

2023

十大科技趨勢

Top Ten Technology Trends of DAMO Academy

10

達摩院

ALIBABA DAMO ACADEMY



人工智能产业链联盟

星主： AI产业链盟主

 知识星球

微信扫描预览星球详情



卷首语

达摩院成立之初，就以人类愿景为使命，立足基础科学、创新性技术和应用技术的研究，致力于探索人类未知与科技前沿。谋定而动方能笃行致远，在纷纷扬扬的潮流喧嚣中，看清未来的走向，需要在科学与技术、产业与应用之间，不断进行思想碰撞、达成共识、创造价值，探寻更具确定性的未来和科技的前进方向。

今年，我们本着回归本质（Back to Basic）的初心，站在产业的角度，审视 2023 年科技趋势，评估那些已经工程化落地、有望近期规模化商用的主流技术，那些具有前沿性的技术探索与实践验证相结合的科技创新产品，以及已经在产业链上下游形成强大生态的应用体系。希望能以此激发科学家、企业家、工程师以及泛科技爱好者的思考与共鸣，共同推动科技创新，实现高水平科技自立自强和全球化发展的互利共赢。

展望 2023，多元技术的协同并进驱动计算与通信的融合、硬件和软件的融合，应用需求的爆发驱动 AI 技术与行业的融合，数字技术与产业生态的融合，企业、个人与政府在安全技术与管理上的融合。科技进步与产业应用双轮驱动的融合创新已成为不可逆转的宏大趋势。

以科技，创未来。我们相信，新一代信息通信技术将为企业和个人创造一个高质量发展与高品质生活共享的未来。我们希望通过分享达摩院在科研实践过程中的洞察、思考和判断，与大家进行思想上的交流和碰撞，为科技的进步贡献一份绵薄之力。

张建锋

阿里巴巴达摩院院长

序言一

每到岁末年初，各国政府、企业及研究机构通常都会做回顾与展望。波谲云诡的国际局势让 2022 年显得更不寻常，新一轮的科技革命和产业变革正在重构全球创新版图。2023 年，科技的走向依旧是世界各国的关注重点，各国在纷纷设立自己的科技战略目标外，还在潜心研究不同技术领域的科技趋势，通过科技占据国际竞争的制高点。科技和产业的联系将会变得更加紧密，新一代信息技术和产业的融合创新发展成为时代主旋律，技术创新将加速转变为现实生产力。

当下，我们需要通过自主研究提出可引导与支撑我国科技和产业发展的技术趋势。只有对前沿技术、颠覆性技术、以重大科技问题为导向的技术趋势及各领域的交叉融合建立深刻理解，才能实现我国整体科技水平从跟跑到领跑的战略性转变。技术没有一劳永逸，永远领先的关键在于持续创新。技术前瞻性预测分析工作难度大，准确的技术预见既需要有工程实践经验的积累和感性认识，也需要有科学理论基础的支撑与理性思维。特别是涉及到多个领域，对科技和产业发展起到关键作用的共性技术，需要产、学、研各界多方面的研究合作和支持。

达摩院是一所致力于开展基础科学和颠覆式技术创新研究和应用的新型研发机构。每年的达摩院十大科技趋势报告是站在科技和产业发展全局的角度，对未来科技的发展方向进行探索，并做出科学、客观、中立的预判。特别欣慰地看到，2023 年的十大科技趋势报告也秉承了这一理念，为科技界和产业界贡献了一份非常有价值、有深度的预见。

邬贺铨

中国工程院院士

序言二

国家提出制造强国、网络强国以来，又提出数字中国，全面推进数字政府、数字经济、数字社会、数字生态建设。数字化已经全面深度融入我们的生产生活，数字化的内容生产与服务也催生了一批全球领先的科技企业，让我们站到了世界科技发展的前沿。我们看到，数据成为生产要素，数字经济成为国家经济重要增长力量和重要组成部分。

在这样一个数据和信息爆炸的急剧变化的数字化社会，年度科技趋势的研究与发布工作并不容易。走过高速发展的时代，在和平与发展依然是主旋律的今天，人们更加关注发展和安全，企业也越来越愿意在不确定的方向上去探索和投入。

云计算的普及和成本的降低，正在让计算更加地“算得快”“算得准”“算得好”，人们越来越感受到数字化生活的便利、高效与美好。达摩院每年发布的十大科技趋势，也已经成为政府、产业、学界以及企业家非常关注的话题，这说明人们依然对互联网科技企业抱有很高的期待。

最近，OpenAI 发布的 ChatGPT 受到广泛关注，它不仅会聊天、编故事、写论文，还能从事写代码、改程序 Bug、构建虚拟机等更复杂的工作。在今年达摩院发布的十大科技趋势中，我们看到一些投资界和产业界已经高度关注的存算一体、芯片先进封装、软硬件的深度融合等科技趋势，也发现了一些有潜在重大应用价值的端网融合的可预期网络、大规模城市数字孪生、基于多模态的预训练大模型等产业趋势，既能引发人们对现在与未来的思考，也能带动人们对科技与产业的投入，其研究是很有价值的。

当然，科技的发展不能做“思想的巨人、行动的矮子”。在此祝愿达摩院在致力于科技趋势研究的同时，也能在基础研究和技术应用上取得更重大的成绩与成就，为国家的高质量发展、人们的高品质生活贡献更多的科技成果与产品服务。

徐愈

中国信息化百人会执委

序言三

1965年4月19日，美国仙童半导体公司一位叫戈登·摩尔的工程师，应邀撰写了一页纸的短文《让集成电路填满更多的元件》，在《电子学》杂志刊出。他也许不会想到，自己的名字日后会家喻户晓。他的预言被完善，称为摩尔定律。67年过去了，这位工程师的预见还在深刻影响信息技术产业，影响我们的生活。

对未来的思考研究判断预言，是人类的思想利器，是照亮前程的灯塔。处在科学与技术迅猛发展变化的时代，人类社会从来没有像今天这样需要穿透未来的思想洞见，从来没有像今天这样需要准确把握科技发展趋势。

预言未来趋势又是一件充满风险、具有挑战性的工作。人们见过戈登·摩尔这样的成功者，也见过误判趋势的失败者。拥有贝尔实验室的AT&T滑向颓势、举国之力的日本五代机项目遭遇惨败，尽管因素复杂，但误判趋势是其中的重要原因。

从这个意义上讲，阿里巴巴达摩院的科技趋势研究项目，重要而具有时代价值。这个项目到今天已经坚持五年，难能可贵。

这项创新性工作的宏大背景，是中国正在建设创新型国家。经过多年的跟随与追赶，我们逐渐在科技领域缩小了与全球领跑者的距离，在有些领域能够并驾齐驱甚至出现领先。与这样的科技发展态势相适应，中国自己的高水平科技趋势发展研究体系应运而生，正成为北京、上海、合肥、粤港澳大湾区、杭州城西科创大走廊等综合性国家科学中心建设中不可或缺的重要内容。

建设创新型国家的一个重要方面，是企业成为科技创新的主体。阿里巴巴已成为深度参与全球科技竞争的重要平台型企业，具备与之相应的强大科技实力，准确把握未来科技发展趋势，正是这一实力的体现。作为致力于探索科技未知，以人类愿景为驱动力，立足于基础科学、创新性技术和应用技术的新型研究机构，达摩院担起了未来科技发展趋势研究的重任。

对于未来科技发展趋势研究这项具有挑战性的高难度工作，阿里巴巴达摩院充分汇集企业内外、国际国内的研究成果和思想资源。尤其是，在中国开创性建立了未来科技发展趋势研究的一套方法论。这一方法论，也会成为其他机构的重要借鉴。

今天，经过全体专家的辛勤工作，《达摩院2023十大科技趋势》研究报告来到我们面前。这是多年来达摩院的一项重要举措，值得肯定与称赞，每年的十大趋势研究报告都通过媒体向群众广为传播，今年还同时举办全球科技趋势论坛，以开放的活动使这一重要研究成果让行业内外共享，让全社会共享，不但影响了科技界和产业界，也惠及全民。

思想和知识，是人类共同的财富。用通俗的语言普及科学技术知识，让广大群众包括青少年不断了解最新科技发展动态与未来趋势，是长期而重要的工作。达摩院向公众的传播，正在潜移默化地促进全民科学素质提高，成为建设创新型国家的基础与力量。

达摩院年复一年的辛勤努力，已经生根开花，在未来必将有更丰硕的成果。

陈宗周

《电脑报》创始人、《环球科学》杂志创始人

方法论阐述

用一种确定的方法,对不确定的未来进行预测预判,是人类的朴素追求。唐纳德·司托克斯在《基础科学与技术创新:巴斯德象限》中把科技从理论研究到应用落地,分为了波尔象限、巴斯德象限、爱迪生象限与皮特森象限,提出了以巴斯德象限为中心的研究思路。以提升技术产业化应用和创新能力为目标,从学术创新、技术突破、产业落地、市场需求等多维度展开,聚焦那些即将实现大规模商业化落地、有望形成产业生态的主流趋势,是今年达摩院十大科技趋势研究的切入点。

在科技趋势的具体研究过程中,我们坚持数据分析与广泛调研相结合,延续了“定量发散,定性收敛”的分析方法,技术上‘Back to Basic’,以产业为出发点,聚焦 ICT 领域,针对计算(存储)、通信(网络)、安全(管理)、应用(行业)等四大领域展开全面的分析。

在定量发散阶段,以论文、专利作为基础数据,我们从计算(存储)、通信(网络)、安全(管理)、应用(行业)四大领域往下细分到子领域,基于论文与专利数量的绝对值与增长率,筛选出学术界与产业界重点关注的细分技术领域。

在定性收敛阶段,以专家访谈与行业洞察为基础,我们对定量发散阶段识别出来的重点领域与技术趋势做深入调研。通过交叉验证去除专家经验中的主观因素,深入挖掘技术概念、面临挑战、突破方向和产业价值,收敛科技趋势的候选清单,形成对趋势的主要判断。

在项目研究过程中,项目组共计访谈了近百位专家,包括科学家、企业家与工程师,并从定量发散到定性收敛,形成了底层技术突破引起的“范式重置”、科技融合触发的“产业革新”、创新应用带来的“场景变换”三大趋势池。

最后,以达摩院专家委员会的专家经验与直觉判断为基础,经过多轮收敛与票选讨论,基于理论关注度、技术可行性、产业化程度、社会价值等因素的综合衡量,我们最后遴选了 2023 最可能的十大科技发展方向,代表着我们对未来愿景的无限憧憬,奉献给新的一年。

目录

范式重置

多模态预训练大模型	p8
基于多模态的预训练大模型将实现图文音统一知识表示,成为人工智能基础设施	
Chiplet	p10
Chiplet 的互联标准将逐渐统一,重构芯片研发流程	
存算一体	p12
资本和产业双轮驱动,存算一体芯片将在垂直细分领域迎来规模化商用	

产业革新

云原生安全	p14
安全技术与云紧密结合,打造平台化、智能化的新型安全体系	
软硬融合云计算体系架构	p16
云计算向以 CIPU 为中心的全新云计算体系架构深度演进,通过软件定义,硬件加速,在保持云上应用开发的高弹性和敏捷性同时,带来云上应用的全面加速	
端网融合的可预期网络	p18
基于云定义的可预期网络技术,即将从数据中心的局域应用走向全网推广	
双引擎智能决策	p20
融合运筹优化和机器学习的双引擎智能决策,将推进全局动态资源配置优化	
计算光学成像	p22
计算光学成像突破传统光学成像极限,将带来更具创造力和想象力的应用	

场景变换

大规模城市数字孪生	p25
城市数字孪生在大规模趋势基础上,继续向立体化、无人化、全局化方向演进	
生成式 AI	p28
生成式 AI 进入应用爆发期,将极大地推动数字化内容生产与创造	

趋势一

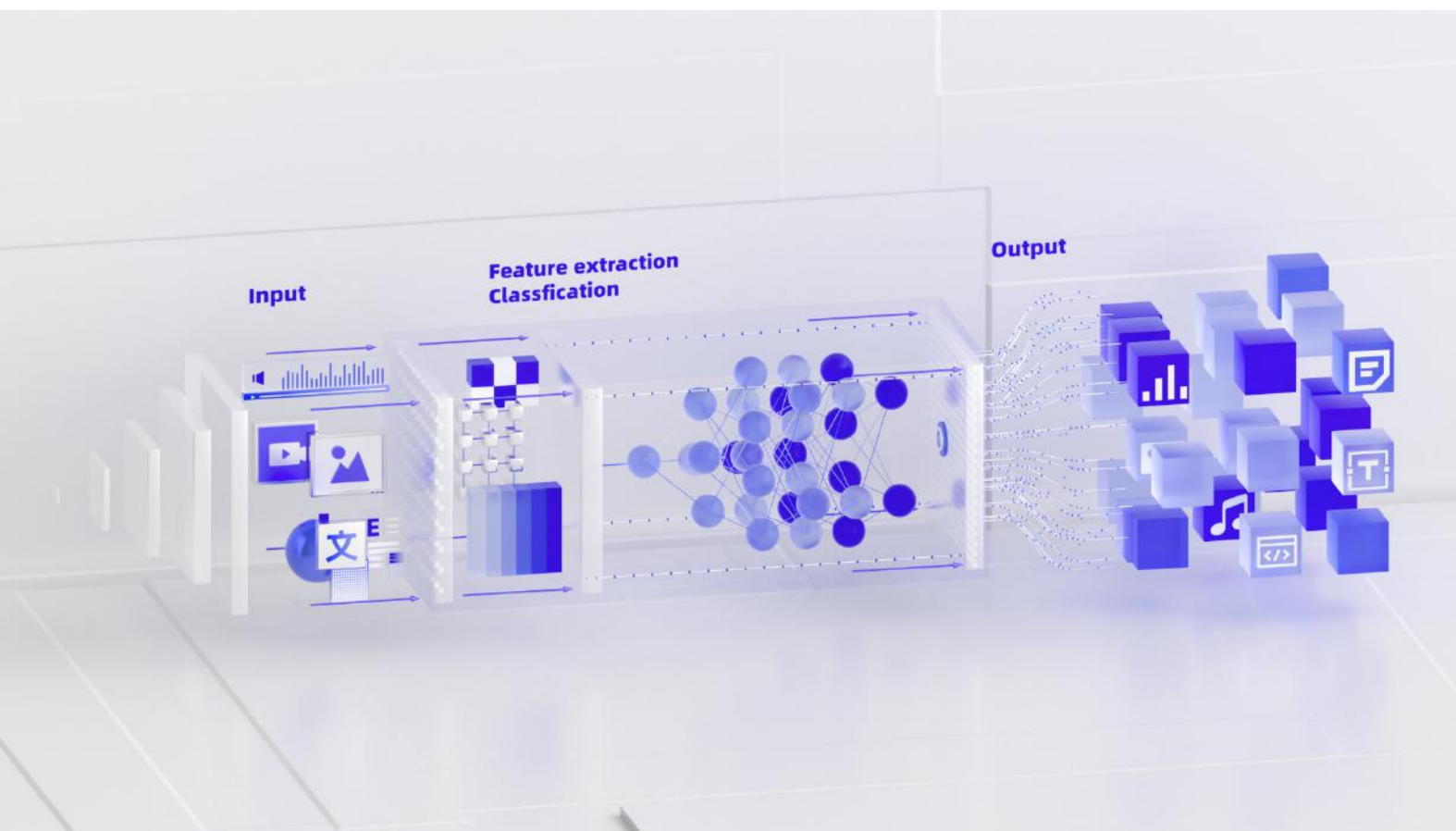
多模态预训练大模型

基于多模态的预训练大模型将实现图文音统一知识表示，成为人工智能基础设施。

摘要

人工智能正在从文本、语音、视觉等单模态智能，向着多种模态融合的通用人工智能方向发展。多模态统一建模，目的是增强模型的跨模态语义对齐能力，打通各个模态之间的关系，使得模型逐步标准化。目前，技术上的突出进展来自于 CLIP（匹配图像和文本）和 BEiT-3（通用多模态基础模型）。基于多领域知

识，构建统一的、跨场景、多任务的多模态基础模型已成为人工智能的重点发展方向。未来大模型作为基础设施，将实现图像、文本、音频统一知识表示，并朝着能推理、能回答问题、能总结、做创作的认知智能方向演进。



趋势解读

基于深度学习的多模态预训练是认知智能快速发展的重要推动力。构建多场景、多任务的预训练大模型将加速模型标准化进程，为人工智能模型成为基础设施创造条件。深度学习模型的不断完善、互联网海量真实数据的积累和生成式预训练的广泛应用，使得人工智能模型在自然语言理解、语音处理、计算机视觉等领域地交叉应用取得显著进展。

2022年，技术上的突出进展来自于 BEiT-3 多模态基础模型，该模型在视觉-语言任务处理上具备出色表现，包括视觉问答、图片描述生成和跨模态检索等。BEiT-3 通过统一的模型框架和骨干网络（backbone）建模，能够更加轻松地完成多模态编码和处理不同的下游任务。另一方面，CLIP（Contrastive Language-Image Pre-training）的广泛应用也促进了多模态模型的技术发展。CLIP 作为基于对比学习的预训练模型，负责从文本特征映射到图像特征，能够指导 GAN 或扩散模型（Diffusion Model）生成图像。在文生图领域，Stable Diffusion 也使用了 CLIP，它能够通过文本提示调整模型，并借助扩散模型改善图像质量。与此同时，开源极大的促进了多模态的融合和预训练模型的发展。通过开源来降低模型使用门槛，将大模型从一种新兴的 AI 技术转变为稳健的基础设施，已成为许多大模型开发者的共识。

多模态预训练模型的发展将重塑人工智能商业模式，并为人们的生产生活方式带来积极影响。对个人而言，类似

CLIP 的多模态模型，将使更多非技术出身的人能够表达自己的创造力，无需再借助工具和编程专业能力。对企业来说，多模态预训练模型将成为企业生产效率提升的关键。商业模式上，具备大数据、算力资源和模型开发能力的科技企业，将会成为模型服务的提供方，帮助企业将基础模型的能力与生产流程融合起来，实现效率和成本最优。

认知智能的发展，不会局限在文本或图像等单一的模态上。未来，如何针对不同模态建立更高效的模型架构和统一的骨干网络，使得大模型能够广泛地支持各种下游任务将成为主要挑战。在此基础上，更多的挑战来自于挖掘不同模态（如图像-文本，文本-自然语言，视频-文本）数据间的相关信息，并巧妙的设计预训练任务，让模型更好的捕捉不同模态信息之间的关联。

语音、视觉和多模态预训练模型将加速人工智能向通用基础模型方向演进。在这个演进过程中，深度学习与强化学习相互促进发展，融合大量行业知识，模型将具备在不断变化的环境中快速适应的灵活性。建立统一的、跨场景、多任务的多模态基础模型会成为人工智能发展的主流趋势之一。随着技术的不断成熟，大模型在开发成本、易用性、开发周期、性能上会更具优势，给产品化和商业化带来更多可能性。

专家点评 |

以 GPT 系列模型为代表的大模型（大规模预训练模型或者基础模型）展现出了超出预期的“理解”和“创造”能力，提示模型尺度与智能之间的正相关关系仍在生效。同时我们对大模型内在机理了解还远远不够，大模型的可解释性和可控性仍很弱，但这正是激励我们继续前行的动力。

文继荣

中国人民大学信息学院院长
高瓴人工智能学院执行院长

多模态预训练模型对图片-文字进行联合表征学习，并扩展到语音，视频等其他模态，在多个多模态相关的任务（理解，检索，生成，问答等）取得明显优势，在广泛的行业应用场景中能够从不同信息源获取知识并进行统一表示学习，可以成为覆盖广泛领域的通用基础模型。

黄非

达摩院语言技术实验室负责人

趋势二

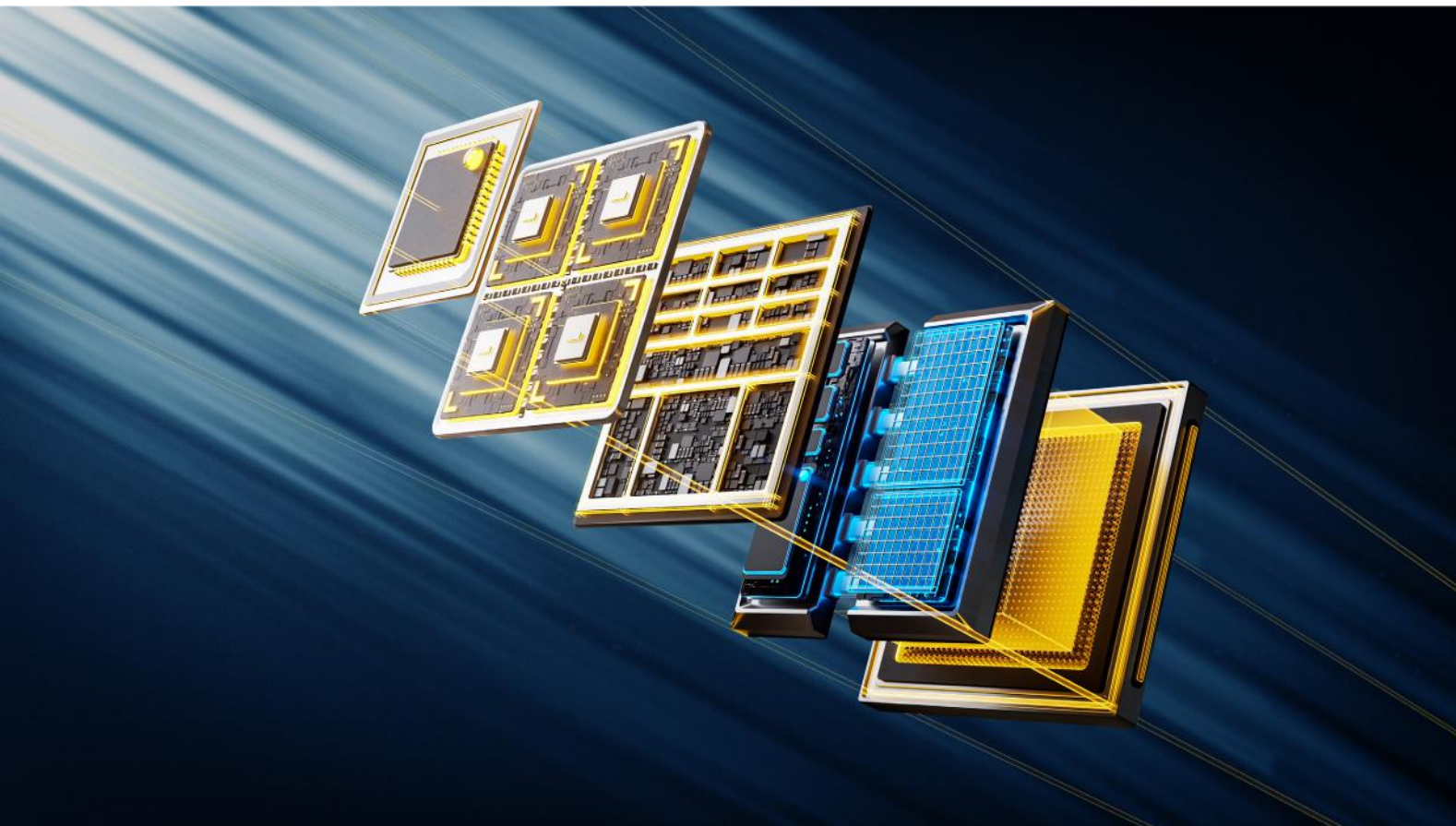
Chiplet

Chiplet 的互联标准将逐渐统一，重构芯片研发流程。

摘要

Chiplet 是硅片级别的“解构-重构-复用”，它把传统的 SoC 分解为多个芯粒模块，将这些芯粒分开制备后再通过互联封装形成一个完整芯片。芯粒可以采用不同工艺进行分离制造，可以显著降低成本，并实现一种新形式的 IP 复用。随着摩尔定律的放缓，Chiplet 成为持续提高 SoC 集成度和算力的重要途径，特

别是随着 2022 年 3 月份 UCle 联盟的成立，Chiplet 互联标准将逐渐统一，产业化进程将进一步加速。基于先进封装技术的 Chiplet 可能将重构芯片研发流程，从制造到封测，从 EDA 到设计，全方位影响芯片的区域与产业格局。



趋势解读

自 1965 年摩尔定律首次被提出以来，集成电路产业一直遵循着摩尔定律向前发展。直到近几年，随着晶体管尺寸逼近材料的物理极限，工艺节点进步的花费已难以承受，芯片性能的提升也不再显著，摩尔定律接近极致。在此背景下，Chiplet（芯粒）技术逐渐崭露头角，有望成为产业界解决高性能、低成本芯片需求的重要技术路线。

Chiplet 创新了芯片封装理念。它把原本一体的 SoC（System on Chip，系统级芯片）分解为多个芯粒，分开制备出这些芯粒后，再将它们互联封装在一起，形成完整的复杂功能芯片。这其中，芯粒可以采用不同的工艺进行分离制造，例如对于 CPU、GPU 等工艺提升敏感的模块，采用昂贵的先进制程生产；而对于工艺提升不敏感的模块，采用成熟制程制造。同时，芯粒相比于 SoC 面积更小，可以大幅提高芯片的良率、提升晶圆面积利用率，进一步降低制造成本。此外，模块化的芯粒可以减少重复设计和验证环节，降低芯片的设计复杂度和研发成本，加快产品的迭代速度。Chiplet 被验证可以有效降低制造成本，已成为头部厂商和投资界关注的热点。

Chiplet 的技术核心在于实现芯粒间的高速互联。SoC 分解为芯粒使得封装难

度陡增，如何保障互联封装时芯粒连接工艺的可靠性、普适性，实现芯粒间数据传输的大带宽、低延迟，是 Chiplet 技术研发的关键。此外，芯粒之间的互联特别是 2.5D、3D 先进封装会带来电磁干扰、信号干扰、散热、应力等诸多复杂物理问题，这需要在芯片设计时就将其纳入考虑，并对 EDA 工具提出全新的要求。

近年来，先进封装技术发展迅速。作为 2.5D、3D 封装关键技术的 TSV（Through Silicon Via，硅通孔）已可以实现一平方毫米 100 万个 TSV。封装技术的进步，推动 Chiplet 应用于 CPU、GPU 等大型芯片。2022 年 3 月，多家半导体领军企业联合成立了 UCle（Universal Chiplet Interconnect Express，通用 Chiplet 高速互联联盟）。Chiplet 互联标准有望逐渐实现统一，并形成开放性生态体系。

面向后摩尔时代，Chiplet 可能将是突破现有困境最现实的技术路径。Chiplet 可以降低对先进工艺制程的依赖，实现与先进工艺相接近的性能，成为半导体产业发展重点。从成本、良率平衡的角度出发，2D、2.5D 和 3D 封装会长期并存；同构和异构的多芯粒封装会长期并存；不同的先进封装和工艺会被混合使用。Chiplet 有望重构芯片研发流程，从制造到封测，从 EDA 到设计，全方位影响芯片产业格局。

专家点评 | ④

Chiplet 技术是提高芯片集成度、节约芯片成本、实现晶粒（die）级可重用的最重要的方法。未来，Chiplet 技术将在高性能计算、高密度计算等领域发挥着重要作用。先进的 Chiplet 技术将继续由代工厂主导，混合使用 2D、2.5D、3D 等先进封装技术将进一步提高产品性价比与竞争力。

王海洋

象帝先计算技术有限公司副总裁

趋势三

存算一体

资本和产业双轮驱动，存算一体芯片将在垂直细分领域迎来规模化商用。

摘要

存算一体旨在计算单元与存储单元融合，在实现数据存储的同时直接进行计算，以消除数据搬移带来的开销，极大提升运算效率，以实现计算存储的高效节能。存算一体非常符合高访存、高并行的人工智能场景计算需求。在产业和投资的驱动下，基于 SRAM、DRAM、Flash 存储介质的产品进入验证期，将优先在低功耗、小算力的端侧

如智能家居、可穿戴设备、泛机器人、智能安防等计算场景落地。未来，随着存算一体芯片在云端推理大算力场景落地，或将带来计算架构的变革。它推动传统的以计算为中心的架构向以数据为中心的架构演进，并对云计算、人工智能、物联网等产业发展带来积极影响。



趋势解读

随着 AI 在各领域的应用逐渐广泛，以深度学习为代表的神经网络算法需要系统高效处理海量的非结构化数据，例如文本、视频、图像、语音等。而传统冯·诺依曼体系下运行的计算机通常包括存储单元和计算单元两部分，数据需要在处理器与存储器之间进行频繁迁移，如果内存的传输速度跟不上 CPU 的性能，就会导致计算能力受到限制，出现“内存墙”“功耗墙”。这就对芯片的并行运算、低延迟、带宽提出了更高的要求。

近年来，产业界领军企业在存算一体的前沿技术研究上持续发力。三星在顶级学术期刊 Nature 上发表了全球首个基于 MRAM（磁性随机存储器）的存内计算研究；台积电在 ISSCC 上合作发表了六篇关于存内计算存储器 IP 的论文，大力推进基于 ReRAM 的存内计算方案；SK 海力士则发表了基于 GDDR 接口的 DRAM 存内计算研究。学术界和产业界普遍认为存算一体有望成为突破算力性能和功耗瓶颈的技术方向之一。特别是在大规模并行计算场景中，例如 VR/AR、无人驾驶、天文数据计算、遥感影像数据分析等，存算一体芯片具备高带宽、低功耗的显著优势。微观上，算力是一个具体的技术指标。算得快（高吞吐、低延迟）、算的准（高精度）、算的省（低成本、低功耗）是对算力的基本要求。存算一体是从微观层面进行架构的优化，面临存储器设计和生产工艺的挑战，需要整个产业链的参与支持。

实现存算一体的技术路径主要有以下三个：技术较成熟的是近存计算，利

用先进封装技术把计算逻辑芯片和存储器封装到一起，通过减少内存和处理单元的路径，以高 I/O 密度来实现高内存带宽以及较低的访问开销。近存计算主要通过 2.5D、3D 堆叠来实现，广泛应用在各类 CPU 和 GPU 上；近期投资热度较高的是存内计算，通过传统的存储介质如 DRAM、SRAM、NOR Flash、NAND Flash 来实现。计算操作由位于存储芯片/区域内部的独立计算单元完成，更适用于算法固定的场景；技术尚处于探索期的是基于非易失性存储器技术做的新型存储原件，比如通过忆阻器 ReRAM 电阻调制来实现数据存储。其他如相变存储器 (PCM)、自旋磁存储器 (MRAM) 等，也作为存算一体新的技术路径。存算一体的计算方式分为数字计算和模拟计算。数字计算主要以 SRAM 作为存储器件，具有高性能、高精度的优势，更适合大算力高效场景。模拟计算通常使用 FLASH、ReRAM 等非易失性介质作为存储器件，存储密度大，并行度高，更适合小算力，计算精度要求不高的场景。

目前，存算一体已经在产业细分领域掀起了创业浪潮，并受到投资界和产业界的关注和投入。存算一体在技术上向着高精度、高算力和高能效的方向发展。在资本和产业双轮驱动下，基于 SRAM、NOR Flash 等成熟存储器的存内计算将在垂直领域迎来规模化商用，小算力、低功耗场景有望优先迎来产品和生态的升级迭代，大算力通用计算场景或将进入技术产品化初期。基于非易失性、新型存储元件的存算一体依赖于工艺、良率的提升，走向成熟预计需要 5-10 年。

专家点评 |

存算一体一直是高效计算的重要技术之一。近年来，万物互联和人工智能的发展加速了存算一体的技术产品化进程，产业界对于存算一体最终的产品形态也在持续探索。未来存内计算产品将以单芯片和 Chiplet 两种形式共存。应用场景的多样性也将从物联网边缘端设备向大算力通用计算领域不断拓展，有望成为 AI 时代主流的计算架构。

李科奕

华夏芯（北京）通用处理器技术有限公司董事长

趋势四

云原生安全

安全技术与云紧密结合，打造平台化、智能化的新型安全体系。

摘要

云原生安全是安全理念从边界防御向纵深防御、从外挂模式向内生安全的转变，实现云基础设施的原生安全，并基于云原生技术提升安全的服务能力。安全技术与云计算由相对松散走向紧密结合，经过“容器化部署”、“微服务化转型”走向“无服务器化”的技术路线，实现安全服务的原生化、精细化、平台化和智能化：

- 以安全左移为原则，构建产品研发、

安全、运维一体化的产品安全体系，增进研发，安全和运维融合协同；

- 以统一的身份验证和配置管理为基础，实现精准授权和动态策略配置；
- 以纵深防御体系为架构，平台级的安全产品为依托，实现精准主动防御，化解传统安全产品碎片化的问题；
- 以安全运营为牵引，实现涵盖应用、云产品、网络等全链路的实时检测、精准响应、快速溯源和威胁狩猎。



趋势解读

随着云计算与各领域深度融合，云上快速迭代、弹性伸缩、海量数据处理等特征要求安全防护体系相应升级，为动态变化、复杂多元的运行环境提供有效的安全防护。

云原生安全是依托云原生理念和技术特性对安全体系进行的优化和重构，通过逐步实现安全技术服务的轻量化、敏捷化、精细化和智能化，来保障云基础设施的原生安全，并形成更强的安全能力。其基本特征包括全链路的DevSecOps安全管理，一体化精细化的身份与权限管控、平台化的纵深防御体系，以及实时化、综合化的可视、可管、可控体系。

云原生安全经历了一系列变迁：从安全保障云原生到云原生赋能安全，内涵不断扩展，逐步形成了一套涵盖基础设施、应用、数据、研发测试、安全运营等在内的防护体系。云原生应用保护平台、面向云原生的攻击面管理平台、云原生威胁检测与响应、云原生事件取证与溯源等，这一系列新型的防护措施也应运而生、快速发展，得到了业界共识。

从管理视角、运营视角和用户视角，可以看到云原生安全的三方面价值：

- 全链路风险可视可控。将安全和合规要求贯穿软件和服务全链路，及时扫描检查关键环节，避免后期处置造成被动，最大程度降低整体风险管控成本。
- 基础设施安全运营闭环高效。安全防护功能融合化，可以实现异常事件响应处置流程的闭环管理；策略

执行自动化，可减少安全运营人员的依赖，降低误操作概率；同时，自动阻断机制可以为应对攻击和修复争取更充分的时间。

- 云上客户资产全面保障。帮助客户全面、实时监测各类数据资产；在身份验证、配置管理、应用运行时监控、数据安全保护等方面提供多元化、灵活调用的安全服务。

实践中，云原生安全也面临着一系列挑战，比如在异构复杂环境中各类数字资产的监控数据如何快速、高质量的采集汇聚；云上各方如何明晰权责，形成开放协同的安全生态等等。

未来 3-5 年，云原生安全将更好的适应多云架构，帮助客户构建覆盖混合架构、全链路、动态精准的安全防护体系。同时配套构建起新型治理体系和专业人才体系；在安全防护效能方面，智能化技术为实现细粒度的访问鉴权、数据安全管控、风险自动识别和处置提供强有力的支持，保障用户顺畅高效使用，提供无摩擦的服务体验。同时，基于云的安全服务形式也将不断创新，云原生的安全托管，以攻促防等形式将逐步发展成熟，成为安全体系的重要组成部分。

专家点评 |

对于未来技术趋势的准确预测，是一件非常难的事情，但作为“随动技术”（reactive tech）的网络安全技术，却不见得特别难。达摩院十大趋势对于网络安全技术的研判经历了反复讨论和调整，从后量子时代密码技术变化，到可信隐私计算的下一步发展；从网络攻防技术的变化，最后落脚到了云原生安全，就是体现了网络安全技术的“后发先至”特点，就像密码技术需要提前应对‘后量子时代’，去回应现在看起来产业化还似乎遥遥无期的量子计算技术一样，对云原生安全的关注，也势必成为新的一年网络安全圈内的热点。

翟起滨

中国科学院信息安全国家重点实验室教授

云原生安全不是特指云原生技术的安全，而是包含云基础设施的原生安全和用云的原生能力形成更加强健、统一、智能的安全能力。

作为云服务提供商，要保证基础设施安全和云产品自身安全，根据服务模式与客户明确安全责任共担的边界。作为安全服务商，可以更充分的借助云的能力，为客户提供平台化、一体化的产品和服务。由于云原生安全的范围广，技术复杂，需要各方更加开放和协同，为云上的用户提供全面高效，可视可管可控的一体化安全服务保障。未来1-2年，可以重点关注云上身份安全和智能化的安全运营中心。

欧阳欣

阿里云首席风险官，阿里云安全产品事业部总经理

趋势五

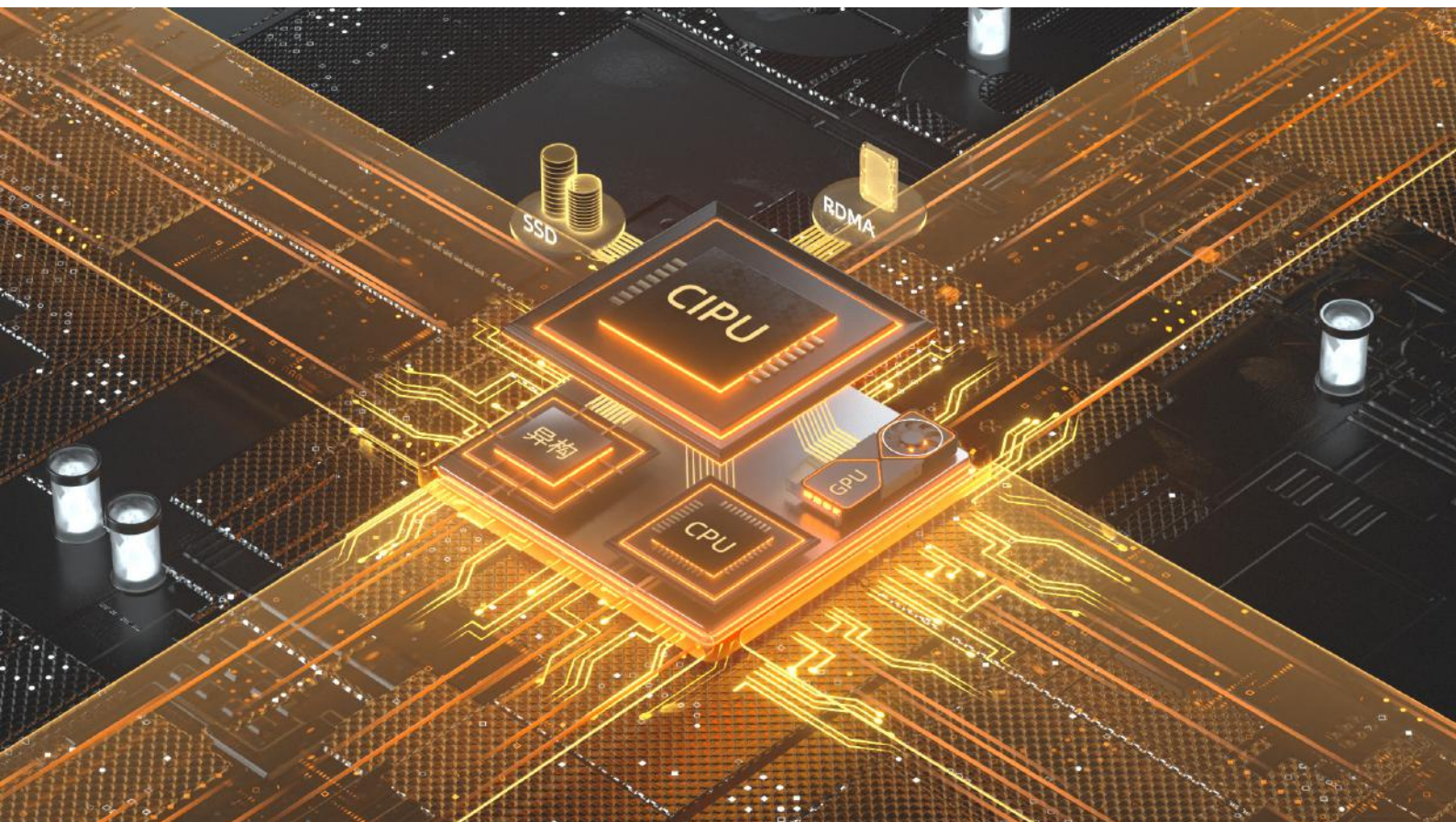
软硬融合云计算体系架构

云计算向以 CIPU 为中心的全新云计算体系架构深度演进，通过软件定义，硬件加速，在保持云上应用开发的高弹性和敏捷性同时，带来云上应用的全面加速。

摘要

云计算从以 CPU 为中心的计算体系架构向以云基础设施处理器 (CIPU) 为中心的全新体系架构深度演进。通过软件定义，硬件加速，在保持云上应用开发的高弹性和敏捷性同时，带来云上应用的全面加速。新的体系架构下，软硬一体化带来硬件结构的融合，接入物理的计算、存储、网络资源，通过硬件资源

的快速云化实现硬件加速。此外，新架构也带来软件系统的融合。这意味着以 CIPU 云化加速后的算力资源，可通过 CIPU 上的控制器接入分布式平台，实现云资源的灵活管理、调度和编排。在此基础上，CIPU 将定义下一代云计算的服务标准，给核心软件研发和专用芯片行业带来新的发展机遇。



趋势解读

随着后摩尔时代的带来，CPU 的性能提升趋近于天花板，数据量的爆发式增长带来极高的数据处理需求。企业在大数据、AI 等数据密集型计算的应用场景越来越多。这两方面的因素导致以 CPU 中心的云计算体系架构碰上了技术瓶颈，无法应对云上时延和带宽的进一步扩展。传统的 CPU 体系架构需要向着软硬一体化的方向迭代升级。

云计算的体系架构发展经历了三个阶段，已经解决了超高并发和大算力的经济性问题。第一阶段在 2010 年左右，以 X86 服务器、互联网中间件为代表的分布式架构技术为主。第二阶段在 2015 年左右，云厂商通过软件定义的方式，构建了虚拟私有云 (VPC) 和计算存储分离的池化架构。在新的池化架构之下，计算存储网络资源可以分别实现按需扩容。

目前，云计算进入第三阶段，引入专用硬件，形成软硬一体化的虚拟化架构，实现了全面硬件加速。这个阶段云计算面临的挑战，是在数据密集计算、云数据中心东西流量越来越大的趋势下，实现云计算单位成本下更高的计算性能，以及更高效的云数据中心管理。而计算效率的提升，还需要回到芯片和系统底层中去。

以 CIPU 为核心的云计算体系架构，在工程实现上主要有以下三方面突破：

首先，是底层硬件结构的融合，带来全面硬件加速。基于 CIPU 的新架构能够向下管理数据中心硬件，配合云操作系统，对计算、存储、网络、安全进行全面加速，把 IDC 真正变成一台高速

总线互联的超级计算机。在用户云上计算最需要的基础云网络和云存储链路上提供更低的延迟、更高的带宽和吞吐。

其次，在全链路实现硬件加速的基础上，创新地实现了 eRDMA，不但能够大规模组网，还能让用户无需修改负载的代码，无感加速，让云上的高性能计算普惠服务化成为现实。

最后，在全新的云数据中心硬件体系架构下，CIPU 和服务器的系统组合，既可以一对多，也能实现多对一，高效满足云上不同计算场景下东西向流量计算配比的灵活需求。

软硬融合的云计算体系架构，保持了软件定义，在分布式架构时期构建的交付敏捷性和灵活度，池化架构时期构建的弹性、可靠性、可用性，还带来了云上应用的全面加速，显著提升了计算性能。

未来三年，云计算向以 CIPU 为中心的全新云计算体系架构深度演进，云上的函数计算，容器服务，数据库服务，大数据服务，AI 等云服务，也将全面被 CIPU 加速。从购买计算资源进行自身应用加速，到全面使用云上服务，用户将获得 CIPU 加速带来的全面体验。从资源到服务，云计算服务的核心价值，很大程度将取决于云厂商能提供的底层计算能力和计算效率。

专家点评 |

软硬件一体化设计是当前计算架构的重要演进方向。尤其在复杂的云计算场景中，软硬件的协同优化与迭代升级更是决定其性能提升的关键。CIPU 是阿里云结合其飞天操作系统，实现对数据中心计算、存储、网络等多维资源进行高效管理和软硬件协同加速的代表性技术。CIPU 在阿里云的成功实践也预示着软硬件一体化的虚拟化架构将引领云计算的技术升级。

任炬

清华大学计算机系副教授

CIPU 有非常多的创新，比较值得一提的是其提供的 eRDMA 能力，通过全面普惠的大规模弹性 VPC 网络，对云上分布式 workload 做高性能网络加速，从缓存，数据库，大数据，AI 等，有 20% 到 80% 不等的 e2e 性能提升，且无需用户修改一行代码。

云计算体系架构基础技术的不断革新，正在推动云上基础计算能力开始大幅超越线下服务器，而企业只要上云就能从云计算资源或云服务中，低成本获得这些还在不断扩大的系统红利。

蒋林泉

阿里云研究员
阿里云神龙计算平台负责人

趋势六

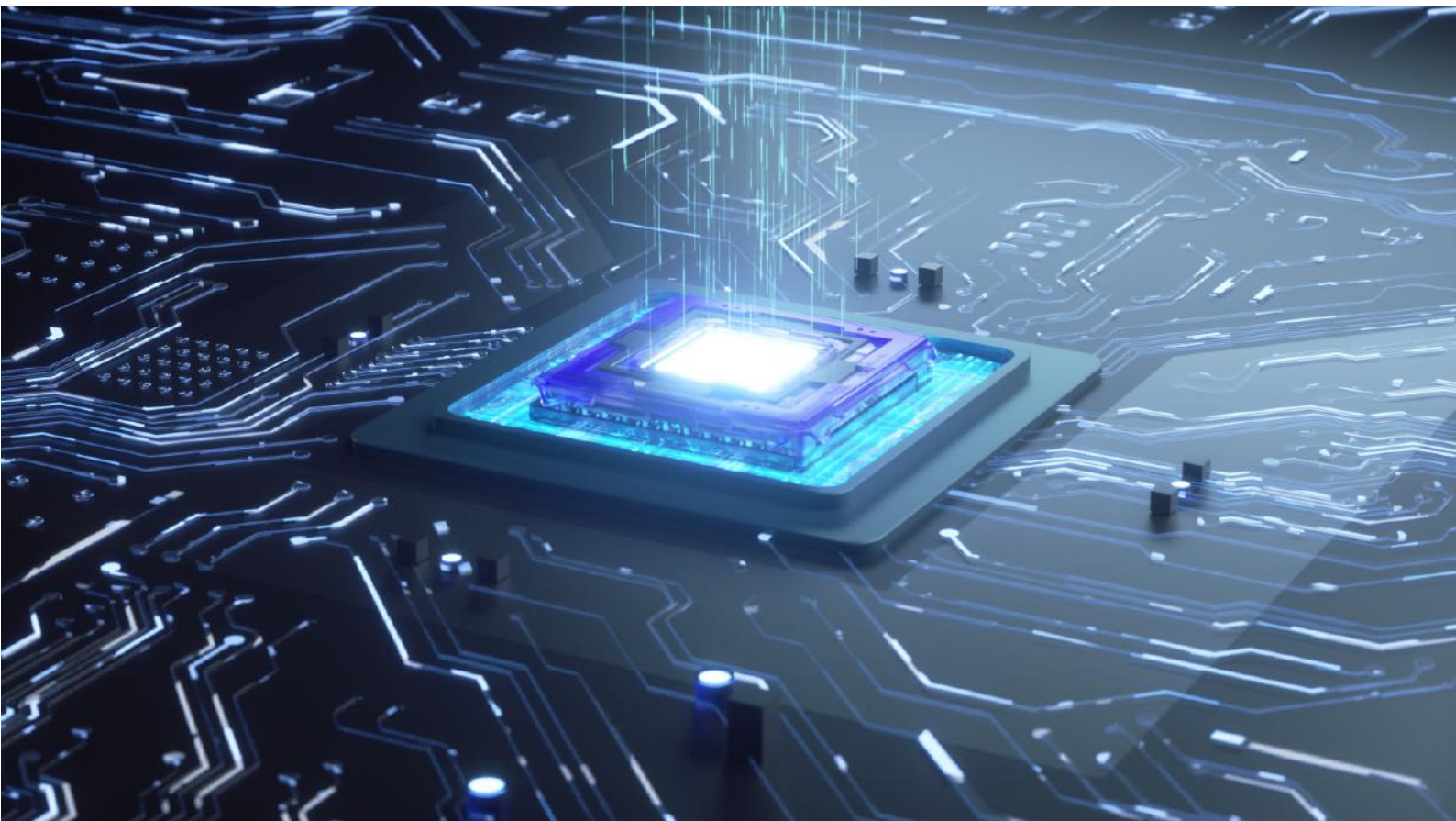
端网融合的可预期网络

基于云定义的可预期网络技术，即将从数据中心的局域应用走向全网推广。

摘要

可预期网络 (Predictable Fabric) 是由云计算定义，服务器端侧和网络协同的高性能网络互联系统。计算体系和网络体系正在相互融合，高性能网络互联使能算力集群的规模扩展，从而形成了大算力资源池，加速了算力普惠化，让算力走向大规模产业应用。可预期网络不仅支持新兴的大算力和高性能计算场

景，也适用于通用计算场景，是融合了传统网络和未来网络的产业趋势。通过云定义的协议、软件、芯片、硬件、架构、平台的全栈创新，可预期高算力网络有望颠覆目前基于传统互联网 TCP 协议的技术体系，成为下一代数据中心网络的基本特征，并从数据中心的局域应用走向全网推广。



趋势解读

可预期网络 (Predictable Fabric) 是由云计算定义，服务器端侧和网络协同的高性能网络互联系统。计算体系和网络体系正在相互融合，它通过高性能网络互联使能算力集群的规模扩展，从而形成大算力资源池，加速了算力普惠化，让算力走向大规模产业应用。可预期网络不仅支持新兴的大算力和高性能计算场景，也适用于通用计算场景，是融合了传统网络和未来网络的产业趋势。

网络的本质是连接。高带宽、低时延、高稳定性、少抖动一直是网络追求的目标。传统 TCP 网络协议栈，虽然在互联网中广泛部署和应用，但是 TCP 协议栈诞生时期的网络带宽和质量已经无法与如今大带宽高质量的数据中心网络相提并论。端侧和网络分层解耦，基于网络黑盒传统尽力而为 (best-effort) 的网络体系结构，无法满足当今大算力池化所需要的高性能网络互联需求。因此，“可预期的”高性能网络架构在大算力需求驱动下应运而生。这对于传统基于“尽力而为”的网络体系提出了新的挑战。

可预期网络以大算力为基本出发点，把端网融合作为实现方式。可预期网络摒弃传统端侧计算、存储和网络分层解

耦的架构，创新地采用端侧和网络侧协同设计和深度融合的思路，构建了基于端网融合的新型网络传输协议、拥塞控制算法、多路径智能化调度、以及芯片、硬件深度定制和卸载等技术的全新算网体系。可预期网络能够大幅度提升分布式并行计算的通信效率，从而构建高效的算力资源池，实现了云上大算力的弹性供给。云计算重新定义的可预期网络技术体系，将对产业链上下游、芯片技术演进产生深远影响，成为算力普惠化的新范式。

算力网络的发展正在经历从 0 到 1 的过程，需要互联网科技公司和运营商共同定义。如果将算力网络作为未来的关键基础设施，它将对网络可预期性提出更高的要求。数字化社会下的算力普惠，将持续驱动数据中心网络向高性能、资源池化的云计算方向发展，这将使网络可预期技术在未来 2-3 年内发生质变，逐渐成为主流技术趋势。

专家点评 | ④

过去十年，网络行业发生最大的变革是互联网公司在超大规模需求驱动下，重新定义了网络系统：以软件为中心的设备白盒替代了传统协议为中心的经典网络。未来十年，在大规模算力需求驱动下，云计算将重新定义下一代高性能网络系统，端网融合架构已经开始触发网络的生态变革，集计算和网络于一体的 DPU 芯片更是雨后春笋般出现，势将推动整个网络向可预期方向演进。

蔡德忠

阿里云智能副总裁兼首席网络科学家

趋势七

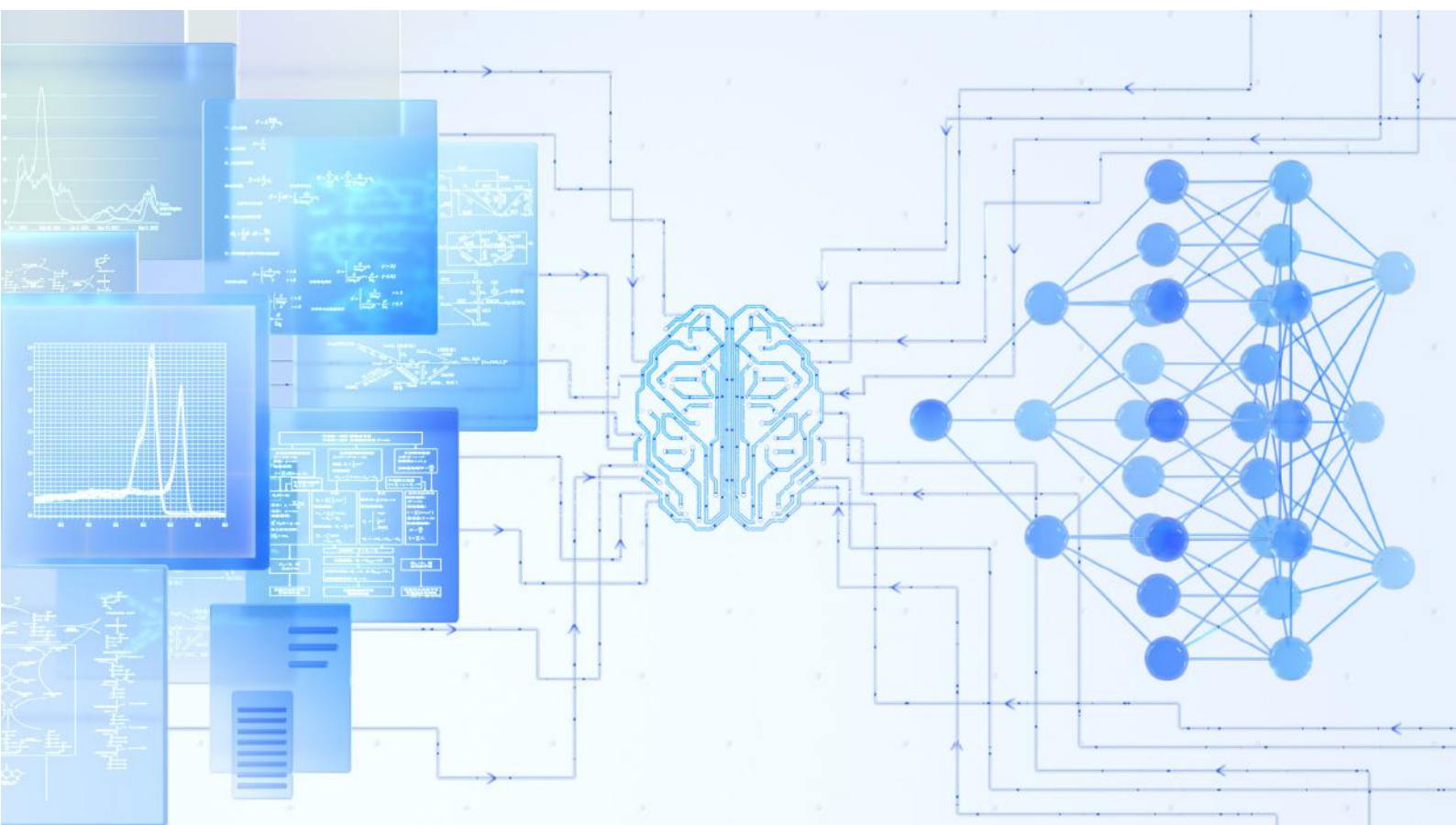
双引擎智能决策

融合运筹优化和机器学习的双引擎智能决策，将推进全局动态资源配置优化。

摘要

企业需在纷繁复杂、动态变化的环境中快速精准地做出经营决策。经典决策优化基于运筹学，通过对现实问题进行准确描述来构建数学模型，同时结合运筹优化算法，在多重约束条件下求目标函数最优解。随着外部环境复杂程度和变化速度不断加剧，经典决策优化对不确定性问题处理不够好、大规模求解响应速度不够快的局限性日益突显。学

术界和产业界开始探索引入机器学习，构建数学模型与数据模型双引擎新型智能决策体系，弥补彼此局限性、提升决策速度和质量。未来，双引擎智能决策将进一步拓展应用场景，在大规模实时电力调度、港口吞吐量优化、机场停机安排、制造工艺优化等特定领域推进全局实时动态资源配置优化。



趋势解读

近年来，全球性突发事件（如疫情、战争、技术管制等）频繁出现，使得外部环境变得更加复杂、不确定性更高；同时，市场不断变化、要求不断提升。企业需在纷繁复杂、动态变化的环境中，快速精准地做出经营决策。

智能决策是综合利用多种智能技术和工具，基于既定目标，对相关数据进行建模、分析并得到最优决策的过程。该过程将约束条件、策略、偏好、目标等因素转化为数学模型，并利用智能技术自动实现最优决策，旨在解决日益复杂、动态变化的经营决策问题（如打车平台派单、充电桩选址、生产排程等问题）。

经典决策优化基于运筹学，起源于二战中的空战规划。它通过对现实问题进行准确描述刻画来构建数学模型，同时结合运筹优化算法，在多重约束条件下求目标函数最优解。基于运筹学的决策优化对数据量的依赖性弱、求解质量较高、可解释性较强，被广泛运用于各类决策场景。

随着外部环境复杂程度和变化速度不断加剧，经典决策优化的局限性愈发突显，主要体现在：一是对于不确定性问题的处理能力不足，二是对大规模问题响应不够迅速。学术界和产业界开始探索引入机器学习，构建数学模型与数据模型双引擎新型智能决策体系。机器学习基于数据驱动模型，模拟出近似解区域，缩小经典方式求解空间，可大幅提升求解效率。机器学习的优势在于可应对不确定性高、在线响应速度快的场景；劣势为学习效率慢、成本高，且求解的质量不够高。由此可以看出，运筹优化和机器学习的结合完美弥补了彼此局限性，极大地提升了决策速度和质量。

双引擎智能决策尚处于起步阶段。众多决策优化场景（如交通领域港口吞吐量优化、机场停机安排等，制造领域工艺优化、产销协同等），开始尝试用双引擎方式在动态变化中快速找到最优解。最典型的、也最具挑战的场景是电力调度场景。电力调度场景转化为智能决策问题可描述为：

- 目标：在满足电网安全稳定运行前提下，降低购电成本或者实现全社会福利最大化，并促进新能源消纳。
- 约束条件：1) 必须满足所有安全约束，包括节点电压、线路与断面热稳定限额；2) 发用电负荷平衡约束；3) 满足物理特性，如机组爬坡、开停机曲线、梯级水电等。
- 决策难点：1) 调度业务非常复杂，涉及海量决策数据：目前省级变量与约束达千万级别；随着新能源快速装机以及引入负荷侧参与调节，直到实现双碳目标全网变量与约束预计将超过十亿级；2) 新能源发电占比将越来越大，其波动性和随机性将对模型驱动的数学优化效率带来极大挑战；3) 机器学习难以保证满足所有安全约束。

双引擎智能决策将机器学习与底层优化技术深度耦合在了一起，在满足各类安全约束条件的情况下，将计算效率提高10倍以上，有望实现秒级调度优化，突破新型电力系统电网调度追风、逐日决策的性能瓶颈。

未来，双引擎智能决策将进一步拓展应用场景，在特定领域实现更多主体、更大范围的资源配置优化，进而推进全局实时动态的资源配置优化。

专家点评 |

近年来，随着全球减碳活动的推广和新能源的发展，能源管理的系统复杂度迅速增加。光伏、风电、储能、微电网、充电桩、电动车在内的各种新能源设备增长迅猛，对发电侧、电网侧和用户侧的能源管理都提出了更高的挑战。传统电力管理需要面对如多能互补、削峰填谷、预测优化、柔性充放等各种新的运营需求。

新的能源智能决策系统将会整合“源网荷储”端的各种信息，对海量的能源数据进行快速分析、自动优化、和实时响应，实现全局和局部的能源资源配置优化，从传统粗放的管理逐渐向精细化和智能化转变。

未来的智慧能源管理将引领一次新的技术革命，带来各种全新的应用和商业机会，连接每个人的生活。双引擎智能决策系统将不断推进能源效率提升和优化，为实现国家的双碳目标奠定坚实的基础。

罗宇翔

普华永道可持续战略与运营合伙人

趋势八

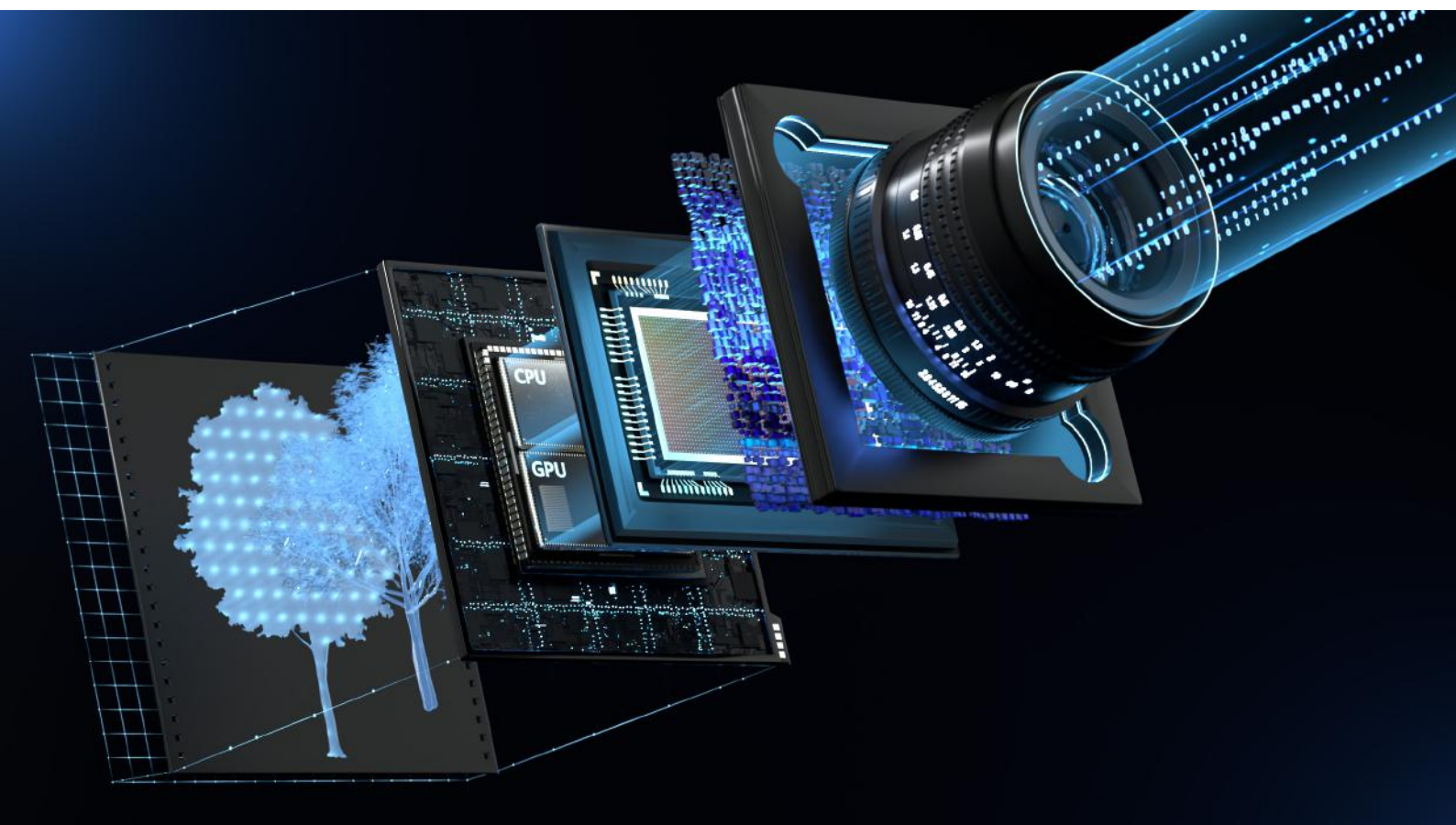
计算光学成像

计算光学成像突破传统光学成像极限，将带来更具创造力和想象力的应用。

摘要

计算光学成像是一个新兴多学科交叉领域。它以具体应用任务为准则，通过多维度获取或编码光场信息（如角度、偏振、相位等），为传感器设计远超人眼的感知新范式；同时，结合数学和信号处理知识，深度挖掘光场信息，突破传统光学成像极限。目前，计算光学成像处于高速发展阶段，已取得许多令人

振奋的研究成果，并在手机摄像、医疗、无人驾驶等领域开始规模化应用。未来，计算光学成像有望进一步颠覆传统成像体系，带来更具创造力和想象力的应用，如无透镜成像、非视域成像等。



趋势解读

传统光学成像建立在几何光学基础上，借鉴人眼视觉“所见即所得”的原理，而忽略了诸多光学高维信息。当前传统光学成像在硬件功能、成像性能方面接近物理极限，在众多领域已无法满足应用需求。例如，在手机摄影领域，无法在保证成像效果的同时缩小器件重量和体积，出现令人诟病的“前刘海”和“后浴霸”的情况；在显微成像领域，无法同时满足宽视场和高分辨率的需求；在监控遥感领域，难以在光线较暗、能见度较低的复杂环境中获得清晰图像……

期才逐步形成。随着信息技术的蓬勃发展，计算光学成像已成为国际研究热点。由于计算光学成像研究内容覆盖范围广，目前还没有一个比较明确的分类方法。按照计算成像技术所解决的应用问题来分类，可以大致分为以下三类：（1）功能提升：对传统方式无法获取的光学信息，如光场、偏振、相干度等进行成像或测量；（2）性能提升：即提升现有成像技术的性能指标，如空间分辨率、时间分辨率、景深、复杂环境鲁棒性等；（3）简化与智能化：通过单像素、无透镜等特定技术简化成像系统，或者以光速实

专家点评 | ④

在过去的十多年里，信息技术的高速发展为光学成像注入了新的生命，计算成像应运而生，悄无声息中颠覆了人类与机器感知世界的方式。从“所见即所得”的一一映射到对高维光场的耦合编码与计算重构，计算成像将光作为信息载体的一部分，模糊了物理世界与数字世界的边界，从而突破了物理约束，见所未见。从此，我们能够捕捉光传播的轨迹，看到千里之外的声音，解析生命活动的奥秘，穿云雾，洞察秋毫。从毫厘微末间的细胞病毒，到广袤宇宙中的第一缕光，计算成像将不断开拓人类的认知边界；从无人系统手机摄影，到工业监测安防监控，计算成像将融入人们生活的方方面面，推动数字经济高速发展。

吴嘉敏

清华大学自动化系
助理教授

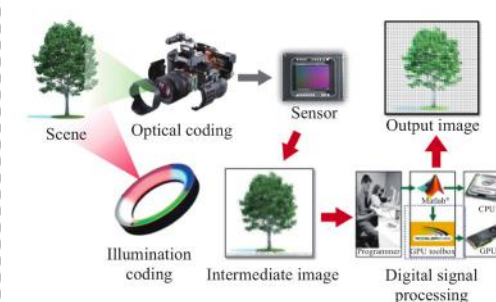
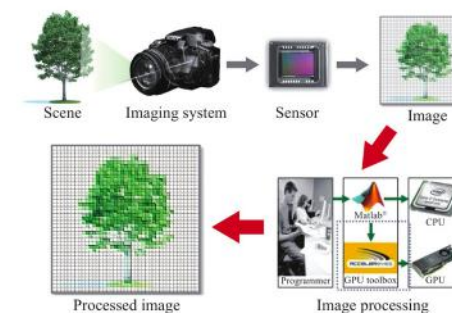


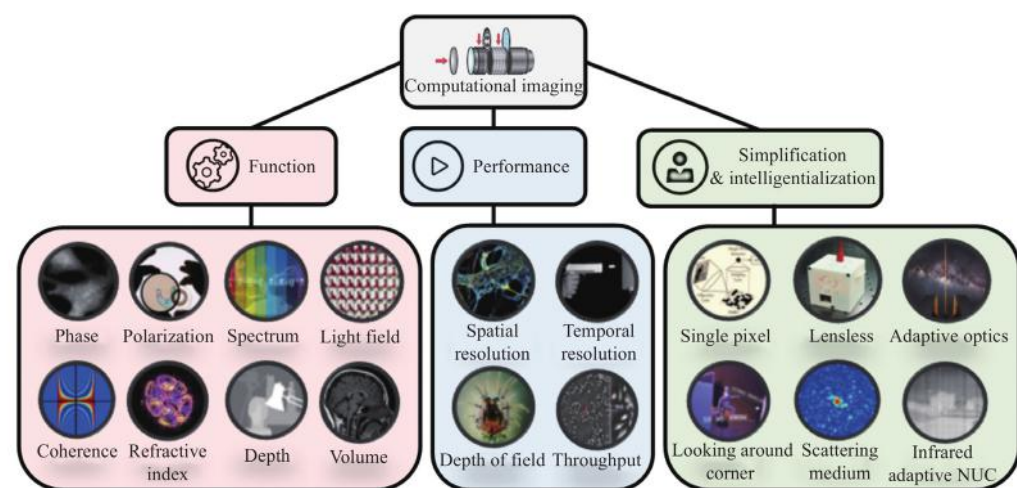
图1: 传统光学成像(左边)VS 计算光学成像(右边)^[22]

随着传感器、云计算、人工智能等新一代信息技术的不断演进，新型解决方案逐步浮出水面——计算光学成像。计算光学成像以具体应用任务为准则，通过多维度获取或编码光场信息（如角度、偏振、相位等），为传感器设计远超人眼的感知新范式；同时，结合数学和信号处理知识，深度挖掘光场信息，突破传统光学成像极限（如图1所示）。

计算光学成像是一个新兴多学科交叉领域，早期概念在上个世纪70年代中

现特定人工智能任务（如图2所示）。

计算光学成像技术现处于高速发展阶段，还需克服诸多挑战：首先，需以传感器为中心重新设计光学系统；其次，由于需要获取多维度光学信息，需引入新型光学器件和光场调控机制，随之而来的是更多的硬件成本和研发/调试时间成本；再次，为了使计算成像硬件和软件有更好的协同，则需重新开发算法工具；最后，对算力要求非常高，对应用设备芯片及其适配性提出更高要求。

图 2: 计算光学成像技术分类^[22]

计算光学成像虽然是一个新兴技术，但已取得了很多人振奋的研究成果（2014 诺贝尔奖——超分辨荧光显微成像、2017 年诺贝尔奖——冷冻电镜），并在手机摄像、医疗、监控、工业检测、无人驾驶等领域开始规模化应用。如在手机摄像领域，主流手机厂商均初步融入了计算光学成像思路，从比拼硬件光学，转而追求硬件加算法的协同；目前手机摄像在相当一部分场景的拍摄效果达到、甚至超过一般单反相机。

未来，计算光学成像将进一步颠覆传统成像体系，带来更具创造力和

想象力的应用。元成像芯片可实现大范围无像差三维感知，有望彻底解决手机后置摄像头突出的问题。无透镜成像（FlatCam）能够简化传统基于透镜的相机成像系统，进一步减小成像系统体积并有望用于各类可穿戴设备。此外，利用偏振成像技术能够透过可见度不高的介质清晰成像，实现穿云透雾。还有非视域成像，能够通过记录并解析光传播的高速过程来对非视域下目标进行有效探测，实现隔墙而视，在反恐侦察、医疗检测等领域具有广泛的应用价值。

趋势九

大规模城市数字孪生

城市数字孪生在大规模趋势基础上，继续向立体化、无人化、全局化方向演进。

摘要

城市数字孪生自 2017 年首度被提出以来，受到广泛推广和认可，成为城市精细化治理的新方法。近年来，城市数字孪生关键技术实现了从量到质的突破，具体体现在大规模方面，实现了大规模动态感知映射（更低建模成本）、大规模在线实时渲染（更短响应时间），以及大规模联合仿真推演（更高精确性）。目

前，大规模城市数字孪生已在交通治理、灾害防控、双碳管理等应用场景取得较大进展。未来城市数字孪生将在大规模趋势的基础上，继续向立体化、无人化、全局化方向演进。



趋势解读

专家点评 | ⚡

经过几年的探索，数字孪生城市发展进入了深水区。党的二十大提出“打造宜居、韧性、智慧城市”等要求，数字孪生城市成为中国式现代化在城市落地的重要手段。今年以来，从工业制造、交通运输、水利流域、能源生产到城市治理，从车间、道路、水域、电厂到楼宇、园区、城市，从产品模型设计、设备远程控制、人机虚实交互到算法仿真推演，数字孪生技术应用在各行各业不断渗透影响深远，大规模动态感知映射、在线实时渲染、以及联合仿真推演技术能力不断提升。然而深水区也暴露出一个关键的问题，就是任何传统引擎，都实现不了数字孪生，数字孪生城市必须打造一个新引擎。因此未来将以大规模“对象实体孪生+业务流程孪生”构建城市孪生体，实现“多源”数据融合重组，“多能”新引擎构建，“多跨”场景实现，城市数字孪生将在大规模孪生体构建基础上，真正实现从量到质的突破。

高艳丽

中国信息通信研究院规划所总工

城市数字孪生自 2017 年首度被提出以来，受到广泛推广和认可。

我们持续追踪城市数字孪生进展，分别在 2019 年及 2021 年的十大科技趋势报告中进行了详细阐述。近年来，中央部委多角度全方位出台数字孪生城市相关政策，旨在进一步推进城市精细化治理新方法。

城市数字孪生旨在以城市为对象，在数字世界建设与物理世界 1:1 的数字映射，进而通过数字映射进行多学科机理与仿真推演，并与物理世界进行实时双向同步。近两年，精准映射、生成渲染、仿真推演等城市数字孪生关键技术实现了从量到质的突破，具体体现在大规模上，分别实现了大规模动态感知映射、大规模在线实时渲染，以及大规模联合仿真推演。

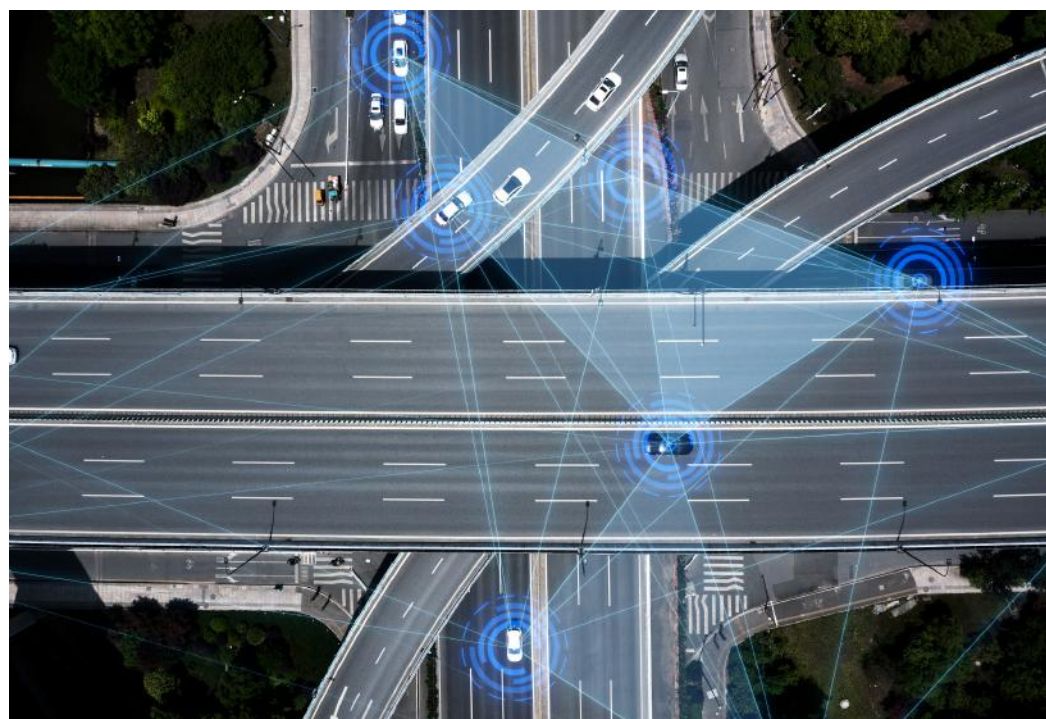


图 3: 城市交通分析

在精准映射方面，与耗人、耗时、耗财力的传统测量测绘方式不同，综合运用遥感、雷达、视觉、定位及其它多种传感器及存量测绘数据，以更低成本实现对城市静态部件，以及动态对象（如人、车等）进行位置、状态等多属性实时感知。未来，通过汇聚城市天空地多维度、各类型传感器数据，再结合 AI 感知能力，可融合提取同一实体的多源异构数据、构建多个实体之间的内在关系，在数字世界中进行大规模、低成本、统一、实时、精准的映射表达。

在生成渲染方面，基于精准映射的数据基础，结合 AIGC（AI Generated Content）与游戏大世界生成 PGC（Professional Generated Content）等技术能力，可实现分层次、分尺寸、多分辨率城市级三维场景模型的自动化生

成，以及支持多人在线、互动式大规模实时渲染。

在仿真推演方面，将多学科、大规模的机理与仿真模型联合作用于同一数字大世界，形成“仿真机理元宇宙”，构建虚实互动、双向调控机制。关键技术包括：1）仿真系统云原生化，基于云原生超算调度和求解器，可大幅缩短仿真计算时间，实现城市级场景、百万级以上实体规模实时计算响应；2）统一接口融合计算，多种机理模型及仿真模型能进行实时融合计算，形成多仿真联合服务能力。

在技术推动和需求牵引的共同作用下，大规模城市数字孪生已在交通治理、灾害防控、双碳管理等应用场景取得较大进展。如在交通治理方面，在城市高精路网、水网、河道、车辆等实体三维建模及实时渲染的基础上（建模成本降低 90%+、时间从数月缩短到数天），

通过联合道路交通流、城市内涝积水、自动驾驶、人群移动等仿真模型，实现对城市大型活动现场人群疏散引导、交通管控策略、天气情况影响、公共交通供给等全方位策略预案的孪生演练与效果评估（针对应急事态做到“1 分钟启动预案”、“5 分钟到达现场”；大型赛事“1 小时疏散”）。

基于数字孪生的智慧城市市场空间十分广阔。据 IDC 预测，到 2025 年智慧城市投资规模将超过千亿美元，5 年复合增长率超过 30%。目前，城市数字孪生面临的瓶颈，在于城市级大规模对象实体孪生以及业务流程孪生的城市孪生体尚未完全搭建起来。城市数字孪生将在大规模特征的基础上，继续向立体化、无人化、全局化方向演进。未来，城市数字孪生既作为城市立体化综合无人系统（无人车、无人机、机器人等）的研发测试环境，同时也是实现全局感知、全局调度的支撑系统。

专家点评 | ⚡

当城市的感知能力，以及以还原、建模、渲染、仿真推演为代表的计算能力发展到一定临界值，城市级别的大规模数字孪生成为可能。构建城市级别的数字孪生，将开启城市领域一个全新的计算范式。城市数字孪生世界不仅仅是城市的可视化渲染，更是城市的众多业务的新的载体，为业务创新提供了支撑。在这里，城市数字孪生不仅精准地捕捉到城市的当下，也全面记录着城市的历史。更重要的是，通过接入多领域的仿真能力，并支持基于云计算的云原生仿真改造，城市数字孪生可以进行大规模的联合仿真推演计算，在一个平台上，多领域多精度的仿真同时进行，并且相互博弈，进而可以从城市的过去推演城市未来发展的可能性。

曾震宇

阿里云智能副总裁、行业解决方案研发部负责人

趋势十

生成式 AI

生成式 AI 进入应用爆发期,将极大地推动数字化内容生产与创造。

摘要

生成式 AI (Generative AI 或 AIGC) 是利用现有文本、音频文件或图像创建新内容的技术。过去一年,其技术上的进展主要来自于三大领域:图像生成领域,以 DALL-E-2、Stable Diffusion 为代表的扩散模型 (Diffusion Model); 自然语言处理 (NLP) 领域基于 GPT-3.5 的 ChatGPT; 代码生成领域基于 Codex 的 Copilot。现阶段的生成式 AI 通常被用来生成产品原型或初稿,应用场景涵

盖图文创作、代码生成、游戏、广告、艺术平面设计等。未来,生成式 AI 将成为一项大众化的基础技术,极大的提高数字化内容的丰富度、创造性与生产效率,其应用边界也将随着技术的进步与成本的降低扩展到更多领域。



趋势解读

生成式 AI 使用各种机器学习算法,从数据中学习要素,使机器能够创建全新的数字视频、图像、文本、音频或代码等内容。它创建出的内容与训练数据保持相似,而非复制。它的发展得益于近年来大模型在基础研究尤其是深度学习上的突破,真实数据的积累和计算成本的下降。在过去的这一年,生成式 AI 将人工智能的价值聚焦到“创造”二字,标志着人工智能开始具备定义和呈现新事物的能力。

过去一年,生成式 AI 的进展主要体现在如下领域:

图像生成领域的进展来自扩散模型 (Diffusion model) 的应用,以 DALL-E-2、Stable Diffusion 为代表。扩散模型是一种从噪声中生成图像的深度学习技术。扩散模型技术的背后,是更精准理解人类语义的预训练模型、以及文本与图像统一表示模型 (CLIP) 的支撑。它的出现,让图像生成变得更具想象力。

自然语言处理 (NLP) 领域的进展来自于基于 GPT3.5 的 ChatGPT (Generative Pre-trained Transformer)。这是一种基于互联网可用数据训练的文本生成深度学习模型,用于问答、文本摘要生成、机器翻译、分类、代码生成和对话 AI。得益于文本和代码相结合的预训练大模型的发展,ChatGPT 引入了人工标注数据和强化学习 (RLHF) 来进行持续训练和优化。加入强化学习后,大模型能够理解人类的指令以及背后的含义,根据人类反馈来判断答案的质量,给出可解释的答案,并对于不合适的问题给出合理的回复,形成一个可迭代反馈的闭环。

代码生成领域的进展来自代码生成系统 AlphaCode 和 Copilot。2022 年 2 月,Deepmind 推出了他们的最新研究成果 AlphaCode。它是一个可以自主编程的系统,在 Codeforces 举办的编程竞赛中,超过了 47% 的人类工程师。这标志着 AI 代码生成系统,首次在编程竞赛中,达到了具有竞争力的水平。基于开源代码训练的 Copilot 开始商业化,作为订阅服务提供给开发者,用户可以通过使用 Copilot 自动补全代码。Copilot 作为一个基于大型语言模型的系统,尽管在多数情况下仍需要人工二次修正,但在简单、重复性的代码生成上,将帮助开发者提升工作效率,并给 IDE (集成开发环境) 行业带来重大影响。

随着内容创造的爆发式增长,如何做到内容在质量和语义上的可控,成为可控式生成,将是生成式 AI 面临的主要挑战。在产业化方面,降成本仍是关键挑战。只有像 ChatGPT 这样的大模型训练成本和推理成本足够低,才有可能规模化推广。此外,数据的安全可控、创作版权和信任问题也需要随着产业化加快逐一解决。

未来三年,生成式 AI 将步入技术产品化的快车道,在商业模式上会有更多探索,产业生态也会随着应用的普及逐步完善。届时,生成式 AI 的内容创造能力将达到人类水平。拥有数据、计算能力、产品化经验的大型科技公司将成为生成式 AI 落地的主要参与者。基于生成模型的计算基础设施和平台会逐步发展起来,模型变成随手可得的服务,客户不需要部署和运行生成式模型的专业技能就可以使用。生成模型将在交互能力、安全可信、认知智能上取得显著进展,以辅助人类完成各类创造性工作。

专家点评 |

生成式 AI 在 2022 年迎来了突破。不管是图片生成,代码生成还是开放域文本生成,在生成内容的质量,逻辑性和安全性方面都有明显的提升。基于 AI 生成技术的应用场景在今后几年会更多涌现。但是,安全可控有伦理责任的生成技术仍然需要重点研发,对于虚假生成内容造成的不良社会影响需要尤其关注。

黄非

达摩院语言技术实验室负责人



人工智能产业链联盟

星主： AI产业链盟主

 知识星球

微信扫描预览星球详情



参考文献

- [1] Gautam Kumar, et al. Swift: Delay is Simple and Effective for Congestion Control in the Datacenter. SIGCOMM 2020.
- [2] Wang, Shuai, et al. Predictable vFabric on Informative Data Plane. SIGCOMM 2022.
- [3] Gibson, Dan, et al. Aquila: A Unified, Low-latency Fabric for Datacenter Networks. NSDI 2022.
- [4] V. Olteanu, et al. An Edge-queued Datagram Service for all Datacenter Traffic. NSDI 2022.
- [5] H. Bao, L. Dong, S. Piao, F. Wei. BEiT: BERT Pre-Training of Image Transformers, Microsoft Research. <https://aka.ms/beit>. arXiv:2106.08254v2,2022.
- [6] X. Pan, T. Ye, D. Han, et al. Contrastive Language-Image Pre-Training with Knowledge Graphs. arXiv:2210.08901,2022.
- [7] C. Saharia, W. Chan, S. Saxena, et al. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. arXiv:2205.11487, 2022.
- [8] R. Zhang, B.Li, W. Zhang, et al. Collaboration of Pre-trained Models Makes Better Few-shot Learner. arXiv:2209.12255,2022.
- [9] W. Ma, M. Zhao, X. Xie, et al. Is Self-Attention Powerful to Learn Code Syntax and Semantics?. arXiv:2212.10017.2022.
- [10] G. Gao et al. Die to Wafer Hybrid Bonding for Chiplet and Heterogeneous Integration: Die Size Effects Evaluation-Small Die Applications. 2022 IEEE 72nd Electronic Components and Technology Conference (ECTC), 2022, pp. 1975-1981, doi: 10.1109/ECTC51906.2022.00310.
- [11] John H. Lau. Recent Advances and Trends in Advanced Packaging. IEEE Transactions on Components, Packaging and Manufacturing Technology (Volume: 12, Issue: 2, February 2022).
- [12] T. Tang, Y. Xi. Cost-Aware Exploration for Chiplet-Based Architecture with Advanced Packaging Technologies. arXiv:2206.07308.2022.
- [13] A. Mullen et al., Gartner Top Strategic Technology Trends for 2022 [EB/OL].2021. <https://www.gartner.com/en/newsroom/press-releases/2021-10-18-gartner-identifies-the-top-strategic-technology-trends-for-2022>.
- [14] Gartner. 中国云安全市场概览 [EB/OL].2022.
- [15] 中国信息通信研究院. 云原生架构安全白皮书. [R], 2021.
- [16] 何宝宏. 云与安全深度融合推动原生云安全发展. [J]. 中国信息安全, 2022.
- [17] 袁曙光. 云安全的未来是云原生安全. [J]. 中国信息安全, 2022.
- [18] 宋胜攀等. 零信任在云原生安全中的应用研究. [J]. 保密科学技术, 2021.
- [19] 世界经济论坛. 数字孪生城市: 框架与全球实践洞察力报告. [R], 2022.
- [20] 中国信息通信研究院, 中国互联网协会, 中国通信标准化协会. 数字孪生城市白皮书. [R], 2021.
- [21] IDC. 中国数字孪生城市市场分析, 2021.[R], 2022.
- [22] 左超, 陈钱. 计算光学成像: 何来, 何处, 何去, 何从? [J]. 红外与激光工程, 2022, 51(2).
- [23] 戴琼海, 赵建林, 司徒国海, 方璐. 计算光学成像专题. [J]. 光学学报, 2020, 40(1).

致谢

外部专家 (以姓氏拼音为序)

陈凯

香港科技大学教授

陈厅

电子科技大学计算机科学与工程学院教授

陈文光

清华大学教授、蚂蚁集团技术研究院院长

陈兴蜀

四川大学网络安全学院院长

陈勋

中国科学技术大学电子工程与信息科学系教授

陈宗周

《电脑报》创始人、《环球科学》杂志创始人

陈辉

中国未来研究会产业创新工作委员会委员

高艳丽

中国信息通信研究院规划所总工

江一舟

复旦大学附属肿瘤医院乳腺外科副主任医师、研究员

李科奕

华夏芯(北京)通用处理器技术有限公司董事长

李世鹏

IEEE fellow、国际欧亚科学院院士

刘益东

中国科学院自然科学史研究所研究员、博士生导师

罗宇翔

普华永道可持续战略与运营合伙人

乔思远

奇安信科技集团股份有限公司副总工程师

任炬

清华大学计算机系副教授

孙茂松

清华大学计算机系党委书记、教授

唐杰

IEEE fellow、清华大学计算机系教授、副系主任

滕皋军

中国科学院院士、东南大学附属中大医院院长

汪晓银

中国未来研究会大数据与数学模型专业委员会主任

王才有

原卫生部统计信息中心副主任

王海洋

象帝先计算技术有限公司副总裁

王军

中国信息安全测评中心原总工程师

王喜文

科技作家、北京华夏互联网智能技术研究院院长

魏小强

奇虎 360 天枢智库高级研究员

文继荣

中国人民大学信息学院院长、教授

武超则

中信建投证券研究所所长

吴飞

浙江大学计算机科学与技术学院教授

邬贺铨

中国工程院院士

吴嘉敏

清华大学自动化系助理教授

谢涛

北京大学计算机学院讲席教授

薛锋

微步在线党支部书记、创始人兼 CEO

徐明伟

清华大学计算机科学与技术系教授、网络技术研究所所长

徐文强

前瞻产业研究院院长

徐愈

中国信息化百人会执委

徐亨

SXR 科技智库创始人及理事长

杨辉

北京邮电大学电子工程学院教授

翟起滨

中国科学院信息安全国家重点实验室教授

邹萍

36 氩研究院院长

张力

中国现代国际关系研究院原副院长

张延川

中国通信学会副理事长兼秘书长

张义荣

奇虎 360 天枢智库总监、网络空间安全高级研究员

朱卫列

中国华能集团有限公司原首席信息师

左晓栋

中国科技大学公共事务学院、网络空间安全学院教授

编纂委员会

指导组

张建锋：阿里巴巴达摩院院长

周靖人：阿里巴巴达摩院副院长、阿里云智能首席技术官

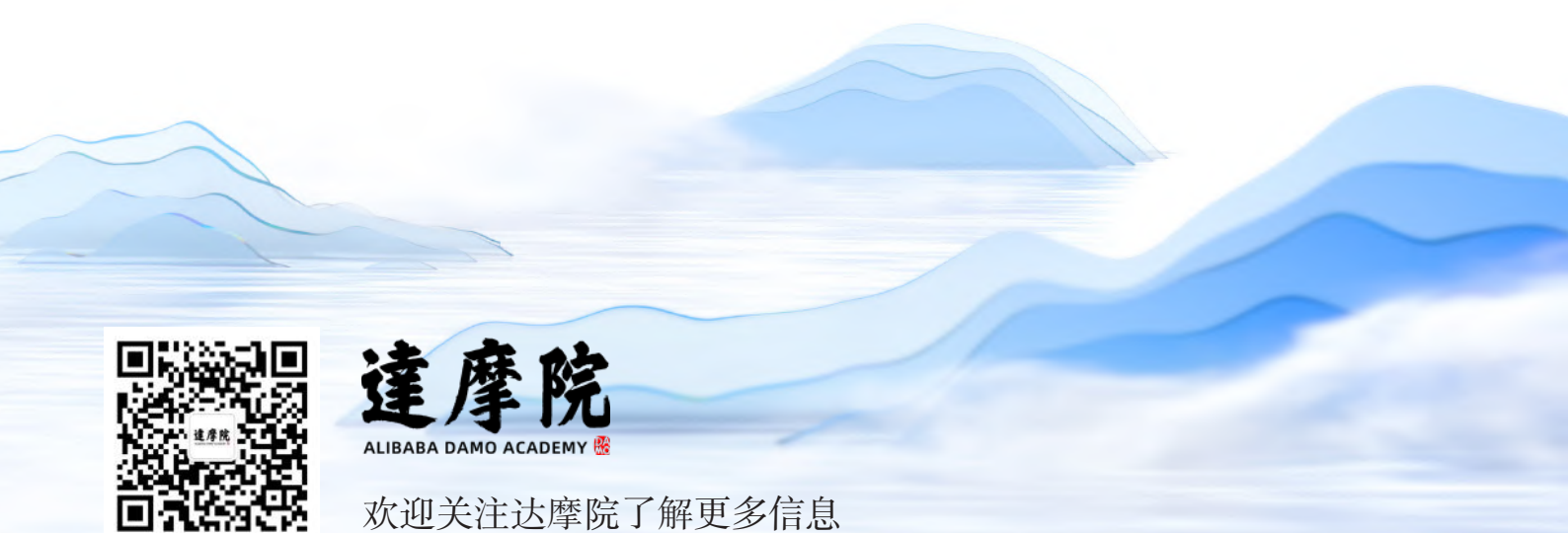
工作组

刘湘雯（组长）：阿里巴巴达摩院秘书长，阿里云智能副总裁

段佳惠、翟恩南、李俊平、李贝、尹泊明、余建平、朱逸菲、范博佳、李博、薛达、王婧璇、刘洋、杨柳林、朱迅垚、钱蓓、陈振华、程弢、肖成、刘湘琳、刘书好、肖剑、谢焯敏、麻芑、秦钊、唐佩、余婧、任妍、刘建强、李亚英

翻译组

邴孝娟、Eric Naik、魏之婷、张宁、林蓉、张紫琦、孙峭、单依依、张彩霞、王娇、樊敏、张佳杰、何婷、陈洁



达摩院
ALIBABA DAMO ACADEMY

欢迎关注达摩院了解更多信息

本内容未经阿里巴巴达摩院（杭州）科技有限公司书面授权认可，任何人不得复制、修改、转载、摘编或以其它任何方式使用本报告的全部或部分内容。