

证券代码：300418

证券简称：昆仑万维

公告编号：2024-059

昆仑万维科技股份有限公司

2024 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）	2,501,811,162.32	2,425,229,544.83	3.16%
归属于上市公司股东的净利润（元）	-389,136,070.45	358,189,359.72	-208.64%
归属于上市公司股东的扣除非经常性损益的净利润（元）	-407,394,274.63	311,942,472.00	-230.60%
经营活动产生的现金流量净额（元）	189,088,081.19	533,113,423.95	-64.53%
基本每股收益（元/股）	-0.32	0.30	-206.67%

	本报告期	上年同期	本报告期比上年同期增减
稀释每股收益（元/股）	-0.31	0.29	-206.90%
加权平均净资产收益率	-2.58%	2.74%	-5.32%
	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）	21,039,773,543.60	21,484,962,358.21	-2.07%
归属于上市公司股东的净资产（元）	15,093,299,461.48	15,174,349,421.64	-0.53%

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	138,696	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
新余盈瑞世纪软件研发中心（有限合伙）	境内非国有法人	15.41%	189,744,943	0	不适用	0
周亚辉	境内自然人	11.76%	144,844,325	0	质押	200,000
李琼	境内自然人	7.82%	96,341,435	0	不适用	0
中国民生银行股份有限公司－华夏中证动漫游戏交易型开放式指数证券投资基金	其他	1.92%	23,633,346	0	不适用	0
香港中央结算有限公司	境外法人	1.78%	21,908,784	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	1.31%	16,116,955	0	不适用	0
#毛金明	境内自然人	1.25%	15,389,803	0	不适用	0
中国工商银行股份有限公司－华泰柏瑞沪深 300 交易型开放式指数证券投资基金	其他	0.82%	10,120,700	0	不适用	0
#毛英俊	境内自然人	0.74%	9,161,000	0	不适用	0
上海浦东发展银行股份有限公司－国泰中证动漫游戏交易型开放式指数证券投资基金	其他	0.59%	7,242,983	0	不适用	0
上述股东关联关系或一致行动的说明	股东周亚辉及新余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）	股东毛金明通过国泰君安证券股份有限公司客户信用交易担保证券账户持有 15,389,803 股，通过普通证券账户持有 0 股，合计持有 15,389,803 股。 股东毛英俊通过广发证券股份有限公司客户信用交易担保证券账户持有 9,161,000 股，通过普通证券账户持有 0 股，合计持有 9,161,000 股。					

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

单位：股

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								
股东名称 (全称)	期初普通账户、信用账户持股		期初转融通出借股份且尚未归还		期末普通账户、信用账户持股		期末转融通出借股份且尚未归还	
	数量合计	占总股本的比例	数量合计	占总股本的比例	数量合计	占总股本的比例	数量合计	占总股本的比例
上海浦东发展银行股份有限公司—国泰中证动漫游戏交易型开放式指数证券投资基金	5,292,333	0.44%	0	0.00%	7,242,983	0.59%	1,284,700	0.10%
中国民生银行股份有限公司—华夏中证动漫游戏交易型开放式指数证券投资基金	16,702,167	1.37%	1,441,000	0.12%	23,633,346	1.92%	406,900	0.03%
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	15,965,946	1.31%	1,048,000	0.09%	16,116,955	1.31%	812,400	0.07%

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

单位：股

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情况					
股东名称 (全称)	本报告期新增/退出	期末转融通出借股份且尚未归还数量		期末股东普通账户、信用账户持股及转融通出借股份且尚未归还的股份数量	
		数量合计	占总股本的比例	数量合计	占总股本的比例
中国建设银行股份有限公司—华安创业板 50 交易型开放式指数证券投资基金	退出	0	0.00%	6,710,695	0.54%
中国农业银行股份有限公司—中证 500 交易型开放式指数证券投资基金	退出	0	0.00%	0	0.00%
上海浦东发展银行股份有限公司—国泰中证动漫游戏交易型开放式指数证券投资基金	新增	1,284,700	0.10%	8,527,683	0.69%

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：

本报告期，公司实现营业总收入 25.02 亿元，同比增长 3.16%。整体毛利率达 78.20%，继续保持在较高水平。公司出海步伐再加速，海外业务收入规模达 22.05 亿元，同比增长 9.02%，占总收入比重达 88.16%，公司以 AI 技术助力产品成功出海，赢得全球市场广泛认可。本报告期，公司继续坚定践行“AI in AGI 与 AIGC”发展战略，持续加大研发投入，上半年公司研发费用为 7.57 亿元，同比增长 115.72%。

本报告期，受资本市场波动下行影响，公司投资业务短期承压，导致归属于上市公司股东净利润转负，在不考虑投资业务影响下，公司亏损幅度将大幅收窄。目前公司 AI 应用加速落地，商业化进程顺利开展，我们相信 AI 业务将成为公司新的增长引擎，为股东创造长期价值。

AI 应用的渗透是一个渐进的过程，要大规模产生收入及利润，可能还需要 1-2 年的过程，但这一时间点必定到来。当前对于公司而言，我们需要做到的是一直在第一梯队，确保在技术创新和市场应用上保持领先地位。目前公司 AI 应用加速落地，商业化进程顺利开展，我们相信 AI 业务将成为公司新的

增长引擎，为股东创造长期价值。

报告期内公司主要业务：

1、AGI 与 AIGC 业务板块

昆仑万维是中国领先的人工智能科技公司，现已完成“算力基础设施—大模型算法—AI 应用”全产业链布局，并构建起由 AI 大模型、AI 搜索、AI 游戏、AI 音乐、AI 社交、AI 短剧组成的多元 AI 业务矩阵。

本报告期，公司在基座大模型迭代、AI 应用发展、商业化变现，三大维度均取得显著进展。

基座大模型侧，公司坚持技术创新，对文本大模型、多模态图文大模型、音乐大模型、3D 大模型、视频大模型进行了全面升级，在性能上均达到行业领先水平，为用户带来前所未有的智能体验。

AI 应用侧，公司多款产品在全球范围内成功落地。报告期内，天工 AI 智能助手日活跃用户（DAU）突破 100 万；截至报告期末，AI 社交产品 Linky 日活跃用户（DAU）突破 50 万；AI 游戏产品《Club Koala》开启 Alpha 测试；AI 音乐方向，2024 年 8 月，公司推出了全球首款 AI 流媒体音乐平台 Melodio 和 AI 音乐创作平台 Mureka；AI 短剧平台 SkyReels 于 2024 年 8 月 19 日正式发布。

商业化侧，天工大模型通过与众多企业如电信运营商、主流手机厂商、金融机构、头部短视频平台、直播平台、教育公司、物业公司等达成商务合作，实现商业化收入。公司海内外 C 端 AI 应用也在加速商业化进程，反哺业务发展。

此外，公司旗下艾捷科芯在研的 AI 大算力芯片项目进展顺利。艾捷科芯致力于推动算力芯片技术的突破，为人工智能、数据中心和高性能计算等领域提供强有力的硬件支持。我们将继续加强研发力度，为实现芯片技术的自主可控贡献力量。

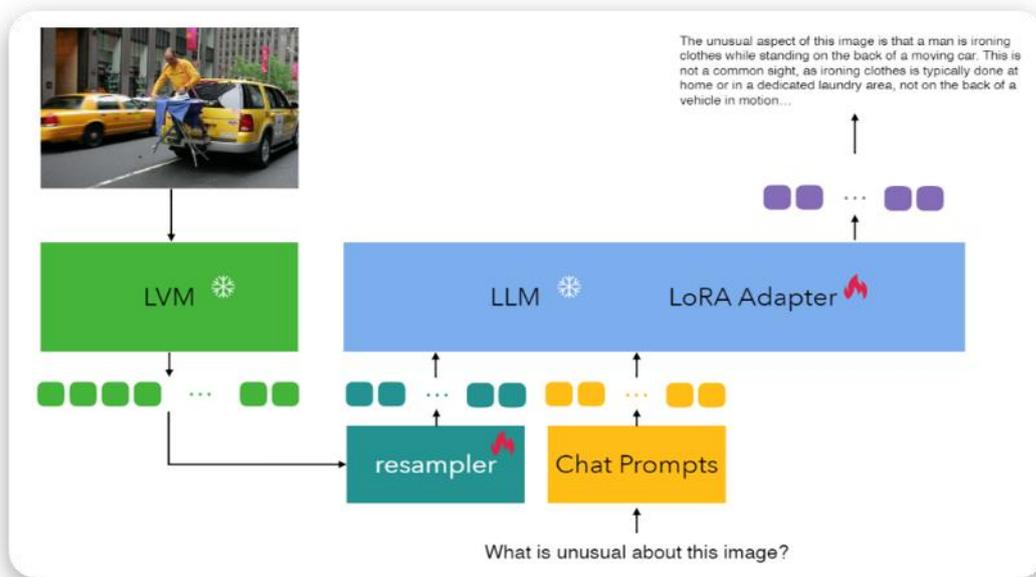
1.1 AI 大模型

公司 AI 大模型布局覆盖天工基座大模型及四大关键垂类模型。具体来说包括语言大模型、多模态图文大模型、音乐大模型、3D 大模型及视频大模型。

报告期内，公司持续迭代升级天工大模型。2024 年 4 月 17 日，正式发布采用业内顶尖 MoE 专家混合模型架构，并拥有 4000 亿参数的大语言模型“天工 3.0”，相较传统大模型，其复杂任务处理能力更强、响应速度更快、训练及推理效率更高、可扩展性也更强。通过 MoE 架构，模型可以将一个复杂的问题分解成多个更小、更易于管理的子问题，并由不同的专家网络分别处理，显著提升了模型训练与推理的性能和效率。相较于“天工 2.0”，新模型各方面性能又一次实现飞跃，“天工 3.0”不仅在语义理解、逻辑推理、通用性、泛化性和学习能力等方面实现了显著的性能提升，其技术知识能力亦实现

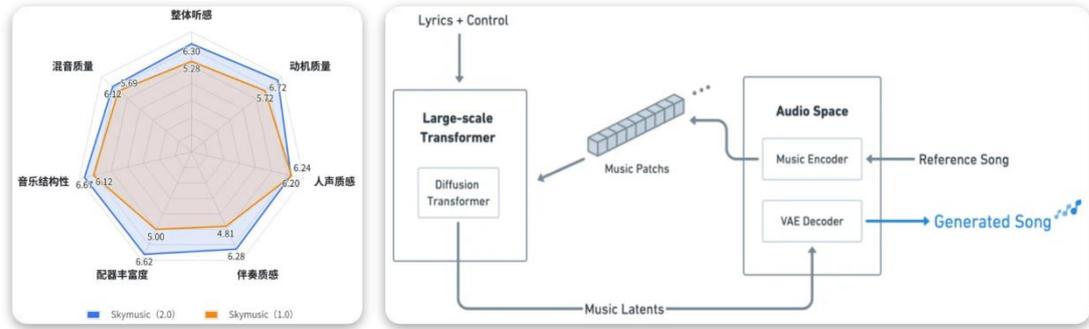
20%的提升，数学、推理、代码和文创能力更是提升超过 30%。

多模态图文大模型方面，公司通过多流结构中多视觉编码器的融合方案，将文本信息与图像特性进行融合，得到的文本图像特征可以被应用到视觉问答、图像描述生成等不同的多模态任务中，从而实现图文内容的理解和推理。在训练过程中，利用不同的数据和策略来训练图文模型的各个模块，以快速高效实现预训练图像模型和语言模型的对齐和升级。该模型以其卓越的图像识别、处理和生成能力，结合深度文本理解，为用户提供了高效、个性化的使用体验。其不仅能精准执行任务，还能创造性地满足用户需求，适用于广告、艺术、教育等多个领域。



昆仑万维多模态大模型 Skywork-MM 架构

音乐大模型方面，公司推出了国内首款公开可用的 AI 音乐 SOTA 模型——“天工 SkyMusic”。该模型采用先进的 Large-scale Transformer 和 Diffusion Transformer 技术，能够生成高质量、具有自然人声和丰富情感表达的双声道立体声音乐，还能够根据用户输入的歌词风格定制歌曲，支持多样的音乐风格，并具备学习颤音、歌剧等歌唱技巧的能力，实现情感丰沛的音乐表达。“天工 SkyMusic”有效降低用户创作门槛，一经推出便收获广泛好评。目前，公司天工 SkyMusic 已迭代至 2.0 版本，相较于上一代模型，SkyMusic 2.0 能够支持长达 6 分钟的 44,100Hz 采样率双声道立体声歌曲，并支持 500 字以上的歌词输入，确保了音乐作品的旋律流畅性和结构完整性。同时，SkyMusic 2.0 在音乐旋律质量、伴奏质量、配器丰富度、音乐结构性和混音质量等维度均有大幅提升，在人声音质、人声自然度和发声可懂度上维持业内领先的 SOTA 水平。



天工 SkyMusic 代际对比及架构示意图

3D 大模型方面，公司技术团队设计了基于 DiT 架构的混合生成方案，充分利用了原生 3D 生成和 2D 升维生成各自的优势。具体来讲，针对几何生成，设计了基于 DiT 架构的端到端生成模型，实现文生 3D 几何和图生 3D 几何；另一方面，针对贴图生成，区别于市面常见的多视角生成算法，设计了渐进式的多视角生成算法，在几何一致性上有显著提升，然后基于这些视角图片即可生成高质量贴图。经过多次版本迭代，公司 3D 生成大模型已达到行业领先水平，并已在多个内部 AI 游戏项目中得到应用。

视频大模型方面，公司采用 3DCausalVAE 加 3D-DiT 的领先架构，使得模型能够高效地解析视频数据中的多维度信息，包括场景、动作、音频等关键要素，从而实现更为精准的视频内容识别与推理。与此同时，该模型还具备出色的生成和编辑功能。模型能够在潜在空间中灵活操控视频内容，实现高质量的视频生成、风格迁移和内容插帧等任务。这些功能为视频创作者和编辑人员提供了强大的创作工具和编辑手段，极大地丰富了视频内容的多样性和创造性。

1.2 AI 搜索

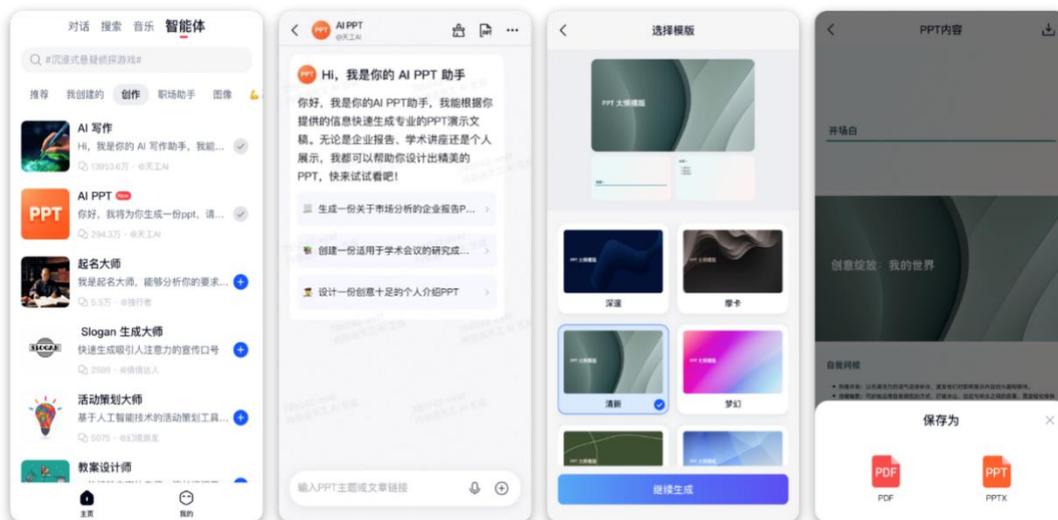
本报告期，随着天工大模型更新到 3.0 版本，公司在“天工 AI 搜索”的基础上，进一步完善产品功能，并发布最新版天工 AI 智能助手。新版助手集成了多轮搜索、图表绘制、研究模式等 AI 能力，提供了一个全面、高效的 AI 生产力工具。它的多轮搜索与综合工具调用能力，使得模型能够独立规划、调用、组合外部工具及整合信息，完成复杂的用户需求。



天工 AI 智能助手展示

新版助手还打造了一个强大的智能体平台，目前，天工 AI 智能体平台已拥有近百万个官方及 UGC 智能体，涵盖学习效率、研究分析、生活娱乐等多个方面。例如备受用户青睐的 AI PPT 智能体，用户仅需简明扼要地提出 PPT 的内容需求，智能体便能够迅速响应，在线生成一份结构完整、可编辑的 PPT 大纲以及详尽的各页内容。此功能不仅实现了智能化的图片检索和模板匹配，还能一键打造出内容充实、逻辑连贯、设计精美的 PPT 文档。用户可以选择在线编辑，或将文档下载至本地进行个性化的调整和分享。

报告期内，天工 AI 智能助手日活跃用户（DAU）突破 100 万，商业化进展顺利，为平台的持续发展奠定了坚实基础。



AI PPT 智能体展示

1.3 AI 游戏

公司目前有包括《Club Koala》、《Project G7》在内的数款 AI 游戏处于研发过程当中。

《Club Koala》是一款基于 AI 打造的 UGC 社交游戏，游戏内部可以分成商业区、游乐区和私人区三大部分，能够为不同游戏乐趣的玩家提供个性化游戏体验。游乐区中，我们为玩家提供了赛车、跑酷、生存等数百种不同类型的小游戏。与此同时，我们还为用户提供了功能强大、上手简单的 AI 游戏编辑器，用户可以轻松将自己脑海中的游戏创意变成现实。商业区中，我们提供了各类公共设施，并构建完善的社交机制，让玩家可以在其中结交志同道合的朋友，也可以去彼此的庄园中造访。在此基础上，玩家还可以购置小岛来建设自己专属的私人庄园，并在其中进行耕种、采摘、钓鱼等活动。与此同时，游戏中还配有大量性格、功能各异的 AI NPC，玩家可以与其开启一段独一无二的奇特故事。

《Club Koala》旨在成为全球玩家的大型线上游乐园，为玩家提供全方位的娱乐和社交体验。产品前期在各类线下游戏展会及 Discord 社区中广受好评。《Club Koala》已于本报告期顺利开启 Alpha 测试，预计于 2024 年 10 月开启 Beta 版测试。

《Project G7》是一款基于 AI 打造的 UGC 开放世界 RPG 游戏，游戏以风格化美术设计、自由探索的开放世界、AI 驱动的剧情和编辑器、以及先进的人物社交系统为核心，构建了三大主要玩法。首先，玩家可以自由探索广阔的城镇，体验驾驶、射击、格斗等多样化的交互方式。其次，Project G7 集成了丰富的社交玩法，通过 RP Agent 技术赋予 NPC 以人格和灵魂，实现与玩家的深度互动。此外，游戏支持 UGC 创作，鼓励玩家分享和展示自己创作的剧情和故事，实现内容的长期供给和沉淀。

1.4 AI 音乐

公司依托天工 SkyMusic 模型的强大能力，面向海内外市场推出多款 AI 音乐产品，以满足不同用户的使用需求。

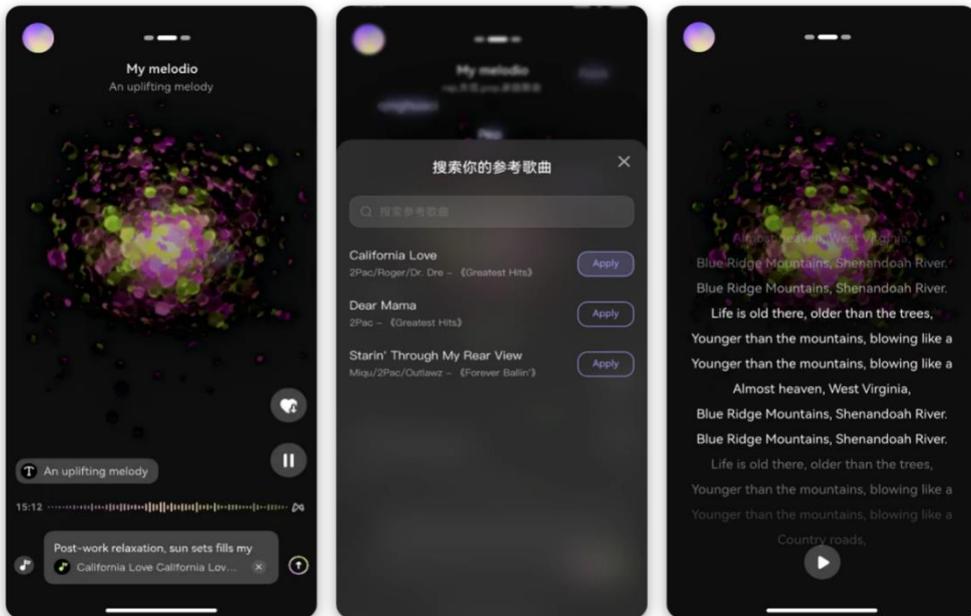
在国内，面向 C 端用户，我们推出了天工音乐服务，用户仅需进入新版天工 AI 助手的音乐频道，输入歌曲名称和歌词（或借助 AI 辅助生成），并选择一首参考曲目设定风格（或是由 AI 智能匹配），之后“天工 SkyMusic”便能自动完成谱曲和混音，创作出个性化的 AI 音乐作品。整个创作流程简单直观，用户可以轻松下载、分享自己的专属歌曲，享受前沿音乐创作技术，挥洒创意，自由表达。



天工音乐功能展示

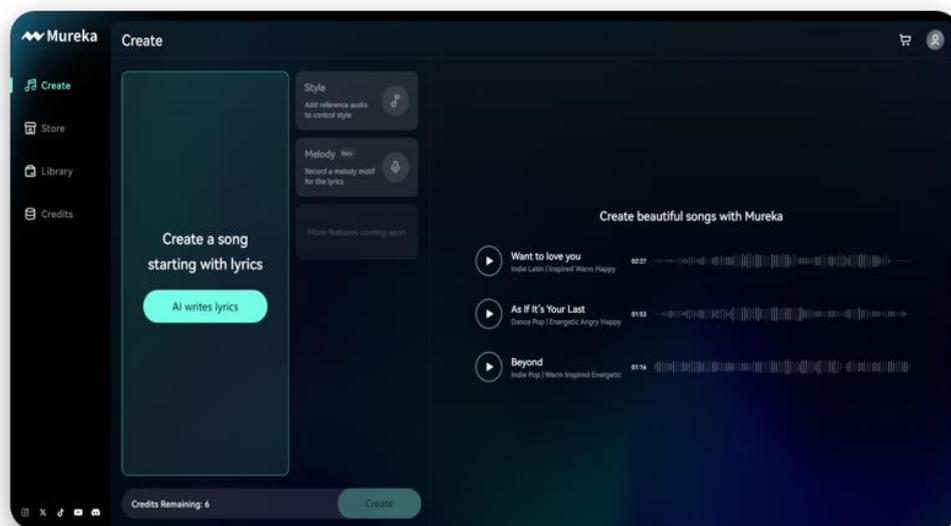
在海外，2024 年 8 月 14 日，昆仑万维正式发布搭载全新自研音乐大模型 SkyMusic 2.0 的全球首个 AI 流媒体音乐平台 Melodio 与 AI 音乐商用创作平台 Mureka，为用户提供更加丰富和个性化的音乐体验。

Melodio 是全球首款个性化 AI 流媒体音乐平台，用户只需根据所处场景或即时心情输入提示词，Melodio 便将持续生成相应风格的定制化无限流式音乐、满足用户个性化情绪表达。用户还可以随时通过简单的指令来调整音乐生成，并对喜欢的片段进行保存、分享满足其社交需求。此外，Melodio 还提供了动态的音乐可视化功能，使用户在享受音乐的同时，也能通过视觉感受音乐的独特律动。



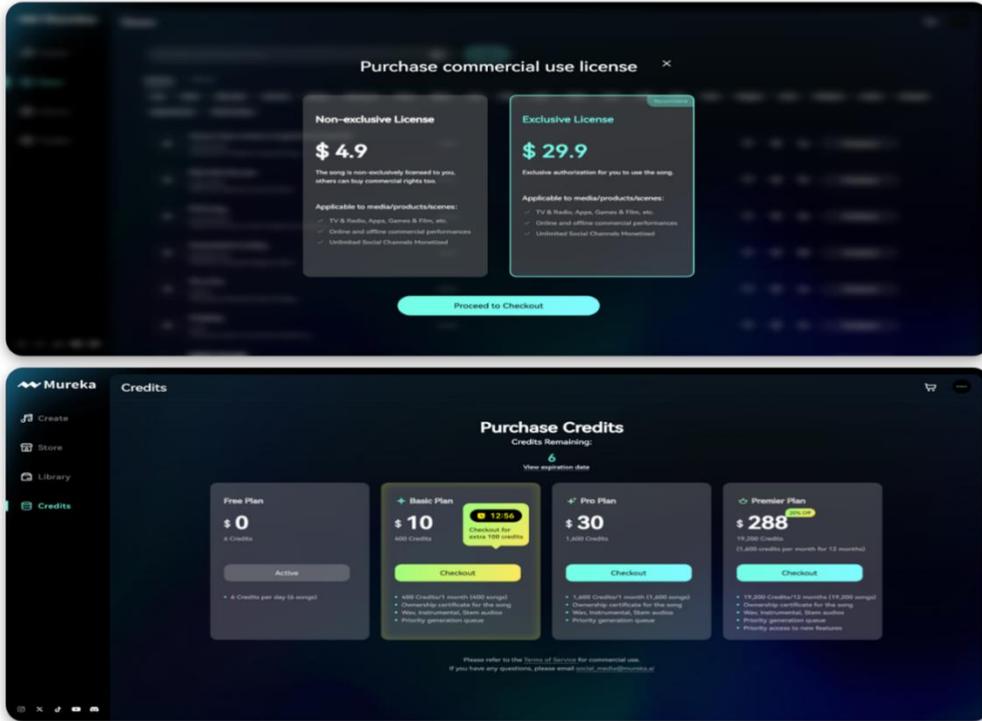
Melodio 产品展示

Mureka 是公司面向海外推出的革命性的 AI 音乐商用创作平台。它不仅为专业艺术家和音乐爱好者提供了一个共同的舞台，还通过其独特的商业模式，为音乐创作和商业化开辟了全新道路。Mureka 能够根据用户输入的歌词和音乐风格偏好，生成具有高度稳定性和可控性的原创音乐，并利用 Style 功能来定制音乐风格。Mureka 的智能算法支持对前奏、主歌、副歌等各个段落的精细调整，确保每首作品都能完美符合创作者的灵感表达。



Mureka 创作界面展示

Mureka 创新的商业模式保证了平台的活力和可持续性。用户在 Mureka 平台上的每一次创作，都需要通过消耗积分来完成。这些积分不仅是用户对平台的贡献，更是他们作品价值的直接体现。当用户完成创作后，可以下载创作证明，并申请在 Mureka 商店中进行展示出售，吸引潜在的买家和听众。除了出售自己的作品，用户还可以在 Mureka 商店购买各种音乐素材，如伴奏和分轨音频等，以进一步丰富和完善音乐作品。而每当用户通过平台出售歌曲获得收益时，Mureka 会按照合理的既定比例进行抽佣。通过这种方式，Mureka 确保了平台、创作者和消费者之间的利益平衡，促进了整个音乐创作生态的健康发展。

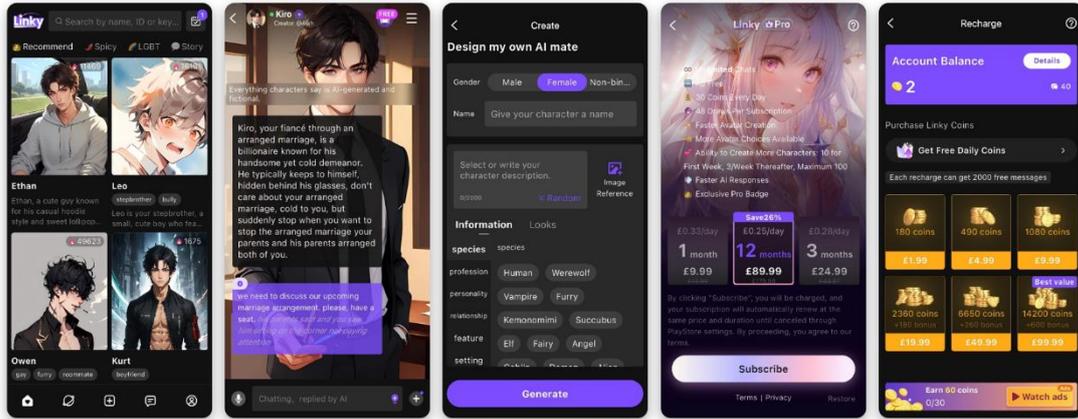


Mureka 商业化展示

1.5 AI 社交

Linky 是昆仑万维面向海外市场推出的新一代 AI 社交产品，致力于通过角色扮演互动和内容消费，为用户提供高质量的情绪价值和内容价值，带来前所未有的社交体验。Linky 已经成功构建了一个 UGC 人设生态闭环，日均 UGC 内容达到上万条，涵盖了丰富的剧情 NPC 及人设内容。这一生态不仅为用户带来了多样化的角色扮演选择，还通过商业化分成体系激励用户持续提升内容质量，确保了平台的活力和持续发展。在技术方面，Linky 的角色大模型和故事大模型能够标签化一键组装大纲剧情和长短篇故事，与用户进行持久的角色扮演互动。目前，产品在 App 内整合了故事、剧本创作和聊天互动功能，支持用户一键生成 Chat 剧本。在人设一致性、长文本记忆、剧情丰富性等方面，Linky 均达到了业内领先水平，为用户提供了更加真实、沉浸式、个性化的社交娱乐体验。

Linky 作为一款全球领先的 AI 社交产品，商业化进展显著，尤其在发达国家和地区表现卓越。报告期内，Linky 日活跃用户（DAU）突破 50 万，用户粘性极强。



AI 社交产品 Linky 功能展示

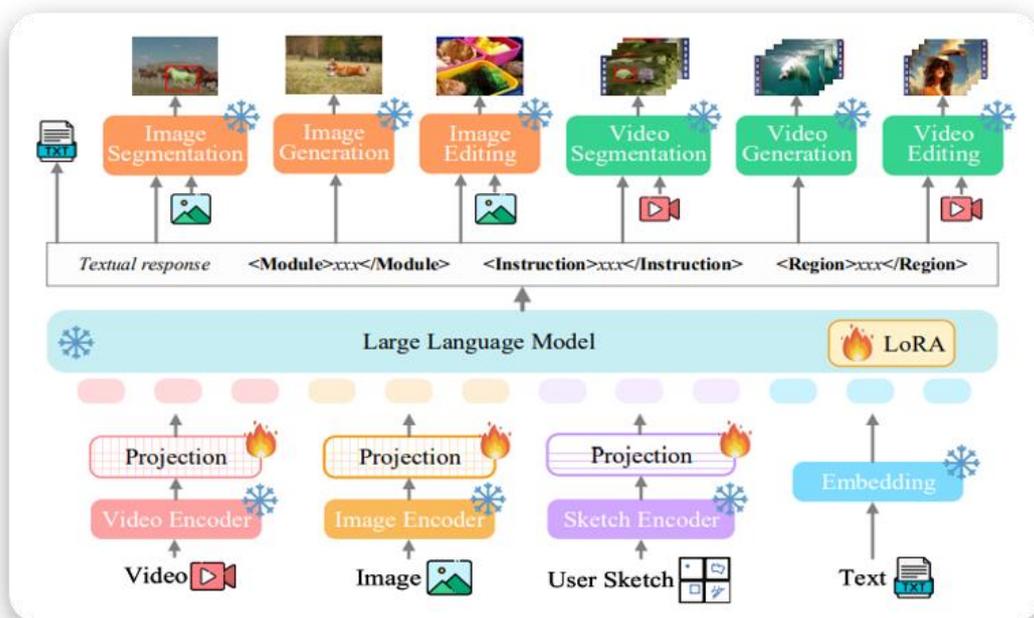
1.6 AI 短剧

2024 年 8 月，我们推出了全球首个集成视频大模型与 3D 大模型的 AI 短剧平台 SkyReels。通过引入先进的 AI 技术，我们大幅简化了剧本创作流程，提高了内容生产的效率和质量，为用户提供了一个从创意构思到视觉呈现的一站式解决方案。SkyReels 平台集成了昆仑万维自研剧本大模型 SkyScript、自研分镜大模型 StoryboardGen、自研 3D 生成大模型 Sky3DGen、以及业界首个将 AI 3D 引擎与视频大模型深度融合的创新平台 WorldEngine。剧本模型是 AI 短剧创作的起点，通过智能分析用户输入的标签化选项，快速组装出符合用户需求的剧情框架。剧本模型亦支持高级剧情大纲输入，能够根据用户给出的大纲自动生成跌宕起伏、引人入胜的剧情内容。此外，模型还支持已有剧本的自动整理与润色，提高创作效率。而分镜模型是将剧本转化为视觉画面的关键环节。模型根据剧本内容和情感走向，智能规划分镜设计，生成分镜图片和对应的文字脚本并自动匹配最合适的镜头语言和视角。分镜模型的智能分析能力，确保了每一帧画面都能够精准传达剧情的情感和张力。同时，用户还可以自由调整分镜、镜头角度、运镜方式等参数，以确保创意的精确呈现。最后视频生成模型是 AI 短剧系统的最终输出环节。它负责将分镜设计转化为实际的视频内容，通过匹配视频模板和应用先进的图像处理技术，创作出视觉效果出众的短剧。这一模型不仅能够保证视频内容的高质量，还能够根据用户的反馈进行实时调整，确保最终作品能够满足用户的期待。

目前 SkyReels 已支持 1080P 60 帧的高清视频输出，单次可生成视频长度达 180 秒，视频生成后用户可一键整合所有创作成果，并可快速导出和分享。平台还支持多种格式下载，方便用户后期编辑，推动“一人一剧”时代加速来临。

1.7 AI 普惠生态建设

本报告期，公司继续坚定履行承诺，通过发布和开源多项技术，进一步推动生态建设。报告期内，昆仑万维与新加坡国立大学、新加坡南洋理工大学团队联合发布了 Vitron 通用像素级视觉多模态大语言模型。Vitron 支持从视觉理解到视觉生成、从低层次到高层次的一系列视觉任务，解决了图像/视频模型割裂问题，提供了一个全面统一静态图像与动态视频内容的理解、生成、分割、编辑等任务的像素级通用视觉多模态大模型，为下一代通用视觉大模型的终极形态奠定了基础。



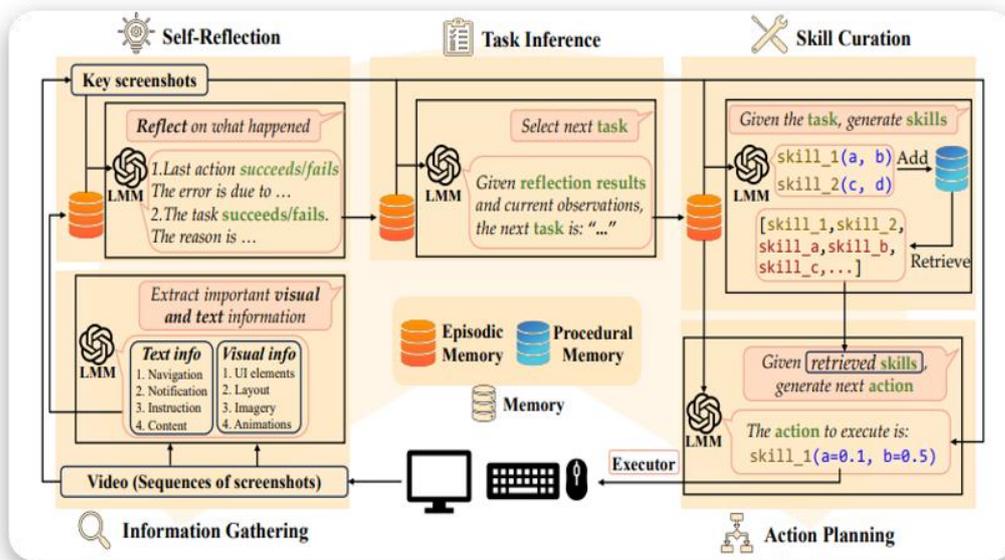
Vitron 框架示意图

公司在报告期内开源了千亿参数级稀疏大模型 Skywork-MoE。这是首个完整应用 MoE Upcycling 技术并落地的千亿参数级开源 MoE 大模型，也是首个支持用单台 4090 服务器完成推理的千亿参数级开源 MoE 大模型。Skywork-MoE 模型的总参数量为 146B，激活参数量 22B，共有 16 个 Expert，每个 Expert 大小为 13B，每次激活其中的 2 个 Expert。在显著提升了模型的推理效率和泛化能力的同时，Skywork-MoE 的推理成本有近 3 倍的下降，大幅降低了使用门槛。

	#AP	#TP	CEVAL	CMMLU	MMLU	GSM8K	MATH	HumanEval
Deepseek-67B	67	67	66.1	70.8	71.3	63.4	18.7	42.7
Qwen1.5-72B	72	72	84.1	83.5	77.5	79.5	34.1	41.5
Llama2-70B	70	70	-	-	68.9	56.8	13.6	29.9
Llama3-70B	70	70	-	-	78.8	82.7	36.7	39.0
Mixtral 8*7B	13	47	-	-	70.6	58.4	28.4	40.2
Mixtral 8*22B	39	141	-	-	77.8	78.6	41.8	45.1
Grok-1	86	314	-	-	73.0	62.9	23.9	63.2
DBRX-Instruct	36	132	-	-	73.7	66.9	-	70.1
Deepseek-V2	21	236	81.7	84.0	78.5	79.2	43.6	48.8
Skywork-13B	13	13	62.1	62.4	62.7	60.2	8.4	18.9
Skywork-MoE	22	146	82.2	79.5	77.4	76.1	31.9	43.9

Skywork-MoE 测评表现

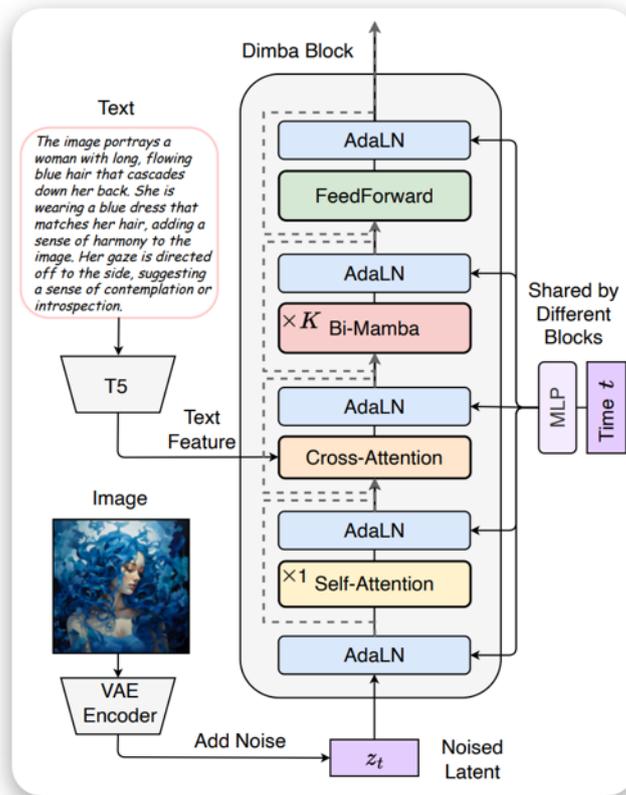
此外，昆仑万维携手北京智源人工智能研究院、新加坡南洋理工大学、北京大学等机构，提出了通用计算机控制框架 Cradle，使 AI Agent 无需训练便能像人一样直接控制键盘鼠标，实现任意开闭源软件交互。Cradle 在进行复杂任务的连续执行和长程规划决策中表现出色，极大地扩展了智能体的交互范围。



Cradle 框架示意图

昆仑万维还开源了用于 Diffusion 的新架构 Dimba，其融合了 Transformer 和 Mamba 的混合架构，

创造出独特的 Dimba 块。Dimba 通过交替堆叠 Transformer 和 Mamba 层，并结合跨注意力层来集成条件信息，优化了图像生成的质量和效率。项目团队采用了高效的数据集构建策略和先进的训练方法，包括质量调优技术和分辨率适应策略，显著提升了图像美感和质量。此外，Dimba 的设计灵活性支持针对不同资源约束和目标场景的定制，实现了在保持高性能的同时减少内存占用。



Dimba 模型架构示意图

同时，作为国内领先的人工智能企业，昆仑万维积极响应国家号召，致力于推动人工智能技术的普惠化发展。公司通过校园公益行动，为师生提供 AI 知识培训和实践操作，激发学生们对 AI 技术的兴趣和探索欲望。昆仑万维还联合众谷公益，为偏远山区儿童带来 AI 系列公益课程，让 338 所学校的 17 万师生享受到 AI 教育，缩小城乡教育差距。

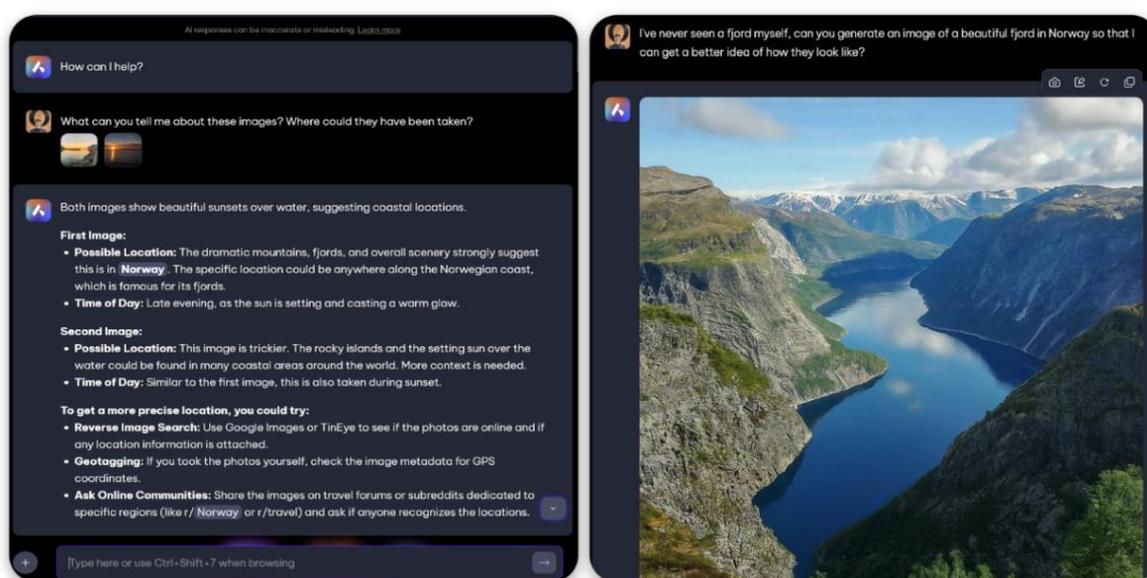
此外，面对 OpenAI 封锁的挑战，昆仑万维迅速推出“一键搬家计划”，助力开发者无缝迁移至国内大模型，并提供 1.5 亿 Tokens 的 API 调用资源。公司还提供专属迁移服务技术支持，包括最新模型能力体验、备案培训及辅助等，确保开发者顺利过渡，支持我国人工智能行业创新发展。

2、海外信息分发与元宇宙业务板块

根据 Opera 2024 年第二季度报告，2024 年上半年，Opera 实现营业收入 2 亿美元，同比增长 17%；实现经调整后 EBITDA 5,152 万美元，同比增长 22%。截至 2024 年第二季度，Opera 全球月活跃用户达 3 亿，同时得益于高 ARPU 地区用户的持续增长，2024 年第二季度 Opera 年化 ARPU 达 1.5 美元，同比增长 25%。

报告期内，Opera 继续引领 AI 浏览器革新，启动“AI Feature Drops”计划，推出一系列创新功能。首先，是“Ask Aria”，它允许用户在浏览网页时通过高亮文本并询问来获取更多信息。无论是深入探索不熟悉的主题，还是快速获取复杂文本的摘要，Ask Aria 都能提供即时帮助。这一功能特别适合像学生和研究人员等需要大量阅读和理解信息的用户。其次是“Aria Voice”，这是一项创新的语音交互功能。它使用户即使在移动中也能与浏览器进行自然语言交流。用户只需点击键盘上的麦克风图标即可激活 Aria Voice，提出问题并听取回答。这项功能不仅提高了 Aria 的可访问性，也为多任务处理和需要免提操作的场景提供了极大的便利。

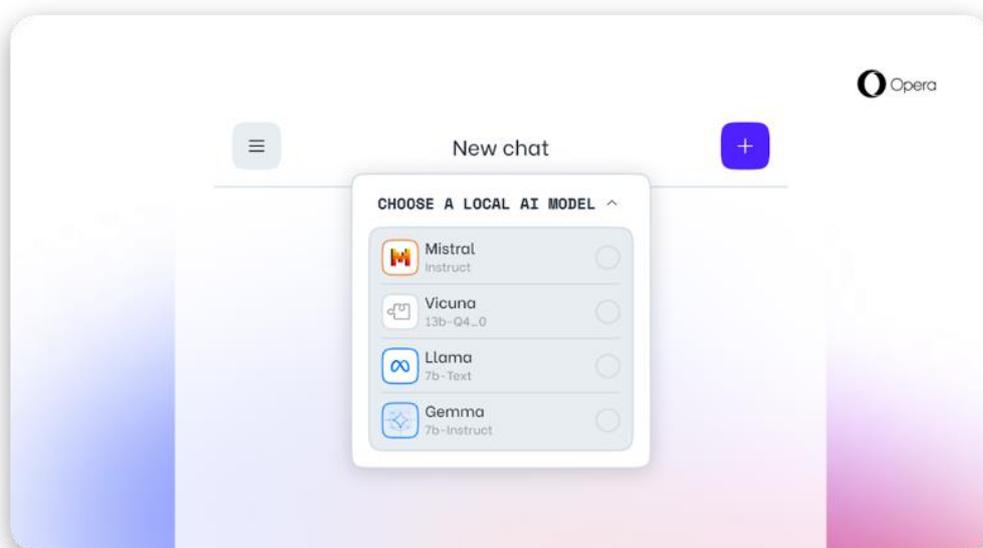
本报告期，Aria 引入了先进的图像生成和理解功能，显著提升了用户体验。借助 Google 的 Imagen2 模型，Aria 能够根据用户的文本提示和描述生成独特图像，为用户提供了一种全新的视觉内容创作途径，丰富了演讲视觉素材和个性化场景图像创作。同时，图像理解功能允许用户上传图片并询问相关问题，Aria 通过识别图片内容并提供相关信息和上下文，帮助用户深入理解图片背后的含义，甚至在解决数学问题或编程难题方面提供支持。此外，我们将图像理解与图像生成相结合，进一步激发了用户的创造力，用户只需上传草图，Aria 便能根据图示和简短文本提示生成新的图像，提供全新的创作体验。



Aria 图像理解、生成展示

Aria 还新增了语音输出功能，让 Aria 能够大声朗读文本答案，非常适合需要辅助功能的人士、多任务处理或喜欢以不同格式获取信息的用户。用户可以通过侧边栏或命令行启动与 Aria 的聊天，并使用新的语音图标来听取回答。这些新功能的推出，不仅展示了 Opera 在 AI 领域的创新能力，也体现了我们对用户需求的深刻理解。我们期待这些功能能够简化用户的浏览体验，并激发更多创新的互动方式。

此外，为满足用户需求，Aria 还引入了本地大语言模型的支持。目前 Aria 能够接入超过 150 种本地 LLM 模型，这些模型来自约 50 个不同的模型家族，包括 Meta 的 Llama、Vicuna、Google 的 Gemma、Mistral AI 的 Mixtral 等。这一创新举措使得用户能够直接在自己的设备上处理需求，而不必将数据发送到远程服务器，从而增强了隐私保护和数据处理速度。用户可以通过 Aria 聊天侧边栏选择本地模式，轻松访问和管理这些本地 LLM，享受更快速、更个性化的 AI 响应。未来我们将继续收集用户反馈，不断优化和完善 Aria 的功能，以提供更优质的服务。



Aria 本地语言模型展示

算力方面，Opera 成功在冰岛凯夫拉维克部署由绿色能源驱动的全新 AI 集群。该集群采用了搭载全新 H100 GPU 的 NVIDIA DGX SuperPOD，训练速度较上一代提升四倍，推理速度较上一代提升 30 倍。在第三方评比中，截至今年六月，该集群已成功跻身世界超级计算机百强，位列第 88 名。此外，Opera 还对基于 Arm 的 Windows PC 进行了优化，新版本相较过去性能提升 4 倍以上，为用户提供更快、更节能的浏览体验。

本报告期，Opera 亦对 GX 版 Aria 进行了全面升级，不仅引入了图像生成、语音输出和图像理解等创新功能，还对聊天功能进行了显著改进。现在，用户可以轻松地通过聊天提示让 Aria 根据描述生成

图像，并且如果初次生成的图像不符合预期，还可以利用“重新生成”按钮来获取新的图像版本。此外，Aria 新增的语音输出功能，可以大声朗读回答，非常适合需要多任务处理或寻求辅助功能的用户。Aria 的图像理解能力也得到了提升，现在能够解析用户上传的图片并据此提供答案。聊天体验也因新增的对话摘要功能而变得更加丰富，使用户能够快速回顾对话要点。同时，Aria 还会在聊天中自动提供信息来源链接，方便用户访问更多相关内容。

此外，Opera GX 浏览器推出 GX Mods Mobile 功能，让用户在 iOS 和 Android 设备上亦可畅享个性化定制界面。这一功能延续了 PC 端 Mods 的成功，提供数千款由社区创造的游戏文化主题 Mods，用户可以轻松通过主菜单选择和激活 Mods，享受定制的背景音乐、动态壁纸等。同时 Opera GX 团队还通过举办动态壁纸设计大赛、与知名 IP 联名合作等方式，不断推进 Mods 生态建设，为用户打造个性化浏览体验。截至本报告期末，GX.store 中 Mod 内容已超过 7,000 款。

Opera GX 还推出全新的“Pixels”体系，用户可以通过完成 GX 生态系统内的任务和挑战来获得 Pixels。Pixels 不仅仅是一种奖励机制，还有一种全新的消费方式。用户可以使用 Pixels 个性化装扮自己的账号，塑造独一无二的个人形象。未来 Pixels 还可以在 GX.games 和 GX.store 中使用，用户可以借此支持 GX 生态内的游戏创作者和 Mods 开发者，通过打赏的方式鼓励他们的创新和努力。通过 Pixels，Opera GX 不仅增强了用户与平台的互动，还为 GX 社区注入了新的活力，推动了创意和共享精神的蓬勃发展。

3、投资业务板块

本报告期，受资本市场波动下行影响，公司投资组合中如爱奇艺等标的股价大幅下滑，进而拖累公司的整体投资收益表现，计入较大亏损。我们将继续密切关注市场动态，优化投资组合结构，并采取必要的风险管理措施，以应对市场的不确定性和潜在波动。

报告期内，公司参与投资的昆仑基金持续把握科技创新的前沿趋势，赋能新质生产力发展。本报告期，昆仑基金成功发掘并投资了国内领先的量子计算应用企业——微观纪元，以期能够促进量子计算技术在更多领域的应用，推动相关产业链的创新与升级。融资方面，欧莱新材已于 2024 年 5 月成功登陆科创板，辐联科技、时的科技、星环聚能、跃赛生物和图湃医疗等企业均顺利完成新一轮融资，展现出强劲发展势头。

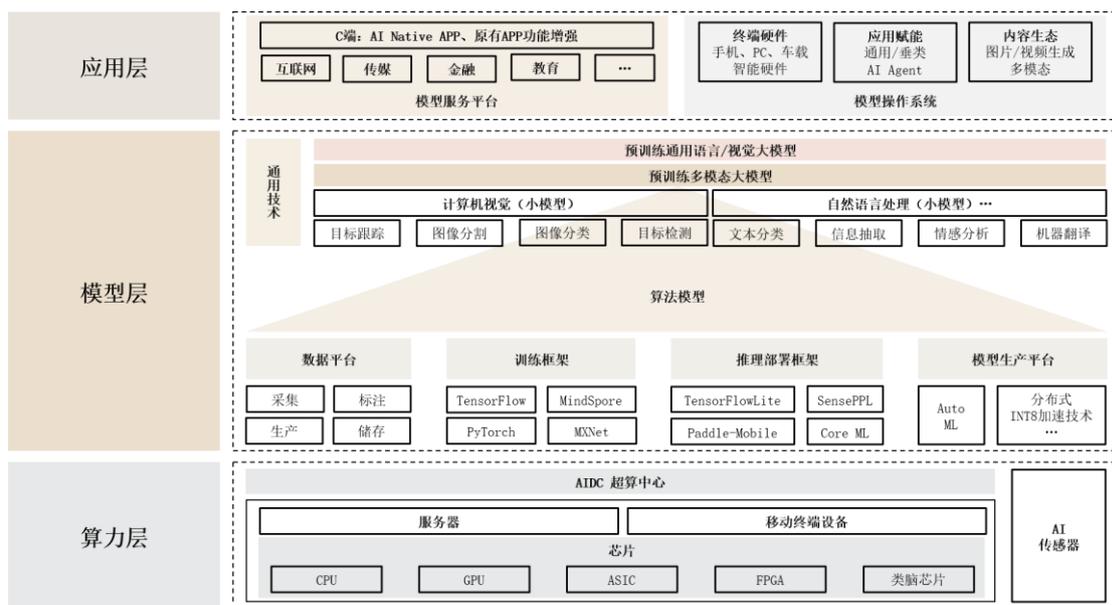
同时，昆仑基金投资组合中其他成员企业亦有亮眼表现。在硬科技领域，小马智行成功在北京地区推出往返大兴机场的自动驾驶收费服务，并在上海中心城区启动了无驾驶人载客服务，自动驾驶商业化应用迈出了坚实的步伐；中科驭数成功发布第三代全自研 DPU 芯片 K2 Pro 等系列新产品，进一步巩固了其在高性能计算领域的领先地位。在新能源领域，纤纳光电的钙钛矿组件通过了国际 IEC TS 63209-

1:2021 多倍加严全序列稳定性认证，为其国际市场拓展奠定了基础；星环聚能高温超导磁体和重复重联方案工程验证取得显著进展，为新能源技术的发展贡献了重要力量。在医疗健康领域，新合生物联合申报的“十四五”项目成功获得国家重点专项立项；征祥医药的 1 类新药玛赛洛沙韦片的上市申请（NDA）已获国家药品监督管理局的正式受理。

未来公司将继续聚焦于新质生产力的培育和发展，通过不断的科技创新和资本投入，推动产业升级和社会进步，为社会创造更多价值。

报告期内公司所处行业情况：

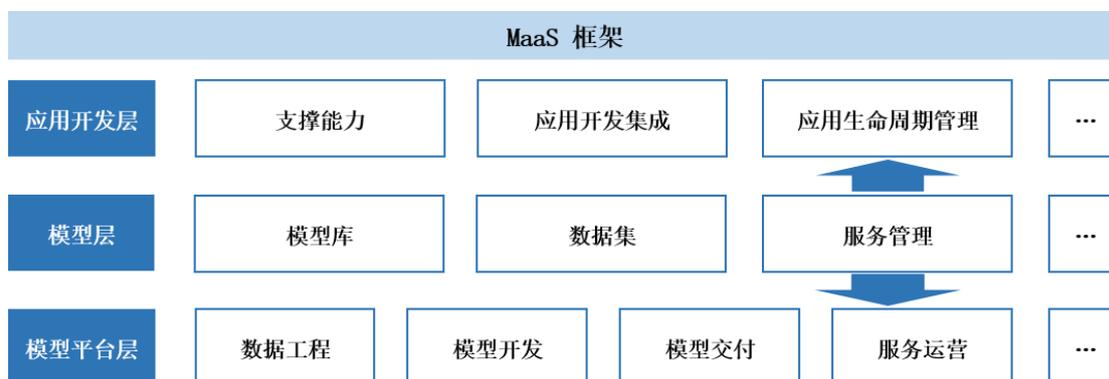
2024 年上半年，人工智能行业继续保持高速发展，在 AI 模型创新和应用落地方面均取得了显著成果。国际上，诸如 OpenAI 的 GPT-4o 和 Google 的 Gemini 系列等大型语言模型不断推陈出新。国内方面，以昆仑万维、字节跳动、阿里巴巴、百度等为代表的科技公司在 AI 大模型技术上亦取得了一系列重要成果。此外，AI 的应用场景不断拓宽，从智能搜索、AI 助手到智能制造、智慧医疗等，AI 技术正逐渐渗透到社会生产和生活的各个方面。特别是在端侧模型、多模态交互和大模型的商业化应用方面，行业展现出了巨大的潜力和活力。国内外的企业和研究机构正通过技术创新和应用实践，推动人工智能行业向着更加智能、便捷和高效的方向发展，整个行业展现出一派生机勃勃的景象。



数据来源：中国信通院、中金公司

模型创新方面，OpenAI 发布视频生成模型 Sora，引起全球广泛关注。该模型能够生成长达一分钟、视觉质量高、并严格遵循用户提示的视频内容，并且能够在单个视频中创建多个镜头，同时保持角色和

在金融、电信等行业的应用已取得积极成效，推动了大模型技术的规模化落地。MaaS 框架由模型平台层、模型服务层和应用开发层组成，提供了从数据处理到模型开发、部署及运营的全流程服务。MaaS 的成功实施，不仅提升了模型服务的效率，也促进了 AI 技术在更广泛场景中的应用。



数据来源：中国信通院

AI 智能体或称为 AI Agent，是 AI 应用的另一种形式，它通过集成到各种应用程序中，为用户提供个性化和智能化的服务。如 GPTs 作为基于大模型的应用开发工具，可通过自然语言方式创建 GPT 应用，国内头部企业也纷纷推出各类应用开发平台，可实现 AI 原生应用开发，并支持构建 AI Agent (智能体)，达到大模型与各组件或系统工具的强互动及关联。据 MarketsandMarkets 预测，至 2028 年全球 AI Agent 市场规模将达到 285 亿美元，2023 至 2028 年复合增长率将达到 43.0%。

根据 QuestMobile 的研究报告，2024 年上半年国内 AIGC 应用同样迎来了爆发式增长，各类 AI 应用陆续推向市场。AIGC App 主要分为 AI 工具和 AI 社交/娱乐两大赛道，形成了细分化的竞争格局。虽然大多数 App 处于早期发展阶段，但已有如天工、豆包等月活跃用户规模在数百万甚至千万量级的 App 出现。QuestMobile 数据显示，目前国内 AIGC 类 App 在 36-40 岁群体间覆盖较好，在年轻用户仍有极大发展空间。



数据来源: QuestMobile

国内外 AI 应用的进展表明, 人工智能技术正成为全球经济增长的新引擎。随着技术的不断成熟和市场的不断扩大, AI 应用正逐步渗透到经济社会的各个领域, 展现出广阔的发展前景和巨大的潜力。无论是在提高生产效率、优化资源配置, 还是在创造新的消费体验、推动社会进步方面, AI 应用都发挥着越来越重要的作用。随着国内外企业和研究机构在 AI 领域的持续投入和创新, 我们有理由相信, 人工智能将为人类社会带来更多的可能性和机遇。