

行业研究

iPhone 时刻的开启，硅基时代的到来

——AI 行业深度报告之一

要点

ChatGPT 引领人工智能进入以 AGI 为代表的新里程碑阶段，拉动算力需求，引领基础设施新周期。人工智能应用场景逐渐增多，模型参数规模和数据量也实现了大幅度增长，为 NLP、CV 等领域带来更强大的表达能力和性能。大模型是目前通往 AGI 的最佳实现方式。算力需求激增，推动算力基础设施迎增长新周期。据 IDC，AI 的应用与普及促使 2022 年我国智能算力规模近乎翻倍，达到 268EFLOPS，超过通用算力规模，2026 年我国智能算力规模将达 1271.4 EFLOPS，22-26 年复合增长率预计达 47.58%；硬件占 AI 支出比重 49.8%，其中最大投资是服务器，占 AI 硬件支出比重 84%以上，AI 将成为未来服务器市场新一轮快速增长的主要推动力。

全球科技巨头大力布局 AI 产业。（1）**微软：投资 OpenAI，Copilot 开启 AI 商业化里程碑。**2023 年 1 月中旬，微软对 OpenAI 投资 100 亿美元，占股达到 49%。微软推出全新的 Copilot 系统，发布 Dynamics 365 Copilot、Microsoft 365 Copilot、Power Platform Copilot 模块。（2）**英伟达：全球 GPU 龙头，产品需求量子价齐升。**全球 GPU 龙头开拓 AI 市场，助力服务器核心算力硬件性能提升。英伟达作为一家以 GPU 为主营业务的半导体设计公司，在 AI 行业发展初期就前瞻布局。FY4Q24 数据中心业务营收达到创纪录的 184 亿美元，环比增长 27%，同比增长 409%。算力需求迅猛增长，服务器 GPU 业务有望迎来量价齐升。（3）**华为：服务器产业深耕多年，打造“鲲鹏+昇腾”双引擎战略布局。**据 counterpoint 数据，2021 年全年华为服务器业务收入 43.43 亿美元，全球市场份额 4.5%。公司坚持计算产业战略，积极布局 AI 服务器领域。

AI 产业链：大模型加速落地，算力空间广阔，应用百花齐放。（1）大模型：国内和海外巨头加速布局，展望未来，随着底层技术逐步革新，基模型和领域大模型持续完善，大模型应用边界不断拓宽。（2）服务器：技术迭代催生新增长点，品牌与白牌之争趋于白热化。超算中心建设加速，服务器需求进一步提升；云边协同发展，增添服务器需求动能；AIGC 推动 GPU 服务器市场持续扩张；服务器品牌与白牌之争白热化，促进服务器市场优胜劣汰。AI 训练服务器方面，单个大语言模型训练驱动 AI 训练服务器市场约 2 亿美元需求。据 IDC 与浪潮信息数据，2023 年中国人工智能服务器市场规模将达到 91 亿美元，同比增长 82.5%，2027 年将达到 134 亿美元，五年复合增长率达 21.8%。（3）光模块：800G 需求释放，中国厂商厚积薄发。AI 大模型发展加速光模块迭代节奏，800G 光模块放量或将超预期：根据 Lightcounting 2021 年的预测，800G 光模块将从 2025 年底开始主导市场，主要原因是 AI 应用等带来的数据流量的增长，超预期的数据中心带宽需求以及光模块厂商技术的迭代。（4）边缘算力：AIoT 智能终端的“大脑”，端侧算法部署的核心。（5）PCB：AI 算力提升驱动行业快速成长。（6）智能终端：AI 赋能硬件入口，加速万物互联。

投资建议：推荐工业富联、大华股份、海光信息、寒武纪、芯原股份。（1）工业富联：消费电子龙头，AI 服务器打开成长空间；（2）大华股份：中国移动入股落地，智慧物联龙头扬帆起航；（3）海光信息：CPU/DCU 领军企业；（4）寒武纪：AI 芯片技术国内领先；（5）芯原股份：稀缺的国产 IP 供应商，Chiplet 布局领先。

风险提示：宏观经济风险、AIGC 发展不及预期风险、市场需求及上游供应不足风险。

电子行业
买入（维持）

作者

分析师：刘凯

执业证书编号：S0930517100002

021-52523849

kailiu@ebsecn.com

分析师：吴春阳

执业证书编号：S0930521080002

021-52523686

wuchunyang@ebsecn.com

分析师：林仕霄

执业证书编号：S0930522090003

021-52523818

linshixiao@ebsecn.com

联系人：孙啸

021-52523587

sunxiao@ebsecn.com

联系人：王之含

wangzhihan@ebsecn.com

行业与沪深 300 指数对比图



资料来源：Wind

投资聚焦

研究背景

近几年来人工智能高速发展，现已经进入了以 AGI 为代表的新里程碑阶段。随着 AI 基础设施逐步完善，深度学习技术不断进步，人工智能应用场景逐渐增多，模型参数规模和数据量也实现了大幅度增长。伴随 AI 技术升级和大模型成熟，AI 绘画与 ChatGPT 的成功破圈，生成式 AI 技术迎来发展拐点，行业关注度大幅提升。

我们区别于市场的观点

1、市场 AI 深度报告针对某一细分领域，我们本篇报告对服务器、GPU、PCB、云端算力、光模块、边缘算力、智能终端等领域进行深度分析。本文测算中国超算服务市场的规模，800G 光模块 2023-2025 年市场规模；此外本文还从智能手机、智能音箱、智能机顶盒、智能家居四个角度分析了 AI 对于智能终端行业的影响。

2、市场认为近两年算力建设完备，未来算力建设动力不足，我们认为受益于 AI 等相关新兴领域的应用以及“东数西算”政策下云计算、超算中心的蓬勃发展，数据计算、存储需求呈几何级数增长，算力需求持续释放带动算力基础设施产业迎来增长新周期。

股价上涨的催化因素

- (1) AI 场景落地超预期
- (2) AI 芯片研发超预期
- (3) 下游应用复苏超预期

投资观点

推荐工业富联、大华股份、海光信息、寒武纪、芯原股份。

工业富联：消费电子龙头，AI 服务器打开成长空间；

大华股份：中国移动入股落地，智慧物联龙头扬帆起航；

海光信息：CPU/DCU 领军企业；

寒武纪：AI 芯片技术国内领先；

芯原股份：稀缺的国产 IP 供应商，Chiplet 布局领先。

目 录

1、 AI: iPhone 时刻的开启，硅基时代的行业特征	7
2、 全球科技巨头大力布局 AI 产业	10
2.1、 微软:投资 OpenAI, Copilot 开启 AI 商业化里程碑	10
2.2、 英伟达:全球 GPU 龙头，产品需求量价齐升.....	12
2.3、 华为：服务器产业深耕多年，打造“鲲鹏+昇腾”双引擎战略布局	14
3、 AI 产业链：大模型加速落地，算力空间广阔，应用百花齐放	16
3.1、 大模型：国内和海外巨头加速布局	16
3.2、 服务器：AIGC 推动 GPU 服务器市场持续扩张	17
3.3、 光模块：800G 需求释放，中国厂商厚积薄发	19
3.4、 边缘算力：AIoT 智能终端的“大脑”，端侧算法部署的核心.....	21
3.5、 PCB：AI 算力提升驱动行业快速成长.....	22
3.6、 智能终端：AI 赋能硬件入口，加速万物互联.....	22
4、 投资建议	24
5、 重点公司推荐	25
5.1、 工业富联（601138.SH）：消费电子龙头，AI 服务器打开成长空间	25
5.2、 大华股份（002236.SZ）：中国移动入股落地，智慧物联龙头扬帆起航	30
5.3、 海光信息（688041.SH）：CPU/DCU 领军企业.....	36
5.4、 寒武纪（688256.SH）：AI 芯片技术国内领先.....	41
5.5、 芯原股份（688521.SH）：稀缺的国产 IP 供应商，Chiplet 布局领先	45
6、 风险提示	50

图目录

图 1: 人工智能的发展	7
图 2: 大模型发展及相关应用实现落地时间表	8
图 3: 中国智能算力规模 (EFLOPS)	9
图 4: Copilot 自动生成邮件	10
图 5: Dynamics 365 Customer Insights	11
图 6: Microsoft 365 Copilot 系统	11
图 7: Power Apps	12
图 8: Power Automate	12
图 9: 2022 年 A100 及未来 H100 各领域应用数量统计	13
图 10: 优秀的大模型能够赋能各行各业开放任务	16
图 11: 大模型行业的国外厂商	16
图 12: 大模型行业的国内厂商	17
图 13: 英伟达 H100/A100 对比	17
图 14: 全球服务器与中国出货量及增速	18
图 15: 全球与中国服务器市场规模及增速 (单位: 亿美元, %)	18
图 16: 2022 年中国服务器市场份额 (按营收)	18
图 17: 2022 年 AI 服务器采购量占比	18
图 18: 数据中心光模块需求演进 (21-26 年数据为预测值)	19
图 19: 传统三层网络架构	20
图 20: 叶脊架构	20
图 21: 以太网光模块销售额增长率 (23-28 年数据为预测值)	20
图 22: SoC 示意图	22
图 23: 2021 年中国智能家居设备市场各品类出货量及同比增长率	23
图 24: 中国 AI 产业链梳理	24
图 25: 公司发展历程	25
图 26: 公司股权结构 (截至 2023 年 11 月)	25
图 27: 2016 年-2023Q1-Q3 公司营业收入	26
图 28: 2016 年-2023 Q1-Q3 公司归母净利润	26
图 29: 2016 年-2023 Q1-Q3 毛利率及净利率	27
图 30: 2016 年-2023 Q1-Q3 期间费用率	27
图 31: 公司业务及产品	30
图 32: 公司发展历程	31
图 33: 公司股权结构 (截至 2023 年三季度)	31
图 34: 2016 年-2023 Q1-Q3 公司营业收入	32
图 35: 2016 年-2023 Q1-Q3 公司归母净利润	32
图 36: 2018Q1-2023Q3 公司单季收入	32
图 37: 2018Q1-2023Q3 公司单季归母净利润	32
图 38: 2016 年-2023 Q1-Q3 毛利率及净利率	33
图 39: 2016 年-2023 Q1-Q3 期间费用率	33

图 40: 公司打造持续升级巨灵人工智能开发平台	33
图 41: 海光信息主要产品简介	36
图 42: 海光信息营业收入及增速	36
图 43: 海光信息归母净利润及增速	36
图 44: 海光信息毛利率及净利率情况	37
图 45: 公司研发费用及其营收占比	37
图 46: 海光信息部分客户	38
图 47: 寒武纪思元 370-S4 智能加速卡	41
图 48: 寒武纪玄思 1000 智能加速器整机	41
图 49: 寒武纪 2019 年-2023 Q1-Q3 营业收入及同比情况	42
图 50: 寒武纪 2019 年-2023 Q1-Q3 归母净利润情况	42
图 51: 寒武纪 2019 年-2023 Q1-Q3 毛利率及净利率情况	42
图 52: 寒武纪 2019 年-2023 Q1-Q3 研发费用及占比情况	42
图 53: 芯原股份业务	46
图 54: 芯原股份 2019 年-2023 Q1-Q3 营业收入及同比情况	46
图 55: 芯原股份 2019 年-2023 Q1-Q3 归母净利润情况	46
图 56: 芯原股份 2019 年-2023 Q1-Q3 毛利率及净利率情况	47
图 57: 芯原股份 2019 年-2023 Q1-Q3 研发费用及占比情况	47
图 58: 基于 Chiplet 的异构架构应用处理器的示意图	48

表目录

表 1: 英伟达主要芯片产品.....	13
表 2: 英伟达 AI 服务器主要产品对比	14
表 3: 华为服务器产品	15
表 4: 工业富联收入拆分.....	28
表 5: 工业富联可比公司估值-PE 估值.....	29
表 6: 工业富联盈利预测与估值简表	29
表 7: 大华股份收入拆分 (单位: 百万元, %)	34
表 8: 可比公司估值-PE 估值	35
表 9: 大华股份盈利预测与估值简表	35
表 10: 海光信息主营业务拆分及预测 (单位: 亿元, %)	39
表 11: 可比公司估值-PS 估值.....	40
表 12: 海光信息盈利预测与估值简表.....	40
表 13: 寒武纪产品布局	43
表 14: 寒武纪主营业务拆分及预测 (单位: 亿元, %)	44
表 15: 寒武纪可比公司估值	45
表 16: 寒武纪盈利预测与估值简表	45
表 17: 芯原股份主营业务拆分及预测 (单位: 亿元, %)	49
表 18: 芯原股份可比公司估值	50
表 19: 芯原股份盈利预测与估值简表.....	50

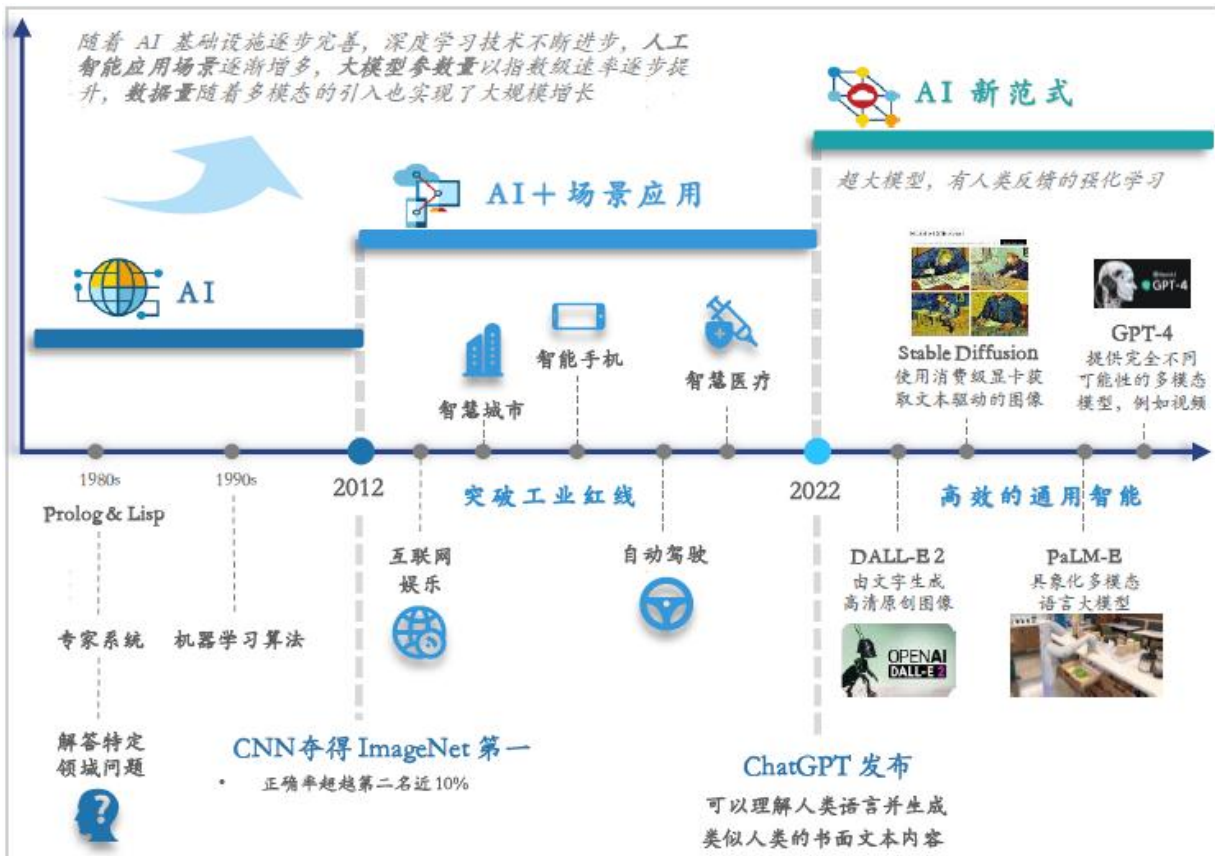
1、AI: iPhone 时刻的开启，硅基时代的行业特征

ChatGPT 引领人工智能进入以 AGI 为代表的新里程碑阶段。随着 AI 基础设施逐步完善，深度学习技术不断进步，人工智能应用场景逐渐增多，模型参数规模和数据量也实现了大幅度增长，为 NLP、CV 等领域带来更强大的表达能力和性能。人工智能发展历程中主要有两大里程碑：

里程碑一：2012 年 CNN 获得 ImageNet 第一，标志着机器视觉识别能力开始逐渐超越人眼识别准确率，开启了人工智能革命。随着深度学习技术不断突破，诞生了一批“AI+场景应用”的专属模型，但是整体研发成本比较高、研发时间比较长。

里程碑二：2022 年 ChatGPT 的出现，掀起了又一波人工智能发展热潮，以大模型+RLHF 为核心的技术落地意味着人工智能开启 AI 新范式。人工智能相关产业开始基于强大的基模型进行发展，通过人类反馈和强化学习不断解锁基模型的能力，以解决海量开放式任务，带来了新的研究范式。

图 1：人工智能的发展

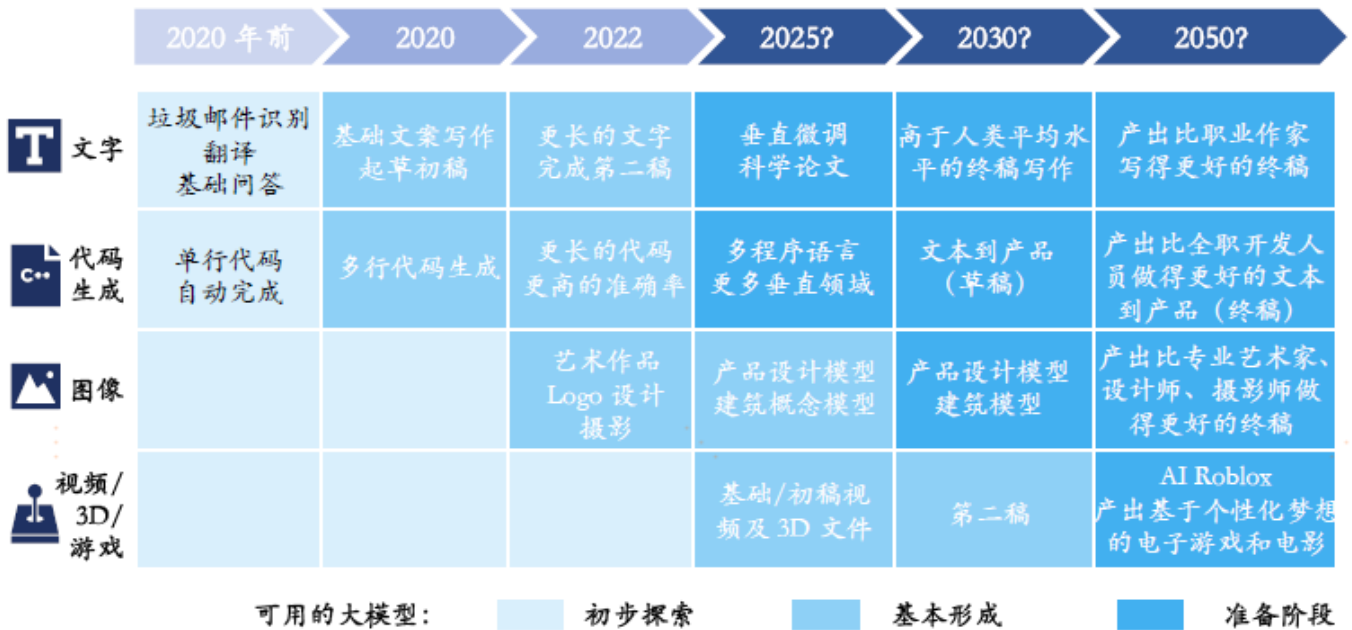


资料来源：沙利文

AGI 技术能够精准识别人类情绪意图、理解人类语言、学习人类知识并进行类脑推理与创造，大模型是目前通往 AGI 的最佳实现方式。以 ChatGPT 为代表的人工智能技术已经具备 AGI 的核心技术和特征，能够自动化地学习任何可以符号

化的知识及信息，不断自我优化，充分理解和流畅表达人类语言，同时逻辑推理能力强，实现了具备一般人类智慧的机器智能。

图 2：大模型发展及相关应用实现落地时间表



资料来源：沙利文

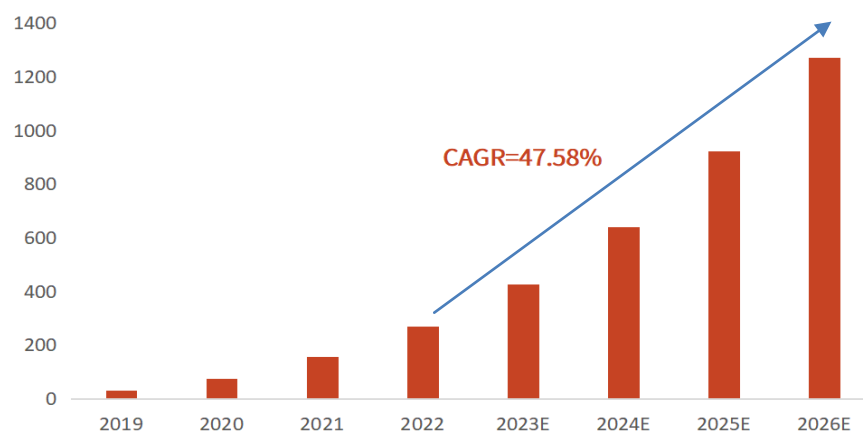
伴随 AI 技术升级和大模型成熟，AI 绘画与 ChatGPT 的成功破圈，生成式 AI 技术迎来发展拐点，行业关注度大幅提升。生成式 AI 是指基于大模型、生成对抗网络 GAN 等人工智能技术，通过已有数据寻找规律，并通过适当的泛化能力生成相关内容的技术，可生成如图像、文本、音频、视频等原创内容变体。例如，以 ChatGPT、Midjourney、文心一格、商汤商量、Codex 为代表的生成式 AI 应用拥有文本语言理解能力、涌现能力以及思维链推理能力，能够完成文学创作、新闻写作、数理逻辑推算、代码生成、图片生成等多项任务。目前，国内电商、游戏、文娱、设计等行业正在积极使用相关的生成式 AI 应用来提高自身工作效率，尤其以文生图应用为主。

生成式 AI 不仅能够增强并加速下游多领域的设计，而且有潜力“发明”人类可能错过的**新设计、新对象**。生成式 AI 有生成大规模、高质量、低成本内容优势，在算力和算法支持下生成大量内容，生成的内容质量将持续超越 UGC 与 PGC。未来有望为各行业提供内容支持并促进其内容繁荣，最大化释放内容生产力。

文字生成属于发展成熟、易于跨界转化的赛道，而跨模态生成赛道的发展潜力最大。生成式 AI 应用根据模态划分为文字生成、音频生成、图像生成、视频生成、跨模态生成。语音合成、文本生成图像属性编辑等技术应用目前较为成熟，跨模态生成、策略生成是高增长潜力的应用场景，在自动驾驶、机器人控制等领域有极高应用价值。

算力需求激增，推动算力基础设施迎增长新周期。受益于 AI 等相关新兴领域的应用以及“东数西算”政策下云计算、超算中心的蓬勃发展，数据计算、存储需求呈几何级数增长，算力需求持续释放带动算力基础设施产业迎来增长新周期。据 IDC，AI 的应用与普及促使 2022 年我国智能算力规模近乎翻倍，达到 268 EFLOPS，超过通用算力规模，2026 年我国智能算力规模将达 1271.4 EFLOPS，22-26 年复合增长率预计达 47.58%；硬件占 AI 支出比重 49.8%，其中最大投资是服务器，占 AI 硬件支出比重 84%以上，AI 将成为未来服务器市场新一轮快速增长的主要推动力。

图 3：中国智能算力规模（EFLOPS）



资料来源：IDC 统计及预测，光大证券研究所

2、全球科技巨头大力布局 AI 产业

2.1、微软:投资 OpenAI, Copilot 开启 AI 商业化里程碑

2023 年 1 月中旬,微软对 OpenAI 投资 100 亿美元,占股 49%。微软与 OpenAI 的合作始于 2019 年 7 月,当时微软宣布向 OpenAI 提供 10 亿美元的投资,并与其建立独家云计算合作关系。此外,微软还将为 OpenAI 提供技术支持和咨询服务,帮助其实现“通用人工智能”的愿景。

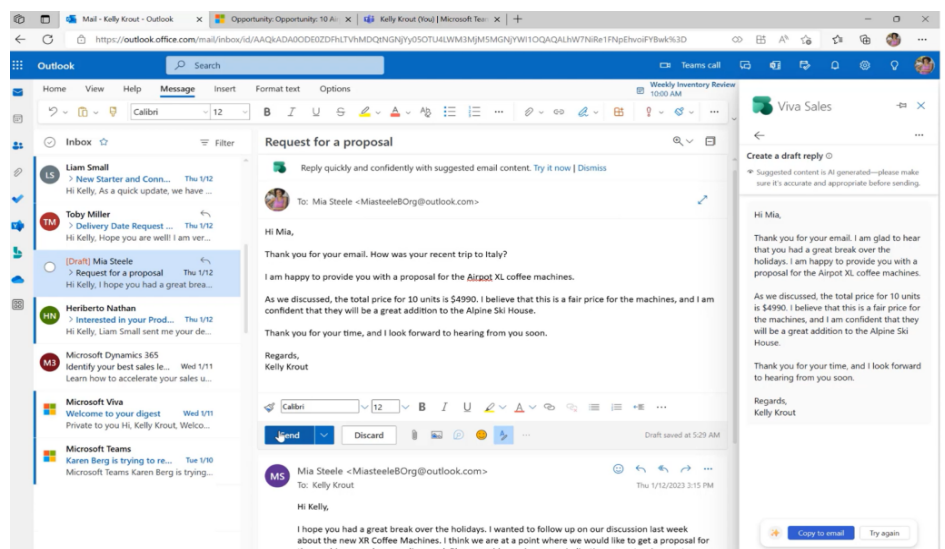
微软将 GPT-4 模型装进了 Office 套件,推出了全新的 Copilot 系统,发布 Dynamics 365 Copilot、Microsoft 365 Copilot、Power Platform Copilot 板块,开启 AI 商业化里程碑。

Dynamics 365 Copilot: GPT4 赋能业务全线

Dynamics 365 Copilot 赋能销售人员随时快速响应。在 Dynamics 365 Sales 和 Viva Sales 中, Copilot 可以编写给客户的电子邮件回复,在 Outlook 中自动生成一个 Teams 会议的总结邮件,从销售的 CRM 系统中自动提取产品、报价之类的细节,由 Teams 通话中总结出的核心要点一并汇总到邮件里。

Dynamics 365 Copilot 助力客服人员时刻对答如流。在 Dynamics 365 Customer Service 中,能够针对聊天对话和电子邮件中的问题,撰写出符合上下文语境的答案。

图 4: Copilot 自动生成邮件



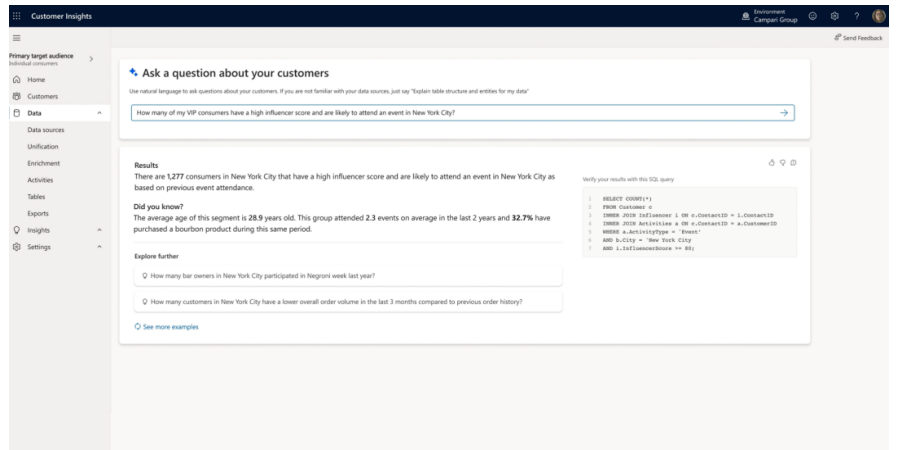
资料来源:微软官方公众号

Dynamics 365 Copilot 升级市场营销体验,轻松找到目标受众。在 Dynamics 365 Customer Insights 和 Dynamics 365 Marketing 中的 Copilot 使营销人员能够简化他们在数据探索、受众细分和内容创建方面的工作流程:

借助 Dynamics 365 Customer Insights 中的 Copilot,可以应对复杂的计算并且准确定位特定客户群体。

利用 Dynamics 365 Marketing 中的 Copilot,市场人员可以创建一个支持检索辅助功能的全新目标客户类别。

图 5: Dynamics 365 Customer Insights



资料来源: 微软官方公众号

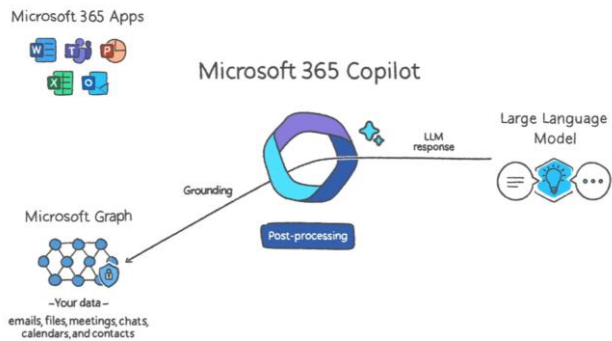
● Microsoft 365 Copilot: 开启全新工作方式

Copilot 是将大型语言模型 GPT-4 内建到 Microsoft 365。在程序里面可以用非常口语化的方式使用 AI，还可以通过各个程序的配合，达到预期效果。

Microsoft 365 Copilot 简化工作流程，释放创造力和生产力：

1. Copilot in Word 能够与用户一起撰写、编辑、总结和创作。
2. Copilot in PowerPoint 能够在创作过程中,通过自然语言命令将想法转化为设计好的演示文稿。
3. Copilot in Excel 能够帮助用户识别趋势,在短时间内创建专业型式的数据可视化。
4. Copilot in Outlook 能够帮助用户整合并管理收件箱。
5. Copilot in Teams 能够直接从对话上下文中提供实时摘要和待办事项。
6. Business Chat 汇集了来自文档、演示文稿、电子邮件、日历、笔记和联系人的数据,能够帮助用户总结聊天内容、撰写电子邮件、查找关键日期,根据其他项目文件制定计划。

图 6: Microsoft 365 Copilot 系统



资料来源: 微软官方公众号

● Power Platform Copilot: 低代码未来前景打开

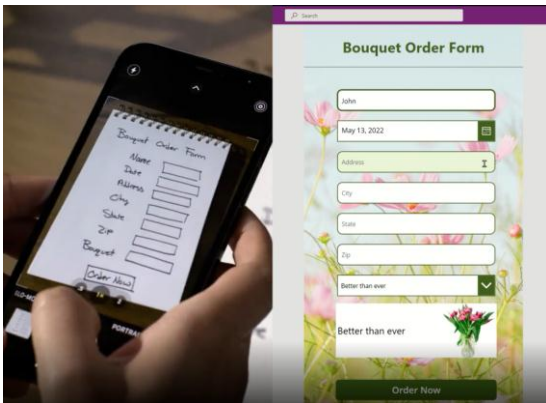
Copilot 是将 AI 结合到各种软件内的功能。Copilot 是 Microsoft Power Platform 的一个新功能，可以在 Power Apps, Power Virtual Agents 和 Power Automate 中基于 GPT 能力提供 AI-powered 的帮助，让制作者可以用自然语言描述他们想要的_{应用}、_{流程}或_{机器人}，然后 Copilot 可以快速完成创建，并提供改进的建议。

(1) Power Apps Copilot 可以让制作者通过自然语言对话的方式来描述并且直接生成想要的应用程序，可以通过和 Copilot 的对话来对应用进行调整，Copilot 也可以为用户提供应用的改进建议来解决问题。

(2) 通过 Power Automate Copilot，可以通过自然语言来生成比以前更复杂的流程，Copilot 还可以以对话的方式不断优化和迭代更新流程，使用户无需专业知识也可以构建想要的流程。

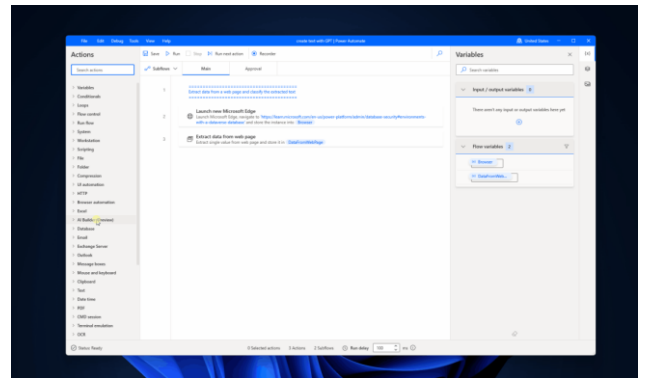
(3) 通过 Power Virtual Agents Copilot，可以直接使用自然语言来描述机器人流程，无需手动一个一个进行创建。

图 7: Power Apps



资料来源：微软官方公众号

图 8: Power Automate



资料来源：微软官方公众号

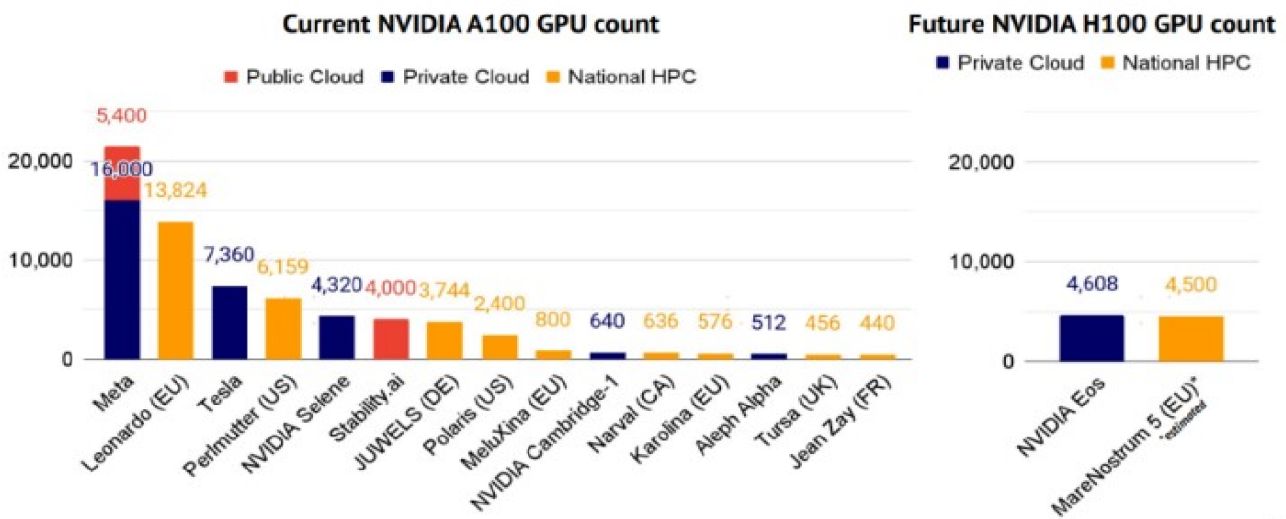
2.2、 英伟达:全球 GPU 龙头，产品需求量价齐升

全球 GPU 龙头开拓 AI 市场，助力服务器核心算力硬件性能提升。英伟达作为一家以 GPU 为主营业务的半导体设计公司，在 AI 行业发展初期就前瞻布局。2022 年公司数据中心产品营业收入 150.05 亿美元，同比增长 41.38%，2020-2022 CAGR 高达 49.70%，增势迅猛。ChatGPT 发布引爆市场、Google AI 对话模型 Bard 开放公测、百度发布“文心一言”大模型等一系列互联网巨头纷纷发力 AI 领域，催生 AI 服务器需求增长。据 TrendForce 预测，2023 年 AI 服务器出货量同比增速将达 37.7%，2024 年将增长 38%。公司有望受益于 AIGC 带来的核心算力硬件性能需求大幅提升，AI 服务器系统业务未来将成为公司数据中心产品营业收入的新增长点。

GPU 性能领先全球，H100 拓展市场领先地位。公司深耕 GPU 芯片多年，据英伟达在 IEEE 会议公布数据显示，从 2012 年的 K20X 到 2020 年的 A100，GPU 的推理性能提高了 317 倍，远超摩尔定律的发展速度。A100 作为当前 AI 服务器 GPU 中的高端产品，性能相比上一代 Volta GPU 提升 20 倍，其 AI 推理性能是 CPU 的 249 倍，打破 16 项性能 AI 记录，在 AI、数据分析、高性能计算上应用广泛，目前仍是 AI 领域所用的主力 GPU，被 AWS、微软 Azure 和百度云等。

算力需求迅猛增长，服务器 GPU 业务有望迎来量价齐升。 AIGC 产品与应用的逐步落地催生了巨大的算力需求，据 OpenAI 测算，自 2012 年以来，全球头部 AI 模型训练算力需求 3-4 个月提升一倍，每年头部训练模型所需算力增长幅度高达 10 倍。作为 AI 时代的算力核心，GPU 需求旺盛。同时，芯片性能的大幅提升带动 GPU 价格不断增长，英伟达 V100/A100/H100 三代 GPU 价格分别为 10000 /15000/36500 美元。作为服务器 GPU 领导者，英伟达将直接受益于 AIGC 大热带来的核心算力硬件性能需求提升，服务器 GPU 业务有望迎来量价齐升。

图 9：2022 年 A100 及未来 H100 各领域应用数量统计



资料来源：Stateof.ai

表 1：英伟达主要芯片产品

英伟达数据中心芯片情况	芯片型号	上市时间	售价 (美元)	售价 (元)	中国供应情况
中低端	A10	2021 年 4 月 13 日	3200	22080	供货
	A16	2021 年 4 月 13 日	3500	24150	供货
	A30	2021 年 4 月 13 日	4700	32430	供货
	A40	2020 年 10 月 5 日	5300	36570	供货
	L40	2021 年 9 月 21 日	7600	52440	供货
高性能	V100	2017 年 5 月 11 日	10000	69000	供货
	A800	2022 年 11 月 8 日	12000	82800	缺货
	A100	2020 年 5 月 15 日	15000	103500	美国政府禁止供应中国
	H100	2022 年 3 月 21 日	36500	251850	

资料来源：Thinkmate, 《财经》，光大证券研究所整理；汇率：1 美元=6.9 人民币

表 2：英伟达 AI 服务器主要产品对比

	NVIDIA DGX A100	NVIDIA DGX H100
GPU	8 个 NVIDIA A100 80GB Tensor Core GPU	8 个 NVIDIA H100 Tensor Core GPU
GPU 显存	共 640GB	共 640GB
性能	5 petaFLOPS AI 10 petaOPS INT8	32 petaFLOPS FP8
NVIDIA® NVSwitch™	6	4x
系统功耗	最大 6.5 kW	最高 10.2kW
CPU	双路 AMD Rome 7742、共 128 个核心、2.25 GHz（基准频率）、3.4 GHz（最大加速频率）	双路 x86
系统内存	2TB	2TB
网络	8 个单端口 NVIDIA ConnectX-7 200Gb/s 的 InfiniBand 端口 2 个双端口 NVIDIA ConnectX-7 VPI2 10/25/50/100/200Gb/s 以太网	8 个单端口 NVIDIA ConnectX-6 VPI4 400Gb/s 的 InfiniBand/ 以太网 2 个双端口 NVIDIA ConnectX-6 VPI2 10/25/50/100/200Gb/s 以太网 4 个 OSFP 端口，提供 8 个单端口 NVIDIA ConnectX-7 网卡 1 个 400Gb/s 的 InfiniBand/ 以太网网卡 1 个 200Gb/s 的 InfiniBand/ 以太网网卡
网络管理		RJ45 接口 10Gb/s 板载网卡 50Gb/s 的以太网可选网卡 RJ45 接口主机基板管理控制器（BMC） 2 个 NVIDIA BlueField-3 DPU BMC（均为 RJ45 接口）
存储	操作系统：2 个 1.92TB M.2 NVMe 驱动器 内部存储：30TB（8 个 3.84TB）U.2 NVMe 驱动器	操作系统：2 块 1.9TB NVMe M.2 硬盘 内部存储：8 块 3.84TB NVMe U.2 硬盘
系统软件	Ubuntu Linux 操作系统 同时支持：Red Hat 企业级 Linux CentOS 化。	DGX H100 系统预安装 DGX 操作系统，该操作系统基于 Ubuntu Linux，包含 DGX 软件堆栈（所有必要软件包和驱动均针对 DGX 优化）。 客户可以选择单独安装 Ubuntu Linux 或 Red Hat Enterprise Linux 以及必要的 DGX 软件堆栈。
运行温度范围	5°C 至 30°C	5°C 至 30°C

资料来源：英伟达官网，光大证券研究所整理

2.3、华为：服务器产业深耕多年，打造“鲲鹏+昇腾”双引擎战略布局

深耕服务器产业多年，打造“鲲鹏+昇腾”双引擎战略布局。华为从 2008 年开始对外提供服务器产品，目前形成了通用计算（鲲鹏）和 AI 计算（昇腾）两大业务布局。由于美国将华为列入实体清单，限制 x86 架构的芯片供给，华为在 2021 年 12 月出售了 x86 服务器业务。原本主营 x86 服务器业务的子公司超聚变分离后仍然稳定发展，2022 年国内市占率位列第三，达到 10.1%。目前华为致力于打造以“Arm 架构的鲲鹏+达芬奇架构的昇腾”为根基的双引擎战略布局，借助自研芯片优势，持续拓展服务器业务布局。

坚持计算产业战略，积极布局 AI 服务器领域。华为创新处理器架构“达芬奇”以匹配算力的增速，推出全场景处理器族，其中面向通用计算的鲲鹏系列与面向 AI 计算的昇腾系列主要应用于服务器产品，已服务政府、金融、运营商、电力、交通、医疗、教育、电力、油气、制造等领域，为全球合作伙伴开发服务器应用及解决方案。其中，采用鲲鹏 920/916 处理器的 TaiShan 服务器包括 2280E 边缘型、1280 高密型、2280 均衡型、2480 高性能型、5280 存储型和 X6000 高密型等产品，适配高性能计算、数据库、云计算等各种应用场景，现已应用于中科大千万亿次“瀚海 20 超级计算系统”。最高配置 2480 高端服务器基于鲲鹏 920 处理器，最高提供 256 核、3.0GHz 主频的计算能力和最多 25 个 SSD 硬盘。此外，昇腾系列 AI 服务器包括 Atlas 800 推理服务器与 Atlas 800 训练服务器，广泛应用于中心侧 AI 推理、深度学习模型开发和训练场景。Atlas 800 推理服务器搭载 64 核架构的鲲鹏 920 处理器，支持 8 张 Atlas 300I 推理卡，整机可提供 640 路高清视频实时分析（1080P 25FPS），Atlas 800 训练服务器达到 4U 高度，最高可提供 2.24 PFLOPS FP16 的算力，为 AI 推理、深度学习模型开发和训练提供超强算力支持。

表 3: 华为服务器产品

机型	通用计算服务器		AI 计算服务器	
	2480 高性能型	2480 高端型	Atlas 800 训练服务器	Atlas 800 推理服务器
形态	2U 机架服务器	2U 机架服务器	4U AI 服务器	2U AI 服务器
处理器型号	4 个鲲鹏 920 处理器	4 个鲲鹏 920 处理器	4 个鲲鹏 920 处理器+8 个昇腾 910 处理器	2 个鲲鹏 920, 最大支持 8 个 Atlas 300I 推理卡
内存插槽	32 个 DDR4-2933 DIMM 插槽	32 个 DDR4-3200 DIMM 插槽	最多 32 个 DDR4 内存插槽, 支持 RDIMM 内存速率最高 3200 MT/s 单根内存条容量支持 16GB/32GB/64GB	32 个 DDR4 内存插槽, 最高 3200 MT/s
本地存储	25*2.5 SAS/SATA ; 8*2.5 SAS/SATA	25*2.5 SAS/SATA ; 8*2.5 SAS/SATA	2 * 2.5 SAS/SATA+3 * 2.5 NVMe 2 * 2.5 SATA+3 * 2.5 NVMe 2 * 2.5 SAS/SATA+6 * 2.5 NVMe 2 * 2.5 SATA+6 * 2.5 NVMe 2 * 2.5 SATA+8 * 2.5 SAS/SATA	25*2.5 SAS/SATA 12*3.5 SAS/SATA 8*2.5 SAS/SATA+12x2.5 NVMe
RAID 支持	支持 RAID 0, 1, 5, 6, 10, 50, 60, 支持超级电容掉电保护	支持 RAID 0, 1, 5, 6, 10, 50, 60 支持超级电容掉电保护	支持 RAID 0, 1, 5, 6, 10, 50, 60	支持 RAID 0, 1, 5, 6, 10, 50, 60
板载网络	1 个灵活 IO 卡, 支持 4*10GE 光口或者 4*25GE 光口; 1 个板载 LOM, 光口或者 4*25GE 光口; 1 个板载支持 4*GE 电口	1 个板载网络插卡, 支持 4*10GE 光口或者 4*25GE 光口; 1 个板载 LOM, 支持 4*GE 电口	支持 8 * 100GE 光口+4* 25GE 光口 /2 * 100GE	
PCIe 扩展	最多 9 个 PCIe 4.0 插槽, 支持 3 个 PCIe 4.0 x16 和 6 个 PCIe x8 插槽	最多 4 个 PCIe 4.0 x8+2 个 PCIe 4.0 x16 标准插槽	最多支持 2 个 PCIe 4.0 扩展插槽	最多支持 9 个 PCIe4.0 PCIe 接口, 其中 1 个为 RAID 扣卡专用的 PCIe 扩展槽位, 另外 8 个为标准的 PCIe 扩展槽位
电源	2 个热插拔 900W 交流或 2000W 交流或 1200W 直流或 1500W 直流电源模块, 支持 1+1 冗余	2 个热插拔 2000W 交流电源模块, 支持 1+1 冗余	4 个热插拔 3 kW/2 kW 交流电源模块, 支持 2+2 冗余	2 个热插拔 900 W 或 2000 W 交流电源模块, 支持 1+1 冗余备份
供电	支持 100~240V AC, 240V DC, -48V DC	支持 220~240V AC, 240V DC	支持 200 ~ 240 V AC, 240 V DC	
风扇	支持 6 个热拔插风扇模组, 支持 N+1 冗余	支持 6 个热拔插风扇模组, 支持 N+1 冗余	支持 8 个热拔插风扇模组, 支持 N+1 冗余	4 个热拔插风扇, 支持 N+1 冗余备份
温度	5°C~40°C	5°C~35°C	工作温度 (风冷) : 5°C~35°C; 工作温度 (液冷) : 5°C~40°C	5°C~40°C
尺寸(宽 x 深 x 高)	447 mm*790 mm*86.1 mm	447 mm*790 mm*86.1 mm	175 mm * 447 mm * 790 mm	447 mm * 790 mm * 86.1 mm

资料来源: 华为官网, 光大证券研究所整理

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/227162010050006052>