亿欧智库 https://www.iyiou.com/research

2025年人工智能算力创新知识产权研究报告

亿欧智库 EO Intelligence

研究报告

目录

CONTENTS

01 人工智能算力创新知识产权现状

2025年，全球人工智能大模型快速发展，人工智能算力需求快速增长，**算力成为当前AI产业发展中备受关注的战略资源。**

算力从技术概念到产业落地需跨越“创新-转化-盈利”的环节，**知识产权正是串联此环节的关键纽带。**

本章以全球视野浅析各国的算力战略部署情况，基于此，分析**全球以及中国的算力知识产权布局态势。**全球人工智能算力专利申请总量已超 114 万件，而中国复合年增速高达 14.72%，展现出强劲的创新动能。

1.1 人工智能算力：驱动数字经济发展的新动能

◆ **人工智能算力是指支撑人工智能算法运行所需的计算资源与处理能力，是衡量计算设备或系统在人工智能任务中性能水平的核心指标。**随着AI大模型技术的持续演进，人工智能算力不仅构成了AI模型训练与应用的物理基础，也成为了驱动数字经济发展的新动能。

技术革新驱动算力需求快速增长

- 规模法则揭示了模型性能随规模提升而增强，降低了研发不确定性，并验证了“大力出奇迹”的有效性。
- 随着人工智能技术高速演进，模型已从小规模算法扩展到千亿、万亿级参数，并由单一文本延展至图像、音频、视频等多模态数据，不断推高算力需求。
- IDC数据显示，2024年全球加速计算服务器市场销售额同比增长160.4%至1317.6亿美元，出货量同比增长38.9%至217.5万台。

算力是数字经济发展的“动力源”

- 人工智能算力基础是赋能千行百业智能化转型的基础动力。随着产业数字化转型加速、人工智能应用场景不断拓展，新兴应用持续涌现。
- 此背景下，数字经济迎来蓬勃发展。IDCA（International Data Center Authority）数据显示，数字经济在2024年约占全球GDP的15%，约合16万亿美元。

图1-1 技术革新推高算力需求，加速服务器市场扩张

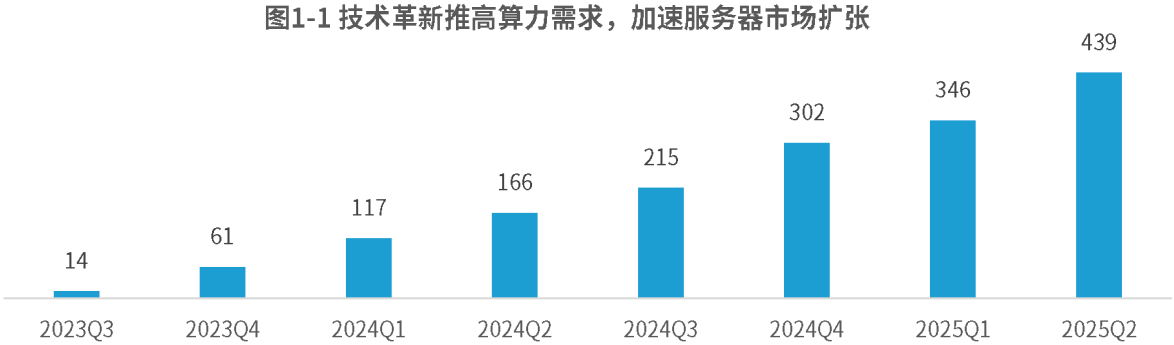

图1-2 中国完成备案并上线提供生成式人工智能服务的大模型数量

◆ 目前，全球各个国家和地区正通过强有力的政策支持或投资激励推动人工智能算力技术研发和产业应用，促进算力产业蓬勃发展。

美国

“星际之门”国家级计划预计将投入5000亿美元用于人工智能算力基础设施建设。

日本

2024年宣布设立规模超过650亿美元的投资基金，用于支持芯片和人工智能行业的发展。

中国

深入实施“人工智能+”行动中指出强化智能算力统筹，支持人工智能芯片攻坚创新与软件生态培育，加快超大规模智算集群技术突破和工程落地。

欧盟

推出一揽子人工智能创新计划，其中包括计划在2025年初启动首批人工智能工厂的建设，为生成式人工智能模型开发提供计算、存储和数据等服务。

图1-3 人工智能算力政策支持或投资激励情况

数据来源：网信办、工信部、亿欧智库

1.2 全球人工智能算力创新知识产权布局态势

◆ OpenAI研究报告显示，自2012年以来，最大规模人工智能训练任务所使用的计算力呈指数级增长，平均每3.5个月翻一番，其增长速度远超摩尔定律所设定的18个月翻倍周期。**人工智能算力已不再只是传统计算资源的补充，而是逐渐演进为支撑数字世界智能化转型、激发数据价值裂变的基石与核心生产力。**

◆ 人工智能算力的发展水平直接决定了数字经济系统能实现的复杂程度与智能高度。知识产权是对算力领域技术创新成果的法律保护与价值锚定。算力的每一次突破都依赖技术创新，知识产权则为技术创新筑牢“护城河”，既保障研发者权益以激发持续创新活力，又通过成果的流转与共享，推动算力技术在全球范围高效转化与应用，成为算力从技术优势迈向产业优势的纽带。因此，**探究全球算力知识产权发展情况，是理解算力产业创新生态与竞争格局的核心切入点。**

◆ 迄今为止，人工智能算力全球有效发明专利超过114万件[注1]，其中有效授权发明专利超过94万件，占比82%；审查中的专利数量超过46万件，表明当前仍然有较大的潜在有效专利储备，3-5年内人工智能算力有效发明专利储备有望突破150万件。美国和中国是当前人工智能算力技术创新和产业应用的主要国家，算力发明专利申请量居全球前二，分别占比28%和25%。

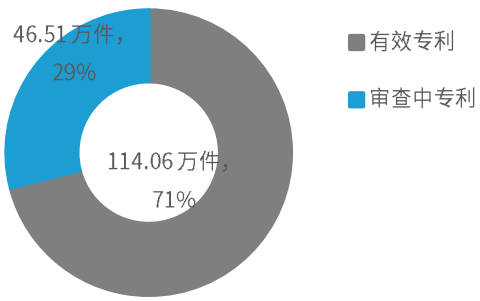

图1-4 全球人工智能算力专利数量情况

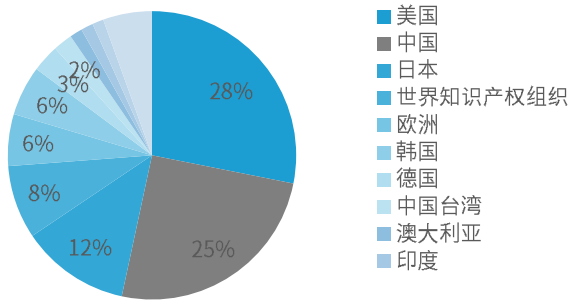

图1-5 全球算力发明专利申请分布

◆ **全球申请趋势上，人工智能算力发明专利申请量逐年提升[注2]，复合增长率为6.3%，技术创新热度持续高涨。**2020年，OpenAI发布GPT-3，让全球看到了AI的巨大商业潜力，各大科技公司和初创企业纷纷投入资源构建自己的算力底座，当年算力发明专利申请增长率达到最大的10%。2023年，全球年度算力发明专利申请量已超过20万件。

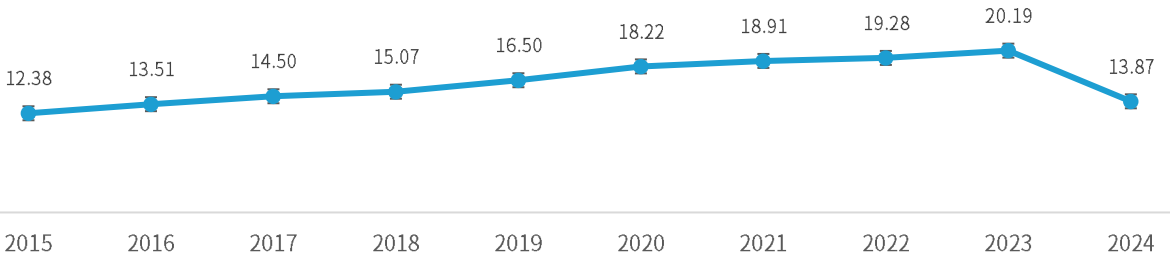

图1-6 全球人工智能算力技术发明专利申请趋势(单位：万件)

注1：人工智能算力的性能主要由底层基础设施硬件决定，因此本报告研究的专利以人工智能算力基础设施硬件为主，涵盖AI芯片、AI服务器、存储、网络等相关技术
注2：2024年数据量下降的原因是专利公开滞后，下文同
注3：根据专利关键词选取口径的不同，检索结果可能存在差异。

数据来源：公开专利数据库、亿欧智库

1.3 中国人工智能算力创新知识产权布局态势

◆ **中国紧跟全球态势，在人工智能算力创新知识产权布局上积极作为。**政策引导与企业投入并重，专利数量增长，布局日益完善。中国人工智能算力发明专利申请趋势上，**申请量逐年提升，复合增长率为14.72%，**远超全球专利申请增速，尤其是2020年专利申请增长率达到22%。2023年，中国年度算力发明专利申请量已超过8万件。

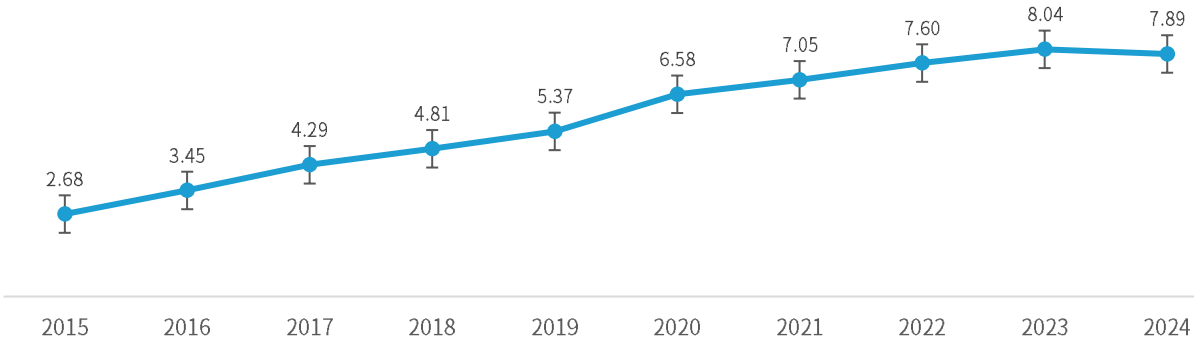

图1-7 中国人工智能算力技术发明专利申请趋势(单位：万件)

◆ 国内的有效授权发明专利中，**形成了以浪潮信息、华为和联想为核心的第一梯队。**

◆ **浪潮信息**致力于发展新一代以系统为核心的计算架构，打造开放、多元、绿色、智能的元脑智算产品和方案，为人工智能的创新和应用，提供强大算力平台。基于开放架构设计，业界率先实现“一机多芯”，引领多元算力生态共进。

◆ **华为**致力于打造基础软硬件平台，提供“鲲鹏+昇腾”多样性算力，通过算、存、网深度融合创新，打造坚实算力底座；坚持硬件开放、软件开源的策略，携手合作伙伴构建开放的计算产业生态。

◆ **联想**致力于构建全栈式计算能力，打造了涵盖服务器、存储、边缘计算及高性能计算的全栈式AI基础设施，为AI应用提供强大算力支撑，推动混合式人工智能的发展。

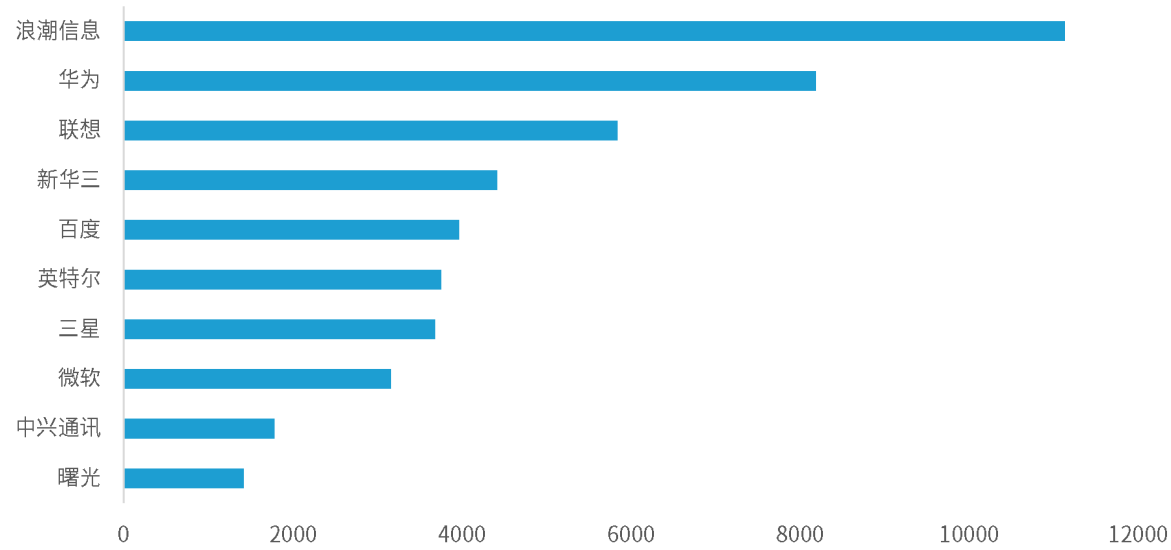

图1-8 中国人工智能算力有效发明专利数量排名（单位：件）

注：根据专利关键词选取口径的不同，检索结果可能存在差异。
数据来源：公开专利数据库、亿欧智库

02 中国人工智能算力创新关键技术专利分析

DeepSeek R1系列模型的发布推动产业摆脱唯算力规模论，促使人工智能算力发展从硬件堆叠的规模扩张阶段迈入增效提质阶段。新的**算力技术创新模式将聚焦芯片级性能突破和系统架构整合优化两大维度，形成以芯片为基础硬件，数据中心为终端的全栈式创新格局。**

此格局下，以**系统架构整合优化的数据中心为大规模训练与推理提供主要支撑**。系统架构各个层面的核心技术优化与协同将决定整体算力的能效表现。基于此，本研究报告选取在系统架构创新的**五大关键技术（异构计算、互连技术、液冷技术、供电技术、系统管理与控制）**作为算力知识产权分析的切入点，以揭示产业升级的技术突破口，并提供专利布局的参考路径。

2.1 异构计算：革新人工智能算力架构范式

◆ 随着摩尔定律失效，传统CPU性能提升速度放缓，难以满足AI、大数据、图形处理等应用对于算力和能效需求的爆炸式增长，迫使计算架构转向异构计算。**异构计算成为极具潜力的发展方向**，它突破传统单一处理器架构局限，在单一系统中整合并协同运用CPU、GPU、ASIC、FPGA、NPU等不同类型处理器，构建多元化、高效能计算生态。异构协同模式能依据任务性质与需求，精准调配计算资源，实现性能与效率双重优化，全方位满足多样化计算需求。

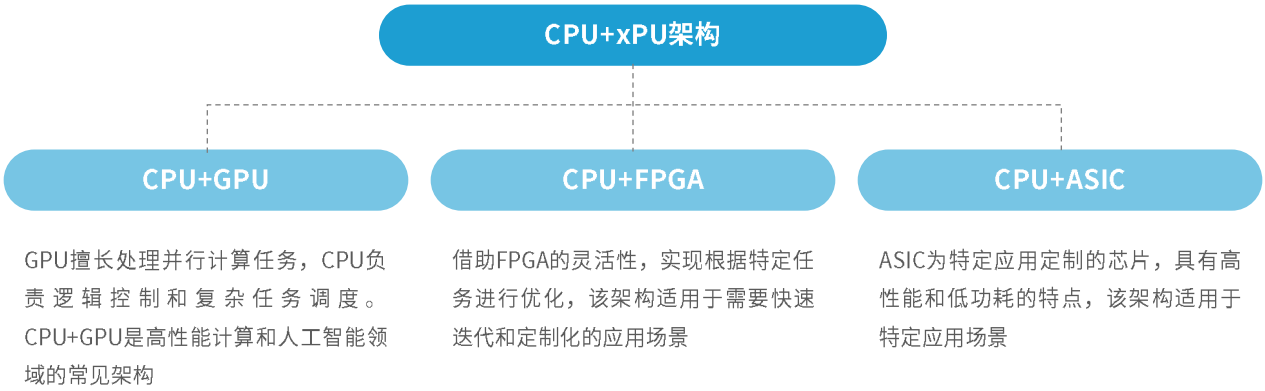

图2-1 人工智能领域主流异构计算架构

◆ 当前，中国市场“CPU+GPU”的异构计算方式是人工智能异构计算主流组合，广泛应用于各类场景，推动行业技术进步创新。中国异构计算发明专利申请呈持续上升态势，2020年专利申请增长率达到最大的30%，2023年中国异构计算发明专利年度申请数量超过1.6万件。

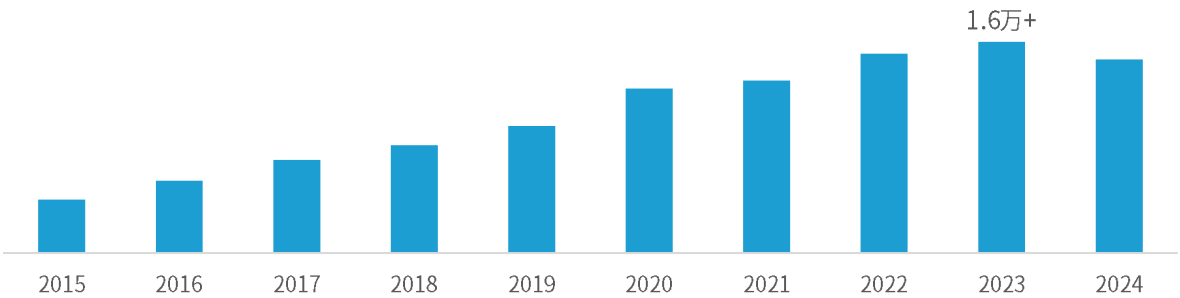

图2-2 中国异构计算发明专利申请趋势

◆ 在异构计算方向的有效授权发明专利数量排名中，**浪潮信息和英伟达分别位居中国第一和第二。**

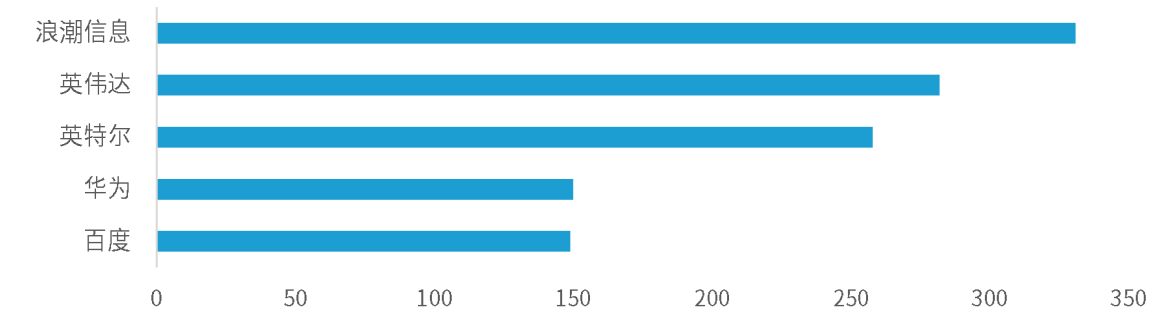

图2-3 中国异构计算有效授权发明专利排名

注：根据专利关键词选取口径的不同，检索结果可能存在差异。
数据来源：公开专利数据库、亿欧智库

2.1 异构计算：革新人工智能算力架构范式

◆ 浪潮信息以“硬件重构+软件定义”的融合架构为核心，实现计算、存储、内存和加速资源的彻底解耦与池化。英伟达深耕芯片架构、先进网络和CUDA并行计算模型，推动全球数据中心现代化改造，使吞吐量实现指数级增长，并降低能耗和成本。

浪潮信息：久久为功，主张软硬协同

浪潮信息凭借“硬件重构+软件定义”的融合架构及池化解耦设计大幅提升大模型推理性能

- **2014年，浪潮信息就超前提出了“硬件重构+软件定义”的融合架构**，这一理念的核心是通过硬件资源的解耦与重构（如计算、存储、网络资源的池化）和软件定义的智能调度，实现数据中心的灵活性和高效性。
- **2023年，浪潮信息发布了融合架构3.0原型系统**，以开创性的系统架构设计实现了计算资源、存储资源、内存资源、异构加速资源等核心IT资源彻底解耦与池化，支持池化资源异步升级、支持细粒度多主机共享高并发存储、亚微秒级远端内存共享访问等特性，可通过软件定义实现“一套系统，N类应用”。
- **2025年2月，浪潮信息推出了专为大模型推理设计的高吞吐推理服务器——元脑R1推理服务器NF5868G8**，业界首次实现单机支持16张标准PCIe双宽卡，基于创新研发的PCIeFabric的16卡全互连拓扑，实现任意两卡P2P通信带宽可达128GB/s，降低通信延迟超60%。

英伟达：CPU+GPU异构架构，打造极致性能

英伟达新一代的Blackwell架构GPU打破生成式AI和加速计算的壁垒，使业界能够基于此构建万亿参数级大语言模型和运行实时生成式AI；Grace CPU具有超强的性能和效率，可以与GPU紧密结合以增强加速计算能力，也可以作为强大而高效的独立CPU进行部署，为现代数据中心带来突破性的CPU性能和效率。

- 随着模型的复杂性激增，人工智能对加速计算和能源效率提出了更高要求。英伟达推出的NVIDIA GB200 NVL72、NVIDIA GB300 NVL72等产品，借助NVIDIA Grace CPU 和 NVIDIA Blackwell GPU 开启计算新时代。
- NVIDIA GH200 Grace Hopper超级芯片等产品实现大规模AI和HPC应用。

数据来源：公开专利数据库、亿欧智库

2.2 互连技术：驱动算力集群迈向高效协同

◆ 随着人工智能的发展，单机硬件性能逐渐逼近物理极限，大模型对于算力芯片的大规模互连提出了高带宽、低时延、高可靠的要求。在**此背景下，Scale-Up（纵向拓展）和Scale-Out（横向拓展）两种互连架构快速发展。**

◆ **Scale-Up升级单节点硬件资源（如CPU、内存、存储等），**侧重于强化单体设备的处理能力，并满足AI训练推理时的高带宽和低时延需求。Scale-Up的核心在于突破了传统网络架构在带宽、时延、可靠性上的瓶颈。在技术多样性、场景差异化、生态竞争及标准化需求等多重因素驱动下，行业内涌现出诸如PCIe、UALink、OISA、 NVLink、CXL等互连协议。

表2-1 Scale-Up互连协议对比

维度	PCIe	UALink	OISA	NVLink	CXL
定位	通用I/O互连标准	加速器间的高性能互连标准	开放GPU互连协议体系	高速GPU互连协议	动态多协议标准
发起者	PCI-SIG 行业标准组织	UALink联盟	中国移动等	NVIDIA	CXL联盟
技术本质	点对点连接分层协议架构	分层协议栈+UPLI接口分离式通信模型	开放GPU卡间互连协议	NVIDIA专有高速GPU互连协议	基于PCIe构建缓存一致性和内存共享机制
特性	高带宽、低延迟、灵活扩展	高带宽、低延迟、开放生态	原生共享内存语义、高带宽、低延迟、开放性	极高带宽、稀疏直接拓扑、多链路聚合	缓存一致性 内存共享与池化 高带宽与低延迟

◆ **Scale-Out强调分布式弹性扩展，通过增加集群内的节点数量提升系统整体的硬件性能与容错能力。**在云计算、大数据处理、分布式数据库等场景中，Scale-Out架构通过多轨多平面拓扑连接实现快速扩容，并根据负载分担和拥塞控制算法达到高带宽和低时延。

◆ **RDMA是Scale-Out架构中节点间高速通信的核心技术支撑。**RDMA具有“零拷贝”直接内存访问、绕过CPU内核和减少协议栈处理等技术优势，可将节点间通信延迟降至微秒级，带宽提升至数百Gbps，同时释放CPU资源用于计算任务。RDMA存在多种协议实现，如InfiniBand、RoCE、iWARP等，需根据场景选择适配协议。

InfiniBand

- 专为RDMA设计的网络，从硬件级别保证可靠传输，具备更高的带宽和更低的时延。
- 需要配套IB网卡和IB交换机。

RoCE

- 基于以太网的RDMA，允许在以太网上实现高性能、低延迟、高吞吐量的数据传输。
- 无需使用特殊的网络硬件设备。

iWARP

- 基于TCP的RDMA网络，利用TCP达到可靠传输。
- 相比RoCE，大量TCP连接会占用大量的内存资源，对系统规格要求更高。

图2-4 Scale-Out互连协议对比

数据来源：公开专利数据库、亿欧智库

2.2 互连技术：驱动算力集群迈向高效协同

◆ **在中国，众多企业在互连技术领域积极开展技术创新并推动标准制定。**中国互连技术发明专利申请呈持续上升态势，2023年中国互连技术发明专利年度申请数量超过1.4万件。

◆ 由中国移动牵头推出的OISA（Omni-directional Intelligent Sensing Express Architecture，全向智感互联）协议，旨在提供GPU卡间高速互连标准，包括大规模GPU对等互连、极致报文格式、数据层流控和重传以及高效物理传输等解决方案。OISA Gen1支持128张GPU通过8个Switch芯片互连，任意卡间互连带宽达到800GB/s，每个Switch芯片支持128个端口、交换容量达到51.2Tb/s。

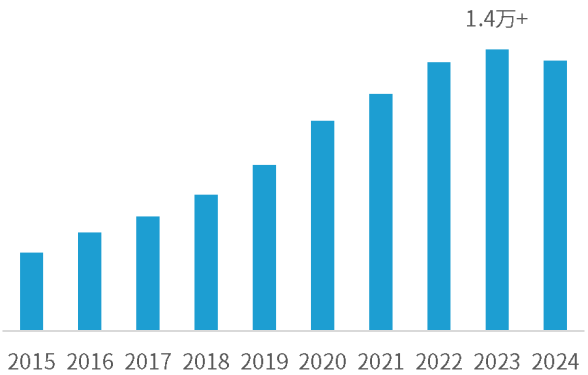

图2-5 中国互连技术发明专利申请趋势

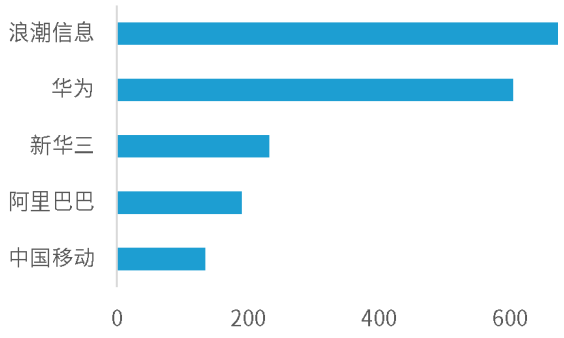

图2-6 中国互连技术有效授权发明专利排名

◆ **在互连技术方向的有效授权发明专利数量排名中，浪潮信息和华为分别位居中国第一和第二。**

浪潮信息：超节点AI服务器“元脑SD200”

2025年8月，浪潮信息发布面向万亿参数大模型的超节点AI服务器“元脑SD200”。

- 该服务器基于自主研发的开放总线交换技术首创多主机三维网格系统架构，实现64路本土GPU芯片高速互连。
- 通过创新远端GPU虚拟映射技术，突破多主机交换域统一编址难题，实现显存统一地址空间扩增8倍，单机可以提供最大4TB显存和64TB内存，为万亿参数、超长序列大模型提供充足键值缓存空间。
- 依托百纳秒级超低延迟链路，构建64卡大高速互连域统一原生内存语义通信，显著提升计算与通信效率。

华为：CloudEngine 16800系列交换机

华为作为全球领先的信息与通信技术解决方案供应商，其在服务器行业凭借全栈自研技术与生态整合能力，已成为全球重要的服务器和数据中心基础设施供应商。

- **华为提供的CloudEngine 16800系列交换机专为大规模GPU集群设计**，实现网络0丢包与E2Eμs级延时，整机最大支持9642Tps交换容量，支持GE、10GE、25GE、40GE、100GE、400GE不同速率互连互接；同时内嵌AI芯片，搭载iLossless算法，可对全网流量进行实时学习训练，动态调整ECN阈值，实现网络自适应优化。

注：根据专利关键词选取口径的不同，检索结果可能存在差异。
数据来源：公开专利数据库、亿欧智库

2.3 液冷技术：助力人工智能算力绿色低碳转型

◆ **液冷技术已成为新型数据中心建设的能效最优解。**随着AI芯片功率密度的持续攀升，传统风冷散热已无法满足高密度算力需求，液冷技术凭借传热效率比空气高6倍、蓄热量高1000倍的散热优势，正在成为主流散热方案。另一方面，在“双碳”战略的大环境下，数据中心绿色低碳和可持续发展也成为“不可逆”的大趋势。

◆ **液冷技术可分为非接触式液冷和接触式液冷两大类。**非接触式液冷主要指**冷板式液冷**，相比传统风冷散热技术可实现60%~90%的能耗降低；接触式液冷包括**浸没式液冷**和**喷淋式液冷**，其中浸没式液冷可完全去除散热风扇，换热能力强，节能效果好。

表2-2 液冷技术对比

维度	冷板式液冷	浸没式液冷	喷淋式液冷
技术架构	包括冷板、CDU、一次/二次循环系统	整机浸入绝缘冷却液腔体	冷却液以喷淋形式直接作用于热源
工作原理	冷却液在冷板流道内与CPU/GPU等部件换热，再通过换热器带走热量	冷却液与所有部件直接接触换热，支持相变换热	通过蒸发吸热散热
改造/维护难度	低，易维护	高，需整体停机维护	中等
服务器渗透率	高	中	低
典型PUE值	1.05–1.15，全液冷可至1.05	1.05–1.10	1.08–1.15
应用场景	适用于高功率密度机柜部署或对现有设施改动小的场景	适用于高功率密度和极限散热需求的机柜或新一代AI服务集群场景	适用于对散热效果要求较高，但又无需大改整体架构的场景

液冷技术存在的瓶颈和未来发展方向

冷板式液冷

存在的瓶颈：冷板式液冷存在成本较高、通用性较差、技术稳定性差的问题。
未来发展方向：1、提高方案经济性；2、提高通用性和可维护性；3、降低系统故障风险，保障安全运行。

浸没式液冷

存在的瓶颈：浸没式液冷存在系统复杂性和运维难度高的问题。液冷剂回收要求较高。
未来发展方向：1、降低液冷介质成本；2、提升维护便携性；3、完善循环和密封设计。

喷淋式液冷

存在的瓶颈：喷淋液冷技术有液滴分布控制、喷淋均匀性以及液体回收循环复杂的缺点。
未来发展方向：1、提升喷淋的精准性；2、优化堵塞与防泄漏设计；3、设计封闭式循环系统。

◆ **中国液冷技术发明专利申请数量增长迅猛。**发明专利申请量从2015年的400+增至2023年的2900+，增长超过6.5倍。数据表明液冷技术越来越成为未来技术创新和产品落地的发展趋势。

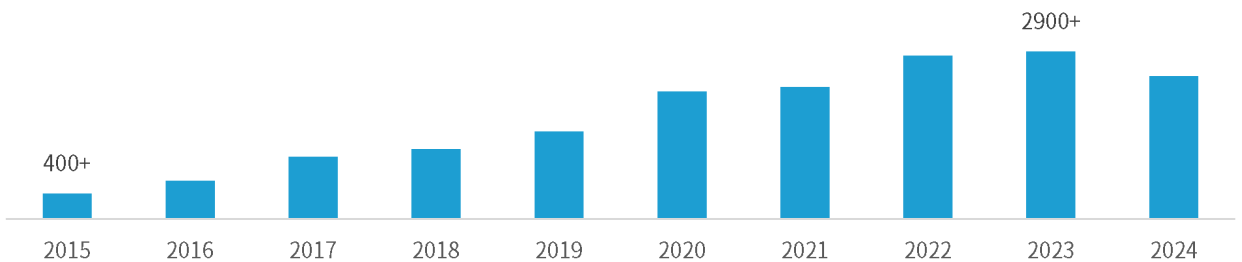

图2-7 中国液冷技术发明专利申请趋势

注：根据专利关键词选取口径的不同，检索结果可能存在差异。
数据来源：公开专利数据库、亿欧智库

2.3 液冷技术：助力人工智能算力绿色低碳转型

◆ **在液冷技术方向的有效授权发明专利数量排名中，浪潮信息和百度分别位居中国第一和第二。**浪潮信息提出"All in 液冷"战略，围绕液冷技术的安全性、高密度部署、节能降耗、运维便捷性四大核心价值，构建了系统化的"专利群落"布局。百度在液冷技术领域推出了全球首款支持OCP OAI标准和液冷的AI计算平台X-MAN 4.0。

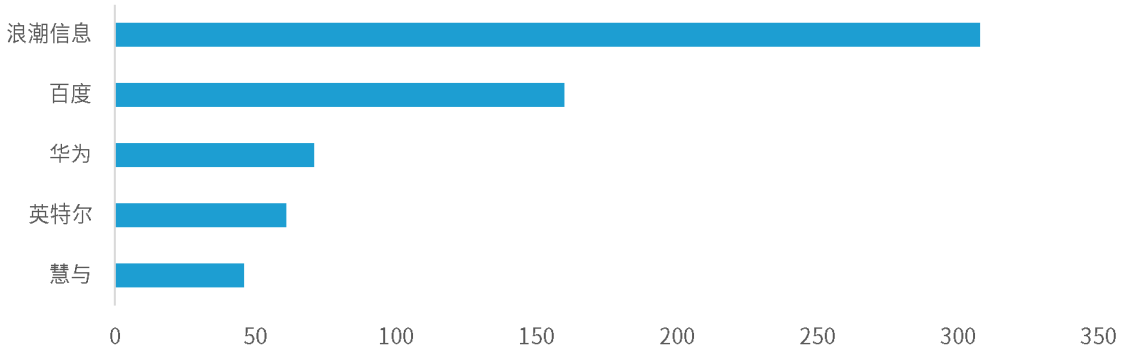

图2-8 中国液冷技术有效授权发明专利排名

浪潮信息：“All in 液冷”战略

浪潮信息元脑服务器第八代平台凭借全栈液冷能力，为节点级、机柜级、数据中心级提供全方位的液冷整体解决方案。

- **在服务器节点方面，**创新开发负压式液冷循环系统，围绕负压循环、排气除湿、防汽化、漏液安全等方面实现技术突破。射流式冷板引入喷射湍流强化换热的理念，较传统冷板而言，换热效率提升25%以上；两相冷板系统解决单路串联均温、多通路流量分配、热流密度极限等难题，为未来AI高密服务器散热提供新的设计思路。
- **针对机柜级液冷，**全新一代元脑液冷整机柜服务器ORS6000G8基于OCP开放标准设计，支持供电、供液盲插，极大简化安装与运维流程。
- **面向数据中心液冷，**元脑服务器第八代平台可搭配智能CDU（冷量分配单元）控制系统，实现冷量和电量的实时按需分配，有效降低CDU的耗电量，助力液冷数据中心实现PUE无限接近1。

百度：AI计算平台X-MAN 4.0

百度在液冷技术领域推出了全球首款支持OCP OAI标准和液冷的AI计算平台X-MAN 4.0。

- 该平台单节点支持8个AI加速器，通过冷板式液冷散热技术显著提升了高密度计算场景的能效，并推动了AI硬件加速模块与系统的标准化设计。
- 百度发布的《天蝎4.0液冷整机柜开放标准》为冷板液冷方案提供了行业规范，其自研的液冷冷源系统专利通过预制化制冷单元与辅助单元的协同设计，优化了冷却介质质量，进一步降低了液冷系统的能耗与成本。

注：根据专利关键词选取口径的不同，检索结果可能存在差异。
数据来源：公开专利数据库、亿欧智库

2.4 供电技术：重塑能源应用，构筑算力基石

◆ **服务器机柜更新迭代加速、性能持续提升的同时，系统能耗激增，亟需供电技术革新。**目前，单卡GPU功耗在250W到1200W之间，服务器单机柜功率从原来10kW攀升至100kW以上，未来将达到1MW以上。这种高功率密度需求对电源效率、供电架构等方面都提出了前所未有的挑战。

功率半导体逐渐成熟

第三代功率半导体SiC（碳化硅）与GaN（氮化镓）逐渐成熟，并凭借其高压、高频、高温下的卓越表现，已成为AI电源设计的新核心。

- 与传统硅器件相比，SiC/GaN具备更高开关频率、更低导通损耗，为实现高功率密度、高效率、小体积提供了关键突破口，尤其适用于HVDC、PFC、DC-DC多级拓扑的高效转换。

供电架构重塑

OCP ORv3等标准推动PSU从服务器节点“解耦”，以集中供电的方式部署至整机柜，形成机柜级别的大功率供电系统。

- 第一代5.5kW PSU，如今正向第三代22kW+演进，输出电压从48V提升至400V，大幅降低传输损耗与电缆体积。集中供电不仅让散热与布线更高效，也更适应AI服务器对功率密度与可维护性的极限要求。

电压平台迎来根本转变

AI服务器集群对能效的极致追求，推动高压直流（HVDC）方案快速普及。

- 相较传统UPS架构的90%-95%系统效率，HVDC系统效率跃升至95%-98%，不仅节能显著，更减少了空间与运维成本；并且HVDC系统全链路故障点数量减少，供电可靠性显著增强。

图2-9 供电技术主要发展趋势

◆ **近十年间，中国在人工智能算力基础设施供电技术方向发展迅速，相关企业积极开展专利布局。**2024年，中国供电技术方向发明专利申请数量已超过1.2万件，较2015年增长了3倍。

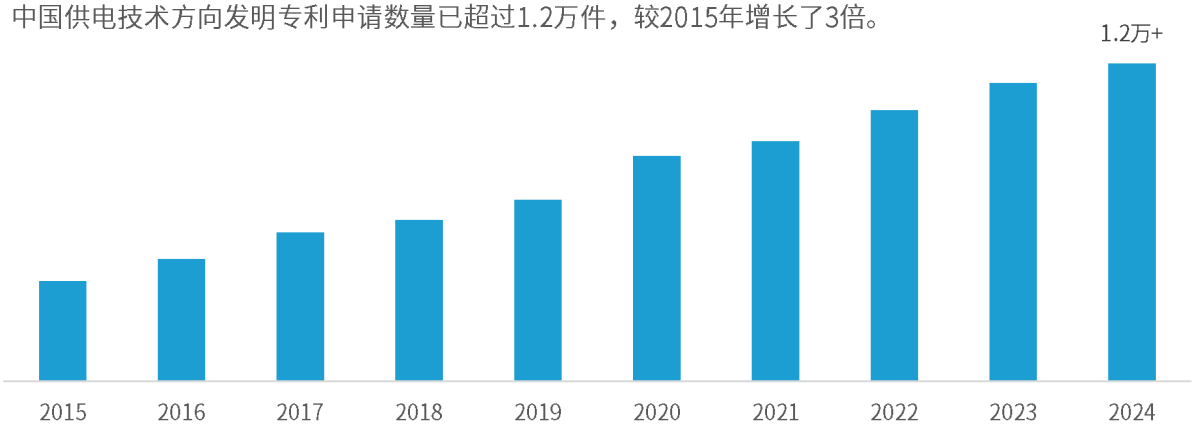

图2-10 中国供电技术发明专利申请趋势

注：根据专利关键词选取口径的不同，检索结果可能存在差异。
数据来源：公开专利数据库、亿欧智库

2.4 供电技术：重塑能源应用，构筑算力基石

◆ **在供电技术方向的有效授权发明专利数量排名中，浪潮信息和台达电子分别位列第一和第二。**浪潮信息的供电技术从单个服务器的电源模块延伸到了整个数据中心的供电架构、能效管理和故障应对。台达电子深耕芯片、服务器和数据中心等供电产品及解决方案，有效提升了算力基础设施的供电效率和可靠性。

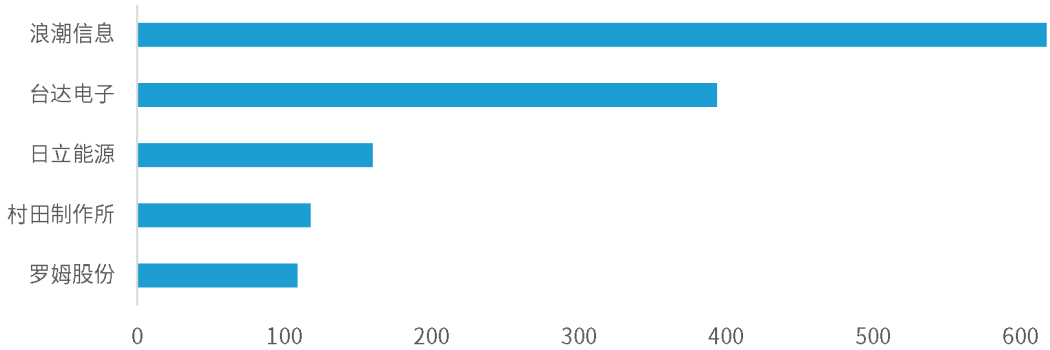

图2-11 中国供电技术有效授权发明专利排名

◆ 在**浪潮信息**最新发布的供电解决方案中，其创新性地采用了高压直流HVDC替换传统UPS设备，从10kV交流电**只需1次交直流转换，实现高压直流一步到柜。**其具有更少转换次数，更方便新能源储能挂载的特点。机柜侧，供电系统采用高效功率半导体氮化镓，采用自研多模组并联的动态均流技术，支持高压直流供电，减少了交流到直流的转化，单柜即可支撑430kW-720kW的高效供电。基于此架构，一座30MW的智算中心每年可节省超千万元电费。

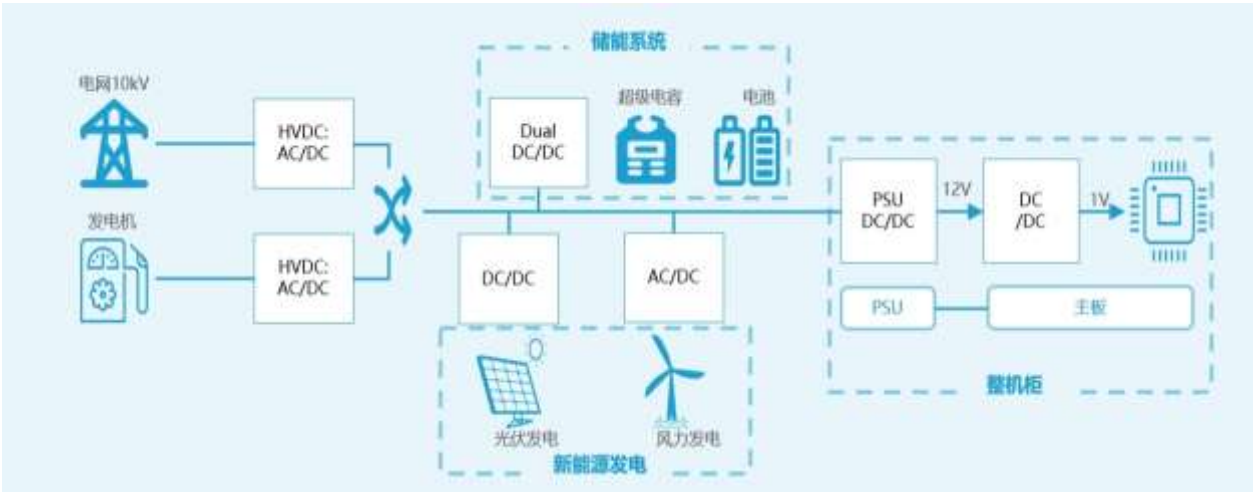

图2-12 HVDC（高压直流）电气架构

◆ **台达电子**作为全球领先的电力电子和热能管理解决方案提供商，其基于长久以来的研发投入和技术积累为AI服务器与数据中心用户提供“从电网到芯片”（From Grid to Chip）全方位解决方案。台达电子提出的“三高”电源方案（高效率、高密度、高功率），覆盖了不断电系统（UPS）、机架式电源（Power Shelf）、电池备援模块（BBU），机架式高功率电容模块（PCS）到板端的DC-DC转换器及功率电感（Power Choke）等电源供应及管理解决方案。

注：根据专利关键词选取口径的不同，检索结果可能存在差异。
数据来源：公开专利数据库、亿欧智库

2.5 系统管理与控制：破解多元算力管理难题

◆ **多元算力时代下，大规模的异构算力设备面临如何兼容多种处理器架构、多种设备协议、不同管理芯片的系统设计层的挑战。**在异构算力设备中，构建覆盖“硬件-软件”的RAS（可靠性、可用性、可维护性）体系，通过故障预警、容错设计、快速修复功能，提升多元算力集群的抗风险能力和运维效率尤为重要。

◆ **在计算机系统底层架构中，BMC（基板管理控制器）和BIOS（基本输入输出系统）是计算机系统管理与控制的重要组成部分。**BMC是在服务器中嵌入的复杂而独立SOC系统，是数据中心集中运维管理IT设备的核心组件，对服务器安全可靠运行、远程集中管理和控制部署至关重要。BIOS是刻在主板ROM芯片上不可篡改的启动程序，BIOS固件在计算机启动过程中发挥着不可或缺的作用，是提高系统稳定性、扩展硬件适配性、修复BUG的关键部分。二者协同支撑算力基础设施从启动到运行的稳定与高效。

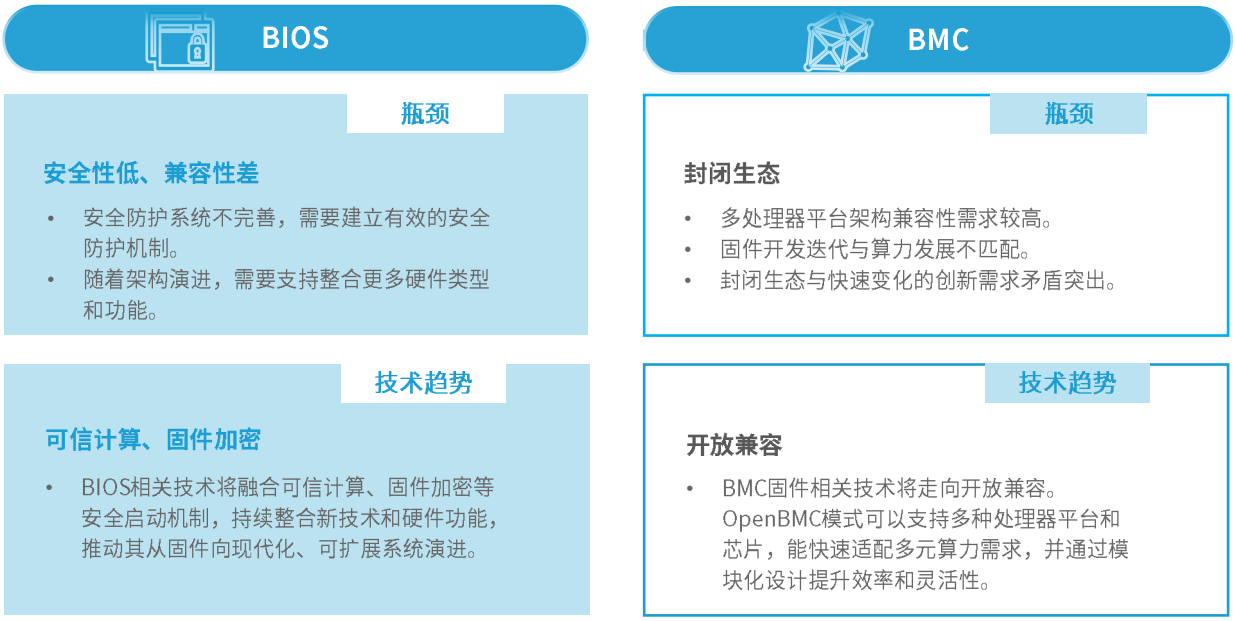

图2-13 系统管理与控制技术瓶颈与发展趋势

◆ **系统管理与控制方向的专利申请整体呈上升趋势，**2022年申请量超过3800件。在系统管理与控制方向的有效授权发明专利数量排名中，**浪潮信息和联想分别位居中国第一和第二。**

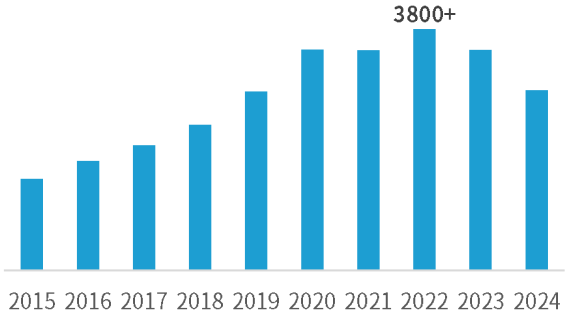

图2-14 中国系统管理与控制发明专利申请趋势

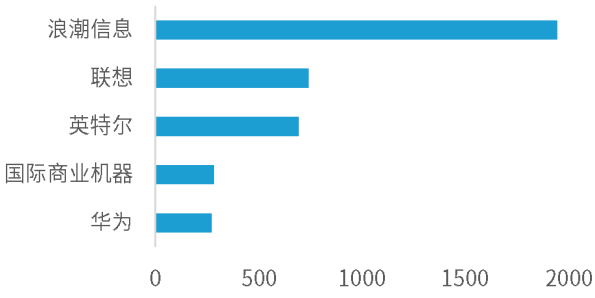

图2-15 中国系统管理与控制有效授权发明专利排名

注：根据专利关键词选取口径的不同，检索结果可能存在差异。
数据来源：公开专利数据库、亿欧智库

2.5 系统管理与控制：破解多元算力管理难题

◆ **更开放先进的BMC固件发展之路——OpenBMC。**传统BMC固件存在着诸多问题，且随着数据中心规模的不断增长，运维需求愈发朝着精细化、定制化的方向发展，业界开始探索更开放先进的BMC固件发展之路——OpenBMC顺势而生。OpenBMC是一个Linux基金会项目，其目标是为BMC生成一个可定制的开源固件堆栈。

◆ **浪潮信息在布局专利的同时，也在积极拥抱OpenBMC，且连续多年在OpenBMC社区开源代码贡献排名中保持全球第5位和中国第1位，**构建了稳定、可靠、安全的开放架构通用服务器产品矩阵。浪潮信息通过分层解耦、模块化设计的OpenBMC方案，在BMC层面实现了软硬件的标准设计，支持服务器产品的快速、稳定迭代，从而更快、更好的满足用户资产信息管理、故障预警、远程管理和批量自动部署等需求。近期，**浪潮信息升级元脑服务器智能固件管理平台InBry。**通过软硬件协同系统优化BMC固件架构，率先支持全球最新BMC双节点管理、以及协处理器多任务管理；平台新增智能内存管理、智能化散热调控、智能故障预警诊断等复杂服务器硬件管理功能，BMC固件管理运维效率提升65%，为大规模数据中心的云计算、大数据、人工智能等多样化的算力场景提供了更强兼容性、更高性能、更稳定、更智能的远程管理能力。

图2-16 元脑服务器智能固件管理平台InBry

◆ **联想服务器通过BIOS和BMC的协同工作，构建了一个从本地硬件初始化到远程智能监控的完整管理生态。**为了打造极致的可靠性，联想推出“双子星”BMC去耦设计，通过双BMC模块实现冗余，主备模块自动切换，确保管理可靠性；通过XClarity Essentials工具，管理员可本地或远程高效配置BMC，提升运维效率。联想ThinkSystem服务器采用XClarity Controller（XCC）作为新一代管理控制器，其整合服务处理器、超级I/O、视频控制器等功能，支持HTML5和可通过XClarity Mobile 进行访问等。

数据来源：公开专利数据库、亿欧智库

03 人工智能算力创新知识产权建议

随着国际竞争局势的加剧，人工智能算力技术创新需聚焦芯片级性能突破与系统级架构优化，保障核心技术安全；还需通过提升专利质量与全球布局，确保国际市场竞争力。

通过核心技术专利深耕、“开源开放+专利共享”生态协同，既能保障企业的核心竞争力，也能推动整个产业摆脱低水平重复建设，推动技术共享和标准统一，降低产业链适配成本，实现从算力规模领先到算力规模与算力生态双领先，最终赋能人工智能产业的高质量、可持续发展。

3.1 聚焦核心技术，构建专利防护

- **人工智能算力领域的技术创新热度已进入增长爆发期。**专利检索数据表明人工智能算力已成为全球人工智能产业竞争的“核心赛道”，相关企业对技术突破的优先级与知识产权保护的重视程度均持续上升，算力创新的“技术-专利”联动已成为未来该产业发展的核心逻辑之一。
- **人工智能算力的产业创新要聚焦“芯片级突破”和“系统级整合”两个层面。**GPU作为人工智能算力的最小核心单元，其性能直接决定算力上限；以“异构计算、AI算力集群”为基础的系统架构，正推动算力基础设施从“单一服务器”向“大规模协同集群”转型。

01

精准布局关键技术

- 围绕芯片级（如多DIE封装、显存带宽优化等）与系统级（如异构技术、互连技术、液冷技术、供电技术、系统管理与控制等）的关键技术，将研发成果及时转化为专利，确保企业核心技术的有效保护。

02

重视专利质量与全球布局

- 专利布局不能仅追求数量，还需要提升专利质量，确保专利权利要求的稳定性和产品的对应性；同时，针对人工智能产业的全球化竞争，中国企业在进入海外市场时要积极推进专利的海外布局，护航海外业务。

3.2 探索“开源开放+专利共享”模式，推动产业协同创新

01

以开源开放构建产业生态

- 人工智能算力的软件层（如算力调度框架、整机系统、BMC等）可采用开源模式，吸引全球开发者参与迭代，快速完善生态；同时，通过“开源许可协议+专利授权”的组合，明确开源技术的专利使用和授权范围，避免开源技术被滥用。

02

推动硬件标准统一化

- 推动算力领域的硬件接口、管理协议、能效评价等方面的标准统一化，形成统一的开放架构，降低产业链上下游的适配成本，加速技术落地。

03

组建产业专利池，推动交叉许可

- 产业链上下游企业可联合组建人工智能算力专利池，通过专利交叉许可实现技术共享，打破“专利壁垒”，降低企业的技术使用成本，形成“技术共享-联合创新-产业升级”的良性循环。

数据来源：政府官网、亿欧智库

亿欧智库

网址：https://www.iyiou.com/research
邮箱：hezuo@iyiou.com
电话：010-53321289

扫码关注亿欧智库
查看更多研究报告

扫码添加小助手
加入行业交流群