

# Chat with RTX 推动模型本地部署，AIPC 内容生态持续丰富

——电子行业跟踪报告

强于大市 (维持)

2024 年 02 月 20 日

## 投资要点:

**英伟达推出 Chat with RTX，本地化部署为其重要特征:** 英伟达发布 AI 聊天机器人 Chat with RTX，用户可在搭载 NVIDIA GeForce RTX 30 系列或更高版本的 GPU（至少 8GB 显存）进行本地部署。Chat With RTX 可以将用户的资料与大语言模型建立连接，从而令大语言模型更具个性化。利用检索增强生成 (RAG)、TensorRT-LLM 和 RTX 加速，该应用具备本地文件及在线视频检索分析功能。同时，本地化部署有望解决数据隐私化的痛点，减少端侧应用对外部网络的依赖。

**AIPC 内容生态持续丰富，产品性能提升有望刺激换机需求:** 行业巨头积极推进 AIPC 应用布局，联想陆续推出了 AI NOW、AvatarMaster 等 AI 应用助手，微软使 Copilot 落地端侧键盘部署并持续更新功能，同时完善订阅会员模式。继微软推出 Copilot 及联想推出 AI 助手之后，英伟达发布本地化部署的 AI 聊天机器人 Chat with RTX，而该应用的本地化部署具备一定配置要求。随着模型参数量不断提升，对载体的算力及存力等均提出更高要求，AI 模型下沉端侧的过程亦会推动大众消费级 PC 产品的性能提升，亦将拉动换机需求，PC 换机主周期有望从 2024 年下半年开始。

**AIPC 行业将从“AI Ready”发展至“AI On”，拉动 PC 市场规模增长:** 目前 AIPC 行业仍处于 AI Ready 的准备阶段，该阶段产品基本具备了对 AI 任务更具针对性的本地混合算力。随着巨头积极推进 AIPC 应用布局，相关内容生态持续丰富，AIPC 将快速渗透至工作、游戏及创作等多个领域，此时 AIPC 行业的发展将进入 AI On 阶段，产品将能够服务于更广泛的通用场景，并且能够实现端边协同计算、跨设备互连接力，以及基于个人数据和使用历史，在边缘私域环境下实现个人大模型的微调训练。据 IDC，在 AIPC 带动下，PC 的应用场景将得到进一步拓展，拉动市场规模进入新一轮增长，中国 PC 市场有望在未来几年保持稳步增长态势，预计台式机、笔记本电脑市场总规模从 2023 年的 3900 万台增长至 2027 年的 5000 万台以上，增幅接近 28%。

**投资建议:** 继联想推出 AI NOW 等应用、微软对 Copilot 的持续更新之后，英伟达作为又一巨头入局 AIPC 应用端，发布个人 AI 助手的本地化部署 Chat with RTX，AIPC 生态内容持续丰富。Chat with RTX 的本地部署具备一定配置要求，推动大众消费级 PC 产品性能提升，同时 AIPC 产品性能提升有望刺激换机需求，加速 PC 换机周期，拉动 PC 市场规模进入新一轮增长。建议关注在 AIPC 领域前瞻布局的整机、芯片及应用厂商，及打入全球 PC 巨头供应链的零部件及代工企业。

**风险因素:** 终端复苏不及预期；行业竞争加剧；AIPC 芯片发展不及预期；AIPC 应用发展不及预期。

## 行业相对沪深 300 指数表现



数据来源: 聚源, 万联证券研究所

## 相关研究

英伟达发布 Chat with RTX，关注 AI 产业链投资机会

基金重仓比例升至高位，AI 及消费电子产业链受关注

英特尔与联电合作布局 12nm，全球 PCB 产值有望回升

分析师:

夏清莹

执业证书编号:

S0270520050001

电话:

075583223620

邮箱:

xiaqy1@wlzq.com.cn

研究助理:

陈达

电话:

13122771895

邮箱:

chenda@wlzq.com.cn

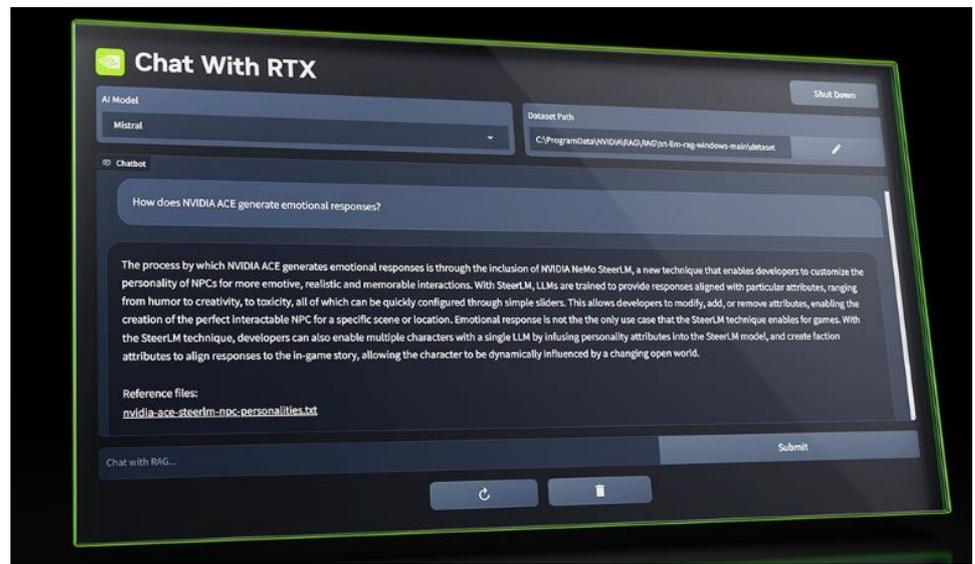
## 正文目录

1 英伟达推出 Chat with RTX，本地化部署为其重要特征.....	3
2 AIPC 内容生态持续丰富，产品性能提升有望刺激换机需求 .....	4
3 投资建议.....	7
4 风险因素.....	7
图表 1: 英伟达推出个性化 AI 聊天机器人 Chat with RTX.....	3
图表 2: 官方演示英伟达 CES2024 发布会及其链接.....	3
图表 3: 官方演示 Chat with RTX 视频检索功能.....	3
图表 4: 2024 年以来行业巨头积极推进 AIPC 应用布局 .....	4
图表 5: Chat with RTX 本地部署的系统要求.....	5
图表 6: 全球 PC 季度出货数据 .....	5
图表 7: AIPC 的渗透率有望在未来几年持续攀升 .....	6
图表 8: 中国个人电脑市场规模预测（单位：百万台） .....	6

## 1 英伟达推出 Chat with RTX，本地化部署为其重要特征

英伟达发布Chat with RTX，支持本地化个性化部署。英伟达发布适用于Windows的聊天机器人Chat with RTX，支持本地化个性化部署，丰富了端侧AI落地应用的形式。用户可在搭载NVIDIA GeForce RTX 30系列或更高版本的GPU（至少8GB显存）进行本地部署，打造个人专属聊天机器人。Chat With RTX可以将用户的资料（文档、笔记、视频或其他数据）与大语言模型建立连接，从而令大语言模型（LLM）更具个性化。利用检索增强生成（RAG）、TensorRT-LLM和RTX加速，用户可以与自定义聊天机器人对话。此外，由于Chat with RTX完全在Windows RTX PC或工作站本地以安全的方式运行，因此用户可以更快获得更满意的结果。

图表1: 英伟达推出个性化AI聊天机器人Chat with RTX



资料来源：英伟达官网，万联证券研究所

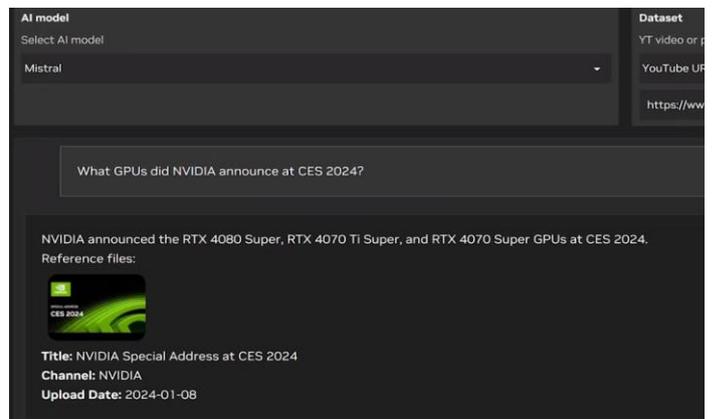
Chat with RTX具备本地文件及在线视频检索分析功能。1) 针对文本内容，Chat With RTX支持多种文件格式，包括文本文件、pdf、doc/docx和xml。用户只需在该应用中指定包含目标文件的文件夹，该应用便会在几秒内将目标文件加载到库中；2) 针对视频内容，用户可以提供YouTube播放列表的URL，然后该应用会自动加载播放列表中的视频的转写内容，让用户能够查询视频中包含的内容。

图表2: 官方演示英伟达 CES2024 发布会及其链接



资料来源：英伟达官网，万联证券研究所

图表3: 官方演示 Chat with RTX 视频检索功能



资料来源：英伟达官网，万联证券研究所

本地化部署有望解决数据隐私化的痛点，减少端侧应用对外部网络的依赖。1) 对于数据隐私化，终端设备处理器对AI引擎的集成使数据得以在本地完成处理和存储，极大增强了终端设备自身的独立性。该模式避免了云端或其它外部服务器对AI任务的参与，减少了外部不可控因素对AI应用的干扰。AI应用的本地化处理减少了个人暴露于云端的数据量，降低了数据在传输过程中被窃取或篡改的风险，分担了云端数据安全压力。2) 就端侧应用对外部网络的依赖而言，终端设备本地的AI算力使应用减少了对外部网络的依赖，即使在离线状态下依然可以进行工作，突破了网络方面对AI应用的工况限制。同时也规避了环境中网络连接状态对任务的影响，避免由于网络由于繁忙或安全方面的问题对数据处理造成影响，这对于一些对实时性要求较高的应用场景尤为重要。

## 2 AIPC 内容生态持续丰富，产品性能提升有望刺激换机需求

行业巨头积极推进AIPC应用布局，丰富AIPC内容生态。行业巨头积极入局，推进AIPC应用布局，联想陆续推出了AI NOW、AvatarMaster、Yoga Creator Zone等AI办公助手，微软使Copilot落地端侧键盘部署并持续更新功能，同时完善订阅会员模式、落地商业应用。继微软推出Copilot及联想推出AI助手之后，英伟达发布本地化部署的AI聊天机器人Chat with RTX，具备本地文件及在线视频检索分析功能。

图表4: 2024年以来行业巨头积极推进AIPC应用布局

推出时间	公司	应用更新	内容描述
1月	微软	Copilot	微软宣布在Windows PC键盘中添加Copilot键，该按键可以直接在Windows 10和Windows 11电脑中打开Copilot生成式AI助手。
1月	联想	AI NOW	以用户自建的设备知识库为基础，打造个性化互动体验。AI NOW将以更加个性化与沉浸式的方式简化工作流程，加强协作。通过使用自然语言，用户将能够检查和更改显示效果或性能等常见设置，搜索和汇总电子邮件及文档，创建会议邀请，并在视频会议期间混合使用实时摄像头和数字替身 (avatar)。
1月	联想	AvatarMaster	是一款全新的AI驱动应用程序，可根据用户个人资料生成3D数字化身，从外表和面部特征到服饰选择，均支持多种个性化功能。用户创建和定制虚拟分身后，可在视频会议、游戏通话以及多个平台之间以动画和流媒体形式展现自己。
1月	联想	Yoga Creator Zone	专为创作者、艺术家等用户设计的联想Yoga Creator Zone软件，助力使用者以简单、私密且安全的方式利用生成式AI。Yoga Creator Zone旨在为想象力赋能，提供图像生成工具，可将基于文本的描述或草图转换为惊艳的图像，无需输入任何复杂的提示、代码或设置。用户只需输入任何预想的内容，系统即可快速生成相应的视觉图像。
1月	微软	Copilot Pro	每月订阅费用为20美元，可解锁访问Word、Excel和PowerPoint等Office常用办公软件中的大模型能力，同时优先访问最新的OpenAI模型。在其之上，用户也可以基于需求构建起属于自己的Copilot GPT。
2月	微软	Copilot 功能更新	微软庆祝Copilot一周年，并推出了几项重大更新，包括全新界面、聊天中的图像编辑、“Designer GPT”集成等等，同时发布了即将到来的超级碗广告。
2月	英伟达	Chat with RTX	适用于Windows的聊天机器人with RTX，支持本地化个性化部署，具备本地文件及在线视频检索分析功能。

资料来源：联想集团，CSDN，机器之心，AI技术新动态，万联证券研究所

Chat with RTX的本地化部署具备一定配置要求，推动大众消费级PC产品性能提升。硬件配置方面，Chat with RTX需要在搭载NVIDIA GeForce RTX 30系列或更高版本的GPU(至少8GB显存)进行本地部署，同时RAM需求16GB或更高，且由于Chat with RTX目前支持Mistral及Meta的Llama 2模型，下载所有必要文件会占用约50GB-100GB的存储空间。操作系统及驱动方面，本地部署需要Windows 11操作系统及535.11或更新版本的驱动。随着模型参数量不断提升，对载体的算力及存力等均提出更高要求，AI模型下沉端侧的过程亦会推动大众消费级PC产品的性能提升。

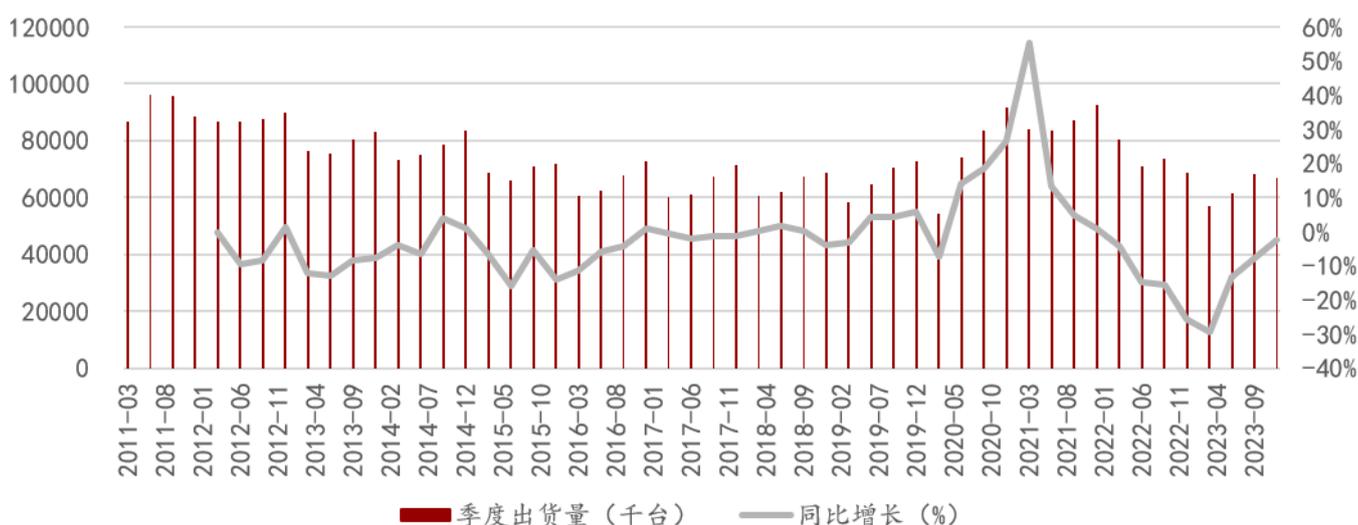
图表5: Chat with RTX 本地部署的系统要求

平台	Windows
GPU	NVIDIA GeForce® RTX 30 或 40 系列 GPU 或配备至少 8GB VRAM 的 NVIDIA RTX™ Ampere 或 Ada Generation GPU
RAM	16GB 或更高
操作系统	Windows 11
驱动	535.11 或更新版本驱动

资料来源: 英伟达官网, 万联证券研究所

PC换机周期将至, AIPC产品性能提升有望刺激换机需求。观察全球PC季度出货历史数据, PC的一轮换机周期为4-5年左右。上一轮换机高峰主要是受到宅经济发展的推动, 而后PC出货量逐渐下滑, 回落至近年来较低区间, 并于2023年下半年企稳回升。随着AIPC产品性能不断提升, 将拉动换机需求, PC换机主周期有望从2024年下半年开始。

图表6: 全球 PC 季度出货数据

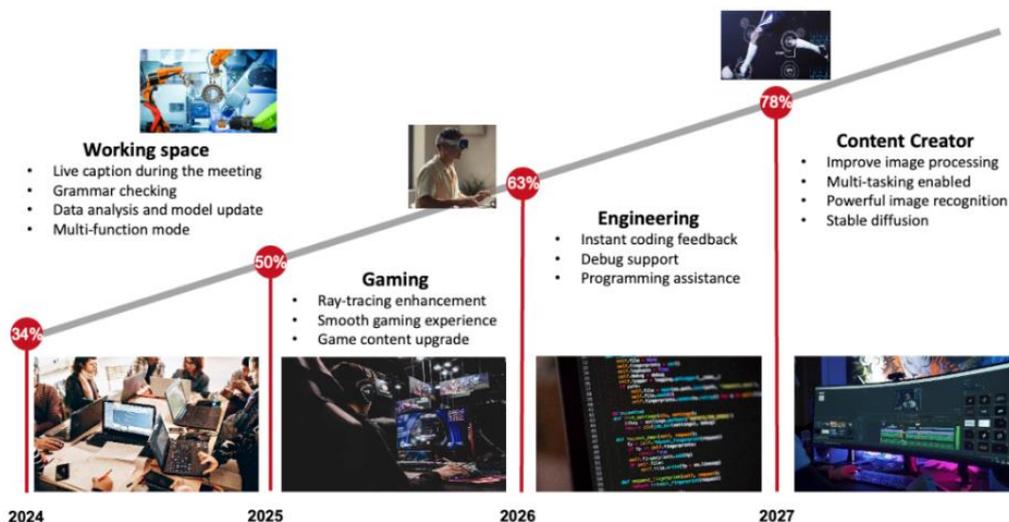


资料来源: iFinD, 万联证券研究所

AIPC行业将从“AI Ready”发展至“AI On”,有效提升个人生产力。目前AIPC行业仍处于AI Ready的准备阶段, 该阶段产品基于对AI任务更具针对性的本地混合算力, 能够为AIPC的软件及服务创新提供基本保障。随着行业巨头积极推进AIPC应用布局, AIPC内容生态持续丰富, AIPC将快速渗透至工作、游戏、工程及创作等多个领域, 成为个人AI助理, 此时AIPC行业的发展将进入AI On阶段, AIPC产品将能够服务

于更广泛的通用场景，并且能够实现端边协同计算、跨设备互联接力，以及基于个人数据和使用历史，在边缘私域环境下实现个人大模型的微调训练。Counterpoint Research预计到2025年，AI笔记本电脑的渗透率有望达到50%。

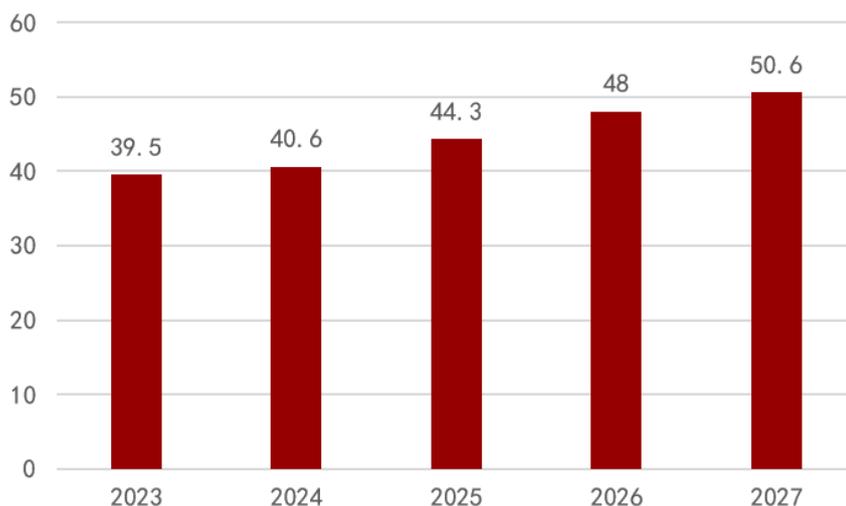
图表7: AIPC 的渗透率有望在未来几年持续攀升



资料来源: Counterpoint Research, 万联证券研究所

AIPC有望拉动PC市场规模进入新一轮增长。2023年的企业支出放缓和去库存进程将带动PC厂商的库存补充需求，并成为企业PC市场的推动力。据IDC，在AIPC带动下，PC的应用场景将得到进一步拓展，拉动市场规模进入新一轮增长，中国PC市场有望在未来几年保持稳步增长态势，预计台式机、笔记本电脑市场总规模从2023年的3900万台增长至2027年的5000万台以上，增幅接近28%。

图表8: 中国个人电脑市场规模预测 (单位: 百万台)



资料来源: 2023年AIPC产业(中国)白皮书, 万联证券研究所

### 3 投资建议

继联想推出AI NOW等系列应用，以及微软对Copilot的持续更新，英伟达作为又一巨头，入局AIPC应用端，发布个人AI助手的本地化部署Chat with RTX，AIPC生态内容持续丰富。Chat with RTX的本地化部署具备一定配置要求，推动大众消费级PC产品性能提升，同时AIPC产品性能提升有望刺激换机需求，加速PC换机周期，拉动PC市场规模进入新一轮增长。建议关注在AIPC领域前瞻布局的整机、芯片及应用厂商，及打入全球PC巨头供应链的零部件及代工企业。

### 4 风险因素

终端复苏不及预期；行业竞争加剧；AIPC芯片发展不及预期；AIPC应用发展不及预期。

## 行业投资评级

强于大市：未来6个月内行业指数相对大盘涨幅10%以上；

同步大市：未来6个月内行业指数相对大盘涨幅10%至-10%之间；

弱于大市：未来6个月内行业指数相对大盘跌幅10%以上。

## 公司投资评级

买入：未来6个月内公司相对大盘涨幅15%以上；

增持：未来6个月内公司相对大盘涨幅5%至15%；

观望：未来6个月内公司相对大盘涨幅-5%至5%；

卖出：未来6个月内公司相对大盘跌幅5%以上。

基准指数：沪深300指数

## 风险提示

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。

## 证券分析师承诺

本人具有中国证券业协会授予的证券投资咨询执业资格并登记为证券分析师，以勤勉的执业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，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收到任何形式的补偿。

## 免责声明

万联证券股份有限公司（以下简称“本公司”）是一家覆盖证券经纪、投资银行、投资管理和证券咨询等多项业务的全国性综合类证券公司。本公司具有中国证监会许可的证券投资咨询业务资格。

本报告仅供本公司的客户使用。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。本报告中的信息或所表述的意见并未考虑到个别投资者的具体投资目的、财务状况以及特定需求。客户应自主作出投资决策并自行承担投资风险。本公司不对任何人因使用本报告中的内容所导致的损失负任何责任。在法律许可情况下，本公司或其关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或类似的金融服务。

市场有风险，投资需谨慎。本报告是基于本公司认为可靠且已公开的信息撰写，本公司力求但不保证这些信息的准确性及完整性，也不保证文中的观点或陈述不会发生任何变更。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。分析师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。

本报告的版权仅为本公司所有，未经书面许可任何机构和个人不得以任何形式翻版、复制、刊登、发表和引用。未经我方许可而引用、刊发或转载的引起法律后果和造成我公司经济损失的概由对方承担，我公司保留追究的权利。

## 万联证券股份有限公司 研究所

上海浦东新区世纪大道 1528 号陆家嘴基金大厦

北京西城区平安里西大街 28 号中海国际中心

深圳福田区深南大道 2007 号金地中心

广州天河区珠江东路 11 号高德置地广场