



计算机

优于大市（首次）

证券分析师

钱劲宇

资格编号：S0120523090002

邮箱：qianjy5@tebon.com.cn

相关研究

微软继续引领全球 AI 浪潮

微软 2023 Ignite 大会纪录

投资要点：

- **亮点一：AI 基础设施的革新。**

微软与 OpenAI 合作，集成 ChatGPT 可以用于搜索、协作、工作和学习工具。推出了 AI 加速器芯片 Azure Maia 和基于 Arm 架构的云原生芯片 Azure Cobalt。Azure Boost 即将正式商用，可将存储和网络进程从主机服务器迁移到专用硬件和软件上，从而提高存储和网络速度。

- **亮点二：扩展微软 Copilot 体验。**

通过更新可以在 Windows 和 Bing 中访问 M365 Copilot; Bing chat 和 Bing Chat Enterprise 更新为 Copilot。

Microsoft Copilot Studio（自己构建 Copilot）：一款低代码工具，用于根据自己的场景构建独立的 copilot，除此之外，微软融合 gpts 进来，可以与 copilot 配合使用，可以构建、测试、发布自定义的 copilot 和 gpt。Copilot Studio 现已推出，与 Copilot for Microsoft 365 的集成现已推出预览版。

- **亮点三：Windows 与 AI 的深度融合。**

2024 年初，Copilot for Microsoft 365 将集成到 Windows 桌面中，允许用户在 Microsoft 365 Graph 内容和上下文中进行 prompts。

推出 window AI studio，集合了 Azure AI Studio 和 Hugging Face 的尖端 AI 开发工具和模型，使得开发人员能够微调、自定义、部署 SLM 在 Windows 本地进行使用。开发人员可以使用工作区中集成的 Prompt Flow 和 Gradio 模板快速测试其微调模型。

英伟达即将更新 TensorRT-LLM for Windows，TensorRT-LLM for Windows 即将通过全新封装接口与 OpenAI 的聊天 API 兼容，这将使数以百计的开发者项目和应用能在 RTX PC 的本地运行，而非云端运行，因此用户可以在 PC 上保留私人数据和专有数据。

报告内容来自大会直播实时翻译，更多详细内容请参考：

1) 官方新闻汇集渠道 <https://news.microsoft.com/ignite-2023-book-of-news/>

2) 开题演讲片段 <https://news.microsoft.com/ignite-2023/>

- **风险提示：** AI 技术落地不及预期、市场需求不及预期、全球供应链风险加剧

附录：

1. 微软 Day 1 keynote 会议内容

最新数据显示，Copilot 已经推动了生产力的真正提升，使用 Copilot 的人在寻找信息上花费的时间更少，他们举行的会议更有效，他们能够在工作文档上进行协作，不管是 Word 文档、电子表格、电子邮件。在本周的 Ignite 上，微软将推出 100 多项新更新，涵盖整个技术栈的每一个层面。微软的端到端 Copilot 技术栈覆盖了基础设施、基础模型、数据工具链，Copilot 等。

1.1. AI 基础设施

微软已经以突破性进展引领了产业发展：例如，与 OpenAI 建立合作伙伴关系，并将 ChatGPT 功能集成于搜索、协作、工作和学习工具中。随着在 AI 领域的进一步发展，微软正在重新构想云基础设施，以确保每一层硬件和软件堆栈都获得优化。

微软的数据中心网络已经是世界上最先进和最广泛的网络之一。为了满足 AI 和未来工作负载的需求，微软推出了 Hollow core fiber 技术，利用空气作为光纤的导光介质，传输速度提升了 47%，在英国数据中心最先使用（去年底微软收购这条赛道的领跑者英国 Lumenisity）。

微软与 Nvidia 的深入合作。微软利用 Nvidia 的 GPU 在云中构建了 AI 超级计算基础设施，OpenAI 利用这一基础设施交付了大型语言模型。上周，亚洲地区提交给 MLperf 基准测试联盟的数据量最大，使用了 10000 个 H100 GPU，是之前记录的三倍，性能超过任何其他云平台。在最新的世界超级计算机前 500 名中，Azure 是公共云中最强大的超级计算机，总体排名第三。

拓展与芯片供应商的合作伙伴关系，为客户提供基础设施选项：

(1) 将在微软智能云 Azure 中添加 AMD MI300X 加速虚拟机 (VMs)。ND MI300 虚拟机 (VMs) 将采用 AMD 最新的 GPU——AMD Instinct MI300X，旨在加速 AI 工作负载的处理，以实现高范围的 AI 模型训练和生成式推理。

(2) 开放预览为英伟达 H100 Tensor Core GPU 打造的全新 NC H100 v5 虚拟机系列，将提高大中型 AI 训练和生成式推理的性能、可靠性和效率。微软还发布了 ND H200 v5 虚拟机系列计划，这是一款 AI 优化虚拟机，配备了即将推出的英伟达 H200 Tensor Core GPU。

(3) OpenAI 的 GPT-4 已经在 MI300X 上运行了。

发布了微软设计的两款芯片：

(1) Microsoft Azure Cobalt：是一款基于 Arm 架构的云原生芯片，针对通用工作负载的性能、功率和成本效益进行了优化。64 bit 128 core。为微软团队、Azure 通信服务以及 Azure SQL 提供动力。明年微软将向客户提供这款产品。

(2) Microsoft Azure Maia: 是一款微软自研 AI 加速器芯片, 用于 OpenAI 模型、Bing、GitHub Copilot 和 ChatGPT 等 AI 工作负载运行云端训练和推理。采用 5nm 制程, 晶体管数量达到 1050 亿个。(微软还给 Maia 重新设计了一套散热系统, 目前主要用于处理微软内部算力需求)

Azure Boost 将正式商用。该系统可将存储和网络进程从主机服务器迁移到专用硬件和软件上, 从而提高存储和网络速度, 使 Azure 成为高性能工作负载的最佳云平台, 同时也增强了安全性。

1.2. 基础模型服务

微软正在将最新的 GPT-4 引入 Azure OpenAI 服务, 包括具有视觉功能的 GPT-4 Turbo。GPT-4 Turbo 提供更低的定价、结构化的 JSON 格式, 以及扩展提示等功能, 现在可以在单个提示中放入 300 页的文本。GPT-4 Turbo 将在 Azure OpenAI 服务中提供, 新模型的 tokens 定价将与 OpenAI 持平。不久后, 还将能够将 GPT-4 Turbo 与 Azure AI Vision 连接起来, 允许使用视频图像和文本进行提示。

微软还将在 Azure OpenAI 服务中引入 GPT-4 的微调功能, 允许使用自己的数据创建 GPT-4 的定制版本。

model catalog 帮助继续拓展生成式 AI 模型的选择和灵活性, 为开发者提供更加全面的开放模型选择。模型即服务 (Model-as-a-Service) 是微软在 Build 开发者大会上发布的模型目录中的新功能, 让专业开发者将能够轻松将 Meta Llama 2、Mistral 即将推出的高级模型以及 G42 Jais 等最新的 AI 模型作为 API 端点集成到他们的应用当中, 还可以用自己的数据定制这些模型, 无需担心 GPU 基础架构的设置和管理, 消除复杂性。

Azure AI Studio 预览版推出, 借助 Azure AI Studio, 可以利用自己的数据构建、训练自己的 “copilot”, 或者使用自己数据来构建其他模型。

Azure AI Search 的功能更新: Vector Search 现已正式商用, 企业可以在其生成式 AI 应用中为每个用户带来高精度的体验。

新的 GPT-3.5 Turbo 模型 (支持 16K token 提示词长度) 将正式商用, GPT-4 Turbo 将于 2023 年 11 月底在 Azure OpenAI 服务中公开预览。GPT-4 Turbo with Vision 即将推出预览版, DALL-E 3 现已在 Azure OpenAI 服务中公开预览, 与 GPT-4 共同推动下一代企业解决方案的发展, 使企业能够利用图像实现高级功能。与 Azure AI 视觉服务一起使用时, GPT-4 Turbo with Vision 可以理解视频以生成文本输出。

微软发布了自己训练的最新开源预训练模型 Phi-2, 参数达到 27 亿, 在数理、逻辑等方面均显著好于上一代产品。

英伟达和微软同台交流, 发布最近合作内容:

(1) 英伟达将发布基于微软智能云 Azure 的 AI 工坊服务 (AI foundry service), 旨在帮助企业和初创公司在 Azure 上加速开发、调整和部署自己的定制 AI 模型。英伟达 AI 工坊服务 (NVIDIA AI foundry service) 汇集了英伟达 AI 基础

模型系列、英伟达 NeMo 框架和工具，以及英伟达 DGX Cloud AI 超级计算和服务这三大要素，为企业构建自定义生成式 AI 模型提供了端到端解决方案。企业可以在 Azure 上利用英伟达 AI Enterprise 软件部署其模型，为生成式 AI 应用——如智能搜索、总结和内容生成等，提供助力。

(2) Nvidia Omniverse 可以在 Azure 云上使用，提供 Omniverse Cloud 服务，根据英伟达官方信息，该一站式服务能够使全球汽车制造商统一其核心产品与业务流程的数字化，从而提高生产速度和运营效率，缩短产品上市时间。

(3) 英伟达与微软一起发布 DirectML 增强功能，用来赋能使用 Llama2 作为基础模型的开发者。

(4) TensorRT-LLM 即将更新，未来会支持更多的大语言模型，在 RTX GPU 8GB 及以上显存的 PC 和笔记本电脑上使要求严苛的 AI 工作负载更容易完成。TensorRT-LLM for Windows 即将通过全新封装接口与 OpenAI 的聊天 API 兼容，这将使数以百计的开发者项目和应用能在 RTX PC 的本地运行，而非云端运行，因此用户可以在 PC 上保留私人 and 专有数据。

1.3. 扩展 Copilot M365 体验

(1) Microsoft Copilot for Microsoft 365: 本月，Copilot for Microsoft 365 正式面向企业开放商用。目前，Visa、BP 集团、本田和辉瑞等客户，以及埃森哲、安永、毕马威、勤达睿和普华永道等合作伙伴都在使用 Copilot。微软将根据从“早期使用计划”(Early Access Program)和其他研究渠道中获得的经验，继续提供新价值。本次会议发布了 PowerPoint、Excel、Microsoft Viva 等套件更新。

- M365 copilot: 将首先在 word、powerpoint 推出个性化功能(可以根据自己的偏好调整风格、语气的偏好等)，之后会在 Microsoft 365 其他的应用中推出。除此之外，word 即将更新利用 copilot 追踪和注释的功能；PowerPoint 即将推出图像库。

- Microsoft Teams Copilot: 现在可以在 Teams 中使用 Copilot 合成长帖子，审查工作中的操作等。预计 2023 年 12 月，智能回顾功能将在 copilot 上线，能够通过提供摘要帮忙回顾错过的会议；Microsoft Whiteboard 中的 Copilot 将在 Teams 会议期间自动捕获和可视化口头讨论点，并将它们组织到 Whiteboard 中的协作空间中，并在所有会议参与者之间共享。

- Microsoft Outlook Copilot: 2024 年初推出，可为即将举行的会议做好准备，梳理邀请详细信息、相关电子邮件和相关文档。

- Microsoft Viva: 发布了 Microsoft Copilot Dashboard (预览链接: <https://insights.cloud.microsoft/#/CopilotDashboard>) 现在处于预览阶段，该仪表板根据 Microsoft 365 应用程序的使用情况显示有多少员工符合条件并准备好从 Copilot 中受益，详细分析了跨应用程序的 Copilot 使用情况，并根据会议、聊天、电子邮件和其他信息提供有关 Copilot 对生产力影响的早期信号。2024 年 1 月推出的预览版包括更多高级功能，例如 Microsoft Viva Insights 中的 Copilot 可以帮助用自然语言查询 Viva Insights 中的数据并生成个性化的报告。

(2) Microsoft Copilot Studio: 一款低代码工具, 用于根据自己的场景构建独立的 copilot, 除此之外, 微软融合 gpts 进来, 可以与 copilot 配合使用, 可以构建、测试、发布自定义的 copilot 和 gpt。Copilot Studio 现已推出, 与 Copilot for Microsoft 365 的集成现已推出预览版。演示案例:
<https://poqhpzqg4o.feishu.cn/minutes/obcnr562k7c5a359l95w3c12>

GPTS 融入 Copilot: 微软也对 OpenAI 上周宣布的 GPS 感到非常兴奋, 它是任何人创建 ChatGPT 定制版本的新方法, 这对工作或家庭中的特定任务更有帮助。今后, 将能够在 copilots 中使用 plugins 和 gpts 来定制您的体验。

(3) Microsoft Copilot for Service: 最新的 Copilot 提供基于用户角色的支持, 帮助企业加快客户服务的 AI 转型。Copilot for Service 包括 Microsoft Copilot for Microsoft 365, 能够利用生成式 AI 扩展现有的联络中心。在客户互动中, 客服人员可以用自然语言向 Copilot for Service 提问, 并从知识库中获得基于数据源的相关洞察, 从而更快、更智能地解决问题。

(4) Copilot in Microsoft Dynamics 365 Guides: 结合生成式 AI 和混合现实能力, 这款 Copilot 可以帮助一线人员在不影响工作流程的情况下, 更快地完成复杂任务和解决问题。这款无需手动操作的 Copilot 将率先用于 HoloLens 2, 帮助服务行业专业人员使用自然语言和人类手势, 通过叠加在设备上的内容和全息图提供交互式指导。

(5) Microsoft Copilot for Azure: 这是一款针对 IT 人员、可简化日常 IT 管理的 AI 伙伴。它不仅仅是一个工具, 更提供了统一的聊天体验, 能够理解用户的角色和目标, 并提高设计、操作和排除应用及基础架构故障的能力。Microsoft Copilot for Azure 可帮助 IT 团队深入了解其使用情况, 释放新的 Azure 能力, 并在云端和边缘协调任务。

1.4. 加强数据与 AI 的联系: Microsoft Fabric

Microsoft Fabric 现已开放使用, 它通过将每个人汇集到一个由 AI 驱动的平台, 在企业级数据基础上统一所有数据资产, 重塑团队处理数据的方式。提供将现有云数据仓库以及数据库从 Cosmos DB 或 Azure SQL DB、Mongo DB 和 Snowflake DB 添加到 Fabric 的方式, 不仅在云上, 而且在任何 Cloud Fabric 上。它们均采用开源 Apache Parquet 格式和 Fabric 原生的 Delta Lake 格式。

Copilot in Microsoft Fabric 还可以与 Microsoft Office 和 Teams 集成, 培养数据文化, 从而在整个企业内加大利用数据价值, 进行创新。

1.5. 版权承诺和安全

微软宣布将 Copilot 版权承诺进一步覆盖至使用 Azure OpenAI 服务的客户。这项新权利将被称为“客户版权承诺”。作为版权承诺扩展的一部分, 微软发布了新的文档, 帮助 Azure OpenAI 服务的客户实施技术措施, 降低侵权性内容的风险。客户需遵守该文档内容, 才能获取相应权利。

Azure AI Content Safety 现已正式商用, 帮助企业检测和减少有害内容, 以提供更好的在线体验。客户可将 Azure AI Content Safety 作为 Azure OpenAI 服

务中的内置安全系统,也可以将开源模型作为 Azure 机器学习中提示工程的一部分,还可以将其看作独立的 API 服务。

Microsoft Sentinel 和 Microsoft Defender XDR (前身为 Microsoft 365 Defender) 将合二为一,打造业界首个统一安全运维平台,并提供嵌入式 Security Copilot 体验。通过内置的生成式 AI 技术,提供一种统一的、强大的体验,专注于以机器的规模和速度抵御威胁,降低环境复杂度以为防御者提供帮助。

此外,嵌入到 Intune、Purview 和 Entra 中的 Security Copilot 扩展功能将帮助 IT 管理员、合规部门和身份管理团队简化复杂的场景。在 Entra 中,身份管理员可以快速排除身份访问故障。在 Purview 中,数据安全相关警报提供更为丰富的上下文信息,助以安全人员更快解决问题。在 Intune 中,IT 管理员可以使用“假设”分析来保持业务持续运行,同时改善治理和合规性。

2. 专题会议: 构建 copilot 和使用 AI studio

关于根据任务类型选择模型的演示从 19:30 开始:

https://poqhpzqg4o.feishu.cn/minutes/obcnsnrj1zifw657715ju5z8?from=from_copypink

AI Studio 预览: <https://ai.azure.com/>

- Microsoft Planner: 简单、灵活、可扩展且智能的工具,用于组织任务、将任务与目标对齐、管理资源和报告结果。它与 Microsoft Teams 集成,具备不同的视图,如列表、看板、时间线功能等。案例:可以管理从项目到程序再到完整的项目组合,并根据用户的需求进行配置,它会建议任务以满足用户计划的目标,通过简单的提示,可以快速从个人待办事项列表创建计划和报告等。

- Azure AI Studio Demonstration: 对 Azure AI Studio 的演示展示了其集成人工智能模型、管理数据和创建人工智能驱动应用程序的能力,强调了数据质量的重要性。在 Azure AI Studio 中提供了微调和检索增强两类选项,检索增强就是 Azure AI search 向量搜索和关键字搜索的结合形成混合搜索,提高了用户检索的使用体验。

-构建 Copilot 案例:一段视频中提及了微软的合作伙伴 Informatica 利用了 Azure OpenAI 创建了他们的 GPT 和 copilot 对应位置 29:45 处
https://poqhpzqg4o.feishu.cn/minutes/obcnsnrj1zifw657715ju5z8?from=from_copypink

- Azure AI Studio 提供了学习和挑战赛模块:通过给学习课程、活动参与来促进社区参与。

-多模态:将 GPT 4 Turbo、最新版本的 GPT 4、具有视觉功能的 GPT 4 Turbo 和 Dolly 3 引入 Azure OpenAI service。可以接收视频,也可以接收图像,用户可以为它提供视频或图像,提出有关材料的问题,例如它可以总结视频内容。案例:上传一段度假时拍摄的视频,没有告诉它具体在哪里,提出问题是提供一份行程和打包清单,我们可以看到他精确地指出视频是在哪个地方进行了徒步旅行,还能为我创建半天的行程和徒步旅行的打包清单。除此之外,还可以利用

Azure AI 语音、自定义合成语言和风格

-Model Benchmarking System: Azure AI Studio will offer a model benchmarking system, allowing users to compare different models using common benchmarks and their own data.

-AI Studio 向边缘端扩展: 简化了在 windows 设备上的生成式 AI 应用开发体验, 开发人员可以微调自定义部署 slm

-案例: Perplexity 的成功案例。Perplexity 旨在改变人们在互联网上检索和获取信息的方式, 其核心是性能和速度。他们广泛利用 Azure AI Studio 进行工作流程, 包括从构思到试验, 直至满意后部署到生产环境。通过 Azure OpenAI 服务, 他们能够控制并独立选择哪个模型在生产环境中提供服务。Perplexity 最初是一个 Slack 机器人, 用户提问, 它回答。他们后来将其与网络浏览器连接, 并增加了搜索互联网的能力, 效果令人惊叹, 他们收到了大量积极反馈, 并基于此开发了官方的 Perplexity 产品。

-创建流程: 对应使用 AI studio 的演示从 18:37 开始:

https://pqohpzqg4o.feishu.cn/minutes/obcnsmd7w5p63qc3i22b475j?from=from_copylink

(1) 创建 Azure AI 资源, 可以帮助链接不同的资源, 这些资源都包括在项目中共用户使用, 例如可以使用 Azure OpenAI、Azure Machine Learning、Azure AI Search。除此之外, 还可以看到项目所依赖的计算基础设施, 可以进行管理, 可以访问 API 及密钥; 还可以查看谁可以访问自己的项目, 也可以进行角色和权限的分配

(2) 演示中, 首先利用 GPT 3.5 turbo 问了一个问题, 模型没有响应, 因为该问题涉及到公司的具体数据, 而该模型尚未根据公司数据进行训练。

(3) 用 Azure AI Search 添加数据, 首先获取数据, 将其添加到整个平台都支持的链接中, 从中提取数据, 然后添加该数据到向量索引。然后问同样的问题的时候, 模型就知道答案了, 还可以准确查看文档中的信息来自何处, 除此之外, 在语言方面, 用西班牙语提问, 模型会根据西班牙语的提问从英文文档中学习, 然后给出西班牙语的回复。

(4) 想问多个问题, 并对答案评估。可以一次输入多个问题或者导入数据集, 可以手动查看的所有不同输出, 也可以将用户的偏好添加到这些输出中。除此之外, 还可以看到正在构建的 insights 的摘要, 构建的数据集也可以导出和共享, 也可以保存结果用于后面的评估

- 数据连接器: 让用户更轻松地连接数据并将其转换为正确的形式。微软拥有连接所有 Azure 数据源的连接器, 连接到它们并将它们引入是非常容易的, 也可以与来自本地、其他云提供商等的数据源进行通信, 以便为这些模型提供您所需的信息, 以便获得它们的最佳输出, 无论是结构化数据还是非结构化数据。

信息披露

分析师与研究助理简介

钱劲宇：德邦证券计算机行业首席分析师，哥伦比亚硕士，曾就职于华福证券、国泰君安，对 AI、网安、云计算等有深度的研究。

分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的职业态度，独立、客观地出具本报告。本报告所采用的数据和信息均来自市场公开信息，本人不保证该等信息的准确性或完整性。分析逻辑基于作者的职业理解，清晰准确地反映了作者的研究观点，结论不受任何第三方的授意或影响，特此声明。

投资评级说明

1. 投资评级的比较和评级标准： 以报告发布后的 6 个月内的市场表现为比较标准，报告发布日后 6 个月内的公司股价（或行业指数）的涨跌幅相对同期市场基准指数的涨跌幅； 2. 市场基准指数的比较标准： A 股市场以上证综指或深证成指为基准；香港市场以恒生指数为基准；美国市场以标普 500 或纳斯达克综合指数为基准。	类 别	评 级	说 明
	股票投资评级	买入	相对强于市场表现 20%以上；
		增持	相对强于市场表现 5%~20%；
		中性	相对市场表现在-5%~+5%之间波动；
		减持	相对弱于市场表现 5%以下。
	行业投资评级	优于大市	预期行业整体回报高于基准指数整体水平 10%以上；
		中性	预期行业整体回报介于基准指数整体水平-10%与 10%之间；
		弱于大市	预期行业整体回报低于基准指数整体水平 10%以下。

法律声明

本报告仅供德邦证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

市场有风险，投资需谨慎。本报告所载的信息、材料及结论只提供特定客户作参考，不构成投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需要。客户应考虑本报告中的任何意见或建议是否符合其特定状况。在法律许可的情况下，德邦证券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并进行交易，还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送，未经德邦证券研究所书面授权，本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。如欲引用或转载本文内容，务必联络德邦证券研究所并获得许可，并需注明出处为德邦证券研究所，且不得对本文进行有悖原意的引用和删改。

根据中国证监会核发的经营证券业务许可，德邦证券股份有限公司的经营经营范围包括证券投资咨询业务。