

2024年中国端侧AI场景应用分析：智能终端与边缘计算的融合创新

China AI Terminal Application Industry

中国AI端末应用业界

报告标签：端侧AI、终端AI

主笔人：饶立杰

报告提供的任何内容（包括但不限于数据、文字、图表、图像等）均系头豹研究院独有的高度机密性文件（在报告中另行标明出处者除外）。未经头豹研究院事先书面许可，任何人不得以任何方式擅自复制、再造、传播、出版、引用、改编、汇编本报告内容，若有违反上述约定的行为发生，头豹研究院保留采取法律措施、追究相关人员责任的权利。头豹研究院开展的所有商业活动均使用“头豹研究院”或“头豹”的商号、商标，头豹研究院无任何前述名称之外的其他分支机构，也未授权或聘用其他任何第三方代表头豹研究院开展商业活动。

团队介绍

头豹是国内领先的行企研究原创内容平台和创新的数字化研究服务提供商。头豹在中国已布局3大研究院，拥有近百名资深分析师，头豹科创网(www.leadleo.com)拥有20万+注册用户，6,000+行业赛道覆盖及相关研究报告产出。

头豹打造了一系列产品及解决方案，包括数据库服务、行企研报服务、微估值及微尽调自动化产品、财务顾问服务、PR及IR服务，研究课程，以及分析师培训等。诚挚欢迎各界精英与头豹交流合作，请即通过邮件或来电咨询。

报告作者



袁栩聪
首席分析师
oliver.yuan@Leadleo.com



饶立杰
行业分析师
arjen.rao@Leadleo.com

头豹研究院

咨询/合作

网址：www.leadleo.com

电话：13080197867（李先生）

电话：15999806788（袁先生）

深圳市华润置地大厦E座4105室

摘要

端侧AI，是人工智能应用的一种实现方式，侧重于将AI的能力下沉到设备端，减少对云服务的依赖，从而实现更快的响应速度、更好的隐私保护和更低的网络需求。2023年中国端侧AI市场规模为1,939亿元，预计2028年其市场规模将为19,071亿元，年均复合增长率为58%。技术进步和边缘计算的出现是端侧AI行业发展的驱动因素，未来端侧AI将与云侧AI结合为混合式AI一同使用，采用混合式AI，可在全球范围内实现成本效益、能效、性能提升、隐私保护、安全性增强和个性化服务。

■ 中国端侧AI行业现状如何？

目前，中国主流科技企业纷纷在端侧AI领域加大投入并取得显著成果，特别是在移动端和操作系统层面上的AI集成与创新。同时海外科技企业推动端侧AI技术的集成和应用更广泛。同时，端侧大模型是端侧AI中游厂商实现落地应用的关键因素之一，众多端侧AI中游厂商选择自研端侧大模型。

■ 中国端侧AI市场规模如何？

2023年中国端侧AI市场规模为1,939亿元，从2018至2023年，其年均复合增长率为116.3%。在2023年之前，端侧AI技术已在多个领域落地应用，其中智能安防和智能车载设备是两个重要的应用领域。AI手机和AI PC开始兴起，两者庞大的市场需求将在未来支撑端侧AI行业迅速发展。预计2028年，中国端侧AI市场规模将为19,071亿元，年均复合增长率为58%。

■ 中国端侧AI发展如何？

混合式AI是指，终端设备和云端服务器相互协作，根据具体场景和时间，合理分配AI计算任务，以提供更优的用户体验，并高效利用计算资源。在特定应用场景中，计算任务主要在终端设备上执行，仅在需要时才将部分任务迁移到云端处理。而在以云端为核心的场景中，终端设备会根据自己的计算能力，在条件允许的情况下分担部分AI计算任务。

目录

◆ 中国端侧AI行业综述	-----	06
• 端侧AI定义及分类	-----	07
• 行业发展历程	-----	08
• 市场规模测算	-----	09
◆ 中国端侧AI行业产业链分析	-----	10
• 产业图谱及说明	-----	11
• 上游产业链分析	-----	12
• 中游产业链分析	-----	13
• 下游产业链分析	-----	14
◆ 中国端侧AI行业分析	-----	16
• 相关政策	-----	17
• 驱动因素	-----	18
• 限制因素	-----	20
• 发展趋势	-----	21
• 竞争格局	-----	22
◆ 中国端侧AI代表性企业	-----	24
• 华为	-----	25
• 商汤科技	-----	26
• 中科创达	-----	27
◆ 方法论及法律声明	-----	28

Contents

◆ Overview of China's AI Terminal Application Industry	06
• Definition and Classification of Intelligent Perception	07
• Industry Development History	08
• Market Size Calculation	09
◆ Analysis of China's AI Terminal Application Industry Chain	10
• Industry Map and Explanation	11
• Analysis of Upstream Industry Chain	12
• Analysis of Midstream Industry Chain	13
• Analysis of Downstream Industry Chain	15
◆ Analysis of China's AI Terminal Application Industry	17
• Related Policies	18
• Driving Factors	19
• Limiting Factors	20
• Development Trends	21
• Competitive Landscape	22
◆ Representative Enterprises of Perceived Intelligence in China	24
• HUAWEI	25
• Sense Time	26
• ThunderSoft	27
◆ Methodology and Legal Statement	28

Chapter 1

中国端侧AI行业综述

- 端侧AI定义及分类
- 行业发展历程
- 市场规模测算

中国端侧AI行业综述——端侧AI定义及分类

端侧AI侧重于将AI的能力下沉到设备端，并涵盖AI驱动系统融合、AI重塑操作系统、端侧AI应用、AI端云协同四层应用，按照模型训练方向的分类方式，端侧AI可以分为5大类

端侧AI行业定义与分类

- **端侧AI**，是人工智能应用的一种实现方式，侧重于将AI的能力下沉到设备端，减少对云服务的依赖，从而实现更快的响应速度、更好的隐私保护和更低的网络需求。与云端大模型相比，端侧AI更强调本地化运行，以满足用户对隐私的需求和个性化的任务需求。值得关注的是，端侧AI行业的应用范围广泛，包括但不限于个人电脑（PC）、智能手机等设备。
- 目前，端侧AI涵盖四层应用：①AI驱动系统融合，实现多设备间无缝衔接与数据共享；②AI重塑操作系统，通过意图识别优化人机互动，提升用户体验；③端侧AI应用，涵盖图像处理、文档摘要等实用功能；④AI端云协同，结合云端大模型，拓展AI应用场景边界。
- 按照模型训练方向的分类方式，端侧AI行业可以分为5大类，包括自然语言处理端侧AI、计算机视觉端侧AI、智能语音端侧AI、全模态处理端侧AI以及科学计算端侧AI。

01

自然语言处理端侧AI

在自然语言处理（NLP）领域的应用意味着用户的语音或文本命令可以直接在用户的设备上被处理和理解，而不是上传到云端处理。目前，许多科技公司都在研发专门针对端侧的NLP解决方案，例如轻量级的NLP模型、边缘计算平台以及优化的运行时环境。

02

计算机视觉端侧AI

在本地设备上处理和分析图像及视频数据，以实现识别、分类、追踪、场景理解等多种视觉任务。与自然语言处理的端侧AI相似，计算机视觉端侧AI也旨在提升隐私性、降低延迟、支持离线操作并优化资源使用。

03

智能语音端侧AI

在本地设备上处理语音识别、语音合成、自然语言理解和语音交互等任务的技术。该技术的发展使得智能设备能在不依赖网络连接的情况下，实现快速、高效且保护隐私的语音控制和交互。

04

全模态处理端侧AI

在边缘设备上同时处理和融合多种类型的数据模态（如文本、语音、图像、视频等）的人工智能技术。该综合处理方式旨在通过跨模态信息的整合，提高决策的准确性和上下文理解能力，为用户提供更为丰富和自然的交互体验。

05

科学计算端侧AI

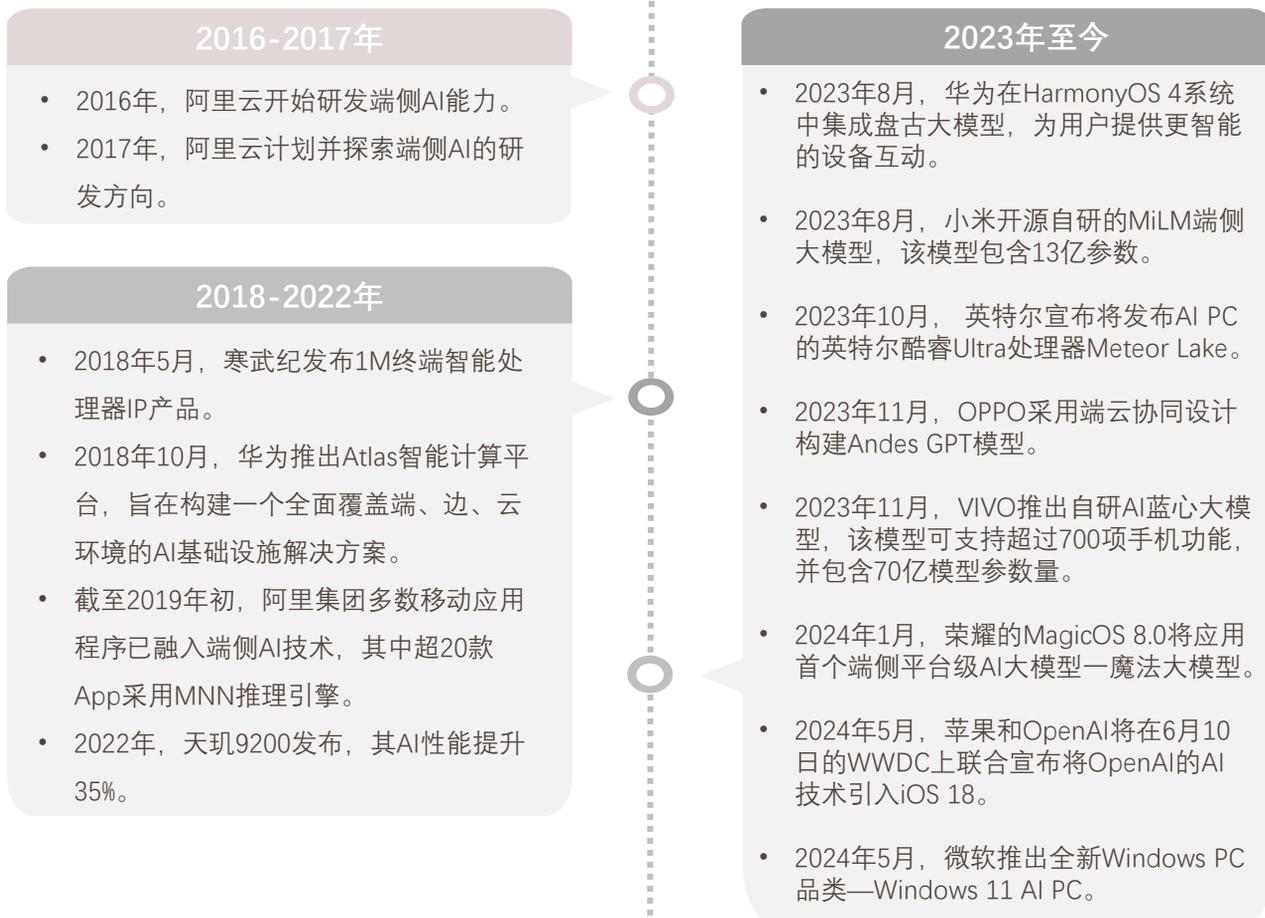
在科研、工程设计、数据分析等专业领域，利用部署在本地设备上的人工智能算法来加速计算过程、优化模型预测或执行复杂数据分析的任务。与通用的端侧AI相比，科学计算端侧AI更加侧重于处理大规模数据集、执行复杂的数学运算、模拟物理过程或解决特定领域的科学问题。

来源：专家访谈，头豹研究院

中国端侧AI行业综述——行业发展历程

中国端侧AI行业经历三个阶段，分别为起步探索阶段、技术研发与初步落地阶段和应用快速拓展阶段

中国端侧AI发展历程



- **中国端侧AI行业经历三个阶段：起步探索阶段、技术研发与初步落地阶段和应用快速拓展阶段。**
- **起步探索阶段：**2016-2017年，中国逐渐进入端侧AI领域，且互联网巨头企业率先发力。这标志着中国正式开启端侧AI领域的探索之旅。
- **技术研发与初步落地阶段：**2018-2022年，中国端侧AI行业在本土创新企业的推动下迅速启动，核心芯片研发、计算平台建设、广泛应用实践及性能持续升级。
- **应用快速拓展阶段：**2023年至今，中国主流科技企业纷纷在端侧AI领域加大投入并取得显著成果，特别是在移动端和操作系统层面上的AI集成与创新。同时海外科技企业推动端侧AI技术的集成和应用更广泛。

来源：企业官网，头豹研究院

中国端侧AI行业综述——市场规模测算

2023年中国端侧AI市场规模为1,939亿元，预计2028年其市场规模将为19,071亿元，年均复合增长率为58%，端侧AI已大规模用于智能安防和智能车载设备，未来AI手机和AI PC将成为主要增长因素

中国端侧AI行业市场规模，2018-2028E

单位：亿元



- 2023年中国端侧AI市场规模为1,939亿元，从2018至2023年，其年均复合增长率为110.3%。

在2023年之前，端侧AI技术已在多个领域落地应用，其中智能安防和智能车载设备是两个重要的应用领域。端侧AI在智能安防领域的应用包括视频监控分析、面部识别、异常行为检测等，以上技术能够实时分析监控数据，提高安防系统的响应速度和准确性，从而增强公共安全。在智能交通和自动驾驶领域，端侧AI技术被用于车辆的环境感知、决策制定和实时控制，包括车道保持辅助、碰撞预警、自动泊车等智能功能，旨在提升驾驶安全和车辆的自动化水平。
- 值得关注的是，自2023年起，AI手机和AI PC开始兴起，两者庞大的市场需求将在未来支撑端侧AI行业迅速发展。其中AI手机通过集成NPU或AI加速器，能实现更快的图像处理、语音识别和自然语言处理等功能，进而提供更加丰富的智能服务，如智能助手、实时翻译、健康监测等，极大地提升用户体验。而AI PC利用AI技术为用户提供强大的计算能力和智能功能，特别是在创意工作、游戏、教育和企业级应用中。例如，AI PC能通过学习用户的行为模式，优化性能，提供个性化的软件和服务。同时，端侧AI的应用将不断拓展到新的领域，如智能家居、健康医疗、工业自动化等。预计2028年，中国端侧AI市场规模将为19,071亿元，年均复合增长率为58%。

来源：专家访谈，头豹研究院

Chapter 2

中国端侧AI行业产业链分析

- 产业图谱及说明
- 上游产业链分析
- 中游产业链分析
- 下游产业链分析

中国端侧AI行业产业链分析——产业图谱

中国端侧AI行业的上游产业链为核心技术和硬件提供商，中游产业链为端侧AI系统集成与软件平台开发商，下游产业链为应用领域与服务

端侧AI产业图谱

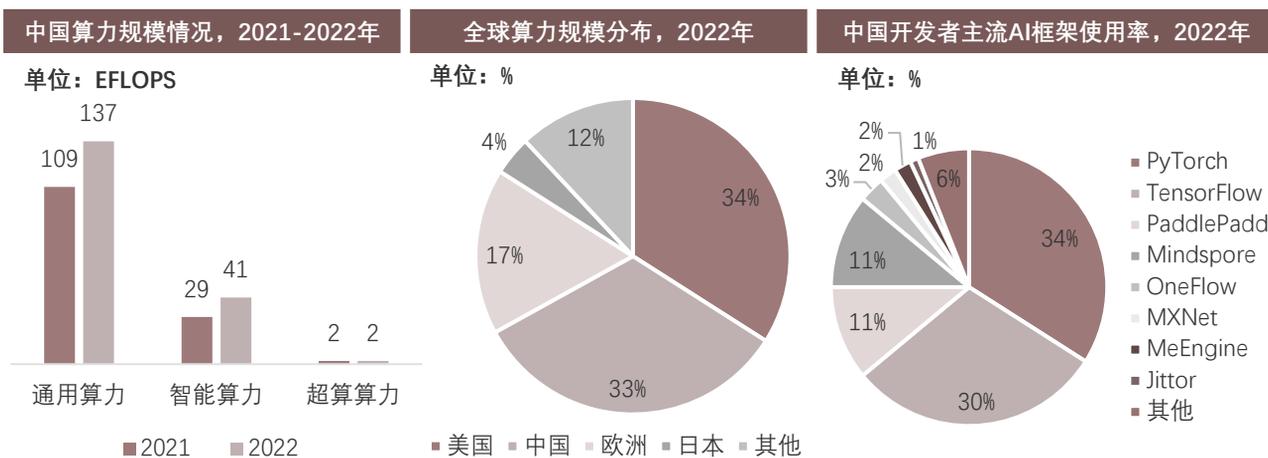


来源：头豹研究院

中国端侧AI行业产业链分析——上游产业链分析

2022年中国通用算力和智能算力分别达137EFLOPS和41EFLOPS，分别占全球比重27%和39%，MindSpore在中国自主研发的AI框架中活跃度最高，且未来NPU将在端侧AI领域广泛应用

中国端侧AI上游产业链分析



- 伴随着人工智能、大数据、云计算等前沿技术的迅速发展，算力作为技术基石，其需求呈爆发式增长。2022年中国通用算力和智能算力分别达137EFLOPS和41EFLOPS，分别同比增长25.7%和41.4%。
- 美国、中国、欧洲、日本在全球算力规模中的份额分别为34%、33%、17%和4%。其中，在基础算力方面，中国以27%份额的排名第二；在智能算力方面，中国以39%的份额排名第一。
- 在选择AI框架时，开发者最重视的因素是框架的易用性，其中有40%的开发者将其列为首要考量，其次性能为第二位选择依据。值得关注的是，目前在中国自主研发的AI框架中，华为的昇思（MindSpore）展现出最高的活跃度。

各类AI芯片对比					中国AI服务器市场规模，2023-2027E	
指标	CPU	GPU	NPU	FPGA	单位：亿美元	
定制化程度	通用型	通用型	专用型	专用型	91	+10.2%
功耗	低	中	低	高	搜索 《2024年中国端侧AI场景应用分析》	
成本	高	高	低	高		
算力	低	中	高	低		

完整版本登录 www.leadleo.com

- NPU是为加速AI推理任务而特别设计的处理器，通过优化的架构实现低功耗运行。由于AI工作负载主要涉及神经网络层的计算，包括标量运算、向量运算和张量运算，以及非线性激活函数的处理，因此未来NPU将在端侧AI领域广泛应用。NPU针对网络运算优化，可通过编程实现特定网络功能，提供比CPU和GPU更高效、低能耗的神经网络处理能力，且相比FPGA成本更优。
- 随着AIGC、云计算技术的发展，超级计算市场有望迎来新的增长。高端服务器市场和GPU服务器市场将持续扩张，推动AI服务器市场的进一步发展。2023年，中国AI服务器市场规模达91亿美元，预计2027年，中国AI服务器市场规模将达134亿美元，年均复合增长率为10.2%。目前，中国长城和百度携手合作，并成功完成一项重要的技术突破。该合作内容针对擎天EF860双路服务器和昆仑AI加速卡的适配测试，并标志着继AI-TF2000飞桨一体机之后，双方在中国AI服务器领域的再次显著进展。

来源：中国通信院，中国知网，头豹研究院

中国端侧AI行业产业链分析——中游产业链分析

2023年中国端侧大模型市场规模达8亿元，预计2028年其市场规模将达79亿元，年均复合增长率为58%，端侧AI企业商业模式可分为5种，且主要参与者可分为4类

中国端侧AI中游产业链分析



中国端侧AI企业商业模式		中国端侧AI行业主要参与者分类		
分类	企业	特征		
01	B2B和B2C: 硬件设备集成销售	拥有全面的行业解决方案, 出色的软件服务以及全局数据平台	互联网型企业	阿里云
02	咨询服务: 行业定制化解决方案		To B、To C 人工智能企业	商汤科技
03	招投标项目: 与政府合作	具备较强的硬件集成能力, 以及提供实时或近实时的处理能力		端侧AI终端场景企业
04	技术授权: 收取许可费		vivo	
05	订阅服务: 订阅软件及增值服务	具备庞大的规模和强大的抗风险能力, 以及整合上下游产业链资源的能力	央企	国家电网
				国家能源集团

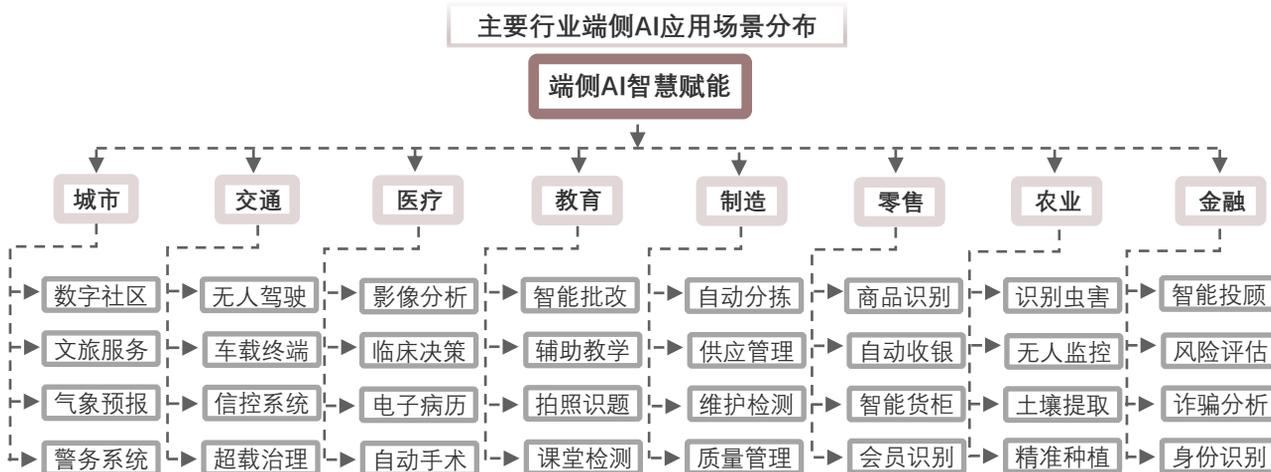
- 端侧AI企业的商业模式选择取决于其技术优势、市场定位、目标客户群体以及行业特点，模式包括销售硬件设备集成、提供行业定制化解决方案、参与政府项目、收取许可费以及订阅和增值费用。
- 端侧AI企业可分为互联网型企业、To B、To C人工智能企业、端侧AI终端场景企业以及央企。其中互联网型企业的优势在于拥有强大的在线服务和用户基础，而端侧AI终端场景企业需与硬件设备紧密结合，提供端到端的解决方案。

来源: 专家访谈, 头豹研究院

中国端侧AI行业产业链分析——下游产业链分析

端侧AI的应用领域广泛渗透到城市管理、交通、医疗、教育、制造业、零售、农业、金融、数字社区以及文旅服务等多个行业，其中智慧安防和智慧车载领域的规模已达1,226亿元和119亿元

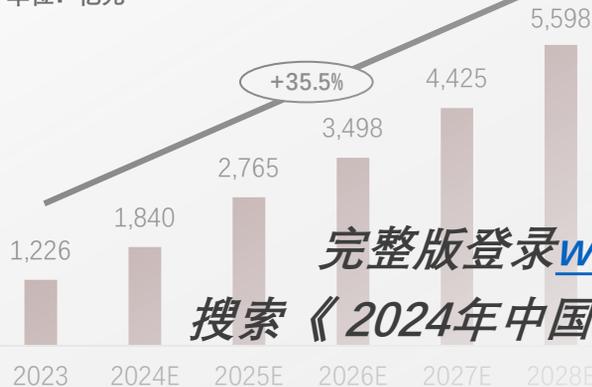
中国端侧AI行业下游应用领域分析



■ 端侧AI的应用领域广泛渗透到城市管理、交通、医疗、教育、制造业、零售、农业、金融、数字社区以及文旅服务等多个行业。端侧AI通过在设备端完成数据处理，提供高速、低功耗、高隐私保护的解决方案，涉及应用如AR体验、游戏交互、美颜及虚拟试妆、智能内容推荐、用户状态感知与智能推送等，以及在汽车领域实现个性化与安全驾驶辅助。此外，端侧AI还在持续推动着工作与生活的变革，例如提升笔记本电脑的智能化水平至可与行业领先品牌竞争的程度，以及在MWC等展览会上展示的最新AI赋能的用户体验创新。端侧AI的应用预示着一个更加智能、便捷且安全的数字化生活时代的到来。

中国智能安防市场规模，2023-2028E

单位：亿元



中国智能车载市场规模，2023-2028E

单位：亿元



完整版登录 www.leadleo.com

搜索 《2024年中国端侧AI场景应用分析》

■ 2023年中国智慧安防市场规模达1,226亿元，同比增长15.5%，预计2028年将达到5,598亿元，复合增长率为35.5%。端侧AI在安防领域的应用，如人脸识别、行为分析等，不仅提高了安防效率，还降低了误报率。

■ 2023年中国智能车载市场规模达216亿元，同比增长15.5%，预计2028年将达到640亿元，复合增长率为40%。端侧AI在车载领域的应用，如智能语音助手、自动驾驶辅助等，提升了驾驶体验和安全性。

来源：深圳新闻网，中国知网，数字信息安防产业技术创新联盟，头豹研究院

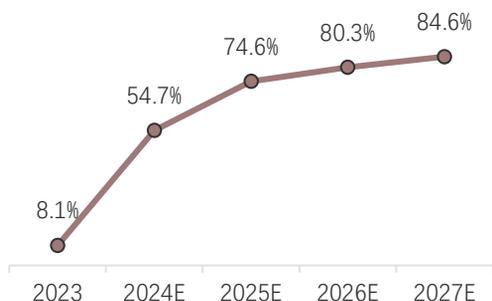
（接上页——下游产业链分析）

未来AI PC和AI手机将作为端侧AI领域的重要载体，预计2027年，中国AI PC占个人电脑的市场比重将达84.6%，AI手机将占51.9%的手机市场份额，同时80%以上的终端设备将具备AI处理能力

中国端侧AI行业下游产品与软件服务分析

中国AI PC市场规模占比，2023-2027E

单位：%



全球已布局AI PC企业

企业	发布日期	AI PC产品	所应用AI技术
苹果	2024.3.4	MacBook Air (M3)	16核神经网络引擎
荣耀	2024.3.18	MagicBook Pro 16	AI硬件、AI助理、AI互联
微软	2024.3.22	Surface Pro 10	AI专属芯片NPU
华为	2024.4.11	MateBook X Pro	盘古大模型

- AI PC集成混合AI算力单元，可本地运行个人大模型，同时创建个性化知识库，进而实现自然语言交互。AI PC不仅是生产力工具，也是个人AI助理，不仅可提高效率，还能简化工作流程，同时保护隐私数据安全。
- 2023年，中国AI PC占个人电脑的市场比重为8.1%，预计2027年其占比将达84.6%。值得一提的是，预计到2027年，AI平板电脑在平板电脑市场的占比将达到75%。
- 未来，惠普、戴尔、联想、华硕等电脑制造商计划与英特尔和AMD的新一代CPU产品发布计划保持同步，并推出搭载新处理器的AI PC。在AI PC的引领下，PC产业将从应用中心转向用户中心，从应用驱动转变为需求驱动。

中国AI手机销量及占比，2023-2027E

单位：亿台，%



全球布局AI手机企业

企业	是否自研	是否落地	应用于AI手机的大模型
vivo	是	是	蓝心大模型
OPPO	与联发科技合作	是	AndesGPT大模型
荣耀	是	是	魔法大模型
华为	是	是	盘古大模型
小米	是	是	XiaomiAISP
苹果	与OpenAI合作	否	ChatGPT
三星	是	是	三星高斯

- AI手机搭载支持生成式AI的智能终端硬件平台，并内嵌专属智能体。AI技术不仅推动摄影技术的发展，生成式AI在手机端侧得到应用，涵盖智能助手和操作系统优化等领域。
- 2023年，中国AI手机的出货量为0.1亿台，同时超过30家手机厂商宣布将AI大模型引入手机中。值得关注的是，预计2027年，中国AI手机的出货量将达1.5亿台，并占中国51.9%的手机市场份额，年均复合增长率为196.8%。
- 预计2024年，中国终端设备市场中超过50%的设备将内置具备AI处理能力的硬件基础。预计2027年，该比例将上升至80%，并进入稳定增长的阶段。

来源：中国知网，白皮书，头豹研究院

Chapter 3

中国端侧AI行业分析

- 相关政策
- 驱动因素
- 限制因素
- 发展趋势
- 竞争格局

中国端侧AI行业分析——相关政策

中国一系列跨领域的政策，为端侧AI行业提供战略支持、算力基础、标准化指导、服务管理规范和应用场景创新的全方位推动力

中国端侧AI行业相关政策

政策名称	《深入实施“东数西算”工程加快构建全国一体化算力网的实施意见》		
发布时间	2023	发布主体	发改委，工信部等5部门
主要内容	该政策指出，以算力高质量发展赋能经济高质量发展为主线，充分发挥全国一体化算力网络国家枢纽节点，协同推进“东数西算”工程		
政策名称	《数字中国建设整体布局规划》		
发布时间	2023	发布主体	中共中央，国务院
主要内容	该政策指出，推进数字社会治理精准化，深入实施数字乡村发展行动，以数字化赋能乡村产业发展、乡村建设和乡村治理。同时普及数字生活智能化，打造智慧便民生活圈、新型数字消费业态、面向未来的智能化沉浸式服务体验		
政策名称	《新产业标准化领航工程实施方案（2023—2035年）》		
发布时间	2023	发布主体	工信部，科技部等4部门
主要内容	该政策指出，研制人形机器人术语、通用本体、整机结构、社会伦理等基础标准。同时研制人形机器人感知系统、定位导航、人机交互、自主决策、集群控制等智能感知决策和控制标准		
政策名称	《生成式人工智能服务管理暂行办法》		
发布时间	2023	发布主体	发改委，工信部，教育部等7部门
主要内容	该政策指出，鼓励生成式人工智能算法、框架、芯片及配套软件平台等基础技术的自主创新，同时完善与创新相适配的科学监管方式，制定相应的分类分级监管规则或者指引，从而倡导人形机器人标准化		
政策名称	《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》		
发布时间	2022	发布主体	工信部，教育部，科技部等6部门
主要内容	该政策指出，在城市管理领域，探索城市大脑、城市物联感知、政务数据可用不可见、数字采购等场景。在交通治理领域，探索交通大脑、智慧道路、智慧停车、自动驾驶出行、智慧港口、智慧航道等场景		

- 中国一系列跨领域的政策，为端侧AI行业提供战略支持、算力基础、标准化指导、服务管理规范和应用场景创新的全方位推动力。这意味着端侧AI将在国家层面得到更加明确的发展路径、更加强大的技术支撑和更加广阔的应用前景，从而加速AI技术的创新应用和产业生态的构建。通过推动端侧AI向更深层次、更广领域发展，经济社会的数字化转型和高质量发展将获得新动能。

来源：各政府部门，头豹研究院

中国端侧AI行业分析——驱动因素

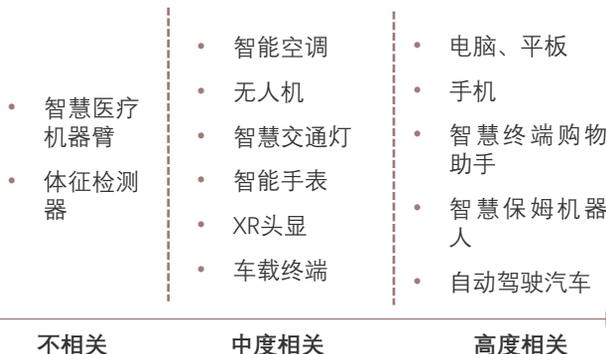
技术进步和边缘计算的出现是端侧AI行业发展的驱动因素，主要在于移动通信技术的迭代、异构结构提高处理性能以及云计算与终端设备之间的空白被填补

中国端侧AI行业发展驱动因素

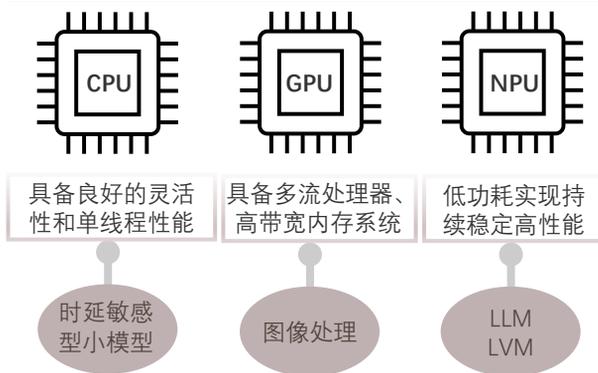
技术进步

移动通信技术的迭代，异构结构提高处理性能

各类终端产品应用端侧AI对通信技术需求程度



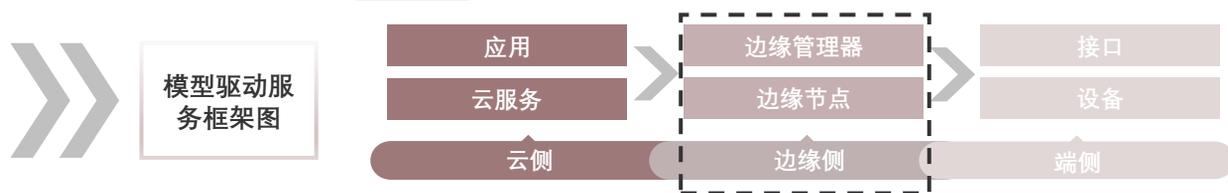
异构结构中不同处理器使用目的



- 移动通信技术主要通过提升数据传输速度、降低延迟和增强连接性，为端侧AI提供更广阔的施展空间和实时交互能力。值得关注的是，手机、电脑、自动驾驶汽车等终端产品需配置6G网络，以完全实现应用需求。主要原因在于，涉及高清视频处理、实时数据分析和复杂模型推理的应用，需极高的数据传输带宽，同时6G网络将提供更低的延迟，进而提高应用的实时性和可靠性。此外，6G网络将采用更加先进的网络架构，包括去中心化和智能化的网络管理，将有助于提高端侧AI产品的运行效率和灵活性。
- 终端异构混合计算资源（CPU、NPU和GPU）是推动AI技术在各个领域广泛应用的关键。异构混合架构通过结合不同指令集和体系结构的计算单元，构建一个本地计算系统。该架构能利用CPU、NPU和GPU等计算设备的优势，针对不同的AI任务提供高效灵活的处理方案。

边缘计算

填补云计算与终端设备之间的空白



- 边缘计算代表分布式系统架构的演进，其主要概念是将数据处理和业务逻辑推向网络的边缘，即靠近数据生成的位置。边缘计算旨在减轻云端服务器的计算压力，并加快物联网设备对数据处理需求的响应时间。可见，边缘计算通过提供低延迟、高隐私、低成本、高可靠性的计算平台，不仅解决端侧AI应用的核心问题，还促进AI技术在更多领域的创新应用和商业落地。预计2025年，中国边缘计算算力将占总算力市场比重的10%，预计2030年，其占比将达70%。

来源：全球移动通信系统协会，专家访谈，头豹研究院

（接上页——驱动因素）

AI Agent的特点在于其自主性，能不依赖持续的人为指令，并依据任务需求自我调整策略，未来AI Agent将在端侧设备上独立运行，成为智能设备的核心组件

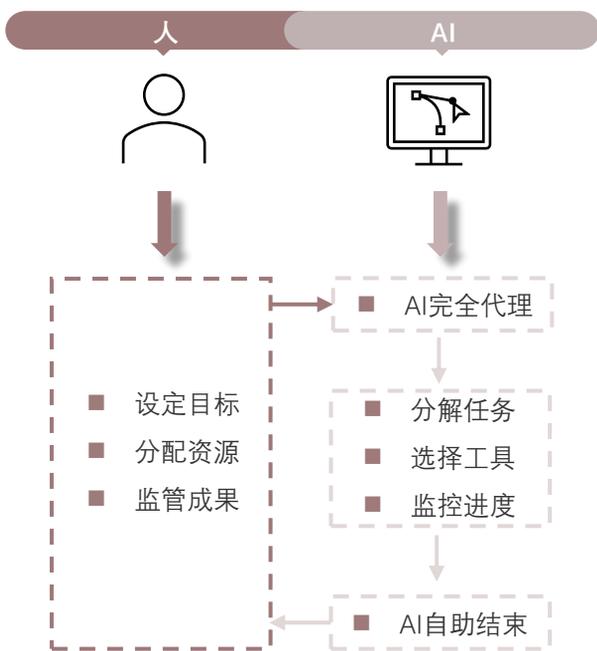
中国端侧AI行业发展驱动因素

AI Agent兴起

自主决策执行，成为端侧AI的核心组件

- AI Agent是一种技术实体，能自主感知周遭环境，通过内在的智能处理进行决策，并执行相应行动以达成特定目的。AI Agent的特点在于其自主性，能不依赖持续的人为指令，并依据任务需求自我调整策略。AI Agent的设计较为高级，一旦接收到一个目标指令，即可启动一系列复杂逻辑，包括但不限于自我学习、工具利用等，来独立推进行动计划直至任务成功完成。

人与AI Agent合作模式图



AI Agent生态框架图



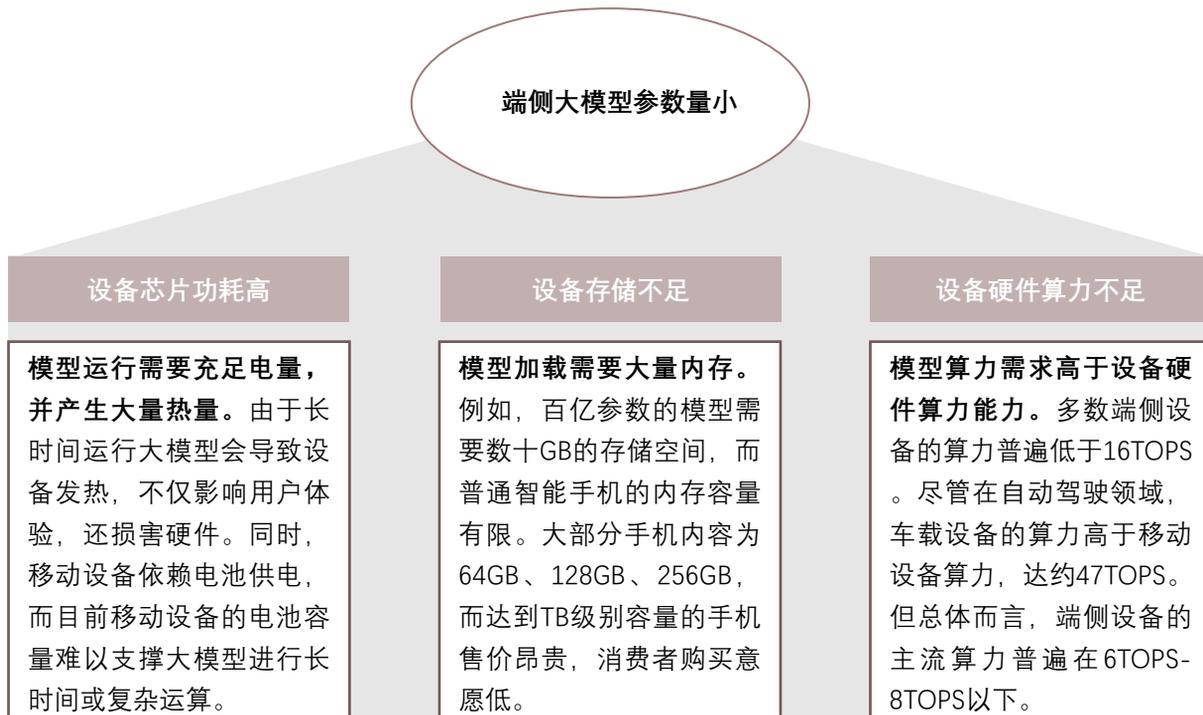
- AI Agent将在端侧设备上独立运行，成为端侧AI的核心组件。以联想的“智能终端AI OS”为例，该系统通过将AI技术集成到个人Agent中，并嵌入智能终端，打造新一代的AI驱动操作系统。该集成方式不仅增强端侧设备的智能功能，还支持跨应用操作、主动服务和自我进化的能力。同时，端侧AI模型的运行效率显著提高，能在短时间内处理大量数据，实现快速响应和低延迟。随着AI技术的不断进步，AI Agent将在端侧AI领域扮演愈发重要的角色。

来源：头豹研究院

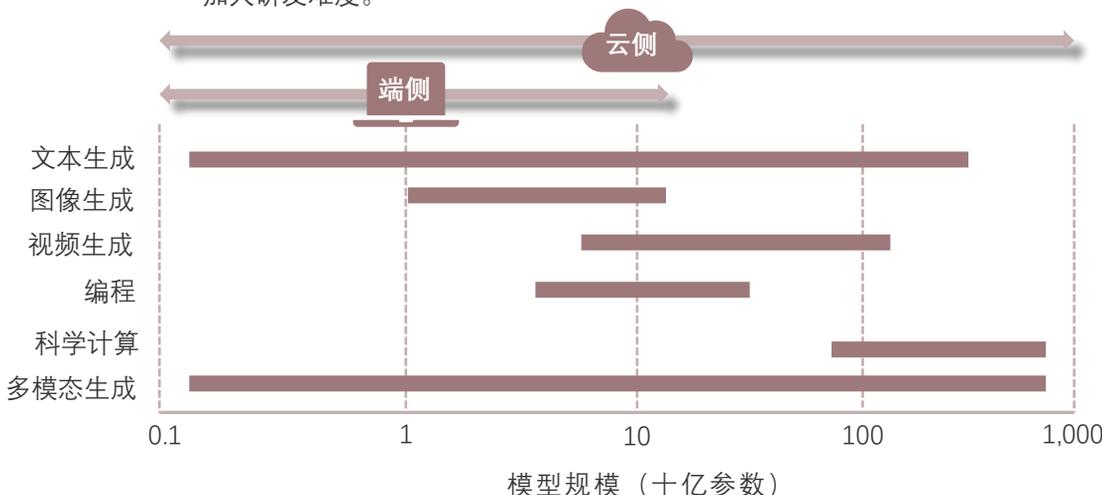
中国端侧AI行业分析——限制因素

端侧AI行业发展的核心限制因素是端侧大模型参数量小，主要原因在于设备芯片功耗高、设备存储不足以及设备硬件算力不足

中国端侧AI行业发展限制因素



■ 端侧AI行业发展的核心限制因素是端侧大模型参数量小，主要原因在于设备芯片功耗高、设备存储不足以及设备硬件算力不足。云端大模型参数量级通常达到上百亿至数千亿，而端侧大模型参数量较小，主流端侧大模型参数量约为20亿。此外，端侧大模型需适应不同性能的设备，在模型设计时需考虑到广泛的硬件兼容性，因此加大研发难度。



来源：白皮书，专家访谈，头豹研究院

中国端侧AI行业分析——发展趋势

未来端侧AI将与云侧AI结合为混合式AI一同使用，采用混合式AI，可在全球范围内实现成本效益、能效、性能提升、隐私保护、安全性增强和个性化服务

中国端侧AI行业发展趋势

中心云、边缘云及端侧关系图



- 混合式AI是指，终端设备和云端服务器相互协作，根据具体场景和时间，合理分配AI计算任务，以提供更优的用户体验，并高效利用计算资源。在特定应用场景中，计算任务主要在终端设备上执行，仅在需要时才将部分任务迁移到云端处理。而在以云端为核心的场景中，终端设备会根据自己的计算能力，在条件允许的情况下分担部分AI计算任务。
- 混合式AI技术将允许开发者和提供商利用边缘设备上的计算资源，从而降低成本。同时，混合式AI能在全球范围内提供高性能、个性化服务、隐私保护和安全保障等优势。

混合式AI的优势

低能耗：边缘设备支持高效AI处理，且能效优于云端。结合处理和数据传输，边缘设备以低能耗运行生成式AI模型。

高隐私性和安全性：端侧AI的设计有助于保护用户隐私，通过本地处理查询和个人信息，避免数据泄露，对于企业或工作环境中保护敏感信息具有重要意义。

低成本：生成式AI搜索的成本是传统搜索方法的10倍。通过将处理任务从云端转向边缘设备，可降低云压力和成本。

强可靠性、性能和时延：在混合式AI的架构下，终端侧AI处理稳定高效，即使云服务拥堵也能保持性能，同时可避免云端高峰时的排队、延迟和拒服问题。

强个性化：混合AI实现个性化体验，数字助手可定制化服务，同时确保隐私安全。此外，个性化应用不仅限于消费者，企业也可利用混合AI标准化代码编写等。

来源：白皮书，头豹研究院

中国端侧AI行业分析——竞争格局

中国端侧AI行业的竞争格局梯队主要根据市场力进行划分，其中包括端侧AI行业解决方案、硬件产品以及软件服务，其中第一梯队包括华为云、阿里云以及商汤科技

中国端侧AI行业竞争格局分析

企业名称	是否上市	主要产品与服务	市场竞争力
华为云	否	HiAI3.0平台是一个具备端侧分布式AI开放能力的平台，将SoC中加入指向AI算力的NPU，完成AI手机的研发	第一梯队
阿里云	否	开源了推理引擎MNN，用于支持端侧AI应用，端侧AI技术紧贴用户在AR特效、搜索推荐等方面有诸多创新应用	第一梯队
商汤科技	是	发布行业首个“云、端、边”全栈大模型产品矩阵，其中包括应用于终端设备的“商汤端侧大模型”，以及“日日新SenseNova 5.0”大模型体系	第一梯队
中科创达	是	提供包括AI手机、AI PC等端侧品类的整体解决方案，发布AI魔方大模型和滴水OS整车操作系统，以及AI开发套件RubikLiteDK	第二梯队
云从科技	是	在人脸识别、图像识别等领域深耕，其技术广泛应用于金融、安防、交通等多个行业	第二梯队
汉王科技	是	计划于2025年推出人形机器人，该人形机器人融合先进的机器人传感技术、智能汽车技术以及采用鲲鹏鸿蒙系统的集成	第二梯队
寒武纪	是	芯片设计专长在于加速AI计算，为端侧设备提供强大的本地推理能力，支持快速响应的AI功能和服务	第二梯队
澜舟科技	否	利用端侧大模型，为端侧设备提供强大的本地推理能力，支持快速响应的AI功能和服务	第三梯队
荣耀	否	自研平台级端侧AI大模型“魔法大模型”，支持AI助手、场景识别等功能，强调个性化体验	第三梯队
vivo	否	发布自研蓝心大模型，在近年来的旗舰机型中广泛应用AI技术，比如AI相机增强、游戏体验优化等	第三梯队

完整版登录www.leadleo.com

搜索《2024年中国端侧AI场景应用分析》



来源：头豹研究院

（接上页——竞争格局）

中国端侧AI行业竞争格局的第一梯队为互联网头部企业和To B龙头企业，第二梯队为大部份AI先行企业，第三梯队为端侧场景应用企业，其中包括车企、手机厂商等

中国端侧AI行业竞争格局分析

- 中国端侧AI行业的竞争格局梯队主要根据市场力进行划分，其中包括端侧AI行业解决方案、硬件产品以及软件服务。第一梯队包括华为云、阿里云以及商汤科技等

完整版登录www.leadleo.com

搜索《2024年中国端侧AI场景应用分析》

- 中国端侧AI行业竞争格局的第二梯队包括中科创达、云从科技以及汉王科技等。第二梯队的企业在端侧AI产品聚焦在少数特定场景，如人脸识别、语音识别等

来源：专家访谈，头豹研究院

Chapter 4

中国端侧AI代表性企业

- 华为
- 商汤科技
- 中科创达

中国端侧AI代表性企业——华为

华为的竞争亮点有领先的5G通信技术、全栈云服务以及“云+端”协同框架，同时拥有广泛的设备覆盖和快速增长的市场份额，以及昇腾AI处理器和Atlas方案支持云端到端侧AI应用

华为企业介绍

- 企业名称：华为技术有限公司
- 总部地址：深圳
- 成立时间：1987年
- 行业领域：计算机、通信和其他电子设备制造业、人工智能

行业解决方案	智慧金融	智慧制造	智慧政务	智慧教育	智慧医疗	智慧交通
	智慧电力	智慧矿业	智慧能源	智慧零售	智慧建筑	智慧地产
产品服务	企业网络	企业光网络	数据存储	计算	智能协作	
	企业无线	企业服务与软件	管理系统	华为云	华为坤灵系列	
昇腾系列产品	AI模块	AI加速卡	智能边缘	AI服务器	AI集群	

✓ 竞争亮点

01

领先5G通信技术

提供高速、低延迟的网络解决方案

02

全栈云服务

提供从IaaS到PaaS的全栈云服务，涵盖计算、存储、数据库、AI、IoT等多种服务

03

“云+端”协同架构

提供基于人工智能芯片及模块的软硬件集成解决方案

华为从2017年推出麒麟970开始，就在其高端智能手机中集成了专用的神经网络处理单元（NPU），作为华为端侧AI技术的基础。后续的麒麟芯片持续迭代，提升AI运算能力，使得设备能更高效地运行复杂的AI算法。此外，华为通过提供开放平台如HiKey开发板以及相关的SDK和工具，鼓励开发者创建更多基于端侧AI的应用和服务。

■ 华为拥有广泛的设备覆盖和快速增长的市场份额

目前，鸿蒙操作系统的生态已经扩展到超过8亿台设备，已有180款设备已经升级到HarmonyOS 4.2版本，其中包括智能手机、平板电脑、耳机以及智能显示屏等多种类型。2024年第一季度，中国智能手机市场出货量达6,926万台，同比增长6.5%。在众多厂商中，华为以17%的市场份额排在中国市场出货量的第二位，同比增长幅度达110%，与排名第一的荣耀市场份额的差距仅为0.1%。而在2023年第四季度，华为的市场份额为13.9%，在市场中排名第四。

■ 华为昇腾AI处理器和Atlas方案支持云端到端侧AI应用

华为利用华为昇腾系列AI处理器及配套的基础软件平台，Atlas人工智能计算解决方案得以构建，该方案涵盖广泛的硬件形态，如Atlas模块、板卡、边缘计算设备、服务器群组等，旨在满足“端侧、边缘、云端”全方位的AI计算需求。其中，AI模块在边缘设备上执行目标识别和图像分类任务，广泛应用于智能摄像头、机器人和无人机等端侧AI应用场景。

来源：中国知网，企业官网，头豹研究院

中国端侧AI代表性企业——商汤科技

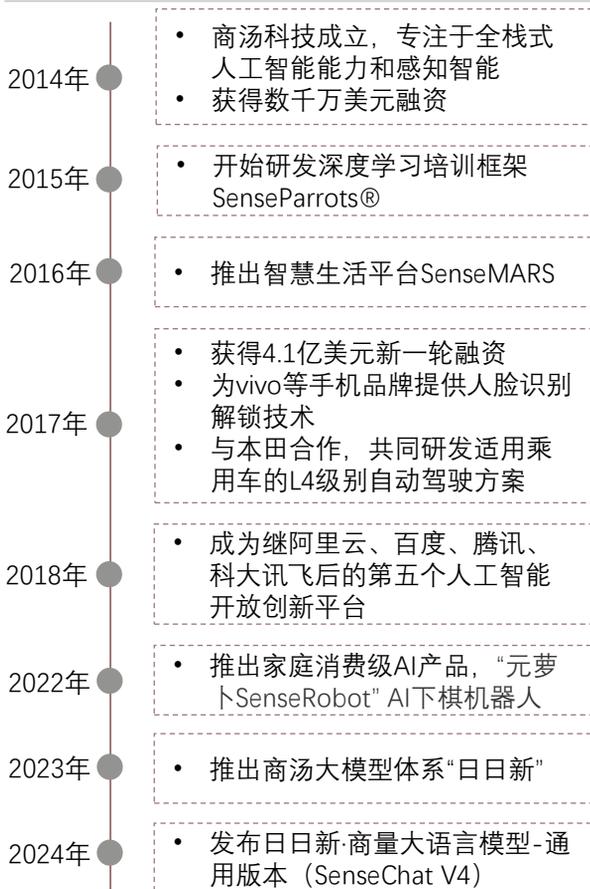
商汤科技成立至今10年，于2023年推出商汤大模型体系“日日新”，同时完成“云端边”全栈布局，依靠超大算力支撑，再结合大模型与行业实现深度融合

商汤科技企业介绍

- **企业名称：**商汤集团有限公司
- **成立时间：**2014年
- **总部地址：**上海
- **行业领域：**人工智能

- 商汤科技的AI视觉技术在汽车行业特别是整车制造业中备受推崇，其凭借全球顶尖的车内（In-Cabin）视觉解决方案及卓越的融合创新能力，使商汤绝影在全球最大的汽车市场中，车内视觉软件领域占据了领导地位。
- 商汤科技的“感知决策一体化自动驾驶通用大模型框架”提高多目标追踪的精度20%，增强车道线识别的准确度30%，减少对目标运动轨迹预测的误差大幅38%，同时降低在驾驶规划上的误差28%。

商汤科技发展历程



商汤科技竞争优势

- 完成“云端边”全栈布局：**为应对金融、编程、医疗、政府等领域边缘计算中AI应用需求的增长，商汤推出企业级大模型一体机，支持千亿级模型加速与知识检索硬件加速的本地部署。该一体机的优势在于即购即用，且相较同类产品节省推理成本80%，检索速度显著提升，减轻CPU负担50%。
- 行业深度融合：**商汤科技运用其“商汤日日新 SenseNova”大模型体系推动产业革新。在金融服务中，商汤可运用数字人技术优化客服和营销，并借力大语言模型新增投资分析、报告生成等功能。在医疗方面，商汤依据大量医学资料创立的“大医”医疗语言模型，能进行导诊、健康咨询等。
- 具有超大算力支撑：**商汤临港人工智能超算中心的3,740P超大算力池是由商汤科技（SenseTime）开发和运营的一个高性能计算基础设施。该超算中心不仅能供应庞大的计算资源，还凭借其尖端网络技术和高性能存储方案，无缝支撑包括自动驾驶、元宇宙等在内的高密集型AI应用场景，确保复杂运算需求的高效满足。

来源：中国知网，企业年报，企业官网，头豹研究院

中国端侧AI代表性企业——中科创达

中科创达与芯片厂商的深度合作，拥有全栈式产品线与AI端侧落地经验，推出的Rubik Avatar数字人解决方案，利用端侧大模型和裸眼3D技术，为服务型场景提供AI数字人

中科创达企业介绍

- **企业名称：**中科创达软件股份有限公司
- **成立时间：**2008年
- **总部地址：**北京
- **行业领域：**高新技术、人工智能、智能物联网

- 中科创达推出的Rubik Avatar数字人解决方案，利用端侧大模型和裸眼3D技术，为服务型场景提供AI数字人，不依赖现有人力即可提供快速、精准的互动问答。该解决方案支持云端和本地部署，本地部署版本可运行在高通高算力平台QCS8550上，进行本地图像渲染和模型推理，提高数据隐私性并降低延迟，同时大幅降低云资源占用和维护成本。

中科创达发展历程



中科创达竞争优势

- 与芯片厂商的深度合作：**中科创达与高通等芯片厂商保持长期紧密的合作关系。该合作不仅为中科创达提供在智能手机、汽车等AI终端应用上的广泛芯片支持，还通过联合实验室、联合开发QRD参考设计等方式，共同开发研究新技术，推动AI相关应用的精准落地。例如，中科创达在搭载高通8系列芯片平台的边缘设备上实现LLaMA-2 130亿参数模型的稳定运行，这显示公司在端侧AI领域的技术实力和落地能力。
- 全栈式产品线与AI端侧落地经验：**中科创达拥有覆盖软件开发到硬件AI芯片的全栈式产品线，并在端侧AI落地方面具备丰富的经验和卡位优势。公司聚焦“AI+”的应用产品研发，不断推动各种AI相关应用的精准落地。例如，中科创达发布的“滴水OS”整车操作系统，支持多域跨域融合，融合AI大模型，已经在中国市场的单SOC舱驾一体硬件环境中成功运行。

来源：企业官网，头豹研究院

方法论

- ◆ 头豹研究院布局中国市场，深入研究19大行业，持续跟踪532个垂直行业的市场变化，已沉淀超过100万行业研究价值数据元素，完成超过1万个独立的研究咨询项目。
- ◆ 研究院依托中国活跃的经济环境，研究内容覆盖整个行业的发展周期，伴随着行业中企业的创立，发展，扩张，到企业走向上市及上市后的成熟期，研究院的各行业研究员探索和评估行业中多变的产业模式，企业的商业模式和运营模式，以专业的视野解读行业的沿革。
- ◆ 研究院融合传统与新型的研究方法，采用自主研发的算法，结合行业交叉的大数据，以多元化的调研方法，挖掘定量数据背后的逻辑，分析定性内容背后的观点，客观和真实地阐述行业的现状，前瞻性地预测行业未来的发展趋势，在研究院的每一份研究报告中，完整地呈现行业的过去，现在和未来。
- ◆ 研究院密切关注行业发展最新动向，报告内容及数据会随着行业发展、技术革新、竞争格局变化、政策法规颁布、市场调研深入，保持不断更新与优化。
- ◆ 研究院秉承匠心研究，砥砺前行的宗旨，从战略的角度分析行业，从执行的层面阅读行业，为每一个行业的报告阅读者提供值得品鉴的研究报告。

法律声明

- ◆ 本报告著作权归头豹所有，未经书面许可，任何机构或个人不得以任何形式翻版、复刻、发表或引用。若征得头豹同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“头豹研究院”，且不得对本报告进行任何有悖原意的引用、删节或修改。
- ◆ 本报告分析师具有专业研究能力，保证报告数据均来自合法合规渠道，观点产出及数据分析基于分析师对行业的客观理解，本报告不受任何第三方授意或影响。
- ◆ 本报告所涉及的观点或信息仅供参考，不构成任何投资建议。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告。在法律许可的情况下，头豹可能会为报告中提及的企业提供或争取提供投融资或咨询等相关服务。本报告所指的公司或投资标的的价值、价格及投资收入可升可跌。
- ◆ 本报告的部分信息来源于公开资料，头豹对该等信息的准确性、完整性或可靠性不做任何保证。本文所载的资料、意见及推测仅反映头豹于发布本报告当日的判断，过往报告中的描述不应作为日后的表现依据。在不同时期，头豹可发出与本文所载资料、意见及推测不一致的报告和文章。头豹不保证本报告所含信息保持在最新状态。同时，头豹对本报告所含信息可在不发出通知的情形下做出修改，读者应当自行关注相应的更新或修改。任何机构或个人应对其利用本报告的数据、分析、研究、部分或者全部内容所进行的一切活动负责并承担该等活动所导致的任何损失或伤害。

业务合作

会员账号

可阅读全部原创报告和百万数据，提供PC及移动端，方便触达平台内容

定制报告/词条

行企研究多模态搜索引擎及数据库，募投可研、尽调、IRPR等研究咨询

定制白皮书

对产业及细分行业进行现状梳理和趋势洞察，输出全局观深度研究报告

招股书引用

研究覆盖国民经济19+核心产业，内容可授权引用至上市文件、年报

市场地位确认

对客户竞争优势进行评估和证明，助力企业价值提升及品牌影响力传播

云实习课程

依托完善行业研究体系，帮助学生掌握行业研究能力，丰富简历履历



业务热线

袁先生：15999806788

李先生：13080197867