



商汤集团股份有限公司

2023全年业绩简报

坚持原创，让 AI 引领人类进步

2024.03

大装置与大模型的协同效应下，2023年生成式AI收入爆发式增长



大装置
运营算力达12,000P




大模型
“模型即服务”



生成式AI
收入达 **11.8 亿元**

200% 收入增长
35% 占集团总收入



年份	收入占比
2022	10%
2023	35%

34亿元

集团营业收入
(同比 -10.6%)

38亿元

回款
(同比 +48.5%)

4,500+ 人 (↓ 11.1%)

集团总雇员数

↓ 10.6%

营运成本

55亿元 (↓ 2.2%)

EBITDA (亏损)²
(同比收窄)

备注:

1. 截至2023年12月31日数据

2. 未经调整后

从算力基础设施到大模型落地能力的前瞻性布局



大装置

12,000P

大装置丰富的算力资源储备



日日新

每三个月升级一代

基础模型技术快速迭代

生成式AI应用

人工智能模型能力

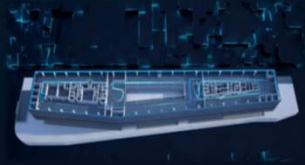
感知智能



智能内容生成



决策智能



智能内容增强



AI 1.0



大语言模型



文生图



数字人生成



3D小物品生成



3D大场景生成



企业 Copilot助手
Raccoon

AI 2.0

面向未来发展的业务分类，生成式AI成为商汤核心增长动能

原有四大业务板块

- 智慧商业
- 智慧城市
- 智慧生活
- 智能汽车

业务重组后 三大业务板块

	2022年 收入	2023年 收入	yoy %
生成式AI	3.9亿元	11.8亿元	↑ 200%
传统AI *	31.2亿元	18.4亿元	↓ 41%
智能汽车	2.9亿元	3.8亿元	↑ 31%

* 智慧城市业务 2022年占集团收入比 29%，2023年占集团收入比 <10%



备注：
1. 截至2023年12月31日数据

生成式AI商业化发展迅速

首批通过

国家备案的八家大模型之一

收入突破 10 亿元

数十个订单金额超千万元的客户

~120 倍

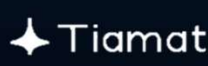
半年内C端调用量增长

国内第一

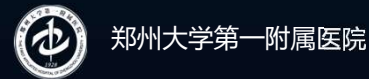
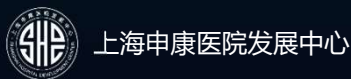
《2023年中国AI开发平台市场报告》
增长指数和创新指数排名¹



公有云



私有部署



公有云服务

私有化部署

模型定制服务

备注：
1. 沙利文联合头豹研究院发布

SenseCore大装置支持万亿参数大模型的全生命周期生产

12,000P (45,000 GPUs)

运营算力规模*

*: 截至24年3月26日

全国一体化智算网络



上海临港AIDC

万卡互联

超大集群
训练

90%+

超大规模集群
线性加速比

无中断

超一个月
连续训练

半小时内

诊断恢复
时长

300%

推理性价比
较一年前提升

自主可控、面向未来的国产异构能力



300+

DeepLink 提供的
标准算子接口

99.5%

可覆盖CUDA支持的
AI大模型计算需求

商汤“日日新”大模型体系发展主要里程碑



“日日新 4.0” 大模型性能全面超越GPT3.5，逼近GPT4



编程能力



推理能力



语言能力



工具调用能力



* 基于 Opencompass full 全集测试

中等开源模型性能世界领先

Search for your model (separate multiple queries with ";") and press ENTER...

Select columns to show

Average ARC HellaSwag MMLU TruthfulQA Winogrande
 GSM8K Type Architecture Precision Merged Hub License
 #Params (B) Hub Model sha

Hide models

Private or deleted Contains a merge/merge Flagged MoE

Model types

pretrained fine-tuned on domain-specific datasets chat models (RLHF, DPO, IFT, ...)
 base merges and moerges ?

Precision

float16 bfloat16 8bit 4bit GPTQ ?

Model sizes (in billions of parameters)

? ~1.5 ~3 ~7 ~13 ~35 ~60 70+

T	Model	Average	ARC	HellaSwag	MMLU	TruthfulQA	Winogrande	GSM8K
●	chargoddard/internlm2-7b-llama	66.94	60.49	80.99	63.16	54.25	79.87	62.85
●	internlm/internlm2-7b	66.68	58.02	81.24	65.24	48.73	83.82	63
●	google/gemma-7b	63.75	61.09	82.2	64.56	44.79	79.01	50.87
●	Qwen/Qwen1.5-7B	61.76	54.18	78.51	61.97	51.08	71.27	53.53
●	Deci/DeciLM-7B	61.55	59.39	82.51	59.76	40.33	79.95	47.38
?	Dans-DiscountModels/Dans-07YahooAnswers-7b	61.07	61.52	83.69	63.52	41.84	78.53	37.3
●	mistralai/Mistral-7B-v0.1	60.97	59.98	83.31	64.16	42.15	78.37	37.83
●	andysalerno/mistral-sft-v3	60.93	61.35	82.23	63.4	48.49	77.66	32.45
●	Qwen/Qwen-7B	59.19	51.37	78.47	59.84	47.79	72.69	44.96

终端小模型性能可与大型模型相媲美，推动边缘运行

端侧模型系列扩充

7B

1.8B

...



支持 **多芯片平台**



成熟AI能力累计搭载 **超过20亿** 部手机



UniAD和多模态大模型率先在 **主流车载芯片** 部署

端侧模型已应用于主流芯片厂商的旗舰芯片



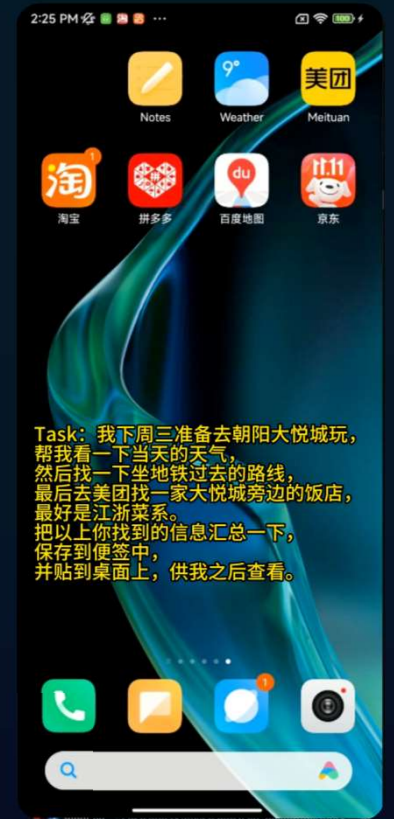
Qualcomm

MEDIATEK

推理速度在7B模型上实现了业界领先的

16 tokens/秒

大语言模型将成为未来智能设备的主要流量入口



Task: 我下周三准备去朝阳大悦城玩，帮我看一下当天的天气，然后找一下坐地铁过去的路线，最后去美团找一家大悦城旁边的饭店，最好是江浙菜系。把以上你找到的信息汇总一下，保存到便签中，并贴到桌面上，供我之后查看。

中国最优秀的代码解释器，展现稳定的工具调用能力

企业Copilot助手：国内领先的编程和数据分析模型



商汤办公小浣熊 Raccoon

Raccoon – Another Comprehensive CO-pilOt Navigator



领先的代码解释器

85.71%

1,000+数据分析测试集上精度
vs GPT-4 84.62%

No.1

SuperClue代码评测排名

- 支持多种文件类型
- 支持跨文件分析
- 支持多类分析、统计、可视化

强大的多模态开放任务处理能力

精准解读并理解图像内容，进行问题解答及互动，同时掌握网络热门话题及其背后深层含义

商量多模态对话

投入千卡训练国内最好、全球前三的文生图模型

2023-01

2023-04

2023-07

2024-01



秒画1.0

0.8B 参数量

输入提示词
生成成二次元图片



秒画2.0

1.4B 参数量

真实的光影效果
丰富的细节内容



秒画3.0

7B 参数量

专业摄影级的图片细节刻画
实现提示词 (Prompt) 补全功能



秒画4.0

10B 参数量

高效数据清理策略增量至10亿+图文对
达到10倍加速效果



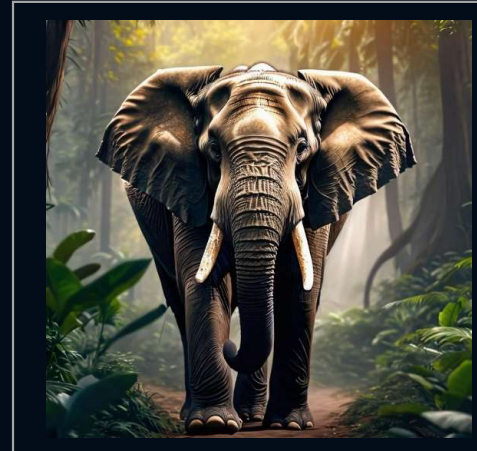
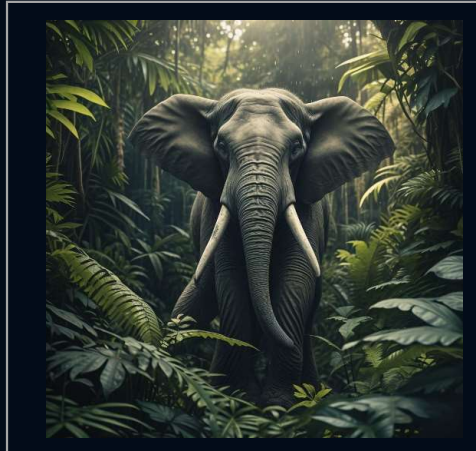
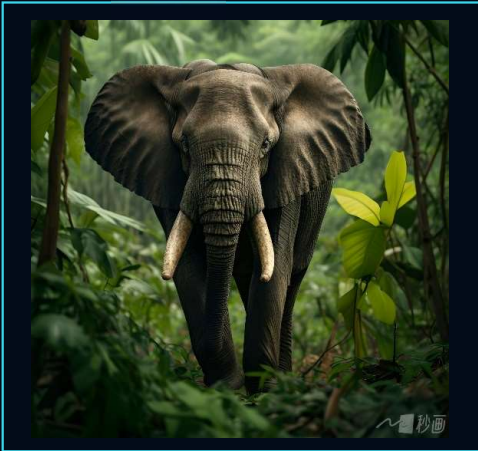
文生图评测案例

秒画

竞品一

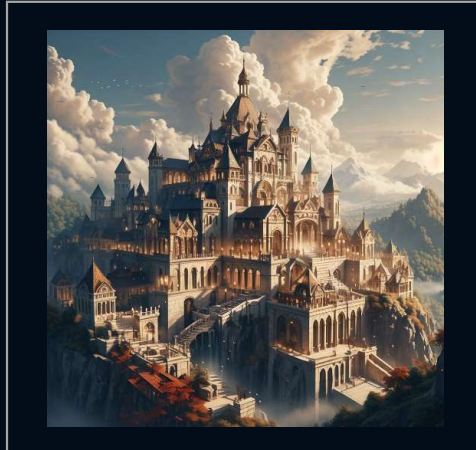
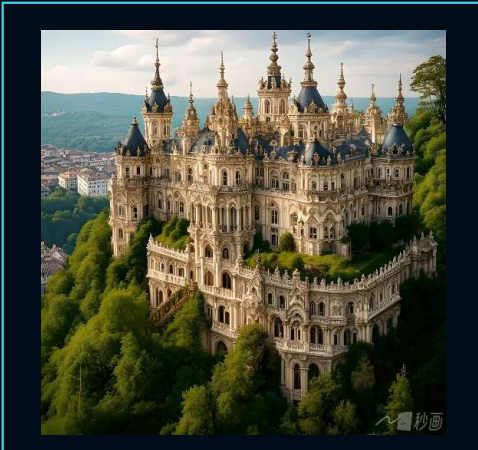
竞品二

写
实



Prompt:
智慧老象, 全身,
在茂密的丛林中,
平静温柔, 细致
逼真, 摄影写实,
国家地理风格

建
筑



Prompt:
欧式宫殿位于山
顶, 山下有许多
城镇, 金碧辉煌,
俯拍

高清流畅的文生视频大模型

基于秒画先进
的技术架构

故事性
多镜头

分钟级4K
视频画质





金融领域



医疗领域



企业
Copilot 助手



拟人交互



移动终端

金融和医疗：大模型能力渗透各应用场景，推动行业变革

推动银行券商的智能运营，和智能化客服体验



数字员工
数字客服

银行内部服务的“数字员工”，服务客户的“数字客服”

金融文档
检索问答

通过RAG检索增强生成技术，在上万量级文档知识库中进行精准搜索融合

基于海量医学知识训练的“商量·大医”大模型



“商量·大医”大模型在互联网问诊、导诊、用药咨询、健康问答等四项任务的自动化评测和执业医师人工评测中，表现优于GPT-4

No.2

MedBench 职业药剂师评测



上海申康医院发展中心



郑州大学第一附属医院



华西医院

拟人交互：C端的广泛应用触达

AI人设模型：让IP角色活化，与“梦中人”交流互动



角色创建与定制

支持快速、高质量IP角色定制

长对话记忆

角色可精准记忆超长对话内容

60%+

用户次日留存率

2,000+

单日最长交互轮数

18小时+

单日最长交互时长

传统AI 持续稳步发展，为生成式AI打下基础

集团战略转型下，传统AI业务经历一系列调整，收入占比下降



持续推动战略转型

提升收入质量

优化现金流

智慧城市、智慧商业、智能终端的坚实基础，推动生成式AI技术快速应用和推广

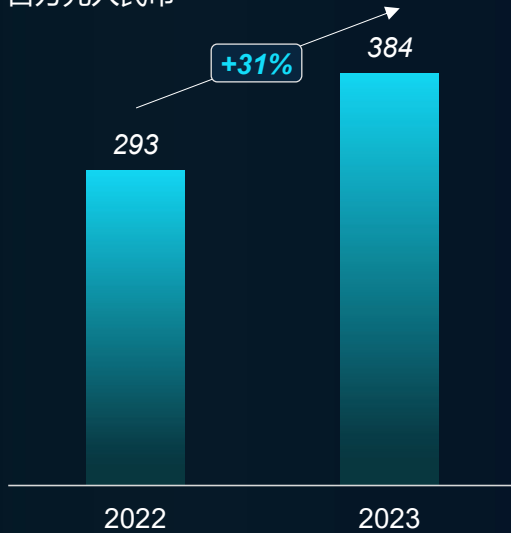


智能汽车：大模型时代汽车智能化的核心供应商，商业化量产持续领跑

总收入稳定增长，量产交付大幅跃升

商汤绝影年收入情况，2022-2023

单位：百万元人民币



195万台

累计交付数量

90+

累计量产车型

1,600万台

新增定点订单

41

新增定点车型

绝影新增
量产数量

49万台

129万台

↑ 163%

智能座舱
单车毛利

同比增长 30%

AGI打造智能汽车新范式

应用层



智能驾驶

L2++ 量产
BEV + 感知决策一体



智能座舱

大模型应用上车
打造智能座舱大脑



数据闭环

集模型训练、自动标注
仿真测试于一体



车路云协同

融合AGI技术的
车路云协同交通体系

模型层

UniAD 端到端感知决策一体模型

通用对话模型 | 代码生成模型 | 逻辑推理模型

多模态感知模型 | 文生图模型 | 人像模型 | 3D物体生成模型

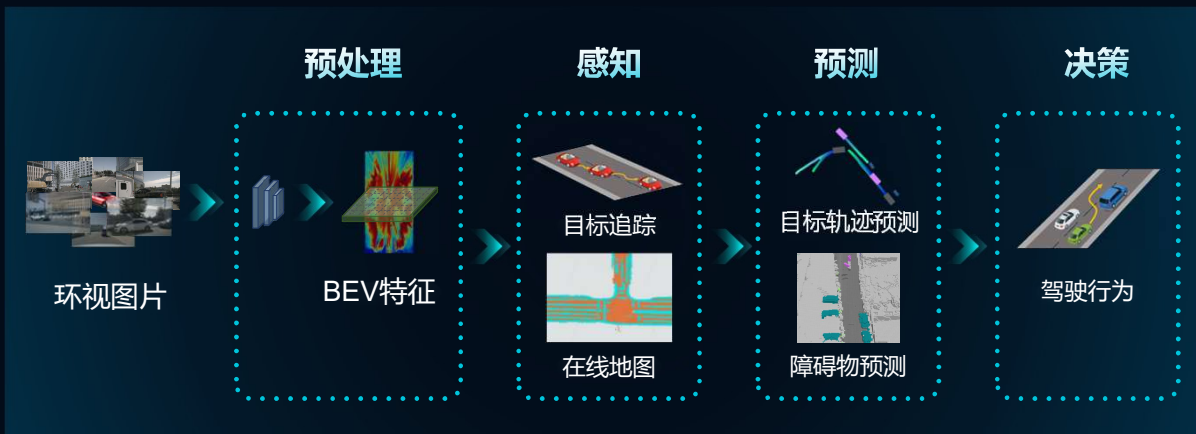
基础设施层

商汤大装置12,000P的丰沛算力集群

生成式AI
传统AI
智能汽车

智能驾驶：打造舱驾融合的全新驾乘体验

➢ UniAD：首个感知决策一体自动驾驶通用模型，荣获2023 CVPR最佳论文，已上车完成闭环



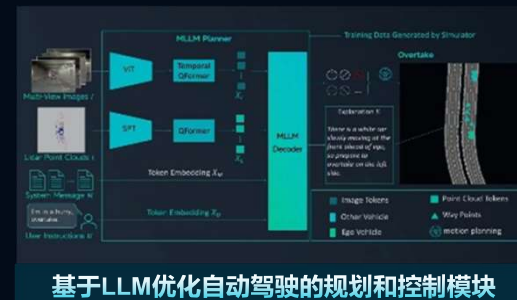
➢ DriveMLM方案

SOTA

基于CARLA*的闭环驾驶

25% ↓

每英里接管次数



➢ 全栈智驾软件算法+数据闭环云服务，广泛适配高中低端芯片，已在众头部厂商量产L2/L2+方案



*备注：CARLA是知名自动驾驶仿真测试平台

智能座舱：技术和市场的引领者，正基于大模型打造全新交互体验

▶ 绝影打造行业领先的车舱大模型产品体系，深化落地与国内外主流车厂的合作 ◀

四大座舱产品矩阵



DMS



OMS



近场感知



座舱大脑

车舱大模型集群



11项行业领先的座舱新功能



健康问诊



旅游规划



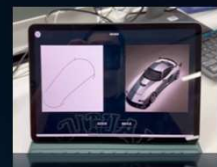
儿童伴读



AI说明书



角色扮演



隔空绘画



创意音绘



新闻智选

国内外主流汽车厂商洽谈大模型需求



.....

2024年战略目标



保持技术领先



推动业务增长



核心业务盈利

凭借大装置+大模型深度协同优势，商汤的生成式AI模型将持续增强，实现更高级别的应用

日日新大模型性能进一步提升，带来高计算密度、
高效能芯片的协同发展

大模型逐渐渗透至生产领域，推动生产力的变革

有效利用多模态数据，打造跨模态的世界模型

端侧大模型在AI手机等智能硬件上重塑用户交互



2024.04

商汤科技
技术交流日

敬请期待

生成式AI

传统AI

智能汽车

财务回顾

2023年全年业绩概览

营业收入

(人民币百万元)



年增长率

毛利

(人民币百万元)



毛利率

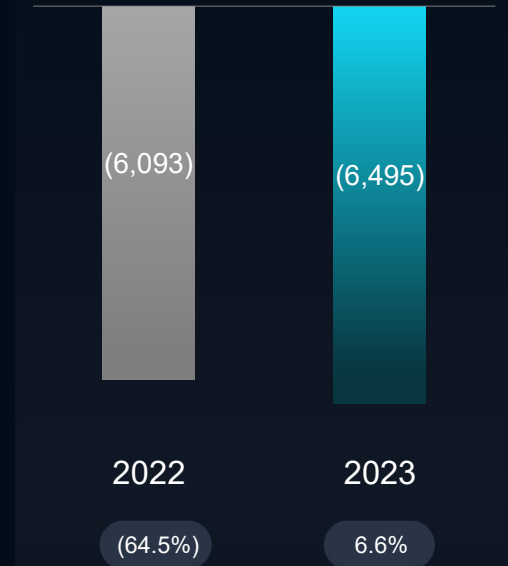
EBITDA

(人民币百万元)

年变动率⁽¹⁾

净亏损

(人民币百万元)

年变动率⁽¹⁾

备注:

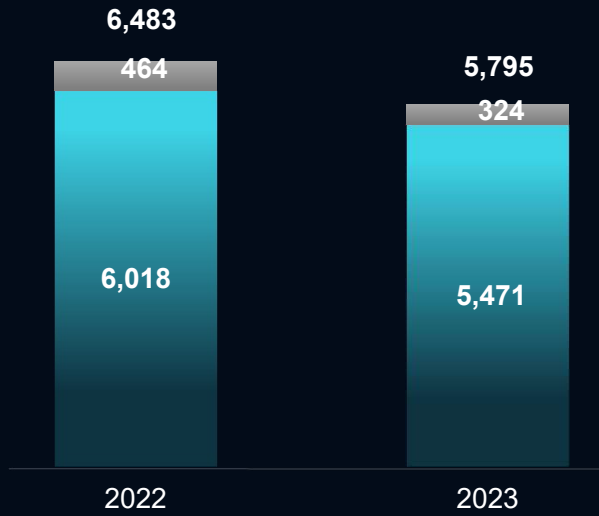
1. 负数代表亏损收窄

运营支出

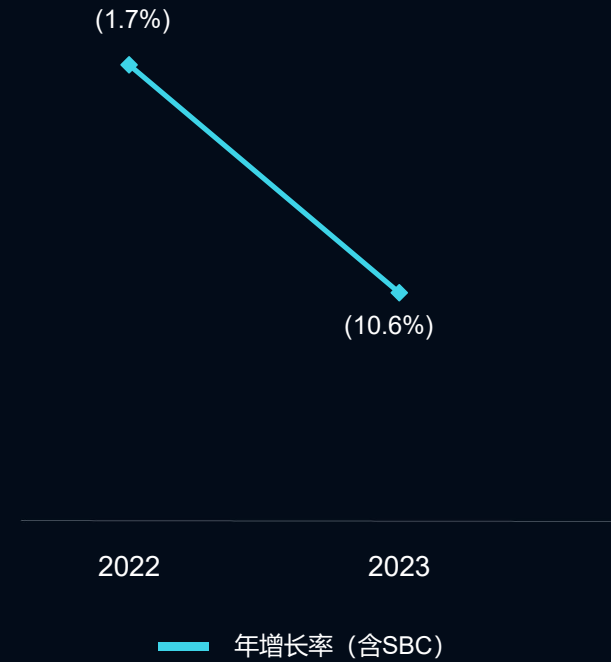
运营支出 (1)

(人民币百万元)

■ 费用 (扣除SBC)
■ SBC



运营支出年增长率



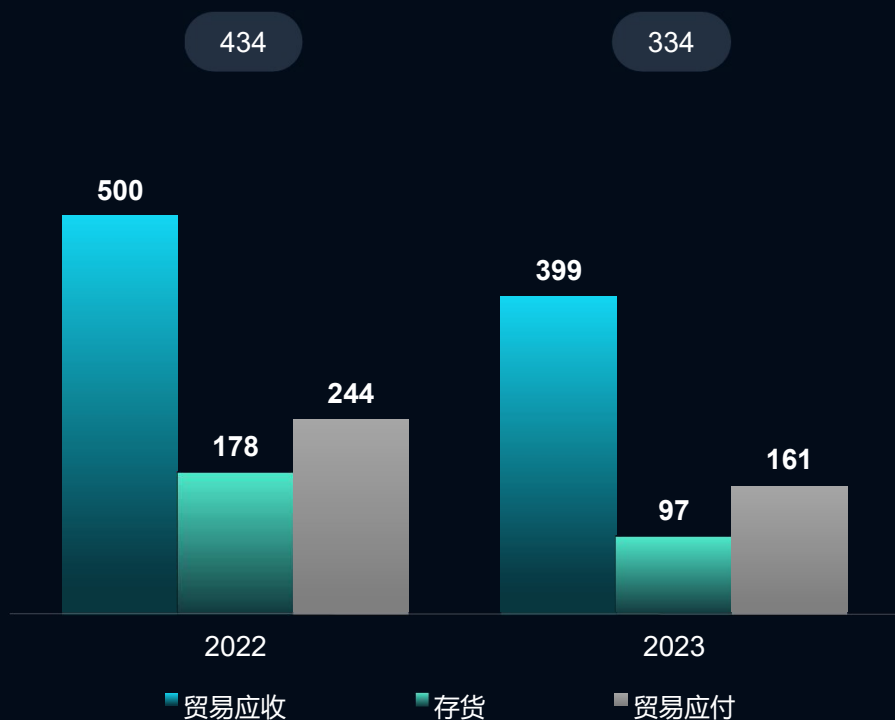
备注:

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 – 以期末值计算⁽¹⁾

现金循环周期 (天) Cash Conversion Cycle



减值拨备占贸易应收额比率 (资产负债表)

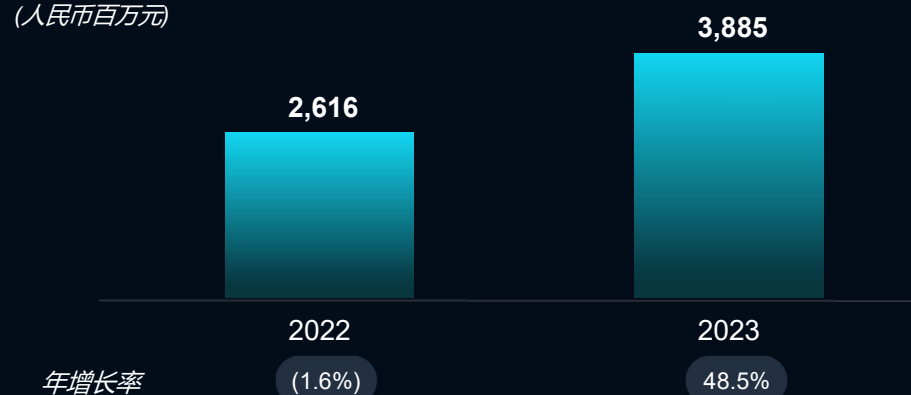
金融资产减值及合约资产亏损净额 (利润表)

(人民币百万元)



贸易应收款回款额

(人民币百万元)



备注:

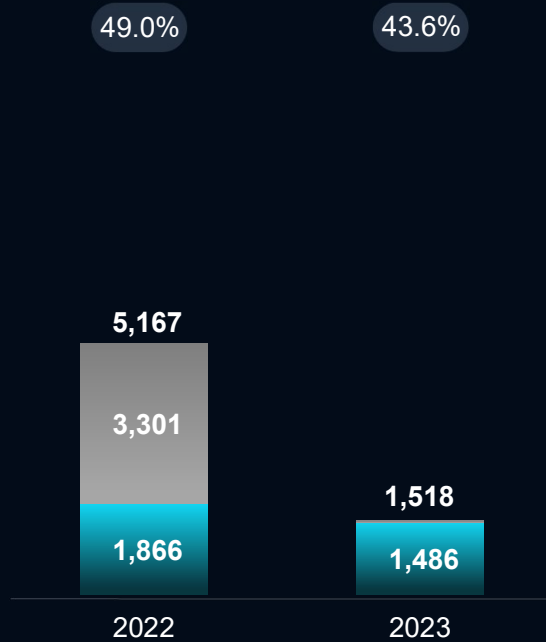
1. 以期末值计算的周转天数等同于期末值除以过去12个月的收入/成本, 并乘以365天 (期间长度为一)

资本支出与现金储备

资本支出⁽¹⁾

(人民币百万元)

■ 其它资本支出
■ 办公楼层相关资本支出



49.0%

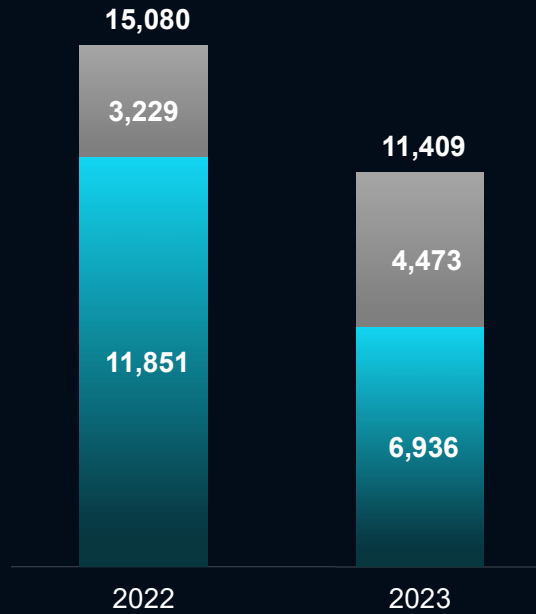
43.6%

其它资本支出占收入比率

总现金储备⁽²⁾

(人民币百万元)

■ 净现金
■ 总借款负债

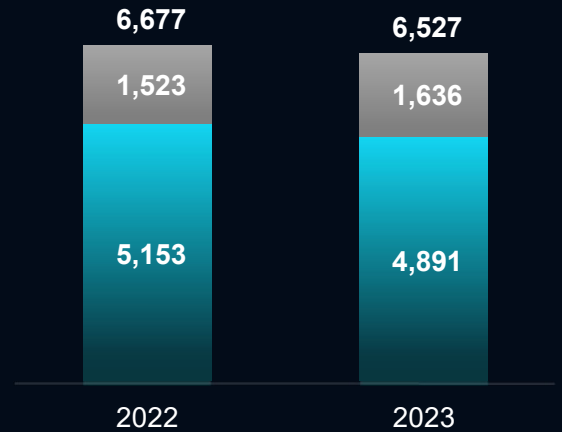


2023年年底未使用银行授信额：~83亿人民币

股权及债券投资余额⁽³⁾

(人民币百万元)

■ 股权投资余额
■ 债券投资余额



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)

Q & A





THANKS!

坚持原创，让 AI 引领人类进步

2024.03