仅供内部参考，请勿外传！

华创证券 HUA CHUANG SECURITIES

证券研究报告

集成电路 2026 年 09 月 22 日

燧原科技-U（688801）深度研究报告

全栈自研筑算力底座，万卡集群启 AI 新篇

- **国产云端 AI 芯片领军企业，八载四代架构筑基科创板新征程。**燧原科技成立于 2018 年，专注云端 AI 芯片研发、设计和销售，是国内领先的通用人工智能基础设施企业。公司坚持 DSA 自主技术路线，已迭代四代架构推出 5 款云端 AI 芯片，构建了覆盖 AI 芯片、加速卡及模组、智算系统及集群、AI 计算及编程软件平台的完整产品体系。2026 年 9 月 11 日，上海证券交易所科创板正式挂牌上市。2022-2025 年营收从约 0.9 亿元增长至 9.9 亿元，复合增长率达 122.33%；2026 年上半年实现营收 11.2 亿元，已超过 2025 年全年水平。公司尚未盈利但亏损呈收窄趋势，2025 年归母净利润-11.6 亿元，同比收窄约 22.94%，AI 加速卡及模组为核心增长引擎。
- **AI 大模型驱动算力需求爆发，国产替代进程加速推进。**AI 大模型训练与多模态应用加速落地，使全球算力需求进入指数级增长阶段。据灼识咨询数据，中国 AI 加速卡市场规模预计将从 2024 年的 2164.77 亿元增长至 2028 年的 8585.22 亿元，云端 AI 芯片占据市场 80%以上份额。从需求结构看，推理场景正成为算力增长的核心引擎，中国推理用 AI 加速卡 2020-2024 年年均增速达 111.06%，预计 2028 年占比将超 70%。从需求端来看，国内云厂商与三大运营商大幅加码 AI 算力资本开支，算力市场需求持续释放。与此同时，美国持续强化高端 GPU 出口限制，在国家政策扶持、算力需求爆发、全球供应链重塑及自主可控加速的多重推动下，高性能 AI 加速芯片行业步入快速成长周期，我国本土厂商在算力自主化大趋势下迎来广阔的发展机遇。
- **GCU 全栈自研，软硬件协同构筑核心壁垒。**公司基于自主指令集独创 GCU-CARE 加速计算单元与 GCU-LARE 片间高速互联，是国内首家自主研发高密度 OAM 模组方案的云端 AI 芯片企业，四代 5 款芯片均一次流片成功。围绕自研硬件构建的全栈软件平台“驭算 TopsRider”已深度适配近千个 AI 模型、覆盖超 300 个应用场景，中等复杂度 CUDA 应用迁移成本仅需约 1 人天。在集群层面，公司已落地国内首个获中国信通院五星级成熟度认证的国产万卡推理集群，形成从单芯片到跨域集群的完整能力体系。公司 IPO 计划募集资金 60 亿元，加码五代/六代芯片，以及先进人工智能软硬件协同创新项目。第五代芯片计划支持 FP4 精度与跨模态生成，第六代计划强化卡间互联与多芯粒异构集成，以应对超大规模模型对算力集群的更高要求。
- **投资建议：**我们预计公司 2026-2028 年营业收入为 35.93 亿元、61.60 亿元、111.30 亿元，对应增速 262.9%、71.4%、80.7%；归母净利润为-5.16 亿元、5.96 亿元、21.53 亿元，对应增速分别为 55.7%、215.6%、261.3%；对应 EPS（摊薄）分别为-1.20 元、1.38 元、5.00 元，首次覆盖给予“推荐”评级。
- **风险提示：**客户集中度较高与关联交易风险、供应链与产品迭代风险、盈利预测关键假设变动。

主要财务指标

	2025A	2026E	2027E	2028E
营业总收入(百万)	990	3,593	6,160	11,130
同比增速(%)	37.1%	262.9%	71.4%	80.7%
归母净利润(百万)	-1,164	-516	596	2,153
同比增速(%)	22.9%	55.7%	215.6%	261.3%
每股盈利(元)	-2.70	-1.20	1.38	5.00
市盈率(倍)	-147	-331	287	79
市净率(倍)	54.3	20.2	18.9	15.3

资料来源：公司公告，华创证券预测　　注：股价为 2026 年 9 月 21 日收盘价

推荐（首次）

当前价：524.00 元

华创证券研究所

证券分析师：吴鸣远

邮箱：wumingyuan@hcyjs.com

执业编号：S0360523040001

证券分析师：周楚薇

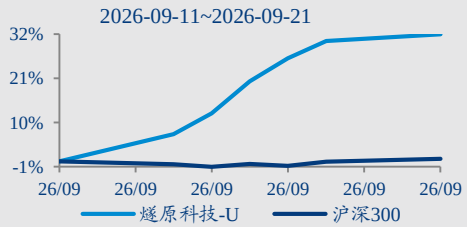
邮箱：zhouchuwei@hcyjs.com

执业编号：S0360526060001

公司基本数据

总股本(万股)	43,035.17
已上市流通股(万股)	1,790.03
总市值(亿元)	2,255.04
流通市值(亿元)	93.80
资产负债率(%)	78.70
每股净资产(元)	6.83
12 个月内最高/最低价	524.00/397.00

市场表现对比图(近 12 个月)



仅供内部参考，请勿外传！

投资主题

报告亮点

本报告深度分析国内云端 AI 算力芯片国产替代发展浪潮，对燧原科技进行全面梳理，阐明公司中长期成长核心驱动力。公司核心投资价值在于：自研 GCU 全栈软硬件体系，依托万卡级集群落地能力、本土完整供应链布局形成差异化壁垒，在大模型算力爆发、海外芯片出口管制收紧的行业机遇下，营收持续高增、亏损稳步收窄，公司预计 2026 至 2027 年进入合并报表盈利兑现窗口。

投资逻辑

基于行业趋势与公司禀赋，燧原科技的投资逻辑核心在于把握国产云端算力替代的历史性窗口，以全栈自主技术承接 AI 算力爆发式增长需求：

1、GCU 自主架构软硬一体化，集群能力行业领先。公司独立 DSA 技术路线，迭代四代五款芯片均一次流片成功；自研 TopsRider 软件大幅降低模型迁移成本，落地国内首个信通院五星级认证国产万卡推理集群，单卡至超节点全链条方案完整。

2、算力需求爆发+国产替代共振，赛道长期高景气。大模型带动推理算力需求快速扩容，海外高端 GPU 管制持续加码，国内 AI 加速卡国产份额持续抬升，本土芯片厂商替代空间广阔。

3、头部客户打底+多元拓展打开增量。腾讯为核心战略股东与第一大客户，在手大额订单支撑业绩基本盘；多家互联网、运营商客户进入测试交付周期，客户结构持续优化。

4、全本土产业链配套，产能交付稳定可控。公司进一步贯通“设计-制造-封测”国产完整链路，提前预付锁定晶圆存储产能，叠加多项国家及地方攻关项目与行业重磅奖项，供应链自主可控优势显著。

关键假设与盈利预测

我们预计公司 2026-2028 年营业收入为 35.93 亿元、61.60 亿元、111.30 亿元，对应增速 262.9%、71.4%、80.7%；归母净利润为-5.16 亿元、5.96 亿元、21.53 亿元，对应增速分别为 55.7%、215.6%、261.3%；对应 EPS（摊薄）分别为-1.20 元、1.38 元、5.00 元，首次覆盖给予“推荐”评级。

仅供内部参考，请勿外传！

目 录

一、国产云端 AI 芯片领军者，商业化落地驱动收入增长 6
（一）专注云端芯片赛道，持续迭代构建全栈算力版图 6
（二）营收高速增长，盈利结构持续优化 9
二、AI 加速芯片市场需求长期向好，国产芯片加速替代突围 12
（一）产业链：芯片设计为核，云端 AI 芯片构筑算力底座 12
1、发展趋势一：云端 AI 芯片主导地位稳固 12
2、发展趋势二：推理落地驱动 DSA 需求高增 14
3、发展趋势三：AI 加速卡基础设施向集群化和综合化演进 17
（二）需求端爆发：GPU 市场规模快速增长，云厂商运营商 AI 相关 CAPEX 提速 17
1、AI 大模型拉动 GPU 需求大幅提升 17
2、国内云巨头与运营商共同加码 AI 算力 CAPEX 19
（三）供给端重构：美国政府再升级高端芯片出口限制，我国芯片代际差持续缩小 21
三、自主 GCU 架构引领，软硬件协同构建 AI 算力核心壁垒 25
（一）硬件筑基，GCU 架构与系统级能力构筑差异化护城河 25
1、GCU 架构专为 AI 负载定制，推理能效优于通用 GPGPU 25
2、四代五芯一次流片成功，产品矩阵覆盖云端训推全场景 25
3、先进封装与超节点产品，将竞争维度延伸至集群级 26
（二）生态立本，软硬件协同与头部客户验证打开商业化空间 28
1、自研“驭算 TopsRider”全栈平台，实现生态自主可控 28
2、头部厂商验证产品能力，成熟工程交付打开商业化空间 28
四、盈利预测 31
五、风险提示 32

仅供内部参考，请勿外传！

图表目录

图表 1 燧原科技发展历程 6
图表 2 燧原科技上市后股权结构图（截至 9 月 11 日） 7
图表 3 公司核心技术团队介绍 7
图表 4 燧原科技产品路线图 8
图表 5 “芯片—软件—集群”三位一体技术架构图 9
图表 6 公司营业总收入 10
图表 7 公司归母净利润 10
图表 8 公司销售毛利率和净利率情况 10
图表 9 公司费用率情况 10
图表 10 公司主营产品营收占比 11
图表 11 公司主营产品毛利占比 11
图表 12 研发费用情况 11
图表 13 研发人员及其占比情况 11
图表 14 人工智能算力基础设施产业链 12
图表 15 AI 芯片分类示意图（按部署位置） 13
图表 16 云端 AI 芯片 vs 边缘 AI 芯片 vs 终端 AI 芯片对比 13
图表 17 2025 年中国 AI 芯片市场结构（按部署位置） 14
图表 18 AI 芯片主流架构对比 14
图表 19 全球 AI 加速卡不同技术路径主要厂商构成情况 15
图表 20 芯片厂商架构对比 15
图表 21 不同架构类型 AI 芯片境内外市场份额对比 16
图表 22 全球 AI 芯片中 GPGPU 架构和 DSA 架构比例趋势 16
图表 23 AI 算力集群三大扩展策略 17
图表 24 中国 AI 与生成式 AI 市场规模预测（2023-2028 年） 18
图表 25 中国人工智能市场支出预测（行业）2026 18
图表 26 全球 AI 加速卡市场规模及预测 18
图表 27 中国 AI 加速卡市场规模及预测 18
图表 28 全球训练和推理 AI 加速卡市场规模 19
图表 29 中国训练和推理 AI 加速卡市场规模 19
图表 30 中国 AI 资本支出情况（单位：十亿美元） 20
图表 31 三大运营商 CAPEX 投入规模及增长趋势（单位：亿元，%） 20
图表 32 近年美国对华芯片部分限制 21
图表 33 我国 AI 芯片相关政策工程（部分） 22

仅供内部参考，请勿外传！

图表 34 全球龙头企业业务和经营情况比较（亿元）.....23
图表 35 2025 年中国 AI 加速卡出货量市场份额.....24
图表 36 GCU-CARE 加速计算单元与 GCU-LARE 片间互联架构.....25
图表 37 邃思产品谱系图.....26
图表 38 智算系统产品.....27
图表 39 服务成熟度检验证书.....27
图表 40 驭算 TopsRider AI 计算及编程软件平台.....28
图表 41 公司第一大客户销售金额和占比.....29
图表 42 潜在客户情况.....29
图表 43 公司全链条供应链布局.....30
图表 44 业务预测（百万元）.....31
图表 45 盈利预测敏感性分析（单位：百万）.....32

仅供内部参考，请勿外传！

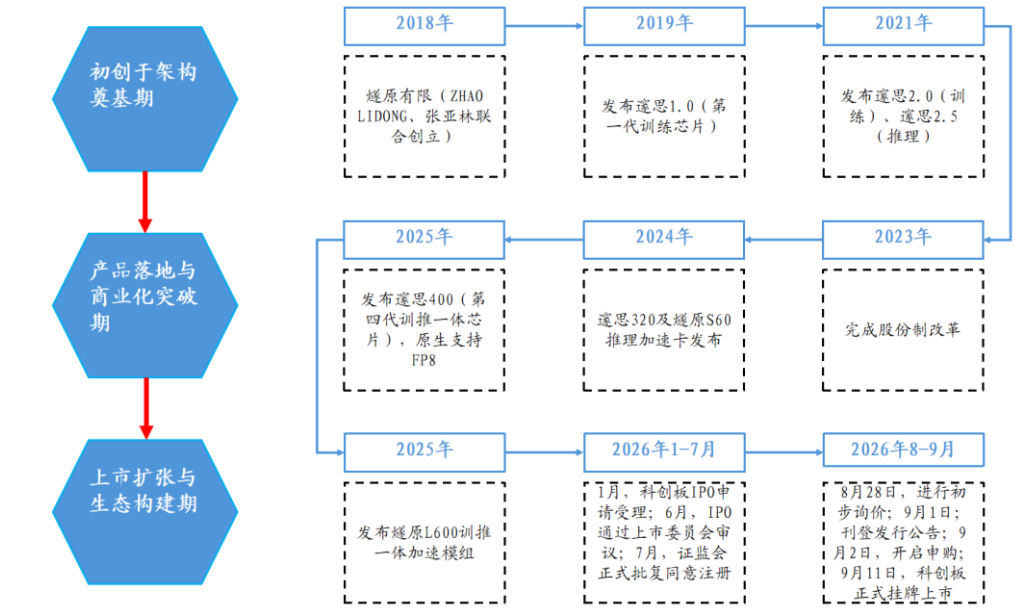
一、国产云端 AI 芯片领军者，商业化落地驱动收入增长

（一）专注云端芯片赛道，持续迭代构建全栈算力版图

云端 AI 芯片自主创领者，八载四代架构筑基科创板新征程。燧原科技成立于 2018 年 3 月 19 日，自创立以来始终专注于云端 AI 芯片的研发、设计和销售，致力于成为“通用人工智能基础设施领军企业”。公司坚持原始创新、自主研发的技术路线，已自研迭代了四代架构 5 款云端 AI 芯片。2026 年 1 月，公司科创板 IPO 申请获受理；6 月 15 日，IPO 通过上市委会议审议；7 月 9 日，证监会正式批复同意注册，审核状态变为注册生效；8 月 28 日，进行初步询价；9 月 1 日；刊登发行公告；9 月 2 日，开启申购；9 月 11 日，上海证券交易所科创板正式挂牌上市。公司发展历程可分为三个阶段：

- **初创与架构奠基期（2018-2021 年）：** 2018 年 3 月，ZHAO LIDONG 与张亚林作为联合创始人共同设立燧原有限。2019 年，公司发布第一代训练芯片“邃思 1.0”。2021 年，公司发布第二代训练芯片“邃思 2.0”及推理芯片“邃思 2.5”，完成训练与推理双产品线的初步布局。
- **产品落地与商业化突破期（2022-2024 年）：** 2023 年，公司完成股份制改造。2024 年，第三代推理芯片“邃思 320”及燧原 S60 推理加速卡发布。2025 年，公司第四代训推一体芯片“邃思 400”发布，原生支持 FP8 低精度计算。
- **上市扩张与生态构建期（2025-2026 年）：** 2025 年，公司基于邃思 400 芯片发布“燧原 L600”训推一体加速模组。2026 年 1 月，公司科创板 IPO 申请获受理；6 月，IPO 通过上市委会议审议，计划募集资金 60 亿元，投向第五代、第六代 AI 芯片系列产品的研发及产业化项目，以及先进人工智能软硬件协同创新项目。

图表 1 燧原科技发展历程



资料来源：燧原科技招股说明书、上海证券交易所公告、iFind、华创证券

公司股权结构多元均衡，治理架构稳定。公司无单一持股 30%以上的股东，不存在控股股东；实际控制人为 ZHAO LIDONG（赵立东）和张亚林。截至 2026 年 9 月 11 日，ZHAO LIDONG 直接持有公司 8.07%的股份，张亚林直接持有公司 8.07%的股份，两人通过燧原汇智与燧原崇英两个员工持股平台间接控制公司 9.18%的股份，合计控制公司 25.32%的

仅供内部参考，请勿外传！

股份。两人已签署一致行动协议，为公司共同实际控制人。腾讯作为战略投资者深度参与公司发展，为公司互联网场景商业化落地提供了重要的生态协同。腾讯科技直接持有公司 17.95%的股份，其一致行动人苏州湃益持有 0.28%，合计持有 18.23%，为公司第一大外部股东。自 2019 年起，腾讯即与公司开展业务合作，双方围绕业务场景适配优化、AI 模型性能调优、配套软件栈完善等关键领域持续深化合作，公司产品已在腾讯大量 AI 业务场景中实现规模化部署。

图表 2 燧原科技上市后股权结构图（截至 9 月 11 日）



资料来源：燧原科技招股说明书，Wind，华创证券

核心管理层深耕芯片产业，技术积淀构筑竞争壁垒。公司核心管理层均拥有国际头部芯片企业的资深研发背景。董事长 ZHAO LIDONG（赵立东）拥有清华大学学士及美国犹他州立大学硕士学位，曾任职于 AMD 并参与创立 AMD 中国研发中心，在芯片设计领域拥有 30 余年行业经验。总经理张亚林拥有复旦大学学士学位，曾任 AMD 中国研发中心技术总监，主导开发了微软 XBOX-ONE 系列主芯片等多款量产产品，拥有 25 年行业经验。核心技术人员方面，硬件芯片部负责人柴菁曾任 AMD 高级芯片研发经理，领导完成四代架构 5 款芯片的设计与量产，均实现一次流片成功；软件研发部负责人罗巍曾任英伟达 CUDA 测试开发和质量验证资深经理，主导自研 TopsRider 全栈软件平台；产品系统部负责人陈松涛曾任 Marvell 产品设计验证资深经理，主导 OAM 模组方案设计及量产。

图表 3 公司核心技术团队介绍

姓名	职务	年龄	履历摘要
ZHAO LIDONG（赵立东）	董事长、首席执行官、董事会秘书	60 岁	燧原科技创始人，美国国籍，清华大学电子工程学学士、美国犹他州立大学电子与计算机硕士。在硅谷工作超 20 年，先后服务于 S3 Inc.（GPU 图形处理器芯片研发）、Juniper Networks（网络安全芯片研发）。2007 至 2014 年任职 AMD，历任计算事业部高级总监、产品工程部高级总监，负责 CPU/GPU/APU 及多个核心 IP 研发，团队规模超千人，参与创立 AMD 中国研发中心。2014 年加入紫光集团，历任锐迪科微电子公司总裁、紫光集团副总裁。2018 年 3 月联合创立燧原科技。
张亚林	董事、总经理	48 岁	燧原科技联合创始人，复旦大学电子学与信息系统专业学士。2000 至 2007 年任职上海奇码数字信息有限公司，担任设计部主管，领导开发 ZJ 系列机顶盒芯片。2008 年加入 AMD，历任资深芯片经理、技术总监，作为全球芯片研发主要负责人之一，领导开发了微软 XBOX-ONE 系列主芯片及全球最大的融合芯片 APU 并一次量产成功。参与创立 AMD 上海研发中心融合芯片部门、AMD 北京研发中心。2018 年 3 月联合创立燧原科技。
姚磊文	董事	44 岁	硕士学历，金融、工商管理专业。2011 年 6 月至今任职腾讯科技（深圳）有限公司投资并购部副总经理。2018 年 7 月起担任燧原科技董事。
张敏	董事、副总经理	39 岁	本科学历。2011 年 7 月至 2018 年 7 月任职超威半导体（上海）有限公司（AMD），历任招聘助理、人力资源经理。2018 年 7 月加入燧原科技任首席人力官。2023 年 9 月起任董事，2025 年 11 月起任副总经理。

仅供内部参考，请勿外传！

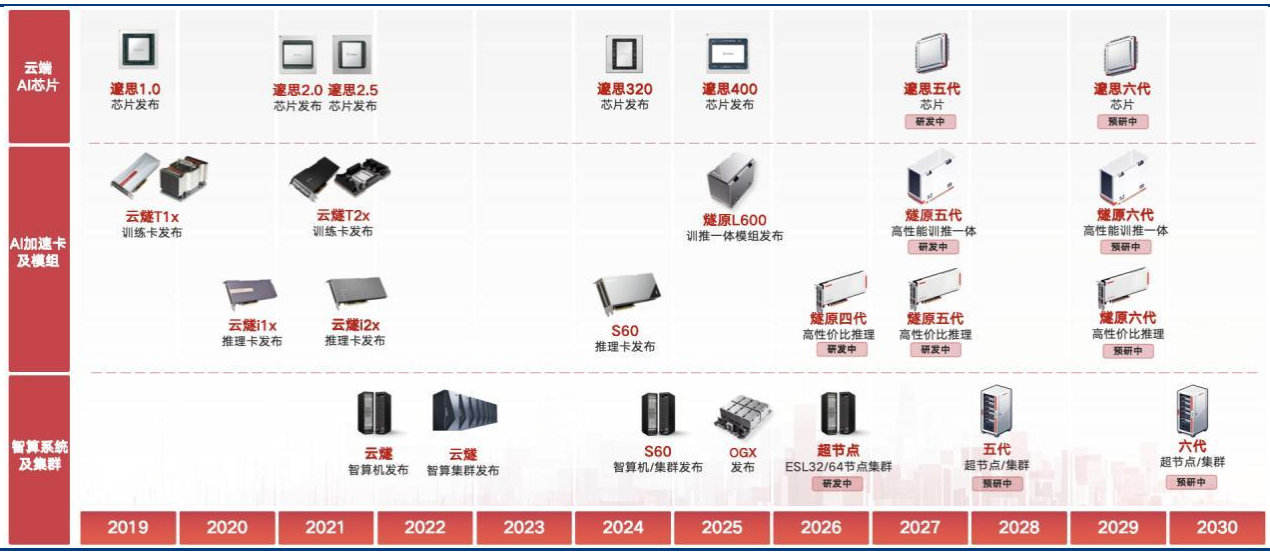
唐伟	董事、财务负责人	51 岁	硕士学历。曾任职中银国际、高盛高华、中金公司投资银行部。2016 年起先后担任四维图新首席财务官、亚信科技首席财务官兼董事会秘书、概伦电子首席财务官兼董事会秘书。2025 年 10 月入职燧原科技，任董事兼财务负责人。
柴菁	副总经理、核心技术人员	45 岁	硕士学历。2007 年 4 月至 2018 年 12 月任职超威半导体（上海）有限公司（AMD），担任高级芯片研发经理。2018 年 12 月加入燧原科技，任硬件芯片部负责人。领导完成公司四代架构 5 款芯片的设计与量产，均一次流片成功。2025 年 11 月起任副总经理。
高笑天	副总经理	44 岁	硕士学历。2007 年 3 月至 2010 年 3 月任职世芯电子（上海）有限公司；2010 年 3 月至 2018 年 6 月任职超威半导体（上海）有限公司（AMD），担任项目经理。2018 年 6 月加入燧原科技，任客户成功与解决方案工程部负责人。2025 年 11 月起任副总经理。
罗巍	核心技术人员	44 岁	本科学历。2007 年 9 月至 2021 年 6 月任职英伟达半导体科技（上海）有限公司，担任 CUDA 测试开发和质量验证资深经理。2021 年 7 月加入燧原科技，任软件研发部负责人。主导自研“驭算 TopsRider”全栈 AI 计算及编程软件平台。
陈松涛	核心技术人员	46 岁	硕士学历。先后任职泰瑞达（上海）、艾萨华科技（上海）、迈威迩电子科技（上海）有限公司（Marvell），担任产品设计验证资深经理。2019 年 3 月加入燧原科技，任产品系统部负责人。主导 OAM 模组方案设计及量产。

资料来源：燧原科技招股说明书、华创证券

四代五芯全场景覆盖，软硬协同构建云端 AI 算力闭环。公司产品路线图涵盖云端 AI 芯片、AI 加速卡及模组、智算系统三大层级，各产品线持续迭代，已形成从芯片到系统解决方案的完整布局。

- **芯片层面，**公司已累计推出五款已发布的芯片产品，分别为邃思 1.0、邃思 2.0、邃思 2.5、邃思 320 以及邃思 400，与此同时，公司下一代芯片产品正持续研发推进中，邃思五代和邃思六代均处于“研发中”状态。
- **加速卡及模组层面，**公司围绕不同应用场景构建了差异化的加速卡产品线。面向训练场景，已推出云燧 T1x 和云燧 T2x 两款训练加速卡；面向推理场景，已推出云燧 i1x 和云燧 i2x 两款推理加速卡；此外，S60 推理卡也已发布，进一步丰富推理产品阵容。
- **智算系统层面，**公司持续推进系统级产品的迭代升级。从早期的智算机产品起步，逐步演进至智算集群，现已完成从智算机到集群的系统化拓展。展望未来，公司已在研超节点 ESL32/64 产品，面向大模型时代的高密度算力需求，进一步向超节点与集群融合方向演进，构建覆盖单机柜到大规模集群的完整算力系统体系。

图表 4　燧原科技产品路线图



资料来源：燧原科技招股说明书

仅供内部参考，请勿外传！

芯片硬件、软件平台与算力集群三位一体，构筑自主可控技术壁垒。公司经过多年积累，形成了芯片及硬件、软件及编程平台和算力集群方案三大类核心技术体系。

- **硬件层面，**公司基于自主指令集，原创自主架构的 GCU-CARE 加速计算单元和 GCU-LARE 片间高速互连技术，公司是国内首家自主研发高密度、高互联带宽 OAM 模组方案的企业，四代 5 款芯片均一次流片成功。
- **围绕自研硬件，公司构建了全栈 AI 计算及编程软件平台“驭算 TopsRider”。**驭算平台是公司完全自主构建的统一 AI 计算及编程软件平台，涵盖了包括固件、驱动程序、编译语言与编译器的开发套件、包含算子库和通信库的 AI 计算扩展库和深度学习框架五个层次的基础软件。驭算具备生态兼容性好、性能卓越及能快速适配前沿模型的特点，使得中等复杂度的 CUDA 应用迁移成本仅需 1 人天。驭算已无缝支持 PyTorch、TensorFlow 等主流深度学习框架及多种分布式训练框架，已深度适配近千个 AI 模型，覆盖互联网及非互联网领域超过 300 个应用场景。
- **在此基础上，公司进一步向集群方案延伸。**截至 26 年 6 月份，千卡、万卡智算中心项目已成功落地，目前已联合客户研发超节点方案，并联合打造具有商业化价值的万卡高速互联集群，在甘肃庆阳建成的万卡集群为国内首个获中国信通院大规模智算集群服务成熟度五星级认证的国产万卡推理集群。

图表 5 “芯片—软件—集群”三位一体技术架构图



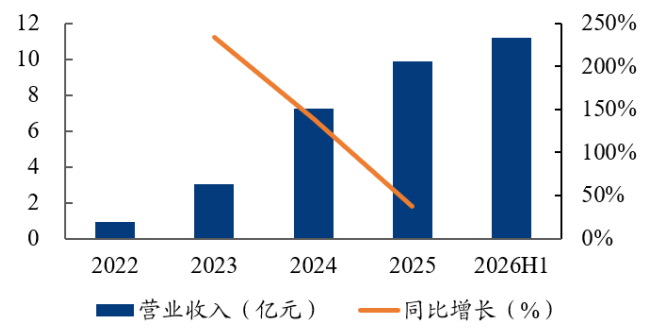
资料来源：燧原科技招股说明书

（二）营收高速增长，盈利结构持续优化

营收高速增长，亏损呈收窄趋势。2022-2025 年公司营业收入由约 0.9 亿元增长至 9.9 亿元，营收规模持续扩张，下游需求爆发为增长提供良好外部环境；2026 年第一季度实现营业收入约 2.87 亿元，下游需求爆发构成核心驱动；公司 2026 上半年实现营收约 11.2 亿元，即上半年收入水平已可覆盖 2025 年全年。盈利端，公司尚未实现盈利但亏损呈收窄态势：2023 年至 2025 年归母净利润分别为-16.65/-15.1/-11.64 亿元，扣非后归母净利润分别为-15.67/-15.03/-11.97 亿元，其中 2025 年归母净利润亏损较上年同期收窄 22.94%，反映经营效率逐步提升。

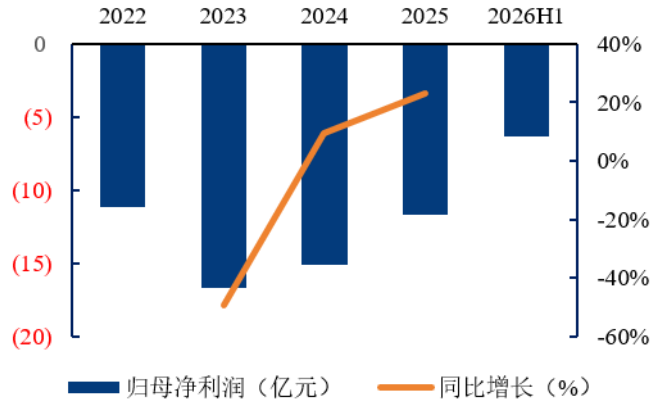
仅供内部参考，请勿外传！

图表 6 公司营业总收入



资料来源：iFind、华创证券

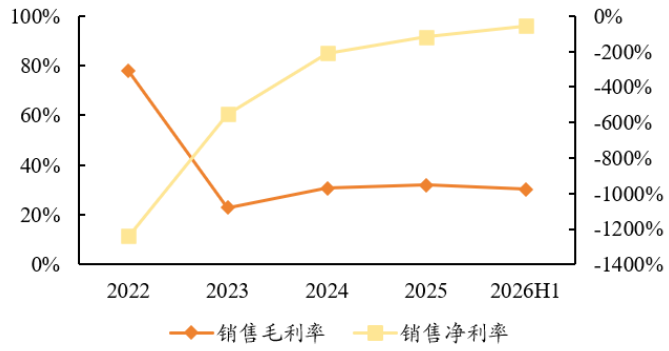
图表 7 公司归母净利润



资料来源：iFind、华创证券

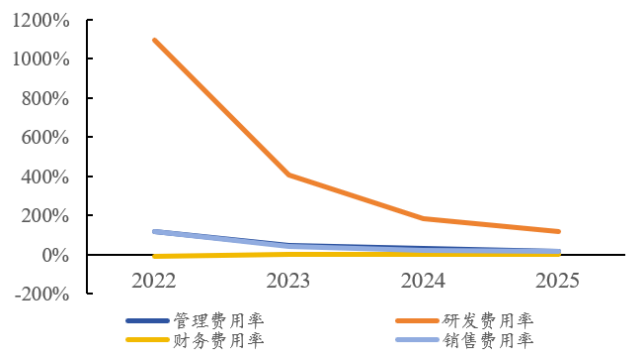
毛利率稳步提升，规模效应优化盈利结构。2023-2025 年，公司主营业务毛利率分别为 22.6%/30.6%/31.8%，伴随产品代际升级与出货放量，主营业务毛利率呈持续改善趋势。净利率方面，2023-2025 年分别为-552.7%/-209.1%/-117.5%，虽尚未实现盈利，但亏损幅度已显著收窄，反映出公司经营效率持续提升与盈利能力逐步修复，公司在 2026Q1 净利率回落到-154.8%，主要源自于公司保持较高强度研发支出，公司在 2026H1 净利率回升至-56.45%，反映出公司盈利能力修复延续。与此同时，公司收入规模快速扩张，规模效应逐步显现，各项费用率呈持续下行态势：2022-2025 年，管理费用率与销售费用率均显著下行，受益于公司收入的快速增长；财务费用率分别为-11.4%/-1.5%/0.8%/0.9%，整体维持低位。毛利率稳定、净利率好转与费用率全面改善共同推动亏损收窄，为盈利修复奠定基础。

图表 8 公司销售毛利率和净利率情况



资料来源：iFind、华创证券

图表 9 公司费用率情况



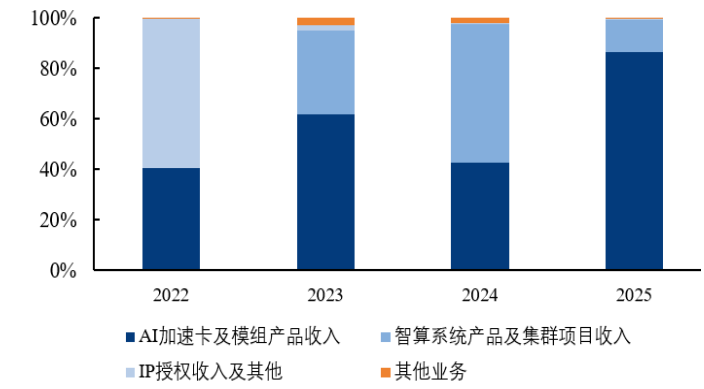
资料来源：iFind、华创证券

收入结构加速向核心产品集中，AI 加速卡及模组已成营收主力。公司主营收入主要分为 AI 加速卡及模组、智算系统及集群两大类。2023 年至 2025 年，AI 加速卡及模组收入占比分别为 61.7%/42.6%/86.5%，已成为公司收入增长的核心驱动力；智算系统及集群收入占比分别为 33.2%/55.2%/13.0%，主要系公司聚焦资源优先满足头部互联网客户采购需求；IP 授权及其他收入占比较小，各期分别为 2.0%/0.4%/0.2%。毛利率方面，2023 年至 2025 年 AI 加速卡及模组毛利率呈现先升后稳的走势，2023 年为 15.5%，2024 年大幅提升至 40.8%，主要系当期主力销售产品由第二代切换为定价更高的第三代 AI 加速卡，该产品性能更优且 AI 模型适配性更广；2025 年该板块毛利率回落至 32.7%，主要原因为头部互

仅供内部参考，请勿外传！

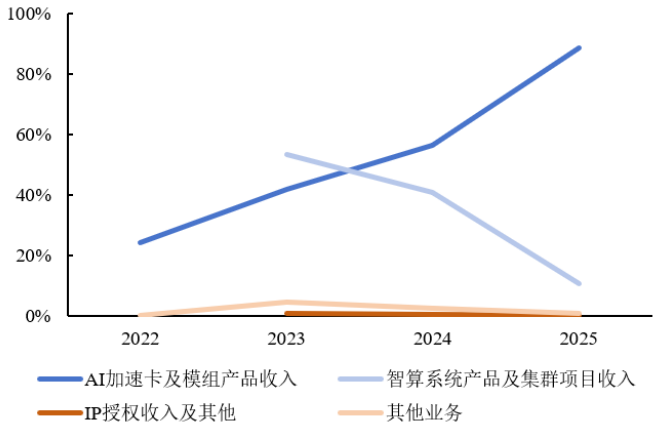
联网客户采购持续放量带来销售价格优惠，同时存储颗粒等核心物料价格波动推升单位成本。

图表 10 公司主营产品营收占比



资料来源：iFind、华创证券

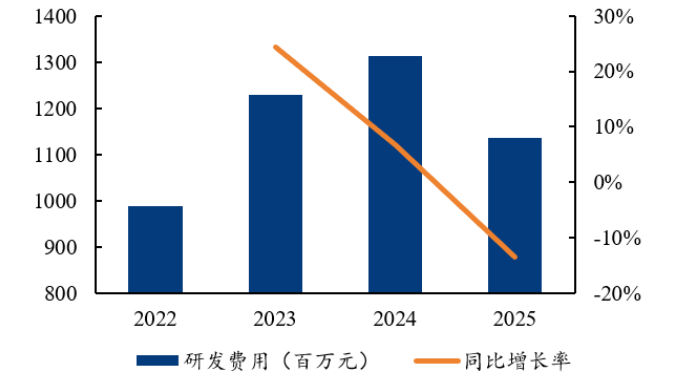
图表 11 公司主营产品毛利占比



资料来源：iFind、华创证券

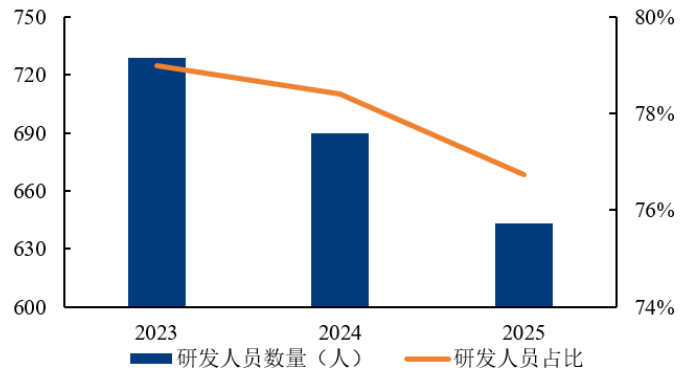
研发投入维持高位，高配比研发团队为技术迭代筑牢根基。2022-2025 年，公司研发费用分别为 9.9 亿元、12.3 亿元、13.1 亿元和 11.3 亿元，整体保持较高规模。其中，2023 年研发费用同比增长约 24.4%，2024 年同比增长约 6.8%，2025 年同比下降约 13.5%，考虑公司产品迭代等因素，公司预计 2026 年公司研发费用将高于往期任一年度规模。截至 2025 年末，公司研发人员为 643 人，占员工总数的 76.7%，高比例的研发人员配置充分彰显公司对技术创新的高度重视，为公司芯片架构的持续迭代及软件平台的不断完善提供了有力支撑。

图表 12 研发费用情况



资料来源：iFind、华创证券

图表 13 研发人员及其占比情况



资料来源：燧原科技招股说明书、华创证券

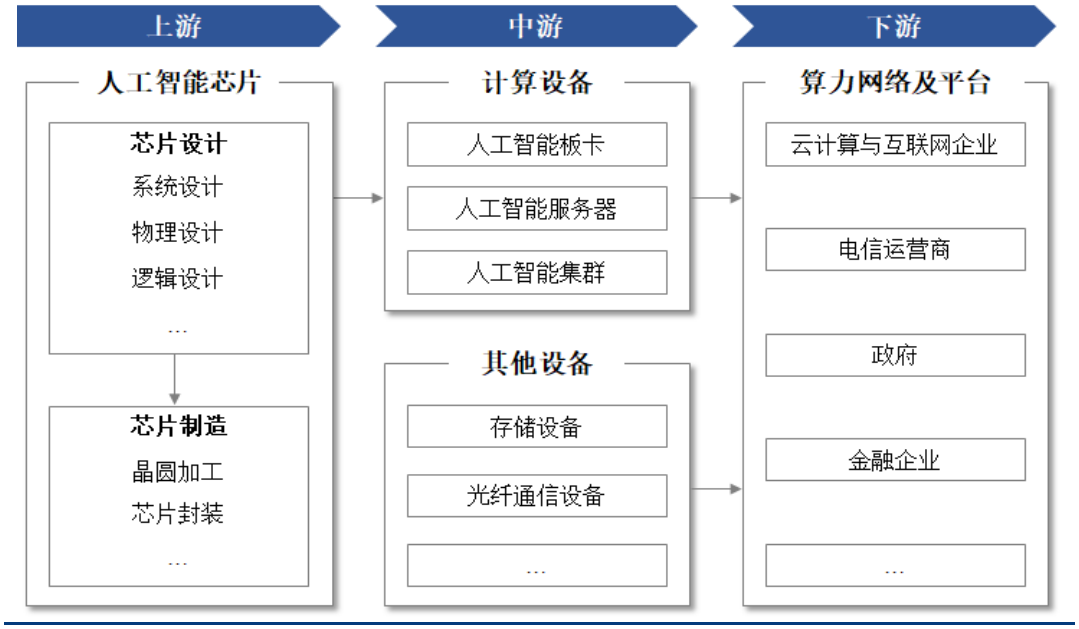
仅供内部参考，请勿外传！

二、AI 加速芯片市场需求长期向好，国产芯片加速替代突围

（一）产业链：芯片设计为核，云端 AI 芯片构筑算力底座

人工智能算力基础设施产业链可分为上游、中游和下游三部分。上游为人工智能芯片环节，涵盖芯片设计和芯片制造两大板块。芯片设计包括系统设计、物理设计和逻辑设计；芯片制造包括晶圆加工、芯片封装等工序，为芯片产品提供基础硬件支撑。**中游为计算设备及其他设备环节。**计算设备包括人工智能板卡、人工智能服务器和人工智能集群等，是算力输出的核心载体；其他设备包括存储设备和光纤通信设备等，保障数据的高效存取与传输。**下游为算力网络及平台环节**，主要由云计算与互联网企业、电信运营商、政府及金融企业等构成，是 AI 算力最终落地与应用的关键出口。整体来看，中游的计算设备制造是产业链的核心环节，决定行业技术路线与竞争格局；上游的芯片设计为中游提供底层硬件基础，下游的应用平台则为中上游提供需求牵引与场景落地空间，三大环节协同推动 AI 算力产业链的完整运转。

图表 14　人工智能算力基础设施产业链



资料来源：燧原科技招股说明书

1、发展趋势一：云端 AI 芯片主导地位稳固

AI 芯片分类：按部署位置划分为云端、边缘端与终端，应用场景各有侧重。AI 加速卡按部署位置可分为云端、边缘端和终端三类，分别对应不同的算力需求层级与应用场景。**1）云端 AI 芯片**部署于 AI 数据中心，在实时联网环境下运行，承担 AI 大模型训练及多用户、多任务场景下的大规模推理任务，对算力密度、内存带宽及互联效率要求最高，是当前 AI 芯片市场竞争的主战场。**2）边缘 AI 芯片**应用于靠近智能终端的边缘服务器，在弱网或断网环境下仍可正常工作，典型应用场景包括工业网关、5G 基站、摄像头集群服务器等，侧重低延迟、低功耗与实时响应能力，是工业互联网和智能制造落地的关键算力节点。**3）终端 AI 芯片**直接嵌入智能手机、智能耳机、智能摄像头等终端设备中，不依赖网络可离线工作，侧重于极低功耗、小尺寸与成本控制，以满足消费级产品的规模化部署需求。三类芯片在算力层级、部署环境与应用场景上形成互补，共同构成了从云端到边缘再到终端的全场景 AI 算力体系。

仅供内部参考，请勿外传！

图表 15 AI 芯片分类示意图（按部署位置）



资料来源：燧原科技招股说明书、华创证券

三类芯片在承载任务、性能特征与功耗上呈现明显的梯度差异。1）云端 AI 芯片处理海量数据，用于 AI 大模型训练和多用户、多任务场景下的大规模推理，性能要求最高、功耗最大（典型功耗为几百瓦至千瓦级），除单卡算力外还需支持多卡高速互联与算力集群扩展，技术复杂度最高，是三类芯片中门槛最高的品类。**2）边缘 AI 芯片**基于局部网络可同时处理多个终端数据，但任务通常局限于单一领域，以推理为主，需要中等算力和低数据延迟性能，一般不要求芯片高速互联与算力集群扩展，技术复杂度适中，典型功耗为几瓦至几十瓦，在性能与功耗之间取得平衡。**3）终端 AI 芯片**处理单个终端数据，只能进行推理任务，算力性能要求最低，强调功耗小（毫瓦至瓦级）、体积小和实时响应，技术复杂度相对较低，是三类芯片中最强调低功耗和小型化的品类。整体来看，从云端到边缘再到终端，三类芯片在算力性能、功耗水平与技术复杂度上呈现逐级递减的梯度分布，分别对应不同层级与应用场景的差异化需求。

图表 16 云端 AI 芯片 vs 边缘 AI 芯片 vs 终端 AI 芯片对比

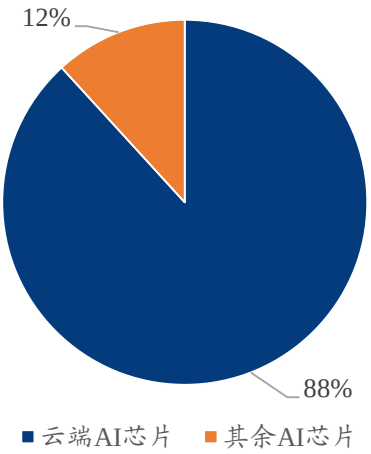
项目	云端 AI 芯片	边缘 AI 芯片	终端 AI 芯片
网络应用场景	应用于 AI 数据中心，实时联网环境下工作	应用于靠近智能终端的边缘服务器，弱网/断网环境下也能工作，例如工业网关、5G 基站、摄像头集群服务器等场景	直接嵌入智能手机、智能耳机、智能摄像头等终端设备，不依赖网络，可离线工作
承载任务	处理海量数据，用于 AI 大模型训练和多用户、多任务场景下的大规模推理场景	基于局部网可同时处理多个终端数据，往往基于单一领域任务，任务以推理为主	处理单个终端的数据，只能进行推理任务
技术/性能特征	性能要求最高，功耗最大，除强调单卡性能外，还需支持多卡高速互联和算力集群扩展，技术复杂度最高	需要中等算力和低数据延迟性能，一般不要求芯片高速互联和算力集群扩展，技术复杂度适中	算力性能要求最低，强调功耗小、体积小和实时响应，技术复杂度相对较低
典型功耗	几百瓦至千瓦级	几瓦至几十瓦	毫瓦至瓦级

资料来源：燧原科技招股说明书、华创证券

云端 AI 芯片主导中国 AI 芯片市场，规模优势显著。根据灼识咨询统计，2025 年中国 AI 芯片市场中，云端 AI 芯片系列产品占据约 88%的主导份额，市场规模约 3,338 亿元；边缘 AI 芯片与终端 AI 芯片两者合计占比约 12%，市场规模合计约 446 亿元。整体来看，云端 AI 芯片凭借其在大模型训练与大规模推理场景中的核心地位，在市场中占据绝对主导份额，其庞大的市场规模与高占比充分体现了云端算力在 AI 产业中的关键支撑作用。

仅供内部参考，请勿外传！

图表 17　2025 年中国 AI 芯片市场结构（按部署位置）

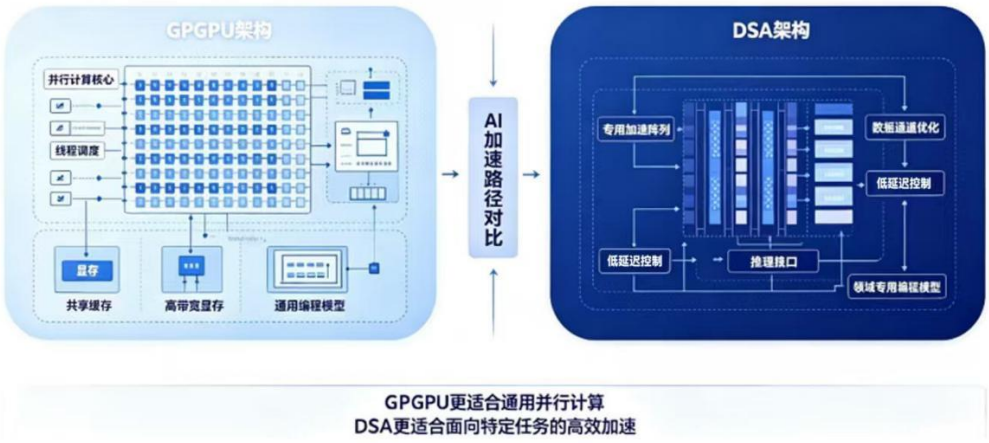


资料来源：燧原科技招股说明书、华创证券

2、发展趋势二：推理落地驱动 DSA 需求高增

AI 芯片架构分野：GPGPU 通用生态主导，DSA 专用优化崛起。目前 AI 芯片的主流架构分为 GPGPU 架构和 DSA 架构两大类。**1）GPGPU 架构**以英伟达为代表，通过强化通用并行计算单元形成了成熟的通用编程生态，可覆盖 AI 计算、科学计算、图形渲染等多个场景，但针对 AI 计算这一特定领域的效率存在一定损耗。**2）DSA 架构（即非 GPGPU 架构）**则针对 AI 并行计算领域进行专用优化，在相同芯片面积上通过牺牲场景通用性来提升特定应用场景下的效率，强化 AI 计算专用加速单元配置，加速效率更高。此外，AI 芯片领域的 ASIC 是一个广义概念，最严格意义的 ASIC 芯片等同于场景完全定制化（如将特定大模型固化在芯片硬件架构中），而 DSA 架构是 ASIC 概念的发展——其基于 AI 并行计算这一领域进行加速优化，但并不固化在某个特定场景或模型，而是兼顾 AI 计算领域内多种算法和模型的通用性。

图表 18　AI 芯片主流架构对比



资料来源：燧原科技招股说明书、华创证券

架构迁移驱动路线分化，GPGPU 生态兼容与 DSA 专用优化两轨并行。AI 芯片架构的迁移趋势引导国内厂商形成两条主流路径的分化。摩尔线程、沐曦股份、壁仞科技、海光信息等选择 GPGPU 路线，以兼容 CUDA 生态为核心策略，在通用计算场景下具备生态

仅供内部参考，请勿外传！

迁移成本低的优势。燧原科技、寒武纪及华为昇腾则选择 DSA/NPU 路线，针对 AI 特定负载进行架构优化，在能效比和单位算力成本上具备差异化竞争力。其中，华为昇腾以自研达芬奇架构起家；寒武纪为 DSA 先行者；燧原科技是国内少数自成立以来始终坚守 DSA 路线、基于自主指令集构建完整 GCU 架构并独立于 CUDA 生态的云端 AI 芯片厂商，通过自研 GCU-CARE 加速计算单元与 GCU-LARE 片间高速互连技术，在推理场景展现出相较 GPGPU 更优的性价比优势。

图表 19 全球 AI 加速卡不同技术路径主要厂商构成情况



资料来源：燧原科技招股说明书

图表 20 芯片厂商架构对比

厂商	架构	技术路线	核心计算单元	互联技术	生态策略
英伟达	Ampere/Hopper/Blackwell	GPGPU	CUDA Core + Tensor Core	NVLink	CUDA 生态
燧原科技	GCU	DSA	GCU-CARE 加速计算单元对标 Tensor Core	GCU-LARE	“驭算 TopsRider”（独立于 CUDA）
华为昇腾	达芬奇架构	NPU	Cube Unit 矩阵计算单元	灵衢 UB（新代际）	CANN
寒武纪	MLUarch03	DSA	张量运算单元	MLU-Link	Cambricon NeuWare
摩尔线程	MUSA（苏堤→春晓→曲院→平湖→花港）	GPGPU	MUSA Processor	夸娥万卡集群	MUSA 生态
沐曦股份	自主 GPU IP 架构	GPGPU	标量+矢量+张量计算单元	MetaXLink 超节点扩展	MXMACA（兼容 CUDA 生态）
壁仞科技	壁立仞架构	GPGPU	SPC 流处理器+专用张量 T-Core	BLink2.0 超节点互连协议	BIRENSUPA 软件平台
海光信息	DCU	GPGPU	通用并行计算架构	HSL 互联总线协议	兼容“类 CUDA”

资料来源：各公司公告及官网、阿里云、新浪财经、鲲鹏昇腾开发者社区、电子工程世界、D&R 社区、人民网、福布斯中国、华创证券

境外 GPGPU 主导，境内 DSA 份额领先。境外市场 GPGPU 架构份额超过 85%，英伟达居主导地位，但 DSA 架构正加速渗透。2025 年 11 月，谷歌发布依靠自研第六代 Trillium TPU（DSA 架构）训练的 Gemini 3 大模型，验证了 DSA 架构支撑万亿参数大模型全流程训练的能力，部分头部互联网厂商及大模型厂商已开始采购谷歌 TPU 产品。境内市场 GPGPU 仍占超过 40%份额，但 DSA 架构占比远高于境外。华为海思为 DSA 龙头及国产出货量最大厂商，寒武纪、昆仑芯、燧原科技等亦为主要参与者。

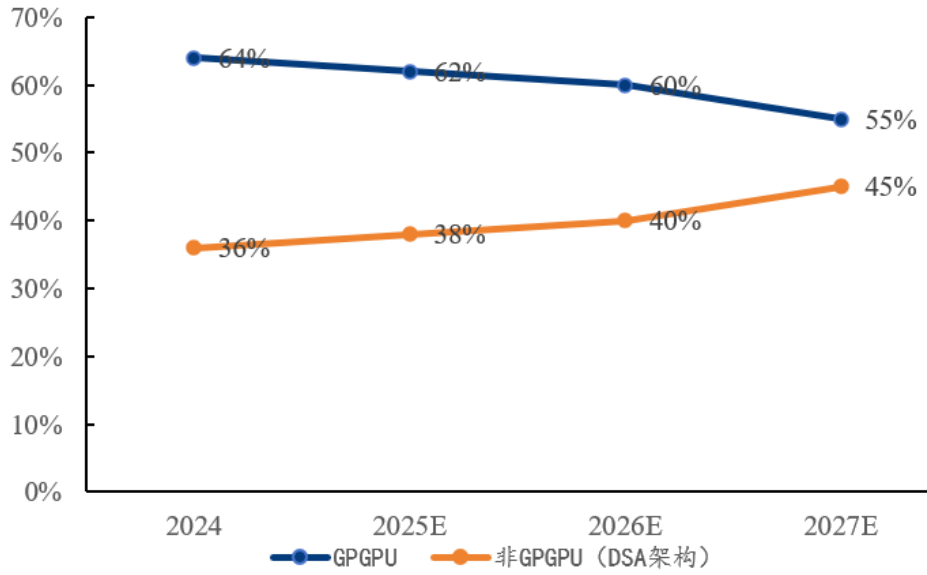
图表 21　不同架构类型 AI 芯片境内外市场份额对比

市场	境外云端 AI 芯片市场		境内云端 AI 芯片市场	
类别	GPGPU	非 GPGPU（DSA 架构）	GPGPU	非 GPGPU（DSA 架构）
2025 年市场规模及占比	1690 亿美元，占比 86%	285 亿美元，占比 14%	1090 亿元人民币，占比 43%	1420 亿元人民币，占比 57%
主要参与者	英伟达、AMD 等	谷歌、亚马逊、微软等自研芯片	英伟达、海光信息、平头哥、摩尔线程、沐曦股份、壁仞科技、天数智芯等	华为海思、寒武纪、燧原科技、昆仑芯等

资料来源：燧原科技第二轮审核问询函的回复、华创证券

DSA 架构份额持续提升，推理性价比优势凸显。AI 大模型推理需求持续提升，具有更高效能、更大数据吞吐量和更低功耗特性的 DSA 架构算力产品加速部署。随着 DeepSeek 等开源模型兴起，AI 底层大模型开发者逐步收敛至少数头部厂商，AI 加速卡所需适配的模型生态趋于集中，市场对 CUDA 生态的依赖有望减弱，能效比与性价比正成为芯片选型的核心指标之一。在此背景下，DSA 架构在推理场景中展现出相较于 GPGPU 更优的性价比优势，市场份额有望持续提升。根据高盛全球投资研究部预测，全球 AI 服务器中非 GPGPU 架构（主要为 DSA 架构）芯片的出货占比或将从 2024 年的 36%逐步提升至 2027 年的 45%。

图表 22　全球 AI 芯片中 GPGPU 架构和 DSA 架构比例趋势



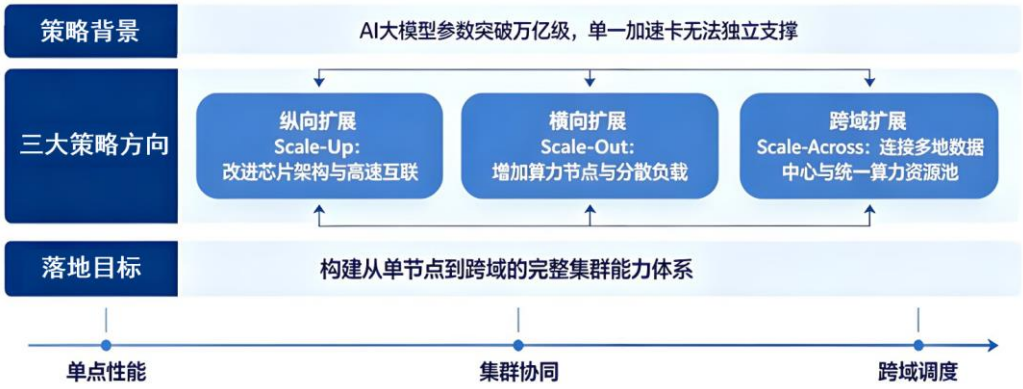
资料来源：燧原科技招股说明书、华创证券

仅供内部参考，请勿外传！

3、发展趋势三：AI 加速卡基础设施向集群化和综合化演进

从单卡算力到系统集群，三大扩展策略协同突破算力瓶颈。随着 AI 大模型参数规模扩展至数千亿甚至万亿级别，单一 AI 加速卡已难以满足超大规模计算需求，行业供应重点正从单点性能提升转向系统化集群解决方案。现代智算集群主要通过三项关键策略协同应对挑战：**一是纵向扩展策略（Scale-Up），**即通过改进 AI 加速卡芯片架构、采用更先进制程提升单卡峰值算力，或通过高速互联架构（如英伟达 NVLink、公司 GCU-LARE）实现芯片间超低延迟通信，提升节点内部数据流通效率。**二是横向扩展策略（Scale-Out），**即通过增加算力节点分散负载、提升处理能力，用于大规模数据集与复杂模型训练场景。但集群规模从千卡扩展至万卡、十万卡时，节点故障几率随之上升，数据中心需配备高效监控体系与先进故障恢复机制，基于智能显存分配、故障点恢复管理等技术确保节点故障时快速响应、最小化停机时间。**三是跨区域扩展（Scale-Across），**即通过高性能网络将地理上分散的多个数据中心连接为统一计算资源池，突破单数据中心物理限制，该策略依赖光通信技术与网络优化算法解决长距离传输的带宽与延迟问题。

图表 23　AI 算力集群三大扩展策略



资料来源：燧原科技招股说明书、华创证券

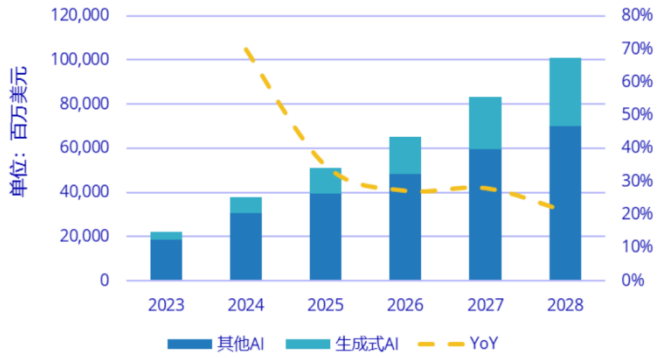
（二）需求端爆发：GPU 市场规模快速增长，云厂商运营商 AI 相关 CAPEX 提速

1、AI 大模型拉动 GPU 需求大幅提升

我国 AI 大模型市场蓬勃发展。从全球市场来看，IDC 数据显示，2024 年全球生成式 AI 市场五年 CAGR 或达 63.8%；到 2028 年，该市场规模将达到 2842 亿美元，占 AI 市场投资总规模的 35%，市场正处于快速扩张阶段。聚焦我国市场，2024 年我国生成式 AI 占 AI 市场投资总规模的 18.9%。伴随生成式 AI 技术的快速发展，这一占比将持续提升，预计到 2028 年，我国生成式 AI 投资占比将达到 30.6%，投资规模超 300 亿美元，2024-2028 年 CAGR 为 51.5%。从长期发展动力来看，AI 技术的创新迭代不断推动应用场景进一步落地，其中 AIGC、多模态、智能决策等热点方向，为市场拓展带来了更多想象力与可能性。同时，“数字化”、“数智化”转型背景下，AI 大模型被政府、企业广泛关注并积极推动，为我国 AI 市场规模的长期增长奠定了基础。从行业应用来看，预测期内 AI 支出主要来自专业服务领域的行业用户、政府和金融行业，合计约占市场总量 50%以上。增长速度来看，银行和地方政府增速最快，2024-2028 年 CAGR 均超 23%。

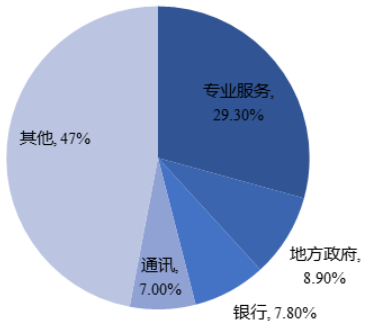
仅供内部参考，请勿外传！

图表 24 中国 AI 与生成式 AI 市场规模预测（2023-2028 年）



资料来源：IDC 中国《全球人工智能和生成式人工智能支出指南》

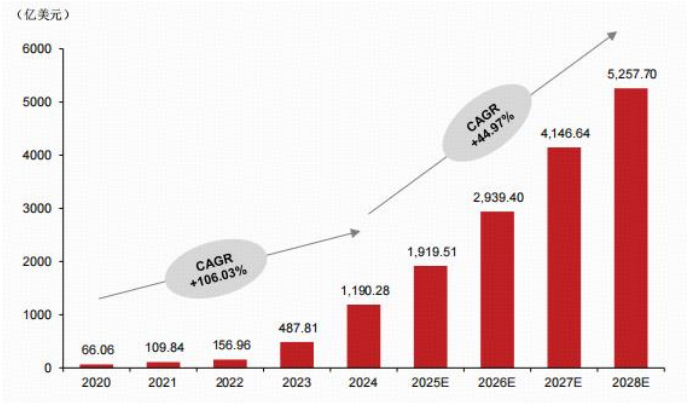
图表 25 中国人工智能市场支出预测（行业）2026



资料来源：中国互联网数据资讯网转引自 IDC 中国、华创证券

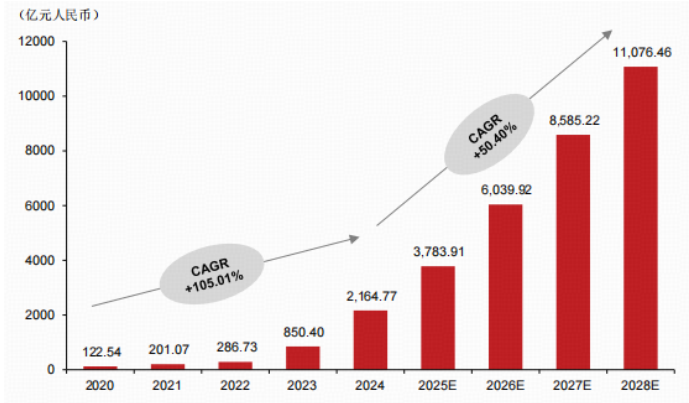
全球 AI 加速卡市场高速扩容，中国需求 2028 年有望占全球 30%。云端 AI 芯片通常以 AI 加速卡形式销售，或进一步集成至智算系统及集群售予客户。在互联网巨头及 AI 大模型企业持续加大人工智能资本开支的驱动下，全球 AI 加速卡市场规模快速增长。据灼识咨询数据，2024 年全球 AI 加速卡市场规模约 1,190.28 亿美元，预计 2028 年将达 5,257.70 亿美元。中国 AI 加速卡市场从 2020 年 122.54 亿元增长至 2024 年 2,164.77 亿元，CAGR 达 105.01%；预计 2028 年将达 11,076.46 亿元，2024 至 2028 年 CAGR 为 50.40%，届时约占全球市场的 30%。增长主要源于 AI 大模型训练与推理对算力的强烈需求，以及国家产业政策与“东数西算”战略对数据中心建设的加速推动。

图表 26 全球 AI 加速卡市场规模及预测



资料来源：灼识咨询，转引自燧原科技招股说明书

图表 27 中国 AI 加速卡市场规模及预测



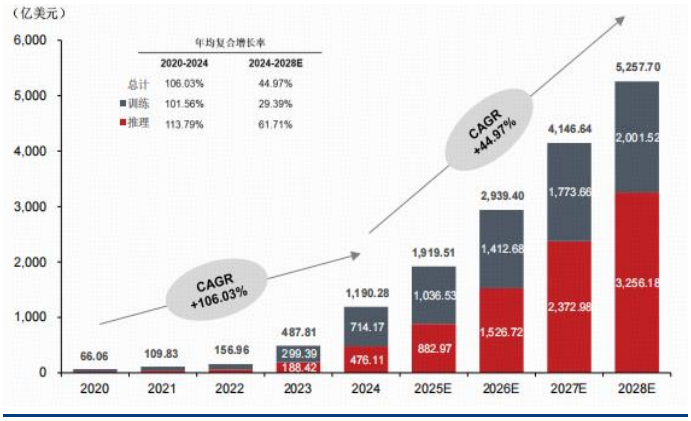
资料来源：灼识咨询，转引自燧原科技招股说明书

推理需求加速释放，中国市场 2028 年推理用 AI 加速卡占比有望超 70%。大模型行业正从大规模训练阶段逐步走向推理落地阶段，预计 2026 年全球 AI 推理对 AI 加速卡的需求将超过 AI 训练场景，中国市场这一趋势更为显著。据灼识咨询预测，全球 AI 加速卡推理用需求将从 2024 年 476.11 亿美元增长至 2028 年 3,256.18 亿美元，复合增长率 61.71%，增速远超整体市场均值。中国市场方面，依托 DeepSeek 等模型性价比提升及丰富产业生态与流量优势，AI 应用实现规模化发展；互联网厂商凭借庞大存量客户群体，AI 应用上线即形成大规模推理需求；DeepSeek、千问等主流模型开源化亦降低了非头部厂商预训练投入门槛，进一步推升推理算力占比。从细分结构看，中国训练用 AI 加速卡

仅供内部参考，请勿外传！

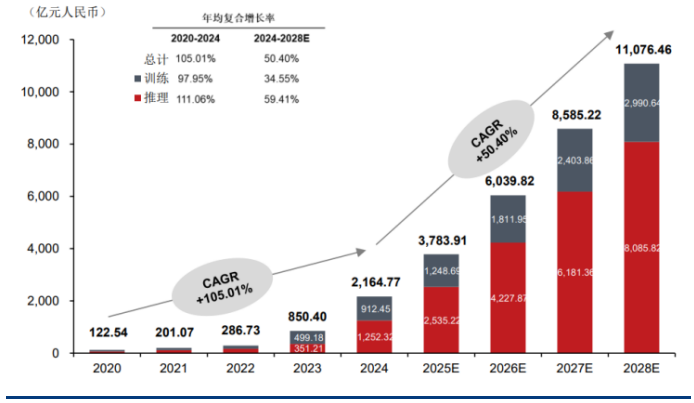
市场规模 2020 至 2024 年复合增速 97.95%，预计 2028 年达 2,990.64 亿元；推理用 AI 加速卡同期年均增速达 111.06%，未来四年有望以 59.41%增速扩张至 8,085.82 亿元，占比提升至超过 70%。

图表 28 全球训练和推理 AI 加速卡市场规模



资料来源：灼识咨询，转引自燧原科技招股说明书

图表 29 中国训练和推理 AI 加速卡市场规模



资料来源：灼识咨询，转引自燧原科技招股说明书

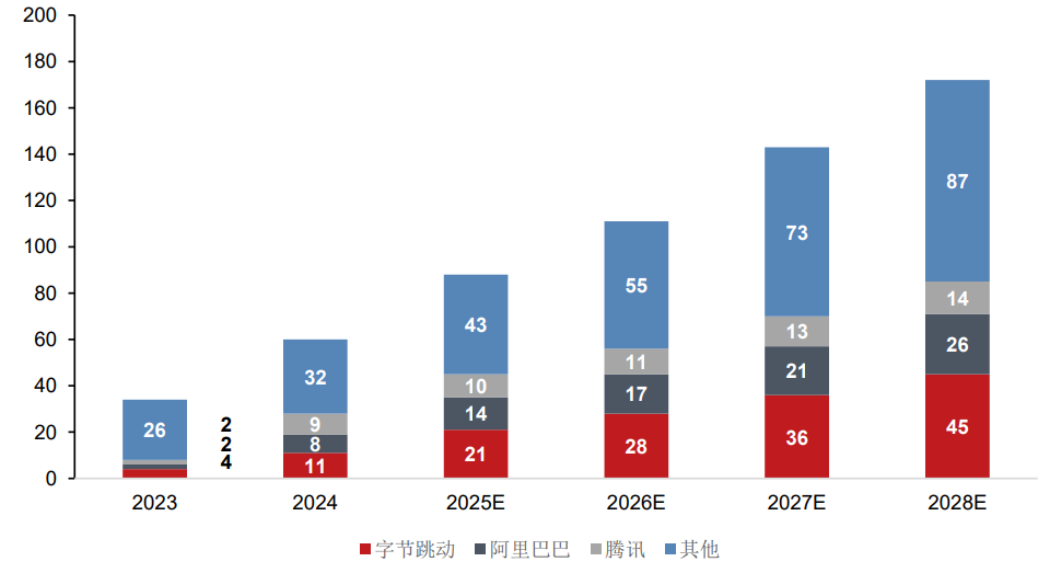
2、国内云巨头与运营商共同加码 AI 算力 CAPEX

国内云厂商 CAPEX 投入提速，AI 算力成为重点布局方向。伴随着大模型参数持续扩张、AIGC 应用爆发及产业智能化进程加速，国内头部互联网厂商 AI 资本支出进入高速增长通道。与此同时，市场呈现显著的头部集中特征，中小企业更倾向于通过云服务弹性租赁算力，而头部互联网企业因自身大模型训练及海量推理需求，正持续加大 AI 算力投入。根据 BERNSTEIN 预测，字节跳动、阿里巴巴和腾讯三大头部互联网厂商的 AI 资本支出将持续攀升，到 2028 年三者合计将占据中国 AI 资本支出的近 50%。从趋势上看，字节跳动自 2024 年起保持 AI 领域高投入节奏，阿里巴巴 2025 年起资本开支显著提速，腾讯则维持稳定增长。在头部厂商持续投入的带动下，国内 AIDC 基础设施建设进入加速期。

- **字节跳动**：公司在 AI 领域持续加大投入，2024 年旗下子公司的火山云太行算力中心整个项目投资额约达 73 亿元，占地约 530 亩，分两期建设，一期已建成投用，二期原预计 2026 年 5 月竣工；
- **阿里巴巴**：2025 年全年阿里的资本性支出约为 1238 亿元，而从 2025 到 2027 年，阿里将投入超过 3800 亿元，用于建设云和 AI 硬件基础设施，总额超过去十年总和，创下中国民营企业在云和 AI 硬件基础设施建设领域有史以来最大规模投资纪录；
- **腾讯**：2025 年，腾讯营业收入为 7517.66 亿元，同比增长约 13.86%，资本性支出为 791.98 亿元，较上一年增长约为 3.18%，虽然增幅不高，但开支维持在一个较高水平。

仅供内部参考，请勿外传！

图表 30　中国 AI 资本支出情况（单位：十亿美元）

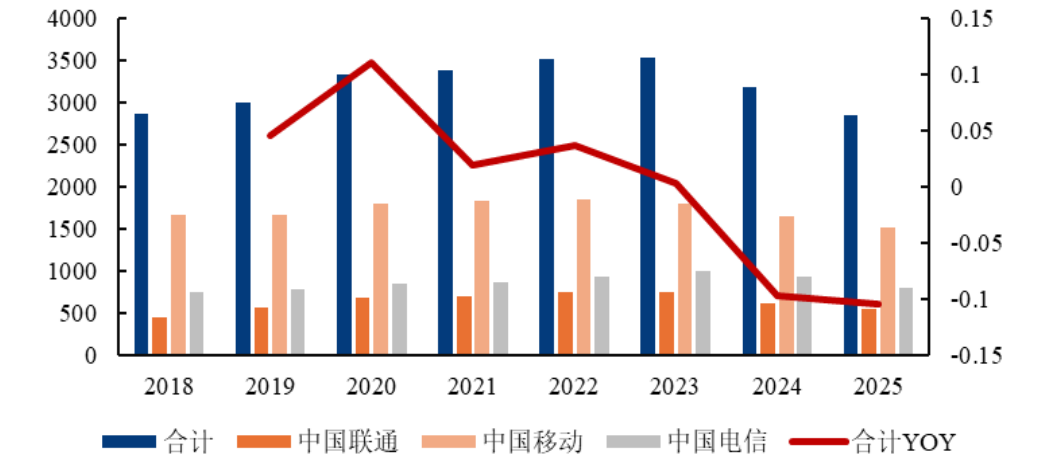


资料来源：BERNSTEIN，转引自燧原科技招股说明书

运营商整体 CAPEX 回落，但算力相关投入逆势增长。2024 年，三大运营商资本开支总额约 3188.7 亿元，较前期略有收缩；2025 年三大运营商投资规模合计降至 2855 亿元，延续小幅压缩趋势。尽管整体 CAPEX 下调，但投资结构持续优化，传统网络建设占比逐步下行，算力与 AI 相关投入则实现逆势提升，仍为运营商适度超前布局的核心方向。

- **中国移动**：2025 年，公司资本开支为 1509 亿，同比下降 8.0%。2026 年公司预计资本开支合计约为 1366 亿元，主要用于确保通信网络领先、加速算力网络建强、探索智能网络创新以及优化端到端感知等方面。
- **中国联通**：2025 年，资本开支 542 亿，同比下降 11.68%，资本开支占主营业务收入比下降至 16%。2026 年资本开支预计在人民币 500 亿元左右，其中算力投资占比超过 35%。
- **中国电信**：2025 年，公司资本开支 804 亿，同比下降约 14%。其中算力基础设施面向长远，加强资源保障，投资人民币 202 亿元。

图表 31　三大运营商 CAPEX 投入规模及增长趋势（单位：亿元，%）



资料来源：各公司财报及公开业绩说明会，华创证券

（三）供给端重构：美国政府再升级高端芯片出口限制，我国芯片代际差持续缩小

美国限制措施不断升级，加速国内 AI 芯片自主可控进程。近年来，针对芯片、先进计算等领域，美国通过出口管制、实体清单，法案等方式出台了一系列限制措施，包括调整高性能芯片受限参数、防止芯片厂商绕过限制等，芯片管制措施持续升级。据摩尔线程招股说明书数据，寒武纪、海光信息等公司的 AI 芯片产业化程度较高，并已在国内市场实现规模化营收。美国高端 GPU 出口限制逐步加强的背景下，中国企业推进自主可控的步伐正不断加快。

图表 32 近年美国对华芯片部分限制

发布机构	发布时间	制裁	制裁内容
BIS	2026 年 5 月 31 日	出口管制	针对总部位于中国内地或中国澳门地区的企业及其境外子公司，对 ECCN 3A090/4A090 等先进 AI 芯片的出口许可实行“推定拒绝”审查，堵截通过第三国子公司规避管制的路径。该要求自 2023 年 11 月起已存在，未因 2025 年 5 月暂停《AI 扩散规则》而改变。
美国国会两党	2026 年 4 月	出口管制	鉴于中国在成熟（旧）制程和半导体基础代工方面的发力，美国国会两党议员于 4 月上旬提出《硬件技术控制多边协调法案》（MATCH Act）。该法案一旦通过，将把对华芯片禁运的矛头从“先进算力/EUV”延伸到“所有通用芯片必需的深紫外线（DUV）光刻机”以及刻蚀设备，并切断相关设备的售后维修服务，旨在彻底瘫痪相关企业的生产线。
BIS	2026 年 1 月 15 日	出口管制	对英伟达 H200、AMD MI325X 等顶级 AI 算力芯片对华出口，由“推定拒绝”转为“逐案审批”。
参议院	2025 年 12 月 4 日	出口管制	参议院小组公布了一项法案，名为（SAFE CHIPS Act）将要求负责出口管制的商务部在 30 个月内拒绝向中国、俄罗斯、伊朗或朝鲜的买家发放任何比目前允许获得的更先进的美国人工智能芯片的许可证申请。
特朗普政府	2025 年 5 月 15 日	出口管制	特朗普政府正式撤销了拜登时期的“全球配额”系统，放宽了对沙特、阿联酋等中东盟国获取 AI 芯片的数量限制，以换取经济利益，但同步加强了对中东地区运往中国的“再出口”审查。
BIS	2025 年 3 月 25 日	实体清单	重点制裁 11 家中国大陆实体和 1 家中国台湾地区实体（包括浪潮集团的多个关联子公司），打击理由是这些企业参与了先进 AI、超级计算机（特别是百亿亿次级计算）的开发，或者试图获取美国原产芯片用于军事现代化。
BIS	2025 年 1 月 13 日	出口管制	将全球划分为三个梯队，对全球各国的算力中心和云厂商设立了“GPU 购买数量上限”，要求严密监控海外算力去向，确保中国无法通过“云租赁”或设立海外空壳公司绕过管制。
拜登政府	2024 年 12 月 2 日	出口管制	由于 AI 模型训练极度依赖显存带宽，美国首次将高带宽内存（HBM 芯片）和相关半导体制造设备纳入全面对华出口限制，并大幅扩大“外国直接产品规则”，凡含有美国技术的海外产品皆受管辖。
BIS	2023 年 11 月 17 日	出口管制	为封杀英伟达专为中国定制的“降配版”芯片（A800、H800），BIS 取消带宽指标，改用“性能密度”为审查标准。这导致连消费级旗舰显卡（如 RTX 4090）也一并被禁。管制范围扩大至全球 44 个受控国家，严防中国通过中东等地转运。
BIS	2022 年 12 月 16 日	实体清单	BIS 将包括寒武纪、长江存储、上海微电子等 36 家中国半导体及 AI 研发企业加入“实体清单”，切断其获取美国芯片设计、制造设备和技术的渠道。
BIS	2022 年 10 月 7 日	出口管制	首次针对高级计算芯片发布出口管制，确立了“总处理性能（TPP）”和“互联带宽”双指标，并首创性地禁止“美国人（US Persons）”在无许可情况下支持中国先进芯片制造。

资料来源：美国联邦公报、BIS、路透社、彭博社、英伟达官网、华创证券

强调“适度超前建设网络设施”，国产算力落地路径清晰。国务院提出深入实施“人工智能+”行动政策，旨在推动人工智能与经济社会深度融合。计划到 2027 年实现人工智能与六大重点领域广泛深度融合，到 2030 年全面赋能高质量发展，到 2035 年全面建成智能经济与社会新阶段，为中国式现代化提供核心支撑。中国信息通信研究院副院长魏亮在 2025 中国算力大会上表示，“国家正按照‘点、链、网、面’体系化推进全国一体化算力网络工作，通过推动优化算力布局、强化技术协同创新、适度超前建设网络设施、丰富算力应用场景，持续提升算网综合供给能力。”下一步，加快突破 GPU 芯片等关键核心技术，扩大基础共性技术供给。同时，东数西算工程在推进算力基建过程中，助力破解区域发展不平衡等问题，为算力基建自主化提供了资源调配与产业协同机遇，推动构建更均衡、更自主可控的算力网络。目前地方已率先落地相关布局，如甘肃庆阳市政府与燧原科技、亿算智能签订《共建国产十万卡算力集群及新质生产力生态圈战略合作框架协议》；河南 2024 年计划投资 568 亿元推进智算中心建设；截至 2025 年 4 月，宁夏标准机架、智算算卡、算力规模分别较 2024 年底新增 2.6 万架、2.3 万张、3.4 万 P，预计 2025 年内新增标准机架 28 万架，强化“东数西算”枢纽功能；标志着我国自主可控的超大规模 AI 算力基础设施建设进入实质推进阶段。

图表 33 我国 AI 芯片相关政策工程（部分）

维度	核心政策/工程	量化目标与要求
人工智能与经济社会融合发展	国务院深入实施"人工智能+"行动政策	2027 年实现人工智能与六大重点领域广泛深度融合；2030 年人工智能全面赋能高质量发展；2035 年，全面建成智能经济与社会新阶段，为中国式现代化提供核心支撑
点、链、网、面体系化推进	算力布局政策体系	持续推动算力网络"点、链、网、面"体系化高质量发展： 将推动完善算力布局政策体系，引导各地合理布局智能算力设施； 加快突破 GPU 芯片等关键核心技术，扩大基础共性技术供给； 面向教育、医疗、能源等重点行业，开展算力赋能专项行动。
算力基建自主化	"东数西算"工程	为算力基建自主化提供了资源调配与产业协同机遇，推动构建更均衡、更自主可控的算力网络；
地方协同布局	省级/市级算力规划	地方政府主导推进，通过项目签约、资金投入等落地： 河南计划投资 568 亿元推进智算中心建设；宁夏 2025 年内预计新增 28 万架标准机架，扩容枢纽节点能力；甘肃庆阳市共建国产十万卡算力集群及生态圈。

资料来源：人民周刊、经济日报、庆阳融媒、华创证券

国内厂商凭借对下游客户需求的快速响应和本地化服务，已在特定领域取得一定市场份额，并持续向新应用领域延伸。从全球龙头来看，英伟达 2025 年实现营收 15,100.33 亿元，净利润 8,396.17 亿元，在中国 AI 加速器市场份额约为 55%，凭借 CUDA 生态与 GPU 架构的持续迭代构筑强大市场地位；AMD 紧随其后，2025 年在中国 AI 加速器市场份额约为 4%。聚焦国内，人工智能芯片市场呈现多元化和快速发展的特征。GPU 作为人工智能芯片的重要组成部分，在人工智能计算市场占据核心地位。近年来，受到美国高性能计算芯片出口管制与国内政策推动等因素影响，海外厂商在中国市场的份额呈现明显下降趋势。国产人工智能芯片公司迎来黄金发展期，以不同技术路径切入市场，各具特色。其中，寒武纪 2025 年实现营收 64.97 亿元、净利润 20.58 亿元，在国内 AI 芯片厂商中已率先实现盈利，市场份额约 2.9%；沐曦股份与公司营收规模相近，2025 年分别实现营收 16.44 亿元和 9.90 亿元，市场份额与燧原科技接近；摩尔线程、壁仞科技和天数智芯营收规模均在 10-15 亿元区间，目前均处于亏损状态，反映出国内 AI 芯片厂商整体仍处于高研发投入的早期阶段。整体来看，国内 AI 芯片企业在营收体量、盈利能力和市场份额上与海外龙头仍存在数量级的差距，但正以不同技术路径加速追赶。

图表 34 全球龙头企业业务和经营情况比较（亿元）

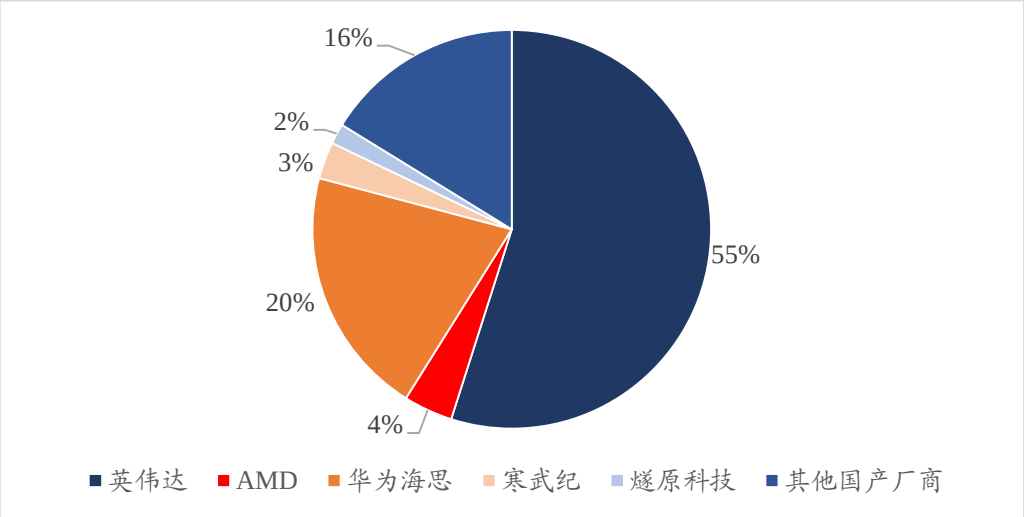
公司名称	主要芯片产品	2025 年营业收入	2025 年净利润	2025 年研发费用	2025 年末总资产	国内市场份额（按销售量计）
英伟达	CPU、图形渲染 GPU 和 GPGPU 架构的 AI 芯片	15100.33	8396.17	1293.48	14461.53	根据 IDC，2025 年英伟达在中国 AI 加速器市场份额约为 55%
AMD	CPU、图形渲染 GPU 和多领域应用的 GPGPU 架构 AI 芯片	2437.13	305.00	569.27	5412.36	根据 IDC，2025 年 AMD 在中国 AI 加速器市场份额约为 4%
寒武纪	DSA 架构的 AI 芯片	64.97	20.58	13.51	134.38	根据 IDC 和公开披露资料，2025 年寒武纪在中国 AI 加速器市场份额约为 2.9%
摩尔线程	图形渲染 GPU 和 GPGPU 架构的 AI 芯片	15.06	-10.01	13.05	153.39	2025 年年度报告未披露销量
沐曦股份	图形渲染 GPU 和 GPGPU 架构的 AI 芯片	16.44	-7.89	10.27	136.75	根据 IDC 和公开披露资料，2025 年沐曦股份在中国 AI 加速器市场份额与燧原科技相近
壁仞科技	GPGPU 架构的 AI 芯片	10.35	-164.93	14.76	59.18	2025 年年度报告未披露销量
天数智芯	GPGPU 架构的 AI 芯片	10.34	-10.04	9.74	39.12	2025 年年度报告未披露销量
燧原科技	DSA 架构的 AI 芯片	9.90	-11.64	11.35	45.22	根据 IDC 和燧原科技销售量，2025 年燧原科技在中国 AI 加速器市场份额约为 1.7%

资料来源：燧原科技招股说明书、华创证券

国产 AI 芯片市场份额与竞争力稳步提升。美国出口管制背景下，国内芯片产业国产化节奏加快，企业协同发展形成良好生态。海关总署数据显示，2025 年 1-12 月，我国处理器及控制器类芯片进口额 2051.2 亿美元，同比增长 6.5%；同期本土出口额 695.1 亿美元，同比增长 22.1%，出口增速较进口高出 15.6 个百分点，出口增速远超进口，显示中国芯片在全球供应链中的地位有所提升。

从市场格局看，国产芯片市场竞争力持续增强。IDC 数据显示，2025 年中国 AI 加速芯片总出货量约 400 万张，其中国产人工智能芯片出货量达 165 万张，市场份额由 2024 年的 34.5%提升至 41.3%，其中华为出货量约 81.2 万颗，位列国产厂商首位；2025 年英伟达仍以 55%的市场份额占据国内市场首位，相较于制裁前约 95%的份额出现显著回落，国产阵营实现明显突破。据科智咨询统计，预计 2026 年国产 AI 加速卡出货量将进一步增长至 230 万张，市场份额有望首次突破 50%达到 51.1%，伴随国产 AI 芯片性能与生态持续完善，本土厂商持续填补市场空白，国内云端 AI 加速器市场竞争格局或将迎来进一步重塑。

图表 35　2025 年中国 AI 加速卡出货量市场份额



资料来源：燧原科技招股说明书、观察者网、华创证券

仅供内部参考，请勿外传！

三、自主 GCU 架构引领，软硬件协同构建 AI 算力核心壁垒

（一）硬件筑基，GCU 架构与系统级能力构筑差异化护城河

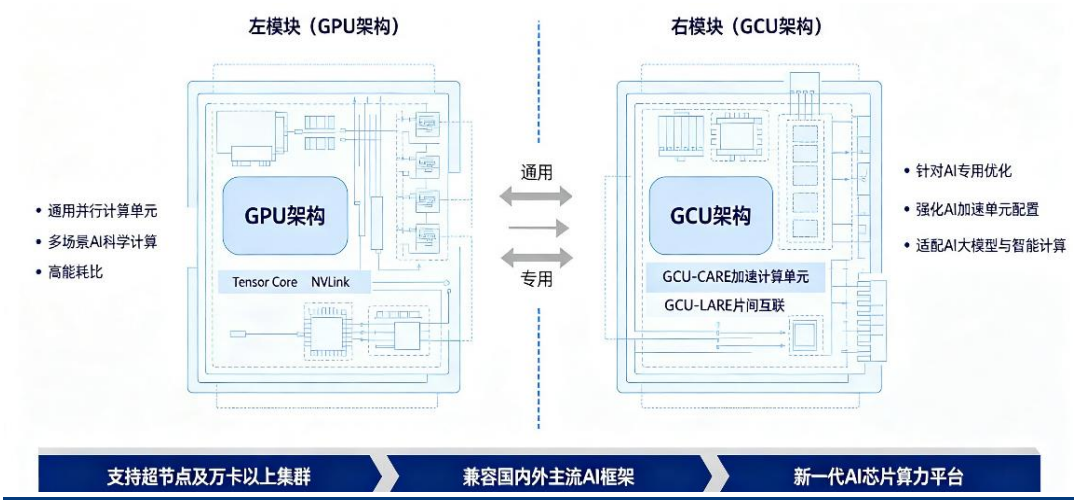
AI 算力竞争已从单芯片纸面参数比拼，升级到整机、集群综合使用效率竞争。燧原科技长期坚持独立自研 GCU 架构，聚焦 AI 训推场景做定向优化，通过"专属计算单元+自研互联技术+成熟量产工程能力"三重优势，在推理赛道形成区别于其他厂商的硬件护城河，并将竞争维度从单芯片延伸至集群级系统能力。

1、GCU 架构专为 AI 负载定制，推理能效优于通用 GPGPU

行业 GPGPU 芯片兼顾图形、科学计算等多场景，大量晶体管资源被通用单元占用，跑大模型推理时算力浪费严重；而燧原 GCU 架构聚焦 AI 训推场景做定向优化，自研两大核心硬件模块：

- **GCU-CARE 加速计算单元：**对标英伟达 Tensor Core，专门针对大模型矩阵运算、多模态推理优化；第四代邃思 400 原生支持 FP8 低精度计算，可显著缩减模型显存占用、提升推理吞吐，适配 DeepSeek 等主流开源大模型规模化部署。
- **GCU-LARE 片间高速互联：**对标英伟达 NVLink，摆脱传统 PCIe 总线带宽、延迟限制，支持多芯片直连组网，可通过网络设备组建超节点及万卡以上大规模算力集群。

图表 36　GCU-CARE 加速计算单元与 GCU-LARE 片间互联架构



资料来源：燧原科技官网、华创证券

2、四代五芯一次流片成功，产品矩阵覆盖云端训推全场景

多代并行、工程验证与路线图清晰，共同驱动产品持续迭代。公司建立了“预研一代、研发一代、量产一代”的多代并行研发模式，将不同代际芯片的研发工作在同一时间轴上并行推进，确保前瞻技术探索、工程化研发与规模化交付有序衔接。成立八年来，公司自研迭代四代架构、推出五款云端 AI 芯片，全部实现一次流片成功，在先进工艺节点下验证了从架构设计到物理实现再到量产交付的全链条工程化能力，有效缩短了迭代周期并控制了流片成本。产品矩阵从早期专注云端训练，逐步拓展至推理、训推一体，形成覆盖全场景的产品谱系。在研产品方面，第五代芯片计划支持 FP4 精度，覆盖文本、音频、图像及视频的跨模态生成；第六代芯片计划强化卡间互联与多芯粒异构集成，以应对超大规模模型的算力集群需求。

仅供内部参考，请勿外传！

图表 37 邃思产品谱系图

芯片名称	芯片介绍	AI 加速卡/模组名称	AI 加速卡/模组介绍	应用场景	图例
邃思 1.0 芯片	2019 年发布 第 1 代 GCU-CARA 计算架构 支持多种计算精度 支持千卡级集群	云燧 T1x 训练系列	支持风冷、冷板液冷方案 板卡功耗：225-300W 支持多卡互联拓扑 PCIe 板卡和 OAM 模组形态	云端训练	
		云燧 i1x 推理系列	支持风冷、冷板液冷方案 板卡功耗：150W PCIe 板卡形态	云端推理	
邃思 2.0 芯片	2021 年发布 第 2 代 GCU-CARA 计算架构 支持多种计算精度 支持千卡级集群	云燧 T2x 训练系列	支持风冷、冷板液冷方案 板卡功耗：300W 支持多卡互联拓扑 PCIe 板卡、OAM 模组形态	云端训练	
邃思 2.5 芯片	2021 年发布 第 2.5 代 GCU-CARA 计算架构 支持多种计算精度	云燧 i2x 推理系列	支持风冷方案 板卡功耗：150W PCIe 板卡形态	云端推理	
邃思 320 芯片	2024 年发布 第 3 代 GCU-CARA 计算架构 支持多种计算精度	燧原 S60 推理加速卡	支持风冷方案 板卡功耗：300W PCIe 板卡形态	云端推理	
邃思 400 芯片	2025 年发布 第 4 代 GCU-CARA 计算架构 支持包括 FP8 在内的多种计算精度 支持万卡及以上集群	燧原 L600 训推一体加速模组	支持风冷、液冷方案 板卡功耗：700W 支持单层最多 128 卡全互联拓扑 OAM 模组形态	云端训练/推理一体	

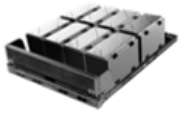
资料来源：燧原科技招股说明书、华创证券

3、先进封装与超节点产品，将竞争维度延伸至集群级

Chiplet 与 OAM 工程积累，推动竞争从芯片迈向集群系统。大模型参数规模从千亿迈向万亿后，单芯片性能提升已无法独立支撑算力需求。公司在 Chiplet 芯粒、高密度 OAM 模组层面形成独家工程积累，解决集群高带宽、高散热痛点：第三代芯片采用同构芯粒封装实现算力翻倍，第四代已落地全国产封装链路；自研 OAM 高密度模组依托 UBB 底座实现多卡直连，搭配冷板液冷耦合设计，可在 95℃高温环境下保障芯片持续满频稳定运行。基于上述硬件底座，公司推出云燧 OGX400 与 ESL 超节点产品，将竞争维度从单

芯片延伸至集群级系统能力。

图表 38 智算系统产品

产品型号	产品介绍	产品应用场景	产品图例
云燧 OGX400	发布时间：2025 年 7 月 算力集成度：单节点 8 卡 存储容量：1.15TB 功耗：6kW	高性能推理与轻量训练	
超节点 ESL32/64	发布时间：待规模量产 算力集成度：单节点 32/64 卡 存储容量：9.2TB 功耗：90kW	大模型预训练与高并行推理	

资料来源：燧原科技招股说明书、华创证券

落地标杆万卡集群，验证规模化服务能力。公司在甘肃庆阳落地国内首个万卡级国产推理智算集群，部署 10016 张 S60 加速卡及 1252 台国产服务器，半年内完成投产，交付效率领先行业。该集群通过中国信通院五星级认证，千卡集群可连续运行 7 天、万卡集群 24 小时无故障，算力利用率稳定在 80%左右，峰值可满负荷运行，持续服务多家头部互联网企业。高速互联与超节点技术的成熟应用，构成了公司从单芯片竞争迈向集群系统竞争的差异化壁垒。

图表 39 服务成熟度检验证书



证书编号：2025VY005491
报告编号：25V01Y000642-001

大规模智算集群服务成熟度检验证书

上海燧原科技股份有限公司
上海市浦东新区金秋路 158 号张润大厦 1 号楼 3 楼

国产万卡推理集群
（五星级）

通过 CTTL-XXYWFF-2025-237《大规模智算集群服务成熟度》的检验

符合性描述：

“经中国信息通信研究院的检验，贵单位的 国产万卡推理集群 通过了 CTTL-XXYWFF-2025-237《大规模智算集群服务成熟度》的检验，符合大规模智算集群服务的集群服务性能、集群资源调度与管理、大模型推理加速和服务性能、服务可用性及集群稳定性、集群可观测性的要求，特发此证。”

（证书有效性信息及详情可见 www.kexinyun.org.cn 及可信云服务号。由于该检验是针对 2025 年 6 月 5 日前该产品/服务情况进行披露，用户应留意在具体使用过程中可能发生的风险。因云计算软件等技术快速发展、产品/服务不断升级迭代、标准规范随行业发展不断更新等原因，当产品或检验标准发生较大变化时，建议本产品/服务进行重新检验。中国信息通信研究院将在 2026 年 6 月组织新一批检验，该证书仅供客户内部使用。）

中国泰尔实验室
中国信息通信研究院
2025年06月17日

地址：北京市海淀区花园北路52号
网址：https://www.kexinyun.org.cn/

地址：北京市海淀区花园北路52号
网址：http://www.caict.ac.cn/

资料来源：燧原科技公众号

仅供内部参考，请勿外传！

（二）生态立本，软硬件协同与头部客户验证打开商业化空间

AI 芯片的长期价值不仅取决于硬件性能，更取决于软件生态成熟度与客户粘性。燧原科技通过自研全栈软件平台降低迁移门槛，依托头部客户真实业务场景验证产品力，并以全链条供应链自主可控保障规模化交付，构建起从"技术可用"到"商业可持续"的完整闭环。

1、自研“驭算 TopsRider”全栈平台，实现生态自主可控

自研 TopsRider 全栈软件平台，从“可用”向“好用”演进。公司独立自主研发的“驭算 TopsRider”全栈 AI 计算及编程软件平台，涵盖驱动、编译器、算子库、工具链等核心组件，通过深度协同自研 GCU 硬件架构，充分释放 DSA 路线在 AI 计算场景下的硬件潜能。在迁移成本方面，平台内置超 1,600 个深度优化的 AI 加速算子，已适配近千款 AI 模型、落地超 300 个应用场景，目前中等复杂度 CUDA 原生应用迁移至 TopsRider 平台仅需约 1 人天，显著降低客户切换成本，这一优势受益于公司与互联网大厂多年在软硬件定制化方面的深耕合作。战略层面，兼容 CUDA 虽短期降低门槛，但长期受制于英伟达生态；公司自建 TopsRider 生态实现技术体系完全自主可控，正从“可用”向“好用”演进，有望成为国产 AI 芯片差异化竞争的核心壁垒。

图表 40　驭算 TopsRider AI 计算及编程软件平台

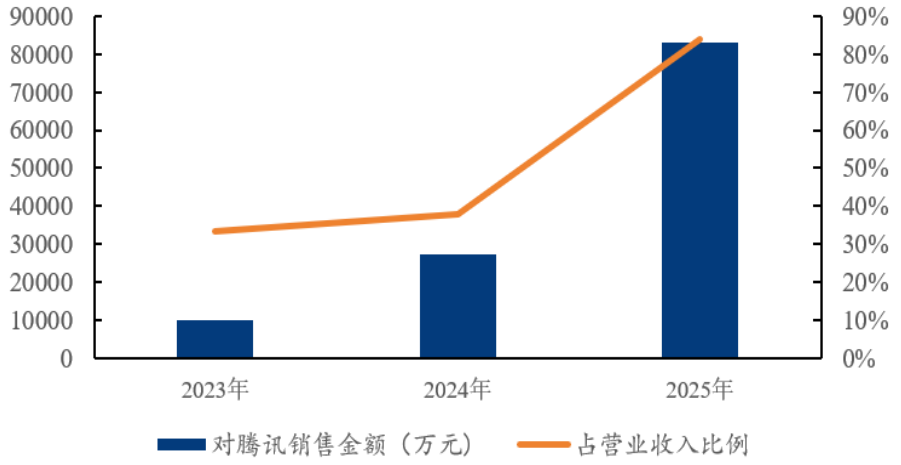


资料来源：燧原科技招股说明书

2、头部厂商验证产品能力，成熟工程交付打开商业化空间

腾讯从核心客户到第一大外部股东，战略协同持续深化。2018 年公司 Pre-A 轮由腾讯领投，此后腾讯持续参与每一轮关键融资，截至燧原科技上市首日，腾讯与其一致行动人苏州湃益合计持股 18.23%，为公司第一大外部股东。双方合作始于 2019 年，围绕业务场景适配优化、AI 模型性能调优、配套软件栈完善等深度协同，公司产品已在腾讯大量 AI 业务场景中实现规模化部署。**从营收数据来看，**2023-2025 年公司对腾讯的销售金额从 1.05 亿元增长至 8.30 亿元，三年实现超 7 倍增长；占营业收入比例从 33.34%攀升至 83.79%，反映出公司在腾讯国产 AI 算力供应链中核心供应商地位的持续巩固。与此同时，腾讯投资董事总经理姚磊文担任燧原科技董事会成员，深度参与公司治理与战略决策；2025 年腾讯分两笔合计支付 3.68 亿元采购预付款，定向用于公司上游供应链备货，进一步强化了双方在供应链层面的协同。

图表 41 公司第一大客户销售金额和占比



资料来源：燧原科技招股说明书、华创证券

多元客户稳步拓展，全链条供应链保障规模化交付。互联网头部厂商 AI 业务具备高并发、大流量、低时延、高稳定性的硬性约束，AI 算力硬件与软件栈深度耦合，客户转换成本极高，存量合作具备极强粘性。公司采取"单点突破、以点带线、以线带面"的策略：依托腾讯高负载场景打磨产品竞争力，再向多元客户圈层拓展。

与此同时，公司积极深化与电信运营商合作，深度参与国家"东数西算"算力枢纽建设，落地甘肃庆阳万卡级推理智算集群、江苏无锡智算中心等多个标杆项目。在非关联头部互联网客户开拓上，多家企业已完成硬件适配与大模型迁移测试，进入集群灰度试运行、小批量试采购环节，按照公司经营规划，2026 年有望实现小规模订单交付，2027 年推进大批量规模化供货。

图表 42 潜在客户情况

类别	公司名称	灰度测试等具体进展	预计业绩
互联网客户	潜在客户 A	第四代产品已通过前期硬件系统测试和模型匹配测试；进入确定灰度测试具体方案阶段，预计于 2026 年年内启动灰度测试	有望在 2026 年 12 月小规模交付，2027 年开始实现大批量交付，毛利率、供货量及对应收入将根据后续商业协商情况确定（下同）
	潜在客户 F	第四代产品已通过硬件系统测试和模型匹配测试，进入确定灰度测试具体方案阶段，预计于 2026 年年内启动灰度测试	有望在 2026 年 12 月小规模交付，2027 年实现大批量交付
	潜在客户 B	第四代产品已通过硬件系统测试和模型匹配测试，客户无灰度测试要求	有望在 2026 年 12 月小规模交付，2027 年分别实现大批量交付
	潜在客户 C	第四代产品已通过硬件系统测试，进入模型测试阶段。由于公司当前研发资源较为紧张，暂缓推进相关测试及商务需求沟通工作	—
	潜在客户 H	第三代产品已通过硬件系统测试和模型自测，正在进行系统级业务部门复测，预计于 2026 年内启动灰度测试	—
非互联网客户	潜在客户 D	第三代产品通过与合作伙伴 A 合作，已实现潜在客户 D 集采项目首次入围及中标并已完成中标后复测验收，合作伙伴 A 已与潜在客户 D 完成框架合同签署	预计 2026 年 6 月前向合作伙伴完成产品交付，已跑通了完整流程，后续与合作伙伴深度合作争取更多市场份额

		第四代产品正在与合作伙伴 A、合作伙伴 B 等合作方共同开发高密度 AI 服务器及超节点服务器，未来将依托合作伙伴在国内运营商 AI 服务器市场的领先地位及深厚渠道资源推进产品测试并进行市场开拓	
	潜在客户 F	第三代产品已完成模型测试，测试结果满足客户要求，与合作伙伴 C 合作已于 2025 年向其实现数百张产品销售。未来公司将与合作伙伴 C、合作伙伴 B 等密切合作，争取在 2026 年实现更大批量销售	—

资料来源：燧原科技招股说明书、华创证券

供应链自主可控，锁产机制与本土产线保障规模化交付。为应对上游晶圆、存储芯片厂商行业集中度高、议价能力较强的特点，公司通过提前锁货并支付较高比例预付款锁定上游产能，保障产品持续批量交付。国产化层面，公司第四代芯片已落地全国产先进封装方案，持续优化超高密度布线、微凸点集成工艺，同时与本土龙头深度合作，打通“本地设计+本地制造+本地封测+本地 EDA/IP”完整产业链路，实现芯片研发到应用全链条自主可控。依托扎实的技术研发实力，公司累计承担 12 项国家及地方科技攻关项目，并获“中国芯”最高奖、吴文俊人工智能科学技术专项一等奖等权威荣誉，全方位完善供应链全链条自主可控能力，为公司多客户规模化订单落地与产品长期稳定交付提供坚实支撑。

图表 43 公司全链条供应链布局

产业链环节	参与方	产业内容
设计端	燧原科技（自研）	自研 GCU 架构、自主指令集、全栈软件平台“驭算 TopsRider”
制造端	上游晶圆厂	提前预付产能保证金以锁定未来产能
	深南电路	板卡加工
	天通精电（天通股份子公司）	板卡加工
封测端	上海先封科技	AI 算力芯片下一代先进封装解决方案
配套端	新华三	东数西算（贵安）算力运营产业联盟正式签约
交付端	燧原科技	第四代芯片已落地全国产先进封装方案，持续优化超高密度布线技术与微凸点集成技术

资料来源：燧原科技招股说明书、金桥集团微信公众号、新华三微信公众号、华创证券

仅供内部参考，请勿外传！

四、盈利预测

基于当前 AI 算力需求的持续爆发及公司产品在云端 AI 芯片领域的领先优势，我们预计公司营收将保持高速增长，2026-2028 年有望实现营收的持续快速扩张。在毛利率方面，受益于产品结构的持续优化与规模效应的逐步释放，综合毛利率有望在 2025 年 31.78%的基础上稳步提升。其中，AI 加速卡及模组作为核心产品线，收入占比已从 2023 年的 61.70%提升至 2025 年的 86.49%，随着三代产品持续放量及四代产品逐步量产，未来有望实现收入持续增长，毛利率虽因客户放量及成本波动有所调整，但整体维持较高水平；智算系统及集群业务目前收入占比有所下降，主要系公司聚焦资源优先满足头部互联网客户需求，未来将根据市场情况灵活推进，毛利率保持稳健；IP 授权及其他业务收入占比较小，对公司整体业绩影响有限。在费用端，随着收入规模的快速扩大，研发、销售及管理费用率有望持续下降。公司盈利路径清晰，经营质量有望持续向好。

图表 44 业务预测（百万元）

		2022	2023	2024	2025	2026E	2027E	2028E
合计	营业收入	90.10	301.19	722.39	990.16	3593.34	6160.25	11129.79
	yoy			139.85%	37.07%	262.9%	71.4%	80.7%
	毛利率		22.97%	30.75%	31.90%	34.76%	42.21%	42.82%
AI 加速卡及模组	营业收入	36.68	185.84	307.56	856.42	3460.00	5985.00	10870.00
	yoy		406.64%	65.50%	178.46%	304%	73%	82%
	毛利率	46.25%	15.51%	40.78%	32.71%	35%	43%	43%
智算系统及集群	营业收入		100.11	398.93	128.24	126.95	167.58	250.53
	yoy			298.49%	-67.85%	-1%	32%	50%
	毛利率		36.73%	22.76%	25.71%	28%	32%	35%
IP 授权及其他	营业收入	53.20	6.15	2.85	1.64	1.56	1.64	1.72
	yoy		-88.45%	-53.57%	-42.46%	-5%	5%	5%
	毛利率	100.00%	6.81%	26.33%	23.09%	25%	25%	25%
其他业务收入	营业收入		9.09	13.05	3.86	4.82	6.03	7.54
	yoy			43.48%	-70.43%	25.00%	25.00%	25.00%
	毛利率		34.73%	39.37%	60.40%	40%	40%	40%

资料来源：iFind、华创证券预测

公司作为国产云端 AI 芯片领军企业，坚持自主可控的 DSA 全栈自研路线，已迭代四代架构 5 款芯片，构建了覆盖 AI 芯片、加速卡、智算系统及软件平台的完整产品体系。公司深度绑定头部互联网客户打磨产品力，2023 至 2025 年营收从 3.01 亿元增长至 9.90 亿元，复合增长率达 81.32%，亏损持续收窄；综上，我们预计公司 2026-2028 年营业收入为 35.93 亿元、61.60 亿元、111.30 亿元，对应增速 262.9%、71.4%、80.7%；归母净利润为-5.16 亿元、5.96 亿元、21.53 亿元，对应增速分别为 55.7%、215.6%、261.3%；对应 EPS（摊薄）分别为-1.20 元、1.38 元、5.00 元，首次覆盖给予“推荐”评级。

仅供内部参考，请勿外传！

五、风险提示

1、客户集中度较高与关联交易风险

公司与第一大客户兼第一大股东腾讯之间形成了深度的业务绑定关系，关联销售占比较高，且近年来呈持续上升趋势。若未来腾讯的采购需求、采购策略或合作关系发生不利变化，将对公司经营业绩产生重大影响。

2、供应链与产品迭代风险

公司产品的晶圆代工、封装测试等环节依赖数量有限的供应商，若上游产能紧张或合作发生波动，可能影响产品的量产交付。同时，AI 芯片行业技术迭代迅速，若公司新一代产品的研发进度、性能表现或市场接受度不及预期，将削弱公司的市场竞争力与行业地位。

3、盈利预测敏感性分析

公司未来的盈利时点主要受毛利率水平及营业收入达成率影响，以收入达成率上下浮动 15%，毛利率上下浮动 5%对公司 2027 年盈利情况进行测算，中性情形下公司 2027 年经营层面盈利约 6.1 亿，若收入达成率或毛利率存在低于中性水平的情况下，公司 2027 年盈亏中枢将下移至亏损约 0.5 亿元至盈利约 3.0 亿元，预计将在 2027 年实现盈利，盈利预测可能不及预期。

图表 45 盈利预测敏感性分析（单位：百万）

毛利率/收入达成率	低收入达成率	中收入达成率	高收入达成率
低毛利率	-46	298	642
中毛利率	216	606	996
高毛利率	478	914	1351

资料来源：iFind，华创证券预测

仅供内部参考，请勿外传！

附录：财务预测表

资产负债表

单位：百万元	2025A	2026E	2027E	2028E
货币资金	2,155	5,560	6,130	6,229
应收票据	5	5	10	24
应收账款	219	906	1,078	1,915
预付账款	388	2,579	2,492	3,182
存货	863	1,763	1,597	2,434
合同资产	0	0	0	0
其他流动资产	312	934	1,201	2,059
流动资产合计	3,943	11,747	12,509	15,843
其他长期投资	0	0	0	0
长期股权投资	0	0	0	0
固定资产	140	731	1,076	1,326
在建工程	0	0	0	0
无形资产	122	460	734	940
其他非流动资产	317	338	360	365
非流动资产合计	579	1,529	2,169	2,631
资产合计	**4,522**	**13,276**	**14,679**	**18,473**
短期借款	126	626	1,026	1,326
应付票据	0	0	0	0
应付账款	52	188	285	509
预收款项	0	0	0	0
合同负债	346	2,875	2,772	3,339
其他应付款	97	110	200	250
一年内到期的非流动负债	90	90	90	90
其他流动负债	170	252	332	539
流动负债合计	882	4,140	4,705	6,052
长期借款	432	632	875	1,169
应付债券	0	0	0	0
其他非流动负债	61	61	61	61
非流动负债合计	493	693	936	1,230
负债合计	**1,375**	**4,834**	**5,641**	**7,282**
归属母公司所有者权益	3,147	8,442	9,038	11,191
少数股东权益	0	0	0	0
所有者权益合计	**3,147**	**8,442**	**9,038**	**11,191**
负债和股东权益	**4,522**	**13,276**	**14,679**	**18,473**

现金流量表

单位：百万元	2025A	2026E	2027E	2028E
经营活动现金流	**-965**	**-2,020**	**960**	**596**
现金收益	-999	-376	1,028	2,793
存货影响	-67	-900	166	-837
经营性应收影响	168	-3,497	-417	-2,530
经营性应付影响	-276	2,759	165	1,048
其他影响	209	-6	18	123
投资活动现金流	**-379**	**-1,083**	**-1,027**	**-1,052**
资本支出	-183	-1,031	-1,009	-1,023
股权投资	0	0	0	0
其他长期资产变化	-196	-52	-17	-29
融资活动现金流	**-22**	**6,507**	**637**	**555**
借款增加	21	700	643	594
股利及利息支付	-19	-45	-66	-83
股东融资	0	0	0	0
其他影响	-24	5,852	60	44

利润表

单位：百万元	2025A	2026E	2027E	2028E
营业总收入	**990**	**3,593**	**6,160**	**11,130**
营业成本	674	2,344	3,560	6,365
税金及附加	1	4	6	18
销售费用	138	162	222	334
管理费用	175	216	246	334
研发费用	1,135	1,401	1,478	1,669
财务费用	9	38	41	74
信用减值损失	-31	-10	-10	-10
资产减值损失	-32	-10	-10	-10
公允价值变动收益	0	0	0	0
投资收益	20	35	35	35
其他收益	19	40	40	40
营业利润	**-1,164**	**-516**	**662**	**2,392**
营业外收入	0	0	0	0
营业外支出	0	0	0	0
利润总额	**-1,164**	**-516**	**662**	**2,392**
所得税	0	0	66	239
净利润	**-1,164**	**-516**	**596**	**2,153**
少数股东损益	0	0	0	0
归属母公司净利润	**-1,164**	**-516**	**596**	**2,153**
NOPLAT	-1,155	-478	633	2,219
EPS(摊薄)（元）	-2.70	-1.20	1.38	5.00

主要财务比率

	2025A	2026E	2027E	2028E
成长能力				
营业收入增长率	37.1%	262.9%	71.4%	80.7%
EBIT 增长率	23.1%	58.7%	247.2%	250.8%
归母净利润增长率	22.9%	55.7%	215.6%	261.3%
获利能力				
毛利率	31.9%	34.8%	42.2%	42.8%
净利率	-117.5%	-14.3%	9.7%	19.3%
ROE	-37.0%	-6.1%	6.6%	19.2%
ROIC	-32.5%	-5.0%	6.5%	18.3%
偿债能力				
资产负债率	30.4%	36.4%	38.4%	39.4%
债务权益比	22.5%	16.7%	22.7%	23.6%
流动比率	4.5	2.8	2.7	2.6
速动比率	3.5	2.4	2.3	2.2
营运能力				
总资产周转率	0.2	0.3	0.4	0.6
应收账款周转天数	126	56	58	48
应付账款周转天数	160	18	24	22
存货周转天数	443	202	170	114
每股指标(元)				
每股收益	-2.70	-1.20	1.38	5.00
每股经营现金流	-2.24	-4.69	2.23	1.38
每股净资产	7.31	19.62	21.00	26.00
估值比率				
P/E	-147	-331	287	79
P/B	54.3	20.2	18.9	15.3
EV/EBITDA	-172	-456	157	57

资料来源：公司公告，华创证券预测

仅供内部参考，请勿外传！

计算机组团队介绍

首席分析师、组长：吴鸣远

上海交通大学硕士，曾任职于东方证券、兴业证券研究所，所在团队于 2020—2022 年连续三年获得新财富最佳分析师第三名，2023 年加入华创证券研究所。

分析师：祝小茜

中央财经大学经济学硕士。曾任职于信达证券。2024 年加入华创证券研究所。

分析师：胡昕安

工学硕士，曾任职于海康威视，2023 年加入华创证券研究所。

助理研究员：周志浩

西安交通大学金融工程学士，克拉克大学金融学硕士，曾任职于众安保险权益投资部，2024 年加入华创证券研究所。

助理研究员：周楚薇

香港中文大学经济学硕士。2024 年加入华创证券研究所。

助理研究员：杨玖祎

香港中文大学硕士。2024 年加入华创证券研究所。

助理研究员：陈科玮

圣路易斯华盛顿大学硕士。2025 年加入华创证券研究所。

华创证券机构销售通讯录

地区	姓名	职务	办公电话	企业邮箱
北京机构销售部	张昱洁	副总经理、北京机构销售总监	010-63214682	zhangyujie@hcyjs.com
	张婷	北京机构销售副总监		zhangting3@hcyjs.com
	刘懿	公募机构销售总监	010-63214682	liuyi@hcyjs.com
	侯春钰	资深销售经理	010-63214682	houchunyu@hcyjs.com
	顾翎蓝	资深销售经理	010-63214682	gulinglan@hcyjs.com
	刘颖	资深销售经理	010-66500821	liuying5@hcyjs.com
	阎星宇	销售经理		yanxingyu@hcyjs.com
	车一哲	销售经理		cheyizhe@hcyjs.com
	张羽驰	销售经理		zhangyuchi@hcyjs.com
	骆振琳	高级销售经理		luozhenlin@hcyjs.com
	刘亚东	销售经理		liuyadong@hcyjs.com
	白嘉琪	销售助理		baijiaqi@hcyjs.com
	吴昱颖	销售经理		wuyuying@hcyjs.com
	龚佳宁	机构销售		gongjianing@hcyjs.com
	朱晞毓	高级销售经理		zhuxiyu@hcyjs.com
	周怡圆	资深销售经理		zhouyiyuan@hcyjs.com
深圳机构销售部	张娟	副总经理、深圳机构销售总监	0755-82828570	zhangjuan@hcyjs.com
	张嘉慧	资深销售经理	0755-82756804	zhangjiahui1@hcyjs.com
	王春丽	资深销售经理	0755-82871425	wangchunli@hcyjs.com
	王越	高级销售经理		wangyue5@hcyjs.com
	汪丽燕	销售经理	0755-83715428	wangliyan@hcyjs.com
	温雅迪	销售经理		wenyadi@hcyjs.com
	胡丁琳	销售经理		hudinglin@hcyjs.com
	付雅琦	销售经理		fuyaqi@hcyjs.com
	王艺嘉	销售助理		wangyijia1@hcyjs.com
上海机构销售部	许彩霞	总经理助理、上海机构销售总监	021-20572536	xucaixia@hcyjs.com
	祁继春	上海机构销售副总监		qijichun@hcyjs.com
	黄畅	保险机构销售总监	021-20572257-2552	huangchang@hcyjs.com
	吴俊	资深销售经理	021-20572506	wujun1@hcyjs.com
	张佳妮	资深销售经理	021-20572585	zhangjiani@hcyjs.com
	郭静怡	高级销售经理		guojingyi@hcyjs.com
	蒋瑜	高级销售经理	021-20572509	jiangyu@hcyjs.com
	吴菲阳	资深销售经理		wufeiyang@hcyjs.com
	朱涨雨	高级销售经理	021-20572573	zhuzhangyu@hcyjs.com
	许馨匀	销售助理		xuxinyun@hcyjs.com
	李凯月	高级销售经理		likaiyue@hcyjs.com
	张豫蜀	销售经理	15301633144	zhangyushu@hcyjs.com
	刘雯	销售经理		liuwen@hcyjs.com
	金可依	销售经理		jinkeyi@hcyjs.com
	姜茜	销售助理		jiangxi@hcyjs.com
	罗兰丹波	资深销售经理		luolandanbo@hcyjs.com
	谢玮颖	资深销售经理		xieweiying@hcyjs.com
广州机构销售部	段佳音	广州机构销售总监	0755-82756805	duanjiayin@hcyjs.com
	王世韬	高级销售经理		wangshitao1@hcyjs.com

	古舞	销售助理		guwu@hcyjs.com
	郭有渔	销售助理		guoyouyu@hcyjs.com
	曹雅博	高级销售经理		caoyabo@hcyjs.com
私募销售组	潘亚琪	机构服务部总经理助理	021-20572559	panyaqi@hcyjs.com
	汪子阳	副总监	021-20572559	wangziyang@hcyjs.com
	江赛专	副总监	0755-82756805	jiangsaizhuan@hcyjs.com
	汪戈	副总监	021-20572559	wangge@hcyjs.com
	宋丹玙	高级销售经理	021-25072549	songdanyu@hcyjs.com
	赵毅	销售经理		zhaoyi@hcyjs.com
	陈孟昕	销售经理		chenmengxin@hcyjs.com
	文焱艺	销售经理		wenyanyi@hcyjs.com
	王秋晓	销售经理		wangqiuxiao@hcyjs.com
	谭雯欣	销售经理		tanwenxin@hcyjs.com
	陈蕾	销售助理		chenlei1@hcyjs.com
	赵方圆	销售经理		zhaofangyuan@hcyjs.com

仅供内部参考，请勿外传！

华创行业公司投资评级体系

基准指数说明：

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/纳斯达克指数。

公司投资评级说明：

强推：预期未来 6 个月内超越基准指数 20%以上；
推荐：预期未来 6 个月内超越基准指数 10%－20%；
中性：预期未来 6 个月内相对基准指数变动幅度在-10%－10%之间；
回避：预期未来 6 个月内相对基准指数跌幅在 10%－20%之间。

行业投资评级说明：

推荐：预期未来 3-6 个月内该行业指数涨幅超过基准指数 5%以上；
中性：预期未来 3-6 个月内该行业指数变动幅度相对基准指数-5%－5%；
回避：预期未来 3-6 个月内该行业指数跌幅超过基准指数 5%以上。

分析师声明

每位负责撰写本研究报告全部或部分内容的分析师在此作以下声明：

分析师在本报告中对所提及的证券或发行人发表的任何建议和观点均准确地反映了其个人对该证券或发行人的看法和判断；分析师对任何其他券商发布的所有可能存在雷同的研究报告不负有任何直接或者间接的可能责任。

免责声明

本报告仅供华创证券有限责任公司（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。

本报告所载资料的来源被认为是可靠的，但本公司不保证其准确性或完整性。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。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。本公司在知晓范围内履行披露义务。

报告中的内容和意见仅供参考，并不构成本公司对具体证券买卖的出价或询价。本报告所载信息不构成对所涉及证券的个人投资建议，也未考虑到个别客户特殊的投资目标、财务状况或需求。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，自主作出投资决策并自行承担投资风险，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。本报告中提及的投资价格和价值以及这些投资带来的预期收入可能会波动。

本报告版权仅为本公司所有，本公司对本报告保留一切权利。未经本公司事先书面许可，任何机构和个人不得以任何形式翻版、复制、发表、转发或引用本报告的任何部分。如征得本公司许可进行引用、刊发的，需在允许的范围内使用，并注明出处为“华创证券研究”，且不得对本报告进行任何有悖原意的引用、删节和修改。

证券市场是一个风险无时不在的市场，请您务必对盈亏风险有清醒的认识，认真考虑是否进行证券交易。市场有风险，投资需谨慎。

华创证券研究所

北京总部	广深分部	上海分部
地址：北京市西城区锦什坊街 26 号恒奥中心 C 座 3A	地址：深圳市福田区香梅路 1061 号 中投国际商务中心 A 座 19 楼	地址：上海市浦东新区花园石桥路 33 号花旗大厦 12 层
邮编：100033	邮编：518034	邮编：200120
传真：010-66500801	传真：0755-82027731	传真：021-20572500
会议室：010-66500900	会议室：0755-82828562	会议室：021-20572522