

证券代码：300418

证券简称：昆仑万维

公告编号：2026-078

昆仑万维科技股份有限公司

2026 年半年度报告摘要

一、重要提示

本半年度报告摘要来自半年度报告全文，为全面了解本公司的经营成果、财务状况及未来发展规划，投资者应当到证监会指定媒体仔细阅读半年度报告全文。

所有董事均已出席了审议本报告的董事会会议。

非标准审计意见提示

☐适用 ☒不适用

董事会审议的报告期利润分配预案或公积金转增股本预案

☐适用 ☒不适用

公司计划不派发现金红利，不送红股，不以公积金转增股本。

董事会决议通过的本报告期优先股利润分配预案

☐适用 ☒不适用

二、公司基本情况

1、公司简介

股票简称	昆仑万维	股票代码	300418
股票上市交易所	深圳证券交易所		
联系人和联系方式	董事会秘书	证券事务代表	
姓名	吕杰	刘娟	
电话	010-65210366	010-65210366	
办公地址	北京市东城区西总布胡同 46 号明阳国际中心 B 座	北京市东城区西总布胡同 46 号明阳国际中心 B 座	
电子信箱	ir@kunlun-inc.com	ir@kunlun-inc.com	

2、主要会计数据和财务指标

公司是否需追溯调整或重述以前年度会计数据

☐是 ☒否

	本报告期	上年同期	本报告期比上年同期 增减
营业收入（元）	5,359,200,218.63	3,733,337,555.52	43.55%
归属于上市公司股东的净利润（元）	1,088,006,458.02	-855,545,912.56	227.17%
归属于上市公司股东的扣除非经常性损益的净利润（元）	-457,611,279.81	-859,208,113.15	46.74%
经营活动产生的现金流量净额（元）	-467,192,371.76	-560,457,758.99	16.64%
基本每股收益（元/股）	0.87	-0.69	226.09%
稀释每股收益（元/股）	0.84	-0.69	221.74%
加权平均净资产收益率	7.92%	-6.19%	14.11%
	本报告期末	上年度末	本报告期末比上年度 末增减
总资产（元）	21,915,605,374.28	20,233,944,014.56	8.31%
归属于上市公司股东的净资产（元）	14,800,936,998.86	13,185,191,363.68	12.25%

3、公司股东数量及持股情况

单位：股

报告期末普通股股东总数	166,380		报告期末表决权恢复的优先股股东总数（如有）	0	持有特别表决权股份的股东总数（如有）	0
前 10 名股东持股情况（不含通过转融通出借股份）						
股东名称	股东性质	持股比例	持股数量	持有有限售条件的股份数量	质押、标记或冻结情况	
					股份状态	数量
北京盈瑞世纪软件研发中心（有限合伙）	境内非国有法人	14.77%	189,744,943	0	不适用	0
周亚辉	境内自然人	11.29%	145,126,025	108,844,518	质押	200,000
李琼	境内自然人	7.50%	96,341,435	0	不适用	0
香港中央结算有限公司	境外法人	1.54%	19,782,117	0	不适用	0
中国工商银行股份有限公司－易方达中证人工智能主题交易型开放式指数证券投资基金	其他	0.44%	5,659,058	0	不适用	0
中国工商银行股份有限公司－易方达创业板交易型开放式指数证券投资基金	其他	0.42%	5,435,750	0	不适用	0
中国工商银行股份有限公司－广发中证传媒交易型开放	其他	0.33%	4,238,695	0	不适用	0

式指数 证券投资 基金						
江苏银 行股份 有限公 司一中 航机遇 领航混 合型发 起式证 券投资 基金	其他	0.31%	3,980,247	0	不适用	0
中国建 设银行 股份有 限公司 一华安 创业板 50 交易 型开放 式指数 证券投资 基金	其他	0.28%	3,563,645	0	不适用	0
广发证 券股份 有限公 司一华 宝创业 板人工 智能交 易型开 放式指 数证券 投资基 金	其他	0.28%	3,560,928	0	不适用	0
上述股东关联关系 或一致行动的说明		股东周亚辉和股东北京盈瑞世纪软件研发中心（有限合伙）为一致行动人。				
前 10 名普通股股东 参与融资融券业务 股东情况说明（如 有）		不适用				

持股 5%以上股东、前 10 名股东及前 10 名无限售流通股股东参与转融通业务出借股份情况

☐适用 ☒不适用

前 10 名股东及前 10 名无限售流通股股东因转融通出借/归还原因导致较上期发生变化

☐适用 ☒不适用

公司是否具有表决权差异安排

☐是 ☒否

4、控股股东或实际控制人变更情况

控股股东报告期内变更

☐适用 ☒不适用

公司报告期控股股东未发生变更。

实际控制人报告期内变更

☐适用 ☒不适用

公司报告期实际控制人未发生变更。

5、公司优先股股东总数及前 10 名优先股股东持股情况表

公司报告期无优先股股东持股情况。

6、在半年度报告批准报出日存续的债券情况

☐适用 ☒不适用

三、重要事项

报告期内经营业绩情况：

报告期内，公司主营业务保持强劲增长态势。2026 年上半年，公司实现营业收入 53.59 亿元，同比增长 43.55%。其中，海外收入 52.03 亿元，同比增长 51.21%。随着 AI 技术商业化进程加速，天工 AI 业务收入规模逐季抬升，成为公司重要增长动能。同期，由于投资业务回暖，公司实现归属于上市公司股东的净利润 10.88 亿元。

报告期内，公司继续围绕 AGI 与 AIGC 核心战略，推进技术研发、产品迭代与全球商业化落地。公司于 2026 年 3 月正式发布 2026 昆仑万维 AGI 战略，全新升级“4+3 战略”，以视频模型、音乐音频模型、游戏世界模型、基座文本与多模态模型四大 SOTA 模型为技术底座，持续支撑 AI 短剧、AI 音乐、AI 游戏三大 AI 原生平台经济体建设。公司先后孵化 AI 大算力 GPU 芯片企业艾捷科芯、机器人本体及大脑企业黎曼动力，支持两家企业在大模型和应用的产业链上下游独立发展，做大做强。

报告期内公司主要业务：

1、天工 AI

报告期内，公司继续围绕人工智能全产业链布局推进研发和产品创新，形成以 SOTA 大模型为技术底座、以全球化应用为牵引的 AI 业务体系。

1.1 大模型

报告期内，公司不断推动大模型技术能力的升级迭代，通过 SOTA 模型体系建设，构筑独具优势的技术护城河。

视频模型方面，2026 年 2 月，公司推出 SkyReels-V4，该模型发布时在国际权威第三方评测机构 Artificial Analysis 的文生视频 Text-to-Video (With Audio) 及图生视频 Image-to-Video (With Audio) 两大榜单中登顶全球榜首。SkyReels-V4 采用音画一体双流联合生成架构，并将生成、修复、编辑等能力整合进统一框架，支持多帧参考、多图片参考、运动参考、视频编辑等多种控制方式，可围绕人物、场景、动作和风格生成更稳定的连续画面，有助于解决 AI 短剧生成中角色走形、场景跳跃、镜头衔接不稳定等行业难点。公司已开放 SkyReels 相关 API 接口，为创作者、开发者和行业客户提供更便捷的顶级 AI 视频生成能力。

音乐音频模型方面，Mureka 系列模型持续迭代。2026 年 1 月，公司发布 Mureka V8，依托自研 MusiCoT 技术体系，在音乐结构理解、编曲完整度、人声表现和整体音质等方面实现升级，发布时即在 Artificial Analysis 的 Vocals (人声) 与 Instrumental (乐器) 两项测评中斩获榜首。

2026 年 3 月，公司进一步推出 Mureka V9，重点提升文本控制、混音质量、人声精准度、生成效率与内容丰富度，使 AI 音乐创作由“能够生成歌曲”进一步走向“能够稳定表达创作意图”。

2026 年 7 月，公司在世界人工智能大会期间发布 Mureka V9.5 与 03 音乐大模型。Mureka V9.5 沿着 MusiCoT 技术路线搭建更具真实感的音乐生成底座，形成“更自然，更真诚”的产品特质；在音乐人评测中，综合表现超越 Suno V5.5 和 Google Lyria 3 Pro 等主流头部竞品模型。03 建立在 V9.5 之上，将音乐思维链由一次性规划升级为持续推演。围绕全曲创作目标反复审视当前表达，回看偏差并修正后续创作决策，使得歌曲的结构更为完整、情绪演进连贯，段落之间的逻辑衔接更加合理。

游戏世界模型方面，公司致力于构建具备长期记忆、持续交互，与开放世界模拟能力的新一代交互式游戏世界模型。2026 年 3 月，公司发布 Matrix-Game 3.0，通过数据、模型和推理三层协同优化，基于 Unreal Engine 构建大规模交互数据引擎，设计带有 Memory 机制的 DiT 架构，并通过少步采样、并行生成、KV Cache、decoder 蒸馏等推理优化，推动游戏世界模型从“可运行的交互 Demo”走向具备工业级性能的操作系统。

2026 年 7 月，公司正式发布 Matrix-Game 3.5，提出将下一帧状态生成与动作生成联合训练框架，推动模型由单向预测下一帧，升级为对状态与动作的联合理解和生成。Matrix-Game 3.5 实现由游戏场景向真实场景扩展，支持多风格动态切换与指令控制，引入 NPC 交互能力，并完成长时记忆能力升级。模型通过 Patch Memory 与 Warped PRoPE 技术，强化长时序记忆、几何一致性建模和相机精准控制，5B

参数模型在 720p 分辨率下可实现单卡 20FPS 实时生成，具备 1 分钟记忆能力，为 AI 游戏、仿真、机器人和通用智能体等方向提供底层技术探索。

基座文本与多模态模型方面，2026 年 5 月，公司发布高性能 Agent 模型 SkyClaw-v1.0，并同步推出轻量化版本 SkyClaw-v1.0-lite。SkyClaw-v1.0 支持百万 token 上下文，面向真实智能体 workflow 重点优化复杂工具调用、多轮任务执行、代码生成、文件编辑、交互式应用构建和研究型数据分析等能力；模型可适配 OpenClaw、Hermes、Nanobot 等 Agent 环境，并兼容主流代码 Agent 框架，具备良好的通用性与生态兼容性。

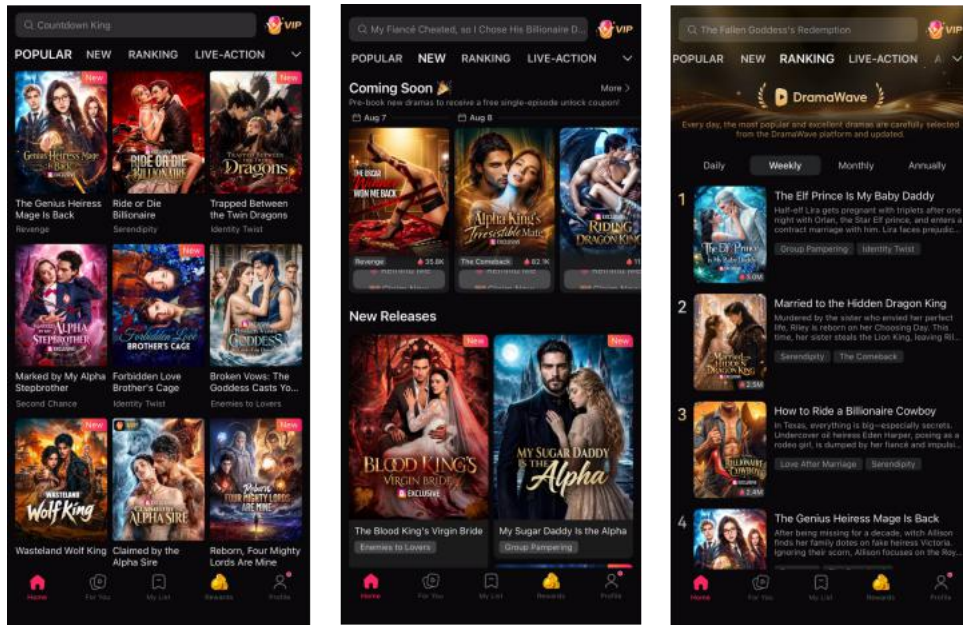
1.2 AI 短剧平台业务

报告期内，公司 AI 短剧平台业务在全球化运营、AI 内容供给扩充及 AI 原生生产体系建设方面持续深化，业务规模与商业化能力持续提升。据全球移动应用数据分析机构 SensorTower 数据显示，2026 年上半年，公司 AI 短剧平台合计 MAU（月活跃用户数）稳居海外短剧市场第一；FreeReels 位列出海短剧 APP 下载量第一名，亦是 2026 年上半年海外短剧赛道唯一一款预估下载量破两亿的产品。2026 年 6 月底，公司 AI 短剧平台业务单月流水超 6500 万美元。

公司现已完成前端分发平台（DramaWave 和 FreeReels）与后端工业化生产体系（SkyProduction）的协同布局。

DramaWave 作为海外付费 AI 短剧平台，深耕用户价值较高的发达国家市场，围绕当地用户偏好持续丰富精品 AI 短剧和真人短剧内容储备，通过题材本地化适配、剧情节奏优化及精细化运营，持续提升付费转化和用户留存，人均单日使用时长超 30 分钟。

依托公司领先的 AI 短剧生产能力，DramaWave 已完成 AI 原生化转型，生产效能持续提升，题材边界大幅扩充。2026 年 7 月，平台月度上新的剧集中，超 90% 内容为 AI 生成，重点打造悬疑冒险、都市情感、奇幻动作等高付费潜力题材，拍摄成本相较于真人剧大幅降低，成片品质达到影视级观看水准。同时，AI 工业化生产体系高效支撑多语种迭代，显著提升平台在海外各区域的内容适配与市场影响力。



DramaWave 界面图

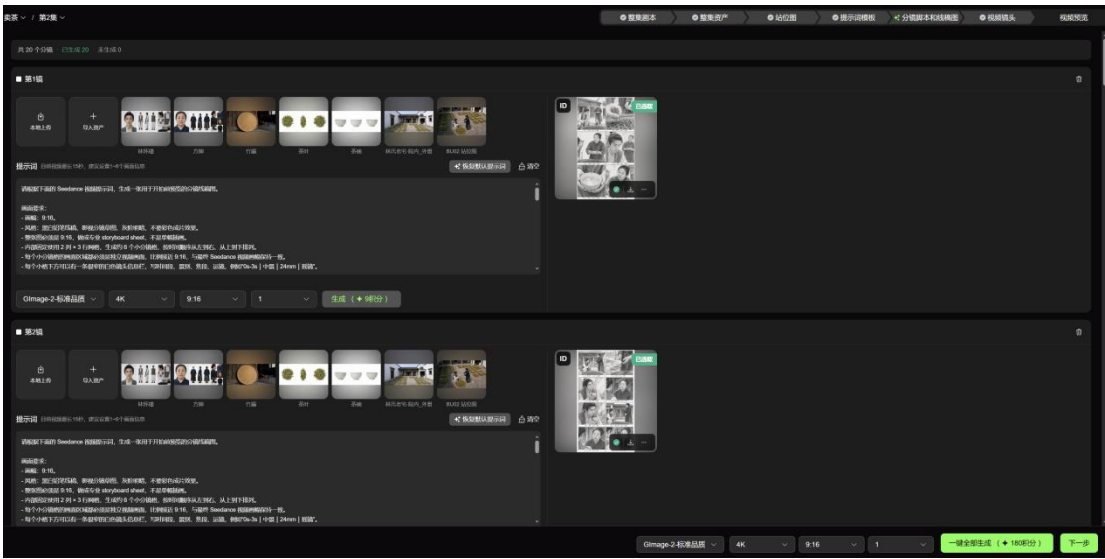


DramaWave 热门 AI 短剧海报

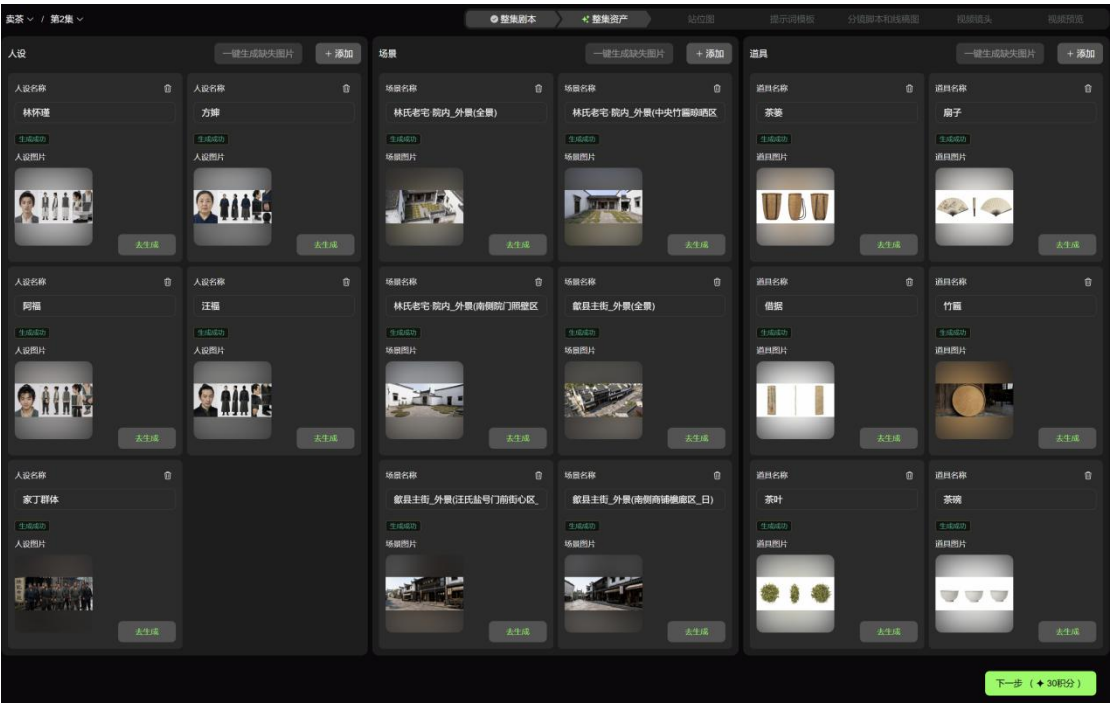
FreeReels 作为海外免费 AI 短剧平台，凭借丰富内容供给、多语种适配、激励机制和更低的使用门槛，在高增长市场快速拓展用户规模，并通过广告变现和内容分发效率提升形成商业化支撑，用户人均单日使用时长超 60 分钟。

报告期内，公司推出天工短剧工作台 SkyProduction，依托 Agent 技术整合 SkyReels、Seedance、可灵等全球顶级视频模型，构建“剧本直达成片”一站式 AI 短剧工业化生产体系。2026 年 7 月，SkyProduction 完成产品重大迭代，上线“Agent 智能分镜+无限画布”双轨创作模式，通过导演 Agent 实现剧本统筹、角色、场景、机位规划，支持多视图资产生成、站位图规划、影视级提示词模板及英文

剧本解析：无限画布配套 720° 全景图推理、多角度与光源编辑、3D 导演台等能力，强化创作者对场景构图、镜头调度的可视化控制，推动 AI 短剧向可控化、专业化生产演进。



示意图：Agent 智能分镜脚本&线稿图生成



示意图：Agent 一键生成资产

1.3 AI 音乐

报告期内，Mureka 保持高频迭代，围绕模型能力、全链路音乐创作、视觉配套产出等方向持续完善产品能力。

音乐创作层面，Mureka 升级 Easy Mode（简易模式），推出对话创作功能，用户通过自然语言描述想法，系统即可在交互中理解并打磨创作思路，一分钟内快速生成完整歌曲。平台新增 Soundtrack（视频图片配乐）功能，用户上传图片或视频，由 AI 自动生成高度契合的原创配乐，解决创作者痛点；同步上线 Remix（混音再创作）功能，支持一键切换曲风、歌词改写和风格切换等轻量化改编功能。

视觉层面，落地 Cover Art（动态封面）功能，支持生成歌曲静态封面与动态封面。同时推出生成音乐视频功能，用户可一键为歌曲生成 MV，结合 Character 功能生成自定义专属 AI 歌手角色，可实现口型同步、背景风格定制，实现从音频创作到音视频成品产出的完整闭环。

依托模型能力、创作工具与平台分发的一体化闭环，Mureka 打通了从“创意构思”到“歌词—人声—角色—歌曲—MV”的完整创作链路，全面释放 AIGC 生产力，推动 AI 时代音乐创作生态革新。

1.4 AI 游戏

报告期内，公司持续推进 AI 技术在游戏中内容生成、智能交互和虚拟世界构建中的应用。公司自研 AI 游戏《猫森小镇》定位为校园生活模拟类开放世界游戏，围绕青春校园主题构建沉浸式互动虚拟世界，并以 AI 驱动的 NPC 深度社交、高自由度玩法和低门槛创作作为核心特色。

报告期内，《猫森小镇》研发工作稳步推进，内测反馈良好。产品新增拍剧编辑器功能，为玩家提供低门槛内容创作和剧情表达环境，完善 UGC 生态支持体系；同时，产品持续升级 AI 能力，NPC 可理解所处环境并根据环境变化做出差异化行为反馈，进一步提升玩家的多模态交互体验。

1.5 AI SuperAgent

2026 年以来，天工超级智能体持续迭代升级，推动 AI 产品从被动问答工具升级为可自主执行任务、实现端到端交付的智能生产力系统。

2026 年 1 月，公司在天工超级智能体上推出 Skywork Video v1.0，将视频生成能力接入智能体 workflow，实现文档、表格、视频一体化创作。

2 月，公司推出适配 Windows 系统的天工 Skywork 桌面版（Skywork Desktop），支持本地化读取和处理 PPT、文档、图片、视频、表格等多格式内容，无需上传云端即可完成信息汇总、内容整理、智能生成和多任务并行处理。同月，基于 OpenClaw 架构重构的 SkyClaw 云端智能体上线，具备文件同步、跨设备多端监控、即时通讯工具交互、云端运行等功能。

3 月，Skywork 在 GitHub 与 Clawhub 上线 PPT、文档、表格、设计、搜索及 AI 音乐创作六大官方技能模块，进一步拓展 AI 生产力边界。

6 月，天工 3.1 版本上线，推出 Skywork Design 与 Dynamic Workflows 两项核心能力。其中，Skywork Design 以 AI 设计画布承载复杂设计项目，支持官网整站、App/Web 原型、资料转设计、参考图复刻等场景，便于用户在可视化空间内完成多页面 UI 项目迭代；Dynamic Workflows 面向大规模任务，由 Agent 先进行动态规划与任务自动拆解，同时再调度数十至上百个并行子 Agent，提升复杂项目的执行效率和交付稳定性。

为进一步增强天工超级智能体的组织协作能力，打通线下真实应用场景，2026 年 7 月，天工升级至 3.2 版本，并上线 Skywork Tags 功能。该功能可接入 Slack、飞书/Lark、钉钉、Discord、Telegram 等即时通讯工具，把 Agent 搬入团队工作环境，基于群聊上下文参与讨论、推进任务执行并沉淀团队协作信息。

同月，公司推出智能硬件产品 Skywork Note，面向会议访谈、客户沟通等场景完成语音采集、录音沉淀与智能整理；同时，依托天工超级智能体的多模态内容生成与处理能力，将非结构化音频内容转化为会议纪要、待办事项、文档草稿、洞察报告等结构化成果，支持后续检索、追溯和复用。Skywork 产品体系已从线上内容创作、办公提效延展至真实物理场景，构建起“线下感知-信息理解-内容生成-任务执行”的 AI 原生生产力平台。

（天工超级智能体 Skywork Super Agents 产品网址：<https://skywork.ai>）

1.6 AI 社交

AI 社交产品 Linky 致力于通过融合虚拟角色互动和 UGC 生态，以 PGC 精品角色与用户创作内容扩充平台内容供给，满足用户的深度陪伴、休闲娱乐、创意表达等多样化社交场景诉求。

报告期内，Linky 持续迭代升级，依托公司大模型技术，持续优化角色人设一致性、对话交互水准、剧情演绎效果及沉浸式互动体验。运营层面，Linky 不断优化投放结构、提升精细化运营能力，用户粘性和 ARPU（每用户平均收入）稳步提升，商业化运营成效显著。

1.7 黎曼动力

2026 年 7 月，子公司黎曼动力推出面向通用具身智能的世界动作模型 Riemann-1.0。该模型面向 Physical AI 场景，定位为新一代 Embodied World Action Model，基于多源具身交互数据训练，融合第一视角人类视频、UMI 示教及异构机器人轨迹，并引入全因果世界动作建模架构与三阶段渐进式具身预训练框架。Riemann-1.0 围绕长时序操作、跨场景泛化和真实部署等具身智能关键问题开展技术验证，在仿真环境与真实机器人任务中展现出对机器人操作、实时自适应控制和高频人机协作等场景的适配能力，实现游戏世界模型由数字内容生成，向物理世界交互的延伸。

2026 年 8 月，黎曼动力与具身数据供应商光轮智能、诺亦腾机器人达成深度合作。黎曼动力将联合光轮智能，推动 Riemann- 1.0 具身世界动作模型、Matrix- Game 3.5 游戏世界模型与光轮智能 EgoSuite 人类数据平台、RoboFinals 评测平台、RoboStack 部署反馈平台的适配及验证；同时携手诺亦腾机器人推进 100 万小时具身数据建设，围绕机器人训练数据、模型研发及真实场景验证开展深度协同，依托高精度动作数据、力反馈数据等多模态具身数据，构建“数据生产-模型训练-真实部署-反馈优化”的技术闭环。通过上述合作，共同夯实具身大脑模型的数据与评测基础设施，推动 Riemann 系列具身大脑模型迭代进化，提升机器人对真实世界的理解、推理与执行能力。

1.8 艾捷科芯

报告期内，AI 算力芯片企业艾捷科芯业务稳步推进，顺利完成新一轮增资扩股，引入包括国家人工智能产业投资基金、长鑫芯聚、启航恒鑫、合肥石溪、余姚石溪、嘉兴算捷、彩讯股份、高榕资本等产业资本与机构投资者，投后估值突破 40 亿元。

作为公司完善 AGI 与 AIGC 战略布局的重要载体，艾捷科芯将充分借助产业资本与市场化投资机构带来的资源赋能，加速芯片产品研发和商业化进程，构筑核心竞争壁垒，为后续登陆资本市场夯实基础。

2、StarMaker 社区

报告期内，StarMaker 聚焦 AI 技术与音乐社交场景的深度融合，持续推进产品升级与精细化运营，丰富平台内容生态，经营表现稳健。

产品迭代方面，StarMaker 围绕用户全链路歌唱体验持续推进产品迭代。通过 AI 音域识别技术实现精准歌曲推荐，结合 AI 智能领唱、AI 美音、罗马音译等功能，有效降低用户演唱门槛。同时，上线 AI 生成个性化声音形象、母亲节祝福视频及宠物跳舞视频等创新玩法，丰富产品可玩性。

运营推广方面，平台兼顾本地化深耕与全球化运营，依托 AI 创新活动形式，有效提升平台热度与用户活跃度。针对核心市场，平台落地斋月主题公益活动，以线上祈福、贺卡互动等形式联动线上线下社区，获得当地用户广泛认可。结合节日与特色主题打造 AI 专属内容活动，同时，精准把握品牌与全球热点，5 月底上线十周年庆典活动，6 月借势世界杯赛事在全球推出歌曲大赛、经典合唱等系列活动，充分调动全球用户的参与热情。

未来，StarMaker 将继续以 AI 创新为驱动，携手全球用户打造 AI 时代高品质音乐社交娱乐社区。

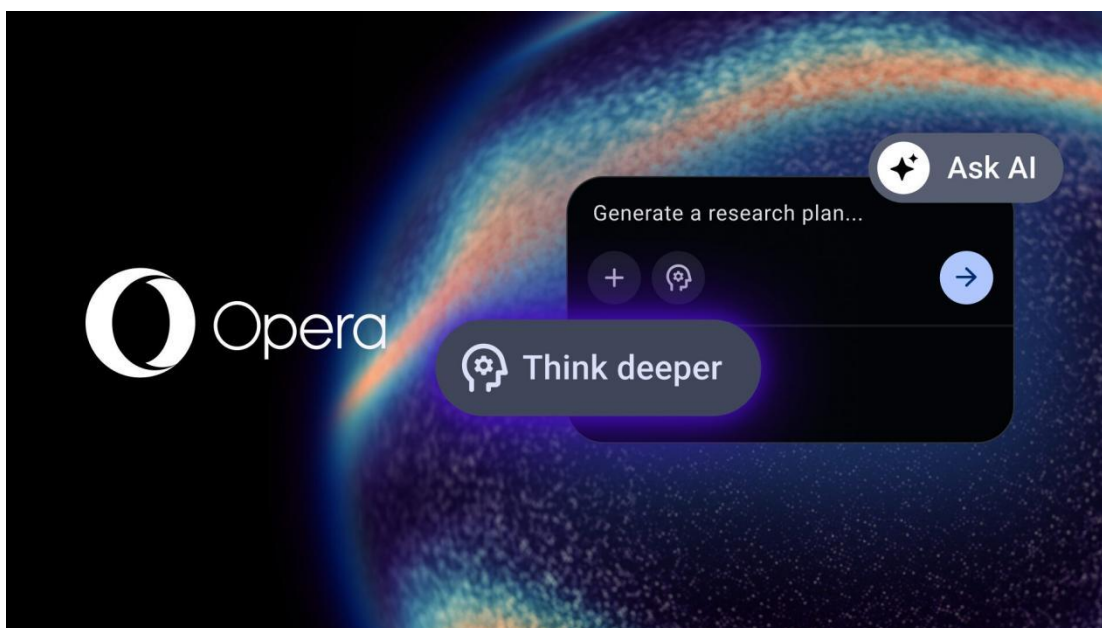
3、海外 AI Agent 浏览器平台 Opera

本报告期，Opera 营业收入与盈利水平均保持稳健增长。2026 年上半年，Opera 实现营业收入 3.54 亿美元，同比增长 24%；经调整后 EBITDA 0.84 亿美元，同比增长 31%。2026 年第二季度，全球月活跃用户达 2.88 亿，其中，Opera GX 游戏浏览器月活跃用户达 0.37 亿，增长势头延续。

报告期内，Opera 持续围绕 AI 原生流量入口进行产品升级，积极探索浏览器在 AI Agent 时代的全新商业价值。

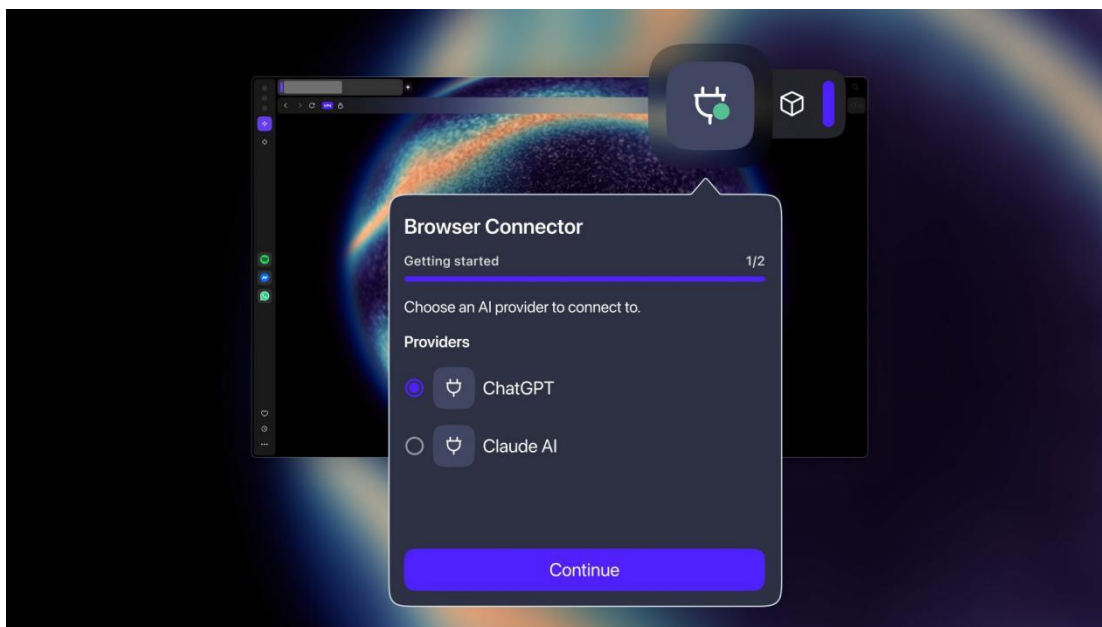
2026 年 1 月，Opera AI 功能全面落地 Opera One、Opera GX 及 Opera Air 三大产品线，提供 YouTube 内容解析、语音与文字转换、多类型文件智能分析、截图快速分享、AI 壁纸生成等功能。同期发布 Opera One R3 版本，围绕标签页、AI 能力和模块化设计三大关键领域进行升级，进一步优化用户浏览与协作体验。

3 月，Opera 安卓端上线 Think Deeper（深度思考）能力，通过强化推理与网络检索机制提升 AI 回答质量和可靠性，将 AI 能力进一步延伸至移动端。同月，Opera GX 正式登陆 Linux，进一步拓展跨平台场景下的覆盖能力。



Opera 的 Think Deeper（深度思考）功能示意图

4 月，Opera 推出 Browser Connector（浏览器连接器）功能，通过 MCP 将 Claude、ChatGPT 等外部 AI 工具连接至浏览器端，用户可在浏览过程中完成问答、创作、分析和任务处理，降低跨应用切换成本。



Opera 的 Browser Connector（浏览器连接器）功能示意图

5 月，Opera 推出开源命令行工具 opera-browser-cli，支持本地 AI 代理通过命令行操控浏览器，可实现网页测试、数据抓取、Web 应用生成、在线表单处理等操作，推动浏览器由信息获取工具向任务执行入口升级。

6 月，Opera GX 完成产品更新，在 AI 辅助、多标签页上下文理解、侧边栏集成、广告拦截、性能与稳定性等方向优化浏览体验。

2026 年第二季度，Opera Android 与 iOS 移动浏览器合计 MAU 在英国同比增长 66%，在美国同比增长 40%，体现其特色浏览器功能和内置 AI 体验在高价值成熟市场的用户吸引力持续提升。

4、其他

报告期内，公司参股投资企业整体运营稳健，发展态势良好，多家企业在业务拓展、技术研发等领域取得阶段性突破。其中，PingCAP Ltd.（平凯星辰）企业级 Agent 平台平凯 Loop 正式公测，迭代 TiDB 内核增强湖仓一体、向量检索能力，完成多项生态互认证，中标多个金融行业标杆项目。沙砾生物 TIL 细胞治疗管线 GT101 注射液推进 II 期临床，GT201 注射液获得中国 CDE 受理、美国 FDA 快速通道认定，并披露早期积极临床数据；体内 CAR-T 业务分拆主体完成融资，全球研发进程加速。

报告期内公司所处行业情况：

2026 年以来，全球人工智能产业从模型能力突破迈向规模化应用和商业化合验证阶段，行业增长由算力基础设施、模型与平台服务、企业软件嵌入式 AI 以及 C 端高频应用共同驱动。根据沙利文预测，2026 年全球大模型平台及大模型原生应用市场规模预达到 465 亿美元，较 2025 年同比增长约 91.36%，AI 基础设施投入同步扩张。

基础模型赛道的迭代逻辑发生明显变化，行业竞争从单纯通用问答能力比拼，转向长上下文、智能体编程、端侧设备操作、多工具协同和智能体自主执行等真实场景下的生产力维度。2026 年上半年，全球头部厂商保持高频版本更新，市场评判标准从单一性能跑分升级为“模型能力+工具生态+完整 workflow”的综合体系能力。

海外厂商持续推进基座模型升级。OpenAI 持续迭代 GPT 系列模型，重点扩容上下文窗口、强化端侧设备操控与智能体编码能力；Anthropic 迭代 Claude 全系模型，持续加码代码调试、长周期智能体任务规划；Google 更新多版本 Gemini 系列通用模型，补强复杂逻辑推理、智能体任务与低成本推理能力。国内厂商同步加速迭代，竞争维度从单纯推理跑分延伸至工程落地、办公自动化、桌面与浏览器原生操作和垂直行业解决方案。国产模型同步加速迭代，昆仑万维、月之暗面、字节跳动、阿里通义千问

等厂商围绕 Agent、多模态和长上下文能力密集更新；DeepSeek、腾讯混元、智谱 AI、MiniMax 等厂商继续强化代码、工具调用、长程任务规划和多模态生成，国产模型迭代逐步进入场景交付和成本效率并重的新阶段。

多模态技术快速成熟，图像、视频、音频、音乐、3D 与交互式世界生成能力持续演进，内容生产正从单点素材生成走向一体化 workflow 生产。AI 真人剧、AI 漫剧已成为多模态技术率先实现商业化落地的核心场景。国内厂商在视频模型实现突破，字节跳动更新 Seedance 2.5 长时序视频模型、阿里上线通义万相 2.7-Video 全模态视频生成体系。昆仑万维推出 SkyReels V4 视频模型、Mureka V9.5 与 O3 音乐模型及 Matrix-Game 3.5 游戏世界模型等多个 SOTA 级垂类模型。

AI 应用迎来重大变化，由简单对话助手升级为可自主拆解任务、联动多工具、完成长周期作业的智能执行主体。产品聚焦办公、浏览器、代码、内容生产等场景落地，MCP 交互协议加速普及，多智能体协同方案逐步商用。行业竞争重心已从基础交互体验，转向可闭环创造业务价值的完整智能 workflow，兼具基座、多模态生成与合规管控全栈能力的企业更具长期竞争优势。

AI 算力与基础设施呈现多元化发展格局，GPU、云厂商自研 ASIC、国产专用算力芯片同步扩产，英伟达、AMD 更新新一代算力硬件，AWS、谷歌、Meta 迭代自研 Trainium、TPU、MTIA 系列芯片并对外商业化交付，国内昇腾、寒武纪等国产芯片批量商用，加速算力自主替代。全球云厂商持续加大数据中心与智算集群资本投入，同步优化算力调度与轻量化租用服务。行业竞争重心由硬件采购转向算力、芯片、调度平台一体化全栈服务能力，算力成本、供给稳定性与软硬件生态成为产业核心比拼维度。