

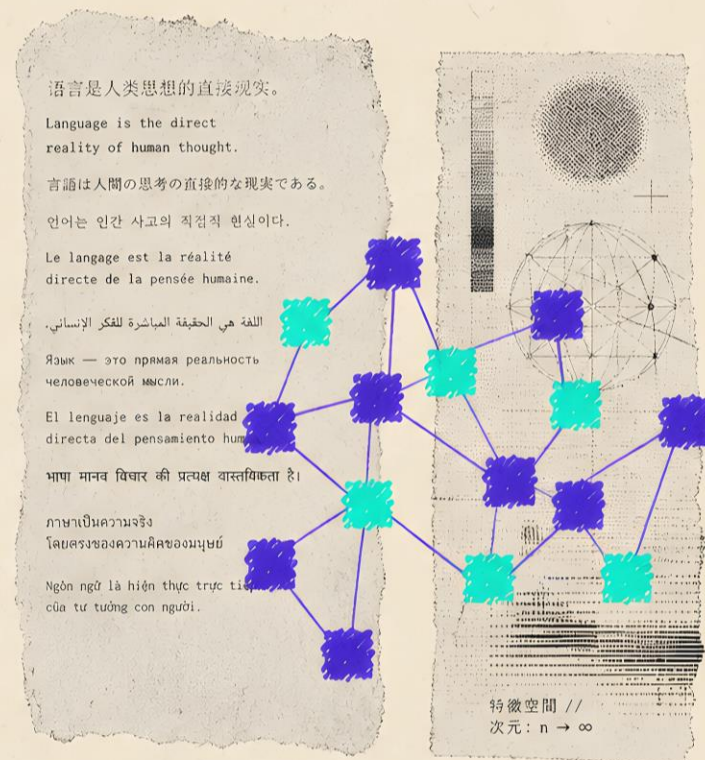


商汤集团股份有限公司

2026 中期业绩简报

坚持原创，让 AI 引领人类进步

2026.08



AI FOR A BETTER TOMORROW



LATENCY FIELD
08.07 / 04:32

t
o
k
e
n
s
...

免责声明

本演示文稿中的信息由商汤科技集团有限公司（“公司”，连同其子公司统称为“集团”）的代表为集团在演示中使用而准备，仅供信息参考。此演示文稿的任何部分均不应成为或依赖于任何合同、承诺或投资决策的基础。

对于本演示文稿中包含的信息或意见的公正性、准确性、完整性或正确性，集团及其董事、监事、高级职员、合伙人、员工、关联方、代理人、顾问或代表均不作任何明示或默示的声明或保证，并且不应对此加以依赖。因使用本演示文稿或其内容或与本演示文稿相关的任何其他原因所引起的任何损失，集团及其董事、监事、高级职员、合伙人、员工、关联方、代理人、顾问或代表概不承担任何责任（无论是由于疏忽还是其他原因）。本文中的信息可能会进行更新、补充、修订、验证和更改，并且此类信息可能会发生重大变化。

本演示文稿基于截至演讲当日的经济、监管、市场和其他条件。应理解为，后续的发展可能会影响本演示文稿中包含的信息，集团及其董事、监事、高级职员、合伙人、员工、关联方、代理人、顾问或代表没有义务更新、修订或确认这些信息。

本演示文稿中传达的信息包含一些前瞻性陈述。这些陈述通常包含“将”、“可能”、“预计”、“预测”、“计划”和“预期”等词语及类似表达。前瞻性陈述由于涉及未来的事件和依赖于未来的情况，因此具有风险和不确定性。可能还存在当前未被认为是重大风险或集团及其顾问或代表未意识到的重大风险。在这些不确定性的背景下，读者不应依赖这些前瞻性陈述。集团不承担更新前瞻性陈述或使其适应未来事件或发展的责任。

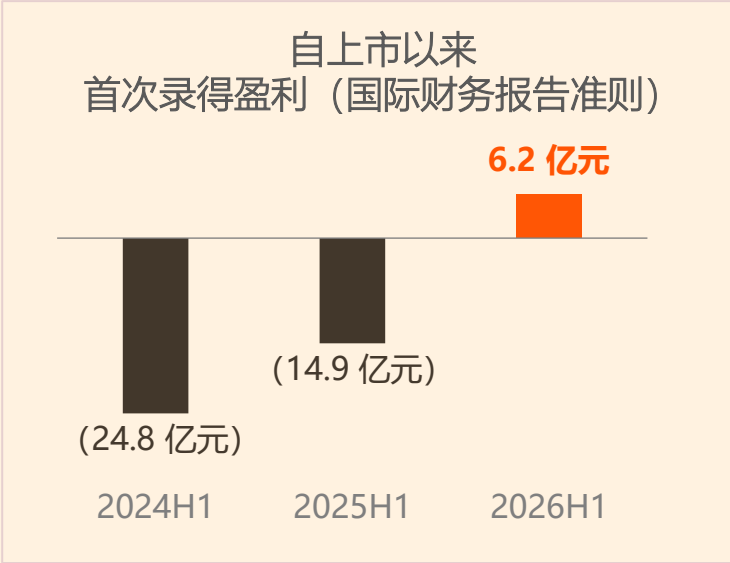
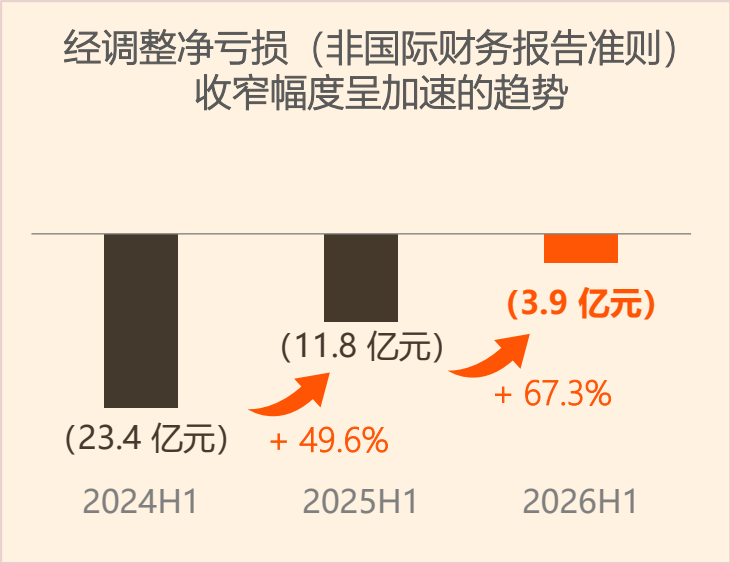
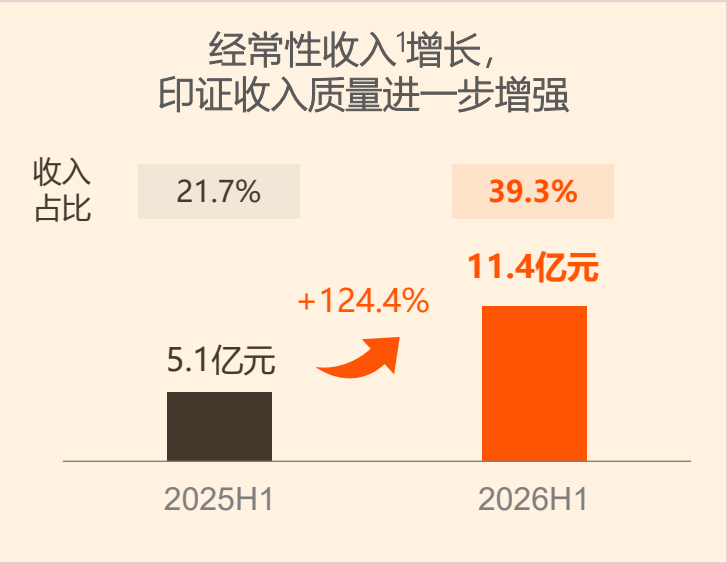
本演示文稿及其包含的信息不构成或不应被视为在任何司法管辖区内销售或认购或招揽或邀请购买或认购公司或其任何子公司或关联方的任何证券的要约的一部分。本文及其包含的信息高度保密，仅供您参考，不得以任何方式全部或部分复制或再分发给任何其他人。

参加本演示文稿会议即表明您承认您将独自负责对市场和集团的市场地位进行评估，并且您将自行进行分析并独自负责形成对集团业务潜在未来表现的看法。

2026年中期业绩亮点



集团总收入	生成式 AI	毛利	毛利率
¥29.1 亿	¥23.3 亿	¥12.1 亿	41.4%
▲ 23.4% YoY	▲ 28.2% YoY 占集团收入 79.9%	▲ 32.9% YoY	▲ 294 基点



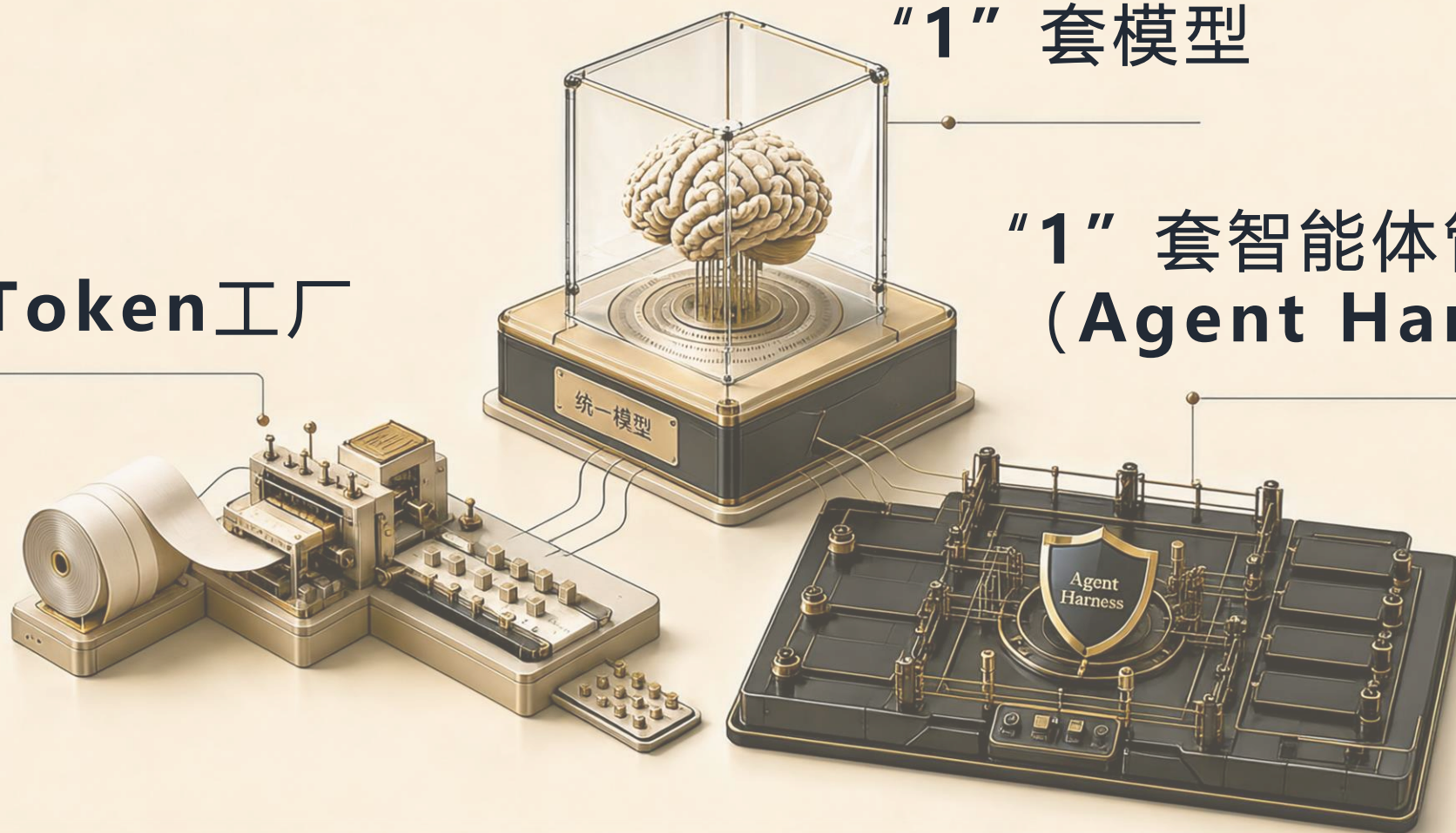
注：
1. 基于报告期内已生效合同，而具备持续续约属性的收入

生成式AI 构建系统级AI能力

"1" 个Token工厂

"1" 套模型

"1" 套智能体管控系统
(Agent Harness)



“1” 套模型 - 统一多模态架构，持续提升智能上限

多模态 LLM



多模态 Agent



物理 AI



原理解生成统一 奠定模型架构基础



U 1 系列开源不到四个月，已在 GitHub 累计获得超过 **1.1 万个** 星标

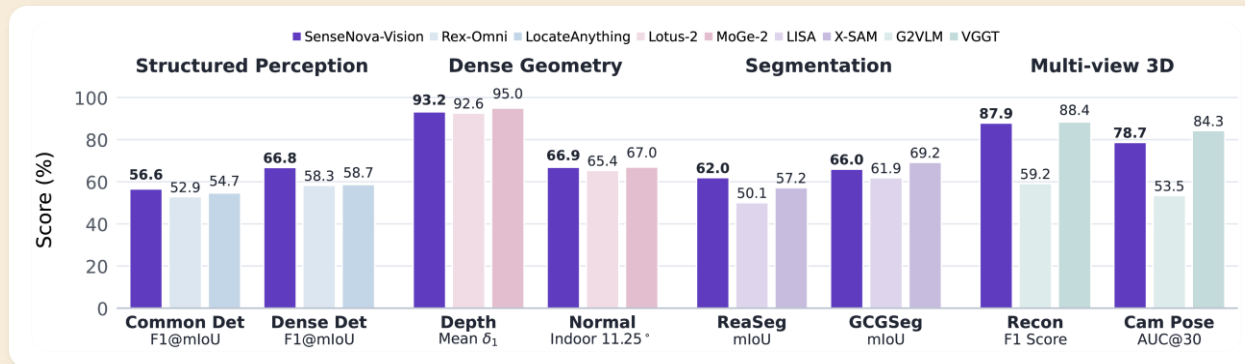
“1” 套模型 – 统一架构同时提升视觉推理与可控生成

SenseNova
Vision

多项 任务¹

超过 Vision Banana
(Google DeepMind)

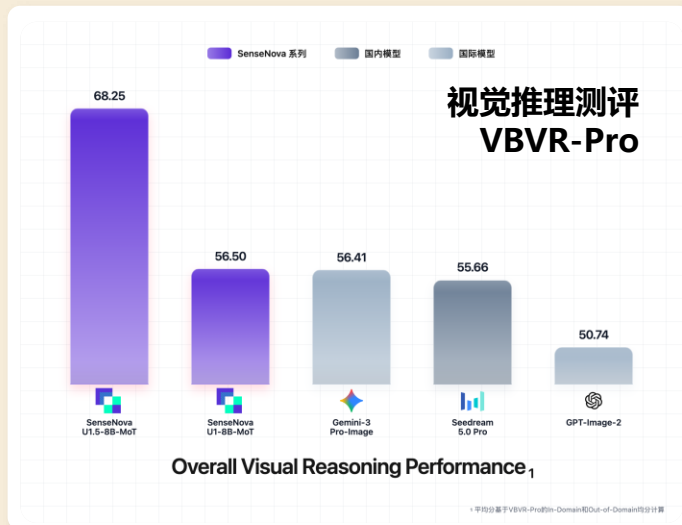
注：1. 视觉理解、稠密几何预测、图像分割和多视角 3D 几何等任务



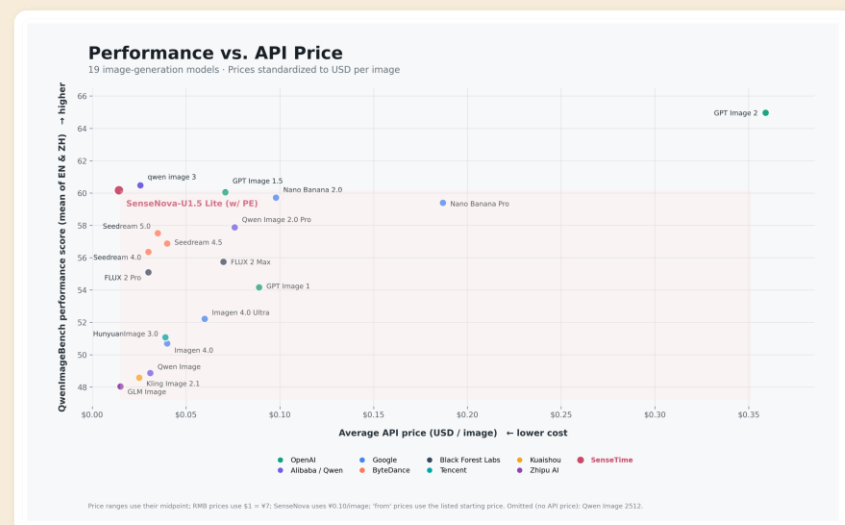
68.3 分

超过 Gemini 3 Pro 和 GPT Image 2

SenseNova
U1.5-Lite



8B-MoT 实现原生 4K 生成
比肩 Nano Banana 2



多模态长程智能体 系统级高价值交付能力

办公场景

SenseNova 6.8-Flash-Lite

质量与交付成本的同步优化

- Token 消耗平均降低约 60%
- 在多项基准测试中全面提升
(ClawEval、ArtifactBench、TeamBench...)
- 领先 DeepSeek-V4-Flash、Gemini 3.5 Flash
和 GPT-5.6 Luna



多模态长程智能体 系统级高价值交付能力

内容创作

SenseNova U1-Pro

规模化与智能体生成

- 从生成单张图片迈向 **交付专业内容成果**
- 创作推理和可靠文字生成
- 比肩GPT Image 2，领先国内多模态模型



即将发布

SenseNova 6.8-Flash

进一步拓展复杂任务和
多智能体协作能力

物理AGI

数字智能 → 物理智能

空间智能

SenseNova SI-8B

8B 参数

八项基准 领先水平

理解空间关系和三维世界

世界模型

Kairos 3.1

8B 参数 - BF16 精度

125ms 推理延迟

对物理世界的理解、生成与预测

理解世界 → 推演未来 → 驱动行动 → 获取反馈 → 迭代模型

1个Token工厂

从算力供给升级为规模化智能生产



算力供给



Token工厂



任务交付



“1” 个Token工厂 - 有力支撑商汤大模型技术摸高和规模化商用

超长上下文并行训练

Agent 模型

2 倍

训练速度提升

大规模强化学习加速

RL后训练阶段

**任务解耦
异构加速**

效率提升

异构混推

LightX2V

6 倍

生成效率提升

国产软硬件适配

规模化部署

1 周

部署周期缩短至

国产生态建设

SenseNova U1携手 **10+ 家** 国产芯片厂商，实现 **Day 0** 适配

Cambricon
寒武纪科技

METAX
沐曦

HYGON
中科海光

摩尔线程
MOORE THREADS

Ascend

sunrise

Enflame
燧原科技

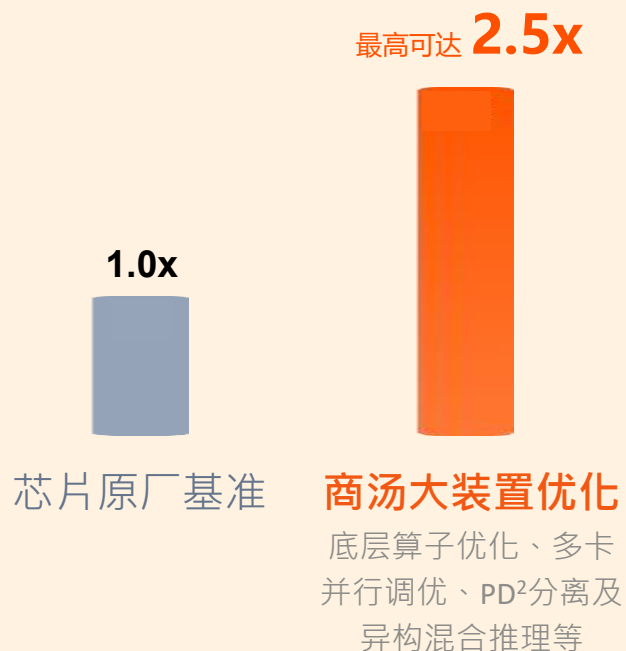
壁仞科技
BIREN TECHNOLOGY

.....

(排名不分先后)

“1” 个Token工厂 - 推动国产算力生态建设和规模化商用

主流国产芯片MFU¹提升



持续拓宽Token服务的全球化版图



注:

1. Model FLOPs Utilization 模型浮点运算利用率, MFU = 模型实际有效浮点运算吞吐 / 集群理论峰值算力 (Achieved model FLOPs per second / Theoretical peak cluster FLOPS)
2. Prefill/Decode

“1” 个Token工厂 – Token 服务规模快速增长

算力运营底座

运营总算力规模达**4.8万P**¹

商业化拓展突破

服务**4家**外部基础模型厂商

高复杂度场景全面落地

AI4S

视频生成

具身智能

世界模型

.....

2.4万亿+

日均Token服务量 | 22x同比激增



注：

1. 截至2026年8月底

“1” 个Token工厂 - 以算电协同持续降低单位智能成本

算电协同 Agent

96%

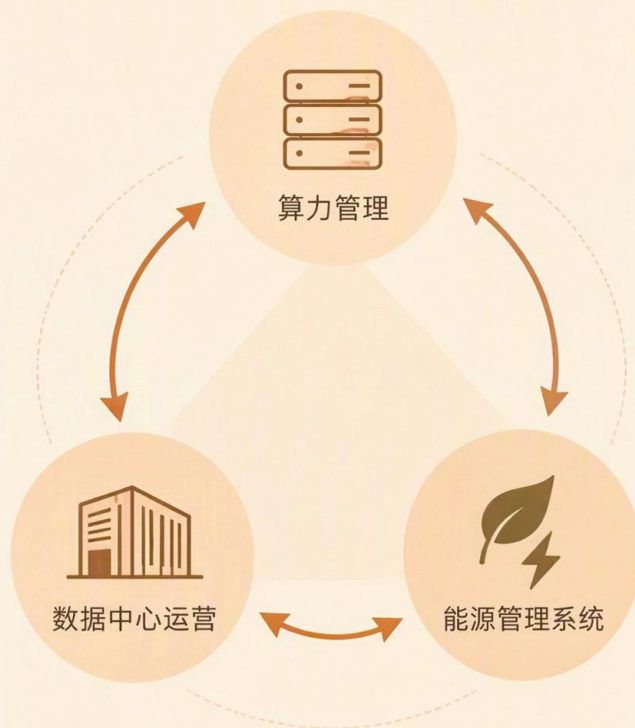
AI 负荷预测准确率

↓ 8%

平均单位用电成本降低

> 1,200 万元

2026 H1 电力累计降本



年碳减排量



每万P年化碳减排能力

24,000 吨

4,000 吨

2025



现在

“1”套智能体管控系统 – 面向企业与个人交付完整任务

多元智能体产品 承接不同用户、行业 and 任务



办公智能体



内容生成智能体



终端交互智能体



智能体管控系统



任务理解



智能调度



可信交付

共享统一底座



模型系统



Token工厂

企业级服务：进入行业 workflow，交付业务结果



公有云服务



API服务



私有企业级部署



FDE落地



商汤小浣熊
Raccoon Work

500%

用户规模增长

600%

企业用户增长

x 10

云服务月活跃
用户规模增长



OPC能力挑战季吸引超过 **65万人** 参与，沉淀超 **8,000** 份真实场景作品



Seko

10,000 分钟

单日视频产量

15亿

累计播放量

10 部

短剧播放量破亿



《告别》

抖音累计全网总播放量破 **亿**



《聚宝仙盆》

红果热度超 **5,000万**

个人助理：让产业级AI能力进入个人生产与生活



用户增长曲线

4,500_w

咋皮家族用户数

90w

25年1月

25年6月

25年12月

26年8月

4500w

“个人CXO” 产品矩阵



咋皮记账
个人财务助手



咋皮健康
个人健康助手



Kapi相机
个人创意助手

视觉AI市场领导地位进一步巩固

20_↑

海外覆盖国家地区

4,500₊

累计客户数

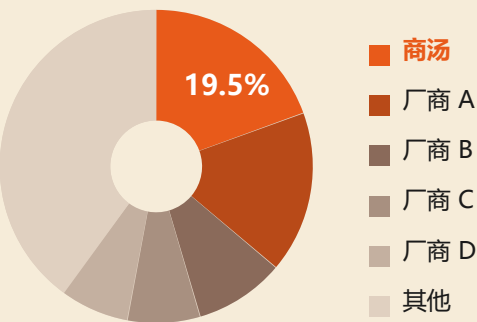
67%

复购率¹

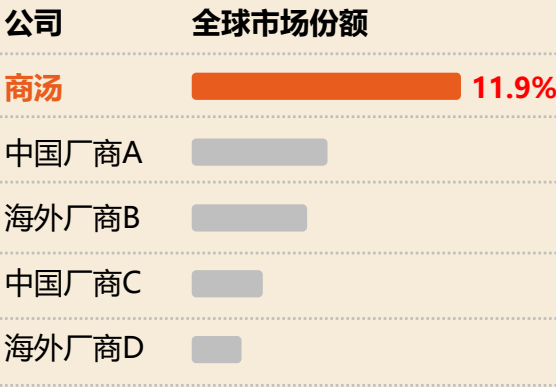
14%

收入增长

连续十年蝉联
中国视觉AI市场第一²



首度登顶 全球及亚太区
视觉AI市场 (Omdia)³



视觉专项科技成果获
科技进步二等奖



生成式AI计算机视觉
前沿技术创新者⁴



注：
1. 计入复购率的客户为本期收入中上一年度采购的客户；2. 《中国AI软件市场半年度追踪报告》，IDC；3. 《Video Analytics Market Share》，Omdia；4. 《Emerging Tech: AI Vendor Race: Tech Innovators Shaping AI Vision Intelligence Solutions》，Gartner。

生态布局 - 连接数字智能与物理智能



具身智能



智能终端



光/量子计算



AI Infra

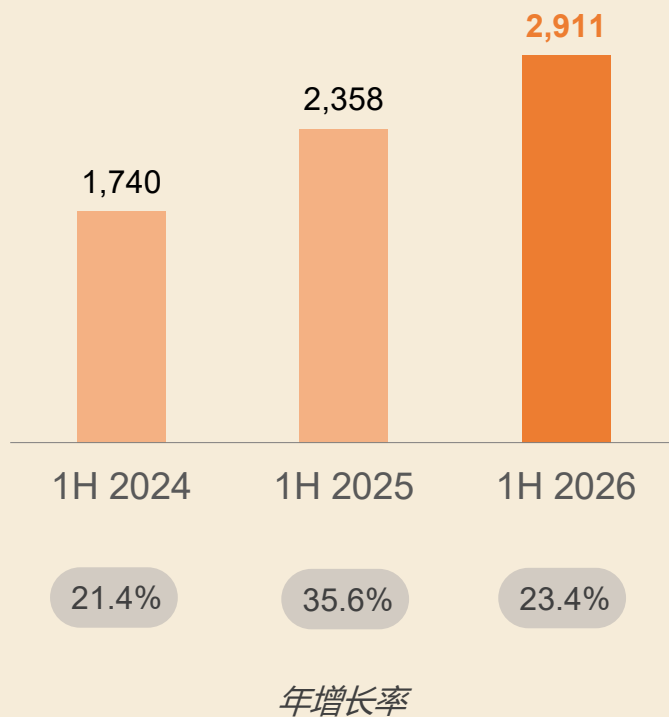
财务回顾

-

收入和毛利

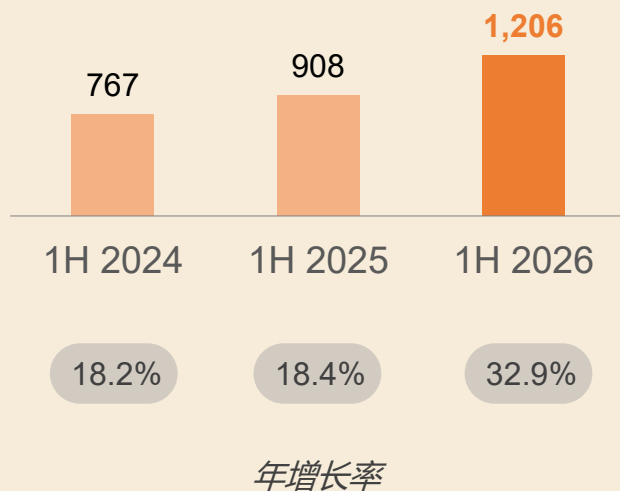
营业收入

(人民币百万元)

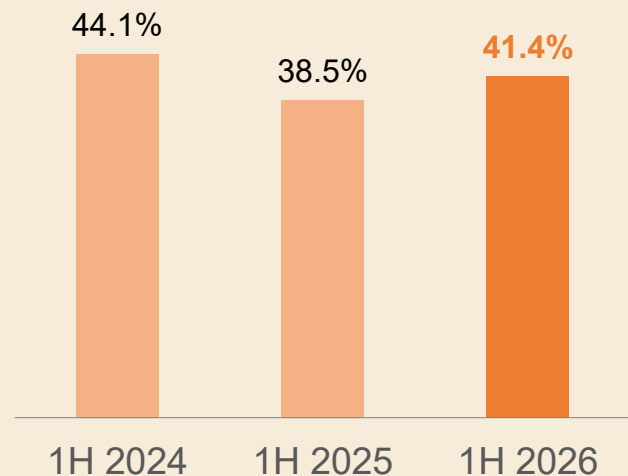


毛利

(人民币百万元)



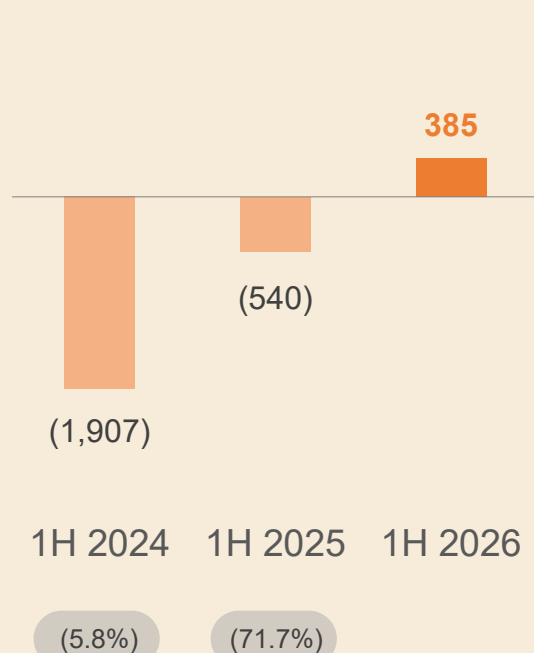
毛利率



利润相关指标

经调整⁽¹⁾ EBITDA

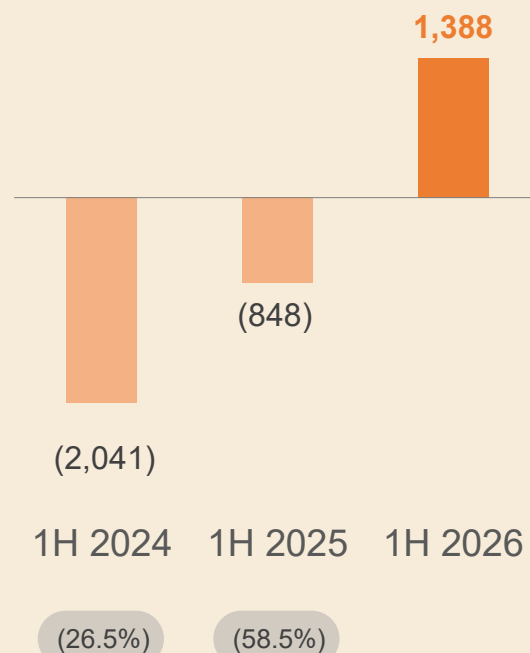
(人民币百万元)



年变动率⁽²⁾

EBITDA

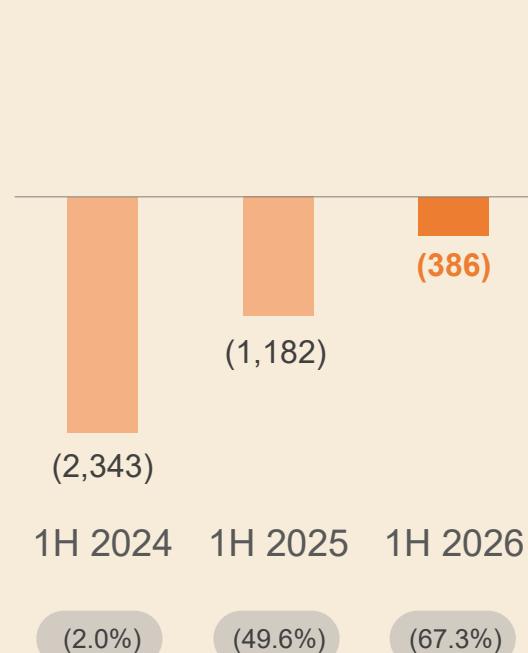
(人民币百万元)



年变动率⁽²⁾

经调整⁽¹⁾净利润 / (亏损)

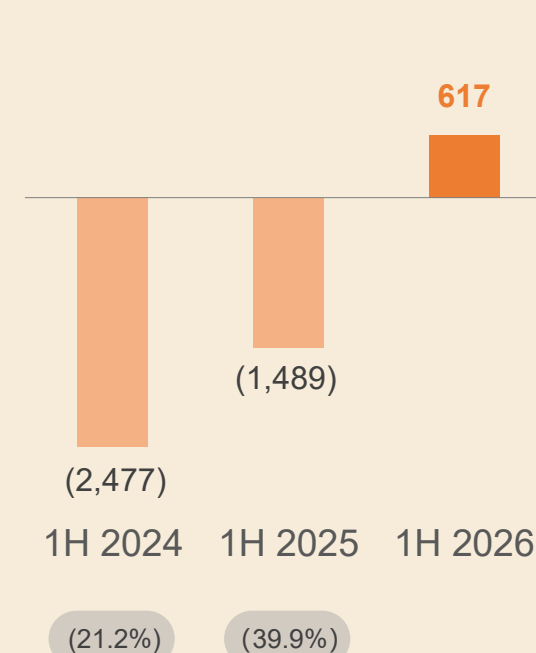
(人民币百万元)



年变动率⁽²⁾

净利润 / (亏损)

(人民币百万元)



年变动率⁽²⁾

备注:

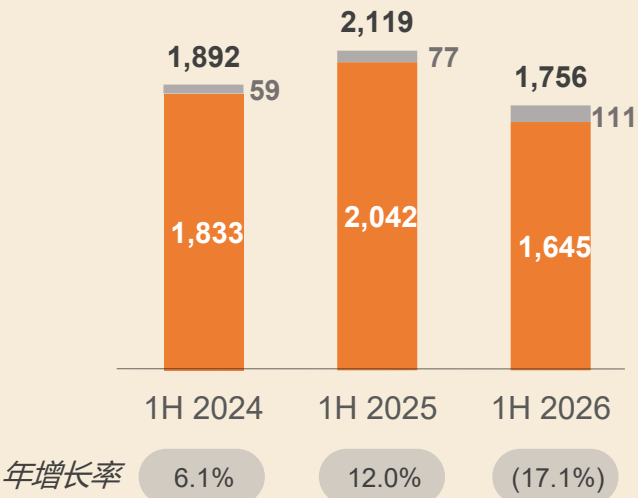
1. 调整项: 优先股的公允价值亏损、以股份为基础的薪酬开支、按公允价值计量且变动计入损益的金融资产的公允价值未实现利得/亏损、归属于有限合伙人的投资基金资产净值、为股权融资发生的专业服务费
2. 负数代表亏损收窄

运营支出

研发费用

(人民币百万元)

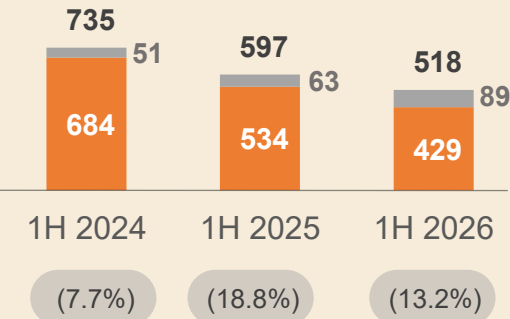
费用 (扣除SBC)
SBC



管理费用

(人民币百万元)

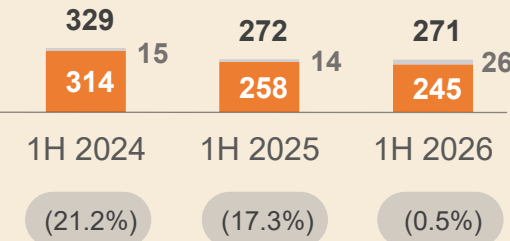
费用 (扣除SBC)
SBC



销售费用

(人民币百万元)

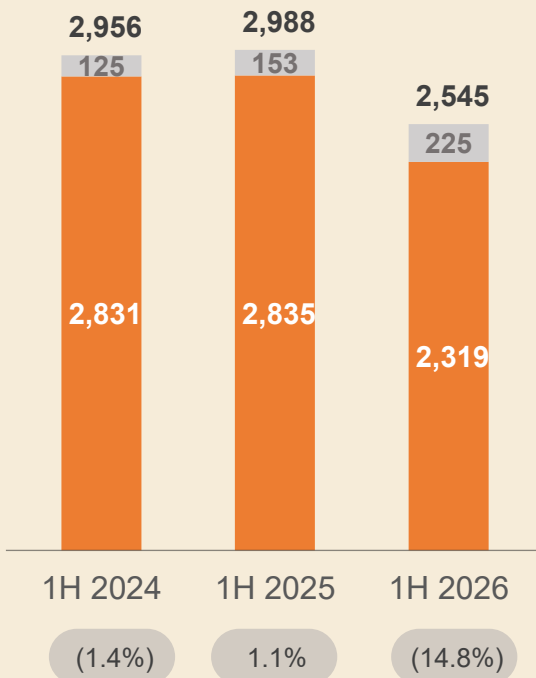
费用 (扣除SBC)
SBC



总运营支出

(人民币百万元)

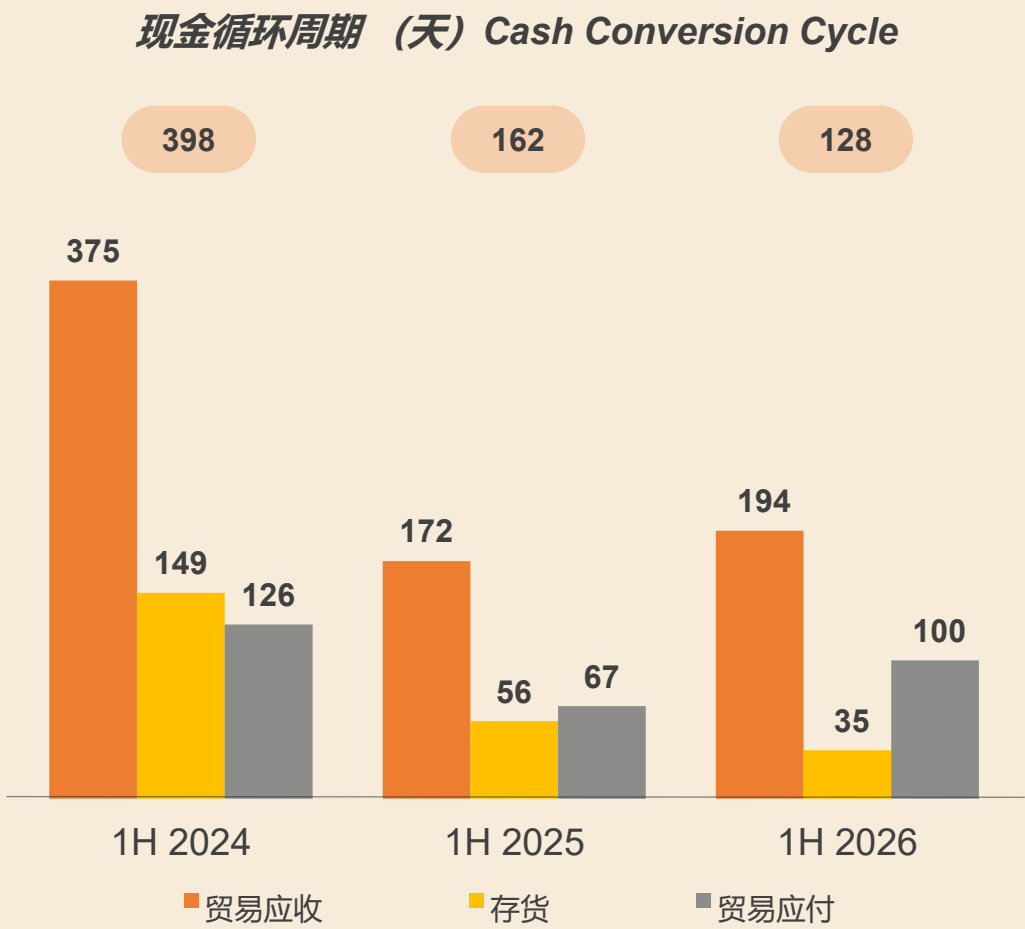
费用 (扣除SBC)
SBC



备注：
1. 运营支出 = 研发费用 + 管理费用 + 销售费用

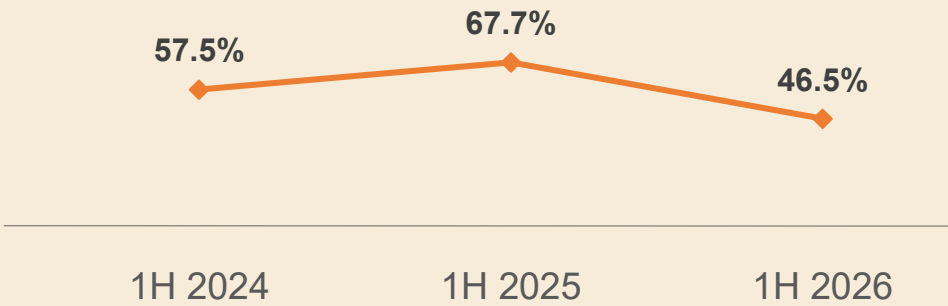
营运资本

周转天数 – 以平均值计算⁽¹⁾

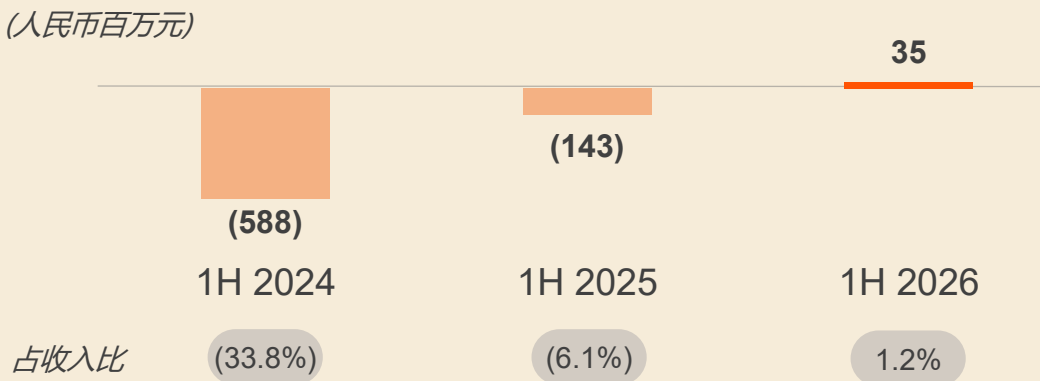


备注:
1. 以平均值计算的周转天数, 等于期初与期末的平均余额除以过去6个月的收入/成本, 并乘以180天

减值拨备占贸易应收额比率 (资产负债表)



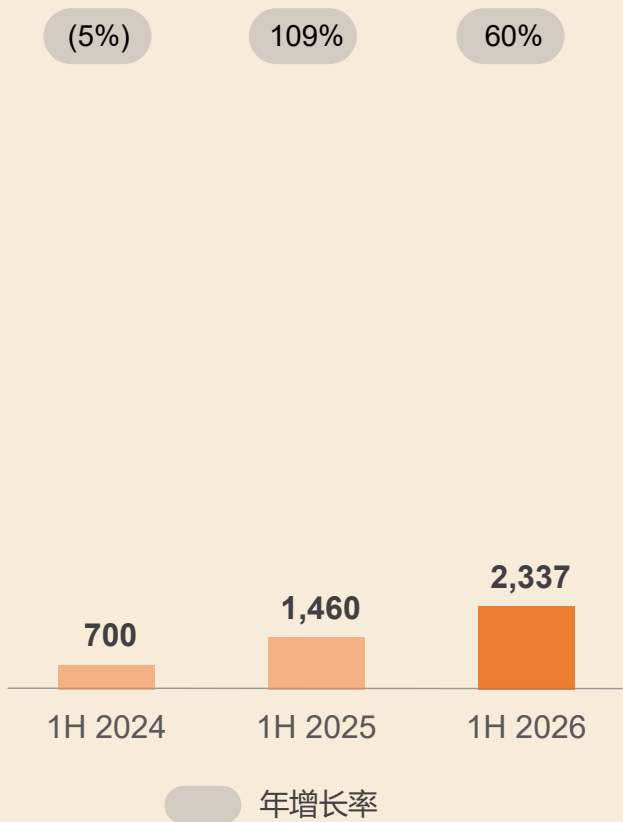
金融资产减值及合约资产亏损净额 (利润表)



资本支出与现金储备

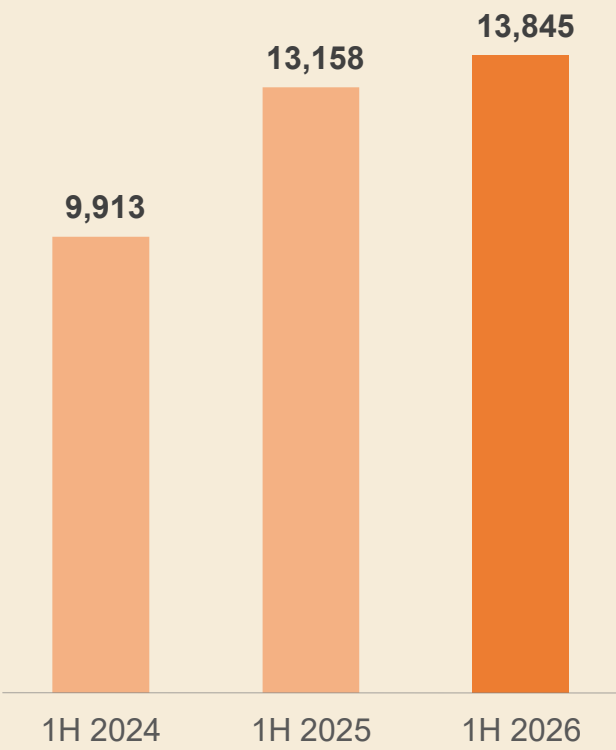
资本支出⁽¹⁾

(人民币百万元)



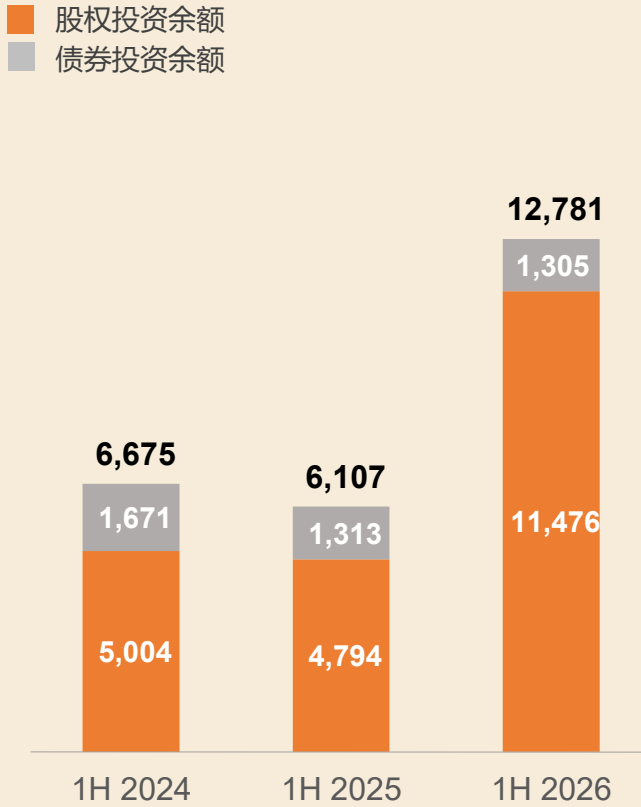
总现金储备⁽²⁾

(人民币百万元)



股权及债券投资余额⁽³⁾

(人民币百万元)



备注:
1. 定义为购买物业、厂房及设备+购买无形资产+购买土地使用权的现金流出
2. 定义为现金和现金等价物+定期存款+结构性存款+受限制现金
3. 定义为按公允价值计量且变动计入损益的金融资产 (非流动资产)

“1” 套模型

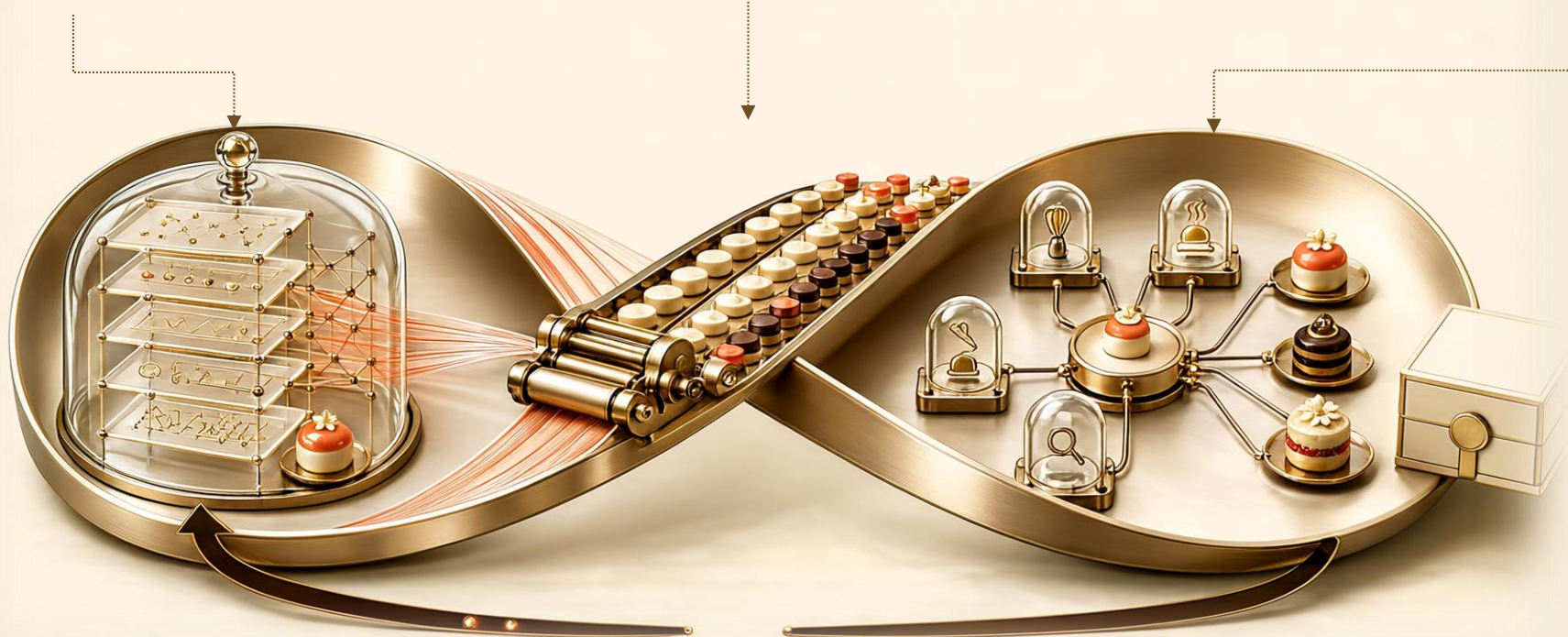
推动多模态智能体
在真实任务中持续进化

“1” 座Token工厂

推动Token工厂
由规模化智能生产迈向 Task 交付

“1” 套智能体管控系统

推动多智能体
协同进入更深层的个人和企业 workflows

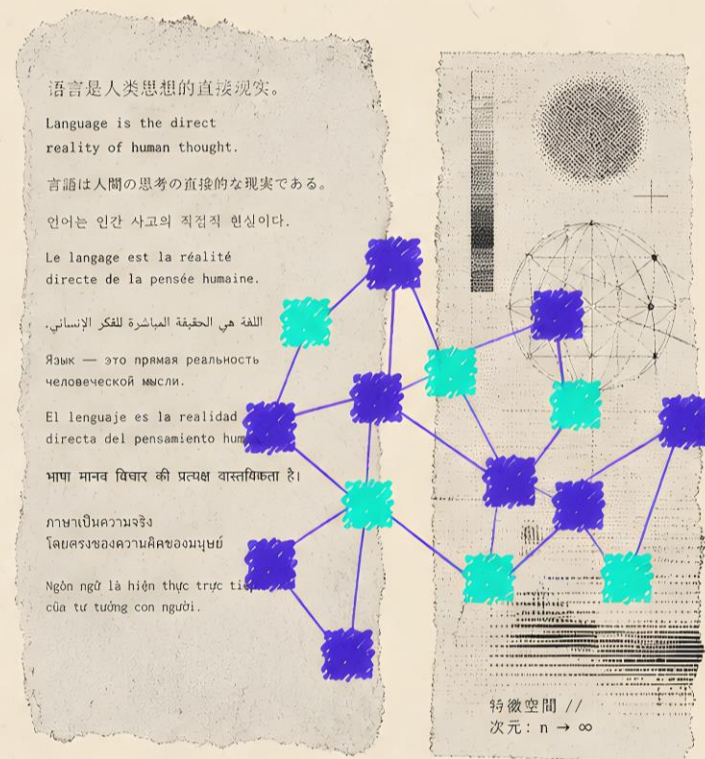


日日新
sensenova

THANKS!

坚持原创, 让AI引领人类进步

2026.08



AI FOR A BETTER TOMORROW