2023 年 03 月 25 日

ChatGPT，英伟达 DGX 引爆 AI“核聚变”

计算机行业

英伟达加速计算助力云上业务场景

英伟达持续赋能加速计算 AI 潮流，在 GTC 会议主题演讲上，英伟达崭新全新的 AI 相关产品助力全球 AI 生态。基础软件，英伟达推出全新加速库；芯片方面，英伟达推出数据中心 Grace CPU，具备高能效、高运行速度等优势；服务器，英伟达推出 DGX 超级计算机；我们认为此次 GTC 大会的重中之重是英伟达推出全新 AI 服务平台，AI 的“Iphone”时刻已经来临，AI foundations 云服务助力初创企业具备拥有生成式 AI 的能力，且已经具备多种生成式 AI 模型和相应案例。**我们认为以英伟达为例，提供全栈 AI 云解决方案，全球底层算力储备充分，AI 赋能千行百业的“燃油”已经充足。**

大模型厂商具备先发优势

我们认为 OpenAI 和百度可以率先发布大模型的根本原因是两者算力平台储备丰厚；**大模型具备先发优势**，大模型或是人工智能的本质是算法的不断迭代学习，而人类反馈强化奖励机制有助于机器更好的理解人类的语境语义，在此基础上，率先发布的算法模型厂商具备先机，以 GPT-4 为例，**GPT-4 的强大是模型本身的“不断进化”**，我们认为 GPT-4 的成功发布与 GPT3.5 的自身算法迭代和用户测评反馈密不可分。百度同样具备大模型的先发优势，已经抢占相应先机，有望通过迭代方式迭代出更加出色的人工智能大模型，从而赋能千行百业。

AI“大爆炸”时代已来临

目前应用方面已有诸多案例，例如 OpenAI 已经接入众多应用端，例如 New Bing、办公软件，微软 Office 365 copilot、工业软件，微软 Dynamics 365 Copilot。**AI 加速赋能应用的时代已经到来**，算力与基础设施的储备已经完善，AI“燃油”已经充足；大模型储备发面，以 Open AI 和百度为例，相关的大模型储备已初具能力，我们判断 AI 加速赋能应用的时代已经到来，以微软相关产品为例，AI 赋能应用的“引线”已经点燃，我们认为率先通过 AI 赋能的应用产品具备“流量”的先发优势，无论是客户数量还是功能体验都会具备优势，**在原有市场中的竞争壁垒更加坚厚，成为解放生产力的 AI 助手。**

投资建议：

我们认为 AIGC 的出世会产生革命性的影响，同时有望赋能千行百业。我们梳理了三条路径图，积极的推荐以下三条投资主线：

评级及分析师信息

行业评级：推荐

行业走势图

分析师：刘泽晶
邮箱：liuzj1@hx168.com.cn
SAC NO：S1120520020002
联系电话：

1）具备算力基础的厂商，受益标的为寒武纪、海光信息、浪潮信息、中科曙光、景嘉微、龙芯中科、神州数码、拓维信息；

2）具备 AI 算法商业落地的厂商，重点推荐科大讯飞、拓尔思，其他受益标的为：海天瑞声；

3）AIGC 相关技术储备的应用厂商，受益标的为：百度、同花顺、三六零、金山办公。

风险提示

核心技术水平升级不及预期的风险；AI 伦理风险；政策推进不及预期的风险；中美贸易摩擦升级的风险。

正文目录

图目录

1.ChatGPT，英伟达 DGX 引爆 AI“核聚变”

1.1. 英伟达加速计算助力云上业务场景

英伟达持续赋能加速计算 AI 潮流：2023 年 3 月 23 日，英伟达 GTC 会议主题演讲开启，是一场全球的科技盛宴，本次大会的宗旨是告诉全球，加速计算是可以实现的。 我们认为其中最重磅的消息，是英伟达展示全新的芯片和系统、加速库、云服务、AI 服务以及助力以助力全球 AI 生态，**我们认为此次 GTC 大会实则为一场全球 AI 盛宴。**在本届 GTC 技术大会上，NVIDIA 创始人兼 CEO 黄仁勋在大会主题演讲中分享了 NVIDIA 加速计算平台如何推动人工智能、元宇宙、云技术、可持续计算的下一波浪潮。

图表 1 英伟达黄仁勋与 Open AI 创始人 Jensen Huang 讨论人工智能

资料来源：英伟达官网，华西证券研究所

基础软件方面，英伟达推出全新加速库：加速库着力解决的是普通计算机无法解决的问题，加速计算并非易事，它需要从芯片、系统、网络、加速库到重构应用的全栈发明。从图形、成像、粒子或流体动力学、量子物理学到数据处理和机器学习，每个经过优化的堆栈都会加速对应应用领域。目前此加速库，已经应用在多个领域，持续开辟 AI 加速的新市场。例如汽车航空空气动力学仿真、量子电路仿真、推荐系统、数字人、优化物流服务、视频处理、基因计算、芯片制造、等方面。

图表 2 英伟达计算加速库

资料来源：英伟达官网，华西证券研究所

硬件架构方面，英伟达推出数据中心 CPU,Grace：加速云数据中心的 CPU 侧重点与过去有着根本性的不同。在 AI 和云服务中，加速计算卸载可并行的工作负载，而 CPU 可处理其他工作负载，比如 Web RPC 和数据库查询。因此 Grace CPU 脱颖而出，其中包含了 72 个 Arm 核心，由超高速片内可扩展的、缓存一致的网络连接，可提供 3.2 TB/s 的截面带宽，内存系统与 LPDDR 低功耗内存购成，这是带有被动冷却功能的计算模组，**与以往 CPU 不同，此款 CPU 可以在云数据中心规模下实现高能效，非常适合云计算应用和科学计算应用。此外，在微服务方面，Grace 的速度比最新一代 x86 CPU 的平均速度快 1.3 倍，而在数据处理中则快 1.2 倍。**

图表 3 英伟达 Grace GPU

资料来源：英伟达官网，华西证券研究所

服务器方面，英伟达推出 DGX(超级计算机)：DGX 配有 8 个 H100 GPU 模组，H100 配有 Transformer 引擎，旨在处理类似令人惊叹的 ChatGPT 模型，8 个 H100 模组通过 NVLINK Switch 彼此相连，以实现全面无阻塞通信。8 个 H100 协同工作，类似一个巨型 GPU。此外，根据介绍，NVIDIA DGX H100 是全球客户构建 AI 基础设施的蓝图，现在已全面投入生产。此款服务器是优势在于：

1、DGX H100 是一个完全集成的硬件和软件解决方案，它包括 NVIDIA Base Command 和 NVIDIA AI Enterprise 软件套件。

2、此款服务器包含极高的新能、网络速度和拓展性，**其架构为生成式 AI、自然语言处理和深度学习推荐模型等最大的工作负载提供了强大的支持。**

图表 4 英伟达 DGXH100

资料来源：英伟达官网，华西证券研究所

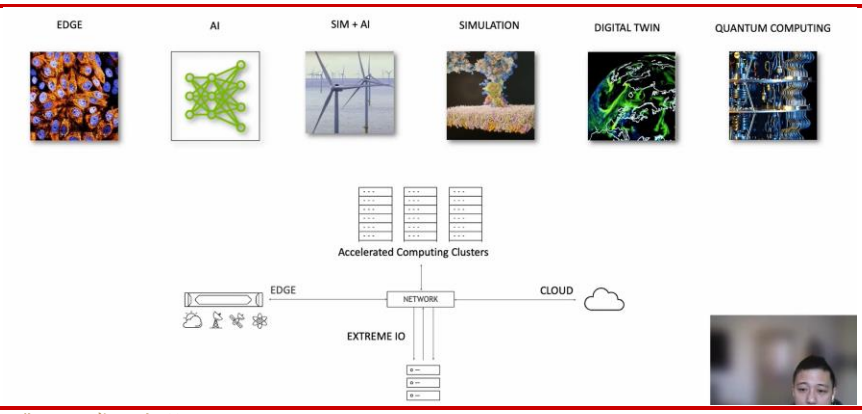
图表 5 英伟达 DGX H100 相关参数

显卡	8 个 NVIDIA H100 Tensor Core GPU
显存	总计 640GB
表现	32 petaFLOPS FP8
NVIDIA® NVSwitch™	4倍
系统电源使用	~11.3kW 最大值
中央处理器	双 56 核第 4 代英特尔® 至强® 可扩展处理器
系统内存	2TB
联网	4x OSFP 端口服务于 8x 单端口 NVIDIA ConnectX-7 VPI - 400Gb/s InfiniBand/以太网 2x 双端口 NVIDIA ConnectX-7 VPI - 1x 400Gb/s InfiniBand - 1x 200Gb/s 以太网
贮存	操作系统：2 个 1.9TB M.2 NVME 驱动器 内部存储：8 个 3.84TB NVMe U.2 驱动器
管理网络	带 RJ45的 10Gb/s 板载 NIC 带 RJ45 的 50Gb/s 以太网可选 NIC 主机 BMC

资料来源：英伟达官网，华西证券研究所

我们认为此次 GTC 大会的重中之重是英伟达推出全新 AI 服务平台：我们已经在《ChatGPT，百度文心一言畅想》中论证，**平台实为模型和算力之间的“桥梁”**，是 AIGC 或大模型生成的必备要素，不论是数据库还是编译器，都需要通过平台来实现资源的合理分配以达到软硬件的最优组合，从而大幅提升模型效率。平台通过调用数据包来适配软硬件之间的结构，来达到模型的最优组合，从而提升模型乃至整个虚拟机的效率。**而英伟达此次的推出的 DGX 云与生成式 AI 服务恰恰印证了我们的观点。**

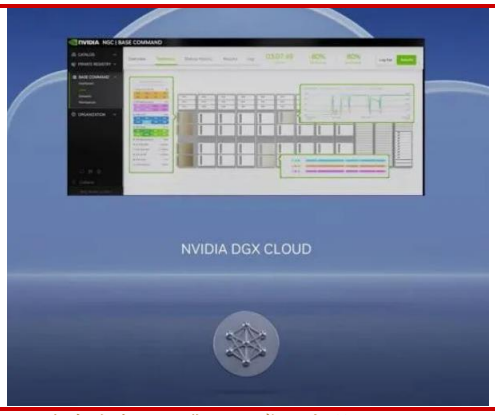
图表 6 英伟达 AI 加速计算上云架构示意图



资料来源：英伟达官网，华西证券研究所

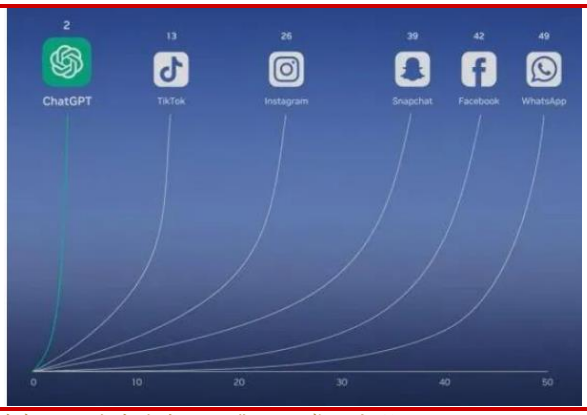
AI 的“Iphone”时刻已经来临：此次发布会上，英伟达 CEO 黄仁勋表示我们正处于 AI 的“iPhone 时刻。初创公司竞相构建具有颠覆性的产品和商业模式，而老牌公司则在寻求应对之法。生成式 AI 引发了全球企业制定 AI 战略的紧迫感。英伟达推出英伟达 DGX Cloud，并与 Microsoft Azure、Google GCP 和 Oracle OCI 合作，为客户提供出色的 NVIDIAAI 以及全球主要的云服务提供商。此外，Oracle Cloud Infrastructure(OCI)将成为首个 NVIDIA DGX Cloud，是一种先进的 DGX AI 超级计算机。

图表 7 英伟达 DRX 云示意图



资料来源：英伟达官网，华西证券研究所

图表 8 英伟达 DRX 云应用示意图



资料来源：英伟达官网，华西证券研究所

英伟达推出英伟达 AI foundations 云服务：此外，黄仁勋表示生成式 AI 将重塑所有行业，生成式 AI 是一种新型计算机，一种我们可以用人类语言进行编程的计算机。这种能力影响深远，每个人都可以命令计算机来解决问题，而之前这是只有计算机程序员才能接触的领域。现在每个人都可以是程序员。生成式 AI 是一种新型计算平台，与 PC、互联网、移动设备和云类似。与之前的计算时代类似，先行者正在打造新的应用，并成立新公司，以利用生成式 AI 的自动化和协同创作能力。此外，英伟达将通过中国云服务商提供 AI 超算能力，中国初创公司也能开发大模型，中国可用的英伟达芯片包括 Ampere 和 Hopper 芯片是 A800 和 H800。

英伟达打造 Aifoundations 云服务，企业可以通过在 NVIDIA DGX Cloud 上的 NVIDIA NeMo 服务快速采用生成式 AI，通过此种云服务能够构建、改进和操作定制的大型语言模型和生成式 AI 模型，目前已知的服务包括：

1、NeMo 服务让企业快速定制基础语言模型；

2、NVIDIA Picasso 服务加速跨图像、视频和 3D 的模拟和创意设计；

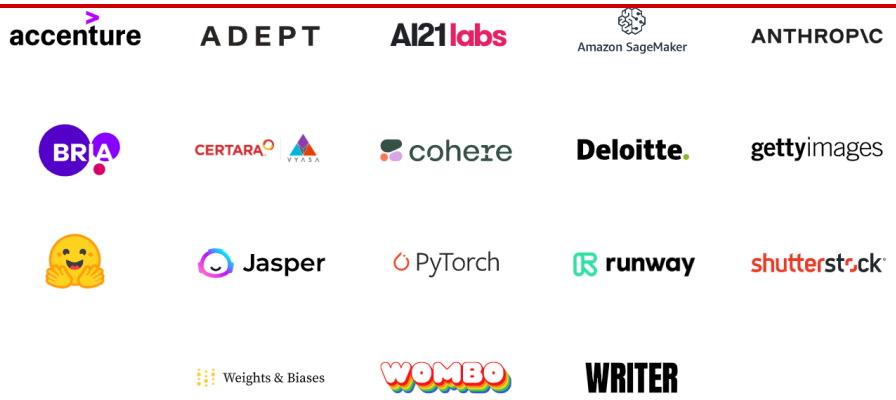
3、NVIDIA BioNeMo ™ 生物学云服务等。

图表 9 英伟达 AI Foundation 能力示意图

资料来源：英伟达官网，华西证券研究所

英伟达 AI foundation 功能强大：目前已经包括大模型训练更加智能，推进生物研究蛋白质工程构建 AI 模型、AT&T 通过英伟达支持智能扬声器和客户呼叫中心、AI 生成式语言应用等典型应用案例。目前已有合作伙伴例如埃森哲、德勤、Pytorch 等。

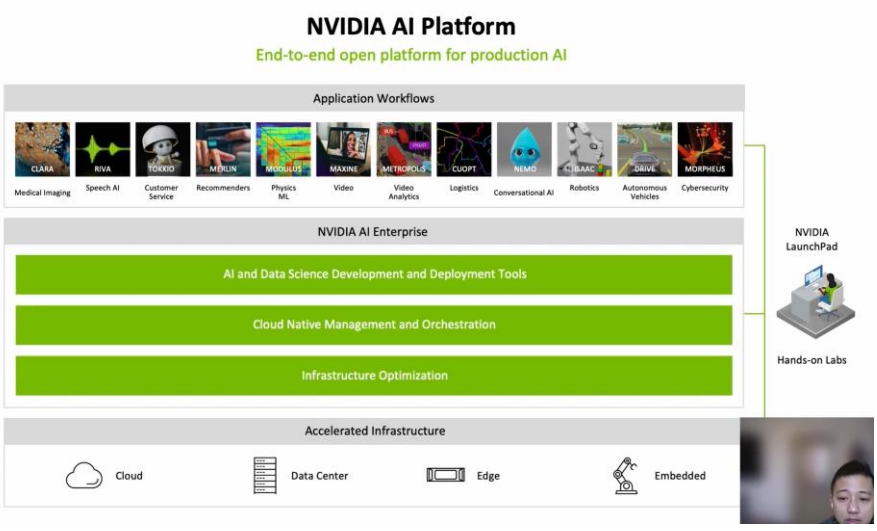
图表 10 英伟达 AI Foundation 合作伙伴示意图



资料来源：英伟达官网，华西证券研究所

英伟达提供全栈 AI 云解决方案，全球 AI+时代已经：英伟达提供的全站云服务包括基础设施、调度、加速库、加速框架、应用框架全套解决方案，上层应用包括自然语言处理、机器视觉、智能驾驶、机器人、生物计算等全行业的应用框架大模型解决方案。我们认为在英伟达的赋能时代下，全球算力储备，加速库与大模型储备充足，在 AI 云服务的赋能下，所有科技企业有望具备生成式 AI 的能力，AI 底层平台框架基本完善，**AI 赋能千行百业的“燃油”已经充足，即将迎来“AI 大爆炸”的全新时代。**

图表 11 英伟达 AI Platform 架构示意图



资料来源：英伟达官网，华西证券研究所

1.2. 大模型具备先发优势，迎接“AI 大爆炸”时代

算力平台储备充分：无论是 Open AI(微软)还是百度在底层“燃料”（算力）部分储备皆已充分。

OpenAI 方面，在 ChatGPT 举世闻名之前，微软“陪跑漫漫长夜”：根据 36 氪和新浪的消息，早在 2019 年，微软斥资几亿美元为 Open AI 的训练打造一台超级服务器，其中包括上万张英伟达 A100，旨意为 ChatGPT 和 New Bing 提供算力基础，

同时，微软还在 Azure 的 60 多个数据中心部署了几十万张 GPU，用于 ChatGPT 的推理。价格上，英伟达 A100 价格约为 15000 美元/张，以一万张为例，AI 芯片造价已经近 1.5 亿美元。此外，根据 36 氪的消息，微软正在打造下一代超级计算机来为为生成式 AI 进步奠定基础。

图表 12 微软超级计算机示意图

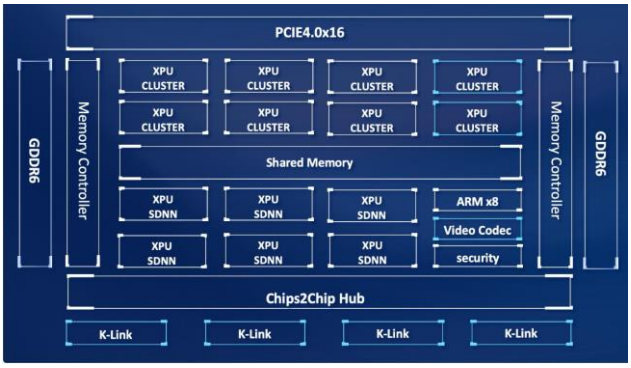
资料来源：CSDN，华西证券研究所

图表 13 微软数据中心示意图

资料来源：CSDN，华西证券研究所

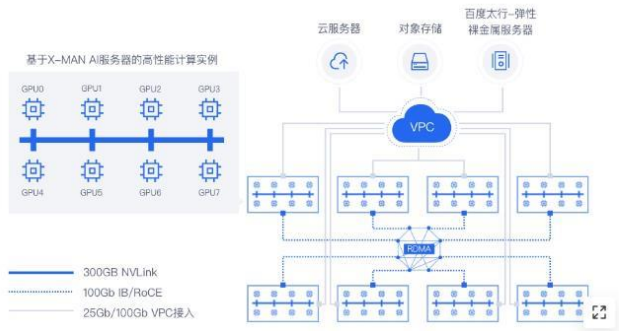
百度方面，百度底层算力技术实力强劲：百度具备建设百度智算中心、百度 AI 异构计算平台的能力，此外自身具备自研的昆仑芯 AI 加速器与昆仑芯 AI 加速卡，具备多重领先优势。**我们认为算力储备是大模型研发的基础，而 Open AI 和百度已经具备相应能力，这也是百度和 Open AI 率先发布大模型的根本原因。**

图表 14 百度昆仑芯云服务器



资料来源：百度智能云官网，华西证券研究所

图表 15 百度百舸 AI 异构计算平台 AI 计算示意图



资料来源：百度智能云官网，华西证券研究所

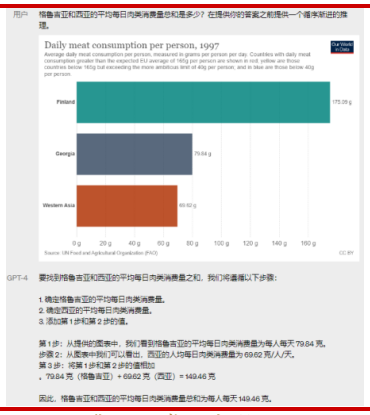
大模型具备先发优势：ChatGPT 的核心是 RLHF（人类反馈的强化模型），这一种独特的奖励机制，我们可以理解成模型可以通过人们的回馈不断强化自身学习，最终得到用户想要呈现的效果，无论是谷歌的 Lamda 还是百度的文心一言皆具有类似的此功能，**我们认为这也是 Open AI 与百度率先发布自身大模型的本质原因。**

GPT-4 的强大是模型本身的“不断进化”：相较于 ChatGPT3.5 而言，**1、模型更加强大，更可靠、更有创意、且更能够理解细微的指令，**例如在各种专业考试中取得优异成绩；**2、文章理解、语义理解能力更强，**例如总结和加工文章，GPT-4 相较于 GPT3.5 更有优势；3、**GPT-4 不仅仅是语言模型，而是多模态融合模型，**GPT-4 可以具备“人类思维”并理解“图片的内容”并进行反馈，比如考试题推理、纸质摘要、图像分析、图像描述等功能。**4、具备极强的复杂推理机制；5、多种语言方面均表现出优越性；我们认为 GPT-4 的成功发布与 GPT3.5 的自身算法迭代和用户测评反馈密不可分。**

图表 16 GPT-4 视觉输入示意图

资料来源：Open AI，华西证券研究所

图表 17 GPT-4 图标推理示意图



资料来源：Open AI，华西证券研究所

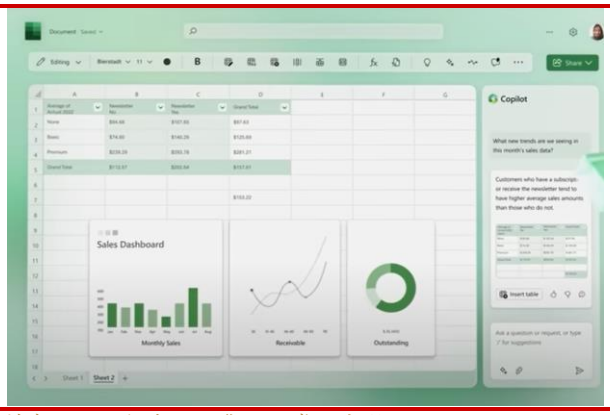
同理，百度同样具备算法的先发优势：我们认为我国 ChatGPT“领头羊”同样具备大模型的领先优势，大模型或是人工智能的本质是算法的不断迭代学习，而人类反馈强化奖励机制有助于机器更好的理解人类的语境语义，在此基础上，率先发布的算法模型厂商具备先机，通过自身的生态和用户数量可以更好的赋能自身模型。总结而言，**率先发布大模型应用的厂商抢占了先机，有望迭代出更加出色的人工智能大模型，更有可能在 AI 的浪潮中率先赋能千行百业。**

图表 18 百度文心一言全景图

资料来源：IDC，华西证券研究所

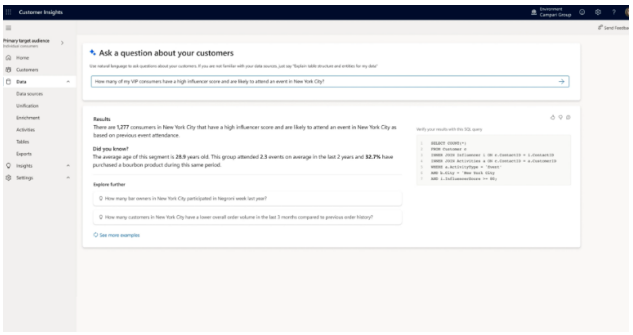
目前应用方面已有诸多案例:OpenAI 已经接入众多应用端，例如 1、New Bing，搜索引擎搜索更便捷、沟通更高效、功能更多元，且 AI 答案的可靠性已得到提升，已经加入 AI 聊天、相关写作、AI 作图等功能；**2、办公软件，微软 Office 365 copilot**，实为解放生产力的 AI 助手，办公软件具备自动编辑、数据可视化分析、提供会议安排、展史精美 PPT 等功能；**3、工业软件，微软 Dynamics 365 Copilot，**AI 融入 CRM 与 ERP，ChatGPT 已经融入商业中心、客服服务、供应链管理、智能订单管理等相关应用。

图表 19 EXCEL Copilot 提供数据分析示意图



资料来源：微软官网，华西证券研究所

图表 20 Coplilot 助力 Marking 精准找到目标受众



资料来源：微软官网，华西证券研究所

AI 加速赋能应用的时代已经到来：全球方面，以科技巨头英伟达为例，**算力与基础设施的储备已经完善，AI“燃油”已经充足**，大模型储备发面，以 Open AI 和百度为例，相关的大模型储备已初具能力，率先发布模型的厂商具备先发优势，有望率先通过反馈奖励模型迭代升级，从而更精准的赋能千行百业，应用方面，随着平台、算力、算法、数据等关键要素储备完成，**我们判断 AI 加速赋能应用的时代已经到来**，以微软相关产品为例，AI 赋能应用的“引线”已经点燃，我们认为率先通过 AI 赋能的应用产品具备“流量”的先发优势，无论是客户数量还是功能体验都会具备优势，**在原有市场中的竞争壁垒更加坚厚，成为解放生产力的 AI 助手。**

2. 投资建议：梳理 AIGC 相关受益厂商

我们认为 AIGC 的出世会产生革命性的影响，同时有望赋能千行百业。我们梳理了三条路径图，积极的推荐以下三条投资主线：

1）具备算力基础的厂商，受益标的为寒武纪、海光信息、浪潮信息、中科曙光、景嘉微、龙芯中科、神州数码、拓维信息；

2）具备 AI 算法商业落地的厂商，重点推荐科大讯飞、拓尔思，其他受益标的为：海天瑞声；

3）AIGC 相关技术储备的应用厂商，受益标的为：百度、同花顺、三六零、金山办公。

3. 风险提示

1、核心技术水平升级不及预期的风险；

2、AI 伦理风险；

3、政策推进不及预期的风险；

4、中美贸易摩擦升级的风险。

分析师与研究助理简介

刘泽晶（首席分析师）： 2014-2015年新财富计算机行业团队第三、第五名，水晶球第三名，10年证券从业经验。

分析师承诺

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，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，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

评级说明

公司评级标准	投资评级	说明
以报告发布日后的6个月内公司股价相对上证指数的涨跌幅为基准。	买入	分析师预测在此期间股价相对强于上证指数达到或超过15%
	增持	分析师预测在此期间股价相对强于上证指数在5%—15%之间
	中性	分析师预测在此期间股价相对上证指数在-5%—5%之间
	减持	分析师预测在此期间股价相对弱于上证指数5%—15%之间
	卖出	分析师预测在此期间股价相对弱于上证指数达到或超过15%
行业评级标准		
以报告发布日后的6个月内行业指数的涨跌幅为基准。	推荐	分析师预测在此期间行业指数相对强于上证指数达到或超过10%
	中性	分析师预测在此期间行业指数相对上证指数在-10%—10%之间
	回避	分析师预测在此期间行业指数相对弱于上证指数达到或超过10%

华西证券研究所：

地址：北京市西城区太平桥大街丰汇园11号丰汇时代大厦南座5层
网址：http://www.hx168.com.cn/hxzq/hxindex.html

华西证券免责声明

华西证券股份有限公司（以下简称“本公司”）具备证券投资咨询业务资格。本报告仅供本公司签约客户使用。本公司不会因接收人收到或者经由其他渠道转发收到本报告而直接视其为本公司客户。

本报告基于本公司研究所及其研究人员认为的已经公开的资料或者研究人员的实地调研资料，但本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载资料、意见以及推测仅于本报告发布当日的判断，且这种判断受到研究方法、研究依据等多方面的制约。在不同时期，本公司可发出与本报告所载资料、意见及预测不一致的报告。本公司不保证本报告所含信息始终保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者需自行关注相应更新或修改。

在任何情况下，本报告仅提供给签约客户参考使用，任何信息或所表述的意见绝不构成对任何人的投资建议。市场有风险，投资需谨慎。投资者不应将本报告视为做出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。在任何情况下，本报告均未考虑到个别客户的特殊投资目标、财务状况或需求，不能作为客户进行客户买卖、认购证券或者其他金融工具的保证或邀请。在任何情况下，本公司、本公司员工或者其他关联方均不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告而导致的任何可能损失负有任何责任。投资者因使用本公司研究报告做出的任何投资决策均是独立行为，与本公司、本公司员工及其他关联方无关。

本公司建立起信息隔离墙制度、跨墙制度来规范管理跨部门、跨关联机构之间的信息流动。务请投资者注意，在法律许可的前提下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的前提下，本公司的董事、高级职员或员工可能担任本报告所提到的公司的董事。