



关于

沐曦集成电路（上海）股份有限公司

首次公开发行股票并在科创板上市

申请文件的

审核问询函的回复

保荐人（主承销商）



华泰联合证券有限责任公司
HUATAI UNITED SECURITIES CO., LTD.

二零二五年八月

上海证券交易所：

沐曦集成电路（上海）股份有限公司（以下简称“公司”、“发行人”或“沐曦股份”）收到贵所于 2025 年 7 月 19 日下发的《关于沐曦集成电路（上海）股份有限公司首次公开发行股票并在科创板上市申请文件的审核问询函》（上证科审〔2025〕113 号）（以下简称“《审核问询函》”），公司已会同华泰联合证券有限责任公司（以下简称“华泰联合证券”、“保荐人”、“保荐机构”）、上海市锦天城律师事务所（以下简称“发行人律师”）、立信会计师事务所（特殊普通合伙）（以下简称“申报会计师”）进行了认真研究和落实，并按照问询函的要求对所涉及的事项进行了资料补充和问题回复，现提交贵所，请予以审核。

关注问题	黑体（不加粗）
关注问题回复、中介机构核查意见	宋体（不加粗）
招股说明书修订内容	楷体（加粗）

如无特别说明，本回复报告中的简称或名词的释义与《沐曦集成电路（上海）股份有限公司首次公开发行股票并在科创板上市招股说明书》中相同。

本回复报告部分表格中单项数据加总与合计可能存在微小差异，均系计算过程中的四舍五入所形成。

目 录

1.关于产品与市场竞争	3
2.关于核心技术	57
3.关于尚未盈利	96
4.关于收入	123
5.关于销售模式	141
6.关于应收账款	149
7.关于采购和供应商	176
8.关于成本和毛利率	183
9.关于股东股权	196
10.关于存货和固定资产	221
11.关于期间费用	258
12.关于股权激励	271
13.关于募投项目	286
14.关于信息披露	305
保荐人总体意见	307

1.关于产品与市场竞争

根据申报材料：（1）发行人产品分为训推一体系列（GPU 板卡、GPU 服务器）、智算推理系列（GPU 板卡）、IP 授权及其他，直接下游为服务器厂商、集成商或智算中心建设方。报告期内，不同产品类型收入占比变化较大；（2）针对不同客户的需求，除交付内嵌 GPU 芯片的板卡/模组外，发行人亦可向客户交付集成多个板卡后的服务器、一体机/工作站，以及由多个服务器、存储设备、网络设备等组成的智算集群，基础算力底座方面相继交付 9 大智算集群；（3）全球范围 GPU 市场形成了由英伟达和 AMD 组成的寡头垄断格局，公司国内竞争对手除通用型计算架构（GPU）企业外，还包括华为海思、燧原科技、昆仑芯等专用型计算架构（ASIC/DSA）芯片设计企业；（4）招股说明书中发行人与国际龙头企业及国内重点企业在市场竞争方面的比对情况做了简要披露，未充分说明公司的竞争地位。

根据公开材料：2025 年 7 月 15 日，英伟达宣布将恢复 H20 GPU 在中国的销售，并推出面向中国市场的全新且完全兼容的 GPU。

请发行人披露：（1）发行人不同产品类型下主要细分产品型号、销量、收入、单价及毛利率情况，与同行业厂商在产品类型丰富度、细分产品类型、主要销售型号数量的比较情况、差异原因。报告期内不同产品类型收入结构变化较大、当前收入集中于曦云 C500 的原因，其他产品的研发、客户开拓及商业化进展，发行人未来主力产品计划；（2）不同产品类型发行人与客户具体合作模式，如智算集群等；未来拟主要推广的产品类型或合作模式，服务器或集群业务是否会与现有客户形成竞争影响合作稳定性；目前交付的智算集群情况、涉及的产品自研外购情况、对应直接客户和终端客户、应用领域等；（3）发行人产品终端客户类型，客户采购后的用途及使用点亮情况，各地对采购销售国产 GPU 的支持政策、客户采购公司产品的主要驱动因素。结合发行人报告期内产品结构变化、客户复购率及持续采购需求、当前国产 GPU 商业化拓展的难点及公司竞争优势等，说明公司主要发展方向及竞争策略；（4）市场主流计算架构（如 GPU、ASIC/DSA 等）以及 GPU 不同产品类型（如计算、图形等）各自对应的市场规模、下游应用领域、行业参与厂商，公司产品覆盖的应用领域产品的市场空间、竞争格局及未来发展趋势；（5）量化分析公司产品在境内市场的占有率情况，

结合发行人与境内外同行业公司在产品布局、性能、业绩规模、市场份额、主要下游客户等情况的比较,分析公司与境外龙头的差距及在境内市场的竞争优劣势。评估境外 GPU 龙头对华产品销售政策变化可能对公司业绩及市场拓展产生的影响。

请保荐机构简要概括核查过程,并发表明确意见。

回复:

一、发行人披露

(一) 发行人不同产品类型下主要细分产品型号、销量、收入、单价及毛利率情况,与同行业厂商在产品类型丰富度、细分产品类型、主要销售型号数量的比较情况、差异原因。报告期内不同产品类型收入结构变化较大、当前收入集中于曦云 C500 的原因,其他产品的研发、客户开拓及商业化进展,发行人未来主力产品计划

1、发行人不同产品类型下主要细分产品型号、销量、收入、单价及毛利率情况

发行人致力于自主研发全栈高性能 GPU 芯片及计算平台,主营业务是研发、设计和销售应用于人工智能训练和推理、通用计算与图形渲染领域的全栈 GPU 产品,并围绕 GPU 芯片提供配套的软件栈与计算平台。报告期内,发行人实现销售收入的 GPU 产品主要包括训推一体系列和智算推理系列,其中训推一体系列主要为 GPU 板卡和 GPU 服务器的销售,具体型号包括曦云 C500 和曦云 C550,智算推理系列 GPU 板卡具体型号包括曦思 N100、曦思 N260。

报告期各期,发行人不同产品类型下主要细分产品型号、销量、收入、单价及毛利率情况如下表所示:

产品类型		2025 年 1-3 月				2024 年			
		收入	销量	单价	毛利率	收入	销量	单价	毛利率
训推一体系列 GPU 板卡	曦云 C500	27,297.01	7,380	3.70	54.44%	31,811.30	6,649	4.78	65.32%
	曦云 C550	3,958.23	661	5.99	68.42%	19,370.09	4,256	4.55	60.50%
	合计	31,255.24	8,041	3.89	56.21%	51,181.39	10,905	4.69	63.50%
训推	曦云 C500	35.00	1	35.00	52.51%	13,659.95	215	63.53	42.43%

产品类型		2025 年 1-3 月				2024 年			
		收入	销量	单价	毛利率	收入	销量	单价	毛利率
一体系列 GPU 服务器	曦云 C550	69.03	1	69.03	39.07%	7,332.18	128	57.28	11.70%
	合计	104.03	2	52.01	43.59%	20,992.13	343	61.20	31.70%
智算推理系列 GPU 板卡	曦思 N100	300.93	1,303	0.23	-0.44%	281.11	482	0.58	6.08%
	曦思 N260	99.82	47	2.12	26.91%	67.96	30	2.27	26.70%
	合计	400.75	1,350	0.30	6.37%	349.08	512	0.68	10.09%
IP 授权		25.00	不适用	不适用	100.00%	100.00	不适用	不适用	100%

注：上表中，收入金额单位均为“万元”，训推一体系列 GPU 板卡、智算推理系列 GPU 板卡销量单位均为“张”、单价单位均为“万元/张（不含税）”，训推一体系列 GPU 服务器销量单位为“台”、单价单位为“万元/台（不含税）”，下同

（续）

产品类型		2023 年				2022 年			
		收入	销量	单价	毛利率	收入	销量	单价	毛利率
训推一体系列 GPU 板卡	曦云 C500	1,546.81	272	5.69	65.48%	-	-	-	-
	曦云 C550	-	-	-	-	-	-	-	-
	合计	1,546.81	272	5.69	65.48%	-	-	-	-
训推一体系列 GPU 服务器	曦云 C500	-	-	-	-	-	-	-	-
	曦云 C550	-	-	-	-	-	-	-	-
	合计	-	-	-	-	-	-	-	-
智算推理系列 GPU 板卡	曦思 N100	1,376.32	2,042	0.67	5.37%	42.64	52	0.82	24.10%
	曦思 N260	-	-	-	-	-	-	-	-
	合计	1,376.32	2042	0.67	5.37%	42.64	52	0.82	24.10%
IP 授权		2,216.67	不适用	不适用	100.00%	-	-	-	-

2、与同行业厂商在产品类型丰富度、细分产品类型、主要销售型号数量的比较情况、差异原因

发行人与同行业可比公司在产品类型丰富度、细分产品类型、主要销售型号数量的比较情况如下：

项目	英伟达	超威半导体	寒武纪	海光信息	景嘉微	摩尔线程	发行人
成立年份	1993 年	1969 年	2016 年	2014 年	2006 年	2020 年	2020 年
产品类型丰富度	GPU 板卡、图形显卡及配套产品，主要应用于数据中心、桌面游戏、专业视觉设计、智能驾驶、机器人等场景	CPU 板卡、GPU 板卡、图形显卡、SoC 等，主要应用于数据中心、个人电脑、工作站等场景	智能芯片及加速卡、智能处理器 IP，主要应用于数据中心、云计算、消费电子等场景	CPU、DCU（GPU）及配套产品、技术服务，其 DCU 产品主要应用于数据中心	图形显控（图形 GPU）、小型专用化雷达领域的核心模块及系统级产品等	AI 智算产品、专业图形加速产品、桌面级图形加速产品、智能 SoC 类产品	GPU 产品及配套软件栈与计算平台，广泛应用于人工智能、通用计算、智慧城市、自动驾驶、云渲染等前沿领域
2024 年度收入构成	数据中心产品：88.27%；游戏 GPU 和相关产品：8.70%；专业视觉设计 GPU：1.44%；智能驾驶 GPU 及相关产品 1.30%；其他业务：0.30%	数据中心产品：48.78%；游戏：10.06%；企业、嵌入式和半定制：13.79%；客户端：27.36%	云端智能芯片及加速卡：99.30%；边缘端智能芯片及加速卡：0.56%；其他：0.15%	高端处理器：99.70%；技术服务 0.30%	图形显控领域产品：52.53%；小型专用雷达领域产品：13.01%；其他：34.46%	AI 智算产品：77.63%；专业图形加速：19.58%；桌面级图形加速：2.48%；其他：0.31%	训推一体 GPU 板卡：68.99%；训推一体 GPU 服务器：28.29%；智算推理 GPU 板卡：0.47%；IP 授权：0.13%；其他：2.11%
主要产品类型对比	人工智能训练、人工智能推理、通用计算、图形渲染	人工智能训练、人工智能推理、通用计算、图形渲染	人工智能训练、人工智能推理	人工智能训练、人工智能推理、通用计算	图形显控	人工智能训练、人工智能推理、通用计算、图形渲染	人工智能训练、人工智能推理、通用计算
主要产品类型收入占比	根据英伟达年报披露，2024 年计算&网络（主要为人工智能训练推理、通用计算相关产品）分部收入占比约为 89%，图形渲染分部收入占比约为 11%	超威半导体数据中心产品主要包括 AI GPU 和服务 CPU，2024 年 AI GPU 收入金额超过 50 亿美元，占全年收入比例约为 20%；未明确披露图形渲染 GPU 收入及占比信息	2024 年、2025 年 1-3 月，寒武纪云端产品线（包括云端智能芯片及板卡、智能整机）收入占比均在 99%以上	海光信息高端处理器收入包括 CPU 及 DCU 收入，未披露其中 DCU 收入最新的占比情况。根据海光信息《首次公开发行股票并在科创板上市招股说明书(注册稿)》，海光信息 2021 年	2024 年图形显控领域产品收入占比为 52.53%	2024 年人工智能训练推理、通用计算相关产品收入占比约为 78%、图形渲染相关产品收入占比为 22%	2024 年人工智能训练推理、通用计算相关产品收入占比约为 98%

项目	英伟达	超威半导体	寒武纪	海光信息	景嘉微	摩尔线程	发行人
				度 DCU 收入占比为 10.34%			
主要产品形态	板卡、服务器等	板卡、服务器等	板卡、服务器、集群	板卡	微处理器、芯片组、板卡等	板卡、集群及一体机	板卡、服务器等
主要产品形态收入占比	未披露	未披露	未披露云端产品线中芯片、板卡和服务器的收入占比，2022 年、2023 年、2024 年和 2025 年 1-3 月，寒武纪智能计算集群系统收入占比分别为 63.46%、85.46%、0.00%和 0.00%	未披露	未披露	2024 年板卡收入占比 49.35%，集群及一体机收入占比 49.76%	2024 年板卡类产品收入占比约为 69%、服务器类产品收入占比约为 28%
主要产品型号	B200、H100、H20、A100	MI325X、MI250X	MLU590、MLU370	BW100、K100	JM9 系列	S4000、S3000、S80	曦云 C500、曦云 C550
主要产品型号收入占比	未明确披露各型号收入及占比，但英伟达在其 2024 年年报中提到 2024 年 Q4 Blackwell 架构 GPU 产品收入占当季数据中心收入的比例约为 31%	AI GPU 收入占比约为 20%，未披露具体型号产品收入占比	未披露	未披露	未披露	未披露	2024 年曦云 C500（含板卡和服务器）收入占比约为 61%、曦云 C550（含板卡和服务器）收入占比约为 36%

注 1：数据来源于 Wind 及公开资料整理，上表中同行业可比公司主要产品类型、形态、型号及收入占比系经检索企业官网、年度报告、新闻媒体等渠道获得，可能与同行业可比公司实际情况存在一定差异。

注 2：英伟达财务报告期间为 2024 年 1 月末至 2025 年 1 月末。

注 3: 英伟达数据中心产品为通用 GPU, 游戏 GPU 和专业视觉设计 GPU 为图形 GPU。

注 4: 景嘉微主要产品为图形 GPU, 主要应用于军工领域。

注 5: 同行业可比公司龙芯中科主要产品为 CPU, GPU 产品占比较小且其 GPU 产品主要应用于 PC 端, 因此上表未予以列示。

2023 年度、2024 年度和 2025 年 1-3 月，发行人训推一体曦云 C 系列产品（含板卡和服务器）收入分别为 1,546.81 万元、72,173.52 万元和 31,359.27 万元，占同期主营业务收入的比例分别为 30.09%、97.28%和 97.87%，最近一年一期收入占比较大。因成立时间、发展战略、业务和产品结构演变过程、所处发展阶段等方面的差异，发行人当前收入主要来源于人工智能芯片产品，而同行业可比公司主营业务产品还包括图形渲染芯片、CPU 芯片等；此外，在人工智能芯片产品型号数量方面，发行人目前主要在售产品型号包括曦云 C500、曦云 C550、曦思 N260 和曦思 N100，其中曦云 C500、曦云 C550 为主力产品，与同行业可比公司主要在售的产品型号数量没有显著差异。发行人当前产品类型丰富度、产品型号数量与当前人工智能芯片行业发展趋势、芯片研发迭代周期客观规律相符，当前收入主要集中于曦云 C 系列产品与人工智能芯片研发特点、行业发展趋势和同行业可比公司情况相符，具有合理性，具体分析如下：

（1）发行人目前产品矩阵、收入结构符合同行业可比公司整体趋势

具体而言，同行业可比公司普遍成立时间早于发行人。部分同行业可比公司在成立初期的主要产品并非人工智能芯片，近年来方才逐渐布局人工智能芯片产品。例如，英伟达成立于 1993 年，早期专注于图形芯片设计业务；超威半导体成立于 1969 年，在早期阶段主要专注于微处理器市场，主要从事 CPU 的研发与销售；海光信息则成立于 2016 年，成立早期专注于高性能通用处理器 CPU 产品的研发和销售。

随着近年来生成式 AI 和大模型领域的市场需求呈现爆发式增长，部分同行业公司逐步调整产品战略，向人工智能芯片相关产品转型，人工智能芯片相关业务的收入占比呈上升趋势。根据英伟达公开披露的财务数据，英伟达 2020 财年数据中心产品（通用 GPU）占营业收入的比例约为 43%，此时 AI 市场需求尚未全面爆发，英伟达的游戏业务仍占主导地位；2022 财年，生成式 AI 概念初显，英伟达发布 H100 GPU，当年数据中心产品业务超越游戏业务成为其最大的业务板块，2023 年，ChatGPT 的出现推动算力需求激增，英伟达算力 GPU 出货量大幅增加，至 2024 财年，英伟达数据中心产品收入占比已达到约 89%。另根据超威半导体公开披露的财务数据，其数据中心产品近年来收入占比也呈现出持续攀升的趋势，从 2022 财年的约 25.6%上升至 2024 财年的约 48.9%。

发行人成立于 2020 年，创始人团队凭借对人工智能芯片行业未来市场潜力的敏锐洞察，在公司设立之初便明确将人工智能训练推理芯片作为核心产品的研发方向。近年来，随着人工智能芯片市场需求的持续大幅增长及发行人核心产品实现量产，发行人训推一体系列产品的销售收入及占比均呈现显著上升态势。与发行人情况相似的是，寒武纪自 2016 年成立以来，便专注于人工智能芯片产品的研发与技术创新，其应用于人工智能场景的云端产品线收入占比也较高，2024 年占比达到 99.30%。

综上所述，同行业可比公司成立时间普遍早于发行人，发展历程较长，产品类型演变经历了较长时间，因此同行业可比公司与发行人在产品类型丰富度方面的差异具有合理性。随着近年来生成式 AI 和大模型领域的市场需求大幅增长，英伟达、超威半导体等可比公司产品重心也呈现出向人工智能芯片产品转变的特征，人工智能芯片产品销售收入占比呈上升趋势。发行人自成立起即专注于人工智能芯片产品的研发，当前核心产品集中在曦云 C500 系列，与行业整体趋势不存在重大差异。

（2）发行人当前主要产品型号数量与人工智能芯片迭代周期客观规律相符

发行人训推一体系列产品主要型号包括曦云 C500、曦云 C550，智算推理系列产品主要型号包括曦思 N100、曦思 N260，曦云 C 系列（包括曦云 C500 及曦云 C550）收入占比较高，是发行人目前主要销售的产品型号。其他同行业可比公司中，英伟达数据中心产品通用 GPU 主力产品主要包括 B200、H100、A100 等；超威半导体 GPU 产品主要型号包括 MI325X、MI250X 等。英伟达和超威半导体同时期大规模销售的主要为其先进架构和制程的最近两代的产品，例如英伟达目前主力销售产品为 B200、H100，同时英伟达基于满足美国出口管制要求的目的，推出了 A800、H800、H20 等中国市场特供版，但均系基于原有芯片架构的功能受限版（A800 基于 A100、H800 和 H20 基于 H100）。寒武纪、海光信息等境内可比公司成立时间较早，其人工智能芯片产品线已经历多轮迭代，产品型号数量略多于发行人。

通常而言，人工智能芯片的研发和迭代周期受到技术发展、市场需求、供应链稳定性等多重因素影响，同行业公司寒武纪在其招股说明书中披露，其云端智能芯片的迭代周期通常在 2 年左右，边缘智能芯片的迭代周期通常会在 2-3 年左

右。英伟达、超威半导体凭借其绝对领先的行业地位、技术实力和稳定的供应链，产品迭代周期也需 2 年左右。以英伟达为例，其 2020 年推出基于 Ampere 架构的核心产品 A100、2022 年推出基于 Hopper 架构的核心产品 H100、2024 年推出基于 Blackwell 架构的核心产品 B200，平均每 2 年完成一次重大架构升级。报告期内，发行人克服供应链限制等外部不利因素，用 3 年时间即实现两款芯片一次流片并成功量产，核心产品曦云 C500 于 2023 年 6 月首批芯片回片，2024 年 2 月正式量产，正式量产至今尚不足两年，下一代基于国产供应链的曦云 C600 系列已于 2025 年 7 月回片。

综上所述，曦云 C500 系列为发行人当前主力销售产品与芯片产品迭代周期的客观规律以及发行人所处发展阶段相符。

3、报告期内不同产品类型收入结构变化较大、当前收入集中于曦云 C500 的原因

报告期内，发行人营业收入主要来自于训推一体系列 GPU 板卡、训推一体 GPU 服务器、智算推理 GPU 板卡以及 IP 授权。总体来看，报告期内，发行人主要产品类型由智算推理系列向训推一体系列转变，产品的具体交付形式方面，2025 年第 1 季度训推一体 GPU 服务器收入占比降低。具体而言，发行人报告期内不同产品类型收入结构变化较大、当前收入集中于曦云 C500 系列的主要原因如下：

（1）核心产品曦云 C500 系列正式量产叠加生成式 AI 市场需求大幅增长，公司训推一体系列产品收入大幅增长并成为公司主要收入来源

报告期内，发行人训推一体系列产品的收入占比从 2023 年的 30.09% 上升至 2025 年第 1 季度的 97.87%，智算推理系列产品收入占比从 2022 年的 100.00% 下降至 2025 年第 1 季度的 1.25%，整体呈现出训推一体系列产品收入占比大幅上升，而智算推理系列产品收入占比大幅下降的趋势。

公司创始团队拥有深厚的 GPU 技术积累及全流程量产经验，在国际领先 GPU 公司负责设计并量产了多款高性能计算 GPU 和高性能渲染 GPU，具备敏锐的市场洞察能力、前瞻性视野和国内稀缺的万卡大集群落地经验。公司成立之初即已确立了重点优先发展应用于训推一体和通用计算的 GPU 芯片和板卡产品的

发展战略，于 2021 年上半年先后完成了曦思 N100 系列和曦云 C500 系列的研发立项。2023 年 4 月，公司智算推理 GPU 芯片曦思 N100 系列正式量产，曦思 N100 芯片从设计、流片到量产，为公司打通了从研发到生产再到销售的全业务流，完善了产品、生产供应链和销售渠道的储备资源，奠定了公司的经营模式和商业模式，也为后续曦云 C500 系列的研发和量产积累了经验。2023 年以来，随着 AI 大模型的技术突破，生成式 AI 商业化进程加速，市场对训推一体 AI 芯片的需求大幅增加。与此同时，公司应用于训推一体和通用计算的曦云 C 系列 GPU 产品进入研发、量产的关键时期，至 2024 年 2 月曦云 C500 系列正式量产，契合了市场需求从传统的 AI 推理场景向更为复杂的 AI 大模型训练与推理一体化场景转变的时机，曦云 C500 系列产品在较短时间内取得下游客户的认可并实现商业化，出货量和销售收入大幅增加，成为公司目前核心主力产品。

（2）曦云 C500 系列产品契合行业需求，综合性能得到客户广泛认可，产品实现大规模出货

近年来全球人工智能算力建设进入加速通道，叠加持续升级的地缘政治风险促使国产替代进一步加速，大模型训练和推理对高性能 GPU 需求迫切。发行人训推一体 GPU 产品的量产契合了下游爆发增长的需求，该系列产品在性能层面对标英伟达 A100，部分应用场景下的核心指标已比肩国际主流产品；通过兼容 CUDA 生态大幅降低客户应用迁移成本与开发门槛，并凭借自研的 MetaXLink 互连技术，精准响应大型客户对集群计算效能的核心诉求，算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心。发行人 GPU 产品深度赋能众多行业应用场景，已率先布局教科研、金融、交通、能源、医疗健康大文娱等行业，产品赋能真实应用场景的竞争力和交付能力得到充分验证。

曦云 C500 系列产品在单卡性能、集群性能及稳定性等多维度方面具有领先的产品表现力，在产品的通用性、易用性等方面得到了下游客户的广泛认可，自正式量产以来快速实现了商业化、规模化落地。因此，发行人曦云 C500 系列产品卓越的产品力和商业化能力与下游大幅增长的行业需求相契合，是该系列产品收入大幅增长并成为公司主要收入来源的核心因素。

(3) GPU 芯片研发投入大、周期长，发行人当前收入集中于曦云 C500 系列具有合理性

GPU 是现代半导体工业中设计最复杂的芯片之一，其研发过程具有典型的“高投入、长周期”特点，一款全新的 GPU 芯片，从立项、设计、验证到最终成功流片并实现量产，通常需要 2 至 3 年，甚至更长的时间，需要巨额的前期投入和流片成本。因此，发行人在公司业务发展的早期阶段，将相对有限的资金、技术和人力投入到旗舰产品的研发，是发行人综合考虑芯片产品研发“高投入、长周期”的客观规律，同时聚焦高价值 AI 训练核心赛道、顺应市场需求和国家政策方向的战略选择，曦云 C 系列产品上市后出货量和销售收入大幅增长，也进一步表明公司战略选择的正确性。此外，包括英伟达在内，同行业公司产品结构也呈现出同一时期内收入主要来源于旗舰产品的特点，例如英伟达 A100 于 2020 年上市后，是英伟达 2021 财年、2022 财年数据中心业务主要的收入来源；H100 于 2022 年 3 月发布后，根据英伟达 2022 年 Q4 业绩电话会，至 2022 年第四季度，英伟达 H100 收入已实现大幅增长并且已远超过 A100，成为其数据中心业务主要收入来源。

(4) GPU 板卡是发行人主要的产品交付形态

具体到训推一体系列内部的产品结构，2024 年和 2025 年 1-3 月，发行人训推一体 GPU 服务器销售收入占比分别为 28.29%和 0.32%，2025 年第 1 季度训推一体 GPU 板卡收入占比上升至 97.55%，训推一体 GPU 服务器收入占比大幅下降，一方面是因为 2025 年第 1 季度期间较短，发行人当期的收入结构不能完全反映发行人总体的收入结构变化趋势，另一方面根据发行人发展战略，发行人核心产品交付形态将以 GPU 板卡为主。

发行人致力于为国家人工智能公共算力平台、运营商智算平台、商业化智算中心等提供基础算力底座，高性能通用 GPU 板卡产品是发行人的核心产品交付形态，发行人下游产业链条参与者还包括服务器整机厂商、智算中心、大模型厂商等。长期以来，国内人工智能芯片市场被英伟达等境外厂商垄断，国产 GPU 上市销售后，下游客户对于国产 GPU 产品的性能以及在集群环境下的可用性、可靠性尚存在疑虑。发行人训推一体 GPU 板卡推出之后，智算集群客户在服务器层面、集群整体层面均需要发行人作为 GPU 板卡原厂提供直接的技术支持，

同时要求发行人以 GPU 服务器整机的形式交付，以降低国产 GPU 智算集群建设过程中的不确定性。因此，在训推一体 GPU 板卡面世后的早期阶段，发行人为了与客户建立良好的合作关系、更直接地响应客户需求，与国内知名服务器 OEM 厂商合作，委托服务器 OEM 厂商将发行人 GPU 板卡集成成为服务器整机，发行人向智算集群客户直接交付 GPU 服务器整机，以满足客户集群建设的要求。

自训推一体 GPU 板卡产品推出以来，发行人已与整机服务器、操作系统、运维管理平台、主流 AI 框架、主流大模型等上下游生态广泛适配，核心 GPU 板卡产品已应用于多个智算集群并稳定运行，GPU 板卡将是发行人未来主要的产品交付形态。

(5) IP 授权非公司重点发展方向，报告期内 IP 授权业务相关收入占比较低

2023 年度，发行人 IP 授权业务收入占总收入的比例为 43.12%。该收入来源于发行人与单个客户签署的一单 IP 授权协议。该笔授权在 2024 年及 2025 年第一季度所实现的收入为该 IP 授权后续的技术支持服务收入。该笔 IP 授权系发行人在产品开发过程中形成的相关 IP，发行人在与该客户接洽过程中，该客户对于相关 IP 的精度、完整度及性能方面较为认可，与其产品开发需求较为契合，因而发行人与该客户建立合作并向其授权相关 IP。IP 授权业务虽与发行人主营业务密切相关，但根据公司发展战略，IP 授权业务非发行人主要业务类型，也不是发行人主要推广的产品类型，未来发行人 IP 授权业务收入预计仍将维持在较低比例。

综上所述，发行人报告期内收入结构变化较大且当前收入主要集中于曦云 C500 产品，系芯片产品研发客观规律、发行人战略选择、曦云 C500 本身较强的市场竞争力等多方面因素共同导致，具有合理性。

4、其他产品的研发、客户开拓及商业化进展，发行人未来主力产品计划

(1) 其他产品的研发、客户开拓及商业化进展

曦云 C500 系列量产以后，发行人其他在研产品主要包括基于国产供应链的训推一体系列曦云 C600、曦云 C700 等，专为生成式 AI 推理市场设计的 Nx 系列，以及用于图形渲染的曦彩 G 系列产品等，上述产品的研发、客户开拓及商

业化进展情况如下：

产品系列	研发进展/客户开拓及商业化进展
曦云 C600 系列、 C700 系列	1、曦云 C600 系列 曦云 C600 是基于国产先进工艺开发的通用 GPU 芯片，构建了从设计、制造到封装测试的国产供应链闭环。曦云 C600 研发项目于 2024 年 2 月立项，已于 2025 年 7 月完成回片并成功点亮，正在进行功能测试，预计于 2025 年底进入风险量产。曦云 C600 产品从目前的工程验证测试到进入风险量产阶段，尚需完成如下工作： （1）测试验证与性能调优 系当前曦云 C600 研发重点，需完成所有的工程验证测试（EVT），并深入开展设计验证测试（DVT）和系统级验证，确保芯片在各种工况下的稳定性、产品可靠性和兼容性。同时，需进行持续的性能调优，以充分发挥硬件潜力，确保达到对标国际旗舰产品的性能指标 （2）软件栈与生态完善 加速 MXMACA 软件栈的优化与适配工作。包括完善编译器、驱动、通信库，确保其对主流 AI 框架和深度推理模型的良好支持，并提升其成熟度和稳定性 （3）供应链与量产准备 构建国产供应链闭环，确保晶圆制造、封装测试等环节的产能、良率和成本控制能满足风险量产及后续规模化生产的需求 （4）客户导入与样板点建设 向潜在客户提供测试样片、技术支持，并共同构建标杆性的行业应用解决方案，有效推动意向订单的获取。曦云 C600 产品已在 2025 年世界人工智能大会（WAIC）正式发布，公司在国家人工智能公共算力平台、互联网、运营商、金融、能源等行业加大了曦云 C600 的推介和交流力度，同时在服务器整机产品的适配和 NRE 开发方面也加大了推广，目前已有多家互联网厂商、金融机构、服务器 OEM 厂商表达对曦云 C600 产品的测试和适配意愿 2、曦云 C700 系列 曦云 C700 是基于国产先进工艺开发的下一代通用 GPU 芯片，拥有更高的性能。曦云 C700 研发项目于 2025 年 4 月立项，目前正处于软硬件购置和产品设计开发阶段，芯片的核心设计、功能验证已大部分完成，正在进行更深入的性能调优，同时基于先进国产工艺，进行物理实现的迭代和优化 曦云 C600 及曦云 C700 两款产品均依托于公司自主创新的核心 GPU IP 及 MXMACA 软件栈，可提供混合精度算力支持，并搭载超高带宽显存，深度融合计算密度与互连系统优化技术，支持从单卡到超大规模集群的全场景加速需求，是针对 AI 训练、推理及通用计算的高效解决方案
曦思 Nx 系列	公司下一代基于先进封装技术的云端大模型推理芯片，专为生成式 AI 推理市场设计。该芯片计划采用公司最新的第三代指令集和先进封装工艺，通过超高显存带宽和大容量显存提高 token 吞吐量，同时保持低功耗和高能效，大幅降低 AI 推理成本 曦思 Nx 系列是本次募投项目之一，项目完成后，相关产品可以广泛应用于国内商业化智算中心，为其提供高性价比的大模型推理解决方案；或者用于 AI PC 等端侧低延迟、独立用户推理场景，支持从百亿到万亿各种参数规模的模型推理。同时，也可以与公司的训练芯片进行互连，提升协同效率
曦彩 G100 系列	面向图形处理场景，内置性能强大的图形处理器，可广泛应用于云游戏、数字孪生、云渲染、影视动画和专业制图等场景，主要应用于云端及边

产品系列	研发进展/客户开拓及商业化进展
	端图形处理，发行人目前已完成曦彩 G100 GPU IP 的设计和验证工作

(2) 发行人未来主力产品计划

短期来看，发行人当前主力产品曦云 C500 系列市场需求良好并持续稳定出货，预计未来 1-2 年内仍将是公司重要的收入来源。截至本回复出具之日，发行人基于国产供应链的曦云 C600 系列已回片，产品良率及各项测试指标良好，正在稳步推进量产进程，预计将于 2025 年底进行风险量产。曦云 C600 系列正式量产后，将成为发行人训推一体系列下一代主力产品，与现有曦云 C500 系列形成优势互补，丰富发行人产品矩阵。

长期来看，随着产品迭代进入“量产一代、在研一代、规划一代”的稳定周期，发行人基于国产供应链的曦云 C 系列产品预计将实现稳定的产品迭代并成为公司稳定的收入来源。发行人第三代高性能通用 GPU 芯片（曦云 C700）研发项目已于 2025 年 4 月立项，为发行人本次募投项目之一，将成为曦云 C600 系列后下一代主力产品。

此外，由于大规模 AI 调用、复杂推理、多步骤推理场景等推理需求的爆发，促使推理算力成为推动产业发展的第二重驱动力，并下沉至不同的应用前景，是未来极具发展潜力的巨大市场。发行人高度重视生成式 AI 推理市场，持续加大国产高能效推理芯片的研发力度，本次募投项目“新一代人工智能推理 GPU 研发及产业化项目”实施完成后，未来曦思 N 系列后续迭代产品的销售将成为发行人营业收入的有力补充。发行人亦十分重视图形渲染芯片日益增长的市场需求，将持续分配资源投入曦彩 G 系列产品的持续研发，未来将形成覆盖训推一体、智算推理和图形渲染等多个领域的产品矩阵。

（二）不同产品类型发行人与客户具体合作模式，如智算集群等；未来拟主要推广的产品类型或合作模式，服务器或集群业务是否会与现有客户形成竞争影响合作稳定性；目前交付的智算集群情况、涉及的产品自研外购情况、对应直接客户和终端客户、应用领域等

1、不同产品类型发行人与客户具体合作模式，未来拟主要推广的产品类型或合作模式，服务器或集群业务是否会与现有客户形成竞争影响合作稳定性

（1）报告期内不同产品类型与客户存在的具体业务模式

报告期内，发行人不同产品类型与客户的主要合作模式如下：

序号	产品类型	直接客户类型	具体合作模式
①	训推一体 GPU 板卡	直销客户-服务器 OEM 厂商	发行人向服务器 OEM 厂商交付 GPU 板卡，以服务器 OEM 厂商出具的签收单确认收入。服务器厂商采购发行人 GPU 板卡产品后组装为 GPU 服务器后对外销售
②		直销客户-其他	发行人向客户交付 GPU 板卡，发行人以客户出具的签收单确认收入，或根据部分客户在销售合同中的要求，以客户出具的集群验收单/验收报告确认收入
③		经销客户	买断式经销，发行人向经销商交付 GPU 板卡，以经销商出具的签收单确认收入。经销商采购发行人板卡产品后向下游客户销售，部分经销商采购发行人产品也存在测试自用或自建集群的情形
④	训推一体 GPU 服务器	直销客户-智算集群	发行人委托服务器厂商将 GPU 板卡加工为 GPU 服务器，或发行人外购服务器机头等组件后自行组装为 GPU 服务器，发行人向智算集群客户交付 GPU 服务器，以客户出具的集群验收报告确认收入
⑤	智算推理 GPU 板卡	直销客户	发行人交付 GPU 板卡，以客户出具的签收单确认收入
⑥		经销客户	买断式经销，发行人交付 GPU 板卡，以客户出具的签收单确认收入
⑦	IP 授权	直销客户	发行人交付 IP，以客户的验收邮件确认收入

1) 训推一体 GPU 板卡的主要合作模式

报告期内，发行人训推一体 GPU 板卡销售业务包括直销模式和经销模式。

直销模式情形下，客户类型主要包括服务器 OEM 厂商和其他类客户，对于服务器厂商客户，发行人需完成服务器 OEM 厂商的标准适配流程，与服务器 OEM 厂商签署 GPU 板卡销售合同，服务器 OEM 厂商采购发行人 GPU 板卡产品并组装为 GPU 服务器后销售给最终用户。对于其他类客户，发行人与客户签署 GPU 板卡的销售合同，发行人以客户出具的签收单确认收入，或根据部分客

户在销售合同中的要求，以客户出具的集群验收单/验收报告确认收入。

经销模式情形下，发行人向经销商交付 GPU 板卡，经销商向其下游客户销售，终端客户包括智算集群以及教科研、金融等领域的垂类用户等。此外，部分经销商基于其自身需求，在采购发行人 GPU 板卡后也存在用于自建集群、自用测试等情形。发行人经销模式均为买断式经销，经销商收到板卡产品后出具签收单，发行人依据签收单确认收入。

2) 训推一体 GPU 服务器的主要合作模式

报告期内，发行人训推一体 GPU 服务销售业务主要为直销模式，客户类型主要为智算集群客户。发行人委托服务器 OEM 厂商将发行人 GPU 板卡产品组装为 GPU 服务器整机，或发行人外购服务器机头等配件后组装为 GPU 服务器整机，最终由发行人直接向智算集群客户交付 GPU 服务器整机。如前所述，该种业务模式系基于早期部分客户的特殊要求，即部分客户出于降低集群建设的不确定性等方面考虑，要求发行人直接交付 GPU 服务器整机以获得发行人作为 GPU 原厂的直接技术支持。

根据具体合同签署方式的不同，该种业务模式可细分为两种类型：其一为发行人与客户签署 GPU 服务器销售合同，并根据合同约定以客户出具的集群验收报告确认收入，例如上海算丰、香港算丰即属于该种业务模式；其二为发行人与客户签署集群销售合同，向客户交付 GPU 服务器以及配套集群设备，以客户出具的集群验收报告确认收入，报告期内，发行人与汇天网络发生一笔集群整体交付的 GPU 服务器销售业务。

3) 智算推理 GPU 板卡的主要合作模式

报告期内，发行人智算推理 GPU 板卡销售业务主要为直销模式和经销模式，发行人向直销客户或经销客户交付 GPU 板卡，以客户出具的签收单确认收入。

4) IP 授权业务的主要合作模式

报告期内，发行人向客户交付 IP、收到客户出具的验收邮件后确认收入。同时，根据 IP 授权合同的约定，发行人在交付 IP 后两年内向客户提供技术支持服务。

(2) 发行人未来拟主要推广的产品类型或合作模式

自成立以来，发行人专注于高性能 GPU 芯片及计算平台的研发与设计，直接下游为服务器厂商、集成商或智算中心建设方，下游应用领域包括教科研、金融、交通、能源、医疗健康、大文娱等。

在产品交付形态方面，发行人未来核心产品交付形态为 GPU 板卡，报告期内 GPU 服务器的销售主要系基于部分客户的特殊要求，未来随着发行人产品市场竞争力提升，下游客户对发行人产品的认可度进一步提升，GPU 服务器整机的销售预计占营业收入的比例较低。

在与客户具体合作模式方面，直销客户将以服务器 OEM 厂商为主，发行人向服务器 OEM 厂商出货 GPU 板卡，由服务器 OEM 厂商按照其标准流程组装为 GPU 服务器向其下游智算集群、集成商交付。发行人未来将减少直接与智算集群客户签署销售合同并承担集群验收义务的业务模式，有利于明确发行人售后维保边界，使得发行人更加专注于 GPU 板卡产品及相关软件生态的核心业务。此外，发行人亦十分重视经销商渠道网络的建设，经销模式将是发行人未来销售模式的重要组成，发行人 GPU 板卡产品具有功能通用性高、下游应用领域广泛等特点，通过经销模式，发行人可以充分发挥经销商区位优势、行业资源的优势，扩大终端客户覆盖范围并提高服务能力。

综上所述，发行人未来主要推广的产品类型为 GPU 板卡，将采取直销和经销并重的销售模式，持续拓展并加深与服务器 OEM 厂商、经销商客户的合作。

(3) 服务器或集群业务是否会与现有客户形成竞争影响合作稳定性

发行人核心产品为 GPU 板卡，与服务器 OEM 厂商、智算集群客户属于行业上下游关系，报告期内发行人直接向智算集群客户销售的 GPU 服务器或交付的集群系根据该等客户的特殊要求，委托服务器厂商加工或外购服务器机头、配件后组装为 GPU 服务器等方式完成交付。如前所述，发行人未来 GPU 服务器或集群业务将逐渐减少，发行人核心的交付产品类型为 GPU 板卡，与现有客户不存在形成竞争从而影响合作稳定性的情形。

2、目前交付的智算集群情况、涉及的产品自研外购情况、对应直接客户和终端客户、应用领域等

一般而言，智算集群由智算硬件基础设施和智算软件基础设施组成。智算集群硬件基础设施通常包括 GPU 服务器、存储服务器、管理服务器、交换机、防火墙等各类计算、存储、网络硬件设备，其中搭载 GPU 板卡的 GPU 服务器是智算集群的核心器件，GPU 板卡通常是智算集群的最小交付单元，同时 GPU 板卡需配套相应的软件栈（包括驱动、编译器、数学库等）方能充分发挥硬件的功能和性能。智算集群软件基础设施主要为智算中心管理软件平台，通常包括算力资源管理调度平台、人工智能平台、算力运营平台和运维监控平台等。从价值构成来看，通常情况下，智算集群硬件基础设施采购成本占智算集群整体采购成本的比例约为 95%（其中 GPU 服务器占比在 70%以上），智算软件基础设施占比约为 5%。

从功能划分来看，智算集群可分为算力资源区和管理区。算力资源区以高性能通用 GPU 服务器为核心，并配套使用高性能 RDMA 网络、高性能分布式存储子系统等设备，组建人工智能算力资源池，构建可满足通用计算、人工智能大模型训练与推理、数据处理等需求的人工智能算力集群；管理区主要部署管理服务器、登录服务器、管理接入交换机、管理汇聚交换机等设备，并在管理服务器上部署虚拟化和云原生容器化融合底座、算力资源管理调度平台、人工智能平台、算力运营平台和运维监控平台，提供智算中心 CPU 及 GPU 资源调度、存储资源调度、算力资源状态监控、大模型训练与推理 workflow、多租户隔离、租户计量计费等相关功能。

发行人致力于自主研发全栈高性能 GPU 芯片及计算平台，主营业务是研发、设计和销售应用于人工智能训练和推理、通用计算与图形渲染领域的全栈 GPU 产品，并围绕 GPU 芯片提供配套的软件栈与计算平台。发行人是国内少数真正实现千卡集群大规模商业化应用的 GPU 供应商，发行人深度构建“1+6+X”生态与商业布局，算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心，区域横跨北京、上海、杭州、长沙、中国香港等地区，为众多人工智能大模型公司、智算科技企业等提供算力支撑，下游应用领域覆盖教科研、金融、交通、能源、医疗健康、大文娱等行业。

发行人核心产品交付形态是 GPU 板卡，MXMACA 软件栈系提高 GPU 易用性的核心，主要与发行人的 GPU 产品配套使用，未单独对外销售。报告期内，在训推一体 GPU 板卡面世早期，为满足部分客户的需求，发行人存在向部分客户交付 GPU 服务器的情形，其中仅有一单以集群整体形式交付的大规模智算集群，即汇天网络项目 F（一期）。

截至报告期末，发行人收入确认金额在 1,000 万元以上，且终端应用于智算集群的 GPU 板卡/服务器销售情况如下：

序号	直接客户	终端客户	具体合作模式	产品类型、型号	集群项目名称	发行人在集群建设中发挥的主要作用	产品自研或外购情况	发行人收入确认时间
1	新华三信息技术有限公司	终端客户 1	①	曦云 C500 板卡	项目 A（一期）	GPU 板卡供应	自研	2024 年
					项目 A（二期）	GPU 板卡供应	自研	2024 年
2	苏州瑞芯智能科技有限公司	终端客户 2	③	曦云 C550 板卡	项目 B	GPU 板卡供应	自研	2024 年
3	超讯通信股份有限公司	终端客户 3	③	曦云 C500 板卡	项目 C（一期）	GPU 板卡供应	自研	2025 年一 季度
4	上海源庐加佳信息科技有限公司	上海源庐加佳信息科技有限公司 ^{注 1}	②	曦云 C500 板卡	项目 D（一期）	GPU 板卡供应，根据销售合同的约定，以集群验收报告作为板卡销售收入确认的依据	自研	2024 年
					项目 D（二期）		自研	2025 年一 季度
5	上海算丰信息有限公司	上海算丰信息有限公司 ^{注 2}	④	曦云 C500 服务器	项目 E（一期）	GPU 服务器供应，根据销售合同约定，负责产品安装、上架、调试和试运行，达到集群验收标准	服务器中 GPU 板卡为自研、服务器其他配件为外购	2024 年
6	汇天网络科技有限公司	汇天网络科技有限公司	④	曦云 C550 服务器及相关集群设备	项目 F（一期）	GPU 服务器及集群其他软硬件设备供应，并负责集群整体交付。组织集群的搭建和调试工作，协助客户完成集群测试验收	服务器中 GPU 板卡为自研，服务器其他配件、集群其他软硬件为外购	2024 年
7	上海华研企业发展（集团）有限公司	终端客户 4	③	曦云 C500 板卡、曦云 C550 板卡	项目 G	GPU 板卡供应	自研	2024 年
8	浙江武珞智慧城市技术有限公司	终端客户 5	③	曦云 C500 板卡	项目 H	GPU 板卡供应	自研	2024 年

序号	直接客户	终端客户	具体合作模式	产品类型、型号	集群项目名称	发行人在集群建设中发挥的主要作用	产品自研或外购情况	发行人收入确认时间
9	中仪英斯泰克进出口有限公司	中仪英斯泰克进出口有限公司	④	曦云 C500 服务器及配套设备	项目 I	GPU 服务器及配套集群设备供应，根据销售合同约定，负责产品安装、上架、调试和试运行，达到集群验收标准	服务器中 GPU 板卡为自研、服务器其他配件为外购	2024 年
10	新华三信息技术有限公司	终端客户 6	①	曦云 C550 板卡	项目 J	GPU 板卡供应	自研	2025 年一 季度
11	北京创世云科技股份有限公司	终端客户 7	③	曦云 C500 板卡	项目 K	GPU 板卡供应	自研	2025 年一 季度
12	香港算丰信息有限公司	香港算丰信息有限公司	④	曦云 C500 服务器	项目 L	GPU 服务器供应，根据销售合同约定，负责产品安装、上架、调试和试运行，达到集群验收标准	服务器中 GPU 板卡为自研，服务器其他配件为外购	2024 年
13	浙江武珞智慧城市技术有限公司	终端客户 8	③	曦云 C500 板卡	项目 M	GPU 板卡供应	自研	2024 年
14	铎云（北京）半导体有限公司	终端客户 9	②	曦云 C500 板卡	项目 N	GPU 板卡供应	自研	2024 年

注 1：项目 D 运营方为上海源庐加佳信息科技有限公司和其同一控制下的上海加佳智云数字科技有限公司

注 2：项目 E（一期）运营方为上海优算丰信息技术有限公司，系上海算丰信息有限公司的关联公司

除上述在报告期内已确认收入的智算集群外，发行人在 2025 年第 2 季度已交付并确认收入的主要集群包括：汇天网络科技有限公司“项目 F（二期）”（128 台 GPU 服务器）、上海算丰信息有限公司“项目 E（二期）”集群（160 台 GPU 服务器）等。

（三）发行人产品终端客户类型，客户采购后的用途及使用点亮情况，各地对采购销售国产 GPU 的支持政策、客户采购公司产品的主要驱动因素。结合发行人报告期内产品结构变化、客户复购率及持续采购需求、当前国产 GPU 商业化拓展的难点及公司竞争优劣势等，说明公司主要发展方向及竞争策略

1、发行人产品终端客户类型，客户采购后的用途及使用点亮情况

发行人主要产品 GPU 板卡及服务器的终端客户即为该等算力基础设施或组件的最终资产持有方，其采购发行人产品的用途除少量零星采购系自用测试之外，大部分则用于搭建具有一定规模的算力集群，以满足 AI 训练/推理、科研等各种计算任务的需要。根据集群用途、建设方背景的不同，发行人产品目前所应用的主要集群项目可进一步分为用于对外提供算力服务的国家人工智能公共算力平台、运营商智算平台和商业化智算中心，以及客户自用的算力集群；而从最终的算力使用方来看，主要包括政府部门、高校、科研机构、金融机构、AI 大模型厂商、中小企业等各类用户，较为分散。

经统计使用发行人报告期内销售产品搭建算力集群且集群规模在 128 卡及以上的主要终端项目情况，具体如下表所示：

单位：张

直接客户	销售产品	数量	终端项目	类型	地点	目前状态	收入确认时间
超讯通信股份有限公司	曦云 C500 板卡	4,096	项目 C（一期）	运营商智算平台	青海	正在进行建设， 预计 2025 年 9 月投入使用	2025 年 1-3 月
上海源庐加佳信息科技有限公司	曦云 C500 板卡	1,024	项目 D（一期）	商业化智算中心	上海	已投入使用	2024 年
		2,048	项目 D（二期）				2025 年 1-3 月
苏州瑞芯智能科技有限公司	曦云 C550 板卡	3,072	项目 B	国家人工智能公共算力平台	杭州	2,048 卡已投入使用； 剩余 1,024 卡由于集群建设 方案调整，预计 2025 年底前投 入使用	2024 年
新华三信息技术有限公司	曦云 C500 板卡	515	项目 A（一期）	国家人工智能公共算力平台	上海	已投入使用	2024 年
		2,052	项目 A（二期）				2024 年
上海算丰信息有限公司	曦云 C500 服务器	1,280	项目 E（一期）	商业化智算中心	上海	已投入使用	2024 年
汇天网络科技有限公司	曦云 C550 服务器	1,024	项目 F（一期）	商业化智算中心	北京	已投入使用	2024 年
浙江武珞智慧城市技术有限公司	曦云 C500 板卡	1,024	项目 H	商业化智算中心	上海	已投入使用	2024 年
上海华研企业发展（集团）有限公司	曦云 C550 板卡	1,024	项目 G	商业化智算中心	香港	一期 256 卡已投入使用； 二期 768 卡因机房改造 未完成，预计 2025 年 10 月投 入使用	2024 年
北京创世云科技股份有限公司	曦云 C500 板卡	312	项目 K	终端客户 7 自用算力集群	深圳	已投入使用	2025 年 1-3 月
新华三信息技术有限公司	曦云 C550 板卡	280	项目 J	终端客户 6 自用算力集群	杭州	已投入使用	2025 年 1-3 月

直接客户	销售产品	数量	终端项目	类型	地点	目前状态	收入确认时间
司							
铎云（北京）半导体有限公司	曦云 C500 板卡	256	项目 N	商业化智算中心	北京	已投入使用	2024 年
浙江武珞智慧城市技术有限公司	曦云 C500 板卡	240	项目 M	某项目自用算力集群	上海	已投入使用	2024 年
香港算丰信息有限公司	曦云 C500 服务器	160	项目 L	商业化智算中心	香港	已投入使用	2024 年
北京创世云科技股份有限公司	曦云 C500 板卡	128	集群 O	商业化智算中心	贵州	已投入使用	2024 年
中仪英斯泰克进出口有限公司	曦云 C500 服务器	128	项目 I	自用算力集群	北京	已投入使用	2024 年

注：为便于统计分析，上表中涉及发行人以服务器形态进行产品交付的，均已折算成对应的 GPU 板卡数量。

由上表可见，截至本回复出具日，除项目尚未建设完毕等情形外，发行人报告期内销售产品已在上述主要终端项目中实际投入使用，运行良好。由上述项目所形成的发行人 2024 年收入合计 63,601.13 万元，其中目前已投入使用的发行人产品对应收入 55,259.54 万元，占发行人 2024 年主营业务收入的比例分别为 85.73%和 74.48%；由上述项目所形成的发行人 2025 年 1-3 月收入合计 24,939.26 万元，其中目前已投入使用的发行人产品对应收入 12,250.99 万元，占发行人 2025 年 1-3 月主营业务收入的比例分别为 77.83%和 38.23%，当期销售产品已投入使用比例较低主要系项目 C（一期）项目（4,096 卡）尚未完成建设所致，该项目预计将于 2025 年 9 月投入运营。

2、各地对采购销售国产 GPU 的支持政策，客户采购公司产品的主要驱动因素

经公开检索，发行人终端客户主要集群项目所在地区发布的有关政策或政府文件中，涉及针对智算中心等主体采购并使用国产算力芯片给予支持的相关内容具体如下：

地区	发布机构	文件名称	内容摘要
北京	北京市经济和信息化局、北京市通信管理局	北京市算力基础设施建设实施方案（2024-2027 年）	对采购自主可控 GPU 芯片开展智能算力服务的企业，按照投资额的一定比例给予支持，加速实现智算资源供给自主可控
北京	北京经济技术开发区管理委员会	北京经济技术开发区关于加快打造 AI 原生产业创新高地的若干政策	对于云服务企业采购国产算力芯片企业相关产品，并在经开区算力调度平台上架对外提供服务的，给予购买合同中的实际采购国产智算芯片服务器部分金额最高 10% 的补贴，每个项目累计不超过 1000 万元
上海	上海市经济和信息化委员会	上海市进一步扩大人工智能应用的若干措施	对自主智能算力设施部署项目，给予最高 10% 建设支持，加速培育人工智能自主生态
上海	上海市通信管理局等 11 个部门	上海市智能算力基础设施高质量发展“算力浦江”智算行动实施方案（2024-2025 年）	对于以智能算力为主，积极部署国产化智能算力资源的相关重点拟建数据中心，适当放宽项目能耗指标要求
上海	中国（上海）自由贸易试验区临港新片区管理委员会	中国（上海）自由贸易试验区临港新片区进一步集聚发展新型算力赋能新质生产力若干政策	鼓励临港新片区企业、高校、科研机构采购并使用关键核心技术自主可控的智算芯片，其在功能、性能、功耗指标上与国际同类产品基本相当，或部分指标达到国际领先水平。依据实际采购智算芯片的发票额，按照不高于 10% 予以补贴，总金额不超过 1000 万元
浙江	浙江省人民政府	关于支持人工智能创新发展的若干措施	倡导政府部门、算力中心、经营主体采购或应用国产软硬件
浙江	杭州市经济和信息化局	市经信局关于市政协十二届三次会议第 23 号提案的答复	鼓励建设供应链安全可控的算力集群，形成市场充分竞争和供应链安全可控的算力体系
青海	青海省数据局等 5 个部门	青海省促进绿色算力产业发展若干措施	对新建、扩建数据中心及绿色算力产业重点项目，按照实际建成的标准机架数和购买算力设备实际投资额，按照投资金额比例不超过 10% 予以奖补
贵州	贵州省大数据发展管理局等 8 个部门	关于促进全国一体化算力网络国家（贵州）枢纽节点建设的若干激励政策	打造人工智能算力高地，支持在贵安新区新建智算中心项目，优先对国产化比例高的项目给予奖励

由上表可见，目前国内各地政府均鼓励倡导建设自主可控的算力基础设施，

支持各类主体采购及使用国产算力芯片，并出台了一些具体补贴政策，但从相关政策所述内容来看，补贴力度总体上较为有限（占项目投资额比例普遍不超过10%），因此政府补贴并非影响投资方采购决策的主导因素。实际上，驱动终端客户采购发行人产品的主要原因包括其自身 AI 场景的算力需求、国际贸易形势、国产替代要求、产品性能及易用性、可靠性等多方面综合考虑，具体说明如下：

（1）发行人产品可满足各场景下的高效算力需求

互联网、教科研、金融、交通、能源等行业均有丰富的 AI 应用场景，终端客户采购算力芯片时，首先是考虑产品的整体性能水平，确保能够实现高效的 AI 大模型训练与推理，同时可能需要兼顾数据处理、科学计算等多种计算任务需求。发行人的曦云 C500 系列训推一体 GPU 具备国内领先的计算能力，拥有丰富的标量、矢量和张量计算单元，支持 FP64、FP32、TF32、FP16、BF16、INT8 等混合精度计算，能够较好地满足人工智能训练/推理、通用计算、科学计算等多元化算力需求，综合性能对标英伟达 A100，并且相较英伟达目前国内可售产品 H20 在大模型训练场景下更具性能优势。

（2）发行人产品符合自主可控的国家发展战略

随着国际贸易形势日益严峻、中美科技博弈持续升级，依赖海外巨头进口 GPU 等 AI 芯片存在着较大的潜在断供风险，加速推进国产替代已成为我国人工智能产业发展的顶层战略和行业共识。例如以项目 A、项目 B 等为代表的国家人工智能公共算力平台均有明确的国产算力部署要求及建设规划，其他一些对安全性要求较高的客户如政府、银行、央国企等也需要采购国产算力芯片。发行人全栈产品基于自主研发的 GPU IP、GPU 指令集和架构，核心技术自主可控，高度契合该等客户的国产化采购需求，并且英伟达产品反复被禁售也进一步为发行人带来了巨大的市场机会。

（3）发行人产品兼容主流软件生态，易用性优势显著

发行人全系列产品采用统一的 GPU 计算架构，兼具通用性和高性能，同时基于自研指令集和 GPU 并行计算引擎打造了 MXMACA 软件栈，能够高度兼容国际主流的 CUDA 生态。对于终端用户而言，兼容 CUDA 可以显著降低现有应用迁移的代码工作量和因兼容性问题导致的开发适配成本、运维成本，以及能够

快速跟踪并适配兼容实时更新的大模型算法及算子。发行人产品的生态兼容性、易用性和迁移效率目前在国内竞品中处于领先地位，获得下游用户的广泛认可。

（4）发行人产品具备较强的集群综合性能

以商业化智算中心为典型的算力服务商客户对算力集群的可扩展性和线性度、利用率、稳定性等有较高要求，集群性能是其选购产品时的重要考虑因素之一。发行人产品拥有国内稀缺的高带宽、超多卡互连能力，基于自主研发的 MetaXLink 高速互连技术可实现 2-64 卡多种互连拓扑及超节点结构，并达到与英伟达 4nm 制程旗舰产品（H200）相当的互连带宽，满足从中小型训练到超大规模集群的差异化需求。此外，使用发行人产品搭建的千卡集群在实际运营中性能表现良好，集群线性度、计算资源利用率、稳定性和异构混训能力行业领先。

3、结合发行人报告期内产品结构变化、客户复购率及持续采购需求、当前国产 GPU 商业化拓展的难点及公司竞争优势等，说明公司主要发展方向及竞争策略

（1）发行人报告期内产品结构变化、客户复购率及持续采购需求

报告期初，发行人仅有一款产品即曦思 N100 智算推理 GPU，主要应用于智慧城市、智慧交通、视频处理等领域。随着发行人第二款产品曦云 C500 于 2023 年 6 月回片、2024 年 2 月正式量产，以及行业趋势和市场需求从传统人工智能向生成式人工智能转变，曦云 C500 系列训推一体 GPU 成为发行人开拓市场的最核心产品。报告期各期，发行人主营业务收入中训推一体产品（含板卡及服务器）的收入占比分别为 0%、30.09%、97.28%和 97.87%，呈大幅上升趋势。

曦云 C500 系列推出市场后，凭借优秀的产品性能、易用性和集群实测表现赢得了客户认可，目前已在多个国家人工智能公共算力平台、运营商智算平台和商业化智算中心实现大规模部署。从报告期内发行人直接客户层面的变化情况来看，2023 年、2024 年和 2025 年 1-3 月由发行人存量客户贡献的收入金额分别为 0 万元、11,861.21 万元和 30,469.91 万元，占各期发行人主营业务收入的比例分别为 0%、15.99%和 95.10%（详见本回复“3.关于尚未盈利/一、/（三）/2、/（2）/4）新客户验证及拓展情况”相关内容说明），客户复购率持续提升。考虑到发行人直接客户中存在较多经销商和服务器厂商，其通常对应多个终端客户，同时

大型智算集群项目具有多批次、分阶段的采购特点，而不同批次采购可能发生在同一年度或期间内，因此从发行人主要终端客户角度按项目口径（对应直接客户的单笔或多笔采购订单）统计客户复购情况相对更具有参考性，具体如下表所示：

单位：张

直接客户	终端项目	销售产品	数量	订单时间
新华三信息技术有限公司	项目 A（一期）	曦云 C500 板卡	515	2024 年 4 月
	项目 A（二期）	曦云 C500 板卡	2,052	2024 年 10 月
	终端客户 1 后续项目	曦云 C550 板卡	4,096	2025 年 4 月
上海算丰信息有限公司、香港算丰信息有限公司	项目 E（一期）	曦云 C500 服务器	1,280	2023 年 12 月
	项目 L	曦云 C500 服务器	160	2024 年 6 月
	项目 P	曦云 C550 板卡	328	2025 年 5 月
	项目 E（二期）	曦云 C500 服务器	1,280	2025 年 6 月
汇天网络科技有限公司	项目 F（一期）	曦云 C550 服务器	1,024	2024 年 8 月
	项目 F（二期）	曦云 C550 服务器	2,048	2025 年 2 月
上海源庐加佳信息科技有限公司	项目 D（一期）	曦云 C500 板卡	1,024	2024 年 7 月
	项目 D（二期）	曦云 C500 板卡	1,024	2024 年 8 月
		曦云 C500 板卡	1,024	2024 年 9 月
超讯通信股份有限公司	项目 C（一期）	曦云 C500 板卡	4,096	2024 年 12 月
	项目 C（二期）	曦云 C500 板卡	4,096	

注：为便于统计分析，上表中涉及发行人以服务器形态进行产品交付的，均已折算成对应的 GPU 板卡数量。

由上表可见，截至本回复出具日，基于前期交付产品的良好使用反馈，发行人主要终端客户普遍在其后续项目（如有）中对发行人产品进行了复购并进一步加大了采购规模。国家人工智能公共算力平台、运营商智算平台和商业化智算中心等算力底座客户是发行人重要的基本盘，发行人一方面密切跟进存量客户的后续项目建设规划，巩固并增加在其扩建项目中所占份额，同时依托市场标杆项目的示范效应和产品口碑，持续开拓该领域的新客户。

目前，曦云 C500 系列仍延续着较好的销售态势，在手订单充沛，下游客户持续采购需求旺盛。随着发行人下一代国产供应链产品曦云 C600 正式推出并稳步启动量产，将与现有曦云 C500 系列形成优势互补，丰富发行人产品矩阵，为客户提供更多灵活选择，满足相关领域的信创需求。在未来较长一段时间内，曦云 C 系列训推一体 GPU 都将是发行人的核心产品线。

(2) 当前国产 GPU 商业化拓展的难点

从行业整体来看，当前国产 GPU 在商业化拓展过程中的主要难点如下：

1) 面临国际巨头的竞争压力，产品性价比存在差距

当前国内 GPU 厂商以初创企业为主，在技术积累、人才密度、经营规模及资金实力各方面与英伟达等国际巨头存在明显差距，叠加制程工艺限制，短期内难以突破其高性价比产品构筑的全球市场壁垒。国内市场方面，由于国际巨头的高端产品被禁售，其可售产品与国产 GPU 的差距在不断被缩小，但考虑到海外政策的不确定性，未来国产 GPU 仍面临和国际巨头高端产品竞争的可能性。

2) 商业客户严苛的服务水平要求及准入门槛拉长产品导入周期

为保障既有业务稳定运行和未来算力长期供给，以互联网企业为代表的商业客户对 GPU 产品的综合性能及稳定性要求极高，而国产 GPU 因新产品众多、综合性能与兼容性需经历多轮严格测试，导致客户采购决策周期较长，商业化进程迟缓。目前国内互联网企业采购 GPU 仍以可售的国际产品为主，国产 GPU 占比较低且多数集中于推理场景，训练场景需要更大规模集群测试的时间与资源投入，因此国产 GPU 产品导入较慢。

3) 国产 GPU 软件生态相对薄弱，客户迁移成本较高

当前人工智能技术尚处于快速迭代阶段，不断涌现的各种大模型算法对于底层算力持续提出了新的需求，然而人工智能芯片的研发周期相对较长，软件适配层面更是需要时间积累。虽然国产 GPU 依托通用型架构优势，具备兼容 CUDA 的理论基础及路径可行性，但生态适配、迭代和培育建设是一项长期、复杂的系统性工程，短时间内预计无法改变国产 GPU 相较国际产品生态薄弱的困境，增加了国产 GPU 商业化拓展的难度。

4) 客户关于能否保障供应链持续稳定有所担忧

近年来，随着地缘政治持续紧张、国际形势愈发复杂，全球半导体产业链核心环节受到较大的行政干预和出口管制。我国集成电路行业整体起步较晚，尤其是在生产制造环节目前与境外顶尖的设备、技术水平还存有显著差距，国产供应商仍需努力追赶。对于互联网企业等部分客户，国产 GPU 的供应链稳定性也是

影响其采购决策的主要考虑之一。

(3) 公司竞争优劣势

发行人秉承严谨务实的工程师文化，在初期的商业拓展过程中，坚持扎实做好每一次具体场景的产品测试，依靠过硬的技术指标及测试数据，逐步获得客户的认可和信任，目前已与多个国家人工智能公共算力平台、运营商智算平台和商业化智算中心建立紧密业务合作，行业知名度与品牌影响力不断提升。发行人产品在通用性、单卡性能、集群性能和 CUDA 兼容等方面均具有国内领先优势，详见本题回复前述内容中关于“客户采购公司产品的主要驱动因素”相关说明，除此之外发行人在市场开拓层面的其他一些竞争优势补充说明如下：

首先，基于对包括 GPU IP、GPU 指令集和架构在内的一整套 GPU 底层核心技术的完全自主掌握，发行人在应对国家人工智能公共算力平台等标杆客户组织的每次产品性能测试比选过程中，能够根据客户的技术标准要求和考察内容变化，快速迭代、优化公司产品的软硬件技术组合方案。这也是相较国内其他友商，发行人持续深耕通用 GPU 领域的技术实力和纵深度体现。

其次，在商业化算力服务领域，发行人可以提供高性价比、多形态（PCIe 板卡、OAM 模组，支持各类组网模式和散热模块）的各种产品选择，帮助商业化智算中心实现更优 PUE（数据中心电源使用效率）的运营水平。此外，发行人联合各大高校、科研院所等生态伙伴，围绕人工智能大模型创新应用，聚焦真实场景的算力需求，帮助智算中心客户主动引流，打造生态共赢圈。

第三，相比其他国内 GPU 厂商，发行人较早地启动了国产供应链布局 and 具体规划，首款国产供应链产品曦云 C600 已完成回片点亮并正式发布，性能测试结果及市场反响良好。依托自主可控、稳定的国产供应链优势，发行人该款产品预计未来将在信创市场实现规模化应用，同时也已受到多家互联网企业的高度关注。

关于竞争劣势，发行人主要是目前在互联网企业客户开拓及产品导入方面，相比部分国内友商进度有所滞后。如前所述，互联网企业的采购决策高度市场化，出于对产品性价比和生态竞争力的考虑，其第一选择仍然是国际产品，而在部分推理场景下可能会使用国产算力芯片，但通常也是优先选择自有产品，或者扶持

其投资的生态链企业。对发行人而言，一方面是暂时缺乏互联网企业股东背景及相应业务支持，另一方面则由于前述友商的产品多数为 ASIC 架构，其通过针对特定任务进行硬件定制化设计可实现较高的能效比，因此在模型和算子相对固化的推理场景下更具优势。

（4）公司主要发展方向及竞争策略

发行人目前正处于从产品优势迈向市场领先优势的阶段，未来将长期坚持“1+6+X”发展战略，不断加大市场开拓力度，具体的竞争策略如下：

持续深耕拓展算力底座客户及市场。针对国家人工智能公共算力平台和商业化智算中心客户，密切跟进存量客户的后续项目建设计划，基于前期交付产品的良好运行与服务，巩固并增加在其扩建项目中所占份额，同时依托市场标杆项目的示范效应，持续开拓新客户。针对运营商客户，争取进入集采短名单，并积极参与其牵头的联合生态开源社区建设，借助运营商大型国有云在特定央国企领域的主导地位，提升公司产品在信创市场的影响力和市场份额。针对互联网企业客户，一方面成立专门的关键客户中心，围绕其需求场景加大资源投入，推进产品优化和测试导入，另一方面持续沟通深入了解互联网客户的最新诉求，主动规划布局公司下一代产品的设计方向。

聚焦六大重点行业进行生态建设。公司将在教科研、金融、交通、能源、医疗健康、大文娱等领域持续积累实际落地经验，与广大软硬件生态伙伴合作击穿相关行业的 AI 应用场景，打造标杆案例，形成可复制的应用解决方案并在行业内大范围商业推广。具体措施层面，主要通过技术开源主动赋能垂直领域的生态伙伴，开展联合科研课题和创新项目合作，同时安排技术团队前移一线市场，深入理解行业客户的实际需求，快速沉淀产品解决方案予以支持。

在商务组织架构方面，公司通过直销、经销与关键客户进行各类市场营销活动。针对直销与关键客户，进一步加强技术前移和本地化服务，以更快触达重大商机。针对渠道合作伙伴，持续提供培训与技术支持，充分发挥其在行业或区域的商务资源及应用技术积累，不断提高经销商的自主商机转化成功率。

（四）市场主流计算架构（如 GPU、ASIC/DSA 等）以及 GPU 不同产品类型（如计算、图形等）各自对应的市场规模、下游应用领域、行业参与厂商，公司产品覆盖的应用领域产品的市场空间、竞争格局及未来发展趋势

1、市场主流计算架构（如 GPU、ASIC/DSA 等）以及 GPU 不同产品类型（如计算、图形等）各自对应的市场规模、下游应用领域、行业参与厂商

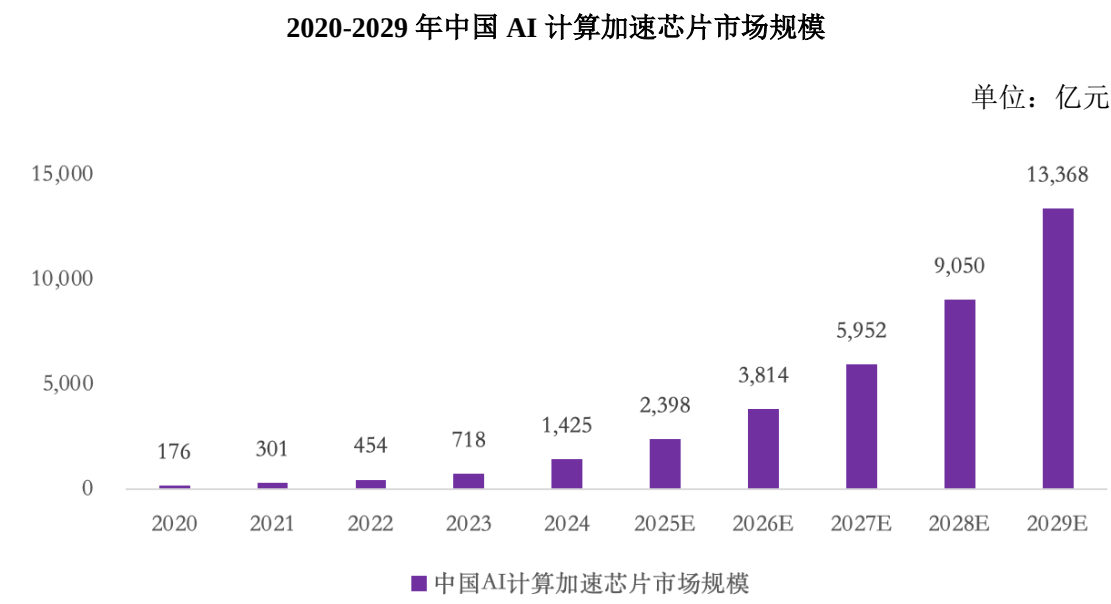
GPU 强大的并行处理能力使其早期多被用于 2D 和 3D 等图形处理场景，例如视频渲染、游戏渲染、专业制图等。随着人工智能的不断发展，自然语言处理、视觉及语音处理等多模态人工智能任务需求使得 GPU 的应用领域逐渐向需要大量重复计算的人工智能训练/推理、复杂科学计算、大数据处理等高性能计算领域的方向持续演进。同时，为满足人工智能特定场景需求，部分云计算公司针对自有的人工智能业务特殊场景开发了专用芯片（一种专为深度学习加速设计的 ASIC/DSA 芯片）。因此，目前市场主流计算架构以及 GPU 主要产品类型的基本情况如下：

架构		类型		功能	下游应用领域	竞争格局	发行人覆盖领域
通用型架构	GPU	图形 GPU		图形渲染	游戏、专业图像、视频渲染等	参与者以专业 GPU 厂商为主，全球竞争格局较为集中，国内市场正处于高速国产替代进程中	曦彩 G 系列
		计算 GPU（以下统称为“通用 GPU”）	训练 GPU/训推一体 GPU	高性能计算	以人工智能训练为主，也可用于人工智能推理、通用计算、科学计算等		曦云 C 系列
			推理 GPU		人工智能推理		曦思 N 系列
专用型架构	ASIC/DSA	人工智能芯片		高性能计算	以人工智能推理为主，部分产品可用于人工智能训练	参与者以科技巨头及其孵化的企业为主，竞争格局相较 GPU 架构更为分散，不同厂商之间差异化竞争	不涉及

（1）GPU 是当前计算芯片的主流架构，具有广阔的市场空间，国产厂商正在逐步提升国内市场份额

通用型架构的 GPU 与专用型架构（ASIC/DSA）芯片均主要应用于人工智能计算领域。随着人工智能技术日趋成熟，人工智能商业化应用将加快落地，推动人工智能芯片市场高速增长。根据弗若斯特沙利文预测，到 2029 年，中国 AI

芯片市场规模将从 2024 年的 1,425.37 亿元激增至 13,367.92 亿元,2025 年至 2029 年期间年均复合增长率为 53.7%。



数据来源：弗若斯特沙利文

通用型架构的 GPU 具备灵活的指令集和精巧的处理器架构，可覆盖人工智能领域多元化应用场景，在模型适配性、应用灵活性、生态兼容性等方面的综合优势突出，是目前主流的人工智能芯片计算架构。根据 IDC 数据，2024 年中国加速服务器市场中 GPU 服务器占据主导地位，约占 70% 的市场份额。根据弗若斯特沙利文数据，2024 年 GPU 架构的人工智能芯片在中国 AI 芯片市场中约占 69.9% 的市场份额，其他 AI 芯片（主要包括 ASIC 及 FPGA）约占 30.1%，预计到 2029 年 GPU 的市场份额将提升至 77.3%。据此测算，2024 年中国 AI 芯片市场中 GPU 的市场规模约 1,000 亿元，以 ASIC/DSA 为代表的其他 AI 芯片的市场规模约 425 亿元。

经过多年竞争与发展，全球 GPU 市场头部化现象显著，整体呈寡头垄断格局，英伟达（NVIDIA）和超威半导体（AMD）两家国外领先厂商基本分割了全球市场，在综合技术实力、销售规模、资金实力、人才团队等方面优势明显。在国内 GPU 市场，国际龙头英伟达仍然占据主导地位，具备较强的市场竞争力和领先优势。当前国际贸易新态势促使国内 GPU 市场需求缺口不断扩大，亟待有竞争力的国产 GPU 来填补。发行人、海光信息、天数智芯、壁仞科技、摩尔线程等国内企业紧跟国际龙头的技术路线，采取 GPU 架构积极参与市场竞争，逐

步提升国产 GPU 的市场渗透率。

其他采用专用型架构(ASIC/DSA)芯片的竞争格局相较 GPU 架构更为分散,不同厂商从不同细分领域切入市场,形成差异化竞争。专用型芯片的设计目的是针对特定场景进行优化,在特定场景下具有能效优势,参与者主要包括华为海思、寒武纪,以及具有特定应用需求的科技巨头所孵化或对外投资的生态链企业,如昆仑芯、平头哥、燧原科技等。

(2) GPU 不同产品类型对应的市场规模、下游应用领域、行业参与厂商

GPU 按照功能及下游应用领域可划分为通用 GPU 和图形 GPU。通用 GPU 主要应用于人工智能计算、通用计算及科学计算等需要执行大量并行任务的应用场景,其中在人工智能计算场景还可进一步细分为训练 GPU 和推理 GPU。图形 GPU 则主要应用于游戏、专业制图、影视动画、数字孪生等下游领域。

根据 Verified Market Research 的数据,2024 年全球 GPU 市场规模为 773.9 亿美元,2030 年有望达到 4,724.5 亿美元,2024-2030 年的复合增长率高达 35.19%,呈现强劲的增长态势。国内市场来看,中国 GPU 产业近年来发展速度较快且取得了显著进步,根据弗若斯特沙利文数据,中国 GPU 市场规模从 2020 年的 384.77 亿元快速增长到 2024 年的 1,638.17 亿元,并有望在 2029 年达到 13,635.78 亿元。关于 GPU 不同细分市场的分析具体如下:

1) 人工智能计算 GPU 市场空间广阔,增速迅猛

人工智能是 GPU 的主要应用场景。人工智能前沿模型的竞争正推动着人工智能水平的迅猛提升,模型迭代时间快速缩短,在性能、效率和应用场景上实现的指数级突破,不仅推动了人工智能从专用领域向通用智能的跨越,其快速迭代的特性更对新一代高性能人工智能计算基础设施提出了迫切需求。根据弗若斯特沙利文数据,中国 AI 智算 GPU 市场规模从 2020 年的 142.86 亿元迅速增至 2024 年的 996.72 亿元,年均复合增长率达 62.5%,预计至 2029 年将提升至 10,333.40 亿元的市场规模,发展势头迅猛。

①人工智能训练算力是大模型智能提升的基石

大模型的发展遵循 Scaling Law 的经验公式,即模型参数量、数据量以及计算资源的增长能得到更好的模型智能。随着人工智能应用的快速兴起,大模型厂

商正在加速探索大模型能力边界，支撑着人工智能训练对高性能算力需求的持续增长。同时，人工智能应用正不断辐射教科研、金融、交通、能源、医疗健康、大文娱等垂直行业场景，中小企业可以通过微调、蒸馏、强化学习等方式参与到人工智能浪潮中，在开源大模型的基础上二次开发后训练的新模型结构或垂直大模型来支撑业务场景，从而为高性能算力需求注入增长动能。

另一方面，人工智能的应用通常面临着语音、图像、视频等多模态数据结构，因此大模型正从只能处理单模态任务的大语言模型向多模态大模型的方向发展，加速人工智能时代的到来。多模态需要更大规模的数据量、更为复杂的模型结构，也将不断产生新的人工智能训练算力需求。

②人工智能推理算力是推动 AI 终端应用大规模落地的关键

人工智能推理任务的高效执行是大模型在终端落地与应用的关键。自动驾驶、智能座舱、具身智能等具有交互特性的生成式人工智能应用需要底层算力设施具备快速响应并处理高并发、大吞吐量、实时的文本、语音、图像等数据的能力。同时，随着人工智能应用对模型的理解能力、处理更为复杂推理任务的能力等提出了更高要求，大模型的思维链（CoT）在不断延长，反哺并推动人工智能推理算力需求蓬勃增长。

在广阔的人工智能计算 GPU 市场中，国内厂商正凭借自身优势不断加速市场渗透，已取得一定商业化进展。根据 Bernstein Research 数据，2024 年中国 AI 加速芯片市场中 GPU（不含 ASIC 芯片公司）的主要参与者包括英伟达、AMD、发行人、摩尔线程。根据 IDC 数据，2024 年中国数据中心加速芯片市场中，出货量在 10,000 张以上的 GPU 公司（不含 ASIC 芯片公司）包括英伟达、发行人和天数智芯。除此之外，海光信息、壁仞科技等公司亦在人工智能计算 GPU 市场有所布局。

2) 图形 GPU 市场保持稳定增长

图形 GPU 主要应用于桌面级的游戏、专业制图等场景，部分高性能图形 GPU 可用于云端渲染。近年来大型 3A 游戏和专业图形渲染等应用领域不断提高对图形处理能力的要求，图形 GPU 作为显卡的核心部件，下游市场对其需求亦稳步提升。根据 Jon Peddie Research 数据，2024 年全球独立显卡出货量为 3,470 万张，

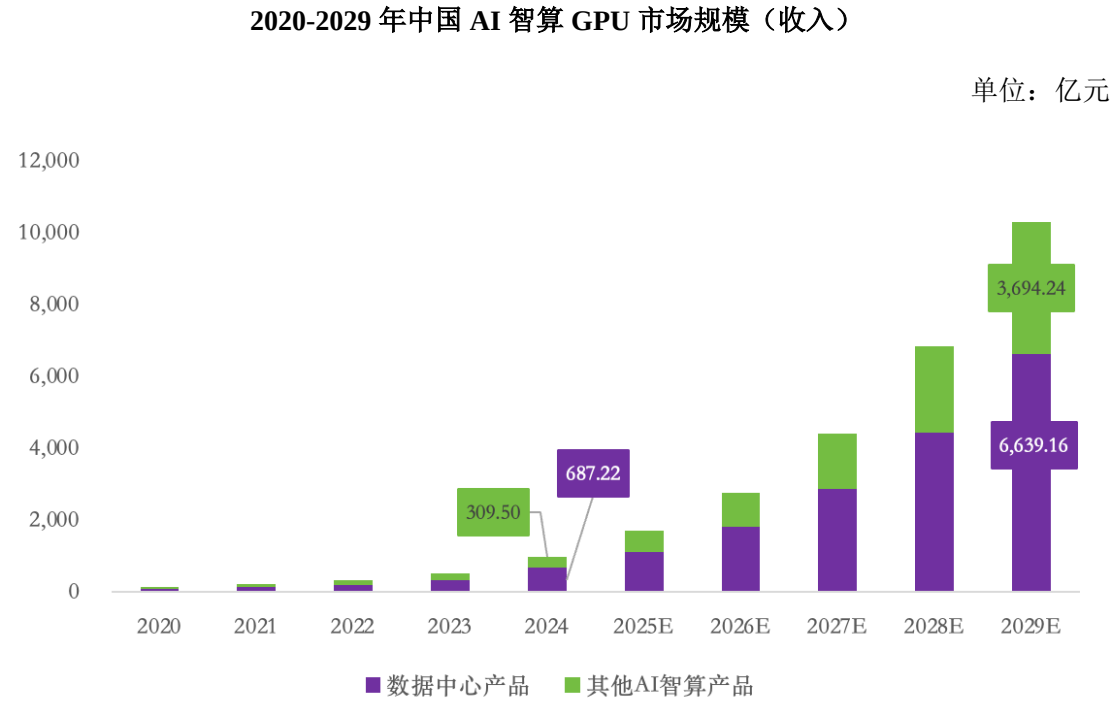
相较去年同期同比增长 12%。根据弗若斯特沙利文数据，中国桌面级 GPU 市场规模从 2020 年的 241.91 亿元增至 2024 年的 641.45 亿元。

图形 GPU 市场的主要参与者包括海外巨头英伟达、AMD，国内供应商主要为景嘉微、摩尔线程、格兰菲智能科技股份有限公司、瀚博半导体（上海）股份有限公司等。发行人在研的图形渲染产品主要为应用于云端渲染的高性能图形 GPU。

2、公司产品覆盖的应用领域产品的市场空间、竞争格局及未来发展趋势

(1) 市场空间

发行人致力于自主研发全栈高性能 GPU 芯片及计算平台，主营业务是研发、设计和销售应用于人工智能训练和推理、通用计算与图形渲染领域的全栈 GPU 产品，并围绕 GPU 芯片提供配套的软件栈与计算平台。发行人的产品目前均主要应用于云端（数据中心）场景。根据弗若斯特沙利文数据，中国 AI 智算 GPU 市场中，数据中心 GPU 产品是过去增速最快的细分市场，其市场规模从 2020 年的 82.00 亿元快速增长至 2024 年的 687.22 亿元，年均复合增长率高达 70.1%，预计未来将以 55.7% 的年均复合增长率高速增长至 2029 年的 6,639.16 亿元。



数据来源：弗若斯特沙利文

（2）竞争格局

根据 Bernstein Research 数据，英伟达和 AMD 在 2024 年中国 AI 加速芯片市场中分别占据 66%、5% 的市场份额，国内企业中华为海思依托华为的雄厚资金、人才储备、商业渠道等，占据约 23% 的市场份额。发行人市场份额约为 1%，在其他国内厂商中位居前列。

从具体产品类型来看，发行人曦云 C 系列产品主要用于人工智能训练（训推一体）和通用计算，也是发行人目前的主力产品；曦思 N 系列产品主要用于人工智能推理；曦彩 G 系列主要用于图形渲染，尚在研发过程中。

在国内人工智能训练市场，发行人主要竞争对手包括英伟达、AMD、华为海思、寒武纪、海光信息、昆仑芯、平头哥、天数智芯、摩尔线程、壁仞科技等。训练芯片对架构设计、存储设计、集群能力、软件生态等综合能力要求较高，尤其随着大模型训练参数量的不断增加、算法结构愈发复杂化，AI 训练任务逐渐需要千卡、万卡甚至更大规模智算集群支撑，因此对计算系统的通信能力提出了较高要求。国内厂商虽然在单卡性能上已取得一定突破，但在集群能力、软件生态方面与国外厂商仍存在较大差距，真正能够将训练芯片投入实际训练任务的国产厂商较少。近年来，受到美国高性能计算芯片出口管制与国内政策推动等因素影响，英伟达在中国市场的份额呈现明显下降趋势。2021 年至 2024 年，英伟达收入结构中来源于中国的比例从 26.42% 下降至 13.11%。国产厂商中，华为海思经过长期技术积淀、全栈技术协同、丰富的人才和客户资源储备，已建立了明显的领先优势。发行人作为国内训练芯片的核心厂商之一，是国内少数真正实现千卡集群大规模商业化应用的 GPU 供应商，并正在研发和推动万卡集群的落地，目前已成功支持 128B MoE 大模型等完成全量预训练。

在国内人工智能推理市场，由于推理芯片设计难度、数据处理规模相对较低，且因推理模型仅涉及前向传播，模型结构相对训练模型更为稳定，部分专用型芯片厂商可以针对特定的推理任务进行深度优化，因此推理芯片市场参与者较多，不同技术路径的人工智能芯片厂商在推理芯片市场各具差异化竞争优势。发行人的智算推理芯片主要应用于云端推理场景，在处理大规模和复杂推理任务时具备较强产品表现。

（3）未来发展趋势

1) 训练芯片与推理芯片需求持续激增，人工智能计算市场国产替代正当时

随着人工智能技术持续突破，全球范围内实现了新一轮 AI 普及，推动全球人工智能竞赛再次加速，下游对算力的需求激增。GPU 公司正迎来结构性市场机遇，传统由超大规模云厂商主导的集中式算力需求模式正逐步演进为由头部云厂商为主、各行各业中小企业为辅的人工智能生态。更多中小企业和开发者借助模型微调、蒸馏、强化学习等技术积极投身人工智能领域，这类庞大的客户群体为 GPU 公司开辟了新的重要增长极。与此同时，在地缘政治不断升级和国家人工智能产业链自主可控需求的驱动下，国内 GPU 市场的国产化率将显著提升，逐步实现对进口产品的替代，这为具有领先技术实力和商业化进展的国产厂商创造了巨大的发展空间。

从人工智能芯片的需求场景来看，当前正处于人工智能技术探索阶段，算力需求主要来源于模型训练。训练芯片相较推理芯片的设计难度和价值量都更高。根据 IDC 数据，训练型 AI 服务器中 GPU 的价值占比可达 73%，而推理型 AI 服务器中 GPU 的价值占比约为 25%。发行人用于训练场景的曦云 C 系列产品已在多元化应用场景商业化落地，未来也将作为发行人的业绩基石，随着市场需求的爆发而快速增长。

未来，随着训练模型的完善与成熟，模型和应用产品逐步投入生产，推理端的人工智能服务器占比将随之攀升，推理需求也将下沉至不同的应用场景。因此，随着 AI 技术的不断进步，推理算力将成为推动产业发展的第二重驱动力。发行人本次募集资金投向之一的“新一代人工智能推理 GPU 研发及产业化项目”专为生成式 AI 推理市场设计，将为发行人注入新的增长动能。

算法创新将推动人工智能技术使用门槛不断降低、赋能千行百业，更多中小企业能够便捷地获取先进技术、积极投身人工智能应用与转型，因此训练芯片和推理芯片的需求都将呈现爆发式增长趋势。据 IDC 预测，预计未来五年内，国内训练和推理算力的年复合增速将分别达到 50%和 190%。基于在人工智能训练市场的市场优势和在人工智能推理市场的技术与产品储备，发行人将在智算产业链国产替代进程中充分受益。

2) 云边端融合发展，通用型计算芯片与专用型计算芯片将共同构建一体化生态

由于人工智能行业处于快速发展期，人工智能算法尚在不断开展前沿技术探索，因此也涌现了处理不同人工智能任务的芯片。关于通用型计算芯片和专用型计算芯片的差异对比情况，具体参见招股说明书之“第五节 业务与技术”之“二 /（三） /3/（1） /3）人工智能芯片技术路径”。长期来看，未来通用型计算芯片和专用型计算芯片将各有侧重，共同构建一体化生态，具体如下：

云端作为资本开支最大、算法迭代最快、计算能力要求最高的需求场景，将成为人工智能训练和推理的“智能大脑”，处理重量级人工智能任务，如通用基础大模型的训练、垂直大模型的复杂推理等。由于算法正处于不断发展演进的过程，模型结构并未固化，迭代创新日趋频繁，快速迭代的算法和大量非传统模型结构的出现需要通用型计算芯片（GPU）完成高效迁移与适配。随着云端智算集群的需求快速扩张以及后训练、推理等阶段新模型结构的快速演变，具有更强通用性、灵活性和可扩展性优势的 GPU 架构将长期成为云端智算的主流方案。

端侧（边缘端、终端）则是综合平衡性能和负载的选择，作为“小脑”和“四肢”处理即时、轻量级人工智能任务，如简单推理，将以更多元的形态进入千家万户，包括 AI 手机、AI PC、AI 眼镜、AI 玩具、AI 机器人等。由于端侧对功耗、负载的敏感度亦新增了对专用型计算芯片的需求，专用型计算芯片可针对固定模型结构及特定应用场景提供更具能效比的解决方案，将成为通用型计算芯片的有效补充。

综上所述，发行人主要产品所处的 GPU 市场空间广阔、成长潜力巨大；在增速强劲的国内人工智能计算市场中，发行人具有领先技术实力和商业化进展，已形成一定市场地位和品牌知名度；在具有稳定需求的图形渲染 GPU 市场中，发行人亦已规划并研发对应的产品线。发行人采用的通用、灵活的 GPU 架构将长期成为云端智算的主流方案，符合行业技术发展趋势。发行人的产品规划与市场需求发展趋势高度匹配，将长期保持增长潜力，不断提高市场份额。

（五）量化分析公司产品在境内市场的占有率情况，结合发行人与境内外同行业公司

在产品布局、性能、业绩规模、市场份额、主要下游客户等情况的比较，分析公司与境外龙头的差距及在境内市场的竞争优劣势。评估境外 GPU 龙头对华产品销售政策变化可能对公司业绩及市场拓展产生的影响

1、量化分析公司产品在境内市场的占有率情况

发行人是国内少数真正实现千卡集群大规模商业化应用的 GPU 供应商，交付的算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心，区域横跨北京、上海、杭州、长沙、中国香港等地区，并深度赋能众多行业应用场景，产品赋能真实应用场景的竞争力和交付能力得到充分验证，在国内具有领先的市场地位。

根据 Bernstein Research 数据，以销售金额口径计，2024 年发行人在中国 AI 加速芯片市场份额约 1%。鉴于不同 AI 芯片厂商产品的单位算力售价比差距较大，从中国 AI 算力规模的角度能够更加公允体现市场份额情况。故按照算力规模（基于 FP16）口径测算，发行人 2024 年市场份额约 1.14%。

2、结合发行人与境内外同行业公司在产品布局、性能、业绩规模、市场份额、主要下游客户等情况的比较，分析公司与境外龙头的差距及在境内市场的竞争优劣势

（1）发行人与境内外同行业公司在产品布局、性能、业绩规模、市场份额、主要下游客户等情况的比较

发行人目前专注于技术门槛更高、市场规模更大、成长空间更广阔的云端高性能通用 GPU，不涉及桌面级/消费级 GPU 产品、CPU 产品等，产品定位清晰、聚焦，有利于公司集中资源实现技术突破和高效产业化。发行人的产品性能、业绩规模、市场份额、客户拓展情况等在国内厂商中均位列前茅。发行人与境内外同行业公司在产品布局、性能、业绩规模、市场份额、主要下游客户等方面基本情况比较如下：

公司简称	产品布局	产品性能			2024 年度 营业收入 （万元）	境内市场份额	主要下游客户
		架构	硬件性能	软件生态			
境外企业							
英伟达	主要产品： GPU、图形显卡及配套产品等，包括数据中心产品、游戏 GPU、专业视觉设计 GPU、智能驾驶 GPU 等业务板块 主要应用领域： 人工智能训练、人工智能推理、通用计算、图形渲染等 主要产品形态： 包括板卡、服务器等	GPU	国际标杆、全球领先	生态竞争能力全球领先，CUDA 生态引领国际市场	93,617,242.83	根据 IDC 数据，2024 年英伟达在中国 AI 加速芯片市场份额约为 70%； 根据 Bernstein Research，2024 年英伟达在中国 AI 加速器市场份额约 66%	云计算科技巨头、AI 大模型厂商、汽车制造商等
AMD	主要产品： CPU、GPU、FPGA、DPU、SoC 等芯片产品，包括数据中心、客户端、游戏及嵌入式等业务板块 主要应用领域： 人工智能训练、人工智能推理、通用计算、图形渲染等 主要产品形态： 包括板卡、服务器等	GPU	全球先进	生态竞争能力较强，ROCm 生态作为开源生态吸引了部分开发者	18,497,901.15	根据 Bernstein Research，2024 年 AMD 在中国 AI 加速器市场份额约 5%	服务器厂商、云计算科技巨头、消费电子终端等
境内企业							
华为海思	主要产品： 人工智能芯片、网络芯片、模拟芯片、MCU、光模块等，包括智慧视觉、智慧媒体、显示交互、智能感知、激光显示等多个业务板块 主要应用领域： 人工智能训练、人工智能推理 主要产品形态： 包括板卡、服务器、集群等	ASIC	国内领先	生态建设较早，自主研发 CANN 异构计算架构，具备一定的生态能力	未披露	根据 Bernstein Research，2024 年华为海思在中国 AI 加速器市场份额约 23%	智算中心、服务器厂商、AI 大模型厂商、运营商、政企客户等
寒武纪	主要产品： 云端、边端智能芯片及加速卡、终端智能处理器 IP 主要应用领域： 人工智能训练、人工智能推理 主要产品形态： 包括板卡、服务器、	ASIC	国内较为领先	自主研发 Cambricon Neuware 基础系统软件，具有云边端一体化开发	117,446.44	根据 Bernstein Research，2024 年寒武纪在中国 AI 加速器市场份额约 1%	大模型算法公司、服务器厂商、人工智能应用公司等

公司简称	产品布局	产品性能			2024 年度 营业收入 (万元)	境内市场份额	主要下游客户
		架构	硬件性能	软件生态			
	集群等，（1）2023 年及以前以集群为主，2023 年智能计算集群系统收入占比为 85.46%；（2）2024 年以来以板卡及服务器为主，2024 年、2025 年 1-3 月云端产品线收入占比分别为 99.31%、99.94%（包括云端智能芯片及板卡、智能整机）			环境			
海光信息	主要产品： CPU、DCU（GPU）及配套产品、技术服务 主要应用领域： 人工智能训练、人工智能推理、通用计算 主要产品形态： 以板卡为主	GPU	国内较为领先	兼容 AMD ROCm 计算生态，积极参与开源社区	916,214.81	未披露 GPU 产品具体商业化情况	服务器厂商、政企客户等
摩尔线程	主要产品： AI 智算产品、专业图形加速产品、桌面级图形加速产品、智能 SoC 类产品等 主要应用领域： 人工智能训练、人工智能推理、通用计算、图形渲染等 主要产品形态： 包括板卡、集群、一体机等，2024 年板卡收入占比 49.35%，集群及一体机收入占比 49.76%	GPU	国内较为领先	依托自研 MUSA 架构搭建国产计算加速生态	43,845.95	根据 Bernstein Research，2024 年摩尔线程在中国 AI 加速器市场份额约 1%	智算中心、政企客户、电商平台等
发行人	主要产品： 训推一体 GPU、智算推理 GPU、图形渲染 GPU 等 主要应用领域： 人工智能训练、人工智能推理、通用计算 主要产品形态： 包括板卡、服务器等，以板卡为主，2024 年板卡收入占比为 69.46%、服务器收入占比为 28.29%；2025 年 1-3 月板卡收入占比为	GPU	国内较为领先	自主研发 MXMACA 软件栈，拥有统一、完整且高效的全栈式工具链，同时高度兼容 CUDA 生态	74,307.16	根据 Bernstein Research，2024 年发行人在中国 AI 加速器市场份额约 1%；基于算力规模测算，发行人市场份额约 1.14%	智算中心、服务器厂商、运营商、政企客户、行业客户

公司简称	产品布局	产品性能			2024 年度 营业收入 (万元)	境内市场份额	主要下游客户
		架构	硬件性能	软件生态			
	98.80%、服务器收入占比为 0.32%						

注 1：资料来源于同行业公司公开披露资料、公司官网、市场研究报告等公开信息。

注 2：部分境内同行业公司未公开披露商业化情况，上表列示了部分市场主流研究机构测算的研究结果。

注 3：英伟达财务报告期间为 2024 年 1 月末至 2025 年 1 月末；AMD 的 GPU 产品主要体现在数据中心及游戏板块；海光信息高端处理器收入包括 CPU 及 DCU（即 GPU）收入。

（2）发行人与境外龙头的差距分析

自 2020 年成立以来，发行人坚持以自主技术创新为导向，持续深耕 GPU 行业，经过“从零到一”全自研的技术沉淀，在国内云端智算市场率先实现了“技术-产品-市场”的产业化应用，产品在部分性能指标上已经接近或达到国际先进水平，但与境外龙头相比仍存在一定差距。

技术水平与产品性能上，境外龙头英伟达长期引领行业技术方向，已形成了深厚的技术底蕴和丰富的行业经验，其产品在硬件架构、单卡算力、互连拓扑、软件生态、大规模集群建设等方面都建立了明显的领先优势；而 AMD 亦紧随其后不断迭代和优化产品性能。

1）在单卡性能方面：经过两个大代际、五个小代际流处理器架构的迭代升级，发行人已形成了全面覆盖计算（人工智能训练、推理、通用计算）和渲染场景的核心 GPU IP，打造了由超过 600 条计算 GPU 指令、800 条渲染 GPU 指令组成自主安全可控的 MXMACA 指令集，单卡算力已接近英伟达主流产品。但受限于工艺制程和研发周期，包括发行人在内的绝大部分国内厂商与英伟达最新产品均仍存在一到两代差距。

2）在集群通信方面：发行人自主研发的 MetaXLink 高速互连技术突破了传统 PCIe 总线在带宽和延迟方面的限制，达到了与英伟达 4nm 制程工艺下旗舰产品（H200）相当的互连带宽性能，是国内厂商中率先将与英伟达 NVLink 技术的差距缩小到一代以内的 GPU 供应商。

3）在软件生态方面：英伟达是软件生态建设的先行者与领导者，全球英伟达 CUDA 开发者已达数百万。而发行人 MXMACA 软件栈截至 2025 年 7 月 31 日的注册用户超过 15,000 人。发行人在软件生态的建设方面与之相比还存在较大的发展空间。

产品布局与商业化上，依托持续的行业深耕，发行人在经营规模上取得了快速增长，但尚处于发展初期阶段，整体产品数量、业绩规模、市场份额、客户数量仍然偏小，相较境外龙头仍存在明显差距。在地缘政治冲突大背景下，境外龙头先进计算芯片产品在国内市场受到一定限制，为国内厂商创造了发展空间。发行人将把握市场机遇，持续进行产品迭代和技术创新，加速市场拓展，进一步丰

富产品矩阵、提高技术创新能力、扩大市场影响力，通过发展先进封装、互连通信等技术、扎根国内需求市场，与境外龙头错位竞争。

（3）发行人在境内市场的竞争优劣势

1) 竞争优势

发行人坚持以技术及产品自主可控为导向，持续深耕 GPU 行业，在专业知识储备深厚、行业经验丰富的研发团队带领下，发行人通过技术创新和设计优化，持续提升 GPU 架构的通用性、GPU 产品的高能效（单卡性能和集群性能）与 GPU 软件生态的易用性，已在境内市场率先实现了“技术-产品-市场”的产业化应用。具体竞争优势如下：

①产品优势：以具备通用性、灵活性的 GPU 架构打造全系列 GPU 产品，产品性能在多个维度领先行业

发行人坚持以自主技术创新为导向，持续深耕 GPU 行业，能够为客户提供具备竞争力的 GPU 产品。根据多家行业权威的国家实验室、计算中心、研究所的测试结果，公司 GPU 产品在通用性、单卡性能、集群性能及稳定性、生态兼容与迁移效率等方面具备较强的核心竞争力，综合技术实力领先行业。

单卡性能方面，公司是国内首批具备先进制程 GPU 芯片设计能力和商业化落地能力的企业之一，全栈 GPU 产品基于自主研发和自主知识产权的 GPU IP、GPU 指令集和架构，单卡性能处于国内第一梯队。

集群性能方面，公司自研的 MetaXLink 具备国内稀缺的高带宽卡间互连能力，可实现 2-64 卡多种互连拓扑，并且在智算集群的线性度、稳定性、利用率等方面具有较强的产品表现。

软件生态方面，公司自主构建的 MXMACA 软件栈不仅拥有统一、完整且高效的全栈式工具链，涵盖应用开发、功能调试和性能调优等核心环节，同时高度兼容 GPU 行业国际主流 CUDA 生态，能够开放拥抱全球开发者丰富的开源成果，具有较高的易用性和迁移效率，在通用性和灵活性上具备独特的竞争力。

具体产品性能优势参见本问询回复报告之“2.关于核心技术”之“一/（二）说明发行人 GPU 主要细分产品的技术先进性”。

②市场优势：快速实现商业化落地，产品赋能真实应用场景的竞争力和交付能力强劲

发行人坚持“1+6+X”生态与商业布局并注重研发成果的产业化，仅用3年时间即实现两款芯片一次流片并成功量产，陆续量产的智算推理曦思N系列GPU产品、训推一体曦云C系列GPU产品推动公司业绩快速爆发，最近三年营业收入年均复合增长率高达4074.52%。2024年，发行人营业收入为74,307.16万元，业绩规模 and 市场份额在国内同行业公司中位列前茅。基于良好的产品使用反馈，发行人主要客户普遍在其后续项目（如有）中对发行人产品进行了复购并进一步加大了采购规模，强劲的客户复购情况与用户粘性将持续推动公司市场份额提升。

发行人是国内少数真正实现多个千卡集群大规模商业化应用的GPU供应商，并正在研发和推动万卡集群落地。凭借突出的产品性能和稳定的供应能力，截至报告期末，发行人GPU产品累计销量超过25,000颗。发行人的产品相继应用部署于10余个智算集群，算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心，区域横跨北京、上海、杭州、长沙、中国香港等地区，并逐渐向更多区域延伸。发行人的产品已广泛适配整机服务器、操作系统、运维管理平台、主流AI框架、主流大模型等上下游生态，通过芯片层、框架层、模型层的深度协同优化，为头部大模型分布式推理提供千卡高性能国产算力，助力“国产算力+大模型产业”实现从技术突破到商业化落地的跨越式发展。

随着人工智能技术的持续突破和模型成本的下探，更多中小企业和开发者借助模型微调、蒸馏、强化学习等技术积极投身人工智能领域，这类庞大的客户群体为发行人拓展了重要的市场机遇。基于通用、灵活的GPU架构，发行人GPU产品深度赋能众多行业应用场景，已率先布局教科研、金融、交通、能源、医疗健康、大文娱等行业：1）在教科研领域，发行人与众多国内知名高校合作建设人工智能产教融合创新算力平台，推动海量实验数据规律识别、虚拟实验仿真、AI辅助教学等，助力AI for Science等前沿科学的突破。2）在金融领域，发行人持续加深与银行、保险公司及基金公司的业务探索，以智能算力支持金融机构数智化转型升级，帮助客户实现大模型在AI数字人、智能问答、信贷评测等实际业务场景中的落地应用。3）在交通领域，发行人深度参与自动驾驶大模型云端算力建设，支撑主机厂自驾模型训练、自动驾驶相关系统仿真，推动交通产业

向高阶自动驾驶迈进。4) 在能源领域，发行人的 GPU 持续赋能能源央国企的大模型产品，推动智慧能源、智慧网关业务，实现能源领域从精准预测到智能决策的闭环优化，助力我国能源的高效生产、精准传输调度、智能存储以及合理消费。5) 在医疗健康领域，发行人积极与医疗机构、医疗研究所合作探索医疗大模型在 AI 辅助诊疗、药物分析、医学成像等业务场景的应用。6) 在大文娱领域，发行人依托底层通用算力优势推动人工智能在文化创意和视觉场景的增质提效。

预计随着产品矩阵不断完善和产品性能持续迭代，发行人将进一步提升市场份额、优化客户结构、提高业绩规模并强化市场影响力，真正推动人工智能赋能千行百业。

③自主可控优势：核心环节全流程自主研发，并率先导入国产供应链

芯片设计方面，发行人是国内少数几家全面系统掌握了 GPU 架构、GPU IP、先进制程 GPU 芯片及其基础系统软件研发、设计和量产等全流程核心技术的企业之一。为保障芯片设计的自主性，发行人成为了国内少数坚持“从零到一”自主研发核心 GPU IP 的集成电路设计企业，经过多年产品设计及落地经验的积累，已形成了全面覆盖计算（人工智能训练、推理、通用计算）和渲染场景的核心 GPU IP，打造了由超过 600 条（XCORE1.0 计算 GPU、XCORE 1.5 计算 GPU）、800 条（XCORE2.0 渲染 GPU）指令组成自主安全可控的 MXMACA 指令集。

制造工艺方面，面对持续升级的地缘政治摩擦，发行人较早启动了国产供应链布局，与国内晶圆厂、封测厂、EDA/IP 等厂商技术合作与攻关，提早规划、设计、导入了基于国产供应链的产品，并率先完成了生产落地，能够确保后续产品高质量及可持续的供应与交付。

④持续创新优势：迅速适应市场需求的技术迭代，彰显突出的持续创新能力

GPU 芯片研发周期较长，而人工智能算法演进速度较快，倒逼算力硬件供应商快速设计出能适配最新模型结构的产品。发行人拥有专业基础扎实、行业经验丰富且稳定的技术团队，核心技术负责人均为 GPU 行业内资深专家，具备敏锐的市场洞察能力和前瞻性视野。在精准的技术研判下，发行人成为国内首批将 FP8、FP4 等低精度数值格式和相关指令集引入后续迭代产品架构的企业，能够

有效支持以 DeepSeek、阶跃星辰、月之暗面等为代表的头部大模型厂商的最新训练与推理需求。

发行人紧密追踪前沿技术趋势，持续推动技术创新与迭代。报告期内，基于自主创新的 DragonFly 互连拓扑，发行人已成功打造 64 卡光互连超节点产品及 3D-Mesh 64 卡超节点产品等。为支持超大规模智算中心、通用计算和数据处理等应用场景对超高互连规模、高带宽域的需求，发行人不断提高智算集群算力密度，突破性实现了单机柜 128 卡超高密度部署，能够支持千亿参数大模型的训练、微调及高效推理。

2) 竞争劣势

发行人虽在境内市场已取得一定市场份额，但因成立时间相对较短、尚处于快速成长阶段，与国内头部企业华为海思相比，在销售规模、销售网络、市场知名度、客户结构、人才规模等方面还存在差距。发行人将进一步引入集成电路和人工智能领域的专业人才，持续加强产品研发并扩充产品矩阵。同时，发行人将坚持“1+6+X”发展战略，加大市场开拓力度，在把握国家人工智能公共算力平台、商用 AIDC 以及运营商、互联网厂商等算力需求基本盘的基础上，充分发挥公司产品通用性优势，深入长尾行业市场，不断提升公司在高性能 GPU 行业的市场地位和影响力。

3、评估境外 GPU 龙头对华产品销售政策变化可能对公司业绩及市场拓展产生的影响

英伟达 H20 系列产品系在 2023 年 10 月美国 BIS 限制出口芯片规格参数的政策出台后，专为中国市场推出的特供版 GPU 产品，其关键参数如下：

公司简称	型号	FP16 (TFLOPS)	FP8 (TFLOPS)	INT8 (TFLOPS)	显存类型	显存大小 (GB)	功耗 (W)	卡间互连 (GB/s)
英伟达	H20	148	296	296	HBM3	96	400	900

2025 年 4 月 15 日，英伟达向美国 SEC 提交的 8-K 文件确认其 H20 系列 AI 芯片及所有达到同等内存带宽、互连带宽能力的电路产品出口至中国及 D: 5 类军火禁运国家将需获得美国政府的许可，标志着 H20 系列产品无法对华销售。2025 年 7 月 15 日，英伟达通过官网宣布美国政府已承诺为 H20 系列产品授予许可证，将恢复向中国销售 H20 系列产品，并将推出面向中国市场的全新 GPU 产

品。截至本回复出具日，英伟达 H20 系列产品已取得许可证。前述英伟达对华产品销售政策的变化不会对发行人业绩及市场拓展产生重大不利影响，具体分析如下：

（1）境外 GPU 龙头对华销售产品与发行人产品的主要客户群体存在较大差异，不构成直接竞争

受限于供应链对算力规格的要求，英伟达 H20 系列产品的算力较低，在人工智能训练任务中进行千亿、万亿级参数量的大规模训练效率有限。但 H20 配备了大容量 HBM3 显存并保留了 NVLink4.0 的卡间高速互连带宽，能加速数据在显存与计算单元间的交换，提升吞吐量，使得大模型推理效率提升。因此 H20 系列产品在人工智能推理场景的应用更为广泛。凭借 CUDA 生态的深度适配、较高的推理效率，英伟达 H20 系列产品主要面向具有丰富的人工智能应用场景需求的互联网厂商。根据 Bernstein Research 的调研结果，2024 年 H20 在中国市场的主要客户为互联网厂商，字节跳动、腾讯、阿里巴巴、百度的 H20 采购量合计占据 87% 的份额。

发行人当前主力产品为用于人工智能训练（训推一体）场景的曦云 C 系列 GPU，凭借较强的供应稳定性与交付能力、产品综合性能，已在国产替代需求明确的场景广泛应用。报告期内，发行人曦云 C500 系列产品的下游客户以国家人工智能公共算力平台、运营商智算平台、商业化智算中心以及教科研、金融等行业客户为主，与英伟达 H20 的主要目标销售市场不存在直接竞争。

（2）发行人 GPU 的产品性能在客观上具备替代境外 GPU 龙头对华销售产品的能力

境外 GPU 龙头对华产品销售政策变化后，H20 系当前英伟达在中国市场能够公开销售的最先进的产品。英伟达 H20 的核心优势在于大容量显存、高速互连带宽、多精度计算能力等。发行人曦云 C500 系列产品的单卡算力已超过 H20，最新一代曦云 C600 系列产品的单卡算力亦领先于英伟达 H20，并具备 FP8 计算单元，能够支持多精度计算。同时，发行人曦云 C600 系列产品配备了大容量显存，超过 H20 的 96GB 版本，可以有效承载更大规模数据与模型、减少数据交换的性能损耗、提升多任务并行执行能力。在卡间互连方面，基于 MetaXLink 高

速互连技术，发行人 GPU 产品的互连带宽达到了与英伟达 4nm 制程工艺下产品相当的性能。因此综合多维度参数对比，发行人 GPU 产品客观上具备替代英伟达 H20 的能力，在人工智能计算场景可以充分发挥高性能优势。H20 相关销售政策变动不会对发行人未来拓客造成重大不利影响。

（3）境外 GPU 龙头对华销售产品无法满足中国 AI 芯片需求缺口

根据英伟达 2026 财年一季报（截至 2025 年 4 月），英伟达为 H20 的库存计提了约 45 亿美元的减值，折合为数量约 30-40 万张 GPU 板卡。根据 IDC 数据，2024 年中国加速芯片市场出货量超过 270 万张，剔除 H20 的库存消耗后中国市场仍存在巨大的需求缺口。

鉴于英伟达已推出新一代 Blackwell 架构，而 H20 为上一代 Hopper 架构，生产计划存在不确定性，未来英伟达在中国市场或将主要销售基于 Blackwell 架构进一步限制性能后的特供产品。Bernstein Research 预测，面对英伟达对华销售产品规格升级空间存在不确定性的情况，国内算力需求方将进一步提升对中国本土 AI 芯片供应商的采购份额，预计到 2027 年，中国本土 AI 芯片供应商的市场份额将提升至 55%。作为国内领先的 GPU 厂商之一，发行人依托独特的竞争优势将快速提升市场份额，既有的市场需求缺口足以支撑发行人的业绩高速增长。

（4）“朝令夕改”风险下，国产替代已成共识

人工智能和集成电路是中美科技战的重要领域。近年来，虽然美国企业基于自身利益与美国政府反复博弈，对华政策“朝令夕改”、反复变动，但美国政府对国内半导体产业的限制整体上呈现持续加剧、不断恶化的态势，极度不稳定的国际贸易政策也为国内 AI 算力需求方带来了极高的断供风险。在此背景下，为保障我国人工智能产业链的稳定性与自主性，GPU 产品国产替代已成共识。发行人当前主要下游行业如政府智算中心、运营商、教科研、金融、能源、医疗健康等亦具有明确的国产化算力建设需求和安全性要求。例如，根据《上海市智能算力基础设施高质量发展“算力浦江”智算行动实施方案（2024-2025 年）》、《上海市关于促进智算云产业创新发展的实施意见（2025 - 2027 年）》，到 2025 年，上海市新建智算中心国产算力芯片使用占比超过 50%，到 2027 年自主可控算力占比进一步提升至 70%。

美国政府对高端芯片的出口管制最早可追溯至 2018 年，在出口管制不断升级的期间，以发行人为代表的中国 GPU 厂商紧抓该关键发展和培育的时间窗口，经过多年的技术积淀已进入大规模商业化阶段。对于国内互联网厂商、大模型厂商等重要算力需求方而言，当前中国高性能 GPU 市场已不再是“无芯可用”的局面，已存在更多具备质价比的选择。因此长期来看，为保障算力供应的稳定，互联网厂商同时采购境外龙头产品与国产厂商基于国产供应链的产品的“双轨制”策略已成大势所趋。发行人已积极与国内互联网厂商进行产品测试与验证，为未来市场拓展奠定了坚实的基础。

（5）产品安全可控广受关注，国产自主自强具有战略意义

GPU 芯片是现代信息产业的基础，其安全性直接关系到国家科技战略发展根基。我国一直高度重视高端算力芯片的安全风险，不断捍卫与坚守数据主权。2024 年 12 月，中国互联网协会、中国汽车工业协会、中国通信企业协会、中国半导体行业协会联合发表声明，强调美国无视国际贸易规则、将国家安全概念泛化，建议国内企业审慎采购美国芯片。

2025 年 5 月，美国政府提出拟议的《芯片安全法案》，要求受出口管制的高性能芯片及相关计算产品，在出口、再出口或境外内部转移前，配备能够实现跟踪和位置验证等功能的“芯片安全机制”；2025 年 7 月 17 日，美国众议院中国问题特别委员会（SCCCP）再次强调了该法案的重要性。《芯片安全法案》如正式实施，将使美国能够随时监控芯片位置、限制运行区域甚至远程切断设备，对我国人工智能产业链的安全可控造成重大不利影响。

英伟达是美国出口管制的关键企业，在英伟达宣称 H20 将恢复向中国销售后，2025 年 7 月 31 日，国家网信办就英伟达 H20 可能存在的“漏洞后门”、“追踪定位”及“远程关闭”等技术问题和安全漏洞进行了约谈，再次引发国内对高端芯片产业链技术自主与国家安全的高度关注。因此，在我国对高端算力芯片安全可控的战略方针下，以发行人为代表的国产替代厂商将加速市场渗透。

综上所述，发行人报告期内主要客户群体与境外龙头存在较大差异，不构成直接竞争；发行人产品性能客观上具备替代英伟达 H20 的能力，在国内高性能算力芯片需求缺口较大以及基于供应稳定性、产品安全性需求带来的国产替代趋

势下，以发行人为代表的高性能 GPU 供应商将迎来黄金发展机遇，境外 GPU 龙头对华产品销售政策变化不会对发行人业绩及市场拓展产生重大不利影响。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构执行了如下核查程序：

1、获取发行人收入成本明细表，分析发行人不同产品类型下主要细分产品型号销量、收入、单价及毛利率情况；查阅同行业可比公司公开披露信息，获取其业务发展历史沿革、主营产品类型、细分产品型号，各业务类型的收入占比等信息，并与发行人业务发展情况、产品结构、收入结构进行对比分析；访谈发行人管理层，了解发行人自成立以来的发展战略，了解发行人报告期内收入结构发生变化的原因、当前收入主要集中于曦云 C500 系列的原因以及未来主要产品规划；访谈发行人研发部门和销售部门负责人，了解发行人其他产品及在研产品研发、客户开拓和商业化进展。

2、获取发行人关于销售模式相关管理制度；访谈发行人销售部门负责人，了解发行人不同产品类型下和客户的具体合作模式，了解发行人未来拟主要推广的产品类型以及与合作客户的合作模式；获取报告期内发行人大额销售合同，了解发行人交付产品的类型，自研或外购的情况；访谈发行人主要直接客户及部分终端客户，了解其采购发行人产品应用于智算集群的情况、智算集群的应用领域情况；获取发行人期后销售收入明细以及在手订单明细，了解发行人期后交付的集群以及在手订单中智算集群的情况。

3、走访发行人主要终端集群项目，查看发行人产品的实际使用情况；检索查阅国内各地关于支持采购及使用国产算力芯片的相关政策文件；获取发行人销售合同台账及相关合同，梳理主要客户的复购情况；访谈发行人管理层及商务负责人，结合行业公开信息，了解当前国产 GPU 商业化拓展难点、公司竞争优势和公司未来市场竞争策略。

4、查阅行业研究报告、期刊论文、国家产业政策等，了解不同计算架构、不同 GPU 产品类型对应的市场规模、下游应用领域、主要参与厂商及其市场份额情况，获取发行人产品路线图，分析发行人所在细分市场的市场空间、竞争状

况。

5、获取发行人销售明细表，根据销量等业务数据量化测算发行人的市场占有率；查阅境内外同行业公司公开披露的文件、参数手册等，了解可比公司产品布局、产品参数、业绩规模、客户结构等情况，从多个维度对比分析发行人与境外龙头的差距及在境内市场的竞争优劣势。

6、访谈发行人技术负责人，了解 GPU 及人工智能行业发展趋势、公司核心竞争力及各产品的先进性与壁垒、境外龙头对华销售政策变动的影响等；查阅境外龙头对华销售政策及销售规模、国内针对芯片安全性的相关政策等，从市场供需角度分析相关政策风险。

（二）核查意见

经核查，保荐机构认为：

1、因成立时间、业务结构演变、发展战略、所处发展阶段等方面原因，发行人在产品类型丰富度、细分产品类型、主要销售型号数量方面与同行业可比公司存在一定差异，差异原因具有合理性。发行人曦云 C500 系列产品于 2024 年 2 月正式量产，契合了生成式 AI 算力需求大规模增长的行业趋势，且产品上市后综合性能、稳定性、易用性和可靠性得到客户充分认可并大规模出货，系发行人当前收入主要集中于曦云 C500 系列的主要原因，发行人当前收入结构与行业趋势、芯片产品研发迭代客观规律相符。

2、发行人未来主要推广的产品类型为 GPU 板卡，将采取直销和经销并重的销售模式，持续拓展并加深与服务器 OEM 厂商、经销商客户的合作。

3、发行人终端客户主要为对外提供算力服务或自用的各类集群的资产持有方，发行人产品已在上述算力集群中实际投入使用；发行人已具体说明各地关于采购使用国产算力芯片的支持政策和客户采购发行人产品的主要驱动因素；报告期内发行人核心产品曦云 C500 系列凭借良好的产品综合性能及易用性获得了主要客户持续复购，发行人已结合当前国产 GPU 商业化拓展难点和公司竞争优劣势，充分说明发行人未来发展方向及市场竞争策略。

4、发行人主要产品所处的 GPU 市场空间广阔、成长潜力巨大；在增速强劲的国内人工智能计算市场中，发行人具有领先技术实力和商业化进展，已形成一

定市场地位和品牌知名度；在具有稳定需求的图形渲染 GPU 市场中，发行人亦已规划并研发对应的产品线；发行人采用的通用、灵活的 GPU 架构将长期成为云端智算的主流方案，符合行业技术发展趋势；发行人的产品规划与市场需求发展趋势高度匹配，将长期保持增长潜力，不断提高市场份额。

5、发行人已补充披露公司产品在境内市场占有率的量化测算结果；相比境外龙头，发行人产品在部分性能指标上已经接近或达到国际先进水平，但与境外龙头相比仍存在一定差距；相比境内同行业公司，发行人在产品性能、商业化进展与交付能力、自主性和持续创新能力等方面具备一定竞争优势；境外 GPU 龙头对华产品销售政策变化不会对发行人业绩及市场拓展产生重大不利影响。

2.关于核心技术

根据申报材料：（1）报告期内，发行人主要收入来源为核心产品训推一体芯片曦云 C500 系列的销售，正在研发曦云 C600 系列和 C700 系列，以及智算推理 GPU 曦思 N 系列、图形渲染 GPU 曦彩 G 系列等新产品；（2）2024 年度，训推一体系列产品合计收入占比为 97.28%，其中 GPU 板卡收入占比 68.99%、GPU 服务器收入占比 28.29%，均属于核心技术收入。GPU 服务器业务为公司根据行业客户的具体应用场景需求，向客户交付集成自身 GPU 板卡后的服务器、一体机/工作站等硬件取得的收入；（3）报告期内，公司向前五名供应商采购晶圆制造、封测、HBM、EDA、IP 授权、技术服务等；（4）报告期各期，公司委外研发费用金额为 6,752.31 万元、2,999.21 万元、9,468.99 万元和 1,136.29 万元；（5）公司自主构建的 MXMACA 软件栈高度兼容 CUDA 生态。

请发行人披露：（1）发行人核心技术的形成及演变过程。报告期内外购技术服务、委外研发的具体内容，是否涉及关键环节或关键核心技术；（2）说明发行人 GPU 主要细分产品的技术先进性；（3）发行人不同细分产品所用芯片、IP 及其他组件的收入构成、外购自研比例、核心技术的具体运用情况，与同行业的对比情况。下游晶圆厂商制程工艺变化对在售、在研产品性能，下游客户采购意愿的影响；（4）MXMACA 软件栈如何实现与英伟达 CUDA 生态兼容、目前的生态建设及用户使用情况，在软硬件设计方面是否存在路径跟随和性能提升瓶颈。

请保荐机构简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）发行人核心技术的形成及演变过程。报告期内外购技术服务、委外研发的具体内容，是否涉及关键环节或关键核心技术

1、发行人核心技术的形成及演变过程

自 2020 年成立以来，发行人始终专注于高性能 GPU 芯片及计算平台的研发与设计，对标海外龙头企业，针对 GPU 架构、IP/SoC 设计、互连通信、先进封

装设计、软件架构设计、驱动程序及基础软件等多个专业领域制定了技术开发路线，并根据大模型技术变革与下游多样化需求不断演进和升级核心技术。经过多年的自主研发与创新实践，发行人已积累形成了 8 大类高性能硬件设计核心技术，6 大类高性能软件设计核心技术。未来，发行人将在此技术积累的基础上持续优化和完善，推进技术开发和产品迭代，实现更高性能 GPU 产品的应用和大规模量产。

相关核心技术的形成及发展演进与迭代情况具体如下：

序号	核心技术名称	技术来源	核心技术形成及演变过程		
			2020-2021 年	2022-2023 年	2024-2025 年
高性能硬件设计技术					
1	高性能 GPU 处理器架构设计技术	自主研发	面向 CNN 卷积模型完成核心 GPU IP（XCORE0.5）的设计，采用 HBM2e 实现 2.5D 封装，并完成曦思 N100 产品的研发设计	面向 LLM 及通用计算完成核心 GPU IP（XCORE1.0）的设计，采用 4 颗 HBM2e 实现 2.5D 封装，实现更高性能及更大显存容量，自研 MetaXLink 实现多卡多种灵活拓扑互连，在多个集群应用中展现出高度稳定性和精度收敛能力	进一步提升面向 LLM、通用计算及图形渲染的核心 GPU IP（XCORE1.5/2.0）性能，升级 HBM3e，通过 Chiplet 多种封装，实现更高性能及更大内存容量，打造多场景下产品支撑能力，提升 MetaXLink 的互连能力，并增加高速 Ethernet 接口以优化互连效率
2	高性能 GPU 流处理器设计技术	自主研发	针对云端及边端的 AI 训练和推理，以及高性能计算等多场景需求，初步构建公司整体产品架构、明确产品定义，并开始研发第一代计算指令集，基本涵盖了主流 GPU 所需求的各类矢量计算、访存以及张量计算功能，支持 FP64、FP32、TF32、FP16、BF16、INT8 等多种精度混合运算，满足 OpenCL 标准，核心算力对标同等工艺水平国际 GPU 龙头英伟达 A100 的水准	完成第一代计算指令集的研发，并在第一代计算指令集基础上，开始研究第一代图形指令集，增加图形渲染处理相关指令，增加对光线跟踪的支持，并满足 DirectX12./Vulkan/OpenGL 图形接口标准；为了进一步提高流处理器处理性能，增强指令发射速率，创新性地采用软件控制指令双发射的方式，在指令集中引入临时寄存器缓冲解决双发射引入的寄存器端口冲突问题，大幅降低了硬件设计复杂度，提高了计算吞吐量和能效比	在第一代计算指令集基础上，研发完成第 1.5 代计算指令集（XCORE1.5），加倍 FP16/BF16 tensor 的矩阵运算算力，增加 Tensor 转置指令，并增加了 FP8/BF8 Tensor 指令及 FP4 转换指令以支持混合精度大模型推理；在第 1.5 代计算指令集基础上研发第 2 代计算指令集（XCORE2.5），全新设计计算指令集编码，大幅减轻了编译器的工作负担并提供了更为灵活的编程方式，同时增加了异步 Tensor 拷贝指令及访存乱序能力，极大提高了流处理器访存性能。完成第一代图形指令集的研发，在增加图形渲染指令之外，采用软件控制双发射方式，增强了指令发射性能，并进一步优化部分指令设计；增加 FP8 Tensor 指令，以支持混合精度大模型训练和推理应用，增加 Tensor 转置指令以加速动态张量格式，加倍 FP16/BF16 张量核算力
3	高性能 GPU 多级缓存结构及内存管	自主研发	缓存方面：为初代计算核心设计采用“本地寄存器+多级缓存+高带宽内存”模式，凭借独特的缓存组网技术，构建起多维度的缓存网络架	缓存方面：完成两级缓存结构的研发，流处理器内置 Level-1 指令和数据缓存、Level-1 常量数据缓存、Level-1 矢量数据缓存，流处理器之间通过 Level-2 数	缓存方面：完成三级缓存结构的研发，在流处理器外增加 Level-1.5 数据缓存，用于同组内流处理器之间交互数据提升数据交换效率。三级缓存组网采用定制化的仲裁器，配

序号	核心技术名称	技术来源	核心技术形成及演变过程		
			2020-2021 年	2022-2023 年	2024-2025 年
	理设计技术		构，提供高带宽、低延迟、低功耗的可靠数据访问服务； 内存方面，为初代计算核心自研了内存管理系统，采用多级页表，灵活支持多规格页表混合使用，配合两级地址缓存实现较高寻址效率；支持 ATS 技术和 PCIe 设备的单根虚拟化，为异构计算场景提供全局内存管理方案，具有较强的扩展性	据缓存进行数据共享，各级缓存及 HBM 和 PCIe 之间通过多维总线组网技术以及负载均衡算法实现各通道的高效数据访问。运用灵活的缓存替换策略、访问延迟隐藏技术以及 QoS 管理机制，大幅缩短数据访问延迟，提高并发场景下的数据访问效率； 内存方面，对内存管理系统进行升级，服务对象数量大幅扩容，为大规模计算核心提供了坚实的支撑，满足高密度算力对内存访问寻址的需求；优化了内存管理单元的服务质量控制机制，让各个计算核心的访存寻址性能始终保持在最佳状态	备多元仲裁策略，为不同 GPU 业务流开辟高效的数据传输通路，实现精细化管理。针对显存访问，对存储介质的访问特性进行专项优化，显著提高其访问效率； 内存方面：针对图形渲染任务，内存管理系统增加了超大内存页与部分驻留纹理的支持，让游戏和应用突破显存容量限制，助力实现超高清纹理的流畅加载和沉浸式视觉体验；创新引入页规格预判机制，避免不必要的缓存条目占用和缓存颠簸；针对更高的核心运行速度和算力规模跃升，对时序和面积进行了优化；增加了安全检查和过滤功能，有力保障了集群用户的数据安全
4	高性能 GPU 系统控制及虚拟化技术	自主研发	在曦思 N 系列产品中使用通用嵌入式 CPU 的方式实现 VPU/NPU 的任务调度，提供了多业务通用的任务调度方案；自主设计 MQL 包格式；完成曦云 C 系列产品任务调度与资源管理的架构设计及模块实现；制定并应用了虚拟化技术的方案，可实现 8 个实例切分	完成任务调度与资源管理设计在曦云 C 系列产品中的应用；实现自主研发的高性能硬件任务调度；在稳定性方面，进一步加强错误检测、保护和异常恢复的机制，实现千卡集群大模型训练长时间无中断运行；在虚拟化方面，进一步优化虚拟化技术方面，可支持 16 个实例切分	实现曦彩 G 系列产品中图形管线相关的任务调度与资源管理设计，为系统分发图形渲染任务，并进行同步及依赖性管理，同时满足计算业务及图形业务的高性能需求；在虚拟化方面，对图形管线的模块进行切分，实现图形渲染的虚拟化，进一步满足云游戏、云桌面及小模型推理的场景需求
5	MetaXLink 高速互连技术	自主研发	构建 MetaXLink IP 架构，实现数据和寄存器通路设计，完成基于仿真器和硬件加速器的验证，并在硅后通过原型验证，实现双芯直连的系统拓扑结构	优化 MetaXLink 架构，在每个 MetaXLink 上引入了双模式，通过增强时延隐藏能力有效提高了传输带宽，并通过在单颗 GPU 芯片上实现 7 个高速 MetaXLink 互连接口，互连通信带宽大幅领先国内平均水平，能够提升模型训练效率；研发节点内 2 卡、4 卡、8 卡	将 MetaXLink 技术与光互连技术相结合，从单节点内 4 卡、8 卡互连扩展到节点外 16 卡、32 卡、64 卡互连，打通了 PCIe Switch Box 跨节点通信技术，成功实现跨机架级别光互连超节点技术落地；进一步优化 MetaXLink 架构和设计，在每个 MetaXLink 上引入了双控制器，可以根据系统需求静态配置，实现

序号	核心技术名称	技术来源	核心技术形成及演变过程		
			2020-2021 年	2022-2023 年	2024-2025 年
				GPU 规模下的多种不同互连拓扑	GPU 之间的高速互连或 GPU 和存储设备之间的高速互连；在设计中加入链路调试子模块，便于硅后快速定位。通过激活更多的硬件设计能力实现包括直通互连、交换机互连及直通和交换机混合互连的方式，支撑不同规模及不同模型切分方式下的互连需求
6	芯粒架构与先进封装设计技术	自主研发	在曦思 N100 系列产品中使用 5 层金属和带有深沟槽硅电容硅转接板，实现了 1 倍光刻下的 CoWoS-S 封装，并达到了 HBM2e 单通道 3.6Gbps 的数据速率	沿用 5 层金属和深沟槽硅电容，采用 4 颗 HBM2e，实现了 2 倍光刻下的 CoWoS-S 封装，并在所有 HBM2e 颗粒上实现单通道 3.6Gbps，成功应用于曦云 C500。完成多芯粒大封装（MCM）架构定义及封装技术，并完成硬件设计和验证	使用 5 层厚铜金属和 2 倍密度的深沟槽硅电容，支持更高速率的 HBM3e，单通道最高速率达 6.4Gbps；同时，将深沟槽硅电容应用于更多种类的高速接口 IP，实现了 2 倍光刻下的 CoWoS-S 封装，以及多芯粒合封的大尺寸封装；丰富芯粒架构，提高芯粒互连带宽，将算力芯粒扩展到高速接口芯粒、存储芯粒以及近存计算芯粒，通过 2.5D CoWoS-S、大尺寸封装和/或 3D 封装，结合硅通孔与混合键合技术，实现更高集成度的算力和互连能力
7	高性能 GPU 动态功耗管理技术	自主研发	内嵌工艺监控器件、温度传感器，电压传感器，根据工艺监控器件的震荡频率和芯片漏电完成工艺偏差自适应校准；根据功耗温度实现性能频率自适应调节	优化微架构设计，降低逻辑深度提升基准频率，提高时钟门控效率并降低运行功耗；自研高速功耗和电源网络抖动监控逻辑，快速响应电源抖动；根据工艺监控器件的实时震荡频率优化老化裕量	进一步优化微架构以匹配工艺库，自研高速电压监控模块和定制环形振荡器，使用逻辑自测试减少工艺偏差裕量；定制数字锁频环自适应电源抖动，采用高速电源管理模块，精细化运行时电压频率调节
8	高性能 GPU 验证方法学及智能验证技术	自主研发	通过搭建基础技术体系和工具链，研发核心算法（包括验证方法学的基础逻辑、优化策略等），建立数据处理的基本框架，开创性构建通用 GPU 设计验证核心平台及验证方法学（MIT1.0），自研验证 IP 及	持续优化核心算法和工具链，构建面向高度自动化和多样数据化的第二代设计验证核心平台（MIT2.0），有效实现多层次混合任务、资源和发布的自动化管理，构建可支撑智能验证的大型定制化数据库架构和部署，实现整个项目的	进一步提升核心平台的性能和功能（MIT2.5），重构了多进程架构并增强了多线程处理能力，实现更快的迭代和高并发的任务执行；引入分布式计算、负载均衡等技术，优化了分布式自动化集群管理；进一步完善大型数据库的结构和功能，支持硬件前

序号	核心技术名称	技术来源	核心技术形成及演变过程		
			2020-2021 年	2022-2023 年	2024-2025 年
			验证自动化(VIP1.0),实现除 PCIe 以外的绝大多数总线协议的突破,超过 60%的验证平台可以自动化生成	功能、性能验证的全数据链高效、检索和管理机制;开发人机交互界面,实现基础的可视化展示与决策支持;针对丰富场景提供定制化验证解决方案,推出自研高性能验证 IP(VIP1.5),将 IP/SoC 的设计验证收敛速度提升 20%以上	后端研发的各类型数据存储和可视化分析需求;开始研发智能验证的“智慧大脑”,实现跨节点的智能化协调与全局优化,显著提升系统的智能水平和自动化能力;推出新一代自研高性能智能验证 IP (VIP2.0),集成智慧处理引擎,支持更丰富的总线协议,同时开始建设围绕核心产品的设计验证生态系统
高性能软件设计技术					
9	GPU 异构计算编程语言与开发环境	自主研发	基于曦思 N100 硅前验证平台,完成底层软件框架设计,以及部分业务支撑软件开发;完成曦云 C500 软件需求分析,基础算法选择和整体架构设计,形成自主研发的 GPU 异构计算编程框架;同时基于先进的 qemu 虚拟仿真平台,在硅前验证阶段开发完整软件技术栈,在打通首个端到端测试的基础上,完成 1,100 余个基础测试案例,及多个算子功能和峰值性能验证;确保 GPU IP 在硅前设计阶段可以和应用场景对接,并大幅缩短软件开发和市场交付时间	完成曦思 N100 配套软件栈的公开发布,每 30-90 天发布一个新版本;持续优化曦思 N100 的多任务性能、高分辨率性能;曦云 C500 形成完整 MXMACA 2.x 版本软件栈,包括驱动技术、虚拟化、人工智能编译器、编程接口、高性能数学库、通信库等;执行大多数重要训练场景的性能优化,超过 6,000 个 CUDA 开源应用的兼容性测试支持;支持 PyTorch、Tensorflow、PaddlePaddle、MindSpore、DeepSpeed、Megatron-LM 等多个 AI 框架,实现约 50 个科学计算任务的迁移和优化;启动曦彩 G100 相关软件的研发	进一步优化提升软件核心技术包括驱动、人工智能编译器、高性能数学库、通信库、软硬件协同优化等技术,软件稳定性和兼容性大幅提升;曦云 C500 系列快速支持覆盖大量客户场景,全面优化深度神经网络、大模型、通用计算等领域的计算性能;逐步开源 MXMACA 软件栈,出版《MXMACA C/C++ 程序设计入门教程》,支持多个开源大模型训练、AI 推理、算子库等应用和软件;建立开发者社区和“沐曦大学”,覆盖中国超过半数的 C9 高校;交付曦彩 G100 相关软件;完成曦云 C600 部分软件开发工作,并开始曦云 C700 软件栈开发工作
10	人工智能编译器与编译技术	自主研发	曦思 N100 系列完成基于 OpenVX 的编译器开发与验证工作;曦云 C500 系列根据架构分析选择 LLVM12 框架作为基础,支持编译器通用功能和曦云 C500 指令集,完成动态内存分配、指令调度、代	实现编译器的全面覆盖测试,并设计实现形式化验证、差分验证;初步实现人工智能指令调度功能,算子融合、动态内存分配、动态调优技术等多种优化技术;初步实现曦彩 G100 的 Vulkan 图形编译器和 MXMACA 编译器	优化 C/C++编译器,实现 LLVM19 的版本升级、动态调优技术、人工智能编译器优化技术、多 IR 支持等;完成曦彩 G100 的 Vulkan、OpenGL、DirecX 编译器交付;完成 triton 编译器支持和优化,发布 Fortran+OpenACC 编译器;研发 OpenXLA 编译器;初步支持 TVM

序号	核心技术名称	技术来源	核心技术形成及演变过程		
			2020-2021 年	2022-2023 年	2024-2025 年
			码结构优化等一系列编译优化技术，支持 2,000 个贯通基础测试用例；完成近 13,000 个编译器测试用例		AI 编译器；完成曦云 C600 C/C++编译器功能完备性并完成大部分性能优化；开始曦云 C700 编译器研发工作；完成各类框架的持续升级工作
11	GPU 高性能通信库设计与优化技术	自主研发	基于曦思 N100，实现 GPU 卡间通信的软件接口基本功能模拟支持，包含地址映射和数据接收和发送等；基于曦云 C500 硬件设计，开展多机多卡先进通信技术研发，形成通信软件的技术架构，确保行业最新规范和发行人硬件通信能力充分对接	初步完成 MCCL 等通信技术的自主研发，支持 2 卡、4 卡、8 卡单节点集合通信，实现并优化曦云 C 系列多节点集群的通信性能，同时支持基于虚拟机的集群部署；完成超过 50 个人工智能模型的训练任务，可支持 130B 大模型训练，通信库具备较强的兼容性、稳定性和高性能	设计多卡通信路径的自动搜索优化，达到硬件链路带宽最优；完成服务器内多卡通信与计算，超节点内通信和集群网络通信的资源平衡和算法优化；完成集合通信技术的生态融合和产品化落地，适配兼容数十种服务器，多款网卡和交换机、多机计算超节点，以及超大集群采用的多种专用高速计算网络拓扑，有力保证大规模 AI 业务的开展。同时，实现发行人 GPU 和其他厂商计算卡间的通信功能兼容，并不断对 DeepSeek、Qwen 等热点应用开展性能优化
12	GPU 高性能数学库设计与优化技术	自主研发	实现数学库高研发效率底层开发框架；完成 DNN、BLAS 数学库的框架设计及部分 GPU kernel 的汇编代码实现	实现对曦云 C500 常见数学库的功能全覆盖，优化 CNN、NLP、LLM、多模态等重点人工智能大模型的训练、推理性能，完成超过 200 个人工智能训练和推理场景的性能初步优化；优化曦彩 G100 的部分算子	完成曦云 C500 的算子库性能全覆盖，支持客户和业界扩展核心场景落地，提供高性能的功能接口；完成曦云 C600 算子库的常见场景性能优化和开源算子库适配；完成曦云 C700 部分算子的性能优化
13	GPU 核心驱动程序	自主研发	完成核心驱动程序支持，包括虚拟化管理机制、内存管理机制、智能任务调度系统、实时状态监控与性能分析工具等；完成用户态驱动程序基础架构设计，支持队列、设备、内存、上下文等基本功能，初步兼容 CUDA 功能，提供了海量人工智能应用可运行的底层软件支持	基于曦云 C500 系列，完成核心技术落地，支持多种用户场景。采用多模块硬件自动容错机制、关键错误监控机制、异构内存管理与页表缓存技术、多种页表大小自适应混合技术等，实现服务器的 24x7 稳定可靠运行；逐步开展生态兼容性，支持十余种操作系统、多款国产 CPU 等	继续深化核心技术，实现分布式显存管理机制与跨节点直接访问技术、动态时分虚拟化技术、高性能计算图技术、计算队列灵活调度技术以及超节点内的细粒度 GPU 调度技术等一系列关键技术，有力保证了关键 AI 应用的性能和用户集群对显卡资源的充分使用；进一步提高生态兼容性，支持 40 余款服务器、多种 X86/ARM CPU，15 种以上操作

序号	核心技术名称	技术来源	核心技术形成及演变过程		
			2020-2021 年	2022-2023 年	2024-2025 年
					系统、60 多个 Linux 内核，基本完成了对国产生态的兼容；交付曦彩 G100 图形驱动，完成 GPU 驱动全品类支持
14	GPU 软硬件协同优化技术	自主研发	完成软硬件协同优化的系统方案设计，为应用开发者设计提供细粒度的硬件运行时性能指标；对接硬件功能完成软件底层设计并设计提供友好的用户调试界面	开发完成软硬件协同优化的完整工具链，实现对曦云 C500 系列软硬件调优过程的完整可视化表达；发布 mcTracer、mcProfiler 等工具，帮助用户在应用运行时快速理解硬件资源的细粒度调度，并做针对性优化；同时开发完成 GPU 时钟级性能追踪器 cycle-trace，以纳秒级精度分析每个 GPU 指令，帮助开展极致性能优化	持续迭代软硬件协同优化工具链，新增算子级别和 AI 框架级别优化工具集：新增算子优化工具 Kernel Selection，AI 框架级工具 Triton 自动调优器和初版 DeepSeek 自动调优器，帮助热点应用获得极致硬件性能；发布 autoTuner，支持大模型推理、大模型训练性能调优，可快速找到最佳并行方案；发布 cuda2maca 工具试用版，提供基于大语言模型对基础算子的自动优化功能

发行人核心技术系经过长期经验积累和技术迭代逐步形成，均为自主研发，不存在来源于其他企业或任何外部第三人的情形。发行人相关核心技术已全面覆盖 GPU 芯片及计算平台关键设计环节，具有独立自主、持续可控的研发与技术迭代创新能力。发行人将立足公司主营业务，紧密围绕 GPU 行业及下游行业发展趋势加强技术研发，借助储备技术不断推动在研产品持续升级迭代，稳固并拓展自身的市场地位。

2、报告期内外购技术服务、委外研发的具体内容，是否涉及关键环节或关键核心技术

报告期各期，发行人技术及咨询服务费金额分别为 1,217.69 万元、750.73 万元、3,904.16 万元及 790.32 万元，其中各期技术服务费金额分别为 741.00 万元、542.27 万元、2,980.53 万元及 643.61 万元。技术服务费主要为发行人采购的用于支持发行人开展研发活动的技术服务，如智能平台软件开发服务、共建实验室费用、集成电路仿真验证费用等。报告期内，发行人各期主要的外购技术服务具体内容、确认的技术服务费及占当期技术服务费总额的比例如下：

单位：万元

技术服务供应商	技术服务内容	技术服务费	占当期技术服务费比例
2025 年 1-3 月			
技术服务供应商 1	智能平台软件开发服务	279.56	43.44%
技术服务供应商 2	集成电路仿真验证服务	229.03	35.59%
技术服务供应商 10	性能优化技术服务费	100.00	15.54%
小计		608.59	94.56%
2024 年度			
技术服务供应商 1	智能平台软件开发服务	1,025.06	34.39%
技术服务供应商 2	集成电路仿真验证服务	710.00	23.82%
技术服务供应商 3	共建实验室技术支持费用	558.94	18.75%
技术服务供应商 8	HTOL 测试服务费	156.96	5.27%
技术服务供应商 6	算子优化技术服务	138.55	4.65%
小计		2,589.51	86.88%
2023 年度			
技术服务供应商 4	IP 维保费用	209.77	38.68%

技术服务供应商	技术服务内容	技术服务费	占当期技术服务费比例
技术服务供应商 9	共建实验室技术支持费用	77.67	14.32%
技术服务供应商 7	IP 维保费用	35.92	6.62%
技术服务供应商 6	算子优化技术服务	35.02	6.46%
技术服务供应商 11	IP 维保费用	31.51	5.81%
小计		389.89	71.90%
2022 年度			
技术服务供应商 5	EDA 硬件验证云技术支持服务	262.99	35.49%
技术服务供应商 2	集成电路仿真验证服务	262.50	35.42%
技术服务供应商 4	IP 维保费用	80.87	10.91%
技术服务供应商 12	编译器测试服务	74.53	10.06%
技术服务供应商 11	IP 维保费用	38.07	5.14%
小计		718.97	97.03%

上述主要外购技术服务商服务是否涉及关键环节或关键核心技术的情况如下：

序号	供应商	外购技术服务具体内容	是否涉及关键环节或关键核心技术	不涉及关键环节或关键核心技术的原因
1	技术服务供应商 1	智能平台软件开发服务	否	曦云 C500 系列产品量产商业化过程中软件栈开发工作量需求较大，发行人外购此技术服务进行算力服务应用层面通用性、执行性技术扩展，不涉及研发关键环节或关键核心技术。
2	技术服务供应商 2	集成电路仿真验证服务	否	Zebu 仿真器是仿真验证工具，在发行人自身的仿真器不够使用时考虑短期外租，属于标准化硬件租赁及常规技术服务，不涉及关键环节或关键核心技术。
3	技术服务供应商 3	共建实验室技术支持费用	否	校企合作进行前沿课题研究，内容限于基础理论研究和应用层技术储备，不涉及芯片研发的关键环节或关键核心技术。
4	技术服务供应商 4	IP 维保费用	否	为外购的 IP 提供常规维保支持服务，服务内容属于技术维护范畴，订阅制付费，不涉及芯片研发的关键环节或关键核心技术。
5	技术服务供应商 5	EDA 硬件验证云技术支持服务	否	发行人租用了 Cadence 的 Palladium 仿真器平台，同时采购对应的仿真算力服务，属于支持研发活动的特定工具及常规验证环境搭建服务，不涉及研发关键环节或关键核心技术。
6	技术服务供应商 6	算子优化技术服务	否	增加 FFT 算子库不常见算子库的性能覆盖度，属于不紧急、不常用的算子优化，不涉及芯片研发的关键环节或关键核心技术。

序号	供应商	外购技术服务具体内容	是否涉及关键环节或关键核心技术	不涉及关键环节或关键核心技术的原因
7	技术服务供应商 7	IP 维保费用	否	为外购的 IP 提供常规维保支持服务，服务内容属于技术维护范畴，订阅制付费，不涉及芯片研发的关键环节或关键核心技术。
8	技术服务供应商 8	HTOL 测试服务费	否	HTOL 是芯片可靠性测试，包括高低温的测试，公司无相关设备故外购技术服务，为单次项目制服务，不涉及芯片研发的关键环节或关键核心技术。
9	技术服务供应商 9	共建实验室技术支持费用	否	校企合作共建实验室是为了培养前沿课题研究、技术储备，不涉及芯片研发的关键环节或关键核心技术。
10	技术服务供应商 10	性能优化技术服务费	否	项目制大语言模型调优服务，属于产品应用层面优化，为了增加曦云 C500 系列产品的普适性，不涉及芯片研发的关键环节或关键核心技术。
11	技术服务供应商 11	IP 维保费用	否	为外购的 IP 提供常规维保支持服务，服务内容属于技术维护范畴，订阅制付费，不涉及芯片研发的关键环节或关键核心技术。
12	技术服务供应商 12	编译器测试服务	否	编译器测试需要较多人力投入，故从外部获取人力资源，属于研发支持，不涉及芯片研发的关键环节或关键核心技术。

报告期内，发行人委外研发费用主要系发行人基于自身研发需求，为实现高效研发、专注于核心技术迭代的目的，将部分非核心的研发工作委托给外部具有相关技术实力及资源的企业、高校或研究所完成。报告期各期，发行人主要委外研发项目基本情况、委外研发费用及占当期委外研发费用总额的比例如下：

单位：万元

委外研发供应商	委外研发项目	委外研发费	占当期委外研发费比例
2025 年 1-3 月			
委外研发供应商 5	多个大模型的推理任务的性能优化开发	450.00	39.60%
委外研发供应商 8	图形 GPU Vulkan CTS Shader 数据集建设、初步功能分析及详细性能分析	156.22	13.75%
委外研发供应商 10	基于沐曦 MXMACA 运算平台的 Triton 算子开发和优化项目	107.54	9.46%
委外研发供应商 3	基于沐曦芯片的人工智能解决方案开发项目	93.50	8.23%
委外研发供应商 10	基于沐曦 MXMACA 运算平台的大模型算法部署优化项目	81.85	7.20%
小计		889.11	78.25%
2024 年度			

委外研发供应商	委外研发项目	委外研发费	占当期委外研发费比例
委外研发供应商 1	后端设计	2,641.51	27.90%
委外研发供应商 1 ^注	后端设计	2,000.00	21.12%
委外研发供应商 4	UBB 开发	850.00	8.98%
委外研发供应商 9	infini-acc 推理引擎	452.83	4.78%
委外研发供应商 3	基于沐曦芯片的人工智能解决方案开发项目	346.50	3.66%
小计		6,290.84	66.44%
2023 年度			
委外研发供应商 1 ^注	后端设计	1,200.00	40.01%
委外研发供应商 6	AlphaSparse 计算库向沐曦的 MXC500 平台移植和优化	481.13	16.04%
委外研发供应商 7	面向同态密码和基因测序的可重构 SIMD 架构设计	242.72	8.09%
委外研发供应商 12	后端设计	240.00	8.00%
委外研发供应商 14	国产芯片在数字农业领域应用的技术研发和应用场景开发	200.00	6.67%
小计		2,363.85	78.82%
2022 年度			
委外研发供应商 1	后端设计及封装设计	3,015.14	44.65%
委外研发供应商 2	芯片后端设计服务	2,550.00	37.76%
委外研发供应商 11	数据中心先进计算架构联合研究项目	300.00	4.44%
委外研发供应商 7	面向同态密码和基因测序的可重构 SIMD 架构设计	242.72	3.59%
委外研发供应商 13	GPU 模糊测试框架研发	221.28	3.28%
小计		6,329.13	93.73%

注：该项目受 2024 年美国政府禁令的影响暂时无法流片，双方已签署补充协议确认终止合同中与曦彩 G100 的相关内容。

上述主要委外研发供应商服务内容及是否涉及关键环节或关键核心技术的情况如下：

序号	委外研发商	委外研发项目	委外研发具体内容	是否涉及关键环节或关键核心技术	不涉及关键环节或关键核心技术的原因
1	委外研发供应商 1	后端设计及封装设计	后端设计及封装设计	否	后端设计及封装设计核心环节由公司自主负责，委外部分为在公司设定的顶层框架下进行的后端模块级别物理设计实现，以及基于芯片外围信

序号	委外研发商	委外研发项目	委外研发具体内容	是否涉及关键环节或关键核心技术	不涉及关键环节或关键核心技术的原因
					息进行的延展性设计及标准化布局布线设计，技术路径明确、流程标准化，不涉及研发的关键环节或关键核心技术。
2	委外研发供应商 1 ^注	后端设计	后端设计	否	后端设计核心环节由公司自主负责，委外部分为在公司设定的顶层框架下进行的模块级别物理设计实现工作，技术路径明确、流程标准化，不涉及研发的关键环节或关键核心技术。
3	委外研发供应商 1	后端设计	后端设计	否	后端设计核心环节由公司自主负责，委外部分为在公司设定的顶层框架下进行的模块级别物理设计实现工作，技术路径明确、流程标准化，不涉及研发的关键环节或关键核心技术。
4	委外研发供应商 2	芯片后端设计服务	后端设计	否	后端设计核心环节由公司自主负责，委外部分为在公司设定的顶层框架下进行的模块级别物理设计实现工作，技术路径明确、流程标准化，不涉及研发的关键环节或关键核心技术。
5	委外研发供应商 3	基于沐曦芯片的人工智能解决方案开发项目	完成算子性能、编译优化、大模型性能优化等软件方案	否	属于软件研发非核心的算子优化测试类环节中不紧急、不常用的部分，技术通用性强，不涉及研发的关键环节或关键核心技术。
6	委外研发供应商 4	UBB 开发	定制化设计和验证适用沐曦产品所要求的 UBB	否	属于行业应用适配及优化类环节，支持公司产品基于外部结构的互联，以及服务器整机的设配，设计路径明确，不涉及研发的关键环节或关键核心技术。
7	委外研发供应商 5	多个大模型的推理任务的性能优化开发	针对多个大模型训练任务和推理任务的优化	否	属于软件研发中特定任务的优化，技术通用性强，不涉及研发的关键环节或关键核心技术。
8	委外研发供应商 6	AlphaSparse 计算库向沐曦的 MXC500 平台移植和优化	AlphaSparse 计算库向沐曦的 MXC500 平台移植和优化	否	属于软件研发非核心的算子优化测试类环节中不紧急、不常用的部分，技术通用性强，不涉及研发的关键环节或关键核心技术。
9	委外研发供应商 7	面向同态密码和基因测序的	从同态密码的需求、基因测序的需求、同态密	否	委托大学、研究院进行技术储备性质的前沿课题研究，涉及不超过 10 条指令，不涉及研发的关键环节或关键

序号	委外研发商	委外研发项目	委外研发具体内容	是否涉及关键环节或关键核心技术	不涉及关键环节或关键核心技术的原因
		可重构 SIMD 架构设计	码算法基本算子在 SIMD 架构上的可重构硬件加速、基因测序算法在 SIMD 架构上的可重构硬件加速等方向,开展技术开发工作		核心技术。
10	委外研发供应商 8	图形 GPU Vulkan CTS Shader 数据集建设、初步功能分析及详细性能分析	Vulkan CTS 的 GPU shader 数据库, 分析 Vulkan CTS 的 Shader 编译及其错误分析, 详细性能分析和优化	否	属于软件研发非核心的算子优化测试类环节中不紧急、不常用的部分, 任务重复性高、技术通用性强, 不涉及研发的关键环节或关键核心技术。
11	委外研发供应商 9	infini-acc 推理引擎	提供优化系统吞吐、支持算子融合、支持显存管理、支持主流量化方法	否	属于软件研发框架级优化测试中不紧急、不常用的部分, 技术通用性强, 不涉及研发的关键环节或关键核心技术。
12	委外研发供应商 10	基于沐曦 MXMACA 运算平台的 Triton 算子开发和优化项目	图像生成相关模型推理的 AI 编译器 Triton 算子优化, vLLM 语言大模型推理的 AI 编译器 Triton 算子优化, 基于 AI 编译器 Triton 支持 Pytorch Benchmarks 推理性能优化	否	属于软件研发非核心的算子优化测试类环节中不紧急、不常用的部分, 不涉及研发的关键环节或关键核心技术。
13	委外研发供应商 11	数据中心先进计算架构联合研究项目	解决服务器 xPU 资源池化、云化过程中遇到的难题, 使服务器 xPU 资源能更加灵活地分配、部署, 并实现无缝对接现有的云服务器生态环境	否	属于其他占比极低的环节, 委托研究机构研究潜在的新型计算设备, 属于技术储备性质的前沿课题研究, 不涉及研发的关键环节或关键核心技术。

序号	委外研发商	委外研发项目	委外研发具体内容	是否涉及关键环节或关键核心技术	不涉及关键环节或关键核心技术的原因
14	委外研发供应商 12	后端设计	后端设计	否	后端设计的核心工作由公司自主负责，委外部分为在公司设定的顶层框架下进行的模块级别物理设计工作，技术路径明确、流程标准化，不涉及研发的关键环节或关键核心技术。
15	委外研发供应商 10	基于沐曦 MXMACA 运算平台的大模型算法部署优化项目	基于 vllm 库的语言大模型部署，基于 HF transformers 库的语言大模型部署	否	属于软件研发非核心的算子优化测试类环节中不紧急、不常用的部分，技术通用性强，不涉及研发的关键环节或关键核心技术。
16	委外研发供应商 13	GPU 模糊测试框架研发	开发面向 GPU 体系的模糊测试框架，并完成验证，形成测试方案并进行后续的技术支持	否	属于软件研发的其他测试环节，为人工密集型工作，技术通用性强，不涉及研发的关键环节或关键核心技术。
17	委外研发供应商 14	国产芯片在数字农业领域应用的技术研发和应用场景开发	以农业应用场景为基础，探索沐曦自研芯片在植保喷施作业、农林信息获取、乡村土地资源调查等场景的应用	否	属于行业应用适配及优化类环节，强化公司在农业领域的扩展，不涉及研发的关键环节或关键核心技术。

注：该项目受 2024 年美国政府禁令的影响暂时无法流片，双方已签署补充协议确认终止合同中与曦彩 G100 的相关内容，序号 3 对应合同为该项目的补充合同。

发行人致力于自主研发全栈高性能 GPU 芯片及计算平台，围绕不同的应用场景，构建丰富的产品矩阵。发行人产品研发的技术流程主要包括硬件研发设计阶段和软件研发设计阶段，具体情况如下：

（1）硬件研发设计阶段

发行人硬件研发设计阶段分为芯片前端设计、中端设计、后端设计，其中：

1) 前端设计阶段中，仅 C 模型算法实现环节存在自主研发和委外研发的情形，仿真验证环节存在自主研发和外购技术服务的情形。其中，C 模型算法实现环节主要由发行人进行自主研发，仅涉及部分测试的内容委托第三方进行，相关内容不涉及公司的关键核心技术；仿真验证环节主要由发行人进行自主研发，仅在其自身仿真器不能满足使用需求时考虑短期外租并由供应商提供技术支持服

务，仿真器属于支持研发活动的特定工具，不涉及关键环节或关键核心技术；

2) 中端设计阶段均由发行人完全自主研发，不涉及外购技术服务或委外研发的情形；

3) 后端设计阶段中的模块布局布线环节存在自主研发和委外研发的情形；后端设计中的其他重要环节，例如全芯片布局、时钟树综合、物理验证等均为自主研发。GPU 芯片研发具有周期长、投入大的特点，其中前端设计及中端设计中的代码开发环节占据了较大的工作量比重。芯片后端设计中的模块布局布线环节，相较于其他环节技术难度较低，为研发过程中的非核心工作，但需要较多经验丰富的人员投入。同时，模块布局布线环节并非持续性的研发需求，而是流片前才需要进行的环节，呈现出一定时间段内“爆发式”的需求。考虑到后端设计在芯片硬件研发过程中的中后期环节，为最大限度地提高公司自有研发人员的人效比，避免后端设计团队在代码开发阶段出现人员闲置，发行人将后端设计中模块布局布线环节中技术难度相对较低但需要较多相关经验人员投入的部分委外研发，有利于发行人将自有技术、研发人员聚焦于价值量更高的核心技术的研发和储备上，亦避免了人力资源的浪费，不涉及公司研发的关键环节及关键核心技术，具有合理性。

(2) 软件研发设计阶段

发行人软件研发设计阶段主要包括固件设计、驱动程序、编译器、数学库、通信库、解决方案功能完善/性能优化等，其中绝大部分环节为发行人自主研发，仅部分涉及测试、调优等技术难度相对较低、且非重要的环节涉及自主研发与外购技术服务、委外研发，相关环节分类及情况如下：（1）不紧急、不常用算子优化测试类环节：重要、常用、紧急算子为自主研发，不紧急、不常用的部分采用外购技术服务或委外研发；（2）行业应用适配及优化类环节：委外部分主要为结合委外研发商在特定行业的积淀提供 ISV 应用支持，外购技术服务部分主要为量产后商业化过程中的平台软件开发及性能技术优化等服务。以上相关外购技术服务、委外研发的内容不涉及研发的关键环节或关键核心技术。

除前述硬件研发设计阶段、软件研发设计阶段涉及的外购技术服务及委外研发外，发行人还存在部分校企合作共建实验室、委托研究机构进行前沿课题研究

等涉及外购技术服务、委外研发，以及外购 IP 涉及的维保支持服务涉及外购技术服务。其中，校企合作共建实验室、委托研究机构进行前沿课题研究等为技术储备性质，不涉及研发的关键环节或关键核心技术；外购 IP 涉及的维保支持服务为技术维护，不涉及研发的关键环节或关键核心技术。

由此，发行人外购技术服务、委外研发所在的研发环节，主要为发行人自主研发，仅将该环节中的一部分技术难度较低、非核心的测试或优化内容委外，以加快整体研发进度。发行人是国内高性能通用 GPU 的领导者，在 GPU 领域处于国内领先水平，当前正处于快速发展阶段，为保持技术先进性和市场竞争力，发行人不断加大技术投入，在团队不懈的技术攻坚下，发行人成为了国内少数几家全面系统掌握了通用 GPU 架构、GPU IP、先进制程 GPU 芯片及其基础系统软件研发、设计和量产核心技术的企业之一。在发行人产品研发不断发展的过程中，发行人将部分技术难度较低的非核心环节委外研发或外购技术服务，可以使发行人聚焦于核心技术研发和迭代，加快产品整体开发进度。

发行人可比公司及其他部分同行业半导体设计公司外购技术服务、委外研发情况如下：

序号	同行业公司	研发费用类别	金额占研发费用比例			主要内容 ²
			T+2 年度	T+1 年度	T 年度 ¹	
1	寒武纪	委托、合作研发支出	0.58%	4.32%	4.79%	委托、合作研发支出主要系公司基于其未来产品定义和设计需求委托供应商 A 进行技术开发，委托供应商 A 研发的技术是智能芯片的组件模块，委托研发的组件模块须在公司的底层技术及其他组件模块的共同支持下方能发挥作用。公司已有芯片产品中有两款（目前正处于样片测试）已使用委托研发模块。
2	海光信息	技术服务费	19.18%	21.00%	3.11%	研发费用中的技术服务费主要为委托技术开发费、芯片后端设计、芯片样片的功能测试、定制版图设计等外包服务等费用。为了缩短产品的研发周期、降低总体研发成本、规避产品的质量风险，公司根据业务需求将部分非核心研发任务以技术服务采购方式，委托给具有专业技术能力的第三方开发完成。委外研发涉及基板开发、与工艺相关的后端设计、产品

序号	同行业公司	研发费用类别	金额占研发费用比例			主要内容 ²
			T+2 年度	T+1 年度	T 年度 ¹	
						测试、测试板卡等环节，其中后端设计环节包括定制计算设计与设计验证服务、PD Flow 设计服务等，产品测试包括处理器验证测试主板测试、主板物理特性测试等。
3	龙芯中科	未明确披露	/	/	/	公司与中科院计算所存在开展合作研发的情况，建立了以“龙芯”为课题的联合实验室，在处理器前瞻设计技术、处理器性能分析优化、基于龙芯处理器的基础软件研发优化等方面开展联合研究工作，属于前瞻理论研究，不涉及具体产品开发。
4	景嘉微	设计费	-	-	9.87%	未披露设计费涉及具体内容
5	摩尔线程	技术服务费	9.40%	14.73%	13.20%	未披露技术服务费涉及具体内容
6	盛科通信	未明确披露	/	/	/	公司将部分后端设计业务委托芯片量产代工商完成，公司能够更加专注于价值量更高的集成电路的产品定义、前端设计和客户支持业务。
7	成都华微	外协费中的技术服务费；委外设计费	9.57%	14.01%	21.07%	外协费中的技术服务主要为电路分析等。 委外设计费主要系为加快研发效率，报告期内公司存在将研发项目部分辅助内容通过委托外部第三方的方式进行，主要包括 IP 采购以及软件开发服务等内容。 公司采购技术服务内容包括电路分析、后端设计服务、软件设计服务、委托开发服务、IP 采购等。
8	中微半导	技术开发费；检测及服务费用	3.09%	3.98%	4.63%	未披露技术开发费、检测及服务费用具体内容。 除技术开发费、检测及服务费用外，截至招股说明书签署日，履行的委托开发及合作开发合同涉及基于中微 MCU 智能产品开发及产学研合作、低功耗蓝牙芯片合作开发、共建“中微半导-哈工大（深圳）人工智能芯片实验室”等。
平均值 ³			6.97%	9.67%	9.45%	-
最大值			19.18%	21.00%	21.07%	-
发行人			13.82%	5.07%	11.57%	-

注 1：成都华微列示数据期间为 2020 年-2022 年，系其年度报告未列示外协费细分构成，故采用招股说明书数据；景嘉微列示数据期间为 2013 年-2015 年，系其仅于招股说明书披露

设计费，故采用招股说明书数据，除列示区间外，其 2011 年度、2012 年度设计费占研发费用的比例分别为 46.94%、38.80%，处于较高水平；其余同行业公司 & 发行人列示数据期间为 2022 年-2024 年；同行业公司年度报告未披露 2025 年一季度研发费用构成情况，故未列示；

注 2: 寒武纪相关信息来自其《关于公司 2020 年年度报告的信息披露监管问询函回复公告》，其他同行业公司信息来自于首次公开发行股票招股说明书及问询回复；

注 3: 平均值及最大值计算剔除未明确披露相关数据的同行业公司。

报告期各期，发行人技术服务费占研发费用的比例为 1.14%、0.78%、3.31% 及 2.95%，委外研发费占研发费用比例分别为 10.43%、4.29%、10.51% 及 5.21%，二者合计占研发费用比例分别为 11.57%、5.07%、13.82% 及 8.17%。同行业公司列示区间各期，涉及技术服务及委外研发科目金额占研发费用的比例平均值区间为 6.97%-9.67%，最大值区间为 19.18%-21.07%，发行人技术服务费及委外研发费占研发费用比例与同行业公司相比处于合理区间。此外，由于芯片研发具有周期长、投入大的特点，对技术服务和委外研发的需求在不同阶段会有所波动，因此其在各期的占比也存在较大差异，发行人的情况符合这一行业特点。

此外，由上表可见，同行业公司亦普遍存在委托后端设计、测试验证、软件优化、行业应用适配、前沿课题研究等外部化研发投入，如海光信息、盛科通信和成都华微等公司也存在将部分后端设计环节委外的情况，与发行人外购技术服务及委外研发内容类似。由此，发行人将部分非核心的后端设计、测试、优化、行业应用适配及前沿课题研究等以外购技术服务或委外研发的形式进行，与其他同行业公司的做法一致，报告期内发行人技术服务费及委外研发费占研发费用的比例相较于同行业公司亦处于合理区间，与行业惯例相符。

综上，发行人涉及外购技术服务、委外研发的环节主要仍为发行人自主研发，相关环节仅非核心、技术难度较低的部分涉及外购技术服务或委外研发，发行人外购技术服务、委外研发的内容不涉及关键环节或关键核心技术。

（二）说明发行人 GPU 主要细分产品的技术先进性

发行人致力于自主研发全栈高性能 GPU 芯片及计算平台，主营业务是研发、设计和销售应用于人工智能训练和推理、通用计算与图形渲染领域的全栈 GPU 产品，并围绕 GPU 芯片提供配套的软件栈与计算平台。报告期内，发行人主要销售产品为曦云 C 系列训推一体 GPU 产品，2022-2024 年及 2025 年 1-3 月，该系列产品合计形成营业收入 105,079.60 万元，占三年一期主营业务收入合计额的

94.31%。发行人训推一体 GPU 板卡的技术先进性分析如下：

1、基于通用、灵活、自主的架构，GPU 产品性能全面提升

GPU 架构是决定 GPU 性能的核心。英伟达、AMD 等国际龙头均基于 GPU 架构进行了数次迭代，从而在市场竞争中持续保持领先的产品性能。发行人全系列产品采用统一的 GPU 计算架构，迄今已自主研发了两个大代际、五个小代际流处理器架构。相比采用 ASIC/DSA 等其他技术路径的 AI 芯片企业，发行人通用、灵活的 GPU 架构能够高效适应云端快速迭代的复杂算法和大量非传统模型结构，从而支持更为广泛的人工智能应用场景。

发行人是国内少数坚持“从零到一”自主研发核心 GPU IP 的集成电路设计企业，经过多年产品设计及落地经验的积累，已形成了全面覆盖计算（人工智能训练、推理、通用计算）和渲染场景的核心 GPU IP，打造了由超过 600 条（XCORE1.0 计算 GPU）、800 条（XCORE2.0 渲染 GPU）指令组成自主安全可控的 MXMACA 指令集，包含标量计算和控制指令、矢量计算指令、张量运算指令、共享内存访问指令、全局内存访问指令、私有内存访问指令、图形处理相关指令等，具有强大的通用性和灵活性。

基于 SIMT（单指令多线程结构）以及 MXMACA 指令集，发行人定义了 GPU 可以执行的多种操作，如算术运算、逻辑运算、内存访问、分支跳转、原子操作、指令并发、异步通信等，通过高精度、低精度和混合精度的灵活组合，能够支持通用基础大模型、语言大模型、图像大模型、多模态大模型、行业大模型、智慧城市、智慧安防、智慧视觉等当前复杂多样的计算需求，亦为未来的技术发展和应用场景提供了坚实的基础。同时，发行人采用创新性双发射架构，通过软件调度机制优化指令流。编译器在编译阶段智能识别指令间操作，精准判断双发射可行性，并动态插入双发射指示位，能够大幅简化硬件复杂度并显著提升了指令发射效率和执行单元利用率。

凭借灵活高性能、高通用性的架构设计，发行人的产品迭代能够快速响应市场变化，技术演进与市场需求同频共振，在产业升级变革中始终保持先发优势。

2、国内先进的计算能力

在算力方面，国外龙头企业始终保持着领先身位，以英伟达为例，其从

Ampere 架构到 Hopper 架构，再到当前最先进的 Blackwell 架构，每一代产品在算力性能上都实现了跨越式提升。受限于工艺制程，国内计算芯片行业的主流算力水平尚处于英伟达 A100 产品阶段，少数厂商通过先进封装（如双芯粒封装）等方式能实现接近英伟达 H100 产品的算力，为国内厂商中的最先进水平。

基于自主研发的 GPU IP 核心 XCORE 架构，发行人推出的训推一体 GPU 算力处于国内第一梯队。以行业内通常用来衡量 GPU 芯片在人工智能训练场景下的算力性能指标 FP16/BF16 为例，发行人曦云 C500/C550 芯片处于英伟达 A100 的算力区间，领先国内大部分厂商；而曦云 C588 芯片在曦云 C550 芯片的基础上实现算力跃升，大幅缩小与英伟达 H100 的差距，在国内处于先进水平。

在计算精度范围方面，计算功能的完备性与精度完整性是支撑多元场景的核心基石。基于全面覆盖计算和图形领域、灵活可扩展的处理器指令集，发行人的产品能同时支持人工智能训练和推理、通用计算、科学计算等场景、实现训推融合，在满足多元化的算力需求的同时，其强大的可扩展性使其也能够适应不断变化的计算场景。发行人拥有丰富的标量、矢量和张量计算单元，支持 FP64、FP32、TF32、FP16、BF16、FP8、INT8、FP6、FP4 等混合精度计算。随着智算中心、通用计算中心融合发展，公司的产品能够同时支持人工智能计算和通用计算，可以有效解决国内在运营的智算中心需要两套分离系统来支持科学计算和人工智能计算的痛点，大幅提升算力集群的多场景融合应用能力。

随着模型参数量的不断提升，更低数据位宽的训练和推理算力可以有效提升计算效率，引导着 GPU 芯片设计向更高效支撑低位宽的方向演进。例如，英伟达 Hopper 架构通过 Transformer 引擎使用 FP8 格式，为万亿参数模型训练提供性能提升方向；以 DeepSeek 为代表的头部大模型厂商亦采用 FP8 等低精度格式训练模型。作为国内少数掌握 FP8 技术、首批将 FP8 计算单元引入 GPU 产品的厂商，发行人将在保证计算精度的同时大幅提高大模型的计算性能。

3、大容量、高带宽、安全可靠的存储能力

存储能力是影响计算单元与存储单元间数据传输效率、单卡算力利用率的重要因素。缓存方面，发行人是全球少数几家掌握了复杂多级缓存结构的 GPU 企业，凭借独特的缓存组网技术，构建起多维度的缓存网络架构，能够针对不同场

景灵活自适应多级缓存控制策略，实现对高速缓存的精细管控、有效提升缓存命中率。同时，基于负载均衡算法、多元的仲裁策略、可配置的优先级控制、突发访问优化、乱序处理优化等定制化策略，发行人 GPU 产品的多级缓存网络为不同 GPU 业务流开辟高效的数据传输通路，并促使海量数据在缓存系统内部总线以及各通道之间实现均衡且高效的流通，达成数据访问效率的最大化，大幅提高 GPU 整体算力利用率。

显存方面，数据规模、模型参数量的指数级增长以及数据处理实时性的要求为 GPU 芯片的显存设计带来较大挑战。发行人训推一体 GPU 芯片均配备了大容量、高带宽的 HBM 内存，显存容量、显存带宽等性能参数在国内同行业企业中均位居前列，能够减少数据读取次数和通信延迟。同时，发行人是全球较早研发超高带宽显存(UBM)控制器接口的企业，将大幅提高 GPU 产品的带宽和容量，高效满足 GPU 高算力下对于高带宽的需求。此外，发行人采取安全校验内存技术，内置 ECC/RAS 多重安全防护模块，能够提高数据完整性和系统稳定性，满足金融、政务、医疗等行业安全规范。

4、国内领先、稀缺的通信能力

随着大模型参数规模指数级增长，单卡有效算力难以应对超大规模的算力需求，高带宽、低延迟的卡间互连技术成为关键竞争要素，对 GPU 集群互连拓扑结构也提出了更高要求，驱动了纵向扩展（Scale-up）与横向扩展（Scale-out）网络架构的技术演进。

发行人具备国内稀缺的高带宽、超多卡互连能力，自主研发的 MetaXLink 高速互连技术支持 GPU 全互连，突破了传统 PCIe 总线在带宽和延迟方面的限制，缩短了 GPU 之间数据传输时的通信时间，为大规模 AI 模型训练提供了充足的通信带宽保证。同时，MetaXLink 根据 GPU 之间数据交互需求优化传输协议并且支持 MetaXLink 端口之间数据直接转发，可以灵活适配大模型发展及其不同切分方式下对于互连延时和带宽的特定需求，满足训练和推理等不同应用场景。基于前述参数对比，发行人 GPU 产品的互连带宽达到了与英伟达 4nm 制程工艺下旗舰产品（H200）相当的性能，代表了国内先进的互连技术水平。

在互连拓扑结构方面，通过创新的架构设计，发行人单颗 GPU 芯片拥有 7

个高速 MetaXLink 互连接口，支持 Full-Mesh、Hybrid Cube Mesh 等多种复杂互连拓扑结构。通过 MetaXLink 各端口不同连接形式，发行人 GPU 产品支持 2 卡、4 卡、8 卡全互连拓扑，为千亿参数规模 AI 大模型应用高频使用到的 AllReduce、AllToAll 等集合通信原语提供单机内的更优拓扑形态与行业内更低时延支持。同时，凭借协议层设计创新，MetaXLink 具备较强的互连拓扑重构能力，能够根据应用需求灵活调整互连拓扑，最大化接口带宽的利用效率。发行人 GPU 产品通过超节点架构可灵活适配 16 卡、32 卡、64 卡等系统规模，满足从中小型训练到超大规模集群的差异化需求，为构建高密度算力和大规模集群、处理更为复杂的人工智能任务提供了关键技术支撑。为进一步提高智算集群算力密度，发行人联合合作伙伴研发并落地了高密度液冷整机柜，突破性实现单机柜 128 卡超高密度部署。丰富的超节点架构和超多卡拓扑结构的设计能力彰显了发行人在 GPU 通信能力方面领先的技术实力，为未来更大规模的 GPU 集群部署奠定了技术基础。

5、易用、通用、开放的软件生态

发行人坚持“自主创新与开放兼容”双轨并行策略，基于自研指令集及 GPU 并行计算引擎打造了自主可控 MXMACA 软件栈，是国内少数能为云端 AI 开发者提供兼顾易用性、高性能、通用性的基础软件栈的企业之一，不断推动软件生态开源共享，助力中国人工智能产业摆脱对国外算力资产的依赖。

易用性方面，发行人自主研发的 MXMACA 软件栈为开发者提供了简单易用的编程语言和完整的 AI 开发工具链，利用自有技术完成对相关开源框架和工具的适配、编译、验证、优化、全面质量测试（包括性能覆盖、功能覆盖等），并自研了调试和性能分析工具，不断降低开发者使用门槛，开发者可以以极低的学习成本使用 MXMACA 软件栈。

高性能方面，发行人自主研发了 GPU 软硬件协同优化技术，从软件实际需求出发，精准定义硬件的功能与性能，通过理论建模、硅前验证、硅后分析等方法，实现软硬件之间的深度协同与优化。基于运行时调优技术、算子融合技术、通信库优化算法、多通道 SDMA 技术、GPU Direct RDMA 技术等，算子利用率和大规模集群通信效能得以大幅提升，确保开发者利用 MXMACA 软件栈可以释放发行人 GPU 的近乎全部性能潜力。此外，发行人自主研发的 AI 编译器能够生成极具高性能的代码，例如使得 HPCG 性能相比英伟达 A100 大幅提高。

通用性方面，发行人 MXMACA 软件栈实现了对 CUDA 生态的高度兼容，一个中等复杂度 CUDA 应用的迁移成本仅需 1 人天，为人工智能、通用计算和大数据处理等领域的现有应用提供了高效迁移至 MXMACA 软件栈的能力，开发者无需过多修改现有代码即可将应用快速部署到公司的 GPU 产品上运行，大幅降低客户应用迁移成本、提升迁移效率。MXMACA 软件栈在 AI 训练框架、推理框架、高性能算子库的支持度和模型的适配度方面具备深厚的积累，支持超过 6,000 个 CUDA 应用、超过 2,200 个高性能算子，与超过 1,000 个模型实现了原生适配。根据中国信通院信息，发行人是首批通过 AISHPerf 基准 DeepSeek 大模型适配评测的公司。基于较强的生态兼容性，发行人 MXMACA 软件栈能够开放拥抱全球开发者丰富的开源成果，具有丰富的人工智能落地场景。

发行人致力于打造中国版 GPU 编程接口标准，坚决推动 MXMACA 软件栈开源共享，通过产学研合作不断助力大模型应用在千行百业加速落地。自 2025 年 2 月 14 日发行人开源社区开放以来，截至 2025 年 7 月 31 日，公司 MXMACA 软件栈注册用户超过 15,000 人，快速增长的开发者数量将为发行人的软件生态贡献新的场景应用并形成正反馈效应。发行人软件生态具体建设情况，参见本问询回复报告之“2.关于核心技术”之“一/（四）/1/（2）目前的生态建设情况”。

6、高线性度、利用率和稳定性的集群性能

在实际人工智能计算场景中，GPU 芯片的效率输出由芯片性能、集群性能和软件生态的易用性共同决定。当前人工智能训练任务呈现超线性、高故障率、低利用率特征，发行人凭借灵活通用的架构、先进的计算能力、存储能力及通信能力、易用且开放的软件生态，构筑了领先的集群性能，能够高效、弹性、稳定、无中断地执行海量参数模型分布式训练任务。

集群线性度方面，发行人攻克了大规模智算集群高效互连难题，硬件层面研发并落地了高带宽、低延迟的超多卡互连拓扑架构，软件层面实现通信路径搜索、分层均衡通信、计算与通信算子融合等，成功在算力、互连通信、软件生态等方面实现了有效均衡，使得发行人的 GPU 产品极具弹性和扩展性。根据第三方研究机构的测试结果，公司在千卡集群上执行 128B 的 MoE 大模型全量预训练的线性度达到了 95%左右，在行业内处于领先水平。

集群计算资源利用率（MFU）方面，通过在架构、存储体系、互连通信等硬件层面多层次、系统性的设计，通信库、DMA 单元、通信与计算相互 Overlap 技术、DualPipeV 技术及其他算子优化和编译器优化等软件层面的设计，基于发行人 GPU 的智算集群在实际大模型训练过程中的 MFU 在 60%-80%左右，能够充分释放 GPU 计算核心算力。同时，发行人依托多种优化策略，能够大幅提高模型训练效率、缩短模型训练周期，实现上千节点的高效协作。

集群稳定性方面，发行人软硬件的高度协同能够实现集群自动调优与动态监测、故障检查并自动恢复、异步持久化等，特别针对 MoE 场景，可以在网络故障场景达到不间断通信，大幅降低恢复开销、提升集群容错能力，确保了大规模集群稳定、无中断地分布式训练，实测在千卡集群上执行 128B 的 MoE 大模型训练任务无故障时间超 14 天（训练任务按需结束）。

集群规模方面，发行人是国内少数真正实现千卡集群大规模商业化应用的 GPU 供应商，并正在研发和推动万卡集群的落地。此外，在多元异构算力的技术趋势下，发行人的 GPU 产品支持构建异构融合计算集群，能够在算子层隔绝异构平台差异，实现集群的弹性扩展，从而有效帮助下游客户在国产替代切换过程中充分挖掘既有资源，提供了扩展现有智算集群、提升整体算力效能的优化方案。

综上所述，发行人的产品与技术整体已达国内先进水平，部分性能指标已接近或达到国际先进水平：发行人全系列产品均采用统一、灵活、通用的 GPU 计算架构和自主研发的核心 GPU IP；发行人训推一体 GPU 算力水平处于国内第一梯队、计算精度覆盖范围广，是国内少数掌握 FP8 技术、首批将 FP8 计算单元引入 GPU 产品的厂商；发行人是全球少数几家掌握了复杂多级缓存结构的 GPU 企业，同时 GPU 产品的显存配置在国内同行业企业中均位居前列；发行人具备国内稀缺的高带宽、超多卡互连能力，互连带宽代表了国内先进的互连技术水平；发行人是国内少数能为云端 AI 开发者提供兼顾易用性、高性能、通用性的基础软件栈的企业之一，MXMACA 软件栈能够高度兼容英伟达 CUDA 生态；发行人的 GPU 产品能够在智算集群实际运营中发挥领先的性能，集群线性度、计算资源利用率（MFU）、稳定性和异构混训能力在国内均具备较强的竞争力。

（三）发行人不同细分产品所用芯片、IP 及其他组件的收入构成、外购自研比例、核心技术的具体运用情况，与同行业的对比情况。下游晶圆厂商制程工艺变化对在售、在研产品性能，下游客户采购意愿的影响

1、发行人不同细分产品所用芯片、IP 及其他组件的收入构成、外购自研比例、核心技术的具体运用情况，与同行业的对比情况

（1）GPU 板卡

发行人为 Fabless 企业，主要从事 GPU 芯片的研发与设计，由晶圆制造厂和封装测试厂完成芯片的生产，并由板卡厂商完成 GPU 芯片和板卡的集成，最终形成完整的 GPU 板卡。板卡作为 GPU 芯片的物理承载与系统集成平台，是将 GPU 核心技术高效转化为可商业化产品不可或缺的基础载体，通常是云端智算产业中的最小交付单元。

在 GPU 板卡成本中，主要包括芯片生产加工成本、HBM 采购成本、封装测试成本等。发行人的核心技术融合贯穿了芯片设计、量产到后续升级优化的全流程，板卡成本中外购的原材料均需基于发行人的核心技术和研发设计才能形成最终的产成品，因此发行人 GPU 板卡均为自研形成，不存在外购原材料后直接使用的情況。发行人在生产制造、HBM 内存采购、封装测试等环节核心技术的具体运用情况及与同行业的比较情况具体如下：

原材料或半成品	发行人核心技术具体运用情况	与同行业的对比情况
生产制造	基于高性能 GPU 验证方法学及智能验证技术： （1）采用成熟先进、多层次的通用 GPU 验证技术，有效保障芯片产品的流片成功率和高质量交付 （2）从设计角度根据器件特点和芯片的性能需求，设计与工艺相匹配的版图，从工艺角度进行光刻方案的探索及优化	发行人仅用 3 年时间即实现两款芯片一次流片并成功量产，相较其他国内厂商流片成功率高、技术产业化速度快；发行人在国产供应链上的布局较早，拥有较强的工艺适配与优化能力
HBM 内存	基于高性能 GPU 多级缓存结构及内存管理设计技术、高性能 GPU 处理器架构设计技术等： （1）架构设计与系统协同：根据外采的 HBM 的容量、显存带宽设计 GPU 内存控制器，确保算力与存储带宽的平衡、HBM 的延迟与 GPU 的时钟频率和工作负载相匹配； （2）接口设计：设计显存控制器接口，定义 HBM 与 GPU 芯片的电气接口、通信协议； （3）物理集成设计：设计 HBM 高密度布线及与 GPU 芯片之间的互连结构；优化 HBM 与 GPU 的协同工作；确保在硅中介层上实现 GPU	在显存容量与显存带宽方面做到了更优的平衡。同时，发行人正不断加强技术开拓，在充分理解国产供应链的能力、供电网络需求、PCB 空间限制、散热能力限制的基础上，通过标准或者先进封装技术，在 2D、2.5D 及 3D 维度继续扩展算力、缓存带宽容量、显存带宽容量和接口带宽，不断提高产品竞争力

原材料或半成品	发行人核心技术具体运用情况	与同行业的对比情况
	与 HBM 的 2.5D 集成	
封装测试	基于芯粒架构与先进封装设计技术，进行包括大尺寸、高密度、高速互连、大功耗的多层基板设计、应用了高密度深沟槽电容技术的硅中介层设计、全链路 SI/PI 仿真技术等	发行人芯粒架构具有较强的通用性和可重构性，支持单芯粒、多芯粒同构或异构封装，标准封装和先进封装多种封装形式，实现不同尺寸和多种功耗规模的封装集成

在形成 GPU 板卡的过程中，发行人亦存在少量外采 IP 的情形，主要包括基础单元类 IP、接口类 IP 等，均为行业内已经较成熟、标准化的通用基础 IP，并不属于 GPU 设计所涉及的关键核心 IP。发行人在外采相关 IP 后仍需基于高性能 GPU 处理器架构设计技术等核心技术进行接口处理模块设计，确保 IP 模块的接口与芯片其他部分的接口兼容；同时将 IP 模块的逻辑设计集成到芯片的整体逻辑设计中，确保逻辑功能的正确性。

同行业公司中，寒武纪、海光信息、摩尔线程等均将 GPU 板卡收入视为核心技术收入，因此发行人将 GPU 板卡收入均视为核心技术收入存在较强合理性。

（2）GPU 服务器

服务器系以 GPU 板卡为核心，在自研 GPU 板卡的基础上结合硬件规格、散热结构、信号连接布局等结构设计，与 CPU、内存、硬盘、网卡等其他外购组件集成而来。相比之下，服务器为相对标准化的产品，市场上可供选择的同类厂商较多，其核心技术更多体现在 GPU 芯片/板卡设计过程中对互连拓扑、通信技术的开发。以英伟达 HGX 服务器为例，其服务器中集成多张 GPU 板卡并基于 NVLink 技术在服务器内实现多 GPU 卡之间的高速互连以最大化 GPU 并行效率。

发行人一方面借助标准 PCIe 服务器快速实现自研高性能 GPU 板卡的商业化落地，为 AI 大模型以及通用计算场景提供算力底座。另一方面，发行人深度参与 OCP 开放标准，将公司自主研发的 MetaXLink 高速互连拓扑转化为可商业化落地的 8 卡高性能服务器系统。结合模型规模快速增长，发行人创新设计了多款支持节点间通过光互连、电互连进行 Scale-Up 域扩展至 16 卡、32 卡、64 卡等的超节点系统，实现国际领先的互连性能，为智算集群提供更高效率的算力。同时，发行人不断积累服务器设计与实现的工程能力，服务器电源往小型化、大功率、高效率、高散热、高可靠和更好的 EMC 特性持续进步，确保 GPU 服务器的功

率密度稳步提升，从而推动超高密度超节点技术的逐渐落地。此外，面对 Serdes 速率提升对系统设计带来的挑战，发行人在服务器中引入更先进的板材、加工工艺、铜缆及光电合封技术等，实现了速率、功率和成本的更优平衡，进而确保 GPU 之间的高速互连。

鉴于 GPU 公司的核心竞争力更多体现在芯片设计与软件生态开发，且未来发行人将继续以 GPU 板卡销售为主要盈利模式，出于谨慎性考虑，发行人仅将 GPU 板卡收入、GPU 服务器中板卡收入及 IP 授权收入认定为核心技术收入。基于 GPU 板卡平均单价与服务器中板卡总数量测算得出，发行人训推一体 GPU 服务器中板卡收入占比在 60%左右。同行业中，英伟达同类型产品 A100 服务器中 GPU 占比约为 60%~70%，发行人与同行业不存在显著差异。

发行人已在招股说明书之“第五节 业务和技术”之“六、发行人的核心技术及研发情况”之“（一）主要产品核心技术情况”之“2、核心技术在主营业务及产品或服务中的应用和贡献情况”中补充修订经调整后的核心技术收入测算过程及占营业收入的比例情况，具体如下：

“发行人核心技术是通过持续积累并自主创新而形成的技术成果，在主营业务及产品或服务中得到广泛应用，助力公司收入规模持续增长。发行人应用核心技术产生的收入及其占营业收入比重情况如下：

项目	符号	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
训推一体 GPU 板卡（万元）	A	31,255.24	51,181.39	1,546.81	-
智算推理 GPU 板卡（万元）	B	400.75	349.08	1,376.32	42.64
GPU 板卡收入（万元）	C=A+B	31,655.99	51,530.47	2,923.13	42.64
训推一体板卡平均单价（万元/张，不含税）	D	3.89	4.69	5.69	-
训推一体服务器中板卡数量（张）	E	12	2,701	-	-
测算后的 GPU 服务器中板卡收入（万元）	F=D×E	46.68	12,667.69	-	-
IP 授权收入（万元）	G	25.00	100.00	2,216.67	-
核心技术产品收入（万元）	H=C+F+G	31,727.67	64,298.16	5,139.80	42.64
营业收入（万元）	I	32,041.53	74,307.16	5,302.12	42.64
核心技术产品收入占比	J=H/I	99.02%	86.53%	96.94%	100.00%

注 1：发行人将 GPU 板卡收入、GPU 服务器中板卡收入及 IP 授权收入认定为核心技术收入。

注 2：报告期内，发行人实现收入的服务器仅包含训推一体服务器。

注 3：因服务器销售合同未明确约定板卡及其他组件的销售金额，仅明确约定服务器中板卡数量，无法准确拆分 GPU 服务器中 GPU 板卡收入。上表 GPU 服务器中板卡收入仅代表测算结果，即测算后的 GPU 服务器中板卡收入=训推一体板卡当期平均单价×当期训推一体服务器中板卡总数量。”

经测算，报告期内，发行人核心技术收入占营业收入的比例分别为 100.00%、96.94%、86.53%、99.02%，核心技术收入占比较高，充分运用核心技术从事研发与经营活动。

2、下游晶圆厂商制程工艺变化对在售、在研产品性能，下游客户采购意愿的影响

下游晶圆厂商制程工艺变化主要影响发行人在研产品线中未来主力销售产品曦云 C600 系列、曦云 C700 系列。曦云 C600 系列、曦云 C700 系列均采用国内先进工艺制程，构建了从设计、制造到封装测试的国产供应链闭环，核心技术自主可控。其中曦云 C700 系列性能接近英伟达 H100，相较发行人当前在售产品的计算能力、存储能力、通信能力及能效比等均大幅提升。曦云 C600 系列基于发行人自主研发的 GPU IP 架构，集成大容量存储与多精度混合算力，支持 MetaXLink 超节点扩展，并内置 ECC/RAS 多重安全防护模块，为金融、政务等关键领域提供高可靠算力基座，能够满足下一代生成式人工智能的训练和推理需求。计算能力方面，曦云 C600 系列通过先进封装等方式实现计算性能升级，因此在计算能力上相较在售产品曦云 C500、曦云 C550 有所提升。同时，曦云 C600 系列增加了对 FP8 数据格式的支持，尤其是 FP8 Tensor 及 Tensor 转置指令，能够显著提高计算效率，更契合当前头部大模型厂商的训练需求。存储能力方面，曦云 C600 系列在曦云 C500/C550 的基础上进一步提升了显存配置，使用最新一代 HBM3e 并大幅提升了显存容量和显存带宽，在模型参数量不断提升的背景下能有效减少数据读取次数、减少数据传输延迟。

然而，受限于工艺制程，曦云 C600 系列的卡间互连带宽略有下降，同时功耗有所提升。基于 MXMACA 软件栈，发行人通过 SDMA 通信引擎、TP Overlap、MoE Comm Overlap、DualPipeV 等优化算法，可以优化通信资源分配、减少流水线空闲、优化传输策略、降低通信开销，从而提高通信效率、缓解理论带宽下降带来的影响，促使集群综合性能提升。功耗方面，功耗较高在一定程度上将增加客户的后续运营成本，但发行人曦云 C600 系列产品应用于云端数据中心，国

内电力成本相对较低，因此相较算力、显存、通信带宽等核心性能指标，功耗并非下游客户在做采购决策时的关键因素。

下游客户在选择 GPU 产品时的主要决策因素包括产品综合性能、国产替代相关政策导向、供应稳定性与安全性等。基于前述分析，发行人通过先进封装技术、互连通信优化等软硬件相结合的措施实现单卡性能和集群性能双轮提升，国产供应链产品曦云 C600 的综合性能相较目前在售产品进一步提高；同时，发行人结合技术发展趋势对计算精度、显存配置等进行了优化，能够充分满足下游客户大模型训练、推理、通用计算、科学计算等不同需求。对于政府智算中心、运营商、金融、能源、医疗健康、军工等政策敏感型、信息安全敏感型客户而言，采购全国产化 GPU 产品符合国产化算力建设相关政策导向、能够避免信息泄露风险，发行人在这类群体中具备较强的客户基础和先发优势。对于互联网大厂、AI 大模型厂商及其他成本敏感型民营企业而言，非国产供应链的产品存在“断供风险”，在符合性能需求的基础上选择具有稳定供应能力的厂商已成为该类群体的重要采购策略。

作为国内高性能 GPU 的领军企业之一，发行人始终将自主创新与产业应用作为发展战略，凭借突出的产品性能和稳定的供应能力，持续与下游客户展开深入合作，已建立了扎实的客户基础。随着曦云 C600 的回片与发布，多家头部互联网大厂或 AI 大模型厂商、国家人工智能公共算力平台、金融、能源、交通等行业客户高度关注曦云 C600 产品，已积极与发行人接洽交流、产品评测等，头部服务器 OEM 厂商亦积极适配曦云 C600 样片，将加速推动曦云 C600 产品的商业化落地。

综上所述，发行人的核心技术融合贯穿了芯片设计、量产到后续升级优化的全流程，GPU 板卡均为自研形成，不存在外购原材料后直接使用的情况；GPU 服务器为相对标准化的产品，其核心技术更多体现在 GPU 芯片/板卡设计过程中对互连拓扑、通信技术的开发，出于谨慎性考虑，发行人仅将 GPU 板卡收入、GPU 服务器中板卡收入及 IP 授权收入认定为核心技术收入；发行人与同行业不存在显著差异；发行人在售、在研产品具有领先的产品性能，基于国产供应链的产品更符合政策导向、能够满足下游客户对供应稳定性与安全性的需求，因此下游晶圆厂制程工艺变化不会对发行人在售、在研产品性能和客户采购意愿造成重

大不利影响。

（四）MXMACA 软件栈如何实现与英伟达 CUDA 生态兼容、目前的生态建设及用户使用情况，在软硬件设计方面是否存在路径跟随和性能提升瓶颈

1、MXMACA 软件栈如何实现与英伟达 CUDA 生态兼容、目前的生态建设及用户使用情况

（1）MXMACA 软件栈如何实现与英伟达 CUDA 生态兼容

作为软件生态建设的先行者与领导者，英伟达于 2006 年推出的通用并行计算架构和编程模型 CUDA 经过多年积淀，已形成了一套完整的开发工具链和强大的生态系统效应。兼容 CUDA 系一项复杂的系统性工程，对硬件架构、代码迁移效率、快速适配与迭代能力和 CUDA 应用兼容覆盖度等方面都提出了极高要求。

在硬件架构方面，软件生态的设计需要与 GPU 硬件结构特性高度耦合才能充分发挥硬件的性能水平。CUDA 生态与英伟达的 GPU 产品高度耦合，为英伟达的 GPU 硬件架构进行了充分优化。因此不同于 ASIC 架构，GPU 架构更为契合 CUDA 生态的结构设计。发行人全系列产品与英伟达均采用统一的通用 GPU 架构，包括与 SIMT 相关的通用指令集、L2 cache 高级操作特性及高性能 GPU 架构设计等，为兼容 CUDA 奠定了坚实的基础。同时，发行人遵循软硬件协同优化的设计理念，硬件架构的设计具有较强的通用性，而 MXMACA 软件栈的设计基于发行人 GPU 产品的硬件架构，避免了因硬件架构差异带来的性能损失。

在代码迁移效率方面，由于底层硬件架构和软件生态组件存在差异，需要大量编译优化技术算法及组合来提高代码迁移效率，减少开发者对代码的调优工作，并确保从 CUDA C++/Triton 等高级语言编译的应用程序性能足以满足客户需求。发行人采用源代码原生兼容 CUDA 生态的方式，支持 CUDA 源代码项目的工程构建自动切换到发行人 MXMACA 自主研发的完整工具链，能够实现用户程序近似零成本迁移，在编译优化方面具有较强的竞争力。同时，MXMACA 软件栈提供了一种简单易用的编程语言（MXMACA C++/MXMACA Fortran 异构编程语言），并兼容主流的 C/C++/Python 异构计算语言，开发者仅需维护一套代码即可，大幅降低开发者的应用门槛，具有较强的易用性。

在适配与迭代能力方面,人工智能技术及应用的高速革新和发展推动 CUDA 生态的版本、框架随之不断迭代。如仅兼容和适配旧版本 CUDA,难以满足最新的人工智能计算任务。MXMACA 软件栈兼容 CUDA SDK 版本覆盖度广,建立了海量应用测试集、随机形式化验证、差分验证、CUDA 随机特性组合的功能验证测试集合,以及全软件栈的自动化回归测试、海量应用测试集,通过可运行、可扩展、可交付的软件实体定义了 CUDA 兼容能力,并确保未来新产品代码开发不会破坏 CUDA 兼容性。

在 CUDA 应用兼容覆盖度方面,如今人工智能计算的应用大多构建在 CUDA 的基础之上,引导着未来人工智能与 GPU 架构的发展方向。CUDA 生态涵盖大量训练与推理框架、应用,以及丰富的高性能算子库以支持应用的覆盖性测试和全面性能调优。MXMACA 软件栈支持超过 6,000 个 CUDA 应用以及多种开源技术,包括 AI 训练框架(PyTorch/PaddlePaddle 等)、AI 推理框架(vLLM/SGLang 等)、高性能算子库(BLAS/DNN 等)和超过 1,000 个开源模型。此外,发行人自主研发的软硬件协同优化技术通过创新的架构设计、算法优化、智能任务调度和内存优化、基于大模型的智能算子优化技术等,能够有效提升 GPU 资源利用效率。

基于 GPU 异构计算编程语言与开发环境、人工智能编译器与编译技术、GPU 高性能通信库设计与优化技术、GPU 高性能数学库设计与优化技术、GPU 软硬件协同优化技术、GPU 核心驱动程序等核心技术,发行人 MXMACA 软件栈具备强大的代码迁移效率、较低的迁移损失、迁移能够获得较高的 GPU 效率、快速适配与迭代能力以及广泛的 CUDA 应用兼容覆盖度,在 API 层面实现了对 GPU 行业国际主流 CUDA 生态的高度兼容,为人工智能、通用计算和大数据处理等领域的现有应用提供了快速迁移至 MXMACA 运算平台的能力。

(2) 目前的生态建设情况

1) 与底层硬件紧密耦合的软件生态,丰富的算子库和工具链能够有效适应复杂多变的大模型新需求

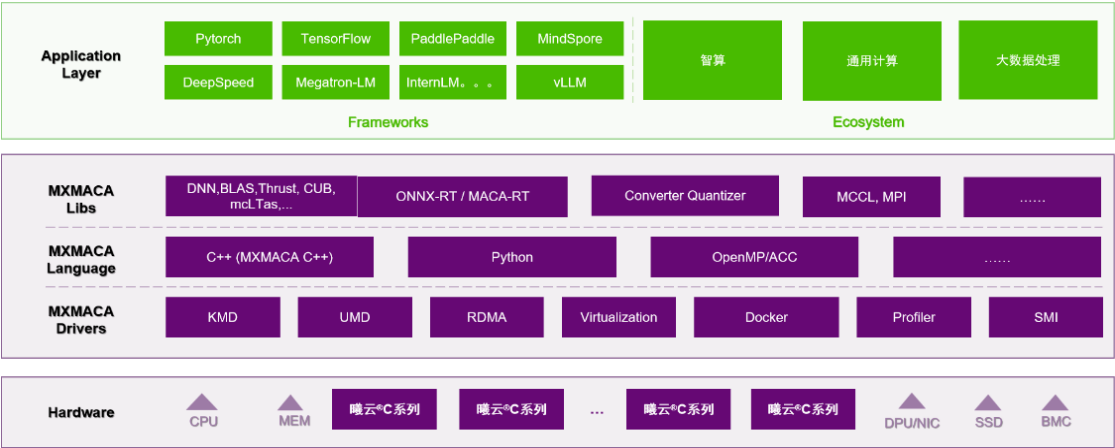
发行人基于“自主创新与开放兼容”双轨并行策略,为全系列 GPU 芯片构建了统一的运算平台及基础系统软件——MXMACA 软件栈,其基于公司自研指

令集以及 GPU 并行计算引擎，集成了主流算法框架、运算库、通信库、操作系统、编程语言、调试和运维管理工具等，为 GPU 产品提供了一套统一、完整且高效的全栈式软件工具链，涵盖应用开发、功能调试和性能调优等核心环节。同时，MXMACA 架构的编程接口在 API 层面实现了对 GPU 行业国际主流 CUDA 生态的高度兼容，为人工智能、通用计算和大数据处理等领域的现有应用提供了快速迁移至 MXMACA 运算平台的能力。发行人坚持推动 MXMACA 软件栈开源共享，不断打造中国版 GPU 开放编程接口标准，推动解决跨 GPU 平台的应用兼容难题。

MXMACA 软件栈按照抽象层次由高到低，可分为四个层级：

- 1) 算法应用层（Application Layer），包括各类 AI 框架、AI 网络模型、AI 应用等；
- 2) 数学库及通信库层（MXMACA Libs），包括用于支持应用层的自研软件库和适配、迁移到 MXMACA 平台的第三方软件库等；
- 3) 编程语言层（MXMACA Language），包括底层编程语言和高级编程语言支持，以及相关编译器和语言转换工具；
- 4) 软件驱动层（MXMACA Driver），包括用于支持内核模式驱动、内存访问、虚拟化和容器、性能分析、代码调试等功能的各种驱动程序和工具。

MXMACA 软件栈框架结构如下图所示：



在算法应用层，发行人构建了统一的运算平台架构和编程模型，原生支持大量 CUDA 开源生态、超过 6,000 个应用，包括 PyTorch、Tensorflow、PaddlePaddle、

JAX 等 AI 训练框架，Megatron-LM、Colossal-AI、Deepspeed、Pai-Megatron-Patch 等大模型训练框架，vLLM、LMDeploy、SGLANG、Transformer、ByteMLPerf 等推理框架，以及大量其他开源库如 FlashAttention、FlashInfer、DGL、deepEP、FlashMLA 等，与 DeepSeek、Llama、Qwen、GLM 等超过 1,000 个模型实现原生适配，改善了高性能异构计算场景中迁移困难、迁移成本高的痛点。

在数学库层，MXMACA 软件栈在 AI 大模型、高性能计算、向量数据库检索等领域，提供高效的标量、向量和矩阵计算接口、卷积计算接口、访存高效的 attention 类算子库、pytorch 优化算子库以及基于硬件特性与业务场景相关的算子库，能够充分释放各产品线硬件潜力；针对 LLM 的多个开源量化算子库（AWQ、GPTQ、GGUF、BitsAndBytes 等）进行了深度性能优化，在提高 LLM 推理的效率的同时持续保持高精度。

在通信库层，随着模型结构逐渐复杂、模型参数量大幅增加，计算集群向更复杂、更大规模发展，从而对集合通信、互连拓扑结构、通信时延等提出了更高要求。发行人基于 GPU 高性能通信库设计与优化技术实现了分层均衡通信、互连拓扑自动识别、通信路径高效搜索、通信资源统一调度等，大幅提升了通信效率、不同硬件架构拓扑的兼容性，提供了低时延高带宽集合通信，并拓展了集群容错和可扩展能力，保障大模型和高性能计算应用在各种集群上的高效协作通信及稳定运行。发行人在千卡集群执行 128B 的 MoE 大模型全量预训练的性能线性度在 95%左右，并且无故障时间超 14 天（训练任务按需结束），在行业内处于领先水平。

在编程语言层，发行人自主研发的 AI 编译器能够将各类不同的深度学习框架的模型代码转换为 GPU 底层硬件指令，通过算子自动调优技术、算子融合、动态内存分配、指令调度、代码结构优化等一系列编译优化技术，使得各类 AI 框架下的模型都能够无缝高效地部署运行在公司 GPU 上。同时，发行人的编译优化技术支持前沿的体系结构设计，能够优化计算和内存访问的调度，充分发挥硬件各流水线峰值性能，大幅提升算子性能。

在软件驱动层，发行人的驱动程序采用创新的异构计算架构，结合精细化的内存管理和智能任务调度算法、虚拟化管理机制、实时监控与纠错机制，一方面实现 GPU 资源的高效分配与动态优化，在多任务/多用户场景下显著提升了 GPU

的资源利用率、任务处理效率；另一方面能够兼容多种硬件架构，并确保 GPU 产品在麒麟、龙蜥、OpenEuler、Ubuntu 等各类操作系统环境中长时间稳定运行。

为促进人工智能应用在多元化场景中的快速落地与规模化部署，MXMACA 软件栈提供了丰富的性能评估工具链、云端运维管理工具等，有助于开发者精准识别性能瓶颈并更高效、灵活地开发、验证与优化，实现了对云边端异构硬件资源的统一管理、动态调度和高效协同，大幅提高了计算资源有效利用率，最高可近 90%。

2) 赋能开发者创新，构建人工智能的“Android”开源时代

发行人致力于打造中国版 GPU 编程接口标准，坚决推动 MXMACA 软件栈开源共享，建立了自由、开放的沐曦开发者社区和论坛平台并逐步开放程序员注册，不断赋能开发者创新和人才培养，是国内少数提供核心软件生态和源代码开源、提供版本 SDK 公开下载渠道、提供英文开发文档资料的 GPU 公司。

2025 年，发行人作为创始成员单位加入由中国计算机协会开源发展委员会（CCF）的 AI Infra 开源工作组，依托 CCF 链接科教资源、产业资源和社会资源等，通过 GPU 课程、大模型应用开发课程等线下课程、在线课程、CCF 大赛、孵化器对接等多种形式，形成产、学、研、用联动的开源创新模式，实现 GPU 应用和大模型应用开发人才的大规模培养。例如，发行人软件生态团队建立了“沐曦大学”，为开发者提供系统化的学习路径，超过数千人已完成在线学习；在浙江大学、电子科技大学、上海科技大学等多所大学展开了 GPU 高性能计算线下课程。发行人出版了《MXMACA /C++程序设计入门教程》，公开了 200 余份开发文档和手册，并已成功引入中国高校教育体系与教学实践，覆盖超过半数的中国 C9 高校，有望打破国外软件生态垄断、支持中国人工智能产业转型升级。

未来，发行人将继续秉承自主开放、合作共赢的发展理念，与硬件设备、智算中心、操作系统、MaaS 服务、AI 框架、大模型、AI 应用方、AI4S 等上下游 AI 生态企业深度联盟与合作，共创安全、稳定、开放的产业生态圈，通过构建完善的生态系统，实现从产品研发到市场的全链条闭环，助力中国人工智能产业自主化突破与长远发展。

(3) 目前的用户使用情况

自 2025 年 2 月 14 日沐曦开发者社区和论坛平台开放以来，截至 2025 年 7 月 31 日，公司 MXMACA 软件栈注册用户超过 15,000 人，网络 API 调用次数超过 532 万次，文件下载量超过 5 万次。持续增长的开发者数量和 MXMACA 软件使用频率增强了发行人产品的客户粘性，保障了公司产品迭代升级与未来市场需求、技术趋势高度吻合。

发行人是国内少数能为云端 AI 开发者提供兼顾易用性、高性能、通用性的软件生态的企业之一，MXMACA 软件栈在 CUDA 兼容性、代码迁移能力、软件性能与丰富度等多个层面获得了下游用户的广泛认可，以稳定、可靠的实测结果扭转了长期以来国产计算芯片难部署的刻板观念。下游用户对 MXMACA 软件栈出具了验证测试报告，部分内容举例如下：

评测机构	具体内容
上海人工智能实验室	“在软件生态方面，曦云 C550 软件栈和开发性支持程度较好，对算子功能整体支持度较完善，基础模型支持完整，算子性能上高于英伟达 A100 芯片，整体性能达到 A100 的 135.2%。”
上海超算中心	“通过严格测试，证明了沐曦 MXMACA 软件栈与 CUDA 及大模型和 HPC 相关生态的开源项目有良好的兼容性。曦云 C500 在 CUDA 生态兼容和性能上的表现都相当亮眼。借助 make_maca 转换工具，无任何修改即可编译并运行 cuda sample 代码。”
上海算法创新研究院	“沐曦的产品能支持绝大多数当前主流结构的模型的训练和推理；工程迁移工作量小，能比较顺利适配原 CUDA 环境上的代码；训练过程中 loss 收敛曲线状况基本能与英伟达同类卡上表现保持近似；推理结果的前后一致性及文本内容与英伟达卡上推理结果基本一致。”
中国科学技术大学	“经过测试，曦云 C500 在精度上满足 CFD 软件的一般要求，在性能上与 A100 处于同一 level，而扩展性则整体优于 A100。此外，使用 maca 软件附带的 cu-bridge 工具，无需任何代码修改就可以将复杂的 HiResX 程序编译成能在曦云 C500 上运行的可执行文件。这也充分体现了曦云 C500 软硬件平台的良好兼容性。”
中国计算技术西部研究院	“标准测试及部分 CUDA 应用随机兼容测试表明，曦云 C500 的 MXMACA CUDA 兼容的功能较为完整，结果与英伟达公司的 A100 相比，性能表现基本相当，部分指标超出。CUDA-examples 等测试编译、执行过程与 CUDA 系统兼容。”
联通（上海）产业互联网有限公司	“通过严格测试，证明了沐曦 MXMACA 软件栈与 CUDA 及大模型相关生态的开源项目有良好的兼容性。曦云 C500 在 CUDA 生态兼容和性能上的表现都相当亮眼。借助 make_maca 编译工具，无任何修改即可编译并运行 CUDA_Sample 代码。”

2、在软硬件设计方面是否存在路径跟随和性能提升瓶颈

英伟达经过多年积淀和持续迭代，已形成深厚的技术底蕴和丰富的行业经验。在硬件性能持续提升的同时，英伟达构建了以 CUDA 为核心的庞大软件生态，

绝大部分主流 AI 框架、人工智能应用等都构建在 CUDA 的基础之上，软件生态的繁荣为英伟达 GPU 硬件平台构筑了强大的竞争护城河。相比之下，国内 GPU 产业起步较晚，发行人自 2020 年成立至今虽已取得一定技术突破，但在硬件性能和软件生态方面与英伟达仍存在一定差距，尚处于市场跟随阶段。发行人结合自身技术基因与竞争优势，扎根国内特定场景进行软硬件性能优化和突破，形成差异化发展路线。

硬件设计方面，摩尔定律的迭代放缓使得海外龙头企业开发新产品面临更大的挑战，发行人已全面系统掌握了 GPU 架构、GPU IP、SoC 设计、封装设计等全流程核心技术，具备“从零到一”的自主创新能力，有望以更快的速度向海外现有产品看齐。当前产品性能提升瓶颈主要体现在工艺制程，发行人深度参与国产先进工艺设计套件（PDK）迭代优化，充分理解并有效解决了先进制程下的设计难点，包括高密度设计，高功耗电源网络设计、高频时钟网络设计以及漏电控制等，能够通过芯粒架构与先进封装设计、设计工艺联合优化、高性能通信库设计等技术，与英伟达形成差异化技术路径并打破工艺制程带来的性能瓶颈。发行人基于国产供应链设计和制造的曦云 C600 芯片一次上电成功，已成功打通了全链路关键路径。同时，发行人深度融入国内 GPU 供应链，将自身技术需求转化为供应链的升级动力，与国内产业链上下游同频共振、共同成长。

软件生态建设方面，发行人自主研发的 MXMACA 软件栈集成了主流算法框架、运算库、通信库、操作系统、编程语言、调试和运维管理工具等，已形成一套统一、完整且高效的全栈式软件工具链。发行人硬件具有独立、专有的通信硬件单元，能够有效实现计算与通信的高度并行。基于公司自身硬件特性及对软件设计的理解，发行人通过在计算通信并行、专家模型优化、显存优化、自动调优与集群训练工具等方面自主创新，有效提升英伟达原生 Megatron-LM 框架在发行人硬件环境中的性能，整体训练速度提升 20%左右、显存可节省 10%~30%、集群计算资源利用率（MFU）可提升 10%~90%。相关软件设计与优化均基于发行人自身硬件特性，因此不存在路径跟随。相较而言，英伟达通过日积月累留存了众多开发者基于其硬件环境开发最新应用，大量 AI 训练开源项目和框架与英伟达 GPU 后端实现了原生适配，在一定程度上制约着其他 GPU 厂商的发展。发行人采取开源策略重构生态规则，致力于构建人工智能的“Android”开源时代，

建立了自由、开放的沐曦开发者社区和论坛平台，不断推动 MXMACA 软件栈开源共享，并推出了在线课程、线下课程、软件开发教材等，降低了 MXMACA 软件栈的使用门槛，将吸引并留存更多开发者基于发行人的硬件环境开发下一代应用、定义未来需求市场，逐步突破生态壁垒。

综上所述，发行人通过通用硬件架构设计、无损代码迁移、快速适配与迭代、丰富的 CUDA 应用兼容覆盖等，实现了对 CUDA 生态的高度兼容；发行人目前已建设了多层次、体系化的软件生态，具有易用性、高性能和通用性，并不断推动软件生态开源共享，获得了下游用户的广泛认可；发行人结合自身技术基因与竞争优势，扎根国内特定场景进行软硬件性能优化和突破，形成差异化发展路线。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构执行了如下核查程序：

1、访谈发行人技术负责人，了解核心技术来源、形成及演变过程；获取并查阅研发项目立项文件、进展文件、外购技术服务合同、委外研发合同、成果文件、付款凭证及发票等相关资料，对比外购技术服务、委外研发内容与关键环节及关键核心技术的关联性。

2、查阅行业研究报告、同行业公司产品参数文档、公司内部调研报告、市场公开报道等，了解 GPU 不同产品技术难易度、市场需求情况、行业技术水平情况，分析发行人技术先进性水平。

3、查阅收入成本表，获取发行人对外采购的原材料情况、成本构成等，并访谈发行人技术负责人，了解核心技术的具体运用情况，分析核心技术收入测算结果的合理性。

4、访谈发行人技术负责人和商务部负责人，了解并分析下游晶圆厂商制程工艺变化对发行人在售、在研产品性能以及下游客户采购意愿的影响。

5、访谈发行人技术负责人，了解公司在软硬件设计方面是否面临路径跟随和性能提升瓶颈。

6、获取发行人 MXMACA 软件栈后台注册用户、API 调用、文件下载等情

况；取得下游用户对 MXMACA 软件栈的验证测试报告。

（二）核查意见

经核查，保荐机构认为：

1、发行人核心技术系经过长期经验积累和技术迭代逐步形成，均为自主研发，不存在来源于其他企业或任何外部第三人的情形；发行人外购技术服务主要为智能平台软件开发服务、共建实验室费用、集成电路仿真验证费用等，委外研发主要为部分后端设计类、非核心的算子优化测试类等非核心的研发工作；发行人涉及外购技术服务、委外研发的环节主要仍为发行人自主研发，相关环节仅非核心、技术难度较低的部分涉及外购技术服务或委外研发，发行人外购技术服务、委外研发的内容不涉及关键环节或关键核心技术。

2、发行人当前主要销售产品为曦云 C 系列训推一体 GPU 产品，整体已达国内先进水平，部分性能指标已接近或达到国际先进水平，在 GPU 架构、计算能力、存储能力、通信能力、软件生态以及集群性能上均具备较强的技术先进性。

3、发行人的核心技术融合贯穿了芯片设计、量产到后续升级优化的全流程，GPU 板卡均为自研形成，不存在外购原材料后直接使用的情况；GPU 服务器为相对标准化的产品，其核心技术更多体现在 GPU 芯片/板卡设计过程中对互连拓扑、通信技术的开发，出于谨慎性考虑，发行人仅将 GPU 板卡收入、GPU 服务器中板卡收入及 IP 授权收入认定为核心技术收入；发行人与同行业不存在显著差异；发行人在售、在研产品具有领先的产品性能，基于国产供应链的产品更符合政策导向、能够满足下游客户对供应稳定性与安全性的需求，因此下游晶圆厂制程工艺变化不会对发行人在售、在研产品性能和客户采购意愿造成重大不利影响。

4、发行人通过通用硬件架构设计、无损代码迁移、快速适配与迭代、丰富的 CUDA 应用兼容覆盖等，实现了对 CUDA 生态的高度兼容；发行人目前已建设了多层次、体系化的软件生态，具有易用性、高性能和通用性，并不断推动软件生态开源共享，获得了下游用户的广泛认可；发行人结合自身技术基因与竞争优势，扎根国内特定场景进行软硬件性能优化和突破，形成差异化发展路线。

3.关于尚未盈利

根据申报材料：（1）报告期内，发行人归母净利润为-77,696.52 万元、-87,115.82 万元、-140,887.94 万元和-23,251.22 万元；截至 2025 年 3 月末，累计未弥补亏损为-104,752.49 万元；经营活动现金流量净额为-66,514.29 万元、-101,731.89 万元、-214,802.33 万元和-53,131.75 万元；（2）报告期内，发行人前五大客户和前五大供应商较为集中，向供应商预付大额款项；（3）2024 年末、2025 年 3 月末、2025 年 4 月末，发行人在手订单金额为 6.99 亿元、7.73 亿元、9.17 亿元，发行人预计 2026 年扭亏为盈。

请发行人披露：（1）结合行业周期、供需关系、市场竞争等行业共性因素、以及发行人所处发展阶段，量化说明发行人报告期内持续亏损的主要影响因素，以及亏损趋势未明显改善的原因及合理性；（2）结合经营活动现金流量、营运资金覆盖、研发投入、客户和供应商集中、国际贸易形势等，分析对发行人未来生产经营的影响；（3）在手订单的对应客户、收入实现可行性；结合不同产品的市场空间变化、发行人市场份额及变动、客户复购、新客户验证及拓展情况等，分析发行人业绩测算的合理性；（4）结合前述情况，说明发行人前瞻性分析相关参数及假设依据是否合理，按照《监管规则适用指引-发行类第 5 号》《监管规则适用指引—发行类第 10 号》要求详细披露原因分析、影响分析、趋势分析、风险因素、投资者保护措施及承诺等事项，并结合研发进度、商业化前景等因素，审慎估计预计实现盈利时间等前瞻性信息，补充披露关于预计盈利时点的敏感性分析，并完善相关重大事项提示。

请保荐机构、申报会计师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）结合行业周期、供需关系、市场竞争等行业共性因素、以及发行人所处发展阶段，量化说明发行人报告期内持续亏损的主要影响因素，以及亏损趋势未明显改善的原因及合理性

报告期内，影响发行人损益的各科目金额及占营业收入比例如下：

单位：亿元

项目	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
	金额	占比	金额	占比	金额	占比	金额	占比
营业收入	3.20	100.00%	7.43	100.00%	0.53	100.00%	0.0043	100.00%
减：营业成本	1.43	44.74%	3.46	46.57%	0.20	37.12%	0.0032	75.90%
营业毛利	1.77	55.26%	3.97	53.43%	0.33	62.88%	0.0010	24.10%
减：税金及附加	0.02	0.77%	0.02	0.26%	0.02	3.79%	0.0034	80.26%
销售费用	0.35	11.06%	1.16	15.60%	0.57	107.45%	0.42	9767.07%
管理费用	1.28	40.04%	1.76	23.63%	1.14	215.93%	0.88	20609.46%
研发费用	2.06	64.36%	8.40	112.99%	6.59	1243.45%	6.20	145379.73%
财务费用	0.00	0.00%	0.25	3.38%	-0.03	-5.17%	-0.06	-1321.72%
股份支付	0.32	9.85%	4.73	63.63%	0.58	108.68%	0.40	9330.73%
加：其他收益	0.03	1.09%	0.17	2.33%	0.13	24.42%	0.02	459.68%
投资收益	0.08	2.42%	0.16	2.10%	0.09	16.87%	0.06	1484.81%
公允价值变动收益	0.02	0.59%	0.00	0.03%	0.00	0.12%	0.0043	101.68%
信用减值损失	-0.17	-5.28%	-0.56	-7.49%	-0.05	-9.26%	-0.01	-159.92%
资产减值损失	-0.01	-0.33%	-1.30	-17.46%	-0.33	-62.79%	-	0.00%
资产处置收益	-	0.00%	-0.17	-2.34%	0.01	2.69%	-0.0003	-8.07%
加：营业外收入	0.00	0.02%	0.00	0.05%	0.00	0.30%	0.0015	35.18%
减：营业外支出	0.01	0.21%	0.02	0.30%	0.02	4.03%	0.02	351.78%
利润总额	-2.32	-72.51%	-14.05	-189.13%	-8.71	-1642.93%	-7.77	-182259.86%
减：所得税费用	0.00	0.05%	0.04	0.48%	0.00	0.11%	-0.0019	-44.76%
净利润	-2.33	-72.57%	-14.09	-189.60%	-8.71	-1643.04%	-7.77	-182215.10%
扣非净利润	-2.19	-68.27%	-10.44	-140.48%	-8.90	-1678.34%	-7.84	-183936.59%

注：表格中销售费用、管理费用、研发费用剔除了股份支付金额，下同。

报告期内，影响发行人损益的主要科目（2024 年占营业收入比例超过 5%）如下：

单位：亿元

项目	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
	金额	占比	金额	占比	金额	占比	金额	占比
营业收入	3.20	100.00%	7.43	100.00%	0.53	100.00%	0.00	100.00%
减：营业成本	1.43	44.74%	3.46	46.57%	0.20	37.12%	0.00	75.90%
营业毛利	1.77	55.26%	3.97	53.43%	0.33	62.88%	0.00	24.10%
销售费用	0.35	11.06%	1.16	15.60%	0.57	107.45%	0.42	9767.07%
管理费用	1.28	40.04%	1.76	23.63%	1.14	215.93%	0.88	20609.46%
研发费用	2.06	64.36%	8.40	112.99%	6.59	1243.45%	6.20	145379.73%
股份支付	0.32	9.85%	4.73	63.63%	0.58	108.68%	0.40	9330.73%
信用减值损失	-0.17	-5.28%	-0.56	-7.49%	-0.05	-9.26%	-0.01	-159.92%
资产减值损失	-0.01	-0.33%	-1.30	-17.46%	-0.33	-62.79%	-	0.00%
净利润	-2.33	-72.57%	-14.09	-189.60%	-8.71	-1643.04%	-7.77	-182215.10%
扣非净利润	-2.19	-68.27%	-10.44	-140.48%	-8.90	-1678.34%	-7.84	-183936.59%

由上表可见，发行人报告期内持续亏损的主要影响因素为：（1）报告期内收入与毛利快速增长，但总体收入规模较小；（2）报告期内研发费用金额较大且持续增长，销售毛利规模不足以覆盖高额的研发支出；（3）报告期内确认了大额的股份支付费用。2024 年，发行人亏损规模扩大的主要原因为：确认了大额股份支付费用、资产减值损失以及信用减值损失，并且研发费用投入增加较多。

发行人报告期内持续亏损的主要影响因素、亏损趋势未明显改善的原因及合理性分析如下：

1、行业共性因素

（1）行业周期

在中美科技博弈的背景下，国产人工智能芯片厂商迎来关键发展机遇，进口替代迫在眉睫。根据 IDC 数据，2024 年，中国本土人工智能芯片品牌渗透率约 30%，出货量达到 82 万张，相较去年同期 15%的国产品牌渗透率已呈明显提升趋势。IDC 数据显示，2024 年中国加速计算服务器市场规模达到 221 亿美元，同比 2023 年增长 134%；到 2029 年，中国加速计算服务器市场规模将超过千亿美元。中商产业研究院预测，2025 年中国人工智能芯片市场规模将增长至 1,530

亿元，2020 年至 2025 年复合增长率达 53%。根据弗若斯特沙利文预测，到 2029 年，中国人工智能芯片市场规模将从 2024 年的 1,425.37 亿元激增至 13,367.92 亿元，2025 年至 2029 年期间年均复合增长率为 53.7%。中国作为全球人工智能芯片市场的重要参与者，预计将与全球人工智能芯片市场同频共振，成长空间巨大。作为人工智能芯片中占比最高的 GPU 芯片，可以预见其在人工智能计算领域具有广阔的发展前景。

我国人工智能芯片行业市场景气度持续向好，但国产人工智能芯片厂商整体起步相对较晚，过去几年大多处于技术突破和产品落地初期阶段。在这一阶段，企业处于大规模量产初期，规模效益尚未充分显现，且由于核心技术积累、产品生态建设、市场认知度等方面均有待提升，仍需要投入大量资源用于技术研发、市场拓展和生态建设等，这使得成本居高不下。同时，国产人工智能芯片厂商在产品性能和市场份额上与国际领先企业存在一定差距，这导致在市场竞争中难以快速抢占大量份额，只能渐进式拓展市场。因此，报告期内发行人持续亏损与国产人工智能芯片企业当前所处的发展阶段及发展特征相符，具有合理性。

（2）供需关系

从需求侧来看，在新一轮人工智能发展浪潮下，国内外智算中心持续建设与扩容，垂直行业场景多点开花，人工智能应用大规模落地。其中教科研、金融、交通、能源、医疗健康、大文娱等与业务流紧密结合的人工智能应用场景有望率先实现商业化落地。2025 年以来，DeepSeek 通过一系列算法优化降低了训练和推理成本，为 AI 应用在终端的普及创造了良好的条件，将推动国内算力市场需求大幅提升。人工智能芯片作为关键的底层算力资产，其市场需求将随下游 AI 应用不断发展而加速释放。

从供给侧而言，人工智能和集成电路是中美科技博弈的重要领域，近年来美国政府从多方面对华展开高端技术封锁，高性能芯片禁令不断升级，国产替代已成为必然趋势。根据 IDC 数据，2024 年国内人工智能芯片市场中，海外厂商占据约 70% 的市场份额，相较去年同期 85% 的市场份额已呈明显下降趋势，国内市场国产替代的空间仍极为宽广。

然而，我国人工智能芯片市场长期由英伟达、AMD 等国际厂商主导，尤其

是英伟达通过 CUDA 生态构建了完整的软硬件技术标准体系，形成行业事实性垄断。国产人工智能芯片厂商整体起步较晚，在产品性能、市场份额等方面与国际厂商存在差距，国产人工智能芯片要进入市场，不仅要在性能上追赶，还面临技术标准适配、用户习惯迁移双重障碍，生态建设需系统性突破。在这个阶段，国产人工智能芯片厂商需要大量资金投入用于技术研发、市场拓展、生态建设等，成本支出持续增加，而收入却因市场开拓难度大、需求释放缓慢等原因难以快速增长。因此，报告期内发行人持续亏损与我国人工智能芯片行业当前行业供需关系相符，具有合理性。

(3) 市场竞争

全球市场来看，GPU 市场头部化现象显著，呈寡头垄断格局。英伟达（NVIDIA）和超威半导体（AMD）两家国外领先厂商基本分割了全球市场，在综合技术实力、销售规模、资金实力、人才团队等方面优势明显。根据 Jon Peddie Research 的数据，独立 GPU 市场方面呈现“一超一强”格局，其中英伟达一家独大，近年来持续维持超 80% 的市场份额，而 AMD 公司则占据剩余近 20% 的市场份额。这种全球市场被高度垄断的局面，使得国产人工智能芯片厂商在走向国际市场、参与全球竞争时，面临着巨大的准入门槛，很难在短时间内突破并获取可观的国际市场份额。

国内市场来看，在人工智能计算领域，发行人主要竞争对手包括国际巨头英伟达、AMD，国内上市公司寒武纪、海光信息，国内未上市公司华为海思、昆仑芯、天数智芯、壁仞科技、燧原科技、摩尔线程等。近年来，受到美国高性能计算芯片出口管制与国内政策推动等因素影响，海外厂商在中国市场的份额呈现明显下降趋势，但凭借其品牌影响力、技术先发优势以及已经建立起的部分用户生态，依然在中国市场占据着重要地位，尤其在一些对人工智能芯片性能要求极高、生态依赖度强的高端应用场景中，仍有较大的话语权。与此同时，国产人工智能芯片公司正迎来黄金发展期，各家企业均在国内积极布局发力，共同争夺市场份额。

在目前国内市场竞争格局下，国产人工智能芯片厂商一方面要与国际巨头的既有优势相抗衡，另一方面还要在国内同行竞争中脱颖而出，需要投入大量资金用于提升技术水平、打造高性能产品、拓展销售渠道以及构建自身的生态体系等。

然而，在市场份额还未充分打开、产品认可度仍有待提升的阶段，这些持续的高额投入难以快速转化为营收，成本支出远超收入，导致过去几年行业普遍处于亏损状态。因此，报告期内发行人持续亏损与我国人工智能芯片行业市场竞争情况相符，具有合理性。

2、发行人所处发展阶段

发行人自 2020 年 9 月成立以来高度重视研发创新，持续高强度进行研发投入，报告期内研发费用分别为 64,752.09 万元、69,862.04 万元、90,089.04 万元和 21,790.41 万元。发行人制定了“量产一代、在研一代、规划一代”的产品研发策略，积极推进主要在研产品的研发工作。发行人智算推理 GPU 芯片曦思 N100 系列、训推一体 GPU 芯片曦云 C500 系列分别于 2023 年 4 月和 2024 年 2 月正式量产。发行人正在研发曦云 C 系列训推一体芯片的第二代产品曦云 C600 系列和第三代产品曦云 C700 系列，其中曦云 C600 系列是基于沐曦自主知识产权核心 GPU IP 架构，构建从设计、制造到封装测试的国产供应链闭环，核心技术自主可控。曦云 C600 集成大容量存储与多精度混合算力，支持 MetaXLink 超节点扩展技术，并内置 ECC/RAS 多重安全防护模块，为金融、政务等关键领域提供高可靠算力基座，满足下一代生成式 AI 的训练和推理需求，性能强劲。曦云 C600 系列已于 2024 年 10 月交付流片，并于 2025 年 7 月回片并成功点亮。曦云 C700 系列基于国产先进工艺开发，并采用发行人新一代 GPU 架构及指令集，相比曦云 C600 系列性能更为强大，同时进一步扩展对 FP4 等低精度的计算支持，定位是未来发行人国产供应链的旗舰产品，对标目前国际主流产品英伟达 H100。发行人在 GPU 架构、显存设计、超节点互连、光互连、芯粒与先进封装设计、设计工艺联合优化、软件设计、软硬件协同优化等方面具有深厚的技术储备，能够通过前述产品研发完成现有核心产品线的持续迭代升级，巩固行业领先地位，同时提升发行人整体研发实力，增强技术储备。

报告期内，发行人销售规模处于快速爬坡阶段，营业收入分别为 42.64 万元、5,302.12 万元、74,307.16 万元和 32,041.53 万元，虽然增长迅速，但整体规模仍然较小，销售毛利规模不足以覆盖高额的研发支出。发行人重视人才激励与长期发展，实施了多轮次的员工股权激励，相应确认了大额的股份支付费用，报告期内股份支付费用分别为 3,978.62 万元、5,762.38 万元、47,283.29 万元和 3,156.31

万元。

此外，2024 年发行人资产减值损失金额、信用减值损失金额分别同比增加 9,642.60 万元、5,074.44 万元，导致 2024 年亏损幅度进一步加大，发行人资产减值损失和信用减值损失增加的主要原因为：（1）2024 年 11 月，美国禁令导致发行人无法继续向晶圆代工厂商下达训推一体曦云 C500 系列产品的生产订单，进而导致发行人光罩等固定资产无法继续使用，因此发行人对相关光罩资产全额计提减值准备；同时发行人部分预付款项相关采购合同也因美国禁令原因无法正常履行，发行人对其中无法收回部分全额计提坏账准备。（2）由于曦思 N100 芯片于 2023 年停产，且考虑到市场环境发生变化，发行人调整了曦思 N100 产品的定价，对 2024 年末结存的曦思 N100 系列的存货计提了存货跌价准备。

2022 年至 2024 年，发行人净利润扣除股份支付、资产减值损失、信用减值损失后情况如下：

单位：亿元

项目	2024 年度	2023 年度	2022 年度
净利润（A）	-14.09	-8.71	-7.77
股份支付（B）	4.73	0.58	0.40
资产减值损失（C）	1.30	0.33	-
信用减值损失（D）	0.56	0.05	0.01
计算结果（E=A+B+C+D）	-7.51	-7.75	-7.36

由上表可见，扣除股份支付、资产减值损失、信用减值损失后，2024 年发行人亏损 7.51 亿元，较 2023 年度亏损额 7.75 亿元呈现收窄趋势。

综上所述，报告期内，发行人始终聚焦核心技术突破，持续进行高强度研发投入以追赶国际领先水平，并稳步推进市场拓展、生态建设以及产品迭代升级工作。受研发投入持续高企、市场生态尚在培育、股份支付及资产减值等多重因素综合影响，报告期内发行人持续亏损且 2024 年亏损幅度有所扩大，这一状况与发行人所处的技术攻坚和市场开拓阶段特征相符，具有合理性。

3、发行人期后经营情况

根据发行人未审财务数据，2025 年 1-6 月，发行人销售收入约 9 亿元，已超过 2024 年全年水平；综合毛利率约 56%，持续保持稳定；亏损趋势有所改善。

未来，随着公司销售规模进一步扩大，盈利能力将得到进一步增强，亏损金额有望持续收窄。

综上所述，发行人报告期内持续亏损及亏损趋势未明显改善，是行业共性因素与自身发展阶段共同作用的结果，具有合理性。行业共性因素方面，国产人工智能芯片企业处于发展初期，国产替代虽带来机遇，但国产厂商整体起步相对较晚，且面临国际巨头垄断与国内同行竞争，供需端存在需求释放慢、生态壁垒高等问题，国产人工智能芯片厂商多呈高研发投入、亏损状态。自身所处发展阶段方面，发行人自 2020 年成立后持续高强度研发，报告期研发费用逐年增长且占营收比例较高，而收入规模虽快速增长但整体仍小，销售毛利难以覆盖成本，叠加大额股份支付费用导致持续亏损。2024 年，发行人受股份支付金额大幅增加、资产减值损失和信用减值损失增加等因素影响，亏损幅度有所扩大。发行人凭借技术、产品和市场优势，持续拓展市场和生态建设，收入规模持续增加，根据 2025 年 1-6 月未审财务数据，发行人销售收入与销售毛利均实现大幅增长，亏损趋势已有所改善，随着销售规模的进一步扩大，未来盈利能力有望逐步增强，亏损金额或将持续收窄。

（二）结合经营活动现金流量、营运资金覆盖、研发投入、客户和供应商集中、国际贸易形势等，分析对发行人未来生产经营的影响

1、经营活动现金流量

报告期内公司经营活动现金流量情况如下：

单位：亿元

项目	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
经营活动现金流入	2.39	8.39	1.35	0.28
经营活动现金流出	7.70	29.87	11.52	6.93
经营活动产生的现金流量净额	-5.31	-21.48	-10.17	-6.65

报告期内，发行人经营活动现金流量净额持续为负，主要原因是发行人销售规模尚在逐步提升过程中，原材料采购、存货备货和研发投入的资金需求较大，此外发行人为应对供应链稳定性风险，对部分原材料进行了战略性储备。

报告期内，随着发行人商业化持续落地，投资者对发行人的认可程度不断提

升，发行人通过股权融资能力持续增强，报告期内，发行人筹资活动现金流量净额分别为 70,402.52 万元、119,697.76 万元、298,890.27 万元和 675,325.54 万元。截至 2025 年 3 月末，发行人账面货币资金、交易性金融资产合计 72.91 亿元，是 2024 年发行人经营活动产生的现金流量净额绝对值的 3.39 倍，发行人货币资金储备充裕，足以支撑发行人持续经营。未来，随着销售规模进一步增加，盈利能力将得到增强，发行人经营活动现金流量净额将进一步得到改善。

2、营运资金覆盖

报告期各期末，发行人营运资金及流动比率情况如下：

单位：亿元

项目	2025 年 3 月末	2024 年末	2023 年末	2022 年末
流动资产（A）	100.31	34.25	12.74	7.42
流动负债（B）	8.20	9.21	3.41	1.12
营运资金（C=A-B）	92.11	25.04	9.33	6.30
流动比率（D=A÷B）	12.23	3.72	3.74	6.60

报告期各期末，发行人营运资金分别为 6.30 亿元、9.33 亿元、25.04 亿元和 92.11 亿元，流动比率分别为 6.60、3.74、3.72 和 12.23，发行人最近一期末营运资金和流动比率大幅增加，发行人短期偿债能力较好。

截至 2025 年 3 月末，发行人货币资金余额为 569,461.33 万元，其中受限货币资金为 198.06 万元，主要是银行承兑汇票、保函的保证金；交易性金融资产余额为 159,688.21 万元，系公司已购买未到期的结构性存款。因此，报告期末发行人可支配资金为 728,951.49 万元，测算如下：

单位：万元

项目	计算公式	金额
货币资金	①	569,461.33
其中：受限货币资金	②	198.06
交易性金融资产	③	159,688.21
可支配资金	④=①-②+③	728,951.49

在其他经营要素保持不变的情况下，以 2024 年为基期，采用销售百分比法测算发行人 2025 年末、2026 年末的经营性流动资产和经营性流动负债，并据此计算未来两年的营运资金需求。经测算，预计发行人 2025 年、2026 年营运资金

需求分别为 194,600.35 万元、362,325.23 万元，合计 556,925.58 万元。因此，发行人可支配资金可以覆盖未来两年营运资金需求，发行人营运资金覆盖情况良好。

3、研发投入

人工智能芯片行业是典型的技术密集型产业，发行人为突破国际生态垄断并实现自主可控目标，始终将研发创新作为企业发展的核心驱动力，研发投入被列为长期战略规划的重要组成部分。报告期内发行人研发费用分别为 64,752.09 万元、69,862.04 万元、90,089.04 万元和 21,790.41 万元，呈现显著增长态势。在不懈的技术攻坚下，发行人成为了国内少数几家系统掌握了先进制程 GPU 芯片及其基础系统软件研发、设计和量产技术的企业之一，深度积累了 GPU IP（包括指令集、微架构等）、GPU SoC、高速互连、GPU 软件等核心技术，打造了自主开放、高度兼容国际主流 GPU 生态（CUDA）的软件生态体系，能够为客户构建软硬件一体的全面生态解决方案，在底层技术上摆脱对国外算力资产的依赖，为推进新质生产力发展提供动力引擎。

发行人制定了“量产一代、在研一代、规划一代”的产品研发策略，发行人第二代高性能通用 GPU 芯片（曦云 C600 系列）研发项目于 2024 年 2 月立项，该系列芯片系基于国产供应链研发设计的新一代训推一体芯片，于 2024 年 10 月交付流片，于 2025 年 7 月回片并成功点亮。发行人第三代高性能通用 GPU 芯片（曦云 C700 系列）研发项目于 2025 年 4 月立项，该系列芯片系基于国产先进工艺开发的通用 GPU 芯片，拥有更高的产品性能，对标目前国际主流产品英伟达 H100。

发行人坚持将技术创新服务于客户，以市场和客户需求为导向确定产品研发方向。未来，发行人将结合市场与客户需求、行业技术发展方向和自身发展战略继续加大研发投入，为发行人的持续创新提供强力保障。持续的研发投入是未来发行人实现技术领先的关键保障，为发行人长期经营发展奠定坚实基础。

4、客户和供应商集中

（1）客户集中情况分析

报告期内，发行人前五大客户主营业务收入占当期主营业务收入总额的比例为 100.00%、91.58%、71.09%、88.35%，发行人客户集中度相对较高，主要系发

行人销售规模处于爬坡阶段，整体收入规模相对较小，而下游客户单笔订单金额较大所致。发行人产品下游主要用途为集成服务器后建设智能算力集群，对于千卡以上的智能算力集群项目，通常投资金额数以亿计，投资规模大、实施周期长，故单个客户采购金额较大，导致在目前阶段发行人客户集中度相对较高。

报告期内，除 2022 年外，发行人不存在对单一客户销售收入占比超过 50% 的情况。2022 年度，发行人产品尚处于业务拓展初期，主要产品尚未实现量产，仅实现对上海东焱信息技术有限公司的少量销售，故当年对该客户销售收入占比较高。因此发行人不存在严重依赖单一客户的情况。

综上所述，报告期内发行人客户集中度较高，主要因销售规模处于爬坡阶段、下游客户单笔订单金额大所致，且除 2022 年外不存在对单一客户销售收入占比超 50% 的情况，未形成严重依赖单一客户的局面。随着未来收入规模扩大、客户结构优化，客户集中度有望下降，当前客户集中不会对发行人未来生产经营的持续性和稳定性产生重大不利影响。

（2）供应商集中情况分析

报告期内，发行人前五大供应商采购金额占当期采购总额的比例为 95.67%、95.01%、91.30%和 93.87%，发行人供应商集中度较高，前五大供应商采购占比相对稳定。其中，公司 D、公司 E、公司 C 为芯片供应商，发行人向其采购晶圆及委托加工服务；公司 A 为 HBM 内存等产品知名供应商，发行人向其采购 HBM 内存。晶圆制造、HBM 内存主要由少数具备先进工艺能力的龙头企业掌握，因此发行人对相关供应商采购金额相对较高，属于芯片设计行业特点。报告期内，发行人不存在对单一供应商采购比例超过 50% 的情况。

截至本回复出具日，报告期期末的大额预付款交付情况良好，未交付部分合同在正常履行中，不存在重大交付风险。为应对供应链安全风险，发行人已提前布局并稳步推进国产供应链替代方案，取得了有效进展。发行人下一代基于国产供应链的曦云 C600 系列已于 2025 年 7 月回片，预计 2025 年底前进入风险量产。

综上所述，芯片设计行业采购往往集中于少数具备先进工艺能力的供应商，导致供应商集中度普遍较高，报告期内发行人供应商集中度较高具有一定行业合理性。为应对供应链安全风险，发行人已提前布局国产替代方案，目前发行人供

应链风险相对可控。报告期内，发行人供应商集中度较高不会对公司未来生产经营的持续性和稳定性产生重大不利影响。

5、国际贸易形势

在高端芯片封锁升级、国产替代加速的大背景下，发行人作为国产 GPU 领域的代表企业，正迎来关键的发展机遇。随着国产人工智能芯片渗透率持续提升、国外厂商中国大陆市场份额明显下滑，发行人凭借自主可控的技术体系与领先的产品性能，处于国产 GPU 赛道的领先地位。未来，发行人有望持续受益于国产化趋势加速带来的广阔市场空间和政策支持，实现业绩的不断增长。同时，为避免受地缘政治等因素影响，发行人为保障原材料供应稳定进行了战略备货，并已提前布局国产替代方案，降低供应链安全风险。

综上所述，发行人经营活动现金流方面，公司通过股权融资积累了充裕货币资金，足以支撑持续经营，且随着销售规模扩大有望改善；营运资金覆盖充足，能满足运营需求；发行人研发投入持续增长助力公司掌握核心技术，并且在研项目进展顺利，为长期发展奠定基础；发行人客户与供应商集中系公司发展阶段及行业特点所致，且未形成对单一客户或供应商的严重依赖；在国际贸易形势下，发行人受益于国产替代趋势，通过战略备货和国产替代布局应对供应链风险。因此，上述因素不会对发行人未来生产经营的持续性和稳定性产生重大不利影响。

（三）在手订单的对应客户、收入实现可行性；结合不同产品的市场空间变化、发行人市场份额及变动、客户复购、新客户验证及拓展情况等，分析发行人业绩测算的合理性

1、在手订单的对应客户、收入实现可行性

截至 2025 年 4 月末，发行人在手订单金额（不含税）9.17 亿元，其中单笔订单金额超过 7,500 万元的客户分别为汇天网络科技有限公司、超讯通信、新华三、公司 3、浙江武珞智慧城市技术有限公司，合计约 7.01 亿元，占总订单金额比例为 76.42%。截至 2025 年 7 月末，上述在手订单已确认收入金额为 3.81 亿元，占在手订单总金额比例为 41.57%，在手订单转化收入情况良好。截至 2025 年 8 月 15 日，发行人在手订单金额（不含税）11.40 亿元，在手订单较充裕。

2、结合不同产品的市场空间变化、发行人市场份额及变动、客户复购、新客户验证及拓展情况等，分析发行人业绩测算的合理性

（1）业绩测算情况

发行人考虑到国际贸易形势变动带来的较大供给缺口和国产替代空间，预计市场份额未来将会持续提升。在国内算力需求持续扩大的背景下，结合公司在手订单情况、下游客户需求、市场开拓情况、产品预计量产时间、产品生命周期与产品销售规划，预计 2025 年至 2027 年公司销售收入持续增长，产品结构由曦云 C500 系列逐步向曦云 C600 系列和曦云 C700 系列升级迭代，综合毛利率保持稳定。

（2）业绩测算合理性

1) 产品市场空间变化

据公司研发和销售规划，2025 年至 2027 年公司主要销售产品为曦云 C 系列训推一体 GPU，主要应用于人工智能计算领域。根据研究机构 IDC 最新发布的报告《2025 年中国人工智能算力发展评估报告》，中国 2024 年 AI 算力规模增量为 308.6 EFLOPS，预计到 2027 年 AI 算力规模增量为 559.6 EFLOPS，复合增长率超过 34%。作为国产 GPU 领军企业之一，发行人产品受益于国产替代加速与下游智算集群建设需求，市场空间持续扩大，为销量增长奠定基础。此背景下，发行人预测销售收入持续增长且销售毛利率保持稳定，与行业高速增长、GPU 产品高壁垒、高附加值特性相符。

2) 市场份额及变动

综合考虑美国芯片封锁下国内市场较大供给缺口和国产替代空间，国内人工智能芯片行业国产化率有望加速渗透。根据 IDC 数据，2024 年中国本土人工智能芯片品牌渗透率约 30%，相较去年同期 15%的国产品牌渗透率已呈明显提升趋势。发行人作为国产替代主力军，已通过自主研发创新、技术积累和商业验证，成功替代国外厂商成为多个智算中心的 GPU 供应商，同时形成并保持了在国产 GPU 中的领先优势。按算力规模口径计算，2024 年发行人市场份额约为 1.14%，随着国产替代深化及产品竞争力提升，预测 2025-2027 年份额将逐步提升，贴合行业渗透节奏。同时，市场份额提升带来的规模效应，有助于发行人维

持毛利率稳定，因此，发行人业绩预测具有合理性。

3) 客户复购

报告期内，发行人旗舰型产品曦云 C500 系列推出市场后，凭借优秀的产品性能、易用性和集群实测表现赢得了客户认可，目前已在多个国家人工智能公共算力平台、运营商智算平台和商业化智算中心实现大规模部署。从报告期内发行人直接客户层面的变化情况来看，2023 年、2024 年和 2025 年 1-3 月由发行人存量客户贡献的收入金额分别为 0 万元、11,861.21 万元和 30,469.91 万元，占各期发行人主营业务收入的比例分别为 0%、15.99%和 95.10%，客户复购率持续提升。考虑到发行人直接客户中存在较多经销商和服务器厂商，其通常对应多个终端客户，同时大型智算集群项目具有多批次、分阶段的采购特点，而不同批次采购可能发生在同一年度或期间内，因此从发行人主要终端客户角度按项目口径（对应直接客户的单笔或多笔采购订单）统计客户复购情况相对更具有参考性，具体如下表所示：

单位：张

直接客户	终端项目	销售产品	数量	订单时间
新华三信息技术有限公司	项目 A（一期）	曦云 C500 板卡	515	2024 年 4 月
	项目 A（二期）	曦云 C500 板卡	2,052	2024 年 10 月
	终端客户 1 后续项目	曦云 C550 板卡	4,096	2025 年 4 月
上海算丰信息有限公司、香港算丰信息有限公司	项目 E（一期）	曦云 C500 服务器	1,280	2023 年 12 月
	项目 L	曦云 C500 服务器	160	2024 年 6 月
	项目 P	曦云 C550 板卡	328	2025 年 5 月
	项目 E（二期）	曦云 C500 服务器	1,280	2025 年 6 月
汇天网络科技有限公司	项目 F（一期）	曦云 C550 服务器	1,024	2024 年 8 月
	项目 F（二期）	曦云 C550 服务器	2,048	2025 年 2 月
上海源庐加佳信息科技有限公司	项目 D（一期）	曦云 C500 板卡	1,024	2024 年 7 月
	项目 D（二期）	曦云 C500 板卡	1,024	2024 年 8 月
		曦云 C500 板卡	1,024	2024 年 9 月
超讯通信股份有限公司	项目 C（一期）	曦云 C500 板卡	4,096	2024 年 12 月
	项目 C（二期）	曦云 C500 板卡	4,096	

注：为便于统计分析，上表中涉及发行人以服务器形态进行产品交付的，均已折算成对应的 GPU 板卡数量。

由上表可见，截至本回复出具日，基于前期交付产品的良好使用反馈，发行

人主要终端客户普遍在其后续项目（如有）中对发行人产品进行了复购并进一步加大了采购规模。对于发行人而言，国家人工智能公共算力平台、运营商智算平台和商业化智算中心等算力底座客户是重要的基本盘。存量客户持续采购，这为发行人的收入增长筑牢了基础，而且稳定的合作关系对维持毛利率水平有益。

4) 新客户验证及拓展情况

发行人持续推进“1+6+X”战略，一方面以国家人工智能公共算力平台、商用 AIDC 及运营商、互联网厂商承建的智算中心等数字算力底座为基本盘，凭产品优势受益于国产算力需求；另一方面借助 GPU 架构优势，围绕六大重点领域和长尾市场拓展，通过行业卡位挖掘新客户群体、拓展市场空间。

报告期内，按新老客户划分的主营业务收入情况如下：

单位：万元、家数

项目		2025 年 1-3 月	2024 年	2023 年	2022 年
新客户	数量	6	31	27	0
	收入金额	1,571.61	62,329.66	5,141.09	-
	收入占比	4.90%	84.01%	100.00%	-
老客户	数量	12	10	0	1
	收入金额	30,469.91	11,861.21	0.00	42.64
	收入占比	95.10%	15.99%	0.00%	100%
总客户数量		18	41	27	1
主营业务收入金额		32,041.53	74,190.87	5,141.09	42.64

注：客户为同一控制下合并；2022 年客户认定为老客户，2023 年至 2024 年及 2025 年 1-3 月新客户为当年首次实现销售收入的客户

报告期各期前 5 大新增客户主营业务收入情况如下：

单位：万元

期间	序号	客户名称	销售收入
2025 年 1-3 月	1	华科智谷（上海）科技发展有限公司	1,136.14
	2	广州祥泽科技有限公司	213.45
	3	北京兆维自服装备技术有限公司	129.74
	4	超聚变数字技术有限公司	49.56
	5	光子算数（北京）科技有限责任公司	21.40
	小计		1,550.29

期间	序号	客户名称	销售收入
	占新客户收入比例		98.64%
2024 年	1	新华三信息技术有限公司	15,095.15
	2	苏州瑞芯智能科技有限公司	13,046.02
	3	香港算丰信息有限公司	10,733.34
	4	汇天网络科技有限公司	7,332.18
	5	浙江武珞智慧城市技术有限公司	6,532.66
	小计		52,739.35
	占新客户收入比例		84.61%
2023 年	1	公司 1	2,216.67
	2	北京创世云科技股份有限公司	1,635.26
	3	上海闪马智能科技有限公司	600.00
	4	中科远见（重庆）科技有限公司	139.51
	5	湖南沐仕杰半导体科技有限公司	116.58
	小计		4,708.02
	占新客户收入比例		91.58%

如上表所示，2023 年和 2024 年发行人产品曦思 N100 系列产品、曦云 C500 系列产品陆续正式量产，新客户销售数量、金额及占比大幅增长。2025 年 1-3 月，新客户收入占比下降，主要原因为，一是发行人产品得到老客户的认可，老客户持续复购；二是发行人通过经销商模式扩大了终端客户覆盖范围，2025 年 1-3 月主营业务收入中经销商模式收入占比为 55.71%，较上年增加 18.29 个百分点。

发行人持续拓展新客户，截至 2025 年 8 月 12 日，累计已有 230 余家客户正在或已测试发行人 GPU 板卡，其中 2024 年 12 月末在测试客户家数为 26 家、2025 年 6 月末在测试客户家数为 43 家，呈现出稳步增长的态势。发行人正在与众多客户进行商务洽谈，客户验证及开拓情况较好。发行人下一代旗舰产品曦云 C600 已于 2025 年 7 月回片并成功点亮，并在 2025 年 7 月 26 日举办的世界人工智能大会上正式发布，发布后市场反响较好，具体参见本问询回复报告之“13.关于募投项目”之“一/（一）三个研发项目的具体研发内容，研发投入资金的具体使用安排，结合曦云 C600、曦云 C700、曦思 N 系列芯片产品的客户开拓、市场销售情况，分析有关募投项目建设的必要性及未来发展安排”。发行人新客户验

证拓展情况良好，新产品研发进展顺利，有力的支撑了收入规模扩张和毛利率维持在合理水平。

综上，公司业绩测算综合考虑了市场空间变化、发行人市场份额及变动、客户复购、新客户验证及拓展情况等因素，具有合理性。

（四）结合前述情况，说明发行人前瞻性分析相关参数及假设依据是否合理，按照《监管规则适用指引-发行类第 5 号》《监管规则适用指引—发行类第 10 号》要求详细披露原因分析、影响分析、趋势分析、风险因素、投资者保护措施及承诺等事项，并结合研发进度、商业化前景等因素，审慎估计预计实现盈利时间等前瞻性信息，补充披露关于预计盈利时点的敏感性分析，并完善相关重大事项提示

1、结合前述情况，说明发行人前瞻性分析相关参数及假设依据是否合理

（1）营业收入和毛利率

营业收入和毛利率前瞻性分析的相关参数及假设依据的合理性已在本问题回复之“（三）/2、结合不同产品的市场空间变化、发行人市场份额及变动、客户复购、新客户验证及拓展情况等，分析发行人业绩测算的合理性”中说明。

（2）期间费用构成及变化情况

报告期内，发行人已自研形成了覆盖计算和渲染场景的核心 GPU IP，打造了超过 600 条（XCORE1.0 计算 GPU）、800 条（XCORE2.0 渲染 GPU）指令组成的 MXMACA 指令集，技术复用性增强，研发效率提升。随着产品迭代进入“量产一代、在研一代、规划一代”的稳定周期，研发投入占比或逐步稳定。此外，经过前期的快速扩张，公司已建立了较为稳定的研发、管理及销售团队，未来公司的管理费用、销售费用的增长也会趋于稳定，公司期间费用率将逐渐趋于稳定。

（3）其他损益

发行人根据报告期内和 2025 年上半年情况，对资产减值损失、信用减值损失、政府补助等其他其他损益进行了合理预测。

（4）损益趋势变化分析

基于前述营业收入、毛利率、期间费用以及其他损益的分析与假设，发行人

预计最早于 2026 年实现扭亏为盈。发行人能否在 2026 年盈利主要受到营业收入增速、毛利率水平以及期间费用的综合影响。如发行人 2026 年营业收入和毛利率情况不及预期，则可能当年无法实现盈利。具体敏感性分析如下：

营业收入 毛利率	下降 10%	不变	上升 10%
减少 5 个百分点	2027 年盈利	2027 年盈利	2026 年盈利
不变	2027 年盈利	2026 年盈利	2026 年盈利
增加 5 个百分点	2027 年盈利	2026 年盈利	2026 年盈利

注：上述营业收入及毛利率是在原有预测基础上变动

发行人从自身经营情况出发，结合产品市场空间、市场份额及变动、客户复购和新客户验证及拓展情况对未来损益情况进行了预测，具有合理性。但是，上述前瞻性信息是建立在推测性假设的数据基础上的预测，具有重大不确定性，投资者进行投资决策时应谨慎使用。

2、按照《监管规则适用指引-发行类第 5 号》《监管规则适用指引—发行类第 10 号》要求详细披露原因分析、影响分析、趋势分析、风险因素、投资者保护措施及承诺等事项，并结合研发进度、商业化前景等因素，审慎估计预计实现盈利时间等前瞻性信息，补充披露关于预计盈利时点的敏感性分析

（1）原因分析

发行人尚未盈利的原因已在招股说明书中“第六节”之“六/（九）/1、原因分析”补充披露如下：

“1、原因分析

报告期内，公司持续亏损，造成累计未弥补亏损金额较大，主要原因为：

（1）中国 GPU 芯片市场曾长期被国外巨头垄断，国产芯片渗透率低

我国 GPU 市场长期由英伟达、AMD 等国际厂商主导，尤其是英伟达通过 CUDA 生态构建了完整的软硬件技术标准体系，形成行业事实性垄断。这种市场格局导致国产芯片面临技术标准适配、用户习惯迁移双重障碍，生态建设需系统性突破，市场拓展呈现渐进式发展特征。同时，国产 GPU 厂商整体起步相对较晚，过去几年大多处于技术突破和产品落地初期的阶段，在核心技术积累、产品生态建设、市场认知度等方面均有待提升，需投入大量资源用于技术研发、市场

拓展和生态建设，导致成本居高不下。

（2）重点行业产品导入需经历技术验证及生态适配周期

公司专注于打造独立自主的高性能通用 GPU 芯片及 AI 计算平台生态，公司智算推理 GPU 芯片曦思 N100 系列、训推一体 GPU 芯片曦云 C500 系列分别于 2023 年 4 月和 2024 年 2 月正式量产。GPU 芯片作为支撑图形渲染、人工智能及高性能计算的核心硬件，其性能表现直接决定了终端设备的运算效能与场景适配能力。因此，在进入国家人工智能公共算力平台、运营商、大型互联网企业等重点行业客户时，需经历严苛的技术验证及生态适配周期。报告期内，公司销售规模处于快速爬坡阶段，营业收入分别为 42.64 万元、5,302.12 万元、74,307.16 万元和 32,041.53 万元，虽然增长迅速，销售毛利规模仍不足以覆盖高额的研发支出。

（3）公司所处 GPU 芯片行业具有极高的技术壁垒与密集的研发投入的行业特点

GPU 芯片行业作为典型资本与技术双密集型领域，公司为突破国际生态垄断并实现自主可控目标，始终坚持自主创新，持续加大核心技术攻坚力度，报告期内公司研发费用分别为 64,752.09 万元、69,862.04 万元、90,089.04 万元和 21,790.41 万元，占营业收入比例分别为 151857.63%、1317.63%、121.24%和 68.01%。

（4）股份支付费用较大

公司重视人才激励与长期发展，实施了多轮次的员工股权激励，相应确认了大额的股份支付费用。报告期内，公司确认的股份支付费用分别为 3,978.62 万元、5,762.38 万元、47,283.29 万元和 3,156.31 万元。”

（2）影响分析

发行人尚未盈利的影响已在招股说明书中“第六节”之“六/（九）/2、影响分析”披露如下：

“（1）对现金流的影响

报告期内，公司经营活动现金流量净额分别为-66,514.29 万元、-101,731.89

万元、-214,802.33 万元和-53,131.75 万元，经营活动现金流量净额为负的主要原因是公司销售规模尚在逐步提升过程中，原材料采购、存货备货和研发投入的资金需求较大。报告期内，随着公司商业化持续落地，投资者对公司的认可程度不断提升，公司通过股权融资能力持续增强。截至 2025 年 3 月末，公司账面货币资金、交易性金融资产合计 729,149.55 万元；报告期各期，公司现金及现金等价物净增加额分别为-11,099.40 万元、284.06 万元、78,910.10 万元、467,138.47 万元，公司具备一定规模的资金储备，现金流状况良好。未来，随着公司销售规模进一步增加，盈利能力将得到增强，经营活动现金流量净额将得到改善。因此，公司尚未实现盈利且最近一期末存在累计未弥补亏损不会对公司现金流产生重大不利影响，公司现金流状况良好，为公司业务拓展、人才吸引、团队稳定性、研发投入、战略性投入、生产经营可持续性带来了有力保障。

（2）对业务拓展的影响

发行人坚持开放共赢的发展理念和“1+6+X”的发展战略，凭借领先的技术实力、长期的行业资源积淀已与产业链上下游建立了良好的合作关系，构成了由智算中心、服务器整机厂、操作系统厂商、大模型厂商、AI 应用方等组成的合作生态，商业化、规模化落地进展领先行业。截至报告期末，公司 GPU 产品累计销量超过 25,000 颗。**发行人的产品相继应用部署于 10 余个智算集群**，算力网络覆盖国家人工智能公共算力平台、**运营商智算平台**和商业化智算中心，横跨北京、上海、杭州、长沙、中国香港等地区，并逐渐向更多区域延伸。发行人 GPU 板卡已搭载于新华三等多家服务器厂商的产品中，在多个国家人工智能公共算力平台、**运营商智算平台**、商业化智算中心中实现规模化应用。在垂类场景中，发行人 GPU 产品深度赋能众多行业应用场景，已率先在教科研、金融、交通、能源、医疗健康、大文娱等行业实现落地应用，产品赋能真实应用场景的竞争力和交付能力得到充分验证。因此，公司尚未实现盈利且最近一期末存在累计未弥补亏损不会对公司业务拓展产生重大不利影响。

（3）对人才吸引和团队稳定的影响

经过多年的发展，公司已凝聚了一支拥有丰富行业经验和研发管理经验，涵盖公司技术研发、管理和销售等各个经营管理环节的人才队伍。报告期内核心技术人员保持稳定，公司人才队伍不断扩大。公司建立了相对完善的薪酬福利体系，

明确人才激励机制。报告期内，公司实施多次员工股权激励，并建立了完善的激励机制，能够有效保障人才吸引力和团队稳定性。虽然股份支付会造成公司产生会计上的利润损失，但并不会造成公司的现金流出。同时，充分的股权激励将有利于统一公司与员工的利益使公司经营发展从中受益。因此，公司尚未实现盈利且最近一期末存在累计未弥补亏损不会对公司团队建设产生重大不利影响。

（4）对研发投入和战略性投入的影响

GPU 芯片行业是典型的技术密集型产业，公司为突破国际生态垄断并实现自主可控目标，始终将研发创新作为企业发展的核心驱动力，研发投入被列为长期战略规划的重要组成部分。报告期内公司研发费用分别为 64,752.09 万元、69,862.04 万元、90,089.04 万元和 21,790.41 万元，呈现显著增长态势。在不懈的技术攻坚下，公司成为了国内少数几家系统掌握了先进制程 GPU 芯片及其基础系统软件研发、设计和量产技术的企业之一，深度积累了 GPU IP（包括指令集、微架构等）、GPU SoC、高速互连、GPU 软件等核心技术，打造了自主开放、高度兼容国际主流 GPU 生态（CUDA）的软件生态体系，能够为客户构建软硬件一体的全面生态解决方案，在底层技术上摆脱对国外算力资产的依赖，为推进新质生产力发展提供动力引擎。因此，公司尚未实现盈利且最近一期末存在累计未弥补亏损不会对公司研发投入和战略性投入产生重大不利影响。

（5）对生产经营可持续性的影响

自成立以来，发行人持续深耕 GPU 和人工智能行业，快速实现了技术的产业化输出。聚焦人工智能计算、通用计算和图形渲染三大领域，发行人基于自主研发的 GPU IP 和统一的 GPU 计算架构，先后推出了用于智算推理的曦思 N 系列 GPU、用于训推一体和通用计算的曦云 C 系列 GPU，以及正在研发用于图形渲染的曦彩 G 系列 GPU，产品性能达到了行业先进水平，在通用性、单卡性能、集群性能及稳定性、生态兼容与迁移效率等方面具备较强的核心竞争力，综合技术实力领先行业。报告期各期，公司营业收入分别为 42.64 万元、5,302.12 万元、74,307.16 万元和 32,041.53 万元，2022-2024 年三年复合增长率为 4074.52%。从行业趋势看，伴随全球大模型参数量级跃升、AI 需求爆发式增长以及国产算力替代加速，高性能 GPU 市场进入战略机遇期。根据 Verified Market Research 的数据，2024 年全球 GPU 市场规模为 773.9 亿美元，2030 年有望达到 4,724.5 亿

美元，2024-2030 年的复合增长率高达 35.19%，呈现强劲的增长态势。根据中商产业研究院数据，2024 年中国 GPU 市场规模约为 1,073 亿元，同比增长 32.96%，国产替代空间显著。基于强劲的市场需求、已验证的商业化路径以及持续优化的产品结构，公司尚未实现盈利且最近一期末存在的累计未弥补亏损不会对公司生产经营可持续性产生重大不利影响。”

（3）趋势分析

发行人尚未盈利的趋势分析已在招股说明书中“第六节”之“六/（九）/3、趋势分析”，披露如下：

“3、趋势分析

收入方面，沐曦股份作为国内高性能通用 GPU 领域的领军企业，其产品矩阵覆盖人工智能训练、推理、通用计算及图形渲染全场景，已形成曦云（训推一体）、曦思（智算推理）、曦彩（图形渲染）三大系列 GPU 产品，性能指标达到国际主流水平。凭借突出的产品性能和稳定的供应能力，截至报告期末，发行人 GPU 产品累计销量超过 25,000 颗。发行人深度构建“1+6+X”生态与商业布局，基础算力底座方面，公司产品相继应用部署于 10 余个智算集群，算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心，区域横跨北京、上海、杭州、长沙、中国香港等地区，并逐渐向更多区域延伸。随着大模型迭代加速、国产算力替代需求激增，公司产品凭借高兼容性（MXMACA 软件生态高度兼容 CUDA）、高集群扩展性（自研 MetaXLink 互连技术支持 2-64 卡多种互连拓扑结构）、高能效比等优势，已率先在教科研、金融、交通、能源、医疗健康、大文娱等垂类行业场景实现布局，商业化进展领先行业。未来，伴随国产供应链版芯片曦云 C600 量产，产品线进一步完善，叠加智算中心建设及运营商、互联网企业采购需求释放，公司收入结构有望进一步多元化，市场份额及营收增速预计保持高位。

成本及费用支出方面，公司采用 Fabless 模式，聚焦芯片研发设计与生态构建，晶圆制造、封装测试等环节外包，生产端成本随国产供应链成熟及规模化量产有望持续优化。报告期内，公司已自研形成了覆盖计算和渲染场景的核心 GPU IP，打造了超过 600 条（XCORE1.0 计算 GPU）、800 条（XCORE2.0 渲染 GPU）

指令组成的 MXMACA 指令集，技术复用性增强，研发效率提升。随着产品迭代进入“量产一代、在研一代、规划一代”的稳定周期，研发投入占比或逐步稳定。此外，经过前期的快速扩张，公司已建立了较为稳定的研发、管理及销售团队，未来公司的管理费用、销售费用的增长也会趋于稳定，公司期间费用率将逐渐趋于稳定。

综上所述，公司产品具备良好的市场前景和广阔的市场空间，市场份额及营收增速预计保持高位，未来随着发行人持续进行成本优化及费用控制，毛利率和期间费用率将趋于稳定，新产品的放量销售将持续为公司带来业绩贡献，公司亏损有望持续收窄，实现扭亏为盈。

上述趋势信息为发行人管理层基于发行人的经营状况、市场情况及趋势作出的预测，受到上述多重因素的影响，该等预测性信息与未来的实际情况可能存在一定的偏差。发行人提醒投资者注意，相关假设的数据基础及相关预测具有较大不确定性，投资者在进行投资决策时应谨慎使用。”

（4）风险因素

发行人已就相关风险因素做出提示，参见公司招股说明书“第二节”之“一/（一）/1、持续亏损和存在累计未弥补亏损的风险”及“第三节”之“一/（二）/1、持续亏损和存在累计未弥补亏损的风险”。

（5）投资者保护措施及承诺

发行人关于投资者保护措施及承诺已在招股说明书中“第六节”之“六/（九）/5、投资者保护措施及承诺”，披露如下：

“5、投资者保护措施及承诺

关于依法落实保护投资者合法权益规定的各项措施具体参见公司招股说明书“第九节”之‘三、存在特别表决权股份、协议控制架构或类似特殊安排，尚未盈利或存在累计未弥补亏损的，关于投资者保护的措施’。

公司于 2025 年 5 月 20 日召开 2024 年年度股东会，审议通过了《关于公司首次公开发行股票前滚存未分配利润处置方案的议案》，同意公司本次发行上市前留存的滚存未分配利润（累计未弥补亏损）全部由公司本次发行上市后的新老

股东按上市后的持股比例共同享有（承担）。”

（6）预计实现盈利时间

发行人预计实现盈利时间已在招股说明书“第六节”之“六/（九）/6、预计实现盈利时间”中披露，关于预计实现盈利时间补充披露如下：

“6、预计实现盈利时间

公司业务规模总体保持增长态势，在训推一体 GPU 产品毛利率保持相对稳定的情况下，公司的主营业务毛利将快速增长，同时随着新产品的研发成功及客户逐步导入，公司盈利能力将得以进一步提升。在主营业务收入保持增长的情况下，公司期间费用占营业收入的比重也将逐步降低。基于公司的测算，公司达到盈亏平衡点的预期时间最早为 2026 年。

公司从自身经营情况出发，结合产品市场空间、市场份额及变动、客户复购和新客户验证及拓展情况，预计公司达到盈亏平衡点的预期时间最早为 2026 年。公司能否在 2026 年盈利主要受到营业收入增速、毛利率水平以及期间费用的综合影响。如公司 2026 年营业收入和毛利率情况不及预期，则可能当年无法实现盈利。具体敏感性分析如下：

营业收入 毛利率	下降 10%	不变	上升 10%
	2027 年盈利	2027 年盈利	2026 年盈利
减少 5 个百分点	2027 年盈利	2026 年盈利	2026 年盈利
不变	2027 年盈利	2026 年盈利	2026 年盈利
增加 5 个百分点	2027 年盈利	2026 年盈利	2026 年盈利

注：上述营业收入及毛利率是在原有预测基础上变动

前述达到盈亏平衡状态时主要经营要素需要达到的水平不构成公司的盈利预测或业绩承诺。上述前瞻性信息是建立在推测性假设的数据基础上的预测，具有重大不确定性，投资者进行投资决策时应谨慎使用。”

3、完善相关重大事项提示

发行人在招股说明书“第二节”之“一/（一）/1、持续亏损和存在累计未弥补亏损的风险”及“第三节”之“一/（二）/1、持续亏损和存在累计未弥补亏损的风险”中补充披露如下：

“1、持续亏损和存在累计未弥补亏损的风险

报告期各期，公司营业收入分别为 42.64 万元、5,302.12 万元、74,307.16 万元和 32,041.53 万元，归属于母公司所有者的净利润分别为-77,696.52 万元、-87,115.82 万元、-140,887.94 万元和-23,251.22 万元，尚未实现盈利。

公司报告期内尚未实现盈利，主要系：（1）中国 GPU 芯片市场曾长期被国外巨头垄断，国产 GPU 芯片渗透率低，面临技术标准适配及用户习惯迁移障碍，市场拓展呈渐进式发展，且国产 GPU 厂商整体起步较晚，处于技术突破和产品落地初期，在多方面有待提升，需投入大量资源用于研发、拓展和生态建设，导致成本居高不下；（2）公司智算推理 GPU 芯片曦思 N100 系列、训推一体 GPU 芯片曦云 C500 系列分别于 2023 年 4 月和 2024 年 2 月正式量产，因公司产品进入重点行业客户需经历严苛的技术验证及生态适配周期，公司销售规模处于快速爬坡阶段，目前收入规模仍然难以覆盖成本费用支出；（3）公司所处 GPU 芯片行业属资本与技术双密集型，有极高技术壁垒与密集研发投入特点，公司为夯实产品竞争优势持续加大研发投入；（4）公司因实施股权激励而确认了较大金额的股份支付费用。

因市场景气度、行业竞争、客户拓展、供应链管理等影响经营结果的因素较为复杂，发行人的营业收入可能无法按预期增长，存在未来一段时期内持续亏损的风险。截至 2025 年 3 月末，公司合并报表、母公司报表未分配利润分别为 -104,752.49 万元、-98,769.23 万元，预计首次公开发行股票并上市后，发行人账面累计未弥补亏损将持续存在，导致一定时期内无法向股东进行现金分红。”

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、申报会计师执行了如下核查程序：

1、查阅发行人财务报表，对发行人报告期内持续亏损的主要影响因素进行量化分析。

2、查阅行业研究报告、同行业公司信息披露文件以及访谈发行人管理层，了解行业周期、供需关系、市场竞争以及发行人发展所处阶段，分析发行人报告期内持续亏损的主要影响因素、亏损趋势未明显改善的原因及合理性。

3、获取发行人 2025 年 4 月至 7 月未审计财务数据，了解发行人期后经营情

况。

4、查阅发行人财务报表、行业研究报告以及与访谈发行人管理层，了解发行人经营现金流量、营运资金覆盖、研发投入、客户供应商集中和国际贸易形势情况，分析对发行人未来生产经营的影响。

5、获取发行人 2025 年 4 月末和 8 月 15 日的在手订单、2025 年 5 月至 7 月末的销售收入清单，了解在手订单的收入实现情况；获取发行人目前正在洽谈的主要订单情况并访谈发行人管理层，了解在手订单金额变动的原因。

6、查阅行业研究报告和同行业公司信息披露文件，获取发行人期后未审财务数据和截止 2025 年 7 月末的最新业绩预测资料，访谈发行人管理层，了解发行人产品市场空间变化、市场份额及变动、客户复购、新客户验证及拓展情况，对业绩预测相关参数进行复核，对发行人前瞻性分析相关参数及假设依据的合理性进行复核。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、发行人报告期内持续亏损及亏损趋势未明显改善原因如下：（1）受行业周期、供需关系、市场竞争等行业共性因素影响，国产人工智能芯片厂商多处于亏损状态；（2）发行人自 2020 年成立后持续高强度研发，报告期研发费用逐年增长且占营收比例较高，而收入规模虽快速增长但整体仍小，销售毛利难以覆盖成本，叠加大额股份支付费用导致持续亏损；（3）2024 年受股份支付金额、资产减值损失、信用减值损失等因素影响，亏损幅度有所扩大。发行人报告期内持续亏损及亏损趋势未明显改善具有合理性。

2、发行人经营活动现金流方面，发行人通过股权融资积累了充裕货币资金，足以支撑持续经营，且随着销售规模扩大有望改善；营运资金覆盖充足，能满足运营需求；发行人研发投入持续增长助力公司掌握核心技术，并且在研项目进展顺利，为长期发展奠定基础；发行人客户与供应商集中系公司发展阶段及行业特点所致，且未形成对单一客户或供应商的严重依赖；在国际贸易形势下，发行人受益于国产替代趋势，通过战略备货和国产替代布局应对供应链风险。因此，上述因素不会对发行人未来生产经营的持续性和稳定性产生重大不利影响。

3、发行人在手订单转化率良好，潜在订单充裕；发行人业绩测算综合考虑了市场空间增长、份额提升、客户复购与新客户拓展等因素，销量与收入增长路径贴合行业趋势，毛利率水平与市场地位相匹配，业绩测算具有合理性。

4、发行人前瞻性分析中营业收入、毛利率、期间费用等相关参数及假设依据具有合理性，但是，上述前瞻性信息是建立在推测性假设的数据基础上的预测，具有重大不确定性，投资者进行投资决策时应谨慎使用。保荐机构和申报会计师已按照《公开发行证券的公司信息披露内容与格式准则第 57 号—招股说明书》《监管规则适用指引-发行类第 5 号》《监管规则适用指引—发行类第 10 号》对发行人未盈利情况、持续经营能力进行了核查，并在招股说明书中详细披露了原因分析、影响分析、趋势分析、风险因素、投资者保护措施及承诺等事项；并结合研发进度、商业化前景等因素，预计发行人最早将于 2026 年实现盈利。发行人已在招股说明书中补充披露关于预计盈利时点的敏感性分析，并完善了相关重大事项提示。

4.关于收入

根据申报材料：（1）报告期内，发行人营业收入为 42.64 万元、5,302.12 万元、74,307.16 万元和 32,041.53 万元，大幅增长，但各类产品平均销售单价均有所下降；（2）发行人产品在进入国家人工智能公共算力平台、运营商、大型互联网企业等重点行业客户时，需经历严苛的技术验证及生态适配周期；（3）对于 GPU 板卡销售业务，于客户签收时确认收入；对于算力集群销售业务，于客户完成集群验收，获取客户验收单时确认收入；对于 IP 销售业务，于客户完成对 IP 的形式验收，获取客户确认邮件时确认收入；（4）2022 年至 2024 年，第四季度收入占比为 100.00%、89.01%、70.06%，2024 年第三季度收入占比 5.66%。

请发行人披露：（1）结合报告期内下游客户需求变动、主要客户的技术验证及适配过程、量产供货时间等，说明报告期内营业收入大幅增长的主要驱动因素；（2）报告期内各类产品整体单价均呈下降趋势的原因及合理性、未来是否会持续下降，与行业周期、趋势、同行业可比公司是否一致；（3）结合各类明细产品或业务交付产品的时点、形式以及客户签收、验收相关流程，说明各产品或业务的收入确认方法与业务实质及同行业可比公司是否一致，是否符合会计准则相关规定；（4）报告期内第四季度收入占比较高、2024 年第三季度收入占比较低的原因及合理性，预计未来收入季节性的趋势。

请保荐机构、申报会计师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）结合报告期内下游客户需求变动、主要客户的技术验证及适配过程、量产供货时间等，说明报告期内营业收入大幅增长的主要驱动因素

1、报告期内下游客户需求变动情况

（1）我国算力总需求增长为国产 AI 芯片提供广阔发展空间

2023 年以来，以 Chat GPT 为代表的生成式 AI 和大语言模型技术浪潮席卷全球，引发了对 AI 算力需求的指数级增长，千亿甚至万亿的参数模型的训练以及海量用户请求的推理服务，极大地促进了市场对于算力的需求。

报告期内，尽管面临外部供应限制，我国对 AI 算力的总体需求并未放缓，反而因各行各业智能化转型而持续高速增长，国内智算中心持续建设与扩容。IDC 数据显示，2020 年中国智能算力规模为 75.0 EFLOPS（基于 FP16 计算，下同），到 2028 年预计将达到 2,781.9 EFLOPS，预计 2020-2028 年复合增长率达到 57.1%。我国积极稳步扩大算力规模，对 AI 算力的刚性总需求为国产 AI 芯片的发展提供了广阔的市场空间。

（2）国产替代政策加速推进国产 AI 芯片市占率提升

报告期内，美国系统升级对我国 AI 芯片出口管制措施，例如 2022 年《先进计算芯片出口管制规则》、2023 年《最终规则》以及 2024 年《人工智能扩散规则》等。与此同时，我国也加速了人工智能产业链“自主可控”的长期战略布局，通过信创等政策，鼓励并要求政府、金融、能源、交通等关键行业的 IT 系统逐步替换为国产化产品。

在国内外政策背景下，国家人工智能公共算力平台、运营商、大型互联网企业等重点行业客户一方面采购英伟达的“特供版”芯片，另一方面也以前所未有的力度启动了对国产 AI 芯片的测试和适配工作。根据 TrendForce 数据显示，我国市场预计外购英伟达、AMD 等芯片的比例会从 2024 年约 63%下降至 2025 年约 42%，而我国本土芯片供应商在国产 AI 芯片政策支持下，预期 2025 年占比将提升至 40%，几乎与外购芯片比例相当。

（3）国产 AI 芯片的综合性能系统性提升进一步促进了下游客户需求的转变

在上述宏观背景的驱动下，下游客户在 GPU 产品采购中的核心决策逻辑已发生显著调整。下游客户对 GPU 厂商的关注焦点从峰值性能、CUDA 生态等传统维度，逐步拓展至政策合规性、供应链稳定性、软件生态成熟度及综合性价比等多元层面；对产品的评估体系也相应从单卡性能、理论算力、硬件采购成本等单一指标，向集群运行效能、实际应用性能、生态协同能力等综合维度转变。

与此同时，国产 GPU 产品在理论性能、实测表现、兼容性及可靠性等关键领域持续突破，已具备支撑下游客户规模化采购的坚实基础。以发行人核心产品曦云 C500 板卡为例，该产品在性能层面对标英伟达 A100，部分应用场景下的

核心指标已比肩国际主流产品；通过兼容 CUDA 生态大幅降低客户应用迁移成本与开发门槛，并凭借自研的 MetaXLink 互联技术，精准响应大型客户对集群计算效能的核心诉求。

由此可见，在外部技术限制与国内政策支持的双重背景下，国产 GPU 厂商在产品性能、兼容性、可靠性及可用性等方面的系统性提升，为下游客户的需求转型提供了进一步的有力支撑。

2、主要客户的技术验证及适配过程、量产供货时间

AI 芯片从设计制造到最终在客户端实现应用，需经过复杂而严谨的技术验证与适配过程，客户在引入国产人工智能芯片产品时，其技术验证与适配过程旨在全方位评估芯片的性能、兼容性、稳定性及在真实业务场景中的应用潜力。具体而言，发行人下游客户技术验证及适配的内容和过程主要包括如下方面：

一是芯片的基础性能与稳定性测试，是评估芯片核心性能的基础环节，客户主要关注的核心指标包括：（1）算力性能（TOPS），系衡量芯片计算能力的首要指标；（2）能效比（TOPS/Watt），是评估芯片经济性的关键指标；（3）功耗与散热，直接关系到智算中心电力和散热配套设施成本以及系统的长期稳定运行；（4）稳定性，通过长时间、高负载的压力测试，监测芯片是否出现宕机、性能衰减或错误率升高等问题。

二是广泛的兼容性适配，决定 AI 芯片能否顺利融入客户现有 IT 生态系统，主要包括：（1）信创兼容性适配，信创兼容要求 AI 芯片必须能与国产 CPU 和国产操作系统等国产软硬件良好适配；（2）主流 AI 框架兼容性：AI 芯片需要支持或兼容主流的深度学习框架，通常需要芯片厂商提供相应的软件开发工具包（SDK）、编译器和优化库，以确保模型可以在其硬件上高效执行；（3）主流模型兼容性：验证 AI 芯片对业界广泛使用的主流模型以及客户所在行业的常用模型的支持度，对于 DeepSeek 和 Qwen 等大语言模型，客户会进行专门的性能测试；（4）CUDA 兼容性：对于非英伟达 AI 芯片而言，客户高度关注芯片厂商是否能够提供高度兼容 CUDA 的迁移方案，以最大程度降低客户的迁移成本。

三是模型性能与适配测试，主要包括：（1）主流大模型性能测试，针对当前热门的大语言模型（如 DeepSeek/Qwen 系列），客户会进行严格的性能对标

测试；（2）小模型测试，对于特定应用场景，客户更关注 AI 芯片在轻量化小模型上的效率、延迟和成本效益；（3）私有模型迁移与适配，部分客户拥有自研的、基于其业务数据训练的私有模型，较为关注从现有英伟达 GPU 平台迁移至新芯片平台的难易程度。

综上所述，国产人工智能芯片的客户导入是一个系统性工程，需平衡性能、兼容性、可靠性与时间成本等多个因素。发行人训推一体系列 GPU 产品推出并正式量产后，在较短时间内完成了下游主要客户的技术验证和适配并实现量产出货，算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心。发行人 GPU 产品深度赋能众多行业应用场景，已率先布局教科研、金融、交通、能源、医疗健康大文娱等行业，产品赋能真实应用场景的竞争力和交付能力得到充分验证。

具体而言，报告期内，发行人主要客户的技术验证及适配过程，量产供货时间如下表所示：

序号	客户名称	报告期内主要实现销售的产品类型	开始测试适配时间	签订首笔订单时间	量产供货时间	开始测试适配至实现量产供货时间
1	新华三信息技术有限公司	训推一体 GPU 板卡	2023 年 11 月	2024 年 3 月	2024 年 5 月	6 个月
2	超讯通信股份有限公司	训推一体 GPU 板卡、智算推理 GPU 板卡	2024 年 4 月	2024 年 4 月	2025 年 1 月	9 个月
3	苏州瑞芯智能科技有限公司	训推一体 GPU 板卡	2024 年 4 月	2024 年 12 月	2024 年 12 月	8 个月
4	上海源庐加佳信息科技有限公司	训推一体 GPU 板卡	2024 年 1 月	2024 年 1 月	2024 年 11 月	10 个月
5	香港算丰信息有限公司	训推一体 GPU 服务器	2023 年 12 月	2023 年 12 月	2024 年 5 月	5 个月
6	汇天网络科技有限公司	训推一体 GPU 服务器	2024 年 3 月	2024 年 8 月	2024 年 11 月	8 个月
7	浙江武珞智慧城市技术有限公司	训推一体 GPU 板卡	2024 年 3 月	2024 年 4 月	2024 年 10 月	7 个月
8	上海华研企业发展（集团）有限公司	训推一体 GPU 板卡	2023 年 12 月	2023 年 12 月	2024 年 6 月	6 个月
9	公司 2	训推一体 GPU 板卡	2023 年 12 月	2024 年 1 月	2024 年 8 月	8 个月

序号	客户名称	报告期内主要实现销售的产品类型	开始测试适配时间	签订首笔订单时间	量产供货时间	开始测试适配至实现量产供货时间
10	北京创世云科技股份有限公司	训推一体 GPU 板卡、智算推理 GPU 板卡	2023 年 7 月	2023 年 9 月	2023 年 12 月	5 个月

注 1：上表中主要客户列示范围为三年一期合计前十大客户

注 2：量产供货时间是指发行人与客户首个 1,000 万元（不含税）以上订单确认收入的时间

注 3：报告期内智算推理板卡产品收入占比相对较小，非发行人报告期营业收入增长的主要驱动因素，因此上表中仅列示训推一体系列产品的开始测试适配时间、签订首笔订单时间和量产供货时间

综上所述，发行人报告期内营业收入的显著增长主要得益于以下几方面因素：

首先，随着人工智能大模型的快速发展，我国对算力的需求呈现爆发式增长，为公司业务带来了强劲的市场需求；其次，受国际形势变化影响，境外供应链限制措施不断收紧，促使下游客户更加依赖国产 GPU 产品，从而推动了国产替代进程的加速；此外，近年来国产 GPU 产品在技术实力、产品性能和性价比等方面持续提升，进一步增强了其市场竞争力，使得下游客户对国产 GPU 产品的采购需求大幅上升。

发行人凭借在产品布局上的前瞻性和市场敏感度，于 2024 年初正式量产训推一体 GPU 产品，推出时间与下游客户需求的转变高度契合。同时，该产品在性能指标上表现优异，成功通过了下游客户较为严格的测试、技术验证及适配流程，显示出发行人产品在实际应用场景下的稳定性和可靠性，从而使得发行人能够迅速打开市场，实现向服务器 OEM 厂商、智算集群客户等下游客户的大规模出货，从而带动了公司营业收入的跨越式增长。

（二）报告期内各类产品整体单价均呈下降趋势的原因及合理性、未来是否会持续下降，与行业周期、趋势、同行业可比公司是否一致

1、报告期内各细分产品的销售均价及变动原因

（1）发行人主要细分产品报告期内销售单价变动总体情况

报告期各期，发行人主要细分产品销售单价报告期内变动情况如下：

单位：万元/张（不含税）、万元/台（不含税）

产品类型	2025 年 1-3 月		2024 年		2023 年		2022 年
	销售 单价	变动率	销售 单价	变动率	销售 单价	变动率	销售 单价
训推一体 GPU 板卡	3.89	-17.18%	4.69	-17.47%	5.69	-	-
其中：曦云 C500	3.70	-22.69%	4.78	-15.87%	5.69	-	-
其中：曦云 C550	5.99	31.57%	4.55	-	-	-	-
训推一体 GPU 服务器	52.01	-15.01%	61.20	-	-	-	-
其中：曦云 C500	35.00	-44.91%	63.53	-	-	-	-
其中：曦云 C550	69.03	20.50%	57.28	-	-	-	-
智算推理 GPU 板卡	0.30	-56.46%	0.68	1.16%	0.67	-17.80%	0.82
其中：曦思 N100	0.23	-60.40%	0.58	-13.47%	0.67	-17.80%	0.82
其中：曦思 N260	2.12	-6.25%	2.27	-	-	-	-

（2）训推一体 GPU 板卡平均销售单价变动分析

1) 曦云 C500 板卡

2023 年、2024 年和 2025 年 1-3 月，发行人训推一体 GPU 曦云 C500 板卡平均销售单价分别为 5.69 万元/张、4.78 万元/张和 3.70 万元/张。

2023 年度，曦云 C500 板卡平均销售单价为 5.69 万元/张，高于 2024 年和 2025 年 1-3 月，主要原因为 2023 年曦云 C500 板卡尚未正式量产，产品尚未全面推向市场，部分客户进行小批量采购，当年实现销售的板卡数量较少，且由于产品处于风险量产阶段，单位成本相对较高，产品定价整体较高。

2024 年度，随着曦云 C500 板卡正式量产，批量生产使得产品的单位成本有所降低，同时产品的市场接受度和客户认可度大幅上升，客户数量及客户单笔订单的采购数量大幅上升，平均销售价格有所下降，符合 GPU 产品从市场导入期到销量大幅增长的成长期销售价格通常会有所下降的客观规律。

2025 年 1-3 月，曦云 C500 板卡平均销售单价为 3.70 万元/张，相较 2024 年度下降幅度为 22.69%，主要受大客户相关订单影响，该客户根据其下游集群建设需要，采购量较大，发行人给予该客户价格优惠政策，进而导致该笔订单平均销售单价较低。若不考虑该客户订单的影响，2025 年 1-3 月曦云 C500 板卡平均销售单价为 4.45 万元/张，相较 2024 年度下降幅度较小。

2) 曦云 C550 板卡

2024 年和 2025 年 1-3 月，发行人训推一体曦云 C550 板卡平均销售单价分别为 4.55 万元/张和 5.99 万元/张，2025 年一季度平均销售单价上升 31.57%。曦云 C550 芯片 FP16 稠密算力、INT8 稠密算力等算力指标高于曦云 C500 芯片，且曦云 C550 板卡采用 OAM 模组，在互联带宽、能效比、扩展性等方面优于曦云 C500 板卡，因此曦云 C550 板卡的整体定价高于曦云 C500 板卡。

2024 年度，发行人曦云 C550 板卡销售单价较低，主要系发行人给予部分采购量较大的客户一定的赠卡优惠，拉低了当年曦云 C550 板卡的平均销售单价。2025 年 1-3 月，曦云 C550 板卡销售不涉及赠送等优惠，平均销售价格上升至正常水平。

(3) 训推一体 GPU 服务器平均销售单价变动分析

1) 曦云 C500 GPU 服务器

2024 年和 2025 年 1-3 月，曦云 C500 GPU 服务器销售单价分别为 63.53 万元/台和 35.00 万元/台，2025 年一季度相较 2024 年下降幅度为 44.91%，主要原因为 2024 年发行人销售的 GPU 服务器应用于智算集群的收入占比约为 93%，单台服务器一般配置 8 张曦云 C500 板卡；而 2025 年一季度发行人仅销售 1 台曦云 C500 GPU 服务器，为定制化的工作站，配置 4 张曦云 C500 系列 GPU 板卡，因此其销售价格明显低于 2024 年曦云 C500 GPU 服务器的平均销售单价。

2) 曦云 C550 GPU 服务器

2024 年，发行人向一名客户销售 128 台曦云 C550 GPU 服务器，平均销售单价为 57.28 万元/台。2025 年 1-3 月，发行人向一名客户销售 1 台曦云 C550 服务器，平均销售单价为 69.03 万元/台。2024 年发行人曦云 C550 GPU 服务器销售单价较低，主要原因为根据相关合同条款安排，发行人未确认或冲减了部分合同收入。

(4) 智算推理 GPU 板卡平均销售单价变动分析

1) 曦思 N100 板卡

报告期内，发行人曦思 N100 系列平均销售单价分别为 0.82 万元/张、0.67

万元/张、0.58 万元/张和 0.23 万元/张，呈下降趋势，主要原因为：曦思 N100 为发行人发布的第一款 GPU 产品，面向传统人工智能场景，随着生成式人工智能的兴起，发行人研发、销售的重心逐渐转移至训推一体曦云 C500 产品线，对于曦思 N100 的定价策略有所调整。

2) 曦思 N260 板卡

2024 年和 2025 年 1-3 月，发行人曦思 N260 板卡产品的平均销售单价分别为 2.27 万元/张和 2.12 万元/张，单价基本保持平稳。曦思 N260 芯片与曦云 C500 芯片基于相同制程工艺生产，其算力性能指标与曦云 C500 芯片较为接近，因此其销售定价高于曦思 N100。但由于曦思 N260 不具备卡间高速互联等功能，主要应用于云端推理，因此其定价低于曦云 C500 系列。

2、各类产品整体单价均呈下降趋势的原因及合理性、未来是否会持续下降，与行业周期、趋势、同行业可比公司是否一致

(1) 报告期内发行人销售单价下降主要受大客户优惠措施、与部分客户阶段性合同安排等短期因素影响，2025 年第 2 季度、在手订单平均销售价格已回升

在报告期内，发行人对部分采购量较大的客户给予了较大优惠措施，此类优惠措施对发行人当期销售价格产生了显著影响。此外，发行人报告期内与部分客户存在阶段性合同条款安排，相关的财务处理进一步影响了主要产品的销售单价。上述较大比例赠卡优惠及合同条款安排，是发行人在新产品上市初期为推广产品、拓展客户而采取的商业安排，属于阶段性销售策略，具有合理性。

2025 年第 2 季度，发行人主要产品的销售价格已有所回升。2025 年第 2 季度，发行人主要产品销售情况如下：

单位：万元、张/台、万元/张、万元/台

产品类型	销售收入	销售数量	平均销售单价（不含税）
训推一体 GPU 板卡	39,421.81	8,291	4.75
其中：曦云 C500	24,267.83	5,454	4.45
曦云 C550	15,153.98	2,837	5.34
训推一体 GPU 服务器	18,806.89	288	65.30

产品类型	销售收入	销售数量	平均销售单价（不含税）
其中：曦云 C500	9,430.09	160	58.94
曦云 C550	9,376.80	128	73.26
智算推理 GPU 板卡	1,132.74	565	2.00
其中：曦思 N100	17.70	40	0.44
曦思 N260	1,115.04	525	2.12

注：上表中 2025 年第 2 季度主要产品销售收入、销售数量为未经审计数据

由上表可见，2025 年第 2 季度，训推一体 GPU 板卡平均销售单价为 4.75 万元/张，相较 2025 年第 1 季度 3.89 万元/张大幅上升，与 2024 年 4.69 万元/张的平均售价相比保持稳定且略有上升；2025 年第 2 季度，发行人部分已在报告期内签署的训推一体 GPU 服务器销售合同验收并确认收入，平均销售价格相较 2024 年和 2025 年 1 季度均有所上升；2025 年第 2 季度，智算推理 GPU 板卡平均单位售价为 2.00 万元/张，相较报告期各期均大幅上升，主要原因为当期售价较低的曦思 N100 板卡销量大幅减少，而售价较高的曦思 N260 板卡销量大幅上升所致。根据发行人 2025 年第 2 季度的销售情况，主要产品销售价格相较 2025 年第 1 季度均有大幅回升，与 2024 年度基本持平或略有上升。

此外，从在手订单情况来看，发行人主要产品平均销售价格保持稳定，未出现明显下降。截至 2025 年 8 月 15 日，发行人预计在 2025 年确认收入的 1,000 万元以上在手订单销售价格情况如下表所示：

订单	产品类型	在手订单金额（万元）	产品数量（张/台）	平均单价（万元/张、万元/台，不含税）
订单 1	曦云 C500 板卡	20,389.38	3,200	6.37
订单 2	曦云 C550 服务器	9,353.44	128	73.07
订单 3	曦云 C500 板卡	9,089.13	2,048	4.44
订单 4	曦云 C500 板卡	1,204.41	292	4.12

（2）发行人主要产品价格变动与行业周期、趋势、同行业可比公司总体一致

2023 年，发行人曦云 C500 板卡处于导入期和风险量产期，产品的单位成本较高，且客户数量以及客户单笔采购量较少，整体产品定价相对较高。2024 年，曦云 C500 系列产品经过技术验证和市场推广，销量大幅增加，价格有所下降，

符合芯片产品的生命周期规律。2025 年 1 季度曦云 C500 系列价格下降系新产品上市后发行人为了实现快速切入市场、提升市场渗透率而做出相应的商业安排，与行业周期、趋势相关度较低，具有合理性。目前，发行人曦云 C500 板卡下游需求良好，新客户拓展及存量客户复购情况良好，在市场同类型产品中仍保持较强的市场竞争力，2025 年第 2 季度销售情况以及在手订单价格情况表明曦云 C500 价格与 2024 年全年相比保持稳定。

报告期内，发行人曦思 N100 系列产品价格呈下降趋势，主要原因为生成式 AI 市场需求的大幅增长致使市场对于传统人工智能推理芯片的需求有所下降，发行人将主要研发和销售资源投入更具市场潜力的曦云 C 系列产品线，下调了曦思 N100 板卡的销售定价。曦思 N100 后续迭代产品曦思 N260 面向生成式人工智能场景，可广泛应用于大模型推理、生成式应用等场景，与当前市场需求较为契合，其价格在 2024 年和 2025 年上半年保持相对稳定。

由于 GPU 产品价格受产品定位、性能与配置、供应链、研发成本等多重因素影响，不同的 GPU 厂商影响其产品价格的主要因素有所不同，且公开渠道无法查询到准确的同行业可比公司产品价格信息，因此难以将发行人 GPU 板卡产品的价格变动与同行业可比公司进行直接比较。

综上所述，发行人各类产品整体单价均呈下降趋势主要系因芯片产品迭代、核心产品放量销售、部分大额客户优惠政策以及部分合同条款安排等多种因素导致，具有合理性。发行人核心产品曦云 C 系列板卡仍保持较强的市场竞争力，从期后销售情况以及在手订单情况可见，发行人主要产品销售价格未出现明显下滑。

（三）结合各类明细产品或业务交付产品的时点、形式以及客户签收、验收相关流程，说明各产品或业务的收入确认方法与业务实质及同行业可比公司是否一致，是否符合会计准则相关规定

1、各类明细产品或业务交付产品的时点、形式以及客户签收、验收相关流程、收入确认方法

根据《企业会计准则第 14 号——收入》（财会[2017]22 号）第四条规定，企业应在履行了合同中的履约义务，即在客户取得相关商品控制权时确认收入。

取得相关商品或服务的控制权，是指能够主导该商品的使用或该服务的提供并从中获得几乎全部的经济利益。报告期内，发行人各类业务交付产品的时点、形式以及客户签收、验收相关流程、收入确认方法如下：

（1）GPU 板卡/服务器销售业务

发行人 GPU 板卡/服务器具有产品技术规格相对固定、质量相对稳定的标准化产品特点，对于一般情形下的 GPU 板卡/服务器销售业务，发行人以快递配送或客户自提方式完成交货。客户收到产品后仅执行数量清点、规格核对、外观检查等程序，不存在测试、验证等实质性验收程序，发行人在取得客户出具的签收单后确认产品的控制权转移并确认收入。

（2）销售合同中约定集群验收条款的 GPU 板卡/服务器销售业务

发行人与部分客户签订的销售合同中约定，发行人向客户交付 GPU 板卡/服务器并经客户签收后，还需完成 GPU 板卡或服务器的安装、上架、调试、试运行等程序，完成集群硬件环境检查、单 GPU 节点基础环境测试、计算网络测试、集群稳定性测试等集群验收流程，待 GPU 板卡/服务器稳定运行且集群性能达到验收标准后，由客户出具集群验收单/验收报告，发行人取得验收单/验收报告后确认产品的控制权转移并确认收入。

（3）IP 授权业务

对于 IP 授权业务中的 IP 授权使用费，发行人向客户的标准交付物包括硬件包、软件包以及相关文档等内容，发行人通过 SFTP/FTP（常用的文件传输协议）向客户进行标准交付物的交付，以客户验收合格后发送的验收邮件为依据确认 IP 授权使用费的收入；对于 IP 授权业务中的 IP 技术支持费，发行人在技术支持服务期限内采用直线法摊销确认收入。

（4）技术服务

报告期内，发行人存在少量技术服务业务，收入占比较小。对于在某一时段内履行履约义务的，公司按履约进度确认收入；不属于某一时段内履行履约义务的，在验收或签收时确认收入。

2、发行人各类型业务收入确认方法与同行业可比公司不存在重大差异，符合企业会计准则的规定

发行人收入确认方法与同行业可比公司对比情况如下：

序号	同行业公司名称	主要业务类型	同行业可比公司公开信息披露的收入确认方法	发行人收入确认政策	与同行业可比公司对比情况
1	寒武纪	云端产品线（云端智能芯片及板卡、智能整机）、边缘产品线、IP授权及软件、智能计算集群系统业务	1、销售商品收入：公司销售商品包括云端智能芯片及板卡销售、边缘智能芯片及板卡销售、智能整机销售、智能计算集群系统销售及基础系统软件销售等，属于在某一时刻履行的履约义务。公司根据与客户签订的合同（订单）约定将相关产品交付客户，经由客户签收或验收确认、已收取价款或取得收款权利且相关的经济利益很可能流入时确认收入； 2、技术服务收入：公司为客户提供技术支持、技术咨询、技术开发等服务内容，属于在某一时刻或某一时段履行的履约义务。公司根据与客户签订的合同（订单）约定向客户提供服务，在已收取价款或取得收款权利且相关的经济利益很可能流入、服务一次性验收完成时或在服务期限内按履约进度确认收入； 3、IP授权收入：IP授权销售业务模式主要分为固定费用模式和按计件模式，其中固定费用模式收入确认的具体方法分别如下：固定费用模式合同中约定授权许可客户享受技术的使用权，属于在某一时刻履行的履约义务，在技术交付客户并验收确认、已收取价款或取得收款权利且相关的经济利益很可能流入时确认收入	1、GPU板卡/服务器销售业务：公司已将GPU板卡/服务器交付客户，经客户签收确认，商品所有权上的主要风险和报酬转移给客户后公司确认收入。合同约定需要安装调试的算力集群业务，公司将算力集群交付客户并安装调试完毕，经客户验收确认，商品所有权上的主要风险和报酬转移给客户后公司确认收入。 （2）GPU IP销售业务：公司已将GPU IP交付客户，经客户验收确认，商品所有权上的主要风险和报酬转移给客户，公司确认收入 （3）服务收入	1、GPU板卡/服务器销售业务：寒武纪云端智能芯片及板卡销售、边缘智能芯片及板卡销售、智能整机销售、智能计算集群系统销售及基础系统软件销售均以客户签收或验收确认收入，与发行人不存在差异 2、IP授权业务：发行人报告期内发生的IP授权业务收入与寒武纪固定费用模式IP授权收入确认方式一致 3、技术服务收入：发行人与寒武纪不存在显著不一致
2	海光信息	海光CPU、海光DCU等	1、产品销售收入的原则：本公司以完成产品交付作为产品销售收入确认时点，在完成产品交付后获取验收单（签收单、自提单）时确认收入； 2、技术服务收入：技术服务是公司为客户提供技术支持、技术咨询、技术开发等服务内容。根据合同约定的验收条款，经客户验收确认后确认收入；或根据合	商品所有权上的主要风险和报酬转移给客户，公司确认收入 （3）服务收入 识别为单项履约义务的售后服务收	1、GPU板卡/服务器销售业务：海光信息产品销售收入以验收单（签收单、自提单）确认收入，与发行人不存在差异 2、技术服务收入：与发行人不

序号	同行业公司名称	主要业务类型	同行业可比公司公开信息披露的收入确认方法	发行人收入确认政策	与同行业可比公司对比情况
			同约定的服务期间，在服务完成收到客户验收单时确认收入	入，于服务期限内按直线分摊确认收入。	存在显著不一致
3	龙芯中科	处理器及配套芯片产品、解决方案	1、芯片产品销售收入、硬件模块产品销售收入：根据合同或订单，公司在产品实际交付并取得经客户签收确认的验收凭据时确认收入； 2、技术服务收入：根据合同或订单，公司在服务已完成并取得经客户确认的终验凭据时确认收入	其他服务收入，对于在某一时段内履行履约义务的，公司按履约进度确认收入；不属于某一时段内履行履约义务的，在验收或签收时确认收入。	1、GPU 板卡/服务器销售业务：龙芯中科芯片产品收入以客户签收确认的验收凭据确认，与发行人不存在显著不一致 2、技术服务收入：与发行人不存在显著不一致
4	景嘉微	图形显控领域产品、小型专用化雷达领域产品、芯片领域产品等	公开披露文件中未明确披露各业务类型的具体收入确认方法		/
5	摩尔线程	AI 智算产品、专业图形加速产品、桌面级图形加速产品、智能 SoC 类产品	1、集群产品、一体机产品、板卡产品、芯片产品销售收入：公司以完成产品交付作为销售收入的确认时点，在完成产品交付后并获取经客户签收确认的凭据(签收单/验收单)时确认收入； 2、技术服务收入：公司向客户提供芯片技术方案研究服务，根据合同或订单，在技术服务已完成并取得经客户确认的终验凭据时确认收入		1、GPU 板卡/服务器销售业务：摩尔线程以客户签收确认的凭据（签收单/验收单）确认收入，与发行人不存在显著不一致 2、技术服务收入：与发行人不存在显著不一致

根据上表，发行人收入确认方法与同行业可比公司同类产品服务的收入确认方法不存在重大差异，发行人各业务类型收入确认方法符合《企业会计准则》的规定。

（四）报告期内第四季度收入占比较高、2024 年第三季度收入占比较低的原因及合理性，预计未来收入季节性的趋势

2022-2024 年度，发行人主营业务收入季节性分布情况如下：

单位：万元

季度	2024 年度		2023 年度		2022 年度	
	金额	占比	金额	占比	金额	占比
第一季度	458.16	0.62%	11.58	0.23%	-	-
第二季度	17,560.50	23.67%	407.48	7.93%	-	-
第三季度	4,196.35	5.66%	145.71	2.83%	-	-
第四季度	51,975.87	70.06%	4,576.32	89.01%	42.64	100.00%
合计	74,190.87	100.00%	5,141.09	100.00%	42.64	100.00%

2022-2024 年度，发行人第四季度收入分别为 42.64 万元、4,576.32 万元和 51,975.87 万元，占全年营业收入的比例分别为 100%、89.01%、70.06%，第四季度收入占比较高。其中，发行人 2022 年仅有的一笔销售收入，合同签订时间为 2022 年 12 月，发行人于当月完成发货并确认收入。2023 年和 2024 年，发行人第四季度收入占比较高的主要原因如下：

第一，公司报告期内核心产品研发成功、实现量产、开展客户导入和销售推广的时间对销售收入的季节性影响较大。发行人智算推理系列产品于 2023 年 4 月实现大规模量产、训推一体系列产品于 2024 年 2 月实现大规模量产。如前文“问题 4.关于收入”之“一/（一）结合报告期内下游客户需求变动、主要客户的技术验证及适配过程、量产供货时间等，说明报告期内营业收入大幅增长的主要驱动因素”回复中所述，发行人新产品上市后需经历较长时间的客户导入过程，客户需对发行人产品进行基础性能测试、兼容性测试等全面测试后方能做出批量购买的决策。此外，不同客户对发行人产品的测试适配周期也不尽相同，例如部分客户在其项目初期对自身 AI 应用的需求处于探索阶段，可能会导致在产品测试过程中测试需求和测试内容发生变更，从而拉长整体的测试适配周期。报告期内，发行人智算推理系列、训推一体系列产品均于当年上半年量产，至下半年或第四季度与主要客户签署大额销售合同并大规模出货，符合 GPU 产品上市后需经历技术验证、客户导入过程方能大规模销售的客观规律，是发行人 2023 年、2024 年第四季度收入占比较高的主要原因。

第二，发行人报告期内总体收入规模较小，客户集中度相对较高，销售收入的季节性分布受大额订单影响较大。2023 年 6 月，发行人与公司 1 签订《知识产权许可协议》，于 2023 年 11 月完成交付并确认收入 2,200 万元，该合同金额

较大，对发行人 2023 年收入的季节性分布影响较大。2024 年 8 月，发行人与汇天网络签署 1024 卡集群销售合同，于 2024 年 11 月取得集群验收报告并确认收入 7,332.18 万元，大幅增加了发行人 2024 年第四季度的销售收入。

第三，部分终端智算集群项目的建设时间影响了发行人收入的季节性分布。例如，发行人于 2024 年底与新华三、瑞芯智能签署 GPU 板卡销售合同，分别应用于项目 A（二期）以及项目 B，对项目建设的进度要求较高。为满足下游集群建设的时间要求，发行人于 2024 年第四季度向新华三、瑞芯智能的发货并确认收入。其中，项目 B 项目时间要求较高，下游客户要求发行人于 2024 年底前完成交货；项目 A（二期）项目时间安排较为紧凑，相关的服务器厂商需进行备货以保障及时交付，发行人按照与新华三订单中的交货期约定，向其完成发货并确认收入。

发行人同行业可比公司收入的季节性分布情况如下：

序号	公司名称	主营业务	第四季度收入占比	关于收入季节性的相关披露情况
1	寒武纪	应用于各类云服务器、边缘计算设备、终端设备中人工智能核心芯片的研发、设计和销售	2022 年、2023 年和 2024 年，寒武纪第四季度收入占比分别为 64.31%、79.46%和 84.22%	寒武纪在其 2025 年向特定对象发行股票审核问询函回复中披露：“2024 年第四季度收入占比较高，主要系公司云端产品线相关产品密集出货所致。公司凭借人工智能芯片产品的核心优势，持续深化与大模型、互联网等前沿领域头部企业的技术合作，云端产品线主营产品经过在客户处测试及适配等流程，并于 2024 年第四季度开始向客户批量出货，实现了大规模商业化落地，2024 年第四季度收入占比较高具有合理性”
2	海光信息	研发、设计和销售应用于服务器、工作站等计算、存储设备中的高端处理器	2022 年、2023 年和 2024 年，海光信息第四季度收入占比分别为 25.47% 、 34.42% 和 33.02%	海光信息在其招股说明书中披露：“2020 年开始，公司营业收入呈现显著的季节性特征，下半年尤其第四季度销售占比较高，主要因为公司主要终端客户，例如金融、电信等行业客户，上半年处于预算审批阶段，以及公司直接客户公司 A、浪潮、同方、联想等服务器厂商，上半年处于投标阶段，采购金额较少，在采购计划和预算审批完成后，客户计划中的项目陆续实施，采购金额逐渐增加，因此客户通常集中在下半年尤其是第四季度完成采购和产品验收”

序号	公司名称	主营业务	第四季度收入占比	关于收入季节性的相关披露情况
3	龙芯中科	处理器及配套芯片的研制、销售及	2022 年、2023 年和 2024 年，龙芯中科第四季度收入占比分别为 34.51% 、 22.03% 和 38.96%	龙芯中科在其招股说明书中披露：“公司销售产品主要客户通常在每年的第一季度制定全年的采购计划并确定预算额，一个自然年度的上半年处于预算编制与审批阶段，新增需求较少，在预算编制结束后，客户计划中的项目陆续实施，需求增加，因此客户通常集中在下半年特别是第四季度完成产品验收。受此影响，公司产品销售收入通常具有上半年较低、下半年尤其是第四季度较高的特点，主要与下游终端产品市场需求有关”
4	景嘉微	高可靠电子产品的研发、生产和销售	招股说明书披露时间较早，2022 年、2023 年和 2024 年，景嘉微第四季度收入占比分别为 36.80%、34.22%和 5.42%	未披露
5	摩尔线程	GPU 及相关产品的研发、设计和销售	2022 年、2023 年和 2024 年，摩尔线程第四季度收入占比分别为 68.87% 、 32.24% 和 35.80%	摩尔线程在其招股说明书中披露：“2022 年四季度销售收入占比较高，主要系公司 2022 年下半年推出第一款量产芯片产品后，实现规模化对外销售。2023 年公司收入季节性不明显。2024 年四季度收入主要来自客户 J，该公司为业内知名经销商，四季度采购较多公司产品，主要由于 2024 年下半年双方开始首次业务合作”

如上表所示，发行人第四季度收入占比高于同行业公司，但同行业公司均普遍呈现出上半年收入占比较低、下半年收入占比较高的特点。此外，寒武纪 2022 年、2023 年和 2024 年第四季度收入占比分别为 64.31%、79.46%和 84.22%，也呈现出较为明显的第四季度收入占比高的特点。

发行人曦云 C500 系列产品于 2023 年下半年开始小批量销售，于 2023 年底与上海算丰签署 160 台曦云 C500 GPU 服务器的大额销售合同项目 E（一期），该合同于 2024 年 5 月执行完毕并确认收入，使得 2024 年第二季度收入占比相对较高。2024 年第三季度，其他主要客户尚处于对发行人产品测试、验证的过程中，尚未签署大规模采购订单或大规模采购订单尚未履行完毕确认收入，因此发行人 2024 年第三季度收入金额较小。

发行人训推一体系列核心产品已在 2024 年第四季度大规模商业化应用，未来随着大模型等人工智能应用对人工智能算力需求的持续旺盛，公司产品预计将

持续稳定地商业化出货。公司 2022-2024 年销售收入集中在第四季度主要受核心产品大规模放量销售时间以及部分大额合同的影响。2025 年一季度，公司实现收入 32,041.53 万元，相较 2024 年一季度同比大幅增长；2025 年第二季度，公司未经审计的营业收入金额约为 5.94 亿元，相较 2024 年第二季度也有大幅增加。未来随着公司产品矩阵逐渐丰富以及持续稳定地大规模出货，公司销售收入的季节性特征将逐渐减弱。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、申报会计师执行了如下核查程序：

1、查询人工智能芯片行业近年来发展趋势、市场需求变化情况以及国产替代相关支持政策，了解国产 GPU 行业近年来发展情况以及下游客户需求变动的具体情况；获取发行人收入成本明细、销售合同明细，了解发行人主要客户下达小批量采购订单时间、大批量采购订单时间；访谈发行人销售部门负责人及主要客户，了解主要客户的技术验证及适配过程、量产供货时间等。

2、获取发行人销售收入明细表，分析发行人各主要产品及主要客户的销售单价，分析价格波动的原因；获取发行人 2025 年第 2 季度销售收入明细，了解期后发行人主要产品价格情况；获取发行人在手订单清单，了解在手订单主要产品销售价格情况；公开渠道查询人工智能芯片市场价格的趋势及影响因素、同行业可比公司主要产品价格情况，向发行人销售部门负责人了解主要产品价格变动与行业周期、趋势、同行业可比公司是否一致。

3、向发行人了解各类业务签收或验收的流程，查阅同行业可比公司公开披露的收入确认方法，分析发行人收入确认方法与同行业可比公司是否一致、是否符合《企业会计准则》的规定。

4、获取发行人收入成本明细表并对发行人销售收入执行细节测试、截止性测试、函证等核查程序，获取合同、签收单、发票、物流、回款等单据，结合销售合同条款、收入确认模式及相关单据，检查发行人收入确认时点的准确性。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、人工智能芯片行业需求大幅增长、国产替代加速等有利因素赋予发行人广阔的市场空间，发行人紧抓下游需求扩张的机遇推出核心产品训推一体系列芯片，在较短时间内完成主要客户的技术验证和适配，快速商业化落地并实现大规模出货，是发行人报告期内营业收入增长的主要驱动因素。

2、发行人报告期内价格下降主要系大额优惠措施、阶段性合同条款安排等短期因素导致，期后销售价格以及在手订单价格已有所回升，发行人主要产品价格的变动与行业周期、趋势和同行业可比公司不存在较大差异。

3、发行人各产品或业务的收入确认方法与业务实质及同行业可比公司一致，收入确认方法符合《企业会计准则》相关规定。

4、发行人第四季度收入占比较高，2024 年第三季度收入占比较低，主要系发行人核心产品客户导入及验证周期、部分大额合同及终端项目建设进度要求等因素导致，具有合理性，发行人收入确认时点准确。

5.关于销售模式

根据申报材料：（1）报告期各期，发行人前五大客户销售金额占比为 100%、91.58%、71.09%和 88.35%，集中度相对较高；（2）报告期各期，经销收入为 0、2,252.23 万元、27,758.94 万元和 17,849.70 万元，直销收入为 42.64 万元、2,888.86 万元、46,431.94 万元和 14,191.82 万元，同行业可比公司海光信息、寒武纪、龙芯中科以直销为主；（3）超讯通信为发行人 2025 年 1 季度第一大客户，双方设立了合资公司。

请发行人披露：（1）结合行业特点、产品特性、下游应用领域等，说明发行人客户高度集中、且经销收入占比较高的原因及合理性，是否符合行业惯例；结合主要客户的毛利贡献情况，说明发行人对主要客户是否存在依赖；（2）发行人与超讯通信的具体合作模式及可持续性，各期合作项目、金额、终端客户名称、最终项目建设及投入使用情况，尚未回款的原因及合理性；双方设立合资公司的背景、未来发展方向，与产品购销交易是否独立开展；各经销商是否存在为发行人囤货的情形。

请保荐机构、申报会计师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）结合行业特点、产品特性、下游应用领域等，说明发行人客户高度集中、且经销收入占比较高的原因及合理性，是否符合行业惯例；结合主要客户的毛利贡献情况，说明发行人对主要客户是否存在依赖

1、结合行业特点、产品特性、下游应用领域等，说明发行人客户高度集中、且经销收入占比较高的原因及合理性，是否符合行业惯例

（1）行业特点、产品特性、下游应用领域

发行人致力于自主研发全栈高性能 GPU 芯片及计算平台，为人工智能计算、通用计算、图形渲染等前沿领域提供高能效、高通用性的算力支撑。发行人所处的集成电路设计产业处于集成电路上游位置，属于集成电路产业的核心环节之一，是国家各项产业政策和发展战略规划的重点区域。随着集成电路行业的迅速发展，

在摩尔定律的推动下，集成电路产品的加工面积成倍缩小，复杂程度与日俱增，呈现出高壁垒、高附加值和高投入的特征。同时，公司全系列产品采用统一的GPU计算架构，能够高效适应云端快速迭代的复杂算法和大量非传统模型结构，支持更为广泛的人工智能应用场景。报告期内，发行人的直接下游主要包括服务器厂商、集成商或智算中心建设方，下游最终使用方包括教科研、金融、交通、能源、医疗健康、大文娱等领域。

（2）发行人客户高度集中的原因及合理性

报告期内，发行人前五大客户的主营业务收入占比分别为 100.00%、91.58%、71.09%和 88.35%，客户集中度相对较高。2022 年度及 2023 年度，公司产品尚处于业务拓展初期，主要产品尚未实现量产，公司将有限的产量向能够实现商业化落地的主要客户进行销售，因此主要客户收入占比较高；2024 年度及 2025 年 1-3 月，随着发行人训推一体芯片曦云 C500 正式量产，该产品陆续获得客户的认可和采购，并在多个国家人工智能公共算力平台、运营商智算平台和商业化智算中心批量应用，由于上述项目单项采购规模较高、投入金额较大，导致相关期间内向主要客户销售集中度较高。

经检索，同行业可比公司各期前五大客户收入合计所占比例如下表所示：

单位：万元

公司名称	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
摩尔线程	未披露	98.16%	97.45%	89.86%
海光信息	未披露	98.16%	99.14%	97.36%
寒武纪	未披露	94.63%	92.36%	84.94%
龙芯中科	未披露	39.17%	37.26%	32.96%
景嘉微	未披露	58.54%	76.82%	65.32%
平均值	-	77.73%	80.61%	74.09%
发行人	88.35%	71.09%	91.58%	100.00%

由上表可见，报告期内发行人前五大客户收入合计所占比例与同行业可比公司不存在重大差异，且与产品结构接近的摩尔线程、海光信息、寒武纪等接近，符合行业惯例。

（3）经销收入占比较高的原因及合理性

报告期内，发行人经销收入占主营业务收入的比例分别为 0%、43.81%、37.42% 和 55.71%，经销收入占比较高，主要原因系：

1) 发行人处于业务拓展期，通过经销模式快速打开市场

发行人成立于 2020 年，自设立起专注于 GPU 业务，2020 年-2022 年期间形成相关技术的原始积累，2023 年以来逐渐将研发优势转换为产品优势，陆续于 2023 年、2024 年量产了曦思 N 系列 GPU 产品、曦云 C 系列 GPU 产品，产品性能方面虽然达到了行业先进水平，但仍处于产品与品牌的拓展期，在品牌影响力、经营规模、销售人员数量及市场服务能力方面存在一定劣势。通过采用经销模式，公司可以充分发挥经销商区位资源、行业资源的优势，扩大终端客户覆盖范围并提高服务能力，是公司结合自身经营需求自主选择的销售策略，符合公司实际情况。

2) 发行人下游应用领域广泛，通过经销模式可快速触达终端业务需求

发行人为通用型计算架构（GPGPU）芯片设计企业，专注于打造独立自主的 GPU 芯片及 AI 计算平台生态，主要业务为研发、设计和销售应用于人工智能训练及推理、通用计算和图形渲染等领域的 GPU 产品并提供配套软件。发行人的直接下游为服务器厂商、集成商或智算中心建设方，下游最终使用方包括教科研、金融、交通、能源、医疗健康、大文娱等领域。因此，发行人产品具有功能通用性高、下游应用领域广泛等特点。报告期内，发行人主要经销商均具有较高市场地位或行业经验，通过经销商客户渠道可快速触达广阔的下​​游应用领域终端客户或最终使用方。

经检索，同行业可比公司中，根据公开披露信息存在经销收入的公司包括摩尔线程、海光信息、寒武纪和龙芯中科，其在报告期内的经销收入及占比情况具体如下：

单位：万元

公司名称	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
	经销收入	占比	经销收入	占比	经销收入	占比	经销收入	占比
摩尔线程	未披露	未披露	25,518.13	59.03%	4,710.68	38.78%	4,518.05	98.56%

公司名称	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
	经销收入	占比	经销收入	占比	经销收入	占比	经销收入	占比
海光信息	未披露	未披露	139,328.51	15.21%	43,736.12	7.27%	52,704.24	10.28%
寒武纪	未披露	未披露	1,463.80	1.25%	3,066.52	4.33%	6,765.57	9.36%
龙芯中科	未披露	未披露	317.34	0.63%	223.23	0.44%	365.73	0.50%
平均值	-	-	41,656.94	19.03%	12,934.14	12.71%	16,088.40	29.68%
发行人	17,849.70	55.71%	27,758.94	37.42%	2,252.23	43.81%	-	-

注：景嘉微曾于 2020 年 3 月披露其全资子公司长沙景美集成电路设计有限公司许可北京神州数码有限公司为“中国大陆地区经销商”，但 2022 年以来各年定期报告中披露营业收入均为直接销售。

由上表可见，发行人报告期内经销收入占比高于同行业可比公司平均值，主要原因系：海光信息、龙芯中科主要产品为 CPU 处理器，由于该类产品属于服务器、工控机等硬件产品的必要配置，主要面向客户为整机厂商与 ODM 厂商，目前国内服务器整机厂商及 ODM 厂商集中度较高、规模效应明显，因此其采用直销为主的业务模式，能够更好地集中销售资源至行业内主要客户；寒武纪的主要产品属于专用型计算架构，2022-2024 年第一大客户的收入占比分别为 60.81%、65.88%和 79.15%，需集中销售、研发资源服务部分重点客户或项目，因此其采取直销为主的业务模式。发行人报告期内经销收入占比与摩尔线程较为接近，其对于经销模式的解释为：“该销售模式（经销）可以使公司更好地专注于产品的设计研发环节，提高公司研发、运营的效率。经销模式符合 Fabless 模式下芯片设计行业的经营模式惯例。”

因此，报告期内发行人经销收入占比高于同行业可比公司具有合理性，与同行业可比公司摩尔线程较为接近，符合行业惯例。

2、结合主要客户的毛利贡献情况，说明发行人对主要客户是否存在依赖

报告期内，发行人前五大客户的毛利额占主营业务毛利的比例分别为 100.00%、96.36%、67.86%和 87.25%。2022 年度及 2023 年度，公司产品尚处于业务拓展初期，主要产品尚未实现量产，整体收入规模较小，其中 2022 年仅实现对上海东焱信息技术有限公司的少量销售，2023 年度对公司 1 有限公司销售主要内容为 IP 授权，毛利率较高，因此报告期内对上述两家客户的毛利占比超过 50%。除上述情况外，报告期内，发行人对单一客户的毛利占比不存在超过

50%的情形。

因此，发行人对主要客户不存在重大依赖。

（二）发行人与超讯通信的具体合作模式及可持续性，各期合作项目、金额、终端客户名称、最终项目建设及投入使用情况，尚未回款的原因及合理性；双方设立合资公司的背景、未来发展方向，与产品购销交易是否独立开展；各经销商是否存在为发行人囤货的情形

1、发行人与超讯通信的具体合作模式及可持续性，各期合作项目、金额、终端客户名称、最终项目建设及投入使用情况，尚未回款的原因及合理性

超讯通信为发行人的全国行业总代理，即一级经销商。2023 年 11 月，发行人与超讯通信签署《全国行业总代理商合作协议》，由超讯通信负责发行人产品和服务在中国移动、中国电信、中国联通、中国能建、中国交通以及“东数西算”算力节点客户处的销售工作。截至本回复出具之日，《全国行业总代理商合作协议》处于有效期内，双方合作关系良好，未来合作具有可持续性。

报告期各期，发行人向超讯通信销售额在 100 万元以上的订单相关信息如下表所示：

单位：万元、张、万元/张							
期间	订单号	合作项目	合同金额	终端客户名称	终端客户用途	最终项目建设及投入使用情况	未回款的原因及合理性
2024 年	880000036	曦云 C500 服务器及配套产品采购	384.59	超讯通信	客户测试	已投入使用	已回款
	880000047	曦云 C500 板卡、服务器及配套产品采购	242.51	终端客户 10（8 张曦云 C500）	客户测试	已投入使用	已回款
				超讯通信（其他板卡及服务器等）	算力租赁和客户测试	已投入使用	已回款
2025 年 1-3 月	8800000076	曦云 C500 板卡及配套产品采购	28,798.38	终端客户 3	算力服务租赁，主要客户为某运营商	已完成服务器生产，待根据终端需求交付	已部分回款，集群建设进度较原计划有所延缓，导致投产的服

期间	订单号	合作项目	合同金额	终端客户名称	终端客户用途	最终项目建设及投入使用情况	未回款的原因及合理性
							务器尚未交付、收款

由上表可见，报告期内发行人与超讯通信合作情况良好，其中 2024 年主要合作项目均已回款，2025 年 880000076 订单尚未回款主要系终端集群项目交付进度有所延迟所致，发行人正在积极与超讯通信沟通并跟进终端项目建设情况，相关回款不存在重大风险。

2、双方设立合资公司的背景、未来发展方向，与产品购销交易是否独立开展

2024 年 6 月，发行人与超讯通信参股设立中能建（北京）智慧科技有限公司（以下简称“中能建智慧科技”），注册资本 3,000 万元人民币，由中国电力工程顾问集团有限公司持股 34%并控股，超讯通信持股 10%，发行人持股 8%。该公司由央企上市公司中国能建（601868.SH）全资子公司中国电力工程顾问集团有限公司牵头，会同发行人、超讯通信等五家企业共同设立，该公司由各股东充分发挥各自优势，致力于为中国能建中电工程分布于“东数西算”八大枢纽节点的算力中心提供算力计算及运营服务，共同打造“算力服务-算力设备-算力技术”生态圈，未来中能建智慧科技将主要为客户提供算力计算及运营服务。

2025 年 2 月，发行人与超讯通信设立合资公司四川讯曦智能科技有限公司（以下简称“四川讯曦”），注册资本 6,000 万元人民币，其中超讯通信持股 56%并控股，发行人持股 18%。该公司为超讯通信布局算力服务器业务的控股子公司，超讯通信自 2019 年起布局算力业务，发展重心逐步向算力业务转型，包括算力服务器销售及算力总集成业务。2024 年以来，超讯通信积极部署自有品牌“元醒”服务器，实现了“元醒”服务器的批量交付。为进一步开展该项业务，超讯通信于 2025 年会同沐曦股份等三家合作伙伴共同投资设立四川讯曦，致力于从事服务器整机生产、销售等业务，服务国家新质生产力建设，未来四川讯曦拟主要从事算力服务器整机生产、销售等业务。

综上所述，上述设立合资公司事项为双方发挥各自优势，进一步加强合作的举措，具有合理性。报告期内，发行人与中能建智慧科技、四川讯曦未发生交易。

根据超讯通信及发行人的确认，相关合作与产品购销交易独立开展，不存在潜在联系。

3、各经销商是否存在为发行人囤货的情形

针对报告期内任意一期收入金额在 100 万元以上的经销商，根据其提供的进销存明细表等材料，报告期内发行人产品向终端客户销售、库存情况汇总如下表所示：

产品类别	报告期内向发行人采购		报告期内对外销售情况		期末库存情况	
	数量	金额	数量	金额	数量	金额
训推一体 GPU 板卡	11,417	45,443.14	10,666	42,018.07	751	3,425.08
智算推理 GPU 板卡	2,160	919.14	1,587	525.02	573	394.12
合计	13,577	46,362.29	12,253	42,543.09	1,324	3,819.20
占报告期内采购比例	100.00%	100.00%	90.25%	91.76%	9.75%	8.24%

注：上表未统计单位价值较低的桥接板等配件。

根据发行人经销商管理相关制度及与经销商签署的协议，报告期内发行人与经销商的合作模式均为买断式经销，不存在代销模式，即经销商采购发行人的产品除质量问题可退回外，其他情况下不具有退货权，由经销商自行负责销售与管理。由于 GPU 产品单价相对较高，大规模采购资金占用成本较大，报告期内公司主要经销商根据其下游客户需求，采取“以销定采”的备货模式。由上表可见，相关经销商报告期内累计销售发行人产品数量及金额占合计采购发行人产品数量及金额的比例分别为 90.25%、91.76%，期末库存数量及金额的比例均小于 10%，实现终端销售比例较高，不存在为发行人囤货的情形。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、申报会计师执行了如下核查程序：

1、查阅发行人所处行业的研究报告及发行人产品介绍；访谈发行人销售负责人，了解发行人客户集中、经销收入占比较高的原因；查阅发行人同行业可比公司公开披露材料；获取并检查了发行人收入成本明细表。

2、查阅发行人与超讯通信签署的《全国行业总代理商合作协议》；访谈超

讯通信，确认各期主要合作项目的相关情况及双方设立合资公司的相关情况；获取并核查报告期内主要经销商的进销存明细表。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、结合发行人行业特点、产品特性、下游应用领域等，发行人客户集中、经销收入占比较高具有合理性，与同行业可比公司相比不存在重大异常情形；发行人对主要客户不存在重大依赖。

2、超讯通信为发行人一级经销商，双方合作情况良好，具有可持续性，各期主要项目合作情况及最终项目建设及投入使用情况良好，部分项目尚未回款主要受终端项目建设进度影响，具有合理原因；设立合资公司为双方发挥各自优势，进一步加强合作的举措，具有合理性，与产品购销交易独立开展，不存在潜在联系；发行人经销商报告期内采购发行人产品的终端销售实现情况良好，不存在为发行人囤货的情形。

6.关于应收账款

根据申报材料：（1）报告期各期，发行人应收账款账面价值分别为 0、3,709.94 万元、47,944.33 万元和 61,506.52 万元，占当期营业收入比例为 0、70.68%、65.40% 和 199.29%；（2）报告期内，发行人对主要经销商华研集团、武珞智慧、瑞芯智能、超讯通信，直销客户新华三、公司 2 等应收金额较大，其中，发行人针对主要经销商均采用统一的信用政策，部分订单约定“背靠背”付款条件；（3）截至 2025 年 3 月末，主要经销商客户超讯通信、武珞智慧、华研集团逾期应收账款金额为 24,681.17 万元，占全部经销商期末逾期金额的 98.05%，预计于 2025 年 6 月底前完成大部分回款；（4）截至 2025 年 5 月 31 日，2023 年至 2025 年 3 月末应收账款回款比例分别为 92.04%、21.96%和 13.13%，各期末坏账准备为 0、37.47 万元、650.27 万元和 2,349.40 万元。

请发行人披露：（1）报告期内应收账款大幅增长的原因及合理性、与营业收入的匹配性，主要客户的实际回款周期、与信用政策匹配性、不同客户是否存在较大差异，是否存在放宽信用政策刺激销售的情形；（2）区分直销/经销客户，列示各期对主要客户的收入金额、应收账款余额及占对应收入的比重、账龄结构、最新期后回款金额及回款比例、逾期金额、逾期比例、逾期原因；（3）报告期各期存在“背靠背”结算条款的应收账款金额及占比、约定“背靠背”结算的原因、相关款项回收和逾期情况、是否存在回款风险；（4）结合应收账款账龄结构、逾期比例较高、期后回款比例较低、客户资信情况以及最新回款等情况，说明报告期内坏账准备计提的充分性。

请保荐机构、申报会计师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

(一) 报告期内应收账款大幅增长的原因及合理性、与营业收入的匹配性，主要客户的实际回款周期、与信用政策匹配性、不同客户是否存在较大差异，是否存在放宽信用政策刺激销售的情形

1、报告期内应收账款大幅增长的原因及合理性

报告期各期末，公司应收账款账面价值分别为 0、3,709.94 万元、47,944.33 万元和 61,506.52 万元。报告期各期末，公司应收账款账面价值大幅增长，主要原因如下：

第一，报告期内，公司核心 GPU 产品量产并实现销售，营业收入增长幅度较大，应收账款金额随营业收入规模的增加而增加，应收账款总体变动趋势与营业收入变动趋势相符。此外，由于发行人第四季度收入占比较大，截至报告期各期末，部分应收账款仍处于信用期内，因此应收账款余额较大。

第二，报告期内，公司与重要客户新华三和瑞芯智能签订了含“背靠背”结算条款的合同，即客户向公司付款前应收到下游用户的项目回款，使公司应收账款回款结算周期延长。2024 年末与 2025 年 3 月末，新华三与瑞芯智能尚未达到“背靠背”回款条件，应收账款余额较大。

第三，报告期内，部分客户受付款流程审批、终端建设进度等因素的影响，存在付款周期超出信用期的情况，导致期末应收账款余额较大，具体情况参见本问询回复报告之“6.关于应收账款/一/（二）区分直销/经销客户，列示各期对主要客户的.....逾期金额、逾期比例、逾期原因”。

2、应收账款大幅增长与营业收入相匹配

2023 年末、2024 年末、2025 年 3 月末，发行人应收账款期末余额占当期营业收入的比例分别为 70.68%、65.40%和 199.29%。2023 年末、2024 年末比例保持相对稳定，应收账款余额与营业收入增长基本匹配。2025 年 3 月末，应收账款余额占营业收入的比例显著升高。一方面，瑞芯智能、武珞智慧与上海华研未形成新的营业收入，前期应收款项尚未收回，另一方面，2025 年 1-3 月收入未进行年化，上述原因导致 2025 年 3 月末应收账款账面余额占当期营业收入的比例显著升高，若进行年化计算，报告期末发行人应收账款余额占当期营业收入的比

例为 66.43%。总体而言，报告期内公司应收账款与当期营业收入较为匹配。

3、报告期内主要客户的实际回款周期、与信用政策匹配性分析

(1) 报告期内主要客户的实际回款周期计算

报告期各期末，主要客户当期实际回款周期列示如下表所示：

单位：万元

年份	序号	客户名称	营业收入	应收账款 平均余额	应收账款 周转率	当期实际 回款周期 (天)
2025 年 1-3 月	1	超讯通信股份有限公司	12,800.17	9,759.99	5.25	69.58
	2	上海源庐加佳信息科技有限公司	9,095.93	-	/	-
	3-1	新华三信息技术有限公司	2,918.69	16,200.91	0.72	506.51
	3-2	新华三信息技术有限公司 (不含背靠背订单)	2,918.69	3,529.81	3.31	110.36
	4	北京创世云科技股份有限公司	2,359.28	366.88	25.72	14.19
	5	华科智谷（上海）科技发展有限公司	1,136.14	141.92	32.02	11.40
2024 年	1-1	新华三信息技术有限公司	15,095.15	7,059.49	2.14	170.70
	1-2	新华三信息技术有限公司 (不含背靠背订单)	3,881.79	723.94	5.36	68.07
	2	苏州瑞芯智能科技有限公司	13,046.02	6,371.00	2.05	178.25
	3	上海算丰信息有限公司	9,403.19	121.85	77.17	4.73
	4	汇天网络科技有限公司	7,332.18	539.74	13.58	26.87
	5	浙江武珞智慧城市技术有限公司	6,532.66	2,546.22	2.57	142.27
2023 年	1	公司 1	2,216.67	795.00	2.79	130.91
	2	上海燃云科技有限公司	1,635.26	767.12	2.13	171.23
	3	上海闪马智能科技有限公司	600.00	-	/	-
	4	中科远见（重庆）科技有限公司	139.51	-	/	-
	5	湖南沐仕杰半导体科技有限公司	116.58	-	/	-
2022 年	1	上海东焱信息技术有限公司	42.64	-	/	-

注 1：当年实际回款周期（天数）=365/（期间收入/应收账款平均余额），2025 年 1-3 月进行年化处理

注 2：若客户的应收款项形成于当年年末，并于次年年初已完成回款，虽然回款周期较短，但因计算公式中“应收账款平均余额=期初期末应收余额平均值”，导致实际回款周期计算

结果偏长，本表中上海燃云科技有限公司、公司 1 属于此种情形，具体分析参见本问询回复报告之“6.关于应收账款/一/（一）/3/（2）与信用政策匹配性分析”。

(2) 与信用政策匹配性分析

报告期内，发行人主要客户整体账龄较短，与信用政策基本相符。报告期内，个别主要客户存在应收款项逾期的情况，主要逾期客户具有以下特点：第一，存在逾期的客户报告期内整体回款情况较好，仅有部分订单的应收款项存在逾期。第二，截至本回复出具日，多个存在逾期的客户应收账款已完成回款，未回款项已有明确的回款计划，不存在重大的回收风险。

报告期内，公司主要客户的实际回收周期与信用政策匹配性分析如下：

年份	客户名称	实际回款周期（天）	信用政策	回款情况分析	与信用政策匹配性分析
2025 年 1-3 月	超讯通信股份有限公司	69.58	报告期内，发行人对超讯通信主要执行预收货款的结算方式，仅少数订单给予发货后 90 天账期。具体而言： 1、2023 年签署的经销协议： 乙方须在经销协议生效后 5 个工作日内支付订金人民币 2,000 万元，后续根据实际情况下达具体订单。 2、2024 年：订单抵扣前期 2,000 万元预付款，少数订单约定 90 天账期（合计订单金额 101.16 万元） 3、2025 年：订单 880000076，订单生效后 2 个工作日内支付订单货款总额的 10%作为首笔货款，剩余 90%尾款于 2025 年 1 月 5 日前付全额货款给甲方。	1、超讯通信于 2023 年末根据经销协议约定向发行人支付预付款 2,000 万元，已用于多笔订单抵扣； 2、超讯通信按期支付报告期内信用期为 90 天的相关订单款项； 3、截至 2025 年 3 月末，超讯通信应收账款余额为两笔逾期订单款项构成，合计应收账款余额为 16,106.73 万元。截至 2025 年 7 月 31 日，上述逾期订单超讯通信已回款 1,736 万元。	超讯通信总体回款周期为 69.58 天，超出信用期约定。具体而言，发行人对超讯通信主要执行预收货款的信用政策，部分订单执行较好，但两笔订单未按期支付预收货款，具体逾期原因参见本问询回复报告之“6.关于应收账款/一/（二）区分直销/经销客户，列示.....逾期金额、逾期比例、逾期原因”。
	上海源庐加佳信息科技有限公司	-	报告期内，发行人对源庐加佳主要执行预收货款的结算方式	报告期内，源庐加佳按照订单约定进行预付货款，未形成期末应收账款，不存在实际回款周期	回款情况较好，与信用政策相匹配
	新华三信息技术有限公司	506.51（含“背靠背”）	报告期内，发行人与新华三共两种结算方式： 1、订单 880500105 约定“背靠背”结算	报告期内，新华三按照订单约定支付货款，其中： 1、背靠背订单，需待新华三与	回款情况较好，与信用政策相匹配

年份	客户名称	实际回款周期（天）	信用政策	回款情况分析	与信用政策匹配性分析
		110.36（不含“背靠背”）	方式 2、其他订单：《框架采购协议》：款项应当在收到产品和/或服务起一百二十（120）天支付，除非双方在订单和/或其他协议中另有约定。	最终客户结算款项后，再与公司结算款项，2025年3月末，尚未达到回款条件。截至本回复出具日，新华三“背靠背”结算的应收账款已回款 10,260 万元。 2、除了约定“背靠背”的订单，其他订单实际回款周期为 110.36 天，短于信用期限。	
	北京创世云科技股份有限公司	14.19	报告期内，发行人对创世云主要为预收货款，仅两个订单给予一定信用期。 具体而言： 1、2024 年，预付款 2,000 万元，后续订单抵扣前期预付款 2、抵扣完毕后订单采取款到发货的方式 3、仅两个订单给予一定信用期： （1）订单 SO880000090，订单生效后支付 30%货款，即 180.54 万元，剩余 70%尾款（421.26 万元）于发货后 90 天内付清 （2）订单 SO880000115，订单金额 312.50 万元，货物交付后 10 天内付清全款	报告期内，创世云均按照订单约定预付货款，2024 年期末无应收账款余额，2025 年 3 月末的应收账款余额即为两笔给予一定信用期的订单，#SO880000115 已经于信用期内支付货款， #SO880000090 的 421.26 万元尚在信用期内，未出现逾期情况。	回款情况较好，周期较短，与信用政策相匹配
	华科智谷（上海）科技发展有限公司	11.40	报告期内，发行人对华科智谷主要为预收货款，具体而言： 订单 880000117 约定，2025 年 3 月 23 日前支付 1,000 万元预付款，发行人为华科智谷开具 2,000 万元发票后，华科智谷立即向甲方支付 1000 万元货款，剩余货款于交付货物后，90 天内支付完成	报告期内，华科智谷已按订单的约定预付 1,000 万元货款，应收款项均在信用期内。报告期后，华科智谷已按订单约定另支付 1,000 万元预付款。2025 年 3 月末，发行人对华科智谷应收账款余额 283.4 万元，截至本回复出具日，华科智谷已完成回款。	回款情况较好，周期较短，与信用政策相匹配
2024 年	新华三信息技术有	170.70（含	参见上文分析	参见上文分析	回款情况较好，与信用政策

年份	客户名称	实际回款周期（天）	信用政策	回款情况分析	与信用政策匹配性分析
	限公司	“背靠背”） 68.07（不含“背靠背”）			相匹配
	苏州瑞芯智能科技有限公司	178.25	1、2024 年签署的经销协议： 本协议生效后 30 天内乙方向甲方支付首笔订单预付款 2,000 万元，在后续订单应支付的货款中分批抵扣，预付款抵扣完，乙方根据项目需求向甲方下单，采用全款提货的方式。 2、订单 880000072：前次已预付的 2,000 万元可等额抵扣本订单金额，抵扣后剩余货款采用和下游客户同比例“背靠背”付款方式，即瑞芯智能收到下游客户比例金额的款项后，10 天内同比例支付给公司	报告期内，瑞芯智能已按合同约定预付 2,000 万元货款。截至本回复出具日，尚未达到背靠背回款条件。	回款情况较好，与信用政策相匹配
	上海算丰信息有限公司	4.73	发行人与上海算丰约定付款分三阶段执行： 第一阶段预付：合同生效后，2023 年 12 月 29 日前支付第一阶段货款 2,118.4 万元。 第二阶段预付：收到第二批服务器后，且第一批服务器集群运行的结果符合约定后的 30 个工作日内，支付货款 3,168 万元 第三阶段回款：全部 160 台服务器调试完成，达到约定的标准后 10 个工作日内，支付货款 5,305.60 万元	2023 年末，发行人对上海算丰的应收款项余额为收入确认前预开发票的增值税税额。2024 年末，发行人对上海算丰无应收款项。具体回款情况如下： 2024 年 1 月 8 日、2024 年 4 月 25 日按照合同约定，分别向发行人支付前两阶段预付款项。2024 年 5 月 23 日，上海算丰出具千卡集群验收报告，发行人于当月确认收入，2024 年 6 月 19 日向发行人支付 5,305.60 万元尾款。	回款情况较好，周期较短，与信用政策整体相匹配
	汇天网络科技有限公司	26.87	报告期内发行人对汇天网络为订单	截至 2024 年 10 月，汇天网络合	已按合同约定预付大部分款

年份	客户名称	实际回款周期（天）	信用政策	回款情况分析	与信用政策匹配性分析
	公司（注1）		#S0880500098，约定付款方式如下： 第一阶段：公司发货前向汇天网络提供该批次服务器的合格证明等文件，汇天网络收到该等文件之日起10日内预付该批次服务器价款总金额的80%，公司收到预付货款后发货。 第二阶段：在本智算中心算力集群建设完成后，甲方或甲方指定的运营单位出租算力服务，当算力服务租金收入达到合同价款的60%，汇天网络于租金收入达到合同价款的60%之日起10日内支付本合同价款总金额的20%。即2,174万元	计向发行人合计预付货款7,368.60万元，但未付清第一阶段款项，剩余款项存在逾期，逾期原因参见本问询回复报告之“6.关于应收账款/一/（二）区分直销/经销客户，列示.....逾期金额、逾期比例、逾期原因”。截至本回复出具日，逾期款项已完成回款。	项，存在小部分款项逾期，期后已完成回款，整体回款周期较短
	浙江武珞智慧城市技术有限公司	142.27	报告期内，发行人对武珞智慧主要为预收货款，此外，订单8800000066在订单生效后90天内付款；其他仅有1个金额较小的订单给予一定信用期，具体为#SO8800000070，于2024年12月16日下单，订单金额19.2万元，给予一个月的信用期，需于2025年1月15日前付清款项。	截至2024年10月，武珞智慧按照订单约定，预付相关款项，合计预付2,289.47万元。2024年11月与12月的两笔订单，合计金额5,092.44万元，武珞智慧未按照合同约定如期付款，存在逾期情况，具体逾期原因参见本问询回复报告之“6.关于应收账款/一/（二）区分直销/经销客户，列示.....逾期金额、逾期比例、逾期原因”。截至本回复出具日，武珞智慧已逾期款项已全额回款。	大部分订单按照合同约定预付回款，存在两个订单存在逾期付款情形，期后已完成回款
2023年	公司1	130.91	1) 协议签署之日起的10个工作日内支付许可方900万（需外加6%增值税） 2) 许可产品交付并验收合格后的10个工作日内支付许可方1500万（需外加6%增值税）；	报告期内，公司1按照合同约定分别于2023年6月，2024年1月付款954万元、1,590万元。截至2025年3月末，尚未达到第三阶段付款条件。	回款情况较好，周期较短，与信用政策整体相匹配

年份	客户名称	实际回款周期（天）	信用政策	回款情况分析	与信用政策匹配性分析
			3) 使用许可产品的芯片流片成功后 10 个工作日内支付许可方 600 万（需外加 6% 增值税）。	由于第二阶段发行人于 2023 年 11 月交付，公司 1 于 2024 年 1 月初完成付款，实际付款周期较短，但因本题计算公式中“应收账款平均余额=期初期末应收余额平均值”导致存在一定计算偏差，导致计算得到的回款周期长于实际回款周期。	
	上海燃云科技有限公司	171.23	报告期内，发行人对上海燃云主要信用政策为“每月最后一个工作日之前一次性付清记账目为上月份的全部货款”，约为 1-2 月的信用期。少数订单约定为预收货款	2023 年，上海燃云整体回款情况较好。2023 年末，发行人对上海燃云应收款项共 1,534.25 万元，其中 1,235.93 万元为 2023 年 12 月确认收入，2024 年 1 月（信用期内）收到货款，回款周期较短，但因本题计算公式“应收账款平均余额=期初期末应收余额平均值”，在确认收入与付款存在跨年的情况下，存在一定计算偏差，导致计算得到的回款周期长于实际回款周期。 其余 298.32 万元为 2023 年 11 月确认收入尚未回款的款项，截至本回复出具日尚未收到货款，出现逾期，逾期原因参见本问询回复报告之“6.关于应收账款/一/（二）区分直销/经销客户，列示.....逾期金额、逾期比例、逾期原因”。	整体回款情况较好，大部分订单回款周期较短，与信用政策相匹配，小部分应收款项出现逾期
	上海闪马智能科技有限公司	-	预收货款	按照合同约定付款，未形成期末应收余额，无回款周期	回款情况较好，与信用政策相匹配

年份	客户名称	实际回款周期（天）	信用政策	回款情况分析	与信用政策匹配性分析
	中科远见（重庆）科技有限公司	-	预收货款	按照合同约定付款，未形成期末应收余额，无回款周期	回款情况较好，与信用政策相匹配
	湖南沐仕杰半导体科技有限公司	-	预收货款	按照合同约定付款，未形成期末应收余额，无回款周期	回款情况较好，与信用政策相匹配
2022 年	上海东焱信息技术有限公司	-	预收货款	按照合同约定付款，未形成期末应收余额，无回款周期	回款情况较好，与信用政策相匹配

注 1：汇天网络科技有限公司已于 2025 年 8 月 15 日完成逾期款项回款

4、报告期内不同客户信用政策不存在较大差异，不存在放宽信用政策刺激销售的情形

报告期内，发行人以“预收货款”的模式为主，同时，由于公司目前仍处于业务快速发展期，针对特定的客户类型、项目类型等因素的考量，采取了特定的信用期限。发行人的信用政策如下：

经销客户方面，2023 年及之前，发行人以“预收货款”方式为主，但由于发行人处于业务发展初期，基于业务拓展的考虑，根据与客户的商业洽谈，给予部分重要经销客户 30 天左右账期。随着发行人业务规模的不断扩大，2024 年以来，发行人对于签约的一级经销商普遍采用“预收货款用于后续抵扣+抵扣完毕后先款后货/申请账期”的收款方式，由经销商先预付一般不低于 2,000 万元的货款用于后续订单的抵扣，待预付款抵扣完毕后优先采用“先款后货”方式进行，若涉及经销商需要一定账期情况，需由经销商向发行人申请，发行人根据经销商项目类型以及战略意义酌情给予 30-90 天不等账期。

直销客户方面，对于一般的直销客户，发行人以“预收货款”方式为主。对于 OEM 厂商，如新华三、公司 2 等，发行人按照行业惯例，给予 90-120 天的信用期。同时，对于战略性项目或重要客户，发行人基于项目特点、业务拓展、维系与客户的长期合作关系的需要等因素考虑，对部分重要直销客户采取了特定的收款策略。如对于分阶段执行的重要项目，一般约定预收部分款项，待达到后续收款条件，再给予一定信用期。对于客户采购的主要用途为千卡以上的算力集群项目，考虑到投资金额大、实施周期长，基于现金流管理、缓解各方资金压力及保证业务稳定性的考虑，发行人允许部分重点项目存在一定的信用期或约定“背靠背”付款模式。

综上，报告期内发行人以预收货款的模式为主，同时，由于公司目前仍处于业务快速发展期，针对特定的客户类型、项目类型等因素的考量，采取了特定的信用期限。总体而言，发行人给予同一客户的信用期保持相对稳定，同类客户、同类项目的信用政策不存在较大差异，不存在放松信用期以刺激销售的情形。

（二）区分直销/经销客户，列示各期对主要客户的收入金额、应收账款余额及占对应收入的比重、账龄结构、最新期后回款金额及回款比例、逾期金额、逾期比例、逾期原因

1、报告期各期发行人对主要直销/经销客户的收入及应收情况

报告期各期，发行人对直销/经销前五大客户的应收账款账龄均在一年以内，并主要集中在 3 个月以内，应收账款账龄结构较短。

报告期各期，发行人对直销/经销前五大客户的收入金额、应收账款余额及占对应收入的比重、账龄结构、最新期后回款金额、回款比例及是否存在逾期情况如下：

单位：万元

年份	销售模式	序号	客户名称	营业收入	应收账款余额	占对应收入的比重	应收账款账龄结构				截至 2025 年 7 月末期后回款	截至 2025 年 7 月末回款比例	是否逾期
							3M 以内	3M 以内占比	3M-1 年	3M-1 年占比			
2025 年 1-3 月 /2025 年 3 月末	直销	1	上海源庐加佳信息科技有限公司	9,095.93	-	-	-	-	-	-	-	-	否
		2	新华三信息技术有限公司	2,918.69	18,282.84	626.40%	3,394.12	18.56%	14,888.71	81.44%	15,886.27	86.78%	否
		3	公司 2	1,072.30	2,175.80	202.91%	1,211.70	55.69%	964.10	44.31%	2,175.80	100.00%	否
		4	铎云（北京）半导体有限公司	601.06	339.60	56.50%	339.60	100.00%	-	0.00%	339.60	100.00%	是
		5	广州祥泽科技有限公司	213.45	241.20	113.00%	241.20	100.00%	-	0.00%	-	0.00%	是
	经销	1	超讯通信股份有限公司	12,800.17	16,105.73	125.82%	14,449.19	89.71%	1,656.54	10.29%	1,736.00	10.78%	是
		2	北京创世云科技股份有限公司	2,359.28	733.76	31.10%	733.76	100.00%	-	0.00%	312.50	42.59%	否
		3	华科智谷（上海）科技发展	1,136.14	283.84	24.98%	283.84	100.00%	-	0.00%	283.84	100.00%	否

年份	销售模式	序号	客户名称	营业收入	应收账款余额	占对应收入的比重	应收账款账龄结构				截至 2025 年 7 月末期后回款	截至 2025 年 7 月末回款比例	是否逾期
							3M 以内	3M 以内占比	3M-1 年	3M-1 年占比			
			有限公司										
		4	北京优创世纪科技有限公司	976.30	-	-	-	-	-	-	-	-	否
		5	上海莹矽科技有限公司	471.67	392.60	83.24%	392.60	100.00%	-	0.00%	392.60	100.00%	是
		合计		31,645.01	38,555.37	121.84%	21,046.02	54.59%	17,509.36	45.41%	17,318.89	44.92%	
2024 年度 /2024 年末	直销	1	新华三信息技术有限公司	15,095.15	14,118.99	93.53%	14,118.99	100.00%	-	0.00%	11,707.89	82.92%	否
		2	上海算丰信息有限公司	9,403.19	-	-	-	-	-	-	-	-	否
		3	汇天网络科技有限公司	7,332.18	1,079.48	14.72%	1,079.48	100.00%	-	0.00%	-	0.00%	是
		4	上海源庐加佳信息科技有限公司	3,881.80	-	-	-	-	-	-	-	-	否
		5	公司 2	3,035.25	5,228.91	172.27%	3,364.91	64.35%	1,864.00	35.65%	5,228.91	100.00%	否
	经销	1	苏州瑞芯智能科技有限公司	13,046.02	12,742.00	97.67%	12,742.00	100.00%	-	0.00%	-	0.00%	否
		2	浙江武珞智慧城市技术有限公司	6,532.66	5,092.44	77.95%	5,092.44	100.00%	-	0.00%	5,092.44	100.00%	是
		3	上海华研企业发展（集团）有限公司	5,487.43	3,483.00	63.47%	3,483.00	100.00%	-	0.00%	3,483.00	100.00%	是
		4	北京创世云科技股份有限公司	1,191.50	-	-	-	-	-	-	-	-	否
		5	超讯通信股份有限公司	749.09	3,414.25	455.79%	3,414.25	100.00%	-	0.00%	1,837.16	53.81%	是
	合计			65,754.27	45,159.06	68.68%	43,295.06	95.87%	1,864.00	4.13%	27,349.40	60.56%	
2023 年度 /2023	直销	1	公司 1	2,216.67	1,590.00	71.73%	1,590.00	100.00%	-	0.00%	1,590.00	100.00%	否
		2	上海闪马智能科技有限公司	600.00	-	-	-	-	-	-	-	-	否

年份	销售模式	序号	客户名称	营业收入	应收账款余额	占对应收入的比重	应收账款账龄结构				截至 2025 年 7 月末期后回款	截至 2025 年 7 月末回款比例	是否逾期
							3M 以内	3M 以内占比	3M-1 年	3M-1 年占比			
年末		3	浙江宇视科技有限公司	18.40	20.79	113.00%	20.79	100.00%	-	0.00%	20.79	100.00%	否
		4	杭州晖众网络科技有限公司	15.93	9.00	56.50%	9.00	100.00%	-	0.00%	9.00	100.00%	否
		5	上海清华国际创新中心	12.30	-	-	-	-	-	-	-	-	否
	经销	1	上海燃云科技有限公司	1,635.26	1,534.25	93.82%	1,534.25	100.00%	-	0.00%	1,235.93	80.56%	是
		2	中科远见（重庆）科技有限公司	139.51	-	-	-	-	-	-	-	-	否
		3	湖南沐仕杰半导体科技有限公司	116.58	-	-	-	-	-	-	-	-	否
		4	北京并行科技股份有限公司	85.00	-	-	-	-	-	-	-	-	否
		5	上海点积实业有限公司	70.20	-	-	-	-	-	-	-	-	否
		合计		4,909.85	3,154.04	64.24%	3,154.04	100.00%	-	0.00%	2,855.72	90.54%	
2022 年度 /2022 年末	直销	1	上海东焱信息技术有限公司	42.64	-	-	-	-	-	-	-	-	否
	合计			42.64	-	-	-	-	-	-	-	-	

注 1：汇天网络科技有限公司已于 2025 年 8 月 15 日完成逾期款项回款

2、报告期各期存在逾期的主要直销/经销客户具体情况分析

报告期内，应收账款逾期的主要原因是终端项目投资金额较高、客户或下游终端客户内部流程审批较慢、终端项目建设进度未完成所致。报告期内，发行人下游主要用途为集成服务器后建设智能算力集群，通常投资规模大、实施周期长，参与主体众多，相关招标、验收及付款流程较长。发行人主要经销商超讯通信、武珞智慧与上海华研虽然具有较高的经营实力与行业地位，但难以实现对全部 GPU 板卡采购资金的垫付，导致存在一定的账款逾期。报告期内，存在逾期的主要直销/经销客户的逾期金额、逾期比例、逾期原因如下：

单位：万元

年度	单位名称	当期营业收入	期末余额	逾期金额	逾期比例	逾期原因
2025 年 1-3 月/2025 年 3 月末	超讯通信股份有限公司	12,800.17	16,105.73	16,105.73	100.00%	因项目建设进度略晚于预期，超讯通信尚未收到下游回款，资金暂时紧张，回款延迟
	铎云（北京）半导体有限公司	601.06	339.60	339.60	100.00%	客户内部付款流程较慢，回款逾期，截至本回复出具日已完成回款
	上海莹矽科技有限公司	471.67	392.60	192.00	48.90%	客户内部付款流程较慢，回款逾期，截至本回复出具日已完成回款
	广州祥泽科技有限公司	213.45	241.20	241.20	100.00%	由于项目原厂商未能按时提供已适配的机头，项目的进度有所延误，导致未能及时回款
2024 年度 /2024 年末	汇天网络科技有限公司	7,332.18	1,079.48	1,079.48	100.00%	因汇天网络公司内部架构变动，导致付款流程延迟，截至本回复出具日已完成回款
	浙江武珞智慧城市技术有限公司	6,532.66	5,092.44	5,092.44	100.00%	前期由于终端集群建设中，资金暂时紧张，回款延迟，截至本回复出具日已完成回款
	上海华研企业发展（集团）有限公司	5,487.43	3,483.00	3,483.00	100.00%	前期由于终端集群建设中，资金暂时紧张，回款延迟，截至本回复出具日已完成回款
	超讯通信股份有限公司	749.09	3,414.25	3,313.09	97.04%	同上述分析
2023 年度 /2023 年末	上海燃云科技有限公司	1,635.26	1,534.25	298.32	19.44%	其部分客户需求发生变化，少量发行人 GPU 板卡周转速度变慢，回款延迟

综上，报告期内主要客户付款逾期主要是由于终端项目投资金额较高、各环节参与主体内部流程审批较慢、终端项目建设进度未完成所致，不存在重大的回收风险。

后续，发行人将进一步加强日常管理，加大回收力度，并密切关注回款状况和逾期催收情况，及时处理逾期账款，并采取有效措施，以提高应收账款的周转速度。

（三）报告期各期存在“背靠背”结算条款的应收账款金额及占比、约定“背靠背”结算的原因、相关款项回收和逾期情况、是否存在回款风险

1、报告期内发行人约定“背靠背”结算的原因

报告期内，与发行人签订“背靠背”结算条款的客户共三家，主要原因如下：

序号	对应客户名称	签署背靠背条款的原因	“背靠背”付款条款主要内容
1	新华三信息技术有限公司	第一，发行人与新华三签署订单时，终端项目的招标时间尚未最终确定，新华三收款时间存在不确定性；第二，由于终端项目投资金额重大，新华三虽然具有较高的经营实力与行业地位，但对全部 GPU 板卡采购资金进行垫付的资金压力较大；第三，发行人与新华三需要共同合作推进终端项目中的超节点技术。因此，基于战略合作、缓解各方资金压力及保证业务稳定性等因素考虑，发行人与新华三就付款条件做了背靠背的约定	双方同意款项应当在收到产品后 120 天支付，或者自新华三收到本次项目货款后 5 个工作日内向发行人支付
2	苏州瑞芯智能科技有限公司	第一，项目 B 投资规模较大，瑞芯智能与下游签署背靠背条款，背靠背条款通过交易链条向上游穿透；第二，终端项目投资金额重大，瑞芯智能虽然具有较高的经营实力与行业地位，但对全部 GPU 板卡采购资金进行垫付的资金压力较大；第三，瑞芯智能为发行人引入终端优质大型项目资源，项目 B 是国家级算力集群。因此，基于战略合作、缓解各方资金压力及保证业务稳定性等因素考虑，发行人与瑞芯智能就付款条件做了背靠背的约定	瑞芯智能已向发行人支付的 2,000 万元预付款中，可等额抵扣订单金额，抵扣后剩余货款采用和下游客户同比例背靠背付款方式，即瑞芯智能收到下游客户比例金额的款项后，10 天内同比例支付给发行人
3	上海爱卿网络科技有限公司	第一，项目 B 投资规模较大，上海爱卿与下游签署背靠背条款，背靠背条款通过交易链条向上游穿透；第二，由于终端项目投资金额重大，上海爱卿对相关采购资金的垫付压力较大。因此，基于战略合作、缓解各方资金压力及保证业务稳定性等因素考虑，发行人与上海爱卿就付款条件做	上海爱卿收到下游客户同比例货款后，10 天内同比例支付给发行人

序号	对应客户名称	签署背靠背条款的原因	“背靠背”付款条款主要内容
		了背靠背的约定	

商业案例中，“背靠背”条款通常出现在大型工程项目或涉及上下游多方参与的市场交易中，该类交易的特点是各个节点的支付流程较长，一方面“背靠背”条款为参与主体提供了一个相对明确的支付条件，从而简化了合同主体对自身支付义务的判断逻辑；另一方面，背靠背条款也满足了参与主体对风险分配的现实需求，通过背靠背条款，交易主体将自身向上游供应商付款的义务直接或间接地绑定于其收到下游客户付款这一条件，从而减轻了自身财务压力及其垫付大量资金造成的现金流压力，因此“背靠背”条款在大型项目中的产生有其商业逻辑与合理性。

发行人与上述三家客户签署“背靠背”结算条款，主要原因系终端项目投资规模大、参与主体多、实施周期长，各环节流程较长。发行人主要经销商虽然具有较高的经营实力与行业地位，但对全部 GPU 板卡采购资金的垫付压力较大，基于现金流管理、缓解各方资金压力及保证业务稳定性的考虑，发行人重点项目约定了“背靠背”付款模式。

2、报告期内发行人“背靠背”结算条款的应收账款情况

报告期各期，存在“背靠背”结算条款的应收账款金额及占比、相关款项回收情况如下：

单位：万元

年度	客户名称	“背靠背”结算条款销售订单号	订单应收账款余额	占当期应收账款的比重	订单-期后回款
2024 年度/2024 年 12 月末	新华三信息技术有限公司	880500105	12,671.10	26.08%	10,260.00
	苏州瑞芯智能科技有限公司	880000072	12,742.00	26.22%	-
	上海爱卿网络科技有限公司	880500151	880.14	1.81%	-
2025 年 1-3 月/2025 年 3 月末	新华三信息技术有限公司	880500105	12,671.10	19.84%	10,260.00
	苏州瑞芯智能科技有限公司	880000072	12,742.00	19.95%	-
	上海爱卿网络科技有限公司	880500151	994.56	1.56%	-

注：上海爱卿网络科技有限公司应收账款余额的变化为开票的影响

3、报告期内发行人“背靠背”结算条款的应收账款未出现逾期情况，不存在重大回收风险

报告期内，不存在“背靠背”相关应收款项逾期情况，也不存在重大回收风险。具体分析如下：

第一，新华三、瑞芯智能、上海爱卿均为行业内知名客户、资信情况良好。新华三是紫光股份旗下领先的数字化解决方案提供商，在智算网络、大模型一体机、全栈 AI 解决方案等领域深度布局，参与多个智算中心建设，是国内人工智能与算力基础设施建设的领军企业。根据“问题 6 一/（一）/3、报告期内主要客户的实际回款周期、与信用政策匹配性分析”，新华三其他订单历史回款情况较好，资信情况较好。瑞芯智能是上市公司迈信林的子公司，专注于人工智能通用应用系统、基础软件开发及算法模型创新。上海爱卿成立于 2020 年，注册资本 1,180 万元，登记状态为存续，无经营异常记录，主要从事算力中心建设、GPU 芯片服务器贸易及集成，在算力及 AI 行业中时间较长，客户、合作伙伴的积累较好。根据公开资料查询，上海爱卿具有高新技术企业资质、曾入选“上海市科技型中小企业”并获“上海市创新型中小企业”认定。综上，新华三、瑞芯智能及上海爱卿作为行业内知名 AI 领域算力设施建设、集成或算力服务公司，其履约能力强、信用良好，相关应收款项不存在重大回收风险。

第二，“背靠背”订单对应的终端项目均为国家级算力集群，终端客户资金实力强，信用情况较好，履约能力较强，相关应收款项不存在重大回收风险。

第三，截至本回复出具日，新华三已回款 10,260.00 万元。发行人商务部门不定期参与新华三、瑞芯智能以及项目其他相关方召开的会议，对齐项目进度，从而及时掌握项目建设的进度和回款安排，一旦达到回款条件，发行人将依据合同约定推进客户付款。发行人后续将持续关注与跟进项目进展，督促应收款项回款。

（四）结合应收账款账龄结构、逾期比例较高、期后回款比例较低、客户资信情况以及最新回款等情况，说明报告期内坏账准备计提的充分性

1、发行人应收账款账龄结构

单位：万元

账龄	2025/3/31		2024/12/31		2023/12/31		2022/12/31	
	金额	比例	金额	比例	金额	比例	金额	比例
3 个月以内	22,203.68	34.77%	44,486.60	91.55%	3,747.41	100.00%	-	0.00%
3 个月至 1 年	41,353.93	64.76%	4,107.99	8.45%	-	0.00%	-	0.00%
1 年以内小计	63,557.60	99.53%	48,594.59	100.00%	3,747.41	100.00%	-	0.00%
1 至 2 年	298.32	0.47%	-	0.00%	-	0.00%	-	0.00%
小计	63,855.92	100.00%	48,594.59	100.00%	3,747.41	100.00%	-	0.00%
减：坏账准备	2,349.40	-	650.27	-	37.47	-	-	-
合计	61,506.52	-	47,944.33	-	3,709.94	-	-	-

由上表可见，报告期内发行人应收账款账龄主要集中在 1 年以内，整体账龄结构与发行人自 2023 年起步的业务实际情况相符，账龄分布合理。截至报告期末，账龄超过一年的仅有发行人对上海燃云应收账款 298.32 万元，金额较小，上海燃云系北京创世云科技股份有限公司（简称“创世云”）的子公司，创世云目前盈利情况较好，上海燃云不存在经营异常的情形，回收风险较小。发行人已按照坏账计提政策计提信用减值损失。

2、发行人与同行业公司坏账计提政策比较

项目	寒武纪	海光信息	龙芯中科	景嘉微	摩尔线程	发行人
3 个月以内	不计提	1%	1%	未披露	1%	1%
3-6 个月	5%	1%	1%	未披露	1%	5%
6 个月至 1 年	5%	5%	5%	未披露	5%	5%
1 至 2 年	10%	15%	10%	未披露	10%	20%
2 至 3 年	30%	30%	50%	未披露	未披露	50%
3-4 年	100%	50%	100%	未披露	未披露	100%
4-5 年	100%	100%	100%	未披露	未披露	100%
5 年以上	100%	100%	100%	未披露	未披露	100%

据上表可见，公司应收账款坏账准备计提比例总体高于同行业可比公司计提比例，计提政策较为谨慎。

3、发行人客户应收账款逾期情况及期后最新回款情况

报告期内，发行人应收账款逾期的主要原因是项目投资金额较高、客户或下游终端客户内部流程审批较慢、终端项目建设进度未完成所致。报告期内，发行人主要客户应收账款逾期情况以及期后回收情况参见本问询回复报告之“6.关于应收账款/一/(二)/2、报告期各期存在逾期的主要直销/经销客户具体情况分析”。

截至 2025 年 7 月末，报告期内发行人多笔金额较大的逾期应收账款已完成收回，期后回款比例已明显提升。截至 2025 年 7 月末，报告期各期末，公司应收账款的期后回款情况如下：

单位：万元

项目	2025 年 3 月末	2024 年末	2023 年末	2022 年末
应收账款余额	63,855.92	48,594.59	3,747.41	-
其中：无“背靠背”条款的应收账款余额	37,562.68	22,301.35	3,747.41	-
期后回款金额	29,924.64	28,412.26	3,449.09	-
其中：无“背靠背”条款期后回款金额	19,664.64	18,152.26	3,449.09	-
期后回款比例	46.86%	58.47%	92.04%	-
其中：无“背靠背”条款的应收账款期后回款比例	52.35%	81.40%	92.04%	-

截至 2025 年 7 月末，公司 2023 年末、2024 年末和 2025 年 3 月末应收账款回款比例分别为 92.04%、58.47%和 46.86%，由于公司与 2024 年第四季度部分大额客户销售合同中存在“背靠背”结算条款，需待该等客户与其下游客户结算款项后，再与发行人结算款项。剔除存在“背靠背”结算条款的应收款项，截至 2025 年 7 月末，公司 2023 年末、2024 年末和 2025 年 3 月末应收账款回款比例分别为 92.04%、81.40%和 52.35%，总体期后回款情况良好，应收账款正在陆续回收过程中，不存在重大回收风险。预计除超讯通信外，发行人其余主要客户的逾期账款将在 2025 年 9 月前基本完成收回。

为保证公司的资金正常周转、往来款项的资金及时收回，发行人制定了《应收账款管理制度》，由财务部负责销售与收款等应收款项的结算与记录，按照公司要求的应收账款合理额度和周转率监控管理，监督管理货款回收。由商务部负责催收货款，商务部销售人员应对其经办的应收账款全程负责，并对客户的经营情况、催付能力进行追踪分析，及时了解客户资金持有量与调剂程度，保证应收

账款的回收。

报告期内，公司各部门严格按照管理制度要求，针对客户的信用状况、按照销售合同及时催收货款、计提坏账准备等具体工作。发行人逾期账款的具体催收措施如下：

类型	催收动作
每周	财务部每周将逾期款项清单发送商务相关责任人及部门负责人，相关责任人反馈逾期原因及回款计划。对于大额重要款项，每周向管理层汇报最新进展。
每月	财务专员每月发送客户对账单，由商务部门与客户对账确认应收款余额，沟通回款时间
不定期	商务部销售专员通过邮件/电话向客户催收
总经理专项催收	针对目前长期未回款的重点逾期客户，由总经理亲自催收

4、逾期客户的资信情况分析

截至本回复出具日，报告期末逾期账款 100 万元以上且仍未回款的客户资信情况如下：

单位：万元

序号	客户名称	成立时间	注册资本	截至本回复出具日仍逾期金额	业务与经营情况分析	与发行人合作的背景	是否破产或被列为失信被执行人
1	超讯通信股份有限公司	1998 年 8 月	15,758.68	14,369.73	上证主板上市公司，2025 年 1-3 月营业收入为 2.66 亿元，归母净利润为 5,007.34 万元	通过他人介绍与发行人初始接洽，基于下游客户板卡采购需求与发行人洽谈合作，2024 年起形成销售订单	否，参见下文分析
2	中仪英斯泰克进出口有限公司	1986 年 2 月	10,000	1,172.95	国有企业，是首批入选国资委“科改示范行动”的 204 家国有科技型企业之一，是中国仪器进出口集团有限公司为进入我国的世界著名仪器、设备制造商及其国内用户提供技术咨询、技术服务以及技术维护而在中关村成立的高新技术企业。根据其官方网站信息，其拥有覆盖全国的 46 家技术服务站点和北京、上海共 3 万平米的保税库。	中国通用技术（集团）控股有限责任公司下属公司，通过他人介绍与发行人初始接洽，基于集群建设需求与发行人洽谈合作，2023 年起形成销售订单	否
3	上海燃云科技有限公司	2022 年 2 月	1,000	298.32	上海燃云科技有限公司系北京创世云科技股份有限公司（以下简称“创世云”）的子公司，创世云是一家以云计算、网络数据优化及智算服务为核心业务的高新技术企业，并有在 A 股上市计划。根据创世云股东福石控股相关公告，创世云 2020 年、2021 年净利润均超 1500 万元。此外，福石控股 2024 年年报披露确认因投资	通过他人介绍与发行人初始接洽，基于下游客户板卡采购需求与发行人洽谈合作，2023 年起形成销售订单	否

序号	客户名称	成立时间	注册资本	截至本回复出具日仍逾期金额	业务与经营情况分析	与发行人合作的背景	是否破产或被列为失信被执行人
					创世云产生的其他综合收益的利得414.5万元，说明创世云目前盈利情况较好，不存在经营异常的情形。		
4	广州祥泽科技有限公司	2021年9月	100	241.20	无公开信息，已出具明确的回款计划说明	通过他人介绍与发行人初始接洽，基于下游客户板卡采购需求与发行人洽谈合作，2024年起形成销售订单	否

2025 年 7 月 31 日，超讯通信发公告称公司及子公司部分银行账户被冻结合计 2,248.41 万元。截至本回复出具日，超讯通信部分银行账户内被冻结的资金已解除冻结。发行人正与超讯通信持续沟通，已要求其出具发行人认可的回款计划。据了解，目前超讯通信计划根据终端要求于 2025 年 9 月末向下游客户交付集成了发行人 GPU 板卡的服务器，预计交付完成验收后收到的货款将足额覆盖对发行人的 1.43 亿债务。届时对该下游客户的收款将采取严格的保护措施以保证对发行人的回款。后续发行人将根据超讯通信的还款计划、实际还款金额、披露的财务数据等综合评价信用风险水平。

经核查，除超讯通信外，截至本回复出具日，报告期末应收账款仍存逾期的客户整体资信情况较好，逾期原因主要系客户内部付款流程较慢所致，未发生破产、被列为失信人等情况，经营及信用情况无重大异常，未发现信用风险显著上升的迹象。发行人报告期内暂未发生因客户经营或信用情况恶化导致的大额损失，应收款项无法收回的风险较低，故未进行单项计提。

综上，发行人应收账款账龄分布合理，坏账准备计提政策设计合理，与同行业可比公司不存在较大差异，逾期客户账龄主要在 1 年以内，且客户经营及信用情况无重大异常，发行人报告期内暂未发生大额信用损失，催收计划得到有效执行。报告期内发行人坏账准备计提充分，符合《企业会计准则》及《监管规则适用指引——发行类第 5 号》的相关要求。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、申报会计师执行了如下核查程序：

1、了解和评价、测试发行人销售与收款循环相关的内部控制设计和运行的有效性。

2、了解公司不同销售模式信用政策，取得发行人针对不同客户的具体信用政策。获取报告期内发行人与经销/直销主要客户的销售合同，查阅主要客户的信用政策，对比分析不同客户之间信用政策的差异及同一客户在报告期内的信用政策是否发生重大变化，结合平均回款周期分析回款情况及是否存在放宽信用政策刺激销售的情形；查询同行业可比公司应收账款坏账计提政策，对比核查发行

人的计提比例与同行业公司相比是否存在重大差异。

3、了解公司不同客户的结算方式，查阅所有涉及“背靠背”结算条款的客户的相关合同，了解采用“背靠背”结算条款是否符合行业惯例。

4、访谈公司相关人员，了解公司业务模式、信用政策情况，与部分客户签订“背靠背”结算条款的原因，分析应收账款规模大幅增长的原因。

5、了解发行人应收账款坏账准备会计政策及计提方法，并评价是否符合《企业会计准则》的相关规定。

6、获取应收账款账龄表、期后回款情况、坏账准备计提政策及坏账准备计提表。复核账龄分类的准确性，核查应收账款坏账计提是否充分。

7、对于逾期客户，询问发行人管理层及销售负责人，了解逾期的原因、是否存在纠纷或潜在纠纷、后续回款计划，结合逾期情况、各类公开信息综合判断信用风险。

8、查询报告期各期经销/直销重要客户的工商企业信用报告，识别相关客户是否存在失信、破产或资金困难等可能导致应收账款无法收回的情形。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、报告期内，发行人应收账款大幅增加具有合理性，与营业收入规模相匹配，主要客户的实际回款周期总体较短，应收账款账龄较短，与信用政策整体较为匹配，同类客户、同类项目的信用政策不存在较大差异，不存在放宽信用政策刺激销售的情形。

2、报告期内，其主要客户付款逾期是由于项目投资金额较高、客户或下游终端客户内部流程审批较慢、终端项目建设进度未完成所致。截至本回复出具日，逾期账款除超讯通信外期后回款情况良好，未回款项已有明确的回款计划，不存在重大的信用风险，不存在特殊条款约定或其他利益安排。

3、报告期内，发行人与特定客户存在“背靠背”结算条款具有合理性，相关款项不存在逾期情况，不存在重大回款风险。

4、报告期内，发行人应收账款账龄主要集中在一年以内，账龄较短。逾期

客户未发生破产、被列为失信人等情况，逾期款项均已回款或有明确回款计划，报告期内，坏账准备计提比例与同行业公司相比计提政策没有重大差异，应收账款坏账准备计提充分。

7.关于采购和供应商

根据申报材料：（1）发行人生产物料采购内容主要为晶圆、HBM 内存、外购硬件及配件、委托加工、电子元器件等；（2）报告期各期，发行人前五大供应商采购金额占当期采购总额的比例为 95.67%、95.01%、91.30%和 93.87%，采购集中度较高；（3）报告期各期末，预付款项金额为 5,178.40 万元、33,122.17 万元、88,986.25 万元和 110,734.37 万元，大幅增长，主要系晶圆、HBM 等原材料供应商在付款方式上均要求较大比例的预付款项。

请发行人披露：（1）区分板卡/服务器列示不同细分产品对应的采购类别、金额、主要供应商；（2）采购主要原材料需要预付大额款项的具体原因及合理性，是否符合行业惯例，预付款项金额与经营性现金流、采购金额的匹配性，预付款项规模及占比与同行业可比公司是否存在差异；（3）发行人预付款相关资金管理方式，是否已建立安全有效的资金管理内控机制。

请保荐机构、申报会计师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）区分板卡/服务器列示不同细分产品对应的采购类别、金额、主要供应商

从公司的采购和生产模式来看，公司向晶圆制造厂采购定制加工生产的晶圆，向封装测试厂采购封装测试服务，向加工厂商采购板卡加工服务，期间公司辅以工艺管理和测试支持。针对部分客户的解决方案需求，公司在使用已经加工完毕的 GPU 板卡产品的基础上，根据客户的具体需求采购相应配套的服务器、存储设备及网络设备等硬件，集成为含发行人 GPU 板卡的服务器或相关产品并向客户交付。因此，在生产 GPU 板卡的层面，公司采购的原材料为晶圆、HBM 内存、委托加工及电子元器件；在已经加工完毕的 GPU 板卡产品基础上，公司为生产服务器进一步采购的原材料为外购硬件及配件。

1、板卡生产对应的晶圆、HBM 内存、委托加工及电子元器件采购金额、主要供应商及其对应的最终原材料供应商名称

在生产 GPU 板卡的层面，公司采购的原材料为晶圆、HBM 内存、委托加工及电子元器件，具体情况如下：

(1) 晶圆

报告期内，公司晶圆的采购金额、主要供应商（报告期各期前 5 大且报告期合计采购金额大于 200 万元，若不足 5 家则全部列示，下同）情况如下：

单位：万元

主要供应商名称	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
公司 D	-	39,766.61	5,668.86	-
公司 E	-	-	2,652.82	260.33
公司 C	1,476.65	537.82	-	-
公司 H	-	333.04	-	-
合计	1,476.65	40,637.47	8,321.68	260.33
占当期晶圆采购总额比例	100.00%	100.00%	100.00%	100.00%

(2) HBM 内存

报告期内，公司 HBM 内存的采购金额、主要供应商情况如下：

单位：万元

主要供应商名称	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
公司 A	9,207.61	32,680.18	6,361.94	800.67
公司 I	881.37	674.60	-	-
公司 J	-	395.86	-	-
合计	10,088.98	33,750.64	6,361.94	800.67
占当期 HBM 内存采购总额比例	100.00%	100.00%	100.00%	100.00%

(3) 委托加工

报告期内，公司委托加工的采购金额、主要供应商情况如下：

单位：万元

主要供应商名称	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
环鸿电子（昆山）有限公司	1,217.34	6,772.83	465.43	-

主要供应商名称	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
公司 B	1,760.55	5,922.59	263.51	-
公司 E	-	34.80	997.61	496.59
合计	2,977.89	12,730.22	1,726.55	496.59
占当期委托加工采购总额比例	98.18%	99.08%	100.00%	100.00%

(4) 电子元器件

报告期内，公司电子元器件的采购金额、主要供应商情况如下：

单位：万元

主要供应商名称	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
深南电路股份有限公司	291.38	691.91	209.81	70.77
环鸿电子（昆山）有限公司	144.95	432.60	116.06	15.73
宝德（上海）科技有限公司	0.00	453.97	212.18	0.00
北京市九州风神科技股份有限公司	19.49	385.29	83.22	0.00
合计	455.82	1,963.76	621.27	86.50
占当期电子元器件采购总额比例	93.15%	87.62%	79.91%	34.51%

2、服务器生产对应的外购硬件及配件采购金额、主要供应商及其对应的最终原材料供应商名称

在已经加工完毕的 GPU 板卡产品基础上，公司为生产服务器或相关产品进一步采购的原材料为外购硬件及配件，其采购金额、主要供应商名称如下：

单位：万元

主要供应商名称	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
万司信息技术（上海）股份有限公司	4,998.67	14,255.76	1,351.50	-
上海罗针信息科技有限公司	-	2,407.51	-	-
上海贝加信息技术有限公司	-	892.03	-	-
上海曦智科技有限公司	-	824.96	-	-
上海神州数码有限公司	-	578.02	-	-
苏州超集信息科技有限公司	5.58	201.96	-	-

主要供应商名称	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
合计	5,004.25	19,160.24	1,351.50	-
占当期外购硬件及 配件采购总额比例	100.00%	96.24%	100.00%	-

(二) 采购主要原材料需要预付大额款项的具体原因及合理性，是否符合行业惯例，预付款项金额与经营性现金流、采购金额的匹配性，预付款项规模及占比与同行业可比公司是否存在差异

1、采购主要原材料需要预付大额款项的具体原因及合理性，是否符合行业惯例

受行业惯例、采购周期、国际形势等因素影响，为保障未来产品供应，公司向晶圆、HBM 内存等重要原材料供应商进行提前订货并预付较高比例的货款，导致预付账款规模较大。芯片设计行业中，采用 Fabless 模式的公司将晶圆制造、封装和测试等环节委托给专业厂商完成。根据行业惯例，企业需通过预付款的采购模式进行采购或锁定产能。综上，发行人相关采购模式符合行业惯例。

2、预付款项金额与经营性现金流、采购金额的匹配性

报告期内，公司现金流量表中购买商品、接受劳务支付的现金与采购总额、预付款项、应付账款变化金额等的匹配性关系如下：

单位：万元

项目	2025 年 1-3 月 /2025 年 3 月末	2024 年度/2024 年末	2023 年度/2023 年末	2022 年度/2022 年末
采购总额 (A)	20,092.30	109,385.72	18,539.14	1,808.23
预付款项期初余额	88,986.25	33,122.17	5,178.40	4,697.12
预付款项期末余额	110,734.37	88,986.25	33,122.17	5,178.40
预付款项增加额 (B)	21,748.12	55,864.08	27,943.76	481.28
应付账款期初余额	12,111.11	7,274.87	4,593.77	5,688.41
应付账款期末余额	16,551.90	12,111.11	7,274.87	4,593.77
应付账款减少额 (C)	-4,440.79	-4,836.24	-2,681.10	1,094.65
应付票据期初余额	3,975.00	-	-	-
应付票据期末余额	-	3,975.00	-	-
应付票据减少额 (D)	3,975.00	-3,975.00	-	-
进项税额	7,000.11	22,321.66	5,213.14	1,654.42

项目	2025 年 1-3 月 /2025 年 3 月末	2024 年度/2024 年末	2023 年度/2023 年末	2022 年度/2022 年末
进项税转出	12.28	117.69	6.19	1.64
增值税进项税额 (E)	6,987.83	22,203.97	5,206.95	1,652.78
A+B+C+D+E	48,362.46	178,642.53	49,008.75	5,036.94
购买商品、接受劳务支付的现金	44,879.38	185,435.66	46,415.95	4,288.15

由上表，预付款项金额与经营性现金流、采购金额具备匹配性。

3、预付款项规模及占比与同行业可比公司是否存在差异

截至报告期末，发行人预付款项规模、占比与同行业可比公司的对比情况如下：

单位：万元

公司	2025 年 3 月末预付 款项	占流动资产合计 比例	预付款项较 2024 年 末增长比例
寒武纪	97,333.00	16.36%	25.69%
海光信息	162,070.28	7.96%	30.67%
龙芯中科	19,273.85	7.97%	21.79%
景嘉微	2,983.60	0.49%	63.21%
摩尔线程	未披露	未披露	未披露
平均值	/	8.20%	35.34%
发行人	110,734.37	11.04%	24.44%

注：摩尔线程尚未披露 2025 年 1-3 月财务数据，根据其招股说明书，2022 年末至 2024 年末，摩尔线程预付账款金额分别为 4,750.47 万元、7,614.62 万元及 56,710.83 万元，占流动资产比例分别为 3.58%、4.79%以及 8.37%，主要系其采购原材料等的预付货款。从其披露情况来看，发行人预付款项占比及增长态势与摩尔线程亦较为可比。

由上表，在当前行业产能紧张的背景下，同行业可比公司亦存在增加采购备货相关的大额预付款的情况，发行人预付款项规模、占比与同行业可比公司不存在显著差异。

（三）发行人预付款相关资金管理方式，发行人已建立安全有效的资金管理内控机制

报告期内，发行人建立了《资金管理制度》《采购管理程序》《（非研发）生产性采购管理程序》《（非研发）生产性供应商管理规定》《（非研发）生产性合格供方评价表》等制度，对供应商管理、采购付款、货物交期管理等进行了

详细规定，从供应商分类管理、预付款审批流程、对供应商及时跟踪及保障货物的及时供应等维度保证预付款项的资金安全。

公司支付预付款项的机制主要包括供应商准入与考评、预付款项审批、货物交期与预付款追踪管理三个方面，有效保障预付款项的安全性，具体如下：

1) 供应商的准入和考评：采购执行过程中，生产性采购相关供应商需要经供应链组进行选择、导入和考评，由研发和工程部门配合进行考评。非生产性采购的供应商均需经过采购部门、总经理助理办公室严格审核，包括材料审查、准入审核、定期评价。

2) 预付审批流程：发行人各类资金支出均需按规定填写相应申请单，具体包括“预付款单”、“无采购订单付款”、“请款单”、“费用报销申请”表单等。不同类型的资金支出审批程序不同，资金支出审批权限按照公司规定执行。其中金额在 5 万元以下（不含）由总经理助理审批，金额在 5 万元以上（含）由总经理审批。

3) 公司在预付账款产生后会及时进行货物交期及预付款追踪管理，主要包括：货物交期管理：采购订单下达前，采购人员须与供应商就订单货物交期达成一致，并严格要求供应商按照采购订单交期如期交货。从采购订单下达到交货前，采购人员须不定时跟催货物的在途状态，确保货物如期交货。对于晶圆和委外封测的采购，运营工程部负责监控供应商（代工厂）良率情况，发现异常及时分析处理，形成书面记录。超过规定信用期限未发货的，采购人员要立即催发货物；对于超过货物到期日仍未发货且无反馈信息的，对接的采购人员需要对逾期欠货供应商的名称、金额、原因等情况做出说明，报采购部门负责人和财务部门审阅、备案，情况严重的逐级提交至总经理审批。

4) 预付款追踪管理：针对预付款，需进行追踪管理业务部门定期向供应商进行对账，将全部拖欠货物按金额大小、拖欠时间长短排序分析，制订催收计划；针对逾期的款项应与债务人协商谈判，并视情况采取资产调查、诉前准备等行动；逾期半年以上的预付账款，至少每月催收一次；对于账龄在三年以上的预付账款，应成立清收小组通过法律诉讼和申请仲裁等措施积极催收。

综上，虽然发行人预付款项金额较大，但发行人已建立了安全有效的资金管

理内控机制，通过执行严格的供应商准入与评价、预付款项审批流程及货物交期及预付款追踪管理，有效保障了预付款项的安全性。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、申报会计师执行了如下核查程序：

1、获取并核查发行人的采购明细表，区分板卡/服务器梳理统计不同细分产品对应的采购类别、金额、主要供应商。

2、获取并核查发行人的采购明细表及预付款项明细表，结合行业特征分析采购主要原材料需要预付大额款项的具体原因及合理性；结合报告期各期采购总额、预付款项金额及经营性现金流量情况分析三者的匹配性，结合同行业可比公司情况分析相关采购模式是否符合行业惯例，预付款项规模及占比与同行业可比公司是否存在重大差异。

3、对发行人预付大额款项采购款供应商实施函证程序并获取回函。

4、访谈采购部门人员，了解相关预付大额款项的交易背景，对其合理性及必要性进行分析，了解大额预付合同的执行进度及后续执行安排，分析是否存在无法回收风险。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、发行人已区分板卡/服务器列示不同细分产品对应的采购类别、金额、主要供应商。

2、发行人已说明采购主要原材料需要预付大额款项的具体原因及合理性；相关采购模式符合行业惯例，预付款项金额与经营性现金流、采购金额具备匹配性，预付款项规模及占比与同行业可比公司不存在重大差异。

3、发行人已建立安全有效的资金管理内控机制，有效保障了预付款项的安全性。

8.关于成本和毛利率

根据申报材料：（1）报告期内，主营业务成本主要包括直接材料、封装测试、板卡加工、制造费用等成本；（2）报告期内，发行人综合毛利率为 24.10%、62.88%、53.43%和 55.26%，高于同行业可比公司均值；（3）发行人产品矩阵包括用于智算推理的曦思 N 系列 GPU 产品（曦思 N100 系列、曦思 N260 系列、曦思 N300 系列），用于训推一体和通用计算的曦云 C 系列 GPU 产品（曦云 C500 系列、曦云 C600 系列），以及用于图形渲染的曦彩 G 系列 GPU 产品（曦彩 G100 系列）；（4）训推一体 GPU 板卡各期毛利率为 65.48%、63.50%和 56.21%，训推一体 GPU 服务器为 31.70%和 43.59%，智算推理 GPU 板卡为 24.10%、5.37%、10.09%和 6.37%；（5）报告期内，直销客户、经销客户毛利率波动较大。

请发行人披露：（1）分产品类型列示各期细分产品主营业务成本的明细构成，不同细分产品的成本构成比重及变动趋势是否存在较大差异并分析原因；（2）分型号列示各期细分产品的销售收入金额及占比、平均销售单价、毛利率，分析毛利率的变动原因及合理性，2023 年至 2025 年 1-3 月综合毛利率高于可比公司均值的原因及合理性；（3）2024 年经销毛利率高于直销的具体原因及合理性，结合产品销售结构、客户等方面的差异，说明直销和经销模式各期毛利率均大幅波动的原因及合理性，与同行业可比公司是否一致。

请保荐机构、申报会计师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）分产品类型列示各期细分产品主营业务成本的明细构成，不同细分产品的成本构成比重及变动趋势是否存在较大差异并分析原因

报告期各期，发行人主要细分产品主营业务成本明细构成、比重及变动情况如下：

单位：元

产品类型	成本构成	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
		单位成本	占比	单位成本	占比	单位成本	占比	单位成本	占比
曦云	直接材料	13,508.90	80.16%	13,175.57	79.41%	15,160.72	77.22%	-	-

产品类型	成本构成	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
		单位成本	占比	单位成本	占比	单位成本	占比	单位成本	占比
C500 GPU 板卡	封装测试	1,094.31	6.49%	1,079.40	6.51%	1,204.06	6.13%	-	-
	板卡加工	1,711.87	10.16%	1,691.38	10.19%	1,906.96	9.71%	-	-
	制造费用	506.04	3.00%	574.64	3.46%	708.23	3.61%	-	-
	其他	31.62	0.19%	70.42	0.42%	652.54	3.32%	-	-
	合计	16,852.74	100.00%	16,591.41	100.00%	19,632.51	100.00%	-	-
曦云 C550 GPU 板卡	直接材料	14,067.42	74.39%	13,229.81	73.59%	-	-	-	-
	封装测试	1,097.04	5.80%	1,080.62	6.01%	-	-	-	-
	板卡加工	3,195.09	16.90%	3,118.31	17.35%	-	-	-	-
	制造费用	507.31	2.68%	502.87	2.80%	-	-	-	-
	其他	43.70	0.23%	45.44	0.25%	-	-	-	-
	合计	18,910.56	100.00%	17,977.04	100.00%	-	-	-	-
曦云 C500 GPU 服务器	直接材料	152,473.70	91.73%	337,911.61	92.38%	-	-	-	-
	封装测试	4,322.76	2.60%	8,410.80	2.30%	-	-	-	-
	板卡加工	6,763.33	4.07%	13,205.84	3.61%	-	-	-	-
	制造费用	1,998.99	1.20%	5,250.25	1.44%	-	-	-	-
	其他	670.36	0.40%	997.74	0.27%	-	-	-	-
	合计	166,229.15	100.00%	365,776.24	100.00%	-	-	-	-
曦云 C550 GPU 服务器	直接材料	381,080.74	90.60%	463,922.62	91.72%	-	-	-	-
	封装测试	8,645.52	2.06%	8,645.52	1.71%	-	-	-	-
	板卡加工	25,179.80	5.99%	25,179.80	4.98%	-	-	-	-
	制造费用	3,997.98	0.95%	3,997.98	0.79%	-	-	-	-
	其他	1,696.18	0.40%	4,052.26	0.80%	-	-	-	-
	合计	420,600.22	100.00%	505,798.18	100.00%	-	-	-	-
曦思 N100 GPU 板卡	直接材料	1,160.88	50.79%	2,518.51	45.98%	3,016.67	47.30%	3,403.89	54.69%
	封装测试	266.83	11.67%	764.81	13.96%	866.31	13.58%	721.71	11.60%
	板卡加工	384.65	16.83%	977.55	17.85%	1,045.05	16.39%	919.40	14.77%
	制造费用	463.92	20.30%	1,179.00	21.52%	1,335.47	20.94%	1,179.00	18.94%
	其他	9.35	0.41%	37.90	0.69%	114.39	1.79%	-	-
	合计	2,285.64	100.00%	5,477.77	100.00%	6,377.89	100.00%	6,224.00	100.00%
曦思 N260 GPU 板卡	直接材料	12,568.96	80.97%	13,803.83	83.13%	-	-	-	-
	封装测试	1,080.69	6.96%	1,077.98	6.49%	-	-	-	-
	板卡加工	1,311.24	8.45%	1,303.48	7.85%	-	-	-	-

产品类型	成本构成	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
		单位成本	占比	单位成本	占比	单位成本	占比	单位成本	占比
	制造费用	499.75	3.22%	354.73	2.14%	-	-	-	-
	其他	62.60	0.40%	66.03	0.40%	-	-	-	-
	合计	15,523.25	100.00%	16,606.05	100.00%	-	-	-	-

1、不同细分产品的成本构成比重差异分析

由上表所示，发行人各细分产品成本构成比重差异主要体现在如下方面：

（1）报告期内，曦云 C 系列板卡直接材料占比整体在 80%左右，曦思 N100 板卡直接材料占比整体在 50%左右，曦云 C 系列板卡产品的直接材料占比高于曦思 N100 系列，主要原因为曦云 C 系列芯片具有更高的设计复杂度、更先进的制程工艺和更大的封装尺寸，晶圆采购成本高于曦思 N100。此外，一颗曦云 C 系列芯片配置 4 颗 HBM，而一颗曦思 N100 芯片配置 1 颗 HBM，因此曦云 C 系列板卡的 HBM 采购成本也高于曦思 N100 芯片。由于曦思 N260 芯片与曦云 C 系列均基于相同的生产工艺制程，因此其直接材料占比与曦云 C 系列更为接近。

（2）曦云 C550 板卡与曦云 C500 板卡相比，直接材料成本占比较低、板卡加工成本占比较高，主要原因为曦云 C550 板卡和曦云 C500 板卡在芯片制造环节的生产制程工艺基本相同，因此两者的直接材料成本、封装测试成本、制造费用无明显差异。但由于曦云 C550 板卡在板卡加工环节采用定制化 OAM 模组，板卡加工环节的委外成本高于曦云 C500 板卡，板卡加工成本占比较高，更高的板卡加工成本间接使得直接材料等其他成本相对占比更低。

（3）曦云 C500 服务器、曦云 C550 服务器的直接材料成本占比均高于相应的板卡产品，主要原因为服务器成本中除板卡成本外，还包括机头、CPU、内存等服务器其他硬件的采购成本，该等硬件采购成本均属于直接材料成本，进而使得 GPU 服务器直接材料成本相对更高。

2、同一细分产品报告期各期主要成本类型变动趋势分析

（1）训推一体 GPU 板卡

整体来看，报告期内曦云 C500 和曦云 C550 板卡各成本类型占比的波动较小。2023 年，曦云 C 系列板卡处于风险量产阶段，晶圆采购量较小且产品良率

相对较低，单位直接材料成本较高；此外，因当期曦云 C 系列产量较小，封装测试、板卡加工等环节的委外采购量较小，单位成本更高。2024 年度，随着曦云 C 系列板卡正式量产，产品良率提升且产量大幅增长，各项成本均呈下降趋势。2025 年 1-3 月，曦云 C 系列板卡单位成本略有上升，主要原因为当期销售的产品芯片均在 2024 年完成生产，一方面 2024 年内曦云 C 系列主要原材料 HBM 受人工智能芯片需求驱动价格上升，另一方面 2024 年美元兑人民币汇率走高，曦云 C 系列以美元计价的晶圆、HBM 采购成本略有上升，进而使得单位直接材料成本略有上升。

（2）智算推理 GPU 板卡

2022-2024 年，曦思 N100 板卡单位直接材料成本及占比均呈下降趋势，其他各类型成本占比波动幅度较小，主要原因为曦思 N100 产品 2023 年 4 月正式量产后，由于采购量增加、产品良率上升，曦思 N100 晶圆单位直接材料成本下降。2025 年 1-3 月，曦思 N100 板卡直接材料成本占比有所上升，主要原因为当期销售的曦思 N100 板卡主要来源于单个客户的单笔订单，发行人根据该客户的需求对该批出售的曦思 N100 板卡进行了部分定制化，成本结构与 2022-2024 年主要销售的型号略有不同。2024 年和 2025 年 1-3 月，曦思 N260 产品销量较小，且各类型成本比重差异较小。

（3）训推一体 GPU 服务器

2024 年和 2025 年 1-3 月，训推一体 GPU 服务器类产品各类型成本比重差异较小。2025 年 1-3 月各项单位成本变动较大，主要系 2025 年 1-3 月销售的 GPU 服务器数量较少，且服务器配置不同所致。

综上所述，报告期内发行人不同产品之间成本构成的差异主要系芯片尺寸、设计复杂程度、制造工艺等因素导致，具有合理性；同一产品在报告期内成本构成比重的变动主要系原材料采购价格波动、产量上升带来规模效应等因素导致，差异较小且具有合理性。

（二）分型号列示各期细分产品的销售收入金额及占比、平均销售单价、毛利率，分析毛利率的变动原因及合理性，2023 年至 2025 年 1-3 月综合毛利率高于可比公司均值的原因及合理性

1、各期细分产品的销售收入金额及占比、平均销售单价、毛利率及其变动原因分析

报告期各期，发行人各细分产品毛利率的总体情况如下：

产品类型	2025 年 1-3 月	2024 年	2023 年	2022 年
训推一体 GPU 板卡	56.21%	63.50%	65.48%	-
其中：曦云 C500	54.44%	65.32%	65.48%	-
其中：曦云 C550	68.42%	60.50%	-	-
训推一体 GPU 服务器	43.59%	31.70%	-	-
其中：曦云 C500	52.51%	42.43%	-	-
其中：曦云 C550	39.07%	11.70%	-	-
智算推理 GPU 板卡	6.37%	10.09%	5.37%	24.10%
其中：曦思 N100	-0.44%	6.08%	5.37%	24.10%
其中：曦思 N260	26.91%	26.70%	-	-

（1）训推一体 GPU 板卡

1）曦云 C500 板卡

报告期内，曦云 C500 板卡毛利率变动的影响因素分析如下：

项目	2025 年 1-3 月	2024 年度	2023 年度
当期毛利率	54.44%	65.32%	65.48%
毛利率较上期变动	-10.88%	-0.16%	-
平均销售单价对毛利率影响	-10.18%	-6.51%	-
平均单位成本变动对毛利率的影响	-0.71%	6.36%	-

注 1：平均销售单价变动对毛利率的影响=（当期单位价格-上期单位成本）/当期单位价格-上期毛利率，下同

注 2：平均单位成本变动对毛利率的影响=当期毛利率-（当期单位价格-上期单位成本）/当期单位价格，下同

2023 年度、2024 年度和 2025 年 1-3 月，发行人曦云 C500 系板卡的毛利率分别为 65.48%、65.32%和 54.44%。2024 年，曦云 C500 板卡的单位成本因产量提升、产品良率提升相较 2023 年有所下降，但当年平均销售单价下降幅度略高

于单位成本，致使 2024 年毛利率略有下降。2025 年 1-3 月，曦云 C500 板卡单位成本变动幅度较小，毛利率的下降主要系销售单价的下降所致。

2) 曦云 C550 板卡

报告期内，曦云 C550 板卡毛利率变动的影响因素分析如下：

项目	2025 年 1-3 月	2024 年度
当期毛利率	68.42%	60.50%
毛利率较上期变动	7.92%	-
平均销售单价对毛利率影响	9.48%	-
平均单位成本变动对毛利率的影响	-1.56%	-

2024 年度和 2025 年 1-3 月，发行人曦云 C550 板卡毛利率分别为 60.50%、68.42%，平均销售单价对毛利率的影响为 9.48%，平均单位成本变动对毛利率的影响为-1.56%，曦云 C550 板卡的毛利率与曦云 C500 呈现出相同的波动特征，即毛利率波动主要受单位售价的影响，2024 年发行人给予曦云 C550 板卡的部分大额采购客户给予一定优惠措施，导致平均单位售价较低。

(2) 训推一体 GPU 服务器

1) 曦云 C500 服务器

报告期内，曦云 C500 服务器毛利率变动的影响因素分析如下：

项目	2025 年 1-3 月	2024 年度
当期毛利率	52.51%	42.43%
毛利率较上期变动	10.08%	-
平均销售单价对毛利率影响	-46.94%	-
平均单位成本变动对毛利率的影响	57.01%	-

2024 年度和 2025 年 1-3 月，发行人曦云 C500 服务器毛利率分别为 42.43%、52.51%，2025 年 1-3 月发行人销售 1 台曦云 C500 服务器，为含 4 张曦云 C500 板卡的工作站，与 2024 年主要销售的曦云 C500 服务器（8 卡）在整机配置、含卡数量等方面均存在较大差异，因此单价和单位成本均较低。

2) 曦云 C550 服务器

报告期内，曦云 C550 服务器毛利率变动的影响因素分析如下：

项目	2025 年 1-3 月	2024 年度
当期毛利率	39.07%	11.70%
毛利率较上期变动	27.37%	-
平均销售单价对毛利率影响	15.02%	-
平均单位成本变动对毛利率的影响	12.34%	-

2024 年度和 2025 年 1-3 月，发行人曦云 C550 服务器毛利率分别为 11.70%、39.07%，2025 年 1-3 月发行人仅销售 1 台曦云 C550 服务器。2024 年发行人实现销售的 128 台曦云 C550 服务器系以集群整体进行交付，除 GPU 服务器外发行人还需交付网络设备、存储设备等其他集群设备，使得分摊到单台 GPU 服务器的单位成本较高；此外，销售合同部分价款尚不满足确认条件，发行人暂未确认相关收入，导致单位销售价格较低。

（3）智算推理 GPU 板卡

1) 曦思 N100 板卡

报告期内，曦思 N100 板卡毛利率变动的影响因素分析如下：

项目	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
当期毛利率	-0.44%	6.08%	5.37%	24.10%
毛利率较上期变动	-6.52%	0.70%	-18.72%	-
平均销售单价对毛利率影响	-143.26%	-14.73%	-16.44%	-
平均单位成本变动对毛利率的影响	136.74%	15.43%	-2.28%	-

报告期各期，曦思 N100 板卡毛利率分别为 24.10%、5.37%、6.08%和-0.44%，曦思 N100 产品是发行人推出的首款产品，其研发和量产过程为后续曦云 C 系列产品的成功量产奠定了基础。2024 年传统人工智能推理芯片市场需求发生变化，发行人研发、销售的重心转移至训推一体系列产品，进而调整了曦思 N100 产品的定价以促进库存的出清。2024 年末，发行人对曦思 N100 板卡相关的存货计提了减值准备，致使 2025 年一季度单位成本因存货跌价转销而大幅降低。报告期内，曦思 N100 板卡单位成本整体呈下降趋势，平均单价的波动对毛利率的影响较大。

2) 曦思 N260 板卡

报告期内，曦思 N260 板卡毛利率变动的影响因素分析如下：

项目	2025 年 1-3 月	2024 年度
当期毛利率	26.91%	26.70%
毛利率较上期变动	0.21%	-
平均销售单价对毛利率影响	-4.89%	-
平均单位成本变动对毛利率的影响	5.10%	-

2024 年度和 2025 年 1-3 月，曦思 N260 板卡毛利率分别为 26.70%和 26.91%，曦思 N260 产品与曦云 C500 系列基于相同的生产工艺和制程，其单位成本与曦云 C500 系列较为接近，但由于其不具备互联功能，不具备训练功能，因此其定价低于曦云 C500 系列。报告期内曦思 N260 板卡销量较小，平均单价、单位成本和毛利率变动幅度较小。

2、2023 年至 2025 年 1-3 月综合毛利率高于可比公司均值的原因及合理性

报告期内，公司与同行业可比上市公司综合毛利率比较情况如下：

可比上市公司	2025 年 1-3 月	2024 年	2023 年	2022 年
寒武纪	55.99%	56.71%	69.16%	65.76%
海光信息	61.19%	63.72%	59.67%	52.42%
龙芯中科	38.00%	31.04%	36.06%	47.09%
景嘉微	32.37%	43.70%	60.32%	65.01%
摩尔线程	未披露	70.71%	25.87%	-70.08%
可比公司均值	46.89%	53.18%	50.22%	32.04%
发行人	55.26%	53.43%	62.88%	24.10%

注：同行业可比公司数据来源于定期报告、招股说明书等公开披露文件

2023 年、2024 年和 2025 年 1-3 月，发行人综合毛利率分别为 62.88%、53.43%和 55.26%，高于可比公司平均值，主要原因如下：

第一，发行人可比公司龙芯中科主营业务为处理器及配套芯片的研制、销售及服务，主要产品为工控类芯片、信息化类芯片及解决方案业务。龙芯中科的产品类型和收入结构与发行人存在较大差异。龙芯中科主营产品中毛利率相对较低的信息化类芯片收入占比提升，导致报告期内龙芯中科毛利率整体低于其他三家可比公司，拉低了发行人同行业可比公司毛利率的平均值。

第二，发行人可比公司景嘉微的主营产品主要包括图形显控领域产品、小型

专用化雷达领域产品以及芯片领域产品。发行人与景嘉微在收入结构、产品类型和客户类型方面存在一定差异。具体而言，景嘉微的芯片领域产品主要为图形处理芯片，但在其收入结构中占比相对较低。此外，景嘉微的主要客户类型为军工集团下属单位，与发行人的客户类型存在较大差异。根据景嘉微 2024 年年报披露，受客户价格审核、成本管控、持续投入研发及市场竞争加剧等因素影响，其毛利率有所下降。因此，发行人的毛利率与景嘉微不具有较强的可比性。

同行业可比公司中，寒武纪主营产品中的云端智能芯片及板卡、智能整机，海光信息主营产品中的 DCU 产品与发行人训推一体 GPU 板卡具有较强相似性。2023 年度、2024 年度和 2025 年 1-3 月，寒武纪和海光信息综合毛利率平均值分别为 64.42%、60.22%和 58.59%。同时期发行人综合毛利率分别为 62.88%、53.43%和 55.26%，2024 年低于寒武纪和海光信息综合毛利率平均值，主要原因为发行人当期毛利率相对较低的 GPU 服务器销售占比较大，2023 年度和 2025 年 1-3 月，发行人综合毛利率略低于寒武纪和海光信息综合毛利率平均值，但整体差异较小。

综上所述，因同行业公司龙芯中科、景嘉微在收入结构、产品类型、客户类型等方面与发行人存在较大差异，导致发行人在 2023 年至 2025 年 1-3 月综合毛利率高于同行业可比公司。发行人综合毛利率略低于与主营业务与发行人更为接近的寒武纪、海光信息，差异较小且差异原因具有合理性。

（三）2024 年经销毛利率高于直销的具体原因及合理性，结合产品销售结构、客户等方面的差异，说明直销和经销模式各期毛利率均大幅波动的原因及合理性，与同行业可比公司是否一致

1、发行人不同销售模式下产品销售结构、毛利率及变动分析

报告期各期，发行人直销/经销毛利率总体情况如下：

项目	产品类型	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
		收入占比	毛利率	收入占比	毛利率	收入占比	毛利率	收入占比	毛利率
直销	训推一体 GPU 板卡	98.47%	63.68%	52.75%	66.70%	-	-	-	-
	训推一体 GPU 服	-	-	43.57%	31.17%	-	-	-	-

项目	产品类型	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
		收入占比	毛利率	收入占比	毛利率	收入占比	毛利率	收入占比	毛利率
	务器								
	智算推理 GPU 板卡	0.02%	72.83%	0.35%	14.85%	23.27%	2.37%	100.00%	24.10%
	IP 授权	0.18%	100.00%	0.22%	100.00%	76.73%	100.00%	-	-
	其他	1.34%	26.61%	3.12%	26.52%	-	-	-	-
	合计	100.00%	63.25%	100.00%	49.86%	100.00%	77.28%	100.00%	24.10%
经销	训推一体 GPU 板卡	96.81%	50.17%	96.14%	60.56%	68.68%	65.48%	-	-
	训推一体 GPU 服务器	0.58%	43.59%	2.75%	45.67%	-	-	-	-
	智算推理 GPU 板卡	2.23%	5.99%	0.68%	5.99%	31.26%	8.24%	-	-
	其他	0.37%	-13.84%	0.44%	5.95%	0.06%	41.31%	-	-
	合计	100.00%	48.90%	100.00%	59.54%	100.00%	47.57%	-	-

报告期内，公司直销客户毛利率整体高于经销商客户，主要原因系经销客户通常承担了部分市场开拓和售后服务职能，公司给予其一定的让利。2024 年发行人直销毛利率低于经销毛利率，主要系当期直销、经销产品结构差异所致，具体分年度分析如下：

2022 年，发行人收入均为直销收入，无经销收入。

2023 年，发行人直销收入中毛利率较高的 IP 授权收入占比为 76.73%，显著提升了当期的直销毛利率；当期经销收入主要来自于 GPU 板卡，且其中毛利率较低的智算推理 GPU 板卡收入占比为 31.26%，导致当期经销毛利率相对较低。

2024 年，训推一体 GPU 板卡正式量产，发行人根据部分直销智算集群客户的需求交付训推一体 GPU 服务器，训推一体 GPU 服务器收入占比大幅上升，当期收入占直销收入的比例达到 43.57%，由于训推一体 GPU 服务器毛利率降低，导致当期直销毛利率相较 2023 年有所下降；此外，2024 年发行人向经销客户销售的主要为毛利率较高的训推一体 GPU 板卡产品，训推一体服务器销售收入占比较低，进而导致 2024 年经销毛利率高于直销毛利率，具有合理性。若不考虑

毛利率较低的 GPU 服务器业务，将 2024 年板卡类收入按直销、经销拆分如下：

单位：万元

项目	2024 年收入	2024 年占比	毛利率
直销-板卡类	24,655.44	33.23%	66.36%
直销-GPU 服务器及其他	21,776.50	29.35%	31.18%
经销-板卡类	26,875.03	36.22%	60.18%
经销-GPU 服务器及其他	883.90	1.19%	40.21%
主营业务收入合计	74,190.87	100.00%	53.48%

由上表可见，若不考虑毛利率较低的 GPU 服务器等业务，2024 年发行人直销-板卡类收入毛利率高于经销-板卡类收入毛利率，具有合理性。

2025 年 1-3 月，发行人直销收入与经销收入的产品结构较为相似，均以训推一体 GPU 板卡的销售为主，且直销收入中训推一体 GPU 服务器收入占比大幅降低，使得当期直销毛利率相较 2024 年度有所回升。由于 2025 年 1 季度发行人给予部分大额采购的经销客户较大力度的赠卡优惠，导致当期经销毛利率低于 2024 年，同时也低于当期直销毛利率。

综上所述，发行人报告期内直销和经销毛利率波动较大的主要原因系各期直销客户、经销客户各期收入结构发生较大变化所致，具有合理性。经销客户方面，因 2023 年经销客户收入结构中有相当比例毛利率较低的智算推理 GPU 板卡，而 2024 年经销客户收入结构中高毛利率的训推一体 GPU 板卡收入占比大幅提升，因此 2024 年经销毛利率有所上升；2025 年 1-3 月经销客户毛利率受大额客户赠卡优惠影响，毛利率相较 2024 年有所降低。直销客户方面，因 2023 年直销收入结构中高毛利的 IP 授权业务占比较大，而 2024 年直销客户收入结构中相对低毛利的 GPU 服务器销售业务占比上升，导致 2024 年直销毛利率下降；2025 年 1-3 月，发行人向直销客户销售的产品类型主要为训推一体 GPU 板卡，而 GPU 服务器的销售大幅下降，导致当期直销客户毛利率有所回升。

2、发行人直销/经销模式毛利率波动情况与同行业可比公司比较情况

报告期内，发行人直销模式、经销模式毛利率波动情况与同行业可比公司比较情况如下：

可比公司名称	2024 年		2023 年		2022 年	
	收入占比	毛利率	收入占比	毛利率	收入占比	毛利率
直销模式						
寒武纪	98.75%	56.87%	95.67%	69.87%	90.64%	65.85%
海光信息	84.79%	61.84%	92.73%	59.05%	89.72%	52.27%
龙芯中科	99.37%	30.86%	99.56%	36.01%	99.50%	47.15%
景嘉微	未披露	未披露	未披露	未披露	未披露	未披露
摩尔线程	40.97%	未披露	61.22%	未披露	1.44%	未披露
可比公司平均值	94.30%	49.86%	95.99%	54.98%	93.29%	55.09%
发行人	62.58%	49.86%	56.19%	77.28%	100.00%	24.10%
经销模式						
寒武纪	1.25%	44.43%	4.33%	56.81%	9.36%	65.87%
海光信息	15.21%	74.18%	7.27%	67.57%	10.28%	53.70%
龙芯中科	0.63%	58.15%	0.44%	43.65%	0.50%	51.81%
景嘉微	未披露	未披露	未披露	未披露	未披露	未披露
摩尔线程	59.03%	未披露	38.78%	未披露	98.56%	未披露
可比公司平均值	5.70%	58.92%	4.01%	56.01%	6.71%	57.13%
发行人	37.42%	59.54%	43.81%	47.57%	不适用	不适用

注：同行业可比公司数据来源于定期报告、招股说明书等公开披露文件，同行业可比公司均未披露 2025 年 1-3 月直销、经销毛利率数据

如上表所示，发行人 2023 年直销毛利率高于同行业可比公司，主要原因为当期发行人 IP 授权直销客户毛利率较高，经销毛利率低于同行业可比公司，主要原因为当期向经销客户销售毛利较低的智算推理 GPU 板卡占比较高。2024 年，发行人直销和经销毛利率与同行业可比公司均值较为接近。总体来看，同行业可比公司虽然同属于芯片设计行业，但由于主营产品、经营模式存在不同，直销和经销收入占比与发行人存在较大差异，其毛利率水平亦存在差异且波动趋势无明显规律。此外，同行业可比公司报告期内普遍存在经销毛利率高于直销毛利率的情形，发行人 2024 年度因直销、经销客户收入结构不一致导致的经销毛利率高于直销毛利率具有合理性。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、申报会计师执行了如下核查程序：

1、获取发行人收入成本明细表，分析发行人不同细分产品的成本构成情况及在报告期内的变动情况；获取发行人标准成本 BOM 表，了解不同细分产品成本构成的差异；向发行人了解不同细分产品成本构成差异的原因及变动原因。

2、获取发行人收入成本明细表，按细分产品类型、主要客户等各维度对销售单价、单位成本、毛利率进行分析，向发行人了解毛利率波动的原因；查阅同行业可比公司公开披露信息，了解同行业可比公司主营业务和产品结构，分析发行人与同行业可比公司毛利率差异的原因。

3、获取发行人收入成本明细表，按直销、经销维度分析产品结构、主要客户对直销、经销毛利率差异的影响；查阅同行业可比公司直销和经销毛利率波动情况。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、报告期内，发行人不同细分产品成本结构差异主要系不同产品生产工艺制程、设计复杂度、芯片尺寸、产品规格等因素导致，具有合理性；报告期内，同一细分产品成本构成比重有小幅波动，主要系原材料采购成本波动、产品良率提升、产量提升带来的规模效应所致，具有合理性，不存在有较大差异的情形。

2、报告期内，发行人主要产品毛利率的变动主要受平均销售单价的影响，毛利率的变动具有合理性；发行人 2023 年至 2025 年 1-3 月综合毛利率高于同行业可比公司平均值主要系部分可比公司毛利率较低所致，发行人毛利率水平与业务结构、产品类型更为类似的寒武纪、海光信息不存在重大差异。

3、发行人 2024 年经销毛利率高于直销毛利率以及报告期内直销、经销毛利率的波动主要系报告期各期直销、经销产品结构不同所致，同行业可比公司直销、经销毛利率波动无明显规律，发行人直销、经销毛利率与同行业可比公司的差异也系产品结构不同所致，具有合理性。

9.关于股东股权

9.1 关于实际控制人

根据申报材料：（1）实控人陈维良直接持有发行人 5.59%的股份，通过持股平台上海骄迈、上海曦骥合计控制发行人 22.94%的股权。发行人董事会由 6 名非独立董事及 3 名独立董事组成，陈维良拥有 4 席非独立董事提名权；（2）2020 年 9 月、10 月，高崧、炫桥资本（任玉会）分别与陈维良签署股权代持协议，陈维良系代持方。2025 年 2 月予以还原；（3）发行人设立及融资过程中，曾与投资方签订对赌协议，并于 2025 年 4 月通过补充协议解除。其中，公司创始人（指陈维良、彭莉和杨建）、创始股东（指陈维良和上海骄迈）作为回购义务人的回购义务终止后附条件可恢复。

请发行人披露：（1）结合陈维良所持股份、上市后股权稀释情况、对上海骄迈和上海曦骥的控制稳定性、拥有的董事会席位等，分析陈维良的控制权稳定性、具体保障措施；（2）高崧、任玉会与陈维良的关系、代持形成的背景，代持还原是否存在纠纷或者潜在纠纷；（3）结合发行人与股东签署的补充协议中，特殊权利条款内容及附条件恢复具体约定，陈维良所持公司股权比例等，说明相关特殊权利条款的清理是否符合《监管规则适用指引——发行类第 4 号》对赌协议的相关规定，对控制权稳定性的影响。

请保荐机构、发行人律师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）结合陈维良所持股份、上市后股权稀释情况、对上海骄迈和上海曦骥的控制稳定性、拥有的董事会席位等，分析陈维良的控制权稳定性、具体保障措施

1、陈维良所持股份、上市后股权稀释情况、对上海骄迈和上海曦骥的控制稳定性、拥有的董事会席位

（1）报告期初至今，陈维良所持股份及对应表决权比例稳定

报告期初至今，在公司历次融资完成后，陈维良所持公司股份的情况如下：

序号	时间点（注）	陈维良持股比例	上海骄迈持股比例	上海曦骥持股比例	合计持股比例
1	报告期初（2022 年 1 月 1 日）	13.45%	31.38%	9.55%	54.39%
2	Pre-B 轮融资第一期（2022 年）	12.52%	29.20%	8.89%	50.60%
3	Pre-B 轮融资第二期（2023 年）	12.14%	28.33%	8.62%	49.10%
4	Pre-B 轮融资第三期（2023 年）	11.91%	27.78%	8.45%	48.14%
5	Pre-B 轮融资第四期、B 轮融资第一期（2023 年）	10.79%	25.18%	7.66%	43.64%
6	B 轮融资第二期至第四期及第二次股权转让（2024 年）	9.60%	22.83%	6.95%	39.38%
7	B 轮融资第五期（2025 年）	8.51%	20.23%	6.16%	34.89%
8	C 轮融资（2025 年）	7.42%	17.64%	5.37%	30.43%
9	Pre-IPO 轮融资（2025 年）	5.59%	13.30%	4.05%	22.94%

注：该时间点指该轮次/期次融资或股权转让完成的时间点及其后的期间，以列示该时间段内陈维良、上海骄迈、上海曦骥的直接持股情况以及陈维良及上海骄迈、上海曦骥合计持股比例

如上表列示，报告期初至今，受多轮融资影响，陈维良及其控制的上海骄迈、上海曦骥合计持有发行人股份的比例逐渐稀释降低，但此前仍持续维持在 30% 以上，且一直为持有发行人股份比例最高的股东，直至 Pre-IPO 轮融资完成，该比例稀释到 30% 以下，但合计仍为持股比例最高的股东，且发行人的其他股东持股比例较低且分散。

（2）本次发行上市后，陈维良及其控制的持股平台合计持股比例仍最高

根据发行人本次发行上市方案，公司本次拟公开发行不超过人民币普通股 4,010 万股股票，按本次发行 4,010 万股股票计算，本次发行上市完成后，陈维良及其控制发行人股份及对应表决权比例的具体情况如下：

序号	股东姓名/名称	持股数量（股）	持股比例
一、陈维良及其控制的持股平台			
1	陈维良	20,129,832	5.03%
2	上海骄迈	47,871,181	11.96%
3	上海曦骥	14,567,737	3.64%

序号	股东姓名/名称	持股数量（股）	持股比例
合计		82,568,750	20.64%
二、其他股东			
1	葛卫东及混沌投资	26,937,920	6.73%
2	经乾二号、经纬创壹号、经纬创荣、经纬厦门、经纬创叁号、经纬创华	18,451,140	4.61%
3	和利国信智芯、南京智沐曦及浦口高科	17,425,988	4.36%
4	红杉瀚辰、红杉雅恒	15,066,795	3.77%
5	引领区基金	9,659,452	2.41%
6	其他 107 名机构投资人股东、自然人股东等	189,889,955	47.46%
三、本次发行公众投资者（按本次发行 4,010 万股计算）			
1	本次发行公众投资者	40,100,000	10.02%
合计		400,100,000	100.00%

本次发行上市后，陈维良及其控制的持股平台合计所持公司股份的比例仍最高，其他股东的持股比例亦同步被稀释，公司股权结构仍较分散，且其他单一股东（包括与其一致行动人）的持股比例与其差距仍较大，本次发行未导致公司股本结构发生重大变化，不会导致陈维良对发行人的控制权发生变化。

（3）陈维良对上海骄迈、上海曦骥具有持续稳定的控制权

上海骄迈、上海曦骥合伙协议中关于执行事务合伙人选任、职权以及更换等相关条款约定基本一致，具体如下：

事项	约定内容
执行事务合伙人选任条件及合伙企业事务执行	第十二条 有限合伙企业由普通合伙人执行合伙事务。
执行事务合伙人的职责和权限	第十五条 执行事务合伙人制定本合伙企业的规章制度，对外代表合伙企业负责本企业的日常经营管理，全体合伙人在此特别同意并不可撤销的授权执行事务合伙人可以对下列事项拥有独立决定权： （一） 改变合伙企业的名称； （二） 改变合伙企业的经营范围、主要经营场所的地点； （三） 执行合伙企业日常事务，办理合伙企业经营过程中相关审批手续； （四） 决定合伙企业对外签订法律文件并代表合伙企业履行； （五） 代表合伙企业处理、解决合伙企业涉及的各种争议和纠纷； （六） 保管及管理本合伙企业所有档案与账簿、合伙人名录等； （七） 代表合伙企业处理合伙企业投资企业的管理、退出、收益分配等事务； （八） 在不违反《合伙企业法》规定的前提下，决定合伙企业的收益分配方案；

事项	约定内容
	(九) 有限合伙人入伙、退伙、持有合伙份额、权益份额及比例变更或合伙企业认缴出资额的增加或减少； (十) 依据所投资企业的增资协议、股东协议、公司章程等组织性文件中的约定，行使本合伙企业的股东权利（包括但不限于投后管理的权利，行使股东表决权及委派或提名董事、监事或高级管理人员），承担股东义务； (十一) 处分合伙企业的不动产； (十二) 转让或者处分合伙企业的知识产权和其他财产权利； (十三) 以合伙企业名义或合伙企业的财产为他人提供担保； (十四) 聘任合伙人或合伙人以外的人担任合伙企业的经营管理人员； (十五) 延长合伙企业的经营期限； (十六) 在合伙企业经营期限内，决定延长全部或部分合伙人在本协议项下的合伙企业认缴出资缴付期限； (十七) 决定将合伙企业收到实缴出资留存部分或对外交易收益用于支付合伙企业的运营、管理及变更、税收缴纳等； 法律法规、本协议或公司与合伙人签署的其他文件中规定的其他职权。 第十六条 全体合伙人一致授权并同意执行事务合伙人行使以下职权： (一) 上述第十五条所述事项变更时涉及市场监督管理部门变更需要签署的相关协议及登记资料，由全体合伙人授权的执行事务合伙人签署。 (二) 上述第十五条所述事项变更时涉及市场监督管理部门变更需要签署的入伙、退伙协议等全部变更登记所需文件（包括但不限于变更决定书、入伙协议、退伙协议、合伙人名册、合伙协议等）由入伙合伙人或退伙合伙人与全体合伙人授权的执行事务合伙人签署。入伙合伙人签署上述变更登记所需文件时，即视为其接受本协议的约束。 (三) 上述第十五条所述事项变更时涉及市场监督管理部门变更需要签署的增资、减资等全部变更登记所需文件（包括但不限于变更决定书、合伙协议、合伙人名册等）由增资、减资的合伙人与全体合伙人授权的执行事务合伙人签署。 (四) 上述第十五条所述事项变更时涉及需要修改合伙协议相关内容，由全体合伙人授权执行事务合伙人单独签署合伙协议补充协议或修改后的合伙协议。
合伙份额转让与新合伙人入伙	第二十二条 合伙人向合伙人以外的人转让其在合伙企业中的全部或者部分合伙份额时，须经执行事务合伙人同意。合伙人之间或合伙人向合伙人以外的人转让在合伙企业中的全部或者部分合伙份额时，在同等条件下，执行事务合伙人或其指定的第三方有优先购买权，其他合伙人 not 享有优先购买权。 第二十三条 经执行事务合伙人同意，合伙企业可以接纳新合伙人入伙，并依法订立书面入伙协议。
普通合伙人(执行事务合伙人)的更换及身份免除	第二十条 执行事务合伙人违反本协议或因故意或重大过失给本合伙企业造成损失或有其他不正当行为且经有权法院或仲裁机构作出裁判或裁决时，经其他合伙人一致决定，可以将其除名，并更换新的执行事务合伙人。 第二十六条 普通合伙人被依法认定为无民事行为能力人或者限制民事行为能力人的，经其他合伙人一致同意，可以依法转为有限合伙人，其他合伙人未能一致同意的，该无民事行为能力或者限制民事行为能力的合伙人退伙。退伙事由实际发生之日为退伙生效日。
有限合伙人与普通合伙人互换程序	第二十八条 有限合伙人和普通合伙人的相互转变程序按照《合伙企业法》有关规定办理。普通合伙人转变为有限合伙人，或者有限合伙人转变为普通合伙人，应当经全体合伙人一致同意。

事项	约定内容
合伙期限	长期

陈维良作为普通合伙人，依据合伙协议的约定担任执行事务合伙人，对上海骄迈、上海曦骥的重要事项拥有独立决定权，包括行使上海骄迈、上海曦骥的股东权利(包括不限于行使股东表决权及委派或提名董事、监事或高级管理人员)，承担股东义务。作为执行事务合伙人，陈维良对合伙人向合伙人以外的人转让份额具有一票否决权，并且享有同等条件下的优先购买权，同时对新合伙人入伙具有独立决定权，能够保障合伙企业的出资人及出资额结构的稳定。此外，仅在执行事务合伙人出现合伙协议约定的重大负面情形且经有权法院或仲裁机构作出裁判或裁决，并经其他合伙人一致决定时，执行事务合伙人才可以被除名和更换，因此，在合伙企业存续期间，陈维良的执行事务合伙人身份能够保持稳定。

基于上述，陈维良作为合伙企业的执行事务合伙人，在合伙企业长期存续期间，依据合伙协议的约定，拥有对上海骄迈、上海曦骥持续稳定的控制权。

(4) 陈维良可以控制的董事会席位

根据公司历次融资相关投资协议、公司章程以及股东会会议文件等资料，报告期初至今，陈维良可以控制的董事会席位情况如下：

期间	董事会成员数	陈维良可以控制的席位	备注
报告期初至2025年3月	5	3	/
2025年3月至今	9 (含3名独立董事)	4	股份有限公司第一届董事会的任期为自2024年12月27日起三年

2025年3月10日，发行人召开2025年第五次临时股东会，审议通过了《关于制定<沐曦集成电路（上海）股份有限公司董事会成员提名规则>的议案》，根据《沐曦集成电路（上海）股份有限公司董事会成员提名规则》规定，董事会成员数量由5人增加至9人，实际控制人陈维良有权提名4位董事，以保证实际控制权的稳定。

因此，报告期初至今，陈维良可以控制的董事会席位始终占非独立董事席位总数的二分之一以上。

2、上市后控制权稳定的具体保障措施

(1) 实际控制人已作出长期维持控制权的安排

实际控制人陈维良及上海骄迈、上海曦骥已出具关于发行人股票锁定期的承诺、关于持股意向及减持意向的承诺函，具体承诺内容如下：

承诺主体	承诺函	承诺内容
陈维良	关于发行人股票锁定期的承诺函	1、本人将严格履行发行人首次公开发行股票招股说明书中披露的股份锁定承诺，自发行人股票在证券交易所上市之日起 36 个月内，不转让或委托他人管理本人直接或间接持有的发行人首次公开发行股票前已发行的股份，也不由发行人回购本人持有的上述股份。 2、在发行人实现盈利前，自发行人股票上市交易之日起 3 个完整会计年度内，不转让或者委托他人管理本人于本次发行上市前已直接或间接持有的发行人股份，也不提议由发行人回购该部分股份；自发行人股票上市交易之日起第 4 个会计年度和第 5 个会计年度内，每年减持的本人于本次发行上市前已直接或间接持有的发行人股份不超过发行人股份总数的 2%。在发行人实现盈利后，本人可以自发行人当年年度报告披露后次日起减持本人于本次发行上市前已直接或间接持有的发行人股份，但应当遵守本承诺函其他规定。 3、发行人上市当年较上市前一年净利润（以扣除非经常性损益后归母净利润为准，下同）下滑 50%以上的，延长本人届时所持股份锁定期限 6 个月；发行人上市第二年较上市前一年净利润下滑 50%以上的，在前项基础上延长本人届时所持股份锁定期限 6 个月；发行人上市第三年较上市前一年净利润下滑 50%以上的，在前两项基础上延长本人届时所持股份锁定期限 6 个月。前述“届时所持股份”分别指本人在发行人上市前取得，上市当年及之后第二年、第三年发行人年报披露时仍持有的股份。 4、在发行人股票上市后 6 个月内，如发行人股票连续 20 个交易日的收盘价均低于首次公开发行股票时的发行价，或者发行人股票上市后 6 个月期末（如该日不是交易日，则为该日后第 1 个交易日）收盘价低于首次公开发行股票时的发行价的，本人在发行人首次公开发行股票前所持有的发行人股份的锁定期限自动延长至少 6 个月。 5、如本人在发行人首次公开发行股票前直接或者间接所持有的发行人股份在锁定期满后 2 年内减持的，减持价格不低于首次公开发行股票时的发行价。若因发行人派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，上述股份价格、股份数量按照证券交易所的有关规定作相应调整。 6、本人担任发行人董事/高级管理人员期间，每年转让的股份不超过本人所持有的发行人股份总数的 25%，本人在离职后 6 个月内，不转让本人所持有的发行人股份。如本人在任期届满前离职的，本人承诺在本人就任时确定的任期内和任期届满后 6 个月内，继续遵守上述对董事/高级管理人员股份转让的限制性规定。 7、本人将严格遵守法律、法规、规范性文件、公司上市的证券交易所业务规则中关于股东持股及股份变动（包括减持）的有关规定，规范诚信履行股东的义务。

承诺主体	承诺函	承诺内容
		8、在本人持股期间，若股份锁定和减持的法律、法规、规范性文件、政策及证券监管机构的要求发生变化，则本人将依据相关法律法规的要求适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。
上海骄迈与上海曦骥	关于发行人股票锁定期的承诺函	1、本企业将严格履行发行人首次公开发行股票招股说明书中披露的股份锁定承诺，自发行人股票在证券交易所上市之日起 36 个月内，不转让或委托他人管理本企业直接或间接持有的发行人首次公开发行股票前已发行的股份，也不由发行人回购本企业持有的上述股份。 2、在发行人实现盈利前，自发行人股票上市交易之日起 3 个完整会计年度内，不转让或者委托他人管理本企业于本次发行上市前已直接或间接持有的发行人股份，也不提议由发行人回购该部分股份；自发行人股票上市交易之日起第 4 个会计年度和第 5 个会计年度内，每年减持的本企业于本次发行上市前已直接或间接持有的发行人股份不超过发行人股份总数的 2%。在发行人实现盈利后，本企业可以自发行人当年年度报告披露后次日起减持本企业于本次发行上市前已直接或间接持有的发行人股份，但应当遵守本承诺函其他规定。 3、发行人上市当年较上市前一年净利润（以扣除非经常性损益后归母净利润为准，下同）下滑 50%以上的，延长本企业届时所持股份锁定期限 6 个月；发行人上市第二年较上市前一年净利润下滑 50%以上的，在前项基础上延长本企业届时所持股份锁定期限 6 个月；发行人上市第三年较上市前一年净利润下滑 50%以上的，在前两项基础上延长本企业届时所持股份锁定期限 6 个月。前述“届时所持股份”分别指本企业在发行人上市前取得，上市当年及之后第二年、第三年发行人年报披露时仍持有的股份。 4、在发行人股票上市后 6 个月内，如发行人股票连续 20 个交易日的收盘价均低于首次公开发行股票时的发行价，或者发行人股票上市后 6 个月期末（如该日不是交易日，则为该日后第 1 个交易日）收盘价低于首次公开发行股票时的发行价的，本企业在发行人首次公开发行股票前所持有的发行人股份的锁定期自动延长至少 6 个月。 5、如本企业在发行人首次公开发行股票前直接或者间接所持有的发行人股份在锁定期满后 2 年内减持的，减持价格不低于首次公开发行股票时的发行价。若因发行人派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，上述股份价格、股份数量按照证券交易所的有关规定作相应调整。 6、本企业将严格遵守法律、法规、规范性文件、公司上市的证券交易所业务规则中关于股东持股及股份变动（包括减持）的有关规定，规范诚信履行股东的义务。 7、在本企业持股期间，若股份锁定和减持的法律、法规、规范性文件、政策及证券监管机构的要求发生变化，则本企业将依据相关法律法规的要求适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。
陈维良、上海骄迈与上海曦骥	关于持股意向及减持意向的承诺函	1、承诺人将严格遵守相关法律、法规和规范性文件的规定以及发行人招股说明书中记载的和承诺人出具的承诺文件中载明的各项锁定期要求，在锁定期内承诺人不会进行任何违反相关规定及股份锁定承诺的股份减持行为。 2、锁定期届满后的两年内，在不违反相关法律、法规、规范性

承诺主体	承诺函	承诺内容
		文件之规定以及承诺人作出的其他公开承诺前提下，承诺人存在适当减持发行人股份的可能。 3、在锁定期届满后，承诺人拟通过包括但不限于二级市场集中竞价交易、大宗交易、协议转让等方式减持所持有的发行人首次公开发行股票前已发行的股份。 4、承诺人减持直接或者间接所持有的发行人股票的价格根据当时的二级市场价格确定，并应符合相关法律法规及证券交易所规则要求；承诺人在发行人首次公开发行股票前直接或者间接所持有的发行人股份在锁定期满后 2 年内减持的，减持价格不低于首次公开发行股票的发行价。若因发行人派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，上述股份价格按照证券交易所的有关规定作相应调整。 5、锁定期届满后，承诺人减持直接或者间接所持发行人股份时，将按照证券监督管理机构及证券交易所的规则及时、准确地履行信息披露义务并配合发行人的信息披露工作，提前将减持意向和拟减持数量等信息以书面方式通知发行人，并由发行人在减持前 3 个交易日对相关减持计划予以公告，通过证券交易所集中竞价交易或者大宗交易方式首次减持的在减持前 15 个交易日对相关减持计划予以公告。 6、在承诺人持有发行人股份期间，若股份减持的法律、法规、规范性文件、政策及证券监管机构的要求发生变化，则承诺人将依据相关法律法规的要求适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。

(2) 其他股东已作出相关承诺

发行人其他股东已出具《关于发行人股票锁定期的承诺函》，持股 5%以上股东及其一致行动人已出具《关于持股意向及减持意向的承诺函》，主要股东已出具《关于不谋求控制权的承诺函》，具体内容如下：

承诺主体	承诺函	承诺主要内容
除上海骄迈、上海曦骥、陈维良之外的其他股东	关于发行人股票锁定期的承诺函	1、自发行人股票在证券交易所上市之日起 12 个月内，承诺人不转让或委托他人管理承诺人持有的发行人首次公开发行股票前已发行的股份，也不由发行人回购承诺人持有的上述股份。对于承诺人所持有的于发行人提交首次公开发行股票申请前 12 个月内作为新增股东所取得的发行人股份，及提交申请前 6 个月内取得的发行人新增股份（含资本公积金转增股本部分），分别自该等股份取得之日[取得之日指相关股份变动完成市场监督管理局变更登记之日。发行人改制为股份有限公司后，通过受让股份方式取得的新增股份部分，取得之日为发行人股东名册记载该等变动之日]起 36 个月内，承诺人不转让或委托他人管理，也不由发行人回购承诺人持有的该等股份。 2、承诺人将严格遵守法律、法规、规范性文件、公司上市的证券交易所业务规则中关于股东持股及股份变动（包括减持）的有关规定，规范诚信履行股东的义务。 3、在承诺人持股期间，若股份锁定和减持的法律、法规、

承诺主体	承诺函	承诺主要内容
		规范性文件、政策及证券监管机构的要求发生变化，则承诺人将依据相关法律法规的要求适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。
葛卫东及混沌投资、经乾二号、经纬创壹号、经纬创荣、经纬厦门、经纬创叁号、经纬创华	关于持股意向及减持意向的承诺函	1、承诺人将严格遵守相关法律、法规和规范性文件的规定以及发行人招股说明书中记载的和承诺人出具的承诺文件中载明的各项锁定期要求，在锁定期内承诺人不会进行任何违反相关规定及股份锁定承诺的股份减持行为。 2、锁定期届满后的两年内，在不违反相关法律、法规、规范性文件之规定以及承诺人作出的其他公开承诺前提下，承诺人存在适当减持发行人股份的可能。 3、在锁定期届满后，承诺人拟通过包括但不限于二级市场集中竞价交易、大宗交易、协议转让等方式减持所持有的发行人首次公开发行股票前已发行的股份。 4、承诺人减持所持有的发行人股票的价格根据当时的二级市场价格确定，并应符合相关法律法规及证券交易所规则要求；承诺人在发行人首次公开发行股票前所持有的发行人股份在锁定期满后 2 年内减持的，减持价格不低于减持时发行人最近一期经审计的每股净资产（若公司最近一期审计基准日后有资本公积转增股本、派送股票或现金红利、增发、配股或缩股等事项导致公司净资产或股份总数发生变化的，每股净资产需相应进行调整）。若因发行人派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，上述股份价格按照证券交易所的有关规定作相应调整。 5、锁定期届满后，承诺人减持所持发行人股份时，将按照证券监督管理机构及证券交易所的规则及时、准确地履行信息披露义务并配合发行人的信息披露工作。在本企业与关联方合计持有公司 5%以上股份时，提前将减持意向和拟减持数量等信息以书面方式通知发行人，并由发行人在减持前 3 个交易日对相关减持计划予以公告，通过证券交易所集中竞价交易或者大宗交易方式首次减持的，在减持前 15 个交易日前对相关减持计划予以公告。 6、在承诺人持有发行人股份期间，若股份减持的法律、法规、规范性文件、政策及证券监管机构的要求发生变化，则承诺人将依据相关法律法规的要求适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。
葛卫东及混沌投资、红杉瀚辰、红杉雅恒	关于不谋求控制权的承诺函	1、除法定一致行动关系外，本企业与其他股东的一致行动关系（如涉及）已经向公司披露。2、本企业认可发行人的实际控制人始终为陈维良。3、本企业系基于财务投资的目的持有公司股权，自持有发行人股份以来，本企业未谋求或与其他股东共同谋求发行人的控制权。4、自本承诺函签署之日起至发行人股票在上海证券交易所科创板上市之日起的 36 个月内，本企业将不会单独或采取与其他主体签订一致行动协议或通过任何其他安排，谋求或共同谋求发行人的控制权。
经乾二号、经纬创壹号、经纬创荣、经纬	关于不谋求控制权的承诺函	1、除法定一致行动关系外，本企业与其他股东的一致行动关系（如涉及）已经向公司披露。2、本企业认可发行人的实际控制人始终为陈维良。3、本企业系基于财务投

承诺主体	承诺函	承诺主要内容
厦门、经纬创叁号、经纬创华、和利国信智芯、南京智沐曦、浦口高科、引领区基金		资的目的持有公司股权，自持有发行人股份以来，本企业未谋求或与其他股东共同谋求发行人的控制权。4、自本承诺函签署之日起至发行人股票在上海证券交易所科创板上市之日起的 36 个月内，本企业将不会单独或采取与其他主体签订一致行动协议或通过任何其他安排，谋求或共同谋求发行人的控制权，亦不会协助或促使发行人实际控制人之外的其他主体通过任何方式谋求发行人的实际控制人地位。

上述承诺内容真实、合法、有效，对承诺人具有约束力，能够保障陈维良对发行人的控制权在本次发行上市后 36 个月内保持稳定。

综上，本次发行不会导致陈维良对发行人的控制权发生变化，发行人已在上海骄迈、上海曦骥、董事会席位层面以及发行人股东持股层面采取了保障控制权稳定的具体措施。

（二）高崧、任玉会与陈维良的关系、代持形成的背景，代持还原是否存在纠纷或者潜在纠纷

1、高崧、任玉会与陈维良的关系、代持形成的背景

根据相关自然人提供的调查表以及访谈结果，高崧、任玉会的相关情况如下：

姓名	与陈维良的关系	代持形成的背景
高崧	朋友	1、交易步骤相对简便，高崧时任华荣汇电子科技（北京）有限公司法定代表人，其个人不愿其对外投资事宜公开披露，要求不显名 2、沐曦有限彼时正在进行天使轮融资，为避免股权转让事宜影响天使轮融资的进度，实际控制人与其协商一致，由实际控制人予以代持
任玉会	市场化投资人，不存在亲属、朋友等亲友关系或其他关联关系	沐曦有限彼时正在进行天使轮融资，为避免股权转让事宜影响天使轮融资的节奏与进度，实际控制人要求并经与其协商一致，由实际控制人予以代持

2、代持还原是否存在纠纷或者潜在纠纷

高崧、任玉会与陈维良代持及解除情况详见发行人招股说明书“第四节 发行人基本情况”之“二、发行人设立情况和报告期内的股本、股东变化情况”之“（六）关于代持及解除情况”。根据代持还原协议、相关股东出具的承诺函以及访谈确认，高崧、任玉会与陈维良的代持已解除，被代持人对代持人于代持期间受托行使股东权利的事实均予以认可，无任何争议、纠纷，各方未发生过包括股权事项在内的任何权利义务争议、纠纷，本次代持还原不存在纠纷或者潜在纠纷。

纷。

综上，高崧、任玉会与陈维良的代持已解除，代持形成具有合理的背景原因，代持还原不存在纠纷或者潜在纠纷。

（三）结合发行人与股东签署的补充协议中，特殊权利条款内容及附条件恢复具体约定，陈维良所持公司股权比例等，说明相关特殊权利条款的清理是否符合《监管规则适用指引——发行类第4号》对赌协议的相关规定，对控制权稳定性的影响。

1、发行人与股东签署的补充协议中，特殊权利条款内容及附条件恢复具体约定

根据公司现行有效的《沐曦集成电路（上海）股份有限公司 Pre-IPO 轮股东协议》（“《Pre-IPO 轮股东协议》”），股东特殊权利约定主要包括股东的反稀释权、优先认购权、优先购买权、共同出售权、投资方的股份转让、回购权、优先清算权、信息权、分红权、共同投资待遇、最优惠条款、合格上市等特殊权利条款。

《Pre-IPO 轮股东协议》由发行人、陈维良、彭莉、杨建、上海骄迈、上海曦骥以及其他股东签署，上述协议关于股东特别权利的具体约定内容如下：

序号	特别权利	相关权利的主要约定内容
1	2.1 反稀释	未经投资方事先书面同意，公司在交割日后拟发行任何新的股权、股份、股票、可转债券、或其他可转换为公司任何股份的认股权利价格不得低于该等投资方对应的增资价格 在经过投资方书面同意的情况下，如果新一轮增资的价格低于投资方增资时的增资价格的，则投资方有权要求公司根据完全棘轮条款对增资价格进行调整，即按照新增资价格调整该投资方应当持有的股份数额，根据本协议约定向该投资方进行股份补偿
2	2.2 优先认购权	公司在交割日后拟开展新一轮增资时，公司应至少提前三十（30）日向投资方发出书面通知，说明交易的所有重要条款和条件（包括但不限于增资金额、认购价格、潜在投资主体信息和付款时间），投资方对此享有优先认购权
3	2.3 优先购买权	如创始股东、创始人拟向任何第三方直接或间接转让、出售、处置其持有的公司股份或者接受预期买方提出购买转让股份的要约，其应向投资方发出书面通知，说明拟转让交易的所有重要条款和条件。投资方有权（但无义务）以相同的价格和付款时间，优先于预期买方或公司的其他创始股东向转让方购买全部或部分转让股份
4	2.4 共同出售权	针对创始股东、创始人作为转让方拟转让其直接或间接持有的公司股份时，如任一投资方决定不行使优先购买权的，则该投资方还有权在股东通知回复期期满前向公司及转让方发出书面通知，要求与转让方

序号	特别权利	相关权利的主要约定内容
		一起以同样的价格、条款和条件向预期买方共同出售其所持有的公司股份
5	2.5 投资方的股份转让	投资方对外转让其持有的全部或部分公司股份的，应提前十五日通知公司并由公司在收到前述通知之日起五日内通知全体股东。投资方持有的公司股份在不出售给公司竞争者及其关联方的前提下，投资方处分其持有的公司股份，无需获得其他股东的同意
6	回购权	如果在本协议签署后发生以下任一回购情形的，Pre-IPO 轮投资方、C 轮投资方、B 轮投资方、Pre-B 轮投资方、A 轮投资方、Pre-A+轮投资方、Pre-A 轮投资方、天使轮投资方有权自回购事件发生后的十二个月内自主选择要求公司和/或创始股东、创始人和/或前述主体指定的第三方按照约定回购该投资方所持有的全部或者部分公司股份： （1）经投资方或其提名任命的公司董事提议公司上市，且公司各项量化指标以及各项条件都已达到上市要求（以独立中介机构包括券商、会计师以及律师出具的专业意见为准），但是由于创始人和/或创始股东的原因导致公司股东会或董事会决议未能通过公司上市的决议； （2）截至 2026 年 3 月 31 日仍未完成合格上市； （3）公司实际控制人发生重大变化，被认定对公司生产经营造成重大不利影响或对上市造成重大不利影响的； （4）公司或创始人、创始股东有犯罪行为或重大违约或其他重大违法违规行，且在收到投资方要求改正的书面通知后三十（30）个工作日内仍没有改正的或仍未消除对公司生产经营或上市可能造成的重大不利影响的； （5）任何公司所拥有或使用的知识产权存在权属争议，从而第三方向公司提起有关侵犯知识产权的诉讼或仲裁，且最终被法院生效判决或仲裁裁决认定公司败诉，且该等败诉对公司生产经营造成重大不利影响或对公司上市造成重大不利影响的； （6）创始人重大违反对前任雇主的竞业限制义务的； （7）集团公司的经营范围和业务性质发生实质性的重大不利变化（具体指其主营业务变更为与 Pre-IPO 轮投资时的业务不同或不相关的业务，且该等变化对公司造成实质性不利的影响，但经公司股东会表决同意的除外）； （8）其他投资方的回购权被触发的情形。
7	2.7 优先清算权	若公司发生清算、结业、解散、关闭或发生出售事件时均视为发生清算事件，经清算后，公司财产在支付清算费用、职工工资、社会保险费用和法定补偿金。缴纳所欠税款后，应当按照 Pre-IPO 轮投资方和 C 轮投资方清算优先额、B 轮投资方清算优先额、Pre-B 轮投资方清算优先额、A 轮投资方清算优先额、Pre-A+轮投资方清算优先额、Pre-A 轮投资方清算优先额、天使轮投资方清算优先额的原则进行份额
8	2.8 信息权	公司应当按照约定向投资方提供相关资料和文件
9	2.9 分红权	公司所有股东（包括投资方）有权按照各自实缴出资比例获得分红
10	2.10 共同投资待遇	在 C 轮领投方、B 轮领投方、Pre-B 轮领投方、A 轮领投方、Pre-A+轮领投方、Pre-A 轮领投方和天使轮领投方持有公司股份期间，C 轮领投方、B 轮领投方、Pre-B 轮领投方、A 轮领投方、Pre-A+轮领投方、Pre-A 轮领投方和天使轮领投方有权在同等条件下优先参与由创始人和/或其配偶发起的任何经营，包括但不限于 GPU 等与公司主营业务相关业务的投资项目或企业
11	2.11 最优惠条款	Pre-IPO 轮投资方和 C 轮投资方的权利一致，且 Pre-IPO 轮投资方和 C 轮投资方的权利至少相当于 C 轮投资之前其他享有优先权的公司股东

序号	特别权利	相关权利的主要约定内容
		所享有的权利，B 轮投资方的权利至少相当于 B 轮投资之前其他享有优先权的公司股东所享有的权利，Pre-B 轮投资方的权利至少相当于 Pre-B 轮投资之前其他享有优先权的公司股东所享有的权利，A 轮投资方的权利至少相当于 A 轮投资之前其他享有优先权的公司股东所享有的权利，Pre-A+轮投资方的权利至少相当于 Pre-A+轮投资之前其他享有优先权的公司股东所享有的权利，Pre-A 轮投资方的权利至少相当于 Pre-A 轮投资之前其他享有优先权的公司股东所享有的权利，天使轮投资方的权利至少相当于天使轮投资之前其他享有优先权的公司股东所享有的权利（如有）。如果公司在将来的融资中对新投资方存在比投资方更为优惠的条款，则该等更优惠条款将自动适用于投资方所拥有的公司股份
12	4.1.2 款项下“对创始人已取得的已解锁股份股权的处置方式”	任一创始人发生如下恶意情形之一的，则公司有权以等同于其成本价或其实缴注册资本或届时股权价值（以孰低者为准）的价格回购其已取得的股权： a) 实施违反法律法规的行为且被以犯罪论处，并最终导致（i）对公司上市造成重大不利影响，或（ii）无法继续为公司提供服务； b) 实施有损公司利益的行为，包括但不限于违反竞业禁止、同业竞争限制、泄露公司秘密等，从而给公司利益造成重大损失的行为； c) 实施其他公司认定的恶意损害公司利益的行为
13	6 公司治理结构安排	部分事项除了需要股东会决议通过外，须经合计持有天使轮投资方 50%及以上的表决权、合计持有 Pre-A 轮投资方 50%及以上的表决权、合计持有 Pre-A+轮投资方 50%及以上的表决权、合计持有 A 轮投资方 50%及以上的表决权、合计持有 Pre-B 轮投资方 50%及以上的表决权、合计持有 B 轮投资方 50%及以上的表决权、合计持有 C 轮投资方 50%及以上的表决权，以及合计持有 Pre-IPO 轮投资方 50%及以上的表决权同意方可通过 公司董事会由五名董事组成，红杉瀚辰有权提名一名董事，天使轮领投方有权提名一名董事（“和利国信智芯董事”），其余三（3）名由陈维良提名，董事会表决部分事项时须经三分之二以上董事通过且必须经 Pre-A 轮董事、和利国信智芯董事的同意方可通过
14	7 合格上市	创始人和公司应尽力争取公司在 2026 年 3 月 31 日前，在投资方及公司共同认可的公共的证券交易市场（含主板、科创板和创业板或其他国内外证券交易市场）上市。合格上市应至少同时满足以下条件： （1）公司满足在投资方认可的境内或境外证券交易市场上市的基本条件； （2）上市前公司估值不低于 211 亿元人民币，且融资规模不低于 21.1 亿元人民币

2025 年 4 月，发行人全体股东签署《关于沐曦集成电路（上海）股份有限公司 Pre-IPO 轮股东协议之补充协议》（以下简称“《补充协议》”），《补充协议》中关于特殊权利条款的终止与附条件恢复的具体约定内容如下：

义务主体	涉及特殊权利	终止约定	是否已终止	附条件恢复约定
发 行 人	“2.6 回购权”	董事会批准报出本次发行上市财务报表日终止	是，2025 年 4 月 30 日，发行人董事会已批准报出本次发行上市财务	不存在，自始不具有任何约束力和法律效力，且该等终止是永久、无条件、不可撤销且不可恢复的

义务主体	涉及特殊权利	终止约定	是否已终止	附条件恢复约定
			报表	
	“2.1 反稀释”、“2.2 优先认购权”、“2.3 优先购买权”、“2.4 共同出售权”、“2.5 投资方的股份转让”、“2.7 优先清算权”、“2.8 信息权”、“2.9 分红权”、“2.10 共同投资待遇”、“2.11 最优惠条款”、4.1.2 款项下“对创始人已取得的已解锁股份股权的处置方式”、“6 公司治理结构安排”、“7 合格上市”	发行人本次发行上市申请被证券交易所正式受理之日终止	是，2025 年 6 月 30 日，发行人本次发行上市申请已被上交所受理	不存在，对公司及各方自始不具有任何约束力和法律效力，且该等终止是永久、无条件、不可撤销且不可恢复的
创始人、创始股东	“2.6 回购权”	发行人本次发行上市申请被证券交易所正式受理之日终止	是，2025 年 6 月 30 日，发行人本次发行上市申请已被上交所受理	如发生以下任一情形（以较早发生的日期为准），公司创始人、创始股东作为回购义务人的回购义务自动恢复效力：（一）公司本次发行上市申请因任何原因被撤回；（二）公司本次发行上市申请被有权证券监管机构及相关主管部门终止、否决或不予（撤销）注册；或（三）有权证券监管机构及相关主管部门向公司核发的本次发行上市批文/注册文件失效
	其他特殊权利条款			不存在

如上表所示，发行人作为回购义务人涉及的对赌协议已于 2025 年 4 月 30 日彻底终止且自始无效，发行人作为义务人的其他特殊股东权利已于 2025 年 6 月 30 日终止，发行人作为义务人的特殊股东权利条款终止后均不可恢复。创始人、创始股东作为回购义务人的条款已于 2025 年 6 月 30 日终止，仅在特定情况下方可恢复，且在本次发行上市审核期间及完成后均不会恢复。

2、相关特殊权利条款的清理是否符合《监管规则适用指引——发行类第 4 号》对赌协议的相关规定，对控制权稳定性的影响

经逐项比对《监管规则适用指引——发行类第 4 号》，发行人相关特殊权利

条款的清理情况如下：

序号	《监管规则适用指引——发行类第 4 号》 核查要求	对照情况
1	发行人是否为对赌协议当事人	否。发行人作为《Pre-IPO 轮股东协议》及《补充协议》的签署方，其作为义务人承担义务的全部特殊权利条款均已终止且约定自始不具有任何约束力和法律效力，且该等终止是永久、无条件、不可撤销且不可恢复的，故发行人不再为对赌协议的当事人
2	对赌协议是否存在可能导致公司控制权变化的约定	以发行人的股东为义务承担主体的对赌条款及其他投资方特殊权利已终止，且本次发行上市在审期间及完成后均不会恢复效力，在此期间不会导致公司控制权发生变化
3	对赌协议是否与市值挂钩	否。以发行人的股东为义务承担主体的对赌条款及其他投资方特殊权利不存在与市值挂钩的约定
4	对赌协议是否存在严重影响发行人持续经营能力或者其他严重影响投资者权益的情形	否。以发行人的股东为义务承担主体的对赌条款及其他投资方特殊权利已终止，且本次发行上市在审期间及完成后均不会恢复效力，在此期间不存在严重影响发行人持续经营能力或者其他严重影响投资者权益的情形
5	回售责任“自始无效”相关协议签订日是否在财务报告出具日之后	否。相关方于《补充协议》中约定发行人为义务承担主体的有关回购的对赌条款“对公司及各方自始不具有任何约束力和法律效力，且该等终止是永久、无条件、不可撤销且不可恢复的”，《补充协议》签订日期为财务报告出具当日，不存在需补充提供协议签订后最新一期经审计的财务报告的情形

如上表所示，发行人股东相关特殊权利条款的清理符合《监管规则适用指引——发行类第 4 号》对赌协议的相关规定，不存在对控制权稳定性有不利影响的情形。

因此，发行人作为义务人的特殊权利条款均已终止且不附恢复条款，创始人、创始股东作为回购义务人的条款亦已终止，且在本次发行上市审核期间及完成后均不会恢复，发行人的控制权清晰、稳定。

综上，发行人相关特殊权利条款的清理符合《监管规则适用指引——发行类第 4 号》的规定，对控制权清晰、稳定不存在不利影响。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、发行人律师执行了如下核查程序：

1、获取并查验发行人历次融资的增资协议、股东协议、公司章程、工商档案以及股东名册（如有）等资料，核查陈维良及其控制的持股平台自报告期初至

今所持发行人股份及对应表决权比例的变化情况、发行人其他股东持股比例及股权结构的情况以及陈维良拥有的董事会席位情况。

2、获取并查验发行人第一届董事会第十一次会议、2024 年年度股东会的会议文件、最新股东名册，核查发行人本次发行上市方案，测算本次发行上市后陈维良所持发行人股份的稀释情况。

3、获取并查验上海骄迈、上海曦骥的合伙协议，核查执行事务合伙人选任、职权以及更换的相关条款的约定内容。

4、获取并查验发行人 2025 年第五次临时股东会会议文件及《沐曦集成电路（上海）股份有限公司董事会成员提名规则》，核查陈维良拥有的董事提名权情况。

5、获取并查验全体股东包括实际控制人及其持股平台作出的关于股票锁定期的承诺、持股 5%以上股东作出的关于持股意及减持意向的承诺函，核查相关股东关于股票锁定安排及减持的承诺情况。

6、获取并查验陈维良作出的关于维持发行人控制权稳定性的承诺函、主要股东作出的关于不谋求控制权的承诺函，核查陈维良以及发行人股东作出的保障实际控制人控制权稳定的具体措施。

7、获取并查验高崧、任玉会的身份证、出资凭证、银行流水、签署的调查表、出具的承诺函，访谈自然人股东本人以及陈维良，核查前述自然人股东与陈维良的关系、代持形成的背景、代持还原是否存在纠纷或潜在纠纷。

8、获取并查验炫桥资本（任玉会）、高崧与陈维良签署的股权转让及委托持股协议，核查委托持股的安排以及代持人对代持解除的相关权利。

9、获取并查验任玉会、莱仙源及炫桥资本的股东与陈维良签署的代持还原协议，以及高崧与陈维良签署的代持还原协议，核查任玉会、高崧与陈维良代持还原的具体安排。

10、获取并查验发行人的《Pre-IPO 轮股东协议》及《补充协议》，核查特殊权利条款及其终止、附条件恢复的具体约定。

11、逐项比对《监管规则适用指引——发行类第 4 号》，核查发行人《Pre-IPO

轮股东协议》中股东特殊权利条款的清理情况。

（二）核查意见

经核查，保荐机构、发行人律师认为：

1、陈维良及其控制的持股平台合计所持发行人股份的比例一直最高，其他股东持股比例较低，股权结构较为分散，且其他单一股东（包括与其一致行动人）的持股比例与其差距仍较大；陈维良作为合伙企业的执行事务合伙人，拥有对上海骄迈、上海曦骥长期持续稳定的控制权；陈维良作为单一股东，控制的董事会席位占多数；实际控制人及其他股东均已作出维持控制权稳定的具体保障措施，陈维良拥有对发行人长期持续且清晰稳定的控制权。

2、高崧与陈维良系朋友关系，任玉会系市场化投资人，高崧、任玉会与陈维良的代持已解除，代持形成具有合理的背景原因，代持还原不存在纠纷或者潜在纠纷。

3、发行人相关特殊权利条款的清理符合《监管规则适用指引——发行类第4号》的规定，对控制权清晰稳定不存在不利影响。

9.2 关于其他股东

根据申报材料：（1）发行人成立于2020年9月，最近一次外部股权融资对应的投后估值为210.71亿元，公司估值增长较快（2）申报前12个月，公司突击入股股东共计82名，其中经纬创壹号、经纬创叁号、河南科投、海南泰达、富海精选二号等多名股东于2025年2月、3月分批次以不同价格入股发行人；

（3）和利国信智芯及其关联方当前合计直接持有发行人4.84%股份，其多次增资入股发行人，于2023年12月以1074.633元每注册资本向共青亚昌宏硕转让其持有的发行人股权，转让价格低于发行人前后两次增资价格；（4）发行人共有非自然人股东119名，其中有18名机构股东无需办理私募投资基金备案或私募投资基金管理人登记手续。

请发行人披露：（1）公司估值增长较快的原因。同一批次增资价格不一致、同一股东间隔时间较短多次以不同价格增资/入股的原因，是否存在相关特殊约定或利益安排；（2）和利国信智芯报告期前多次增资入股发行人，后对外转让

股权的原因、合理性、真实性、股权转让价格公允性；（3）结合《非上市公司监管指引第4号》等有关规定，分析18名机构股东是否需穿透计算股东人数，公司是否存在股东超过200人的情形。

请保荐机构、发行人律师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）公司估值增长较快的原因。同一批次增资价格不一致、同一股东间隔时间较短多次以不同价格增资/入股的原因，是否存在相关特殊约定或利益安排

1、公司估值增长较快的原因

自公司设立以来，其历次外部融资的入股价格由120.71元/注册资本增加至1,429.56元/注册资本，历次融资估值水平增长较快，主要系公司在发展过程中不断创新和积累核心技术、丰富并完善GPU产品线、持续增强产品影响力和客户认可度，经营业绩实现跨越式增长，公司在历次融资中结合同行业可比公司估值水平，经与投资方公平协商一致确定估值。公司估值上涨较快具有合理性，具体如下：

（1）业务实现跨越式增长，营业收入增速迅猛

近年来，受美国高性能GPU/AI芯片出口管制、产业扶持政策密集出台以及国内自主可控市场发展等因素影响，海外厂商如英伟达（NVIDIA）、超威半导体（AMD）等在中国市场的份额呈下降趋势，国产人工智能芯片公司迎来黄金发展期。此外，下游市场需求爆发增长为发行人业务发展提供了良好的外部环境。在新一轮人工智能发展浪潮下，国内外数据中心持续建设与扩容，互联网、金融、医疗、汽车、教育等领域多点开花，人工智能应用大规模落地。

在上述背景下，发行人自成立至今的业务规模呈快速增长趋势，营业收入实现跨越式增长。报告期各期，公司分别实现营业收入42.64万元、5,302.12万元、74,307.16万元和32,041.53万元，业务及收入增长趋势显著。

(2) 研发与产品优势协同发力，高性能 GPU 产品接连进入量产阶段

发行人在设立之初即专注 GPU 业务，确定了通用型和统一的 GPU 计算架构、软件定义硬件的软硬件协同研发体系，在发展初期经过“从零到一”全自研的技术沉淀，发行人在硬件、软件层分别积累了大量核心技术，形成核心技术体系并全面应用到公司产品中，研发与产品优势开始协同发力。

发行人首款智算推理芯片曦思 N100 于 2022 年 1 月交付流片，并于 2023 年 4 月正式量产。曦思 N100 芯片的设计、流片到量产，为公司打通了从研发到生产再到销售的全业务流，完善了产品、生产供应链和销售渠道的资源储备。2024 年 2 月，发行人训推一体芯片曦云 C500 正式量产，陆续在多个国家人工智能公共算力平台、运营商智算平台和商业化智算中心批量应用。同时，基于对国际政策形势的研判，发行人提前布局国产供应链，并基于国产供应链研发设计了新一代训推一体芯片曦云 C600，于 2024 年 10 月交付流片。因此，发行人从设立至今已量产的产品线覆盖了人工智能训练及推理、通用计算等计算场景，快速提升产品影响力和客户认可度，进一步提高公司历次融资的估值水平。

(3) 在手订单增速较快，未来业绩预期积极向好

发行人主要产品训推一体 GPU 板卡、训推一体 GPU 服务器于 2024 年开始实现大规模量产、销售。截至 2024 年末、2025 年 3 月末、2025 年 4 月末，发行人在手订单金额为 6.99 亿元、7.73 亿元、9.17 亿元，累计在手订单金额储备充足且增速较快，在公司融资过程中能够向投资者传递积极的业绩预期。

综上，发行人估值快速增长，主要与公司发展过程中的研发技术沉淀、产品量产进度、经营业绩与在手订单增长以及行业发展的增量需求等因素相关，具有合理性。公司历次融资的估值水平及入股价格均结合实际经营状况，经由交易各方公平协商一致确定，定价合理且具有公允性。

2、同一批次增资价格不一致、同一股东间隔时间较短多次以不同价格增资/入股的原因，是否存在相关特殊约定或利益安排

发行人同一批次增资价格不一致、同一股东间隔时间较短多次以不同价格增资入股的原因如下：

（1）同一批次增资价格不一致的原因

发行人外部投资人的增资入股价格均为参考发行人市场估值并经与各方协商一致确定，不存在同一轮次融资增资价格不一致的情形，发行人同一批次增资价格不一致的，主要系因不同轮次融资时间相近，为简化相关登记手续，发行人合并办理了市场监督管理部门的登记手续所致，具体情况如下：

时间/批次	价格	对应轮次
2023 年 12 月	1,088.8406 元/注册资本	Pre-B 轮融资第四期
	1,264.2583 元/注册资本	B 轮融资第一期

（2）同一股东间隔时间较短多次以不同价格增资/入股的原因

如本题之（一）/1 回复所述，发行人的估值增长较快，间隔时间较短的背景下，估值有较大幅度增长；同时，GPU 市场发展良好，发行人业务发展势头强劲，为扩大公司规模，发行人基于公司发展战略分别于 2025 年 2 月、3 月进行了 C 轮、Pre-IPO 轮融资；根据发行人全体股东彼时签署的股东协议的约定，现有的外部投资人股东具有优先认购权，即发行人在交割日后拟开展新一轮增资时，投资方享有优先认购权；与此同时，经纬创壹号、经纬创叁号、河南科投、海南泰达、富海精选二号等多名股东持续看好发行人及其所在行业的发展前景而追加投资。

因此，同一股东间隔时间较短多次以不同价格增资/入股具有商业合理性。

综上，发行人同一批次增资价格不一致、同一股东间隔时间较短多次以不同价格增资/入股的原因具有商业合理性，发行人与投资方不存在相关特殊约定或利益安排。

（二）和利国信智芯报告期前多次增资入股发行人，后对外转让股权的原因、合理性、真实性、股权转让价格公允性

1、和利国信智芯及其关联方的基本情况

（1）和利国信智芯

和利国信智芯成立于 2019 年，为非投向某一特定项目的创业投资基金，主要业务为创业投资，投资领域主要为半导体领域。

（2）浦口高科

浦口高科成立于 2020 年，有限合伙人为南京浦口开发区高科技投资有限公司，系管理人为该单一有限合伙人专门设立的非投向某一特定项目的创业投资基金，主要业务为创业投资，投资领域主要为半导体、材料、高端装备等泛科技领域。

（3）南京智沐曦

南京智沐曦成立于 2023 年，有限合伙人为扬州沐芯创业投资合伙企业（有限合伙）（以下简称“沐芯创投”）、扬州和利智矽股权投资合伙企业（有限合伙）（以下简称“和利智矽”），普通合伙人及基金管理人为和利创业投资管理（苏州）有限公司（以下简称“和利创投”）。沐芯创投成立于 2024 年 7 月，系投资发行人的专项基金，和利智矽成立于 2023 年，主要业务为股权投资，系和利创投募集设立的第二期专门投资半导体领域的创业投资基金，该基金采取结构化投资的安排，通过设立 SPV（特殊目的实体）进行具体项目的投资，南京智沐曦即为投资发行人设立的 SPV。

2、上述主体历次增资入股发行人的原因

序号	增资方	时间	轮次	原因
1	和利国信智芯	2020 年 12 月	天使轮融资	作为早期投资人，看重创始团队的行业经验，看好 GPU 行业的未来发展前景，作为领投资方参与了发行人的天使轮融资；后续企业发展持续向好，投资人依据其基金的投资策略行使了优先认购权，作为跟投方，进行了追加投资
2		2021 年 1 月	Pre-A 轮融资	
3		2021 年 2 月	Pre-A+轮融资	
4		2021 年 6 月	A 轮融资第一期	
5	浦口高科	2021 年 6 月	A 轮融资第一期	为单一有限合伙人设立的基金，经管理人推荐，投资人看好发行人的发展并进行投资
6	南京智沐曦	2025 年 2 月	B 轮融资第五期	经管理人推荐，投资人看好发行人的发展和未来的发展空间
7		2025 年 3 月	C 轮融资	
8		2025 年 3 月	Pre-IPO 轮融资	

3、和利国信智芯对外转让股权的基本情况

作为创业投资基金，和利国信智芯综合考虑基金投资期限、资金回流、实现基金的分配收益率的现实需求，结合投资退出策略，于 2023 年 8 月拟转让其持

有的部分发行人股权。彼时，发行人股东共青亚昌宏硕认可并看好公司的快速发展，意向追加投资，行使了优先购买权。经双方协商一致，参考 Pre-B 轮融资入股价格及发行人估值并有所折让，确定和利国信智芯以 1,074.63 元/注册资本向共青亚昌宏硕转让其所持有的发行人 3.7222 万元注册资本。本次股权转让均已经过双方内部有权决策机构的批准，系双方的真实意思表示，相关股份转让款均已支付完毕。

综上，和利国信智芯报告期前多次增资入股发行人，后对外转让股权具有商业合理性，交易真实，价格公允。

（三）结合《非上市公众公司监管指引第 4 号》等有关规定，分析 18 名机构股东是否需穿透计算股东人数，公司是否存在股东超过 200 人的情形。

1、股东人数的计算依据

根据《非上市公众公司监管指引第 4 号——股东人数超过二百人的未上市股份有限公司申请行政许可有关问题的审核指引》（以下简称“《指引》”）的相关规定，股份公司股权结构中存在通过“持股平台”间接持股的安排以致实际股东超过二百人的，在依据《指引》申请行政许可时，应当已经将代持股份还原至实际股东、将间接持股转为直接持股，并依法履行了相应的法律程序。《指引》所称“持股平台”是指单纯以持股为目的的合伙企业、公司等持股主体；由此，非单纯以持股发行人为目的即有实际经营业务的持股主体可以不进行还原或转为直接持股。

此外，《指引》规定，以依法设立的员工持股计划以及已经接受证券监督管理机构监管的私募股权基金、资产管理计划和其他金融计划进行持股，并规范运作的，可不进行股份还原或转为直接持股。

根据《私募投资基金监督管理暂行办法》的相关规定，以合伙企业、契约等非法人形式，通过汇集多数投资者的资金直接或者间接投资于私募基金的，私募基金管理人或者私募基金销售机构应当穿透核查最终投资者是否为合格投资者，并合并计算投资者人数。但是（1）社会保障基金、企业年金等养老基金，慈善基金等社会公益基金（2）依法设立并在基金业协会备案的投资计划（3）中国证监会规定的其他投资者投资私募基金的，不再穿透核查最终投资者是否为合格投

资者和合并计算投资者人数。

根据《证券期货法律适用意见第 17 号》相关规定，依法以公司制企业、合伙制企业、资产管理计划等持股平台实施的员工持股计划，员工人数不计算在内；参与员工持股计划时为公司员工，离职后按照员工持股计划章程或协议约定等仍持有员工持股计划权益的人员，可不视为外部人员；新《证券法》施行之前（即 2020 年 3 月 1 日之前）设立的员工持股计划，参与人包括少量外部人员的，可不做清理，在计算公司股东人数时，公司员工部分按照一名股东计算，外部人员按实际人数穿透计算。

根据上述规定及相关审核要求，按照以下标准对发行人最终股东人数进行穿透计算：

（1）已备案私募基金或私募基金管理人、自然人、有限公司（拥有实际经营业务）按 1 名股东计算；

（2）依法以公司制企业、合伙制企业、资产管理计划等设立的员工持股平台，且不含外部人员按 1 名股东计算。2020 年 3 月 1 日后设立的员工持股平台，公司员工与外部股东均按标准（1）穿透计算；

（3）其他类型主体需穿透至标准（1）所列主体计算。

2、18 名机构股东及其他股东的穿透计算股东人数情况

按前述标准，发行人 18 名无需办理私募投资基金备案或私募投资基金管理人登记手续的机构股东中 13 名需穿透计算，穿透后股东人数与其他 5 名无需穿透的机构股东合计人数 75 名；发行人其他 106 名股东中，包括 5 名自然人股东和 101 名已完成私募基金或私募基金管理人备案的股东，无需穿透计算，股东人数计 106 名，发行人股东穿透后人数合计为 180 名（陈维良为直接自然人股东，同时为上海骄迈的普通合伙人，去重算 1 名股东），不存在超过 200 人的情形。

综上，18 名机构股东根据相关法律法规的规定进行穿透后股东数量合计 75 名，发行人全部股东穿透后合计人数为 180 名，不存在股东超过 200 人的情形。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、发行人律师履行了以下核查程序：

1、获取并查验发行人增资协议/投资协议、股权转让协议、相关款项支付凭证及工商档案，核查历次增资及股权转让的情况、优先认购权的享有及行使情况。

2、获取并查验发行人审计报告、在手订单情况、主要型号的研发量产进度以及发行人出具的说明，核查发行人估值增长较快的原因。

3、获取并查验和利国信智芯的营业执照、工商档案、合伙协议以及填写的调查表并与其进行访谈，核查和利国信智芯的基本信息、出资结构、关联关系的情况以及历次增资、对外转让股权的原因。

4、获取并查验和利国信智芯与共青亚昌宏硕签署的股权转让协议、本次股权转让的转让款支付凭证并与其进行访谈，核查和利国信智芯对外转让股权的合理性、真实性、转让价格公允性。

5、获取并查验发行人机构股东的工商档案、填写的调查表、合伙协议、公司章程、股东出具的承诺函及说明性文件并查询基金业协会网站公示的私募投资基金管理人或私募投资基金信息，核查机构股东的出资结构、主营业务、备案情况。

（二）核查意见

经核查，保荐机构、发行人律师认为：

1、发行人估值快速增长，主要与公司发展过程中的研发技术沉淀、产品量产进度、经营业绩与在手订单增长以及行业发展的增量需求等因素相关，其历次增资及股权转让的估值及定价依据具备合理性；发行人不存在同一批次增资价格不一致、同一股东间隔时间较短多次以不同价格增资/入股具有商业合理性，发行人与投资方之间不存在相关特殊约定或利益安排。

2、和利国信智芯报告期前多次增资入股发行人，后对外转让股权具有商业合理性，交易真实，价格公允。

3、18名机构股东根据相关法律法规的规定进行穿透后股东人数合计75名，

发行人全部股东穿透后合计人数为 180 名，不存在股东超过 200 人的情形。

10.关于存货和固定资产

根据申报材料：（1）报告期各期末，发行人存货账面余额为 1,767.44 万元、18,730.50 万元、86,844.49 万元和 89,031.46 万元，增幅较大，主要包括原材料、半成品、库存商品、发出商品和委托加工物资，原材料主要为晶圆、HBM、印制电路板和电子元器件等；（2）报告期各期末，存货跌价准备余额为 0、0、9,166.06 万元和 8,806.52 万元，主要系 2024 年曦思 N100 系列产品市场需求发生变化、销售单价下降，对相关原材料、半成品和库存商品计提减值准备；（3）2023 年和 2024 年末，专用设备类固定资产减值准备金额为 2,786.91 万元、6,592.70 万元，2023 年主要系曦思 N100 系列产品停产，对该系列产品的光罩等相关生产设备全额计提减值准备。2024 年主要系美国政府禁令导致公司无法继续向晶圆代工厂商下达曦云 C 系列产品的生产订单，对公司存放在晶圆代工厂商处的光罩等固定资产全额计提减值准备。

请发行人披露：（1）按照明细种类列示报告期内原材料，按产品型号列示半成品、库存商品、发出商品和委托加工物资的金额、库龄、是否有订单覆盖、后续结转销售情况、计提跌价准备余额，是否存在库龄较长、长期呆滞存货，说明原因及合理性、期后领用或处置情况；（2）结合生产周期、备货政策等因素，说明报告期末存货余额快速增长结构变化较大的原因及合理性，与报告期内采购种类及金额、成本结转的匹配性，与同行业可比公司是否存在重大差异；（3）曦思 N100 正式量产后短期内即停产的原因及合理性，期后销售及结转情况、相关存货跌价准备计提是否充分；市场需求发生变化的具体过程，存货跌价准备计提时间晚于相关固定资产减值准备的原因及合理性，结合减值迹象出现的时间，说明相关资产减值准备计提的及时性；（4）结合各型号产品的市场需求、销售单价及毛利率变动、订单覆盖等情况，说明其他类型产品是否存在已被迭代、积压难以销售等情形，相应存货跌价准备计提的充分性；（5）结合与可比公司存货跌价准备具体计提政策及比例、存货库龄结构、产销率、毛利率水平比较情况，进一步说明发行人计提比例的合理性、是否和同行业可比公司之间存在较大差异；（6）2024 年存放在供应商的固定资产全额计提减值准备的具体情况后续处置情况，是否存在其他固定资产涉及相关情形、资产减值准备计提的充分性。

请保荐机构、申报会计师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）按照明细种类列示报告期内原材料，按产品型号列示半成品、库存商品、发出商品和委托加工物资的金额、库龄、是否有订单覆盖、后续结转销售情况、计提跌价准备余额，是否存在库龄较长、长期呆滞存货，说明原因及合理性、期后领用或处置情况

1、按照明细种类列示报告期内原材料

报告期内，发行人原材料账面余额明细种类列示如下：

单位：万元

原材料明细种类	2025 年 3 月末	2024 年末	2023 年末	2022 年末
晶圆	3,987.45	2,958.16	2,561.27	84.78
HBM	12,274.28	2,515.34	2,568.41	583.86
电子元器件	2,576.52	850.15	516.34	561.90
外购硬件及配件	428.22	428.66	-	-
合计	19,266.46	6,752.31	5,646.02	1,230.54

2、按产品型号列示半成品、库存商品、发出商品和委托加工物资的金额、库龄、是否有订单覆盖、后续结转销售情况、计提跌价准备余额

发行人产品分为训推一体系列和智算推理系列两大类，分类说明如下：

(1) 训推一体系列

1) 金额、库龄及占比情况

报告期各期末，发行人训推一体系列半成品、库存商品、发出商品和委托加工物资金额、库龄及占比情况如下：

单位：万元

2025 年 3 月 31 日								
存货类型	合计		0-180 天		181-360 天		361-720 天	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
半成品	25,861.24	3,263.60	25,860.09	3,263.60	-	-	1.15	-
库存商品	19,948.21	-	19,764.95	-	156.20	-	27.06	-
发出商品	10,745.74	-	4,998.67	-	5,747.07	-	-	-
委托加工物资	1,490.67	-	1,490.67	-	-	-	-	-
小计	58,045.86	3,263.60	52,114.38	3,263.60	5,903.27	-	28.21	-
占比	100.00%	100.00%	89.78%	100.00%	10.17%	-	0.05%	-
2024 年 12 月 31 日								
存货类型	合计		0-180 天		181-360 天		361-720 天	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
半成品	35,616.56	3,159.23	35,615.41	3,159.23	-	-	1.15	-
库存商品	20,777.94	-	20,707.32	-	58.85	-	11.77	-
发出商品	9,073.57	-	9,073.57	-	-	-	-	-
委托加工物资	3,600.10	-	3,600.10	-	-	-	-	-

合计	69,068.17	3,159.23	68,996.40	3,159.23	58.85	-	12.92	-
占比	100.00%	100.00%	99.90%	100.00%	0.09%	-	0.02%	-
2023 年 12 月 31 日								
存货类型	合计		0-180 天		181-360 天		361-720 天	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
半成品	3,482.75	-	3,482.75	-	-	-	-	-
库存商品	233.71	-	233.71	-	-	-	-	-
发出商品	2,074.31	-	2,074.31	-	-	-	-	-
委托加工物资	-	-	-	-	-	-	-	-
合计	5,790.77	-	5,790.77	-	-	-	-	-
占比	100.00%	-	100.00%	-	-	-	-	-

2022 年末，发行人无训推一体系列相关半成品、库存商品、发出商品和委托加工物资。2023 年末，训推一体系列产品处于风险量产阶段，未形成规模化生产及库存积累，因此期末无一年以上长库龄存货，符合研发阶段存货流转的正常特征。2024 年末、2025 年 3 月末，一年以上长库龄存货账面余额分别为 12.92 万元、28.21 万元，占期末存货余额的比例分别为 0.02%、0.05%，金额极小，占比极低。其中，长库龄库存商品为桥接板，系训推一体系列板卡配售零件，金额较小，且具有不易变质、损毁、可保存期限较长的特点，仍具备继续加工或销售的可能性，报告期内及期后正常销售，不存在需计提跌价准备的情形。

2024 年末、2025 年 3 月末，发行人对训推一体系列 0-180 天库龄半成品分别计提存货跌价准备 3,159.23 万元、3,263.60 万元，系发行人基于各生产阶段良率，对不良品充分计提单项存货跌价准备。

综上所述，报告期各期末，训推一体系列相关存货库龄基本在 1 年以内，部分库龄超过 1 年的存货保质期较长，报告期内及期后正常销售，训推一体系列存货不存在长期呆滞的情形。

2) 存货订单覆盖、结转销售情况

报告期各期末，发行人训推一体系列相关存货订单覆盖情况如下：

单位：万元

项目	2025 年 3 月末	2024 年末	2023 年末
半成品、库存商品、发出商品、委托加工物资账面余额（A）	58,045.86	69,068.17	5,790.77
在手订单估计的成本价（B）	34,514.07	24,756.32	5,717.39
订单覆盖率（B/A）	59.46%	35.84%	98.73%
库存商品余额（C）	19,948.21	20,777.94	233.71
库存商品订单覆盖率（B/C）	173.02%	119.15%	2446.36%

注：在手订单估计的成本价=在手订单产品数量*期末库存商品单位账面余额，在手订单为报告期各期末时点数据，下同

报告期各期末，发行人训推一体系列相关存货期后结转销售情况如下：

单位：万元

存货类型	2025 年 3 月末		2024 年末		2023 年末	
	期末余额	结转率	期末余额	结转率	期末余额	结转率
半成品	25,861.24	50.21%	35,616.56	87.05%	3,482.75	93.13%
库存商品	19,948.21	38.00%	20,777.94	54.49%	233.71	97.76%
发出商品	10,745.74	93.18%	9,073.57	100.00%	2,074.31	100.00%
委托加工物资	1,490.67	100.00%	3,600.10	100.00%	-	-
合计	58,045.86	55.25%	69,068.17	79.63%	5,790.77	95.78%

注：存货结转销售数据截至 2025 年 6 月 30 日

2023 年末、2024 年末和 2025 年 3 月末，训推一体系列半成品、库存商品、发出商品、委托加工物资账面余额总额在手订单覆盖率分别为 98.73%、35.84% 和 59.46%，训推一体系列相关库存商品在手订单覆盖率分别为 2446.36%、119.15% 和 173.02%。总体而言，发行人训推一体系列相关存货在手订单覆盖率不高。

2023 年末、2024 年末和 2025 年 3 月末，训推一体系列半成品期后销售结转率分别为 93.13%、87.05%、50.21%，库存商品期后销售结转率分别为 97.76%、54.49%、38.00%，呈下降趋势，一方面，结转销售统计时点距离各报告期末越来越近，另一方面，受到发行人提前进行备货的政策影响，结转销售率下降。

发行人为 Fabless 芯片设计企业，产品采购、生产周期较长，晶圆制造、封测代工等主要供应商长期处于产能饱和状态，且受代工厂生产排期、原材料供应、国际贸易环境变化等因素影响，生产周期和生产进度均具有不稳定性。2024 年，核心产品曦云 C500 系列正式量产，且生成式 AI 市场需求的井喷式增长使得市场对于人工智能训练推理芯片的需求大幅增加，基于对市场潜在需求的预判，同时为避免因生产延迟导致客户订单违约，发行人进行战略性生产备货，以此作为保障产品按时交付的重要策略。经查询已上市的采用 Fabless 模式经营的芯片设计企业，该等企业在手订单覆盖率普遍较低，如寒武纪、复旦微电、钜泉科技等。

总体看来，报告期各期末，发行人训推一体系列存货订单覆盖率未达到 100%，主要系公司根据产品生产排期、应对国际贸易环境的不确定性、市场需求增长等因素的考虑，进行前瞻性备货所致，具有合理性。发行人为充分满足下游销售需求和保证及时交付，提前备货比例较高，因此短期内销售结转率不高，但从长期来看，2023 年年末各类型存货期后销售结转率均在 90%以上，训推一体系列存

货期后销售、结转情况良好。

(2) 智算推理系列

1) 金额、库龄及占比情况

报告期各期末，发行人智算推理系列半成品、库存商品、发出商品和委托加工物资金额、库龄及占比情况

单位：万元

2025 年 3 月 31 日								
存货类型	合计		0-180 天		181-360 天		361-720 天	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
半成品	7,064.90	4,178.99	1,539.53	-	2,776.74	2,107.59	2,748.63	2,071.40
库存商品	4,654.24	568.26	3,808.32	35.83	44.13	27.98	801.79	504.45
发出商品	-	-	-	-	-	-	-	-
委托加工物资	-	-	-	-	-	-	-	-
合计	11,719.14	4,747.25	5,347.84	35.83	2,820.87	2,135.57	3,550.42	2,575.85
占比	100.00%	100.00%	45.63%	0.75%	24.07%	44.99%	30.30%	54.26%
2024 年 12 月 31 日								
存货类型	合计		0-180 天		181-360 天		361-720 天	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
半成品	7,056.16	4,178.99	4,291.30	2,095.33	83.65	63.00	2,681.21	2,020.67
库存商品	3,961.93	1,032.17	2,413.07	71.31	637.82	401.18	911.04	559.67
发出商品	5.91	-	-	-	-	-	5.91	-
委托加工物资	-	-	-	-	-	-	-	-

合计	11,024.01	5,211.16	6,704.37	2,166.64	721.48	464.19	3,598.16	2,580.34
占比	100.00%	100.00%	60.82%	41.58%	6.54%	8.91%	32.64%	49.52%
2023 年 12 月 31 日								
存货类型	合计		0-180 天		181-360 天		361-720 天	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
半成品	6,120.91	-	6,116.82	-	4.09	-	-	-
库存商品	1,166.89	-	1,114.20	-	52.68	-	-	-
发出商品	5.91	-	-	-	5.91	-	-	-
委托加工物资	-	-	-	-	-	-	-	-
合计	7,293.70	-	7,231.02	-	62.68	-	-	-
占比	100.00%	-	99.14%	-	0.86%	-	-	-
2022 年 12 月 31 日								
存货类型	合计		0-180 天		181-360 天		361-720 天	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
半成品	524.45	-	524.45	-	-	-	-	-
库存商品	12.45	-	12.45	-	-	-	-	-
发出商品	-	-	-	-	-	-	-	-
委托加工物资	-	-	-	-	-	-	-	-
合计	536.90	-	536.90	-	-	-	-	-
占比	100.00%	-	100.00%	-	-	-	-	-

发行人智算推理 GPU 芯片曦思 N100 系列于 2023 年 4 月正式量产，2022 年末、2023 年末尚未形成一年以上库龄相关存货。2024 年，受市场需求下降、市场竞争加剧影响，曦思 N100 系列产品销量有所下降，存货库龄整体有所增加，2024 年末、2025 年 3 月末长库龄存货占比较高。2024 年末、2025 年 3 月末，发行人按照成本与可变现净值孰低对曦思 N100 系列相关存货计提存货跌价准备，具体计提方法详见本问题回复之“（三）曦思 N100 正式量产后短期内即停产的原因及合理性..... 结合减值迹象出现的时间说明相关资产减值准备计提的及时性”。2025 年以来，发行人积极应对市场变化，采取降价促销等销售策略，曦思 N100 销量有所提高，2025 年 3 月末 1 年以上长库龄存货有所减少。

2) 存货订单覆盖、结转销售情况

报告期各期末，发行人智算推理系列相关存货在手订单覆盖情况如下：

单位：万元

项目	2025 年 3 月末	2024 年末	2023 年末	2022 年末
半成品、库存商品、发出商品、委托加工物资账面余额（A）	11,719.14	11,024.01	7,293.70	536.90
在手订单估计的成本价（B）	64.12	9.45	18.89	-
订单覆盖率（B/A）	0.55%	0.09%	0.26%	-
库存商品余额（C）	4,654.24	3,961.93	1,166.89	12.45
库存商品订单覆盖率（B/C）	1.38%	0.24%	1.62%	

报告期各期末，发行人智算推理系列相关存货期后销售结转情况如下：

单位：万元

存货类型	2025 年 3 月末		2024 年末		2023 年末		2022 年末	
	期末余额	结转率	期末余额	结转率	期末余额	结转率	期末余额	结转率
半成品	7,064.90	2.07%	7,056.16	23.75%	6,120.91	60.03%	524.45	100.00%
库存商品	4,654.24	18.52%	3,961.93	44.44%	1,166.89	82.73%	12.45	100.00%
发出商品	-	-	5.91	100.00%	5.91	100.00%	-	-
委托加工物资	-	-	-	-	-	-	-	-
合计	11,719.14	8.60%	11,024.01	31.22%	7,293.70	63.69%	536.90	100.00%

注：存货结转销售数据截至 2025 年 6 月 30 日

报告期各期末，智算推理系列相关存货订单覆盖率较低，主要原因为：一方

面，该系列产品整体交付周期较短，销售节奏较快，各期期末时点尚未确认收入订单金额少；另一方面，发行人基于不稳定的供应链和潜在的市场需求提前备货，因此各期末存货余额远大于在手订单金额。

从期后结转销售情况来看，2023 年，曦思 N100 产品正式进入量产阶段，销售情况良好，因此期后销售结转率较高。2024 年以来，曦思 N100 系列产品销量有所下滑，因此 2024 年末、2025 年 3 月末的存货期后销售结转率相应有所下降。2025 年以来，发行人积极应对市场变化，及时调整销售策略，采用降价促销等多种方式使得曦思 N100 销量大幅提高。

综上所述，截至 2024 年末、2025 年 3 月末，发行人智算推理系列 1 年以上相关存货占比较高，主要受曦思 N100 产品因市场需求变化导致的销路不畅影响，发行人已根据成本与可变现净值孰低的原则对曦思 N100 相关存货计提了充分的跌价准备。发行人曦思 N100 系列已进入产品生命周期末期，且晶圆加工环节已停产，发行人已调整曦思 N100 产品生产销售策略，预计未来曦思 N100 产品的期后结转销售情况将不断提升。

3、是否存在库龄较长、长期呆滞存货，说明原因及合理性、期后领用或处置情况

报告期各期末，存货库龄按存货类型列示如下：

单位：万元

2025年3月31日										
存货类型	合计		0-180天		181-360天		361-720天		720天+	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
原材料	19,266.46	795.67	16,289.51	38.47	1,428.21	6.14	1,524.27	751.07	24.48	-
半成品	32,926.14	7,442.59	27,399.62	3,263.60	2,776.74	2,107.59	2,749.78	2,071.40	-	-
库存商品	24,602.45	568.26	23,573.27	35.83	200.33	27.98	828.85	504.45	-	-
发出商品	10,745.74	-	4,998.67	-	5,747.07	-	-	-	-	-
委托加工物资	1,490.67	-	1,490.67	-	-	-	-	-	-	-
小计	89,031.46	8,806.52	73,751.73	3,337.90	10,152.35	2,141.71	5,102.90	3,326.92	24.48	-
占比	100.00%	100.00%	82.84%	37.90%	11.40%	24.32%	5.73%	37.78%	0.03%	-
2024年12月31日										
存货类型	合计		0-180天		181-360天		361-720天		720天+	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
原材料	6,752.31	795.67	4,847.05	38.47	158.83	8.99	1,721.91	748.21	24.51	-
半成品	42,672.72	7,338.23	39,906.71	5,254.56	83.65	63.00	2,682.36	2,020.67	-	-
库存商品	24,739.88	1,032.17	23,120.39	71.31	696.68	401.18	922.81	559.67	-	-
发出商品	9,079.48	-	9,073.57	-	-	-	5.91	-	-	-
委托加工物资	3,600.10	-	3,600.10	-	-	-	-	-	-	-
合计	86,844.49	9,166.06	80,547.82	5,364.33	939.16	473.18	5,332.99	3,328.55	24.51	-

占比	100.00%	100.00%	92.75%	58.52%	1.08%	5.16%	6.14%	36.31%	0.03%	-
2023年12月31日										
存货类型	合计		0-180天		181-360天		361-720天		720天+	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
原材料	5,646.02	-	5,597.59	-	14.02	-	34.41	-	-	-
半成品	9,603.66	-	9,599.57	-	4.09	-	-	-	-	-
库存商品	1,400.60	-	1,347.91	-	52.68	-	-	-	-	-
发出商品	2,080.22	-	2,074.31	-	5.91	-	-	-	-	-
委托加工物资	-	-	-	-	-	-	-	-	-	-
合计	18,730.50	-	18,619.39	-	76.70	-	34.41	-	-	-
占比	100.00%	-	99.41%	-	0.41%	-	0.18%	-	-	-
2022年12月31日										
存货类型	合计		0-180天		181-360天		361-720天		720天+	
	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备	账面余额	跌价准备
原材料	1,230.54	-	1,178.65	-	51.89	-	-	-	-	-
半成品	524.45	-	524.45	-	-	-	-	-	-	-
库存商品	12.45	-	12.45	-	-	-	-	-	-	-
发出商品	-	-	-	-	-	-	-	-	-	-
委托加工物资	-	-	-	-	-	-	-	-	-	-
合计	1,767.44	-	1,715.55	-	51.89	-	-	-	-	-
占比	100.00%	-	97.06%	-	2.94%	-	-	-	-	-

2022 年，发行人处于起步阶段，无一年以上长库龄，不存在库龄较长、长期呆滞存货的情形。

2023 年末，一年以上长库龄存货为 34.41 万元原材料，均为电子元器件。由于电子元器件保质期较长，通常 3-5 年，报告期内及期后，该部分长库龄电子元器件正常投产，不存在长期呆滞情形。

2024 年末，一年以上长库龄存货账面余额为 5,357.50 万元。包括：一年以上长库龄原材料账面余额为 1,746.42 万元，其中，库龄在 720 天以上的原材料账面余额共 24.51 万元，均为电子元器件，报告期内及期后正常投产；库龄在 361-720 天的原材料账面余额共 1,721.91 万元，主要系曦云 C500 系列研发初期 A0 版本晶圆和 ASIC die，因受 2023 年 10 月美国对华出口管制系列政策影响，算力超过许可无法继续生产，相关存货生产放缓，随着发行人转向国产产线替代生产，相关 A0 版本晶圆逐步投产，但 966 颗账面余额为 467.95 万元的 ASIC die，因已超过产品保质期，不满足量产使用标准，发行人已全额计提跌价准备。2024 年末，一年以上长库龄半成品账面余额为 2,682.36 万元，其中 2,681.20 万元为曦思 N100 系列相关半成品；2024 年末，一年以上长库龄库存商品共 922.81 万元，其中 911.04 万元为曦思 N100 系列相关库存商品，对于上述长库龄半成品和库存商品，发行人已充分计提存货跌价。

2025 年 3 月末，一年以上长库龄存货账面余额为 5,127.38 万元，与 2024 年末长库龄存货构成基本一致。2025 年 3 月末，长库龄存货较 2024 年末减少 230.12 万元，主要原因系：一是 65 片长库龄 A0 版本晶圆中已有 16 片完成投产；二是曦思 N100 系列相关长库龄库存商品实现进一步销售。

综上所述，报告期内发行人长库龄存货主要为曦思 N100 系列相关存货和曦云 C500 系列研发初期 A0 版本相关存货。对于曦思 N100 系列相关存货，发行人采用降价销售等方式，积极应对市场变化。2025 年一季度，曦思 N100 产品销量较 2024 年显著增加，相关长库龄存货有所减少。对于曦云 C500 系列研发初期 A0 版本相关存货，报告期内发行人已逐步投产。总体来看，长库龄存货期后领用情况较好。报告期各期末发行人长库龄存货占比较低，分别为 0.00%、0.18%、6.17%、5.76%。除曦思 N100 系列相关存货外，发行人基本不存在长期呆滞存货。

（二）结合生产周期、备货政策等因素，说明报告期内存货余额快速增长结构变化较大的原因及合理性，与报告期内采购种类及金额、成本结转的匹配性，与同行业可比公司是否存在重大差异

1、报告期内发行人存货余额及结构变化情况

报告期各期末，存货余额结构情况如下：

单位：万元

项目	2025 年 3 月末		2024 年末		2023 年末		2022 年末	
	金额	占比	金额	占比	金额	占比	金额	占比
原材料	19,266.46	21.64%	6,752.31	7.78%	5,646.02	30.14%	1,230.54	69.62%
半成品	32,926.14	36.98%	42,672.72	49.14%	9,603.66	51.27%	524.45	29.67%
库存商品	24,602.45	27.63%	24,739.88	28.49%	1,400.60	7.48%	12.45	0.70%
发出商品	10,745.74	12.07%	9,079.48	10.45%	2,080.22	11.11%	-	-
委托加工物资	1,490.67	1.67%	3,600.10	4.15%	-	-	-	-
合计	89,031.46	100.00%	86,844.49	100.00%	18,730.50	100.00%	1,767.44	100.00%

报告期末，各类存货账面余额按产品型号分类变动情况如下：

单位：万元

存货类别	产品型号	2025 年 3 月末			2024 年末			2023 年末			2022 年末
		金额	变动额	变动率	金额	变动额	变动率	金额	变动额	变动率	金额
原材料	训推一体系列	6,888.35	3,182.21	85.86%	3,706.13	1,006.54	37.28%	2,699.60	2,698.30	207561.47%	1.30
	智算推理系列	358.88	34.82	10.74%	324.06	67.87	26.49%	256.19	-289.02	-53.01%	545.21
	通用	12,019.24	9,297.12	341.54%	2,722.11	31.88	1.19%	2,690.23	2,006.20	293.29%	684.03
	小计	19,266.46	12,514.16	185.33%	6,752.31	1,106.29	19.59%	5,646.02	4,415.48	358.83%	1,230.54
半成品	训推一体系列	25,861.24	-9,755.32	-27.39%	35,616.56	32,133.81	922.66%	3,482.75	3,482.75	-	-
	智算推理系列	7,064.90	8.74	0.12%	7,056.16	935.25	15.28%	6,120.91	5,596.46	1067.11%	524.45
	小计	32,926.14	-9,746.58	-22.84%	42,672.72	33,069.06	344.34%	9,603.66	9,079.21	1731.19%	524.45
库存商品	训推一体系列	19,948.21	-829.73	-3.99%	20,777.94	20,544.23	8790.46%	233.71	233.71	-	-
	智算推理系列	4,654.24	692.30	17.47%	3,961.93	2,795.05	239.53%	1,166.89	1,154.44	9274.08%	12.45
	小计	24,602.45	-137.43	-0.56%	24,739.88	23,339.28	1666.38%	1,400.60	1,388.15	11151.58%	12.45
发出商品	训推一体系列	10,745.74	1,672.18	18.43%	9,073.57	6,999.25	337.43%	2,074.31	2,074.31	-	-
	智算推理系列	-	-5.91	-100.00%	5.91	-	-	5.91	5.91	-	-
	小计	10,745.74	1,666.27	18.35%	9,079.48	6,999.25	336.47%	2,080.22	2,080.22	-	-
委托加工物资	训推一体系列	1,490.67	-2,109.43	-58.59%	3,600.10	3,600.10	-	-	-	-	-
	智算推理系列	-	-	-	-	-	-	-	-	-	-
	小计	1,490.67	-2,109.43	-58.59%	3,600.10	3,600.10	-	-	-	-	-
合计		89,031.46	2,186.98	2.52%	86,844.49	68,113.99	363.65%	18,730.50	16,963.06	959.76%	1,767.44

报告期各期末，原材料账面余额分类变动情况如下：

单位：万元

存货类别	产品型号	2025 年 3 月末			2024 年末			2023 年末			2022 年末
		金额	变动额	变动率	金额	变动额	变动率	金额	变动额	变动率	金额
晶圆	训推一体系列	3,951.34	1,029.29	35.22%	2,922.05	372.82	14.62%	2,549.23	2,549.23	-	-
	智算推理系列	36.11	-	0.00%	36.11	24.07	200.00%	12.04	-72.75	-85.80%	84.78
	通用	-	-	-	-	-	-	-	-	-	-
HBM	训推一体系列	-	-	-	-	-	-	-	-	-	-
	智算推理系列	-	-	-	-	-	-	-	-	-	-
	通用	12,274.28	9,758.94	387.98%	2,515.34	-53.08	-2.07%	2,568.41	1,984.55	339.90%	583.86
电子元器件	训推一体系列	2,026.34	1,670.92	470.12%	355.42	205.06	136.37%	150.36	149.06	11466.48%	1.30
	智算推理系列	322.77	34.82	12.09%	287.95	43.80	17.94%	244.16	-216.27	-46.97%	460.43
	通用	227.41	20.63	9.98%	206.78	84.96	69.74%	121.82	21.65	21.61%	100.17
外购硬件及 配件	训推一体系列	428.22	-0.45	-0.10%	428.66	428.66	-	-	-	-	-
	智算推理系列	-	-	-	-	-	-	-	-	-	-
	通用	-	-	-	-	-	-	-	-	-	-
合计		19,266.46	12,514.16	185.33%	6,752.31	1,106.29	19.59%	5,646.02	4,415.48	358.83%	1,230.54

2、结合生产周期、备货政策等因素，说明报告期内存货余额快速增长，结构变化较大的原因及合理性

（1）生产周期、备货政策

发行人主要产品的代工从下单到交付的周期平均为 8 个月左右，与国际头部 GPU 厂商接近。在高端芯片代工厂产能饱和的情况下，发行人产品的生产周期在比肩国际一流产品的基础上具有一定的波动性。

发行人基于对各型号 GPU 产品未来市场趋势及主要客户需求的前瞻性评估，并结合主要供应商产能状况及各生产环节周期的综合考量，公司科学制定存货采购计划，并据此有序开展备货与生产安排。发行人核心原材料晶圆、HBM 采购周期（从下单至交付）平均为 6 个月左右，采购周期较长且不可避免的受到国际供应政策和供应商产能分配的影响，具有一定的不确定性，因此发行人提前备货比例较高。

（2）各期存货余额快速增长结构变化较大原因及合理性

1）2022 年

2022 年度，发行人核心产品均尚未进入量产阶段，因此整体存货规模较小且主要为原材料，库存商品占比极低。

2）2023 年

2023 年末，存货总额较 2022 年末增加 16,963.06 万元，较上期增加 959.76%，这一变动主要源于发行人产品生产阶段的阶段性跨越。2022 年，发行人核心产品尚未进入量产阶段，因此存货以原材料为主，而进入 2023 年后，产品量产进程加速推进，曦思 N100 系列于 2023 年 4 月正式量产，曦云 C500 系列也于同年 8 月进入风险量产阶段，随着两款核心产品逐步从研发阶段逐步转向规模化生产的过程，发行人为匹配产能逐步释放与市场需求，同步加大了原材料采购、半成品及库存商品备货力度，使得原材料、半成品、库存商品等各类型存货余额均呈现大幅度增长。随着生产阶段的推进且产品生产周期较长，存货结构也发生较大变化，2023 年末半成品占存货的比例大幅提升，原材料占比相应下降。

3) 2024年

2024年末,存货总额较2023年末增加68,113.99万元,较上期增加363.65%,其中原材料增加1,106.29万元,占比由30.14%下降至7.78%,半成品增加33,069.06万元,比例由51.27%小幅下降至49.14%,库存商品增加23,339.28万元,占比由7.48%上升至28.49%,发出商品增加6,999.25万元,占比由11.11%小幅下降至10.45%。这一变动主要源于发行人核心业务的规模化扩张,核心产品训推一体GPU芯片曦云C500系列于2024年2月正式量产,并于当年形成规模化生产,直接带动半成品、库存商品、发出商品的显著增长。而原材料相较于整体生产规模增幅较小,主要是由于发行人基于上游供应链变动情况及生产计划安排,2024年年末处于原材料大量消耗但尚未完成下一轮采购补充周期,形成短期库存空窗期,因此虽然原材料金额较上期增加1,106.29万元,但在存货总额中的占比由30.14%下降至7.78%,这一结构变化与供应链波动下的物料周转节奏高度匹配。

4) 2025年3月末

2025年3月末,存货总额较2024年末增加2,186.98万元,较上期增加2.52%,总体波动较小。其中,原材料期末余额较上期增加12,514.16万元,较上期增加185.33%,原材料占比由7.78%上升至21.64%,占比大幅上升,主要原因为发行人于2025年一季度集中大量备货原材料,而2024年期末处于原材料备货空窗期,因此原材料期末余额较上期大幅增加。半成品、产成品期末余额分别下降9,746.58万元、137.43万元,较上期分别下降22.84%、0.56%,主要原因为2025年一季度发行人整体产销率为132.02%,远高于2024年度产销率39.82%,因此半成品、产成品的期末余额和占比较上期均有所下降。

综上所述,报告期内,发行人存货余额快速增长,主要系发行人处于快速发展阶段,业务规模不断扩大,随着核心产品智算推理GPU芯片曦思N100系列、训推一体GPU芯片曦云C500系列分别于2023年4月和2024年2月正式量产,存货余额相应大幅增加,具有合理性。报告期内存货结构波动较大,主要受供应链、生产周期、发行人备货政策、产品投产进度、产销率等因素综合影响,具有合理性。

3、与报告期内采购种类及金额、成本结转的匹配性

报告期各期，发行人存货与主营业务成本勾稽一致，具体如下：

单位：万元

项目	2025 年 1-3 月	2024 年	2023 年	2022 年
存货期初余额（A）	86,844.49	18,730.50	1,767.44	-
本期采购额（B）	20,092.30	109,385.72	18,539.14	1,808.23
制造费用（C）	239.95	3,880.12	2,180.07	125.68
其他出入库净额（D）	-3,373.39	-10,784.58	-1,960.15	-134.11
存货期末余额（E）	89,031.46	86,844.49	18,730.50	1,767.44
产品销售成本 （F=A+B+C+D-E）	14,771.89	34,367.27	1,795.99	32.36
加：运费成本（G）	28.20	144.97	41.13	-
减：因销售转销存货跌价准备 （H）	463.91	-	-	-
主营业务成本（I）	14,336.18	34,512.24	1,837.12	32.36
差异（J=F+G-H-I）	-	-	-	-

注：其他出入库净额主要为研发、外购硬件及配件采购净额法调整和销售样卡的领用等。

其中，生产物料采购内容主要可分为晶圆、HBM 内存、外购硬件及配件、委托加工、电子元器件等。报告期内，各类型生产物料采购的金额和占比如下：

单位：万元

类别	2025 年 1-3 月		2024 年		2023 年		2022 年	
	金额	占比	金额	占比	金额	占比	金额	占比
晶圆	1,476.65	7.35%	40,637.47	37.15%	8,321.68	44.89%	260.33	14.40%
HBM 内存	10,088.98	50.21%	33,750.64	30.85%	6,361.94	34.32%	800.67	44.28%
外购硬件及配件	5,004.25	24.91%	19,908.20	18.20%	1,351.50	7.29%	-	-
委托加工	3,033.08	15.10%	12,848.13	11.75%	1,726.55	9.31%	496.59	27.46%
电子元器件	489.34	2.44%	2,241.28	2.05%	777.47	4.19%	250.65	13.86%
合计	20,092.30	100.00%	109,385.72	100.00%	18,539.14	100.00%	1,808.23	100.00%

报告期内，发行人生产物料采购主要为晶圆、HBM。由于主要原材料采购周期及供应链具有不稳定性，发行人备货政策受市场环境影响变化较大。2025 年 1-3 月，发行人集中大量备货 HBM，因此 HBM 采购比例有所增加，同时，受美国禁令影响，发行人晶圆采购节奏有所放缓，因此晶圆采购比例有所降低。

总体来看，报告期内，发行人采购种类结构较为稳定。

4、与同行业可比公司是否存在重大差异

报告期内，发行人与同行业可比上市公司存货余额波动比较情况如下：

单位：万元

可比公司	2025 年 3 月末		2024 年末		2023 年末		2022 年末
	余额	变动率	余额	变动率	余额	变动率	余额
寒武纪	302,552.12	48.02%	204,396.49	493.11%	34,461.78	-30.96%	49,916.27
海光信息	未披露	未披露	553,227.88	396.05%	111,525.97	-1.64%	113,385.34
龙芯中科	未披露	未披露	105,908.32	6.91%	99,066.74	29.85%	76,293.85
景嘉微	未披露	未披露	74,489.20	6.47%	69,962.25	12.00%	62,468.93
摩尔线程	未披露	未披露	71,856.97	119.43%	32,747.01	12.96%	28,988.95
平均值	/	/	201,975.77	204.39%	69,552.75	4.44%	66,210.67
中位数	/	/	105,908.32	119.43%	69,962.25	12.00%	62,468.93
发行人	89,031.46	2.52%	86,844.49	363.65%	18,730.50	959.76%	1,767.44

报告期内，发行人存货余额大幅增长，各期存货期末余额较上期分别增加 16,963.06 万元、68,113.99 万元、2,186.98 万元，分别增长 959.76%、363.65%、2.52%，主要原因为发行人产品处于量产及销售规模快速爬坡阶段。

2022 年，发行人处于起步阶段，尚无量产产品，因此存货规模较小。2023 年，曦思 N100 正式启动量产，同时曦云 C500 进入风险量产阶段，发行人提前备货，因此 2023 年末存货余额较 2022 年末大幅增加。由于同行业可比公司发展较为成熟，生产销售较为稳定，因此存货余额波动不具有可比性。

2024 年末，发行人存货余额同比增加 363.65%，主要原因系：一方面，核心产品曦云 C500 系列正式进入量产阶段，存货规模相应增加；另一方面，2024 年 AI 算力、数据中心建设等需求激增，为保证生产的连续性以及应对可能的市场需求，发行人进行战略备货。同行业可比公司中，寒武纪、海光信息、摩尔线程同样面临 AI 芯片市场需求激增的情况。寒武纪为保障供应链稳定及满足市场需求提前备货，其存货余额较 2023 年年末增加 493.11%。海光信息为应对更大规模的市场需求及供应链稳定，存货余额较 2023 年年末增加 396.05%，2024 年第四季度账面价值达到 54.25 亿元，环比增长 39%，连续五个季度增长。摩尔线程

也从 2023 年开始进行战略性备货，随着公司业务规模的扩大，存货金额总体呈上升趋势，2024 年存货余额较 2023 年年末增加 119.43%。上述情况表明，在当前 AI 算力、数据中心建设需求旺盛的行业环境下，同行业部分企业基于与发行人相似的市场判断和战略规划，也采取了增加存货储备的策略。而龙芯中科的主营产品为 CPU、景嘉微的主要产品为图形 GPU，与发行人产品类型存在差异，其存货余额变化相对而言受到 AI 算力需求激增的驱动较小。

综上，发行人存货规模快速增长具有合理性，与同行业公司的备货策略、存货规模增长幅度具有可比性，不存在重大差异。

（三）曦思 N100 正式量产后短期内即停产的原因及合理性，期后销售及结转情况、相关存货跌价准备计提是否充分；市场需求发生变化的具体过程，存货跌价准备计提时间晚于相关固定资产减值准备的原因及合理性，结合减值迹象出现的时间说明相关资产减值准备计提的及时性

1、曦思 N100 正式量产后短期内即停产的原因及合理性

曦思 N100 系发行人推出的首款核心产品，定位为面向云端和边端多种传统 AI 应用场景的智算推理 GPU，可广泛适配于智慧城市、智慧交通、云计算终端、短视频智能处理、视频直播实时分析等多元化应用领域。

在市场需求方面，近年来，随着人工智能技术的深度渗透及行业应用的规模化落地，人工智能领域对高性价比边侧推理算力芯片的需求呈现持续扩容态势。2023 年初，国内智慧城市建设进入攻坚阶段，智慧安防、智能交通等细分场景对多路视频编解码、实时智能分析的算力需求尤为迫切；同时，在国产信息技术自主可控政策导向下，具备自主知识产权的国产算力芯片替代需求显著提升。

在产品竞争方面，曦思 N100 在决定量产阶段，对标市场主流产品英伟达 T4 等竞品，已从算力性能、视频处理能力、国产化适配等多个维度构建起差异化竞争优势。

2023 年初，综合考量上述市场需求增长态势、产品竞争优势及成本控制能力，经审慎测算，曦思 N100 在目标应用场景下具备一定的市场需求和明确的盈利空间。据此，发行人于 2023 年 4 月正式量产。2023 年度，曦思 N100 板卡实际销售 2,042 张，平均单价 6,467.03 元/张（不含税），实现毛利约 77.67 万元，

量产决策及经营成果与预期一致，具备商业合理性。

2023 年末，生成式 AI 和大模型领域的市场需求大幅增长，发行人研发、生产重心转移至性能更优、场景适配更广的曦云 C500 系列训推一体产品，曦思 N100 不再进行晶圆加工阶段生产。该停产决策系发行人顺应行业技术迭代趋势、优化产品结构的主动调整，与市场环境变化及长期发展战略高度契合，具备商业合理性。

曦思 N100 芯片作为发行人首款产品，其设计、流片到量产，为发行人打通了从研发到生产再到销售的全业务流，完善了产品、生产供应链和销售渠道的资源储备，奠定了发行人的经营模式和商业模式，具有重要意义。

2、期后销售及结转情况、相关存货跌价准备计提是否充分

报告期各期末，发行人曦思 N100 相关存货期后销售及结转情况如下：

单位：万元

存货类型	2025 年 3 月末		2024 年末		2023 年末		2022 年末	
	期末余额	结转率	期末余额	结转率	期末余额	结转率	期末余额	结转率
原材料	289.26	0.00%	289.26	0.00%	256.19	0.00%	545.21	100.00%
半成品	5,525.37	2.62%	5,525.37	2.62%	6,120.91	60.03%	524.45	100.00%
库存商品	903.65	4.65%	1,666.90	49.46%	1,166.89	82.73%	12.45	100.00%
发出商品	-	-	5.91	100.00%	5.91	100.00%	-	-
合计	6,718.28	2.78%	7,487.44	13.02%	7,549.90	61.53%	1,082.11	100.00%

报告期各期末，发行人曦思 N100 相关存货跌价准备分库龄计提情况如下：

单位：万元

年份	项目	账面金额	0-180 天	181-360 天	361-720 天	720 天+
2025 年 3 月末	存货金额	6,718.28	57.73	2,827.01	3,833.54	-
	跌价金额	5,036.51	35.83	2,141.71	2,858.97	-
	跌价比例	74.97%	62.06%	75.76%	74.58%	-
2024 年末	存货金额	7,487.44	2,878.55	730.47	3,878.42	-
	跌价金额	5,500.42	2,166.64	473.18	2,860.60	-
	跌价比例	73.46%	75.27%	64.78%	73.76%	-
2023 年末	存货金额	7,549.90	7,487.22	62.68	-	-
	跌价金额	-	-	-	-	-

年份	项目	账面金额	0-180 天	181-360 天	361-720 天	720 天+
	跌价比例	-	-	-	-	-
2022 年末	存货金额	1,082.11	1,082.11	-	-	-
	跌价金额	-	-	-	-	-
	跌价比例	-	-	-	-	-

曦思 N100 产品于 2023 年 4 月进入正式量产阶段，2023 年度该产品实现销售收入 1,320.57 万元，不含税销售均价为 6,467.03 元/张，期末最后一笔销售的不含税单价为 7,000 元/张。发行人在测算曦思 N100 相关存货可变现净值时，对于处于各个生产工序的在产品及半成品，基于标准 BOM 清单按工序准确核算至完工状态所需发生的继续加工成本；在确定估计售价时，综合考虑了 2023 年度实际销售均价及期末最后一笔合同销售单价等近期实际交易价格，该等价格能够公允反映资产负债表日该产品的市场价值。经测算，发行人曦思 N100 相关存货（包括原材料、在产品、库存商品等）的可变现净值均高于其账面成本，未出现可变现净值低于成本的情形，因此 2023 年末发行人未对曦思 N100 相关存货计提跌价。2024 年上半年曦思 N100 产品的销售价格、毛利率也未有明显下滑，印证了 2023 年末未计提存货减值的合理性。

2024 年末，受市场需求结构调整、主流竞品低价倾销等因素影响，曦思 N100 产品销量大幅下降，期末订单价格已接近成本价，发生减值迹象。对于原材料，预计不再投产，全额计提跌价准备；对于半成品，以评估售价作为估计售价，减去根据标准 BOM 清单计算的按工序核算至完工状态所需发生的继续加工成本，减去估计的销售费用以及相关税费，计算半成品可变现净值，按照成本高于可变现净值的差额计提存货跌价准备；对于产成品，其中不良品全额计提存货跌价，有订单售价部分以订单售价作为估计售价，无订单售价的产成品以预计售价作为估计售价，以估计售价减去估计的销售费用以及相关税费后作为可变现净值，按照成本高于可变现净值的差额计提存货跌价准备。发行人综合分析市场竞争环境、市场需求趋势、产品竞争劣势，参考部分主流竞品定价在 2,500~3,200 元/张之间，计划以 2,500 元/张（含税）低价清库存，故以 2,500 元/张（含税）作为预计售价。结合期后销售情况来看，2025 年一季度发行人以单价最低 2,500 元/张（含税）的销售策略进行销售，实现销量 1,303 张，销量较 2024 年大幅上升，印证了 2024 年末以 2,500 元/张（含税）作为预计售价的谨慎性和合理性，相关存货跌价准备

计提充分。

3、市场需求发生变化的具体过程，存货跌价准备计提时间晚于相关固定资产减值准备的原因及合理性，结合减值迹象出现的时间说明相关资产减值准备计提的及时性

2023 年末，生成式 AI 和大模型领域的市场需求大幅增长，发行人研发、生产重心转移至性能更高、应用场景更广泛的训推一体曦云 C500 系列产品，曦思 N100 芯片不再进行晶圆加工阶段生产，发行人对晶圆加工工序相关的光罩、模具等固定资产全额计提减值准备。

《企业会计准则第 8 号——资产减值》规定，存在“资产已经或者将被闲置、终止使用或者计划提前处置”这一迹象，表明资产可能发生了减值。资产存在减值迹象的，应当估计其可收回金额。可收回金额的计量结果表明，资产的可收回金额低于其账面价值的，应当将资产的账面价值减记至可收回金额，减记的金额确认为资产减值损失，计入当期损益，同时计提相应的资产减值准备。可收回金额应当根据资产的公允价值减去处置费用后的净额与资产预计未来现金流量的现值两者之间较高者确定。

由于曦思 N100 芯片晶圆加工阶段已停产且该工序相关固定资产不具有其他变现价值，因此相关固定资产公允价值减去处置费用后的净额与资产预计未来现金流量的现值均为 0，即可回收金额为 0，发行人应当对相关固定资产全额计提减值。发行人基于资产使用情况变化，判断资产存在可能发生减值的迹象，对相关资产进行减值测试，测算可收回金额为 0，故对相关固定资产全额计提减值准备，具有及时性、合理性。

2023 年末，曦思 N100 芯片晶圆加工阶段停产系发行人基于市场需求变化、产品迭代而非存货销售情况所作出的公司战略调整。生成式 AI 大模型领域的井喷式增长促使市场对于传统人工智能推理芯片的需求下降，但曦思 N100 板卡的预计销售价格未明显下滑。《企业会计准则第 1 号——存货》规定，资产负债表日，存货应当按照成本与可变现净值孰低计量，存货成本高于其可变现净值的，应当计提存货跌价准备。2023 年末，曦思 N100 相关存货可变现净值高于成本，故未对存货计提跌价准备。2024 年上半年曦思 N100 板卡的销售价格、毛利率也

未有明显下滑，印证了 2023 年末未计提存货减值的合理性。

2024 年末，国内智能视频图像分析加速卡的市场需求下降，同时英伟达的 T4 产品市场上仍有大量囤货，对相同市场定位的曦思 N100 形成了一定冲击。因此，2024 年末，曦思 N100 系列市场需求有所下滑，产品价格处于劣势，相关存货存在减值迹象。发行人从主流竞品对比、市场需求趋势、产品核心竞争劣势等多个角度对曦思 N100 产品销售价格作出评估，计划以 2,500 元/张（含税）进行销售。2024 年末，发行人以 2,500 元/张（含税）作为估计售价，以估计售价减去至完工时估计将要发生的成本、估计的销售费用以及相关税费确定可变现净值，经测算，曦思 N100 产品可变现净值低于存货成本，应当计提存货跌价准备。2025 年一季度，曦思 N100 以最低 2,500 元/张（含税）作为单价进行销售，销量大幅提高，发行人以 2,500 元/张（含税）作为估计售价具有合理性。

综上，存货跌价准备计提时间恰当，存货跌价准备计提时间晚于相关固定资产减值准备具有合理性。

（四）结合各型号产品的市场需求、销售单价及毛利率变动、订单覆盖等情况，说明其他类型产品是否存在已被迭代、积压难以销售等情形，相应存货跌价准备计提的充分性

1、各型号产品销售单价、毛利率变动、订单覆盖情况

报告期内各型号产品销售单价（不含税）、毛利率变动情况如下：

单位：万元/张

产品型号	2025 年 1-3 月		2024 年		2023 年		2022 年	
	销售单价	毛利率	销售单价	毛利率	销售单价	毛利率	销售单价	毛利率
曦云 C500	3.70	54.44%	4.81	65.32%	5.69	65.48%	-	-
曦云 C550	5.99	68.42%	4.55	60.50%	-	-	-	-
曦思 N100	0.23	-0.44%	0.58	6.08%	0.65	5.88%	0.82	24.10%
曦思 N260	2.12	26.91%	2.27	26.70%	-	-	-	-

报告期内各期末各型号产品半成品、库存商品、发出商品、委托加工物资账面余额订单覆盖情况如下：

单位：万元

产品型号	2025 年 3 月末		2024 年末		2023 年末		2022 年末	
	期末余额	覆盖率	期末余额	覆盖率	期末余额	覆盖率	期末余额	覆盖率
曦云 C500	28,225.45	89.60%	41,036.50	58.38%	5,783.30	74.97%	-	-
曦云 C550	29,820.42	30.93%	28,031.67	2.85%	7.47	18495.13%	-	-
曦思 N100	6,429.03	0.03%	7,198.18	0.13%	7,293.70	0.26%	536.90	
曦思 N260	5,290.11	1.18%	3,825.82	-			-	-

2、其他类型产品是否存在已被迭代、积压难以销售等情形，相应存货跌价准备计提的充分性

（1）训推一体系列

从市场需求来看，训推一体系列产品对标英伟达 A100，可广泛应用于智算训练与推理、通用计算、AIforScience 等场景。中商产业研究院预测，2025 年中国人工智能芯片市场规模将增长至 1,530 亿元，2020 年至 2025 年复合增长率达 53%。中国作为全球人工智能芯片市场的重要参与者，预计将与全球人工智能芯片市场同频共振，成长空间巨大。作为人工智能芯片中占比最高的 GPU 芯片，可以预见其在人工智能计算领域具有广阔的发展前景。训推一体系列产品性能达到了国际上同类型主流高端芯片的水平，在国内处于领先地位。有着较强劲的市场需求。

从毛利率情况来看，2023 年、2024 年和 2025 年 1-3 月，曦云 C500 板卡毛利率分别为 65.48%、65.32%和 54.44%，2024 年和 2025 年 1-3 月，曦云 C550 板卡毛利率分别为 60.50%和 68.42%。整体而言，曦云 C 系列维持较高毛利率水平，不存在被迭代和减值迹象。

从订单覆盖情况来看，报告期内，训推一体系列相关存货整体订单覆盖率分别为 0、98.73%、35.84%、59.46%。训推一体系列在手订单覆盖率未达到 100%，主要系公司处于产品生产排期、应对国际贸易环境的不确定性、市场需求增长等因素的考虑，进行前瞻性备货，符合行业特性，具有商业合理性。报告期末，随着发行人训推一体系列相关产品销售规模放量，存货整体订单覆盖率相比 2024 年末明显提升，不存在已被迭代、积压难以销售等情形。

综上，训推一体系列产品不存在被迭代、积压难以销售的情况，报告期内维

持较高毛利率，除特殊事项外，无需计提存货跌价。

（2）智算推理系列

曦思 N100 系列相关分析详见本问题回复之“（三）曦思 N100 正式量产后短期内即停产的原因及合理性.....结合减值迹象出现的时间说明相关资产减值准备计提的及时性”。除曦思 N100 板卡外，发行人智算推理系列在售产品为曦思 N260 板卡。

从市场需求来看，由于大规模 AI 调用、复杂推理、多步骤推理场景等推理需求的爆发，促使推理算力成为推动产业发展的第二重驱动力，并下沉至不同的应用前景，是未来极具发展潜力的巨大市场。曦思 N260 系列系国产人工智能推理 GPU，相较曦思 N100 系列，曦思 N260 系列应用领域主要从面向传统人工智能场景升级迭代面向生成式人工智能场景，可广泛应用于大模型推理、生成式应用等场景，市场需求旺盛。

从销售单价、毛利率来看，2024 年和 2025 年 1-3 月，N260 销售单价分别为 2.27 万元/张和 2.12 万元/张，毛利率分别为 26.70%和 26.91%，均保持稳定。

从订单覆盖情况来看，报告期各期末曦思 N260 系列相关存货订单覆盖率较低，主要系统计时点、产品交付周期短等因素影响。根据 2025 年第 2 季度发行人未经审计财务数据，曦思 N260 板卡实现收入约 1,115.04 万元，平均销售单价约为 2.12 万元/张（不含税），价格保持稳定，期后销售情况良好。

整体来看，报告期内，智算推理系列产品除曦思 N100 产品出现销量下滑情况以外，其他型号产品均不存在被迭代、积压难以销售的情形，报告期内维持稳定的毛利率水平，存货跌价准备计提充分。

（五）结合与可比公司存货跌价准备具体计提政策及比例、存货库龄结构、产销率、毛利率水平比较情况，进一步说明发行人计提比例的合理性、是否和同行业可比公司之间存在较大差异

1、与同行业公司存货跌价计提政策比较

同行业可比公司中，寒武纪、海光信息、景嘉微、摩尔线程与发行人的存货跌价计提政策一致，龙芯中科除按照成本高于可变现净值的差额计提存货跌价准

备外，同时对库龄 3 年以上的存货全额计提存货跌价准备。整体而言，发行人与同行业公司存货跌价计提政策不存在明显差异，符合行业惯例。

可比公司	存货跌价计提政策
寒武纪	资产负债表日，存货采用成本与可变现净值孰低计量，按照成本高于可变现净值的差额计提存货跌价准备。直接用于出售的存货，在正常生产经营过程中以该存货的估计售价减去估计的销售费用和相关税费后的金额确定其可变现净值；需要经过加工的存货，在正常生产经营过程中以所生产的产成品的估计售价减去至完工时估计将要发生的成本、估计的销售费用和相关税费后的金额确定其可变现净值；资产负债表日，同一项存货中一部分有合同价格约定、其他部分不存在合同价格的，分别确定其可变现净值，并与其对应的成本进行比较，分别确定存货跌价准备的计提或转回的金额。
海光信息	资产负债表日，存货应当按照成本与可变现净值孰低计量。当存货成本高于其可变现净值的，应当计提存货跌价准备。可变现净值，是指在日常活动中，存货的估计售价减去至完工时估计将要发生的成本、估计的销售费用以及相关税费后的金额。产成品、库存商品和用于出售的材料等直接用于出售的商品存货，在正常生产经营过程中，以该存货的估计售价减去估计的销售费用和相关税费后的金额，确定其可变现净值；需要经过加工的材料存货，在正常生产经营过程中，以所生产的产成品的估计售价减去至完工时估计将要发生的成本、估计的销售费用和相关税费后的金额，确定其可变现净值；为执行销售合同或者劳务合同而持有的存货，其可变现净值以合同价格为基础计算，若持有存货的数量多于销售合同订购数量的，超出部分的存货的可变现净值以一般销售价格为基础计算。计提存货跌价准备后，如果以前减记存货价值的影响因素已经消失，导致存货的可变现净值高于其账面价值的，在原已计提的存货跌价准备金额内予以转回，转回的金额计入当期损益。按照组合计提存货跌价准备的组合类别及确定依据、不同类别存货可变现净值的确定依据
龙芯中科	资产负债表日，存货采用成本与可变现净值孰低计量，按照单个存货成本高于可变现净值的差额计提存货跌价准备。直接用于出售的存货，在正常生产经营过程中以该存货的估计售价减去估计的销售费用和相关税费后的金额确定其可变现净值；需要经过加工的存货，在正常生产经营过程中以所生产的产成品的估计售价减去至完工时估计将要发生的成本、估计的销售费用和相关税费后的金额确定其可变现净值；资产负债表日，同一项存货中一部分有合同价格约定、其他部分不存在合同价格的，分别确定其可变现净值，并与其对应的成本进行比较，分别确定存货跌价准备的计提或转回的金额。本公司存货采用成本与可变现净值孰低计量，按照单个存货成本高于可变现净值的差额计提存货跌价准备，同时对库龄 3 年以上的存货全额计提存货跌价准备。
景嘉微	期末对存货进行全面清查后，按存货的成本与可变现净值孰低提取或调整存货跌价准备。产成品、库存商品和用于出售的材料等直接用于出售的商品存货，在正常生产经营过程中，以该存货的估计售价减去估计的销售费用和相关税费后的金额，确定其可变现净值；需要经过加工的材料存货，在正常生产经营过程中，以所生产的产成品的估计售价减去至完工时估计将要发生的成本、估计的销售费用和相关税费后的金额，确定其可变现净值；为执行销售合同或者劳务合同而持有的存货，其可变现净值以合同价格为基础计算，若持有存货的数量多于销售合同订购数量的，超出部分的存货的可变现净值以一般销售价格为基础计算。期末按照单个存货项目计提存货跌价准备；但对于数量繁多、单价较低的存货，按照存货类别计提存货跌价准备；与在同一地区生产和销售的产品系列相关、具有相同或类似最终用途或目的，且难以与其他项目分开计量的存货，则合并计提存货跌价准备。以前减记存货价值的影响因素已经消失的，减记的金额予以恢复，并在原已计提的存货跌价准备金额内转回，转回的金额

可比公司	存货跌价计提政策
	计入当期损益。
摩尔线程	于资产负债表日，存货按照成本与可变现净值孰低计量，对成本高于可变现净值的，计提存货跌价准备，计入当期损益。可变现净值，是指在日常活动中，存货的估计售价减去至完工时估计将要发生的成本、估计的销售费用以及相关税费后的金额。计提存货跌价准备时，按单个存货项目计提。
发行人	资产负债表日，存货应当按照成本与可变现净值孰低计量。当存货成本高于其可变现净值的，应当计提存货跌价准备。可变现净值，是指在日常活动中，存货的估计售价减去至完工时估计将要发生的成本、估计的销售费用以及相关税费后的金额。产成品、库存商品和用于出售的材料等直接用于出售的商品存货，在正常生产经营过程中，以该存货的估计售价减去估计的销售费用和相关税费后的金额，确定其可变现净值；需要经过加工的材料存货，在正常生产经营过程中，以所生产的产成品的估计售价减去至完工时估计将要发生的成本、估计的销售费用和相关税费后的金额，确定其可变现净值；为执行销售合同或者劳务合同而持有的存货，其可变现净值以合同价格为基础计算，若持有存货的数量多于销售合同订购数量的，超出部分的存货的可变现净值以一般销售价格为基础计算。计提存货跌价准备后，如果以前减记存货价值的影响因素已经消失，导致存货的可变现净值高于其账面价值的，在原已计提的存货跌价准备金额内予以转回，转回的金额计入当期损益。

2、与同行业公司存货跌价计提比例比较

发行人与同行业可比公司存货跌价准备计提比例比较情况如下：

项目	2025 年 3 月末	2024 年末	2023 年末	2022 年末
寒武纪	未披露	13.21%	71.16%	42.48%
海光信息	未披露	1.93%	3.69%	3.40%
龙芯中科	未披露	15.05%	2.27%	2.27%
景嘉微	未披露	13.12%	8.93%	4.55%
摩尔线程	未披露	9.95%	16.03%	13.30%
可比公司均值	-	10.65%	20.42%	13.20%
发行人	9.89%	10.55%	0.00%	0.00%

同行业可比公司中，寒武纪的存货跌价准备计提比例明显高于其他公司。剔除寒武纪之后，2022 年末、2023 年末以及 2024 年末，其余四家同行业可比公司存货跌价准备计提比例平均值分别为 5.88%、7.73%和 10.01%。发行人 2022 年末和 2023 年末存货跌价准备计提比例低于同行业，主要系当年末发行人存货库龄较短且不存在明显减值迹象，2024 年末，发行人存货跌价准备计提与同行业平均水平接近。

同行业可比公司中，寒武纪 2022 年、2023 年提存货减值比例较高的原因为部分产品已处于生命周期末期，进一步对外销售的可能性较小，与发行人 2024

年末、2025 年 3 月末对曦思 N100 相关存货计提减值的原因和计提政策具有一定的相似性。2024 年末、2025 年 3 月末，曦思 N100 相关存货跌价计提分别为 73.46%、74.97%，与寒武纪 2023 年末存货跌价计提比例相近，具有一定的可比性。

3、与同行业库龄结构比较

报告期内，发行人存货库龄结构如下：

单位：万元

年份	项目	合计	0-180 天	181-360 天	361-720 天	720 天+
2025 年 3 月末	金额	89,031.46	73,751.73	10,152.35	5,102.90	24.48
	占比	100.00%	82.84%	11.40%	5.73%	0.03%
2024 年末	金额	86,844.49	80,547.82	939.16	5,332.99	24.51
	占比	100.00%	92.75%	1.08%	6.14%	0.03%
2023 年末	金额	18,730.50	18,619.39	76.70	34.41	-
	占比	100.00%	99.41%	0.41%	0.18%	-
2022 年末	金额	1,767.44	1,715.55	51.89	-	-
	占比	100.00%	97.06%	2.94%	-	-

同行业可比公司中，除寒武纪外，均未披露报告期内存货库龄结构。寒武纪在其 2025 年向特定对象发行股票审核问询函回复中披露了存货的库龄情况，2022 年末、2023 年末、2024 年末和 2025 年 3 月末，寒武纪存货库龄结构如下表所示：

单位：万元

存货库龄	2025 年 3 月末		2024 年末		2023 年末		2022 年末	
	金额	占比	金额	占比	金额	占比	金额	占比
1 年以内	277,129.38	91.60%	179,490.19	87.81%	11,392.01	33.06%	39,592.18	79.32%
12-18 个月	1,997.60	0.66%	2,012.23	0.98%	9,274.92	26.91%	6,281.54	12.58%
18 个月以上	23,425.14	7.74%	22,894.07	11.20%	13,794.85	40.03%	4,042.55	8.10%
合计	302,552.12	100.00%	204,396.49	100.00%	34,461.78	100.00%	49,916.27	100.00%

报告期各期末，寒武纪 1 年以上库龄存货占比分别为 20.68%、66.94%、12.19% 和 8.40%，同期发行人 1 年以上库龄存货占比分别为 0、0.18%、6.17%和 5.76%，低于寒武纪。

综上所述，发行人对长库龄存货计提跌价充分，长库龄存货占比不存在大幅高于同行业可比公司的情形。

4、与同行业产销率比较

报告期内，发行人产销率与同行业可比公司对比情况如下：

公司名称	产品类型	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
寒武纪	智能芯片及加速卡	未披露	155.50%	11.43%	37.33%
海光信息	高端处理器	未披露	75.35%	153.29%	107.83%
龙芯中科	处理器及配套芯片	未披露	86.10%	59.75%	31.43%
景嘉微	计算机、通信和其他电子设备制造业	未披露	95.65%	84.84%	102.11%
摩尔线程	服务器级	未披露	56.00%	44.00%	32.00%
可比公司均值		-	93.72%	70.66%	62.14%
发行人	训推一体 GPU 板卡	132.96%	41.78%	18.84%	-
	智算推理 GPU 板卡	126.64%	17.69%	45.49%	21.67%
	总计	132.02%	39.82%	39.00%	21.67%

2022 年-2024 年末，发行人产品平均产销率分别为 21.67%、39.00%、39.82%，低于行业平均水平 62.14%、70.66%、93.72%，主要原因系发行人处于前期业务快速拓展阶段，销售逐步起量，为应对旺盛的行业下游需求，采取了较高比例的提前备货策略。2025 年 3 月末，随着市场需求释放，发行人产销率提升至 132.02%。分产品而言：

报告期各期末，发行人训推一体系列产品产销率持续提升，2025 年一季度达 132.96%，产销率变化情况与发行人备货政策、生产节奏把控以及市场需求增长的情况相一致。报告期内，训推一体系列产品销售规模不断扩大，除特殊事项外，相关存货未见减值迹象，因此存货跌价比例较低具有合理性。

报告期各期末，发行人智算推理系列产品产销率具有一定波动。2023 年 4 月，智算推理系列曦思 N100 产品进入量产阶段，销售情况良好，因此产销率较 2022 年上升至 45.49%，符合业务发展初期市场拓展的正常表现。2024 年，受市场需求大幅下降影响，产品产销率相应下降至 17.69%。2025 年一季度，一方面，发行人曦思 N100 产品已停产，另一方面，发行人对曦思 N100 产品采取降价销售策略，销售量大幅上升，产销率也随之大幅上升至 126.64%。发行人于 2024

年末对智算推理系列曦思 N100 相关存货充分计提减值，与产品产销率变化具有一致性。

可比公司中，寒武纪 2021-2024 年末的产销率为 83.60%、37.33%、11.43%、155.50%，其中，2022 年末和 2023 年末，寒武纪的产销率显著降低，但 2024 年末产销率明显提升，主要原因为寒武纪 2024 年生产量较上年减少 88.90%，而销售量增长 51.05%，整体产销量增加至 155.50%，寒武纪 2022 年以来产销率的波动情况与发行人曦思 N100 的产销率波动具有一定的可比性，根据本问询回复报告之“10.关于存货和固定资产/一/（五）/2、与同行业公司存货跌价计提比例比较”分析，寒武纪存货计提减值与发行人对曦思 N100 相关存货计提减值的原因和计提政策具有一定的可比性。

5、与同行业毛利率水平比较

报告期内，发行人与同行业可比上市公司综合毛利率比较情况如下：

可比上市公司	2025 年 1-3 月	2024 年	2023 年	2022 年
寒武纪	55.99%	56.71%	69.16%	65.76%
海光信息	61.19%	63.72%	59.67%	52.42%
龙芯中科	38.00%	31.04%	36.06%	47.09%
景嘉微	32.37%	43.70%	60.32%	65.01%
摩尔线程	未披露	70.71%	25.87%	-70.08%
可比公司均值	46.89%	53.18%	50.22%	32.04%
发行人	55.26%	53.43%	62.88%	24.10%

2022 年，发行人毛利率为 24.10%，低于同行业平均水平 32.04%，主要原因为当期发行人主要产品均未正式量产，收入产品结构与同行业公司存在较大差异，不具有可比性。2023 年度、2024 年度和 2025 年 1-3 月，发行人毛利率分别为 62.88%、53.43%和 55.26%，高于同行业平均水平 50.22%、53.18%、46.89%，而 2024 年末，发行人存货跌价准备计提与同行业平均水平接近，因此总体而言，发行人的存货跌价计提较为谨慎。

根据本题“（四）/1、各型号产品销售单价、毛利率变动、订单覆盖情况”的分析，报告期内，除了曦思 N100 外，其他产品曦云 C500、曦云 C550、曦思 N260 产品整体毛利率水平良好，因此，2024 年以来，发行人对 N100 计提大额

减值，除特殊事项外，未对曦云 C500、曦云 C550、曦思 N260 产品相关存货计提减值，计提较为谨慎、合理。

(六) 2024 年存放在供应商的固定资产全额计提减值准备的具体情况后续处置情况，是否存在其他固定资产涉及相关情形、资产减值准备计提的充分性

1、2024 年存放在供应商的固定资产全额计提减值准备的具体情况后续处置情况

报告期各期末，发行人涉及减值的固定资产情况如下：

单位：万元

项目	2025 年 3 月末	2024 年末	2023 年末	2022 年末
账面原值	11,983.13	11,983.13	5,541.79	-
累计折旧	5,390.43	5,390.43	2,754.87	-
账面余额	6,592.70	6,592.70	2,786.91	-
减值准备	6,592.70	6,592.70	2,786.91	-
账面价值	-	-	-	-

2023 年末，发行人固定资产减值准备余额为 2,786.91 万元，系曦思 N100 相关专用设备减值：曦思 N100 产品系公司推出的首款智算推理芯片产品，后续公司持续加大研发投入推出性能更高、应用场景更广泛的训推一体产品曦云 C 系列，2023 年末由于产品迭代原因，发行人停止曦思 N100 芯片晶圆加工阶段的生产，致使该工序相关固定资产无法继续使用且不具有其他变现价值，因此相关固定资产公允价值减去处置费用后的净额与资产预计未来现金流量的现值均为 0，即可回收金额为 0，发行人对相关固定资产全额计提减值。

2024 年末固定资产减值准备余额为 6,592.70 万元，其中新增 3,805.79 万元系曦云 C500 相关光罩减值：2024 年 11 月，美国禁令导致发行人无法继续向晶圆代工厂商下达训推一体曦云 C500 系列产品的生产订单，进而导致发行人光罩等固定资产无法继续使用，因此发行人对相关光罩资产全额计提减值准备。

上述减值资产暂无处置计划，在供应商未强制要求发行人处置的情况下，相关减值资产仍由供应商保管。

2、是否存在其他固定资产涉及相关情形、资产减值准备计提的充分性

发行人其他固定资产不涉及相关情况，固定资产减值迹象分析如下：

资产减值迹象	公司具体情况	是否存在减值迹象
资产的市价当期大幅度下跌，其跌幅明显高于因时间的推移或者正常使用而预计的下跌	除上述减值资产外，公司固定资产均处于正常使用状态，其资产的市价当期无大幅度下跌情况	否
企业经营所处的经济、技术或者法律等环境以及资产所处的市场在当期或者将在近期发生重大变化，从而对企业产生不利影响	除上述美国禁令导致无法向境外供应商下达曦云 C500 芯片生产订单外，无其他经济、技术或者法律等环境重大变化	否
市场利率或者其他市场投资报酬率在当期已经提高，从而影响企业计算资产预计未来现金流量现值的折现率，导致资产可收回金额大幅度降低	市场基准利率并未发生大幅上调	否
有证据表明资产已经陈旧过时或者其实体已经损坏	经盘点，各期末固定资产不存在损毁或陈旧过时	否
资产已经或者将被闲置、终止使用或者计划提前处置	经盘点，除上述存放在供应商处的已计提的减值资产外，各期末固定资产不存在被闲置、终止使用或者计划提前处置	否
企业内部报告的证据表明资产的经济绩效已经低于或者将低于预期，如资产所创造的净现金流量或者实现的营业利润（或者亏损）远远低于（或者高于）预计金额等	除曦思 N100 产品已计提减值外，公司产品 C500 毛利率正常。	否
其他表明资产可能已经发生减值的迹象	曦云 C600 与曦彩 G100 研发投入正常，相关设备无减值迹象	否

综上所述，除已计提减值的光罩等固定资产外，发行人其余曦云 C 系列相关固定资产仍可用于正常生产，曦云 C 系列作为主要产品正常生产销售，研发投入与毛利正常，无减值迹象。下一代基于国产供应链的曦云 C600 系列已于 2025 年 7 月回片，预计 2025 年底前进入风险量产，相关资产未发现减值迹象；曦彩 G100 处于在研阶段，尚未流片，无减值迹象。此外，通过现场盘点查看发行人的办公用固定资产与通用固定资产实际状况，未见闲置或损坏，未发现减值迹象。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、申报会计师执行了如下核查程序：

1、获取发行人报告期各期末分产品型号、存货类型、库龄的存货明细表，选取样本检查库龄划分是否准确。

2、获取发行人报告期各期末在手订单明细，复算各类存货订单支持率情况。

3、获取发行人截至 2025 年 6 月 30 日存货出入库明细表，了解发行人各类存货期后生产领用、销售情况。

4、访谈发行人相关部门负责人，了解发行人长库龄存货的形成原因、期后领用处置情况，评估其形成的合理性。

5、访谈发行人相关部门负责人，了解发行人存货备货政策、主要原材料的采购周期、主要产品的生产周期、存货变动原因、备货必要性等情况；结合对各类存货与主营业务成本进行勾稽，评估存货快速增长、结构变化较大的合理性；获取并查阅同行业可比公司的招股说明书、年度报告等公开资料，了解同行业可比公司存货变化与发行人是否存在差异以及差异形成的原因。

6、了解和评估发行人存货跌价准备计提政策的合理性，并检查计提政策是否得到一贯执行；了解存货可变现净值的确认依据，检查是否符合实际经营情况及企业会计准则的相关规定；查询同行业可比公司存货跌价准备计提政策，评价存货跌价准备政策的合理性。

7、访谈发行人相关部门负责人，了解各型号产品的市场需求、产品竞争迭代情况，评估存货跌价准备计提的充分性。

8、了解固定资产对应产品、项目基本情况以及行业相关政策；对企业高管进行访谈，了解美国相关政策对公司生产经营以及对长期资产减值的影响；对长期资产进行减值测试，检查固定资产对应产品毛利率与研发投入情况；进行盘点现场查看固定资产实际状况。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、发行人曦思 N100 产品因市场需求变化、产品迭代等原因存在库龄较长情形，发行人已于 2024 年末对曦思 N100 产品计提了充分的跌价准备。除曦思 N100 产品之外，发行人其他产品相关存货长库龄比例较低，不存在因市场需求变化、产品竞争力下降导致的长期呆滞存货的情形。

2、发行人存货规模与备货政策、原材料采购周期、产品生产及销售周期相

匹配，报告期各期末存货快速增长、结构变化较大具有合理性；报告期内，存货余额变化与采购种类及金额、成本结转相匹配，与同行业可比公司不存在重大差异。

3、曦思 N100 正式量产后短期内即停产具有合理性，期后销售及结转情况较好，相关存货跌价准备计提充分。存货跌价准备计提时间晚于相关固定资产减值准备具有合理性，资产减值准备计提具有及时性。

4、除曦思 N100 外，发行人其他类型产品不存在被迭代、难以销售的情形，存货跌价准备计提充分。

5、发行人存货跌价计提比例合理，与行业公司存货跌价计提政策不存在明显差异。

6、发行人已说明 2024 年存放在供应商的固定资产全额计提减值准备的具体情况后续处置情况，其余固定资产对应产品毛利率与研发投入正常，不存在因产品迭代停产、政策变化等原因导致无法使用的情况，发行人固定资产减值准备计提充分。

7、就财务报表整体的公允反映而言，发行人存货跌价准备的计提在所有重大方面符合企业会计准则的相关规定，发行人固定资产减值准备的计提在所有重大方面符合企业会计准则的相关规定。

11.关于期间费用

根据申报材料：（1）报告期各期，发行人期间费用中职工薪酬占比较高；（2）研发费用主要由职工薪酬、折旧和摊销、委外开发、股份支付、材料及工程试制费等构成；折旧和摊销金额为 11,650.12 万元、12,235.34 万元、13,306.10 万元和 3,999.91 万元，以正式量产来划分折旧摊销归属研发费用或成本；（3）研发费用中，委外研发费用金额为 6,752.31 万元、2,999.21 万元、9,468.99 万元和 1,136.29 万元；（4）各期销售费用中，服务费为 404.04 万元、340.79 万元、2,217.21 万元和 787.36 万元；（5）报告期各期，计入管理费用的服务费分别为 3,625.38 万元、4,450.06 万元、8,510.42 万元和 9,181.51 万元，主要包括融资顾问费、律师费、技术支持服务费等。

请发行人披露：（1）报告期内研发、销售、管理人员的平均数量、人均薪酬、人员结构的变动情况，人均薪酬是否与同行业可比公司、当地人均薪酬水平一致；（2）进入正式量产阶段对于客户订单的具体数量标准、与行业惯例是否相符，结合各研发项目正式量产前后的客户订单情况，说明相关标准的客观性和准确性；（3）委外研发金额大幅波动的原因及合理性、与研发项目的匹配性；（4）管理费用中服务费增长的原因。

请保荐机构、申报会计师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）报告期内研发、销售、管理人员的平均数量、人均薪酬、人员结构的变动情况，人均薪酬是否与同行业可比公司、当地人均薪酬水平一致

1、报告期内研发、销售、管理人员的平均数量、人均薪酬、人员结构的变动情况

报告期各期，发行人销售、管理、研发人员的平均数量、人均薪酬、人员结构情况如下：

人员类别	项目	2025 年 1-3 月	2024 年	2023 年	2022 年
研发人员	平均数量（人）	682	672	602	542

人员类别	项目	2025 年 1-3 月	2024 年	2023 年	2022 年
	人均薪酬（万元/年）	76.06	73.28	74.77	71.64
	人员结构占比	77.50%	79.91%	82.69%	85.22%
销售人员	平均数量（人）	101	78	51	29
	人均薪酬（万元/年）	69.16	70.94	70.89	78.83
	人员结构占比	11.48%	9.27%	7.01%	4.56%
管理人员	平均数量（人）	97	91	75	65
	人均薪酬（万元/年）	74.65	68.72	69.83	57.86
	人员结构占比	11.02%	10.82%	10.30%	10.22%

注 1：上表平均数量(人)为发行人当期各月各类人员人数之和/月份数取整后的平均人员数；人均薪酬=当期各类人员薪酬总金额/当期各类人员平均数量，2025 年 1-3 月人均薪酬为年化数据，下同；人员结构占比=各职能部门平均数量/销售、管理、研发部门平均数量之和；

注 2：上表管理人员人均薪酬不包含发放的离职补偿金。

（1）研发人员

报告期各期，发行人研发人员平均数量分别为 542 人、602 人、672 人及 682 人，呈逐年增长趋势，主要系发行人为进一步提升技术创新能力、增强公司自主研发实力，2022 年以来扩招研发人员，与发行人营业收入规模逐年增长的趋势及持续加大研发投入以构建核心技术体系的战略相符。发行人人员结构以研发人员为主，报告期各期研发人员结构占比超过 77%，较为稳定，反映出公司以技术驱动为核心的发展模式。

报告期各期，发行人研发人员人均薪酬分别为 71.64 万元/年、74.77 万元/年、73.28 万元/年及 76.06 万元/年，整体呈上升趋势，主要系公司为持续吸引和保留国内外高端研发人才而提升研发人员薪酬，以保持行业竞争力所致，具有合理性。

（2）销售人员

报告期各期，发行人销售人员平均数量分别为 29 人、51 人、87 人及 101 人，人员结构占比分别为 4.56%、7.01%、9.27%及 11.48%，平均数量及人员结构占比呈逐年增长趋势，主要原因系发行人业务进入快速增长期，为适应业务规模的快速扩张、进一步拓展公司业务、提升市场销售能力，2022 年以来扩充建设销售人员队伍，与发行人营业收入规模逐年增长的趋势一致。

报告期各期，发行人销售人员人均薪酬分别为 78.83 万元/年、70.89 万元/年、70.94 万元/年及 69.16 万元/年，其中 2023 年以来销售人员人均薪酬基本保持稳定，但低于 2022 年，主要系公司自 2023 年起加大了销售团队建设力度，在保留核心资深销售人员的同时，引进了一批具有行业经验且富有开拓精神的年轻销售人才。由于新进人员薪酬水平普遍低于原有资深销售人员，且其在销售团队中的占比逐步提升，从而对整体人均薪酬产生了一定稀释效应，符合公司处于业务扩张期的实际情况，具有合理性。

(3) 管理人员

报告期各期，发行人管理人员平均数量分别为 65 人、75 人、91 人及 97 人，呈逐年上升趋势，系发行人为支撑公司业务规模的快速扩张和组织管理的复杂性提升，适度扩充了管理团队，与公司发展阶段相匹配，与发行人营业收入规模逐年增长的趋势一致。报告期各期，发行人管理人员占比保持在 10%左右，较为稳定。

报告期各期，发行人管理人员人均薪酬分别为 57.86 万元、69.83 万元、68.72 万元及 74.65 万元，整体呈增长趋势。其中，发行人 2022 年管理人员人均薪酬较低，主要系发行人发展初期管理人员的薪酬结构所致；2023 年至 2024 年，发行人管理人员人均薪酬水平保持稳定并略有增长，2025 年 1-3 月人均薪酬较高，主要系公司管理人员薪酬水平随着公司业务规模的扩张及经营业绩的提升而增长，具有合理性。

综上所述，报告期内，发行人研发、销售、管理人员数量呈逐年上升趋势，与公司所处快速增长期扩充人才队伍的情况相符；报告期内发行人人员结构整体较为稳定，销售人员占比随公司业务扩张呈上升趋势，具有合理性。报告期各期，发行人研发人员、管理人员人均薪酬呈上升趋势，系公司为吸引和留住高水平人才而提升薪酬待遇所致，发行人销售人员人均薪酬略有下降，系发行人报告期内扩充销售人员队伍，新入职员工入职当年薪酬水平低于往年入职员工所致，具有合理性。

2、报告期内研发、销售、管理人员人均薪酬是否与同行业可比公司人均薪酬水平一致

(1) 研发人员

2022 年-2024 年，发行人研发人员人均薪酬处于同行业中游水平，整体高于同行业可比公司平均值，但与行业内具有较强研发实力的可比公司薪酬水平较为接近，与同行业没有明显的差异。这表明发行人对标行业头部企业，以吸引和留住最顶尖的芯片设计和研发人才，保障核心技术领先和团队稳定，具有合理性。

(2) 销售人员

2022 年-2024 年，发行人销售人员人均薪酬整体高于同行业可比公司平均值，主要系发行人产品面向数据中心、云服务等高端市场，对销售人员的行业经验、技术理解及客户资源要求较高，发行人业务进入快速增长期，提供较高的薪酬待遇以高效激励销售团队、快速提升市场份额所致，具有合理性。

(3) 管理人员

2022 年-2024 年，发行人管理人员人均薪酬处于同行业中游水平，整体高于同行业可比公司平均值，主要系发行人管理团队由具有深厚行业背景和国际化视野的资深专家组成，发行人向其提供较高薪酬主要基于其战略规划能力、运营管理经验和对公司长期价值的贡献，有助于发行人在激烈的市场竞争中取得优势，具有合理。

综上所述，报告期各期，发行人研发人员、销售人员、管理人员人均薪酬整体高于同行业可比公司平均值，主要系发行人对标行业领先企业，提供有竞争力的薪酬待遇以吸引高价值人才驱动公司业绩增长，提高公司市场竞争力，具有合理性。

3、报告期内研发、销售、管理人员人均薪酬是否与当地人均薪酬水平一致

发行人根据子公司或参股公司最近一年营业收入、总资产或净资产任一单个指标占发行人同期合并财务报表相应指标比重超过 5%作为重要性依据认定，共有 4 家重要子公司，分别为：沐曦灵智科技（杭州）有限公司、沐曦集成电路（南京）有限公司、沐曦科技（成都）有限公司和沐曦科技（北京）有限公司。

报告期内，发行人各地销售人员、管理人员、研发人员平均薪酬与当地薪酬差异情况如下：

单位：万元/年

地区	2025 年 1-3 月		2024 年		2023 年		2022 年	
	发行人平均薪酬	当地平均薪酬	发行人平均薪酬	当地平均薪酬	发行人平均薪酬	当地平均薪酬	发行人平均薪酬	当地平均薪酬
上海	88.60	未披露	80.97	未披露	81.11	14.77	76.93	14.62
杭州	64.67	未披露	56.59	16.72	67.02	16.17	-	15.72
南京	70.68	未披露	63.63	16.57	61.32	15.97	57.58	15.52
成都	72.84	未披露	61.40	12.78	59.96	12.71	52.44	12.00
北京	85.80	未披露	80.24	未披露	82.63	22.46	77.93	17.85

注：各市平均薪酬来源于当地统计局发布的城镇非私营单位在岗职工平均工资数据；沐曦灵智科技（杭州）有限公司 2022 年尚未成立，故发行人无 2022 年杭州平均薪酬。

由上表数据可知，报告期内，发行人母公司及重要子公司的人均薪酬均高于当地平均薪酬，主要系：

（1）从行业和人才结构来看，发行人作为一家专注于高性能 GPU 芯片研发与设计的企业，所处集成电路设计行业属于高技术密集型领域，对人才的专业技能和经验要求较高。报告期各期，发行人研发人员占比超过 77%，该等人员从事高技术门槛工作，薪酬水平远高于普通行业及岗位的平均值；发行人销售人员需具备深度的行业知识和客户资源，管理人员多为具备多年行业经验的资深专家，其市场价值和对发行人战略发展的贡献决定了其高薪酬水平；

（2）从市场竞争环境来看，发行人所处的集成电路设计行业属于高技术、高投入、高风险领域，其人才薪酬水平普遍高于社会平均水平，发行人为吸引和保留国内外顶尖研发、销售、管理，需提供具有国际市场竞争力的薪酬待遇，特别是在与国际及国内同行业公司的人才竞争中，高薪酬是保障团队稳定、技术领先和企业持续发展的重要手段；

（3）从薪酬构成与激励机制来看，发行人已建立了完善的薪酬管理制度，员工薪酬构成包括基本工资、绩效奖励、股权激励等，薪酬支付均基于岗位价值、个人能力、工作资历、绩效考核及市场薪酬水平进行综合评估，符合发行人的内部治理要求和行业惯例；

（4）各地统计局发布的平均工资是该市各行业所有类型人员平均工资，此

处列示的发行人平均薪酬为研发人员、销售人员、管理人员的平均薪酬，此类人员属于高收入群体，平均薪酬较高。

由此，报告期内，发行人母公司及重要子公司的人均薪酬均高于当地平均薪酬，是由发行人所处的行业特殊性、市场竞争、人才结构以及激励机制等多种因素共同作用的结果，具备充分的商业合理性与必要性。

（二）进入正式量产阶段对于客户订单的具体数量标准、与行业惯例是否相符，结合各研发项目正式量产前后的客户订单情况，说明相关标准的客观性和准确性

发行人研发项目的流程主要分可行性与计划阶段、产品开发阶段、新产品导入阶段和量产阶段。其中，新产品导入阶段主要由发行人研发部门主导，公司将芯片设计版图交付晶圆制造厂流片，同时持续开发配套基础软件；在芯片回片后，经过对芯片各项功能和性能进行验证和测试，包括芯片点亮、功能验证、产品系统级验证等，公司评估产品风险量产质量标准，评审后进入风险量产阶段。

进入风险量产阶段后，发行人持续对研发产品进行测试，收集性能、可靠性、封装风险、封测良率等数据，确保其各项指标满足产品需求和设计预期，并组织公司内部相关部门进行评审，形成测试报告与评审结论。

发行人芯片产品所处风险量产阶段与大规模正式量产阶段最主要的区别是风险量产生产规模较小，主要是为了测试和优化生产工艺且产品不会大量投放市场，主要用于内部测试和客户验证，或小批量售卖；大规模正式量产生产规模较大，且会全面投放市场，以满足客户需求并实现商业目标。故为了确保正式量产阶段能够持续、稳定地生产出符合要求的芯片和板卡，发行人在评估产品是否进入正式量产阶段前，每个研发产品必须要完成针对于芯片的各项性能测试，GPU板卡的工程验证测试（EVT）、设计验证测试（DVT）、生产验证测试（PV）、系统验证测试（SVT）以及配套软件的各项功能的测试，只有在前述测试全部完成，评审通过后，研发产品才具有进入大规模正式量产的可行性。同时，综合考虑已获取的客户订单情况，以验证产品的商业化前景。最终，管理层组织市场部、运营部、研发部门等对产品的客户需求、市场前景、供应链产能匹配度以及整体功能性能达标情况进行综合评审。当确认产品整体功能及性能达到设计要求、质

量要求，并结合客户订单情况经管理层评审确认后，产品即进入正式量产阶段。由此，发行人在评估进入正式量产阶段的过程中，客户订单主要为发行人产品通过客户验证、具有商业化前景的佐证，不设定具体的数量标准；各项必须完成的测试是能否进入大规模量产的主要依据以及是发行人一贯执行的客观标准。

进入正式量产阶段的认定标准在半导体设计企业中普遍体现为产品整体功能、性能、可靠性等指标满足设计预期，具备稳定批量生产能力，并获得一定市场验证。同行业公司进入正式量产阶段的标准如下：

公司简称	进入量产阶段的标准定义
寒武纪	样品阶段：样品试生产并对晶圆和芯片进行电气、时序、功能等方面的测试，对芯片板卡进行硬件测试，确保芯片功能和性能符合设计要求；软件项目针对所适配的芯片和应用场景，验证软件的功能和性能符合设计要求； 发布阶段：当芯片、加速卡及配套基础系统软件产品的整体功能及性能达到设计要求及质量要求时，公司委托代工厂进行大规模生产，最终形成可以面向市场开展销售的产品。
海光信息	实施阶段：从事芯片产品的设计与实现的相关技术开发，包括芯片前后端设计、基础软件开发、流片、硅后验证测试等。支持迭代研发的芯片产品的整体功能和性能达到设计要求及预期指标。 结项阶段：芯片产品测试报告汇总分析，产品定型与量产。完成项目结项评审。
景嘉微	主要产品的研发流程分为以下四个阶段：论证阶段、方案阶段、工程研制阶段、设计定型阶段。公司在完成集成电路版图的设计后，将版图交由晶圆制造厂商，由晶圆制造厂商按照公司版图生产出对应晶圆后，再交由封装和测试厂商进行晶圆的封装和测试工作，公司取得芯片成品后视芯片订单的具体情况通过自有测试车间或者委托第三方厂商进行测试，最后销售给客户。
龙芯中科	3A4000 芯片及配套 7A1000 桥片完成在 BV08（客户）桌面产品上的验证测试工作，进入常态化量产阶段.....
摩尔线程	芯片、软件和系统研发同步进行，管理层在各里程碑和最终项目交付前对项目进展及测试结果进行风险评审，决定是否进入下一阶段或投产。量产产品需通过量产测试，通过测试的产品将按计划逐步投放市场。
晶丰明源	风险量产阶段属于正式量产前的必要环节，具体指产品在多个客户端批量试生产并进行试用，期间如存在问题需要在多个客户端重新进行验证，如验证失败，则终止该产品生产。
杰华特	最终经由晶圆代工厂和封装测试厂的配合完成样品的生产、封装、测试，再经公司及下游应用厂商评估确认，达到量产标准。 量产认证阶段：经客户试用确认合格的产品，将进行市场推广，并开展小批量生产。应用市场部和销售部将协同公司各部门以及外部经销商，完成产品在下游目标客户处的准入工作，最终实现芯片的量产。

注：表述来源为各公司公告。

由此可见，同行业内公司普遍结合内部评审及客户验证/订单情况确定产品进入正式量产阶段，结合客户订单作为正式量产标准已在半导体设计行业中形成普遍做法，发行人判断产品进入正式量产阶段的标准符合行业惯例。

发行人报告期内共两款芯片产品进入正式量产阶段，分别是智能推理芯片曦思 N100 于 2022 年 1 月交付流片，并于 2023 年 4 月正式量产；训推一体芯片曦云 C500 于 2022 年 12 月交付流片，并于 2024 年 2 月正式量产。各研发产品量产前后客户订单数量（折算 GPU 芯片数量）情况如下表：

单位：颗

型号	量产前订单数量	正式量产时间点	量产后当年订单数量
曦思 N100	1,310	2023 年 4 月	1,068
曦云 C500	1,758	2024 年 2 月	27,758

由于研发产品的类型不同（如曦思 N100 为智能推理芯片、曦云 C500 为训推一体芯片）、应用场景不同（如曦思 N100 主要应用于云端及边端推理、视频转码等，曦云 C500 主要应用于云端智算、通用计算）导致不同类型研发产品的客户订单采购数量差异较明显，故发行人无法用统一的具体数量标准界定研发项目产品进入正式量产流程，而是结合订单及市场前景综合评估。如上表所示，各研发产品正式量产前发行人获取的曦云 C500 系列产品订单数量大于曦思 N100 系列产品，主要系训推一体芯片产品主要应用大模型训练等场景，单次订单采购量较大。

对于发行人高性能 GPU 芯片产品，客户订单不仅是市场需求的体现，更反映产品经过客户测试验证后的商业接受度。发行人客户在采购前一般需完成性能测试、系统验证、软硬件适配等环节，因此订单的获取本身即体现产品商业可行性。发行人在产品功能、性能、封装良率等技术指标达标基础上，结合客户订单数量作为补充判断标准，有助于确保产品不仅技术成熟，也具备市场接受度，从而更加客观、精准地判断量产时点，具有审慎性和稳健性。

此外，发行人研发产品是否进入正式量产对公司营业成本和研发费用归集产生的影响主要是各芯片产品专用的光罩和 IP 授权的折旧摊销费用，在正式量产前光罩与 IP 授权的折旧摊销计入研发费用，正式量产后则计入生产成本。但发行人进入正式量产的时点对公司营业成本和研发费用归集的准确性产生的影响较小，主要系：①对于风险量产形成的样卡，测试性能较为稳定，具有一定的经济价值，公司将该部分样卡按照《企业会计准则解释第 15 号》的要求，从当期研发费用中结转对应的材料、折旧摊销费用至当期存货，作为库存商品核算；②

研发产品正式量产当年光罩与专用 IP 授权的折旧摊销金额占研发费用比例不超过 2%，占比较低，影响较小，具体如下：

单位：万元

2024 年度				
对应产品	归集科目	光罩	专用 IP 授权	合计占当期研发费用比例
曦云 C500	研发费用	978.99	334.19	1.46%
	生产成本	2,158.97	1,457.66	4.01%
2023 年度				
对应产品	归集科目	光罩	专用 IP 授权	合计占当期研发费用比例
曦思 N100	研发费用	611.59	548.78	1.66%
	生产成本	917.39	823.16	2.49%

注：曦思N100产品2023年度计入生产成本的金额较小主要系曦思N100芯片于2023年11月基本进入停产状态，公司于当月全额计提光罩及IP的减值准备所致。

综上所述，基于 GPU 芯片研发周期长、投入高等行业特性，发行人以风险量产过程中收集的性能、可靠性、封装风险、封测良率等数据作为支撑，在此基础上结合客户订单支撑、市场前景综合评估，由管理层组织多部门进行交叉评审后进入正式量产，充分考虑了 GPU 芯片高研发投入、长周期及应用场景多样化的特点，符合行业惯例。由于不同芯片产品的订单规模受应用场景和客户需求影响较大，发行人未设定统一订单数量标准，而是综合考虑多种因素，并最终通过跨部门综合评审确定产品进入量产阶段，发行人对于进入正式量产阶段的标准选择客观、准确。

（三）委外研发金额大幅波动的原因及合理性、与研发项目的匹配性

报告期各期，公司委外研发金额分别为 6,752.31 万元、2,999.21 万元、9,468.99 万元和 1,136.29 万元，委外研发金额波动的原因系发行人研发项目为高度复杂且周期较长的过程，各研发项目在不同的研发阶段对外部研发服务的需求存在差异，报告期各期执行及确认研发费用的委外研发项目不同所致。

报告期各期，发行人委外研发确认研发费用前五大项目基本情况及与研发项目的匹配性如下：

单位：万元

委外研发供应商	委外研发项目名称	对应研发项目	委外研发费	占当期委外研发费比例
2025 年 1-3 月				
委外研发供应商 5	多个大模型的推理任务的性能优化开发	曦云 C500	450.00	39.60%
委外研发供应商 8	图形 GPU Vulkan CTS Shader 数据集建设、初步功能分析及详细性能分析	曦彩 G100	156.22	13.75%
委外研发供应商 10	基于沐曦 MXMACA 运算平台的 Triton 算子开发和优化项目	曦云 C500	107.54	9.46%
委外研发供应商 3	基于沐曦芯片的人工智能解决方案开发项目	通用	93.50	8.23%
委外研发供应商 10	基于沐曦 MXMACA 运算平台的大模型算法部署优化项目	曦云 C500	81.85	7.20%
小计			889.11	78.25%
2024 年度				
委外研发供应商 1	后端设计	曦云 C600	2,641.51	27.90%
委外研发供应商 1 注	后端设计	曦彩 G100	2,000.00	21.12%
委外研发供应商 4	UBB 开发	曦云 C500	850.00	8.98%
委外研发供应商 9	infini-acc 推理引擎	通用	452.83	4.78%
委外研发供应商 3	基于沐曦芯片的人工智能解决方案开发项目	通用	346.50	3.66%
小计			6,290.84	66.44%
2023 年度				
委外研发供应商 1 注	后端设计	曦彩 G100	1,200.00	40.01%
委外研发供应商 6	AlphaSparse 计算库向沐曦的 MXC500 平台移植和优化	曦云 C500	481.13	16.04%
委外研发供应商 7	面向同态密码和基因测序的可重构 SIMD 架构设计	曦云 C500	242.72	8.09%
委外研发供应商 12	后端设计	曦云 C500	240.00	8.00%
委外研发供应商 14	国产芯片在数字农业领域应用的技术研发和应用场景开发	通用	200.00	6.67%
小计			2,363.85	78.82%
2022 年度				
委外研发供应商 1	后端设计及封装设计	曦云 C500	3,015.14	44.65%
委外研发供应商 2	芯片后端设计服务	曦云 C500	2,550.00	37.76%
委外研发供应商 11	数据中心先进计算架构联合研究项目	曦思 N100	300.00	4.44%
委外研发供应商 7	面向同态密码和基因测序的可重构 SIMD 架构设计	曦云 C500	242.72	3.59%

委外研发供应商	委外研发项目名称	对应研发项目	委外研发费	占当期委外研发费比例
委外研发供应商 13	GPU 模糊测试框架研发	曦思 N100	221.28	3.28%
小计			6,329.13	93.73%

注：该项目受 2024 年美国禁令的影响暂时无法流片，双方已签署补充协议确认终止合同中与曦彩 G100 的相关内容。

由上表可见，发行人 2022 年委外研发金额较高的原因系委外研发供应商 1 曦云 C500 研发项目后端设计及封装设计确认研发费用 3,015.14 万元、委外研发供应商 2 曦云 C500 研发项目芯片后端设计服务确认研发费用 2,550.00 万元所致；2024 年委外研发金额较高的原因系委外研发供应商 1 曦彩 G100 研发项目后端设计确认研发费用 2,000.00 万元、委外研发供应商 1 曦云 C600 研发项目后端设计确认研发费用 2,641.51 万元、委外研发供应商 4 曦云 C500 研发项目 UBB 开发确认研发费用 850.00 万元所致。

发行人前述委外研发涉及研发项目主要时间节点如下：

研发项目	立项	流片	正式量产
曦思 N100	2021 年 4 月	2022 年 1 月	2023 年 4 月
曦云 C500	2021 年 6 月	2022 年 12 月	2024 年 2 月
曦云 C600	2024 年 2 月	2024 年 10 月	/
曦彩 G100	2023 年 2 月	/	/

注：“/”表示截至本回复出具日，尚未进入该阶段。

根据发行人研发项目流程，流片属于新产品导入阶段，立项至流片之间属于产品开发阶段，包括芯片整体架构设计、芯片前端设计、芯片后端设计、基础系统软件研发，以及各技术领域设计需求的测试验证等。结合上表，发行人曦云 C500 研发项目于 2021 年 6 月立项、2022 年 12 月流片，期间处于产品开发阶段，因此于 2022 年该项目委外研发确认研发费用金额较大与该研发项目 2022 年处于产品开发阶段相符，具有合理性。发行人曦彩 G100 研发项目于 2023 年 2 月立项，目前尚未流片，曦云 C600 研发项目于 2024 年 2 月立项、2024 年 10 月流片，该两个研发项目 2024 年委外研发确认研发费用金额较大与其 2024 年处于产品开发阶段相符，具有合理性。曦云 C500 研发项目于 2024 年委外研发确认的研发费用金额较大系发行人在曦云 C500 量产后，为了提升 GPU 产品与主流服务器的兼容性而委托新华三定制化设计和验证适用其产品的 UBB，以满足多样化的

市场和客户需求，与研发项目所处阶段相符，具有合理性。

由此，发行人报告期内委外研发金额波动系各研发项目在不同的研发阶段对外部研发服务的需求存在差异，报告期各期执行及确认研发费用的委外研发项目不同所致，具有合理性；委外研发项目主要为满足发行人曦云 C500、曦云 C600、曦思 N100 研发项目及通用研发需求，与研发项目及其所处研发阶段具有匹配性。

（四）管理费用中服务费增长的原因

报告期内，发行人管理费用中服务费分别为 3,625.38 万元、4,450.06 万元、8,510.42 万元和 9,181.51 万元，持续增长，主要原因系融资财务顾问费增长所致。

报告期内，公司进行了多轮股权融资且融资金额持续增大，为了加快融资进程，公司聘请第三方机构为公司提供融资中介服务，多家 A 股上市公司在上市前均存在聘请融资财务顾问的情况，符合市场惯例。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、申报会计师执行了如下核查程序：

1、访谈发行人人事部门负责人，了解发行人人力资源政策和薪酬制度，了解报告期内员工人数和薪酬变动情况，查阅发行人报告期内的员工花名册、工资表、社会保障明细表等资料，了解人员划分依据，核查工资表人员划分是否正确，核查员工的数量、人员结构和工资水平是否与发行人的生产经营相匹配，分析其人员变动原因；通过检索同行业可比公司的招股说明书、定期报告、临时报告等公告，了解发行人同行业可比公司的平均薪酬水平，并分析发行人各部门员工薪酬波动的原因及其合理性；通过网络查询获取上海市、杭州市、南京市、成都市和北京市在岗职工平均工资，比较发行人母公司及重要子公司平均薪酬与五个市在岗职工平均工资的差异情况并分析其合理性。

2、获取发行人与研发相关的《项目管理程序》《系统开发验证程序》等内部控制制度；通过访谈发行人技术负责人，了解发行人研发体系、研发流程、进入正式量产阶段的判定标准；查询同行业可比公司和其他半导体设计型企业公开披露文件；获取发行人收入成本表，整理报告期内发行人正式量产产品在正式量

产前及正式量产后的订单数量，并进一步分析发行人进入正式量产阶段判定标准的客观性。

3、获取研发项目立项文件、进展文件、委外研发合同、成果文件等相关资料，检查发行人委外研发的具体情况和对发行人研发项目对应阶段的相关性、匹配性，分析委外研发波动较大的原因及合理性。

4、取得发行人管理费用-服务费明细表，分析报告期内管理费用中服务费增长的原因。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、报告期内，发行人研发、销售、管理人员的员工人数变动与平均薪酬变动与实际经营情况相匹配，发行人研发、销售、管理人员人均薪酬略高于同行业可比公司平均水平，系发行人对标行业领先企业，提供有竞争力的薪酬待遇以吸引高价值人才所致，具有合理性；发行人母公司及重要子公司的人均薪酬均高于当地平均薪酬，系发行人所处的行业特殊性、市场竞争、人才结构以及激励机制等多种因素共同作用的结果，具备充分的商业合理性与必要性。

2、发行人以风险量产过程中收集的性能、可靠性、封装风险、封测良率等数据作为支撑，在此基础上结合客户订单支撑、市场前景综合评估，由管理层组织多部门进行交叉评审后进入正式量产，未设置进入正式量产阶段的对于客户订单的具体数量标准。该判断标准充分考虑了 GPU 芯片高研发投入、长周期及应用场景多样化的特点，符合行业惯例；发行人报告期内两款量产产品进入正式量产阶段遵循该等判断标准，正式量产前已获得相应客户订单佐证产品商业化前景，发行人判断产品进入正式量产阶段的标准具有客观性、准确性。

3、发行人报告期内委外研发金额波动系各研发项目在不同的研发阶段对外部研发服务的需求存在差异，报告期各期执行及确认研发费用的委外研发项目不同所致，具有合理性。委外研发项目主要为满足发行人各研发项目及通用研发需求，与研发项目及其所处研发阶段具有匹配性。

4、报告期内，发行人管理费用中服务费持续增长，主要原因系融资财务顾问费增长所致，具有合理性。

12.关于股权激励

根据申报材料：（1）报告期各期，发行人确认的股份支付费用为 3,978.62 万元、5,762.38 万元、47,283.29 万元和 3,156.31 万元，上海曦骥与上海骄迈为发行人的直接持股平台；（2）上海骄迈上层为 21 个间接持股平台与陈维良直接持股，包括 18 个员工持股平台及 3 个外部投资平台，被激励对象包括 14 名公司在职员工与 4 名顾问；（3）2020 年 12 月，发行人以上海曦骥作为员工持股实体实施股票期权计划，截至报告期末，累计向员工授予 631,053,509 份期权，另向 3 名公司顾问授予 39,784,525 份期权，激励对象（除实控人外）根据所签署《期权授予协议》中约定的分期兑现条件及行权价格将部分期权转换为限制性股权，实控人获授的期权立即兑现；（4）2025 年 3 月，所有期权激励的在职员工及公司顾问进行加速行权，公司当月另向部分员工及顾问新授予限制性股权；（5）2024 年和 2025 年 1-3 月，“其他符合非经常性损益定义的损益项目”金额为 -37,569.90 万元和 -2,300.30 万元，系无服务期、无业绩考核而一次性计入当期损益的股份支付费用。

请发行人披露：（1）2020 年股票期权计划中激励对象（除实控人外）约定的具体分期兑现条件及行权价格，实控人获授的期权立即兑现、与其他激励对象不一致的原因及合理性；2025 年 3 月所有期权激励的在职员工及公司顾问进行加速行权的原因；（2）员工持股平台上层股东人数较少的原因、合理性、未来安排；（3）历次股份支付的授予日、公允价值、服务期等各参数确定的依据及合理性，股份支付费用归集分摊的具体过程、2024 年大幅增长的原因及合理性、对应的主要人员，会计处理是否符合企业会计准则相关规定；（4）2024 年和 2025 年 1-3 月计入非经常性损益股份支付费用对应的股权激励对象及行权情况。

请保荐机构、发行人律师、申报会计师对上述问题简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

(一) 2020 年股票期权计划中激励对象（除实控人外）约定的具体分期兑现条件及行权价格，实控人获授的期权立即兑现、与其他激励对象不一致的原因及合理性；2025 年 3 月所有期权激励的在职员工及公司顾问进行加速行权的原因

1、2020年股票期权计划中激励对象（除实控人外）约定的具体分期兑现条件及行权价格，实控人获授的期权立即兑现、与其他激励对象不一致的原因及合理性

(1) 2020 年股票期权计划中激励对象（除实控人外）约定的具体分期兑现条件及行权价格

2020 年 10 月，沐曦有限股东会召开会议并作出决议，全体股东同意上海曦骥认缴沐曦有限的新增注册资本 59.6450 万元，同时沐曦有限与全体股东签署了《天使轮股东协议》，约定上海曦骥所持股份预留为员工期权池。

2020 年 12 月，沐曦有限董事会召开会议并作出决议，审议通过《员工持股计划》。《员工持股计划》规定，上海曦骥为员工持股实体，其持有公司的股权为期权来源，《员工持股计划》项下各激励对象行使并认购每一单位期权份额的对价即行权价格均为 1 元/股（对应沐曦有限 1 元注册资本）。

《员工持股计划》项下激励对象区分员工、顾问，发行人对员工实施的股权激励区分为入职激励以及绩效激励两种情形，均对应不同的分期兑现条件，员工及顾问的具体分期兑现条件如下：

序号	激励类型	激励对象	分期兑现条件
1	入职激励	员工	自授予日起为发行人服务满两年兑现 50%、满三年兑现剩余 25%、满四年兑现剩余 25%；同时约定激励员工在上一年度的业绩达到业绩考核标准的前提下，方可兑现当期期权，若激励员工在上一年度的业绩未达到约定的业绩考核标准。则激励员工的当期期权自动取消
2	绩效激励	员工	自授予日起为发行人服务满一年兑现 40%、满两年兑现剩余 30%、满三年兑现剩余 30%
3	顾问激励	顾问	自授予日起为发行人服务每满 6 个月兑现 25%；或自授予日起为发行人服务每满 12 个月兑现 50%

(2) 实控人获授的期权立即兑现、与其他激励对象不一致的原因及合理性

2021 年 8 月，沐曦有限董事会召开会议并作出决议，同意公司设立期权执行委员会。根据《员工持股计划》规定，期权执行委员会决定及执行期权分配、授予等事宜，并有权修订、变更或取消期权兑现条款和条件。

2024 年 10 月 31 日，发行人期权执行委员会就员工持股计划项下尚未授予的预留激励股权作出决议，同意一次性授予实控人部分期权，实控人获授期权豁免适用《员工持股计划》项下关于期权兑现、行权条件以及期权终止的规定，该等授予期权可立即兑现且实控人有权行权并实缴。2025 年 3 月 31 日，发行人激励股权执行委员会将截至报告期末剩余的限制性股权全部授予实控人陈维良，同样豁免适用《员工持股计划》项下关于限制性股权解锁、终止的规定，在实控人获授限制性股权后，该等限制性股权即全部解锁。

发行人实控人陈维良获授的期权立即兑现，未约定分期兑现条件，主要系陈维良作为公司创始股东、董事长、总经理，以其深厚的行业经验、清晰的战略眼光和卓越的领导力，为公司在高性能 GPU 领域奠定了技术研发基础，推动了产学研合作与生态构建，拓展了关键的战略合作伙伴关系，并积极参与行业对话提升品牌影响力。发行人实控人上述贡献为公司成功经营和业绩提升打下坚实的基础，并且其已与公司形成利益与风险共担的深度绑定关系。综合考虑实控人过往对公司的贡献和未来发展意愿，并经公司期权执行委员会决策同意，不再对实控人获授期权设置分期兑现条件。

对于公司员工、顾问等其他激励对象，由于公司实施员工持股计划的主要目的为完善公司治理结构、健全激励机制、调动优秀员工积极性、增强优秀员工对实现公司稳定、持续及快速发展的责任感和使命感，相关持股人员均为各业务或管理部门员工、外部顾问，公司设立一定期限的服务期安排有助于达到激励效果，同时将个人与公司的未来发展进行一定程度绑定，保障公司治理结构的稳定与未来发展的可持续性，具有合理性。

经查询科创板上市公司披露的员工持股计划，存在实施员工持股计划中实际控制人与其他员工服务期安排不一致的情形，具体如下：

序号	上市公司	涉及员工持股计划及服务期安排
1	杰华特 (688141.SH)	在 2020 年 11 月实施的第二期股权激励中，实控人未约定服务期，其他授予对象服务期认定为授予日至 2025 年底

序号	上市公司	涉及员工持股计划及服务期安排
2	联影医疗 (688271.SH)	为配合发行人股份制改造以及上市计划，并奖励实控人过往对公司发展所做出的贡献，公司于 2020 年 3 月及 2020 年 12 月分别将届时尚未授予的全部剩余股份一次性授予实控人，且不带任何服务期限限制；公司其他激励对象均设置服务期
3	先锋精科 (688605.SH)	公司在实施股权激励时，为除实际控制人及其一致行动人外的被激励员工设置了一定的服务期限

综上，实控人获授期权立即兑现主要系发行人考虑其与公司利益深度绑定、个人过往贡献以及未来发展意愿等因素的综合安排，公司其他激励对象约定相应分期兑现条件能够进一步调动激励对象的工作积极性，有利于实现公司稳定、持续及快速发展，相关安排符合《员工持股计划》规定，符合科创板上市公司操作惯例，具有合理性。

2、2025 年 3 月所有期权激励的在职员工及公司顾问进行加速行权的原因

2025 年 1 月，发行人与华泰联合证券签订《首次公开发行股票辅导协议》并在中国证监会上海监管局辅导备案，明确启动上市工作。基于《首次公开发行股票注册管理办法》《上海证券交易所科创板股票上市规则》等相关法律法规对发行人股权清晰的要求，发行人于 2025 年 3 月 5 日召开第一届董事会第六次会议并作出决议，审议通过了《关于期权加速行权及选任新一届执行委员会的议案》及其附件《沐曦集成电路（上海）股份有限公司员工持股计划（2025 修订）》，允许员工持股计划项下全部激励对象对其有效且尚未兑现的期权加速兑现并行权，激励对象通过加速行权取得的股权为限制性股权，并将原期权的兑现/行权条件、兑现日转为解锁条件、解锁日，前述限制性条件未发生实质变化，对在职员工及公司顾问仍然起到激励作用。

（二）员工持股平台上层股东人数较少的原因、合理性、未来安排

1、员工持股平台上层股东人数较少的原因、合理性

除 3 个外部投资平台外，上海骄迈上层持股平台均为员工持股平台，被激励对象包括 14 名公司在职员工与 4 名公司顾问。

上海骄迈上层 18 个员工持股平台均是 1 名员工或顾问为有限合伙人，公司实控人陈维良为普通合伙人并担任执行事务合伙人的出资结构，各员工持股平台上层股东人数较少，该架构设置的主要原因如下：

(1) 具体授予激励股权数量在核心员工之间存在保密需求；

(2) 发行人考虑管理成本和效率，减少各持股平台层面有限合伙人权益相关事项的沟通成本和协调难度；

(3) 进一步强化实控人在持股平台中的主导地位，即在任一有限合伙人发生变动的情形下，能够有效保障实控人依据《员工持股计划》的规定及以其执行事务合伙人的身份，及时有效地依法依规进行处理，保障上海骄迈层面合伙份额及对应发行人股权的权属清晰稳定，进而强化陈维良对发行人的控制权及稳定性。

依据《合伙企业法》第六十一条规定，有限合伙企业由二个以上五十个以下合伙人设立，有限合伙企业至少应当有一个普通合伙人。上海骄迈上层各员工持股平台股东人数较少及相应架构，符合《合伙企业法》的规定。基于上述，员工持股平台上层股东人数较少具有合理性。

2、员工持股平台未来安排

(1) 规范运作安排

员工持股平台区分核心员工与顾问，核心员工所在持股平台主要依据其签署的《限制性股权授予及代持协议》及其补充协议进行规范运作，顾问所在持股平台主要依据《员工持股计划》及顾问签署的配套期权/限制性股权授予协议及其补充协议（如有）进行规范运作。同时，员工持股平台全体合伙人依据合伙协议的约定行使权利、承担义务。除前述外，发行人、员工持股平台以及激励对象之间不存在其他关于激励股权特殊安排的约定，员工持股平台的主要功能在于作为发行人实施股权激励以及激励对象间接持有激励股权的载体。

(2) 转让限制与锁定期安排

《限制性股权授予及代持协议》以及公司《员工持股计划》对转让限制及锁定期的规定如下：

《限制性股权授予及代持协议》	《员工持股计划》
在目标公司上市之前，除可通过依法继承的方式转让已解锁的限制性股权外，未经实际控制人或目标公司以书面方式同意，特定授予对象不得以任何方式直接或间接出售、转让、让予、质押、抵押、赠与、设定产权负担或以其他方式处分其按照本协议取得并持	在公司上市之前，除可通过依法继承的方式转让已兑现期权和期权份额及已解锁的限制性股权份额外，未经执行委员会以书面方式同意，参与人不得以任何方式直接或间接出售、转让、让予、质押、抵押、赠与、设定产权负担或以其他方式处分其按照本计划及《期权授

《限制性股权授予及代持协议》	《员工持股计划》
有的限制性股权（无论是否完成解锁）。若特定授予对象的继承人继承已解锁的限制性股权，继承人应同时承继本协议规定的、该等限制性股权的各项限制和条件，并遵守本协议的规定。 目标公司上市后，特定授予对象持有的限制性股权应根据相关证券监管机构及证券交易所的要求，并按照目标公司确定的规则在规定期限（以下简称“锁定期”）内予以锁定。	予协议》《限制性股权授予协议》取得并持有的激励股权或激励份额。若参与人的继承人继承已兑现期权和期权份额/已解锁的限制性股权和限制性股权份额，继承人应同时承继本计划及《期权授予协议》《限制性股权授予协议》规定的、该等激励股权或激励份额的各项限制和条件，并遵守本计划及《期权授予协议》《限制性股权授予协议》的规定。 公司上市后，参与人持有的激励份额以及相关权益应根据相关证券监管机构及证券交易所的要求，并按照执行委员会确定的规则在规定期限（“锁定期”）内予以锁定。锁定期的截止日应不得早于员工持股实体持有的公司股份所适用的锁定期。

如上表所列示，本次发行上市前及上市后的法定限售期内，激励对象取得的激励股权均被锁定，处分受限；本次发行上市后，激励对象对激励股权的处置亦需遵守届时法律法规以及发行人制定的减持管理办法的规定，合法合规并有序地进行减持。

截至本回复出具日，上海骄迈作为公司控股股东已出具《关于发行人股票锁定期的承诺函》，其上层员工持股平台遵守《限制性股权授予及代持协议》以及《员工持股计划》的转让限制及锁定期安排，其持有的激励股权或激励份额转让受限并予以锁定，不存在其他特殊安排。

（三）历次股份支付的授予日、公允价值、服务期等各参数确定的依据及合理性，股份支付费用归集分摊的具体过程、2024 年大幅增长的原因及合理性、对应的主要人员，会计处理是否符合企业会计准则相关规定

自公司设立以来，公司股份支付涉及的事项包括：（1）发行人通过上海曦骥持股平台向除实控人外的员工授予期权/限制性股权；（2）实控人向核心员工授予上海骄迈持股平台的限制性股权；（3）发行人向公司顾问授予期权/限制性股权；（4）发行人通过上海曦骥持股平台向实控人授予期权/限制性股权。上述事项的授予日、公允价值、服务期等各参数确定情况以及股份支付归集分摊过程、对应的主要人员、会计处理情况具体如下：

1、发行人通过上海曦骥持股平台向除实控人外的员工授予期权/限制性股权

发行人股东于 2020 年 10 月共同签署《天使轮股东协议》，并由发行人董事

会于 2020 年 12 月审议通过《员工持股计划》，约定上海曦骥所持发行人的全部股权作为公司实施员工期权激励的股权来源，激励方式为期权，由期权执行委员会按照《员工持股计划》的规定决定期权分配、授予等事宜。

2025 年 3 月 3 日，发行人召开第一届董事会第六次会议并作出决议，审议通过《关于期权加速行权及选任新一届执行委员会的议案》，允许员工持股计划项下全部激励对象对其有效且尚未兑现的期权加速行权，激励对象通过加速行权取得的股权为限制性股权。自董事会作出决议之日起，员工持股计划项下的激励方式增加限制性股权，新授予的激励对象获授的均为限制性股权。员工获授期权及限制性股权的期权行权价格/股票认购价格均为 1 元/股。

截至报告期末，发行人通过上海曦骥持股平台共向除实控人以外的公司员工共授予 8 个批次期权以及 1 个批次限制性股权，剔除截至报告期末已离职员工，上述激励对象涉及 730 名公司员工。发行人通过上海曦骥持股平台授予期权/限制性股权的授予日以员工签署授予协议日期为依据；期权公允价值主要参照银信资产评估有限公司出具的《股票期权公允价值估值报告》，按照 Black-Scholes 模型对期权在每个授予日的公允价值进行评估，限制性股权公允价值参考授予日最近一次外部投资者增资价格，具体信息如下：

序号	授予批次	授予日/估值基准日	定价依据	估值结论
1	第一批期权激励	2021 年 9 月 24 日	评估报告银信咨报字（2024）第 06011 号	发行人于估值基准日 2021 年 9 月 24 日授予股票期权，第一、二、三期的股票期权公允价值分别为：692.87 元/股、692.89 元/股、692.92 元/股
2	第二批期权激励	2022 年 9 月 28 日	评估报告银信咨报字（2024）第 06012 号	发行人于估值基准日 2022 年 9 月 28 日授予股票期权，第一、二、三期的股票期权公允价值分别为：1,087.88 元/股、1,087.91 元/股、1,087.93 元/股
3	第三批期权激励	2023 年 3 月 1 日	评估报告银信咨报字（2024）第 06013 号	发行人于估值基准日 2023 年 3 月 1 日授予股票期权，第一、二、三期的股票期权公允价值分别为：1,087.86 元/股、1,087.89 元/股、1,087.92 元/股
4	第四批期权激励	2023 年 12 月 28 日	评估报告银信咨报字（2024）第 06014 号	发行人于估值基准日 2023 年 12 月 28 日授予股票期权，第一、二、三期的股票期权公允价值分别为：1,263.30 元/股、1,263.33 元/股、1,263.35 元/股
5	第五批期权激励	2024 年 4 月 28 日	评估报告银信咨报字（2025）第 06024 号	发行人于估值基准日 2024 年 4 月 28 日授予股票期权，第一、二、三期的股票期权公允价值分别为：1,263.30 元/股、1,263.32 元/股、1,263.34 元/股

序号	授予批次	授予日/估值基准日	定价依据	估值结论
6	第六批期权激励	2024年4月28日	评估报告银信咨报字（2024）第06029号	发行人于估值基准日2024年4月28日授予股票期权，第一、二、三期的股票期权公允价值分别为：1,263.28元/股、1,263.30元/股、1,263.32元/股
7	第七批期权激励	2025年3月14日	由于该部分授予期权在授予当月立即转换为限制性股权，其公允价值参考授予日最近一次外部投资者增资价格1,429.56元/股	
8	第八批期权激励	2025年3月28日		
9	限制性股权激励	2025年3月28日	公允价值参考授予日最近一次外部投资者增资价格1,429.56元/股	

注：公司2024年4月28日同期授予两个批次期权激励，其中第五批期权激励为员工入职激励，第六批期权激励为员工绩效激励

根据上述期权/限制性股权的授予日及公允价值，发行人在服务期内对股份支付费用进行分摊，具体参数及会计处理如下：

服务期	员工被授予期权/限制性股权的兑现或解锁期限分为入职激励以及绩效激励两种情形，根据《期权授予协议》或《限制性股权授予协议》约定，具体为（1）自授予日起为发行人服务满两年兑现/解锁50%、满三年兑现/解锁剩余25%、满四年兑现/解锁剩余25%；或（2）自授予日起为发行人服务满一年兑现/解锁40%、满两年兑现/解锁剩余30%、满三年兑现/解锁剩余30%。 根据公司董事会2020年12月审议通过的《员工持股计划》以及2021年1月修订稿，《员工持股计划》参与人主动提出辞职或主动停止为公司提供服务时，公司有权终止授予参与人期权，宣布全部已经授予但尚未兑现、以及已经兑现但尚未行使完毕的期权失效；若任何期权已经行使完毕，公司或受托人有权自行或指定第三方以等同于行权价格或参与人实缴出资额或届时参与人持有的公司股权价值（以孰低者为准）减去已经取得的收益的价格回购参与人因行权而购买的全部期权份额及其所对应的员工持股实体合伙份额，并要求参与人从员工持股实体退伙，实质上构成对被授予对象的隐含服务期、等待期要求。上述条款未明确回购权的期限，发行人于2025年3月与激励对象补充签署《<期权授予协议>之补充协议》并修订《员工持股计划》，明确全部授予期权加速行权，原期权的兑现/行权条件、兑现日转为解锁条件、解锁日，同时约定公司合格上市后，发行人仅对激励对象已被授予但尚未解锁的限制性股权拥有上述回购权利。 因此上述获授期权/限制性股权的激励对象实际服务期自授予日起，至解锁日与公司合格上市日孰晚者；公司预期于2027年3月31日实现上市，发行人以上述原则确定股权激励服务期。
股份支付归集分摊	构成以权益结算的股份支付，并在服务期内按照激励对象所在的职能部门及承担的职能角色进行分摊
对应的主要人员	剔除截至报告期末的已离职员工，激励对象涉及730名公司员工
会计处理	1、员工离职的处理 员工作为激励对象离职时，均由上海曦骥将期权或限制性股权全部回购收回，前期已确认计提的股份支付费用在离职时冲回。未来授予员工时，再根据授予时点的权益公允价值重新分摊确认股份支付费用 2、期权加速行权转换为限制性股权的处理

	公司于 2025 年 3 月通过加速行权将已授予尚未行权的股票期权转换为限制性股权，限制性股权的锁定期、限售期等限制条件未发生实质性改变，但转换后限制性股权价值较期权价值更低，根据《企业会计准则讲解（2010）》的相关规定，应当如同该变更从未发生，不能减少股份支付公允价值总额。因此期权加速行权转为限制性股权，没有产生权益工具公允价值的增加，不考虑修改后的可行权条件，不影响原股份支付的会计处理，股份支付总额未发生变动 3、股份支付金额确认 公司自授予当月开始按照服务期分摊计提股份支付费用，报告期各期计提股份支付费用 1,709.63 万元、3,197.36 万元、3,992.72 万元与 509.29 万元并相应确认资本公积
会计处理是否符合企业会计准则规定	是

2、实控人向核心员工授予上海骄迈持股平台的限制性股权

在沐曦有限设立时，由陈维良主要出资设立的上海骄迈向发行人认缴出资 196 万元，2020 年 9 月至 2023 年 12 月期间，陈维良与发行人 18 名核心员工签署《限制性股权授予及代持协议》并向其授予一定数量的上海骄迈出资份额，对应发行人的限制性股权并在满足解锁条件时分步解锁，限制性股权授予的认购价格均为 1 元/股。

实控人向核心员工授予限制性股权的授予日以授予协议签署日为依据，其公允价值的确认均参考授予日最近一次外部投资者增资价格，具体参数及会计处理如下：

授予日	2020 年 9 月 14 日、2020 年 9 月 15 日、2020 年 9 月 16 日、2020 年 9 月 27 日、2020 年 9 月 28 日、2020 年 10 月 8 日、2020 年 10 月 16 日、2021 年 6 月 28 日、2021 年 9 月 18 日、2021 年 12 月 24 日、2023 年 3 月 15 日、2023 年 12 月 20 日
公允价值	授予日为 2020 年 9 月至 10 月期间的限制性股权公允价值为发行人同期最近一次的外部增资价格 120.71 元/股； 授予日为 2021 年 6 月至 12 月期间的限制性股权公允价值为发行人同期最近一次的外部增资价格 693.82 元/股； 授予日为 2023 年 3 月 15 日的限制性股权公允价值为发行人同期最近一次的外部增资价格 1,088.84 元/股； 授予日为 2023 年 12 月 20 日的限制性股权公允价值为发行人同期最近一次的外部增资价格 1,264.26 元/股
服务期	根据《限制性股权授予及代持协议》约定，被授予的限制性股权解锁条件为：自交割日起为发行人服务满两年解锁 50%、满三年解锁剩余 25%、满四年解锁剩余 25%；或自交割日起为发行人服务满一年解锁 50%、满两年解锁剩余 25%、满三年解锁剩余 25% 同时《限制性股权授予及代持协议》约定被授予对象“主动提出辞职或主动停止为公司提供服务”时，发行人及其实控人有权自行或指定第三方以等同于行权价格或员工实缴注册资本或届时股权价值（以低者为准）减去已经取得的收益的价格回购其购买的全部限制性股权

	（包括已经解锁的部分和未解锁的部分），实质上构成对被授予对象的隐含服务期、等待期要求。上述条款未明确回购权的期限，发行人于 2025 年 3 月与激励对象补充签署《<限制性股权授予及代持协议>之补充协议》，进一步补充并明确以授予日至预计上市日为回购权期限，激励对象已解锁的限制性股权在公司上市后由激励对象享有。因此上述人员实际服务期自授予交割日起，至解锁日与公司合格上市日孰晚者，公司预期于 2027 年 3 月 31 日实现上市，发行人以上述原则确定股权激励服务期
股份支付归集分摊	构成以权益结算的股份支付，并在服务期内按照激励对象所在的职能部门及承担的职能角色进行分摊
对应的主要人员	公司 18 名核心员工，其中 4 名员工已离职
会计处理	1、员工离职的处理 员工获授上海骄迈持股平台的限制性股权后选择从公司离职，如由实控人回购股份，在员工离职时冲减前期已计提的股份支付费用；如经协商后保留离职员工享有的部分限制性股权，则对保留股权的股份支付费用进行加速一次性确认并计入非经常性损益 2、实控人回购的处理 由于上海骄迈为实控人实际出资企业，授予员工的股权来自于实控人，因此实控人回购不会重新构成股份支付 3、股份支付金额确认 公司自授予当月开始按照服务期分摊计提股份支付费用，报告期各期计提股份支付费用 1,916.74 万元、2,095.21 万元、4,545.70 万元与 41.76 万元并相应确认资本公积
会计处理是否符合企业会计准则规定	是

3、发行人向公司顾问授予期权/限制性股权

根据公司制定的《员工持股计划》安排，员工持股计划适用于在公司（以及公司未来不时设立之附属或关联公司）任职的全职员工、顾问和外部专家。截至本回复出具日，公司曾向 3 名公司顾问授予员工持股平台的期权份额，相关期权份额均于 2025 年 3 月加速行权转换为限制性股权，公司另向 1 名公司顾问直接授予限制性股权，公司顾问获授期权及限制性股权的期权行权价格/股票认购价格均为 1 元/股。

发行人向公司顾问授予期权/限制性股权的授予日以顾问实际为公司提供服务的起始日期为参考；期权公允价值参照银信资产评估有限公司出具的《股票期权公允价值估值报告》，按照 Black-Scholes 模型对期权在每个授予日的公允价值进行评估，限制性股权公允价值参考授予日最近一次外部投资者增资价格，具体参数信息及会计处理如下：

授予日	2022 年 1 月 1 日、2023 年 1 月 31 日、2024 年 1 月 1 日、2025 年 3 月 1 日
-----	--

公允价值	根据银信评估出具的评估报告（银信咨报字（2025）第 06011 号），发行人于估值基准日 2022 年 1 月 1 日授予股票期权，授予相关顾问的第一、二、三、四期的股票期权公允价值分别为：1,087.85 元/股、1,087.86 元/股、1,087.87 元/股、1,087.89 元/股； 根据银信评估出具的评估报告（银信咨报字（2025）第 06012 号），发行人于估值基准日 2023 年 1 月 31 日授予股票期权，授予相关公司顾问的第一、二期的股票期权公允价值分别为：1,087.86 元/股、1,087.89 元/股； 根据银信评估出具的评估报告（银信咨报字（2025）第 06013 号），发行人于估值基准日 2024 年 1 月 1 日授予股票期权，授予相关公司顾问的第一、二、三、四期的股票期权公允价值分别为：1,263.27 元/股、1,263.28 元/股、1,263.29 元/股、1,263.30 元/股；授予相关公司顾问的第一、二期的股票期权公允价值分别为：1,263.28 元/股、1,263.30 元/股； 发行人于 2025 年 3 月 1 日授予限制性股权的公允价值参考授予日最近一次外部投资者增资价格 1,429.56 元/股
服务期	根据发行人与公司顾问签署的《期权授予协议》或《限制性股权授予协议》约定，期权/限制性股权的兑现或解锁期限有以下类型：（1）自授予日起为发行人服务每满 6 个月兑现/解锁 25%；（2）自授予日起为发行人服务每满 12 个月兑现/解锁 50%。 发行人《员工持股计划》的相关安排适用于公司顾问的股权激励，因此参考公司员工获授股权激励的隐含服务期、等待期，公司顾问的实际服务期为自授予日起，至解锁日与公司合格上市日孰晚者；公司预期于 2027 年 3 月 31 日实现上市，发行人以上述原则确定股权激励服务期。
股份支付归集分摊	构成以权益结算的股份支付，并在服务期内按照激励对象所承担的职能角色进行分摊
对应的主要人员	4 名公司顾问
会计处理	公司激励的外部顾问目前不存在股权激励终止情形，公司自授予当月开始按照服务期分摊计提股份支付费用，报告期各期计提股份支付费用 352.26 万元、469.80 万元、1,176.95 万元与 304.91 万元并相应确认资本公积
会计处理是否符合企业会计准则规定	是

4、发行人通过上海曦骥持股平台向实控人授予期权/限制性股权

鉴于公司申请首次公开发行上市的发展规划，发行人计划在上市申报前完成全部股权（包括直接层面和间接层面）的权属确认，公司期权执行委员会于 2024 年 10 月 31 日作出决议，同意一次性授予实控人 297,404,731 份期权份额，授予后实控人可立即兑现并行权；激励股权执行委员会于 2025 年 3 月 31 日作出决议，同意一次性授予实控人 16,102,226 份限制性股权，授予后实控人即全部解锁。实控人获授期权及限制性股权的期权行权价格/股票认购价格均为 1 元/股。

发行人向实控人授予期权/限制性股权的授予日以执行委员会决议日期为依据；由于期权授予后实控人立即兑现行权，因此其期权公允价值与限制性股权的

公允价值确定依据一致，均参考授予日最近一次外部投资者增资价格，具体参数信息及会计处理如下：

授予日	2024 年 10 月 31 日、2025 年 3 月 31 日
公允价值	发行人于 2024 年 10 月 31 日授予实控人期权的公允价值参考授予日最近一次外部投资者增资价格 1,264.26 元/股； 发行人于 2025 年 3 月 1 日授予实控人限制性股权的公允价值参考授予日最近一次外部投资者增资价格 1,429.56 元/股
服务期	实控人获授期权/限制性股权后，豁免适用《员工持股计划》项下关于期权兑现与终止、限制性股权解锁与终止的规定，未设置服务期
股份支付归集分摊	构成以权益结算的股份支付，在授予期间一次性确认股份支付费用
对应的主要人员	公司实控人
会计处理	公司在授予期间一次性确认股份支付费用，2024 年度及 2025 年 1-3 月期间分别计提股份支付费用 37,569.90 万元与 2,300.30 万元并相应确认资本公积
会计处理是否符合企业会计准则规定	是

报告期各期，公司确认的股份支付费用分别为 3,978.62 万元、5,762.38 万元、47,283.29 万元和 3,156.31 万元，2024 年大幅增长主要系当期一次性确认向实控人的股份支付费用 37,569.90 万元，同时随着公司业务规模的不断发展，为充分调动员工的工作积极性并吸引优秀人才加入公司，发行人对股权激励对象的累计范围和力度在逐渐扩大。因此 2024 年股份支付费用大幅增长符合公司实际发展情况，具有合理性。

（四）2024 年和 2025 年 1-3 月计入非经常性损益股份支付费用对应的股权激励对象及行权情况

发行人 2024 年和 2025 年 1-3 月计入非经常性损益股份支付费用分别为 37,569.90 万元与 2,605.79 万元，对应的股权激励对象为公司实控人以及两名离职员工，具体情况如下：

1、公司实控人确认非经常性损益股份支付费用

公司于 2024 年 10 月 31 日及 2025 年 3 月 31 日向实控人授予一定份额的期权及限制性股权，上述股权激励未约定相应服务期，实控人获授期权及限制性股权后立即行权并已实缴认购股权，具体情况详见本问回复“（三）/4、发行人通过上海曦骥持股平台向实控人授予期权/限制性股权”。

根据中国证监会《监管规则适用指引——发行类第 5 号》相关规定：“股份

立即授予或转让完成且没有明确约定等待期等限制条件的，股份支付费用原则上应一次性计入发生当期，并作为偶发事项计入非经常性损益。”发行人于 2024 年度及 2025 年 1-3 月期间就向实控人股权激励一次性确认股份支付费用 37,569.90 万元与 2,300.30 万元，并计入非经常性损益。

2、两名离职员工确认非经常性损益股份支付费用

2020 年 9 月，程刚、孙尔俊在公司成立之初入职并各自与实控人陈维良签署了《限制性股权授予及代持协议》，分别获授对应沐曦有限当时注册资本 12.6 万元的上海骄迈合伙份额，两人于 2020 年 9 月至 10 月期间向实控人陈维良支付限制性股权认购款项，认购完成后其限制性股权由陈维良代为持有。2024 年 8 月，程刚、孙尔俊因个人原因从公司离职。

2025 年 2 月 10 日，发行人召开 2025 年第二次临时股东会并作出决议，同意保留程刚、孙尔俊两名离职员工部分限制性股权，符合员工持股计划的相关规定。2025 年 3 月 21 日，程刚、孙尔俊分别与陈维良签署了《<限制性股权授予及代持协议>之补充协议》，其各自获授的 8.4 万元合伙份额及对应限制性股权由陈维良以原价回购，各自获授的剩余 4.2 万元合伙份额及对应限制性股权仍由其保留。

根据《限制性股权授予及代持协议》安排，程刚、孙尔俊两名离职员工获授的全部限制性股权存在服务期设定，具体情况可详见本问回复“（三）/2、实控人向核心员工授予上海骄迈持股平台的限制性股权”。公司于 2025 年 3 月修改程刚、孙尔俊保留的激励股权计划，同意两人离职后就其保留股权享有处分权，不再适用于服务期的约束，因此就保留激励股权的剩余未分摊股份支付金额一次性加速确认在当期，构成所属期间的非经常性损益。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、发行人律师、申报会计师主要执行了以下程序：

1、获取并查阅发行人历次《员工持股计划》及其决策文件、天使轮融资协议、员工持股计划文件，核查发行人员工持股计划的制订以及期权预留情况。

2、获取并查阅股权激励对象签署的《限制性股权授予及代持协议》《期权授予协议》等相关授予文件，了解发行人不同激励对象的激励类型、分期兑现条件及行权价格。

3、获取并查验发行人期权执行委员会以及激励股权执行委员会的决议、相关股东会决议，核查发行人授予实际控制人激励股权的决策程序履行情况、实际控制人通过历次股权激励取得的发行人的股份比例。

4、获取并查验发行人出具的关于实际控制人激励股权授予的相关说明，了解授予实际控制人激励股权且允许其立即兑现/解锁的原因背景。

5、查询科创板上市公司披露的员工持股计划，分析科创板上市公司员工持股计划中实际控制人与其他员工服务期安排不一致的操作惯例。

6、获取并查验发行人本次发行上市的辅导协议文件，核查发行人办理本次发行上市的辅导备案时间。

7、查阅《首次公开发行股票并注册管理办法》《上海证券交易所科创板股票上市规则》等相关法律法规，核查相关规定对发行人股权清晰的要求。

8、获取并查验发行人第一届董事会第六次会议，核查发行人员工持股计划的修订情况以及加速行权的原因背景。

9、获取并查验上海骄迈上层员工持股平台的工商档案、合伙协议，对发行人实际控制人进行访谈，了解各持股平台上层股东人数较少的原因、合理性。

10、查阅《合伙企业法》关于合伙企业合伙人人数的规定，核查员工持股平台人数构成的合法合规性。

11、获取并查验上海骄迈上层员工持股平台激励对象签署的期权/限制性股权授予协议及其补充协议、合伙协议以及《员工持股计划》，核查规范运作安排、转让限制及锁定期安排的相关规定。

12、查阅公司股权激励授予明细表及相关文件，分析股份支付归集分摊与会计处理的匹配性、合理性及其对公司未来经营业绩的影响，核查计入非经常性损益股份支付费用对应的股权激励对象及行权情况。

（二）核查意见

经核查，保荐机构、发行人律师、申报会计师认为：

1、实控人获授期权立即兑现主要系发行人考虑其与公司利益深度绑定、个人过往贡献以及未来发展意愿等因素的综合安排，公司其他激励对象约定相应分期兑现条件能够进一步调动激励对象的工作积极性，有利于实现公司稳定、持续及快速发展，相关安排符合《员工持股计划》规定，符合科创板上市公司操作惯例，具有合理性。2025 年 3 月所有期权激励的在职员工及公司顾问进行加速行权，系发行人基于相关法律法规对本次发行上市之发行人股权清晰的要求所作的安排。

2、员工持股平台人数较少主要系发行人出于激励授予情况保密、提高管理效率以及强化控制权的考量，符合相关法律法规规定，具有合理性；各员工持股平台根据《员工持股计划》、激励股权授予协议及其补充协议的相关规定及约定进行规范运作，激励对象所持激励股权转让受限并遵照证券监管机构要求进行锁定，未来不存在其他特殊安排。

3、公司历次股份支付的均按照企业会计准则及相关规定进行股份支付归集分摊与会计处理，授予日、公允价值、服务期等各参数确定具备合理性。公司 2024 年股份支付费用大幅增长主要系当期一次性确认向实控人的股份支付，同时发行人在 2024 年对股权激励对象的累计范围和力度逐渐扩大。

4、2024 年和 2025 年 1-3 月计入非经常性损益股份支付费用对应的股权激励对象为实控人以及两名离职员工，其对应授予的股权激励已进行有效行权或认购。

13.关于募投项目

根据申报材料：（1）发行人拟募集 390,366.59 万元用于新型高性能通用 GPU 研发及产业化项目（C600、C700 芯片）、新一代人工智能推理 GPU 研发及产业化项目（N 系列芯片）及面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目。报告期内，公司主要产品收入集中于 C500，其他产品尚未形成收入或收入贡献较少；（2）发行人募投研发项目中 IP 授权费用、委外研发费用投资金额较大，C600、C700 研发项目中委外研发费用均占比近 20%；（3）截至 2025 年 3 月末，公司账面货币资金、交易性金融资产合计 72.91 亿元。

请发行人披露：（1）三个研发项目的具体研发内容，研发投入资金的具体使用安排，结合 C600、C700、N 系列芯片产品的客户开拓、市场销售情况，分析有关募投项目建设的必要性及未来发展安排；（2）发行人募投项目中 IP 授权费用、委外研发费用的具体内容，与自研 IP、自身研发活动的异同，是否涉及公司核心技术或涉及核心研发生产环节；（3）结合公司目前可支配资金情况，说明募集资金规模是否合理。本次募投项目预计收益情况，折旧摊销、期间费用等对发行人未来经营业绩及财务数据的具体影响，并针对性进行重大事项提示。

请保荐机构、申报会计师简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）三个研发项目的具体研发内容，研发投入资金的具体使用安排，结合 C600、C700、N 系列芯片产品的客户开拓、市场销售情况，分析有关募投项目建设的必要性及未来发展安排

1、三个研发项目的具体研发内容，研发投入资金的具体使用安排

（1）新型高性能通用 GPU 研发及产业化项目

本项目为发行人在研项目，系研发基于国产先进工艺的发行人曦云 C 系列训推一体芯片第二代产品 C600 和第三代产品 C700，具体研发内容主要为进一步完善公司在高性能 GPU 架构优化、高集成度高能效 SoC 设计、国产先进工艺物理设计、可靠性验证测试、性能功耗协同优化、弹性可扩展互连系统构建、自主

软件栈设计等方面的关键核心技术，最终完成相关产品开发并实现量产销售。

研发投入资金的具体使用安排：

1) 第二代高性能通用 GPU 芯片（曦云 C600）研发项目

本项目投资总额为 136,977.30 万元，其中拟使用募集资金投入金额为 84,101.70 万元，具体情况如下：

单位：万元

序号	项目	投资金额	占比	使用募集资金投入
1	设备购置	21,157.29	15.45%	20,483.00
2	软件购置	2,580.52	1.88%	2,010.00
3	IP 授权费用	5,670.00	4.14%	-
4	人员工资	42,080.00	30.72%	34,643.00
5	委外研发费用	27,374.39	19.98%	18,914.10
6	流片费用	7,548.70	5.51%	5,196.60
7	租赁费用	3,023.02	2.21%	2,855.00
8	基本预备费	3,283.02	2.40%	-
9	铺底流动资金	24,260.36	17.71%	-
项目总投资		136,977.30	100.00%	84,101.70

本项目于 2024 年 2 月立项，截至本次发行上市董事会前已发生费用支出以及投资总额中的基本预备费、铺底流动资金部分，将使用发行人自有资金进行投入。本项目各项投资支出的具体内容说明如下：

单位：万元

序号	项目	投资金额
设备购置		
1	研发设备	15,665.00
2	测试设备	3,718.27
3	网络设备	800.00
4	服务器/工作站/DIY 平台	550.90
5	硬件仿真器	423.12
合计		21,157.29
软件购置		
1	EDA	570.52

序号	项目	投资金额
2	研发软件工具	2,010.00
合计		2,580.52
IP 授权费用		
1	高速接口类 IP	4,007.50
2	存储类 IP	1,662.50
合计		5,670.00
人员工资		
1	芯片研发人员	14,160.00
2	软件研发人员	27,920.00
合计		42,080.00
委外研发费用		
1	芯片/软件设计委外研发—子公司	20,560.00
2	芯片/软件设计委外研发—第三方	2,408.10
3	后端开发	3,059.70
4	服务器合作 NRE	1,000.00
5	其他	346.59
合计		27,374.39
流片费用		
1	光罩	3,693.90
2	晶圆	700.00
3	封装/测试/板卡治具及 NRE	2,914.80
4	可靠性测试	240.00
合计		7,548.70
租赁费用		
1	EMU 租赁	523.02
2	外部算力租赁	2,500.00
合计		3,023.02

2) 第三代高性能通用 GPU 芯片（曦云 C700）研发项目

本项目投资总额为 204,015.07 万元，其中拟使用募集资金投入金额为 161,818.06 万元，具体情况如下：

单位：万元

序号	项目	投资金额	占比	使用募集资金投入
1	设备购置	23,452.50	11.50%	23,452.50
2	软件购置	7,671.56	3.76%	7,671.56
3	IP 授权费用	8,750.00	4.29%	8,750.00
4	人员工资	60,480.00	29.64%	60,017.00
5	委外研发费用	40,700.00	19.95%	40,700.00
6	流片费用	17,927.00	8.79%	17,927.00
7	租赁费用	3,300.00	1.62%	3,300.00
8	基本预备费	4,868.43	2.39%	-
9	铺底流动资金	36,865.58	18.07%	-
项目总投资		204,015.07	100.00%	161,818.06

本项目于 2025 年 4 月立项，截至本次发行上市董事会前已发生费用支出以及投资总额中的基本预备费、铺底流动资金部分，将使用发行人自有资金进行投入。本项目各项投资支出的具体内容说明如下：

单位：万元

序号	项目	投资金额
设备购置		
1	研发设备	16,843.50
2	测试设备	4,018.00
3	网络设备	1,200.00
4	服务器/工作站/DIY 平台	600.00
5	硬件仿真器	791.00
合计		23,452.50
软件购置		
1	EDA	3,288.56
2	研发软件工具	4,383.00
合计		7,671.56
IP 授权费用		
1	高速接口类 IP	4,500.00
2	存储类 IP	1,500.00
3	基础单元类 IP	1,700.00
4	微控制器 IP	250.00

序号	项目	投资金额
5	其他 IP	800.00
合计		8,750.00
人员工资		
1	芯片研发人员	14,400.00
2	软件研发人员	46,080.00
合计		60,480.00
委外研发费用		
1	芯片/软件设计委外研发—子公司	28,720.00
2	芯片/软件设计委外研发—第三方	8,500.00
3	服务器合作 NRE	3,000.00
4	其他	480.00
合计		40,700.00
流片费用		
1	光罩	11,500.00
2	晶圆	3,000.00
3	封装/测试/板卡治具及 NRE	3,127.00
4	可靠性测试	300.00
合计		17,927.00
租赁费用		
1	EMU 租赁	800.00
2	外部算力租赁	2,500.00
合计		3,300.00

(2) 新一代人工智能推理 GPU 研发及产业化项目

本项目系研发公司下一代基于先进封装技术的云端大模型推理芯片 Nx，具体研发内容主要为针对该产品特点和市场需求，改进公司现有的高性能 GPU 流处理器、SoC 架构、芯粒架构等设计方案，同时加强 GPU 软硬件协同优化，以及与国内厂商合作研发 3D DRAM 堆叠技术、混合键合先进封装工艺等内容，实现新的显存解决方案，最终完成产品开发及商业化。

本项目投资总额为 57,813.35 万元，其中拟使用募集资金投入金额为 45,305.64 万元，具体情况如下：

单位：万元

序号	项目	投资金额	占比	使用募集资金投入
1	设备购置	8,074.00	13.97%	8,074.00
2	软件购置	1,448.64	2.51%	1,448.64
3	IP 授权费用	6,550.00	11.33%	6,550.00
4	人员工资	14,400.00	24.91%	14,400.00
5	委外研发费用	5,740.00	9.93%	5,740.00
6	流片费用	9,093.00	15.73%	9,093.00
7	基本预备费	1,359.17	2.35%	-
8	铺底流动资金	11,148.54	19.28%	-
项目总投资		57,813.35	100.00%	45,305.64

本项目投资总额中的基本预备费、铺底流动资金部分，将使用发行人自有资金进行投入。本项目各项投资支出的具体内容说明如下：

单位：万元

序号	项目	投资金额
设备购置		
1	研发设备	7,256.00
2	测试设备	218.00
3	服务器/工作站/DIY 平台	600.00
合计		8,074.00
软件购置		
1	EDA	1,448.64
合计		1,448.64
IP 授权费用		
1	高速接口类 IP	4,500.00
2	存储类 IP	1,500.00
3	基础单元类 IP	300.00
4	微控制器 IP	250.00
合计		6,550.00
人员工资		
1	芯片研发人员	4,800.00
2	软件研发人员	9,600.00
合计		14,400.00

序号	项目	投资金额
委外研发费用		
1	芯片设计委外研发—第三方	3,500.00
2	服务器合作 NRE	2,000.00
3	其他	240.00
合计		5,740.00
流片费用		
1	光罩	6,000.00
2	晶圆	700.00
3	封装/测试/板卡治具及 NRE	2,093.00
4	可靠性测试	300.00
合计		9,093.00

(3) 面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目

本项目主要是围绕前沿芯片技术研发和前沿 GPU 系统研发两大方向，开展行业前沿技术及基础技术的研究工作。其中，前沿芯片技术研发具体包括但不限于超高带宽显存、先进工艺芯粒架构、设计工艺联合优化、GPU 光互连技术等研发内容；前沿 GPU 系统研发具体包括但不限于大功率 POL 电源设计与优化、GPU 系统散热技术研究、scale up 互连方式优化改进、超节点服务器系统交换架构的演进与优化等研发内容。

本项目投资总额为 102,115.42 万元，其中拟使用募集资金投入金额为 99,141.19 万元，具体情况如下：

单位：万元

序号	项目	投资金额	占比	使用募集资金投入
1	场地费用	1,076.49	1.05%	1,076.49
2	设备购置	34,842.20	34.12%	34,842.20
3	软件购置	10,030.00	9.82%	10,030.00
4	IP 授权费用	4,500.00	4.41%	4,500.00
5	人员工资	24,000.00	23.50%	24,000.00
6	委外研发费用	6,000.00	5.88%	6,000.00
7	流片费用	14,420.00	14.12%	14,420.00
8	租赁费用	4,272.49	4.18%	4,272.49

序号	项目	投资金额	占比	使用募集资金投入
9	基本预备费	2,974.24	2.91%	-
项目总投资		102,115.42	100.00%	99,141.19

本项目投资总额中的基本预备费部分，将使用发行人自有资金进行投入。本项目各项投资支出的具体内容说明如下：

单位：万元

序号	项目	投资金额
场地费用		
1	新办公场所租赁费用	1,076.49
合计		1,076.49
设备购置		
1	研发设备	10,907.20
2	测试设备	20,710.00
3	服务器/工作站/DIY 平台	3,225.00
合计		34,842.20
软件购置		
1	EDA	8,000.00
2	通用类软件	1,850.00
3	测试/仿真软件	180.00
合计		10,030.00
IP 授权费用		
1	接口类 IP	1,500.00
2	高速接口类 IP	1,000.00
3	存储类 IP	1,000.00
4	基础单元类 IP	1,000.00
合计		4,500.00
人员工资		
1	芯片研发人员	10,400.00
2	软件研发人员	13,600.00
合计		24,000.00
委外研发费用		
1	芯粒委外研发	1,200.00
2	先进封装合作研发	1,200.00

序号	项目	投资金额
3	电路设计/失效分析服务	2,100.00
4	服务器合作 NRE	1,300.00
5	其他	200.00
合计		6,000.00
流片费用		
1	光罩	10,300.00
2	晶圆	1,120.00
3	封装/测试/板卡治具及 NRE	3,000.00
合计		14,420.00
租赁费用		
1	数据中心机柜租赁	4,272.49
合计		4,272.49

2、结合 C600、C700、N 系列芯片产品的客户开拓、市场销售情况，分析有关募投项目建设的必要性及未来发展安排

（1）曦云 C600

曦云 C600 系列是发行人基于国产工艺开发的首款通用 GPU 产品，核心技术及供应链自主可控，且在制程切换的情况下实现了性能升级，集成大容量显存与多精度混合算力，支持 MetaXLink 超节点扩展技术，并内置 ECC/RAS 多重安全防护模块，为金融、政务等关键领域提供高可靠性算力基座，满足下一代生成式 AI 的训练和推理需求。延续发行人目前的“1+6+X”生态与商业布局，算力基础设施中的国家人工智能公共算力平台、运营商智算中心以及垂直行业中金融、能源、教科研等央国企主导领域将是曦云 C600 系列的主要应用场景，该等场景对于信创要求较高，因此国产供应链产品得以大规模应用；而曦云 C600 在算力、计算精度、显存和互连等方面均具有较强竞争力，搭配高度兼容 CUDA 生态的 MXMACA 软件栈，产品综合优势显著。

曦云 C600 于 2025 年 7 月完成回片并点亮，根据初步测试表现，产品各方面性能符合设计预期。发行人前期已与各潜在客户就曦云 C600 进行推介和交流，并在 2025 年世界人工智能大会（WAIC）上正式发布了该款产品，获得市场广泛关注和积极反响。目前，上海、杭州等地的人工智能公共算力平台、部分头部互

联网企业以及多家商业银行高度关注曦云 C600 产品，均与发行人沟通并要求在产品验证流程结束后第一时间进行试用测评。此外，多家国内头部 OEM 厂商积极适配曦云 C600 系列产品，要求发行人提供样卡资源，相关定制化机型的 NRE 开发合同已在执行中。综上所述，从市场开拓、合作推进情况及存量客户延续性来看，发行人曦云 C600 产品未来的销售订单预计较有保障。

(2) 曦云 C700

曦云 C700 系基于更高制程的国产先进工艺开发，并采用发行人新一代 GPU 架构及指令集，同时将进一步扩展对 FP4 等低精度的计算支持，相比曦云 C600 系列整体性能更为强劲，定位是未来发行人国产供应链的旗舰产品，全面对标目前国际主流 GPU 产品英伟达 H100。因此，凭借更加高效的训练及推理效率，曦云 C700 产品广泛适用各类计算场景，在曦云 C600 信创市场需求的基础上更深入应用于互联网等领域，客户范围和销售规模也将相应扩大。

(3) Nx

Nx 是发行人专为生成式 AI 推理市场规划设计的一款产品，将通过三维堆叠技术及混合键合先进封装工艺，实现大容量显存和超高显存带宽，大幅提高 token 吞吐量的同时保持低功耗和高能效，显著降低 AI 推理成本。此外，Nx 可与发行人曦云 C 系列训推一体芯片互连，组成 Prefill-Decode 分离的推理系统，提升协同工作效率。未来，Nx 产品将广泛应用于国内商业化智算中心，为其提供高性价比的大模型推理解决方案，或者也可以用于 AI PC 等端侧低延迟、独立用户推理场景，支持从百亿到万亿各种参数规模的模型推理，市场前景可期。

综上所述，各募投产品具有良好的市场前景，曦云 C600 的客户开拓工作正在稳步推进。通过本次募投项目，将帮助发行人完成现有核心产品线的持续迭代升级和国产供应链切换，进一步巩固行业竞争优势，同时增强公司整体研发实力、丰富技术储备，是发行人实现未来发展目标的重要基础，具有充分的实施必要性。

(二) 发行人募投项目中 IP 授权费用、委外研发费用的具体内容，与自研 IP、自身研发活动的异同，是否涉及公司核心技术或涉及核心研发生产环节

1、IP 授权

发行人各个募投项目投资金额中有关 IP 授权费用的具体内容，以及所采购 IP 是否涉及发行人核心技术或核心研发环节的说明具体如下：

采购 IP 名称	金额（万元）	IP 类型	IP 具体内容或实现功能	不属于 GPU 设计关键核心 IP 的说明
新型高性能通用 GPU 研发及产业化项目—曦云 C600 子项目				
IP 许可 1	1,333.50	高速接口类	高速串行计算机扩展总线接口（PCIe）	标准接口，开放协议，业界通用
IP 许可 2	2,254.00	高速接口类	以太网物理接口、通用锁相环模块	标准接口，开放协议，业界通用
IP 许可 3	420.00	高速接口类	以太网接口控制器、远程直接内存访问 IP	标准接口，开放协议，业界通用
IP 许可 4	1,662.50	存储类	高带宽显存物理接口及其控制器	标准接口，开放协议，业界通用
合计	5,670.00			
新型高性能通用 GPU 研发及产业化项目—曦云 C700 子项目				
Memory Controller/PHY	1,500.00	存储类	高带宽显存物理接口及其控制器	标准接口，开放协议，业界通用
PCIe Controller/PHY	3,000.00	高速接口类	高速串行计算机扩展总线接口（PCIe）	标准接口，开放协议，业界通用
UCIe Controller/PHY	1,500.00	高速接口类	通用芯粒互连扩展总线接口（UCIe）	标准接口，开放协议，业界通用
PLL、展频、时钟传输单元	200.00	基础单元类	锁相环及其展频模块、单/双端时钟传输单元	通用基础类 IP 在国产工艺上的移植
基础 IP	500.00	基础单元类	PVT 监测单元、静态数据存储器件	通用基础类 IP 在国产工艺上的移植
数字微控制器 IP	250.00	微控制器	负责芯片的上电初始化、任务调度、功耗管理、中断及异常处理等	通用基础类 IP
数字寄存器总线 IP	150.00	总线类	提供芯片的寄存器读写访问通路	通用基础类 IP

采购 IP 名称	金额（万元）	IP 类型	IP 具体内容或实现功能	不属于 GPU 设计关键核心 IP 的说明
数字加密 IP	300.00	安全类	实现中国商用密码算法，保障芯片信息安全	通用基础类 IP
数字 PPA 相关定制化 IP	350.00	基础功耗管理类	上电复位单元、数字延时器、低压差线性稳压器	通用基础类 IP 在国产工艺上的移植
Foundry IP 定制	1,000.00	基础单元类	数字电压域通用输入输出单元、电平转换模块、同步器、多比特锁存器	通用基础类 IP 在国产工艺上的移植
合计	8,750.00			
新一代人工智能推理 GPU 研发及产业化项目				
Memory Controller/PHY	1,500.00	同上	同上	同上
PCIe Controller/PHY	3,000.00	同上	同上	同上
UCIe Controller/PHY	1,500.00	同上	同上	同上
PLL、展频、时钟传输单元	200.00	同上	同上	同上
基础 IP	100.00	同上	同上	同上
数字微控制器 IP	250.00	同上	同上	同上
合计	6,550.00			
面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目				
超高带宽显存—基础 IP	1,000.00	存储类	高带宽显存物理接口及其控制器	标准接口，开放协议，业界通用
超高带宽显存—3D 相关 IP	1,000.00	接口类	对准标记、带有硅通孔的三维接口单元、双倍速率源同步总线	晶圆厂通用方案，标准接口或协议
先进工艺芯粒架构—UCIe（12nm）	500.00	高速接口类	通用芯粒互连扩展总线接口（UCIe）	标准接口，开放协议，业界通用
先进工艺芯粒架构—UCIe（7nm）	500.00	高速接口类	通用芯粒互连扩展总线接口（UCIe）	标准接口，开放协议，业界通用
先进工艺芯粒架构—基础 IP	500.00	接口类	数字电压域通用输入输出单元、双端口差分时钟传输接口单元、DC-DC 电源转换模块	通用方案

采购 IP 名称	金额（万元）	IP 类型	IP 具体内容或实现功能	不属于 GPU 设计关键核心 IP 的说明
设计工艺联合优化—基础 IP	1,000.00	基础单元类	优化芯片数字逻辑的功耗、频率、面积的基础单元库	通用基础类 IP 在国产工艺上的移植
合计	4,500.00			

由上表可见，发行人对外采购的 IP 主要是行业内已经较为成熟、标准化的通用基础类 IP 和接口类 IP，并不属于 GPU 设计所涉及的关键核心 IP。发行人自研构建 GPU 完整核心 IP 阵列，包括但不限于全数值精度矩阵/矢量计算单元、指令调度单元、数据存取单元和着色/几何/纹理/光追单元等渲染专用硬件，以及多级缓存系统、全局任务调度器、片上互连总线等核心 IP。此外，发行人自主研发大量基础功能模块，例如安全初始化与芯片管理模块、芯片中断与异常控制模块、直接显存和系统内存访问模块、电源抖动自适应时钟模块、存储单元优化等各类型 IP，覆盖 GPU 芯片设计全方位需求。

2、委外研发

发行人各个募投项目投资金额中有关委外研发费用的具体内容，以及委外研发是否涉及发行人核心技术或核心研发环节的说明具体如下：

委外研发项目	金额（万元）	委外工作的具体内容	不涉及公司核心技术或核心研发环节的说明
新型高性能通用 GPU 研发及产业化项目—曦云 C600 子项目			
芯片设计委外研发—子公司	8,240.00	由于本项目实施主体为发行人母公司，同时部分全资子公司的研发人员参与本项目，因此母公司通过委外形式与子公司进行费用结算	如前所述，实际非外部委托，不适用
软件设计委外研发—子公司	12,320.00	同上	同上
芯片设计委外研发—第三方	408.10	封装设计开发	主要是基于芯片的 ubump 和 pinlist 等外围信息做延展性设计，不属于核心研发工作
软件设计委外研发—第三方	2,000.00	部分非核心算子的优化、大模型适配调优	仅为增强产品生态鲁棒性和泛化性，非核心工作
后端开发	3,059.70	依据发行人提交的模块网表，完成模块的后端设	后端设计的关键环节包括全芯片布局、高速时

委外研发项目	金额（万元）	委外工作的具体内容	不涉及公司核心技术或核心研发环节的说明
		计以及模块级别的设计验证工作	钟、布局布线和全芯片的签核流程等均由发行人负责，委外部分主要是模块级别的后端实现工作，不涉及发行人核心技术或核心研发环节
服务器合作 NRE	1,000.00	服务器通用基板（UBB）开发	UBB 主要是支持发行人产品基于外部结构的互连以及服务器整机适配，非发行人核心研发内容
硅中介层试制	106.59	发行人独立设计硅中介层后委外进行制造	相关设计均由发行人完成
第三方测试和认证	160.00	电磁兼容测试、安规测试、可靠性测试、RoHS 化学认证、震动和冲击实验	第三方实验室按照国际标准完成相关测试和认证
PI/SI 仿真委外	80.00	PA 和 OAM 卡的高速信号 SI 仿真	芯片 PI/SI 模型及 PCB 板文件均由发行人提供，仅输出仿真报告和设计优化建议
合计	27,374.39		
新型高性能通用 GPU 研发及产业化项目—曦云 C700 子项目			
芯片设计委外研发—子公司	9,600.00	同上	同上
软件设计委外研发—子公司	19,120.00	同上	同上
芯片设计委外研发—第三方	5,000.00	依据发行人提交的模块网表，完成模块的后端设计以及模块级别的设计验证工作	同上
软件设计委外研发—第三方	3,500.00	同上	同上
服务器合作 NRE	3,000.00	同上	同上
第三方测试和认证	320.00	同上	同上
PI/SI 仿真委外	160.00	同上	同上
合计	40,700.00		
新一代人工智能推理 GPU 研发及产业化项目			
芯片设计委外研发—第三方	3,500.00	依据发行人提交的模块网表，完成模块的后端设计以及模块级别的设计验证工作	同上
服务器合作 NRE	2,000.00	同上	同上

委外研发项目	金额（万元）	委外工作的具体内容	不涉及公司核心技术或核心研发环节的说明
第三方测试和认证	160.00	同上	同上
PI/SI 仿真委外	80.00	同上	同上
合计	5,740.00		
面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目			
超高带宽显存—存储核心芯粒委外研发	1,200.00	根据超高带宽显存的特性定制/优化存储核心芯粒	存储芯粒的研发设计不是发行人核心技术
先进工艺芯粒架构—混合键合封装合作研发	1,200.00	使用混合键合技术进行异构芯粒和晶圆的堆叠，完成设计规则定义及验证	由发行人拆解应用场景，并提取具体的需求指标给到合作方（核心研发环节），合作方主要负责生产工艺的研发
设计工艺联合优化—电路设计服务	1,200.00	结合 GPU 的需求以及国产工艺特点，研发专用标准单元及定制电路	由发行人完成 GPU 对于定制单元、定制电路的需求分析（核心研发环节），合作方仅负责设计单独电路，对整体支持方案未触达
设计工艺联合优化—失效分析服务	900.00	对各环境下的芯片失效做根因分析，并反馈给工厂做相应的改善或拦截，提升产品质量	该分析过程中所需的芯片结构、电路设计、版图和 DFT 测试程序以及失效理论分析均由发行人负责，委外研发工作主要是使用专业检测仪器完成失效根因确认
GPU 超节点系统设计—服务器合作 NRE	1,300.00	超节点服务器硬件系统设计、整机功能测试、集群性能测试等	超节点服务器的研发需要基于发行人自研的 MetaXLink 和高速以太网口等卡间互连技术，整体设计方案由发行人内部软硬件部门协同打造，委外部分仅为服务器硬件实现，且委外研发成果及知识产权归属于发行人
GPU 超节点系统设计—第三方测试和认证	200.00	电磁兼容测试、安规测试、可靠性测试、RoHS 化学认证、震动和冲击实验	第三方实验室按照国际标准完成相关测试和认证
合计	6,000.00		

剔除发行人内部母子公司之间委托工作内容后，曦云 C600、曦云 C700 研发项目中委外研发费用占项目总投资金额的比例分别为 4.97%和 5.87%，占比较低。发行人在产品研发过程中，涉及委外的主要是芯片后端设计、软件测试优化、服务器适配等

环节一些技术难度、重要性较低或者非公司主业的工作内容，不涉及发行人核心技术或核心研发环节；发行人将该等工作进行委外，有助于发行人聚焦核心技术研发迭代，加快产品整体开发进度，符合行业惯例。发行人持续深耕 GPU 领域，坚持自主可控的发展路线，在 GPU 核心计算架构、互连架构和高性能 SoC 系统集成等方面积累了丰富的丰富经验，通过自主研发形成了具备较强行业竞争力的完整核心技术体系。

（三）结合公司目前可支配资金情况，说明募集资金规模是否合理。本次募投项目预计收益情况，折旧摊销、期间费用等对发行人未来经营业绩及财务数据的具体影响，并针对性进行重大事项提示

1、结合公司目前可支配资金情况，说明募集资金规模是否合理

发行人本次募集资金 390,366.59 万元，募资规模系根据募投项目资金使用需求以及本次发行比例、预计市值情况综合测算确定。

（1）可支配资金情况

截至 2025 年 3 月 31 日，发行人货币资金余额为 569,461.33 万元，其中受限货币资金为 198.06 万元，主要是银行承兑汇票、保函的保证金；交易性金融资产余额为 159,688.21 万元，系公司已购买未到期的结构性存款。因此，报告期末发行人可支配资金为 728,951.49 万元，测算如下：

单位：万元

项目	计算公式	金额
货币资金	①	569,461.33
其中：受限货币资金	②	198.06
交易性金融资产	③	159,688.21
可支配资金	④=①-②+③	728,951.49

（2）未来三年营运资金需求

在其他经营要素保持不变的情况下，以 2024 年为基期，根据发行人 2025 年、2026 年及 2027 年预计销售收入，采用销售百分比法测算发行人 2025 年末、2026 年末及 2027 年末的经营性流动资产和经营性流动负债（假设各项资产和负债占营业收入的比例保持不变，与 2024 年末水平一致），据此计算未来三年的营运资金需求。

经测算，预计发行人 2025 年、2026 年和 2027 年的营运资金需求分别为 194,600.35 万元、362,325.23 万元和 292,049.82 万元，三年合计为 848,975.40 万元，因此发行人目前账面可支配资金 728,951.49 万元将主要用于填补未来三年日常经营活动所产生的新增流动资金缺口，而本次发行上市募集资金则用于投入公司相关研发项目，募集资金用途中不包括补充公司流动资金以及各募投项目的铺

底流动资金，募资规模具有合理性。

2、本次募投项目预计收益情况，折旧摊销、期间费用等对发行人未来经营业绩及财务数据的具体影响，并针对性进行重大事项提示

发行人本次募投项目中，新型高性能通用 GPU 研发及产业化项目和新一代人工智能推理 GPU 研发及产业化项目在相关募投产品完成研发并投入市场后，预计能够获得良好的经济效益。经测算，各项目的内部收益率情况如下：

序号	募投项目	内部收益率（税后）
1	新型高性能通用 GPU 研发及产业化项目—曦云 C600 子项目	20.65%
2	新型高性能通用 GPU 研发及产业化项目—曦云 C700 子项目	24.16%
3	新一代人工智能推理 GPU 研发及产业化项目	21.35%

面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目为纯研发类项目，并不直接产生经济效益。

本次募投项目实施过程中，将产生一定的研发费用，包括购置设备、软件及 IP 等资产而形成的折旧摊销额，以及支付职工薪酬、委外研发等费用化支出。但随着募投项目的逐步达产，发行人未来经营规模及盈利水平亦将有所提升，前述折旧摊销、期间费用对发行人未来经营业绩的影响有限。本次募投项目是发行人未来主要的研发方向，相关费用支出为发行人未来主要的研发投入，因此该等项目的实施不会对发行人经营业绩产生额外的不利影响。

针对以上事项，发行人已在招股说明书“第二节 概览”之“一、重大事项提示”之“（一）重大风险提示”及“第三节 风险因素”之“一、与发行人相关的风险”之“（二）财务风险”补充披露如下内容：

“8、募投项目新增较大费用支出的风险

本次募集资金投资项目实施过程中，将产生较大金额的折旧摊销和其他费用支出，对发行人各年度经营业绩有直接影响。由于募投项目的建设、完工及产生效益需要一定的时间周期并且可能存在各种不确定性，若未来发行人所处市场环境等因素发生重大不利变化，导致募投项目无法实现预计效益，则前述新增的折旧摊销、期间费用等将对发行人未来经营业绩构成较大不利影响。”

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构、申报会计师执行了如下核查程序：

1、查阅发行人募投项目可行性研究报告及其中在研项目的立项资料，访谈发行人管理层，了解募投项目的具体研发内容、资金使用安排和募投产品未来市场需求情况、目前客户接洽情况。

2、获取并复核募投项目资金使用明细、效益测算的相关数据底稿，分析测算合理性和准确性。

3、获取发行人关于募投项目支出中涉及外采 IP 和委外研发的具体内容说明，访谈发行人技术负责人，了解相关 IP 和委外内容是否涉及发行人核心技术或核心研发环节。

4、结合盈利预测数据，测算发行人未来三年营运资金需求缺口，论证本次募资规模合理性。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、发行人已具体说明本次募投项目的研发内容、资金使用安排以及募投产品未来市场需求情况，本次募投项目将帮助发行人完成现有核心产品线的迭代升级和国产供应链切换，并进一步增强发行人整体研发实力与技术储备，具有必要性。

2、发行人已说明各募投项目中 IP 授权费用、委外研发费用的具体内容，发行人对外采购的 IP 和委外研发内容不涉及发行人核心技术或核心研发环节。

3、结合发行人目前可支配资金、未来三年营运资金需求、募集资金用途等情况，本次募集资金规模具有合理性；本次募投项目预计效益良好，折旧摊销、期间费用等对发行人未来经营业绩影响较为有限，发行人已就相关事项在招股说明书中进行了补充风险提示。

14.关于信息披露

根据申报材料：（1）招股说明书重大事项提示及风险因素中部分内容的针对性不强，如“技术人才流失或储备不足的风险”“规模扩张带来的管理风险”等；（2）招股说明书中关于发行人技术水平、市场竞争状况等说明不够充分。

请发行人按照《公开发行证券的公司信息披露内容与格式准则第 57 号——招股说明书》等规则要求：（1）结合公司实际情况梳理“重大事项提示”“风险因素”各项内容，突出重大性、增强针对性，充分披露风险产生的原因和对发行人的影响；（2）结合发行人所处细分领域的行业发展状况、技术水平、市场竞争格局以及发行人自身情况等，对招股说明书“业务与技术”章节进行完善和针对性披露。

请保荐机构简要概括核查过程，并发表明确意见。

回复：

一、发行人披露

（一）结合公司实际情况梳理“重大事项提示”“风险因素”各项内容，突出重大性、增强针对性，充分披露风险产生的原因和对发行人的影响

发行人已按照要求结合公司实际情况梳理了招股说明书“第二节 概览”之“一、重大事项提示”和“第三节 风险因素”各项内容，突出重大性、增强针对性，充分披露风险产生的原因和对发行人的影响。

（二）结合发行人所处细分领域的行业发展状况、技术水平、市场竞争格局以及发行人自身情况等，对招股说明书“业务与技术”章节进行完善和针对性披露

发行人已结合本次问询之“1.关于产品与市场竞争”及“2.关于核心技术”相关回复在招股说明书“第五节 业务与技术”章节就发行人所处细分领域的行业发展状况、技术水平、市场竞争格局以及发行人自身情况进行补充完善和针对性披露。

二、中介机构核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构执行了如下核查程序：

- 1、查阅招股说明书，比对相关章节披露内容的具体修改情况。
- 2、查阅发行人所在行业的相关产业政策、所属细分领域的行业研究报告以及同行业公司的公开资料。
- 3、访谈发行人技术负责人有关发行人技术水平、竞争地位的情况。

（二）核查意见

经核查，保荐机构认为：

- 1、发行人已按照要求对招股说明书中重大事项提示和风险因素相关内容进行了修改完善，增强风险提示的针对性。
- 2、发行人已结合自身情况在招股说明书中补充披露发行人所处细分领域的行业发展状况、技术水平、市场竞争格局等。

保荐人总体意见

对本回复材料中的发行人回复（包括补充披露和说明的事项），本保荐机构均已进行核查，确认并保证其真实、完整、准确。

（以下无正文）

发行人董事长声明

本人已认真阅读《关于沐曦集成电路（上海）股份有限公司首次公开发行股票并在科创板上市申请文件的审核问询函的回复》的全部内容，确认回复的内容真实、准确、完整，不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担相应法律责任。

发行人董事长：



陈维良

沐曦集成电路（上海）股份有限公司



2025年8月27日

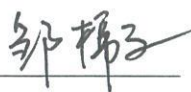
（本页无正文，为《关于沐曦集成电路（上海）股份有限公司首次公开发行股票并在科创板上市申请文件的审核问询函的回复》之签章页）

沐曦集成电路（上海）股份有限公司



（本页无正文，为华泰联合证券有限责任公司《关于沐曦集成电路（上海）股份有限公司首次公开发行股票并在科创板上市申请文件的审核问询函回复》之签章页）

保荐代表人：


邹棉文


孙琪

华泰联合证券有限责任公司

2025 年 8 月 27 日



保荐人法定代表人声明

本人已认真阅读沐曦集成电路（上海）股份有限公司本次审核问询函回复的全部内容，了解问询函回复涉及问题的核查过程、本公司的内核和风险控制流程，确认本公司按照勤勉尽责原则履行核查程序，审核问询函回复不存在虚假记载、误导性陈述或者重大遗漏，并对上述文件的真实性、准确性、完整性、及时性承担相应法律责任。

保荐人法定代表人：


江禹

华泰联合证券有限责任公司

2025年8月27日

（本页无正文，为立信会计师事务所（特殊普通合伙）《关于沐曦集成电路（上海）股份有限公司首次公开发行股票并在科创板上市申请文件的审核问询函的回复》之签字盖章页，我们仅对审核问询函中需要发行人会计师进行核查的事项发表核查意见。）



中国注册会计师：李新民

A handwritten signature in black ink, appearing to read "李新民".



中国注册会计师：戴志敏

A handwritten signature in black ink, appearing to read "戴志敏".



中国·上海

二〇二五年八月二十七日

(本页无正文,为《上海市锦天城律师事务所《关于沐曦集成电路(上海)股份有限公司首次公开发行股票并在科创板上市申请文件的审核问询函的回复》之律师签署页,我们仅对审核问询函中需要发行人律师进行核查的事项发表核查意见)

上海市锦天城律师事务所(盖章)



负责人:



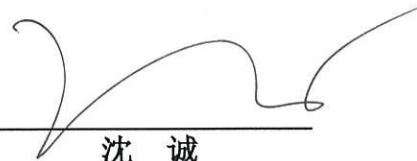
沈国权

经办律师:



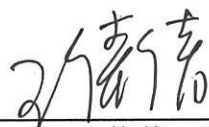
王立

经办律师:



沈诚

经办律师:



王倩倩

2021年8月27日

上海·北京·深圳·杭州·苏州·南京·成都·重庆·太原·青岛·厦门·天津·济南·合肥·郑州·福州·南昌·西安·广州·长春·武汉·乌鲁木齐·长沙·海口·昆明

地址: 上海市浦东新区银城中路 501 号上海中心大厦 9/11/12 层, 邮编: 200120
电话: (86) 21-20511000; 传真: (86) 21-20511999
网址: <http://www.allbrightlaw.com/>