

东方港湾春风三号私募证券投资基金

投资者月度报告——2026 年 8 月

一. 基金概况

名称	东方港湾春风三号私募证券投资基金	备案编码	SGM544
运作方式	开放式	类型	私募证券投资基金
管理人	深圳东方港湾投资管理股份有限公司	托管人	国泰海通证券股份有限公司
成立日期	2019-05-14	到期日	2034-05-13
投资经理	黄海平	投资策略	股票策略/股票多头

二. 净值月报

估值日期：2026 年 08 月 31 日

产品名称\类别	单位净值	单位累计净值	产品份额（份）	产品净资产
东方港湾春风三号私募证券投资基金	0.316	0.504	58,633,311.24	18,530,476.08

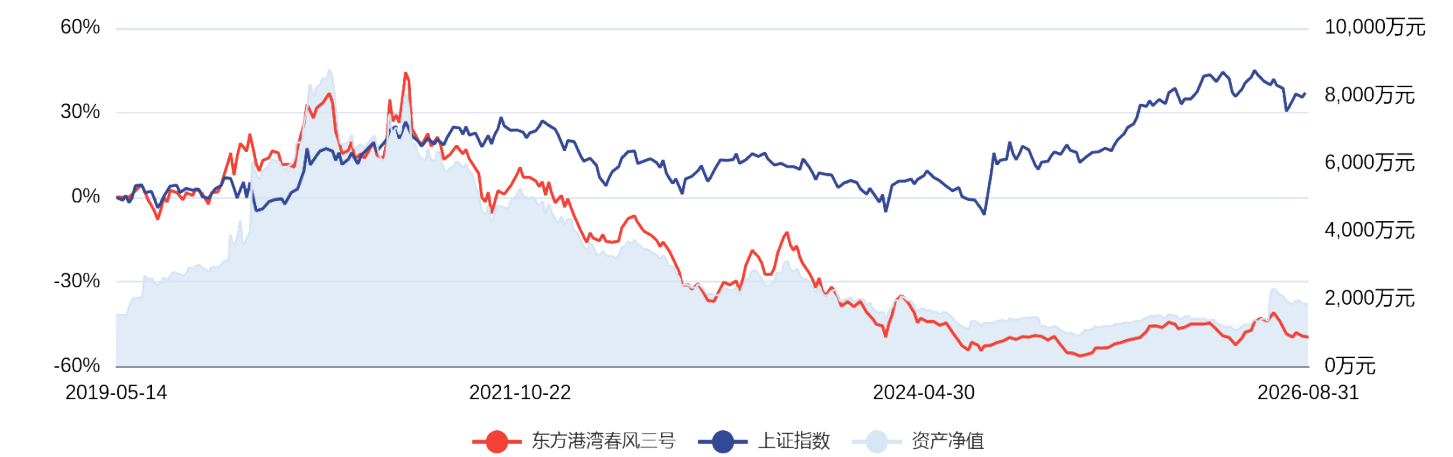
三. 历史业绩

年份	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	全年
2019					0.00%	2.70%	-4.67%	0.31%	0.92%	1.61%	-3.18%	9.13%	6.40%
2020	1.32%	7.89%	-3.18%	2.97%	-4.42%	11.63%	8.78%	2.57%	-16.08%	-1.44%	2.40%	2.55%	12.12%
2021	7.06%	-1.95%	-6.73%	1.42%	-1.50%	-1.02%	-16.55%	0.00%	4.93%	3.64%	-4.42%	-2.25%	-17.91%
2022	-5.82%	-4.38%	-9.56%	1.19%	-2.21%	8.87%	-4.56%	-6.08%	-4.93%	-17.67%	-2.17%	-10.06%	-45.82%
2023	13.20%	0.00%	14.23%	0.17%	2.94%	7.72%	-11.22%	-5.44%	-8.35%	-13.16%	-1.86%	-1.19%	-6.94%
2024	-22.84%	29.91%	5.28%	-12.76%	-3.92%	-4.08%	-17.56%	-1.03%	-0.35%	2.44%	4.08%	2.94%	-24.28%
2025	-2.86%	-5.56%	-12.46%	0.40%	9.06%	5.78%	5.80%	5.81%	10.37%	0.83%	-3.84%	1.71%	13.33%
2026	3.64%	-13.24%	-9.35%	17.18%	14.96%	5.36%	-23.49%	0.00%					-11.48%

风险提示及免责声明

本报告仅供本产品的持有人接收，任何媒体、网站或个人不得转载、链接、转帖或以其他方式复制发表。在任何情况下，本报告中提及的公司仅为阐述产业逻辑举例，不作任何投资建议，本报告所载内容仅供接收者参考之用，亦不视为任何形式的投资操作建议。接收者不应单纯依靠本报告而取代个人的独立判断，对于因使用、引用、参考本报告内容而导致的投资损失、风险及纠纷的，东方港湾不负任何责任。

净值回报及资产净值走势图

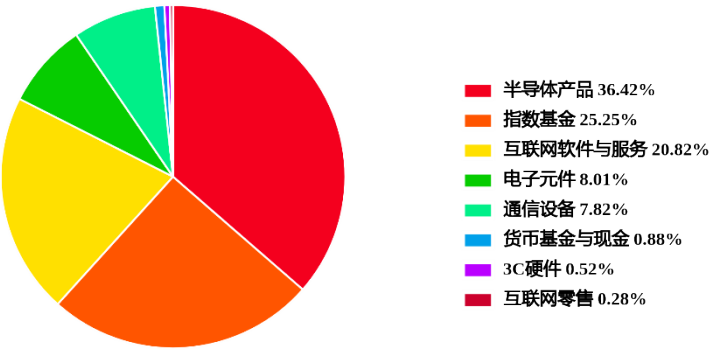


四. 业绩分析

业绩分析

项目	本产品	上证指数
成立以来收益率	-49.60%	38.24%
相对指数收益率	-87.84%	\
本月收益率	0.00%	4.02%
本年收益率	-11.48%	0.44%
复合年化收益率	-8.95%	4.53%
夏普比率	-0.0519	0.0363
最大回撤（成立以来）	-70.41%	-26.98%

持仓投资组合



注：回报使用分红不投资计算方法，累计净值为单位净值加过往每单位分红净值。

风险提示及免责声明

本报告仅供本产品的持有人接收，任何媒体、网站或个人不得转载、链接、转帖或以其他方式复制发表。在任何情况下，本报告中提及的公司仅为阐述产业逻辑举例，不作任何投资建议，本报告所载内容仅供接收者参考之用，亦不视为任何形式的投资操作建议。接收者不应单纯依靠本报告而取代个人的独立判断，对于因使用、引用、参考本报告内容而导致的投资损失、风险及纠纷的，东方港湾不负任何责任。

五. 主要持仓逻辑

备注：按 wind 行业标准划分

行业名称	持仓比例	持仓逻辑
半导体产品	36.42%	全球 AI 产业进展迅速，RL 的进步推动模型的思考能力的加深和泛化，带动各种垂直 agent 的落地，反过来也促进了推理所需算力的爆发。未来跨模态统一的模型研发，也会带动训练范式的新变化，算力的需求会进一步打开。英伟达的产品目前还是通用性能遥遥领先，在最强的社区生态下算子的优化也是比竞争者更加有优势。虽然近期市场将英伟达简单粗暴的归类为 OpenAI 阵营并且认为这个阵营在谷歌的进攻下会输，但是大模型的进步并非如此二元和线性，同时除开 OpenAI 外，英伟达的芯片对于大部分公司来说也是最具性价比的产品。 公司是全球半导体制造领域的绝对龙头，具备不可替代的技术护城河和全球价值链定价权。其领先的制程工艺（如 3nm、2nm）远超对手，是苹果、英伟达、AMD 等高端芯片的唯一代工选择。随着 AI 算力需求爆发，先进封装（CoWoS、SoIC）成为关键瓶颈，而台积电的产能掌控力赋予其结构性议价优势。同时，美国、日本、德国建厂保障地缘分散。在去库存周期尾声、全球电子需求复苏背景下，台积电有望重回成长轨道。公司财务稳健、毛利率高企，ROE 常年维持 25%以上，具有长期现金流能力。当前估值仍处历史中枢偏下位置，具备较好安全边际与向上弹性，是 AI 时代产业核心的“锚型资产”。 公司 2026 年核心转折点在于 18A 工艺验证成功与量产爬坡，PantherLake/CoreUltraSeries3 系列 AIPC 处理器已获市场认可，服务器 CPU 渠道反馈显示 2026 年基本售罄，有提价空间。公司 Foundry 业务逐步去风险，潜在外部大客户（如苹果/高通）订单可能在下半年落地。AIPC 市场爆发+数据中心需求回暖，叠加成本纪律改善与资产负债表优化，市场开始重新给予估值修复溢价。尽管短期供给瓶颈与执行风险仍存，但 18A 成功已大幅降低技术死亡风险，2026 年是“证明年”。 公司作为 AI 产业基础设施的基石，深度受益于全球 AI 推进的大浪潮和中国本土的自主可控战略。在中国，半导体不仅是商业实体，更是国家科技自立自强的关键环节。政府的巨额补贴和政策倾斜，确保了这些企业拥有充足的资金进行研发和扩产，以满足日益增长的国内市场需求。 公司在 AI 时代的核心逻辑是数据中心业务爆发式增长，特别是 MI400 系列 GPU 加速器有可能在推理市场获得认可（目标 2027 年达双位数甚至 20%）。与 OpenAI 等战略合作逐步兑现高 ASP 收入，叠加服务器 CPU 几乎售罄，未来 agenticAI 驱动的 CPU 需求结构性加速。cpu 的基本面上行周期，GPU 提供弹性，是二线硬件比较有前景的公司。 公司深度受益于“AI 端侧落地+全球存储芯片产能紧缩”的双重景气周期。公司通过“跳代研发”与自主产能扩建，已进化为“DRAM 研发设计+晶圆制造+封测一体化”

风险提示及免责声明

本报告仅供本产品的持有人接收，任何媒体、网站或个人不得转载、链接、转帖或以其他方式复制发表。在任何情况下，本报告中提及的公司仅为阐述产业逻辑举例，不作任何投资建议，本报告所载内容仅供接收者参考之用，亦不视为任何形式的投资操作建议。接收者不应单纯依靠本报告而取代个人的独立判断，对于因使用、引用、参考本报告内容而导致的投资损失、风险及纠纷的，东方港湾不负任何责任。

		的 IDM 型存储龙头。依托在主流智能终端及服务器供应链中的稳固地位，以及在 DDR5、LPDDR5/5X 及高带宽内存（HBM）领域的突破性布局，公司展现出极强的国产替代弹性与技术护城河。在高端存储国产化率加速提升与国产算力基础设施建设计划扩张的大背景下，该股兼具自主可控溢价与存储周期上行红利，具备显著的长期价值。 公司是全球存储芯片三巨头之一，正处于“AI 驱动的存储超级周期”与“利润率重塑”的双重共振期。凭借领先的 1β/1γ 制程技术，在 2026 财年一季度交出了惊人的业绩。公司 HBM3E 已获得英伟达核心认证且 2026 年产能已提前售罄，同时在 1.6T 光模块及超大规模 SSD（如 245TB 系列）领域展现出极强的定价权。AI 应用正从训练向大规模推理端迁移，美光作为算力底座不可或缺的“供血者”，兼具周期向上的爆发力与技术领先的护城河。 公司是“AI 网络中枢+定制算力芯片”双重领跑者的极强防御性与高增长潜能。构建了极为深厚的技术壁垒，且已锁定谷歌、Meta 等核心客户至 2029 年的定制化订单。凭借极高的运营效率（经营利润率超 66%）及至 2027 年 AI 收入破千亿美元的清晰愿景，公司兼具半导体龙头的确定性与软件定义基础设施的溢价红利。
指数基金	25.25%	AI 作为通用目的技术革命,是一个比互联网还大的时代革命,且发展速度非常快,这会深刻改变人类的思维方式、工作方式和生活方式,改造几乎所有行业的商业模式,大幅降低生产三要素中人力成本,同时作为平台型技术衍生出更多的行业技术。一切都得益于多模态大模型在算力快速发展之下的能力提升。美股纳指在过往 50 多年来的复合回报超过 9.5%,随着 AI 对各个行业的赋能,美股公司未来数年的增长和成本下降,会带动指数进入新一轮的快速成长期。且大企业从芯片到大模型到应用,也在此起彼伏地推进 AI 发展,通过指数组合的形式可以降低非系统性风险,平衡先行者与潜力者的关系。 作为全球首只聚焦存储芯片的专题 ETF，该 ETF 精准卡位以高带宽内存（HBM）、DRAM 及 NAND Flash 为核心的数据存储与传输赛道。随着 AI 大模型训练与推理对高带宽、大容量存储需求的指数级递增，HBM 已成为 AI 服务器的核心产能瓶颈。在供给端高度垄断、需求端强劲拉动双重作用下，存储行业迎来量价齐升的超级周期，赋能该 ETF 极高的业绩弹性与供需议价权。在 AI 算力往“存算一体”演进与存储行业供需格局持续改善的大背景下，该标的兼具高 beta 弹性与存储寡头垄断红利，具备显著的跨周期配置价值。 该标的深度受益于“全球 AI 算力浪潮+半导体硬件代际升级”的景气周期。全面覆盖芯片设计、晶圆代工与设备以及模拟与功率芯片等核心环节。依托对美股半导体头部巨头的高集中度配置，该 ETF 展现出极高的业绩确定性与全球技术护城河。凭借全产业链领军企业的技术壁垒、规模效应与自由现金流生成能力，SOXX 既享有 AI 集群建设带来的高 Beta 盈利弹性，又兼具非 AI 芯片周期复苏的下行防御韧性，是拥抱全球科技硬件变革与算力基建长周期确定性的绝对龙头标的。

互联网软件与服务	20.82%	公司是全球搜索龙头，长期投资价值根植于其难以撼动的生态系统和前瞻性的战略布局。首先，其核心的搜索和广告业务，凭借强大的网络效应和用户习惯，构建了极深的护城河。这部分业务持续贡献着稳健的现金流，为公司的高研发投入提供了坚实基础。此外，公司在人工智能（AI）领域也具有深厚积累，同时在全面发力。从基础研究到应用层面，无论是赋能现有业务（如搜索、广告、安卓生态），还是开辟新的增长曲线（如 Waymo 自动驾驶、AI 大模型 Gemini），AI 正成为驱动公司下一轮增长的核心引擎。考虑到 AI 技术的颠覆性潜力及其在各行各业的渗透，公司无疑占据了浪潮之巅的有利位置。此外，公司的云业务虽然起步相对较晚，但增速很快，其在数据分析和 AI 基础设施方面具备独特优势，未来在企业级市场的拓展空间依然广阔。综合来看，公司凭借其核心业务的壁垒、AI 领域的领导地位以及云计算的增长潜力，具备穿越经济周期的韧性和持续创造价值的能力。
电子元件	8.01%	公司深度受益于“AI 算力爆发+新能源车智能化”的双重景气周期。公司通过整合索尔思光电与 GMD，已进化为“AI PCB+高速光模块+车载精密件”的全能型平台商。依托在苹果、特斯拉等核心供应链中的稳固地位，以及在 800G/1.6T 光模块领域的领先布局，公司展现出极强的盈利弹性与技术护城河。在算力基建持续扩张与汽车电子渗透提升的大背景下，该股兼具行业龙头溢价与跨赛道增长红利，具备显著的长期价值。
通信设备	7.82%	AI 的发展为 PCB 产业带来了显著的投资机遇。AI 训练和推理服务器、数据中心等基础设施对高密度、高层数、高速率的高端 PCB 需求激增。这些复杂 PCB 需要更高的技术门槛和更严格的制造工艺，从而提升了相关企业的毛利率和市场壁垒，其价值将随着 AI 算力需求的提升而水涨船高。选择那些在高频高速板材、HDI（高密度互连）和高层数板领域有技术优势和产能布局的 PCB 企业，是捕捉 AI 时代硬件浪潮的关键。

六. 投资观点

过去两个月，半导体硬件股票经历了较为深度的调整，市场忧声四起。但我们倾向于这是一种发展中的再平衡，让 AI 的发展变得更健康，而不是系统性风险的前兆。

这种再平衡，来自半导体企业未来 3 年超高的业绩预期和因此形成的资金拥挤度，与支撑预期的弱现实之间的矛盾，我们在 5-6 月的观点中提到过我们观察到的这种分裂。**但这种分裂，可以通过股价与现实变化来弥合，而这也正是过去两个月正在发生的事：**一方面，半导体板块普遍经历了巨大的回撤，其中存储行业更是下跌超过 60%。虽然我们担心的风险还都还没发生，但价格调整已提前将风险进行了计价。而另一方面，相比股价更重要的是，现实也在同步发生着变化。

一、AI 正在创造新需求

事实上，投资者所有担忧的核心，本质是怀疑 AI 投资究竟创造了什么增量价值。价值是指可以生产出的被人所需要的东西。当前 AI 更多是带来了生产力的巨大飞跃，这点相信已经少有质疑。但我们生产了再多的 AI 视频，更快地做出一版 PPT 或 Excel，APP Store 里出现再多 AI 编程出来的应用，也并没有创造出更多的需求：人们每天看视频的时间没法大幅提升，再多 PPT 也无法带来更多的生意，Excel 模型做得再好也没有让投资收益突飞猛进，手机常用的 APP 也并没有因此增加。

人工智能，作为一项通用目的技术会让整个行业都受益，成本飞速下降，产量大幅提升，而总需求如果没有显著增加，结果只会让行业变得极度内卷，所有人分到的蛋糕变少了，利润也变薄了。这就是今天发生在 AI 短视频和 AI 小游戏身上的故事，最后受益的往往只有分发渠道，吃到了广告增长的红利。这也是当前市场担心 AI 投资有效性的根本原因。

一项技术的发明通常都会被用以满足旧需求，像 AI 视频、AI 编程、AI 修图和 AI 办公软件，因为满足这些需求通常不需要对生产关系做出任何改变：团队人员、产品形态、分发渠道、收费模式以及市场接受度都无需变动，所以最容易发生。

但故事到这里还没结束。事实上，AI 技术正在逐渐进入各种任务场景，而对应的需求可能将拥有非常高的“需求弹性”。这大体可以划分成 3 个层次：

首先，AI 会把过去因为成本太高而不存在的需求，变成经济上可行的需求。这主要集中在现代服务业上，例如教育、医疗、法律、咨询、金融理财等。当价格大幅下降，产能上限打开后，人们的需求会被彻底点燃，谁不想拥有这些专属服务。当前的 AI 聊天框里，已经出现了这些专业服务的雏形，

风险提示及免责声明

本报告仅供本产品的持有人接收，任何媒体、网站或个人不得转载、链接、转帖或以其他方式复制发表。在任何情况下，本报告中提及的公司仅为阐述产业逻辑举例，不作任何投资建议，本报告所载内容仅供接收者参考之用，亦不视为任何形式的投资操作建议。接收者不应单纯依靠本报告而取代个人的独立判断，对于因使用、引用、参考本报告内容而导致的投资损失、风险及纠纷的，东方港湾不负任何责任。

需求是非常大的。但仅仅是刻在参数里的专业知识和互联网的专业文档，最多只能叫“搜索”，不能叫“服务”。以“AI 教育”服务为例，除了“知识问答”外，还需要 AI 老师根据学生的学习阶段和目标，给出具体学习方案；在学习过程中做动机管理，生成有趣的交互界面，保持学生足够的学习动力；监测并保留学习成果，甚至链接外部任务给予更多成就感。这就需要 Agent 有成熟的教学 Skill，能够生成实时的互动界面，有足够的数据库支持，还能够主动与学生进行互动，等等。这正是当前 Harness 系统在进化的方向。这些需求会以私人 AI 医生，私人 AI 老师，私人 AI 法律顾问，私人 AI 理财师等产品形态出现在市场上，满足过去没法被满足的需求。虽然我不清楚这样的服务会集中在一个通用 Agent 的入口，还是会以垂直 Agent 的形式出现。

其次，AI 会把原来人们因为能力不足而做不到的事情，降低门槛变成真实存在的需求。这主要集中在创作领域，例如软件工具、视频制作、游戏制作、多人项目组织、3D 打印等。过去人们只能从别人制作的标准化产品中做出选择，来满足自己的部分需求，AI 降低了生产工作的门槛之后，自我创作以满足需求（而非通过市场交易），会变得更加司空见惯。就像消费级 3D 打印正在把轻量制造普及化一样，这对于商品经济来说也是一种颠覆性创新。因此，过去只属于 B 端的创意工具，在 AI 的赋能下也会进入 C 端市场，消费级 3D 打印机、Vibe Coding 和 AI 视频制作 Agent 的普及化，只是这个趋势的开端。本月，GPT-6 问世后，又带火了建模软件 Blender 软件工具，掀起了一波 3D 场景建模热潮，把游戏制作的门槛又拉低了一个层级。或许不久后，组建 AI 团队进行创业可能也是一种消费，因为门槛和成本都可能会被 AI 大幅压低。所以，我们或许不用过于在意这些生产力的提升是否有对应的需求，可能生产本身即消费。

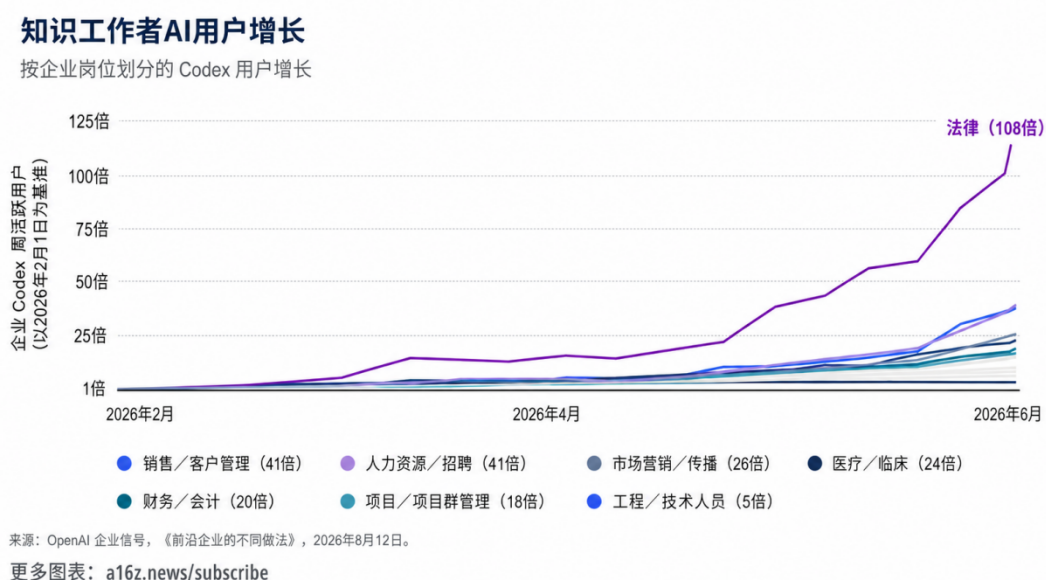
再者，AI 会创造出现在人们自己还不知道的需求，这才是真正的消费革命。这可能出现在各行各业。例如，游戏行业可能会借助 AI 视频生成技术实时生成剧情、人物和场景，改变传统的预设机制，创造出真正的“生成式”游戏。本月，英伟达发布了新的游戏显卡算法——DLSS 5，以实时生成的方式，改变游戏场景人物的光照和材质表现，直接让许多老游戏的画面出现了质的飞跃，逼近现实画面。同时，本月 GPT 6 Astra 的问世，也让我们进一步看到了生成式游戏的雏形。就像 AI 视频行业的发展，从 2023 年粗糙的“威尔史密斯吃面条”，到今天视频平台上充斥的各种 AI 作品，AI 生成式游戏这种新物种，可能也会在 AI 成本下降到一定程度后走进消费者的视野。再比如，与传统游戏和视频的体验式需求不同的是，当 AI 生成式技术发展到一定程度后，人们可能会萌生出“拥有一个世界”的收藏需求。它不是统一剧情的游戏，而是由你定规则、持续演化并铭记历史的世界。你建的港口、引入的物

风险提示及免责声明

本报告仅供本产品的持有人接收，任何媒体、网站或个人不得转载、链接、转帖或以其他方式复制发表。在任何情况下，本报告中提及的公司仅为阐述产业逻辑举例，不作任何投资建议，本报告所载内容仅供接收者参考之用，亦不视为任何形式的投资操作建议。接收者不应单纯依靠本报告而取代个人的独立判断，对于因使用、引用、参考本报告内容而导致的投资损失、风险及纠纷的，东方港湾不负任何责任。

种、经历的事件，都会永久影响它。可邀友共居，也可与家人共育。你不必每天进去，也不必每天完成任务。它的价值可能在于另一种感受：世界上有个地方，因我参与而成为现在的样子，就像有人养花、维护庭院、收藏物品一样。这种游戏的需求会超越游戏者时间的限制，或许也打开一个更大的需求空间。

改变需求这件事，需要组织架构、产品形态、分发渠道、收费模式与消费者培育的共同进步，与数值世界里快速进步的模型能力相比，更需要时间，但方向是确定的。最近两个月最大的变化，就是市场上主流的大模型产品形态，都已经完成了初步的 Agent 化——从回答问题，走向调用工具执行多步任务，并在不断扩大任务的范围和完成度。在必争的编程任务以外，GPT Codex 现在可以在人类的指挥下，做出超越我们团队分析师的 excel 财务模型。Claude Tag 开始脱离个体工作环境，在 Slack 团队协作办公软件中，利用已授权的工具和数据完成任务，变成了团队的一员。腾讯的 Workbuddy 也开始接入微信的环境，帮你阅读和总结群聊的信息，省去大量“爬楼”的时间。Agent 化是开启 AI 新需求的重要基石，也是市场一直呼唤的“除了 Coding 以外还能干什么”的答案。（见下图）



二、Agent 普及的不竭动力：任务环境交互经验与 RSI

Agent 的普及扩散才刚刚开始，持续进步的动力来自 RSI（递归式自我改进）的新范式。过去，大语言模型的进步源于对人类已有互联网数据的压缩，所以市场一度特别担心数据和预训练见顶。而今

风险提示及免责声明

本报告仅供本产品的持有人接收，任何媒体、网站或个人不得转载、链接、转帖或以其他方式复制发表。在任何情况下，本报告中提及的公司仅为阐述产业逻辑举例，不作任何投资建议，本报告所载内容仅供接收者参考之用，亦不视为任何形式的投资操作建议。接收者不应单纯依靠本报告而取代个人的独立判断，对于因使用、引用、参考本报告内容而导致的投资损失、风险及纠纷的，东方港湾不负任何责任。

天，Agent 的进化来源，是 AI 在真实任务环境里进行试错探索所积累下的经验。我们在 25 年 7 月的观察《人类数据 VS. AI 经验》一文中写到过：“Agent 一旦落地于真实任务环境中，以第一人称的视角去执行任务并与环境产生互动反馈，所积累的数据将更动态、更高质量、更具定制性且源源不断，其价值远超传统训练集，更加接近生命的本质。”今天产业界称之为 RSI，在今年 6 月份 Anthropic 发布的论文《When AI builds itself》后，被广为讨论。我们观察到，Agent 在 LLM 和 Harness 系统两层架构中，都开始出现了在人类帮助下自己改进自己的现象。

在 LLM 侧的模型训练中，DeepMind、Anthropic 和 OpenAI 都已经在做 RSI 的“早期工程形态”，虽然还没有进入严格意义上无人干预的递归自我改进，但半自动化 RSI 机制已经通过以下至少 5 个方面，不断做着自我改进：

1. **推理经验数据的再训练**：模型在实际执行任务过程中产生大量推理与行动轨迹，通过环境反馈和 Verifier 对这些轨迹进行验证、筛选和压缩。一部分经验可以直接沉淀为 Harness 的 Skill/Memory，在推理时复用；另一部分则被转化为 SFT、RL、偏好学习或蒸馏数据，通过后训练更新模型参数；经过更大规模积累后，也可能进入下一代模型的预训练或 mid-training 数据。于是更强模型能够产生更高质量的经验和训练数据，进一步训练出更强的下一代模型，形成递归式改进循环。

2. **AI 验证器**：让更强的模型帮忙“出题、判卷、找错误”，用这些反馈筛选训练数据、训练下一代模型。新模型变强后，再帮助改进“题目”和“判卷方法”，形成“评得更准—训练得更好—再评得更准”的循环。

3. **自动化研究探索**：人类给出研究目标、资源边界和初始评价标准，AI Agent 自主提出假设、编程、运行实验、调参、Debug、分析结果和不断优化。截至 2026 年 5 月，Anthropic 进入生产代码库的代码中，超过 80% 已经由 Claude 编写。同时，Claude 会在保持正确性的前提下，对训练 AI 模型的代码进行提速。2025 年 5 月 Opus 4 平均优化约 3 倍训练速度；而 2026 年 4 月内部 Mythos Preview 已达约 52 倍。

4. **攻防式自我改进**：通过攻击模型与防守模型的自博弈，不断发现更强攻击并将其转化为安全训练数据和评价集，使模型攻击能力与防御能力共同升级，提升模型的安全度和价值观对齐。

5. **训练基础设施优化**：AI 能优化训练代码和算力调度，提升算力使用效率；还能辅助尝试模

风险提示及免责声明

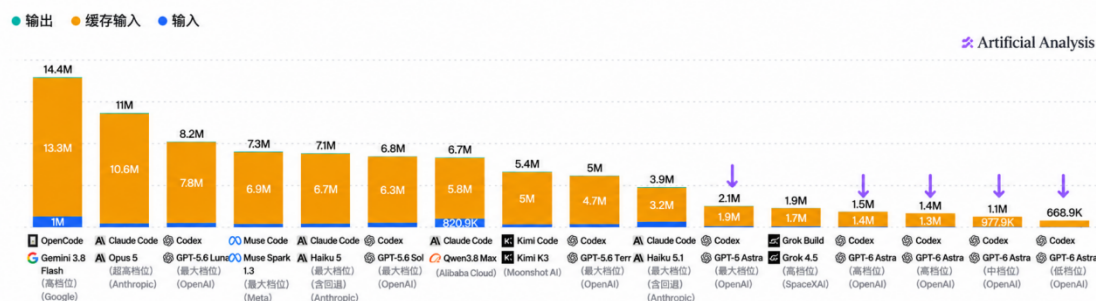
本报告仅供本产品的持有人接收，任何媒体、网站或个人不得转载、链接、转帖或以其他方式复制发表。在任何情况下，本报告中提及的公司仅为阐述产业逻辑举例，不作任何投资建议，本报告所载内容仅供接收者参考之用，亦不视为任何形式的投资操作建议。接收者不应单纯依靠本报告而取代个人的独立判断，对于因使用、引用、参考本报告内容而导致的投资损失、风险及纠纷的，东方港湾不负任何责任。

型结构、参数和数据配比，找到更优的推理方案。同等资源下可完成更多实验或训练出更好模型，并用更低的成本来推理。

而在 Harness 侧，递归式自我改进的范式，是 Agent 会从真实任务的成功与失败中，提炼环境规律、用户偏好和工具正确用法，沉淀为可复用的 Memory 与 Skill；并不断根据上下文管理效率，来调整记忆压缩和使用的机制；同时，调整不同大小模型路由方式，以不断提升任务完成的准确率和速度；同时，不断降低任务完成的循环步数和工具使用率，以降低任务完成成本，也是最近发布的 Meta Spark 1.3、Gemini 3.8 flash 和 GPT 6 的重要改进方向（见下图）。这些 Harness 层面的自我改进，即使底层 LLM 参数保持不变，Agent 的整体能力也可能不断提升。

GPT 6降低单任务运行成本与时间

每项任务的平均输入、缓存和输出 Token 数



① 不同提供商的提示缓存命中率会因路由而显著不同，这会实质性改变实际成本。

Agent 在 RSI 自我改进之路上能走多快多远，最核心的“燃料来源”是“可验证的环境交互经验”。Agent 必须不断进入真实或高保真的任务环境，采取行动，并从环境中获得可验证的反馈，再把这些反馈转化为可保留的能力改进。这与人类学习的范式如出一辙。所以未来 AI 厂商的竞争，可能得看谁能够让模型接触到更广的任务环境，获得最多的高质量“行动—反馈”数据，并将其转换成新的 Agent 能力。同时这种环境交互经验，也是消费者使用 Agent 时的“个性化”来源，能够形成一定的竞争壁垒。

三、迷雾渐散去

当 Agent 任务不断普及，使用成本不断下降时，新的消费需求才会被激发出来。而这正是过去两个月，正在发生的事实。在这个过程中，生产关系的变化要更慢也更复杂，所以软件行业最终的产品

风险提示及免责声明

本报告仅供本产品的持有人接收，任何媒体、网站或个人不得转载、链接、转帖或以其他方式复制发表。在任何情况下，本报告中提及的公司仅为阐述产业逻辑举例，不作任何投资建议，本报告所载内容仅供接收者参考之用，亦不视为任何形式的投资操作建议。接收者不应单纯依靠本报告而取代个人的独立判断，对于因使用、引用、参考本报告内容而导致的投资损失、风险及纠纷的，东方港湾不负任何责任。

形态、行业地位、组织架构何去何从，属实难以预判。但 AI 生产力能够创造的价值，却是相对确定的：我们可能正在经历一次历史罕见的硬件创新进步浪潮，算力存力等硬件能力的升级，加上算法与模型的进步，让同样投资金额下所生产的 AI Token 数量每年以数十倍在增长（参考 B300 相对于 H200 的 token 相对产出量），同时每 token 的智能含量也在飞升。这使得我们所讲的 Agent 普及度和所创造的价值不断提升，也是当前算力价格一路不断上升背后最核心的驱动因素（如下图）。如果模型层因为竞争而没法出现长期垄断的玩家，那么技术进步所创造的价值，可能会被消费者和半导体硬件所瓜分。



长期投资的道路，80%的时间都是寂寞和痛苦的，即便是长期牛股，实际上涨的时间一般不会超过 20%。在过去两个月的阴霾时段，半导体的深度调整以及 Agent 自我递归的成长，让市场得到了事实与预期之间的再平衡。临近月底，英伟达在业绩会上提出了 2027 年 70% 的增长预期，也让未来的迷雾又散开了几分。一切似乎又在重回轨道。

我们一如既往地看好 AI 时代的发展趋势，所投资公司的竞争力依旧牢固，我们一直以满仓状态在迎接这个新时代。我们判断从 GPT-6 开始算，AI 算是正式完全踏入了 Agent 时代。Agent 会在许多行业里开始普及，会衍生出许多 AI 职业服务（而不仅是工具），这是我们投资信心的重要来源。不要辜负一个时代的馈赠！

风险提示及免责声明

本报告仅供本产品的持有人接收，任何媒体、网站或个人不得转载、链接、转帖或以其他方式复制发表。在任何情况下，本报告中提及的公司仅为阐述产业逻辑举例，不作任何投资建议，本报告所载内容仅供接收者参考之用，亦不视为任何形式的投资操作建议。接收者不应单纯依靠本报告而取代个人的独立判断，对于因使用、引用、参考本报告内容而导致的投资损失、风险及纠纷的，东方港湾不负任何责任。