

AI 端侧应用系列报告（三）

苹果手机的三层成长逻辑

行业研究 · 行业专题

投资评级：优于大市（维持）

证券分析师：熊莉

021-61761067

xionglil@guosen.com.cn

S0980519030002

- **三层次模型调用保障用户隐私，实现更强大、更高效且高度优化的 AI。** AI部署方式：设备端与云端相结合，本地小模型、私有云苹果大模型、外加gpt-4o 三层次模型调用。云端和本地终端处理方式效率不同，但仍能提供相近体验，并在成本、能耗、隐私与安全、个性化等方面带来优势；同时，通过优秀的5G连接技术确保信息在端到端之间高效传输。
- **苹果在AI领域耕耘三层逻辑，巩固高端市场地位。** 1) 换机潮：随着AI大模型与智能手机结合的趋势已经日渐明朗。AI介入的软硬一体化战略即将推出更智能、更个性化的设备，有望引发下一股大规模的换机潮，众多金融服务公司给出了对于苹果收益的利好预测。2) “新苹果”税：原先在用户购买下载App时，苹果公司会通过其支付系统收取一定比例的费用，这部分分成在更隐私更个性化的AI语境下必然继续优化改变。3) 硬件扩展：最新发行的M4芯片包含一个16核神经处理单元，NPU性能为38 TOPS，比M2芯片快了整整一倍，比A11仿生芯片强了近60倍。同时，随着AI大模型变得更精简、可在设备上运行并专注于推理任务，苹果自研开发AI推理芯片，实现芯片升级。
- **风险提示：**技术研发不及预期、市场竞争加剧、下游需求不及预期。

01

苹果AI的部署方式

02

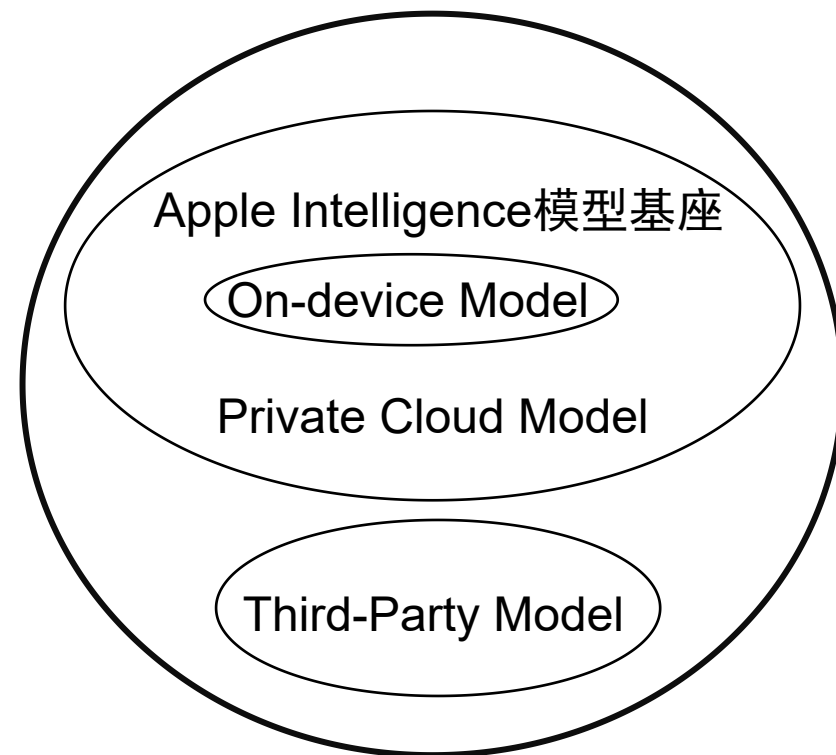
苹果在ai领域的三层逻辑

03

风险提示

◆ AI手机的三层模型

- 第一层为Apple Intelligence模型基座中的自有模型（On-device），安装在用户设备上，通常具有较小的模型尺寸，以适应设备上的计算和存储限制，成为交互新UI。
- 第二层为Apple Intelligence模型基座中的私有云模型（Private Cloud），是储存在Apple私有云服务器上的较大模型，处理更复杂的任务（目前还未命名）。
- 第三层为第三方模型（Third-Party），由其他公司或组织开发的模型，Apple可能与之合作以提供服务或功能，处理剩下约10%响应需求。



On-Device Model
功能收敛，不做aigc

Private Cloud Compute Model
处理复杂任务，自研芯片推理

Third-Party
外部大模型

优先级1

优先级2

优先级3

01

苹果AI的部署方式

02

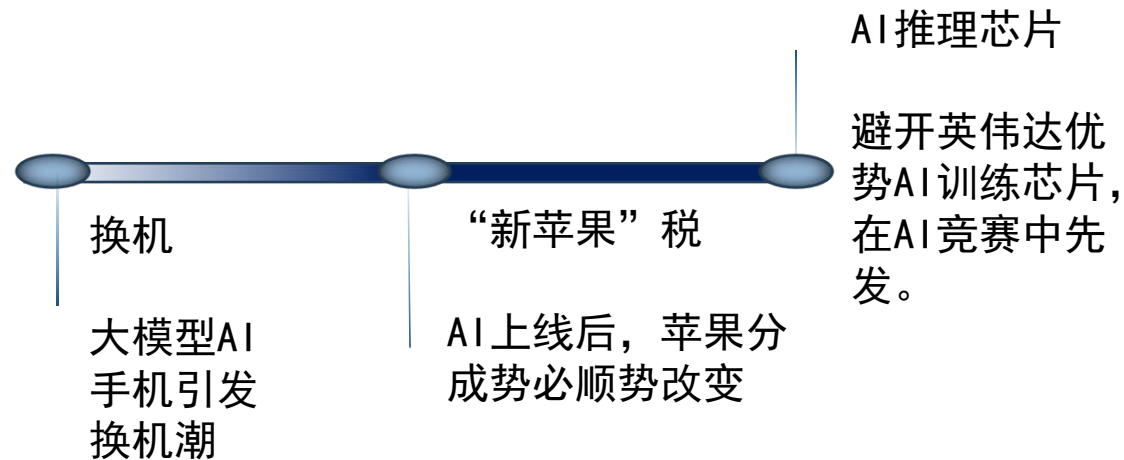
苹果在ai领域的三层逻辑

03

风险提示

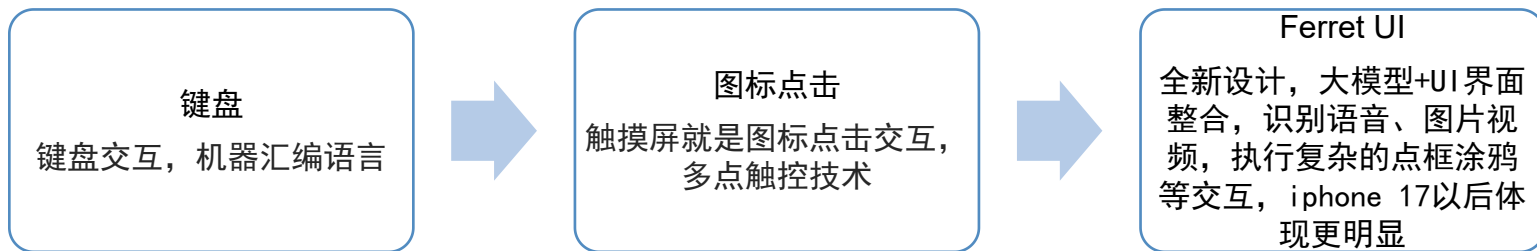
◆ 苹果AI的三层逻辑

- 第一逻辑为换机，AI大模型与智能手机结合的趋势已经日渐明朗。苹果将于明年发布的iPhone16有望成为其首款AI手机，有望引发一股大规模的换机潮，众多金融服务公司给出了对于苹果收益的利好预测。
- 第二层为“新苹果”税，苹果用户通过苹果App Store进行App付费下载或购买数字商品及服务时，苹果公司会通过其支付系统收取一定比例的费用，这部分分成在苹果的垄断地位下显得尤为突出，也遭受过诸多审查。AI上线后苹果税也必然相应改变调整。
- 第三层为硬件扩展，其中以AI推理芯片为主要落脚点。过去十年间，苹果公司积极推进iPhone、iPad、Apple Watch和Mac电脑的自研芯片设计，新发行的M4芯片大有可为。而与此同时，苹果公司已在自主开发可以在数据中心服务器中运行AI软件的推理芯片，避开英伟达有望为其在AI竞赛中带来一项关键优势。

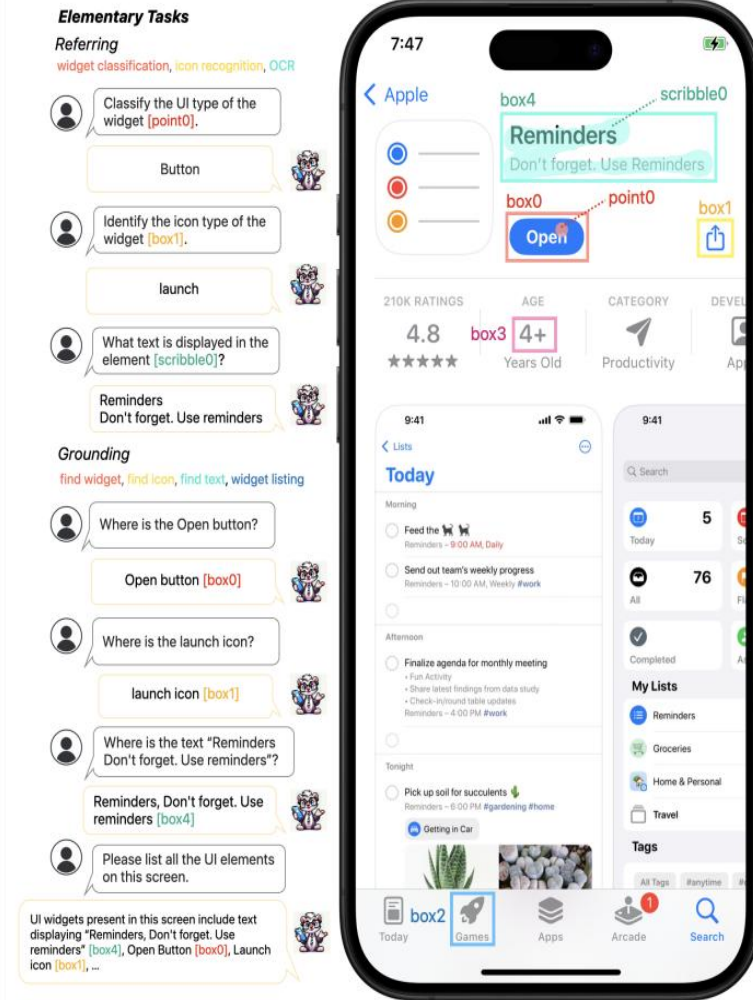


第一逻辑：换机-Intelligence

- ◆新siri是一个Ferret UI，全新的UI交换界面，从图标点击进入了新的交互方式领域。



- ◆此UI的目的就是去训练大模型去视觉上理解手机屏幕UI界面，自动化用户操作中的感知以及交互功能。它专注于屏幕内的特定 UI 元素，为更高级的任务打下基础，例如将自然语言映射到UI屏幕上的操作，执行更高级的推理功能。
- ◆referring tasks和grounding tasks，即引用（可以让用户以自由形状圈定屏幕上的某个区域）和定位（大模型输出回答时正确地选定相关区域的能力）。

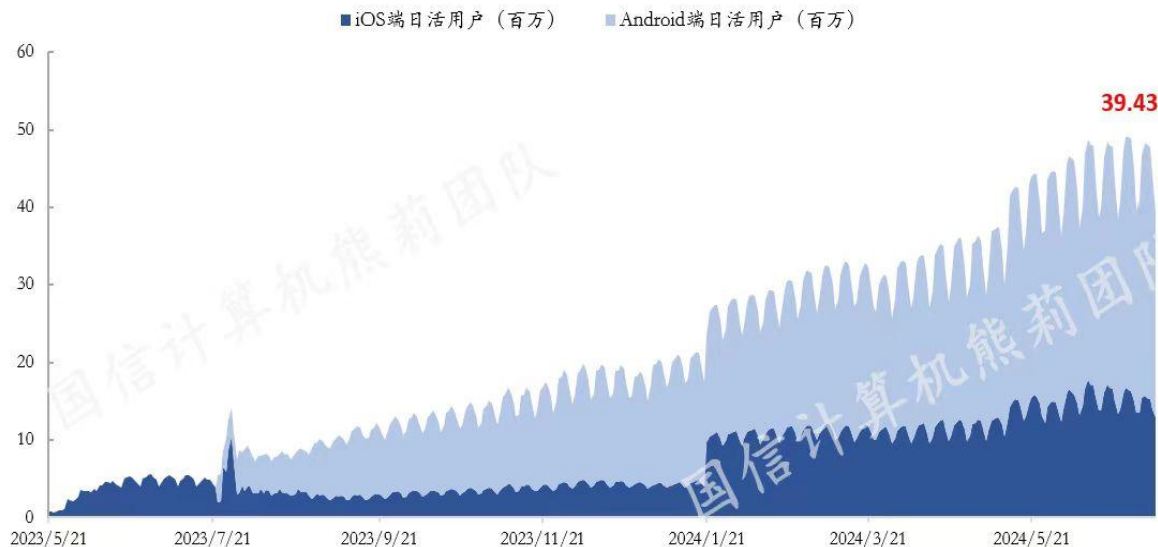


第一逻辑：换机-Openai

- ◆ Siri18 可以直接调用Openai 模型，不需要账号登录。Siri 会判断查询是否有助于转发到 ChatGPT，并询问是否允许共享，用户可以提出与文本、文档、照片、PDF 等相关的问题。
- ◆ 目前GPT-4o发布之后，ChatGPT日活攀升，月活从2.6亿涨到3亿，客户体验反馈良好。

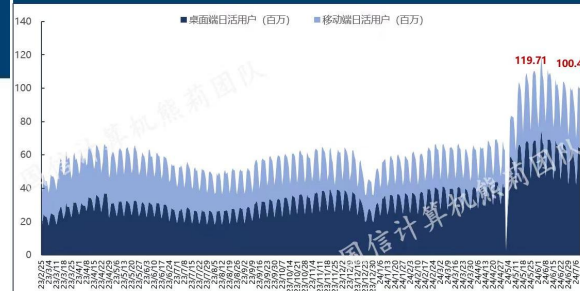
【国信计算机】ChatGPT APP端数据跟踪

2024年7月14日更新



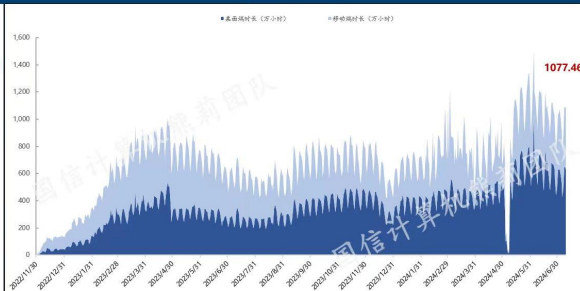
【国信计算机】ChatGPT访问量数据跟踪

2024年7月14日更新



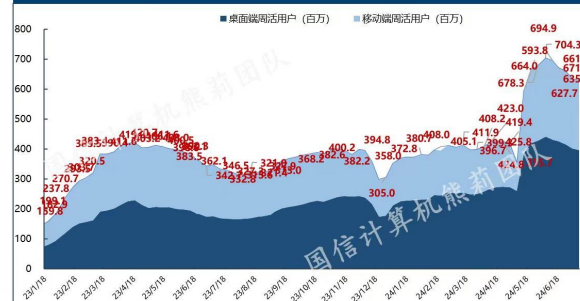
【国信计算机】ChatGPT访问时长数据跟踪

2024年7月14日更新



【国信计算机】ChatGPT周活数据跟踪

22.12.14-2024.7.10 (最后统计周为7.4-7.10)



【国信计算机】ChatGPT平均访问时长数据跟踪



资料来源: Openai官网、国信证券研究所整理

请务必阅读正文之后的免责声明及其项下所有内容



Openai 的ChatGPT触达C端用户 远不及苹果14亿用户。没有GPT-4o 这个agent模型，月活用户停留在2.6亿半年之久。



GPTs暂时宣告不成功，微软云下架了GPTs相关产品。微软发布公告，Copilot GPTs 将于7月10日起停服，用户已经创建的GPTs将被清空。微软给出的官方解释是公司战略调整正在将GPT的重点转向商业和企业场景，背后原因可能是缺乏商业回报。



Google也在积极准备接入苹果，和Openai位置一样，提供云端大模型服务。Mark Gurman最新透露苹果公司将于今年秋季宣布与Alphabet旗下的谷歌的大模型Gemini 建立合作关系，届时Gemini 模型将有望接入苹果的 Apple Intelligence。



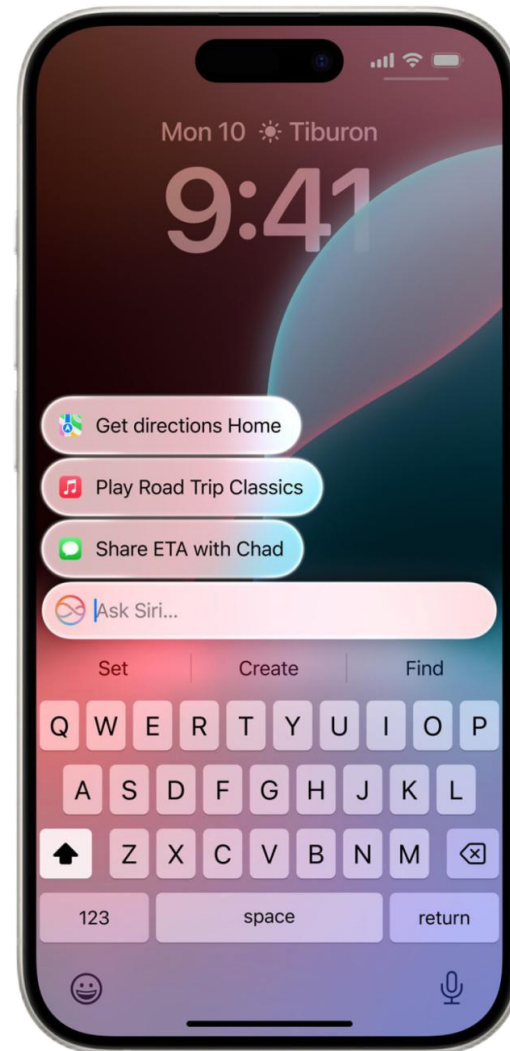
Meta积极洽谈苹果，准备云端提供大模型服务。苹果与人工智能初创公司 Anthropic的潜在合作关系也在考虑之中，不过，苹果仍在考虑是否整合Meta公司的人工智能服务。

◆新机上线更强Siri

- 今年的iphone 15pro iphone16 升级之后，Siri18可以直接链接云端的GPT-4o。AI普适性的生活功能，今年就能实现，但跨app内容组合功能 要等iphone 17之后才有体现。
- Siri 现在可以为用户提供全面的设备支持，无论他们身在何处，都能解答关于 iPhone、iPad 和 Mac 操作的数千个问题。比如，用户可以学习如何在 Mail 中安排邮件、如何从浅色模式切换到深色模式等各种操作。

◆预计换新后苹果服务业绩猛增

- 预计2025年和26年iphone销量达到4.98亿，单财年增增长22%。Apple Intelligence技术的推出将激发潜在的“换新”需求，从而开启“多年升级周期”，带动出货量强劲上涨。预计在25财年和26财年，66-69%的iPhone出货量将是新机型，比24财年高出2-6个百分点，比22财年之前的峰值高出3个百分点，推动iPhone ASP年度增长4-5%。
- Apple Intelligence功能将成为苹果设备多年升级周期的“显著催化剂”。受益于AI技术，Woodring预计，苹果的平均售价（ASP）年增长率将达5%。



第二逻辑：新苹果“税”

◆ 苹果生态紧密度高，苹果对个人客户，对内容商提供数据保护

- 个人隐私模糊处理清洗
- 内容商调用但保证不用客户数据训练
- 未经同意，不得将数据接入大模型，如ChatGPT
- “端侧模型+云上模型”的方式，使用信息都会集中存到Apple服务器上而非个人



◆ 跨app任务窗口

- “Hi siri, 帮我看看我下午的飞机不能起飞概率有多高？如果高于65%，麻烦帮我换高铁”
- 授权商旅网，航旅纵横，航空公司，12306，支付第三方

◆ Intelligence做任务接收，单一指令，内容商接入API，做交互

- 复杂指令，接受任务，云端拆解任务
- 调取不同内容数据库，给予解决方案
- AI拆解任务、调用APP/API接口等一些能力的支持



第二逻辑：新苹果“税”

- 苹果和大模型厂商合作分成

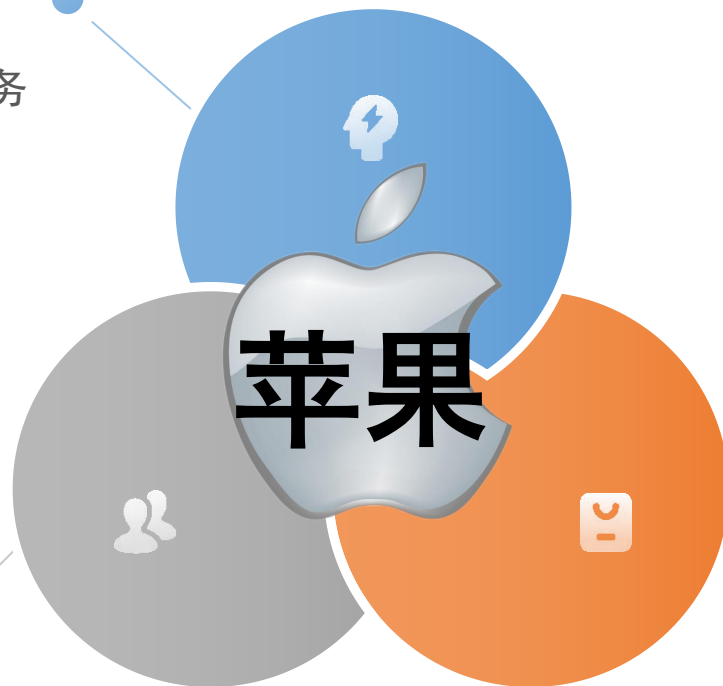
客户

付钱者，提出要求，提出任务



内容厂商

内容厂商，接入生态，根据自身情况，提高不同程度的调取服务。



大模型

提供大模型能力，最终按照调用tokens数量进行结算，最开始免费，培养客户习惯



◆不同厂商之间采取不同合作方式

- 苹果的目标是通过AI代理改变人机交互方式，将数据和隐私信息统一管理，以更全面地理解用户需求。尽管AI代理会降低用户在某些APP上的时长，但苹果的AI功能将替代部分工具型APP，实现更高效、更个性化的用户体验

◆数据商

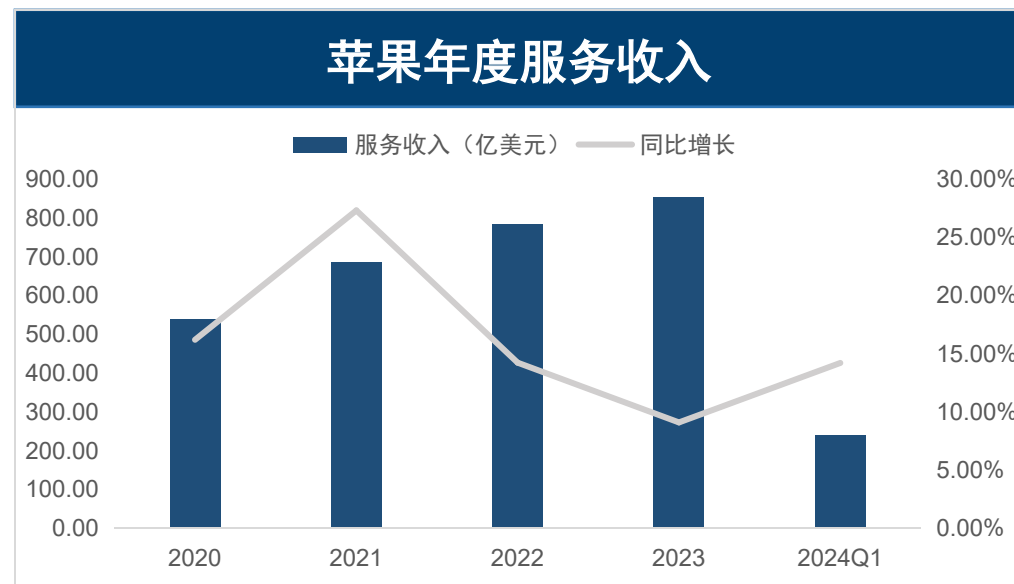
- 接入Intelligence，被调取，应用服务苹果分成大头

◆内容服务商

- 本身内涵数据和服务，微信和亚马逊等，提供较成熟解决方案，siri接入它们的服务端口，内容商占大头。苹果的终端覆盖优势，内容商的格局未来可能都会有大变化

◆苹果最新财报

- 单季度服务收入238亿美金，相对去年同期上升14.16%
均摊14亿用户，一年一台机贡献68美元
- 若新Intelligence后续能成功，议价能力继续攀升





安卓生态的问题

- 模型与芯片的适配问题，高通gen4等算子特别高，但是对于内存，显存等如何和模型适配，还需要磨合。
- 模型厂商如何删除参数，保留哪些模型，留哪些功能适配芯片，也需要磨合。
- 安卓体系手机主要市场在韩国和中国，没有多模态大模型做支撑，没有往这方向的新UI研发方向靠。

HarmonyOS 鸿蒙生态的问题

- 正在迁移，这个工程量本身很大
- 后续如何引入内容厂商加入新UI生态
- 云端模型选择则是各家手机厂商要做的事

◆未来，Google+高通给出了一个一体化方案

- GPU+NPU+DRM +模型，赋予手机厂商新UI，从而助力崛起

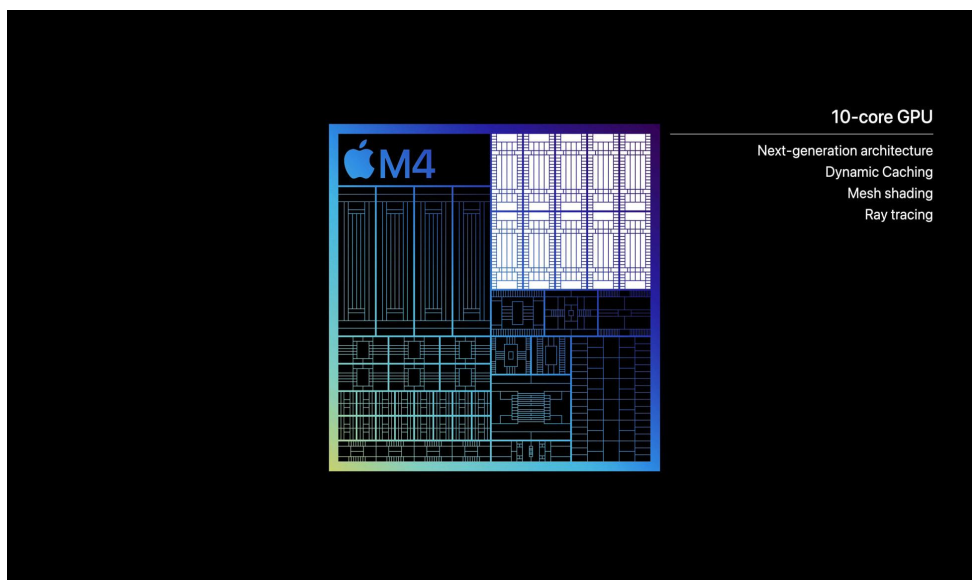


◆ Privide Cloud

- 多核心CPU：M4拥有一个全新的CPU，最多包含10个核心，包括性能核心和效率核心。
- 10核GPU：M4的GPU基于M3芯片中引入的下一代GPU架构，首次为iPad带来动态缓存、硬件加速的光线追踪和网格着色技术。

◆ 最快的神经引擎

- M4拥有Apple迄今为止最快的神经引擎，每秒能够执行高达38万亿次运算，速度超过当今任何AI PC的神经处理单元。



- ◆ M5预计会使用TSMC的封装技术（packaging technology），目的在于提高消费者的Mac以及AI服务器和云计算工具的性能表现。

第三逻辑：AI推理芯片

◆ 苹果开发自研推理ai芯片，但今年不会投放大量芯片

- 当前，AI市场主要集中在使用大数据训练大语言模型的[训练]阶段，英伟达成为这一领域的主要受益者。
- 然而，随着AI大模型变得更精简、可在设备上运行并专注于推理任务，芯片制造商的市场重心将转向[推理]，即模型应用
- 展望产业发展趋势，AI算力负载有望逐步从训练向推理端迁移，从而降低AI芯片门槛。

◆ 苹果要求大模型厂商将模型放在自己的私有云

- 保证客户和内容商数据安全，具有相对大的体量。

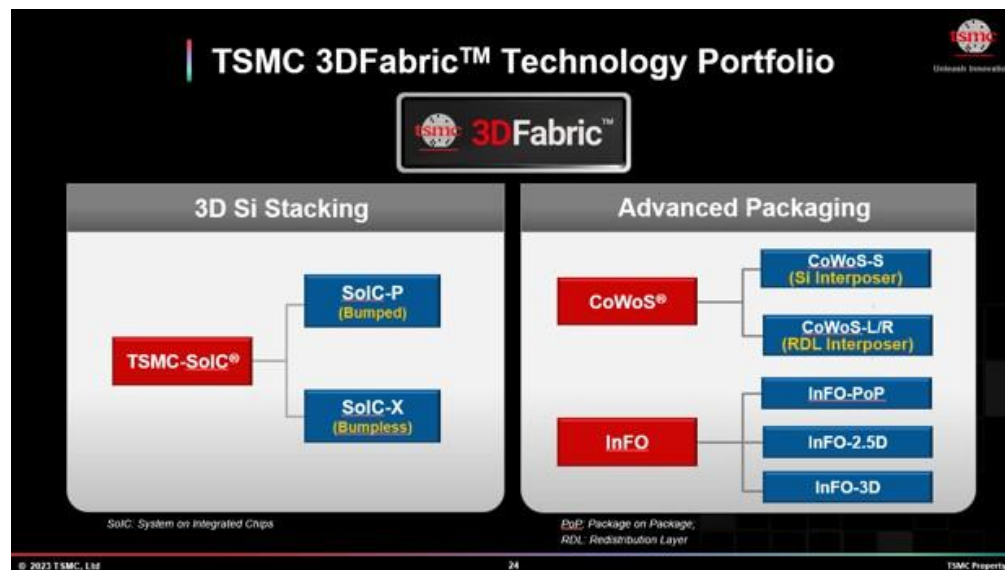
预计 A18 Pro 的芯片尺寸将比 A18 大 15-20%，两款处理器均采用台积电的 N3E 节点



第三逻辑：AI推理芯片

- ◆ 苹果未来将在M系列的芯片上，例如M5 Pro/Max/Ultra等处理器上采取台积电（TSMC）的SoIC-X的3D封装技术。台积电预计2025年及以后将大幅度扩大SoIC产能。
- ◆ 该系列芯片将用于Mac和一些即将推出的苹果私有云计算的基础设备上
- ◆ 苹果疑似想要分割M系列芯片中的CPU和GPU为不同的芯片（chiplets），具体细节暂未明了

台积电先进封装技术组合 -3DFabric

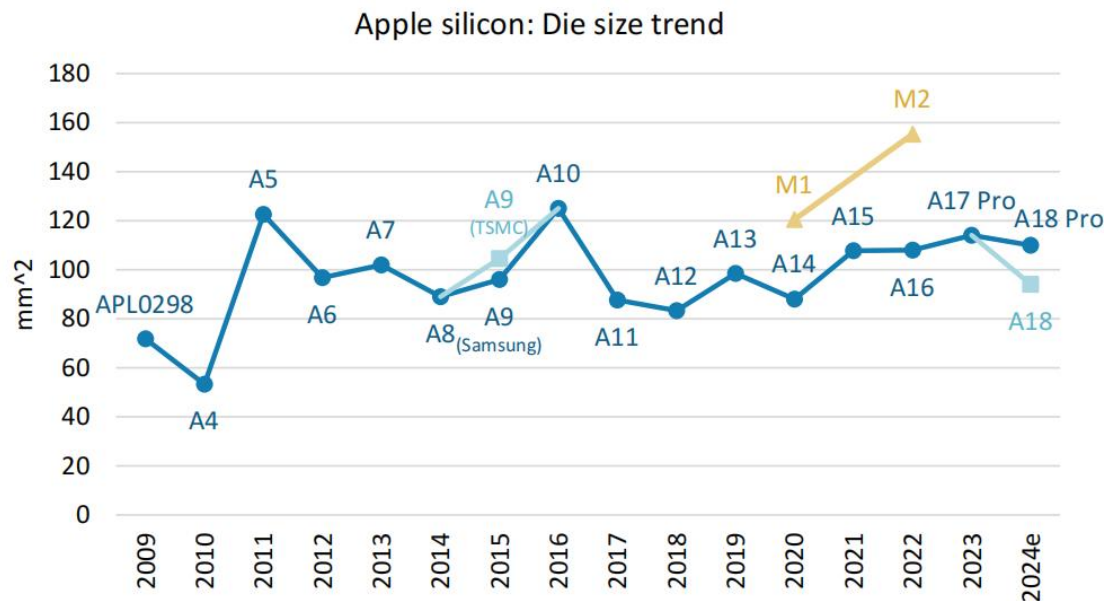


资料来源：苹果官网、国信证券研究所整理

第三逻辑：AI推理芯片

- ◆关于“M5芯片”的最早提及的地方在官方Apple code当中。
- ◆苹果采用SoIC工艺主要目的是为了在有限的性能牺牲下节约成本。
- ◆缩小SRAM（静态随机存取存储器）变得越来越有挑战性，在3nm或2nm工艺节点上制作I/O功能性价比低。chiplet成为解决问题的答案——将不同的组件分成不同工艺节点上的更小的芯片，以便更先进的节点主要用于计算功能，而内存（SRAM）和模拟（I/O）功能可以用相对更成熟的节点（例如5nm）制造。

预计 A18 Pro 的芯片尺寸将比 A18 大 15-20%，两款处理器均采用台积电的 N3E 节点



资料来源：苹果官网、国信证券研究所整理

◆ 芯片端

- 2017年，苹果发布A11处理器，首次搭载神经网络引擎处理器单元NPU，打开移动终端的AI时代。
- 之后，苹果A系列芯片不断升级，2020年开始发布的M系列芯片更带来了突破性的机器学习性能。
- 2024年5月，苹果全新发布M4芯片，包含一个16核神经处理单元，NPU性能为38 TOPS，比M2芯片快了整整一倍，比A11仿生芯片强了近60倍。

◆ 平台端

- 配备 A17 Pro 芯片或任何 M 系列芯片的设备即将能够使用Apple intelligence，包括iPhone 15 Pro, iPhone 15 Pro Max, M1或更高版本的Mac, iPad Pro 和iPad Air。
- 即将推出的iPhone 16系列等也将支持Apple Intelligence。
- 公司将在未来一年里陆续更新Apple Intelligence功能，并支持更多的语言和设备。



01

苹果AI的部署方式

02

苹果在ai领域的三层逻辑

03

风险提示

- **技术研发不及预期：**目前ai手机技术尚未完善，在模型迭代、数据集、算力等角度存在技术不及预期的可能。
- **市场竞争加剧：**目前该领域面临激烈的价格竞争和毛利率下行压力、竞争对手对技术进步的快速采用以及消费者对价格的高敏感等特点。
若市场竞争加剧，将对相关公司营收带来不利影响。
- **下游需求不及预期：**整体宏观经济环境存在不确定性，若下游需求复苏不及预期，可能导致相关产业链公司业绩承压。

国信证券投资评级			
投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后6到12个月内的相对市场表现，也即报告发布日后的6到12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数（000300.SH）作为基准；新三板市场以三板成指（899001.GSI）为基准；香港市场以恒生指数（HSI.HI）作为基准；美国市场以标普500指数（SPX.GI）或纳斯达克指数（IXIC.GI）为基准。	股票投资评级	优于大市	股价表现优于市场代表性指数10%以上
		中性	股价表现介于市场代表性指数±10%之间
		弱于大市	股价表现弱于市场代表性指数10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业投资评级	优于大市	行业指数表现优于市场代表性指数10%以上
		中性	行业指数表现介于市场代表性指数±10%之间
		弱于大市	行业指数表现弱于市场代表性指数10%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。



国信证券

GUOSEN SECURITIES

国信证券经济研究所

深圳

深圳市福田区福华一路125号国信金融大厦36层

邮编：518046 总机：0755-82130833

上海

上海浦东民生路1199弄证大五道口广场1号楼12楼

邮编：200135

北京

北京西城区金融大街兴盛街6号国信证券9层

邮编：100032