

软通动力（301236）

软通动力：软硬算一体化平台成型，Token 工厂打开算力运营新空间

买入（首次）

2026 年 08 月 25 日

证券分析师 王紫敬

执业证书：S0600521080005

021-60199781

wangzj@dwzq.com.cn

盈利预测与估值	2024A	2025A	2026E	2027E	2028E
营业总收入（百万元）	31,316	35,090	39,879	45,477	51,387
同比（%）	78.13	12.05	13.65	14.04	13.00
归母净利润（百万元）	180.38	206.11	295.68	559.12	982.38
同比（%）	(66.21)	14.27	43.45	89.10	75.70
EPS-最新摊薄（元/股）	0.17	0.20	0.29	0.54	0.95
P/E（现价&最新摊薄）	213.38	186.73	130.17	68.84	39.18

投资要点

- 软通动力为国内领先的全栈智能化产品与服务提供商，由传统 IT 服务商向“软件、硬件、算力、模型、应用”一体化平台迈进。公司长期深耕 ICT、互联网、运营商、金融、高科技与制造等行业；2024 年完成同方计算机及智通国际相关业务整合，业务形态由人力密集型软件服务向软件服务、计算硬件和智算运营三位一体转变。
- AI 应用由模型训练向规模化推理扩散。截至 2025 年 6 月，我国计算设备智能算力规模按 FP32 口径达到 782EFlops，同比增长 96%，智能算力已成为新增算力的主要组成部分。伴随智能体、长上下文、多轮交互和工具调用逐步进入生产环境，算力需求由阶段性预训练向持续性推理扩散，算力服务的评价指标也将由理论 P 值、卡数和峰值吞吐，逐步转向单位 Token 成本、首字延迟、并发能力、稳定性及服务等级保障。
- 公司现有算力规模超 1800P，怀来智算中心一期拟新增 3316.15P 训练与推理算力，项目建成后理论算力规模有望超过 5100P。怀来项目平均测算毛利率为 38.03%、税后内部收益率为 10.14%；公司已接洽的潜在客户预计算力需求约 1300P—1600P，占项目新增产能约 39%—48%。
- 北京壹号词元工厂投产，标志着公司由传统卡时租赁进一步升级为标准化 Token 产能运营。2026 年 6 月 9 日，公司“词元工厂计划”首个标杆项目北京壹号词元工厂在京点亮，一期每日 Token 产能达到 1.4 万亿，并开源面向智能体长时运行场景的词元工厂性能基准及 LoadGen 2.0 评测框架。项目通过算力调度、KV Cache 复用、全栈观测和推理工程优化，将底层硬件转化为具备明确服务质量、可弹性扩展的 Token 供给。
- 头部大模型厂商正式签约，算力业务由资源建设进入真实负载验证和商业运营阶段。公司与某国内头部大模型厂商签署智算服务协议，为其提供基于北京壹号词元工厂的 Token 推理服务，覆盖大模型推理加速、高性能集群适配、差异化 SLA 保障及行业 AI 应用落地。模型厂商将真实推理业务部署于词元工厂，能够直接提高算力消纳确定性，并验证公司在长链路推理、多轮交互和高并发环境下的调度与缓存管理能力。
- 盈利预测与投资建议：我们预计公司 2026—2028 年归母净利润分别为 2.96、5.59 和 9.82 亿元，对应 2026-2028 年 PE 分别为 130x、69x、39x。在公司相关业务增长及盈利改善逐步兑现的前提下，可在可比公司估值基础上给予一定的成长性及全栈能力溢价，首次覆盖给予“买入”评级。
- 风险提示：智算项目建设进度不及预期，算力客户及利用率不及预期，Token 服务合同经济性不及预期，Token 产能与实际销量存在差异，算力租赁价格下降，硬件业务竞争加剧，存货及现金流风险，芯片供应与技术迭代风险，异构适配和平台能力不及预期，AI 投入回报不及预期。

股价走势



市场数据

收盘价(元)	37.26
一年最低/最高价	32.88/65.50
市净率(倍)	2.89
流通 A 股市值(百万元)	29,275.53
总市值(百万元)	38,571.15

基础数据

每股净资产(元,LF)	12.89
资产负债率(% ,LF)	58.40
总股本(百万股)	1,035.19
流通 A 股(百万股)	785.71

相关研究

- 《软通动力(301236)：2024 年三季度点评：Q3 业绩亮眼，鸿蒙产业加速》
2024-10-27
- 《软通动力(301236)：2024 年中报点评：明确新业务战略，华为产业迎来催化》
2024-09-01

内容目录

1. 软通动力：IT 服务龙头，软硬一体迈向全栈智能	4
1.1. IT 服务龙头，战略并购完成软硬一体布局	4
1.2. 计算硬件驱动收入扩张，盈利能力仍处转型底部	5
2. 算力租赁加速升级，AI PC 进入规模渗透期	6
2.1. 推理需求持续扩张&算力供给模式升级，智算运营行业 β 向上	6
2.1.1. 算力租赁连接芯片供给与 AI 应用，商业模式持续升级	6
2.1.2. AI 需求由训练向推理扩散，算力消耗由项目制转向持续运营	7
2.1.3. 供给快速增加，利用率与真实客户成为行业核心壁垒	7
2.2. AI PC 进入规模渗透阶段，产品形态由 NPU 加速向端侧智能体演进	8
3. 成长性分析：资源力+产品力+客户力，智算运营进入兑现期	11
3.1. 现有算力超过 1800P，怀来项目打开规模扩张空间	11
3.1.1. 全国节点初步形成，算力业务由建设向运营延伸	11
3.1.2. 怀来 3316P 项目扩产，算力规模有望接近现有三倍	11
3.2. 服务器+调度平台协同，构筑算力运营底座	11
3.2.1. 服务器产品能力完善，掌握算力基础设施入口	11
3.2.2. 天元平台解决异构调度，软件能力提高有效算力产出	12
3.3. 北京壹号词元工厂投产，算力租赁向 Token 生产升级	12
3.3.1. 一期日产能 1.4 万亿 Token，形成首个标准化推理工厂	12
3.3.2. 开源性能基准，尝试建立 Token 工厂评价体系	12
3.4. 头部模型厂商签约，真实负载验证商业模式	13
3.4.1. 模型厂商部署推理服务，词元工厂进入商业运营阶段	13
3.4.2. 客户签约提高利用率确定性，但合同经济性仍待披露	13
3.4.3. “模型厂商造脑、词元工厂生产 Token”，形成产业专业化分工	14
3.5. 客户+资源+平台三力协同，形成区别于纯算力租赁商的优势	14
3.5.1. 存量大客户id提供算力需求入口	14
3.5.2. 国产与国际算力异构布局，覆盖差异化需求	14
3.6. 智算业务收入占比较低，毛利贡献有望快速提升	14
4. 盈利预测与投资评级	15
4.1. 盈利预测	15
4.1.1. 软件与数字技术服务	15
4.1.2. 计算产品与智能电子	15
4.1.3. 数字能源与智算服务	15
4.2. 估值分析	16
4.3. 投资建议	17
5. 风险提示	17

图表目录

图 1: 软通动力构建软硬一体化新发展模式.....4

图 2: 2021-2026 年公司营业收入变化 5

图 3: 2021-2026 年公司归母净利润变化 5

图 4: 北京壹号词元工厂集成了算力调度和 KV Cache 复用算法 7

图 5: 算力调度与经营管理平台架构..... 8

图 6: Windows 11 AI+ PC 9

图 7: 全球 AI PC 出货量以及市场份额 9

图 8: Windows ML 堆栈图 10

表 1: 算力租赁服务类型.....6

表 2: 公司业务拆分预测表..... 16

表 3: 可比公司 PE 估值表..... 17

1. 软通动力：IT 服务龙头，软硬一体迈向全栈智能

1.1. IT 服务龙头，战略并购完成软硬一体布局

软通动力深耕软件和信息技术服务二十余年，完成计算机业务并购后，实现由软件服务商向全栈计算平台的关键跨越。公司成立于 2005 年，早期主要为大型企业提供咨询、软件开发、数字技术服务和数字化运营，逐步形成覆盖 ICT、金融、互联网与运营商、高科技与制造等行业的大客户体系。2022 年公司登陆创业板，2024 年完成同方计算机及智通国际相关业务整合，获得国产 PC、消费级 PC、服务器、AI 工作站、智能终端和海外渠道能力，为业务发展开创新的增长空间。2025 年坚定推进“软硬一体全栈智能”战略，加速出海打造增长新引擎。

并购前，公司收入主要来自软件与数字技术服务，商业模式具有轻资产、人员密集和项目制交付等特征；并购后，公司收入规模快速跨越 300 亿元，同时进入计算硬件和智算基础设施市场。公司不再仅承接客户数字化项目，而是能够同时提供服务器及终端设备、基础软件、智算集群、模型平台和行业应用，单一客户的可覆盖价值量显著扩大。

图1：软通动力构建软硬一体化新发展模式



数据来源：软通动力官网，东吴证券研究所

公司当前形成软件与数字技术服务、计算产品与智能电子、数字能源与智算服务三大核心业务，持续打造中国领先的全栈智能化产品与服务能力。三类业务并非简单并列：软件业务负责进入客户业务流程、识别 AI 需求并交付应用；计算产品业务提供服务器、AI PC、工作站和智能终端等硬件载体；数字能源与智算服务负责算力中心建设、算力调度、算力订阅及 Token 供给，由此形成“软件获取需求—硬件建设底座—算力持续运营—应用反向拉动调用”的闭环。

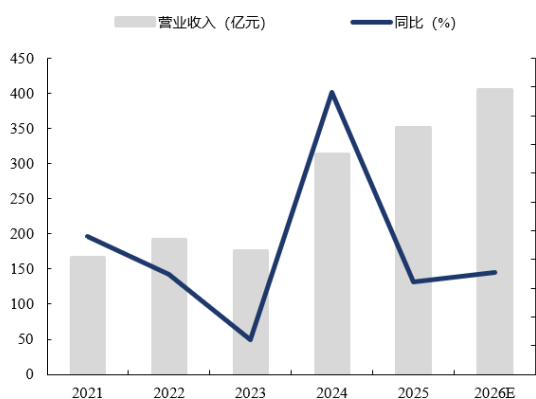
智算业务成为公司第二成长曲线。2025 年公司 AI 相关业务收入达到 184.66 亿元，占营业收入 52.6%；2026 年一季度 AI 相关业务收入 45.14 亿元，同比增长 39%。公司

聚焦智算运营核心赛道，是国内首批“调度平台+基建+生态”的一体化运营服务商。随着公司北京壹号词元工厂投产、与头部大模型厂商签订算力供应协议，智算服务与算力租赁业务将持续赋能公司盈利水平升级。

1.2. 计算硬件驱动收入扩张，盈利能力仍处转型底部

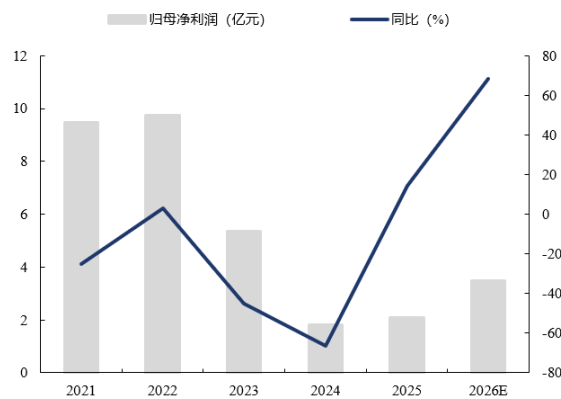
战略并购推动公司收入增长，但低毛利硬件业务占比提升，使利润率明显低于并购前水平。2023—2025 年，公司分别实现营业收入 175.81、313.16 和 350.90 亿元，2024 年收入同比增长 78.13%，主要来自计算机业务并表；同期归母净利润分别为 5.34、1.80 和 2.06 亿元，2025 年归母净利率仅 0.59%。

图2：2021-2026 年公司营业收入变化



数据来源：iFind，东吴证券研究所

图3：2021-2026 年公司归母净利润变化



数据来源：iFind，东吴证券研究所

计算产品与智能电子业务是收入增长的主要动力。2025 年该板块收入 158.09 亿元，同比增长 22.36%，但毛利率仅 4.82%。公司旗下软通计算机稳居业内第一梯队，围绕昇腾、鲲鹏产品布局超节点服务器、液冷服务器、AI 服务器及液冷工作站等产品，广泛覆盖国产替代重点行业，整体市场份额领先。旗下智通国际经营成果再创新高，采用 OEM 与 ODM 双轨经营模式，拥有面向消费市场的“机械革命”和面向商用市场的“恒悦”两大品牌；是 NVIDIA 全球核心伙伴，AMD 大中华区第二大 OEM，英特尔 2025 年度钛金级合作伙伴。

硬件业务仍是公司全栈智能战略不可或缺的环节。公司拥有服务器整机、AI 工作站、终端和液冷设备能力，可以直接参与智算中心硬件交付；服务器进入客户后，还可带动操作系统、模型部署、集群运维和行业应用收入。其战略价值并不只体现在当前设备销售毛利，而在于为算力运营提供设备基础和客户入口。

软件业务保持稳健，但传统项目制交付的利润率仍有修复空间。2025 年软件与数字技术服务收入 189.61 亿元，同比增长 4.64%，毛利率 16.50%，较 2024 年有所下降。公司通过 AI 编码、自动测试、智能运维、MaaS 平台和 FDE 交付模式提升单人产出，

若上述能力能够形成标准化产品和重复性收费，软件业务利润率有望逐步恢复；若 AI 主要用于内部提效，则收入模式仍将以人力和项目交付为主。

2. 算力租赁加速升级，AI PC 进入规模渗透期

2.1. 推理需求持续扩张&算力供给模式升级，智算运营行业β向上

2.1.1. 算力租赁连接芯片供给与 AI 应用，商业模式持续升级

算力租赁是将 AI 芯片、服务器及配套基础设施形成的计算能力转化为可计量、可交易服务的一种商业模式，主要包括物理服务器租赁、智算实例租赁和计算时长租赁。算力租赁连接上游 AI 芯片和服务器供给与下游模型训练、微调和推理需求：算力运营商将分散的计算、存储和网络资源池化，形成可以统一管理、切分和调度的算力资源，再以卡时、台月、整机、集群等方式向外提供。对算力需求方而言，租赁模式使其可以按照项目周期和实际任务量获取算力，不必为阶段性或波动性的计算需求自行建设同等规模的芯片、服务器、机房、网络和供电基础设施，从而将前期集中投入转化为随使用规模发生的算力服务支出。

表1：算力租赁服务类型

指标名称	描述
机房租赁服务	客户租赁整个机房，用于部署服务器和网络设备。
机柜租赁服务	客户租赁机柜空间，放置服务器和网络设备。
物理机租赁服务	客户直接租赁物理服务器，可定制硬件配置。
云主机租赁服务	客户租赁虚拟化的计算资源，动态扩展或缩减。
机时租赁服务	客户按需租赁计算时间，按使用时间支付费用
云际协同服务	可支持多算力中心跨地域、跨主体协同提供计算服务。
支持实例类型	可提供算力服务实例是否丰富，统计支持的实例类型，包含如虚拟机、容器、函数、裸金属、超算实例、智算实例等

数据来源：国家数据局，东吴证券研究所

传统算力租赁收入可以简化为：

算力租赁收入 = 可运营算力规模 × 可售时间 × 资源负载率 × 单位租赁价格

随着算力应用由基础资源调用向模型训练和规模化推理延伸，算力租赁正在由单纯提供设备和计算时间，向“算力资源+开发环境+模型服务+运维服务”升级。基础形态主要提供机房、机柜和裸金属服务器租赁，更高层级则增加虚拟化资源、跨中心调度、预置操作系统、开发环境、基础模型和应用服务，使客户能够缩短算力选型、环境配置和模型部署流程。

算力计费单位也在由卡时和核时进一步延伸至 Token。工信部提出推行按卡时、核时及 Token 计费等灵活付费模式，产业实践中算力服务亦开始从按照硬件占用时长收费，向按照实际模型调用量和 Token 消耗量交付。

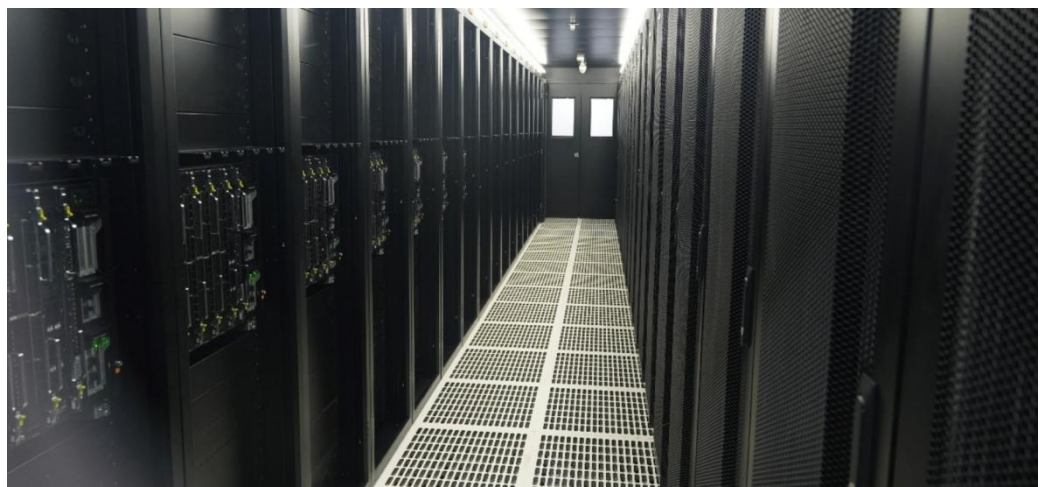
Token 工厂不是算力租赁的替代品，而是建立在算力租赁和集群运营之上的更高层商业形态。底层仍需要服务器、芯片、电力和机房，但评价标准由理论算力转向单位时间能够稳定输出多少有效 Token，以及这些 Token 能否满足首字延迟、吞吐、正确率和可用性要求。软件调度和推理优化能力因此成为算力资产回报率的重要组成部分。

2.1.2. AI 需求由训练向推理扩散，算力消耗由项目制转向持续运营

大模型产业初期的算力需求主要来自预训练，应用普及后推理需求有望成为更长期、更分散的增长来源。训练通常要求高性能的大规模集群，推理分布于办公、编程、搜索、金融、工业、汽车、智能终端和智能体等场景，推理应用覆盖个人用户、企业客户及多个行业场景，使用频率更高，并伴随业务量长期增长。

智能体进一步放大了推理基础设施的复杂度。普通对话通常是一问一答，而智能体需要持续规划、调用工具、读取长上下文并反复生成中间结果，同一个任务可能产生多轮、长链路 Token 消耗。高并发任务还会产生大量 KV Cache，对显存管理和集群调度提出更高要求。北京壹号词元工厂聚焦 Agentic Serving 场景，针对智能体长时运行、长周期和多轮次负载，引入动态负载仿真、算力调度及 KV Cache 复用能力，反映公司对推理需求结构变化的判断。

图4：北京壹号词元工厂集成了算力调度和 KV Cache 复用算法



数据来源：软通动力官网，东吴证券研究所

单位 Token 价格下降并不必然导致算力需求下降，反而可能通过场景扩张释放更大调用量。当推理成本降低后，原本因成本过高而无法落地的客服、营销、研发、工业分析和个人智能体场景具备商业可行性；若调用量增长快于单位价格下降，推理算力总体市场仍将扩张。

2.1.3. 供给快速增加，利用率与真实客户成为行业核心壁垒

国内智算供给快速扩张，但不同地区、不同芯片和不同任务之间仍存在结构性错配。同样标称为 1P 的算力，在不同计算精度、芯片架构、显存容量、集群互联和软件框架下，实际模型产出可能存在较大差异。客户真正需要的不是理论峰值，而是与其模型匹

配、能够稳定运行的有效算力。

图5：算力调度与经营管理平台架构



数据来源：国家数据局，东吴证券研究所

行业竞争的核心正由“能否获得算力卡”转向以下三项能力：

第一，**异构适配能力**。能够兼容不同芯片、操作系统、模型框架和推理引擎，降低客户迁移成本。

第二，**运营调度能力**。能够根据任务时延、模型大小和成本要求，动态分配资源，提高显存、缓存和设备利用率。

第三，**需求消纳能力**。能够获得长期真实业务负载，而不是仅依赖短期算力转售和中间商订单。

软通动力的差异化恰好集中于第三项。公司原有大型行业客户和软件团队能够识别AI应用需求，计算硬件业务能够提供服务器和集群，天元平台和词元工厂负责持续运营。

软通动力不是算力规模最大的企业，但有望成为交付链条较完整的企业。公司既可以向客户出售服务器，也可以建设和运营集群，还能够提供模型部署、Token服务和行业应用。竞争优势在于通过软硬协同提高有效Token产出，并将算力嵌入客户实际业务。

2.2. AI PC 进入规模渗透阶段，产品形态由 NPU 加速向端侧智能体演进

AI PC 是在传统 CPU 和 GPU 基础上进一步引入 NPU 等专用 AI 计算单元，使部分模型推理和 AI 任务能够直接在终端本地运行的新一代个人计算设备。NPU 能够并行处理 AI 工作负载，并减少相关任务对 CPU 和 GPU 资源的占用。微软 2024 年推出 Copilot+ PC 时，将 40+ TOPS NPU 算力作为其核心硬件门槛之一，并联合高通、英特

尔、AMD 以及主要 PC 厂商推动相关产品落地，AI PC 由概念验证逐步进入标准化硬件平台阶段。

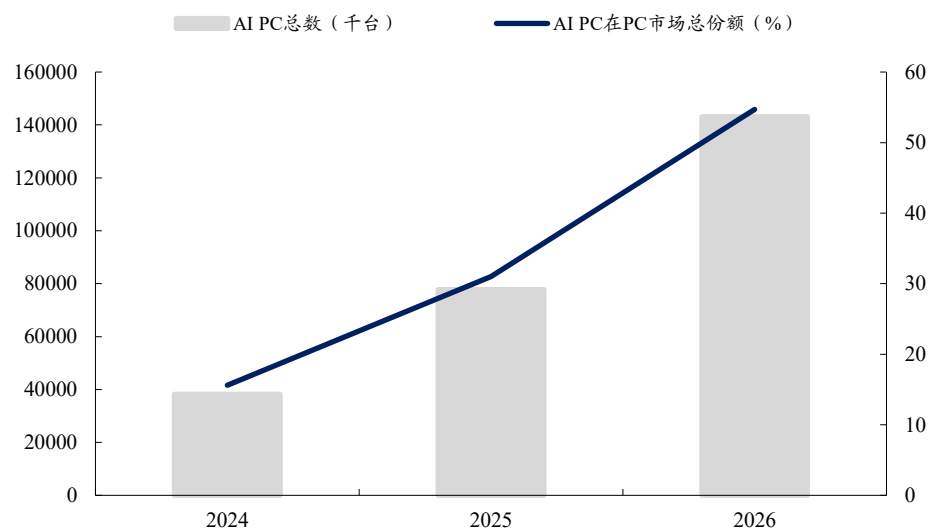
图6: Windows 11 AI+ PC



数据来源：Windows 官网，东吴证券研究所

AI PC 渗透率正在快速提升，PC 产业逐步由传统换机周期进入 AI 能力普及周期。根据 Gartner 预测，全球 AI PC 出货量将由 2024 年的 3815 万台增至 2025 年的 7779 万台，对整体 PC 市场的渗透率由 15.6% 提升至 31.0%。2026 年 AI PC 出货量预计进一步达到 1.43 亿台、占整体 PC 市场 54.7%，其中 AI 笔记本渗透率预计达到 58.7%。

图7: 全球 AI PC 出货量以及市场份额

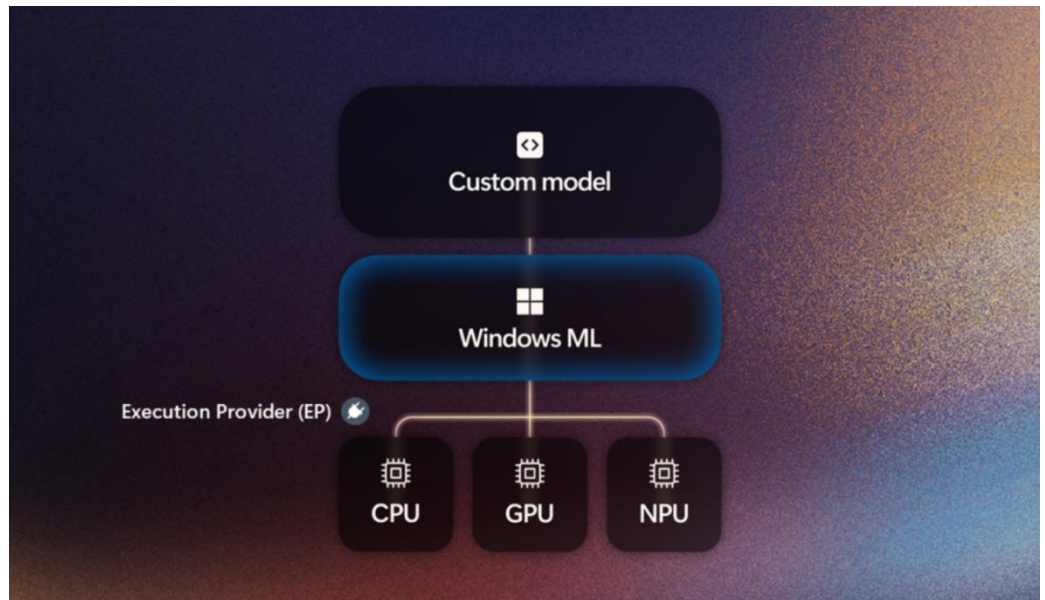


数据来源：Gartner，东吴证券研究所

随着硬件逐渐普及，AI PC 的竞争重点正在由单纯提高 NPU TOPS 向端侧模型运

行能力和软硬件生态协同延伸。Windows ML 已支持开发者在 CPU、GPU 和 NPU 之间部署端侧模型，并根据终端硬件选择适配的推理后端，操作系统开始承担模型部署、硬件适配和 AI 算力调用的统一平台角色。端侧推理可以减少部分数据上传和云端调用，同时改善响应时延和隐私保护，但复杂模型和高强度任务仍需要云端算力，PC 侧 AI 架构正在由单纯本地推理进一步向本地与云端协同的混合 AI 模式发展。

图8: Windows ML 堆栈图



数据来源：Windows 博客，东吴证券研究所

在此基础上，AI PC 的产品形态进一步由“AI 功能增强”向“智能体 PC”演进。2026 年以来，产业侧开始将 PC 上的 AI 智能体定义为能够完成思考、调度、执行、交互和记忆等多步骤任务的软件系统，通过本地中小模型处理文档理解、检索、总结及中间规划等任务，并在需要更复杂能力时调用云端大模型。这种架构意味着 PC 不再只是承载单点生成式 AI 应用，而逐步成为能够理解用户意图、调用本地应用和云端服务并持续执行工作流的个人智能入口。

国内 AIPC 产业政策和标准体系也在加快完善。深圳市《加快推进人工智能终端产业发展行动计划（2025—2026 年）》明确将人工智能计算机列为重点产品方向，提出推动计算机成为由端侧大模型赋能的生产力工具和多平台协作的个人工作站。2026 年 4 月，我国正式发布《人工智能终端智能化分级》系列国家标准，其中 GB/Z 177.4-2026 专门面向微型计算机，人工智能终端智能化水平由 L1 响应级、L2 工具级逐步提升至 L3 辅助级和 L4 协同级，AIPC 开始形成统一的智能化评价框架。国务院《关于深入实施“人工智能+”行动的意见》进一步提出，到 2027 年新一代智能终端、智能体等应用普及率超过 70%，到 2030 年超过 90%，并明确提出大力发展人工智能电脑等新一代智能终端，为 AIPC 中长期普及提供政策支撑。

当前 AI PC 仍处于“硬件渗透快于应用价值兑现”的发展阶段，产业增长既受到

AI 功能升级驱动，也受到传统 PC 换机周期和供应链环境影响。2026 年 PC 产业面临 DRAM、NAND 等存储器供应趋紧和价格上涨压力，Omdia 统计美国 2026Q1 价格低于 500 美元的 PC 出货量同比下降 18.7%，同时厂商更多将资源配置向高价值 AIPC 倾斜，成本压力可能加快 AI 能力向中高端产品集中。后续 AIPC 产业发展的核心变量将由“是否搭载 NPU”进一步转向端侧模型能力、智能体实际使用频率、操作系统与应用生态适配，以及本地计算能够在多大程度上降低云端 Token 成本并形成用户可感知的生产力提升。

3. 成长性分析：资源力+产品力+客户力，智算运营进入兑现期

3.1. 现有算力超过 1800P，怀来项目打开规模扩张空间

3.1.1. 全国节点初步形成，算力业务由建设向运营延伸

公司已经具备超过 1800P 的现有算力规模，并通过自持资源和合作资源构建多节点算力网络。公司算力业务覆盖智算中心规划、设计、建设、运营和调度，采用核心资源自持与合作算力接入相结合的方式。相较全部自建的重资产模式，混合模式既可以控制核心节点服务质量，又能够减少跨区域扩张所需资本投入。

公司原有智算布局包括平潭异构算力池、北京及环京节点、韶关公共算力平台等，服务内容已经由智算中心建设向算力租赁、算力订阅和平台运营延伸。公司承担的不只是机房和设备交付，还包括异构适配、统一调度、资源监控、模型部署和持续运维。

3.1.2. 怀来 3316P 项目扩产，算力规模有望接近现有三倍

怀来智算中心一期是公司未来三年最重要的算力增量项目。项目拟新增 3316.15P 训练与推理算力，相当于公司现有约 1800P 规模的约 1.8 倍；建成后公司理论算力规模将超过 5100P。项目位于张家口数据中心集群，能够承接京津冀地区训练、离线推理和批量数据处理需求。

项目成熟期盈利水平明显高于公司现有业务。怀来项目平均测算毛利率为 38.03%，税后内部收益率为 10.14%，显著高于公司数字能源与智算服务当前毛利率。差异主要来自现有板块仍包含项目建设、数字能源和平台投入，而怀来测算更多反映算力进入稳定运营后的成熟收益。

客户消纳仍是怀来项目的核心风险。公司当前接洽潜在需求约 1300P—1600P，仅覆盖新增产能约 39%—48%，且“接洽需求”与正式合同、最低使用量承诺之间存在差异。怀来项目能否达到测算回报率，取决于投产节奏、客户合同期限、平均利用率和算力租赁价格，而非单纯的建设进度。

3.2. 服务器+调度平台协同，构筑算力运营底座

3.2.1. 服务器产品能力完善，掌握算力基础设施入口

公司完成计算机业务整合后，具备服务器、AI 工作站、PC 和智能终端研发制造能力，能够由设备供应延伸至集群建设和运营。智算中心并非 AI 芯片的简单堆叠，还涉及服务器结构、互联网络、存储、供电、散热及运维体系。公司拥有硬件设计和集成交付能力，可以围绕自有智算项目持续优化产品，并通过客户硬件订单导入后续软件和运营服务。

公司硬件业务当前毛利率较低，但 AI 服务器、液冷服务器和超节点设备的技术含量、客户黏性和后续服务价值高于传统 PC。若高价值产品占比提升，并带动操作系统、模型部署和运维收入，公司硬件业务毛利率有望逐步改善。

3.2.2. 天元平台解决异构调度，软件能力提高有效算力产出

异构调度平台是公司由重资产租赁向平台型运营商升级的关键。不同芯片在指令集、算子、显存、通信协议和模型框架方面存在差异，客户难以直接在多个资源池之间自由迁移。公司通过统一资源抽象、任务调度、模型适配和计量管理，将不同节点纳入同一算力服务体系。

平台具备两方面经济价值：一是提高自有设备利用率，减少单一节点闲置；二是接入外部合作资源，在不全部承担硬件资本开支的情况下扩大可运营规模。若天元平台能够按照调度量、调用量或运营服务持续收费，公司智算业务将由项目建设收入逐步转向平台运营收入。

3.3. 北京壹号词元工厂投产，算力租赁向 Token 生产升级

3.3.1. 一期日产能 1.4 万亿 Token，形成首个标准化推理工厂

北京壹号词元工厂正式点亮，标志着公司算力业务由出售底层资源向交付标准化 Token 产能升级。2026 年 6 月 9 日，公司词元工厂计划首个标杆示范项目在北京投产，一期每日 Token 产能达到 1.4 万亿，满负荷理论年产能约 511 万亿 Token。项目聚焦智能体服务负载，集成算力调度、KV Cache 复用、全栈观测及推理优化能力。

传统算力租赁中，客户购买 GPU 节点或卡时后仍需自行完成模型部署、推理引擎配置、算子适配和运维；词元工厂则由公司承担底层硬件到模型推理之间的工程环节，客户直接获得具备 SLA 保障的 Token 服务。公司商业模式由“出租生产设备”向“生产并交付数字产品”延伸。

词元工厂的核心价值在于提高同等硬件的有效产出。对算力运营商而言，采购更多芯片只是扩大理论产能；通过批处理、请求调度、KV Cache 复用、算子融合和通信优化，才能提高单卡 Token 吞吐并降低单位成本。若公司能够持续提升相同设备上的 Token 产量，其软件能力可以转化为超越硬件租金的利润来源。

3.3.2. 开源性能基准，尝试建立 Token 工厂评价体系

公司同步开源词元工厂性能基准和 LoadGen 2.0，试图将 Token 服务由非标准化项目转化为可比较的工业化产能。该基准针对智能体长时运行、多轮状态保持、请求间隔变化和上下文膨胀等特征，评价对象不再只是模型单次跑分，而是系统在真实推理负载下的吞吐、延迟和稳定性。

开源基准短期未必直接形成收入，但具备两方面战略意义：一是帮助客户比较不同算力服务的有效产出，降低采购决策难度；二是使公司积累的调度和运维经验转化为行业标准影响力。若公司定义的指标逐渐成为客户采购 Token 服务的重要参考，其先发运营数据和工程经验将形成一定壁垒。

3.4. 头部模型厂商签约，真实负载验证商业模式

3.4.1. 模型厂商部署推理服务，词元工厂进入商业运营阶段

6 月 25 日智算服务协议是词元工厂投产后首个公开披露的头部模型客户验证。根据协议，该模型厂商将大模型推理服务部署于北京壹号词元工厂，软通动力提供标准化、弹性可扩展并具备 SLA 保障的 Token 推理服务，合作范围覆盖推理加速、高性能集群优化、服务质量保障和行业 AI 应用落地。

真实业务负载是对词元工厂更有效的产品验证。头部模型厂商的任务包含长链路推理、多轮交互和高并发调用，对缓存、显存、任务调度及服务稳定性的要求明显高于实验室压力测试。公司能够在真实场景中积累负载数据、调整调度算法，并不断降低单位 Token 成本。

3.4.2. 客户签约提高利用率确定性，但合同经济性仍待披露

本次签约最重要的意义是获得长期潜在算力消纳入口，而不是立即贡献多少收入。模型厂商面向企业和开发者提供 API 和应用服务，理论上能够形成高频、持续的推理需求；相比短期卡时订单，其业务负载与模型调用量共同增长，有利于提高词元工厂利用率和设备全生命周期回报。

但公司仅披露了合作内容，未披露客户名称、合同金额、服务期限、Token 采购量、最低消费承诺和结算价格。因此，目前无法准确测算该协议对 2026 年收入和利润的贡献，也不能根据模型厂商特征推断其具体身份。

我们认为，本次协议应主要反映为三项边际变化：

第一，商业模式从概念走向客户验证。词元工厂不再只是内部建设方案，而是已经有模型厂商将真实推理服务部署其中。

第二，算力利用率预期改善。模型厂商持续调用需求有望为工厂提供基础负载，降低资产闲置风险。

第三，公司估值属性发生变化。算力业务不再仅按重资产租赁评估，而应增加推

理平台、软件优化和 Token 运营价值。

3.4.3. “模型厂商造脑、词元工厂生产 Token”，形成产业专业化分工

词元工厂使模型厂商能够减少大规模推理集群的自建投入，将资本和研发资源集中于模型迭代与应用生态。软通动力负责服务器、芯片适配、推理框架、资源调度、缓存管理和 SLA，模型厂商负责算法、模型和产品，两者共同形成从模型能力到产业应用的完整链条。

该模式能否复制取决于公司能否继续获取第二家、第三家模型客户。单一客户能够证明产品可用，但不足以证明平台具备广泛复制性；当多个模型厂商和行业客户共同使用词元工厂后，公司才能通过多租户调度提高峰谷互补，并降低客户集中度。

3.5. 客户+资源+平台三力协同，形成区别于纯算力租赁商的优势

3.5.1. 存量大客户提算力需求入口

客户渠道是软通动力区别于新进入算力租赁企业的核心优势。公司长期服务通信、金融、互联网、制造和能源等大型客户，在客户内部已经拥有软件开发、数字化运营和基础设施团队。公司可以从现有数字化项目中识别 AI 需求，再将服务器、模型平台、算力和行业应用交叉销售。

纯算力租赁商通常需要在设备投产后重新寻找客户，而软通动力可以由咨询和软件项目提前锁定需求。对金融、制造和央国企客户而言，数据合规、私有化部署、国产化适配和持续运维通常比单纯卡时价格更重要，公司全栈交付模式更符合此类客户采购习惯。

3.5.2. 国产与国际算力异构布局，覆盖差异化需求

不同客户和模型对芯片生态、模型兼容性和成本要求不同，异构资源比单一芯片路线更有利于提高需求覆盖率。公司硬件体系重点围绕国产计算生态建设，同时通过异构调度接入多类型资源。其价值不在于押注某一种芯片，而在于将不同资源统一转化为客户可直接使用的训练、推理和 Token 服务。

异构布局也提高了运营难度。不同芯片需要不同算子、通信库和推理引擎，公司必须持续投入适配与运维。若天元平台和词元工厂能够实现跨架构资源抽象，上述复杂性将由进入门槛转化为竞争壁垒。

3.6. 智算业务收入占比较低，毛利贡献有望快速提升

数字能源与智算服务当前收入规模很小，但成熟算力运营的毛利率显著高于计算硬件业务。2025 年该板块收入仅 3.01 亿元，占总收入 0.86%，毛利率 10.63%；怀来项目成熟期平均测算毛利率达到 38.03%。随着自持算力投产、Token 服务放量和平台收入占比提升，板块利润贡献有望明显快于收入增长。

我们预计数字能源与智算服务收入由 2025 年的 3.01 亿元增长至 2028 年的 24.00 亿元，三年复合增速接近 100%；毛利率由 10.6% 提升至 30.0%，毛利由 0.32 亿元增长至 7.20 亿元。2028 年该业务收入占比预计仅 4.7%，但毛利贡献占比有望超过 11%，成为公司利润率修复的重要驱动力。

4. 盈利预测与投资评级

4.1. 盈利预测

我们预计公司 2026—2028 年营业收入分别为 398.79、454.77 和 513.87 亿元，同比增长 13.7%、14.0% 和 13.0%。其中，计算硬件仍是收入增量的主要来源，智算运营则是毛利增量和估值扩张的主要来源。

4.1.1. 软件与数字技术服务

公司存量大客户基本盘稳健，AI 咨询、MaaS 平台、智能体开发和 FDE 交付有望贡献新增收入；但传统项目制业务仍面临人员成本和价格竞争，因此对收入及毛利率保持审慎。

我们预计 2026—2028 年该业务实现收入 202.88、219.11、236.64 亿元，同比增长 7%、8%、8%；毛利率分别为 16.5%、16.9%、17.2%。

4.1.2. 计算产品与智能电子

公司 PC、服务器、AI 服务器、AI 工作站和智能终端产品持续放量，信创需求和智算基础设施建设推动服务器业务增长。伴随 AI 服务器、液冷设备和高价值企业级产品占比提升，硬件毛利率有望逐步修复。

我们预计 2026—2028 年该业务实现收入 189.71、221.96、253.03 亿元，同比增长 20%、17%、14%；毛利率分别为 5.5%、5.8%、6.2%。

4.1.3. 数字能源与智算服务

现有算力节点持续运营，怀来项目逐步建设，北京壹号词元工厂进入客户验证阶段。考虑头部模型厂商合同金额尚未披露，我们不单独上调 2026 年收入，但在 2027—2028 年逐步计入 Token 服务、算力订阅和平台运营放量。

我们预计 2026—2028 年该业务实现收入 6.00、13.50、24.00 亿元，同比增长 99%、125%、78%；毛利率分别为 18.0%、24.0%、30.0%。

我们预计公司 2026—2028 年归母净利润分别为 2.96、5.59 和 9.82 亿元。2026 年公司仍处于硬件扩张、AI 研发和智算项目投入期，利润改善幅度有限；2027 年以后，随着算力利用率提高、Token 业务放量、硬件毛利率修复及费用率摊薄，利润增速有望显著高于收入增速。

表2: 公司业务拆分预测表

单位: 百万元	指标	2025A	2026E	2027E	2028E
软件与数字技术服务	营业收入	18960.5	20287.8	21910.8	23663.7
	yoy	4.6%	7.0%	8.0%	8.0%
	毛利率	16.5%	16.5%	16.9%	17.2%
计算产品与智能电子	营业收入	15809.1	18970.9	22195.9	25303.4
	yoy	22.4%	20.0%	17.0%	14.0%
	毛利率	4.8%	5.5%	5.8%	6.2%
数字能源与智算服务	营业收入	300.9	600.00	1350.0	2400.0
	yoy	16.2%	99.4%	125.0%	77.8%
	毛利率	10.6%	18.0%	24.0%	30.0%
其他业务	营业收入	19.4	20.0	20.0	20.0
	yoy	12.9%	3.1%	0.0%	0.0%
	毛利率	48.0%	48.0%	48.0%	48.0%
合计	总收入	35089.9	39878.7	45476.7	51387.0
	yoy	12.1%	13.7%	14.0%	13.0%
	总毛利率	11.2%	11.3%	11.7%	12.4%

数据来源: Wind, 东吴证券研究所测算

4.2. 估值分析

采用 PE 估值法对公司进行估值, 选择**宏景科技**、**中软国际**、**润建股份**作为可比公司。可比公司均在政企数字化、人工智能应用及算力服务等业务环节与软通动力存在一定交集: 宏景科技以智慧城市业务为基础, 业务范围已延伸至算力服务, 可提供 AI 大模型训练与推理、数据治理以及运营运维等服务; 中软国际围绕 AIERP、AI 鸿蒙、Token 运营和 AI 原生应用开发形成全栈全场景 AI 产品与服务体系, 其中 Token 运营覆盖算力、模型和智能体一体化能力; 润建股份以数字化智能运维为基础, 业务覆盖通信网络、数字化、IDC 及新能源等领域, 并依托算力服务、数据服务和人工智能行业模型开发平台拓展 AI 业务。软通动力当前形成软件与数字技术服务、计算产品与智能电子、数字能源与智算服务三类核心业务, 并全面推进“软硬一体、全栈智能”战略, 以 AI 基础设施、计算智能、场景智能和终端智能为四大核心板块。

与可比公司相比, 公司不仅具备软件和数字技术服务能力, 还覆盖服务器、AI 终端、算力基础设施建设和智算运营等环节, 业务链条相对完整。公司已完成约 33.48 亿元向特定对象发行, 募集资金主要投向京津冀软通信创智造基地、AIPC 智能制造基地及怀来智算中心一期等项目。募投项目有望提升 AIPC 等计算产品的生产规模及智能化水平, 并增强训练与推理算力供给能力, 为智算业务后续扩张提供支撑。与此同时, 公司 AI Infra 层面的算力收入过去主要依赖硬件销售, 随着企业级 AI 应用及智能体需求增长, 商业模式逐步向持续运营服务、Token 资源消耗和平台使用费等模式延伸。可比公司 26-28 年的 PE 均值分别为 49x、27x、15x, 公司 26-28 年的 PE 分别为 130x、69x、39x, 考虑公司预计伴随募投项目实施及效益释放, 盈利能力和经营业绩有望进一步提

升。公司 AI 业务已形成较大收入规模，并围绕 AI 基础设施、计算智能、场景智能和终端智能形成较完整布局，在软件服务、服务器及 AI 计算产品、AIPC 等智能终端以及智算基础设施与运营环节具备较完整的全栈能力，首次覆盖给予“买入”评级。

表3: 可比公司 PE 估值表

股票代码	股票简称	收盘价（元）	EPS（元/股）			PE		
			2026E	2027E	2028E	2026E	2027E	2028E
301396.SZ	宏景科技	186.1	2.0	4.1	10.0	93	45	19
00354.HK	中软国际	3.2	0.2	0.3	0.4	15	12	9
002929.SZ	润建股份	53.7	1.4	2.2	3.3	40	24	16
	均值					49	27	15
301236.SZ	软通动力	37.2	0.3	0.5	0.9	130	69	39

数据来源：iFind（截至 2026 年 8 月 22 日，可比公司预测数据来自 iFind 一致预期，软通动力预测数据为东吴证券测算），东吴证券研究所

4.3. 投资建议

软通动力已经由传统 IT 服务商转型为国内少数同时覆盖服务器、基础软件、智算中心、异构调度、Token 服务和行业应用的全栈智能平台。公司现有 1800P 以上算力和怀来 3316P 新增项目提供资源基础，北京壹号词元工厂将算力交付单位由卡时升级为 Token，头部模型厂商签约则首次验证了真实商业负载。

短期看，公司硬件收入增长较快，但低毛利、存货和现金流压力仍将制约利润表现；中期看，怀来项目和词元工厂有望推动智算收入和毛利率同步提升；长期看，公司若能够形成“模型厂商提供模型、软通动力生产 Token、行业客户持续调用”的分工模式，算力业务将由重资产租赁升级为软件定义的 AI 生产运营。

我们预计公司 2026—2028 年归母净利润分别为 2.96、5.59 和 9.82 亿元，对应 2026/2027/2028 年 PE 分别为 130x、69x、39x。考虑公司当前仍处于转型投入阶段，2026Q1 利润端受到转型投入增加影响，短期估值水平较高。但 AI 业务较快增长、募投项目逐步落地以及智算业务由硬件销售向运营服务、Token 资源消耗和平台使用费延伸，有望带来业务结构优化和中期盈利弹性。因此，在相关业务增长及盈利改善逐步兑现的前提下，可在可比公司估值基础上给予一定的成长性及全栈能力溢价，首次覆盖给予“买入”评级。

5. 风险提示

1、**智算项目建设进度不及预期。** 怀来项目设备采购、机房建设、网络接入或验收延期，将影响新增算力投产和收入确认节奏。

2、**算力客户及利用率不及预期。** 公司接洽的 1300P—1600P 需求尚未完全覆盖怀来新增产能，且潜在需求不等同于正式保底合同，利用率偏低可能导致折旧和运维成本压制利润。

3、Token 服务合同经济性不及预期。 公司尚未披露头部模型厂商协议金额、服务期限、保底 Token 数量和结算价格，协议可能更多体现战略合作，实际收入贡献存在不确定性。

4、Token 产能与实际销量存在差异。 北京壹号词元工厂 1.4 万亿 Token 日产能为产能指标，若客户调用不足、模型负载变化或服务要求提高，实际可计费产出可能低于规划。

5、算力租赁价格下降。 芯片供给增加、推理效率提升和行业竞争加剧可能推动单位算力和单位 Token 价格下降，对项目回报率造成压力。

6、硬件业务竞争加剧。 PC 和服务器的业务毛利率较低，若行业价格竞争加剧、原材料上涨或客户验收推迟，公司整体利润率可能低于预测。

7、存货及现金流风险。 硬件备货和智算项目建设占用大量资金，若销售回款放缓，可能增加借款、融资成本和存货减值压力。

8、芯片供应与技术迭代风险。 AI 芯片更新速度较快，设备可能面临性能落后和减值；海外芯片政策变化也可能影响部分集群的建设和运营。

9、异构适配和平台能力不及预期。 不同芯片、模型和推理框架的适配难度较高，若天元平台和词元工厂无法持续提高有效 Token 产出，将削弱公司相较纯硬件租赁商的优势。

10、AI 投入回报不及预期。 公司持续投入 MaaS、智能体、Token 平台和算力调度，若相关产品未能形成规模化商业收入，可能继续压制短期盈利能力。

软通动力三大财务预测表

资产负债表(百万元)	2025A	2026E	2027E	2028E	利润表(百万元)	2025A	2026E	2027E	2028E
流动资产	22,357	27,216	28,173	29,817	营业总收入	35,090	39,879	45,477	51,387
货币资金及交易性金融资产	6,445	7,991	7,313	7,502	营业成本(含金融类)	31,159	35,372	40,156	45,015
经营性应收款项	7,468	9,286	10,290	11,241	税金及附加	164	187	214	242
存货	7,710	7,369	7,808	8,128	销售费用	1,091	1,216	1,364	1,516
合同资产	0	1,795	1,955	2,107	管理费用	1,557	1,735	1,910	2,081
其他流动资产	734	776	807	839	研发费用	906	1,037	1,182	1,336
非流动资产	5,057	7,251	8,630	9,226	财务费用	148	125	127	139
长期股权投资	127	127	127	127	加:其他收益	135	152	159	170
固定资产及使用权资产	2,051	3,448	4,422	4,819	投资净收益	(10)	0	0	0
在建工程	128	728	1,028	1,128	公允价值变动	105	0	0	0
无形资产	958	1,071	1,177	1,276	减值损失	(280)	0	0	0
商誉	208	208	208	208	资产处置收益	17	8	5	5
长期待摊费用	116	116	116	116	营业利润	33	366	687	1,233
其他非流动资产	1,469	1,552	1,552	1,552	营业外净收支	(28)	(5)	(5)	(5)
资产总计	27,414	34,467	36,803	39,043	利润总额	6	361	682	1,228
流动负债	14,222	17,389	18,806	20,110	减:所得税	(172)	65	123	246
短期借款及一年内到期的非流动负债	5,780	5,879	5,879	5,879	净利润	177	296	559	982
经营性应付款项	4,207	6,878	7,808	8,628	减:少数股东损益	(29)	0	0	0
合同负债	1,750	1,795	1,955	2,107	归属母公司净利润	206	296	559	982
其他流动负债	2,486	2,837	3,164	3,497	每股收益-最新股本摊薄(元)	0.20	0.29	0.54	0.95
非流动负债	2,508	3,303	3,803	4,003	EBIT	72	486	809	1,367
长期借款	1,103	1,903	2,403	2,603	EBITDA	647	675	1,130	1,741
应付债券	0	0	0	0	毛利率(%)	11.20	11.30	11.70	12.40
租赁负债	146	146	146	146	归母净利率(%)	0.59	0.74	1.23	1.91
其他非流动负债	1,259	1,254	1,254	1,254	收入增长率(%)	12.05	13.65	14.04	13.00
负债合计	16,731	20,692	22,610	24,114	归母净利润增长率(%)	14.27	43.45	89.10	75.70
归属母公司股东权益	10,386	13,477	13,896	14,631					
少数股东权益	297	297	297	297					
所有者权益合计	10,683	13,774	14,193	14,929					
负债和股东权益	27,414	34,467	36,803	39,043					

现金流量表(百万元)	2025A	2026E	2027E	2028E	重要财务与估值指标	2025A	2026E	2027E	2028E
经营活动现金流	(242)	396	887	1,432	每股净资产(元)	10.90	13.02	13.42	14.13
投资活动现金流	(991)	(2,339)	(1,700)	(970)	最新发行在外股份(百万股)	1,035	1,035	1,035	1,035
筹资活动现金流	599	3,506	137	(274)	ROIC(%)	13.11	2.02	2.99	4.74
现金净增加额	(712)	1,545	(677)	189	ROE-摊薄(%)	1.98	2.19	4.02	6.71
折旧和摊销	575	189	321	374	资产负债率(%)	61.03	60.04	61.44	61.76
资本开支	(976)	(2,279)	(1,700)	(970)	P/E(现价&最新股本摊薄)	186.73	130.17	68.84	39.18
营运资本变动	(1,156)	(241)	(217)	(151)	P/B(现价)	3.41	2.86	2.77	2.63

数据来源:Wind,东吴证券研究所,全文如无特殊注明,相关数据的货币单位均为人民币,预测均为东吴证券研究所预测。

免责声明

东吴证券股份有限公司(以下简称“本公司”)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。

本研究报告仅供符合中国证监会《证券期货投资者适当性管理办法》规定的专业投资者使用,且接收人须为本公司认定并建立服务关系的专业机构投资者使用,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。本报告不面向普通投资者发布或传播。任何非专业机构投资者收到本报告后,应立即退回并销毁,不得依据本报告内容做出任何投资决策。

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>