

## 计算机行业点评报告

# CPU 涨价，CPU 有望在 Agent 时代迎来大周期

增持（维持）

2026 年 02 月 01 日

证券分析师 王紫敬

执业证书：S0600521080005  
021-60199781

wangzj@dwzq.com.cn

证券分析师 王世杰

执业证书：S0600523080004  
wangshijie@dwzq.com.cn

### 投资要点

- **CPU 供应不足，供应商预计将上调价格：**受超大规模云厂商需求驱动，英特尔与 AMD 计划于 2026 年上调服务器 CPU 价格 10%-15%。为确保利润，英特尔将核心先进制程产能大规模从消费端转向服务器，导致全球 PC 交付保证率大幅下滑，低端 PC 市场将面临供应不足。
- **AI 热潮抬升原材料，CPU 成本上升：**DRAM 生产转向 HBM，消耗更多晶圆；与 NAND 需求攀升、交货期延长，库存告急，挤占 CPU 晶圆材料供给。CPU 部件 PCB 应用及加工材质的转变，使得钻针使用寿命缩短，消耗量暴增；CCL 采用的树脂体系、玻纤布与铜箔匹配复杂，新进入者良率提升缓慢与客户认证周期长，导致有效产能释放缓慢；二者纷纷涨价，带动 CPU 价格上涨。
- **AI 推理和 Agent 发展迅速，拉动高端多核 CPU 需求：**CPU 负载正从“人类节奏”转向“机器节奏”。Agentic AI 远不是单次推理，而是动态推理+多步决策+外部工具调用的循环，这比传统大模型调用更耗资源、负载更复杂、成本更高。这种资源调用增长，加之为了安全防范而产生的高频沙箱隔离开销，使得 CPU 资源消耗呈现指数级放大。Deepseek 提出 Engram 模块，提升 CPU 用量。
- **CPU 产能从消费级向服务器级转移：**在英特尔 2025 年第四季度电话会议上，公司财务官表示“正在尽可能将产能转向数据中心以满足强劲需求”，表明在供应受限情况下英特尔优先满足服务器端需求。KeyBanc 等机构报告也指出，2026 年服务器 CPU 产能已基本售罄，公司将部分产能从 PC 端调配到服务器 CPU 以缓解供不应求。AMD 在 2025-2026 年的服务器 CPU 市场份额快速增长，表明其正在将资源与产线重心从传统消费端向服务器、数据中心倾斜。
- **投资建议：**AI Agent 落地速度正逐渐加速，其带来的沙箱创建、负载卸载等需求有望大幅拉动 CPU 的需求。但同时，全球算力供应链产能紧张，多个环节涨价，使得 CPU 制造成本增加。两方面影响下，我们认为 CPU 产业有望迎来大周期。
- **相关标的：**CPU：澜起科技、海光信息、广合科技、龙芯中科、中国长城等。数据库：星环科技（基于 ARM 优化，与 NV-GPU-GraceCPU 适配）。
- **风险提示：**Agent 落地进展不及预期；产能供应问题缓解。

### 行业走势



### 相关研究

《商业航天为何会成为 2026 年的主线？》

2026-01-28

《迎接“数据要素价值释放年”》

2026-01-18

## 内容目录

1. CPU 供给紧张，供应商上调价格 .....	4
2. 涨价背后的原因：AI Agent 即将爆发 .....	4
2.1. AI 热潮抬升原材料，CPU 成本上升 .....	4
2.2. AI 推理和 Agent 发展，拉动高端多核 CPU 需求 .....	5
2.3. Deepseek 提出 Engram 模块，提升 CPU 用量 .....	8
2.4. CPU 产能从消费级向服务器级转移 .....	8
3. 投资建议 .....	9
4. 风险提示 .....	9

## 图表目录

图 1: 英特尔 25 年财报 (分部门) .....	4
图 2: agentic AI 结构 .....	5
图 3: 进化后的 Agentic AI 与现有 AI 评分对比 .....	5
图 4: 五大 agentic AI 的运行时延主要构成 .....	6
图 5: 在大规模数据集情况下, CPU 的动态能耗占比变得相当显著 (占 44%) .....	6
图 6: 运行沙箱时的系统调用延迟测量 .....	8

## 1. CPU 供给紧张，供应商上调价格

**英特尔将晶圆供给转向数据中心和人工智能服务器。**1月22日，据路透社消息，英特尔表示难以满足其用于人工智能数据中心的服务器芯片需求。23日发布的财报显示，数据中心和人工智能（DCAI）产品季度增长9%，全年增长5%，而其客户端计算事业部（CCG,销售 Core 处理器、Arc GPU 及其他消费产品）季度下降7%，全年下降3%。出于利润考虑，英特尔正在将其内部晶圆供应转向盈利更高的 DCAI 部门，而营收相对不如意的 CCG 部门则依靠外部晶圆供应。

图1：英特尔 25 年财报（分部门）

业务部门收入及趋势	2025 第四季度 <sup>1</sup>	相比2024 第四季度	2025 <sup>1</sup>	相比2024
<b>英特尔产品：</b>				
客户端计算事业部（CCG）	\$82亿	下降7%	\$322亿	下降3%
数据中心和人工智能事业部（DCAI）	\$47亿	上升9%	\$169亿	上升5%
英特尔产品总收入	\$129亿	下降1%	\$491亿	下降1%
英特尔代工	\$45亿	上升4%	\$178亿	上升3%
其他所有	\$6亿	下降48%	\$36亿	下降1%
部门间抵消	\$(43)亿		\$(177)亿	
净总收入	\$137亿	下降4%	\$529亿	持平

数据来源：英特尔公众号，东吴证券研究所

**英特尔 18A 芯片良率提升缓慢，影响 CPU 出货量。**18A 工艺引入了新的晶体管设计和更高效的电力传输方式，但迄今仅交付了一小部分符合英特尔质量标准的芯片。2025 年夏天的报告显示，18A 生产线上出厂的芯片中，只有 10%符合英特尔的需求。英特尔称良率每个月提升 7%-8%，CEO 陈立武表示，“虽然良率提升速度符合内部计划，但仍低于期望水平。”据英特尔首席财务官称，良率预计将在 2026 年底达到英特尔期望的成本水平，2027 年将达到行业标准。

**服务器 CPU 预计涨价。**Trendforce 预计 2026 年全服务器出货年成长为 12.8%，A 服务器出货年增 28%以上，因此 CPU 供需紧张，将推动相关芯片涨价。英特尔投资者加贝利基金的分析师米卡托·马基诺表示：“到 2026 年，服务器 CPU 的价格上涨幅度至少达两位数。”同时，Keybanc 数据显示，由于超大规模云服务商大幅采购，英特尔与 AMD 在 2026 全年的服务器 CPU 产能已基本售罄。为了应对供需极端失衡并确保后续供应稳定，两家公司均计划将服务器 CPU 价格上调 10-15%。

## 2. 涨价背后的原因：AI Agent 即将爆发

### 2.1. AI 热潮抬升原材料，CPU 成本上升

CPU 部件紧缺，带动价格上涨。

**CPU 部件 PCB、CCL 产能不足。**沙利文大中华区执行总监谢书勤表示，AI 服务器等领域 PCB 高多层板的应用及加工材质的转变，使得钻针使用寿命缩短，消耗量暴增，估计在原来的基础上增长了四五倍。高阶 CCL 采用的树脂体系、玻纤布与铜箔匹配复杂，新进入者良率提升缓慢与客户认证周期长，导致有效产能释放缓慢。叠加上游超薄铜箔与高端玻纤布产能有限，形成阶段性紧缺格局。

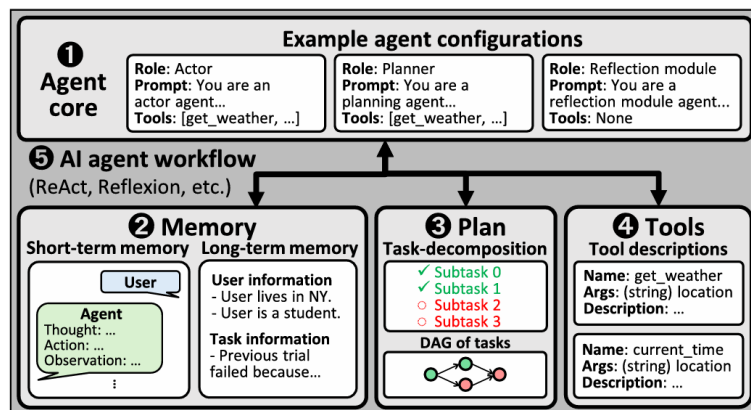
**PCB、CCL 涨价，带动 CPU 价格上升。**因玻纤布等原料供应紧张、价格飙升，日本半导体材料厂 Resonac 宣布自 3 月 1 日起调涨铜箔基板（CCL）、黏合胶片等印刷电路板（PCB）材料售价、涨幅达 30%以上。去年底，建滔集团也向客户发涨价函，表示原物料成本已难以吸收，新接单 CCL 价格全面调涨 10%。

## 2.2. AI 推理和 Agent 发展，拉动高端多核 CPU 需求

**Agentic AI 对 CPU 的影响不是线性的，而是乘数级的。**

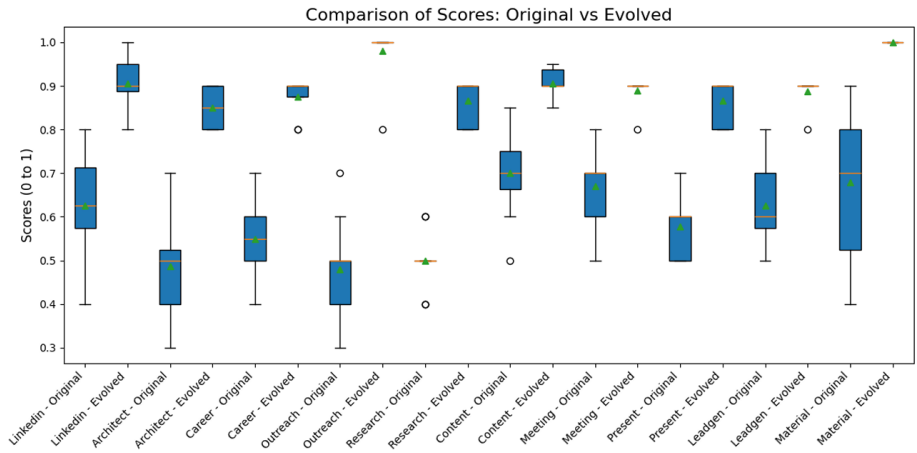
Agentic AI 与当前的 AI 主要区别在于，Agentic AI 能够通过自主生成和测试假设来改进系统配置，从而在无需人工干预的情况下实现最佳性能。

图2: agentic AI 结构



数据来源：《The Cost of Dynamic Reasoning: Demystifying AI Agents and Test-Time Scaling from an AI Infrastructure Perspective》，东吴证券研究所

图3: 进化后的 Agentic AI 与现有 AI 评分对比



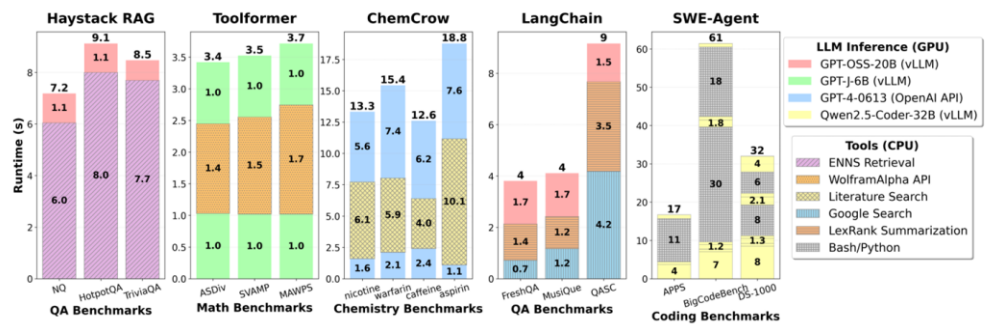
数据来源：《A Multi-AI Agent System for Autonomous Optimization of Agentic AI Solutions via Iterative Refinement and LLM-Driven Feedback Loops》，东吴证券研究所

**Agentic AI 工作会大幅提升 CPU 负载。**

Agentic AI 远不是“单次推理”，而是动态推理+多步决策+外部工具调用的循环，这比传统大模型调用更耗资源、负载更复杂、成本更高。动态决策结构使得资源消耗与任务长度呈非线性关系增长，即：如果 Agent 需要更多步骤、更多内部查询/操作，则计算资源（特别是 CPU 和能耗）会成倍增长。

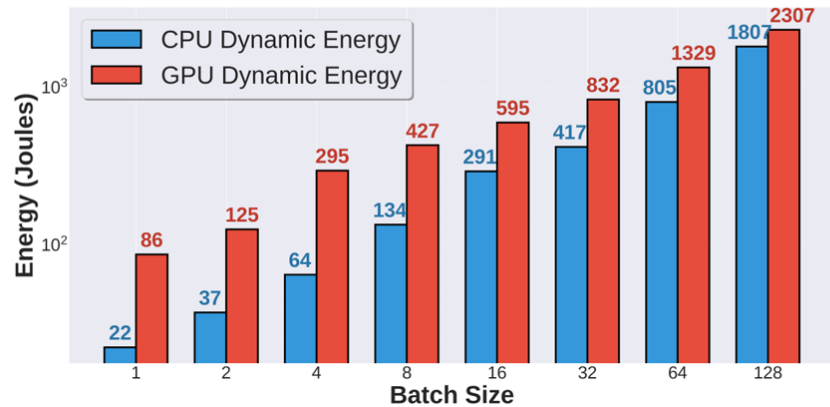
Agentic AI 工作负载很大程度上依赖 CPU。智能体框架通常包含决策编排器、工具调用、外部 API、数据库访问等复杂逻辑，这些操作往往在 CPU 上执行。在多种典型 agentic workload（如 Toolformer、Langchain、ChemCrow 等）中，在 CPU 上处理工具调用可能占总延迟的 84.5%-90.6%。CPU 动态能耗在大批量 agent 处理中可占总能耗的近 44%，说明工作负载对 CPU 的依赖度极高。

图4：五大 agentic AI 的运行延迟主要构成



数据来源：《A CPU-Centric Perspective on Agentic AI》，东吴证券研究所

图5：在大规模数据集情况下，CPU 的动态能耗占比变得相当显著（占 44%）



数据来源：《A CPU-Centric Perspective on Agentic AI》，东吴证券研究所

### 沙箱运行带来 CPU 荷载指数爆发。

**Agentic AI 运行过程中需要沙箱保护电脑系统。**沙箱是一种安全隔离策略和模型，其核心目标是在一个受限的、封闭的环境中运行不受信任的程序或代码。OpenAI、Anthropic、Microsoft 等头部大厂均在操作指引文档里提醒用户在运行 Agentic AI 时需要创建沙箱来保护电脑。如果在运行 Agentic AI 时不创建沙箱，则可能面临多种风险。

**1.基础设施安全失控风险。**AI 生成的代码存在植入高危系统指令(如删除根目录 `rm -rf /*`)、恶意系统调用或无限资源耗尽循环的潜在可能。在传统物理服务环境中，此类代码将直接攻击主机操作系统 CPU、内存等基础设施，导致数据损失等后果。

**2.跨环境执行失配风险。**AI 生成的代码通常深度耦合特定的运行时库、依赖版本及系统配置。当代码迁移时，将面临底层 CPU 架构、操作系统内核及系统环境的显著差异。这种在本地运行成功但在他处失败的问题是 CPU 系统执行上下文的不一致性导致，严重阻碍了部署的确定性与可靠性。

**3.并发资源干扰与隔离失效风险。**在多 AI 代理 (Agent) 任务高并发执行的场景下，若运行时隔离机制不彻底，单个任务中的缺陷将通过共享操作系统内核进行资源争夺。这种干扰会劫持其他良性任务乃至关键系统服务的 CPU 时间片与内存资源，可能引发级联故障，导致整体服务性能雪崩式下降。

**Agentic AI 中沙箱会显著提升 CPU 工作量，增加系统时延。**有研究对 `fork()`、`execve()`、`exit()` 等系统调用的延迟进行测量，结果显示，在沙箱机制下，系统调用的延迟明显提高(每次 `syscall` 增加数微秒开销)：`fork()` 增加约  $6.8 \mu\text{s}$ ；`execve()` 增加约  $1.2 \mu\text{s}$ ；`exit()` 增加约  $5.9 \mu\text{s}$ 。这在大量频繁短生命周期任务(例如 agent 调用多个工具/脚本)场景下可被放大为显著 CPU 工作量。另有研究表明函数执行的冷启动延迟在大量并发沙箱启动时会严重拖慢整体性能，说明隔离机制对 CPU 调度和资源管理存在高开销。DecentJS 的 proxy-based sandbox 在 JavaScript 基准测试中表现出  $8\times$  至  $32\times$  的执行开销，远高于原生执行。

图6: 运行沙箱时的系统调用延迟测量

	fork()	execve()	exit()	wait()
<b>total latency (microseconds)</b>	169	375	145	-
<b>overhead (microseconds)</b>	6.8	1.2	5.9	11.2
<b>overhead (% of total)</b>	4.0	0.3	4.1	-

数据来源:《A Flexible Containment Mechanism for Executing Untrusted Code》, 东吴证券研究所

### 2.3. Deepseek 提出 Engram 模块, 提升 CPU 用量

1月12日, DeepSeek 联合北京大学发布《Conditional Memory via Scalable Lookup》, 推出 Engram 模块, 可以用 CPU 内存部分替代 GPU 显存作为运行载体。

在现有的架构中, 模型记忆知识的方式是隐式的。当模型需要回答“法国的首都是哪里”时, 它无法像人类查阅字典那样直接检索出“巴黎”这个词, 而是需要通过多层注意力机制 (Attention) 和前馈神经网络 (FFN) 进行复杂的矩阵运算, 实际上是在通过计算来模拟检索过程。

这种机制造成了巨大的资源浪费: 模型不得不动用宝贵的推理算力去“死记硬背”大量固定的事实性知识 (如人名、地名、固定搭配等), 这不仅效率低下, 还挤占了模型处理复杂逻辑推理的“脑容量”。

针对这一点, DeepSeek 提出了“条件记忆”的概念, 并设计了实体模块 Engram 来加以实现。Engram 的设计灵感源自自然语言处理 (NLP) 领域经典的 N-gram (N 元语法) 模型, 但在深度学习语境下进行了现代化改造。传统的 N-gram 依靠统计词汇共现频率来预测下一个词, 而 Engram 则将其转化为一种可学习的、基于哈希 (Hash) 的向量检索系统。

DeepSeek 在实验中成功演示了将一个高达 1,000 亿参数的 Engram 嵌入表完全存储在廉价的主机内存 (CPU DRAM) 中, 而非昂贵的 GPU 显存里。在模型计算前序层级时, 系统通过 PCIe 通道异步地将所需的记忆数据搬运至 GPU。

Engram 模块推广后, 预计将进一步提升 CPU 需求。

### 2.4. CPU 产能从消费级向服务器级转移

在英特尔 2025 年第四季度电话会议上, 公司财务官表示“正在尽可能将产能转向

数据中心以满足强劲需求”，表明在供应受限情况下英特尔优先满足服务器端需求。KeyBanc 等机构报告也指出，2026 年服务器 CPU 产能已基本售罄，公司将部分产能从 PC 端调配到服务器 CPU 以缓解供不应求。AMD 在 2025-2026 年的服务器 CPU 市场份额快速增长，表明其正在将资源与产线重心从传统消费端向服务器、数据中心倾斜。

### 3. 投资建议

AI Agent 落地速度正逐渐加速，其带来的沙箱创建、负载卸载等需求有望大幅拉动 CPU 的需求。但同时，全球算力供应链产能紧张，多个环节涨价，使得 CPU 制造成本增加。两方面影响下，我们认为 CPU 产业有望迎来大周期。

相关标的：

**CPU：**澜起科技、海光信息、广合科技、龙芯中科、中国长城等。

**数据库：**星环科技（基于 ARM 优化，与 NV-GPU-GraceCPU 适配）。

### 4. 风险提示

**Agent 落地进展不及预期。**Agent 落地中创建沙箱的需求是导致 CPU 涨价的重要原因，如果 Agent 发展不及预期，会影响 CPU 的需求和涨价节奏。

**产能供应问题缓解。**如果下游晶圆厂产能大幅扩产或者良率提升，会影响 CPU 涨价的节奏。

## 免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司（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，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。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。

在法律许可的情况下，东吴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。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中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。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

## 东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证 50 指数），具体如下：

公司投资评级：

- 买入：预期未来 6 个月个股涨跌幅相对基准在 15% 以上；
- 增持：预期未来 6 个月个股涨跌幅相对基准介于 5% 与 15% 之间；
- 中性：预期未来 6 个月个股涨跌幅相对基准介于 -5% 与 5% 之间；
- 减持：预期未来 6 个月个股涨跌幅相对基准介于 -15% 与 -5% 之间；
- 卖出：预期未来 6 个月个股涨跌幅相对基准在 -15% 以下。

行业投资评级：

- 增持：预期未来 6 个月内，行业指数相对强于基准 5% 以上；
- 中性：预期未来 6 个月内，行业指数相对基准 -5% 与 5%；
- 减持：预期未来 6 个月内，行业指数相对弱于基准 5% 以上。

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应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所  
苏州工业园区星阳街 5 号  
邮政编码：215021

传真：（0512）62938527

公司网址：<http://www.dwzq.com.cn>