

云计算蓝皮书

(2025 年)

中国信息通信研究院

2025年7月

版权声明

本蓝皮书版权属于中国信息通信研究院,并受法律保护。
转载、摘编或利用其它方式使用本蓝皮书文字或者观点的,
应注明“来源:中国信息通信研究院”。违反上述声明者,
本院将追究其相关法律责任。



前 言

云计算作为人工智能时代的关键基础设施，是人工智能深度赋能实体行业的必要基座，也是加快形成新质生产力，推动我国数字经济走向新发展阶段的重要助力。过去一年，以 DeepSeek 为代表的人工智能大模型取得突破性进展，推动“人工智能+”消费级应用拉开发展序幕，普惠易用、一体化调度的智能计算能力促使云计算焕发新质，加速推进全球云战略布局与产业智能化发展，并呈现以下新特点：

一是各国加速推进云计算战略，夯实人工智能时代核心竞争力。

美国积极推进“云智能”战略，大力促进云计算与 AI 技术深度融合。欧盟更新战略部署，推动云计算基础设施建设，以加速提升国家技术竞争力。亚非拉国家持续加大云计算领域投资，以提升云技术和应用水平，缩小与先进国家间的云计算发展差距。中国持续强化政策引领，推动云计算技术融合发展与行业应用深化，助力产业智能化转型升级。

二是消费级 AI 应用激活全球云计算市场，我国云计算产业韧性显著。2024 年，全球云计算市场规模为 6929 亿美元，同比增长 20.2%，未来几年，人工智能向全社会的深度渗透将为云计算产业带来持续需求，预计到 2030 年全球云计算市场规模将接近 2 万亿美元。2024 年，我国云计算市场规模达 8288 亿元，同比增长 34.4%，保持较高增速，以云为基座的数智市场格局已经形成。随着量子计算、区块

链、人工智能与云计算的融合革新，云计算的边界将进一步扩展，预计到 2030 年我国云计算市场规模将突破 3 万亿元。

三是云计算驱动 AI 服务范式创新，激发人工智能时代新质生产力。随着人工智能技术的不断革新，人工智能开始走入千行百业，从工作场景走入生活场景，社会对算力的普惠化、场景化、生态化需求愈发凸显。针对人工智能时代的新需求，云计算服务模式正在加速向人工智能+转化，呈现出 AIaaS、AIPaaS、MaaS、AISaaS、AIMSP 等全产业链进化趋势，协同推进 AI 技术从基础设施到商业价值的全链路转化，形成人工智能时代的新质生产力范式。

中国信通院自 2012 年起持续跟踪云计算发展态势，本蓝皮书是中国信息通信研究院自《云计算白皮书（2012 年）》后第 11 次发布云计算相关研究成果。本蓝皮书立足产业新发展、新变化、新需求，聚焦人工智能时代下全球云计算发展重点，以独创性分析框架深入阐释云计算政策环境、市场发展及技术热点，以前瞻性视角研判云智融合、全栈创新等未来趋势，为政府决策、产业升级和技术突破提供参考。

目 录

一、全球云计算发展概述.....	1
（一）各国加速推进云计算战略，构建人工智能时代云计算竞争力.....	1
（二）全球云计算市场稳定增长，人工智能引发新一轮需求激增.....	3
（三）云计算技术持续深度融合，成为智能化创新发展的引擎.....	6
二、我国云计算发展概述.....	14
（一）中央与地方协同发展云计算，助力产业智能化转型升级.....	14
（二）我国云计算市场保持高速增长，AI 云成为厂商破局关键.....	16
（三）“云+人工智能”双轮驱动转型新范式，行业智能化应用加速普及...18	
三、云计算驱动 AI 服务范式创新，激发人工智能时代新质生产力.....	24
（一）AIaaS 革新智能基础设施架构，助力 AI 模型实现高效训练推理....	26
（二）AIPaaS 重塑 AI 开发服务新范式，加速 AI+应用快速规模化落地.....	29
（三）MaaS 提供全流程服务能力，成为大模型工程化落地重要载体.....	33
（四）AISaaS 驱动企业级应用智能化，大幅提升业务场景运行效率.....	36
（五）AIMSP 聚焦 AI 云专业技术服务，打通端到端大模型应用交付体系	40
四、发展展望.....	44

图 目 录

图 1 全球云计算市场规模及增速（亿美元）4

图 2 2024 年全球各区域云计算市场规模占比.....5

图 3 2024 年全球主要厂商云计算业务营收（亿美元）6

图 4 中国云计算市场规模及增速（亿元） 16

图 5 中国公有云细分领域市场规模及增速（亿元） 17

图 6 2024 年中国公有云 IaaS 厂商市场占比..... 18

图 7 智算云体系架构图.....25

一、全球云计算发展概述

（一）各国加速推进云计算战略，构建人工智能时代云计算竞争力

美国积极推进“云智能”战略，促进云计算与 AI 技术深度融合，强化云计算技术应用过程中的安全保障体系。在布局云计算和人工智能融合方面，美国积极行动，致力于构建先发优势，从而持续引领全球技术发展潮流。自 2018 年将“云优先（Cloud First）”政策升级为“云智能（Cloud Smart）”战略以来，美国在云基础设施建设上的投入已累计超 1000 亿美元。2025 年 1 月，特朗普宣布启动“星际之门”项目，该项目计划联合 OpenAI、软银、甲骨文等科技巨头，持续加大对下一代 AI 所需的物理与虚拟基础设施建设的投资力度，其中涵盖数据中心以及云计算平台等关键领域。在保障云应用安全方面，美国更新发展路线图，推动各机构更加安全、高效地运用云技术。2024 年 5 月，美国总务管理局对“联邦风险和授权管理计划（FedRAMP）”的 2024-2025 年路线图进行了细化，明确强调各部门在使用云技术时，对关键信息的安全防御举措，以及云端迁移过程中的安全保障工作。特别强调要建立自动化流程，用于关键信息的输入、运用以及安全评估和审查结果的复用，以此减轻参与者的负担，加快云解决方案的实施进程。美国政府希望通过采用自动化流程和安全云服务，以释放资源赋能创新，从而高效提供公共数字服务。

欧盟更新战略部署，推动云计算基础设施建设，以加速构建国

家技术竞争力。一方面，欧盟长期以来视云计算为国家科技创新和数字经济时代的重要支撑。《2024 年数字欧洲工作计划》继续追加欧盟在数字领域的资金投入，为包括人工智能、云计算以及网络安全在内的数字解决方案提供了 7.627 亿欧元的资金。其中，重点强调要投入高性能计算、通信基础设施，安全运营中心网络等云计算基础设施的建设。**另一方面**，欧盟还关注数字技术发展及相关人才培养。2024 年 3 月，欧盟委员会通过“地平线欧洲”的第二个战略计划，即《“地平线欧洲” 2025-2027 年战略计划》。该计划延续了“地平线欧洲”框架计划中推动科研创新的核心目标，将数字化转型定为开展研究与创新资助的战略方向之一，大力发展云计算、人工智能、半导体等新兴技术，旨在进一步明确未来三年在关键领域的研发方向，加强欧洲的技术主权和竞争力。

其他地区国家推动各行业深化用云，凭借深度云化赋能业务实现提质发展。亚太地区国家继续提升云计算技术和应用水平。日本大力推动云计算技术在政务领域广泛应用，以此提高政府服务效率。2024 年 7 月，日本总务省发布的当年《信息通信相关现状报告》，明确云服务在社会经济活动中的重要性，在公共部门积极推广云计算应用，以提升政府服务的效率和质量。韩国多年来持续加大在云计算领域的投入，计划通过精进云技术抢占智能化发展先机。2024 年便投资 1219 亿韩元支持本国云计算产业发展，旨在培育跨行业和跨社会的云技术，推进云计算与人工智能的发展。**东南亚地区积极推**

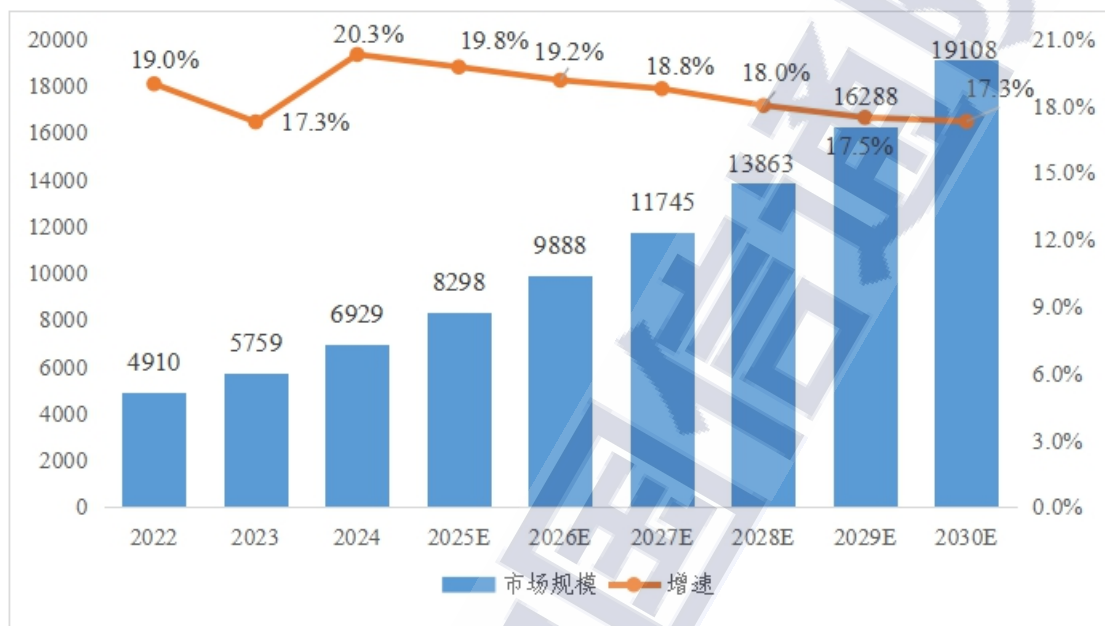
进以云计算为基础的数字化转型。泰国在 2024 年提出“云优先”（“Cloud First”）政策，并将其作为泰国数字化转型的核心目标，鼓励政府及各行业加速上云，以促进泰国经济转型升级，实现数字泰国的目标。中东地区国家推动云计算领域的创新和开放。2024 年 2 月，卡塔尔政府发布了《数字议程 2030》（Digital Agenda 2030），旨在推动卡塔尔电子政务的发展，促使政府各部门积极利用云计算等信息技术，打破信息孤岛，实现跨部门、跨领域的协同服务。2024 年 9 月，土耳其发布《2025-2027 年中期计划》，提出增加 AI 研究中心数量，并与国际共享计算基础设施，推动云计算在技术协作和供应链协同方面发挥更重要的作用。拉美地区国家加大云计算领域投资，努力缩小与云计算技术先进国家间的发展差距。2024 年 9 月 25 日，微软宣布未来三年内将投资 13 亿美元在墨西哥建设云计算和人工智能基础设施，旨在改善连通性，促进中小企业采用人工智能技术。12 月 5 日，智利政府发布《2024-2030 年国家数据中心战略规划》，明确将数据中心和云计算作为发展重点，促进数据中心投资，提出建立共享的国家多云服务以支持数字政府转型。

（二）全球云计算市场稳定增长，人工智能引发新一轮需求激增

从整体来看，全球云计算市场保持平稳增长。Gartner 数据显示¹，2024 年以 IaaS、PaaS、SaaS 为代表的全球云计算市场规模达 6929

¹ 《Forecast: Public Cloud Services, Worldwide, 2022-2028, 3Q24 Update》Gartner 2025 年 3 月。

亿美元，同比增速 20.3%。未来几年，随着 AI 模型训练对 IaaS 消费的拉动，以及应用类模型服务在 SaaS 侧不断完善，对云计算的需求持续增加，预计到 2030 年全球云计算市场规模将接近 2 万亿美元。

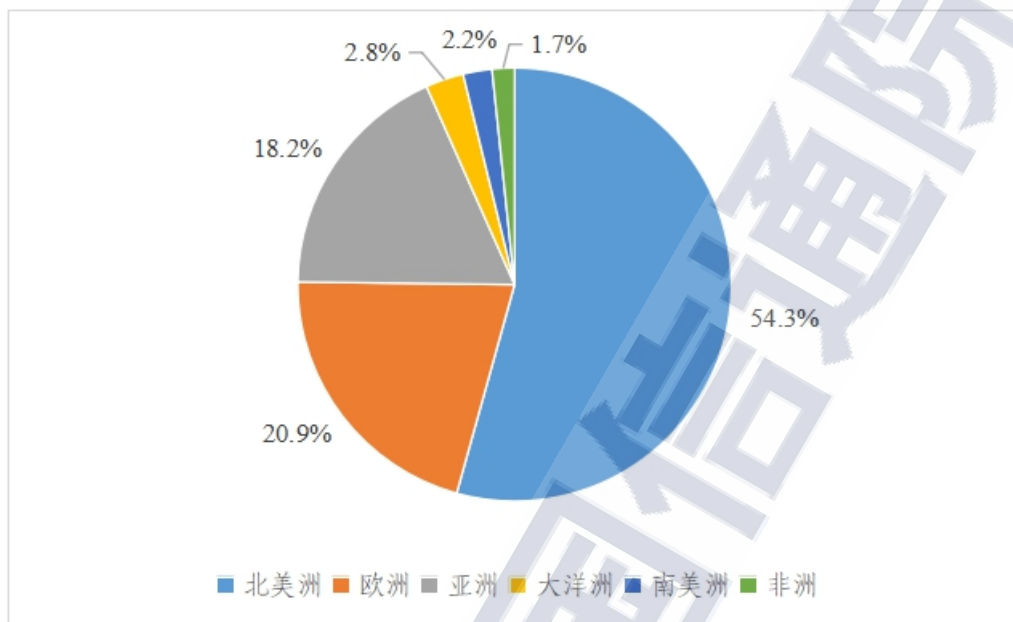


来源：Gartner，2025 年 3 月

图 1 全球云计算市场规模及增速（亿美元）

从区域层面来看，各大洲市场份额维持现有格局，拉美和中东有望成为新热点区域。2024 年，北美洲以 54.3% 的市场份额占据主导地位；欧洲和亚洲依托国家政策驱动、成熟云市场环境、AI 技术发展分别占据 20.9% 和 18.2% 的市场规模；大洋洲和南美洲分别受高成熟度和高价格敏感度限制，整体占比 5%；非洲市场受北非地区云计算发展影响，市场稳中有进，占比达 1.7%。拉美和中东地区则在 AI 算力需求、地缘政治和国际化布局的多维驱动下持续释放潜力，市场规模增长均超过 20%。随着多家头部云服务商在墨西哥、阿联

酋等地“开服建厂”，以及沙特“2030 愿景”和阿联酋“数据本地化”等利好政策的推动，区域增长动力强劲。

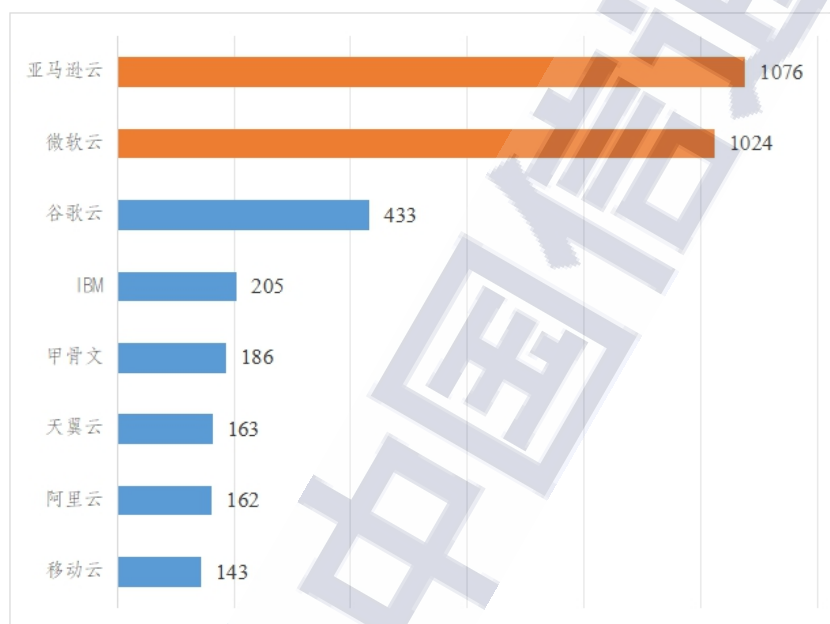


来源：Gartner，2025 年 3 月

图 2 2024 年全球各区域云计算市场规模占比

从厂商层面来看，第一梯队突破千亿美元里程碑持续领跑，第二梯队加码 AI 云投入保持快速追赶。2024 年，亚马逊云、微软云依靠 AI 转型策略，通过“AI 研发投资+合作”的 Cloud for AI 深度融合战略分别以 1076 亿美元和 1024 亿美元营收稳居头部地位，但随着加速发展，原有非 AI 业务转型和底层硬件基础设施瓶颈显露，长期增长面临考验。第二梯队厂商则纷纷加大 AI 投入，沿着头部路径加快追赶。谷歌云、阿里云等互联网原生企业聚焦全栈式 AI 云服务战略，全面加码 AI 工具及产品研发。IBM、甲骨文等传统软件厂商主攻优势业务 AI 转型，将混合云 AI 服务、AI 赋能企业级云数据

库作为核心战略。天翼云、移动云等电信运营商倾向从 AI 算力支撑切入，积极布局算力基础设施。当前，第一梯队与第二梯队厂商均以人工智能技术为核心驱动力布局未来增长赛道，差异化发展路径已形成。但随着 AI 云服务在商业场景中的规模化应用突破，叠加创新型服务生态的持续构建，或将成为弥合梯队间资源差距的关键杠杆。



来源：公开资料整理

图 3 2024 年全球主要厂商云计算业务营收（亿美元）

（三）云计算技术持续深度融合，成为智能化创新发展的引擎

经过多年发展，云计算技术不断迭代升级，不仅构筑起支撑各类智能新技术创新发展的关键基石，更是凭借强大的算力与广泛的兼容性，与众多领域深度交融，催生出一系列全新的技术发展方向，极大地拓展了智能应用的边界，重塑各行业智能化转型的格局。

基础设施层面，一云多算聚焦“云+人工智能”协同发展的需求，实现一个云平台对多种算力资源高效协同调度、管理和数据打通。随着人工智能技术的发展，企业催生出多元化的业务场景，多种算力难协同、数据分散难治理等技术难点日益凸显，在此背景下一云多算应运而生。它是指通过云的弹性按需的能力，凭借集约化的服务模式 and 运营体系，以“一朵云”的方式将智能算力、高性能算力、通用算力等多种资源池融合，提供统一调度、管理、运维运营等操作的技术能力。当前，一云多算技术呈现出以下几个特点：一是资源方面，多算融合技术灵活弹性满足不同业务场景下的算力需求，促进数据共享和流通。智能算力、高性能算力、通用算力计算任务需要共享整个数据的生产过程，涵盖从生产、处理、训练到推理的各个环节。同时，不同的计算任务以及不同的运行时态能够灵活地进行云网络和高性能网络切断，以此确保安全性以及网络的互通性。例如，亚马逊云的高性能计算平台能够提供多元异构算力支持，满足客户不同算力的使用需求。二是调度方面，协同调度技术打破了单类型的算力性能瓶颈，支撑复杂业务场景下的多任务并发处理能力。企业在进行数据生产、加工以及训练和推理等工作过程中，会涉及不同的任务类型以及与之相适应的不同算力形态，例如 CPU（Central Processing Unit，中央处理器）、GPU（Graphics Processing Unit，图形处理器）等。当前，不同的算力呈现出“烟囱型”状态，在进行调度时需要协同调度技术调度不同的算力形态。如阿里云、

腾讯、百度云、天翼云、火山引擎等均实现在一张网络内统一调度多种算力。三是管理方面，一云统管技术从用户视角出发，提升多算场景管理和运营效率。在平台运维方面，一云统管技术引入了多算运维工具，针对大模型训练、推理等需求进行性能优化和报错诊断，提升了系统的稳定性和可靠性。在平台运营方面，一云统管技术通过智能工单服务解决了用户找人难、排查难的问题，并以用户视角进行运营流程设计、经营分析和计量计费模块，帮助用户更精细化地管理多算场景，提升整体运营效率。例如，谷歌云的 Anthos 平台集成了丰富的监控、日志记录和诊断工具，帮助企业实现智能化、自动化的运维操作。

运行安全层面，云+应用运行安全备受关注，面向业务构建监管控一体化的运行安全保障能力渐成趋势。近年来，随着企业云化转型的持续深化，云+应用运行的稳定性愈发重要，确保业务连续性成为企业发展的关键因素。为保护业务稳定安全运行，需要实现从应用、云平台到基础设施的全面技术贯通和责任共担，建立监控共享、统一管理、同步处置的云+应用运行安全保障能力。一是业务侧全景监控能力方面，云+应用运行安全强调业务全链路监控和分析能力建设。在故障发现环节，通过建立统一监控平台，结合全链路追踪、智能化观测等技术，实现云服务、应用服务与基础设施监控数据的实时采集、关联分析和可视化展示，打破层级间的信息孤岛；在故障定位环节，通过对全链路数据进行全面收集并深入分析，结合实

时日志聚合、性能指标关联分析和智能算法等技术，实现故障快速定位。例如，腾讯云可观测平台基于指标、链路、日志、事件等全类型监控数据，提供了一体化监控解决方案；亚马逊云 CloudWatch 与 X-Ray 的组合实现了业务侧全景监控与故障的快速定位。二是全生命周期管理和运维方面，云+应用运行安全强调从业务逻辑态、部署态、运行态三个维度开展统一规划与管理。在设计环节，着重强调从设计阶段便将应用、云服务和基础设施相融合，进行框架搭建以及运行环境规划；在部署过程中，促使应用、云服务与基础设施的部署及测试同步推进；在运行时期，构建集监控、管理、处置于一体的运行安全保障体系。例如，阿里云推出了应用运维平台面向应用提供“云+应用”全链路管理能力。三是统一事件处置能力方面，云+应用运行安全强调构建包含方案选择、事件处置、工具建设的全链路事件处置体系。在处置方案选择阶段，通过综合考量各层次解决方案的效果、处理速度、潜在风险等多方面因素，选取最适配的方案，缩短事件对业务的影响时间，缩小影响范围；在事件处置阶段，通过构建应用、云服务与基础设施事件的统一调度与处理机制，整合应用、云服务与基础设施事件源，结合自动化 workflow，实现跨层级事件的快速响应与闭环处理；在工具建设阶段，构建穿透式运维平台，集成各层级服务的 API 接口，实现应用和云服务同步处置，避免处置不及时、数据不一致、操作步骤繁琐等问题。

优化治理层面，云卓越架构助力企业，围绕上云应用场景搭建

适配的云上架构，充分释放云计算的业务价值。Flexera《2024 年云现状调查报告》显示，75%以上的企业在云成本管理、资源治理、安全合规等方面面临挑战。云卓越架构是云服务商基于企业上云用云经验形成的最佳实践，聚焦安全、稳定、运营、成本、性能五大维度，保障企业上云用云业务价值充分释放。当前，云卓越架构发展呈现以下特点：

一是技术融合方面，云卓越架构融合多种不同领域的关键技术筑牢企业用云基线。云卓越架构采用先进的加密、身份认证和访问控制技术，保障数据和应用的安全；运用负载均衡技术、冗余备份技术以确保系统持续稳定运行；借助容器、微服务等原生技术达成资源弹性扩展及高效运营；凭借资源动态调配技术、成本分析技术达成成本优化；通过缓存技术、分布式计算技术提升系统性能。例如，华为云在云卓越架构中融合了先进的安全防护技术，保障企业数据安全。

二是智能应用方面，云卓越架构结合智能化技术促进用云效益提升。云卓越架构利用 AI 算法分析大量数据来识别异常与潜在威胁；结合 AI 技术对故障数据进行分析，快速精准诊断故障原因；借助自动化运维技术、智能监控技术提升运营效率；运用 AI 技术实时监控、分析云支出状况，并提出优化建议；运用 AI 算法实现资源自动调配以实现性能提升。例如，亚马逊云卓越架构利用 Cost Explorer 工具深度分析历史消费数据，为用户生成费用预测模型与优化建议。

三是场景落地方面，云卓越架构通过不断迭代优化在不同业务场景形成最佳实践。云卓越架构从安全、稳定、

运营、成本、性能五个维度出发，通过工具和度量模型评估场景落地方案与预期的差距，并提供改进指导建议，不断迭代生成各业务场景的最佳实践，助力企业发挥云计算效益。例如，阿里云卓越架构建立安全责任、成本优化等模型对企业上云后的方案进行评估，指导企业优化上云资源利用效益。

行业应用层面，行业云平台正与 AI 深度融合，如同行业智能应用的强力孵化器，加速向 AI 行业云的进阶蜕变。当前，传统行业智能化升级仍面临设备及系统间数据未打通、场景复杂未形成规模复制推广等挑战。行业云平台作为破局关键枢纽，向下提供行业算力服务和统一开发，向上沉淀垂直领域业务知识，能够破解 AI 技术与传统行业业务场景深度融合的壁垒，加速 AI+行业发展，释放新旧动能转换的乘数效应。当前，行业云平台在产品升级、生态拓展、持续运营等方面呈现出新变化。**一是产品服务方面**，行业云平台由提供垂直行业云服务转向提供行业模型与智能应用。建设方发挥自身业务经验优势，依托先进的基础设施，构建面向垂直行业的算力服务平台，整合数据资源，解决传统行业数据流通难题，形成行业级模型服务，提升行业级 AI 应用开发效率。例如，晋云工业互联网平台以云计算为基石，深度整合煤炭行业数据，构建起煤炭行业大模型，并在此基础上衍生出智能开采辅助决策、智能通风监测等行业级模型服务产品。**二是生态系统方面**，行业云平台以新型云计算“平台经济”的形式集聚生态伙伴，发挥平台效应，孵化高价值行

业场景方案。合作共建一站式服务的企业应用市场，深度链接供需双方，加速行业云产品及服务的交易，实现行业智能化建设方案的快速推广与低成本复制，进而形成具有行业及地方特色的数字产业集群。例如，中化化工行业云作为化工行业业务协同生态平台，吸纳云上应用服务伙伴，助力企业在供应链中与上下游各方实现协同共享。三是持续运营方面，IT 团队正从传统运营向智能运营模式升级，借助 AI 技术，加速构建智算云服务台等智能化运营能力，以智能问答、智能检测、智能报表生成等手段提升运营效率。例如，IBM 借助 Cloud Pak for AIOps 助力荷兰银行实现智能化运维转型。通过模型对 IT 系统日志进行自动分析，精准预测故障并生成修复方案，提升事件响应效率的同时，有效减少人为操作失误，高度契合提升行业云平台运营效率的目标。

互联互通层面，算力互联网成为企业“上云用算”的核心要素，有力支撑灵活、可扩展的算力服务生态体系构建。人工智能时代，大模型训推、边缘计算等任务式应用爆发，驱动分布式算力调度需求激增。各类算力服务商积极参与算力互联网技术创新，构建算力标准化、服务化的大市场和算力相互连接、灵活调用的一张逻辑上的网。一是设施网络层方面，算力互联网基于计算和传输的物理资源，构建分布式、灵活的算力基础设施，支持动态调度与分配。服务商在设施网络层通过创新应用大规模 GPU 集群高性能通信、弹性网络、全光网络、长距离高性能通信传输等技术，系统提升算力基

基础设施的并行计算能力和整体性能。例如，中国移动牵头发起 X-Link 开放卡间互联技术生态，打造开放式架构超节点，提升算力互联效率。二是资源互联层方面，算力互联网可以实现异构算力的智能感知、实时发现、按需使用，推动跨地域资源协同调用。在资源汇聚中，服务商通过算力标识网关、API 接口通信等方式实现基于统一算力标识规则的资源汇聚；在服务调用中，服务商通过开放接口、互联协议等方式，实现跨服务商算力资源互操作。例如，亚马逊云、微软云、谷歌云、阿里云等国内外算力服务提供商支持采用 IaC（Infrastructure as code，基础设施及代码）、SCP（Standard Communication Protocol，标准通信协议）等方式对资源进行管理和配置，允许用户通过标准方式定义和管理基础设施。三是应用服务层方面，算力互联网能够保障计算任务及数据可实现资源匹配和高效执行，提供按需计算服务，促进智能化应用落地。服务商在应用服务层构建结构化的算力应用描述语言，该语言与现有主流算力应用兼容，通过规范算力应用描述的标签、属性及内容，确保算力应用信息在不同环境下都能被准确兼容解析。服务商针对 CUDA、OpenCL、ROCm 等国内外主流的 GPU 计算框架进行兼容适配，实现统一编程模式、统一 API 接口、统一并行计算、统一硬件兼容。

二、我国云计算发展概述

（一）中央与地方协同发展云计算，助力产业智能化转型升级

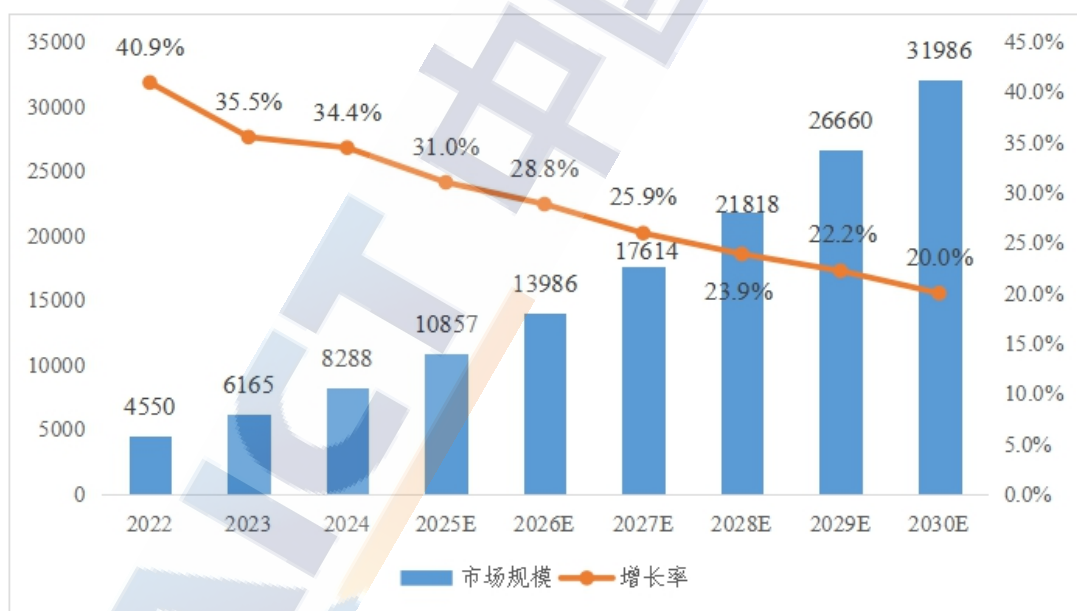
国家层面，持续强化政策引领，推动云计算由基础应用向新技术融合发展转变，加速拓展行业应用深度。一方面，从推动云计算发展的政策来看，2017-2020 年期间，国家相继出台的诸如《云计算发展三年行动计划（2017-2019 年）》《推动企业上云实施指南（2018-2020 年）》等多项政策，侧重于促进云计算技术应用、夯实云计算基础设施，近两年出台的政策更注重利用新兴技术赋能国计民生领域，以促进各行业智能化转型。例如，在《关于推进新型城市基础设施建设打造韧性城市的意见》中，提出利用物联网、云计算、大数据等实现系统平台、家居产品互联互通，加快构建跨终端共享的统一操作系统生态；在《关于数字贸易改革创新发展的意见》等中，提出加快发展物联网、云计算、人工智能、区块链等领域对外贸易；上述两项政策标志着云计算发展正由基础技术应用向赋能城市建设发展、数字贸易改革等方面转变。另一方面，从云计算应用的政策来看，近两年侧重于持续深化云计算在垂直行业领域的应用。例如，在《工业和信息化部等三部门关于促进环保装备制造业高质量发展的若干意见》中，提出充分利用人工智能、大数据、云计算等新一代信息技术，提升环保装备设计、生产、使用等各环节数字化智能化绿色化水平；在《医药工业数智化转型实施方案

（2025-2030 年）》中，提出鼓励建设一批高性能云计算平台、区块链、数据中心等信息基础设施，支撑医药企业“智改数转网联”；上述两项政策标志着云计算应用正由互联网、金融等行业向制造、医疗等传统行业持续渗透。

地方层面，各省市积极响应国家号召和布局，塑造具有地方特色的重点行业创新应用格局。一方面，各地依托国家政策引导，围绕智算技术创新和集群建设，加快人工智能与云计算深度融合。例如，上海市在《关于促进智算云产业创新发展的实施意见（2025-2027 年）》中，提出打造若干综合型智算云平台 and 一批垂直型智算云平台，形成一批智算云标杆应用；天津市在《算力产业发展实施方案（2024-2026 年）》中，提出重点推动提升智能算力规模，加快构建多元算力互联互通和统一服务的算力交易平台，促进算力供给、调度、使用、结算智能化。**另一方面，**各地围绕当地产业应用特色，积极推动云计算融合创新从制造业向广告、建筑等其他行业的渗透与应用。例如，北京市在《促进数字广告业高质量发展的指导意见》中，提出着力以大数据、云计算、大模型、数字人等技术赋能数字广告发展；江苏省在《关于推进全省智能建造发展的指导意见》中，提出加强物联网、大数据、云计算、人工智能、区块链等新一代信息技术在建筑领域中的融合应用。

（二）我国云计算市场保持高速增长，AI 云成为厂商破局关键

从整体来看，我国云计算市场保持较高增长态势。据中国信通院统计，2024 年我国云计算市场规模达 8288 亿元，较 2023 年增长 34.4%，仍保持较高增长速度，标志着以云计算为基座的数智市场格局已初步形成。其中，公有云市场规模 6216 亿元，同比增长 36.6%；私有云市场规模 2072 亿元，同比增长 29.3%。随着量子计算、区块链、人工智能等技术与云计算的融合革新，云计算的市场边界将进一步扩展，预计在“十五五”期间仍保持 20% 以上的增长，到 2030 年我国云计算市场规模有望突破 3 万亿元。

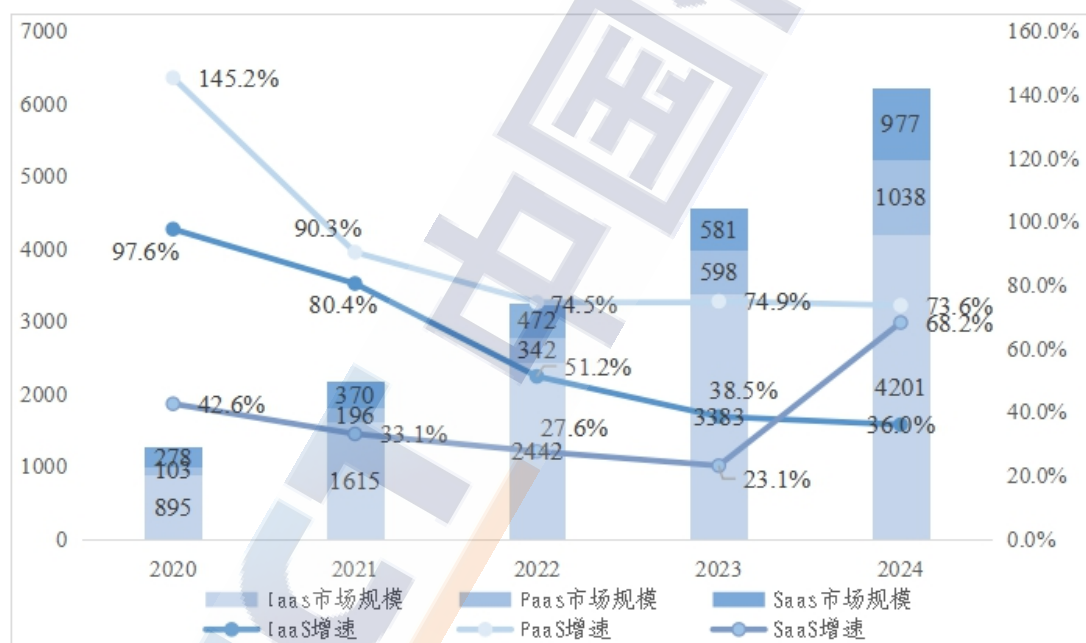


来源：中国信息通信研究院，2025 年 4 月

图 4 中国云计算市场规模及增速（亿元）

从细分领域来看，智算服务和智能体是 IaaS 和 SaaS 市场增长

的主要驱动力。增量方面，2024 年我国公有云 IaaS 市场规模达 4201 亿元，同比增长超 800 亿元，其中智能算力服务需求成为增长的最主要因素。增速方面，我国公有云 SaaS 市场率先实现增长反弹，企业级应用的裂变及智能体的爆发推动市场增长率由 23.1% 上涨至 68.2%，随着 AI 云基础设施支撑及 AI 云应用场景落地，预计未来 3-5 年将形成“技术+生态”双轮驱动格局。此外，公有云 PaaS 增长相对较弱，但也受 AI 开发平台和出海业务推动，2024 年市场规模突破千亿元。



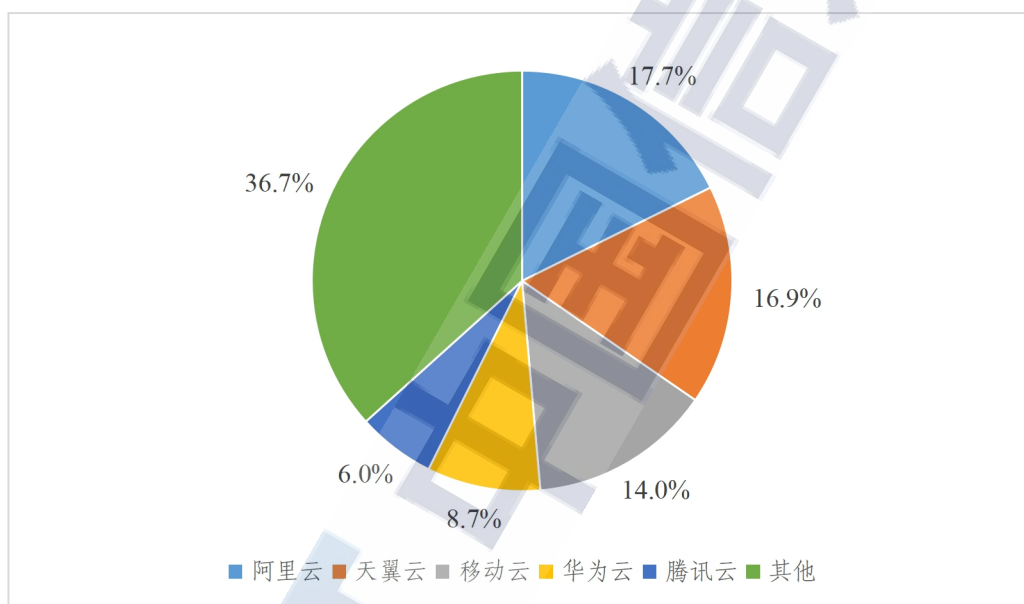
来源：中国信息通信研究院，2025 年 4 月

图 5 中国公有云细分领域市场规模及增速（亿元）

从厂商层面来看，头部格局基本形成，中腰部厂商试图借助细分领域打破格局。据中国信通院调查统计²，阿里云、天翼云、移动

² 市场规模为 2024 年全年数据统计，主要依据企业财报、人员访谈、可信云评估、历史数据等得出。对于市场数据不明确的领域，只发布头部企业整体情况，不做具体排名。

云、华为云、腾讯云占据中国公有云 IaaS 市场份额前五³；公有云 PaaS 方面⁴，阿里云、百度云、华为云、腾讯云、天翼云、移动云处于领先地位。头部厂商通过在 AI 云布局上的先发优势拉开差距，中腰部厂商保持积极追赶态势。为打破市场格局，中腰部厂商通过深化垂直行业 AI 云、聚焦 AI 提升的高性能计算场景，发展云生态联合，服务企业出海业务等差异化战略在细分领域打开局面。



来源：中国信息通信研究院，2025 年 4 月

图 6 2024 年中国公有云 IaaS 厂商市场占比

（三）“云+人工智能”双轮驱动转型新范式，行业智能化应用加速普及

去年，《云计算蓝皮书（2024 年）》全面分析了我国各行业上云的梯队分层效应，并从云建设、云发展和云应用等方向详细分析了

³ 因为 IaaS 和 CDN 是两种业态，需要分别获得互联网资源协作服务业务牌照和内容分发网络业务牌照，所有 IaaS 不包括 CDN 收入，只统计计算、存储、网络等基础资源服务收入。

⁴ 排名不分先后。

各梯队上云发展特点及成果。今年，本蓝皮书将以用云为切入点，从行业类型出发，挑选政务、交通两个具有典型性的上云行业，深入分析其云上智能化应用特点；从企业规模出发，分别对央国企、中小企业的用云方式进行剖析，旨在发挥重点行业“头雁”效应，推动人工智能全行业创新赋能，构筑云计算新发展格局。

政务方面。“一体化”与“智能化”成为各地政务云建设的关键词，“融智焕新一朵云”为数字政府发展奠定“基础+全栈”的基石。各地数字政府一体化综合改革加速，强调三融五跨的管理和服务水平综合提升，对政务云提供的支撑能力提出了更为全面的要求。整体来看，政务云发展呈现以下四点特征：**一是政策导向不断集中，引导一体化数智底座持续夯实。**为深入贯彻国家“人工智能+”战略部署，深化数智赋能，各地近期重点聚焦基础设施智能化建设，部署新要求，提出新思路。山东省强调夯实一体化数智平台能力支撑，完善政务“一张网”“一朵云”一体化支撑能力，提升云网功能、性能和调度管理能力。重庆市聚焦算力支撑能力，精细管理在大模型训练、推理过程所需的算力资源，更有效强化智能算力资源供给。**二是技术融合不断提速，应用成效持续优化。**政务云的核心能力正在从以通用计算为核心升级为以智能计算和数据为核心，依托于政务云的大规模分布式计算的统一调度分发服务能力，推动算力与 AI 大模型的深度融合，释放政务云新潜能，为政务云的高效运营和智能化发展提供有力支撑，有力提升政务云对上层业务的赋能作用。

三是场景应用更趋智能，赋能服务深度与广度提升。随着人工智能技术应用深度的提升，政务云场景化服务能力更趋智能，在政府服务与城市治理等典型场景下实现深度赋能。在政务服务领域，政务云支撑智能客服、数字人客服等应用，极大简化行政流程，提升政府办公效率；在城市治理方面，政务云作为“交互枢纽”“计算枢纽”“服务枢纽”和“数据枢纽”，加速各类城市治理数据的融合。**四是生态格局渐进细化，产业合作模式不断创新。**从建设内容看，政务云依旧是各服务商提供服务的核心锚点，政务大数据、政务大模型、政务应用、政务安全、泛行业政务应用等方面均有服务商布局，各服务商相互合作，逐步细化数字政府总体发展框架指导下的产品赛道，共同构建创新、清晰、共赢的生态共同体。

交通方面。交通云由“基础数字底座”向“全局智能引擎”加速迈进，“云+人工智能”赋能场景智慧应用成为交通云智升级新抓手。交通云作为支撑行业转型发展和业技融合创新的关键基石，承载着人工智能、大模型等关键技术，推动交通行业云智升级迈入新台阶。整体来看，交通云发展呈现出以下三大特点：**一是政策精准布局培育行业转型动能，公路民航等领域加快核心系统云改智转。**近年来，围绕行业转型发展，交通领域各项政策频繁出台，已形成以《交通强国建设纲要》为核心，各细分领域精准施策的“1+N”立体政策支持体系，为全行业转型升级提供明确方向及行动指南。围绕数字化、智能化等关键发展要旨，公路、民航等细分领域云智改造

渐入“深水区”。公路领域，各地高速集团锚定路网运行系统云原生改造，推动公路大模型应用从基础的“查政查数”向进阶的“问策问措”实现智慧升级；民航领域，东部机场集团加速云上大模型+物联网赋能智慧机场建设，加快万物智联向航班运控、航路规划、保障资源调度等机场核心系统改造升级。

二是业技融合创新释放云智乘数效应，核心场景价值为先加速技术赋能智慧交通。近年来，各地围绕交通细分场景发展痛点与业务需要，结合云、网、数、智、安等技术加快车路云一体化、自动驾驶、智能网联车、交通大模型等试点示范迭代升级，交通云愈发由信息化实践模板向数智化创新引擎角色升级，融合知识库搭建、RAG 微调、Agent 开发等技术，充分发挥云智一体乘数效应，加快公路工程勘设、城市路网指挥、机场运行保障、港口航道闸口、物流货运线路等核心生产运营场景应用智能化发展，锚定效率与服务两大业务价值显现，不断提升公路、水路、民航、港口、物流等细分领域的效能，加码交通行业智能化应用成效。

三是行业共识反哺生态共建共享，协同交通产业链上下游优化整体效能提升。面对 DeepSeek、具身智能、低空经济等现象级行业话题热点，交通大模型一体机、智能体开发、定制化产品服务与解决方案愈发涉及产业全链条、全要素、全过程参与，促使交通产业在接口开放、数据交互、标准对齐等方面加速共识形成，联合上下游产品供需方，使业技适配需求与规模持续扩大，提升生产效能。

央国企方面。央国企用云重点从“数字化基础设施建设”转向“智能化云上应用”，央国企级行业智能化应用成为新蓝海。国民经济增长的“顶梁柱”、科技创新的“国家队”，是将智能化与政策和产业融合、带动全产业链智能化发展的重要推动力，也是我国产业升级、产业结构优化的重要根基。整体来看，央国企用云发展呈现三大特征：一是政策聚焦以云提升科技创新能力，深化“人工智能+”专项行动。国务院国资委、工业和信息化部、国家发展改革委等部门 2024 年开展中央企业人工智能专题推进会、“人工智能+”专项行动深化部署会等一系列央国企行业智能化行动，将人工智能、云计算等重点领域纳入“十五五”央国企重点发展规划，明确技术研发、应用推广等方面的目标与任务，发布《关于促进企业数据资源开发利用的意见》等政策，要求企业大力推广云计算、边缘计算、大数据分析等平台服务，支持企业开发和使用智能化工具。二是上云成为筑牢央国企智能化应用根基的重要路径。过去一年，央国企围绕上云完成了资源集约化、数据价值化、安全体系化、技术敏捷化的“四化”工作。中国铁建等跨地域央国企通过多云融合策略，实现了各地铁路系统零散云计算资源的智能化统管。中国信科、广东粤数等服务型央国企将业务数据和业务实时情况与云相结合，以数据驱动服务向智能化迭代，提升了邮件、即时通信工具等应用的信息流转率与故障处理能力。天翼云、联通云等运营商构建“云-边-端-管”的智能化云安全防护体系和安全应用，极大地提升了云安

全能力。国家电网将深度学习、5G、边缘计算等云计算技术逐渐应用到电网检测、产线控制、输电等生产环节，通过技术迭代加速生产应用智能化。三是基于云平台构建行业大模型，持续释放央国企的产业“压舱石”价值。央国企依托成熟云平台建设，可为大模型提供充分的计算资源、数据存储、弹性扩展及部署推理等基础条件。能源领域，中国石油、国家能源集团等通过能源大模型，整合生产过程中涉及的地理、能源等数据，并实现基于不同场景优化能源调度运营与决策效率。制造领域，中国宝武、中国船舶集团等通过制造大模型整合战略、市场和技术数据，构建辅助决策链，并在生产层面实现全流程协同。农业、水利、烟草等传统领域的多家央国企陆续发布行业大模型并逐渐赋能到业务领域，极大地扩充了央国企智能化应用的深度与广度。

中小企业方面，企业上云从“业务上云”逐步深化为“云智业融合”。作为市场经济的活力源泉，中小企业是稳定经济增长、促进就业创业的关键力量，更是促进产业生态多元化发展，推动数字化普惠创新的重要基石。整体来看，中小企业上云及智能化转型进程呈现四大特点：一是政策加码中小企业数字化转型，推动人工智能创新赋能。2024 年，工业和信息化部发布《中小企业数字化赋能专项行动方案（2025-2027 年）》，计划通过城市试点推进数字化改造，加速 AI 应用，到 2027 年实现中小企业上云率超过 40%。深圳市推行“企业出一点、服务商让一点、政府补一点”的支持原则。金华市

出台资金扶持细则，激励各方参与数字化改造。二是中小企业上云普及率整体提升，从业务上云向云上智能化过渡。随着人工智能、大数据等新兴技术的应用落地，企业对算力需求激增，上云成为中小企业应用新技术的最佳选择。部分数字化程度较高的企业已在云端运营大数据分析、人工智能应用类服务，实现了从基础业务上云向云上智能解决方案的升级。畅捷通财报显示，其 2024 年云订阅收入同比增长 38%，新增付费企业用户 10.7 万家。三是中小企业探索智能化升级，SaaS 集成 AI 显著提高业务效率。生成式 AI 兴起后，SaaS 服务商纷纷融入 AI 功能。中小企业借助企业级 SaaS “小快轻准” 优势，快速在多场景部署应用 AI 能力。如钉钉的“魔法棒”功能助力提升办公效率，腾讯会议 AI 小助手实现会议管理的优化，金蝶小微 AI 开单助手提高开单效率至 90%、记账效率提升 10 倍。四是行业链式上云趋势明显，链主引领中小企业智能化升级。中小企业智能化转型深受产业链协作影响，链主完成自身数字化转型后，带动全生态上云赋智，构建云上数智生态。以工业制造为例，链主搭建云平台供应链管理系统，与链中企业实时共享库存、订单等关键数据。链上中小企业也随需求上云提升数字化水平，响应智能化规划生产与物流安排，从而实现整条产业链云上数智化转型。

三、云计算驱动 AI 服务范式创新，激发人工智能时代新质生产力

随着人工智能技术的不断革新，人工智能开始走入千行百业，

从工作场景走入生活场景，社会对算力的普惠化、场景化、生态化需求愈发凸显。针对人工智能时代的全新需求，云计算服务模式迎来了关键转型期，正以迅猛之势向“人工智能+”方向加速转化。传统的云计算架构已难以满足人工智能对算力、算法和数据处理的复杂要求，全新的智算云体系架构应运而生。在这一架构中，AIaaS（智算云基础设施服务）、AIPaaS（智算云平台服务）、MaaS（大模型服务）、AISaaS（智算云应用服务）、AIMSP（智算云管理技术服务）等创新模式不断涌现，展现出全产业链的进化态势。这些模式紧密协作，从提供底层算力支撑的基础设施，到算法模型的开发与部署，再到面向用户的各类智能化应用，协同推进 AI 技术实现从基础设施建设到商业价值创造的全链路转化。这不仅重塑了企业的生产运营模式，更为整个社会构建起智能时代的新质生产力范式，为经济发展与社会进步注入强大动能。

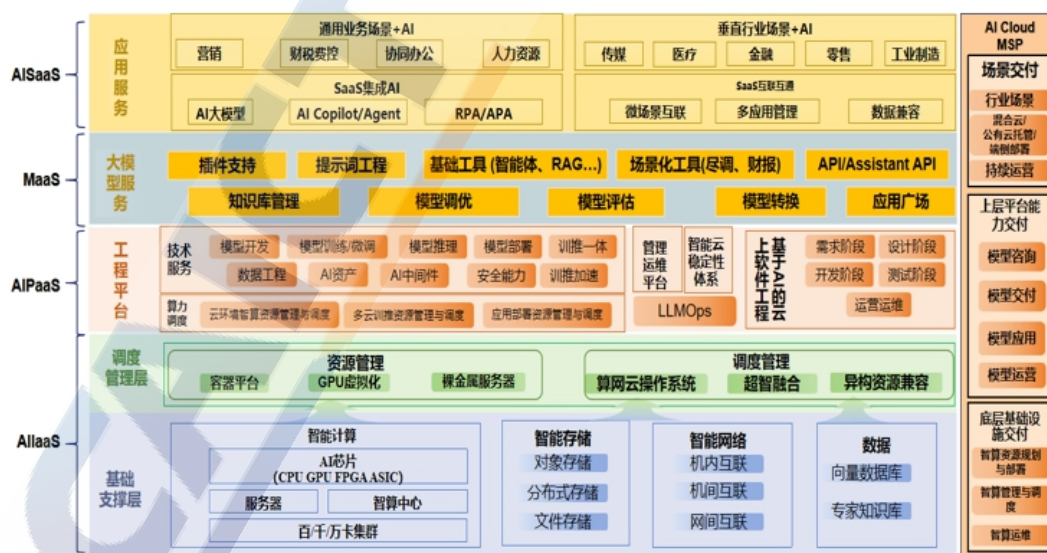


图 7 智算云体系架构图

（一）AIaaS 革新智能基础设施架构，助力 AI 模型实现高效训练推理

随着人工智能技术加速渗透千行百业，AIaaS 作为企业智能化转型的核心引擎，正通过多架构算力融合破解场景碎片化挑战，依托智能调度技术实现异构资源的高效协同，最大化智算效能。同时，安全与能效的双重保障推动软硬件协同优化，构建高可靠、可持续的智能化基础设施，为千行百业提供敏捷、高效的 AI 服务支撑。

AIaaS 以异构智算资源为核心，AI 推理需求占比逐步递增，与训练需求差距逐渐缩小。随着 DeepSeek 等先进大模型的快速发展，企业和开发者正加速将这些模型应用于决策支持、数据分析和流程自动化等关键场景。这一趋势推动 AI 推理需求呈现爆发式增长，为 AIaaS 带来广阔的发展空间和创新机遇。AIaaS 通过持续优化异构计算架构、提升资源调度效率，为各类 AI 应用场景提供更加强大、灵活的基础设施支持。一是 AIaaS 极致优化异构计算资源供给与输出。视频推理场景需要高能效比的 ASIC（Application-Specific Integrated Circuit，专用集成电路）芯片，而金融风控场景则依赖 FPGA（Field-Programmable Gate Array，现场可编程门阵列）实现微秒级响应，AIaaS 可动态调配 CPU、GPU、NPU（Neural Processing Unit，神经网络处理单元）等多元计算资源以应对算力需求差异。例如，AWS 的 Nitro 系统结合 FPGA 实例和自研芯片，为机器学习训练与推理提供异构资源池，支持用户混合调用不同加速器。微软

Azure 部署了包含 Habana Gaudi 芯片和 GPU 的混合算力池，针对自然语言处理等场景优化能效比，部分任务较传统方案能耗降低 40%。

二是 AIaaS 支持计算资源的智能调度与弹性扩展。智算云平台通过虚拟化技术和自动化资源管理，异构计算资源在推理场景下能够实现无缝扩展。例如，亚马逊云 Inferentia2 芯片支持千亿参数模型的秒级部署，能够快速响应业务需求的变化。**三是软硬件协同优化提升上层 AI 应用效能。**不同计算架构在特定任务中的性能表现存在差异，这种差异不仅体现在算力效率上，更直接影响系统的能耗比和响应速度。目前，业界正积极探索异构计算方案，通过灵活组合不同计算单元来匹配多样化的工作负载。例如，阿里云“磐久”服务器通过“CPU+GPU+FPGA”异构架构，在推荐系统推理任务中实现了单位算力功耗降低 40% 的能效优化。

AIaaS 打造智能异构算力调度引擎与弹性资源编排系统，大幅提升计算资源利用率和推理任务响应速度。云计算平台通过动态编排技术和推理即服务模式，能够显著提升算力资源的利用率和推理任务的执行效率。**一是 AIaaS 采用动态编排技术激活算力资源潜能。**传统静态资源分配导致推理集群平均利用率不足 40%，而智能调度系统通过时序预测算法，可动态聚合碎片化算力资源。例如，阿里云 PAI 平台采用时空联合调度策略，在图像识别推理任务中实现 90% 以上的资源复用率，推理成本下降 57%。**二是 AIaaS 催生“推理即服务”AI 交付新模式。**大模型推理存在显存占用高、冷启动慢等痛

点，这些痛点引发了“模型托管+算力优化”新型融合服务。例如，亚马逊云 Inferentia2 芯片内置神经元核心硬件加速引擎，支持千亿参数模型秒级部署；华为云 ModelArts 推出自适应批处理技术，吞吐量较通用方案提升 20 倍。三是 AIaaS 使用智能调度算法优化资源分配。AI 推理工作负载呈现明显的峰谷波动特征，AIaaS 通过实时预测负载变化并动态调整资源配比，保障服务质量的同时最大化资源利用率。例如，腾讯云 StarLake 采用智能调度算法，根据推理任务负载动态分配资源，集群整体能效提升 30%。

AIaaS 驱动算力与效能平衡，开创“越智能越节能”的新范式。

AIaaS 通过硬件级安全架构、液冷技术和软硬件协同优化，能够实现安全与能效的平衡，从而为 AI 推理任务提供绿色、安全的计算环境。一是 AIaaS 硬件级架构增强边缘 AI 场景隐私保护。边缘 AI 推理场景下构建端到端的加密传输体系，并在硬件层面实现可信执行环境（TEE），确保数据从采集到推理的全生命周期安全。传统软件加密方案往往导致性能损耗，而硬件级安全架构能够显著降低这种损耗。例如，英伟达 BlueField-3 DPU 通过数据卸载与硬件级隔离技术，降低隐私计算的性能损耗，性能表现显著优于传统软件加密方案。二是创新性融合液冷技术与异构计算架构。AIaaS 通过液冷技术降低能耗，能够大幅降低数据中心处理高计算密度 AI 训推任务产生的能耗。例如，阿里云“磐久”服务器通过“CPU+GPU+FPGA”异构架构，在推荐系统推理任务中实现能效优化，单位算力功耗降

低 40%。微软 Azure 部署浸没式液冷系统，数据中心冷却能耗减少 95%，为大模型推理提供绿色算力支持。三是软硬协同优化构建算力与效能平衡点。AIaaS 依托能效与安全双轮驱动的平衡机制，在 AI 应用安全和数据中心能效之间构建平衡点。在实现高效能 AI 推理的过程中，确保资源调度具备毫秒级响应能力的同时构筑严格的硬件级隔离防线，为上层 AI 应用安全提供可靠保障。例如，腾讯云 StarLake 采用智能调度算法，根据推理任务负载动态分配资源，集群整体能效提升 30%，同时通过零信任架构确保数据安全。

（二）AIPaaS 重塑 AI 开发服务新范式，加速 AI+应用快速规模化落地

随着人工智能技术的快速发展，大模型、生成式 AI 等技术不断取得突破性进展，AI 应用场景日益丰富，对算力的需求也呈现爆发式增长。AI 云平台服务作为推动 AI 大规模应用的核心，正在成为智能算力高效利用和 AI 技术升级的关键支撑。AIPaaS 依托大规模异构智能算力的融合调度，为 AI 模型开发、训练、部署和推理提供了高效、便捷的云化服务，成为企业“人工智能+应用”落地的催化剂。

AIPaaS 降低了 AI 应用的开发门槛与技术成本，成为企业智能化转型的核心引擎。AIPaaS 通过技术架构创新与资源整合，显著降低 AI 应用开发门槛与成本，成为企业实现智能化升级的核心驱动力。一是云服务架构释放了 AI 开发的算力潜能。AIPaaS 凭借高性能、低能耗、灵活稳定且拓展性强的服务模式，通过智能调度算法实现

算力资源的动态分配与优化，可针对训练、推理、微调等多样化任务类型匹配最佳计算资源组合，满足大模型对高性能算力大幅提升的需求。AI 云平台通过这种差异化服务需求，使企业无需关注底层硬件适配细节，即可专注业务逻辑开发，实现多样化 AI 任务。例如，字节跳动扣子平台不仅内置大量行业模板，更通过可视化算法配置界面与自动化超参调优工具，例如，字节跳动扣子平台内置大量行业模板，企业通过“模板复制、个性化调整”即可完成应用部署，开发效率大幅提升。二是 AIPaaS 全栈工具链驱动了人工智能应用开发范式。AI 云平台内置的自动化模型调优组件、可视化数据标注工具及预训练模型库，同时集成混合精度训练、分布式梯度压缩等开发算法，在保障模型精度的前提下显著缩短迭代周期，使得开发人员聚焦业务场景创新，快速生成适配的行业解决方案。例如，天翼云智算平台提供从模型部署到监控的一站式工具，支持快速集成 CRM 数据并生成个性化营销方案。三是标准化接口打通了生态协同壁垒。基于云服务构建的 API 市场与模型仓库，可以实现跨平台算法组件的即插即用，更通过安全技术保障数据主权与模型溯源，使企业在调用第三方 AI 服务时仍能保持核心资产的安全可控，加速形成开放共赢的智能生态。例如，钉钉 AIPaaS 开放 API 接口，支持与 ERP、CRM 等系统无缝对接，形成“平台+生态”的智能服务网络。

AIPaaS 打造从数据到应用的全流程技术路线，构建企业规模化落地智算服务的核心枢纽。企业依托 AIPaaS 实现数据价值挖掘、模

型敏捷开发与服务的规模化落地，最终形成“智算即服务”的核心竞争力。

一是打通数据全生命周期管理。基于 AIPaaS 平台构建数据湖仓一体化架构，实现从数据采集、清洗到特征工程的全流程自动化处理。例如，通过预置数据标注工具与联邦学习机制，在确保隐私合规前提下将模型训练数据准备效率提升 60% 以上。阿里云 PAI 平台通过“数据湖治理、模型开发、服务编排”三阶段贯通，助力企业快速构建 AI 应用。

二是重塑 AI 训推流程。AI 云平台提供从 TensorFlow/PyTorch 框架适配、MLOps 流水线搭建到模型服务化部署的全栈应用开发部署套件，结合预训练大模型与低代码开发能力，大幅缩短 AI 应用开发周期，AIPaaS 层的平台化能力赋能企业基础设施智能化升级。例如，腾讯云 TI 平台构建模型训练到推理加速的全托管服务，支持主流框架一键迁移至国产算力环境，大幅降低异构算力使用门槛，加速行业解决方案沉淀。

三是加速 AI 应用规模化落地。当前，AI 应用正从技术突破向场景渗透加速演进。AIPaaS 依托云服务架构，不仅为大模型训练提供弹性扩展的智算服务，还能在推理阶段根据实时请求量自动扩容，将典型 AI 服务的响应延迟降低至毫秒级，成为加速 AI 应用规模化落地的重要推动力。例如，华为云 ModelArts 提供开箱即用的大模型服务，通过预置 Prompt 模板和自动超参调优，企业可快速构建智能 Agents。

AIPaaS 提供稳定的算力环境，保障模型训练推理平稳运行。AIPaaS 通过多层次的监控、调度、巡检等措施，确保了 AI 模型的训练过程

全流程可管可控，在硬件故障概率高的情况下，做到自动容灾和断点续训。

一是在资源调度方面，AIPaaS 支持弹性扩缩容、训练调度、推理调度、监报告警等多种能力，在模型推理过程中可以根据业务需求自动增加或减少计算资源，并支持对模型训练任务的资源分配与执行顺序进行合理安排，以最大化资源利用率、优化训练时间，并确保多任务环境下的公平性和任务稳定性，同时具备对计算资源的全面监控。例如，移动云使用自研 `lava` 调度器，进行资源统一纳管及训练调度，实现资源统一分配、任务统一调度，解决资源碎片化问题。

二是在模型训练保障方面，AIPaaS 提供隐患巡检、自动诊断、容错容灾、检查点保存、断点续训等稳定性保障能力，在训练任务开始前以及训练任务中检测集群健康度，输出检测报告，自动剔除问题机器，自动补齐资源，在任务训练过程中可以实时监测，识别和定位问题。例如，百度百舸 AI 异构算力管理解决方案通过自动二分通信测试，为隐式硬件故障诊断提效。

三是在训练集群发生错误或故障时，AIPaaS 能够确保训练任务继续运行，并保证任务功能和性能的一致性，以及根据训练任务的不同场景和需求，提供多种保存策略，提升训练过程中任务的连续性以及系统的整体稳定性。在训练过程中由于意外中断而停止后，能够自动从检查点继续恢复训练。例如，移动云在某车企智慧驾驶案例中，提供了千卡集群规模的、长达数月级的稳定训练服务，对于千亿参数模型的检查点实现了秒级保存，并且可以灵活地接入到该车企的智慧驾驶业务系统中。

（三）MaaS 提供全流程服务能力，成为大模型工程化落地重要载体

在人工智能技术快速发展的当下，大模型工程化落地已从技术验证阶段转换到产业应用量产阶段，MaaS 正从大模型部署优化、应用开发工具链完善与计价模式灵活性提升三方面，显著降低企业 AI 应用构建的门槛与成本，加速大模型在各行业的深度渗透与价值释放。

MaaS 通过优化大模型部署与模型二次开发工具链，降低企业大模型推理与定制化门槛。在优化大模型部署方面，MaaS 能够满足不同模型对各类芯片（如 GPU、NPU 等）以及主流 AI 推理框架（如 vLLM、SGLang 等）的适配需求，实现 AIaaS 与框架特性的精准匹配。这种匹配不仅提高了模型的运行效率，还有效降低了大模型推理服务的延迟，提升幅度可达 30% 以上。MaaS 通过与 AIPaaS 在优化资源分配和任务调度等功能上相结合，还能够显著提升大模型推理阶段的单卡 token 吞吐量。模型部署层面的优化，不仅使 MaaS 服务商能够更高效地利用计算资源，同样大幅提升了企业侧用户在使用大模型时的服务质量。例如，火山引擎 MaaS 平台在优化部署 DeepSeek-R1 时，通过分离式推理架构，在 AIaaS 层深度适配 GPU 集群，释放硬件并行能力，推理吞吐率较开源版本提升 5 倍，输出 token 时延缩短 40%；在 AIPaaS 层，整合机器学习平台实现弹性调度，显著提升单卡 token 效率。在大模型二次开发工具链方面，MaaS 聚焦大模型在数据、提示词、微调、评估等关键环节的模型工具链

配备。MaaS 平台通过多模态数据标注实现结构化数据自动清洗与标注，结合低代码微调框架支持模型的调参，完成千亿参数模型的二次定制化开发流程。这帮助企业能够基于通用大模型，融合行业知识图谱与业务特性，完成企业专属大模型的开发，进而响应特定业务场景的需求。例如，工商银行通过 MaaS 平台大模型二次开发工具链，融合其金融知识图谱与信贷业务的风控规则、流程逻辑等核心业务特性，构建起千亿参数级的金融大模型，并深度应用在公信贷业务领域，实现了信贷业务全流程的智能化赋能，显著提升其业务效率与风险管控能力。

MaaS 通过扩展全栈工具链组件与垂直领域结合，持续提升企业 AI 应用的开发效率与场景适配性。在应用开发工具链方面，以 Manus 为代表的通用人工智能（AGI）正在对 MaaS 平台的组件种类、功能深度及跨模态处理能力提出挑战。MaaS 需通过整合互联网搜索引擎、检索增强生成、代码解释器、视觉-语言模块、多模态数据标注工具及领域专用算法库等核心组件，构建起覆盖 AI 应用构建与编排的全生命周期工具链。以腾讯云 TI 平台为例，其内置的“多模态 workflow 引擎”通过拖拽式界面配置，可实现从 OCR（Optical Character Recognition，光学字符识别）、知识库检索、代码生成到结果可视化的全链路自动化流程，显著降低了企业在完成多模态应用开发时的技术门槛。在垂直领域方面，MaaS 应用工具链的功能深度与场景适配能力决定了 AI 应用在业务场景下的商业价值。这种能力体现在企

业用户基于 MaaS 服务对与行业知识库的深度整合、智能体框架的灵活调用及标准化 API 的生态衔接，进而驱动不同行业的场景应用快速落地。在零售行业，企业用户通过微软云 Azure OpenAI、IoT Hub、Azure Machine Learning 等 MaaS 核心功能组件，构建个性化搜索系统，通过用户行为数据与商品特征的实时分析，实现精准推荐与智能客服响应，进而提升零售企业客单价，同时降低客服人力成本。

MaaS 通过灵活的计价模式与推理服务成本优化，使得企业能够更精细化地管理其大模型成本。计价模式方面，MaaS 服务商普遍采用基于 Token 的计费机制，按输入输出及缓存命中计量。这种按需付费的机制，将固定投入转化为弹性支出，企业用户无需提前投入大量资金押注于某个特定规模的模型或基础设施，以低成本迅速试错，精准选型出性能与效益兼备的模型，加速大模型工程化落地的进程。国际主流服务商除按 Token 计费外，还提供订阅制等多元化收费模式，其 API 调用收入占比约 30%，订阅服务占 70%。相较而言，我国 MaaS 服务商以 Token 计费为主，具备灵活的时段差异化定价能力，且百万级别单位 Token 的成本显著低于国际市场。这主要得益于国产算力基础设施优化、大模型生态迭代加速及产业规模化效应，推动了 MaaS API 推理服务的边际成本下降。MaaS 推理服务成本优化方面，以 DeepSeek V3/R1 为代表的开源大模型的架构创新与算力效率提升正重构行业定价体系。头部厂商通过稀疏化计算、混合专家架构、教师模型蒸馏等技术，将大模型 API 平均推理成本

压缩至原先成本的 12%-15%，技术突破与成本优化的双向加速正在重构 MaaS 服务的成本经济模型。这种成本的大幅下降，如阿里云 Qwen-Long API 价格降至 GPT-4 的 1%，智谱 AI GLM-4 推理成本较年初降低 90%，使得企业级 AI 应用开发预算从千万级降至百万级。据测算，基于开源模型的 MaaS 服务可使智能问答、工业质检等场景的生产级别大模型研发迭代周期从 6 个月压缩至 3 周，实现 AI 应用敏捷开发的范式迁移。

（四）AISaaS 驱动企业级应用智能化，大幅提升业务场景运行效率

AISaaS 从“理论可行”向“商业有效”迈进，通过普惠化、场景化、生态化发展路径，积极融入生成式 AI、机器学习等人工智能技术，持续提升服务能力，成为人工智能深入赋能行业场景，深化生产级实践的重要手段。

AISaaS 驱动产品形态创新，企业级 SaaS 融合 AI 促进企业个性化、自动化、数智化。AI 与 SaaS 的融合使 SaaS 服务商不仅关注云服务的功能性和易用性，也开始探索融合 AI 技术所能带来的服务效率、个性化、智能化提升，为企业用户提供更为便捷和高效的服务。一是 AISaaS 提供个性化服务提升用户使用体验。AISaaS 根据用户的企业角色、历史操作习惯以及偏好等数据，以自动推荐功能模块、优化界面布局等方式动态适配用户需求，显著提高用户体验。二是 AISaaS 自动化处理机械化任务提高人效。AI 技术的引入使 SaaS 可

通过学习用户操作行为，以自动化组件辅助自然人处理如客服、财会等高重复度的任务，大幅提升工作效率。如某财税管理服务商在 SaaS 中融合结合机器学习等 AI 模块，将单人记账的效率翻倍提升，从每人每天 200-400 笔提高到 800-1000 笔。**三是 AISaaS 驱动挖掘数据价值赋能决策。**AI 与 SaaS 的结合使企业能够深度联动分析不同板块的数据，挖掘潜在价值并识别风险，以自然语言处理（NLP）技术为例，与 SaaS 结合后，企业便能对大量非结构化数据进行高效解析。在汽车涂装环节中，先通过产线中的物联网设备采集生产数据，再汇总到模拟仿真 SaaS 模拟生产流程，最后定位并优化影响生产效率的主要因素，提高净产量。

AISaaS 助力企业用户低门槛应用 AI 功能，企业级 SaaS“小快轻准”特点促进持续进化。通过标准化、模块化的产品设计，深度融合垂类大模型与 SaaS 优势，发挥企业级 SaaS 的轻量化、快速部署、低成本特性，助力 AI 从实验室走向广泛的企业场景。**一是轻量化设计推动中小企业普及应用 AI。**企业级 SaaS 的轻量化特性使 AI 技术能够以低成本、低复杂度的方式落地，尤其对于资源有限的中小企业，通过少量资金和人力即可快速部署 AI 功能。如服装行业中，基于流水线管理服务串联生产环节，提高生产效率 25%。**二是 SaaS 部署模式能够快速响应市场的变化，及时调整策略。**企业级 SaaS 的快速部署特性使 AI 技术能够迅速适配不同行业场景和市场变化，相较于传统本地化部署的软件，应用 SaaS 模式无需复杂的硬件配置和长

时间调试，企业可以按需订阅、即时启用，并根据市场趋势和业务需求动态调整功能模块。三是低成本推动垂直场景应用 AI，挖掘落地价值。如医疗健康领域中，某健康检查服务机构通过数亿人次的体检数据、医学文献以及专家知识库，训练了一个专注于健康管理的大语言模型，并根据新的行业标准持续进行迭代，形成面向健康管理 AISaaS 服务，大幅降低了人工重复操作，并提高了健康管理报告的反馈质量，数据显示，截至 2025 年 2 月，已累计生成超十万份报告，报告准确率超 90%。

AISaaS 提供垂直化、场景化孵化环境，加速 AI 技术生产级落地。AISaaS 作为 AI 技术场景化应用的重要载体，为不同行业提供灵活、高效的智能解决方案，显著推动 AI 技术的生产级应用。一是 AISaaS 从通用业务向垂直场景突破。AISaaS 向大规模商业化突破的首要步骤是挖掘垂直行业场景的独特需求，并挖掘传统行业大规模积累数据资产价值，针对性提升 AISaaS 面向垂直场景的处理能力，明显提升用户的应用质量，从而为下一步传播推广 AISaaS 提供切实案例。例如，某医疗预约服务 SaaS 与大模型融合，基于 200 万条医患对话数据训练并由医护人员反馈调优，结合医院排班规则和患者历史就诊记录，将线上预约匹配时间从平均 8 分钟缩短至 30 秒，患者预约后等待时间减少 60%。二是从技术突破到可持续创新。根据 SaaS 服务商积累的行业经验，结合 AI 技术突破推出能力更为先进的服务，再基于用户数据反哺优化服务，实现供需互利的正向循环。

如 Adobe Creative Cloud 集成 Adobe Stock、Adobe Firefly、Adobe Sensei 等生成式 AI 工具和模块，使 Adobe 数字媒体业务 2024 年营收攀升至 31.3 亿美元，同比增长 10%。且随着全球数百万用户在平台上高频次使用各项服务，源源不断产生的创作数据、操作数据持续涌入 Adobe 的数据库，为 AI 能力的持续迭代提供基础，形成“雪球效应”。三是 AISaaS 从单点突破到生态协同。随着 AI 在各类场景中的优势显现，各行各业对于前沿技术的接纳程度和需求大幅提高，使 AISaaS 提供商开始探索技术栈融合、产业链协同等生态化发展策略以实现互惠共赢。如钉钉、飞书、腾讯企点等 SaaS 服务商纷纷开放各自应用市场，吸引细分场景的 SaaS 提供商进驻，共同为用户“一键触达”的从通用企业管理到垂直业务专精的一站式 AISaaS 服务。

AISaaS 结合 Agent 实现智能跃迁，驱动 SaaS 软件架构和服务范式，引领产业向实时化、智能化、生态化深度演进。AISaaS 与 Agent 本质上是技术赋能与应用形态的共生关系。通过两者深度融合，共同推动企业 IT 服务从单一 SaaS 服务向智能化、实时化、生态化的 AISaaS 持续演进。一是从静态规则到动态智能。传统 SaaS 参考行业经验和数据预设静态规则和流程对输入进行处理，而融合智能体后可通过大模型自适应学习，额外参考多模态实时数据优化输出，实现动态升级。例如，2024 年“双十一”期间，某商家以电商 SaaS 智能体矩阵为基础，组建了贯穿电商经营全流程的多智能体系，根据历史销售数据、市场热点和用户行为等信息实时更新营销方案，

帮助其工作效率提升 75% 以上。二是从 CURD 驱动到目标驱动。传统 SaaS 以数据增删改查（CURD）为核心，自然人仍需学习使用不同系统；而融合智能体系统后，SaaS 能够主动理解用户意图，以完成目标为导向执行任务，大幅降低学习成本。例如，Anthropic 发布的 Computer use 智能体能够通过理解屏幕内容，直接点击和操作系统界面；OpenAI 智能体 Operator 可以分析用户指示，自动完成餐厅订位、购买日常用品、预订比赛门票等日常任务。三是从孤立系统到生态互联。传统 SaaS 系统之间需要通过定制开发接口进行集成，实施周期较长。智能体的应用可能打破传统 SaaS 间的数据和互操作屏障。通过模仿人类操作、文本识别、标准化协议等方式基于多智能体协同网络快速构建跨 SaaS 的工作流，快速实现跨平台协作。

（五）AIMSP 聚焦 AI 云专业技术服务，打通端到端大模型应用交付体系

在人工智能云服务领域，AIMSP 指全栈全生命周期的 AI 云专业技术服务，依托人才、工具、流程、技术等服务要素，为企业提供算力层、数据层、模型层“咨询、交付、集成、运营”端到端的技术服务。AIMSP 致力于构建完备的 AI 云管理服务能力体系，填补了大模型应用交付落地过程中“最后一公里”的生态闭环缺口，正逐步成为企业智能化转型进程中的核心枢纽。

AIMSP 专注于提供智算基础设施全栈交付服务，打通智算中心从规划建设到商业化运营的完整链条。当前，算力资源碎片化、跨

架构调度效率低下以及运营成本居高不下等问题，已成为行业内普遍面临的难题。AIMSP 创新性地构建了“资源适配、效能优化、运维运营”三位一体的全栈解决方案，实现了从基础设施交付到商业价值转化的闭环式管理。在资源适配与容量规划方面，AIMSP 借助异构硬件兼容性验证和场景化建模分析技术，系统性地规划算力资源适配方案，并预留具备弹性的冗余空间，以此有效满足 AI 业务动态扩展的需求。中国电信、中国移动等通信运营商，通过开展异构硬件兼容性验证并构建弹性算力池，为所属区域内的客户提供了灵活且可靠的算力支持。在资源效能优化方面，AIMSP 打造了“硬件加速-系统调度-算法协同”的三层调优体系，深度融合异构计算核心加速策略与智能调度算法，构建起跨层联动的资源效能模型，确保算力供给能够与 AI 业务需求实现动态平衡。例如，阿里云、华为云、腾讯云等厂商在智算集群交付相关项目中，通过动态调整资源池划分策略，充分释放算力资源效能，从技术维度实现降本增效。在资源运维运营支撑方面，AIMSP 依托智能运维平台，构建起风险预警和快速响应机制，借助多级保障体系实现故障精准定位和业务连续性管理，建立起以运维数据为驱动的资源迭代优化闭环。例如，重庆电信在某区域型智算中心建设项目中，借助定制化运维平台搭建面向 GPU 资源及光模块等硬件设备的全链路监控体系，将故障定界定位周期缩短至 40%。

AIMSP 着力构建数据治理与知识工程全流程服务体系，为行业

领域专属模型开发提供高质量云数据集治理服务。当前，数据质量良莠不齐、知识关系梳理难度大、数据与模型迭代优化脱节等问题，严重制约了行业模型的发展。AIMSP 提供数据治理标准化、知识图谱工程化、数据智能分析等核心能力，实现了对行业领域专属模型开发数据集赋能。

一是数据治理与工程化处理方面，AIMSP 提供覆盖多源数据采集、智能清洗标注、分层存储架构搭建以及自动化特征工程的全流程服务，为后续的数据分析和模型构建筑牢基础。例如，北京电信基于行业数据智能标注基地，推出一套完整的高质量数据集生成与智能标注服务，集成了数据标注、特征工程、云端存储管理等多项功能，具备自定义标注模板、领域知识库构建以及多模态数据处理等能力，有力保障了数据的高质量和可用性。

二是行业知识库构建与管理方面，AIMSP 依据领域数据构建结构化知识图谱，并运用自然语言处理、信息抽取等技术，对知识关系进行精细化管理，进而提升模型的智能化水平和决策准确性。例如，谷歌云依托 Vertex AI 平台打造的端到端数据处理流水线，能够快速构建包含产品目录、客户画像等内容的领域知识库，助力企业高效构建专属模型。阿里云通过知识图谱平台完成建模，一站式提供高质量知识管理专业技术服务，保障客户数据安全。

三是数据可视化与智能分析方面，AIMSP 建立覆盖数据可视化与决策支持的全链路服务体系，通过集成趋势分析、异常检测及预测算法模型，打通从数据洞察到迭代优化的闭环流程。例如，Tableau、Power BI、帆软等厂商借助

基于智能化 BI 工具为行业客户提供自动化数据处理、可视化分析的一站式服务，显著提升大数据分析模型训练的效率，增强业务决策准确性。

AIMSP 提供大模型全栈工程化交付服务，加速推动企业 AI 应用实现场景化的端到端落地。Gartner 在年度技术成熟度曲线报告中多次提及“AI 工程化”，并将其列为重要的战略技术趋势之一。AIMSP 通过标准化的工程流程和智能化的工具链，对 AI 云大模型工程化交付体系进行重构，有效攻克了模型开发周期长、部署成本高以及跨团队协作效率低等行业难题。一是模型咨询与规划方面，AIMSP 通过开展需求调研、场景可行性分析以及技术成熟度评估，为客户提供 AI 场景拆解、模型选型设计、数据治理规划以及算力资源匹配的体系化咨询方案，助力客户规划 AI 云能力建设的落地实施路径。例如，阿里云、亚信科技等均推出了基于云平台的大模型工程化交付专业技术服务，涵盖从行业需求诊断、模型架构设计到算力资源优化的全流程规划。二是模型集成与部署方面，AIMSP 依据客户的技术架构和业务场景需求，提供模型适配调优、异构环境部署、工程化封装以及全链路测试服务，确保 AI 模型与业务系统能够高效协同运行。例如，华为云基于昇腾 AI 基础软硬件栈，为制造企业提供跨平台部署解决方案，实现了模型在边缘设备与云端服务器之间的高效协同推理。例如，浪潮云依托海若 AgentStore 平台，针对政企客户多模态算力适配、数据安全合规等需求，提供云端模型私有化部

署、异构环境兼容及全流程安全管控服务。三是模型运营与优化方面，AIMSP 构建了模型全生命周期的智能化运营管理体系，提供监控告警、迭代更新、资源调度以及成本管控等核心服务，实现了模型版本迭代自动化、算力资源动态优化以及推理性能实时监控。例如，微软、百度云、腾讯云等厂商通过构建基于 AI 云的自动化监控、动态资源调度、智能模型迭代及多技术融合的运营体系，实现了模型性能优化、成本降低与效率提升的闭环管理。

四、发展展望

2025 年，“十四五”规划步入收官之年，数字中国建设持续稳步推进。云计算作为数字产业发展的重要底层支撑，正从单一的基础资源供给，向智能时代的核心引擎加速转变，逐渐成为重塑全球数字竞争格局的战略性基础设施。展望“十五五”，云计算将与人工智能深度融合，推动数字技术迈向系统性变革，形成智能化、场景化、全栈化的技术底座。聚焦云计算产业发展，将呈现以下特征：

市场规模方面，人工智能助力云计算保持快速增长态势。随着人工智能发展，云端大模型推理和训练需求持续高涨，AI 大模型用云的交付形态将发生根本性变革，从传统算力租赁转向模型即服务模式。同时，智能体服务等新兴商业模式的出现，也将带动云计算上下游产业链协同发展。预计在“十五五”期间，在人工智能大模型拉动下，我国云计算市场规模仍将保持 20%以上的增长。

服务模式方面，云智融合催生云计算全新服务体系。随着智能

计算逐渐成为算力结构的主要组成，传统的通用云计算正加速与智算融合，升级成为可服务于人工智能技术和应用发展的智算云，推动云计算服务体系向“AI+”演进，促使全新的智算云服务模式应运而生。预计在“十五五”期间，AIaaS、AIPaaS、MaaS、AISaaS、AIMSP将逐渐成为云计算主流服务模式，共同助力智能服务的普惠应用。

行业应用方面，云计算+人工智能深度赋能千行百业。目前，我国工业制造、医疗、交通等行业已广泛应用云计算和人工智能技术，实现了业务流程优化与效率提升。随着智能体、边缘计算、大模型等新技术的不断突破，预计在“十五五”期间，云计算和人工智能将不再局限于单点应用，而是深度融入行业全链条，从研发设计、生产制造到运营管理，推动各行业向数智化全面升级。

面向“十五五”时期，云计算将深度融入数字经济主战场，加速整合科技创新资源，为社会发展、产业变革以及人类生产生活注入强大动力，助力国家筑牢在全球数字经济领域的竞争新优势。

中国信息通信研究院

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62300072

传真：010-62304980

网址：www.caict.ac.cn

