

# 新一代智能终端蓝皮书

——从“人工智能+终端”到人工智能终端

(2025 年)

中国信息通信研究院

2026年3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,编者将追究其相关法律责任。



## 前 言

中国信息通信研究院在《新一代智能终端蓝皮书（2024年）》中提出，“新一代智能终端”是基于信息技术，以强感知、强计算、强交互、强体验为特征，能够执行多元化复杂任务，为用户提供强智能服务的新型智能终端。一年来，以大模型为核心的人工智能技术正引发终端智能化的二次革命。新一代智能终端已实现从“人工智能+终端”到“人工智能终端”的历史性跨越。

当前，人工智能终端已全面迈入规模化商用阶段。以AI手机、AI PC、智能可穿戴、人形机器人、智能网联汽车为代表的多元产品形态深度融入日常生活与产业场景，成为新一轮科技革命和产业变革的重要载体。《国务院关于深入实施“人工智能+”行动的意见》（国发〔2025〕11号）明确要推动智能终端“万物智联”，培育智能产品生态，大力发展智能网联汽车、人工智能手机和电脑、智能机器人、智能家居、智能穿戴等新一代智能终端。彰显了AI终端在服务高质量发展、创造高品质生活中的时代使命。

在这一进程中，人工智能终端展现出鲜明的“四新”特征——认知协同、场景预见、意图驱动、服务共生。这四大特征源于主动感知理解、多模态自然交互、智能化服务和自主学习进化四大核心能力，共同构成了AI终端从被动工具向主动伙伴演进的能力框架，标志着人机关系正发生从“工具使用”向“伙伴协作”的根本性转变。

支撑这一转变的是全栈技术的深度融合。硬件层面，

“CPU+GPU+NPU”异构计算架构成为行业标配，存储性能突破“存储墙”瓶颈，感知交互系统迈向多模态融合；软件层面，操作系统与 AI 深度融合构建端云协同智能中枢，终端智能体重新定义服务模式，安全与隐私体系为可信体验提供坚实保障。这种软硬协同演进与端云协同的技术重构，为“四新”特征的落地提供了系统性支撑。

在此战略指引下，本报告立足国家发展高度，深入剖析新一代智能终端的发展现状、技术特点等，旨在揭示 AI 终端如何通过全栈创新支撑“四新”能力落地，为相关领域的决策者、研究者及从业者提供全面深入的产业洞察与前瞻指引。

报告编制过程中得到了业内相关企业的大力支持，在此特别感谢华为终端有限公司、荣耀终端有限公司、维沃移动通信有限公司、欧珀广东移动通信有限公司、北京小米移动软件有限公司、北京三星通信技术研究有限公司、高通无线通信技术（中国）有限公司、联想（北京）有限公司、上海无问芯穹智能科技有限公司等。由于智能终端相关技术仍处于快速变化发展阶段，我们对新一代智能终端的认识还有待持续深化，报告中如存在不足之处，敬请读者批评指正。

# 目 录

一、政策与技术驱动，新一代智能终端全面迈向智能原生.....	1
（一）顶层设计层面，发展人工智能终端成为全球共识.....	4
（二）产业生态层面，人工智能终端迈入规模化商用阶段.....	5
（三）标准体系层面，健全人工智能终端测评体系.....	6
（四）产品体验层面，人工智能终端智能化水平持续提升.....	8
二、“四新”特征凸显，人机交互范式深刻演进.....	22
（一）更懂你：多模态智能交互升维至认知协同.....	22
（二）更贴心：主动服务进化为场景预见.....	23
（三）更省事：应用架构重塑为意图驱动.....	24
（四）更连贯：设备互联迈向场景化服务共生.....	25
三、硬件底座全栈演进，夯实 AI 终端核心能力.....	26
（一）芯片架构持续演进，强化端侧 AI 核心算力.....	27
（二）存储带宽与容量持续升级，破解“存储墙”瓶颈.....	32
（三）多模态感知交互体系升级，迈向感知与表达一体化.....	33
（四）通信与 AI 双向赋能，共筑 AI 终端“四新”特质基础.....	44
（五）能效管理多路径协同，平衡性能与续航矛盾.....	45
四、软件系统智能重构，定义可信服务新范式.....	48
（一）操作系统与 AI 深度融合，构建端云协同智能中枢.....	48
（二）终端智能体定义服务新范式，开启服务找人时代.....	50
（三）安全与隐私体系筑牢，构建可信 AI 终端基石.....	56
五、以 AI 手机和 AI PC 为突破口，加快实现“人工智能+”终端普及目标.....	60
（一）强化政策牵引，优先支持 AI 手机与 PC 成为：“人工智能+”落地主阵地.....	61
（二）加快标准统一，打通“一次开发、多端部署”的技术底座.....	61
（三）突破关键技术，全面提升端侧智能性能与能效平衡.....	62
（四）构建安全可信生态，推动开源协同与用户权益保障并重.....	62

## 图目录

图 1 AI 终端全栈技术架构与“四新”能力演进全景图.....	4
图 2 AI 手机辅助通话场景测试结果.....	12
图 3 AI 手机影像处理 L3 级任务完成度测试结果.....	12
图 4 AI 手机影像处理子场景测试结果.....	13
图 5 AI 手机智能问答场景测试结果.....	14
图 6 手机处理器技术发展技术图谱.....	28
图 7 商用 AI PC 端侧芯片支持大模型能力.....	32
图 8 手机显示屏四阶段演进技术.....	34
图 9 智能手机全面屏解决方案.....	34
图 10 手机光学摄像头技术的发展.....	37
图 11 2023—2025 年主流旗舰手机技术参数水平.....	38
图 12 端侧本地模型耗电占比.....	46
图 13 终端快充技术功率进阶历程.....	47

## 表 目 录

表 1 手机代表产品发布情况汇总.....	10
表 2 AI PC 代表产品发布情况汇总.....	16
表 3 AI 赋能的可穿戴设备发布情况汇总.....	21
表 4 新一代 AI 终端交互范式演进对比.....	22
表 5 硬件底座演进及其对“四新”能力的支撑.....	26
表 6 主流手机芯片厂商最新旗舰芯片关键 IP 配置汇总表.....	28
表 7 2024—2025 年主流厂商处理器芯片型号及核心技术参数.....	29
表 8 电声技术发展.....	40
表 9 麦克风在智能终端人机交互的作用.....	41
表 10 MEMS 扬声器 vs 传统扬声器.....	43
表 11 2025 年各主流品牌热门机型的电池典型容量.....	46
表 12 软件系统重构及其对“四新”能力的支撑.....	48

## 一、政策与技术驱动，新一代智能终端全面迈向智能原生

当前，新一代智能终端正朝着以大模型为基础、智能体为核心的智能原生方向发展，最终将实现万物智联、协同智慧的全新范式。在这一进程中，新一代智能终端的内涵已发生深刻演进——人工智能不再是终端的附加功能，而是深度融入芯片、操作系统、感知交互与连接架构的智能内核。由此，新一代智能终端已全面具象化为人工智能终端。

人工智能终端是指具备主动感知理解、多模态交互、智能化服务和自主学习进化等能力，结合了端侧和云端的人工智能能力，独立或协同完成特定人工智能任务的终端产品。

人工智能终端不仅能够接收用户的明确指令，还能主动感知用户需求、理解使用场景，并通过语音、图像、手势、文字等多种方式进行自然流畅的交互。更重要的是，它能基于对环境和用户行为的理解，提供个性化、前瞻性的服务，甚至在使用过程中不断优化自身表现，实现“越用越聪明”的效果。

当前，AI 终端已进入规模化商用阶段，典型产品形态包括 AI 手机、AI PC、智能可穿戴设备（如智能手表、AR/VR 眼镜）、人形机器人以及智能网联汽车等。这些设备不再是孤立的工具，而是嵌入生活与工作全场景的智能节点，正在深刻改变人机关系的本质。本报告聚焦在 2025 年智能化服务水平进一步提高的 AI 手机、AI PC 和日益成熟的 AI 眼镜等可穿戴设备，深入分析其 AI 能力增强背后

的全栈能力演进——硬件协同突破及软件系统性升级。

在实际应用中，人工智能终端展现出“四新”的鲜明特征：**认知协同、场景预见、意图驱动、服务共生**。这四大特征并非凭空而来，而是建立在**主动感知理解、多模态自然交互、智能化服务、自主学习进化**四大核心能力基础之上的自然演进。

上述“四新”特征的实现，离不开底层全栈技术的系统性突破。AI 终端的发展不是单一技术的胜利，而是**硬件、软件、算法、安全与生态协同演进**的结果（图 1）。

在硬件层面，传统的 CPU（中央处理器）已无法满足 AI 计算的高并发、低延迟需求。如今，“**CPU+GPU+NPU**”异构计算架构已成为行业标配。其中，NPU（神经网络处理器）专为 AI 任务设计，能高效处理图像识别、语音合成、大模型推理等复杂运算。同时，存储技术也取得关键突破，通过高速缓存、存算一体等方案，有效缓解了长期制约性能的“存储墙”问题。此外，感知与交互硬件全面升级，高清摄像头、高灵敏度麦克风阵列、毫米波雷达、触觉传感器等多模态感知单元协同工作，为终端提供了“眼耳鼻舌身”般的综合感知能力。

在软件层面，操作系统正经历一场深刻的 AI 化重构。新一代 AI 操作系统不仅调度计算资源，更构建了一个端云协同的智能中枢。本地端负责实时响应和隐私敏感任务，云端则提供强大的模型训练与知识更新能力。两者无缝衔接，既保障了响应速度，又实现了持

续进化。与此同时，**终端智能体（Agent）**概念兴起，每个 AI 终端都可视为一个具备目标、规划与执行能力的智能体，能自主调用应用、协调服务、完成复杂任务。例如，用户只需说“帮我准备明天的出差”，智能体就能自动订票、打包行李清单、同步会议资料，全程无需人工干预。

安全与隐私是 AI 终端普及的前提。为此，行业构建了多层次可信体系：从硬件级的安全芯片，到软件层的数据加密与权限管理，再到用户可控的隐私设置，确保智能服务在“可用不可见”的原则下运行。只有让用户放心，AI 终端才能真正融入生活。

2025 年，在国家“人工智能+”战略和全球技术创新浪潮的双重推动下，AI 终端产业迎来爆发式增长。一方面，产品形态更加多元：AI 手机强调个人助理能力，AI PC 聚焦生产力提升，可穿戴设备关注健康与沉浸体验，人形机器人探索家庭与公共服务场景，智能网联汽车则成为移动的第三生活空间。另一方面，应用场景全面渗透：从智能家居、智慧办公，到智能制造、智慧医疗、智慧教育，AI 终端正成为连接数字世界与物理世界的桥梁。

尤为关键的是，产业生态日趋成熟。芯片厂商、终端制造商、操作系统开发者、大模型公司、应用服务商等各方紧密协作，形成了从底层技术到上层应用的完整链条。开源社区、标准制定、开发者工具等基础设施的完善，也加速了创新落地。

可以预见，在不远的将来，每个人身边都将拥有一个或多个 AI

终端，它们了解你的习惯、预判你的需求、守护你的隐私、陪伴你的成长。这场由 AI 终端引领的人机协作新时代，已然全面开启。



来源：中国信息通信研究院

图 1 AI 终端全栈技术架构与“四新”能力演进全景图

## （一）顶层设计层面，发展人工智能终端成为全球共识

一是全球主要经济体正围绕人工智能终端展开战略部署。美国于 2025 年 7 月发布《赢得竞赛：美国人工智能行动计划》，明确将智能网联汽车、机器人列为人工智能将在物理世界带来广泛创新的领域，驱动汽车制造商和技术公司投资于 AI 驱动的车载计算平台、传感器融合系统等；欧盟《人工智能法案》聚焦于自动驾驶、工业设备与消费电子三大终端领域，通过严格监管确保 AI 技术安全融入社会应用。法案将前两者的 AI 系统划为“高风险”类别，要求企业遵循最高安全标准并完成强制性 CE 认证，以保证技术可靠落地。同时，法案对消费电子中的通用大模型施加透明度规定，并限制情

感识别等敏感应用，以在发展中保护公民权利。这些举措共同彰显了欧盟在推进终端智能化过程中“可信 AI 优先”的战略导向；日本以“社会 5.0”战略推动服务机器人等终端在社会场景落地；韩国则聚焦 AI 芯片等强化技术主权。

二是国家人工智能战略持续深化，发展人工智能终端已成为该战略的重要实施环节。2025 年《政府工作报告》强调要“支持大模型广泛应用，大力发展智能网联新能源汽车、人工智能手机和电脑、智能机器人等新一代智能终端”。国务院《关于深入实施“人工智能+”行动的意见》（国发 202511 号）明确提出，“大力发展智能网联汽车、人工智能手机和电脑、智能机器人、智能家居、智能穿戴等新一代智能终端，打造一体化全场景覆盖的智能交互环境”。这标志着新一代智能终端不仅是技术创新的产物，更是国家培育新质生产力、构建现代化产业体系战略支点的，其发展被赋予了服务高质量发展、创造高品质生活的时代使命。相较国外，我国依托超大规模市场、全球最完整的智能终端产业链及制度化的安全治理框架，在人工智能终端领域自然形成了“场景驱动、生态协同、安全可控”的系统性发展路径；在强化技术创新的同时，正加快构建覆盖芯片、系统、应用、数据的全链条治理体系，为全球 AI 产业演进贡献兼具创新活力与制度韧性的中国实践。

## （二）产业生态层面，人工智能终端迈入规模化商用阶段

一是全球人工智能终端供应链趋于成熟，我国实现人工智能终

**端产业关键技术突破。**全球产业链上下游企业正围绕端侧大模型、异构算力、智能体操作系统等构建全栈能力，从芯片（高通、联发科、华为昇腾等）、操作系统（鸿蒙、澎湃 OS、ColorOS）到端侧大模型与智能体框架，人工智能终端产业链各环节已形成协同发展的良好生态和成熟的技术栈。我国厂商在关键赛道实现系统性突破，逐步建立起与全球巨头并跑甚至领跑的能力，推动行业走向多极化、融合化的跨设备智能新阶段。

二是新一代智能终端产业生态呈现爆发式增长，AI 手机、AI PC、AI 眼镜等人工智能终端产品加快走进千家万户。Canalys 预测，2025 年全球 AI 手机渗透率将达到 34%，AI PC 出货量将超 1 亿台，占 PC 出货总量的 40%以上，2025 年上半年国内可穿戴腕带设备出货量达 3390 万台，同比增长 36%。IDC 预测，2025 年全年 AI 眼镜市场出货量预计达到 1,451.8 万台，到 2029 年全球智能眼镜年销量有望突破 5500 万副，市场规模将达千亿元级别。据 Counterpoint 统计预测，随着用户对健康监测需求升级，加速推动消费电子需求和技术迭代（如传感器精度提升、AI 算法优化），预计到 2034 年将达到 4317 亿美元，复合年均增长率（CAGR）为 19.59%。

### （三）标准体系层面，健全人工智能终端测评体系

中国计算机行业协会人工智能产业工作委员会下设的人工智能终端工作组，联合全国信息技术标准化技术委员会人工智能分技术委员会（SAC/TC28/SC42）、全国音频、视频及多媒体系统与设备

标准化技术委员会（SAC/TC242）下设的人工智能终端标准组，围绕人工智能终端智能化水平不统一、用户认知模糊等产业痛点，广泛汇聚国内产学研用各方力量，系统开展《人工智能终端智能化分级》系列国家标准研制工作。该系列标准由全国信息技术标准化技术委员会提出并归口，以《第1部分：参考架构》为顶层指引，《第2部分：总体要求》为通用基准，分品类制定了移动终端、微型计算机、电视等多类终端的专项标准，首次构建了覆盖“感知、认知、执行、记忆、学习”五大能力域，区分L1响应级、L2工具级、L3辅助级的可量化、可测试、场景化的智能化分级框架与测评方法。该标准体系不仅填补了我国在AI终端评价领域的空白，为产品设计、用户选型、市场监管和政策制定提供了权威技术依据，更通过推动产业从“参数堆砌”向“能力导向”转型，有效引导人工智能终端高质量发展，加速端侧智能与端云协同生态成熟，对提升我国在全球AI终端产业中的竞争力、保障技术主权与供应链安全具有重要战略意义。

与此同时，电信终端产业协会（TAF）和中国通信标准化协会（CCSA）等协会积极协同推进配套标准体系建设。TAF聚焦场景化落地和交互，在智能服务分级方面，已发布《人工智能终端智能服务能力分级指南》，并正在制定《人工智能手机智能服务能力分级场景实践指南》系列，推动分级标准向具体用户体验转化以及终端智能体交互；在智能体交互方面，已经发布《智能终端意图框架总

体技术要求》、《智能终端意图框架接口技术要求》，同步开展《面向移动智能终端的智能体通信协议技术要求》系列和《手机智能体技术研究报告》的研究，推动终端智能体同应用、其他智能体和设备间交互。CCSA TC11 则重点布局生成式 AI 在终端侧的技术要求与安全治理，立项《生成式人工智能终端智能服务技术要求和测试方法》标准，编制《人工智能 安全治理 人工智能终端安全技术要求》标准草案，同步开展《生成式 AI 和大模型在终端侧的应用和标准化需求研究》课题，夯实技术底座与合规边界。

为落实国家“人工智能+”行动战略，2025 年 12 月，工业和信息化部组织筹建中国人工智能终端行业协会，助力我国人工智能终端产业高质量发展。

#### （四）产品体验层面，人工智能终端智能化水平持续提升

AI 手机、AI PC 已实现初步商用并形成代表性产品，AI 眼镜、AI 耳机在健康监测、多模态交互、轻量化 AI 部署等方面形成独特技术范式，人形机器人、智能网联汽车等系统复杂度高、场景融合性强的产品则正处于加速商用落地的关键节点。从消费终端到产业终端、从单体智能到系统智能，人工智能终端产品智能化水平正不断提高，加速催化新应用落地。

##### 1. AI 手机智能化服务水平已达到智能助理级<sup>1</sup>

<sup>1</sup> AI 手机智能化服务水平分级依据为《T/TAF 289—2025 人工智能终端智能服务能力分级指南》。该指南将 AI 手机智能化服务水平分为 L1（智能响应级）、L2（智能辅助级）、L3（智能助理级）、L4（智能协同级）、L5（自主智能级）5 级。

（1）智能服务能力持续演进，智能化水平达到智能助理级

2025 年，AI 手机智能化服务能力与垂直场景深度融合，“场景化”“主动化”服务成为重构用户体验、定义市场新价值的战略锚点。2024 年 AI 手机市场整体处于探索与尝鲜期，AI 功能多以单点演示为主，如生成式图片创作、文本生成、内容总结等，仅有少部分机型支持复杂任务自动化执行。2025 年，手机厂商战略重心已从单点功能的迭代优化转向场景智能的深度创新，产品升级策略不再仅局限于硬件参数的提升或独立 AI 功能的堆砌，更多转向智能服务生态的构建与复杂 AI 应用场景的渗透，通过开放底层 AI 能力与第三方服务深度融合，不断探索更多场景化的创新应用，代表产品及应用见表 1。AI 手机智能化服务能力持续演进，智能化水平达到智能助理级。例如，在出行、购物、办公等垂直场景中，手机可整合多源数据与服务，提供一站式主动解决方案。智能助理级（L3 级）服务主要基于系统级智能体，能够理解用户模糊意图，识别用户情感状态，编排多步骤任务执行路径，实现任务闭环操作，并对任务状态进行评估。智能服务等级越高，智能化水平程度越高。随着技术的演进、终端产品的迭代、应用场景的升级等多种变化，将会出现更高等级的智能服务。

表 1 手机代表产品发布情况汇总

公司名称	代表产品	发布时间	模型名称	部署方式	芯片	代表性 AI 应用
小米	17Pro Max	2025.09	MiLM	端云协同	骁龙 8 至尊版	一键问屏、智能提醒
维沃	X300 Pro	2025.10	蓝心大模型	端云协同	天玑 9500	图像风格化、智能消除、通话摘要、文本总结
欧珀	Find X8 ultra	2025.04	安第斯大模型	端云协同	骁龙 8 至尊版	一键问屏、AI 识万物、AI 翻译、AI 灵感成片、小布记忆
华为	Pura 80 Pro	2025.06	盘古大模型	端云协同	/	AI 连续翻译、AI 隔空传送、智能提醒、通话摘要
荣耀	Magic 8 pro	2025.10	魔法大模型	端云协同	骁龙 8 至尊版	AI 追色、YOYO 玩图、一键领券购物，一键美食服务
中兴、字节跳动	努比亚 M153	2025.12	豆包大模型	端云协同	高通骁龙 8 至尊版	一键打车、购物、领取优惠券、多模态实时交互问答
苹果	iPhone 17	2025.09	/	端云协同	苹果 A19	日程创建、文本问答
三星	Galaxy Z Fold7	2025.07	/	端云协同	骁龙 8 至尊版	多模态输入/输出、智能拖放、即圈即搜、实时字幕翻译、语聊世界

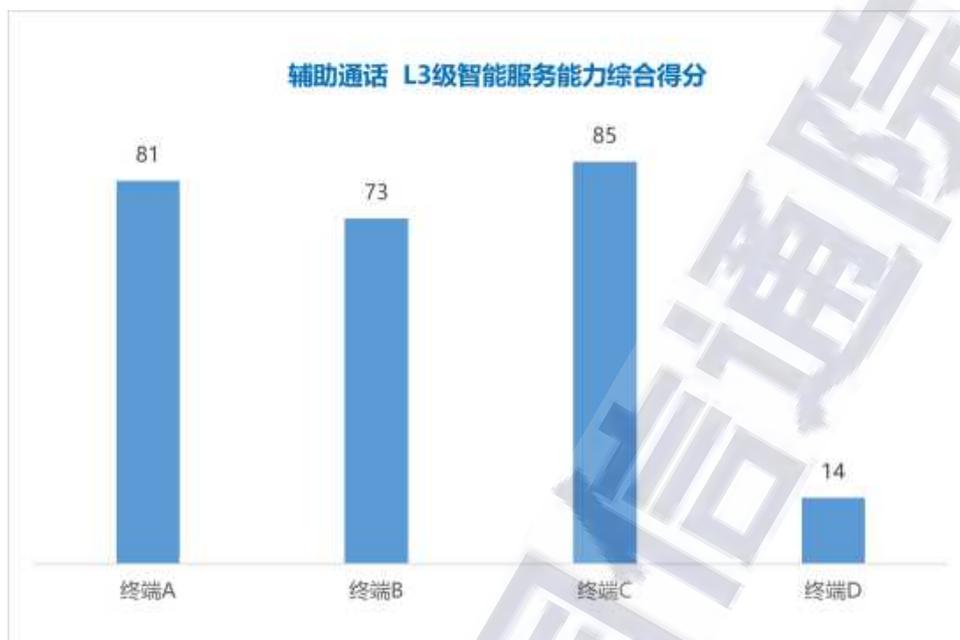
来源：公开信息整理

依据终端服务场景特点差异，将智能服务初步划分为基础型智能服务和扩展型智能服务。基础型智能服务主要以系统工具应用为核心，依托终端厂商对操作系统底层及系统工具的深度优化，实现能力的快速演进。代表性场景包括用机操控、辅助通话、影像处理、智能问答、内容创作、系统优化。此类服务因垂直整合程度高、生态可控性强，2025 年智能化水平提升显著，L3 级场景覆盖度更高。

相比 2024 年，辅助通话和影像处理两大场景应用实现突破性进展。辅助通话代接能够进行智能识别与自动代接，支持多轮智能应答、自定义开场白、多场景回复等，确保重要信息不遗漏且能有效拦截干扰电话，4 款测试终端的辅助通话 L3 级智能服务能力综合得分情况见图 2。通话摘要能够自动提取通话中的关键信息（如时间、地点、任务、联系人等），生成结构化文本摘要，支持用户快速查阅核心内容，并可同步至日历、待办事项或分享至其他应用。影像处理层面，终端不仅能够完成智能消除、自由扩图、图片风格化、一键成片等复杂功能，还能通过自由指令对图片进行消除等操作，极大提升了用户处理图像的便捷性。6 款测试终端的影响处理测试结果见图 3 和图 4。

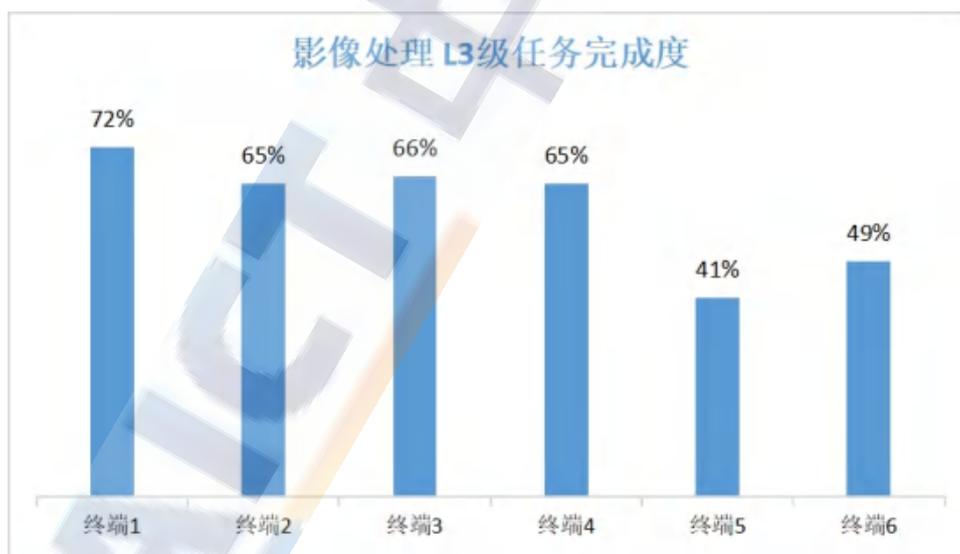
基础型智能服务已成为当前旗舰产品的标配能力，用户体验也正从“可用”逐步迈入“好用”阶段。**扩展型智能服务**代表了智能生态的融合方向，需终端与第三方应用通过 API 调用、服务组件化、生态互联等方式协同完成，该类服务更关注复杂意图理解、跨应用任务自动化执行。代表性场景包括生活服务、交通出行、智慧办公、通讯社交、休闲娱乐、学习教育。目前，在生活服务、交通出行、智慧办公等领域的部分子场景中，L3 级智能服务已初步落地并显现价值，但仍面临生态壁垒高、接口标准不一、动态权限管理等挑战。未来，随着终端厂商与应用厂商协同加深、AI 中间件能力增强以及模型轻量化部署进一步成熟，该类服务将具备更广泛的场景覆盖与

更流畅的用户体验，成为下一代智能服务竞争的核心高地。



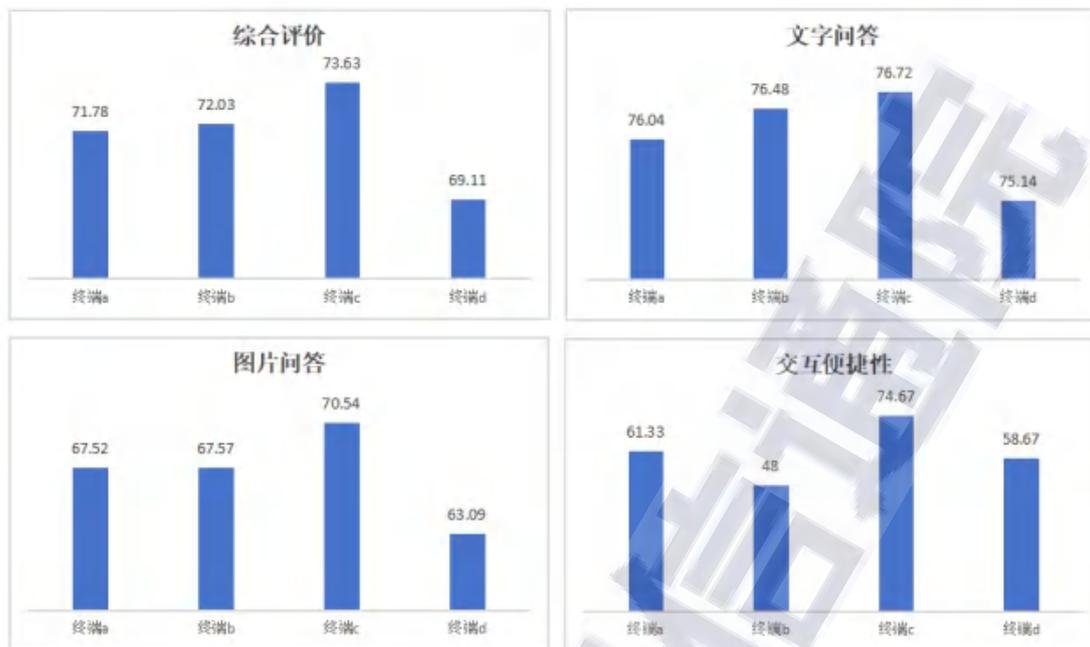
来源：中国信息通信研究院

图 2 AI 手机辅助通话场景测试结果



来源：中国信息通信研究院

图 3 AI 手机影像处理 L3 级任务完成度测试结果

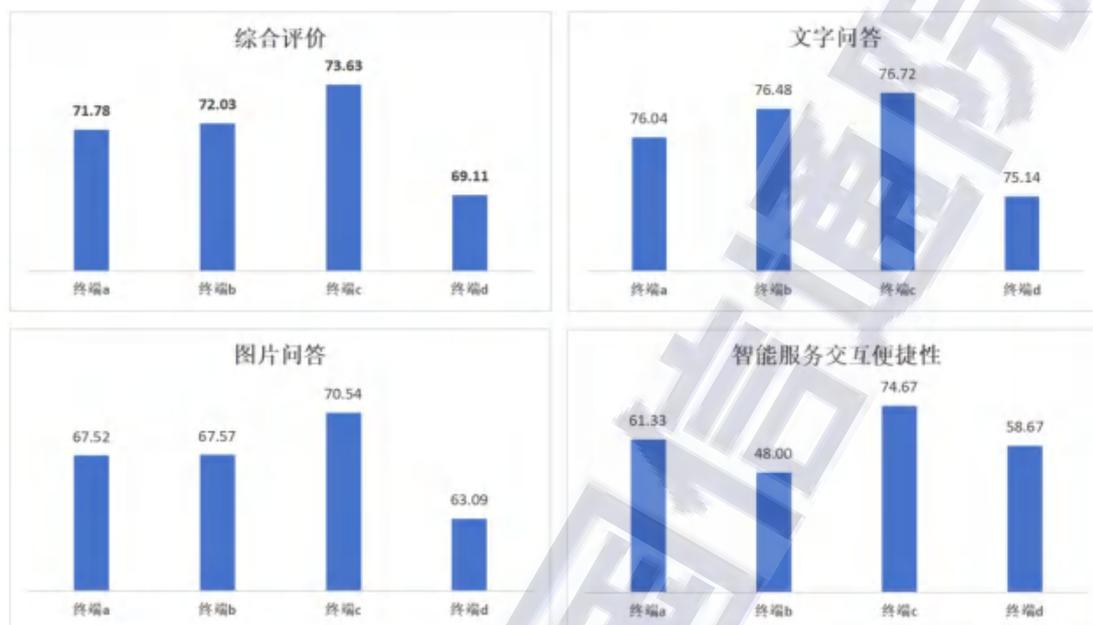


来源：中国信息通信研究院

图 4 AI 手机影像处理子场景测试结果

基于强大的端侧计算与云端协同能力，智能手机不仅实现了高效精准的文字问答和知识检索，更在图像理解、拍照识别和多模态实时交互方面取得突破性进展。随着人工智能技术的加速落地与持续迭代，智能手机正日益成为用户获取信息与服务的核心入口。其能力边界已从早期的简单信息检索，全面扩展至深度融合多模态交互的智能新阶段。拍照识别、圈选识屏、多模态实时交互等应用的落地，使得用户能够通过自然语言、相机和跨应用指令，随时随地获取所见即所得的答案、服务与操作支持，显著提升了信息获取的效率和场景应用的广度。中国信息通信研究院测试数据显示，4款主流旗舰手机在知识回答准确性、多模态识别能力等方面整体表现良好，但在图像理解、交互便捷性等能力上仍存在差异（图 5）。其中，部

分机型在拍照识别与圈选识屏等场景具备一定的优势，展现出较强的实用价值，而在复杂交互对话与多模态融合能力方面仍有优化空间。



来源：中国信息通信研究院

图 5 AI 手机智能问答场景测试结果

## （2）“自动化任务”重构任务处理范式，重塑人机交互体验

“自动化任务”重构复杂任务处理范式，将传统“多轮交互、分布执行”的繁琐流程压缩为单一指令驱动的端到端闭环执行，实现从用户意图表达至目标落地的全链路自动化运转。在自动化任务中，用户仅需通过一句指令表达意图，手机便能理解用户意图并自动拆解任务逻辑，构建从需求提出到结果交付的全服务流程闭环。任务自动化执行已从简单的指令执行演进为深度场景化的智能服务，如“一句话修图”“一句话导航”“一句话打车”等为代表的端侧高频服务已在不同终端产品实现落地应用。该类服务显著降低了用

户在复杂任务中的认知负荷与操作时延，重塑了移动设备的核心交互范式与服务价值。随着屏幕交互、智能体互联网等技术的日趋成熟，自动化任务将覆盖更广泛的场景，显著提升移动终端在复杂服务场景下的智能响应能力。

未来，竞争焦点将转向生态开放性与场景覆盖深度，终端厂商、开发者和服务提供商将共同构建下一代“感知-决策-服务”闭环体验，助力“人工智能+”消费提质政策的落地推广。

## 2. AI PC “生产力工具”能力持续进阶

当前，AI PC 凭借端侧智能革新，深度融合硬件架构创新、软件生态重塑与多场景应用拓展，正从初步探索迈向成熟商用阶段，成为赋能个体生产力跃升和驱动行业数字转型的关键力量（表 2）。相较于 2024 年，随着 AI PC 硬件算力显著提升、智能体能力增强、AI 场景应用拓展，AI PC “生产力工具”能力持续进阶。例如最新发布的高通骁龙 X2 Elite NPU 算力已达 80 TOPS，较上代提升 78%（相关 AI PC 终端预计于 2026 年上半年上市）；智能体可通过一句话指令识别模糊意图、自主规划并执行多步骤任务，如 OpenAI 的 Operator 智能体能自动完成网页操作、订票、填表等复杂任务；应用场景方面从单一功能辅助升级为全流程办公生产力中枢，支持本地知识库检索与业务数据深度分析，实现会议、办公、创作、行业等丰富场景应用。

表 2 AI PC 代表产品发布情况汇总

公司	代表产品	发布时间	搭载处理器	搭载模型	智能体	代表性 AI 应用
联想	YOGA Air 14 Aura AI 元启版	2025.2	英特尔酷睿 Ultra 7 处理器等	天禧大模型	天禧 AI	适配办公、协作与娱乐等场景；AI 操控、AI 搜索、AI 翻译、AI 笔记、AI 服务等五大功能
联想	ThinkBook 16p AI 元启版	2025.4	英特尔酷睿 Ultra 7/9, AMD 锐龙 9 8945 HX 处理器等	天禧大模型	天禧 AI	AI 文档处理与协同工作等办公应用；AI 优化使用体验与硬件性能等设备控制；AI 赋能设计与内容生产效率等专业应用
联想	ThinkBook Flip AI PC 概念机	2025 年世界移动通信大会	英特尔酷睿 Ultra 7 处理器等	天禧大模型	天禧 AI	适配办公、协作与娱乐等场景；AI 操控、AI 搜索、AI 翻译、AI 笔记、AI 服务等五大功能
荣耀	荣耀 MagicBook Pro 14	2025.2	英特尔酷睿 Ultra 5/9 处理器	魔法大模型	YOYO 智能体	在搜索、阅读、创作和笔记四大应用场景下，支持端侧智慧搜索、隐私投屏、声纹降噪以及荣耀笔记等 AI 功能
荣耀	荣耀 MagicBook Art 14	2025.7	英特尔酷睿 Ultra 7 处理器	魔法大模型	YOYO 智能体	支持语音指令，可完成系统功能、自研应用、第三方应用以及跨设备复杂任务，如搜索文件等
华为	HUAWEI MateBook Pro / Fold	2025.5	麒麟 X90 处理器（根据外界推测）	盘古大模型等	小艺系统级智能体	小艺慧记、小艺知识空间、小艺文档助理、小艺设备专家、小艺翻译等 AI 能力
小米	REDMI Book Pro 2025 系列	2025.2	英特尔酷睿 Ultra 7/5 处理器	MiMo 模型	小爱助手	AI 视频内容搜索、AI 文档处理等 AI 能力；支持小米生态系统
微软	Windows Surface Pro 12 / Surface Laptop 13	2025.7	高通骁龙 X Plus 处理器	Phi-Silica 模型	Microsoft 365 版 Copilot AI 代理	Recall 功能、Cocreator 共同创作、Windows Studio 图像视频处理、实时字幕和自动超级分辨率

来源：公开信息整理

全球主要企业加快 AI PC 布局步伐，抢占新一轮产业变革竞争制高点。国际层面，谷歌、微软、Meta、苹果、高通、英伟达等头部主体在端侧模型、芯片算力、终端赋能等领域持续发力。微软在 AI 功能上持续推进，如升级 Windows Copilot Runtime 到 Windows AI Foundry，为开发者提供更好的模型部署环境，提供由内置 AI 模型支持的即用型 AI API 等新特性，提升 Copilot+PC 的 AI 应用和性能。国内层面，联想提出“天禧 AI，一体多端”战略，围绕天禧个人超级智能体，加速 AI PC、AI 平板等多端 AI 赋能应用、智能体生态建设和统一 AI 交互体验；华为推出基于鸿蒙 OS 的 AI PC，打造生态体系，推出小艺智能体开放平台，构建操作系统、鸿蒙应用/元服务与智能体之间交互协同范式；荣耀升级 YOYO 智能体，支持语音指令实现复杂任务，构建基于智能体的人机交互和 AI 服务流转的跨端生态。

当前 AI PC 软硬件技术持续升级，赋能 AI 能力提升。一方面硬件基础层面，AI PC 芯片架构、算力、工艺、存储提升，以更高效支持大模型运行和赋能；另一方面 AI PC 软件基础创新优化以提高推理性能，AI 智能体应用加速赋能。

一是软件基础层面，通过和硬件协同优化不断突破端侧性能提升与能效控制，以在有限硬件条件下实现高效推理。推理性能上，通过高性能并行解码与算子融合，提升大模型的响应速度；资源管理上，通过显存精细化调度与异构资源优化，缓解内存占用压力；能效方面，借助量化与稀疏化技术等，实现低功耗与高精度的平衡。

例如，荣耀提出 HONOR Turbo X 技术，通过全域 AI 调校，覆盖芯片至应用层和 NPU 驱动优化，从内到外实现软硬件系统调优能力，实现离电性能不减、下载速度翻倍等突破。无问芯穹 Mizar 推理加速引擎为联想 AI PC 等提供软硬协同的优化路径，提升推理性能的同时优化能耗。

二是软件应用层面，AI PC 智能化交互和个性化服务能力提升，智能体“超级应用”新生态加速形成。AI PC 智能化交互实现从“指令响应”向“意图预判”转化，能够更深刻理解用户潜在需求，自动完成或辅助完成办公、生活等常规应用场景任务；例如微软 Edge 加入 Copilot 模式，通过分析当前页面内容、浏览历史及用户行为模式，预测下一步操作需求。同时，大模型驱动应用生态向 AI 终端智能体生态的演进速度加快，对软件的应用架构与模式进行全新重构。智能体“超级应用”加速赋能，能够实现跨平台、跨场景的智能服务，例如可直接调用 Windows Copilot 实现“文档智能总结+实时会议翻译”。国内方面，联想“天禧个人超级智能体”作为用户专属的面向 AI 终端交互的形态，采用端云一体的混合架构，具备对用户意图的主动感知、任务自主规划与编排执行等能力。荣耀 YOYO 助理智能体支持自然语义理解和操控，提升创作与跨端互传效率，如新增 AI PPT 生成、代码创作工具和实时翻译智能体，并通过“一句话的事”智控语音指令完成系统级操作。此外，AI PC 通过垂域智能体、专业辅助工具或插件，能够辅助用户完成设计、金融等专业

领域的复杂任务，辅助用户提升效率。

### 3. AI 可穿戴设备从功能附加到能力重构

相较于 2024 年，AI 眼镜从音频眼镜、拍摄眼镜已逐渐升级到支持画面显示的 AR+AI 眼镜，部分产品已实现双目全彩显示和健康指标检测，提供更多元的能力和服务，覆盖更广泛的应用场景（表 3）。AI 眼镜以独特的硬件特性——集摄像头、麦克风、显示屏与传感器于一体，具备了视觉、听觉、触觉等多模态感知能力，为 AI 提供了丰富的数据输入通道，成为 AI 可穿戴设备的热门发展方向。2025 年开年以来，AI 眼镜市场增长势头强劲，小米、阿里巴巴、雷鸟创新、Rokid、XREAL、影目科技、李未可等企业纷纷推出 AI 眼镜新品，普遍接入 Deepseek、通义千问、讯飞星火、Kimi 等主流大模型，实现语音交互、环境识别、实时翻译、场景理解与记忆辅助等高阶功能。高通骁龙 AR1、恒玄 2700 等专用芯片的成熟，为眼镜端侧提供了强大的 NPU 算力支持，使得本地化 AI 推理成为可能，显著降低了延迟并提升了隐私安全性。未来，AI 眼镜将进一步向“超轻量化”与“超低功耗”不断迭代发展，重量控制在 30 克以内、续航能力提升将成为关键突破点。同时，Micro LED 等新型显示技术的成熟，将推动 AR 与 AI 深度融合，实现虚实交融的沉浸式交互体验，真正成为用户随身的智能助手与信息入口。

相较于 2024 年，智能手表与手环在健康监测方面大幅升级，新增无创血糖趋势估算、血压连续校准及情绪压力分析功能，并依托

端侧 AI 实现更精准的健康预警（表 3）。智能手表和手环作为最早普及的产品类别，已从初级的运动监测工具发展为集健康管理、通信、支付于一体的综合载体。例如，苹果 Watch Series 和华为 WATCH GT 系列通过 AI 算法实现心率异常检测、睡眠质量分析和压力评估，部分产品获得医疗器械认证，初步具备医疗级诊断能力。华为 WATCH D 引入 AI 算法实现血压、血氧、睡眠质量甚至血糖趋势的无创监测。AI 模型对多维度生理数据的长期追踪与分析，能够提供个性化的健康预警与运动指导，从“感知智能”迈向“认知智能”。

相较于 2024 年，AI 耳机不再局限于降噪与语音助手，而是整合情境感知与语义理解，可在会议中实时生成摘要、区分说话人角色（表 3）。AI 耳机正从传统的音频播放设备，进化为具备环境感知、语音理解与主动服务能力的智能感知设备。以苹果 AirPods Pro、索尼 LinkBuds 及国产 TWS 新品为代表，AI 耳机已普遍集成主动降噪、语音唤醒、实时翻译和会议纪要等功能。通过嵌入式 AI 协处理器或专用 AI-NPU 架构，如炬芯科技 ATS323X、恒玄科技智能音视频 SoC，耳机能够在极低功耗下运行神经网络算法，实现语音增强与环境音识别。未来，AI 耳机将不仅服务于听觉，更将成为环境感知与情绪识别的入口，结合上下文理解与个性化推荐，提供真正“懂你”的智能服务，成为用户全天候的私人助理。

表 3 AI 赋能的可穿戴设备发布情况汇总

产品名称	产品类别	发布时间	公司名称	代表性 AI 应用
雷鸟 V3 AI 拍摄眼镜	AI 眼镜	2025 年第一季度	雷鸟创新	搭载定制多模态 AI 大模型，支持语音助手、AI 拍摄、翻译、物体识别
Rokid Glasses	AI+AR 眼镜	2025 年第三季度	Rokid	接入通义大模型，支持语音交互、AI 问答、识物、翻译
小米 AI 眼镜	AI 眼镜	2025 年第二季度	小米	对标 Ray-Ban Meta，集成 AI 助手、摄像头、音频功能，免提交互
Oakley Meta Vanguard 运动 AI 智能眼镜	AI 眼镜	2025 年 9 月	Meta-Oakley	集成 Meta AI 语音助手，实现免提的 3K 高清视频录制、实时环境问答和运动数据记录
Meta Ray-Ban Display	AI+AR 眼镜	2025 年 9 月	Meta-RayBan	集成 AI 语音助手，提供语音交互功能，支持全彩 AR 显示
华为 WATCH 5	AI 手表	2025 年 6 月	华为	搭载“腕上小艺”AI 助手，融合盘古与 DeepSeek 大模型，支持 AI 健康监测、智慧交互、全场景互联。
Apple Watch Series 11	AI 手表	2025 年 9 月	苹果	搭载 watch OS 和 S10 芯片，支持 Apple Intelligence 健康监测与个性化建议。
小米手环 9 Pro	AI 手环	2024 年 10 月	小米	小米澎湃 OS 赋能，支持 NFC 智能切卡、身体信息触发米家自动化、AI 运动数据分析。
讯飞 AI 会议耳机	AI 耳机	2025 年 5 月	科大讯飞	录音转写、实时翻译、AI 脑图生成、摘要、离线翻译、AI 嘴替、语音唤醒助理、深度多会议问询。
AirPods Pro 3	AI 耳机	2025 年 9 月	苹果	Apple Intelligence 驱动的实时语音翻译功能

来源：公开信息整理

## 二、“四新”特征凸显，人机交互范式深刻演进

人工智能与智能终端的技术融合，不是简单的功能叠加，而是人机关系的根本性重塑。终端正从听从指令的工具进化为懂你、帮你、伴你的伙伴。本章将从交互、服务、应用、互联四个维度，解读 AI 终端如何实现从被动响应到主动智能的跨越（表 4）。

表 4 新一代 AI 终端交互范式演进对比

核心维度	传统终端	AI 终端	带来的直观感受
交互方式	单一模态 (点按、语音)	认知协同 (眼神+语音+手势)	更懂你：像真人一样交流，无需死记指令
服务模式	被动响应 (用户点击触发)	场景预见 (基于习惯主动推荐)	更贴心：还没开口，服务已准备好
应用架构	流程驱动 (人适应软件)	意图驱动 (软件适应人)	更省事：一句话搞定跨软件复杂任务
设备关系	简单连接 (点对点传输)	服务接力 (跨端无缝流转)	更连贯：服务跟着人走，不受设备限制

来源：中国信息通信研究院

### （一）更懂你：多模态智能交互升维至认知协同

传统的通过语音、触控、视觉等彼此割裂的交互模式，已难以满足用户对自然、高效、智能体验的核心诉求。AI 终端通过跨模态语义融合（Semantic Integration）、隐式意图（Implicit Intention）的精准识别和交互的连续性与记忆能力，实现了认知协同的交互模式。这一能力源于端云协同的 AI 架构：终端基于端侧 AI 能力，结合个人知识图谱与实时情景感知，初步实现低延迟、高隐私的意图推理；云端大模型则提供海量知识与复杂规划能力，二者协同构成端云混

合智能架构。作为 AI 终端的重要标志，认知协同不仅重塑了人机交互体验，更是全场景智能交互环境构建的核心支撑。

尽管认知协同交互已取得长足进步，但仍面临诸多技术挑战：一是在复杂场景下，多模态信号的异构性与时空对齐难题，导致跨模态语义融合准确率不高。二是当前大模型仍易出现幻觉或知识盲区，使得跨模态推理结果失真，影响任务执行的可靠性。三是多模态感知依赖对用户行为、生物特征与环境数据的全面采集，如何在保障个性化服务的同时，通过透明可控的授权机制划定数据使用边界、避免隐私过度暴露，是赢得用户信任的关键。

## （二）更贴心：主动服务进化为场景预见

传统的指令响应式服务，已难以满足用户对主动、连贯、个性化体验的核心诉求。AI 终端服务能力正从高效的指令响应能力，向更具智慧的场景预见能力跃迁。这一能力源于一套融合全域情景感知与行为意图推理的系统级智能架构：全域情景感知是通过融合多源传感数据、协同跨应用数据与感知网络状态，构建用户物理与数字世界的动态视图；意图推理是基于构建的个人知识图谱，运用事件关联与因果推理算法将用户隐式意图精准转化为主动服务指令，并完成跨应用任务执行。场景预见是 AI 终端情境感知与行为预判的核心体现，是构建以用户为中心的主动服务生态，实现“服务找人”的核心驱动力。

然而，AI 终端的场景预见能力仍面临如下挑战：一是动态复杂

场景下，情景感知的完整性与实时性不足，导致预测失准。二是行为模型依赖长期数据积累，冷启动用户或低频场景下样本匮乏，易引发服务误判。三是若缺乏用户可控的授权机制，过度主动的服务极易滑向侵扰式服务。因此，平衡主动服务与用户主权的关键，需将是否接受主动服务的决定权交还给用户。

### （三）更省事：应用架构重塑为意图驱动

传统的应用开发模式将 AI 作为附加功能，用户需在预设流程中逐步操作，体验割裂且效率低下，这一模式已难以支撑认知协同交互和场景预见服务的新需求。AI 终端正在推动应用架构从流程驱动向意图驱动演进。其核心是构建以大模型为大脑的智能体，实现自主拆解任务、调度跨应用服务、执行端到端闭环的能力。这一能力依赖两大技术支柱：一是操作系统正与 AI 深度融合，构建统一的感知、记忆与操控能力。二是端云协同的计算架构兼顾响应速度、隐私保护和功能强度。端侧大模型保障低延迟与隐私安全，处理高频、敏感任务；云端大模型提供海量知识与复杂规划能力。架构重塑标志着应用从功能容器进化为服务引擎，是支撑智能体生态落地的关键基础。

然而，架构重塑的深入推进仍面临三重挑战：一是跨应用操作的生态壁垒，不同厂商的服务接口与权限体系尚未统一，导致智能体调度能力受限。二是模型幻觉与任务规划偏差，大模型在复杂逻辑推理中易产生错误，影响任务执行可靠性。三是系统级 AI 的安全

风险，深度集成的原生 AI 一旦被恶意利用，可能触发越权操作，需构建覆盖输入过滤、输出核查与权限动态管控的纵深防御体系。

#### （四）更连贯：设备互联迈向场景化服务共生

传统设备互联多依赖私有协议和封闭生态，以点对点连接与指令交互为基础，但因缺乏统一标准，导致用户在不同品牌设备间频繁手动切换，以支撑一体化全场景覆盖的智能交互环境。AI 终端正推动万物智联从设备互联向场景化服务共生跃迁。这一能力源于端云协同架构与开放互操作标准的双重驱动。在端侧，操作系统通过资源虚拟化与分布式任务调度框架，将多设备整合为统一计算平台。在生态侧，以分布式软总线技术等跨设备互操作标准，结合意图驱动和 MCP 协议等，正逐步打破数据与服务的生态壁垒。场景化服务共生标志着 AI 终端从单机智能迈向群体智能的关键一步，是构建全域智能生态、实现“人工智能+”行动战略目标的核心支撑。

然而，其规模化落地仍面临三重挑战：一是生态割裂，跨品牌的数据格式、服务接口与权限体系上的不兼容，使得跨生态的意图级服务流转难以无缝实现。二是端云协同架构的可靠性瓶颈，多设备协同共同完成感知、决策与服务时，系统复杂度呈指数级上升，其可靠性不仅关乎体验，更关乎生命与安全。三是用户隐私自主意识提升，对授权机制提出更高要求，授权复杂、服务价值难以满足用户期待时，导致技术难以转化为实际服务。因此，亟需完善商业互信、技术协同、用户赋权的治理体系，才能实现真正的一体化全

场景覆盖的智能交互。

### 三、硬件底座全栈演进，夯实 AI 终端核心能力

2025 年，终端硬件技术的竞争已超越了单纯的参数堆砌，转向了对 AI 体验的全面支撑。芯片不再只是算得快，更要算得巧；传感器不再只是拍得清，更要看得懂。本章将从芯片、存储、感知、连接与能效五个维度，解析硬件如何通过底层进化，为 AI 终端的“四新”能力提供坚实的物理支撑（表 5）。

表 5 硬件底座演进及其对“四新”能力的支撑

硬件系统	核心变化	带来的直观价值	支撑“四新”特征
计算芯片	从“通用计算”到“专人专岗” (CPU 主导→CPU+GPU+NPU 异构协同)	<b>算得更快更省电：</b> 专门处理 AI 任务，让手机能本地运行大模型，反应无延迟。	<b>认知协同</b> (保障实时交互流畅)
存储系统	从“仓库扩容”到“智能调度” (单纯加容量→存算一体+高速通道)	<b>存取零等待：</b> 不仅装得多，还能预判你会用什么数据，提前调取，让服务切换如丝般顺滑。	<b>场景预见</b> (支撑主动服务预加载)
感知交互	从“记录信息”到“理解环境” (高清录制→多模态感知理解)	<b>拥有“数字五官”：</b> 摄像头和麦克风能像人眼人耳一样，不仅记录画面声音，还能看懂场景、听懂情绪。	<b>认知协同</b> (实现多模态自然交互)
网络连接	从“高速管道”到“通感一体” (追求网速→连接与感知融合)	<b>时刻在线不掉线：</b> 网络能感知你的位置和动作，协同周边设备，确保服务像接力赛一样不中断。	<b>服务接力</b> (保障跨端服务连续)
能效管理	从“加大电池”到“精打细算” (电池扩容→AI 动态能效平衡)	<b>续航更持久：</b> 在算力暴增的同时，通过 AI 智能分配电量，解决“高智商”带来的“高能耗”焦虑。	<b>全场景支撑</b> (所有体验的基础)

来源：中国信息通信研究院

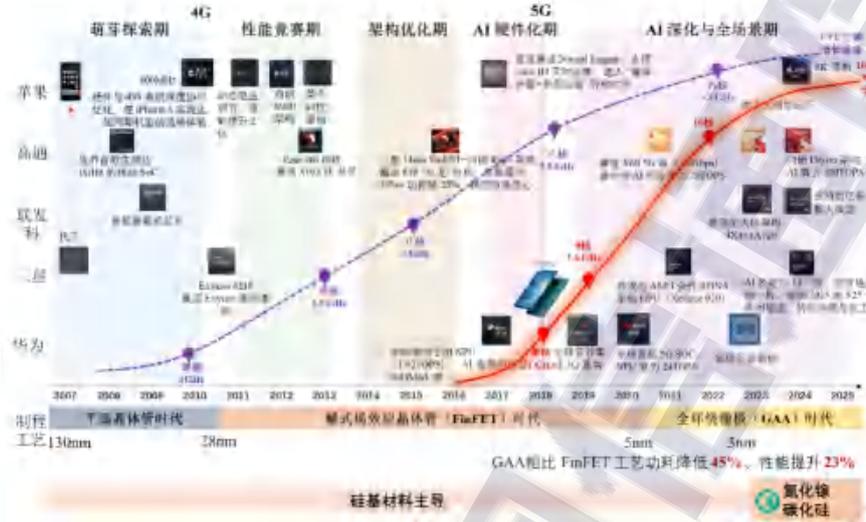
## （一）芯片架构持续演进，强化端侧 AI 核心算力

### 1. 手机处理器架构持续优化，功能集成全场景化

2025 年，手机处理器围绕制程工艺、微架构优化、异构算力协同、存储带宽优化与能效管理等持续演进，为全场景 AI 体验提供底层技术支撑。制程工艺已迈入 3nm-4nm 时代，台积电 N2 工艺的量产带来同等性能下 10—30% 的功耗优化，为高负载 AI 任务提供能效基础。处理器性能演进全面转向架构驱动，各厂商通过不同技术路径持续提升算力：苹果采用自研 Firestorm 等核心优化微架构；高通推进 Oryon 架构迭代；联发科采用 ARM 最新 C1-Ultra 大核；三星优化其定制核心，均实现 5% 以上的主频提升与 30% 以上的缓存扩容。

在功能集成方面，“CPU+GPU+NPU”的异构计算架构已成为行业标配。CPU 计算效率提升显著，通过大容量缓存与统一内存架构优化，显著降低数据搬运开销；GPU 引入硬件级光线追踪技术大幅提升阴影、反射与全局光照的真实感；NPU 已成为端侧 AI 推理的核心引擎，峰值算力突破 100TOPS，部分产品已支持 BitNet 1.58bit 超低比特推理与存算一体架构，使 4K 文生图等复杂 AI 任务在终端实时运行成为可能（图 6）。存储性能同步跃升，LPDDR5X 内存带宽突破 8533 Mbps，UFS 4.1 存储随机读取性能达 500K IOPS，配合统一内存架构有效缓解“存储墙”瓶颈。能效管理采用多级动态电压频率调节、场景感知功耗调度与 AI 驱动的预测模型，将整机功耗控制在可持续运行阈值内。处理器通过软硬协同调度机制，实现 NPU、

GPU、DSP 与 CPU 之间的动态负载分配与任务卸载，成为支撑多模态感知、实时语义理解与情境智能决策的关键基础设施（表 6）。



来源：公开数据整理

图 6 手机处理器技术发展技术图谱

表 6 主流手机芯片厂商最新旗舰芯片关键 IP 配置汇总表

厂商	芯片	CPU	GPU	NPU	其他主要 IP 核
苹果	A19 Pro	自研 6 核： 2×4.26GHz+4×2.60GHz	Apple10 架构 5/6 核	16 核神经引擎 +神经加速单元	ISP、N1 无线芯片、安全模块
高通	骁龙 8 Elite Gen 5	Oryon CPU： 2×4.6GHz+6×3.62GHz	新一代 Adreno	Hexagon NPU (性能提升 37%)	Spectra AI ISP、X85 5G 调制解调器、Wi-Fi 7
联发科	天玑 9500	1×C1-Ultra 4.21GHz+3×C1-Premium 3.5GHz+4×C1-Pro 2.7GHz	Mali-G1 Ultra (12 核)	NPU 990 (算力提升 110%)	Imagiq 1190ISP、UFS 4.1
三星 (Samsung)	Exynos 2500	1×X925 3.2GHz+ 2×A725 2.74GHz+5×A725 2.36GHz+2×A520 1.8GHz	Xclipse 950 (RDNA3)	24K MAC NPU(2-GNPU+2-SNPU) (算力提升 39%)	5G NR、GNSS、多频调制解调器、ISP

来源：公开信息整理

芯片架构的技术跃迁，是 AI 终端实现“四新”特征的算力基石。在认知协同层面，NPU 与 CPU、GPU 的协同调度，使多模态感知和实时语义理解成为可能；在场景预见层面，低功耗 AI 算力支持终端持续感知与预测，实现主动服务；在意图驱动层面，混合精度计算架构支撑端侧大模型的高效运行，为意图理解与任务拆解提供基础；在服务共生层面，能效平衡的异构计算为跨设备任务协同提供支撑，使终端群能够智能分配计算负载。芯片技术从参数竞争到体验驱动的转变，为“四新”特征的全面实现提供了坚实的物理基础。

## 2.PC 处理器算力与能效突破，增强 AI 应用支撑能力

当前随着大模型向端侧发展，AI PC 进入爆发阶段，AI 功能从基础辅助转向 AI 生产力工具演进，落地覆盖办公、创作、行业等场景应用。在 AI 技术驱动下，PC 处理器由 CPU 主导、GPU 协同向集成 NPU 发展，从单一任务向算力协同、功能互补的一体化架构演进，同时算力和工艺制程等持续突破，支持更强的端侧 AI 能力，驱动 PC 能力跃升，为 AI PC 场景应用夯实技术底座（表 7）。

表 7 2024—2025 年主流厂商处理器芯片型号及核心技术参数

发布年份	品牌	型号	制程	AI 算力	核数/线程
2025 年	高通	骁龙 X2 Elite 系列	第三代 N3P 3nm	NPU 算力：最高 80TOPS	12/16 核
2024 年	高通	骁龙 X Elite 系列	台积电 4nm	NPU 算力：持续 45 TOPS	8 核 Oryon CPU

发布年份	品牌	型号	制程	AI 算力	核数/线程
2025 年	英特尔	酷睿 Ultra 200 H 系列	台积电 N3B 3nm	综合 AI 算力达 99 TOPS	8 个英特尔 Xe 核
2024 年	英特尔	酷睿 Ultra 200V 系列 (Lunar Lake)	台积电 N3B 3nm	NPU 算力: 48 TOPS, 综合 AI 算力达 120TOPS	8 核 (4P 核+4E 核)
2024 年	英特尔	酷睿 Ultra 200S 系列 (Arrow Lake - S)	台积电 N3B 3nm	NPU 算力: 配备第三代 NPU	最大 24 核 (8P 核 +16E 核) /24 线程
2025 年	AMD	Ryzen Threadripper 9000 系列	台积电 N4P 4nm	具备 AI 处理能力 (具体算力未公开)	核心数线程数大幅提升
2024 年	AMD	锐龙 AI Max +395	台积电 4nm FinFET 制程工艺	NPU 算力: 50 TOPS	16 核 (全性能核, Zen5 架构)/32 线程
2025 年	华为	麒麟 X90	未明确公布	AI 算力未明确公布	核心数线程数未明确公布

来源：公开信息整理

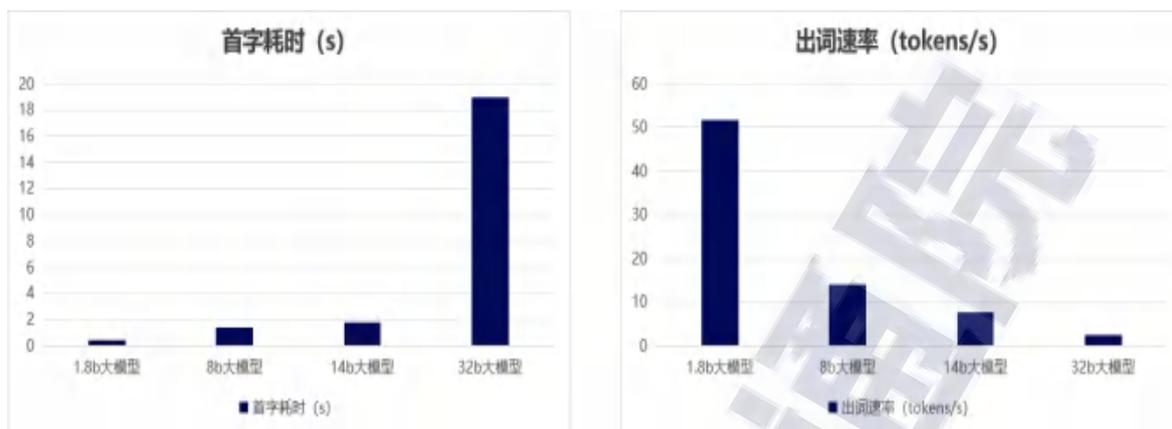
在该阶段，一是“CPU+GPU+NPU”异构协同成为 AI PC 处理器主流架构，AI 算力显著提升。这一阶段标志性事件是，2023 年英特尔第 14 代酷睿（Meteor Lake 架构）首度将 NPU 集成入 PC 处理器，AI 任务能效比为上代产品的 2.5 倍。此外，高通、英伟达逐步推出 ARM 架构 PC 处理器，渗透 AI PC 芯片市场。例如，高通发布骁龙 X Elite 芯片，可在 ARM PC 上流畅运行 130 亿参数大模型，实现百毫秒级实时翻译，功耗仅 1.5W（NPU 单独运行的极限值）。

二是先进工艺制程进一步推动算力与能效双重突破，主流制程进入 3nm-4nm 时代。例如，台积电 3nm 工艺（N3B/N3E）应用于英

特尔酷睿 Ultra 200 系列、苹果 M4 芯片，4nm 工艺用于高通骁龙 X Elite、AMD 锐龙 AI Max+395。NPU 在算力提升的同时进一步降低功耗，如苹果 M4 的 NPU 在 38TOPS 算力下，功耗仅 4W，较 M1 的神经网络引擎（16TOPS/3W）实现“算力翻倍+功耗微增”。

三是 PC 芯片 AI 软件栈提升主流 AI 模型和框架等兼容适配能力，以充分发挥端侧 AI 应用支撑能力。PC 芯片 AI 软件栈作为连接硬件算力与 AI 应用的桥梁，其兼容适配能力直接决定了端侧 AI 应用的性能表现和用户体验。如英特尔推出 OpenVINO™ 工具套件，通过统一的硬件抽象层和 API 接口，支持 PyTorch、TensorFlow 等模型转换、硬件算力灵活调度等，缩短端侧模型开发周期。

AI PC 大模型支持能力与端侧 AI PC 芯片强相关，当前商用 AI PC 主流芯片以英特尔、高通、AMD 为主，受限于 AI PC 设备计算和存储能力，在端侧能够支持 2B\3B\7B\13B 等大小模型（13B 未消费普及）。中国信息通信研究院的检测数据表明，在性能中等的 AI PC 芯片上模型参数由 2B 增大至 32B，模型性能和体验明显下降；当前主流商用 AI PC 端侧模型主要为 1.5-7B（图 7）。



来源：中国信息通信研究院

图 7 商用 AI PC 端侧芯片支持大模型能力

未来 AI PC 芯片将在架构、算力、制程、软件栈等方面持续协同演进，以支持更强的模型推理性能。随着 AI PC 芯片制程工艺向 18A、14A 演进，突破 2nm 以下工艺水平，NPU 算力有望突破 100TOPS，进一步推动 AI PC 性能提升和场景赋能。

## （二）存储带宽与容量持续升级，破解“存储墙”瓶颈

内存与存储两大技术体系协同演进，为突破制约终端 AI 性能的“存储墙”提供技术支撑，直接赋能“四新”特征的落地。在认知协同层面，高带宽、大容量内存是实现低延迟、高并发 AI 推理的先决条件。内存是直接与 CPU、GPU 和 NPU 交互的存储单元，其带宽决定了大模型权重加载与数据传输的效率。LPDDR5X 速率突破 8533Mbps，单通道容量从 16GB 向 32GB 迈进，使端侧设备能够常驻百亿参数级大模型，确保多模态交互的流畅体验。在场景预见层面，存储技术从简单的容量扩展升级为智能数据调度。UFS4.0 接口

带宽达 4.2GB/s，3D NAND 堆叠层数突破 300 层，推动手机存储迈入 1TB 时代。新一代存储通过智能分区与预加载机制，基于用户行为预测模型加载优先级，实现个人知识图谱常驻与意图预测，支撑主动服务体验。在意图驱动与服务共生层面，统一内存架构实现 CPU、GPU、NPU 间的数据共享，为跨应用任务编排提供低延迟数据通道。

存内计算（CIM）技术通过架构创新，将部分计算单元集成到存储单元中，减少数据搬运开销，在特定 AI 负载场景下大幅提升能效比。当前，CIM 技术已在部分高端终端芯片中实现应用，为突破传统计算架构瓶颈提供了新思路。随着技术成熟，CIM 有望在能效敏感的边缘 AI 场景发挥更重要作用，为 AI 终端的持续进化开辟新路径。

### （三）多模态感知交互体系升级，迈向感知与表达一体化

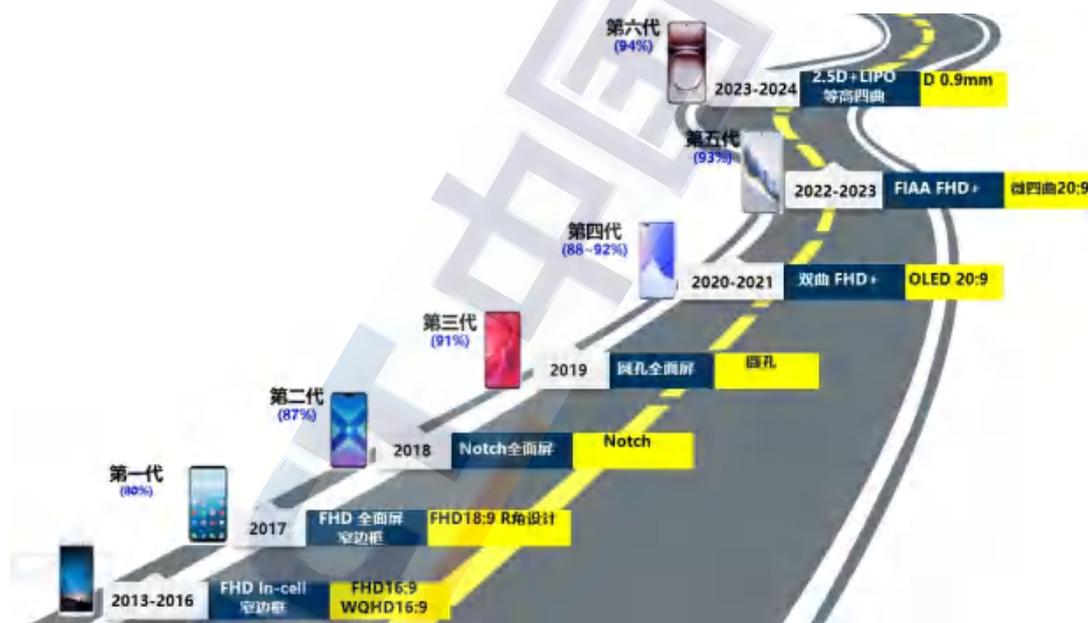
#### 1. 手机显示体验从参数竞争到健康护眼与智能交互的融合演进

手机显示屏历经四代演进，已从单一显示工具，演进为融合视觉体验、健康护眼与智能交互的核心人机界面（图 8）。它正朝着能主动感知、理解与决策的“认知界面”发展，旨在成为全场景智能的支柱，这标志着显示技术正式步入认知协同体验驱动的新纪元。折叠屏、屏下摄像等形态创新重塑交互边界。其中，全面屏通过屏内挖孔，窄边框，等高四曲等技术发展，如天马 OLED 屏通过下边框弯折技术，将屏占比推升至 94%，接近全屏显示的理想形态（图 9）。

	1990s	2000s初期	2007-2016	2019至今
功耗	功耗极低	功耗低	功耗中等	平衡功耗体验与功耗
分辨率	—	QVGA(240x320) HVGA(480x320)	QVGA(320x240) VGA(480x320) 768p(1280x800) 744p(1242x1440)	2160p(3840x2160) 从4:3到2K/1K/10
色彩	单色	实现彩色，色彩项目响应器 一色彩饱和度和对比度 和响应速度提升	广视角和色彩准确度高 液晶成为行业基础	广色域DCI-P3为旗舰机标配 BT 2023未来应用预期
屏幕刷新率	—	60Hz	刷新率90Hz / 2017年前 峰值120Hz巨幅刷新率主导 (2017年) IPS-LCD优化液晶分子排列	高普及 (2020至2021) / 90/120Hz成为旗舰机标配/部分机型 144Hz 在优化(2022年至今) LTPO 1-120Hz 自适应/165Hz 超高频
材质等特性	液晶/电阻 触控式背光或无背光	背光技术 TFT-LCD技术通过为每个像素 提供单个液晶像素	LCD追求极致薄/ OLED自发光 OLED追求对比度与寿命创新 窄边框、刘海屏等设计	OLED技术成为主流 折叠、卷曲等柔性形态及屏下摄像头技术等
TFT基板	a-Si 非晶硅	a-Si 非晶硅	LTPS 低温多晶硅	LTPO 低温多晶硅/氧化物
成果	从通话工具向 可观察数据跨越	催生移动互联网应用 奠定早期智能手机基础	显示性能快速提升 推动视觉体验革新	形态自由、交互丰富和显示完整 高度沉浸与体验驱动
代表机型	摩托罗拉200	诺基亚7250	iPhone 4 / IPS-LCD里程碑 / 三星Galaxy S系列 / OLED引领者	三星 S25 Ultra(高分辨率 3120 x 1440 (Quad HD+)) 一加15(165Hz超高频先行者)
	液晶+TFT-LCD阶段	IPS与OLED技术争锋的时代	高刷新率刷新率显示为特征驱动的阶段	

来源：公开信息整理

图 8 手机显示屏四阶段演进技术



来源：天马解决方案

图 9 智能手机全面屏解决方案

当前，由用户长时间近距离使用手机引发的视疲劳、干眼及近视等视健康问题备受关注，用户体验已从对单一参数的性能追逐，演进为对综合显示品质与视觉健康保护能力协同优化的系统性要求。

主流协同优化中的**智能技术已形成以硬件为基础、软件为调节、用户引导为辅助、行业创新为动力的多层次智慧护眼路径。**

**硬件调光与感光：**采用无频闪设计（DC/高频 PWM 调光），低亮度下自适应启用类 DC 调光。结合环境光传感器，实现亮度与色温的自动调节。**软件算法与干预：**通过护眼模式集成“20-20-20”用眼提醒，夜间算法调节色温以缓解节律干扰。智能 HDR 技术通过场景识别与内容分析，优化亮度映射，保留明暗细节。**用户引导与习惯培养：**系统引导用户开启自动亮度与护眼模式，并建议适宜观看距离及环境光照度。针对未成年用户，强化时长限制与坐姿距离提醒，推动主动行为改变。**行业协作与标准建设：**通过产学研合作，推动智能显示设备视觉健康评价体系与技术验证，联合眼科机构建立健康评估体系，开展医工结合与人因工程研究。如天马高端 OLED 屏（简称天马天工屏）通过硬件优化与智能算法协同，将多项智慧护眼技术集成构建智慧护眼体系，为移动终端视觉健康提供可行的技术解决方案。

显示屏对视觉健康的影响是学术界与产业界持续关注和研究关注的焦点，其各项工程参数与眼部健康之间的长期效应，仍处于深度探索阶段。未来 AI 赋能下的显示认知协同新纪元，将向**智能感知交互、柔性形态创新与个性化视觉健康**三大方向纵深发展。

**首先，屏幕将发展为更智能的多维感知与交互入口。**通过集成更多微型传感器，屏幕本身成为高敏感知层。屏下摄像头、屏下 3D

结构光等技术的成熟，将实现视觉无干扰的全面屏。AI 视觉算法进一步使隔空手势、注视感知等无接触交互成为可能，实现“注视唤醒、移开熄屏”等直观操作，提升交互自然度与系统能效。

其次，屏幕形态将持续突破物理边界，向柔性化、无缝化认知协同演进，其发展深度依赖 AI 技术。具体表现为：第一，AI 通过动态交互、画质优化与像素补偿等技术，解决折叠、卷曲等形态交互与显示难题；第二，屏幕载体将从刚性玻璃向柔性薄膜与微投影融合演进，实现无处不在的显示；第三，AI 将作为顶层中枢，协同管理多设备的显示任务，实现理解环境、用户身份和任务的情境感知与无缝显示。

最后，智慧护眼将进入 AI 驱动的个性化定制阶段，实现“千人千屏”。显示品质与视觉健康保护能力协同优化的核心在于 AI 对显示内容、环境光线与用户状态三大维度的融合感知与动态调节。即 AI 通过融合感知内容、环境与用户状态，动态调节屏幕参数等实现从被动优化到主动响应的个性化护眼，构建从实时干预到长期视觉健康管理的闭环，代表了“主动感知-智能响应”的新一代健康显示方向。

## 2. 光学摄像头实现从基础成像到全场景智能体验的跃迁

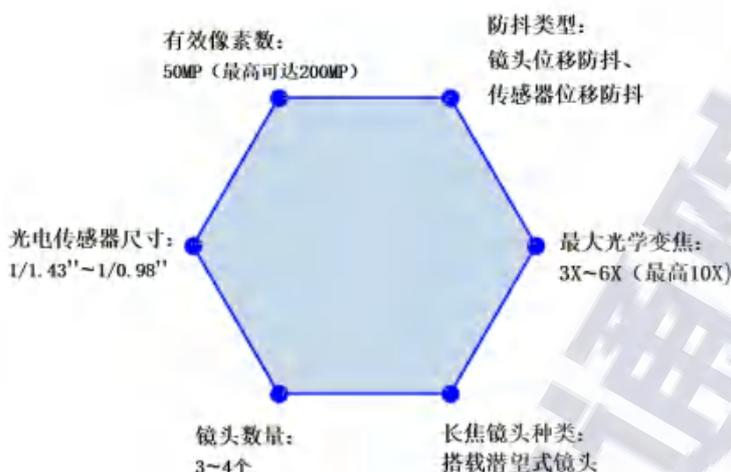
经过近 20 年的发展，当前智能终端光学摄像头技术已进入硬件技术高度成熟的阶段（图 10）。以智能手机为例，主流旗舰机型普遍配备 3—4 颗摄像头，为实现多场景、多角度的拍摄提供了硬件基

础。光电传感器尺寸取得重大突破，一部分机型的传感器尺寸突破至 1/0.98 英寸，使得手机在光线捕捉能力上大幅提升。而像素规格也稳定在 50MP 左右的高水平范围，而三星在 Galaxy S24 Ultra、Galaxy S25 Ultra 上应用了高达 200MP 的图像传感器，通过像素合并技术进一步增强感光能力，从而提升低光下拍摄效果。在变焦能力上，潜望式长焦镜头得到广泛应用，普遍能够实现 3-6 倍的光学变焦，如 2024 年的 OPPO Find X7 Ultra 能够实现 6 倍光学变焦，使远距离捕捉的影像更加清晰。在防抖技术方面，几乎所有旗舰机都采用了镜头位移防抖方案（图 11）。



来源：公开信息整理

图 10 手机光学摄像头技术的发展



来源：公开信息整理

图 11 2023—2025 年主流旗舰手机技术参数水平

在硬件逐渐趋同的背景下，计算摄影与多摄像头系统深度融合，AI 算法加持实现专业级影像效果，标志着智能终端摄像头正式进入全场景的智能影像时代，凸显 AI 终端场景预见与服务共生的核心特征。在场景识别方面，手机的 AI 芯片搭载场景识别模型，通过主动感知理解能力精准判断夜景、逆光、舞台、烟火等各类拍摄场景，自动切换拍摄参数以适配不同环境实现了从被动响应到场景预见的人机交互范式升级。在画质优化方面，依托超广、主摄、长焦、潜望式长焦等多摄系统，实现全焦段覆盖以满足不同距离拍摄需求，是硬件架构与 AI 能力深度协同的架构重塑体现；在极低照度环境下，通过 AI 驱动的多帧原始数据降噪与高动态范围融合算法，来提升画面亮度，把控色彩表现，提升暗光下的成像质量；通过时空域 AI 降噪降低图像中的噪点，使视频画质更加纯净；利用高帧速率捕获与短曝光相结合，配合 AI 运动矢量预测，在拍摄目标快速运动的场

景下更清晰地捕捉动态影像，提升运动抓拍的成功率；EIS 与 OIS 的协同防抖，在各种动态拍摄场景中，用以保持画面的平稳流畅、降低抖动带来的画质损失。智能终端光学摄像头技术通过算法与硬件的认知协同，正持续优化全场景成像体验，以真正实现从功能叠加到范式重构的质变。

未来智能终端摄像头将向高适应性传感器技术、更高集成度镜头组和更强计算摄影能力的方向发展。首先，为满足用户对更高画质和优质低光表现的需求，传感器将向更大尺寸演进，同时较小的像素尺寸结合像素合成技术进一步提升高分辨率与高感光度的平衡；其次，潜望式长焦技术将突破现有光学变焦限制，有望实现更高的光学变焦倍数，并借助 ALoP 等新型折叠光路技术进一步压缩模组，顺应终端设备轻薄化设计趋势；最后，计算摄影将与 AI 深度协同，通过多帧合成、实时场景感知等算法提升画质，同时 AR 功能将逐步成为高配主流，推动摄像头从单纯的影像采集工具向环境感知与交互核心演进，最终实现真正的全场景智能视觉体验。

### 3. 麦克风从通话拾音工具演进为智能终端实现声学感知与交互的核心器件

功能机时代以麦克风为拾音工具仅能满足窄带通话需求。智能机的兴起，使 MEMS 麦克风逐渐成为主流，推动了高清语音通话的普及。麦克风阵列和 AI 算法结合，支撑起超高清语音、语音交互等多种应用场景，成为 AI 感知交互核心器件之一（表 8）。

表 8 电声技术发展

阶段	时间	核心驱动力	听筒/扬声器	麦克风特征	代表机型
第一阶段 (功能机时代)	-2007	窄带语音通信	-定制化设计,满足基本通话需求 -单听筒、单扬声器配置 -窄带语音 (300Hz-3.4kHz),音质粗糙	-以驻极体麦克风 (ECM) 为主 -单麦克风设计 -抗干扰能力弱,易受噪声影响	诺基亚 1100、小灵通 (PHS)、摩托罗拉 V3
第二阶段 (智能手机萌芽)	2007-2010	智能手机革命+3G 普及+多媒体初现	-向微型化、低功耗转型 -支持更宽频响 (100Hz-8kHz) -初步支持音乐播放	-MEMS 麦克风逐渐成为主流 -单麦克风为主,集成度更高 -出现基础数字降噪算法 (如 DSP 处理)	iPhone 4、Moto XT800、HTC Hero
第三阶段 (移动互联网爆发)	2011-2016	4G 普及+多媒体消费+语音助手兴起	-支持全频带音频 (20Hz-20kHz) -采用复合振膜、大腔体设计提升音质 -双扬声器立体声初现	-多麦克风阵列成为主流 (双麦、三麦、四麦) -波束成形 (Beamforming) 技术应用 -智能算法应用成熟 (如回声消除、噪声抑制)	iPhone 4s/5/6、三星 Galaxy S3/S6、Moto XT910
第四阶段 (AIoT 与沉浸体验)	2017-至今	5G+AI 深度融合 EVS 超高清语音编码 TWS/可穿戴爆发	-MEMS 扬声器开始商用 (超小体积、高精度) -超线性扬声器技术普及 (大振幅、低失真) -多声道立体声 (双扬、四扬) 成旗舰标配 -频响扩展,支持超高频 (50Hz-14kHz) 语音通话	-AI 深度赋能: 端侧 AI 模型实现实时语音增强 -智能声学模组 (集成 AI 降噪、唤醒、定位) -多模态融合: 骨传导、超声波辅助拾音 -高信噪比 (>70dB)、一致性好	各品牌旗舰机标配三麦双扬, TWS 耳机, 折叠屏手机
未来展望 (智能感知时代)	2026-2030	多模态交互无损音频流媒体隐私与可持续发展	-超宽频响应 (20kHz 以上) -新型低功耗材料 (如石墨烯振膜、压电 MEMS) -环保制造工艺与可回收设计	-智能感知麦克风: 支持情感计算、健康监测 (呼吸、咳嗽分析) -无感交互: 骨传导+超声波+环境声感知融合 -内置隐私保护机制 (本地处理、语音加密) -自适应环境建模与主动降噪	---

来源: 公开信息整理

麦克风作为智能终端感知世界的“耳朵”，其性能直接影响终端的智能化水平与用户体验。智能降噪与语音增强 AI 算法，突破传统降噪技术限制，能精准区分人声与交通、风声等背景噪声，有效提取清晰人声，保证嘈杂环境下终端的录音和通话质量。AI 算法结合麦克风阵列，实现对说话者位置的精准定位，并利用波束成形技术聚焦目标声源，抑制干扰，使智能音箱、会议终端等设备实现声源定位与远场拾音。AI 算法赋能麦克风不仅能“听”，还能“懂”，可识别婴儿啼哭、报警声等特定声音事件，并触发智能响应，增强设备环境音识别与场景理解能力。AI 与麦克风技术的深度融合，持续拓展智能终端应用边界，推动人机交互向更自然、更智能的方向演进（表 9）。

表 9 麦克风在智能终端人机交互的作用

作用	应用场景
语音输入与交互	语音助手、微信语音
通话与通信	电话通话、视频会议或语音聊天
录音功能	录音机、会议记录、课堂笔记等
环境声音感知	智能监控设备、主动降噪
语音识别与处理	采集原始音频信号，后续语音识别（ASR）、声纹识别、降噪、回声消除等算法处理的基础
辅助功能	为残障人士提供语音控制、语音输入、实时语音转写等服务

来源：公开信息整理

麦克风技术正向智能化、高信噪比、强抗干扰方向演进。首先从单麦克风向阵列化应用发展，融合降噪算法精准适配远场、高噪

等复杂场景；其次 AI 算法深度赋能，显著提升多语种和方言的识别精度与交互体验；最后从独立功能到生态入口，通过视觉、触觉等传感器融合实现多模态交互，成为智能设备的交互中枢，支撑情感计算、健康监测等高级服务。

#### 4. 扬声器成为智能终端个性化输出与智能化交互的接口

扬声器作为与麦克风功能相反的声输出器件，在智能终端中的作用已从单一声音播放，演变为连接用户、设备与环境的桥梁。空间音频、N'BASS®虚拟后腔等技术重塑智能终端音频体验，标志着扬声器从听清迈向沉浸式智能交互新时代。

当前 AI 已成为扬声器产业演进的核心驱动力。双扬声器与 AI 算法的深度融合，显著提升了通话音质和交互体验。MEMS 技术应用于扬声器（表 10），凭借高保真、微型化等优势，逐渐用于 TWS 耳机等可穿戴场景；扬声器阵列广泛应用于车载、XR 等场景，通过 AI 感知技术实时追踪用户头部位置，动态调整声场，实现 3D 音效和空间音频，给用户带来沉浸式体验。AI 通过感知用户听力特征、播放内容类型及所处环境，动态调整扬声器的输出，定制个性化音频体验。以 MPEG-H 音频技术为例，AI 不仅支持用户自定义声场细节，还能实时分析播放内容与环境噪声，自适应优化音量和音效，确保任何场景下的最佳听感。

表 10 MEMS 扬声器 vs 传统扬声器

扬声器类型	压电 MEMS 扬声器	传统扬声器
工作原理	逆压电效应，驱动压电结构形变，带动硅振膜振动发声。	依靠线圈、磁铁驱动振膜振动发声。
核心组件	压电 MEMS 芯片、硅振膜、控制电路等	磁铁、音圈、振膜、悬边、导磁柱等
频率响应	可达 40kHz 以上，支持超高清音频和无损播放	一般为 20Hz-20kHz(人耳范围)
响应速度	极快，比传统技术快达 150 倍，瞬态响应优异，声音更精准	相对较慢，受限于音圈和振膜的机械惯性
相位一致性	高度一致，多单元协同工作时声场精准，适合阵列设计	单元间存在微小差异，大规模阵列需额外校准
音质特点	高清晰度、高保真、低失真、高分辨率，细节还原能力强	音色温暖自然，低频表现较好
尺寸与重量	极小、超薄可做到毫米级，体积缩小约 40%，重量约 20%	体积较大，尤其是大尺寸全频单元
功耗	低功耗，有助于延长续航	功耗较高，尤其在大音量输出时显著
制造工艺	半导体工艺	传统机械组装工艺
典型应用场景	TWS 耳机、AR/VR 设备、助听器等小型可穿戴设备	智能手机、蓝牙音箱、电视音响、车载音响、家用音响系统
主要缺点	当前大功率输出能力有限，低频较差；初期成本高	体积大、一致性差、难以微型化

来源：公开信息整理

未来，扬声器技术演进围绕小型化、高性能、低功耗三大核心需求展开，如支持 20Hz-40kHz 超宽频，满足无损音频传输需求；通过材料创新（如压电陶瓷）降低功耗。AI、物联网和材料科学的进步，推动扬声器成为全场景智慧生活的关键组件。

## （四）通信与 AI 双向赋能，共筑 AI 终端“四新”特质基础

随着生成式人工智能与大语言模型技术的快速发展，通信与 AI 正从独立技术演进为深度融合、协同优化并相互增强的系统性支撑架构。一方面，先进通信技术为人工智能突破终端算力瓶颈、实现端边云协同提供了基础支撑；另一方面，AI 技术正全面渗透至通信技术的各个层面，优化用户连接体验。二者的协同进化，共同为 AI 终端“四新”特质筑基赋能。

新一代通信技术为 AI 终端“四新”特质提供了关键基础支撑。5G 和 Wi-Fi7 等技术带来的吉比特每秒级峰值速率和毫秒级超低时延，突破了传统网络瓶颈对 AI 应用实现的制约，是保障“四新”特质得以流畅体验的坚实前提。AI 引擎可预判任务需求，将时延敏感型任务留存终端，将中等任务调度至边缘计算节点，或将复杂推理任务通过高速网络上传至云端。同时，它还为联邦学习等分布式 AI 技术提供了基础，支持在保护隐私的前提下完成模型协同训练。

AI 赋能的智能连接是“四新”特质得以落地的关键体验保障。AI 终端对连接的稳定性、智能性和场景感知力提出了极高要求。AI 的核心赋能作用在于提升终端的场景感知精度与网络优化能力，推动连接体验从被动响应向主动预判的范式演进。针对电梯、地铁、高铁等复杂场景，AI 终端能够智能化预判并优化连接。AI 算法还可实时监测蜂窝与 Wi-Fi 网络状态，并基于当前应用类型对网络质量

的需求进行链路选择与聚合，保障用户在网络切换过程中的无缝连接与流畅体验。除宏观调度外，AI 算法还深入信号处理的微观层面。通过智能天线管理，AI 可依据用户使用状态与环境干扰实时优化多天线工作模式，动态选择信号收发性能最优天线组合。同时，AI 可智能识别环境中存在的同频段干扰源，优化算法调整信道与发射功率有效规避干扰，提升室内信号覆盖与穿透性能。此外，AI 模型基于用户行为制定个性化网络优化策略，提升网络搜索效率并增强连接稳定性。

面向 6G 时代，通信与 AI 的融合将迈向更深层次的“内生智能”阶段，为“四新”特质的持续演进提供动力。首先，通感一体化将成为标准功能。通过高精度定位、手势识别、呼吸心跳监测与环境建模等技术，为“四新”特质提供全新物理感知维度，使终端成为连接物理世界与数字世界的关键接口。其次，智能重塑环境将成为可能。以 AI 驱动的波束成形和可重构智能表面等新技术，将为未来“四新”特质的演进提供按需定制的转变连接能力，有望在智慧城市等场景中实现全覆盖的稳定连接与超高精度定位。

### （五）能效管理多路径协同，平衡性能与续航矛盾

能效是制约大模型研发与端侧智能普及的重要因素。骁龙 8 Gen 3 NPU 的瞬时功耗可达 5W 左右。中国信息通信研究院的检测数据表明，在没有经过适配优化的情况下，不同终端在端侧运行大模型的功耗占比差异较大（图 12）。



来源：中国信息通信研究院

图 12 端侧本地模型耗电占比<sup>2</sup>

AI 终端的能耗矛盾，本质是算力需求指数增长与能效提升线性演进之间的结构性失衡。电池扩容是解决这个问题的有效手段之一。以手机为例，2023 年起随着硅碳负极、半固态电池等技术逐步落地，2025 年主流品牌热门机型电池容量已经普遍超过了 4800mAh，个别产品甚至达到了 8300mAh（表 11）。但受终端体积所限，电池容量无法持续扩充。举例而言，使用硅碳负极技术的电池能量密度不低于 600wh/L，一个 161.3mm（长）× 75.3mm（宽）× 9.35mm（厚）的终端，估算理论电池最大容量也不高于 17900mAh，同时厂商还要综合考虑轻薄手感、电池寿命、散热表现和安全性等多重因素。

表 11 2025 年各主流品牌热门机型的电池典型容量

品牌名称	电池名称	产品名称	电池容量
荣耀	青海湖电池	X70	8300mAh
华为	硅碳负极电池	Pura 80 Ultra	5700mAh
小米	金沙江电池	17 Pro Max	7500mAh

<sup>2</sup> 使用不同品牌 6 台手机与 6 台平板电脑作为测试载体，部署开源本地模型“口袋 AI”(1.5B)，通过网络调研梳理 200 个高热度 AI 咨询问题构建标准化测试题库，依托专业 AI 能耗测试工具发起自动化问询，使模型持续运行 6 小时。全程监测设备能耗参数、表面发热特性等关键指标，分析该过程口袋 AI 平均每小时耗电占终端总电量的耗电比。

品牌名称	电池名称	产品名称	电池容量
欧珀	冰川电池	Find X8 Ultra	6100mAh
维沃	蓝海电池	Y500	8200mAh
苹果	--	Iphone17 pro Max	4823mAh
三星	--	S25 Ultra	5000mAh

来源：公开信息整理

因此快充技术成为弥补续航短板、缓解用户电量焦虑的核心手段。过去十年，我国快充技术实现了跨越式发展，充电功率从最初普遍的 5W 升级至如今 240W 的高功率快充阶段（图 13）。为进一步解决充电技术互通功率低、协议互不兼容和资源浪费等问题，2021 年中国信息通信研究院牵头组建的 UFCS 联盟推动我国快充技术逐步走向统一，2025 年不同品牌产品互通功率达到 40W，进一步提升了用户使用便捷性。



来源：公开信息整理

图 13 终端快充技术功率进阶历程

未来 AI 不仅是能耗的消耗者，更是能效的优化者。通过任务调度、模型压缩与场景感知，实现“该省则省、该强则强”的动态能效平衡，从根本上提升单位能量的智能产出比，实现能用平衡。

#### 四、软件系统智能重构，定义可信服务新范式

2025 年，终端操作系统正从终端软硬件资源的管理员进化为管理用户意图的大管家。通过与 AI 的深度融合，操作系统不再被动等待指令，而是具备了主动思考与执行的能力。本章将探讨操作系统、智能体与安全体系如何重构软件范式，重新定义人与服务的连接方式（表 12）。

表 12 软件系统重构及其对“四新”能力的支撑

软件系统	核心变化	带来的直观价值	支撑的“四新”特征
操作系统	从“资源调度”到“智能中枢” (管理 APP → 端云协同+意图理解)	变身“超级管家”： 系统能听懂你的潜台词，协调大小模型（端云协同），既保护隐私又聪明能干。	认知协同 (理解复杂意图)
智能体	从“功能堆砌”到“自主执行” (用户点击 APP → AI 调服务)	服务主动找人： 不用在几十个 APP 里来回跳转，说句话（如“定出差行程”），智能体自动帮你搞定一切。	意图驱动 (一语即达) 场景预见 (主动服务)
安全体系	从“被动防御”到“可信基石” (查杀病毒 → 隐私计算+端侧闭环)	守住“隐私红线”： 确保 AI 越聪明越安全，敏感数据不出设备（端侧处理），让用户敢用、放心用。	全场景支撑 (构建信任基础)

来源：中国信息通信研究院

##### （一）操作系统与 AI 深度融合，构建端云协同智能中枢

操作系统与 AI 深度融合的目标，是构建一个能主动感知、理解

用户意图，自主编排资源完成复杂任务的“智能中枢”。华为、小米（Xiaomi）、高通等企业的分析报告明确指出，仅靠云侧，无法满足用户对智能终端的即时响应、离线可用以及个人隐私安全的需求，仅靠端侧算力，无法支撑高频、复杂的 AI 推理与规划任务，“端云协同”是构建强大端侧推理服务的必然选择和基础架构。

### 1.端云协同架构变革操作系统软件栈，实现智能中枢范式重塑

为克服早期云端 AI 带来的高延迟、隐私风险和弱网失效等瓶颈，全球主流操作系统厂商正一致转向端云协同的软件技术架构。端云协同架构混合了多种灵活组合，能根据场景动态分配 AI 计算工作负载，以优化成本、能耗、性能、隐私和个性化体验，将强大的 AI 能力植入操作系统底层。在新一代 NPU 芯片、轻量化模型算法（如 MoE）和端云协同的共同加持下，AI 原生操作系统（AIOS）成为可能。未来，OS 将可能进化为智能的 AI 任务调度器，通过对 CPU、GPU、NPU 等异构资源进行精准、动态分配，最终实现从芯片到用户体验的完整闭环，将传统被动的应用启动器重塑为主动感知、预测并服务于用户意图的智能中枢。

### 2.新场景新服务驱动生态重构，人工智能终端操作系统产业竞争白热化

在终端操作系统与 AI 深度融合这一战略赛道上，市场呈现出多

极化、生态融合的激烈竞争格局。国外厂商方面，苹果以自研芯片+OS+硬件的垂直整合闭环，通过 Apple Intelligence 深度调用系统底层数据，巩固用户黏度；谷歌以 Gemini 大模型全面赋能 Android 生态，巩固其主导权；微软则以 Copilot 重塑 Windows 生态，定义 AI PC 新标准；国内厂商方面，华为鸿蒙、小米澎湃 OS、OPPO ColorOS、vivo BlueOS 等均强调跨生态互联的全场景协同。例如，华为鸿蒙系统与盘古、DeepSeek 大模型深度融合，配合 YOYO/小艺智能体实现高效端云协同与准确意图理解。拥有庞大用户基础的国内外平台型企业正加速将 AI 能力与操作系统深度捆绑，以构建新的“生态护城河”。

中国信息通信研究院正牵头制定操作系统智能化水平、端云协同能力等系列标准，并围绕计算效率、响应时间等核心指标提出性能测试方法，系统性地引导和推动终端侧智能计算能力的创新与应用发展。

## （二）终端智能体定义服务新范式，开启服务找人时代

终端智能体运行于用户终端设备，依托终端的环境感知、算力、存储与安全能力，借助操作系统作为能力编排与治理的枢纽，完成用户意图的理解、决策与自主执行。通过端—云协同的智能算法与模型，终端智能体实现从“人找服务”向“服务找人”的根本性转变，成为与操作系统深度融合的人机交互界面，有望重塑应用生态与商业模式。

## 1.技术路径演进：从接口调用到自主决策的跨越

全球范围内，智能体的技术实现路径呈现多元化发展态势，从依赖固定接口的工具增强型应用，逐步向具备自主认知能力的智能体应用演进，形成了多层次的技术体系。

**一是终端智能体入口多样化，以自然语言方式进行交互。**终端智能体的入口是用户和智能体应用交互的界面。智能助手是智能体应用常见的交互入口，用户可以在智能助手界面通过语音，文字等自然语言交互方式同智能体进行交互。终端厂商智能助手通常是系统级的入口，作为操作系统的一部分可以充分调用终端硬件能力，用户可以使用按键或助手应用图标进入智能助手，服务和结果则可以通过搜索栏、负一屏等系统不同界面触达给用户，如华为的小艺、荣耀的 YOYO，小米的小爱同学、欧珀的小布、维沃的 Jovi 等。对于应用厂商，智能助手通常在应用中实现，用户点击对应功能图标或在应用的文本框输入即可快速获取所需服务，如百度的小度、航旅纵横的小横等。

**二是以大语言模型为大脑，通过端云协同为用户提供高效服务。**终端智能体核心是使用强大的大语言模型作为“大脑”，通过学习调用外部工具来扩展自身能力。首先，大语言可以理解用户的自然语言输入，并根据用户问题进行推理，进行任务编排和规划，决定需要调用哪些外部工具或服务；然后执行行动，即调用该外部工具或服务；最后，观察调用返回的结果，并基于新信息进行下一轮推

理，直至问题解决。由于终端智能体工作场景复杂，使用单独一个大语言模型无法很好满足用户需求，通常采用端云协同方式，在端侧和云侧部署多个大语言模型实现协同服务。通常对于算力要求较低，且涉及用户隐私数据的任务通过端侧“小”模型进行处理；对于数据运算较高的推理任务则使用云侧“大”模型进行处理。如在 Android 系统中，涉及文本、图片处理等数据量较小的任务使用端侧部署的自研轻量级大模型（如魔法大模型、蓝心大模型等）处理；如果涉及音视频、垂域知识、推理等负责任务则使用云端部署的大规模大模型，此外还会与第三方合作，使用包括 DeepSeek 或通义千问等一系列模型进行处理。

三是接口调用和屏幕理解成为终端智能体操作外部工具的主要技术方案。终端智能体通过调用外部工具或应用实现服务闭环。目前主流的实现方案包括调用 API 接口方式和识别图形用户界面(GUI)的屏幕理解方式实现。对于调用 API 接口方式，由于外部各种应用接口多样，目前业界倾向于提供一套标准的智能体通信协议实现智能体同外部工具或应用的对接，新的协议不仅能适应智能体以大模型驱动并使用自然语言交互的特点，也能降低终端厂商和应用厂商的开发成本，避免产生“M×N”的重复适配。目前关注度极高的智能体相关协议和标准包括，由 Anthropic 公司推出的 MCP 协议适用于智能体对工具的调用，谷歌推出的 A2A 协议和 AP2 协议则适用于智能体与智能体之间的协作以及智能体交易；在移动操作系统领域，

由中国信息通信研究院牵头，联合国内终端厂商和应用厂商共同提出的智能终端意图框架标准<sup>3</sup>、IOS 系统的 App Intent 接口、Android 系统的 appfunctions API 等。

对于屏幕理解方式则是让智能体像人一样查看页面，通过点击、滑动、输入等方式和外部应用进行交互。屏幕理解方式对大模型的视觉理解能力、算力要求较高，实现难度较大。微软公司的 OmniParser 方案就是一款基于纯视觉的屏幕解析工具，其核心功能是通过屏幕截图识别用户界面中的可交互元素（如按钮、输入框、图标等），并生成结构化数据，进而驱动大语言模型像人类一样操控终端设备。

## 2.产业生态构建：从产品形态到商业模式的转型

智能体的兴起正在推动现有应用生态发生变革，一个以“用户意图”（Intent）为核心的新生态正在萌芽，未来应用形态、商业模式都将发生重大变化。

一是应用形态正在向交互拟人化、功能原子化、服务主动化发展，开发范式也随之改变。传统应用采用键盘或者触摸滑屏方式进行操作，而智能体应用则可以通过用户的一句话，大语言模型自动理解用户意图执行任务。这就使软件从传统的“规则+经验”，向“模型+数据”为主的新体系转变，开发活动将以模型的质量、性能和持续迭代为中心。

<sup>3</sup> [TTAF 282—2025] 智能终端意图框架总体技术要求; [TTAF 283—2025] 智能终端意图框架接口技术要求

另一方面在智能体应用中，各传统应用的服务可以被解构为一个原子化功能组件，智能体可以根据需要灵活组合。平台方将提供智能体运行框架，开发者的角色将从开发一个完整的 App，转向开发能被智能体高效调用的、功能单一的“微服务”或“技能”，并通过被调用的频率获得收益分成。

智能体应用还可以利用情景感知功能，根据用户使用习惯和偏好主动提供服务。如 Microsoft Outlook 在 AI 加持下已不是日历表，更像是一位工作秘书，能根据用户的日常行程自动建议最佳会议时间，甚至还能识别出哪些会议不必要参加。

二是从“应用孤岛”到“意图驱动”的转变显著。传统模式下，用户需要打开不同的 App 来完成不同任务，形成了数据和体验的“孤岛”。智能体则旨在打破这些孤岛，用户只需用自然语言表达一个最终意图（如“规划一次周末的家庭出游”），智能体便能自主地在后台完成调用航旅 App 比价、调用点评 App 订餐厅、调用地图 App 规划路线等一系列动作。在这种意图驱动的模式下，能够最精准理解并最高效执行用户意图的智能体平台将有望吸引更多流量。荣耀在 MagicOS 9.0 发布会上，通过 YOYO 演示“一句话下单 2000 杯咖啡”，展示了智能体对传统 App 交互模式的颠覆能力，这考验的是终端智能体基于用户设备数据，对用户意图的主动理解，再进行任务拆分、执行，完成对自有或者第三方应用的调度能力。

### 3. 标准规范探索：构建开放协同安全的智能体生态

智能体已成为推动新一轮科技革命与产业变革的关键力量，但是在快速发展的同时也带来了互联互通、基准评估、行业应用及安全治理等挑战，亟需构建科学、系统、前瞻的标准体系，推动构建开放协同安全的智能体生态。

**一是聚焦智能体交互协议标准，破除互联互通壁垒。**智能体在通信协议、接口规范、数据格式等方面缺乏统一标准，形成了严重的“生态孤岛”效应，难以实现高效协作与能力聚合，推动智能体交互标准建设是打通壁垒、构建统一生态前提。智能体通信接口标准化，制定统一的服务发现、消息传递、会话管理等协议，确保不同智能体之间能够进行对话。智能体语义理解标准化，建立面向任务和知识框架，使智能体能够准确理解指令、上下文及协作意图，实现语义层面的互操作。智能体行为描述标准化，规范能力描述、服务注册与发现机制，使其能自主识别并调用其他智能体的服务，形成跨平台、跨场景的群体智能网络，为构建大规模、复杂任务的自动化解决方案奠定基础。目前，中国信息通信研究院已经联合主要终端企业、应用企业共同制订智能终端意图框架相关标准，未来还将继续从国际标准到团体标准各个层面，研究制订终端智能体发布平台、通信、交互、安全、支付等相关标准，不仅能兼容国际先进协议还能自主演进，为我移动生态繁荣起到支撑作用。

**二是建立智能体基准评估标准，科学衡量能力水平。**智能体能力日趋复杂，其能力高低不能仅凭个别案例主观判断，必须建立客

观、公平、可量化的基准测试标准。构建多维度能力评估指标体系，全面涵盖其任务完成效率、复杂环境适应性、逻辑推理能力、多模态理解能力、长程记忆与上下文关联能力、人类意图对齐程度以及能耗效率等关键性能指标。开发标准化测试环境与数据集，包括模拟仿真平台、对抗测试场景和覆盖各行各业的挑战性任务集，确保评估的全面性与公正性。建立分级分类认证体系，根据评估结果对智能体进行能力分级，为市场采购、用户选型和行业准入提供清晰、可靠的依据，从而引导产业资源向高水平技术集中，避免低水平重复建设。

**三是推动构建智能体应用标准，加速行业深度融合。**智能体的主要应用在于赋能千行百业，但其与具体行业的深度融合面临场景差异大、流程复杂、合规要求高等挑战。通用技术标准无法直接满足垂直领域的特定需求，智能体行业应用标准是推动其规模化落地的关键环节。提炼智能体在部署集成、任务流程管理、人机协作界面等方面的共性要求。重点在智能制造、智慧医疗、智慧教育等关键领域，牵头制定与行业业务流程、数据规范、安全法规紧密衔接的应用标准。通过应用标准，降低企业集成智能体的技术门槛与成本，确保其应用过程符合行业监管要求，真正实现降本增效和价值创造。

### **（三）安全与隐私体系筑牢，构建可信 AI 终端基石**

AI 终端在“四新”演进过程中，在为用户带来更自然、精准、

主动的服务体验同时，也面临 AI 滥用、不安全模型使用及隐私泄露等问题。为应对这些挑战，我国发布国家标准 GB 45438-2025《网络安全技术 人工智能生成合成内容标识方法》和 GB/T 45654-2025《网络安全技术 生成式人工智能服务安全基本要求》，为 AI 内容标识与生成式 AI 服务安全提供合规指引。

终端企业正加速构建覆盖芯片、操作系统、应用及云服务的全链路安全防护体系，确保 AI 终端在安全可控基础上实现高质量发展。同时，中国信息通信研究院正推进生成式 AI 产品与服务安全和用户权益保护测评标准制定，构建涵盖模型产品、信息处理、服务响应及典型场景的测评体系，并持续优化“智御”个人信息保护大模型能力，为 AI 技术规范应用与创新提供技术支撑。

### 1.构建覆盖应用、模型与数据的端云协同 AI 终端纵深防御体系

为保障终端 AI 安全，行业正构建覆盖应用、模型算法、数据的纵深防御体系，系统性筑牢 AI 终端安全屏障。

**强化 AI 终端应用权限安全管控。**对 AI 应用遵循权限最小化原则，按场景化应用服务需求动态分配细粒度权限，避免过度授权；依托轻量化内容分析工具实时筛查 AI 应用操作指令，避免恶意输入导致权限滥用。

**加固 AI 终端模型协同推理安全与准确性。**结合语义分析引擎与动态规则库解析 AI 输入和输出，拦截 AI 协同推理的非安全内容或

恶意指令，加固 AI 协同推理安全；引入事实核查知识库对 AI 输出内容进行真实性校验，提升 AI 协同推理准确性。

**建立 AI 终端多维度数据安全防护。**在 AI 模型训练阶段，采用数据增强、审查过滤加固预训练数据，对检索增强生成的上下文记忆数据实施访问控制，对个人信息、模型参数等数据均通过加密、完整性校验防泄露和篡改，保护用户隐私和数据安全。

## 2.保障 AI 终端智能体感知交互的融合安全

随着智能体逐步成为实现“场景预见”的核心载体，业界正从感知、交互等层面构建融合安全保障，提升用户信任水平。

**加强智能体感知数据隐私安全。**对屏幕信息、语音等感知数据实施上下文感知授权，避免意图感知过程采集过多信息；依托轻量化智能过滤引擎实现感知数据的语义遮蔽、模糊化处理，保护意图感知数据不包含用户隐私信息。

**保障智能体协同交互安全。**通过建立界面语义校验和完整性检测，确保智能体交互界面安全；以协议安全扫描和白名单等方式支持 MCP 交互协议安全，增强智能体与工具交互调用安全；加强智能体的细粒度权限管控，确保智能体与操作系统交互权限安全。

**完善智能体安全治理标准，防范系统性风险。**针对智能体自主决策特性，构建覆盖底层技术安全（训练数据安全、模型鲁棒性、代码漏洞防护）、行为安全与伦理对齐（制定行为“红线”标准）、社会安全与治理（建立研发-部署-运行全生命周期追溯审计机制，

明确责任主体）的完整体系，防范个人权益侵害、系统运行风险及社会性危害。

### 3.建立 AI 端云协同的数据隐私防护新范式

端云协同是当前 AI 终端“场景化服务共生”的主要技术路线，针对云端存在数据隐私泄露风险，行业正积极建立“安全可信、用后即焚、随时可验证”的数据隐私防护新范式。

**实施数据内容端到端加密。**将设备端的安全隐私标准扩展至云端，用户数据通过端到端加密方式将用户隐私数据安全传输给云端节点。

**建立硬件锚定信任根。**基于硬件信任根对云端节点执行安全启动链验证固件完整性，构建硬件隔离的安全处理空间，确保 AI 模型处理用户数据的过程与 AI 云服务相互隔离，防止用户隐私数据泄露。

**确保云端数据用后即焚。**用户隐私数据仅在完成 AI 模型推理计算的必要周期内暂存数据，计算结束后立即按预设规则删除，确保用户隐私数据“用后即焚”，有效降低泄露风险。

**构建云端安全验证。**对 AI 云服务生成云端软硬件环境、日志等关键信息的可验证安全记录值，在用户终端传输数据前远程验证云端节点可信度，保障用户隐私数据仅传递给可信的 AI 云服务。

**实现账号与路由匿名化。**对用户请求信息去除身份信息，并在网络路由过程中隐藏用户 IP 地址，确保用户身份、请求内容与路由路径无法关联。

## 五、以 AI 手机和 AI PC 为突破口，加快实现“人工智能+”终端普及目标

今年是“十四五”规划收官之年，也是“十五五”谋篇布局之年。国务院《关于深入实施“人工智能+”行动的意见》（国发〔2025〕11号）明确提出“到2027年，率先实现人工智能与6大重点领域广泛深度融合，新一代智能终端、智能体等应用普及率超70%……到2030年，我国人工智能全面赋能高质量发展，新一代智能终端、智能体等应用普及率超90%”的目标。《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》明确要求全面实施“人工智能+”行动，以人工智能引领科研范式变革，加强人工智能同产业发展、文化建设、民生保障、社会治理相结合，抢占人工智能产业应用制高点，全方位赋能千行百业。加强人工智能治理，完善相关法律法规、政策制度、应用规范、伦理准则。完善监管，推动平台经济创新和健康发展。

AI终端凭借其贴近用户、直连场景、实时响应的独特优势，正成为技术落地的关键入口和价值释放的核心载体。加快构建性能优异、协同开放、安全可信的AI终端生态，对实现“人工智能+”从技术突破走向规模应用具有决定性意义。作为我国人工智能产业落地的先导领域和重要主战场之一，AI手机、AI PC等产品已率先具备规模化落地条件的现实基础，建议从政策引导、标准统一、技术突破和生态协同四个维度系统推进，重点加快在AI手机和PC领域

的率先普及。

### **（一）强化政策牵引，优先支持 AI 手机与 PC 成为：“人工智能+”落地主阵地**

AI 手机、AI PC 是“人工智能+”行动落地的重点载体，在教育、医疗、办公、养老、出行等高频民生和产业场景中优先部署试点示范项目。通过消费补贴、政府采购、行业准入等方式，鼓励用户换机升级；同时制定面向 AI 终端的专项产业扶持政策，对集成端侧大模型、具备智能体能力的设备给予研发激励和市场推广支持。建立以用户体验和实际效能为导向的评估机制，推动“好用、实用、爱用”的 AI 终端加速进入千家万户。

### **（二）加快标准统一，打通“一次开发、多端部署”的技术底座**

聚焦 AI 手机与 PC 共性需求，优先制定端侧大模型部署、量化压缩、推理接口、能效管理等关键标准，推动主流芯片厂商开放 NPU 加速能力并实现接口兼容。加快构建统一的智能体协同协议（如意图框架、MCP、A2A 等），确保不同品牌、不同操作系统的 AI 终端在任务编排、服务调用和权限管理上具备互操作性。同步推进 AI 终端操作系统、跨设备服务流转等中国方案纳入 IEEE、3GPP 等国际标准体系，提升我国在 AI 终端生态中的话语权。

### （三）突破关键技术，全面提升端侧智能性能与能效平衡

建设面向 AI 手机和 PC 的端云协同技术验证平台，构建覆盖办公、通信、健康管理等典型场景的高质量测试数据集，加速混合模型（云端大模型+端侧轻量化模型）与终端硬件的深度适配。重点优化端侧大模型的推理效率、响应速度和功耗表现，推动操作系统原生集成 AI 调度、记忆管理和智能体运行环境。同步建立 AI 终端能效标识制度，将典型 AI 负载下的能耗纳入产品评价体系，引导产业向高性能、低功耗、长续航方向升级。

### （四）构建安全可信生态，推动开源协同与用户权益保障并重

依托龙头企业和开源社区，打造面向 AI 手机与 PC 的开源大模型、工具链和智能体开发平台，降低中小企业创新门槛。同步构建覆盖芯片、系统、应用、算法和数据的全栈安全检测机制，制定 AI 终端安全、智能体行为规范及用户隐私保护标准。建立动态化、场景化的 AI 终端智能化水平测评体系，定期发布“AI 终端体验指数”和“安全可信评级”，以透明化评估引导市场良性竞争，增强用户对 AI 终端的信任感和使用意愿。

## 术语中英文对照表

中文术语	英文术语
智能体	Agent
人工智能	Artificial Intelligence (AI)
波束成形	Beamforming
比特网络	BitNet
中央处理器	Central Processing Unit (CPU)
信道状态信息	Channel State Information (CSI)
端云协同	Cloud-Edge Synergy
认知协同	Cognitive Collaboration
存算一体	Compute-in-Memory (CIM)
数字信号处理器	Digital Signal Processor (DSP)
图形处理器	Graphics Processing Unit (GPU)
图像信号处理器	Image Signal Processor (ISP)
意图框架	Intent Framework
大模型	Large Language Model (LLM)
低功耗双倍数据速率内存	LPDDR (Low Power Double Data Rate)
微机电系统	Micro-Electro-Mechanical Systems (MEMS)
混合专家模型	Mixture of Experts (MoE)
模型幻觉	Model Hallucination
调制解调器	Modem
多模态交互	Multimodal Interaction
神经网络处理器	Neural Processing Unit (NPU)
操作系统	Operating System (OS)
光学图像防抖	Optical Image Stabilization (OIS)

中文术语	英文术语
有机发光二极管	Organic Light-Emitting Diode (OLED)
个人知识图谱	Personal Knowledge Graph (PKG)
私有云计算	Private Cloud Compute (PCC)
可重构智能表面	Reconfigurable Intelligent Surface (RIS)
场景化预见	Scenario-based Prediction
场景化服务共生	Scenario-based Service Symbiosis
智能终端	Smart Terminal
芯片系统	System on Chip (SoC)
通用闪存存储	UFS (Universal Flash Storage)
统一内存架构	Unified Memory Architecture (UMA)

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：010-62300393

传真：010-62304980

网址：[www.caict.ac.cn](http://www.caict.ac.cn)

