次中国互联网络发展状况统计报告》<sup>30</sup>，截至2022年12月，全国互联网医疗用户占网民整体规模的比例为34.0%，农村地区互联网医疗用户占农村网民规模比例为21.5%。预计在“到2030年，远程医疗覆盖省市县乡四级医疗卫生机构”<sup>31</sup>的政策目标驱动下，优质医疗资源还将进一步扩容并持续下沉，将带动医疗领域数据中心、云数据中心等算力基础设施建设，实现公共型算力的低成本普及。

<sup>29</sup> 高校网络服务情况和教育信息化需求调研报告（2022年度），中国教育和科研计算机网CERNET，2023年3月

<sup>30</sup> 第51次中国互联网络发展状况统计报告，中国互联网信息中心，2023年3月

<sup>31</sup> 《“健康中国2030”规划纲要》，国务院，2016年10月



# 05



# 应变： 需求洞察篇

## ICT行业

### 算力需求洞察

ICT创新往往“牵一发而动全身”，需要“通信+计算+存储”三位一体协同升级，朝着云、物、大、智、虚等方面不断进化，相关算力需求呈现多元化、低时延、灵活迁移等显著特征：



#### 计算场景丰富多样，多元算力需求显著提升

各类智能化应用渗透率不断提高，智能化愈发多样化、个性化，相关算力需求大规模增长的同时，呈现出多元化趋势，在通用算力之外，智能计算、边缘计算、安全计算等需求高企。



### 5G网络尚在持续部署，计算时延要求日趋严格

以目前的网络建设水平，中长距离数据传输仍面临较大时延，算力调度等严重受限，如“东数西算”工程中网络时效性仍较弱；此外，复杂多样的计算任务需要运行稳定且高效的网络作为支撑，对算力提出了更高要求。



### 智能终端走向泛在分布，要求算力可灵活迁移

手机、电脑、智能家居设备、可穿戴医疗健康设备等多样化智能终端不断普及，传统以云为主的集中存储和计算模式愈发难以满足“跨终端”“零等待”的用户体验需求，现有场景要求基于终端设备实时、多端、无缝访问用户数据，实现高效一致的跨终端计算。

## 算力应用场景

### 案例一

#### 某国际电子产品解决方案公司——基于云计算的数据湖解决方案

“十四五”数字经济发展规划要求，数字技术自主创新能力显著提升，数字化产品和服务供给质量大幅提高，产业核心竞争力明显增强，在部分领域形成全球领先优势。预计在这一政策目标下，实体经济发展对ICT集成服务供应商的需求将不断增加，拥有广泛行业知识和定制服务能力的企业竞争优势尤为突出，算力也将成为相关企业提升服务能力不可缺少的基础性资源。

案例企业是一家提供综合设计、制造、供应链和产品管理服务的制造解决方案供应商，业务涉及医疗、包装、航空航天等领域。员工数量多达几十万名，分布在全球的数十个国家的上百个厂区，作为劳动力密集型企业，实现劳动力优化是其业务降本增效的关键课题。

#### 业务需求

案例企业在利用庞杂劳动力数据，实现端到端的人力资源管理优化过程中，逐渐意识到优化现有计算能力的必要性：

**现有算力不足以处理数亿级的业务数据量。**案例企业在全世界拥有26万名员工，如此量级的员工每天都会产生数亿级的业务数据，包括考勤打卡数据、生产效能数据等，这些业务数据不仅数量庞大，数据质量层面也会出现大量错综复杂的问题，经评估，案例企业现有算力服务无法有效应对。

**计算能力无法支撑端到端的劳动力优化解决方案。**从劳动力招聘入职、岗位分配、在岗培训，到员工绩效管理、奖惩管理、出勤管理、生产效能管理，再到最终的员工离职管理，劳动力优化需在人力资源管理的每一个环节中进行，而案例企业相对缺乏端到端劳动力优化的云计算能力。

## 解决方案

案例企业借助外部服务商搭建了人力资源管理整体数据模型，通过云计算建立数据集成与处理机制，实现数据互联互通，持续推进劳动力优化解决方案。

**实现端到端场景的人力资源数据建模，为云计算方案实施奠基。**通过梳理端到端的劳动力优化场景需求，对分布在各大业务系统和主数据系统的数据要素进行探源，完成了人力资源全生命周期数据的有效建模，为后续数据分析奠定了扎实基础。

**利用云计算技术建立数据集成与处理机制，实现数据互联互通。**打通各大数据来源系统与数据湖平台的数据传输机制，设计基于业务需求导向的数据模型，通过云计算构建自动化、智能化的数据集成与处理的整体解决方案，进而实现数据互联互通。

## 案例企业收益

案例企业借助该数据湖解决方案，提高了底层计算能力，使得数据可用性、响应速度等大幅提升，有效优化了用户体验和全流程人力资源管理业务。

**数据可用性提升，全流程人力资源管理业务得以优化。**在端到端数据互联互通的基础上，原本庞大且复杂的数据得到了有效利用，可快速响应人力资源管理的全流程业务需求，人均效率提升约20%。

**数据响应速度提升，用户体验得以优化。**数据应用平台的响应速度提升了8倍左右，响应时间由分钟级缩减至数秒级，明显提高了平台的用户好感度和粘性。

# 制造业

当前，中国制造业数字化转型已取得不错成效，智能制造试点示范项目生产效率平均提高45%、产品研发周期平均缩短35%、产品不良品率平均降低35%，涌现出了离散型智能制造、流程型智能制造、网络协同制造、大规模个性化定制、远程运维服务等新模式新业态<sup>32</sup>。

不过，与高质量发展的要求相比，智能制造发展仍存在供给适配性不高、创新能力不强、应用深度广度不够等问题，相关算力需求呈现高效化、低成本、绿色化等显著特征：

## 算力需求洞察



### 为赋能业务流程及打造增量价值，算力高效率需求上升

目前许多装备制造业企业相对缺乏完善灵活的生产控制计划，且生产计划与采购计划未能有效结合，“研-产-供”等业务流程急需优化，全价值链运营及决策效率低下，数字化运营及智能技术的应用对算力等基础设施提出了更高要求。



### 产线周期急需优化，低成本的推进需以算力为支撑

工厂各种活动数字化成本较高、周期较长，企业在产品、工艺、产线等方面研发与验证工作急需逐渐降低成本，需要生产系统建模与仿真技术的支持。降低试错成本、缩短产线周期需要海量算力的普遍应用。



### 绿色智能的数据与计算设施正加快布局，推动“绿色计算”是当务之急

加快节能低碳技术研发，加快先进绿色技术产品应用，鼓励应用高密度集成的高效IT设备，提升能源利用效率，需要持续提升能源高效清洁利用水平，引导新型数据中心向新能源发电侧建设，全面提升用能效率。发展“绿色计算”，即散热和节能用能技术是当务之急。

<sup>32</sup> 《“十四五”智能制造发展规划》，工信部等八部门，2021年12月21日，[http://www.gov.cn/zhengce/zhengceku/2021-12/28/content\\_5664996.htm](http://www.gov.cn/zhengce/zhengceku/2021-12/28/content_5664996.htm)

## 算力应用场景

### 案例二

#### 冠捷科技——边缘算力落地AI视觉质检场景

在万物互联、万物智联的时代背景下，显示器作为人机交互的核心界面承载着构建人机融合新关系的重要使命，当前显示器行业正由“量产为王”阶段，加速转向“技术为王”“质量为王”阶段。以往单纯依靠人工人眼的质检易造成质检结果不稳定，因此，利用智能技术解决质检问题是行业迫在眉睫的需求。基于机器视觉算法的智能化质检可以在显示器生产过程中承担“质量守门员”和“效率加速器”的角色。

冠捷科技是一家液晶显示器及电视智能制造企业，在液晶显示器领域，其产量长期保持全球领先，年产量接近6,000万台，全球市占率超30%；在液晶电视领域，其产销量基本稳居全球前五，全球市占率约为6%。作为行业内的代表性企业，其计划在2024年、2026年实现数字化工厂建设和智能化工厂转型。

#### 业务需求

2020年以来，居家办公带来的显示器需求增长和疫情造成的用工紧张之间的矛盾爆发，促使冠捷科技开始寻求智能化的屏幕质检方案。在这一转型进程中主要出现了以下两类算力需求：

**AI质检模型训练需要大量生产现场数据，要求补充边缘算力。**冠捷科技推进智能化质检取代传统人力质检，目的在于降低人力成本、提高质检效率和准确率。但这一过程中，需要大量生产现场的数据喂养AI质检模型，如何有效采集、整合、处理分散在不同设备、车间、产线的数据成为搭建AI质检模型的首要难题。这使得在生产现场大规模部署边缘计算类硬件成为必要之举。

**显示器缺陷种类多、成因复杂，要求算力支撑AI质检模型加速迭代。**AI质检模型训练过程中不仅需要结合现场侧的多类结构化和非结构化数据，还需要融合大量从业人员的知识经验，做到持续迭代，尤其当前显示器工艺升级速度加快，正倒逼模型加速优化，如何实现底层算力和上层模型迭代需求的精准匹配堪称是行业级难题。

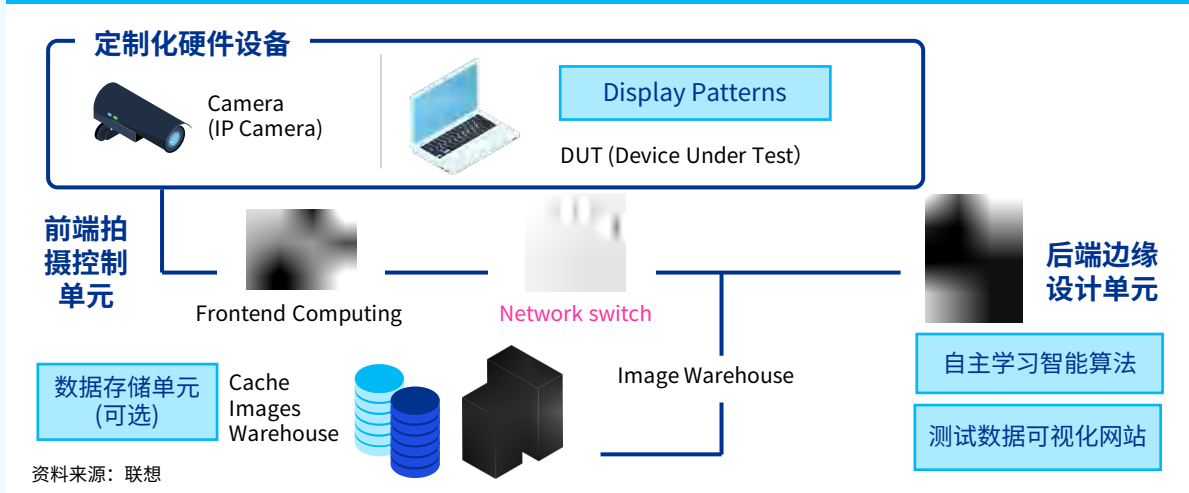
## 解决方案

为保障AI质检在实际生产过程中发挥实效，冠捷科技与联想共同探索总结了一套以“边缘算力+AI视觉质检”为核心的智能解决方案：

**智能边缘类设备保障算力供给基础。**为补充边缘算力，联想集团联合生态合作伙伴专门在冠捷科技的工业生产现场搭建了大型工作站，以工作站+工控机的模式保障规模化算力的稳定供给，底层计算架构主要采用“CPU+GPU”模式。

**软硬件一体化解决方案助力AI模型持续迭代。**为保障AI质检模型实时有效，联想在提供边缘硬件设备的基础上，进一步为冠捷科技提供了机器视觉算法和Edge AI小样本算法，并将自适应技术加入算法中，对不良风险进行分析、预测预警、回溯，通过软硬件一体化的解决方案，底层算力得以更加动态精准地满足场景化需求。

图 18 | AI视觉质检解决方案整体架构



## 案例企业收益

AI质检的有效落地不仅帮助冠捷科技降本增效，也为制造行业的智能化转型提供了良好范例。

- **节约人力成本：**在AI视觉质检方案的助力下，冠捷科技显示屏质检环节人力成本降低约50%；
- **提高检测效率和准确率：**漏失率控制在200ppm（百万分之一）之内，过杀率达到6‰（行业平均水平约在10‰），整体检测效率和准确率提升约30%；
- **提高生产现场效率：**得益于边缘计算硬件设备的投入，实时类计算任务时延一般控制在毫秒级；
- **智能制造解决能力沉淀：**基于“边缘算力+AI视觉质检”的有益探索和成功经验，冠捷科技与联想共同探索总结了一套智能制造解决方案，即在充分保障边缘算力供给的基础上，算力和算法保持动态精准适配，实现AI质检模型的持续迭代。





### 案例三

#### 某世界500强家电企业——高效算力助力智能工厂

中国家电制造是“中国智造”走向世界的一张耀眼“名片”，在制造业数字化转型中，相关企业率先探索，积极创新，再次扛起了大旗。近年来，海尔、美的等大型家电制造企业在不断驱动自身业务转型的过程中，卡位工业互联网，以自研工业互联网平台赋能其他制造业企业数字化转型，打造了大量具有推广价值的典型案例。

案例企业从家电领域起步，已逐步转型为覆盖智能家居、楼宇科技等诸多创新业务的全球化科技集团，在全球拥有数百家子公司、数十个研发中心和主要生产基地，近年来，全面数字化和全面智能化是该企业的核心发展战略。

#### 业务需求

案例企业建设智能工厂的实践中，一度面临算力总量不足和缺乏边缘算力建设模式的双重挑战。

**工厂现场设备互联产生大量数据资源，催生算力需求。**案例企业在搭建工业互联网平台时，需要借助工艺卡等获取各类生产设备的数据，如注塑机的各类工艺参数、运行数据等，进一步地，打通全流程数据要求实现各类设备的互联互通，必然会产生海量非结构化数据，导致算力需求激增。

**质检等场景亟需以算力取代人力，边缘算力需求尤为迫切。**以质检场景为例，家电产品质检后的不良品会继续在产线中流转，需要工人在特定位置搬出，不仅存在劳动强度大、效率低的问题，而且人力为主的工作模式，导致产能信息不透明，给产线管理者监督和评价工人工作带来了阻碍。一方面需要以机械臂等边缘智能设备进一步取代人力劳动，另一方面需要强化边缘数据采集，使得面向工业现场的边缘计算需求尤为突出。

## 解决方案

案例企业通过搭建以边缘计算为基础的工业互联网平台，实现了设备数据的实时通讯和上传，以更充沛且高效的算力赋能智能工厂。

**建设以边缘计算为基础的工业互联网平台。**通过设备监控仪，向下连接设备，向上连接SCADA（监控和数据采集系统，Supervisory Control And Data Acquisition）和工业互联网平台，提供标准统一的结构化数据，进行数据计算和本地分析，提供人机互动和SCADA的各种核心应用，基于IoT数据实现AI和边缘数据应用。

**强化数据实时分析和定时存储，提高算力利用率。**建成了相关服务器，采集产线运行时间、物料到达时间和设备运行各项参数，协同扫描枪的扫码打标，实现生产节拍计算和报警规则响应；对数据进行分时过滤，将更为重要的数据上传到平台，减少对网络带宽的压力。并对来自设备层的实时状态数据、报警、故障等信息做定时存储，完成设备层与信息系统之间的数据交互。

## 案例企业收益

**在以边缘算力为主的新增算力支撑下，工业现场智能化水平明显提升。**边缘算力的补充，提高了设备信息化水平，设备联机率提升至82%，工艺质量、生产过程、管理设备等数据可视化程度大幅提升。

**高效算力与实时数据分析相互驱动，实现业务降本增效。**降本方面，设备联机率大幅提升，为设备故障诊断、预测性维护等智能化应用奠定了良好基础，每年可节约800万元以上的设备运维费用。增效方面，生产节拍从15s缩短至13s，整体效率提升约13%，质检场景中不良品处理响应提升30%以上。

## 案例四

### 中国钢研——云计算、高性能计算、智能计算协同打造数字化科研平台

中国钢铁工业经历了从小到大、从弱渐强的发展历程，正迎来“由大到强”的机遇期，技术创新和绿色发展是两大重要方向，尤其是在高端材料、绿色低碳等工艺技术基础研究和应用研究方面，还需不断提高自主创新能力，必然离不开数字技术在钢铁工业的集成创新和融合应用。

中国钢研科技集团有限公司（简称“中国钢研”）是我国金属新材料研发基地、冶金行业重大关键与共性技术的创新基地、国家冶金分析测试技术的权威机构，是国家首批103家创新型企业试点单位之一，是中关村科技园首批100家创新型企业之一，有19个国家级中心和实验室、25个省级中心以及8个产业技术创新联盟依托中国钢研设立。目前，中国钢研拥有各类科技成果5,000余项。中国钢研立足于主责主业，发挥优势、强化协同，重点聚焦于“共性关键和前沿技术研发”“重点材料及制品”等领域。

#### 业务需求

中国钢研在进行新材料研发模式创新时，“原理+数据”的新方法论相较于“文献+试错”的传统方法，需要更强大的算力及数据资源管理能力，在此期间，中国钢研主要面临以下业务需求：

**现有算力无法支持大规模且高度复杂的调度、开发和计算活动。**当时中国钢研的整体信息化程度较低，全图形模式的调度技术、算法开发和大型计算任务的并行处理需要依靠算力来完成，原有算力难以实现上述需求，仍需依托各个大学完成。

**缺乏可供远程使用的算力资源。**供研发部门使用的图形工作站机器设备成本高、设备体积大且使用场景受限（难以远程使用），机器闲置造成算力无法被充分利用。

#### 解决方案

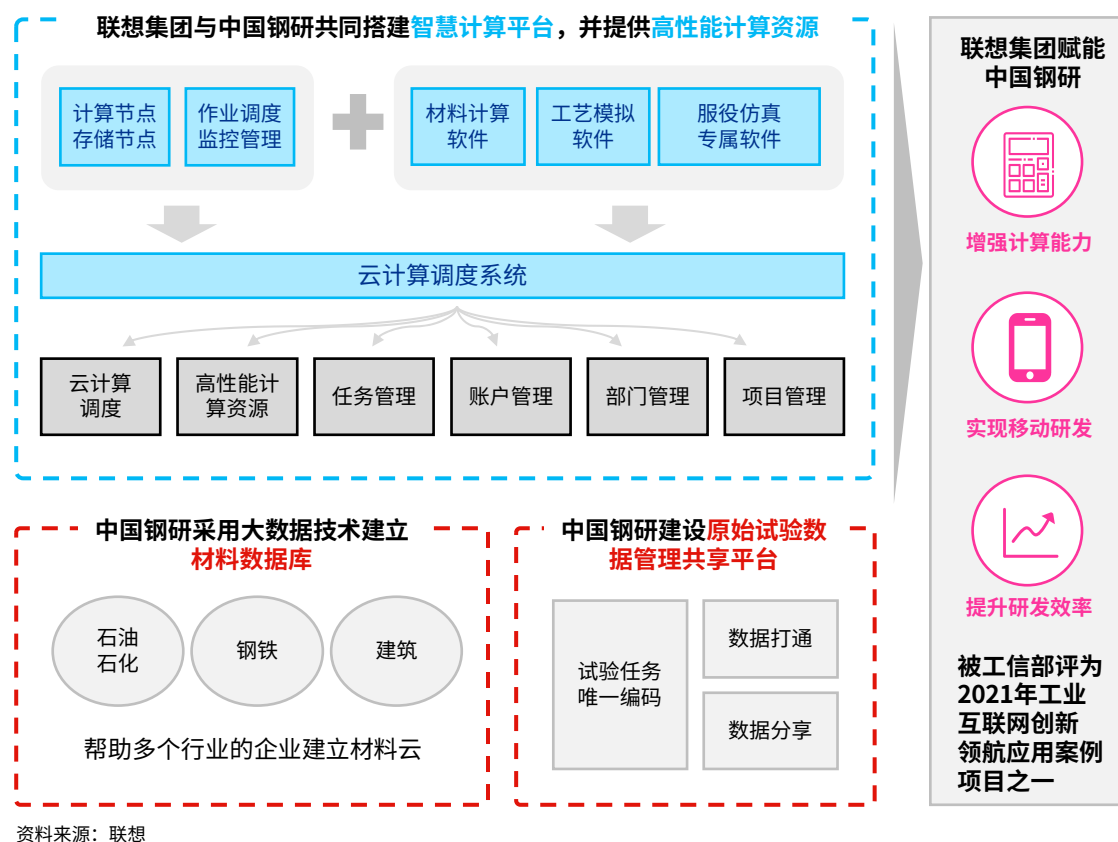
联想集团借助自身成熟的仿真设计一体化解决方案，助力中国钢研在三个月内完成了智慧计算平台的主要建设任务，并在众多行业得到成功应用，主要包含以下内容：

**提供充足的高性能计算资源。**联想集团为中国钢研提供了充足的高性能计算资源。集成管理了20余种国际知名计算和仿真软件，包括第一性原理计算软件4种、相图相场计算软件4种、多物理计算软件1种等，覆盖纳观、微观、介观、宏观的全尺度计算资源，具备材料设计、组织性能预测、工艺模拟、服役仿真全流程计算能力。

**建立了钢研云计算系统。**多数软件通过云计算平台均可实现基于浏览器的远程计算、移动计算及排队计算，算力融合的模式解决了中国钢研远程无法使用算力、算力闲置的问题。科研人员可随时随地登录个人账户，建立虚拟项目组，远程启动建模设计软件，并在电脑、Pad、手机等多种固定或移动终端上提交计算任务、获取计算数据。

**大数据材料数据库：**中国钢研采用大数据技术建立材料数据库，材料数据库集材料和工艺专业知识、计算工程师经验于一体，针对具体材料和应用场景的微应用方案，可由具体计算案例改造实现，并将这一能力开放给多家石油石化、钢铁、建筑等传统行业企业，帮助建立材料云。

图 19 | 联想为钢研提供的智能解决方案



## 案例企业收益

联想与中国钢研合力建设平台的应用不仅将助力钢铁行业数字研发能力到达国内领先水平，也为我国高端装备、重大工程领域的进步提供着技术支撑。

**强大计算能力带来了传输、运算和处理能力的提升。**工业互联网材料云计算平台已于2021年7月开放试运行、11月正式发布启用。在3个月试运行期内，数万人次访问、完成了4,000余次各类软件的计算任务，并可方便地实现部门、团队及个人的计算任务统计。可以支撑255万亿次峰值运算速度、2PB数据存储容量，支持同时运行10个以上大型计算任务、200个以上并发设计和数据任务、1万台以上科研设备数据接入，缓解了中国钢研算力不足的问题。

**给布局战略新兴领域的行业带来深远影响力。**中国钢研的仿真设计一体化方案是高端装备制造领域的典型创新应用，未来有望服务众多战略新兴领域，走向全球科技前沿。

## 案例五

### 某中国500强新能源企业——基于绿色、高性能计算的数据资产化解决方案

新能源行业数字化转型深入推进，数据和算力发挥的作用越来越重要，在合理的ESG战略指导下，利用绿色、高性能算力管理数据资产，正成为相关企业重点布局方向。

案例企业是中国某500强新能源企业，作为全球可信赖的清洁能源战略合作伙伴，致力于推动能源变革，让人人可负担、可靠、可持续的能源惠及全球，构建“可持续更美好”的未来。案例企业深度聚焦能源开发、能源装备、能源服务、能源应用四大领域，以强大科研创新和领先业务实践，将可再生能源的效率提升至新高度。

#### 业务需求

案例企业在推动数据资产化过程中，面临着计算能耗过高、数据质量较差导致算力难以有效赋能等问题：

**计算能耗过高与可持续发展战略相矛盾。**随着业务系统处理的数据和请求越来越多，数据量快速攀升，节能降耗成为企业必须面对的问题。系统需要处理大量的实时数据、连接着庞大的数据存储系统、同时还要进行大量的图形处理、音频处理、视频编解码等，计算能耗过高，不符合可持续发展战略的方向要求。

**低质量历史数据占用存储空间，造成算力浪费。**大量的历史留存冷数据无法被有效识别及处理，形成数据“包袱”。而现有的算力能力很难及时将这些数据“包袱”变成数据“金矿”，会占用大量的存储空间，造成一定的成本浪费。

#### 解决方案

**提高绿色计算能力，提升能源利用效率。**针对不同等级的数据，制定、优化算力调度和分配方案，集中优质算力解决重要业务问题，减少不同优先级业务算力的同等投入；利用集团大数据平台的能力赋能业务单元、子公司的数据分析应用，节约子公司的资源投入，在保证服务稳定的前提下，合理分配算力资源，提高资源利用率，降低服务器的用量，从而减少碳排放。

**提升高性能计算能力，保证数据在大数据平台各层间全链路存储、加工、流转时的高效利用。**改变之前信息系统烟囱式建设、数据分散存储的状态，以逻辑模型的方式将原本分散、重复、不一致、不规范的数据设计结构进行梳理、优化和重构，并在大数据平台上实现数据的规范整合、存储，搭建基于高性能计算的大数据应用平台，对历史数据进行清洗、挖掘、建模、深度挖掘，使算力有效赋能前端业务，减少大数据平台的冗余存储和重复计算。

## 案例企业收益

案例企业在数据资产盘点和大数据平台建立的过程中，不仅实现了业务的降本增效，也实现了计算基础和计算能力的提升。

**推进可持续发展战略和能耗成本的降低。**通过构建绿色计算能力，帮助案例企业实现近百张表单的下线或降低更新频率，从而提高了算力使用的合理性和有效性。同时通过实施集团算力有效分配策略，发挥集团算力优势赋能业务单元、子公司，从而减少全集团能效成本的浪费，更好地推进集团ESG战略。

**大幅推进案例企业计算能力的跃迁，提升了数据可用性。**一体化数据资产体系为大数据平台建设赋能，贯穿数据入库、整合存储、共性指标统一加工、分析应用等全流程，实现单点加工、多处使用，大幅推进计算能力的提升，数据的优化及重构也提升了数据的可用性。

## 案例六

### 基于云计算的IaaS规划及管理驾驶舱应用项目

案例企业是一家以节能减排、环境保护为主业的中央企业，发展至今已成为我国节能环保领域规模大、专业全、业务覆盖面广、综合实力强的旗舰企业。

#### 业务需求

案例企业在信息化系统运营过程中，曾存在算力基础薄弱且发挥价值不足、没有建立“两地三中心”的容灾模式<sup>33</sup>等诸多痛点。同时，在集团财务部门的数据使用过程中，也出现了缺少大数据平台支撑、数据分析工具落后、没有可视化工具等诸多痛点。

**缺乏相应的数据应用以发挥算力的价值。**管理驾驶舱作为将经营数据可视化，将管理策略具象化的数字化解决方案，可以在数据分析、展示、洞察、治理方面提供清晰的表述和展示，在支持管理决策方面起着关键作用。由于算力缺乏，相关数字底座建设不完全，无法充分发挥管理驾驶舱的功能，通过分析、风险预警活动来辅助科学决策。

**云计算能力欠缺，无法满足上云需求。**案例企业的基础设施建设依托外部数据中心建设，目前还处于单一机房运行阶段，存在云计算能力薄弱，算力无法灵活调度，且数据存储缺少容灾备份的设备和工具等问题。急需加强对算力基础设施建设的投入，大规模扩展其现有的基础架构。

<sup>33</sup> 两地三中心是指同城双中心加异地灾备，是一种商用容灾数据备份解决方案；两地是指同城、异地；三中心是指生产中心、同城容灾中心、异地容灾中心。

## 解决方案

针对目前所面临的痛点，案例企业采用了构建“两地三中心”云平台的数字化解决方案，在提升算力的基础上实现了算力的灵活调度，并构建了异地容灾机制。

**基于云计算能力构建云平台。**基于异地数据机房，构建案例企业自用的基础设施体系。在此基础上搭建的云平台可以实现计算、存储、网络、安全资源的灵活调度，从而提高设备的利用率，降低机房的自用成本。

**通过提升算力实现实时数据分析及动态数据更新。**通过算力的提升，管理驾驶舱的功能得以最大化提升，企业的内外部经营相关数据均实现了可视化展示。在财务管理层面，财务数据的可视化和实时更新提升财务部人员的工作效率，降低了数据统计的难度和复杂度。

## 案例企业收益

通过基于云计算的IaaS规划及管理驾驶舱项目，客户从整体上提高了信息化建设水平和集团财务数据的可视化水平。降低设备重复采购率、人员的工作效率，提高了数据采集、分析和应用的能力，同时也实现了数据赋能管理层有效决策辅助的价值。

**节约了采购成本，实现降本增效。**服务器和存储设备采购量下降14.8%，大幅度节约了采购成本；信息化人员的设备管理范围提升23%，集团财务数据汇总提交时间降低了50%（由原来的8小时，降低为4小时）；负责数据统计分析的人员减少了13%，有效提升了人员效率。

**提高了企业的数据管理水平，及科学化管理决策的能力。**企业从原有的人工汇总和处理数据的工作方式转变为智能化数据汇总的工作方式，大大提高了客户管理决策的效率和准确性。





# 汽车行业

## 算力需求洞察

汽车厂商之间的竞争正在从“马力”主导的速度、操控竞争，转向“算力”主导的数据、软件竞争，汽车整体产业链合作也从“链式”逐步转化为“网状”，从车载计算平台到车企数据中台再到产业链级算力中心，算力需求全面爆发，呈现出大型化、弹性化、集约化等显著特征：



### 汽车全产业链算力需求扩张，要求大规模、弹性化算力供给

汽车智能化升级过程伴随着汽车全产业链条的数字化发展，研发端、生产制造端、运营端、营销端、车端、出行端等各环节的数据将被打通，算力需求将扩大至汽车产业链全环节，要求提供更大规模且调度灵活的算力资源，倒逼算力供给侧积极推动算力弹性部署。



### 车企自建算力平台渐成趋势，要求实现低成本、集约化建设

面向车载计算平台以及全产业数字化转型带来的巨量算力需求，越来越多的车企开始选择自建数据中心等算力平台，一方面有利于积极应对汽车行业越来越严格的数据管理要求，另一方面有利于算力供不应求的市场环境下保障自给自足，更有望与生态合作伙伴共享算力。不过，车企自建算力平台也意味着漫长的建设周期和巨额的投入，亟需实现低成本、集约化建设。



## 算力应用场景

### 案例七

#### 吉利汽车——基于汽车仿真计算云平台的集团级算力共享化实践

汽车消费者的个性化需求日新月异，产品创新速度正成为车企制胜关键，公开数据显示，新车研发周期已经从5-7年缩短至2-3年<sup>34</sup>，越来越多的车企开始借助数字孪生、仿真模拟等技术辅助产品设计，提高研发效率，使得高性能、高弹性的算力需求成倍增加。

2021年吉利汽车发布“智能吉利2025”战略，并宣布未来5年内将持续投入1,500亿元用于研发领域，加速科技成果转化落地，旗下各品牌未来5年将推出超过25款全新智能新能源产品，为此，吉利汽车正积极自建高性能计算中心、智算中心等算力底座，面向全集团乃至全产业链提供充足的算力资源。

#### 业务需求

近年来，吉利汽车不断推动产品研发环节“从实到虚，以虚促实”的数字化变革，相关算力需求增长迅猛，但在建设算力中心、推动算力共享过程中也出现了不少痛点。

**自建算力中心成本高，要求实现集约化建设：**单个算力中心建设成本较高，出于经济性考虑，吉利汽车研究院作为面向全集团提供整车研发支持的重要部门，承担起了组建集团级算力中心的任务，要求实现大规模、高性能算力的集约化建设。

**要求面向全集团子品牌实现算力共享，但缺乏合理模式：**完成算力集约化建设的下一步，便是将算力资源共享给各子品牌，尽管云化、虚拟化等技术能实现算力资源池化，保证算力能以虚拟产品的形式流通，但还需设计合理的算力定价和运营模式，以实现集团和子品牌之间算力的合理结算。

<sup>34</sup> 汽车研发周期缩短、OTA普及，车企如何应对研发数字化转型挑战？，极客公园，<https://www.geekpark.net/news/293469>

## 解决方案

联想以“裸金属管理+可视化运维+个性化计量计费”的整体解决方案为吉利汽车搭建汽车仿真计算云平台，推动集团范围内的算力规模化建设和共享化运营。

**裸金属管理保障算力灵活、弹性供给：**汽车仿真计算云平台采用裸金属架构，虚拟化软件直接安装在服务器上，综合了传统服务器租用稳定性与云上资源高度弹性两大优势，实现了算力的即买即用，快速交付。

**数字孪生系统推动算力协同运维：**基于数字孪生技术搭建算力中心管理平台，可以帮助业务专家实时掌握机柜容量、用电量等数据，对数以千计的算力节点的实现从上架到变更的全生命周期管理。此外，由于数字孪生技术理解门槛较高，联想出动专家资源帮助吉利汽车业务人员掌握性能调优技能，实现了协同运维。

**个性化计量计费模式助力算力高效结转：**基于汽车仿真计算云平台的灵活算力供给，不同子品牌获取算力资源时，可结合自身品牌定位、算力使用方式等获取不同的折扣率，既保证了各子品牌能以低于外采算力的价格获取算力资源，又充分保证了集团内算力的合理高效结转。

## 案例企业收益

- **提升研发效率：**汽车仿真计算云平台提供的算力资源主要面向汽车产品研发数字化需求，可全面提升从集团到各子品牌的研发效率；
- **获取澎湃算力：**汽车仿真计算云平台项目分三期推进，预计建设完成后，总体算力规模可进入全球算力中心Top200行列。
- **对外输出算力：**预计随着个性化计量计费模式在集团内部跑通后，吉利企业有望利用充沛的算力资源，面向全产业链生态合作伙伴赋能。

## 案例八

### 某国内领先汽车公司——基于云计算的分析云平台项目

汽车行业数字化和智能化趋势的发展带来了海量数据，但数据规模和复杂性也对企业的运营分析提出了新的挑战，算力逐渐成为不可忽视的驱动力量。算力资源池化平台可以满足算力资源的灵活分配及调度，不仅能大幅提升资源利用率、降低总拥有成本（TCO），更能赋能汽车行业的智能化升级。

案例企业为一家国内规模领先的汽车上市公司，努力把握产业发展趋势，加快创新转型，正在从传统的燃油车生产和销售企业，向新能源汽车、互联网汽车的商业化推进，并积极开展智能驾驶等技术的研究和产业化探索。

#### 业务需求

案例企业在企业运营管理的过程中，虽然积累了大量的业务数据，但是由于数据质量和业务系统的复杂性，运营部门依然面临着“有数不能用”的尴尬境地。

**算力不足导致运营数据的分析结果不全面。**大量业务数据在企业运营管理中会持续生成，但算力不足导致业务运营数据分析结果往往面临不全面、不清晰等问题，容易造成管理决策的误判。

**有效的工具及技术的缺乏导致云计算能力不足，数据缺乏互联互通。**企业的各类数据存储在不同的业务系统中，需要将这些不同系统的数据整合起来才能进行分析。在以往的管理报告和运营报告出具过程中，缺乏有效的工具和技术手段，往往需要耗费大量的时间和人力资源，且在整合过程中容易出现数据标准不同且格式不统一等问题，需要额外的人力进行清理和整理。

## 解决方案

针对业务需求，案例企业采用以数据平台为核心的数据应用架构，同时配合数据治理，对企业的业务数据进行综合管理。

**加强算力供给来处理大量数据。**通过对应用部门的需求梳理，在加强算力布局的同时展开数据治理，包括：数据资产盘点、数据标准、数据组织规划及权责分配等，形成了企业级数据资产，实现了“看的清、管的住、用的好”的数据应用目标。

**通过建设私有云+内存计算平台，在确保数据安全的前提下实现数据互联互通。**私有云+内存计算平台使用户能既满足企业高度的安全权限需求，又能够达到更高的计算能力，同时结合内存计算，提供给用户顺畅的数据应用体验。

## 案例企业收益

基于云计算+内存计算的分析平台为企业提供了数据应用的新机遇，帮助企业快速获取数据并加以分析，从而提高数据分析的效率，优化业务运营，提高商业竞争力。

**助力企业数据分析与决策。**基于云计算的分析平台可以自动处理大量业务数据，为其他的数据分析和决策提供更快速的响应。

**赋能业务，提升运营效率。**借助数据分析平台，快速整合前端市场信息和业务需求，并将其转化为优化产品设计的方案，并同步生成生产计划协同生产，大大缩短了交付周期，提升整体运营的效率。

# 教育行业

## 算力需求洞察

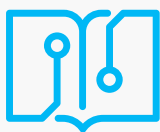
2021年7月，“双减”政策落地，强调着眼建设高质量教育体系，强化学校教育主阵地作用，几乎同时，国家教育部等六部门发布《关于推进教育新型基础设施建设构建高质量教育支撑体系的指导意见》，提出加快推进教育新基建，构建高质量教育支撑体系。面向教育高质量发展需求，服务于教育新基建的公共型算力成为刚需，而在学校这一教育主阵地中，服务高校科研的高性能算力需求潜力巨大。

### 教育新基建持续倒逼公共型算力供给



中国计划到2025年，基本形成结构优化、集约高效、安全可靠的教育新型基础设施体系<sup>35</sup>。其中，构建新型数据中心是重要手段，强调支持省级教育行政部门通过混合云模式建设教育云，为本地区教育机构提供便捷可靠的计算存储和灾备服务，并鼓励区域和高校共享高新性能计算资源和人工智能算力资源。预计在政策强力推动下，面向各级院校和各类教育机构的公共型算力需求会持续高涨。

### 高校科研计算强力带动高性能算力需求



高校是国家战略科技力量的重要组成部分，当前国内许多高校正提速建设高性能计算中心，此类计算中心的规模通常在“百节点、万核心”，日常使用率达到80%以上，结合2022年11月公布的第60届全球超级计算机TOP500榜单来看，上海交通大学、中南大学、清华大学、南京大学这四所中国高校的高性能中心成功上榜，建设规模已进入千万亿次级别（P级别）。预计在“双一流”建设背景下，越来越多高校将加大高性能算力投入，以加快取得突破性研究成果，保持优势学科的学术水平领先。

<sup>35</sup> 教育部等六部门关于推进教育新型基础设施建设构建高质量教育支撑体系的指导意见，  
[http://www.moe.gov.cn/srcsite/A16/s3342/202107/t20210720\\_545783.html](http://www.moe.gov.cn/srcsite/A16/s3342/202107/t20210720_545783.html)

## 算力应用场景

### 案例九

#### 上海交通大学——校级公共计算平台“交我算”助力科研攻关

“交我算”是上海交通大学的校级公共计算平台，主要包括云平台、人工智能计算平台、高性能计算平台、ARM超算平台和杨元庆科学计算中心五大计算平台和科学大数据平台。

其中，助力上海交通大学登顶中国高校算力排名之首的关键设备——“杨元庆科学计算中心”中的大型科学计算机“思源一号”，拥有共计60,566个CPU核，双精度峰值超过了6PFLOPS。“思源一号”于2021年底上线，现已切实助力上海交通大学的师生开展基础科研，回顾其初期建设痛点、中间建设过程和最终落地价值，有望为中国高校建设公共高性能计算平台提供有益参考。

#### 建设痛点

**数据驱动科研成为新范式，考验底层数据管理能力：**随着学术界越来越多地使用量化方法解决科研问题，数据驱动科研已逐渐成为趋势，而科研数据具有来源复杂、数据类型广泛、存储形式多样等特点，并且数据共享正在成为科研成果传播的重要形式，对高校计算平台的数据存储、动态监测、安全防范等能力提出了极大考验。

**高性能计算平台工程复杂、建设周期长，需充分考虑各类环境限制：**为保证高性能计算平台的各类设备能安全平稳运行，在工程规划初期就需要考虑建筑与结构、电气、空气调节与给排水、网络与布线系统、智能化系统等环境因素。以上海交通大学为例，在建设“思源一号”之前，已建有云平台、人工智能平台等计算平台，如何协调好新增算力和原有算力的布局极其考验建设者的智慧。

**高性能计算平台散热难、能耗高，需充分结合高校节能降耗目标：**高校承担着复杂的科研和教育功能，水电消耗量本身就比较高，结合中国地质大学测算数据来看，2017年中国教育领域的碳排放高达2.46亿吨，占全国总量的2.5%<sup>36</sup>。根据上海市发改委《关于印发上海市2021年节能减排和应对气候变化重点工作安排的通知》，上海交通大学节能减排工作考核目标为：学校单位建筑面积能耗增幅控制在1%以内。这就意味着，高性能计算平台除了要充分满足科研计算所需的容量和性能要求外，还要兼顾节能减排。

<sup>36</sup> 全国政协委员王焰新：建议高校推动碳中和校园建设，中国新闻网，2022年3月4日，<https://www.chinanews.com.cn/gn/2022/03-04/9692041.shtml>

## 解决方案

“思源一号”具备高计算密度、绿色节能、能源可循环利用三大技术特色。

**搭建高计算密度系统，极大节省平台基建成本：**“思源一号”由联想集团负责建造和交付，共使用14个机柜，占地面积约300平方米，相较通用机房节约了80%的机柜数量和90%的建造面积，极大节省了基建成本。

**采用温水水冷技术，节约能耗成本40%以上：**“思源一号”采用了联想“海神”温水水冷技术，系统PUE约在1.1，相较传统风冷技术，温水水冷使用50-60°的温水充分吸收大型计算机运行时产生的热量，并利用水的自然降温进行散热，减少了系统风扇功耗，可节约42%左右的电力成本。

**实现余热回收再利用，每年可实现950吨的额外碳补偿：**在利用温水散热的基础上，“思源一号”计算系统还实现了余热回收，在冬季供暖、食堂热水加热、地下室除湿等场景中，使用余热可以充分取代电力、燃气等能源，每年可实现950吨的额外碳补偿。

## 案例企业收益

- **助力科研攻关：**“思源一号”主要在李政道研究所实验楼投入，已有效助力上海交通大学实现科研成果突破，涉及生物学、计算机科学、天文学等众多前沿领域；
- **计算能力达到国际领先水平：**“思源一号”的计算能力不仅领先一众国内高校，也超过了哈佛、剑桥等国际名校，在世界舞台上展示了中国高校的科研支撑能力。
- **温水水冷技术具有普遍推广价值：**“思源一号”所采用的温水水冷技术节能降耗成效显著，已经成功推广至马来西亚气象局、哈佛大学文理学院计算中心、北京大学高性能计算系统“未名1号”等多领域。



# 医疗行业

## 算力需求洞察

在人口老龄化加剧与医疗资源有限等问题日益凸显的背景下，中国医疗信息化正在加速推进，尤其是随着中国公共卫生工作重心逐步由防控感染转为医疗救治，医院信息化工作将脱离快速应急响应状态，回归促进医院高质量发展的主航道，围绕健康医疗大数据规范应用和“互联网+医疗健康”创新发展等任务，对算力服务提出了安全可靠、灵活调度等要求。

### 医疗大数据质量及安全管理需求日益突出，安全可靠成为算力硬要求



目前医院的数字化转型进程中面临着临床业务数据分散、患者隐私泄露、数据未授权使用、搜集病例数据耗时太多、数据难以复用以及数据标准不一等问题。随着医院和医生的数据意识加强，数字化场景日益丰富，实现医疗大数据精细化管理、加快健康医疗数据安全体系建设，强化数据安全监测和预警等成为必选项，必然要求更加安全可靠的算力。

### 进一步优化医疗资源跨时空配置，需以可灵活调度的算力为支撑



互联网医疗充分打破了医疗资源分布的时空限制，实现了医疗资源的合理利用和精准服务，当前互联网医疗应用已经从简单的医患跨时空对接，逐步过渡到线上线下一体化阶段，数字化诊疗服务将陪伴患者从门诊到住院、从院内到院外、从线上到线下的全病程，患者可以通过各类医疗智能硬件和软件应用快速、精准地获取医疗服务，对算力提出了可灵活调度的要求。

## 算力应用场景

### 案例十

#### 湖州市中心医院——可订阅式算力服务落地慢病管理场景

长期以来，国人的健康观念是“以治病为中心”，大量的医疗资源集中投入在急病和重症领域，慢病没有得到足够的重视，而事实上，随着人口老龄化的不断加剧，中国的流行病学模式已经从传染病向慢性非传染性疾病转变，慢病管理服务的需求不断上扬。

湖州市中心医院是浙江省湖州市的一家三级甲等综合性医院，承担着湖州市及省内、外部分周边地市约300万人口的医疗、保健任务。近年来，湖州当地政府不断加大对慢病防治的关注度，2021年《浙江省卫生健康事业发展“十四五”规划》明确要求“将重大慢病过早死亡率控制在8.5%以下”。在此背景下，湖州市中心医院将养老和健康管理作为重点发展方向，积极推动慢病科研和管理工作。

#### 业务需求

湖州市中心医院计划建设专门的慢病管理中心联合实验室和慢病管理系统，但面临算力资源不足和缺乏合理算力建设模式的双重挑战。

**重塑数据采集流程，原有存储和计算资源相对不足：**湖州市中心医院原有HIS（医院信息系统，Hospital Information System）主要记录了患者从入院到出院的数据，相对缺少患者在院外的数据，这就导致慢病管理数据不全，湖州市中心医院计划重塑数据采集流程，实现院内院外数据一体化，其本地数据中心无法承担激增的存储和计算资源需求，亟需补充额外算力。

**公有云模式不利于数据安全，亟需兼具灵活性和安全性的算力服务：**针对补充额外算力这一问题，湖州市中心医院发现虽然公有云可以提供灵活、可扩展、快速部署的算力服务，但存在安全隐患，为此，需要兼具灵活性与安全性的解决方案。

#### 解决方案

联想为湖州市中心医院提供了基于“端-边-云-网-智”技术架构的智慧医疗解决方案，在TruScale服务模式，湖州市中心医院能以“按需购买，按量付费”的形式订阅算力等IT资源。

**快速搭建IT基础架构，提供稳定可靠的算力底座：**以“端-边-云-网-智”技术架构为基础，联想利用高性能服务器快速搭建了的IT基础设施平台，以满足湖州市中心医院不断增长的数据存储和计算资源需求，并定期将数据备份到磁带库以实现数据保护和灾难恢复，确保数据安全。

**创新服务模式，减轻医院IT投入和管理负担：**联想主要以TruScale服务模式进行项目交付，底层逻辑在于打破软硬件及服务分散采购和投建运管分段实施的传统模式，整合为订阅式、一站式全包的服务模式，能充分保证算力等IT资源的灵活弹性扩展，在商业模式上实现“按需付费”，减轻了湖州市中心医院自建数据中心的成本投入和后续IT运维负担。

图 20 | 在“端-边-云-网-智”各层级，TruScale服务都有对应的解决方案



资料来源：联想

### 案例企业收益

- 慢病数据库逐步建成：**在全新IT基础架构的支撑下，湖州市中心医院从200万份门诊数据和8万份住院数据中精准筛选出了慢病管理相关数据，数据维度包括慢病诊断人数、慢病类型、慢病地图、年龄分布等，覆盖7万多名慢病患者、12大类慢病病种。
- 切实助力医院慢病管理：**在IT基础设施以上，湖州市中心医院还与联想合作设计并开发了用于慢病跟踪管理的定制化管理平台，全面接入原有HIS系统，通过该平台，湖州市中心医院已实现对IgA肾病、肺结节、帕金森三种慢病患者的管理。

# 金融行业

## 算力需求洞察

无论是信贷、支付、理财等传统金融服务，还是供应链金融、智能投顾、数字货币等新型服务形态，背后都离不开算力的支撑，在金融服务敏捷化转型趋势下，算力向各类场景渗透的同时，也将朝着弹性灵活、安全可靠的方向持续进化。



### 敏捷服务成为金融机构核心竞争力，亟需弹性灵活的算力

随着金融服务不断走向线上化、场景化，敏捷高效的服务能力已成为金融业竞争焦点，金融机构往往需要在短时间内上线新应用，且在具体应用场景中还存在大量高并发、低时延计算需求，要求实现底层算力的灵活供给，根据实际所需资源量实现按需匹配。



### 数据安全问题不容忽视，安全存储、安全计算不可或缺

在《网络安全法》《数据安全法》《个人信息保护法》等法规相继出台、金融监管部门不断强化执法力度等严监管趋势下，分布式存储、多方安全计算等能够实现“数据不出本地”“数据可用不可见”的技术正不断走向落地，推动安全算力融入数字金融服务全流程和金融业务全链条。

## 算力应用场景

### 案例十一

#### 中信建投——基于超融合架构的安全算力助力敏捷转型

中信建投是一家大型综合证券公司，并于2019年启动智能客户服务平台建设，先后完成了智能外呼、智能质检、智能电话客服、员工赋能平台等系统的建设工作，目前该平台已具备语音识别、语音合成、语义理解、知识图谱等多项基础AI能力。

## 业务需求

随着智能客户服务平台上接入的应用不断增多，中信建投意识到要避免只重应用不注重基础平台能力建设带来的烟囱系统、架构落后难改造、新业务响应慢等问题，提出要构建“薄前台、厚中台、强核心”的数字化金融客户服务体系架构，亟需强化IT架构以提高智能客户服务平台的稳定性、安全性。

**集中式IT架构不利于保障业务连续性：**伴随业务创新越来越快，智能应用的不断增多，敏态业务快速响应的需求越来越强烈，整体平台的高可靠性、连续性问题日益明显，集中式IT架构存在固有的性能瓶颈且弹性差，难以充分保证上层应用稳定连续运行。

**业务安全保障工作难度大、运维成本较高：**为保障业务安全运行，中信建投的异地容灾业务应用必须与业务系统达到1:1的配置，运维难度和成本随着新型应用不断增多而抬升。

## 解决方案

中信建投一直采用“自研+外购”的方式推动智能客户服务平台建设，针对IT架构改造和构建异地灾容系统等问题，中信建投积极引入了联想的解决方案，充分利用微服务、云原生等技术降低了运维成本。

**搭建超融合架构，夯实数字底座：**联想超融合架构实现了整个基础架构软硬件规格的统一，为后续的扩展、升级明确了方向，轮询的系统升级机制极大降低了系统风险。此外，超融合架构采用扁平化、分布式设计，大大缩短采购、供货和测试时间，满足了敏态业务快速上线的需求。

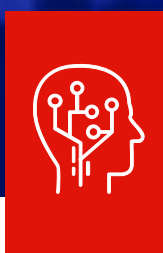
**改造灾备系统，提高易管理、易拓展等功能：**联想根据中信建投业务需求，将灾备机房划分为2个中心集群，每个集群运行虚拟化软件以及相应的业务系统。在底层存储方面，主要由HX分布式池提供存储资源，并通过超融合管理平台对服务器硬件、虚拟化软件、存储进行统一管理和维护，保证后期可进行无缝横向扩展。

## 案例企业收益

**降本增效成果显著：**在“薄前台、厚中台、强核心”的体系架构支撑下，中信建投智能服务平台年咨询量增加超60%，自动问答与辅助问答次数占服务总量的60%以上，智能外呼系统日均外呼9.7万通，按人工每人每天处理160件回访任务计算，平均每工作日可节省27个人工。

**充分保证了敏捷业务响应：**随着平台稳定性不断提升，智能IVR（交互式语音应答，Interactive Voice Response）系统可以做到7\*24小时无人值守，日均服务量占呼叫中心日均总服务量的18%，在2022年疫情封控期间，有效地实现了对客户的无间断服务。

# 06 筹谋： 趋势展望篇



## “普惠”算力助力充分释放数据红利， 中国数字经济体量有望突破百万亿元

当前，数据对提高生产效率的乘数作用不断凸显，是最具时代特征的生产要素，与其他生产要素相比，数据具有可复制、非消耗、边际成本接近于零等新特性，打破了自然资源有限供给对增长的制约，能够为经济转型升级提供不竭动力。

中国网民数量、数据资源、数字化应用场景全球领先，14亿人口所产生的数据资源将支持中国数字经济发挥规模经济、范围经济和网络经济优势。“普惠”算力作为数字经济的基础设施，可以强有力地推动跨层级、跨地域、跨部门数据汇聚融合、共享交换和开发应用，提升数据要素的整体流通效能和经济社会价值。数据+“普惠”算力的双向驱动，将充分盘活数据资产，释放“数据红利”，为中国数字经济的发展提供坚实基础。据中国（深圳）综合开发研究院技术团队预测，2020-2025年，中国数字经济年均增速将保持在15%左右。到2025年，数字经济规模将突破80万亿元，占GDP比重达55%。到2030年，数字经济体量有望突破百万亿元<sup>37</sup>。

<sup>37</sup> 2030年我国数字经济体量有望突破百万亿元——各地为何争相布局数字经济，[http://paper.ce.cn/jjrb/page/1/2020-11/23/07/2020112307\\_pdf.pdf](http://paper.ce.cn/jjrb/page/1/2020-11/23/07/2020112307_pdf.pdf)



## “普慧”算力推动大小模型协同进化，通用人工智能将真正走向落地

中国是世界人工智能重要领军国家之一，各级政府、科研机构、企业等都在积极推动AI大模型发展。例如，北京市经济和信息化局发布的《2022年北京人工智能产业发展白皮书》中，明确提出支持头部企业打造对标ChatGPT的大模型，着力构建开源框架和通用大模型的应用生态。此外，智源研究院、清华大学等单位联合研发了全球最大的智能模型“悟道2.0”，模型参数规模达到了1.75万亿<sup>38</sup>，国内各大科技公司也纷纷推出了自家的大模型产品，AI大模型的竞赛正在打响。数据、算法、算力三大核心要素是各类大模型竞争焦点，现阶段投入商用的大模型参数量基本在千亿级，以百度的“文心一言”为例，其模型参数量为2,600亿。公开资料显示，目前在追赶GPT3.5的中国团队约有10个，能够拥有或持续获取算力将是决定未来竞争格局的关键，对比当前国际头部大模型团队1万张H100（约10万A100）的算力规模来看，中国未来能拥有此等算力体量的团队可能不超过2个<sup>39</sup>。

<sup>38</sup> 《2022年北京人工智能产业发展白皮书》重磅发布，北京市人民政府，2022年2月14日，[https://www.beijing.gov.cn/ywdt/gzdt/202302/t20230214\\_2916514.html](https://www.beijing.gov.cn/ywdt/gzdt/202302/t20230214_2916514.html)

<sup>39</sup> 陆奇最新演讲全文实录：大模型带来的新范，凤凰网科技，2023年5月11日，<https://tech.ifeng.com/c/8Pi1FcSOTLF>

值得注意的是，AI大模型的真正价值最终将体现在具体场景中，未来产业竞争将从“规模”转向“应用”。目前大模型训练都在云端实现，所用算力是中心化的，精度要求高且成本高，意味着大模型落地必然会面临能耗和性能平衡的难题。破局之道在于大小模型协同进化，即在利用大参数训练完大模型之后，通过高精度压缩，将大模型转化为端侧可用的小模型，大模型相当于超级大脑，小模型相当于垂直领域专家，共同推动AGI落到实处。而这一过程中，算力由中心化走向泛在化，由通用化走向智能化，恰恰也是“普惠”算力释放价值的过程。

“普惠”算力作为通用人工智能的核心动力，既是帮助科技公司搭建AI模型的“卖铲人”，又是检验算力公司创新能力、制造能力的“试金石”，更是推动中国迈向科技强国的“加速器”。在全球算力供给告急的背景下，“普惠”算力正在迎来规模化、产业化发展机遇。

## “普惠”算力要求安全不容有失， 数据、算法、算力三位一体



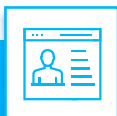
### 数据安全：已上升至国家立法层面

中国已将数据安全提升至国家立法的层面，《国家安全法》《网络安全法》《数据安全法》及《个人信息保护法》等共同构筑了数据安全保护的法律法规框架体系。推动算力发展过程中，数据流动和安全发展的矛盾将逐渐显现，必须严守数据安全防线，规避数据泄露、数据贩卖、数据滥用等风险。《全国一体化大数据中心协同创新体系算力枢纽实施方案》强调，“加强对基础网络、数据中心、云平台、数据和应用的一体化安全保障，提高大数据安全可靠水平。”

具体来说，可从数据要素全生命周期加强安全防范。在数据传输流转阶段，可以利用区块链分布式、透明性、可追溯性和不可篡改性等特征，促进数据流动过程透明化，不断增强数据流动的时效性和安全性。在数据存储阶段，可以不断提高相关数据设施的容灾能力，合理评估并积极应对不同类型、不同级别数据的泄露风险；在数据访问阶段，要加快数据清权确权，建立数据权责清单，让各使用主体以规治数、按权用数、决策循数。在数据使用、共享阶段，要落实事前可预防、事中可阻断、事后可溯源的全方位数据安全，实现数据“可用不可见”。







## 算法安全：伦理规范问题不可忽视

任何一项新技术或应用的诞生，都必然伴随着风险和争议，公开资料显示，ChatGPT已经引起美国、欧盟等关于人工智能立法的讨论，实际上AI决策难以解释、不可问责一直是棘手难题，导致了无人驾驶、手术机器人等难以落地，数字金融欺诈、大数据杀熟、生物特征识别等更是不断引发信任危机。随着ChatGPT等生成式AI走向应用，AI在“可解释性”上面临的挑战可能更加严峻，例如，现阶段ChatGPT有能力完成各种基于文本的任务，但它不可避免地会产生一些与事实不符的内容，原因是ChatGPT在回答问题时只是根据前一个单词“预测下一个单词”，进而生成答案，这一过程绕不开神经网络运行机理产生的黑箱效应，运行过程难以解释，部分运行结果的说服力也将大打折扣，某种程度上，AI缺乏“可解释性”带来的“不可信”，正在阻碍AI大规模应用，也是算力发展必须克服的难关。

中国正在以算法安全可信、高质量、创新性发展为导向，建立健全算法安全治理机制，构建完善算法安全监管体系，推进算法自主创新，《关于加强互联网信息服务算法综合治理的指导意见》《互联网信息服务算法推荐管理规定》等政策文件出台，中国开始迈向对算法进行事前事中全流程、动态规制与监管的“算法法”时代。



## 算力安全：战略地位将进一步提高

对比数据安全、算法安全来看，未来中国算力安全的战略地位有机会且有必要进一步提升。在国家政策层面，算力安全受到的重视程度相对不足，多以地方政府级别的政策为主，例如，上海市经济信息化委发布了《上海市推进算力资源统一调度指导意见》，提出鼓励建设国产自主可控、安全可靠的算力基础设施和基于国产自主可控的算力应用生态。2023年1月，成都发布全国首个算力产业政策《成都市围绕超算智算加快算力产业发展的政策措施》，直接提出鼓励智算中心建设国产自主可控、安全可靠的人工智能算力基础设施和技术路线生态，打造全球领先的人工智能计算平台、城市智脑平台等，提供普惠算力服务。

保障算力安全需要国家从顶层设计发力，推动各界共同努力。2022年8月，美国政府限制英伟达向中国出口GPU计算芯片A100和H100，而高端GPU计算芯片作为算力核心一旦受限，中国算力产业发展将受到明显影响。具体来说，高端芯片、数据中心、网络等基础设施是算力产业可持续发展的重要保障，要以核心技术自主创新为牵引，加快研发突破，通过产学研联合攻关增强算力产业链的稳定可控性，不断提升算力的自主可控水平。此外，中国超大型、大型数据中心仍有较大发展空间，需坚持适度超前原则，以建代用，尽可能提高算力基础设施的承载能力。

<sup>40</sup> 周杰，柯沛等，ChatGPT：潜力、前景和局限，信息与电子工程前沿FITEE（2023）

# “普惠”算力发展任重道远， 绿色化、场景化、市场化三措并举

## 举措一

### 算力绿色化关乎可持续发展，亟需提高单位能耗的计算能力

能量和信息是科技发展的主线，计算的本质是利用能量对信息进行处理加工，信息越多，计算就越复杂，需要的能量也越多，当能量增长跟不上近乎无限扩张的信息时，提高能量利用率将成为必然选择。

从能量和信息转换的角度来说，不断提高单位能耗的计算能力是算力发展的重要底层驱动力之一，而公开资料显示，目前大约80%的电能消耗只是在维护计算机状态，只有不到20%的电能消耗在做计算。此外，“双碳”目标背景下，数据中心等能耗大户带来的碳排放问题亟需得到解决，根据开放数据中心委员会测算，2020年中国数据中心能耗总量为939亿千瓦时，碳排放量为6,464万吨，预计到2030年，中国数据中心能耗总量将达到3,800亿千瓦时左右，碳排放增长率将超过300%。

因此，算力绿色化不仅符合科技发展的基本规律，更关乎人类命运共同体前途，已成为全产业链的关键课题。算力绿色化核心目标是提高每瓦功耗所产生的算力，主要发展思路包括提高绿色能源使用占比、采用创新型制冷技术降低数据中心散热过程中能耗、综合管理IT设备提高算力利用效率等。例如，在用能方面，可以提高光伏等新能源系统发电占比；在降耗方面，可以进一步推动液冷、温水水冷取代传统风冷；在提效方面，可以利用服务器虚拟化等技术提高设备使用率。

<sup>41</sup> 电商数据中心如何降低碳排放？多家头部互联网企业开放绿色计算专利技术，21经济网，2022年11月9日，<https://www.21jingji.com/article/20221109/herald/e775972cc330fc6d49a00f97a980500c.html>

<sup>42</sup> 《数据中心算力碳效白皮书（2022年）》，开放数据中心委员会，2022年10月

## 举措二

### 算力场景化面临知识壁垒，亟需业务、数据、算力高效联动

算力要落地到具体场景中才能真正转化为生产力，而行业愈是细分、业务愈是复杂，算力场景化落地面临的“知识壁垒”就越高。例如，大量行业垂直类软件要求高度凝练总结行业经验、最佳实践等知识，这一知识软件化的过程，经常伴随着非结构化和半结构化数据占比不断升高、数据负载从单一负载转向混合负载等结果，会使得数据流通、压缩存储、加密解密等对算力的需求趋向复杂化和定制化，正倒逼着算力服务商不断提高行业理解力、强化数据资产沉淀、提高算力针对场景化需求的适配度等。

可以预见，随着各行业全要素、全流程、全场景迈向数字化和智能化，数据架构和业务架构将不断向更强大的算力平台上迁移，实现“业务、数据、算力”之间的高效联动，是保证算力资源按需匹配、精准赋能的必要前提。

## 举措三

### 算力市场化仍需加强，亟需多方共同推动算力统一交易

当前算力需求增长确定性高，算力供给规模相对有限，通过市场化交易可以大幅提升算力资源配置效率，是破解算力供需矛盾，推动算力产业生态进入正向循环的关键。各方仍在积极探索算力交易模式，亟需解决算力归属复杂、难以度量定价、区域发展不平衡等问题。

针对算力归属复杂的问题，主要原因在于当前规模化算力基础设施大多由多方共建，要求算力建设方明确划分算力资源占比，保证算力产权可溯源；针对算力难以度量定价的问题，可以提倡由行业管理部门和龙头类企业带头建立统一标准，为算力定价提供可靠依据；针对算力区域发展不平衡问题，在“东数西算”工程稳步推进过程中，各级地方政府部门可以结合区域算力资源禀赋和市场环境，“因地制宜”建设算力基础设施，积极融入全国算力一体化网络，大力建设算力供需适配平台、算力交易平台等，畅通区域间算力流通、共享、按需分配。

## 关于毕马威中国

毕马威中国在三十一个城市设有办事机构，合伙人及员工超过15,000名，分布在北京、长春、长沙、成都、重庆、大连、东莞、佛山、福州、广州、海口、杭州、合肥、济南、南京、南通、宁波、青岛、上海、沈阳、深圳、苏州、太原、天津、武汉、无锡、厦门、西安、郑州、香港特别行政区和澳门特别行政区。在这些办事机构紧密合作下，毕马威中国能够高效和迅速地调动各方面的资源，为客户提供高质量的服务。

毕马威是一个由独立的专业成员所组成的全球性组织，提供审计、税务和咨询等专业服务。毕马威国际有限公司（“毕马威国际”）的成员所以毕马威为品牌开展业务运营，并提供专业服务。“毕马威”可以指毕马威全球网络内的独立成员所，也可以指一家或多家毕马威成员所。

1992年，毕马威在中国内地成为首家获准中外合作开业的国际会计师事务所。2012年8月1日，毕马威成为四大会计师事务所之中首家从中外合作制转为特殊普通合伙的事务所。毕马威香港的成立更早在1945年。率先打入市场的先机以及对质量的不懈追求，使我们积累了丰富的行业经验，中国多家知名企业长期聘请毕马威提供广泛领域的专业服务（包括审计、税务和咨询），也反映了毕马威的领导地位。

## 关于联想集团

联想集团是一家成立于中国、业务遍及180个市场的全球化科技公司。联想聚焦全球化发展，持续开发创新技术，致力于建设一个更加包容、值得信赖和可持续发展的数字化社会，引领和赋能智能化新时代的转型变革，为全球数以亿计的消费者打造更好的体验和机遇。

联想作为全球领先ICT科技企业，秉承“智能，为每一个可能”的理念，持续研究、设计与制造全球最完备的端到端智能设备与智能基础架构产品组合，为用户与全行业提供整合了应用、服务和最佳体验的智能终端，以及强大的云基础设施与行业智能解决方案。

联想每年为全球用户提供数以亿计的智能终端设备，包括电脑、平板、智能手机等，同时积极推动全行业“设备+云”和“基础设施+云”的发展，以及智能化解决方案的落地。目前，联想是全球第一大PC厂商、第三大服务器厂商，并11次高居全球算力500强榜单榜首。

面向新一轮智能化变革的产业升级契机，联想提出智能变革战略，围绕智能物联网（Smart IoT）、智能基础架构（Smart Infrastructure）、行业智能与服务（Smart Verticals & Services）三个方向成为行业智能化变革的引领者和赋能者。联想将继续投资于技术创新与社会价值、深化以服务为导向的转型，强化同一个联想和数字化，并在未来持续将服务与解决方案打造成联想新的核心竞争力。

目前，联想核心业务由三大业务集团组成，分别为专注智能物联网的IDG智能设备业务集团、专注智能基础设施的ISG基础设施方案业务集团及专注行业智能与服务的SSG方案服务业务集团，在全球约有77,000名员工。2022/23财年，联想集团的整体营业额已达4240亿人民币。

# 联系我们



## 康勇

首席经济学家  
毕马威中国  
邮箱: k.kang@kpmg.com  
电话: 010-8508 7198



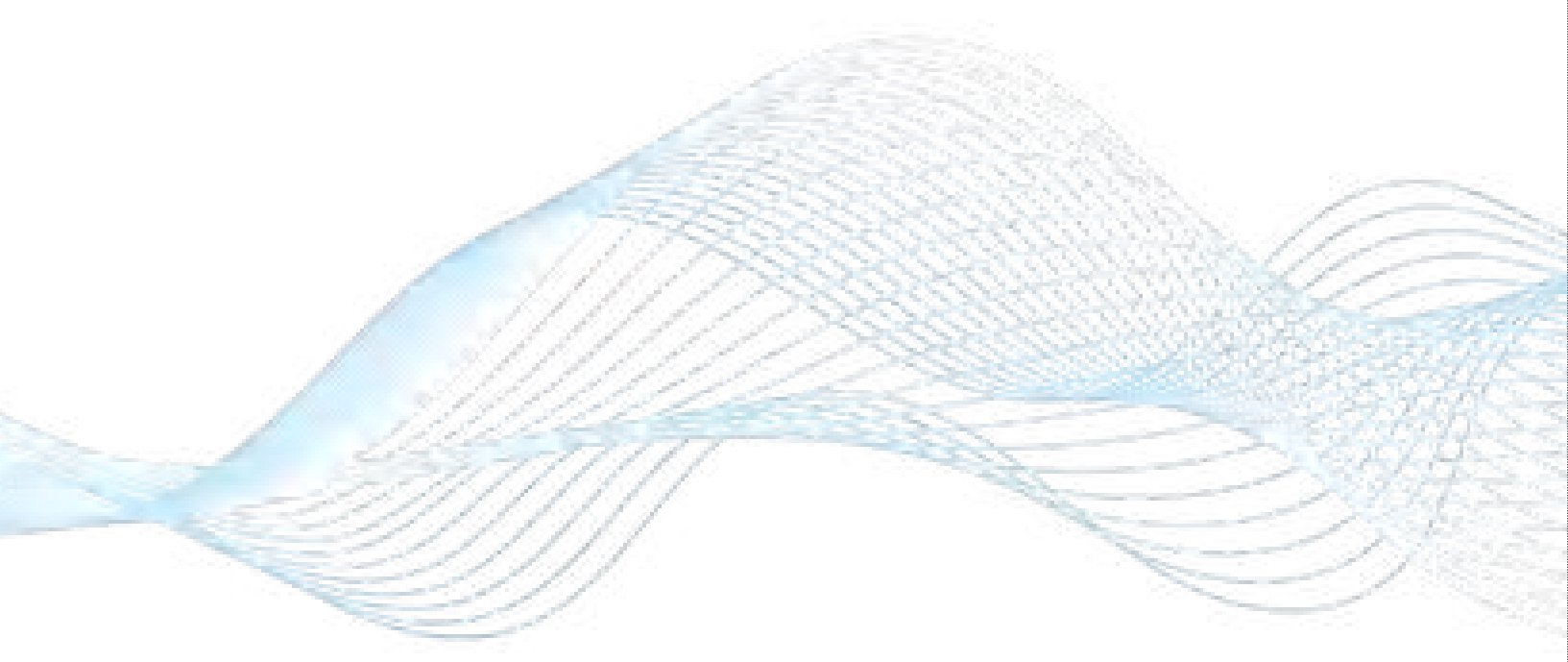
## 张庆杰

数字化赋能咨询服务主管合伙人  
毕马威中国  
邮箱: qingjie.zhang@kpmg.com  
电话: +86 10 8508 4069



## 陈俭德

科技、通信及媒体行业主管合伙人  
毕马威中国  
邮箱: daniel.chan@kpmg.com  
电话: +86 21 2212 2168



[kpmg.com/cn/socialmedia](https://kpmg.com/cn/socialmedia)



如需获取毕马威中国各办公室信息，请扫描二维码或登陆我们的网站：  
<https://home.kpmg.com/cn/en/home/about/offices.html>

本刊物所载资料仅供一般参考用，并非针对任何个人或团体的个别情况而提供。虽然本所已致力提供准确和及时的数据，但本所不能保证这些数据在阁下收取本刊物时或日后仍然准确。任何人士不应在没有详细考虑相关的情况及获取适当的专业意见下依据本刊物所载资料行事。

© 2023 毕马威华振会计师事务所（特殊普通合伙）— 中国合伙制会计师事务所，毕马威企业咨询（中国）有限公司— 中国有限责任公司，毕马威会计师事务所— 澳门特别行政区合伙制事务所，及毕马威会计师事务所— 香港特别行政区合伙制事务所，均是与英国私营担保有限公司— 毕马威国际有限公司相关联的独立成员所全球性组织中的成员。版权所有，不得转载。在中国印刷。

毕马威的名称和标识均为毕马威全球性组织中的独立成员所经许可后使用的商标。