

智能计算引领者，AI芯片生态构筑宽广护城河

投资要点

- **推荐逻辑:** 1) 大模型和生成式 AI 带动数据中心 AI 芯片需求快速增长，英伟达数据中心 GPU 市占率超 80%，超强算力&通信性能+CUDA 生态+DGX Cloud AI 云服务构筑公司宽广的护城河；2) 游戏业务重回正向增长。随着显卡需求反弹，带动游戏业务回升；3) 公司业绩高速增长，未来三年净利润年复合增速 137%，估值显著低于同业平均水平。
- **AI 芯片驱动高速增长，超强硬件性能+软件生态构筑强劲护城河。**通过对硬件和软件进行集成，英伟达加速计算统一平台提供了端到端的解决方案。借助基于 GPU、DPU 和 CPU 新一代架构构建的 NVIDIA 加速计算平台，或重塑 AI 时代的数据中心。随着生成式 AI 和大语言模型需求不断增长，众多厂商竞相部署 AI 算力，带动公司数据中心芯片大幅增长。公司 A100、H100、H200、L40S 等明星产品线供不应求，订单能见度已至 2024 年；加速网络平台 Spectrum-X 旨在提高 AI 云的性能和效率，基于 DGX Cloud 的 AI 超级计算服务及 AI 代工服务为长尾市场提供定制大模型和生成式 AI 的服务。英伟达数据中心 GPU 市占率超 80%，超强算力&通信性能+CUDA 生态+DGX Cloud AI 云服务构筑公司宽广的护城河，预计未来三年收入 CAGR 高达 88%。
- **游戏业务重回增长轨道，Avatar 云引擎为游戏 AI 赋能。**游戏业务在经历四个季度的下调整之后，近两个季度业绩加速回暖，同比增速达到 21.7%和 81.5%。公司为游戏推出定制化的 AI 模型铸造服务 Avatar 云引擎，使用人工智能驱动的自然语言交互，通过给 NPC 赋予 AI 来改变游戏。随着显卡需求反弹，公司游戏业务有望持续回升。
- **智驾渗透率提升，汽车业务体量站上新台阶。**随着智能驾驶渗透率提升，公司汽车业务体量从 2023 财年开始呈现上升趋势，近五个季度单季收入超过 2.5 亿美元。公司与联发科在智能座舱领域开展合作，该系统集成了用于 AI 和图形的新 GPU chiplet IP，预计 2026~2027 年投入量产。
- **盈利预测与评级:** 公司是全球 AI 芯片龙头企业，预计未来三年归母净利润年复合增速为 137%。结合业绩增速，给予公司 2025 财年 35xPE，目标价 667.1 美元，对应市值 1.65 万亿美元。首次覆盖，给予“买入”评级。
- **风险提示:** AI 芯片需求或不及预期；游戏业务复苏或不及预期；产能扩产或不及预期；竞争加剧的风险；高端产品对华禁运政策加码的风险。

指标/年度	FY2022A	FY2023A	FY2024E	FY2025E	FY2026E
营业收入(百万元美元)	26914.00	26974.00	59677.23	93841.56	118573.21
增长率	61.40%	0.22%	121.24%	57.25%	26.35%
GAAP 净利润(百万元美元)	9752.00	4368.00	28117.21	47081.77	58069.36
增长率	125.12%	-55.21%	543.71%	67.45%	23.34%
Non-GAAP 净利润(百万元美元)	11259.00	8366.00	30737.00	50206.00	61420.00
增长率	79.37%	-25.69%	267.40%	63.34%	22.34%
每股收益 EPS	3.95	1.77	11.38	19.06	23.51
净资产收益率	36.65%	19.76%	55.99%	48.39%	37.38%
PE	117.94	263.32	40.91	24.43	19.81

数据来源: 公司公告, 西南证券

西南证券研究发展中心

分析师: 王湘杰
执业证号: S1250521120002
电话: 0755-26671517
邮箱: wxj@swsc.com.cn

分析师: 杨镇宇
执业证号: S1250517090003
电话: 023-67563924
邮箱: yzyu@swsc.com.cn

相对指数表现



数据来源: Wind

基础数据

52 周区间(美元)	140.31-504.09
3 个月平均成交量(百万)	43.98
流通股数(亿)	24.70
市值(亿美元)	11502

相关研究

目 录

1 智能计算引领者，GPGPU 开创者.....	1
2 AI 芯片驱动高速增长，超强硬件性能+软件生态构筑强劲护城河.....	3
3 游戏业务重回增长轨道，Avatar 云引擎为游戏 AI 赋能.....	8
4 全球智驾芯片龙头，汽车业务体量站上新台阶.....	10
5 盈利预测与投资建议.....	12

图 目 录

图 1: 英伟达重大发展历程	1
图 2: 英伟达硬件产品线	1
图 3: 英伟达软件产品线	1
图 4: 英伟达营收 (亿美元) 及增速	2
图 5: 英伟达 GAAP/non-GAAP 净利润 (亿美元) 及增速	2
图 6: 英伟达各业务收入 (亿美元)	3
图 7: 英伟达各业务收入占比	3
图 8: 全球 GPU 市场规模 (亿美元)	3
图 9: 中国 GPU 市场规模 (亿美元)	3
图 10: NVIDIA 加速计算平台	4
图 11: NVIDIA 加速系统与传统系统的对比	4
图 12: NVIDIA DGX A100 组件配置	4
图 13: DGX GH200 连接 NVLink 交换机系统的拓扑结构	6
图 14: NVIDIA DGX GH200 性能指标	6
图 15: 数据中心专属 CPU GRACE	6
图 16: 众多合作伙伴在英伟达 CUDA 环境下开发应用	7
图 17: 英伟达 CUDA 的部分加速库和开发工具	7
图 18: 英伟达游戏业务业绩情况 (亿美元)	8
图 19: 英伟达游戏业务产品线	8
图 20: 技嘉、微星&华擎月度收入 (亿新台币)	9
图 21: 技嘉、微星&华擎月度收入同比增速	9
图 22: 英伟达汽车业务业绩情况 (亿美元)	10
图 23: NVIDIA DRIVE 涵盖从云端到车端的创新	10
图 24: NVIDIA DRIVE 端到端解决方案	11
图 25: 英伟达 DRIVE 合作伙伴生态系统	12

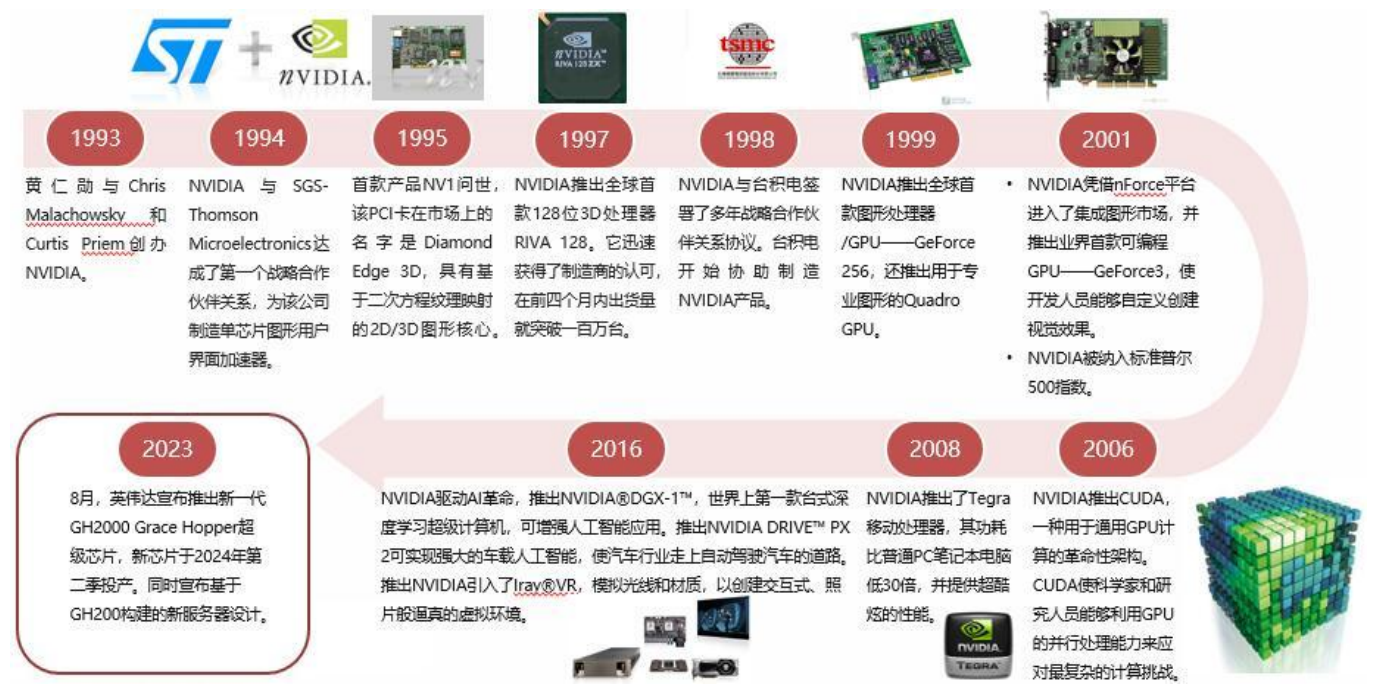
表 目 录

表 1: 英伟达重要股东持股	2
表 2: 英伟达 GPU 性能指标: A100&H100.....	5
表 3: NVIDIA L40S GPU 规格	5
表 4: BlueField-2 DPU & BlueField-3 DPU 性能指标	7
表 5: 英伟达 AI 超级计算服务	8
表 6: GEFORCE RTX 30 系列的规格	9
表 7: GEFORCE RTX 40 系列的规格	9
表 8: 英伟达自动驾驶芯片代际	11
表 9: 分业务收入	13
表 10: 可比公司估值情况	13
附: 财务报表	14

1 智能计算引领者，GPGPU 开创者

英伟达成立于 1993 年，是全球 GPU 龙头企业。英伟达早期借助 CUDA 库的构建和完善，显著降低了 GPU 编程门槛，并将 GPU 推向通用计算市场，引领了 GPGPU 的进化之路。英伟达借助软件业务形成的 AI、Omniverse、Drive Hyperion 等平台，不断壮大其计算生态，进一步夯实了公司的业务护城河，数据中心、智能驾驶、AI、元宇宙等业务将打开公司新的成长空间。

图 1：英伟达重大发展历程



数据来源：英伟达中国公众号，西南证券整理

英伟达的产品分为硬件和软件。硬件产品分为游戏和娱乐、笔记本电脑和 workstation、云和数据中心、网络、GPU 和嵌入式系统六大板块；软件产品有应用框架、应用和工具、游戏和创作、基础架构和云服务板块五大板块。

图 2：英伟达硬件产品线

硬件	软件
游戏和娱乐 GeForce 显卡 游戏笔记本电脑 G-SYNC 显示器	笔记本电脑和 workstation NVIDIA RTX 桌面工作站 专业笔记本电脑中的 NVIDIA RTX NVIDIA RTX 驱动的 AI 工作站 云和数据中心 概览 DGX 系统 NVIDIA OVX EGX 平台 IGX 平台 DRIVE Constellation 网络 概览 DPU 以太网 InfiniBand GPU GeForce NVIDIA RTX / Quadro 数据中心 Titan RTX 嵌入式系统 Jetson DRIVE AGX Clara AGX

数据来源：公司官网，西南证券整理

图 3：英伟达软件产品线

应用框架	应用和工具	游戏和创作	基础架构	云服务
AI 推理 - Triton 汽车 - DRIVE 云端 AI 视频流 - Maxine 数据分析 - RAPIDS 医疗健康 - Clara 高性能计算	智能视频分析 - Metropolis 元宇宙应用 - Omniverse 推荐系统 - Merlin 机器人 - Isaac 语音 AI - Riva 电信 - Aerial	应用目录 NGC 目录 NVIDIA Broadcast App NVIDIA NGC 3D 工作流 - Omniverse Modding - RTX Remix 数据中心 GPU 监控 NVIDIA RTX Experience NVIDIA RTX 桌面管理器 RTX 加速的创意应用程序 视频会议 AI 工作站	GeForce Experience NVIDIA Broadcast App 动画 - Machinima Modding - RTX Remix 边缘部署管理 推理服务 IO 加速 软件 虚拟 GPU	云游戏 云原生支持 NeMo BioNeMo Private Registry Base Command Fleet Command

数据来源：公司官网，西南证券整理

截止 2023 年 9 月 30 日，英伟达前五大股东持股合计 6.83 亿股，占比 27.6%。其中，Vanguard、贝莱德集团、FMR LLC（富达投资集团）、道富四家机构持股占比分别为 8.25%、7.28%、4.97%和 3.59%。创始人黄仁勋持股近 8668 万股，占比 3.51%。

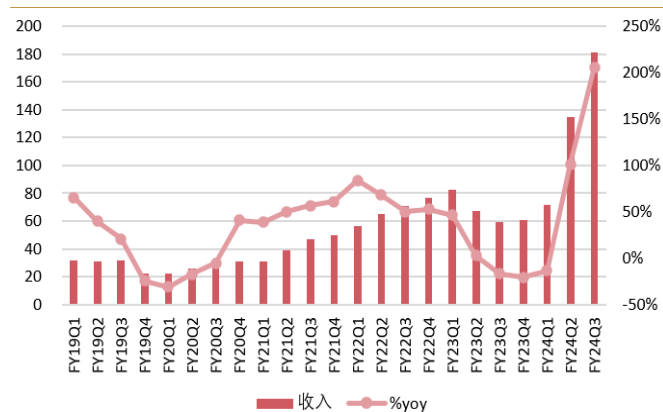
表 1：英伟达重要股东持股

重要股东	持股数量（亿股）	持股比例
Vanguard	2.04	8.25%
贝莱德集团	1.80	7.28%
FMR LLC	1.23	4.97%
道富集团	0.89	3.59%
黄仁勋	0.87	3.51%

数据来源：Bloomberg，西南证券整理

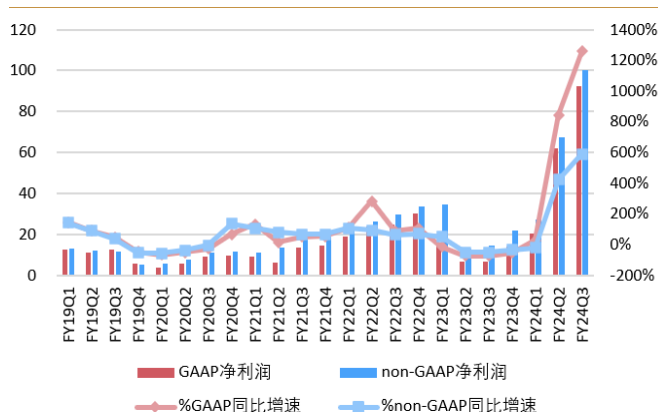
今年以来 AI 大模型和生成式 AI 迎来爆发，对 GPU 的需求大幅提升，英伟达 2024 财年呈现快速增长态势。FY24Q2 营收 135 亿美元，同比增长 101%，环比增长 88%，大超市市场预期；FY24Q3 收入 181 亿美元，同比增长 206%，实现了近五年来最大单季度同比涨幅。FY24Q3 GAAP 净利润 92.4 亿美元，同比增长 1259%，环比增长 49.4%；non-GAAP 净利润 100.2 亿美元，同比增长 588%，环比增长 48.7%。净利润的增长也超市市场预期，实现近五年来单季度最大涨幅。

图 4：英伟达营收（亿美元）及增速



数据来源：公司公告，西南证券整理

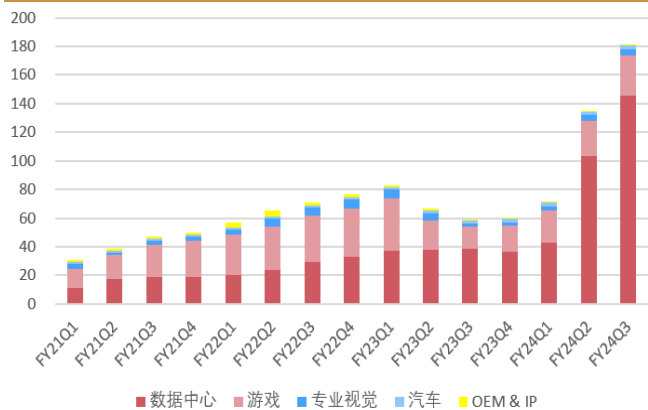
图 5：英伟达 GAAP/non-GAAP 净利润（亿美元）及增速



数据来源：公司公告，西南证券整理

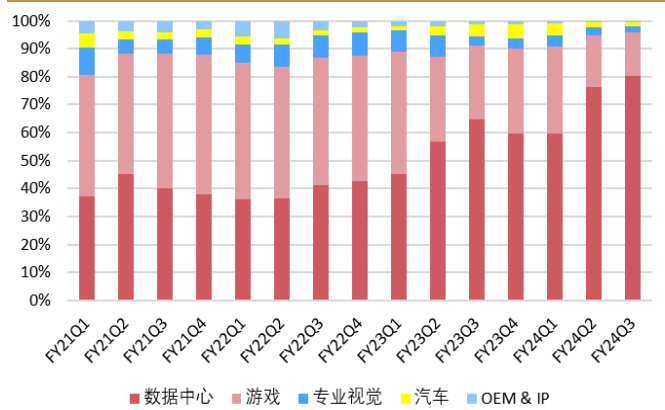
英伟达业务结构可分为数据中心、游戏、专业视觉、汽车、OEM&IP 五大板块。其中，数据中心和游戏为公司主要收入来源。FY24Q3 数据中心业务营收 145 亿美元，同比增长 278.7%，环比增长 40.6%，收入占比超过 80%，成为英伟达收入和利润增长的最主要贡献。对生成式 AI 和大语言模型的需求不断增长，众多云公司竞相部署 AI 芯片，带动了公司的数据中心芯片大幅增长。

图 6：英伟达各业务收入（亿美元）



数据来源：公司公告，西南证券整理

图 7：英伟达各业务收入占比



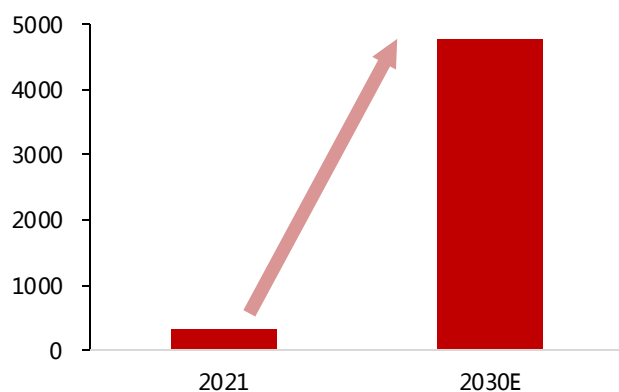
数据来源：公司公告，西南证券整理

2 AI 芯片驱动高速增长，超强硬件性能+软件生态构筑强劲护城河

人工智能发展离不开算力的支撑，随着模型的参数量和复杂程度不断增长，算力基础设施的供给水平将直接影响应用落地及其迭代的进程。目前全球 AI 服务器占 AI 基础设施市场的八成以上，是 AI 基础设施的主体。IDC 预计 2025 年全球 AI 服务器市场规模达 266 亿美元。2020-2025 年 CAGR 为 18.9%，远超全球通用服务器市场增速。

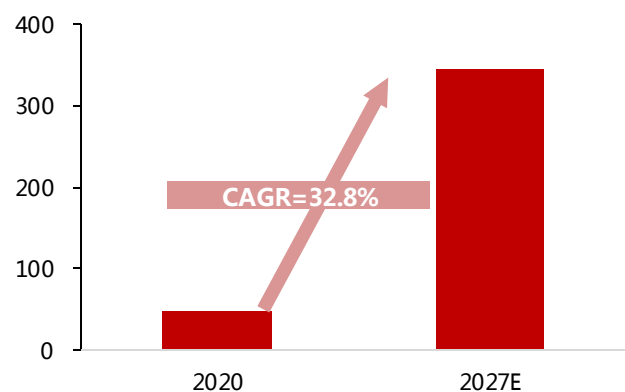
GPU 是目前主流的 AI 算力芯片。根据 Verified Market Research，2021 年全球 GPU 市场规模 334.7 亿美元，预计到 2030 年约 4773.7 亿美元，CAGR 达 33.3%；2020 年中国 GPU 市场规模 47.39 亿美元，预计到 2027 年市场规模达 345.6 亿美元，CAGR 为 32.8%，国内外 GPU 市场前景广阔。

图 8：全球 GPU 市场规模（亿美元）



数据来源：Verified Market Research，西南证券整理

图 9：中国 GPU 市场规模（亿美元）

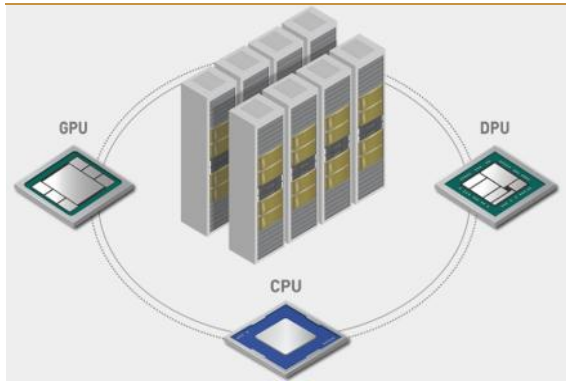


数据来源：Verified Market Research，西南证券整理

英伟达的数据中心解决方案整体提供了 NVIDIA 加速计算统一平台，从 AI、数据分析，到高性能计算（HPC），再到渲染，数据中心都是攻克某些重要挑战的关键。端到端的 NVIDIA 加速计算平台对硬件和软件进行了集成，可为企业构建强大而安全的基础设施蓝图，支持在现代化工作负载中实施从开发到部署的操作。

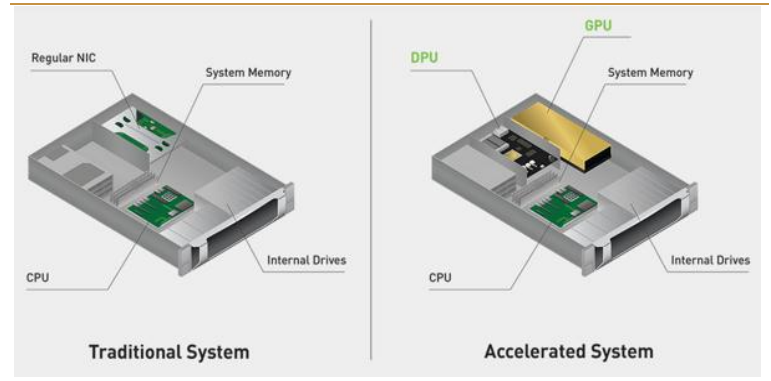
借助基于 GPU、DPU 和 CPU 三种新一代架构构建的 NVIDIA 加速计算平台，英伟达或重塑 AI 时代的数据中心。NVIDIA 加速计算平台可提供支持从用于改善业务预测的数据分析、到自动驾驶汽车使用的 AI、再到用于医疗诊断的高级可视化等多种应用的基础架构。采用 NVIDIA 加速系统的每个服务器和工作站都将配备计算加速器，为当今使用的现代应用（包括 AI、可视化和自主机器等）提供支持。其中许多系统还会配备 DPU，可加速对云原生和云计算框架至关重要的网络、存储和安全服务。

图 10：NVIDIA 加速计算平台



数据来源：英伟达官网，西南证券整理

图 11：NVIDIA 加速系统与传统系统的对比



数据来源：英伟达官网，西南证券整理

数据中心 GPU 方面，公司产品线涵盖了 A100、H100、L40、L40S、GH200 等产品。A100 采用 NVIDIA Ampere 架构，是 NVIDIA 数据中心平台的引擎。A100 性能比上一代 Volta 提升了 20 倍，并可划分为 7 个 GPU 实例。A100 提供 40GB 和 80GB 显存两种版本，A100 80GB 将 GPU 显存增加了一倍，并提供超快速的显存带宽（超过 2 TB/s），可处理超大型模型和数据集。2048 个 A100 可在一分钟内成规模地处理 BERT（某自然语言处理模型）之类的训练工作负载，A100 的推理吞吐量是 Intel Xeon Gold 6240 CPU 的 249 倍。

图 12：NVIDIA DGX A100 组件配置



数据来源：英伟达官网，西南证券整理

H100 配备第四代 Hopper 架构和 Transformer 引擎，与 A100 相比，H100 的综合技术创新可将大型语言模型速度最大提高 30 倍。对比 A100 GPU，训练方面，H100 可为多专家 (MoE) 模型提供高 9 倍的训练速度；推理方面，H100 可将推理速度提高 30 倍，并提供超低的延迟；而在 HPC 应用方面的性能则提升高达 7 倍。H100 CNX 将 H100 的强大功能与 NVIDIA ConnectX-7 智能网卡的先进网络功能相结合，加速 GPU 驱动的 I/O 密集型工作负载。

表 2：英伟达 GPU 性能指标：A100&H100

GPU	A100	H100
算力 (FP64/TF32/FP16/INT8)	19.5 TFLOPS /312 TFLOPS / 624 TFLOPS /1248 TOPS	67 TFLOPS /989 TFLOPS / 1979 TFLOPS /3958 TOPS
显存	40 GB / 80 GB	80GB
显存带宽	2039 GB/s	3.35TB/s
最大功耗	400 W	最高可配置到 700 W
互联技术	NVLink: 600 GB/s PCIe 4.0: 64 GB/s	NVLink: 900GB/s PCIe 5.0: 128GB/s
多实例 GPU	最大为 7 MIG	最大为 7 MIG

数据来源：英伟达官网，西南证券整理

英伟达 L40 系列由 Ada Lovelace 架构提供支持，为 GPU 加速数据中心工作负载提供神经网络、可视化、计算和 AI 功能。最新一代的 L40S，其 Tensor 性能为 1466 TFLOPS，RT Core 性能为 212 TFLOPS，单精度浮点运算性能 91.6 TFLOPS。NVIDIA Ada Lovelace 架构提供第四代 Tensor Core、第三代 RT Core、CUDA Core 核心、Transformer 引擎、DLSS 3 的支持。

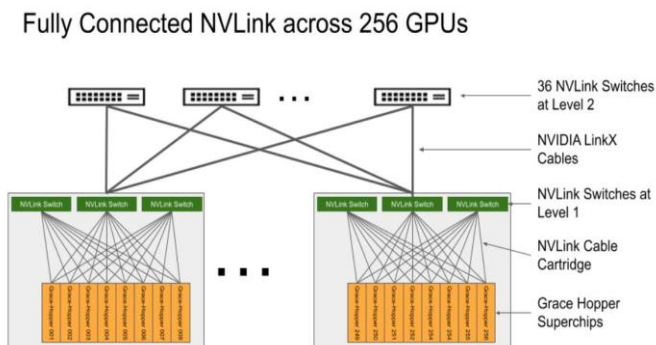
表 3：NVIDIA L40S GPU 规格

NVIDIA L40S GPU 规格	
算力 (FP32/TF32/FP16/FP8)	91.6 TFLOPS /366 TFLOPS /733 TFLOPS /1466TOPS
显存	带错误校验的 48GB GDDR6
显存带宽	864GB/s
最大功耗	350 W
互联技术	PCIe 4.0x16: 双向 64 GB/s

数据来源：英伟达官网，西南证券整理

DGX GH200 是将 256 个 NVIDIA Grace Hopper 核心完全连接到单个 GPU 中的新型 AI 超级计算机。NVIDIA DGX GH200 支持万亿参数 AI 大模型训练，能处理大规模推荐系统、生成式人工智能和图形分析处理 TB 级模型，并为大型 AI 模型提供 144 TB 的共享内存和线性可扩展性。H200 在显存上做了大幅升级，是市面上第一款带 HBM3e 显存的 GPU，其显存容量达到 141GB，显存带宽高达 4.8 TB/s。

图 13: DGX GH200 连接 NVLink 交换机系统的拓扑结构



数据来源: 英伟达官网, 西南证券整理

图 14: NVIDIA DGX GH200 性能指标

DGX GH200 Technical Specifications	
CPU and GPU	256x NVIDIA Grace Hopper Superchips
CPU Cores	18,432 Arm® Neoverse V2 Cores with SVE2 4X 128b
GPU Memory	144TB
Performance	1 exaFLOPS
Networking	256x OSFP single-port NVIDIA ConnectX®-7 VPI with 400Gb/s InfiniBand 256x dual-port NVIDIA BlueField®-3 VPI with 200Gb/s InfiniBand and Ethernet 24x NVIDIA Quantum-2 QM9700 InfiniBand Switches 20x NVIDIA Spectrum™ SN2201 Ethernet Switches 22x NVIDIA Spectrum SN3700 Ethernet Switches
NVIDIA NVLink Switch System	96x L1 NVIDIA NVLink Switches 36x L2 NVIDIA NVLink Switches
Management Network	Host baseboard management controller (BMC) with RJ45
Software	NVIDIA AI Enterprise (optimized AI software) NVIDIA Base Command (orchestration, scheduling, and cluster management) DGX OS / Ubuntu / Red Hat Enterprise Linux / Rocky (operating system)
Support	Comes with three-year business-standard hardware and software support

数据来源: 英伟达官网, 西南证券整理

CPU 方面, 英伟达宣布推出首款面向 AI 基础设施和高性能计算的数据中心专属 CPU——NVIDIA Grace, 由两个 CPU 芯片通过最新一代 NVLink-C2C 技术互联组成。Grace 基于最新的 ARM v9 架构, 单个 socket 拥有 144 个 CPU 核心, 可提供 1TB/s 的内存带宽, 利用纠错码 (ECC) 等机制提供当今领先服务器芯片两倍的内存带宽和能效, 兼容性亦十分突出, 可运行 NVIDIA 所有软件堆栈和平台, 包括 NVIDIA RTX、HPC、Omniverse 等。

图 15: 数据中心专属 CPU GRACE



数据来源: 英伟达官网, 西南证券整理

DPU 方面, 英伟达截止目前已经推出了三代 DPU, 目前广泛应用的主要是 BlueField-2 和 BlueField-3 DPU。BlueField-2 DPU 将 NVIDIA Mellanox ConnectX-6 Dx 网络适配器与 ARM 核心阵列相结合, 可提供具有完整软件可编程性的专用硬件加速引擎, 速度可达 200 Gb/s。BlueField-2 可针对从边缘到核心数据中心和云的云网络、存储、网络安全、数据分析、HPC 和人工智能提供灵活且高性能的解决方案, 同时降低总体拥有成本。

BlueField-3 DPU 是英伟达首款以线速处理软件定义网络、存储和网络安全的 DPU, 速度可达 400 Gb/s。BlueField-3 支持客户组织从云到核心数据中心再到边缘计算, 构建软件定义硬件加速的 IT 基础设施。

表 4: BlueField-2 DPU & BlueField-3 DPU 性能指标

DPU	BlueField-2	BlueField-3
网络接口	以太网或 InfiniBand, 最大速度 200 Gb/s	以太网或 InfiniBand, 最大速度 400 Gb/s
ARM 核心	8 个 ARM v8 A72 核心 (64 位)	16 个 ARMv8.2+A78 Hercules 核心 (64 位)
内存	单 DDR4 DRAM 控制器, 8GB/16GB 板载 DDR4	双 DDR5 5600MT/s DRAM 控制器, 16GB 板载 DDR5
存储	NVMe 和 VirtIO-blk, NVMe-oF 和 NVMe/TCP 加速	M.2/U.2 接口, 用于直连存储
AI 加速	GPUDirect, GPUDirect Storage	HPC/AI All-to-All 引擎, GPUDirect/Storage
软件	DOCA 软件开发套件	DOCA 软件开发套件

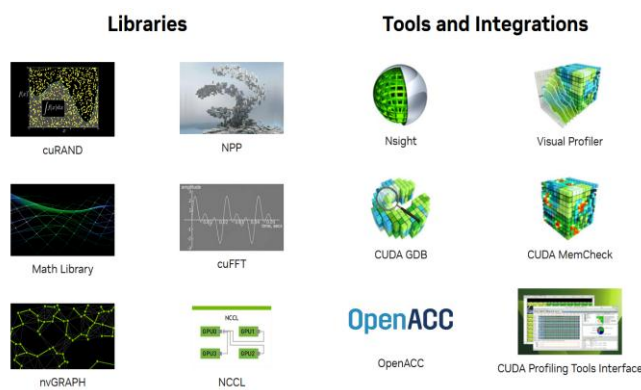
数据来源: 英伟达官网, 西南证券整理

CUDA 是由英伟达开发的用于 GPU 上的通用计算的并行计算平台和编程模型。借助 CUDA, 开发人员能够利用 GPU 的强大性能, 显著加速计算应用程序。在 GPU 加速的应用程序中, 工作负载的顺序部分在优化了单线程性能的 CPU 上运行, 而应用程序的计算密集部分则在成千上万个 GPU 核心上并行运行。使用 CUDA 时, 开发人员使用当下主流的编程语言, 如 C、C++、Fortran、Python 和 MATLAB 等, 并通过一些基本的关键字的扩展来表达并行性。英伟达的 CUDA 提供了开发 GPU 加速应用程序所需的 Toolkit, CUDA Toolkit 包括 GPU 加速库、编译器、开发工具和 CUDA 运行时库等。

自 2006 年诞生以来, CUDA 生态系统迅速发展。使用 CUDA 开发的成千上万个应用程序已经部署到嵌入式系统、工作站、数据中心和云中的 GPU 上, 合作伙伴包括了微软、Adobe 等全球头部科技企业。CUDA 在包括图像处理、深度学习、数值分析和计算科学等广泛的领域得到加速应用。

图 16: 众多合作伙伴在英伟达 CUDA 环境下开发应用


数据来源: 英伟达官网, 西南证券整理

图 17: 英伟达 CUDA 的部分加速库和开发工具


数据来源: 英伟达官网, 西南证券整理

AI 超级计算平台服务方面, 英伟达推出 DGX Cloud 的 AI 超级计算服务, 与微软 Azure、谷歌 OCP、Oracle OCI 合作, 通过 Web 浏览器就能访问, 以便企业为生成式 AI 和其他开创性应用训练先进的模型。DGX Cloud 实例的起步价为每个实例每月 36999 美元。其每个实例都具有 8 个 NVIDIA H100 或 A100 80GB Tensor Core GPU, 每个节点共有 640GB 的 GPU 内存。DGX Cloud 提供了专用的 DGX AI 超级计算集群, 并配备了 NVIDIA AI 软件 (加速库套件)。

英伟达推出了全新云服务及代工厂 NVIDIA AI Foundations, 使企业能构建、改进、运营使用其专有数据训练的、用于特定领域任务的定制大模型和生成式 AI。

表 5：英伟达 AI 超级计算服务

英伟达 AI 超级计算服务	
NeMo	文本生成模型构建服务，提供 80 亿~5300 亿个参数的模型，会定期更新额外的训练数据，帮助企业为客服、企业搜索、聊天机器人、市场情报等生成式 AI 应用进行模型定制。
Picasso	视觉语言模型构建服务，具有先进的文生图、文本转视频、文本转 3D 功能，可为产品设计、数字孪生、角色创建等使用自然文本提示的应用快速创建和定制视觉内容。这些素材导入 Omniverse 可构建逼真的元宇宙应用和数字孪生仿真。
BioNeMo	生命科学服务，提供 AI 模型训练和推理，加速药物研发中最耗时和成本最高的阶段，可加速新蛋白质和治疗方法的创建以及基因组学、化学、生物学和分子动力学研究。
ACE	英伟达发布一项游戏定制 AI 模型代工服务，即 Avatar Cloud Engine for Games。该服务可用在软件 and 游戏中构建和部署定制的语音、对话和动画 AI 模型。

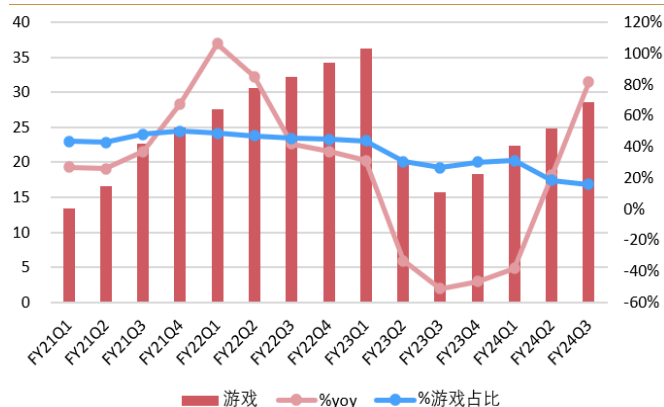
数据来源：英伟达 GTC2023，西南证券整理

在与生成式 AI 相关企业的合作过程中，英伟达与 Adobe 合作开发新一代先进生成式 AI 模型；与 Getty Images 合作训练文生图、文本转视频基础模型；与 Shutterstock 合作，训练从简单文本提示中创建生成式 3D 模型(Edify-3D)，将创作时间从几小时减少到几分钟；与三菱联合发布了将用于加速药物研发的日本第一台生成式 AI 超级计算机 Tokyo-1 等。

3 游戏业务重回增长轨道，Avatar 云引擎为游戏 AI 赋能

游戏业务曾是英伟达的第一大业务，在 2021-2022 财年占据了 46% 以上的收入比例。不过随着游戏和加密货币市场需求疲软，2023 财年以来公司游戏业务经历了大幅调整，FY23 游戏业务同比下滑 27% 至 90.7 亿美元，占收入比例降至 33.6%；FY24Q1 游戏业务收入 22.4 亿美元，同比减少 38%。FY24Q2-Q3 游戏业务见底回升，分别实现营收 24.9 亿美元、28.6 亿美元，同比增长 21.7% 和 81.5%。在经历四个季度的下跌调整之后，游戏业务重新实现正向增长。目前游戏业务收入占比为 15.8%。

图 18：英伟达游戏业务业绩情况（亿美元）



数据来源：公司公告，西南证券整理

图 19：英伟达游戏业务产品线



数据来源：英伟达官网，西南证券整理

英伟达游戏显卡产品主要有 GEFORCE RTX 30 系列和 GEFORCE RTX 40 系列等。GEFORCE RTX 30 系列采用第 2 代 NVIDIA RTX 架构——NVIDIA Ampere 架构，搭载专用的第 2 代 RT Core、第 3 代 Tensor Core 以及 SM 多单元流处理器，可带来逼真的光线追踪效果和先进的 AI 性能，支持 NVIDIA DLSS，AI 加速性能出色，支持 NVIDIA Reflex，系统延迟极低。

表 6: GEFORCE RTX 30 系列的规格

产品	GeForce RTX 3090 Ti	GeForce RTX 3090	GeForce RTX 3080	GeForce RTX 3070	GeForce RTX 3060
CUDA 核心数量	10752	10496	8960 / 8704	5888	3584
加速频率 (GHz)	1.86	1.70	1.71	1.73	1.78
显存容量	24GB	24GB	10GB/12GB	8GB	8GB/12GB
显存类型	GDDR6X	GDDR6X	GDDR6X	GDDR6	GDDR6
显卡功率 (W)	450	350	350 / 320	220	170

数据来源: 英伟达官网, 西南证券整理

GEFORCE RTX 40 系列 GPU 由更高效的 NVIDIA Ada Lovelace 架构提供动力支持: 采用新型 SM 多单元流处理器将性能功耗比最高提升至 2 倍; 采用第四代 Tensor Core, 与仅使用传统的图像渲染方式相比, 采用 DLSS 3 时, 性能最高提升至 4 倍; 采用第三代 RT Core, 将光线追踪性能最高提升至 2 倍, 在性能和 AI 驱动图形领域实现质的飞跃。

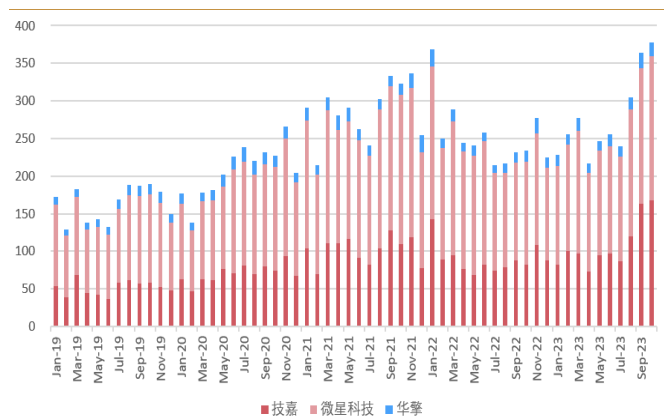
表 7: GEFORCE RTX 40 系列的规格

产品	GeForce RTX 4090	GeForce RTX 4080	GeForce RTX 4070	GeForce RTX 4060
CUDA 核心数量	16384	9728	5888	3072
加速频率 (GHz)	2.52	2.51	2.48	2.46
显存容量	24GB	16GB	12GB	8GB
显存类型	GDDR6X	GDDR6X	GDDR6X	GDDR6
显卡功率 (W)	450	320	200	115

数据来源: 英伟达官网, 西南证券整理

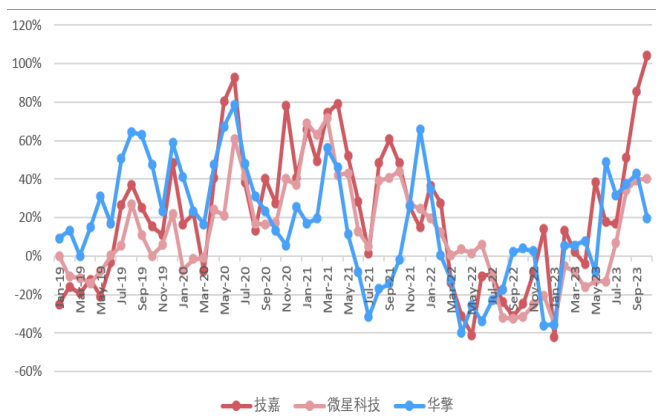
英伟达游戏业务合作伙伴众多, 其中台湾的几家板卡厂商如技嘉、微星、华擎等, 部分产品将英伟达的 GEFORCE RTX 等系列的产品加工成为完整的游戏显卡用以出售。三家板卡厂商的月度收入在经过 2021 年的高点以后, 2023Q3 前收入趋于平稳, 2023 年 8-10 月收入开始大幅反弹, 近两个月三家收入整体已提升至历史最高水平, 从侧面反映英伟达游戏业务已恢复正常。

图 20: 技嘉、微星&华擎月度收入 (亿新台币)



数据来源: 公司公告, 西南证券整理

图 21: 技嘉、微星&华擎月度收入同比增速



数据来源: 公司公告, 西南证券整理

英伟达的游戏定制 AI 模型代工服务，即 Avatar Cloud Engine for Games (ACE, Avatar 云引擎)，可用来在软件和游戏中构建和部署定制的语音、对话和动画 AI 模型。借助 ACE，游戏中的普通 NPC 可以变成能够发起对话的互动角色。英伟达引入了 NeMo SteerLM，这项新技术使开发者能够定制游戏 NPC 的个性，以实现更情感丰富和逼真的交互，开启了未来游戏 NPC 的新篇章。

4 全球智驾芯片龙头，汽车业务体量站上新台阶

英伟达是全球自动驾驶芯片平台的头部企业。随着智能驾驶渗透率的提升，公司汽车业务体量从 2023 财年开始呈现上升趋势，近五个季度单季收入体量超过 2.5 亿美元。

图 22：英伟达汽车业务业绩情况（亿美元）



数据来源：公司公告，西南证券整理

图 23：NVIDIA DRIVE 涵盖从云端到车端的创新



数据来源：英伟达官网，西南证券整理

英伟达的自动驾驶 Drive 平台是面向自动驾驶汽车的端到端解决方案，开启了软件定义自动驾驶汽车的时代。英伟达可提供完整的智能汽车硬件和软件堆栈，将使用新的软件和服务商业模式，向软件定义汽车过渡。Drive 平台包括 DRIVE Hyperion（自动驾驶汽车开发平台）、DRIVE SDK（构建部署自动驾驶功能的模块化开放式平台，包括高度自动化监督驾驶“DRIVE AV”和 AI 驾驶舱“DRIVE IX”）、DRIVE Sim（为开发自动驾驶提供模拟场景的仿真平台）、NVIDIA DGX（DNN 训练平台）。

图 24: NVIDIA DRIVE 端到端解决方案



数据来源：英伟达官网，西南证券整理

DRIVE Orin SoC 是英伟达现在主要的智能车辆中央计算机，为自动驾驶功能、置信视图、数字集群以及 AI 驾驶舱提供动力支持。借助可扩展的 DRIVE Orin 产品系列，开发者只需在整个车队中构建、扩展和利用一次开发投资，便可从 L2+级系统一路升级至 L5 级全自动驾驶汽车系统。目前已超过 35 家整车厂和 robotaxi 客户采用 Orin 芯片。

DRIVE Thor 是英伟达新一代集中式车载计算平台，可在单个安全、可靠的系统上运行高级驾驶员辅助应用和车载信息娱乐应用。DRIVE Thor 超级芯片借助英伟达新的 CPU 和 GPU 突破，可提供出色的 2000 万亿次浮点运算性能，同时降低总体系统成本。Thor 计划于 2025 年开始量产。

表 8: 英伟达自动驾驶芯片代际

代际	芯片型号	算力 (TOPS)	功耗 (W)	制程
第一代	Tegra X1	2	10	20nm
第二代	Tegra Parker	8	15	16nm
第三代	Tegra Xavier	20	20	16nm
第三代	Drive Xavier	30	30	12nm
第四代	Drive AGX Orin	254	45	7nm
第五代	Thor	2000	——	——

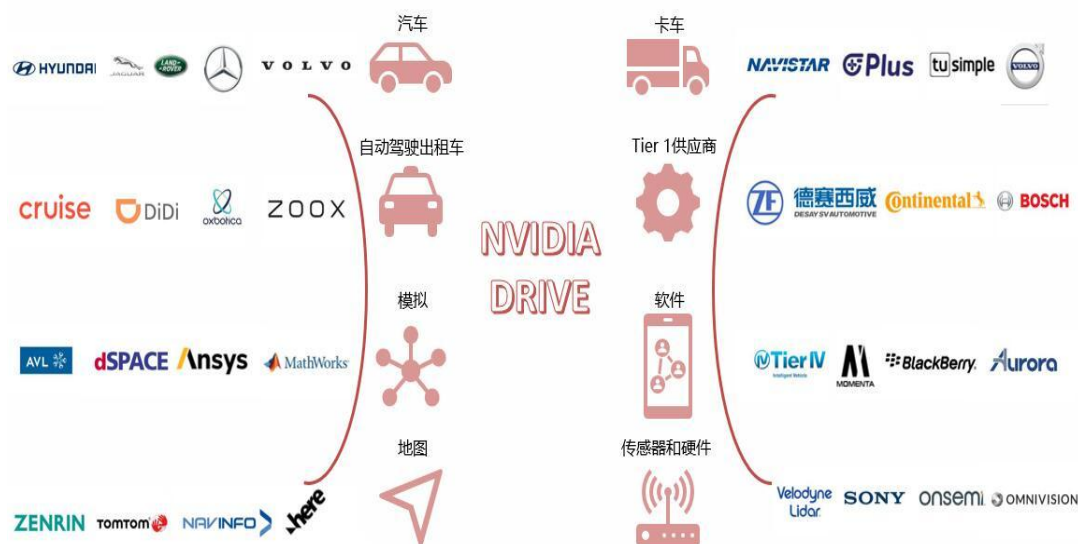
数据来源：英伟达官网，西南证券整理

2023 年 5 月 29 日，英伟达与联发科宣布将在智能座舱领域开展合作。双方将共同为新一代智能汽车提供解决方案，联发科将开发集成英伟达 GPU 芯粒 (chiplet) 的汽车 SoC，搭载 NVIDIA AI 和图形计算 IP，该芯粒支持互连技术，可实现芯粒间流畅且高速的互连互通。英伟达与联发科合作的首款芯片将用于智能座舱，预计 2025 年问世，并在 2026 年至 2027 年投入量产。

2023 年 8 月，原小鹏汽车自动驾驶副总裁吴新宙加入英伟达，领导英伟达自动驾驶团队。作为智驾领域的顶尖人才，吴新宙拥有深厚的学术背景和丰富的工程管理经验，是小鹏汽车在智能驾驶领域取得突破并站在行业前沿的核心人物。吴新宙的到来或弥补英伟达在汽车自动驾驶芯片和软件领域的短板，他在推动英伟达的智能汽车业务发展方面将起到关键作用。

英伟达还与全球数百家汽车制造商、一级供应商、开发商以及研究所通力合作，携手将 GPU 技术和 AI 相结合，从而转变深度学习、自然语言处理和姿势控制技术，改变人们的驾驶方式，并助力车辆实现自动驾驶。英伟达 DRIVE 合作伙伴生态系统涵盖 8 个细分行业，包括比亚迪、捷豹路虎、联发科、梅赛德斯-奔驰、理想、小鹏、蔚来、小马智行、图森未来、沃尔沃等重点合作伙伴。

图 25：英伟达 DRIVE 合作伙伴生态系统



数据来源：英伟达官网，西南证券整理

5 盈利预测与投资建议

关键假设：

假设 1：大模型训练和推理对 AI 加速卡需求持续提升，英伟达高端 GPU 市占率维持在 80% 以上。

假设 2：下游需求复苏，新产品上市带动游戏显卡 ASP 提升，游戏业务未来三年重回增长态势。

假设 3：智驾渗透率持续提升，智能汽车业务未来三年复合增速达到 27% 左右。

基于以上假设，我们预测公司 2024-2026 财年分业务收入如下表：

表 9：分业务收入

单位：百万美元		FY2023A	FY2024E	FY2025E	FY2026E
数据中心	收入	15,005.00	46,404.46	77,727.48	99,949.76
	增速	41.38%	209.26%	67.50%	28.59%
游戏	收入	9,067.00	10,333.66	12,536.80	14,407.29
	增速	-27.24%	13.97%	21.32%	14.92%
专业视觉	收入	1,544.00	1,519.45	1,823.34	2,005.67
	增速	-26.86%	-1.59%	20.00%	10.00%
汽车	收入	903.00	1,107.89	1,411.01	1,850.40
	增速	59.54%	22.69%	27.36%	31.14%
其他	收入	455.00	311.77	342.94	360.09
	增速	-60.84%	-31.48%	10.00%	5.00%
合计	收入	26,974.00	59,677.23	93,841.56	118,573.21
	增速	0.22%	121.24%	57.25%	26.35%

数据来源：Wind，西南证券

我们选取了半导体产业链的 3 家可比公司，2023-2024 年平均估值为 119 倍 PE、41 倍 PE。公司是全球 AI 芯片龙头企业，大模型训练与推理带动 AI 芯片需求爆发，预计公司未来三年归母净利润年复合增速为 137%，当前股价下公司 2024-2025 财年对应估值分别为 41xPE、24xPE，相比同业平均水平，公司显著低估。结合业绩增速，给予公司 2025 财年 35xPE，目标价 667.1 美元，对应市值 1.65 万亿美元。首次覆盖，给予“买入”评级。

表 10：可比公司估值情况

股票代码	证券简称	股价 (美元)	市值 (亿美元)	PE			
				2022A	2023E	2024E	2025E
AMD.O	AMD	118.38	1912	79	211	60	38
AVGO.O	博通	913.12	3769	17	26	22	20
INTC.O	英特尔	41.92	1767	14	-355	42	20
平均值				37	119	41	26
NVDA.O	英伟达	465.66	11502	263	41	24	20

数据来源：Wind，西南证券整理。注：平均值剔除负值影响

附：财务报表

资产负债表 (百万美元)	FY2023A	FY2024E	FY2025E	FY2026E	利润表 (百万美元)	FY2023A	FY2024E	FY2025E	FY2026E
货币资金	13296.00	34278.61	74150.25	126325.39	销售收入	26974.00	59677.23	93841.56	118573.21
应收账款	3827.00	9156.82	14398.96	18193.76	销售成本	11618.00	16619.60	23938.57	30858.17
预付款项	791.00	1138.96	1791.01	2263.02	销售和管理费用	2440.00	2780.96	3941.35	5608.51
存货	5159.00	5599.95	8066.06	10397.62	研发费用	7339.00	8354.81	10041.05	13043.05
其他流动资产	0.00	0.00	0.00	0.00	营业利润	4229.00	31327.37	55189.32	68063.89
流动资产总计	23073.00	50174.34	98406.28	157179.78	其他非经营损益	-48.00	265.00	201.00	253.00
长期股权投资	0.00	0.00	0.00	0.00	税前利润	4181.00	31592.37	55390.32	68316.89
固定资产	3807.00	3969.57	4075.47	4193.38	所得税	-187.00	3475.16	8308.55	10247.53
无形资产	1676.00	1813.33	1950.67	2071.33	净利润	4368.00	28117.21	47081.77	58069.36
其他非流动资产	12626.00	13873.00	16447.00	19458.00	少数股东损益	0.00	0.00	0.00	0.00
非流动资产合计	18109.00	19655.90	22473.14	25722.71	归属于母公司股东的净利润	4368.00	28117.21	47081.77	58069.36
资产总计	41182.00	69830.25	120879.42	182902.49	EBITDA	5720.00	33107.93	57027.93	70098.09
应付账款	1193.00	2219.74	3508.15	4913.64	NOPLAT	4411.00	28251.12	47149.64	57956.11
其他流动负债	4120.00	4984.30	7179.30	9254.52	EPS(元)	1.77	11.38	19.06	23.51
流动负债合计	6563.00	7204.04	10687.44	14168.16	Non-GAAP 净利润	8366.00	30737.00	50206.00	61420.00
长期债务	9703.00	9703.00	9703.00	9703.00					
其他非流动负债	2815.00	2705.00	3189.00	3662.00					
非流动负债合计	12518.00	12408.00	12892.00	13365.00	财务分析指标	FY2023A	FY2024E	FY2025E	FY2026E
负债合计	19081.00	19612.04	23579.44	27533.16	成长能力				
归属母公司权益	22101.00	50218.21	97299.98	155369.34	销售收入增长率	0.22%	121.24%	57.25%	26.35%
少数股东权益	0.00	0.00	0.00	0.00	EBIT 增长率	-58.85%	666.47%	73.93%	22.93%
股东权益合计	22101.00	50218.21	97299.98	155369.34	EBITDA 增长率	-49.48%	478.81%	72.25%	22.92%
负债和股东权益合计	41182.00	69830.25	120879.42	182902.49	税后利润增长率	-55.21%	543.71%	67.45%	23.34%
					盈利能力				
					毛利率	56.93%	72.15%	74.49%	73.98%
现金流量表 (百万美元)	FY2023A	FY2024E	FY2025E	FY2026E	净利率	16.19%	47.12%	50.17%	48.97%
税后经营利润	4418.15	27881.36	46910.92	57854.31	ROE	19.76%	55.99%	48.39%	37.38%
折旧与摊销	1544.00	1100.10	1356.76	1661.43	ROA	10.61%	40.27%	38.95%	31.75%
财务费用	-5.00	415.46	280.84	119.76	ROIC	23.17%	125.81%	149.16%	137.22%
其他经营资金	-316.15	-4227.69	-4876.90	-3117.65					
经营性现金净流量	5641.00	25169.23	43671.63	56517.85	估值				
投资性现金净流量	7375.00	-2521.15	-3519.15	-4222.95	P/E	263.32	40.91	24.43	19.81
筹资性现金净流量	-11617.00	-1665.46	-280.84	-119.76	P/S	42.64	19.27	12.26	9.70
现金流量净额	1399.00	20982.61	39871.64	52175.14	P/B	52.04	22.90	11.82	7.40

数据来源：公司公告，西南证券

分析师承诺

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，报告所采用的数据均来自合法合规渠道，分析逻辑基于分析师的职业理解，通过合理判断得出结论，独立、客观地出具本报告。分析师承诺不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接获取任何形式的补偿。

投资评级说明

报告中投资建议所涉及的评级分为公司评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 个月内的相对市场表现，即：以报告发布日后 6 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。其中：A 股市场以沪深 300 指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以恒生指数为基准；美国市场以纳斯达克综合指数或标普 500 指数为基准。

公司评级	买入：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅在 20% 以上
	持有：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅介于 10% 与 20% 之间
	中性：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅介于 -10% 与 10% 之间
	回避：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅介于 -20% 与 -10% 之间
	卖出：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅在 -20% 以下
行业评级	强于大市：未来 6 个月内，行业整体回报高于同期相关证券市场代表性指数 5% 以上
	跟随大市：未来 6 个月内，行业整体回报介于同期相关证券市场代表性指数 -5% 与 5% 之间
	弱于大市：未来 6 个月内，行业整体回报低于同期相关证券市场代表性指数 -5% 以下

重要声明

西南证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会核准的证券投资咨询业务资格。

本公司与作者在自身所知情范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

《证券期货投资者适当性管理办法》于 2017 年 7 月 1 日起正式实施，本报告仅供本公司签约客户使用，若您并非本公司签约客户，为控制投资风险，请取消接收、订阅或使用本报告中的任何信息。本公司也不会因接收人收到、阅读或关注自媒体推送本报告中的内容而视其为客户。本公司或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或争取提供投资银行或财务顾问服务。

本报告中的信息均来源于公开资料，本公司对这些信息的准确性、完整性或可靠性不作任何保证。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、价值及投资收入可升可跌，过往表现不应作为日后的表现依据。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，本公司不保证本报告所含信息保持在最新状态。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，投资者应当自行关注相应的更新或修改。

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。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，本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。

本报告及附录版权为西南证券所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用须注明出处为“西南证券”，且不得对本报告及附录进行有悖原意的引用、删节和修改。未经授权刊载或者转发本报告及附录的，本公司将保留向其追究法律责任的权利。

西南证券研究发展中心

上海

地址：上海市浦东新区陆家嘴 21 世纪大厦 10 楼

邮编：200120

北京

地址：北京市西城区金融大街 35 号国际企业大厦 A 座 8 楼

邮编：100033

深圳

地址：深圳市福田区益田路 6001 号太平金融大厦 22 楼

邮编：518038

重庆

地址：重庆市江北区金沙门路 32 号西南证券总部大楼 21 楼

邮编：400025

西南证券机构销售团队

区域	姓名	职务	座机	手机	邮箱
上海	蒋诗烽	总经理助理、销售总监	021-68415309	18621310081	jsf@swsc.com.cn
	崔露文	销售经理	15642960315	15642960315	clw@swsc.com.cn
	谭世泽	销售经理	13122900886	13122900886	tsz@swsc.com.cn
	薛世宇	销售经理	18502146429	18502146429	xsy@swsc.com.cn
	岑宇婷	销售经理	18616243268	18616243268	cyryf@swsc.com.cn
	汪艺	销售经理	13127920536	13127920536	wyyf@swsc.com.cn
	张玉梅	销售经理	18957157330	18957157330	zymf@swsc.com.cn
	李煜	销售经理	18801732511	18801732511	yfly@swsc.com.cn
	卞黎旸	销售经理	13262983309	13262983309	bly@swsc.com.cn
	龙思宇	销售经理	18062608256	18062608256	lsyu@swsc.com.cn
	田婧雯	销售经理	18817337408	18817337408	tjw@swsc.com.cn
	阙钰	销售经理	17275202601	17275202601	kyu@swsc.com.cn
北京	魏晓阳	销售经理	15026480118	15026480118	wxyang@swsc.com.cn
	李杨	销售总监	18601139362	18601139362	yfly@swsc.com.cn
	张岚	销售副总监	18601241803	18601241803	zhanglan@swsc.com.cn
	杨薇	高级销售经理	15652285702	15652285702	yangwei@swsc.com.cn
	王一菲	销售经理	18040060359	18040060359	wyf@swsc.com.cn
	姚航	销售经理	15652026677	15652026677	yhang@swsc.com.cn
	胡青璇	销售经理	18800123955	18800123955	hqx@swsc.com.cn

	张鑫	销售经理	15981953220	15981953220	zhxin@swsc.com.cn
	王宇飞	销售经理	18500981866	18500981866	wangyuf@swsc.com
	路漫天	销售经理	18610741553	18610741553	lmtyf@swsc.com.cn
	马冰竹	销售经理	13126590325	13126590325	mbz@swsc.com.cn
广深	郑龔	广深销售负责人	18825189744	18825189744	zhengyan@swsc.com.cn
	杨新意	销售经理	17628609919	17628609919	yxy@swsc.com.cn
	龚之涵	销售经理	15808001926	15808001926	gongzh@swsc.com.cn
	丁凡	销售经理	15559989681	15559989681	dingfyf@swsc.com.cn
	张文锋	销售经理	13642639789	13642639789	zwf@swsc.com.cn
	陈紫琳	销售经理	13266723634	13266723634	chzlyf@swsc.com.cn
	陈韵然	销售经理	18208801355	18208801355	cyryf@swsc.com.cn