

**关于深圳云豹智能股份有限公司
首次公开发行股票并在创业板上市申请文件的
审核问询函之回复报告**

保荐人（主承销商）



广东省深圳市福田区中心三路 8 号卓越时代广场（二期）北座

深圳证券交易所：

根据贵所于 2026 年 7 月 7 日下发的《关于深圳云豹智能股份有限公司首次公开发行股票并在创业板上市申请文件的审核问询函》（审核函〔2026〕010100 号）（以下简称“问询函”），中信证券股份有限公司（以下简称“保荐人”、“保荐机构”或“中信证券”）作为深圳云豹智能股份有限公司（以下简称“公司”、“发行人”或“云豹智能”）首次公开发行股票并在创业板上市的保荐人，已会同发行人、发行人申报会计师天健会计师事务所（特殊普通合伙）（以下简称“申报会计师”）、发行人律师北京市中伦律师事务所（以下简称“发行人律师”）、发行人评估机构沃克森（北京）国际资产评估有限公司（以下简称“评估机构”），本着勤勉尽责、诚实守信的原则，就问询函所提问题逐条进行了认真核查、讨论及回复，具体情况如下文，请予审核。

除另有说明外，本回复报告所用简称或名词释义与《深圳云豹智能股份有限公司首次公开发行股票并在创业板上市招股说明书（申报稿）》中的释义相同。

本问询函回复中的字体：

问询函所列问题	黑体（加粗）
对问询函所列问题的回复	宋体
对招股说明书的引用	楷体（不加粗）
对招股说明书的修改、补充	楷体（加粗）

本回复报告中若出现总计数尾数与所列数值总和尾数不符的情况，均为四舍五入所致。

目录

问题 1.关于主营业务与核心竞争力	3
问题 2.关于盈利预期与前瞻性信息	51
问题 3.关于关联交易	82
问题 4.关于实际控制人与股东情况	100
问题 5.关于收入与客户	106
问题 6.关于毛利率	111
问题 7.关于存货	115

问题 1.关于主营业务与核心竞争力

申报材料显示：发行人为率先实现 400Gbps 高性能全功能 DPU 芯片技术突破与规模化商用的国产芯片厂商。DPU 在算力集群中的配置价值提升。发行人产品分为全功能 DPU 及 AI DPU 两类，分别对标英伟达 BlueField 系列及 ConnectX 系列。

请发行人披露：

（1）结合已有及在研 DPU 产品在云计算、AI 智算等典型下游场景中的具体功能，分析 DPU 产品对各类下游使用场景的必要性及重要性。

（2）结合推理、AI Agent、AI 编程等典型场景下，可由 DPU 卸载的 CPU 任务占比及 DPU 卸载效率、提高集群综合算力效率比例等，分析 DPU 提升数据中心整体性能、降低集群算力综合成本的具体体现，客户采用相关方案的成本收益情况对发行人产品开拓的影响。

（3）全功能 DPU 与 AI DPU 的功能及适用场景；结合下游 AI 行业发展趋势及不同 AI 应用领域对两类产品需求情况、全功能 DPU 及 AI DPU 产品使用数量与 CPU、GPU、服务器、节点数量之间的配比关系，分析产品销量增长潜力，下游对发行人产品需求变化趋势，发行人研发方向与下游需求的匹配性。

（4）当前发行人 DPU 产品与数据中心主流 CPU/GPU 硬件、网络架构和软件生态之间的适配情况，现有产品与现有客户的适配情况，导入新客户所需适配周期及适配难度，结合其他客户验证及在手订单情况，分析开拓及适配其他客户的能力；行业主流 CPU/GPU 厂商自研 DPU 的情况，采用发行人 DPU 产品与搭配其自研 DPU 的综合性能、性价比对比情况，发行人 DPU 产品面临的竞争情况。

（5）结合不同下游应用领域、下游行业主要客户合作情况，分析发行人市场拓展能力。

请保荐人说明核查依据、过程，并发表明确意见。

发行人披露：

一、结合已有及在研 DPU 产品在云计算、AI 智算等典型下游场景中的具体功能，分析 DPU 产品对各类下游使用场景的必要性及重要性

（一）公司已有和在研 DPU 产品情况

公司是国内领先的数据处理器（DPU）芯片设计企业，专注于 DPU 及相关产品的研发、设计和销售。作为率先实现 400Gbps 高性能全功能 DPU 芯片技术突破与规模化商用的国产芯片厂商，公司具备对标国际巨头的技术实力。公司 DPU 创新地实现了通用可编程能力，能够提供包括弹性网络与存储、虚拟化、安全及低时延 RDMA 等功能在内的一站式解决方案。公司 DPU 作为支持人工智能与数字经济产业发展的关键网络基座，承担了低时延传输、存储、数据管控、AI 加速及安全等核心职责，广泛应用于 AI 智算、云计算、存储、通信、网络安全及边缘计算等关键战略领域。

公司已围绕 DPU 核心功能、应用场景构建了软硬件一体化的全栈产品体系，在产品布局上，云霄 1.0 系列全功能 DPU 为多核异构 SoC 产品，集成 64 个自研 RISC-V 微处理器核和 16 个 Arm V9 架构 Neoverse N2 高性能核，采用创新的层级化可编程芯片架构实现 DPU 的通用可编程性，以灵活适配大规模组网的复杂计算场景；风驰 2.0 系列 AI DPU 重点面向智算网络优化，在简化架构、降低功耗的基础上，进一步优化大规模组网场景下 RDMA 性能、可编程拥塞控制与乱序重排功能，能更好适配 AI 智算集群中 GPU 跨服务器高速互联的网络需求，同时也支持私有云、边缘计算、存储等场景对高性能 RDMA 网卡的需求；公司同步基于 Chiplet 技术和风驰 2.0 芯片研发最高支持 2*400Gbps 的风驰 2.5 系列 AI DPU。此外，公司已启动研发的云霄 2.0 系列与风驰 3.0 系列为新一代最高支持 800Gbps/1.6Tbps 网络带宽的高性能 DPU 产品。公司产品广泛应用于 AI 智算、云计算、存储、通信、网络安全及边缘计算等关键战略领域。

公司已有和在研 DPU 主要产品如下：

系列名称	产品状态	网络带宽	主要应用场景	产品介绍	产品示意图
云霄 1.0	量产	400Gbps	AI 智算、云计算、通信、边缘计算、AI 存储、智能交换机/网关等	16 个服务器级 Arm V9 Neoverse N2 CPU 核； 64 个自研 RISC-V 多线程微处理器； 最高支持 2*200GE 端口； 灵活高性能可编程引擎； 低时延 RoCEv2； 支持 PCIe Gen5.0	
风驰 2.0	流片成功	400Gbps	AI 智算、云计算、通信、通用服务器、存储服务器等	最高支持 1*400GE 或 2*200GE 端口； 支持低时延 RoCEv2 及高级拥塞控制算法； 支持逐包负载均衡及包喷洒等高级特性； 基于 RISC-V 可编程数据路径可实现对 GSE 的支持； 支持 PCIe Gen5.0	
风驰 2.5	在研	800Gbps	AI 智算、云计算、通信、通用服务器、存储服务器等	通过 Chiplet 技术最高支持 2*400GE 或 4*200GE 端口； 支持低时延 RoCEv2 及高级拥塞控制算法； 支持逐包负载均衡及包喷洒等高级特性； 基于 RISC-V 可编程数据路径可实现对 GSE 的支持； 支持 PCIe Gen5.0	-
云霄 2.0	在研	800Gbps	AI 智算、云计算、通信、边缘计算、AI 存储、智能交换机/网关等	最多 32 个服务器级 Arm V9 Neoverse N2 CPU 核； 96 个第三代自研 RISC-V 多线程微处理器； 最高支持 1*800GE 或 2*400GE 端口； 支持接收端驱动的基于信用的流量控制及用户自定义拥塞控制算法	-
风驰 3.0	在研	800Gbps/ 1.6Tbps	AI 智算、云计算、通信、通用服务器、存储服务器等	支持 2*800GE/4*400GE /1*800GE/2*400GE 等端口； 支持接收端驱动的基于信用的流量控制及用户自定义拥塞控制算法；	-

系列名称	产品状态	网络带宽	主要应用场景	产品介绍	产品示意图
				支持选择性重传，支持包截断、逐流多路径、逐包负载均衡、包喷洒、乱序接收和重排等	

公司主要产品适用场景及主要功能如下：

DPU 类别	主要产品系列	场景需求	具体功能
全功能 DPU	云霄系列	云计算前端网络；多租户虚拟化平台；AI 智算；AI 存储网络	设备虚拟化与网络虚拟化，弹性网络、低时延 RDMA、存储、安全、管控的卸载及加速，可编程数据平面硬件卸载，KV-Cache 缓存管理及加速
AI DPU	风驰系列	AI 智算；云计算后端网络；高性能训练/推理集群	大规模 RDMA 网络通信，多路径负载均衡，可编程拥塞控制，快速重传能力，存储卸载

全功能 DPU 侧重在大规模多租户的云网络架构中扮演重要角色，主要用于支撑弹性网络、云存储、虚拟化、安全以及数据管控等基础设施功能的运行，通过软硬件协同实现对复杂数据与网络处理的全场景应用，是数据中心、智算集群的重要数据面和控制面组件。AI DPU 侧重面向 AI 智算集群的后端网络场景，围绕大模型训练与推理过程中产生的庞大上下文数据流，承担集群协同、高带宽低时延 RDMA 传输、可编程拥塞控制与数据调度等关键功能，为大规模 AI 算力集群的高效运行提供支撑。

（二）DPU 在云计算场景中的具体功能、必要性及重要性

1、云计算场景概述

云计算场景是 DPU 产品最核心的下游应用领域之一。在数据中心的场景中，云服务商通过虚拟化技术将物理服务器的计算、网络 and 存储资源池化，以弹性裸金属、虚拟机和容器等形式向租户提供云服务。该场景下，云服务商的核心需求之一是构建弹性裸金属、虚拟机及容器云服务“三合一”统一架构。每台物理服务器通常需同时运行多租户业务应用和大量基础设施服务（包括网络虚拟化、存储虚拟

化、安全隔离、资源调度等），传统架构中这些基础设施服务由 Host CPU 以软件方式处理，随着网络带宽从 10G/25G 向 100G/400G 演进，基础设施开销占用的 CPU 算力比例持续攀升，成为制约数据中心效率提升的关键瓶颈。

2、云计算场景中 DPU 的具体功能

公司云霄系列全功能 DPU 在云计算场景中实现管控、存储、网络和安全的全面卸载，具体功能如下：

（1）网络虚拟化与协议卸载

虚拟交换机（OVS）卸载：云霄 DPU 将 OVS 网络业务从 Host CPU 卸载至 DPU，实现数据包解析、匹配和转发的 DPU 加速，显著降低 OVS 网络业务对 CPU 算力的消耗。

Overlay 协议支持：全面支持 VXLAN、GRE、GENEVE 等网络协议的硬件卸载，实现虚拟网络与物理网络的高效隔离与互通，支持多租户网络环境下的大规模网络虚拟化部署。

Virtio 虚拟化加速：通过硬件加速 Virtio 虚拟化框架，在提供更高性能和更低时延的同时，使数据中心从传统网卡向 DPU 的迁移更加平滑，保障业务连续性。

弹性网络：支持弹性网卡功能，按需动态分配网络带宽和虚拟网络接口资源，满足云租户对网络资源的弹性伸缩需求。

（2）存储虚拟化、卸载及加速

NVMe-oF 存储协议卸载：通过硬件加速 NVMe over Fabrics 协议，将网络存储虚拟化为本地高性能 NVMe 设备，使远程存储性能接近本地存储水平，显著提升云存储服务的吞吐量和延迟表现。

Virtio-blk 存储虚拟化加速：通过硬件加速 Virtio-blk 虚拟化框架，实现块设

备虚拟化的性能提升，支持云盘启动功能，使虚拟机和容器从远程存储启动时获得接近本地存储的性能体验。

弹性存储：支持弹性存储功能，按需动态分配存储资源，并支持针对不同弹性存储实施差异化策略，满足云租户对存储资源的灵活配置需求。

RDMA 存储访问：支持 RoCE v1&2 协议，通过 RDMA 技术实现高吞吐、低延迟的远程存储访问，支持高性能存储网络连接。

（3）虚拟化资源统一管理

弹性裸金属、虚拟机和容器云服务“三合一”统一架构：云霄 DPU 为数据中心提供“三合一”统一架构和运维解决方案，全面解决弹性裸金属、虚拟机和容器三种不同场景的资源虚拟化和统一管理，实现不同类型业务上云及无感知迁移。

热插拔与热迁移：支持热插拔、热迁移，可实现在客户业务不中断的情况下进行资源迁移，保证在宕机时业务快速恢复，支持业务的弹性扩展和无感迁移。

热升级：支持在不中断业务的情况下对 DPU 固件进行热升级和迭代，确保基础设施持续优化不影响业务连续性。

（4）安全隔离与零信任架构

安全服务域隔离：云霄 DPU 为每台服务器提供独立的安全服务域，通过与服务器的应用域隔离实现零信任架构（Zero Trust，指一种以资源保护为核心的网络安全理念，认为对资源的访问，无论主体和资源是否可信，主体和资源之间的信任关系都需要从零开始，通过持续环境感知与动态信任评估进行构建，从而实施访问控制），即使 Host CPU 被入侵，DPU 上的安全控制平面仍可独立运行，防止攻击向整个数据中心扩散。

加密加速：支持 IPSEC 和 TLS 的硬件加密加速，执行数据传输加密，在不

影响其他加速功能的前提下保障数据传输安全。

安全生命周期管理：支持信任根、安全启动、安全固件升级及经过身份验证的容器和应用程序生命周期管理，构建从芯片到应用的完整安全链条。

服务质量(QoS)保障：为不同租户和不同业务流提供高标准服务质量(QoS)保证，确保关键业务的网络和存储性能不受其他租户影响，保障多租户环境下的服务等级协议（SLA）承诺。

3、DPU 对云计算场景的必要性

（1）消除“数据中心税”的刚性需求

根据谷歌 2015 年研究数据，在传统数据中心架构中，约 30%的 CPU 算力被用于网络协议处理、存储虚拟化、安全服务等基础设施开销，该部分开销被称为“数据中心税”。随着云计算网络带宽从 10G/25G 向 100G/400G 演进，以软件方式处理基础设施协议栈的 CPU 开销呈非线性增长趋势。例如，在 400Gbps 带宽下，处理网络数据包接收、协议栈解析、数据包转发等需要消耗数十个 CPU 核心。在摩尔定律放缓的背景下，CPU 算力增长已无法匹配网络带宽和数据量的增长速度，二者形成日益扩大的“剪刀差”。DPU 通过将基础设施任务从 CPU 卸载，是消除“数据中心税”、释放 CPU 算力用于核心业务的刚性需求。

（2）多租户安全隔离的架构性需求

云计算场景中，多租户共享同一物理服务器资源，安全隔离是云服务商的底层架构需求。传统基于软件的安全隔离方案存在“被攻破即扩散”的风险，一旦 Host CPU 的操作系统或虚拟机管理器被入侵，攻击者可能获取整个服务器的控制权。DPU 通过将安全控制平面独立于 Host CPU 运行，构建零信任架构，即使主机被入侵，DPU 上的安全策略仍可独立执行，有效防止攻击横向扩散。该安全隔离能力是云服务商实现多租户安全承诺的架构性基础，基于 DPU 实现是优于纯软件方案的新一代安全方案。

（3）虚拟化资源统一管理的业务连续性需求

大型云数据中心同时运行弹性裸金属、虚拟机和容器三种计算形态，传统方案中，不同计算形态的网络和存储管理相互割裂，导致运维复杂度高、迁移成本大。DPU 通过统一的硬件卸载层，实现三种计算形态的资源一体化管理和性能加速，支持数据中心基础设施在不同计算形态间的弹性迁移。该能力对于云服务商降低运维成本、提高云基础设施利用率和保障业务连续性具有不可替代的作用。

（4）弹性资源调度的效率需求

客户对云计算算力需求存在动态波动，云计算的核心价值之一在于算力资源的弹性伸缩。DPU 支持弹性网卡和弹性存储功能，能够按需动态分配网络和存储资源，使云服务商可以根据业务负载动态调整资源配额，提升资源利用效率。传统基于软件的弹性资源调度方案受限于 CPU 处理能力，难以在高带宽场景下实现毫秒级的资源弹性响应，DPU 的硬件加速是实现算力资源的高效弹性调度的必要保障。

4、DPU 对云计算场景的重要性

（1）提升数据中心整体算力效率

公司云霄系列 DPU 通过将网络、存储、安全等基础设施任务全面卸载至 DPU，单台服务器可释放数十个 CPU 核心用于运行租户业务应用，直接提升数据中心的可售卖算力。以一台配置双路 64 核（共 128 核）CPU 的服务器为例，卸载 30% 的基础设施开销后，可释放约 38 个 CPU 核心用于业务计算（ $128 \times 30\% = 38.4$ ），从而提升了可售卖算力，对云服务商的收入具有直接的正面影响。

（2）降低数据中心总体拥有成本（TCO）

DPU 通过硬件加速提升基础设施处理效率，在相同性能需求下可减少所需的 CPU 核心数量，降低服务器采购成本；同时，功耗降低有助于降低数据中心的电力和散热成本。根据英伟达发布的 BlueField DPU 能效白皮书，使用 DPU

来卸载和加速网络、安全、存储或其他基础架构功能以及控制平面应用，可将服务器耗电量显著降低。节省的电量会随着服务器负载的增加而增加，在大规模数据中心中累计节省的运营成本十分可观。

（3）增强云服务安全性与合规性

DPU 提供的硬件级安全隔离能力，为云服务商满足数据安全法规等监管要求提供了底层技术支撑。在金融、政务等对数据安全要求极高的行业云场景中，DPU 的零信任架构和安全服务域隔离能力是云服务商获取相关资质认证和客户信任的重要技术基础。公司 DPU 产品已在工商银行、招商银行、南方电网等金融、电力行业头部客户中实现出货，验证了其在高安全要求场景中的重要性。

（4）支撑数据中心向新一代架构演进

数据中心正在“以 CPU 为中心”向“CPU（通用计算）+GPU（加速计算）+DPU（数据处理）”三大主力芯片协同的新型架构演进，DPU 作为该架构转型的核心基础设施，使云服务商能够在不影响现有业务的前提下，实现数据中心网络、存储和安全架构的平滑升级。公司云霄 DPU 通过支持 Virtio-net/Virtio-blk 虚拟化硬件加速，使数据中心从传统网卡向 DPU 迁移更加平滑，业务无感知迁移，降低了云服务商的架构升级风险和成本。

（三）DPU 在 AI 智算场景中的具体功能、必要性及重要性

1、AI 智算场景概述

AI 智算场景是 DPU 产品增长最快的下游应用领域之一。随着大模型参数规模从数十亿向万亿级演进，AI 训练和推理对算力规模的需求呈指数级增长，单一 GPU 服务器的算力已远不能满足需求，万卡乃至十万卡 GPU 集群成为 AI 大模型训练与推理的标配基础设施。在超大规模 GPU 集群中，GPU 之间的数据通信成为制约计算效率的关键瓶颈——若网络无法以足够低的延迟和足够高的吞吐完成 GPU 间的数据传输，GPU 将处于等待状态，造成昂贵的 GPU 算力资源的闲置浪费。

在该背景下，DPU 在 AI 智算场景中的角色从云计算场景的“基础设施卸载”升级为“算力网络核心枢纽”。全功能 DPU 继续承担 CPU 网络、存储等基础设施任务卸载，同时 AI DPU 为 GPU 集群提供高带宽、低延迟、零丢包的网络互联支撑，直接提升 AI 智算集群线性加速比和整体计算效率。

2、AI 智算场景中 DPU 的具体功能

（1）高性能 RDMA 网络与 GPU Direct 数据传输

通过 DPU 支持 RoCEv2 协议实现高性能 RDMA 网络，使 GPU 能够跨服务器直接读写对端 GPU 内存，无需经过 CPU 和操作系统复杂协议，大幅降低数据传输延迟与 CPU 开销，为数据并行和模型并行训练与推理提供高性能通信通道。同时，DPU 可在 GPU 与存储系统之间建立直接数据传输路径，绕过 CPU 内存中转实现训练与推理数据的零拷贝传输，对 AI 训练与推理中图片、视频、文本等大规模数据的批量传输具有关键加速作用，避免 CPU 因中断和数据复制承受过重负担。

（2）大规模集群网络拥塞控制

在大规模 GPU 集群中，大量 GPU 同时进行参数同步和梯度聚合会产生网络拥塞，若拥塞控制不当将导致丢包重传，严重降低训练与推理效率。公司风驰系列 AI DPU 针对智算网络场景优化了拥塞控制机制：支持自适应路由，根据网络实时负载动态选择最优传输路径，避免热点链路过载，提升集群网络的整体吞吐量和均衡性；在此基础上，DPU 支持基于优先级的流量控制（PFC）和显式拥塞通知（ECN），构建高性能网络基础设施，保障 AI 智算数据传输的零丢包和低延迟。

（3）AI 训练与推理数据存储加速

AI 大模型训练过程中，需要频繁加载训练数据和写入检查点及训练日志，推理任务启动时也需加载大模型参数并对 KV-Cache 进行持久化。DPU 通过 NVMe-oF 硬件加速将远程 NVMe 存储经高速网络虚拟化为 GPU 服务器的本地

存储，满足大规模训练数据加载与推理模型加载的高吞吐带宽需求；同时依托高吞吐网络连接保障训练过程中检查点的持续高速写入，避免因 I/O 瓶颈导致训练中断风险；此外，DPU 的高吞吐、低延迟存储访问能力支持同一服务器内多个 GPU 高效访问相同模型参数或训练数据集，减少数据冗余传输。

（4）智算网络协议卸载与 CPU 减负

在 AI 智算场景中，除 GPU 间通信外，集群管理、任务调度、存储、安全等基础设施服务仍需消耗 CPU 算力。公司云霄系列全功能 DPU 将这些基础设施任务从 Host CPU 卸载至 DPU，使 CPU 可以全力配合 GPU 进行训练与推理任务的调度和管理。在智算集群中，每台 GPU 服务器配置一颗全功能 DPU 即可释放数十个 CPU 核心，整个集群累计释放的 CPU 算力可用于支撑更大规模的训练任务编排和集群管理。在 AI 推理场景中，CPU 资源被 Agent 生命周期管理、多 Agent 调度、沙盒隔离、RAG 检索、Tool call 路由等消耗，CPU 需要将网络协议、存储 I/O、KV-Cache 搬运等基础任务卸载给 DPU，以保证 CPU 计算资源配置到更关键的编排调度、支撑 GPU 满负荷运行任务。AI 推理场景下，DPU 的价值因 CPU 资源紧张而愈发重要。

公司 DPU 支持全调度以太网标准等，已在客户腾讯、中国移动商用部署，验证了公司 DPU 产品在智算中心场景中的适用性和商用价值。

3、DPU 对 AI 智算场景的必要性

（1）GPU 集群网络互联的不可替代性

AI 大模型训练与推理的核心瓶颈已从单卡算力转向集群网络互联效率。在数据并行计算中，每个 GPU 节点在每次迭代后需同步梯度数据；在张量并行和流水线并行推理中，GPU 间同样需要高频传输激活值和中间张量。若网络延迟过高或带宽不足，GPU 将处于等待状态，导致集群线性加速比下降（AI 集群训练线性加速比是衡量 GPU 集群训练效率的核心指标，定义为集群实际训练吞吐量与单卡训练吞吐量之比）。根据 Haiquan Wang 等学者《Hiding Communication

《Cost in Distributed LLM Training via Micro-batch Co-execution》一文的研究成果，以千卡集群训练千亿参数大模型为例，每次梯度同步需传输数百 GB 数据，若网络吞吐不足，通信时间可能占总训练时间的 30%-50%。DPU 通过 RDMA 和 GPU Direct 技术提供高吞吐、低延迟的 GPU 间数据传输通道，是保障 GPU 集群训练与推理效率的不可替代的基础设施组件。

（2）高可靠网络传输的刚性需求

AI 训练与推理对网络丢包极度敏感，在大规模集群中丢包对集群整体效率的影响被指数级放大。高可靠传输能力成为大规模智算网络的刚性需求。公司 AI DPU 针对智算网络场景优化的拥塞控制算法、选择性重传等可靠性传输机制，能够大幅降低丢包对训练和推理效率的影响，是支撑大规模 GPU 集群稳定运行的必要技术保障。

（3）CPU 算力释放的规模效应

在 AI 智算中心中，每台 GPU 服务器通常配置多颗高性能 GPU，但 CPU 核心数量有限。无论在训练还是推理场景中，若 CPU 需同时处理网络协议栈、存储虚拟化、安全业务等基础设施任务，将挤占本应用于 GPU 训练任务调度和管理的 CPU 算力。在万卡集群（约 1,250 台 8 颗 GPU 搭配 2 颗 64 核 CPU 的智算服务器）规模下，若每台服务器因基础设施开销损失 30% 的 CPU 算力，整个集群将损失约 4.8 万个 CPU 核的等效算力（ $1,250 \times 2 \times 64 \times 30\% = 4.8$ 万）。在推理场景中，高并发请求的负载均衡、KV-Cache 管理和动态批处理调度同样需要 CPU 参与，DPU 卸载基础设施任务后可释放 CPU 算力用于推理调度优化。DPU 通过卸载这些基础设施任务，在集群规模下产生显著的算力释放效应，是提升 AI 智算中心整体效率的必要手段。

（4）检查点写入与推理数据加载的数据安全需求

AI 大模型训练通常持续数天至数周，训练过程中需定期保存检查点以防止训练中断造成的从头开始。检查点数据量巨大，写入过程中若 I/O 性能不足将导

致训练暂停。在推理场景中，模型参数的快速加载决定了推理服务的冷启动时间，KV-Cache 的持久化和热迁移也依赖高速存储 I/O。DPU 通过 NVMe-oF 和高速 RDMA 网络加速检查点写入与模型加载，保障训练过程的数据安全性、连续性和推理服务的快速恢复能力，该功能对于降低 AI 训练的试错成本、保障训练可靠性以及提升推理服务可用性具有必要性。

4、DPU 对 AI 智算场景的重要性

（1）提升 AI 集群训练与推理效率

DPU 通过高性能 RDMA 网络和 GPU Direct 等技术，降低 GPU 间通信延迟、提升数据同步吞吐，直接提升集群线性加速比。在推理场景中，张量并行和流水线并行推理同样依赖 GPU 间高频通信，DPU 的低延迟传输能力可降低推理首 Token 延迟并提升吞吐量。对于动辄投入数亿元的 AI 大模型训练项目和高并发推理服务，DPU 带来的效率提升具有显著的经济价值。

（2）支撑超大规模 AI 集群扩展

随着 AI 大模型参数规模持续增长，GPU 集群规模从千卡向万卡、十万卡扩展，集群网络的管理复杂度和拥塞风险急剧上升。公司在研的 800Gbps/1.6Tbps AI DPU 将提供更高带宽和更先进的拥塞控制能力，支撑更大规模的 GPU 集群互联。DPU 作为智算网络的核心组件，其性能和功能直接决定了 AI 智算集群的可扩展规模上限，对于 AI 产业的长期发展具有重要战略意义。

（3）保障 AI 训练与推理数据安全与训练可靠性

DPU 提供的硬件级安全隔离和高速检查点写入能力，为 AI 训练与推理过程中的数据安全和可靠性提供底层保障。DPU 的安全隔离能力在金融、政务等对数据安全要求极高的 AI 应用场景中，可防止云端训练数据在多租户智算中心中的泄露风险，并可在推理场景中防止模型参数和用户请求的跨租户泄露；高速检查点写入功能则保障训练中断后可快速恢复，降低因硬件故障或网络中断导致的训练成本损失。

（4）降低 AI 智算中心建设成本

DPU 通过提升 GPU 利用率和集群训练效率，在相同训练任务需求下可减少所需的 GPU 数量，直接降低 AI 智算中心的硬件采购成本。在推理场景中，DPU 卸载网络和存储任务后可提升单服务器可部署的推理并发数，降低单位推理请求的硬件成本。此外，DPU 释放的 CPU 算力可减少服务器 CPU 配置要求，进一步降低硬件成本。

（四）DPU 在其他下游场景中的具体功能、必要性及重要性

1、存储场景

在分布式存储场景中，DPU 作为存储控制器处理数据移动、加密和 RDMA 任务，可替代传统网卡实现更高性能的存储网络连接。公司 DPU 支持 NVMe-oF、Virtio-blk 等存储协议的硬件加速，支持高吞吐存储访问，能够满足全闪存数据中心和 AI 驱动的存储性能需求。随着存储全面转向闪存技术，存储网络带宽需求从 10G/25G 向 100G/400G 及更高带宽快速演进，传统以 CPU 处理存储协议栈的方案已无法跟上闪存存储的性能增长。DPU 通过硬件加速存储协议处理，使存储网络性能与闪存存储性能相匹配，是构建高性能分布式存储系统的必要组件。

2、网络安全场景

在网络安全场景中，DPU 提供独立的安全服务域，支持加密加速、防火墙规则卸载、入侵检测与防御等功能。公司 DPU 通过硬件级安全隔离和加密加速，为数据中心提供零信任安全架构的基础设施支撑。随着网络攻击手段日趋复杂和监管合规要求不断提升，传统基于软件的安全方案在处理能力和隔离性方面面临瓶颈。DPU 通过独立的安全控制平面和硬件加密引擎，即使在 Host CPU 被入侵的情况下仍能保障安全策略的执行，是构建新一代数据中心安全体系的重要技术基础。公司云霄 DPU 已在南方电网、工商银行、招商银行等对网络安全要求极高的行业客户中实现出货，验证了其在网络安全场景中的实用价值。

3、通信与边缘计算场景

在 5G 通信和边缘计算场景中，DPU 可用于 5G RAN（无线接入网）的精密定时加速、电信云网络功能虚拟化（NFV）卸载等应用。公司 DPU 支持流量整形、电信云加速等功能，可满足通信运营商对低延迟、高吞吐网络处理的需求。5G 网络和边缘计算对网络处理的实时性和确定性延迟有严格要求，传统 CPU 方案难以在保证业务应用算力的同时满足通信协议处理的实时性需求。DPU 通过专用硬件加速通信协议处理，释放 CPU 算力用于业务应用，是 5G 和边缘计算基础设施的重要组件。公司已与中国移动、中国联通等国内头部电信运营商开展合作，验证了 DPU 产品在通信场景中的适用性。

综上所述，公司已有及在研 DPU 产品在云计算、AI 智算等典型下游场景中具有不可替代性及重要战略价值，具体体现在：

第一，在云计算场景中，DPU 通过全面支持弹性裸金属、网络虚拟化、存储虚拟化、安全隔离和资源管理功能，消除占 CPU 算力约 30%的“数据中心税”，是多租户安全隔离、虚拟化资源统一管理和弹性资源调度的架构性基础。公司云霄 1.0 系列 DPU 已在行业头部客户腾讯云、移动云实现大规模部署，验证了 DPU 对云计算场景的重要性。

第二，在 AI 智算场景中，DPU 通过高性能 RDMA 网络、GPU Direct 数据传输、无损网络拥塞控制等功能，直接决定 GPU 集群的训练线性加速比和可扩展规模，是大规模 GPU 集群不可或缺的网络核心枢纽。公司风驰 2.0 AI DPU 已流片并加速推向市场，在研的 800Gbps/1.6Tbps AI DPU 将进一步支撑超大规模 AI 集群扩展，契合 AI 大模型训练与推理对高性能智算网络的迫切需求。

第三，在存储、网络安全、通信与边缘计算等场景中，DPU 通过硬件加速存储协议处理、提供零信任安全架构、支撑 5G 通信网络功能卸载，满足各场景对高性能、低延迟、高安全性的专业化需求，公司产品已在南方电网、中国移动、工商银行等头部客户中实现多场景商用。

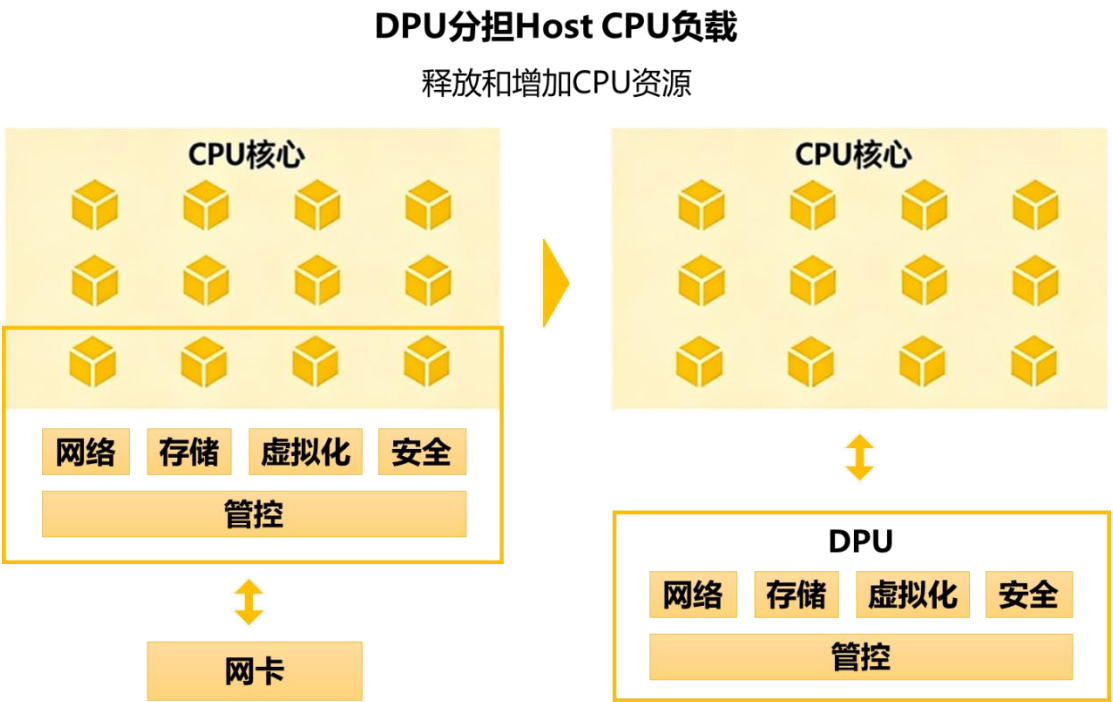
第四，从行业发展趋势看，根据弗若斯特沙利文数据，中国 DPU 市场规模预计从 2025 年的 497.58 亿元增长至 2030 年的 1,290.91 亿元。在国产替代背景下，公司作为全功能 DPU 国产独立厂商第一名，与国际龙头企业先进产品的代差已缩小至一代，其 DPU 产品对各类下游使用场景的必要性及重要性将随市场规模增长和技术迭代持续提升。

二、结合推理、AI Agent、AI 编程等典型场景下，可由 DPU 卸载的 CPU 任务占比及 DPU 卸载效率、提高集群综合算力效率比例等，分析 DPU 提升数据中心整体性能、降低集群算力综合成本的具体体现，客户采用相关方案的成本收益情况对发行人产品开拓的影响

（一）DPU 提升数据中心整体性能的具体体现

1、可由 DPU 卸载的 CPU 任务占比及 DPU 卸载效率

随着 AI 等前沿产业快速发展，计算需求呈爆炸式增长，云化基础设施平台基于 CPU 完成前端网络数据转发的传统模式出现瓶颈，CPU 串行计算难以在高并发网络数据转发工作中发挥最大计算能力。正如 CPU 相比 GPU 在处理简单重复的高并发任务时效率较低，CPU 亦不擅长网络虚拟化、云存储、低时延传输、安全软件部署以及查表、转发等以数据为中心、I/O 密集的功能。DPU 针对 I/O 密集型任务进行硬件级优化，可将网络、存储、虚拟化、安全、管控等功能从 CPU 卸载至 DPU 上进行处理，能够有效降低 CPU 负载、提高数据中心运行效率。



在 AI 场景中，CPU 承担的任务可按功能划分为网络通信处理、存储 I/O 管理、数据预处理与序列化、安全隔离与加密、虚拟化与资源调度等类别。

在 AI 推理场景中，除上述数据中心基础设施开销外，推理引擎本身的 CPU 调度开销同样显著。根据 UC Berkeley WukLab 的研究，vLLM 推理框架的 CPU 调度开销可占总推理时间的 50% 以上，主要来源于输入张量构建、Token 解码和元数据准备等任务。

AI 场景下，可由 DPU 卸载的 CPU 任务及 DPU 卸载效率情况如下：

可卸载任务	预估开销占比	功能描述	DPU 卸载效率
网络通信处理	约 20%~35%	API 请求路由、TCP/IP 协议栈、RDMA 数据传输、推理结果序列化与网络发送、Agent 工具调用的网络请求路由、多 Agent 间通信、外部 API 交互的数据传输	高（CPU 开销降低 80%）
存储 I/O 管理	N/A	KV-Cache 读写及跨节点共享、模型参数加载、训练数据预取、NVMe-oF 协议处理	高（IOPS 提升 2 倍、读写延迟降低 30%）
安全隔离与加密	N/A	数据传输加密、多租户隔离、	高（硬件加密引

可卸载任务	预估开销占比	功能描述	DPU 卸载效率
		访问控制策略执行	擎性能提升 5 倍)
虚拟化与资源调度	虚拟化约 20%~30% 调度约 50% (含 tensor 预处理、 detokenization、元数 据准备)	vSwitch 转发、资源调度、QoS、 弹性扩缩容	高 (vSwitch 延迟 降低 90%、支持 热迁移停顿缩短 90%)

数据来源：英伟达、中国移动、Wuklab、Revankar、Kanev 等研究、Starsurge、CSDN

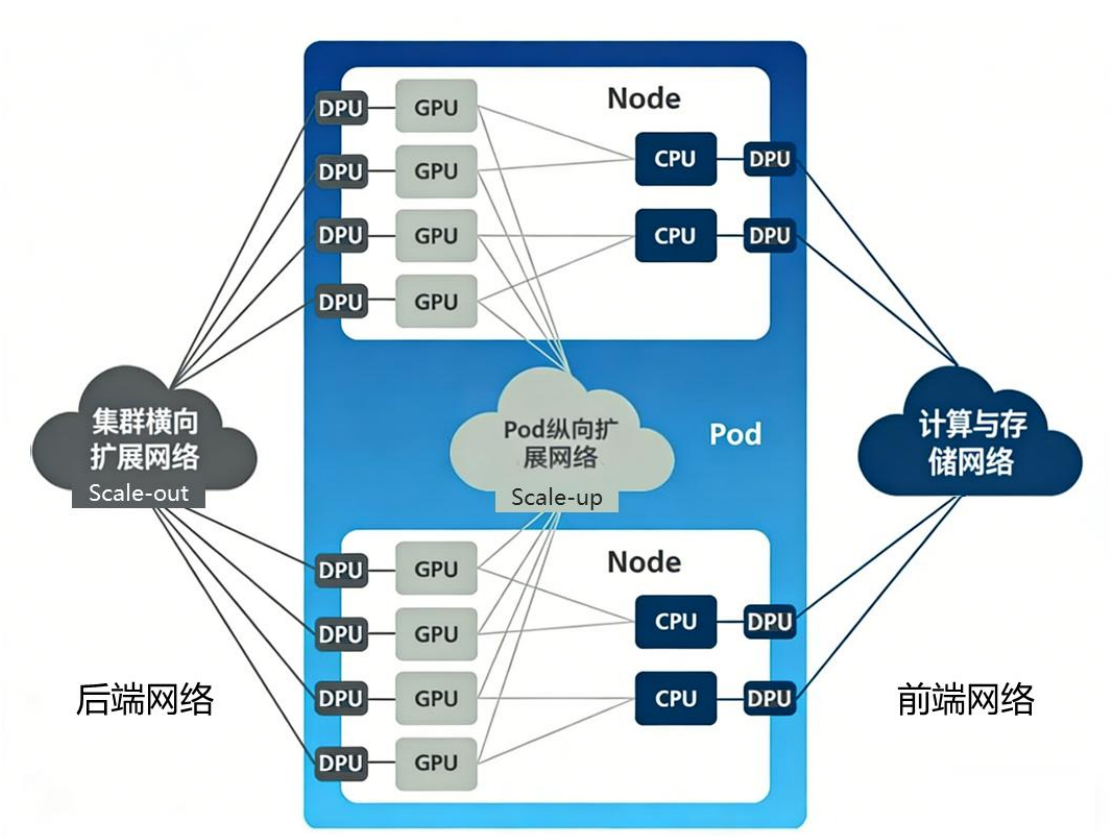
学术研究成果和行业应用数据均表明，网络通信、存储 I/O、安全加密及虚拟化调度四类基础设施任务可由 DPU 实现高效卸载。

该等 DPU 卸载价值已在公司客户规模化商用中得到充分验证：依托于公司云霄 DPU 的加持，腾讯第九代云服务器实例提供 6750 万 PPS 的吞吐能力，网络延迟降低 35%；存储性能实现跨越式升级，吞吐性能比第八代服务器提升最高达 200%，磁盘密度提升最高达 100%，数据加解密性能提升最高达 100%；虚拟化适配方面，芯片 Memory ECC 校验覆盖率达 100%，实例不可恢复故障率趋近于零。

2、提高集群综合算力效率比例

随着 AI 模型参数持续增长，单卡算力和显存受限，在全球 AI 算力竞赛下，迫使训练集群规模持续扩大，网络性能成为瓶颈。AI 大模型以 GPU 集群分布式训练为基础，但随着算力集群规模扩大，GPU 集群内部产生大规模互联需求，训练过程中各集群计算节点间需要频繁地进行参数同步。当 GPU 数量、性能相同时，集群算力受制于网络耗时。在智算中心架构下，同时采用 Scale-out 和 Scale-up 两个独立的网络基础设施协同解决上述问题成为当下主流趋势，Scale-up 高速网络解决同一服务器或服务器机柜内小规模 GPU 高速互联，Scale-out 高速网络解决不同 AI 服务器或机柜之间的大规模 GPU 高速互联，实现超大规模智算中心各节点 GPU 之间的通信。带有 RDMA 功能的 DPU，有效支持高吞吐、低时延网络通信，日益成为 AI 服务器 Scale-out 后端网络的标配。在传统的 TCP/IP 通信方式下，收发数据需要 CPU 多次参与、速度较慢。而 RDMA 技术可以绕过操作系统 CPU 内核，让一台服务器可以直接访问另外一台服务器的内存，时延

性能显著改善。借助 DPU 实现零拷贝、内核旁路、无 CPU 干预效果及针对 AI 大规模集群优化的可编程拥塞控制，AI 应用直接与远端服务器进行数据交互和内存读写，消除了 GPU 跨节点通信网络带宽和时延瓶颈，提高了资源利用率和训练效率。

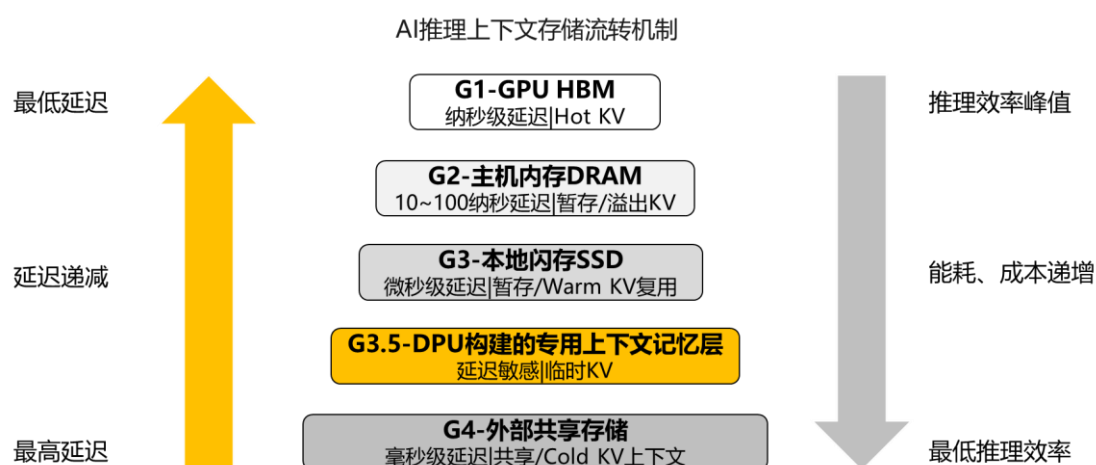


在 AI 推理场景中，CPU 主要承担数据预处理、多模态数据适配、集群硬件调度、推理数据传输与 KV-Cache 管理等辅助任务。在推理集群引入 DPU 能够有效节约 CPU 在前述网络通信等基础设施任务上的开销，从而令 CPU 专注于核心计算任务，同时提升 GPU 利用率。GPU 利用率提升的核心机制在于：DPU 卸载网络通信和数据预处理任务后，GPU 无需等待 CPU 完成数据传输即可开始计算，减少了“GPU 等数据”的空闲周期。同时，DPU 通过 RDMA 技术实现 GPU Direct 数据传输（绕过 CPU 内存拷贝），进一步缩短数据供给延迟。

DPU 在提升集群综合算力效率方面的综合价值已获得多家厂商及实测数据验证。根据英伟达与 F5 的 2026 年联合测试数据，在经 Tolly Group 验证测试中，

通过英伟达 BlueField-3 DPU 卸载和加速，实现了 CPU 开销降低 80%（将传统 CPU 负载均衡所需的 12 核心降至 2 个核心），Token 吞吐量提升 40%、首令牌时间（TTFT）缩短 61%，整体请求延迟降低 34%；根据网络方案商 Starsurge 公布的数据，配置 DPU 后，AI 集群的 CPU 网络通信开销可降低高达 80%，GPU 利用率最高可提升 30%，任务完成时间缩短 20%-40%。根据中国电信天翼云数据，某科研机构实验显示，引入 DPU 后，千卡级 AI 集群的 GPU 利用率从 70%提升至 90%，训练效率提高 40%。

此外，在 KV-Cache 管理方面，DPU 的作用日益重要。随着大模型上下文窗口持续增长，GPU 显存容量成为推理成本的重要瓶颈。英伟达于 2026 年 GTC 发布的 Vera Rubin 系统中，BlueField-4 系列 DPU 承担 KV-Cache 管理与硬件加速任务，在 GPU 高速 HBM 与外部存储之间构建“记忆层”，实现推理吞吐量提升最高达 5 倍。



根据华为公布的数据，华为 OceanDisk 1800 智能盘框通过 DPU 硬件卸载 KV 语义，在 Qwen-32B 模型实测中，KV-Cache 命中率从 15%提升至 90%，首 Token 延迟从 7.43 秒降至 1.31 秒，端到端推理成本降低 30%。

3、AI Agent、AI 编程等进阶推理场景中 DPU 的效用

在 AI Agent 场景中，由于 CPU 负载显著高于纯推理场景，DPU 卸载的价值更大。DPU 将网络通信、存储 I/O 和安全隔离等基础设施任务卸载后，CPU 资

源可专注于 Agent 任务规划、逻辑决策和代码执行等核心业务逻辑，直接提升 Agent 并发处理能力和响应速度。

AI 编程场景（如 GitHub Copilot、Cursor、Claude Code 等）本质上是 AI 推理与 Agent 能力的综合应用，其 CPU 开销兼具推理场景的数据预处理特征和 Agent 场景的多步骤交互特征。在 AI 编程工作流中，CPU 需承担代码库索引与检索、代码 Tokenization、编译环境管理、代码执行沙箱维护、版本控制交互、多文件上下文管理等任务。

AI 编程场景对存储 I/O 的需求尤为突出，大型代码库的索引和检索涉及大量随机读写操作，传统架构下 CPU 需深度参与存储协议处理。根据英伟达数据，在架构中引入 BlueField-3 DPU 将大量同时运营的虚拟机在线迁移造成的网络中断时间缩短 90%；根据 CSDN 技术博客文章，DPU 通过 SPDK 硬件卸载存储虚拟化 and 数据校验任务，可将块存储 IOPS 提升 2 倍、读写延迟降低 30%，显著加速代码库检索和编译流程。

（二）DPU 降低算力综合成本的具体体现

DPU 降低集群算力综合成本主要体现在硬件采购成本、能耗成本、运维及故障成本等。根据中国电信天翼云数据，以 1,000 台服务器规模为例测算使用 DPU 的成本收益情况如下：

1) 硬件方面：相同业务负载下所需服务器数量减少 30%，网络 PPS 提升 100%，存储 IOPS 提升 200%，硬件采购成本降低约 1.2 亿元。

2) 能耗方面：部署 DPU 后，网络、存储和安全任务被卸载至硬件，CPU 负载显著降低，单台服务器功耗从 800W 降至 500W，年电费减少约 400 万元，年碳排放量减少约 2,000 吨。

3) 运维方面：DPU 通过硬件级故障隔离和自动化运维，显著降低运维成本。网络故障平均修复时间从 4 小时缩短至 30 分钟，业务中断损失减少 90%，运维团队减少 50%，年人力成本节约 800 万元，运维效率提升约 3 倍。

（三）客户采用相关方案的成本收益情况对发行人产品开拓的影响

客户采用 DPU 提升了数据中心整体性能、降低了集群算力综合成本，具体体现如上所示。所对应的主要收益体现如下：

一方面，云服务商可以提升可售卖算力比例、降低 TCO。DPU 将网络、存储和安全任务从主机 CPU 卸载后，租户购买的 vCPU 不再被网络开销侵蚀，可售卖资源增加。以双路 96 核 CPU 服务器为例，卸载 30%基础设施开销后，可售卖算力提升约 43% ($((192 \times 30\%) / (192 \times 70\%)) = 42.86\%$)，直接转化为云服务收入增长。

另一方面，对于 AI 智算中心，DPU 提升了 GPU 利用率、进而降低单位 Token 推理成本。根据英伟达公布的数据，BlueField-4 STX 参考架构与传统 CPU 架构相比，Token 吞吐量提升最高达 5 倍。以千卡集群为例，若 DPU 使有效吞吐量提升 15%，所需 GPU 数量可减少约 13% ($15\% / (100\% + 15\%) = 13.04\%$)，按照英伟达 H100 GPU 约 18-25 万元/颗计算，可节省 GPU 采购成本约 2,340 万元-3,250 万元 ($1,000 \times 13\% \times 18 = 2,340$ 万元，同理 25 万元单价时为 3,250 万元)。

公司云霄 DPU 已在头部云服务商腾讯、头部电信运营商中国移动等客户实现规模化部署，截至 2026 年 6 月末，云霄 DPU 累计销量超 11 万颗，表明公司客户对 DPU 方案具有高度认可和持续的采购需求。客户采用相关方案的成本主要为 DPU 硬件及配套软件采购成本。客户采用 DPU 方案实现的显著收益，对公司产品开拓产生以下积极影响：

1、DPU 提升性能的优异表现使得市场需求持续扩大，公司作为行业领军企业将受益于此

根据弗若斯特沙利文数据，全球 DPU 市场规模已从 2021 年的 649.93 亿元增长至 2025 年的 1,964.91 亿元，预计到 2030 年将进一步增长至 4,362.39 亿元；中国市场预计在 2030 年达到 1,290.91 亿元，成为 AI 基础设施中增长最快的细分领域之一。客户对 DPU 降低 TCO 的逐步重视和实际收益验证，将推动 DPU 从

“可选项”向“必选项”转变，持续扩大公司的市场空间。

公司是目前国内少数实现高性能 DPU 大规模量产与商用落地的独立厂商之一，公司 400Gbps 全功能 DPU 产品对标英伟达主力在售的 BlueField-3 系列 DPU。客户采用 DPU 方案的实际成本收益验证了公司产品的技术成熟度和商业价值，为公司持续开拓市场提供了有力支撑。

2、公司产品矩阵随客户需求场景增加而持续拓展、客户结构持续优化

基于客户在不同场景下的成本收益需求，公司持续拓展产品矩阵：首款云霄 1.0 系列 400Gbps 全功能 DPU 已在腾讯、中国移动等头部客户实现规模化商用；第二款面向 AI 网络场景优化的风驰 2.0 系列 400Gbps AI DPU 已成功流片、正加速推向市场；新一代 800Gbps 全功能 DPU 和 800Gbps/1.6Tbps AI DPU 正在研发中。公司本次 IPO 拟募资 30.35 亿元，分别投入新一代全功能 DPU 研发及产业化项目、新一代 AI DPU 研发及产业化项目和下一代高性能 DPU 核心技术研发项目，以持续满足客户在 AI 推理、Agent 和编程等场景中的成本优化需求。

公司产品已在头部云服务商腾讯完成大规模部署，同时已成为中国移动、中国联通等国内头部电信运营商、国产知名 AI 大模型厂商的重要供应商或战略合作伙伴，并在金融、电力等行业头部客户完成项目落地。随着客户对 DPU 成本收益认知的深化，公司有望持续拓展行业客户覆盖范围。

3、Agent 时代 DPU 需求加速释放，公司 DPU 产品迎结构性增长机遇

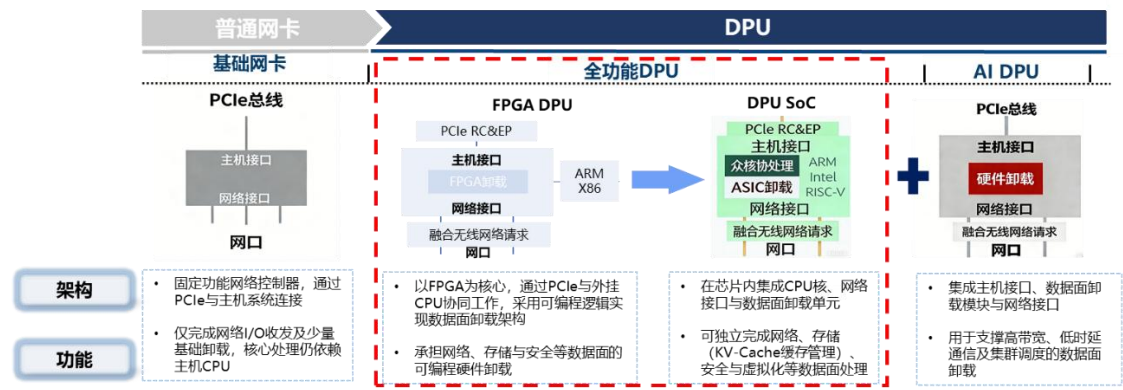
随着 AI 产业从训练主导转向推理与 Agent 驱动，DPU 正从过去的数据中心可选项变为 AI 基础设施的关键组件。在 Agent 时代，AI 系统从训练转向高频推理与持续运行，对 CPU 的算力需求持续上升，GPU 利用率亦越来越依赖网络效率，DPU 承担大规模集群间的数据调度与网络卸载任务，直接影响 CPU 和 GPU 利用率。英伟达于 2026 年 GTC 发布的 Vera Rubin 系统中，BlueField-4 系列全功能 DPU 和 ConnectX-9 系列 AI DPU 已成为与 CPU、GPU、NVLink 交换机并列的“第四根支柱”，进一步验证了 DPU 在 AI 基础设施中的战略地位。这一趋势

为公司 DPU 产品的市场开拓提供了强有力的行业动能。

三、全功能 DPU 与 AI DPU 的功能及适用场景；结合下游 AI 行业发展趋势及不同 AI 应用领域对两类产品需求情况、全功能 DPU 及 AI DPU 产品使用数量与 CPU、GPU、服务器、节点数量之间的配比关系，分析产品销量增长潜力，下游对发行人产品需求变化趋势，发行人研发方向与下游需求的匹配性

（一）全功能 DPU 与 AI DPU 的功能及适用场景

数据处理器（Data Processing Unit，简称“DPU”）芯片是一种以数据处理为核心的通用处理单元，是对计算资源相关的低时延网络、弹性存储、安全和管控的协同处理平台。DPU 是由传统网卡（NIC）逐步演进而来。网卡经历了由基础功能卸载向高算力、高可编程数据处理的演进过程，形态由普通网卡、具备简单卸载能力网卡逐步升级为具备高性能、通用可编程卸载能力与低时延 RDMA 能力的 DPU 网卡；同时在面向云原生架构、全场景数据处理需求的全功能 DPU 基础上，结合 AI 智算集群后端网络优化的特定需求，演进为支持超大规模 GPU 高速互联的 AI DPU。



资料来源：弗若斯特沙利文

全功能 DPU 在大规模多租户的云网络架构中扮演重要角色，主要用于支撑弹性网络、云存储、虚拟化、安全以及数据管控等基础设施功能的运行，通过软硬件协同实现对复杂数据与网络处理的全场景应用，是数据中心、计算集群的重要数据面和控制面组件。此外，英伟达在新一代全功能 DPU BlueField-4 增强了 KV-Cache 缓存管理与硬件加速能力，赋能推理上下文记忆存储平台，同步引领

全功能 DPU 在 AI 智算场景中的应用迈向新前沿。

相比之下，AI DPU 侧重面向 AI 智算集群的后端网络场景，围绕模型训练与推理过程中产生的庞大上下文数据流，承担集群协同、高带宽低时延 RDMA 传输、可编程拥塞控制与数据调度等关键功能，为大规模 AI 算力集群的高效运行提供支撑。

全功能 DPU 与 AI DPU 的主要功能及适用场景如下：

DPU 类别	主要产品系列	场景需求	具体功能
全功能 DPU	云霄系列	云计算前端网络；多租户虚拟化平台；AI 智算；AI 存储网络	设备虚拟化与网络虚拟化，弹性网络、低时延 RDMA、存储、安全、管控的卸载及加速，可编程数据平面硬件卸载，KV-Cache 缓存管理及加速
AI DPU	风驰系列	AI 智算；云计算后端网络；高性能训练/推理集群	大规模 RDMA 网络通信，多路径负载均衡，可编程拥塞控制，快速重传能力，存储卸载

1、云计算场景中 DPU 的功能

公司云霄系列全功能 DPU 在云计算场景中实现管控、存储、网络和安全的全面卸载，具体功能如下：

（1）网络虚拟化与协议卸载

虚拟交换机（OVS）卸载：云霄 DPU 将 OVS 网络业务从 Host CPU 卸载至 DPU，实现数据包解析、匹配和转发的 DPU 加速，显著降低 OVS 网络业务对 CPU 算力的消耗。

Overlay 协议支持：全面支持 VXLAN、GRE、GENEVE 等网络协议的硬件卸载，实现虚拟网络与物理网络的高效隔离与互通，支持多租户网络环境下的大规模网络虚拟化部署。

Virtio 虚拟化加速：通过硬件加速 Virtio 虚拟化框架，在提供更高性能和更低时延的同时，使数据中心从传统网卡向 DPU 的迁移更加平滑，保障业务连续性。

弹性网络：支持弹性网卡功能，按需动态分配网络带宽和虚拟网络接口资源，满足云租户对网络资源的弹性伸缩需求。

（2）存储虚拟化、卸载及加速

NVMe-oF 存储协议卸载：通过硬件加速 NVMe over Fabrics 协议，将网络存储虚拟化为本地高性能 NVMe 设备，使远程存储性能接近本地存储水平，显著提升云存储服务的吞吐量和延迟表现。

Virtio-blk 存储虚拟化加速：通过硬件加速 Virtio-blk 虚拟化框架，实现块设备虚拟化的性能提升，支持云盘启动功能，使虚拟机和容器从远程存储启动时获得接近本地存储的性能体验。

弹性存储：支持弹性存储功能，按需动态分配存储资源，并支持针对不同弹性存储实施差异化策略，满足云租户对存储资源的灵活配置需求。

RDMA 存储访问：支持 RoCE v1&2 协议，通过 RDMA 技术实现高吞吐、低延迟的远程存储访问，支持高性能存储网络连接。

（3）虚拟化资源统一管理

弹性裸金属、虚拟机和容器云服务“三合一”统一架构：云霄 DPU 为数据中心提供“三合一”统一架构和运维解决方案，全面解决弹性裸金属、虚拟机和容器三种不同场景的资源虚拟化和统一管理，实现不同类型业务上云及无感知迁移。

热插拔与热迁移：支持热插拔、热迁移，可实现在客户业务不中断的情况下进行资源迁移，保证在宕机时业务快速恢复，支持业务的弹性扩展和无感迁移。

热升级：支持在不中断业务的情况下对 DPU 固件进行热升级和迭代，确保基础设施持续优化不影响业务连续性。

（4）安全隔离与零信任架构

安全服务域隔离：云霄 DPU 为每台服务器提供独立的安全服务域，通过与服务器的应用域隔离实现零信任架构（Zero Trust，指一种以资源保护为核心的网络安全理念，认为对资源的访问，无论主体和资源是否可信，主体和资源之间的信任关系都需要从零开始，通过持续环境感知与动态信任评估进行构建，从而实施访问控制），即使 Host CPU 被入侵，DPU 上的安全控制平面仍可独立运行，防止攻击向整个数据中心扩散。

加密加速：支持 IPSEC 和 TLS 的硬件加密加速，执行数据传输加密，在不影响其他加速功能的前提下保障数据传输安全。

安全生命周期管理：支持信任根、安全启动、安全固件升级及经过身份验证的容器和应用程序生命周期管理，构建从芯片到应用的完整安全链条。

服务质量(QoS)保障：为不同租户和不同业务流提供高标准服务质量(QoS)保证，确保关键业务的网络和存储性能不受其他租户影响，保障多租户环境下的服务等级协议（SLA）承诺。

2、AI 智算场景中 DPU 的功能

（1）高性能 RDMA 网络与 GPU Direct 数据传输

通过 DPU 支持 RoCEv2 协议实现高性能 RDMA 网络，使 GPU 能够跨服务器直接读写对端 GPU 内存，无需经过 CPU 和操作系统复杂协议，大幅降低数据传输延迟与 CPU 开销，为数据并行和模型并行训练与推理提供高性能通信通道。同时，DPU 可在 GPU 与存储系统之间建立直接数据传输路径，绕过 CPU 内存中转实现训练与推理数据的零拷贝传输，对 AI 训练与推理中图片、视频、文本等大规模数据的批量传输具有关键加速作用，避免 CPU 因中断和数据复制承受

过重负担。

（2）大规模集群网络拥塞控制

在大规模 GPU 集群中，大量 GPU 同时进行参数同步和梯度聚合会产生网络拥塞，若拥塞控制不当将导致丢包重传，严重降低训练与推理效率。公司风驰系列 AI DPU 针对智算网络场景优化了拥塞控制机制：支持自适应路由，根据网络实时负载动态选择最优传输路径，避免热点链路过载，提升集群网络的整体吞吐量和均衡性；在此基础上，DPU 支持基于优先级的流量控制（PFC）和显式拥塞通知（ECN），构建高性能网络基础设施，保障 AI 智算数据传输的零丢包和低延迟。

（3）AI 训练与推理数据存储加速

AI 大模型训练过程中，需要频繁加载训练数据和写入检查点及训练日志，推理任务启动时也需加载大模型参数并对 KV-Cache 进行持久化。DPU 通过 NVMe-oF 硬件加速将远程 NVMe 存储经高速网络虚拟化为 GPU 服务器的本地存储，满足大规模训练数据加载与推理模型加载的高吞吐带宽需求；同时依托高吞吐网络连接保障训练过程中检查点的持续高速写入，避免因 I/O 瓶颈导致训练中断风险；此外，DPU 的高吞吐、低延迟存储访问能力支持同一服务器内多个 GPU 高效访问相同模型参数或训练数据集，减少数据冗余传输。

（4）智算网络协议卸载与 CPU 减负

在 AI 智算场景中，除 GPU 间通信外，集群管理、任务调度、存储、安全等基础设施服务仍需消耗 CPU 算力。公司云霄系列全功能 DPU 将这些基础设施任务从 Host CPU 卸载至 DPU，使 CPU 可以全力配合 GPU 进行训练与推理任务的调度和管理。在智算集群中，每台 GPU 服务器配置一颗全功能 DPU 即可释放数十个 CPU 核心，整个集群累计释放的 CPU 算力可用于支撑更大规模的训练任务编排和集群管理。在 AI 推理场景中，CPU 资源被 Agent 生命周期管理、多 Agent 调度、沙盒隔离、RAG 检索、Tool call 路由等消耗，CPU 需要将网络协议、存

储 I/O、KV-Cache 搬运等基础任务卸载给 DPU，以保证 CPU 计算资源配置到更关键的编排调度、支撑 GPU 满负荷运行任务。AI 推理场景下，DPU 的价值因 CPU 资源紧张而愈发重要。

3、存储场景中 DPU 的功能

在分布式存储场景中，DPU 作为存储控制器处理数据移动、加密和 RDMA 任务，可替代传统网卡实现更高性能的存储网络连接。公司 DPU 支持 NVMe-oF、Virtio-blk 等存储协议的硬件加速，支持高吞吐存储访问，能够满足全闪存数据中心和 AI 驱动的存储性能需求。随着存储全面转向闪存技术，存储网络带宽需求从 10G/25G 向 100G/400G 及更高带宽快速演进，传统以 CPU 处理存储协议栈的方案已无法跟上闪存存储的性能增长。DPU 通过硬件加速存储协议处理，使存储网络性能与闪存存储性能相匹配，是构建高性能分布式存储系统的必要组件。

4、网络安全场景中 DPU 的功能

在网络安全场景中，DPU 提供独立的安全服务域，支持加密加速、防火墙规则卸载、入侵检测与防御等功能。公司 DPU 通过硬件级安全隔离和加密加速，为数据中心提供零信任安全架构的基础设施支撑。随着网络攻击手段日趋复杂和监管合规要求不断提升，传统基于软件的安全方案在处理能力和隔离性方面面临瓶颈。DPU 通过独立的安全控制平面和硬件加密引擎，即使在 Host CPU 被入侵的情况下仍能保障安全策略的执行，是构建新一代数据中心安全体系的重要技术基础。公司云霄 DPU 已在南方电网、工商银行、招商银行等对网络安全要求极高的行业客户中实现出货，验证了其在网络安全场景中的实用价值。

5、通信与边缘计算场景中 DPU 的功能

在 5G 通信和边缘计算场景中，DPU 可用于 5G RAN（无线接入网）的精密定时加速、电信云网络功能虚拟化（NFV）卸载等应用。公司 DPU 支持流量整形、电信云加速等功能，可满足通信运营商对低延迟、高吞吐网络处理的需求。5G 网络和边缘计算对网络处理的实时性和确定性延迟有严格要求，传统 CPU 方

案难以在保证业务应用算力的同时满足通信协议处理的实时性需求。DPU 通过专用硬件加速通信协议处理，释放 CPU 算力用于业务应用，是 5G 和边缘计算基础设施的重要组件。公司已与中国移动、中国联通等国内头部电信运营商开展合作，验证了 DPU 产品在通信场景中的适用性。

(二)结合下游 AI 行业发展趋势及不同 AI 应用领域对两类产品需求情况、全功能 DPU 及 AI DPU 产品使用数量与 CPU、GPU、服务器、节点数量之间的配比关系，分析产品销量增长潜力

1、下游 AI 行业发展趋势

2026 年以来，AI 产业正经历从“AI 训练”向“AI 推理”与“AI Agent”的结构性转变。德勤预计全球推理负载占 AI 算力比例 2026 年增长至约 2/3。随着 AI 技术范式从“对话生成”向“自主任务执行”跃迁，AI Agent 成为行业核心焦点。佐治亚理工学院学者 2025 年研究指出：在 Agent 工作负载中，CPU 端工具处理所占用时间占总延迟的 50%至 90.6%，GPU 在相当长时间内处于闲置等待状态。随着 Agent 应用的普及，系统对 CPU 调度、网络卸载、数据传输的需求将大幅增加，直接推高 DPU 作为基础设施卸载核心的价值。

2、不同 AI 应用领域对两类产品需求情况

AI 场景对 DPU 的需求具备以下特征：（1）需求爆发性强：AI 训练场景对 DPU 的需求侧重于消除网络瓶颈、减少 GPU 等待时间，要求 DPU 具备高带宽、低时延特性。随着 AI 训练从千卡迈向万卡，集群网络互联需求指数级增长，相对应 AI DPU 部署规模及单卡性能要求显著提高，网络带宽由 400Gbps 向 800Gbps 乃至 1.6Tbps 迈进；（2）价值量升级：AI 推理场景对 DPU 的需求侧重于加速数据访问、扩展有效内存，要求 DPU 提供高 I/O 吞吐、低访问延迟、内存扩展功能。随着大模型上下文窗口持续增长，GPU 显存容量成为推理成本的重要瓶颈，全功能 DPU 在 KV-Cache 管理方面的作用日益凸显。英伟达最新一代 BlueField-4 系列全功能 DPU 在 Vera Rubin 系统中为每颗 Rubin GPU 动态分配 16TB 专用上下文空间，将单 Token 推理成本降低最高达 90%；（3）场景延展，

从训练延伸到推理、Agent 等全周期场景，每个环节均需两类 DPU 支撑。

3、全功能 DPU 及 AI DPU 产品使用数量与 CPU、GPU、服务器、节点数量之间的配比关系

当前，AI 服务器向超节点形态演进，是 AI 大模型算力需求与现有集群架构效率瓶颈之间矛盾的必然结果。AI 服务器架构主流形态从“GPU 绝对主导”向“CPU+GPU+DPU+网络”全系统协同均衡演变。以英伟达历代 AI 服务器经典架构为例，全功能 DPU 及 AI DPU 产品在服务器中与 CPU、GPU 的配比关系如下：

AI 服务器配置	Hopper DGX H100	Blackwell GB300 NVL72	Vera Rubin VR200 NVL72
发布时间	2022 年	2025 年	2026 年
计算节点/托盘数量	1	18	18
CPU 数量（张/机柜）	2	36	36
GPU 数量（张/机柜）	8	72	72
全功能 DPU 数量（张/机柜）	2	18	18
AI DPU 数量（张/机柜）	8	72	72

如表所示，每台单节点通算/AI 服务器通常在前端网络配置 1 或 2 张全功能 DPU，而 AI 服务器后端网络中 AI DPU 与 GPU 的配比关系长期稳定在 1:1。

受以英伟达为代表的国际巨头引领，国内 AI 基础设施正从传统（GPU）8 卡 AI 服务器向 32 卡、128 卡等更大规模的“超节点”架构演进，后端网络对 AI DPU 的需求随之线性增长。根据华泰证券研究报告，中科曙光 ScaleX640 通过 640 张 GPU 堆叠，华为 910C 的 CM384 超节点最大卡数 384，阿里磐久 128 卡超节点服务器单柜支持 128 卡。

除上述计算托盘外，英伟达 16 个计算机柜组成的最新一代计算集群 DGX SuperPOD 还另外搭配 32 个存储托盘，每个存储托盘使用 4 颗 BlueField-4 系列全功能 DPU，集群存储托盘共使用 128 颗全功能 DPU。

4、产品销量增长潜力分析

DPU 作为释放 CPU 及 GPU 算力、加速网络互联、提升计算集群效能的数据中心基础设施关键组件，已广泛应用于 AI 智算、云计算、存储、通信、网络安全及边缘计算等关键战略领域。未来，随着全球新一代算力基础设施的规模化建设、人工智能与工业互联网等新兴领域需求增长，DPU 市场规模预计将持续扩大。中国 DPU 市场增长主要驱动因素如下：

（1）AI 智算集群对低时延与大规模组网能力需求提升，凸显 DPU 在网络基础设施中的价值

在大规模 AI 训练与推理场景中，随着算力集群规模的快速增长，其对高带宽、低时延及稳定通信环境的依赖显著增强，通信效率逐步成为制约整体算力利用效率的重要因素。特别是在执行 All-Reduce 这类数据“接力”汇总并分发操作时，硬件和软件层面的通信调度如果不够高效，会直接导致 GPU 等待数据，不仅拉低 AI 训练效率，也会让集群扩展受限。在 DPU 侧承载网络处理、通信加速及相关卸载任务，有助于降低 CPU 与 GPU 的非计算任务负载，确保大规模集群的高带宽、低时延及稳定通信环境，提升算力资源利用效率。DPU 在 AI 智算基础设施中的配置价值持续增加、日益凸显。

（2）算力池化、资源虚拟化、弹性裸金属云服务及多租户架构加速普及，持续推动 DPU 需求增长

随着算力资源集中部署与池化程度不断提高，多租户并行运行成为 AI 智算与云计算基础设施的主流形态，对资源隔离、性能确定性及统一调度能力提出更高要求。在该架构下，CPU 除承担核心计算任务外，还需负责大量网络、存储与安全等基础设施管理功能，负载压力持续上升。为降低 CPU 资源占用并提升系统整体运行效率，相关数据面功能逐步向 DPU 卸载，叠加云厂商提供弹性裸金属服务的强需求，推动网络、存储、安全、虚拟化和管控等能力向 DPU 侧迁移，从而带动 DPU 在数据中心中的配置比例与应用深度持续提升。

（3）数据中心向更高传输速率、芯片向先进封装技术方向演进，推动 DPU 产品持续升级

随着数据中心网络带宽需求快速增长，通信速率不断向更高速率代际演进，对低时延、高稳定性及能效表现提出更高要求。在此背景下，DPU 需在高速接口支持、数据处理吞吐能力及系统集成度等方面持续升级，以适配高带宽、低时延的数据中心运行环境。同时，先进封装技术的应用有助于提升芯片集成度与互连效率，改善功耗与性能密度表现，使 DPU 在有限空间与功耗约束下实现更高性能输出，进一步强化其在数据中心基础设施体系中的技术价值。

（4）地缘政治升级、国家层面政策支持加码，为国内 DPU 行业发展提供良好外部环境

集成电路是中美科技博弈的重点领域，在地缘政治冲突大背景下，DPU 作为数据中心基础设施的重要芯片，国产替代必要性突出、国内市场空间及前景巨大。近年来，美国政府对美国半导体产业的限制持续升级。2022 年，美国商务部工业与安全局（BIS）发布《出口管制条例》，限制先进芯片对华出口；2023 年，美国 BIS 出台了《先进计算芯片规定》（《AC/S IFR》）和《半导体生产物项管制新规》（《SME IFR》），进一步收紧芯片对华出口；2025 年，美国 BIS 发布《关于可能适用于先进计算芯片及其他用于训练 AI 模型商品的管制的政策声明》《关于通用禁令 10（GP10）对中华人民共和国（PRC）先进计算芯片适用的指南》《关于防止先进计算芯片转移的行业指南》，从使用范围、供应链制裁等角度进一步加强了对先进芯片和相关技术的出口管制，将出口管制风险进一步延伸至产业链的各个参与方。

我国对集成电路产业及 DPU 芯片高度重视，从政策和资金等层面大力支持 DPU 发展。2023 年工信部等六部门联合发布《算力基础设施高质量发展行动计划》，强调针对智能计算、超级计算和边缘计算等场景，开展数据处理器（DPU）、无损网络等技术升级与试点应用，实现算力中心网络高性能传输。2026 年，工信部等八部门联合发布《“人工智能+制造”专项行动实施意见》，提出推动智能芯片软硬协同发展，支持突破高端训练芯片、端侧推理芯片、人工智能服务器、

高速互联、智算云操作系统等关键技术，鼓励企业基于业务特征实现云边端算力协同，整合多元异构芯片资源，计划通过计算、存储、网络优化升级，加快推动已有数据中心转型智算中心。

在“东数西算”“算力网络”“人工智能+”等政策引导下，国家持续推动算力基础设施建设与算力资源的统筹布局，强调提升算力系统整体运行效率与自主可控水平。相关政策鼓励通过技术手段优化计算、网络与存储之间的协同关系，提高基础设施运行效率与安全性。作为统筹计算、网络与存储的关键组件，DPU在基础设施升级、算力协同及系统效率提升方面的功能定位与政策导向高度契合，为中国 DPU 产业发展提供了稳定且持续的外部支撑。

基于上述行业发展趋势，公司产品销量增长潜力主要体现在以下方面：

1、AI 智算集群规模扩张：随着大模型训练从千卡迈向万卡集群，集群网络互联需求指数级增长。2026 年《政府工作报告》中已明确提出加快实施超大规模智算集群。中国移动计划在长三角（嘉善）智算中心投资 50 亿元，打造万卡级智算中心，并提出到 2028 年底，在 AI 领域总体投入翻倍增长，并探索建设 10 万卡智算集群。在当前主流智算服务器架构下，国内各地万卡集群规划建设形成了若干万张 AI DPU 需求缺口。公司风驰系列 AI DPU 为支持 GPU 万卡互联的核心芯片，销量增长将直接受益于上述集群规模扩张趋势。

2、国产替代加速：弗若斯特沙利文数据，当前英伟达等海外厂商在中国 DPU 细分市场的市场占有率超 70%，仍占据主导地位。公司率先实现了国产 400Gbps 高性能全功能 DPU 芯片的技术突破与规模化商用，缩小了国产 DPU 与国际领先产品的代差，正在逐步实现高性能全功能 DPU 芯片国产替代。随着国家数据安全与产业链稳定要求的提升，公司 DPU 作为国产 DPU 领域领先产品，市场渗透率有望持续提升。

3、推理需求爆发、Agent 应用普及：德勤预计全球推理负载占 AI 算力比例 2026 年增长至约 2/3。推理场景对 DPU 的 KV-Cache 管理、存储加速、网络卸载需求远超训练场景，AI Agent 导致 CPU 调度负载大幅增长，推动全功能 DPU 的

基础设施卸载需求，将为公司全功能 DPU 产品带来新的增长空间。

4、产品迭代驱动：公司新一代 800Gbps/1.6Tbps DPU 产品已投入研发并逐步推向市场，将进一步适配下一代 AI 数据中心需求。产品代际升级将带来单颗价值量提升及新增市场机会。

报告期内，公司营业收入从 2023 年 17 万元增长至 2024 年 3,635 万元，再到 2025 年 3.70 亿元，呈现爆发式增长态势。截至 2026 年 6 月末，DPU 产品累计销量超 11 万颗，在头部云服务商、电信运营商实现大规模商用，为未来增长提供了坚实的基础。根据弗若斯特沙利文数据，中国 DPU 市场预计 2030 年达到 1,290.91 亿元，中国市场将可以为公司提供百亿级的营收增长空间。

综上所述，公司 DPU 产品销量具备较强的增长潜力，下游客户对公司产品需求呈现长期增长趋势。

（三）下游对发行人产品需求变化趋势，发行人研发方向与下游需求的匹配性

下游对公司产品需求变化趋势呈现如下特征：

（1）需求爆发性强：AI 训练场景对 DPU 的需求侧重于消除网络瓶颈、减少 GPU 等待时间，要求 DPU 具备高带宽、低时延特性。随着 AI 训练从千卡迈向万卡，集群网络互联需求指数级增长，相应对 AI DPU 部署规模及单卡性能要求显著提高，网络带宽由 400Gbps 向 800Gbps 乃至 1.6Tbps 迈进；

（2）价值量升级：AI 推理场景对 DPU 的需求侧重于加速数据访问、扩展有效内存，要求 DPU 提供高 I/O 吞吐、低访问延迟、内存扩展功能。随着大模型上下文窗口持续增长，GPU 显存容量成为推理成本的重要瓶颈，全功能 DPU 在 KV-Cache 管理方面的作用日益凸显。英伟达最新一代 BlueField-4 系列全功能 DPU 在 Vera Rubin 系统中为每颗 Rubin GPU 动态分配 16TB 专用上下文空间，将单 Token 推理成本降低最高达 90%；

(3) 场景延展，从训练延伸到推理、Agent 等全周期场景，每个环节均需两类 DPU 支撑。

公司本次 IPO 募投项目涉及的三个研发方向如下：

募投项目	研发方向	对应下游需求	匹配分析
新一代全功能 DPU 研发及产业化项目	升级现有全功能 DPU，提升性能（800Gbps）与功能完整性	云计算、AI 推理、AI Agent 场景下算力基础设施卸载刚性需求	云计算、AI 推理、AI Agent 市场持续发展，全功能 DPU 需求稳定增长
新一代 AI DPU 研发及产业化项目	开发面向智算网络场景的 AI DPU 产品，支持更高带宽互联（800Gbps/1.6Tbps）	AI 训练 / 推理 / Agent 场景爆发性需求	AI 智算需求高速增长，AI DPU 为核心增长需求
下一代高性能 DPU 核心技术研发项目	研发高性能 DPU 相关的芯片、软件等核心技术、前沿科技	万卡集群及超大规模推理前瞻性需求	产品路线图与行业带宽升级趋势匹配

公司研发方向与下游需求高度匹配，具体分析如下：

1、网络带宽升级匹配：行业正从 400Gbps 向 800Gbps/1.6Tbps 迈进，公司现有产品已实现 400Gbps 规模化商用，新一代 800Gbps/1.6Tbps 产品已投入研发，与行业带宽升级节奏匹配。

2、场景延展匹配：从云计算场景的基础设施卸载，到 AI 训练与推理场景的 GPU 高速互联，再到 AI 推理/Agent 场景的 KV-Cache 管理、RDMA 网络加速、多 Agent 调度网络支持、推理服务多租户隔离，公司产品线覆盖了 AI 产业全周期需求。全功能 DPU 主要应用于云计算场景，同时在 AI 智算前端网络、Agent 调度、存储加速等场景中发挥重要作用；AI DPU 主要应用于 AI 智算后端网络的 GPU 间高速通信。公司产品矩阵与下游需求场景高度匹配。

3、国产替代诉求匹配：我国 DPU 存在高度依赖进口的问题。近年来，美国对先进芯片对华出口的限制持续升级，下游客户国产替代需求迫切。公司 DPU 产品在保证性能的同时提供更具市场竞争力的价格，积极满足国内云服务商、运营商、AI 大模型厂商国产替代的迫切需求。

综上所述，公司下一代产品的研发方向与网络带宽升级、场景延伸、国产替代等下游需求高度匹配。

四、当前发行人 DPU 产品与数据中心主流 CPU/GPU 硬件、网络架构和软件生态之间的适配情况，现有产品与现有客户的适配情况，导入新客户所需适配周期及适配难度，结合其他客户验证及在手订单情况，分析开拓及适配其他客户的能力；行业主流 CPU/GPU 厂商自研 DPU 的情况，采用发行人 DPU 产品与搭配其自研 DPU 的综合性能、性价比对比情况，发行人 DPU 产品面临的竞争情况

（一）当前发行人 DPU 产品与数据中心主流 CPU/GPU 硬件、网络架构和软件生态之间的适配情况，现有产品与现有客户适配情况

公司 DPU 产品适配兼容情况如下：

维度	适配兼容情况
CPU 硬件	x86 架构：英特尔、超威半导体、海光等 Arm 架构：英伟达、华为鲲鹏等
GPU 硬件	英伟达、超威半导体、昆仑芯、天数智芯、登临科技等已适配 燧原科技、沐曦股份等国产卡正在适配
操作系统	海外主流：RHEL、Ubuntu、Centos、Rocky Linux、Windows 等 国产信创：麒麟、龙蜥、欧拉、中科方德、中国移动 BC-Linux、Clinux 等
网络协议	RoCEv1&2、NVMe-oF、VLAN、VXLAN、GRE、GENEVE、L2VPN 等
网络标准	支持 GSE 全调度以太网
软件生态	DPDK、SPDK、OVS 等主流开源框架，支持 SM2/SM3/SM4 等国密算法及其他主流加密算法

公司首款全功能 DPU 云霄 1.0 与现有客户适配情况如下：

现有客户	适配场景
腾讯、中国移动、新浪	云计算场景的网络卸载、存储虚拟化、安全隔离、运维管理等功能； 智算集群的网络虚拟化、多租户隔离、存储加速、安全等功能
中国移动、欣诺通信	通信、网络安全、NPD（网络入侵检测系统）场景等
国产知名大模型厂商、锐捷网络	AI 智算及存储场景

现有客户	适配场景
南方电网	电力行业数据中心的安全隔离、网络加速等场景
工商银行、招商银行	金融行业数据中心的网络安全、数据加密、业务隔离等场景

（二）导入新客户所需适配周期及适配难度，结合其他客户验证及在手订单情况，分析开拓及适配其他客户的能力

经过数年深耕，公司已构建良好的应用生态和品牌影响力。公司 DPU 产品已在头部云服务商腾讯实现规模化导入、多场景商用；同时，公司也已成为中国移动、中国联通等国内主要电信运营商以及国产知名 AI 大模型厂商、互联网企业、通信设备商等客户的重要合作伙伴，并在金融、电力等行业的关键客户拓展方面已取得进展、完成项目落地。通过与前述各领域头部客户的深度合作，公司已在产品适配性、生态构建及解决方案落地方面积累了经验，形成了相应的先发优势与竞争壁垒，有助于公司新客户开拓及导入。

DPU 作为数据中心基础设施的关键网络硬件，下游客户（尤其是头部云服务商、电信运营商等）对产品的技术验证、兼容性测试及场景适配要求严苛，验证周期通常较长。具体的验证周期根据客户行业、技术生态及产品具体需求有所不同，云服务商、电信运营商及大型金融机构的验证周期通常为 6-12 个月，AI 大模型公司、通信设备商的验证周期通常为 6-9 个月；完成验证至量产出货的周期通常为 3-6 个月。

目前，公司 DPU 产品已在头部云服务商腾讯完成大规模部署与多场景商用，并在电信运营商中国移动、国产知名 AI 大模型厂商、新浪、欣诺通信、锐捷网络、南方电网、工商银行、招商银行等多领域头部客户实现销售出货，同时积极推动公司全功能 DPU、AI DPU 两款产品在其他头部云服务、电信运营商、垂类行业客户的测试验证、量产导入，加速拓展下游市场与丰富客户结构，逐步构建“客户开拓-订单增长”的良性循环。

截至 2026 年 6 月 30 日，公司在手订单金额为 4.94 亿元，在手订单充足。

公司开拓及适配其他客户的能力主要体现在以下方面：

第一，技术架构先进，技术平台可复用。公司 DPU 产品对标国际前沿技术指标和技术路线，具有的通用可编程特性能够最大程度发挥 DPU 的功能，广泛兼容各种协议并支持多种及应用场景。公司已建立完整的 DPU 软硬件研发平台，包括芯片设计、软件开发工具、软硬一体化能力等，可快速适配新客户需求。

第二，行业导入经验积累丰富。公司已在云计算、AI 智算、通信、金融、电力等多个行业完成产品验证，具备跨行业导入经验。上述行业客户的实际应用中的需求亦反哺了公司下一代产品设计研发。

第三，产品研发路线图清晰。公司下一代产品网络速率将从当前 400Gbps 向 800Gbps 乃至 1.6Tbps 迭代，未来将持续满足客户的升级需求。

第四，国产替代政策红利。在国产替代背景下，国内数据中心客户对国产 DPU 产品的需求持续增长。公司产品支持 SM2/SM3/SM4 等国密算法，具备拓展国产信创市场的能力。

（三）行业主流 CPU/GPU 厂商自研 DPU 的情况，采用发行人 DPU 产品与搭配其自研 DPU 的综合性能、性价比对比情况，发行人 DPU 产品面临的竞争情况

1、行业主流 CPU/GPU 厂商自研 DPU 情况

根据公开数据，主流 CPU/GPU 厂商英伟达、超威半导体、英特尔、华为海思存在自研 DPU 芯片及相关产品的情况，具体如下：

网络带宽	英伟达	超威半导体	英特尔	华为
100Gbps	ConnectX-5 (2015 年) BlueField-1 (2017 年)	Pensando 第一代 (2020 年)	-	SP923Q (2025 年)
150Gbps	-	-	-	SP923H (2025 年)
200Gbps	ConnectX-6 (2019 年) BlueField-2	-	Mount Evans FPGA-Oak Springs Canyon	SP925D (2025 年) AI DPU:

网络带宽	英伟达	超威半导体	英特尔	华为
	(2020 年)		(2022 年)	SP230 (2026 年)
400Gbps	ConnectX-7 BlueField-3 (2021 年)	Pensando Elba (2022 年) Pensando Giglio Pensando Salina 400 (2024 年) Pollara 400 AI NIC (2025 年)	Mount Morgan (2026 年)	AI DPU: SP560 (2026 年)
800Gbps	ConnectX-8 (2025 年) ConnectX-9 BlueField-4 (2026 年)	-	-	AI DPU: SP560 (2026 年)

上述厂商的具体情况如下：

(1) 英伟达 (NVDA.O)

英伟达成立于 1993 年，总部位于美国，主要从事 GPU 的设计和制造，产品应用于游戏、专业可视化、数据中心和自动驾驶等领域，在 AI 领域处于全球领先地位。旗下 BlueField 系列、ConnectX 系列为 DPU 领域产品。

2020 年 4 月，英伟达完成了对全球 DPU 行业的领军企业 Mellanox 的收购。Mellanox 成立于 1999 年，是一家面向服务器、存储和超聚合基础设施的端到端以太网和 InfiniBand 智能互联解决方案与服务的供应商，曾于 2007 年在美国纳斯达克证券交易所上市，股票代码：MLNX。

(2) 超威半导体 (AMD.O)

超威半导体成立于 1969 年，总部位于美国，是高性能与自适应计算领域的领先企业，致力于提供优质的产品和服务，助力客户解决各种重大的挑战，产品应用于数据中心、嵌入式系统、游戏和 PC 等市场。

2022 年 4 月，超威半导体完成了对 DPU 公司 Pensando 的收购。Pensando

成立于 2017 年，是一家专注于数据中心边缘进行高性能、软件定义数据处理的公司。其核心产品是分布式服务平台，该平台以 DPU 为载体，旨在将云服务商级别的网络、安全、存储和遥测功能下放到每一台企业服务器中。此外，超威半导体旗下 Xilinx 品牌的 FPGA 芯片亦广泛应用于各厂商 FPGA DPU 设计方案中。

（3）英特尔（INTC.O）

英特尔成立于 1968 年，总部位于美国，是全球领先的半导体设计制造企业，产品涵盖处理器、主板、网络接口控制器、闪存、图形处理器和嵌入式处理器等，此外，英特尔还提供晶圆制造和先进封装的代工服务。与谷歌共同开发 DPU SoC 型 IPU Mount Evans，是英特尔首款高性能全功能 DPU，集成了多核 Arm 处理器和硬件加速单元，应用于数据中心服务器。FPGA 型 DPU 产品 IPU Oak Springs Canyon 基于 Agilex 系列 FPGA，提供灵活的可编程性，适合快速演进的协议和功能。英特尔旗下已分拆品牌 Altera 的 FPGA 芯片广泛应用于各云厂商自研 FPGA DPU 设计方案中。

（4）华为（海思）

华为海思成立于 2004 年，为行业客户与开发者提供芯片、器件、模组和板级解决方案，业务覆盖联接、智慧视觉、智慧媒体、显示交互、MCU、智能感知、模拟、光模块、激光显示等多个领域。华为旗下 SP900 系列是基于华为海思 Hi1822 处理器的 DPU 网卡，支持存储卸载、网络卸载、管理面卸载，最高支持 2*100GE 网络接口。

2、采用发行人 DPU 产品与搭配其自研 DPU 的综合性能、性价比对比情况

根据公开数据，公司云霄 1.0 全功能 DPU 与主流厂商同代际产品在综合性能、性价比对比如下：

参数指标	云豹 云霄 1.0	英伟达 BlueField-3	超威半导体 /Pensando Giglio	英特尔 Mount Morgan E2200
Arm CPU	V9: Neoverse N2 16x CPU Core	V8: Cortex-A78 16x CPU Core	V8: Cortex-A72 16x CPU Core	V9: Neoverse N2 24x CPU Core
RISC-V CPU	64x Hyper-Threading RISC-V	16x Hyper-Threading RISC-V	-	-
网络带宽	400Gbps	400Gbps	400Gbps	400Gbps
网络接口	112G SerDes	112G SerDes	56G SerDes	112G SerDes
PCIe 接口	PCIe Gen5 x20	PCIe Gen5 x32	PCIe Gen4 x16	PCIe Gen5 x32
内存接口	2CH DDR5 @4800MT/s	2CH DDR5 @5600MT/s	2CH DDR5 @5600MT/s	4CH LPDDR5 @6400MT/s
内存带宽	2x38.40 GB/s	2x44.80 GB/s	2x44.80 GB/s	4x51.20 GB/s
数据平面 可编程性	P4 Pipeline+64x Hyper-Threading RISC-V	P4 Pipeline+16x Hyper-Threading RISC-V	P4 Pipeline+144 MPU	P4 Pipeline、FXP 数据包处理器
RDMA 支持	RoCEv2 with Programmable DCQCN	RoCEv2 with PCC	RoCEv2	RoCEv2&Falcon

DPU 的网络带宽以及采用的工艺节点是衡量产品性能水平的重要指标，其中工艺节点是区分各代 DPU 的重要标志，也是能综合体现 DPU 技术先进性的核心指标之一；网络带宽（最高端口速率）代表 DPU 的数据吞吐能力，是评价 DPU 硬件资源承载能力的主流标准；接口传输速率主要体现为 SerDes 接口、DDR 接口等内置固化模块的性能指标，反映了 DPU 的数据传输效率性能。

公司云霄 1.0 全功能 DPU 芯片在处理器性能、网络带宽及可编程性等方面达到英伟达 BlueField-3 DPU 的同等水平，并在虚拟端口消息速率、虚拟设备数量等虚拟化关键指标上进一步优化。公司 DPU 产品在实现同等网络带宽和主要功能的同时，成本显著低于国际竞品，具备明显的性价比优势。因此，公司 DPU 产品具备适配主流服务架构的能力、对标国际主流产品的综合性能且具备更优越的性价比。

此外，在公司基于英伟达 H20 GPU 搭建的真实测试环境中，AI 网络集合通信测试结果显示，公司风驰 2.0 系列 AI DPU 与英伟达 ConnectX-7 在网络延迟、

带宽测试等结果基本持平，表明公司在研 AI DPU 与英伟达同等产品在 GPU 跨服务器通信性能方面不存在重大差异。

3、发行人 DPU 产品面临的竞争情况

DPU 行业呈现高度集中的态势，其中英伟达收购 Mellanox 后凭借强大的生态优势、深厚的技术积累以及稳固的客户群体，长期稳居行业头部，并在采用自研 ASIC/SoC 架构实现单芯片高度集成数据处理能力的 DPU 市场保持领先地位。根据弗若斯特沙利文数据，以 2025 年收入计算，英伟达等海外厂商在中国 DPU 上述细分市场的市场占有率超 70%，仍占据主导地位。

近些年，国内企业积极布局算力网络基础设施。一方面，国产独立厂商如云豹智能快速崛起，在实现市场份额突破的同时，不断提升产品技术实力，并与国际大厂展开直接竞争，推动行业格局持续演变；云豹智能坚定选择行业领先、高难度的 SoC 技术路线，已经成为国内率先实现 400Gbps 高性能全功能 DPU SoC 芯片技术突破与规模化商用的芯片厂商；另一方面，部分国产厂商及云服务商采取基于海外商业 FPGA 芯片的自研方案，加速了 400Gbps 以下 DPU 产品快速落地，共同推动行业格局持续演变。根据弗若斯特沙利文数据，不考虑云服务商自研产品，以 2025 年收入计算，公司在中国全功能 DPU 市场排名国产独立厂商第一。

公司 DPU 产品具备对标国际龙头厂商在售主力产品的技术实力，但从综合实力来看，公司目前在产品线丰富度、技术生态完善度、前沿产品研发进度等方面，相比国际龙头企业尚存在差距，仍需要通过持续的研发投入和技术创新，逐步缩小与国际顶尖公司的差距。然而，相比国内同行业公司，公司产品无论在技术路线先进性，还是在芯片架构、工艺制程、最高网络带宽、通用可编程性等技术指标方面，都具有明显的领先优势，也奠定了公司作为国内 DPU 行业头部厂商的市场地位。

从 DPU 购买决策权看，国内云服务商、运营商等下游客户在服务器核心芯片配置中起主导和决定性作用。下游客户通常基于自身业务需求、供应链稳定性、

综合性价比等因素进行自主选择。上述 DPU 选型、整机适配过程并非由第三方 CPU/GPU 公司单方面主导。CPU/GPU 厂商出于自身利益考量，通过生态绑定、捆绑销售自有品牌 DPU，虽然能够缩短采购流程、形成闭环生态内的网络协同效应，但另一方面也制约了客户选择更贴近自身网络生态需求、更具灵活性与性价比产品的机会，导致客户选型的机会成本大幅提升。特别是大型云服务商与 AI 公司更重视网络生态的灵活开发与自主可控，且当前地缘政治环境亦加剧了供应链安全的担忧。相应地，公司 DPU 产品采用开放生态、具备较高的通用可编程性，能够灵活适配下游客户网络生态与主流服务器架构、支持热迁移，具备较强的竞争实力，已成为腾讯、中国移动等头部云服务商、电信运营商的优先选择。

从外部竞争环境来看，受当前地缘政治及供应链安全影响，CPU/GPU 国产化趋势明显。中国企业从英伟达等国际 AI 芯片转向本土供应商，助推国产替代目标。国际巨头通过生态绑定、捆绑销售等方式形成的市场先发优势，在国内市场存在挑战。

国内主流 CPU/GPU 独立厂商如寒武纪、海光信息、摩尔线程、沐曦股份、燧原科技等未公布自研 DPU 计划。公司产品作为国产高性能 DPU 的代表，有望填补国内市场空白。华为海思已发布的 SP925 系列全功能 DPU 最高支持 200Gbps 网络带宽，在综合性能和生态开放性方面与公司云霄系列 DPU 存在差异，预计不会构成直接市场竞争。

从内在竞争实力来看，公司 DPU 产品相较竞争对手具备以下优势：第一，开放生态优势。公司采用开放标准协议设计，兼容绝大多数 CPU 和 GPU 服务器，而英伟达推崇 GPU-NVLink-CUDA-DOCA 封闭生态策略，客户捆绑购买存在生态锁定风险。第二，国产替代优势。公司 DPU 产品支持 SM2/SM3/SM4 国密算法，满足信创场景安全合规要求，而国外厂商不支持国密算法。第三，场景适配优势。公司 DPU 产品基于通用可编程灵活性支持多场景应用，已在多个行业完成验证，具备跨行业适配能力。第四，本地化服务优势。公司提供定制化客户开发服务，在当前 AI 基础设施持续激烈变革的过程中，能够快速响应客户一线业

务需求。

综上，虽然以英伟达为首的国际芯片厂商凭借强大的生态优势、深厚的技术积累以及稳固的客户群体获得了先发优势。但公司作为国产 DPU 头部企业，在产品、技术、生态、价格竞争方面均具有较强的竞争实力。

五、结合不同下游应用领域、下游行业主要客户合作情况，分析发行人市场拓展能力

公司不同下游应用领域、下游行业主要客户合作情况如下：

应用领域	所在行业	主要客户	合作情况
云计算	云服务商	腾讯	基于公司云霄 1.0 DPU 的腾讯玄灵网卡定位于云主机的性能加速，将原来运行在 Host CPU 上的虚拟化、网络/存储 IO 等功能下移，利用 DPU 专门处理网络通信和存储 IO 相关的任务，如 VPC 网络加速、虚拟化加速、存储云盘加速等，实现了 Host CPU 的零占用。 依托于玄灵网卡的架构加持，腾讯第九代云服务器实例提供 6750 万 PPS 的业界顶尖吞吐能力，原生支持 IPv6，网络延迟降低 35%，为金融高频交易、直播等场景提供微秒级确定性时延；连接数进一步提升 50%，可承载百万级实时并发请求；存储性能实现跨越式升级，吞吐性能比第八代实例提升最高达 200%，支持更高的并发访问和数据传输；磁盘密度提升最高达 100%，数据加解密性能提升最高达 100%； 同时在虚拟化适配方面，已率先实现玄灵网卡全子系统支持 RAS 特性，芯片 Memory ECC 校验覆盖率达 100%；结合故障预测与热逃生技术，实例不可恢复故障率趋近于零
	电信运营商	中国移动	中国移动先后发布了基于公司云霄 1.0 全功能 DPU 芯片的大云磐石 DPU 网卡（云业务）、智算琢光 DPU 网卡（智算网络）、5G 灵耀网卡（通信网络）共三款合作产品。其中，大云磐石 DPU 通过三级可编程芯片架构、硬件级网络流表管理及高性能存储协议栈等关键技术创新，实现了高达 5000 万 PPS 的网络转发性能，超 100 万 IOPS 存储性能，能够全面推动数据中心性能、效率进阶升级，支撑云主机 TCO 降低 13% 以上。磐石 DPU 已列入中国移动当前 PC 服务器、智算服务器的集采配置清单
	电信运营商	中国联通	中国联通已发布基于公司云霄 1.0 全功能 DPU 芯片的“联通大衍” DPU 产品，具备千万条流表容量和亿级 PPS 吞吐等行业顶尖能力，同时具备软硬件协同、数据硬件加密、RDMA 增强

应用领域	所在行业	主要客户	合作情况
			等先进功能，助力联通云实现软硬一体化架构升级
	互联网厂商	新浪	公司于 2025 年 6 月中标并在下半年完成量产，向新浪供应全功能 DPU 网卡，用于其私有云业务部署和裸金属服务
AI 智算	云服务商	腾讯	基于公司云霄 1.0 DPU 的腾讯玄灵网卡已适配并量产部署于腾讯自研 AI 服务器，用于支持其云端调度的智算业务
	电信运营商	中国移动	公司与中国移动合作发布的智算琢光 DPU 芯片是首颗全量支持 GSE 标准的 DPU 芯片，采用 200G 端口速率，集成报文容器喷洒和 DGSQ 拥塞控制机制等技术，与主流交换芯片完成对接验证。基于该芯片搭建的 GSE 网络性能较传统 RoCE 网络提升 30% 以上，显著提高 GPU 节点通信效率
	金融	招商银行	基于总行及分行金融业务 AI 推理任务本地化部署需求，公司作为国产供应商优胜者入围，已完成云霄 1.0 DPU 适配测试及小批量出货
通信	电信运营商	中国移动	公司与中国移动合作发布的“灵耀”智能网卡面向网络云 NFV/SDN 架构 vSwitch 卸载加速，创新性提出流表聚合、报文追踪等电信级卸载方案，支持电信级百万流表及微秒级超低时延，通过多层级热保护等定制化硬件设计应对极限工况，具备高可靠、高性能、易集成、易迭代四大特性，以及自主、标准和开放三大优势
网络安全及边缘计算	通信	欣诺通信	采购公司 DPU 芯片主要用于自身网络安全产品（NPD 网络分流器），使用公司 DPU 的高性能 Arm 核以及基础网络功能
	电力	南方电网	基于电网边缘云卸载加速、搭建分布式存储系统等需求，其网络及存储产品供应商采购公司 DPU 产品

由上表可见，公司 DPU 产品已在头部云服务商腾讯完成大规模部署及商用，同时公司已与国内头部电信运营商中国移动和中国联通、AI 大模型厂商、互联网公司、金融及电力行业头部客户深度合作。上述客户的认可是公司技术实力和市场拓展能力的体现。

通过与前述各领域头部客户的深度合作，公司已在产品适配性、生态构建及解决方案落地方面积累了经验，形成了相应的先发优势与竞争壁垒，有助于公司新客户开拓及导入。公司正在积极推动公司全功能 DPU、AI DPU 两款产品在其他头部云服务、电信运营商、垂类行业客户的测试验证、量产导入，加速拓展下游市场与丰富客户结构，逐步构建“客户开拓-订单增长”的良性循环。

公司的市场拓展能力主要体现在如下方面：

第一，技术架构先进，技术平台可复用。公司 DPU 产品对标国际前沿技术指标和技术路线，具有的通用可编程特性能够最大程度发挥 DPU 的功能，广泛兼容各种协议并支持多种及应用场景。公司已建立完整的 DPU 软硬件研发平台，包括芯片设计、软件开发工具、软硬一体化能力等，可快速适配新客户需求。

第二，行业导入经验积累丰富。公司已在云计算、AI 智算、通信、金融、电力等多个行业完成产品验证，具备跨行业导入经验。上述行业客户的实际应用中的需求亦反哺了公司下一代产品设计研发。

第三，产品研发路线图清晰。公司下一代产品网络速率将从当前 400Gbps 向 800Gbps 乃至 1.6Tbps 迭代，未来将持续满足客户的升级需求。

第四，国产替代政策红利。在国产替代背景下，国内数据中心客户对国产 DPU 产品的需求持续增长。公司产品支持 SM2/SM3/SM4 等国密算法，具备拓展国产信创市场的能力。

中介机构核查程序及核查意见：

一、核查程序

保荐人履行了以下核查程序：

1、查阅弗若斯特沙利文等第三方机构出具的行业报告，谷歌、UC Berkeley 等单位的研究成果，英伟达等竞品公开信息及新闻报道，核实主流 DPU 产品性能参数、DPU 使用效果、DPU 与服务器匹配关系，分析 DPU 行业变化趋势、进行竞品对比；

2、查阅《招股说明书》及发行人说明文件，了解发行人 DPU 产品功能、应用场景、研发方向及与下游需求匹配性，调研公司技术路线图及产品迭代规划，了解客户开拓策略、适配周期及行业拓展计划，分析发行人竞争优势及市场拓展能力；

3、查阅公司产品适配/兼容性清单，了解公司产品适配进展；

4、查阅发行人销售明细表及在手订单,获取发行人与主要客户的合作协议,了解公司产品客户适配情况、商业化情况;

5、实地访谈发行人主要客户,了解公司产品销售背景、应用场景及用户评价。

二、核查意见

经核查,保荐人认为:

1、DPU 产品对下游使用场景具有必要性及重要性;

2、DPU 能够提升数据中心整体性能、降低集群算力综合成本,客户采用相关方案的成本收益对发行人产品开拓具有积极影响;

3、发行人产品销量增长具备潜力,发行人研发方向与下游需求具有匹配性;

4、发行人 DPU 产品采用开放生态已适配主流硬件、软件并已在头部客户完成导入,公司在产品所处竞争环境中具有竞争优势;

5、发行人具备开拓及适配其他客户或市场的能力。

问题 2.关于盈利预期与前瞻性信息

申报材料显示：发行人预计最早可在 2028 年实现盈利。根据弗若斯特沙利文数据，全球 DPU 市场规模预计到 2030 年将增长至 4,362.39 亿元；中国 DPU 市场规模预计到 2030 年将增长至 1,290.91 亿元。

请发行人披露：

（1）结合新客户开拓、客户验证及在手订单情况、客户验证至量产出货周期情况、相关客户数据中心相关投资规模、全功能 DPU 及 AI DPU 在数据中心建设中的价值量占比等，分析报告期后预计销量规模增长幅度的依据。

（2）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，就预计盈利时点进行敏感性分析，并完善相关重大事项提示。

请保荐人、申报会计师说明核查依据、过程，并发表明确意见。

发行人披露：

一、结合新客户开拓、客户验证及在手订单情况、客户验证至量产出货周期情况、相关客户数据中心相关投资规模、全功能 DPU 及 AI DPU 在数据中心建设中的价值量占比等，分析报告期后预计销量规模增长幅度的依据。

（一）新客户开拓、客户验证及在手订单情况、客户验证至量产出货周期情况

经过数年深耕，公司已构建良好的应用生态和品牌影响力。公司 DPU 产品已在头部云服务商腾讯实现规模化导入、多场景商用；同时，公司也已成为中国移动、中国联通等国内主要电信运营商以及国产知名 AI 大模型厂商、互联网企业、通信设备商等客户的重要合作伙伴，并在金融、电力等行业的关键客户拓展方面已取得进展、完成项目落地。通过与前述各领域头部客户的深度合作，公司已在产品适配性、生态构建及解决方案落地方面积累了经验，形成了相应的先发优势与竞争壁垒，有助于公司新客户开拓及导入。

DPU 作为数据中心基础设施的关键网络硬件，下游客户（尤其是头部云服务商、电信运营商等）对产品的技术验证、兼容性测试及场景适配要求严苛，验证周期通常较长。具体的验证周期根据客户行业、技术生态及产品具体需求有所不同，云服务商、电信运营商及大型金融机构的验证周期通常为 6-12 个月，AI 大模型公司、通信设备商的验证周期通常为 6-9 个月；完成验证至量产出货的周期通常为 3-6 个月。

目前，公司 DPU 产品已在头部云服务商腾讯完成大规模部署与多场景商用，并在电信运营商中国移动、国产知名 AI 大模型厂商、新浪、欣诺通信、锐捷网络、南方电网、工商银行、招商银行等多领域头部客户实现销售出货，同时积极推动公司全功能 DPU、AI DPU 两款产品在其他头部云服务、电信运营商、垂类行业客户的测试验证、量产导入，加速拓展下游市场与丰富客户结构，逐步构建“客户开拓-订单增长”的良性循环。

截至 2026 年 6 月 30 日，公司在手订单金额为 4.94 亿元，在手订单充足。

（二）相关客户数据中心相关投资规模、全功能 DPU 及 AI DPU 在数据中心建设中的价值量占比

1、客户数据中心相关投资规模显著增长

中国 DPU 市场的下游需求主要来自云服务商、电信运营商及 AI 大模型、电力、金融等其他行业。其中，云服务商是中国 DPU 市场最核心需求来源。云服务商在多租户、虚拟化及云原生架构下运营大规模数据中心，对网络、存储、安全、虚拟化、数据管控等数据面及控制面能力具有高度一致且持续增长的需求；电信运营商是中国 DPU 市场的重要下游客户与算力基础设施建设主体。作为云网融合与新型信息基础设施的重要参与者，电信运营商持续加大对 AI 智算与云计算基础设施的投入，以满足政府及行业客户对安全可控、高可靠算力服务的需求；此外，AI 大模型、电力、金融等其他行业客户构成中国 DPU 市场中不断扩展的多元化需求来源。

近年来，头部云服务商、电信运营商和 AI 大模型公司持续加强数据中心等

算力基础设施建设，资本开支规模不断提升。根据伯恩斯坦（Bernstein）2025年12月发布的报告，至2028年中国云服务商与电信运营商AI相关资本支出复合增长率将达到25%。云服务商方面，根据美银证券报告，腾讯2026-2028年资本开支预计将分别达到1,850亿元、2,250亿元和2,500亿元；根据TrendForce资讯，字节跳动2026年计划资本开支将超过2,000亿元；根据大和证券研究，百度2026年资本开支预计将达到230亿元。电信运营商方面，根据上市公司财报及业绩发布会，中国移动2026年算力网络资本开支预计为378亿元，同比增长62.4%；中国联通2026年算力相关资本开支预计约为175亿元。AI大模型公司方面，2025年以来国内AI大模型公司相继完成融资，算力基础设施建设将成为重要投资方向。

随着下游行业客户的资本开支规模增长，其贡献的DPU市场规模亦加速扩张。根据弗若斯特沙利文报告，2025年中国云服务商贡献的DPU市场规模约283.62亿元，预计到2030年将以26.08%的年复合增长率增长至903.64亿元；电信运营商是中国DPU市场的重要下游客户与基础设施建设主体，2025年中国电信运营商贡献的DPU市场规模达到约104.49亿元，预计到2030年将以19.83%的复合年增长率增长至258.18亿元。

综上所述，云服务商、电信运营商及其他行业等客户数据中心相关投资规模预计将显著增长，为公司提供了广阔的下游市场空间与持续的业务发展机遇。

2、全功能DPU及AIDPU在数据中心建设中的价值量占比

全功能DPU及AIDPU主要用于数据中心建设中的服务器及相关网络设备。根据美国银行全球研究（BofA Global Research）、TrendForce等研究机构的相关报告，目前服务器仍是数据中心建设中最核心的投资环节，价值量占比超过60%。根据弗若斯特沙利文的测算，全功能DPU及AIDPU合计约占服务器及相关网络设备价值量的7.5%左右。由此估算全功能DPU及AIDPU在相关客户资本支出中的占比可达4.5%左右。

结合相关企业的资本支出数据，其DPU产品采购规模估算如下：

单位：亿元

行业	客户名称	2026 年预计资本支出规模	估算 DPU 采购规模
云服务商	腾讯	1,850	83.25
	字节跳动	2,000	90.00
	百度	230	10.35
电信运营商	中国移动	378	17.01
	中国联通	175	7.88
AI 大模型公司	国产知名 AI 大模型厂商	100	4.50
-	合计	4,733	212.99

注：腾讯、字节跳动、百度、中国移动、中国联通资本支出规模为 2026 年预测值；国产知名 AI 大模型厂商资本支出数据系根据其融资规模估算。

综上，估算上述云服务商、电信运营商、AI 大模型公司等公司合作的 DPU 采购合计需求超过 210 亿元，为公司 DPU 产品销售提供了广阔的市场空间。

（三）报告期后预计销量规模增长幅度的依据

2026-2028 年，公司预计主要产品销量持续增长。报告期内，公司首款全功能 DPU 产品于 2024 年下半年开始量产出货，推动公司 2025 年营业收入同比增长超过 900%。2026 年 1-6 月，公司营业收入同比增长超过 1,500%（未经审计），呈快速增长趋势。

公司基于全功能 DPU 产品打造标杆效应后推出 AI DPU 产品，将进一步打开增长空间，公司预计预测期（2026 年至 2028 年，下同）内营业收入仍将保持快速增长态势。目前，公司 AI DPU 产品风驰系列已成功流片，正在加速推向市场，当前推进测试验证或样机出货的客户包括腾讯、中国移动、国产知名 AI 大模型厂商、锐捷网络等，进行产品对接的客户包括云服务商 B、云服务商 A、电信运营商 A 等，该产品预计在 2027-2028 年实现放量销售。具体分析如下：

1、DPU 市场规模持续增长，国产替代空间逐步释放，公司产品增长空间显著

人工智能和数字经济产业的快速发展将推动算力需求的爆发式增长，进而带动大规模 AI 智算集群与云计算基础设施的建设投入，DPU 市场规模预计将持续增长。根据弗若斯特沙利文报告，中国 DPU 市场规模由 2021 年的 150.44 亿元增长至 2025 年的 497.58 亿元，期间年复合增长率达 34.86%；预计到 2030 年将增长至 1,290.91 亿元，是算力基础设施领域中增长最为迅速、成长空间最为广阔的细分市场之一。

根据弗若斯特沙利文报告，目前中国 DPU 市场中海外头部厂商如英伟达等占有较高市场份额，整体国产化率处于较低水平，具备核心技术自主可控能力的 DPU 产品在中国市场中的整体占比不足 20%。在全球经济环境复杂化和科技竞争不断加剧的背景下，实现核心芯片的自主可控已成为保障国家产业安全和科技自立自强的关键战略任务。DPU 作为与 CPU、GPU 并列的数据中心“三大主力芯片”之一，存在高度依赖进口的问题，国产替代需求迫切。随着 AI 智算、云计算、存储、通信、网络安全及边缘计算等关键战略领域对 DPU 的需求进一步提升，国产替代空间将加速释放。

在 DPU 市场规模持续增长的背景下，公司作为中国全功能 DPU 市场国产独立厂商排名第一的厂商，未来增长空间显著。

2、公司产品矩阵持续丰富，全功能与 AI DPU 产品加速拓展业务增长空间

公司围绕 DPU 核心功能、应用场景构建了软硬件一体化的全栈产品体系，旨在为客户提供完整的软硬件协同解决方案。在产品布局上，公司依据不同应用场景对产品性能、功能与成本的差异化需求，推出了全功能 DPU 与 AI DPU。

国内云服务商、电信运营商、AI 厂商等下游客户在服务器核心芯片配置中起主导和决定性作用，前述客户更加重视网络生态的灵活开发与自主可控。公司全功能与 AI DPU 产品支持通用可编程、具备较高的生态兼容性，同时提供更具

市场竞争力的价格，可有效满足国内云服务商、电信运营商、AI 厂商对网络生态灵活性和国产替代的迫切需求。

全功能 DPU 方面，公司第一款全功能 DPU 云霄 1.0 系列自研芯片为多核异构 SoC 芯片，采用创新的层级化可编程芯片架构实现 DPU 的通用可编程性，灵活适配云端大规模组网的复杂计算场景，在处理器性能、网络带宽及可编程性等方面达到英伟达 BlueField-3 DPU 的同等水平，并在虚拟端口消息速率、虚拟设备数量等虚拟化关键指标上进一步优化。公司 DPU 产品在实现同等网络带宽和主要功能的同时，价格显著低于国际竞品，具备明显的性价比优势，销量快速增长。该产品于 2024 年下半年开始大规模量产出货，推动公司 2025 年营业收入同比增长超过 900%，2026 年 1-6 月营业收入同比增长超过 1,500%（未经审计）。

AI DPU 方面，公司推出了 AI DPU 产品风驰 2.0 系列，其重点面向 AI 智算场景优化，在简化架构、降低功耗的基础上，进一步强化大规模低时延 RDMA 加速、可编程拥塞控制与乱序重排等功能，更好适配 AI 智算集群中 GPU 跨服务器高速互联的后端网络需求，同时也支持私有云、边缘计算、存储等场景对高性能 RDMA 网卡的需求。根据弗若斯特沙利文数据，中国 AI DPU 市场有望受益于人工智能产业的快速发展，预计 2030 年市场规模超过 940 亿元，市场空间广阔。在公司基于英伟达 H20 GPU 搭建的真实测试环境中，公司风驰 2.0 系列与英伟达 ConnectX-7 在通信延迟、带宽方面测试结果基本持平，具有与国际巨头竞品对标的基本技术实力，价格更具竞争力且更符合当前国产化及自主可控趋势。上述 AI DPU 产品已成功流片，正在加速推向市场，预计在 2027-2028 年实现放量销售。

综上所述，公司通过构建全功能 DPU 与 AI DPU 的产品矩阵，不仅实现了在核心技术指标上对标国际巨头竞品的基本技术实力，更凭借显著的性价比优势、生态灵活性以及对国产自主可控需求的深度契合，完成了首款全功能 DPU 的规模化量产与业绩爆发式增长。未来，随着风驰 2.0 系列 AI DPU 产品的逐步放量销售，公司将进一步全面切入高速扩张的智算中心及人工智能市场，持续拓宽业务增长边界。

3、公司获头部客户认可，推进规模化商业落地，在手订单充足

在客户开拓方面，公司凭借突出的产品技术实力与生态适配能力，逐步获得下游头部客户的广泛认可，加速推进规模化商业落地。公司 DPU 产品已在头部云服务商腾讯实现大规模部署、多场景商用，打造了标杆效应；同时，公司也已成为中国移动、中国联通等国内主要电信运营商以及国产知名 AI 大模型厂商、互联网企业、通信设备商的重要合作伙伴，并在金融、电力等行业的关键客户拓展方面已取得进展、完成项目落地。

通过与前述各领域头部客户的合作，公司已在产品适配性、生态构建及解决方案落地方面积累了丰富的丰富经验，形成了较强的先发优势与竞争壁垒，将持续推进产品在各领域头部客户的规模化商业落地。截至 2026 年 6 月 30 日，公司在手订单金额为 4.94 亿元，在手订单充足。上述头部客户积累为公司后续业务规模的持续扩张与业绩转化奠定了坚实基础。

综上所述，受益于下游核心客户数据中心投资规模扩大以及国产替代空间的加速释放，DPU 市场规模持续提升。公司作为国内 DPU 领域的领先厂商，通过构建全功能 DPU 与 AI DPU 的全栈产品矩阵，在核心性能上实现了对标国际巨头竞品的技术实力，全功能 DPU 产品已顺利通过下游头部客户验证并进入规模化量产阶段，AI DPU 新产品亦成功流片并加速推进商业化。目前，公司产品已成功导入云服务商、电信运营商及 AI 大模型等领域的行业头部客户，在手订单储备充沛。整体上看，公司报告期后预计销量规模增长幅度符合自身增长态势及市场发展趋势，具有合理性。

二、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，就预计盈利时点进行敏感性分析，并完善相关重大事项提示。

（一）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

受全球存储市场价格上涨等因素影响，公司 DPU 板卡 2026 年 1-6 月销售平均单价较 2025 年明显提高，同时腾讯等主要客户对 DPU 产品采购需求持续增长，上述量价提升共同推动公司 2026 年 1-6 月营业收入保持高速增长，同比增速超过 1,500%（未经审计）。基于以上市场变化以及公司实际经营情况，公司前瞻性分析相关参数及假设依据如下：

1、营业收入预测

公司主要产品为全功能 DPU 和 AI DPU，公司基于该等产品的行业市场空间、公司整体业绩变化及产品规划等，对上述产品收入进行预测，预计 2026-2028 年的公司营业收入整体增速趋于平缓。

预测期内，公司主要产品收入整体呈现持续增长态势。全功能 DPU 销售收入保持平稳增长，仍为预测期内的重要收入来源；与此同时，AI DPU 产品预计增长较快，收入占比显著提升，将成为公司重要的业务增长点。

（1）销量预测

公司目前的销量预测建立在广阔的市场需求、清晰的产品放量周期及充分的客户开拓基础之上，具体如下：

市场需求方面，随着人工智能和数字经济产业的快速发展将推动算力需求的爆发式增长，进而带动大规模 AI 智算集群与云计算基础设施的建设投入，头部云服务商、电信运营商和 AI 大模型公司持续加强数据中心等算力基础设施建设，

资本开支规模不断提升。DPU 作为与 CPU、GPU 并列的数据中心“三大主力芯片”之一，存在高度依赖进口的问题，国产替代需求迫切。

产品放量周期方面，公司 DPU 产品在基础设施卸载、算力协同及系统效率提升方面的功能定位与政策导向及市场需求高度契合，报告期内销量快速增长。报告期内，公司首款全功能 DPU 产品从 2024 年下半年开始大规模量产出货，预计未来三年销量仍将保持快速增长；公司第二款 AI DPU 产品风驰系列已成功流片，正在加速推向市场，预计在 2027-2028 年实现放量销售。

客户开拓方面，公司已完成腾讯、中国移动等头部客户的技术验证、多场景商用与规模化部署，打造了标杆效应，将持续拓展更多头部客户，并在云服务、电信运营商以及人工智能、电力、金融等应用领域加速渗透，公司产品未来销量增长具有良好的客户基础。

公司客户开拓及销售规模增长的具体依据参见本回复之“问题 2/一、结合新客户开拓、客户验证及在手订单情况、客户验证至量产出货周期情况、相关客户数据中心相关投资规模、全功能 DPU 及 AI DPU 在数据中心建设中的价值量占比等，分析报告期后预计销量规模增长幅度的依据”的相关内容。

（2）销售单价

公司基于各主要产品的近期销售单价、预测期内的市场竞争趋势以及原材料价格走势预计 2026-2028 年产品销售单价。

全功能 DPU 方面，公司产品自 2024 年开始销售，2025 年实现放量销售。2026 年 1-6 月，因全球存储价格上涨，推动公司全功能 DPU 销售单价上升。预测期内，公司 2026 年产品单价以最近一期销售单价为基准，2027 年、2028 年结合竞品格局及原材料价格走势等情况，谨慎预计单价逐年回落。

预测期内，公司上述产品的价格总体呈下降趋势，具有谨慎性及合理性。

2、毛利率预测

预测期内，公司预计综合毛利率持续上升，主要系高毛利率的 AI DPU 产品收入占比提升所致。

（1）全功能 DPU

报告期内，公司首款全功能 DPU 产品于 2024 年量产、在 2025 年放量销售。2026 年 1-6 月，受板卡对应存储配件价格持续上涨行情的影响，公司全功能 DPU 销售毛利率为 23.68%（未经审计），预测 2026 年全年毛利率将继续受存储涨价影响略有下滑；2027 年及 2028 年随着存储价格预期回落及公司规模效应进一步显现，产品毛利率将有所回升。

（2）AI DPU

公司风驰 2.0 系列 AI DPU 已成功流片，预计在 2026 年下半年量产销售，该产品定位于当前需求快速增长的 AI 智算集群后端网络市场。基于芯片架构简化带来的生产良率和产出率提升，以及 AI DPU 板卡无需集成较多高价值元器件，该款产品的定价弹性更高、毛利率也相对较高。预测期内，预计公司 AI DPU 毛利率总体相对平稳。

3、期间费用预测

预测期内，随着公司收入的快速增长，预计公司期间费用整体规模较报告期增长，同时规模效应下期间费用率总体将呈下降趋势，与报告期内趋势一致。其中，2027 年公司预测研发费用较高，主要系公司 2027 年计划完成两个项目的流片，相关光罩费用较大。

公司销售费用和管理费用主要包括职工薪酬、股份支付、宣传及推广费和咨询服务费等，公司主要基于未来相关部门人员设置、日常经营情况及股权激励分摊，预测销售费用和管理费用，其整体规模将会增长。公司研发费用主要包括职工薪酬、折旧与摊销、光罩费、技术服务和股份支付等，公司主要根据研发项目

计划和需求，综合考虑未来研发人员配置及股权激励相关股份支付费用分摊等因素进行预测。

综上所述，公司前瞻性分析相关参数及假设依据系紧密结合期后实际经营与市场变化情况制定。其中，销量预测基于算力基础设施建设带来的广阔市场需求、清晰的产品流片量产周期及头部客户的开拓进展；销售单价预测充分考量了近期销售均价、竞品价格、供应链原材料走势及产品结构变化；毛利率与期间费用预测符合公司规模效应释放及经营规划的实际情况，具有合理性。

（二）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

公司已在招股说明书“第六节”之“十、（八）报告期尚未盈利、最近一期期末存在未弥补亏损的分析”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，具体如下：

“1、原因分析

（1）DPU 市场长期由国外巨头主导，国产芯片渗透率提升面临渐进式挑战

中国 DPU 市场长期被国外巨头占据，国产自主可控芯片渗透率较低。国际大厂凭借先发技术优势、成熟的软件生态以及深厚的客户基础，在国内市场占据主导地位。近年来，包括公司在内的国内企业正加速技术追赶与产品迭代，积极推动国产化替代进程，但整体而言，国产 DPU 在技术标准适配、客户使用习惯迁移及完整产业生态构建等方面仍需系统性突破，市场拓展呈现出渐进式发展特征。报告期内，公司业务尚处于规模放量前期，销售收入暂未能覆盖高额的研发与市场投入。

（2）重点行业客户导入需经历严苛的技术验证及生态适配周期

DPU 作为 AI 智算、云计算等关键领域算力基础设施的核心网络基座，客户对其性能、可靠性以及与计算系统、软件生态的兼容性要求极高。公司产品在进入大型云服务商、电信运营商、金融及电力等国家重点行业客户的过程中，需完成多层次、长周期的技术验证、产品测试认证及软硬件生态适配。报告期内，公司正处于客户导入与产品规模化应用的推进阶段，收入及毛利规模目前尚不足以覆盖持续的高额经营投入。

(3) 公司所处 DPU 行业技术壁垒高，需持续进行高强度研发投入

DPU 属于高端网络芯片，技术壁垒高，产品研发涉及复杂架构设计、工艺制程、多模块集成、算法优化、软硬件协同等多个高难度领域，具有技术密集型和经验密集型的特征。报告期内，公司持续维持高强度的研发投入，各期研发费用金额分别为 63,065.91 万元、57,172.48 万元、55,526.46 万元，主要用于扩大研发团队、购买 IP 授权及 EDA 工具等。高额的研发投入是导致公司报告期亏损的主要原因。

(4) 实施股权激励导致股份支付费用较高，系吸引与保留核心人才的必要投入

DPU 行业是高端人才密集型行业，专业人才的竞争异常激烈。为建立并维持一支行业顶尖的研发、管理及销售团队，确保公司的长期技术竞争力与业务拓展能力，公司实施了覆盖范围较广的股权激励计划。根据企业会计准则的要求，公司对设置服务期的股份支付费用在激励对象服务期内进行分摊，确认为经常性损益，对于未设置服务期的股份支付费用，一次性确认为非经常性损益。报告期各期，股份支付金额分别为 4,700.70 万元、5,364.71 万元和 71,167.69 万元，对

报告期内的净利润产生了显著影响。该等措施虽在会计层面影响了当期利润，但有效实现了核心人才与公司长远发展的利益绑定，是公司持续创新和经营发展的重要保障。

2、影响分析

（1）对公司现金流的影响

报告期内，为支持核心技术研发、产品流片验证及市场生态建设，公司经营活动现金流出保持较高水平，报告期各期经营活动产生的现金流量净额分别为-50,469.46万元、-59,795.88万元和-70,595.03万元。同时，得益于公司良好的发展前景与技术实力，获得了知名投资机构的资金支持，融资活动为公司提供了较为充沛的营运资金。截至2025年12月31日，公司货币资金及交易性金融资产金额合计29.71亿元，流动资金储备相对充足，能够满足当前研发投入、业务拓展及日常运营的资金需求。未来，随着公司产品商业化进程加速及销售收入增长，经营活动现金流预计将逐步改善。

综上，公司尚未盈利及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。

（2）对公司业务拓展的影响

公司是率先实现400Gbps高性能全功能DPU芯片技术突破与规模化商用的国产芯片厂商，在芯片架构、性能及关键应用落地方面已建立起一定的先发优势和技术壁垒。公司DPU产品已成功导入部分行业头部客户并实现大规模部署，验证了其技术可行性与商业价值。凭借领先的技术实力、产品竞争力、对客户复杂应用需求的深刻理解以及持续的生态建设，公司正逐步扩大市场影响力。当前亏损

状态主要源于公司前期战略性研发投入与市场培育投入,是公司把握产业升级与国产替代机遇的主动选择,为公司推动市场国产替代、获取客户订单、拓展应用场景并持续迭代产品筑牢坚实基础。

综上,公司尚未盈利及存在累计未弥补亏损未对公司业务拓展产生重大不利影响。

(3) 对公司人才吸引和团队稳定性的影响

公司已凝聚了一支在芯片设计、软硬件协同及终端应用等领域经验丰富的核心团队,并在报告期内保持稳定。公司通过提供有竞争力的薪酬福利、广阔的职业发展平台以及股权激励计划,构建了多层次的人才激励与保留机制。虽然股份支付费用在财务报表上体现为亏损,但该机制有效激发了团队积极性,保障了人才队伍的长期稳定性与创造力。因此,公司尚未盈利及存在累计未弥补亏损未对人才吸引和团队稳定性产生重大不利影响。

(4) 对公司研发投入的影响

持续、高强度的研发投入是 DPU 行业保持竞争优势的必然要求,也是公司的核心战略。报告期内,公司研发投入规模始终保持在较高水平,确保了核心技术自主可控与不断突破,推动公司 DPU 产品持续向更高性能、更前沿技术方向演进,为实现国内 DPU 市场加快国产替代筑牢坚实基础。公司深刻认识到研发投入对自身长期价值的关键作用,并坚定实施以技术研发为核心的发展战略。因此,当前尚未盈利及存在累计未弥补亏损未改变公司既定的研发投入计划与技术创新节奏,未对研发投入产生重大不利影响。

(5) 对公司生产经营可持续性的影响

从市场前景看,人工智能和数字经济产业的快速发展推动了算力需求的爆发式增长,从而带动大规模 AI 智算集群与云计算基础设施的持续建设投入,为 DPU 行业带来了广阔且确定的市场空间。从公司自身看,公司拥有国内领先的技术优势与商业化能力,首款 DPU 产品已获得下游头部客户验证及大规模商用,市场竞争力与影响力不断提高,商业化进程稳步加快。公司账面资金充裕,能够保障未来一定期间的战略投入与正常运营。综合判断,公司尚未盈利且存在累计未弥补亏损不会对生产经营的可持续性产生重大不利影响。

3、趋势分析

全球数字化转型与人工智能浪潮正驱动算力基础设施发生根本性变革, DPU 作为与 CPU、GPU 并列的数据中心“第三颗主力芯片”,其战略价值与市场需求日益凸显。根据弗若斯特沙利文预测,中国 DPU 市场规模未来几年将保持高速增长,预计到 2030 年将由 2025 年的 497.58 亿元增长至 1,290.91 亿元,国产化替代进程也将同步加速。这为具备核心技术与产品竞争力的国产厂商提供了历史性发展机遇。

公司作为国内 DPU 领域的先行者,已完成了从 0 到 1 的技术与产品突破,正处在从 1 到 N 的商业化规模扩张阶段。根据弗若斯特沙利文报告,公司是率先实现 400Gbps 高性能全功能 DPU 芯片技术突破与规模化商用、具备对标国际巨头能力的国产芯片厂商。以 2025 年收入统计,公司在中国全功能 DPU 市场排名国产独立厂商第一。

凭借领先的产品技术及生态服务能力,公司首款 DPU 产品已在头部云服务商腾讯完成大规模部署及多场景商用,并与国内头部电信运营商中国移动和中国联

通、国产知名 AI 大模型厂商以及金融及电力行业头部客户深度合作及出货。公司已在产品适配性、生态构建及解决方案落地方面积累了丰富的丰富经验，形成了相应的先发优势与竞争壁垒。

随着公司第二款高性能 AI DPU 产品逐步量产交付、重点行业客户覆盖深度与广度不断拓展、产品规模效应逐步显现以及生态体系日趋完善，公司营业收入有望实现持续快速增长。同时，随着收入规模的扩大，研发及市场费用率预计将逐步下降，公司盈利能力将得到改善。

上述趋势信息为公司管理层基于自身经营状况、市场情况作出的趋势预测，受多重因素影响，该等预测性信息与未来的实际情况可能存在一定的偏差。公司提醒投资者注意，相关假设基础及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

发行人已在本招股说明书“第二节 概览”之“一、重大事项提示”之“（一）特别风险提示”和“第三节 风险因素”之“一、与发行人相关的风险”中披露了发行人“尚未盈利且存在累计未弥补亏损风险”。

5、投资者保护措施及承诺

发行人已在本招股说明书“第九节 投资者保护”之“一、本次发行完成前滚存利润的分配安排和已履行的决策程序”披露了发行人在本次发行上市后，公司的累计亏损由本次发行完成后的新老股东按其所持股份比例共同承担。发行人已在本招股说明书“第十二节 附件”之“附件二、与投资者保护相关的承诺”披露了发行人相关主体就减持股票做出的相关承诺。”

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）特别风险提示”和“第三节 风险因素”之“一、与发行人相关的风险”中披露了发行人“尚未盈利且存在累计未弥补亏损风险”，具体如下：

“公司所处的 DPU 及相关高端网络/智算芯片领域，具备研发周期长、技术门槛高、产品验证及规模化导入周期较长等特点。报告期各期，公司扣除非经常性损益后归属于母公司所有者的净利润分别为-67,270.04 万元、-61,964.26 万元和-56,160.26 万元，尚未实现盈利；截至 2025 年末，公司所有者权益为 470,425.71 万元，累计未分配利润为-116,940.97 万元，存在累计未弥补亏损。公司可供股东分配的利润为负值，预计短期内无法进行利润分配，对投资者的投资收益造成一定影响。

公司尚未盈利且存在累计未弥补亏损的主要原因如下：（1）中国 DPU 市场长期被国外巨头占据，国产芯片渗透率较低，技术标准适配、用户习惯迁移及生态建设尚需突破，市场拓展呈渐进式发展；（2）公司产品进入重点行业客户需经历严苛的技术验证及适配周期，报告期内公司产品导入处于持续推进中；（3）DPU 芯片属于高端网络芯片，技术壁垒高，涉及复杂芯片架构设计、多模块集成、算法优化、软硬件协同等多个高难度领域。为保持技术领先性并与国际巨头竞争，公司须保持较高的研发投入，包括维持较大的研发团队以及支付较高的流片费用、IP 授权费用、EDA 工具费用等。报告期内，公司研发费用持续处于高位；（4）公司所属的 DPU 行业是人才密集型行业，高端人才的争夺异常激烈。为吸引和留住核心技术人员及管理骨干，公司实施了覆盖范围较广的股权激励，并相应确认了较大金额的股份支付费用。

随着公司启动新一代高性能 DPU 产品研发，IP 及流片费用、设计成本等研发投入预计将维持在高位。若未来公司不能持续提高产品的市场渗透率、加快重点客户产品导入以扩大收入规模、合理保持产品盈利水平，或出现下游客户需求波动、产品交付与验收周期延长、供应链成本上升、核心研发项目进展未达预期等，均可能导致公司亏损进一步扩大或盈利时间推迟，进而对持续经营能力及投资者回报产生不利影响。公司可供股东分配的利润为负值，预计短期内无法进行利润分配，对投资者的投资收益造成一定影响。”

发行人已在招股说明书“第九节 投资者保护”之“一、本次发行完成前滚存利润的分配安排和已履行的决策程序”披露了发行人在本次发行上市后，公司的累计亏损由本次发行完成后的新老股东按其所持股份比例共同承担，具体如下：

“一、本次发行完成前滚存利润的分配安排和已履行的决策程序

根据公司 2026 年第三次临时股东会批准，公司本次发行上市前的滚存利润由发行后新老股东按照本次发行后的股份比例共同享有，累计未弥补亏损由发行后的新老股东按照发行后的持股比例相应承担。”

发行人已在招股说明书“第十二节 附件”之“附件二、与投资者保护相关的承诺”披露了发行人相关主体就减持股票做出的相关承诺，具体如下：

“（一）关于公司上市后的股份锁定、持股及减持意向的承诺

1、实际控制人的承诺

公司实际控制人萧启阳和张学利承诺：

“1、自发行人股票上市交易之日起 36 个月内，本人不得转让或者委托他人

管理本人直接和间接持有的发行人首次公开发行股票前已发行的股份，也不得提议由发行人回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本人亦将同等地遵守上述锁定承诺。

2、发行人上市后6个月内，如发行人股票连续20个交易日的收盘价均低于发行价（如在此期间除权、除息的，将相应调整发行价），或者上市后6个月期末收盘价低于发行价（如在此期间除权、除息的，将相应调整发行价），本人所持有的发行人股票锁定期自动延长6个月。

3、发行人上市时未盈利的，在发行人实现盈利前，本人自发行人股票上市之日起3个完整会计年度内，不减持首次公开发行前所持有的公司股份，本人在前述期间内不再担任发行人董事/高级管理人员的，仍将继续遵守前述承诺；自发行人股票上市之日起第4个会计年度和第5个会计年度内，每年减持的首次公开发行前所持有的公司股份不得超过公司股份总数的2%（在计算减持比例时，本人及一致行动人所持发行人股份合并计算，下同），并应当符合相关法律法规和规范性文件的规定。发行人实现盈利后，可以自当年年度报告披露后次日起减持首次公开发行前所持有的公司股份，但本人亦同时遵循其他限售安排和自愿锁定承诺及相关法律法规和规范性文件的规定。

4、锁定期届满后，本人在就任公司董事或高级管理人员时确定的任职期间内和任期届满后六个月内，每年转让的公司股份不超过本人所持有公司股份总数的25%；离职后半年内，不转让本人所持有的公司股份。

5、本人在发行人首次公开发行前所持有的公司股份在锁定期满后两年内减

持的，减持价格不低于本次发行时的发行价格（如在此期间除权、除息的，将相应调整减持底价）。

6、如若本人拟在锁定期届满后减持股票的，本人在作为公司 5%以上股东期间将提前 3 个交易日予以公告；同时本人将认真遵守中国证监会、深圳证券交易所关于股东减持的相关规定，审慎制定股票减持计划，通过合法方式进行减持，并按照深圳证券交易所的规定及时、准确地履行信息披露义务。

7、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，则本人愿意自动适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。

8、若本人未履行上述承诺，本人同意承担因违反上述承诺而产生的法律责任。”

2、实际控制人一致行动人的承诺

公司实际控制人的一致行动人云豹创芯、云豹创智承诺：

“1、自发行人股票上市交易之日起 36 个月内，不转让或者委托他人管理本企业直接和间接持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业亦将同等地遵守上述锁定承诺。

2、发行人上市时未盈利的，在发行人实现盈利前，本企业自发行人股票上市之日起 3 个完整会计年度内，不减持首次公开发行前所持有的公司股份；自发行人股票上市之日起第 4 个会计年度和第 5 个会计年度内，每年减持的首次公开

发行前所持有的公司股份不得超过公司股份总数的 2%（在计算减持比例时，本企业及一致行动人所持发行人股份合并计算，下同），并应当符合相关法律法规和规范性文件的规定。发行人实现盈利后，可以自当年年度报告披露后次日起减持首次公开发行前所持有的公司股份，但本企业亦同时遵循其他限售安排和自愿锁定承诺及相关法律法规和规范性文件的规定。

3、如若本企业拟在锁定期届满后减持股票的，本企业在作为公司 5%以上股东期间将提前 3 个交易日予以公告；同时本企业将认真遵守中国证监会、深圳证券交易所关于股东减持的相关规定，审慎制定股票减持计划，通过合法方式进行减持，并按照深圳证券交易所的规定及时、准确地履行信息披露义务。

4、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，则本企业愿意自动适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。

5、若本企业未履行上述承诺，本企业同意承担因违反上述承诺而产生的法律责任。”

3、其他持股 5%以上的股东承诺

持有公司 5%以上股份的股东深圳腾讯及其关联方广西腾讯、苏州湃益承诺：

“1、自发行人股票上市交易之日起 12 个月内，不转让或者委托他人管理本企业持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业亦将同等地遵守上述锁定承诺。

2、对于本企业通过发行人本次发行上市提交申请前 6 个月内增资扩股所持有的股份，该等股份自取得之日（取得之日按照公司股东名册记载该等变动之日或完成工商变更之日（如适用）孰晚原则确定，分次取得的，分别自取得之日起 36 个月内，本企业不转让或者委托他人管理本企业持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业亦将同等地遵守上述锁定承诺。

3、如若本企业拟在锁定期届满后减持股票的，本企业在作为公司 5%以上股东期间将按照届时适用的法律法规要求进行公告（如适用）；同时本企业将认真遵守中国证监会、深圳证券交易所关于股东减持的相关规定，审慎制定股票减持计划，通过合法方式进行减持，并按照深圳证券交易所的规定及时、准确地履行信息披露义务。

4、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，则本企业愿意自动适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。

5、若本企业未履行上述承诺，本企业同意承担因违反上述承诺而产生的法律责任。”

4、申报前 12 个月新增股东的承诺

(1) 共青城金棠、杭州麦智、景源基石、点睛伍号

公司股东共青城金棠、杭州麦智、景源基石和点睛伍号承诺：

“1、自发行人股票上市交易之日起 12 个月内，不转让或者委托他人管理本企业持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业亦将同等地遵守上述锁定承诺。

2、对于本企业所持有的本次发行上市提交申请前 12 个月内作为新增股东取得的发行人股份，自该等股份取得之日（取得之日按照公司股东名册记载该等变动之日或完成工商变更之日（如适用）孰晚原则确定）起 36 个月内（分次取得的，分别自取得之日计算锁定期限），本企业不转让或者委托他人管理本企业持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业亦将同等地遵守上述锁定承诺。

3、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，则本企业愿意自动适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。

4、若本企业未履行上述承诺，本企业同意承担因违反上述承诺而产生的法律责任。”

（2）其他申报前 12 个月新增股东

除上述股东外，公司其他申报前 12 个月新增股东承诺：

“1、自发行人股票上市交易之日起 12 个月内，不转让或者委托他人管理本企业/本人持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配

股等除权或除息事项而相应增加的部分，本企业/本人亦将同等地遵守上述锁定承诺。

2、对于本企业/本人所持有的本次发行上市提交申请前 12 个月内作为新增股东取得的发行人股份，自该等股份取得之日（取得之日按照公司股东名册记载该等变动之日或完成工商变更之日（如适用）孰晚原则确定）起 36 个月内（分次取得的，分别自取得之日计算锁定期限），本企业/本人不转让或者委托他人管理本企业直接和间接持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业/本人亦将同等地遵守上述锁定承诺。

3、对于本企业/本人通过发行人本次发行上市提交申请前 6 个月内增资扩股所持有的股份，自该等股份取得之日（取得之日按照公司股东名册记载该等变动之日或完成工商变更之日（如适用）孰晚原则确定）起 36 个月内（分次取得的，分别自取得之日计算锁定期限），本企业/本人不转让或者委托他人管理本企业直接和间接持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业/本人亦将同等地遵守上述锁定承诺。

4、在本企业/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，则本企业/本人愿意自动适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。

5、若本企业/本人未履行上述承诺，本企业/本人同意承担因违反上述承诺

而产生的法律责任。”

5、其他股东的承诺

(1) 瀚辰创业

公司股东瀚辰创业承诺：

“1、自发行人股票上市交易之日起 12 个月内，本企业不转让或者委托他人管理本企业持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。

2、如公司于 2026 年 6 月 17 日前（含当日）向证券交易所正式申报本次发行上市的申请文件（以下简称“首发申报”），则本企业因发行人 2025 年 12 月 17 日实施资本公积转增股本所新增取得的 707.7244 万股股份，自该等股份取得之日（取得之日指新增股份完成工商变更登记手续之日，即 2025 年 12 月 17 日）起 36 个月内，本企业不转让或者委托他人管理本企业持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。如公司于 2026 年 6 月 17 日后进行首发申报的，则本企业所持前述股份仅按照本承诺函第 1 项执行，即自发行人股票上市交易之日起 12 个月内，本企业不转让或者委托他人管理本企业持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。

3、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，则以届时有效的规定为准。

4、若本企业未履行上述承诺，本企业将承担相应的法律责任。”

(2) Dahlia、True light

公司股东 Dahlia、True light 承诺：

“1、自发行人股票上市交易之日起 12 个月内，本企业不转让或者委托他人管理本企业持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业亦将同等地遵守上述锁定承诺。

2、对于本企业通过发行人本次发行上市提交申请前 6 个月内增资扩股所持有的股份，自该等股份取得之日（取得之日按照公司股东名册记载该等变动之日或完成工商变更之日（如适用）孰晚原则确定）起 36 个月内（分次取得的，分别自取得之日计算锁定期限），本企业不转让或者委托他人管理本企业持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业亦将同等地遵守上述锁定承诺。

3、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，则本企业愿意自动适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。

4、若本企业未履行上述承诺，本企业同意承担因违反上述承诺而产生的法律责任。”

(3) 其他股东

除上述（1）至（2）项所述股东外，公司的其他股东承诺：

“1、自发行人股票上市交易之日起 12 个月内，本企业不转让或者委托他人管理本企业持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业亦将同等地遵守上述锁定承诺。

2、对于本企业通过发行人本次发行上市提交申请前 6 个月内增资扩股所持有的股份，自该等股份取得之日（取得之日按照公司股东名册记载该等变动之日或完成工商变更之日（如适用）孰晚原则确定）起 36 个月内（分次取得的，分别自取得之日计算锁定期限），本企业不转让或者委托他人管理本企业直接或间接持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权或除息事项而相应增加的部分，本企业亦将同等地遵守上述锁定承诺。

3、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，则本企业愿意自动适用变更后的法律、法规、规范性文件、政策及证券监管机构的要求。

4、若本企业未履行上述承诺，本企业同意承担因违反上述承诺而产生的法律责任。”

6、未直接持股的董事、高级管理人员的承诺

公司未直接持股的董事刘庆敏、高级管理人员陆遥承诺：

“1、自发行人股票上市交易之日起 12 个月内，不转让或者委托他人管理本人持有的发行人首次公开发行股票前已发行的股份，也不提议由发行人回购该部分股份。对以上锁定的股份因发生派息、送股、资本公积转增股本、配股等除权

或除息事项而相应增加的部分，本人亦将同等地遵守上述锁定承诺。

2、发行人上市后6个月内，如发行人股票连续20个交易日的收盘价均低于发行价（如在此期间除权、除息的，将相应调整发行价），或者上市后6个月期末收盘价低于发行价（如在此期间除权、除息的，将相应调整发行价），本人所持有的发行人股票锁定期自动延长6个月。

3、发行人上市时未盈利的，在发行人实现盈利前，本人自发行人股票上市之日起3个完整会计年度内，不减持首次公开发行前所持有的公司股份，本人在前述期间内离职的，仍将继续遵守前述承诺；公司实现盈利后，可以自当年年度报告披露后次日起减持首发前股份，但本人亦同时遵循其他限售安排和自愿锁定承诺及相关法律法规和规范性文件的规定。

4、锁定期届满后，本人在就任公司董事或高级管理人员时确定的任职期间内和任期届满后六个月内，每年转让的公司股份不超过本人所持有公司股份总数的25%；离职后半年内，不转让本人所持有的公司股份。

5、本人在发行人首次公开发行前所持有的公司股份在锁定期满后两年内减持的，减持价格不低于本次发行时的发行价格（如在此期间除权、除息的，将相应调整减持底价）。

6、本人将认真遵守中国证监会、深圳证券交易所关于股东减持的相关规定，审慎制定股票减持计划，通过合法方式进行减持，并按照深圳证券交易所的规定及时、准确地履行信息披露义务。

7、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，则本人愿意自动适用变更后的法律、法规、规

范性文件、政策及证券监管机构的要求。

8、若本人未履行上述承诺，本人同意承担因违反上述承诺而产生的法律责任。” ”

（三）就预计盈利时点进行敏感性分析，并完善相关重大事项提示

公司已在招股说明书“第六节”之“十六、（三）3、公司盈利时点的敏感性分析”中补充披露如下：

“3、公司盈利时点的敏感性分析

公司能否于 2028 年实现盈利主要受收入及综合毛利率影响，针对上述因素的敏感性分析如下：

实现盈利年度		收入变动比例						
		-20%	-10%	-5%	不变	+5%	+10%	+20%
毛利率 变动百 分点数	-10	2029 年	2029 年	2028 年	2028 年	2028 年	2028 年	2028 年
	-5	2029 年	2028 年	2028 年	2028 年	2028 年	2028 年	2028 年
	不变	2028 年	2028 年	2028 年	2028 年	2028 年	2028 年	2028 年
	+5	2028 年	2028 年	2028 年	2028 年	2028 年	2028 年	2027 年
	+10	2028 年	2028 年	2028 年	2028 年	2027 年	2027 年	2027 年

注：以上时间为公司扣非后净利润首次为正的时间”

针对公司尚未盈利及前瞻性信息，公司已在招股说明书“第二节”之“一、（三）公司报告期内尚未实现盈利的特别事项”中披露如下：

“（三）公司报告期内尚未实现盈利的特别事项

随着全球新一代算力基础设施的规模化建设、人工智能与数字经济等新兴领域需求增长，DPU 市场规模预计将持续增长。根据弗若斯特沙利文数据，中国 DPU 市场规模由 2021 年的 150.44 亿元增长至 2025 年的 497.58 亿元，预计到 2030

年将增长至 1,290.91 亿元，为公司发展提供了广阔的市场空间。此外，公司产品在头部云服务商的规模化商用已充分验证了其技术成熟度与生态兼容性，正加速渗透至更多云服务商、电信运营商、AI 大模型公司及金融、电力等行业客户领域。随着公司产品矩阵和应用场景进一步丰富，客户导入和验证进一步推进，公司预计未来收入将持续增长。公司结合自身产品开发、客户导入和验证进度，以及行业市场空间等因素，预计最早可在 2028 年实现盈利。

公司上述有关未来营业收入增长、产品放量周期、财务数据变动、预计盈利时间等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。”

中介机构核查程序及核查意见：

一、核查程序

保荐人、申报会计师履行了以下核查程序：

1、查阅发行人审计报告及财务报表，获取发行人收入成本明细，访谈发行人管理层，了解公司报告期内收入、成本、毛利率以及期间费用变动情况，了解和分析发行人尚未盈利的原因及合理性；

2、获取发行人盈利预测资料，复核前瞻性信息关键假设及预测数据的合理性；

3、查阅相关行业研究报告和公开披露数据，了解行业发展情况及市场规模增长情况，查询主要公司客户 CAPEX 资本支出数据，了解全功能 DPU 及 AI DPU 在资本支出中的价值量占比；

4、获取发行人在手订单情况，访谈发行人管理层，了解发行人主要产品情况、客户开拓及验证导入进展情况；分析发行人销售规模增长及收入预测的合理性；

5、获取发行人报告期内及 2026 年 1-6 月的收入成本明细、在手订单情况，

访谈发行人财务人员，了解并分析发行人产品单价、成本及毛利率的预测依据及合理性；

6、获取发行人期间费用预测明细，访谈发行人管理层，了解并分析发行人各期间费用预测的完整性及合理性；

7、获取发行人实际控制人、董监高及核心技术人员的相关承诺；

8、复核发行人盈利时点的敏感性分析，并在此基础上进一步分析发行人前瞻性信息的合理性和谨慎性。

二、核查意见

经核查，保荐人、申报会计师认为：

1、发行人已构建良好的应用生态和品牌影响力，客户开拓、客户验证情况良好且在手订单充足，腾讯等主要客户数据中心相关投资规模显著增长；结合全功能 DPU 及 AI DPU 在数据中心建设中的价值量占比等情况，公司预测销量及收入增长符合自身增长态势及市场发展趋势，具有合理性。

2、发行人前瞻性分析相关参数及假设依据具有合理性，发行人按照《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，谨慎估计公司最早可在 2028 年实现盈利，发行人亦已在招股说明书中披露预计盈利时点的敏感性分析，并完善相关重大事项提示。

问题 3.关于关联交易

申报材料显示：腾讯控股及相关方为发行人关联方，报告期内向发行人采购 DPU 产品并大规模部署于数据中心服务器中，2025 年销售金额为 3.52 亿元。目前国内暂无其他厂商可实现同等级别 DPU SoC 芯片量产及商业化落地。

请发行人披露：

（1）与腾讯业务合作的稳定性，产品的定制化程度、可编程性情况、产品在腾讯数据中心应用过程中积累经验的可复用性，发行人对其他非关联客户的开拓有效性；预计与腾讯控股及相关方交易金额、占营业收入比例的未来变化趋势。

（2）发行人对腾讯的重大依赖是否可能对发行人产生重大不利影响，关联交易是否影响发行人的独立性，单一客户重大依赖对发行人持续经营能力的影响，并进一步完善风险提示。

请保荐人、申报会计师说明核查依据、过程，并发表明确意见。

发行人披露：

一、与腾讯业务合作的稳定性，产品的定制化程度、可编程性情况、产品在腾讯数据中心应用过程中积累经验的可复用性，发行人对其他非关联客户的开拓有效性；预计与腾讯控股及相关方交易金额、占营业收入比例的未来变化趋势。

（一）公司与腾讯业务合作的稳定性

公司主营业务为 DPU 及相关产品的研发与销售。腾讯是国内领先的云服务提供商，其持续向公司采购 DPU 产品，并在数据中心服务器中实现大规模部署。公司与腾讯的合作关系稳定且可持续，主要依据如下：

1、从商业价值角度，公司产品技术先进，可帮助腾讯明显提升算力基础设施性能并优化成本，实现较大收益

公司作为国内领先的 DPU 芯片设计企业，长期聚焦关键核心技术的攻关，

不断强化技术创新能力，坚持核心技术自主可控发展路线，通过自主研发形成了具备较强竞争力的核心技术体系，在架构设计、芯片设计、系统软件、验证测试领域形成了若干核心技术，为公司 DPU 产品在性能、质量、易用性、迭代效率等方面形成市场显著竞争力奠定坚实基础。公司云霄 1.0 全功能 DPU 芯片在处理器性能、网络带宽及可编程性等方面达到英伟达在售主力产品 BlueField-3 DPU 的同等水平，并在虚拟端口消息速率、虚拟设备数量等虚拟化关键指标上进一步优化，同时具备显著性价比。

腾讯在算力基础设施规模化部署公司 DPU 产品能够带来显著的性能提升与成本优化收益：一是算力效率提升，通过将网络、存储、安全等基础设施负载从 CPU 卸载至 DPU 处理，释放大量 CPU 核心算力用于承载业务应用，有效提升服务器算力利用率；二是整体成本降低，在同等业务承载能力下减少服务器部署数量，降低数据中心 CAPEX 投入，同时通过硬件加速提升单位算力产出，优化整体 TCO；三是功耗显著下降，DPU 专用处理引擎相较通用 CPU 处理同类负载具备更高的能效比，有助于降低数据中心整体功耗与运营成本，契合绿色数据中心的发展方向。

综上，公司凭借自主可控的核心技术体系与对标国际一线的产品竞争力，能够持续为腾讯算力基础设施创造性能提升、成本优化、功耗降低的多重商业价值，双方合作建立在坚实的技术契合与商业价值基础之上，具备较强的稳定性与可持续性。

2、从市场需求角度，目前阶段腾讯稳定攀升的 DPU 需求为双方合作提供坚实基础

DPU 作为释放 CPU 及 GPU 算力、加速网络互联、提升计算集群综合效能的数据中心基础设施关键组件，已广泛应用于 AI 智算、云计算、存储、通信、网络安全及边缘计算等关键战略领域。未来，随着全球新一代算力基础设施的规模化建设、人工智能与工业互联网等新兴领域需求增长，DPU 市场规模预计将持续扩大。

近年来，中国 DPU 市场规模持续增长，根据弗若斯特沙利文数据，中国 DPU 市场规模由 2021 年的 150.44 亿元增长至 2025 年的 497.58 亿元，期间年复合增长率达 34.86%；预计到 2030 年将增长至 1,290.91 亿元，是 AI 智算与云计算基础设施领域中增长最为迅速、成长空间最为广阔的细分市场之一。对比全球市场，2025 年中国 DPU 市场规模占全球约 25.32%，预计到 2030 年，占比将提升至约 29.59%，中国市场在全球 DPU 行业地位逐步提升。

腾讯作为国内领先的云服务商，面临着数据中心大规模扩容及算力效率提升的迫切需求，急需高性能国产 DPU 产品以优化其算力基础设施。一方面，腾讯云服务规模位居国内第一梯队，公有云与 AI 智算业务的持续增长驱动服务器部署规模不断扩大，对作为服务器核心网络硬件的 DPU 需求同步攀升；另一方面，在供应链安全日益重要的背景下，降低对境外芯片供应商的依赖、实现核心基础设施自主可控已成为头部云厂商的核心战略。此外，通过 DPU 卸载基础设施负载、提升算力利用率、降低整体 TCO 与功耗，也是云厂商优化运营效率的关键路径。

综上，公司与腾讯的合作系基于持续的业务需求，具备坚实的长期合作基础。

3、从合作发展角度，双方合作渊源深厚，技术粘性与产品迭代共同保障长期稳定

从历史合作维度看，公司与腾讯自 2021 年起即开展合作，合作渊源深厚、协同程度高。双方基于在云化网络与 DPU 领域的深度产业理解，就公司首款全功能 DPU 产品开展研发合作，共同打造面向腾讯下一代算力基础设施和多元化业务场景需求、并在性能、成本以及可编程灵活性方面具备核心竞争力的 DPU SoC 产品。公司于 2023 年完成该款 DPU SoC 的研发设计，2023 年 7 月正式流片，同年 10 月芯片回片并逐步完成功能测试、性能调优、软件开发、量产导入等工作，2024 年下半年开始实现量产，整体开发与验证周期长达 2 至 3 年。

在长期合作过程中，双方深度磨合、彼此熟悉对方的技术路线与业务需求，公司产品已充分适配腾讯云的网络架构与软件生态。若腾讯更换 DPU 供应商，

需重新经历产品定义、联合研发、适配验证、量产导入、规模化部署的完整周期，面临较高的时间成本、技术成本与切换风险，此外，根据致尚科技的公开信息披露，英伟达的 BlueField-3 的平均单价为 1.50 万至 3.00 万元，远高于公司产品平均价格，国内暂无其他厂商可实现同等级别 DPU SoC 芯片量产及商业化落地。因此，腾讯不存在随意替换公司产品的动机。

从当前及未来合作维度看，公司与腾讯保持深度稳定的协同合作，持续联合推进多款新一代 DPU 的开发验证，在研迭代产品持续筑牢双方合作根基。公司 2025 年 6 月已启动云霄 2.0 全功能 DPU 研发，与腾讯深度对齐基于 800G 网络带宽的云网络及新一代 AI 智算架构下多场景、多功能产品需求与性能标准、技术规格，进一步在该 DPU 产品开展技术合作。

综上所述，公司与腾讯的合作渊源深厚，技术粘性与产品迭代共同保障长期稳定。

4、报告期后公司与腾讯的合作情况

报告期后，公司向腾讯及相关方销售收入持续保持高速增长。截至 2026 年 6 月 30 日，公司对腾讯及相关方的在手订单充足，未来合作稳定。

综上，公司与腾讯的业务合作具有稳定性。

（二）产品的定制化程度、可编程性情况、产品在腾讯数据中心应用过程中积累经验的可复用性，发行人对其他非关联客户的开拓有效性

1、产品定制化程度

公司 DPU 产品针对腾讯存在一定程度的适配性开发，但该等定制化主要针对场景及软硬件生态的适配，而非芯片核心架构与底层技术，具体情况如下：

（1）DPU 芯片

公司 DPU 芯片整体采用通用化架构设计，核心处理器、网络引擎、加速模

块及底层技术路径均为通用配置，支持客户基于硬件的灵活可编程，因此广泛适配云计算、AI 智算、通信、存储等多领域客户需求。

（2）DPU 板卡

硬件设计维度：网卡的硬件设计相对标准，公司对外销售通常提供标准配置的通用 DPU 板卡，部分客户因自身需求会在接口规格、散热设计等方面进行微调。公司通常将 DPU 板卡硬件开发工作委托 ODM 厂商实施，公司负责提供设计规格、产品需求定义及测试验证方案，对开发方向与最终成果进行全流程管控与审验。公司 DPU 板卡硬件设计的核心目标是实现板卡形态、接口规格与客户服务器部署环境的匹配，属于板卡产品物理形态层面的适配，不涉及 DPU 底层技术路线的调整。

软件开发维度：由公司软件研发团队自主完成，构建基于 DPU 芯片与板卡的一体化软件方案，包括底层驱动 SDK、客户二次开发接口及技术支持体系等。软件开发的确保 DPU 板卡与客户云架构、网络协议及智算网络体系的兼容适配，属于软件生态层面的调整优化，底层驱动框架与核心软件具有通用性，部分适配客户云业务自研架构的软件 API 接口存在差异。

因此，公司 DPU 板卡仅在软硬件层面进行物理形态与生态层面的适配，但不涉及芯片核心架构与底层技术路线的调整。

2、产品可编程情况

公司自主研发的全功能 DPU 是国内少数基于自研 RISC-V 和 Arm 最新一代 V9 指令集混合架构的 DPU SoC 芯片。该产品搭载创新的“层级化可编程”的芯片设计架构，即通用可编程、嵌入引擎可编程、领域可编程，分别对应芯片的通用功能、内部特定引擎功能以及垂直领域功能。通过此架构，成功将用户业务层与底层基础资源管理层分离开来，实现了以 DPU 为中心，对服务器内的计算、网络和存储资源进行统一调配和安全控制。以层级化可编程创新架构，公司 DPU 产品实现了从芯片架构设计到产品部署端到端全流程自主可控，面向异构计算持

续迭代，推动 DPU 架构面向云数据中心持续演进。

基于该“层级化可编程”芯片架构，该产品集成 64 个自研 RISC-V CPU 核，并采用创新的多线程技术，实现数据处理和卸载的可编程性；同时集成 Arm 最新一代 V9 架构的高性能 Neoverse N2 核以实现更通用的可编程性及对 Linux 开源软件生态的支持，从架构上真正实现 DPU 的通用可编程性。公司 DPU 芯片还支持可编程 RDMA 技术，能够灵活适配大规模组网复杂场景所需的高带宽、低时延传输及可编程拥塞控制算法。

3、产品在腾讯数据中心应用过程中的积累经验的可复用性，发行人对其他非关联客户开拓的有效性

公司与腾讯之间的合作协议中未约定任何排他性条款、技术成果限制使用条款或对公司开拓其他客户的限制性条款，公司在与其合作过程中形成的通用技术能力、产品迭代成果及工程化经验，均可自由应用于其他客户的产品开发与市场拓展，不存在法律或合同层面的限制。

结合前述产品定制化程度与可编程性特点，公司在腾讯数据中心应用过程中积累的技术经验与工程能力具备充分的可复用性。同时，公司商业化初期与腾讯开展深度战略合作具备显著长期价值，能够为公司全域市场开拓形成强力支撑，具体表现如下：

一是腾讯云作为国内头部云服务商，覆盖通用算力、AI 智算、存储、网络安全等全业务场景，其全维度业务需求反馈与产品生态协同，能够持续反向优化公司 DPU 的产品定义、功能迭代与性能打磨；适配腾讯的标准化 DPU 解决方案可覆盖国内绝大多数云厂商通用业务场景，为公司拓展其他云服务商、电信运营商、AI 大模型公司等客户奠定产品基础；

二是高端 DPU 芯片从设计落地至大规模量产存在极高综合门槛，腾讯严格的量产标准化规范、现网质量管控要求，有效推动公司 DPU 产品以高标准完成灰度测试至规模化商业落地的全流程，并实现高品控标准下的十万卡级出货。依

托在腾讯大规模现网落地沉淀的通用技术方案与量产管控能力，相关标准化体系、技术经验可快速复制至其他云服务商、电信运营商、AI 大模型公司等客户，形成权威商用标杆背书，充分提振客户批量采购合作信心，释放行业示范引领效应；

三是腾讯是国内云服务行业前沿技术生态的核心引领方，双方在 ODCC 超节点、算力池化等下一代算力架构、行业标准制定层面同步开展联合研发，有助于公司 DPU 产品深度融入行业主流技术体系，加速全行业市场渗透与产品推广。

4、公司的其他客户开拓情况

如上所述，公司自成立初期便与腾讯展开产品技术合作，依托公司与腾讯合作中沉淀的多场景 DPU 解决方案、通用技术与量产管控能力、标准化体系等产品化、商业化能力与经验，公司具备向其他客户开拓的能力。

经过数年深耕，公司已构建良好的应用生态和品牌影响力。公司 DPU 产品已在头部云服务商腾讯实现规模化导入、多场景商用；同时，公司也已成为中国移动、中国联通等国内主要电信运营商以及国产知名 AI 大模型厂商、互联网企业、通信设备商等客户的重要合作伙伴，并在金融、电力等行业的关键客户拓展方面已取得进展、完成项目落地。通过与前述各领域头部客户的深度合作，公司已在产品适配性、生态构建及解决方案落地方面积累了经验，形成了相应的先发优势与竞争壁垒，有助于公司新客户开拓及导入。

DPU 作为数据中心基础设施的关键网络硬件，下游客户（尤其是头部云服务商、电信运营商等）对产品的技术验证、兼容性测试及场景适配要求严苛，验证周期通常较长。具体的验证周期根据客户行业、技术生态及产品具体需求有所不同，云服务商、电信运营商及大型金融机构的验证周期通常为 6-12 个月，AI 大模型公司、通信设备商的验证周期通常为 6-9 个月；完成验证至量产出货的周期通常为 3-6 个月。

目前，公司 DPU 产品已在头部云服务商腾讯完成大规模部署与多场景商用，并在电信运营商中国移动、国产知名 AI 大模型厂商、新浪、欣诺通信、锐捷网络、南方电网、工商银行、招商银行等多领域头部客户实现销售出货，同时积极

推动公司全功能 DPU、AI DPU 两款产品在其他头部云服务、电信运营商、垂类行业客户的测试验证、量产导入，加速拓展下游市场与丰富客户结构，逐步构建“客户开拓-订单增长”的良性循环。

报告期内，公司积极拓展新客户，非关联方客户销售金额分别为 17.32 万元、248.62 万元和 1,767.22 万元，非关联客户销售规模呈快速增长趋势。2026 年 1-6 月，公司对非关联方客户销售金额继续保持快速增长趋势。截至 2026 年 6 月 30 日，公司在手订单金额为 4.94 亿元，在手订单充足。

（三）预计与腾讯控股及相关方交易金额、占营业收入比例的未来变化趋势

综合公司产品技术实力、腾讯 DPU 采购需求增长、双方长期稳定的合作基础及新产品合作与验证进展，预计公司与腾讯及相关方的交易金额未来将持续增长，具体分析详见本题回复之“（一）公司与腾讯业务合作的稳定性”。

公司预计关联交易占比将呈现短期保持高位、长期逐步下降的趋势，具体分析如下：

1、短期而言，公司关联交易规模保持增长

短期来看，公司对腾讯的销售金额预计将持续增长，收入占比仍将维持在较高水平，符合行业规律与公司发展战略：

一是腾讯 DPU 采购需求较大且持续增长。腾讯作为国内头部云服务商，亦是国内算力基础设施投资主力，2025 年资本开支达到 1,128.75 亿元，2026 年一季度资本开支达到 319 亿元，预计全年资本开支将维持增长态势；腾讯作为公司最早完成产品适配并实现大规模商用的头部客户，采购需求将伴随自身云业务与 AI 算力扩张持续释放，公司当前 DPU 供应占其整体采购需求的比例尚低，未来份额增长仍有较大空间。因此，腾讯短期内仍将是公司收入的核心来源。

二是国内 DPU 市场长期被英伟达等海外头部厂商占据，公司第一款 400Gbps 高性能全功能 DPU 芯片量产实现了国内行业的首次突破，但在市场竞争中仍面

临国际巨头成熟产品的较大压力。基于此背景，公司在首款 DPU 产品上主动采取“以点带面、重点突破”的市场策略，以腾讯为云服务商领域“种子客户”，集中资源向种子客户倾斜，确保公司首款产品在头部客户快速上量，在打磨商业化能力的同时积累核心客户资源并形成品牌背书、市场影响力，为后续加速国内 DPU 市场国产替代、赢取更多市场份额筑牢基础。

三是头部客户验证放量存在客观周期。DPU 作为 AI 智算、云计算等关键领域算力基础设施的核心网络基座，客户对其性能、可靠性以及与自身算力基础设施、网络生态的兼容性要求极高。公司产品在进入大型云服务商、电信运营商、金融及电力等国家重点行业客户的过程中，需完成多层次、长周期的技术验证、产品测试认证及软硬件生态适配，整体验证与放量周期较长。公司当前正在积极拓展其他头部客户的过程中，相关业绩释放需要一定时间。

2、长期而言，公司关联交易占比将会下降

长期而言，经过数年深耕，公司已构建良好的应用生态和品牌影响力，未来有望通过持续的客户拓展和产品拓展，增加对非关联客户的销售规模，逐步降低关联销售占比。

客户拓展方面，公司 DPU 产品除在腾讯实现多场景、规模化商用外，也已成为中国移动、中国联通等国内主要电信运营商以及国产 AI 大模型公司、互联网企业、通信设备商等客户的重要合作伙伴，并在金融、电力等行业的关键客户拓展方面已取得进展、完成项目落地。通过与前述各领域头部客户的深度合作，公司已在产品适配性、生态构建及解决方案落地方面积累了经验，形成了相应的先发优势与竞争壁垒，具体情况详见本题回复之“（二）/4、公司的其他客户开拓情况”。

产品拓展方面，公司在首款云霄 1.0 全功能 DPU 实现大规模部署的经验基础上加速产品升级迭代，已启动研发支持 800G 网络带宽及更高性能的下一代云霄 2.0 全功能 DPU 产品。同时，公司重点面向智算网络优化、适配 GPU 高速互联的风驰 2.0 AI DPU 产品已完成芯片流片及测试、正加速推向市场，将成为公

司第二条重要产品线。相较全功能 DPU 产品需适配客户前端网络架构和系统生态，AI DPU 重点在于支持行业通用的智算网络协议与算力架构，因而具备更高的产品普适性，市场开拓更快速，有利于公司进一步拓宽市场边界、丰富客户类型，持续弱化对单一领域、单一客户的依赖，稳步优化整体客户结构，降低关联方交易占比。随着多代际、多场景产品矩阵的持续完善，公司产品将逐步实现对云计算、AI 智算、通信、存储、网络安全等多元应用场景的深度覆盖，进而带动客户群体的多元化拓展，持续优化客户结构、降低单一客户集中度风险。

综上，公司预计关联交易占比将呈现短期保持高位、长期逐步下降的趋势。

二、发行人对腾讯的重大依赖是否可能对发行人产生重大不利影响，关联交易是否影响发行人的独立性，单一客户重大依赖对发行人持续经营能力的影响，并进一步完善风险提示。

（一）发行人对腾讯的重大依赖是否可能对发行人产生重大不利影响，单一客户重大依赖对发行人持续经营能力的影响

1、报告期内，公司对腾讯销售收入占比较高，是符合行业客观规律与各自发展阶段需要的双向选择

（1）行业客观规律维度

DPU 产品作为高性能服务器的关键组件，云计算是最主要的应用场景之一。在我国云服务市场高度集中的竞争格局下，国产 DPU 厂商优先服务头部云服务客户，符合行业客观发展规律。

从下游需求结构来看，根据弗若斯特沙利文统计数据，云服务厂商是中国 DPU 市场最核心的需求来源，2025 年其贡献的国内 DPU 需求占比超过 50%，决定了 DPU 厂商的客户结构向云服务领域倾斜。从下游竞争格局看，我国云服务市场呈现“巨头主导”特征，2025 年第一季度国内前三大云厂商（腾讯、阿里巴巴、字节跳动）合计市场份额超 60%。下游高度集中的市场结构，亦使 DPU 厂商需集中资源深耕头部优质客户，符合行业经营特点。腾讯作为国内云服务及

AI 应用领域的领军企业,是公司作为国产高端 DPU 稀缺供应商的优先合作对象,双方合作系在下游高集中度市场格局下,头部云服务厂商与核心国产 DPU 芯片企业基于市场化原则形成的战略合作,具有合理的商业逻辑。

从行业发展规律来看,高端芯片企业在发展初期优先通过头部“种子客户”完成产品打磨与标杆案例建设,与下游头部客户形成示范效应,再向全行业推广及商业化落地,是行业通行的成功路径。海光信息、燧原科技等算力芯片企业在商业化初期均呈现高度依赖单一或少数种子客户的特征,表明客户集中度高在高端芯片企业发展初期具有一定普遍性。

综上,公司发展初期对腾讯的销售收入占比较高,是下游需求结构、市场竞争格局、产业链上下游关系、企业发展规律共同决定的客观结果,具有充分的商业合理性,符合 DPU 行业下游客户集中的客观现状和企业发展规律。

(2) 双方发展阶段维度

从腾讯角度来看,采购高性能国产 DPU 是布局下一代数据中心架构、建设高效算力基础设施、保障供应链安全的必要选择。一方面,随着国内主流云服务商网络向 400Gbps 及以上带宽演进,传统 FPGA DPU 方案在功耗和成本方面的约束逐步显现,腾讯自研的银杉 1.0 FPGA 网卡难以满足自身云基础设施升级需要,采用符合行业主流发展趋势的 DPU SoC 方案成为必由之路;另一方面,在算力硬件供应链安全日益重要的背景下,采购国产高性能 DPU 以替代海外商业 FPGA 芯片及英伟达网卡产品,是腾讯降低国际供应商依赖、丰富供应链体系的关键战略举措。

国内 DPU 行业发展处于初期阶段,具备成熟技术与产品落地能力的第三方芯片厂商十分稀缺。公司作为率先实现 400Gbps 高性能全功能 DPU 芯片技术突破与规模化商用的国产芯片厂商,产品技术路线与性能对标国际大厂主流水平,同时具备显著性价比。基于芯片架构与自主软件栈的通用可编程性与灵活性,公司全功能 DPU 能够充分满足腾讯面向多类场景、功能的新一代网络基础设施升级需求。基于公司云霄 1.0 DPU 芯片的玄灵智能网卡助力腾讯第九代云服务器实

现了 6,750 万 PPS 的业界顶尖吞吐性能、较上一代服务器提升最高达 200%，同时网络延迟降低 35%、连接数进一步提升 50%，可承载百万级实时并发请求，存储性能实现跨越式升级，支持更高的并发访问和数据传输，磁盘密度提升最高达 100%、数据加解密性能提升最高达 100%。因此，采购公司 DPU 产品是腾讯当前升级云化网络、加大高性能算力建设投入、优化成本效益、保障供应链安全的必然选择。

从公司角度来看，公司成立至今尚不足六年，首款 DPU 芯片刚完成量产与大规模商用，正处于从技术产品闭环向规模化销售跨越的关键成长期。在发展初期阶段，因内外部资源有限，公司主动采取“以点带面、重点突破”的市场策略，将技术支持、产能交付等资源聚焦于国内头部云服务厂商腾讯，是快速实现产品量产与商业化的合理选择。

腾讯基于自身需求与公司建立长期深度合作，在公司 DPU 芯片研发阶段积极开展场景、功能需求对接与技术合作，并在量产导入阶段加速软件生态与服务器硬件的适配，双方形成了深度的战略协同，腾讯由此成为公司早期战略合作的“种子客户”，公司亦成为腾讯 DPU 的核心供应商，双方具备长期稳定合作的基础。因此，公司报告期内收入主要由腾讯贡献，是新品导入期聚焦核心资源、与战略客户深度协同的结果，也是公司基于自身发展阶段的合理选择。

综上所述，公司与腾讯在 DPU 领域长期稳定的深度合作是符合行业客观规律、基于各自发展阶段需要的战略性、市场化双向选择。

2、报告期内，公司与腾讯业务合作稳定且具有持续性

公司作为国内领先的独立 DPU 厂商，首款高性能全功能 DPU 关键性能指标对标英伟达在售主力产品 BlueField-3，是国内少数达到国际大厂同等水平的 DPU SoC 产品，具有较强国产替代价值。在中美竞争加剧、供应链安全面临挑战的背景下，腾讯作为国内头部云服务厂商，其在云计算和 AI 业务领域的快速扩张对 DPU 供应的稳定性、自主可控性提出了更高要求，选择头部国产 DPU 厂商作为主力供应商，是基于市场化原则与产品技术性能做出的审慎选择。报告期内，公

司与腾讯业务合作保持稳定，并已建立长期战略合作关系，期后在手订单充足，业务合作具有较强的可持续性，具体情况详见本题回复之“一/（一）公司与腾讯业务合作的稳定性”。

3、公司产品通用性已获验证，多领域标杆客户落地驱动收入结构优化

公司 DPU 产品向腾讯大规模销售及商用，已充分验证公司产品的技术成熟度、品质可靠性及生态兼容性，形成了显著的市场示范效应，为公司跨行业、跨场景拓展客户奠定了坚实基础。报告期内，公司已在电信运营商、AI 大模型公司、其他互联网厂商、金融及电力等行业的头部客户取得实质性突破，收入结构持续优化，具体情况详见本题回复之“一/（二）/4、公司的其他客户开拓情况”。

综上，报告期内公司对腾讯销售收入占比较高，系双方基于云网络架构升级、DPU 产品技术演进与价值提升、供应链自主可控需求所形成的战略性、市场化双向选择，符合 DPU 行业客观规律与各自发展阶段需要。同时，公司与腾讯保持长期稳定的战略合作关系，向腾讯销售具备较强的稳定性和可持续性。此外，依托与腾讯合作的示范效应以及产品技术成熟度和生态兼容性，公司已成功拓展多个其他领域的标杆客户，未来客户收入结构将持续优化。因此，报告期内公司对腾讯销售收入占比较高，不会对公司正常经营发展和持续经营能力产生重大不利影响。

（二）关联交易是否影响发行人的独立性

1、公司与腾讯及相关方合作的必要性及合理性

公司与腾讯及相关方之间的关联销售交易具有充分的必要性及合理性，具体体现在以下方面：

从采购方需求看，采购 DPU 是腾讯及相关方布局下一代数据中心架构、响应信息技术应用创新趋势的必要投入。在供应链安全日益重要的背景下，采购国产高性能 DPU 是降低对国际供应商依赖的关键战略举措。公司作为率先实现 400Gbps 高性能全功能 DPU 芯片技术突破与规模化商用的国产芯片厂商，产品

技术路线与性能对标国际主流水平，能够满足腾讯 AI 智算与云计算场景对算力基础设施的网络升级需求，是国内稀缺的高性能全功能 DPU 供应商。

从销售方发展看，公司 DPU 产品经头部厂商腾讯大规模商用，充分验证了产品的技术成熟度、品质可靠性及生态兼容性，有助于形成强大的市场示范效应。目前，公司已与头部电信运营商、AI 大模型公司、其他互联网厂商、金融及电力行业头部客户开展业务合作，客户拓展取得实质性进展。

从产业协同看，双方战略合作有助于深化产业链上下游协同，共同构建更安全、高效的国产算力生态，符合行业发展趋势与国家政策导向。

综上，公司与腾讯及相关方之间的关联销售交易具有必要性及合理性。

2、公司与腾讯及相关方合作的定价合理性

针对 DPU 芯片业务，公司向腾讯及相关方和其他客户销售的 DPU 芯片单价基本一致。相较 DPU 芯片，DPU 板卡需根据不同客户的服务器型号、接口规格、性能配置及网络生态需求进行设计和生产，是一种差异化的软硬件解决方案，因此产品单价不具备直接可比性。公司 DPU 板卡产品采用客户差异化议价机制，实行动态调价、友好协商的谈判原则，定价综合考虑市场目标、需求规模，以成本为基础，叠加合理预期毛利综合确定，该定价机制公允且具备商业合理性。

3、相关交易均已履行关联交易决策程序

公司现行有效的《公司章程》《股东会议事规则》《董事会议事规则》《关联交易管理制度》规定了股东会、董事会在审议有关关联交易事项时关联股东、关联董事回避表决制度及其他公允决策程序，且有关议事规则及决策制度已经公司股东会审议通过，符合有关法律、法规、规章和规范性文件的规定。

公司第一届董事会第八次会议、2026 年第三次临时股东会审议通过了《关于确认公司关联交易的议案》，对公司报告期内发生的需审议的关联交易予以确认，关联董事及关联股东回避表决。公司独立董事对上述关联交易议案发表了肯

定意见，认为该等事项内容合法、有效，符合公司和全体股东的利益，不存在损害中小股东的利益的情况。

综上，报告期内公司与腾讯及相关方的关联销售交易是基于行业特性和双方市场地位形成的正常商业合作，具有必要性、合理性和公允性，且均已履行必要的内部决策程序。公司在资产、人员、财务、机构等方面均独立于腾讯，双方的业务合作是基于市场化选择，该等关联交易不影响公司的独立性。

（三）进一步完善风险提示

发行人已在招股说明书“第三节 风险因素”之“（二）当前客户集中度较高及单一客户重大依赖的风险”“（三）关联销售占比相对较高的风险”“（四）市场拓展与客户多元化不及预期的风险”对上述事项进一步完善风险提示如下：

“（二）当前客户集中度较高及单一终端客户重大依赖的风险

公司 DPU 产品是算力基础设施的核心网络基座，广泛应用于 AI 智算、云计算、存储、通信、网络安全及边缘计算等关键战略领域，当前终端客户包括腾讯、中国移动、新浪、南方电网等。报告期内，公司向前五名客户的销售金额占当年营业收入的比例分别为 100.00%、99.35%和 92.87%，客户集中度较高，与公司所处的发展阶段和下游市场集中度特征相关；按终端客户统计，报告期内，公司向第一大终端客户腾讯的直接及间接销售金额占当年营业收入的比例分别为 0%、93.16%和 95.22%，占比较高。公司当前尚处于商业化爬坡期，采取与下游头部客户深度合作的“种子客户”战略符合科技创新企业的成长特征，通过与种子客户的大规模商用合作，有利于公司产品技术打磨以及面向全市场加速推广。公司与腾讯保持长期稳定的战略合作关系，向腾讯销售具备较强的稳定性和可持续性。

如果腾讯等主要客户的业务需求、生产经营情况、经营战略、技术路线、价格等商业条款发生重大不利变化，公司新产品及新技术不能继续得到腾讯等主要客户的认可等，则公司来自腾讯等主要客户的收入及毛利将可能受到影响，从而对公司的经营业绩产生不利影响。

（三）关联销售占比较高的风险

截至 2025 年 12 月 31 日，腾讯通过相关主体合计持有公司 19.78%的股权，为公司重要关联方之一。报告期内，公司对腾讯直接及间接销售金额占当期营业收入的比例分别为 0%、93.16%和 95.22%，占比较高。前述与公司所处的发展阶段和下游市场集中度特征相关。公司预计未来一段时期内对腾讯销售占比较高的情形仍将持续。

公司与腾讯保持了长期稳定的战略合作关系，并积极拓展新客户，非关联销售金额在报告期内呈现快速增长趋势。若未来新客户收入增速未能超过来自腾讯的收入增速，将造成关联销售占比保持较高水平的风险。

（四）市场拓展与客户多元化不及预期的风险

虽然公司产品在腾讯的规模化商用验证了其技术成熟度，并在技术创新、产品研发及市场拓展等方面取得了一定成绩，且已积极开拓多类下游头部客户并取得进展，但不同的大型云服务商和电信运营商往往拥有差异化数据中心网络架构和软件生态，对公司 DPU 产品的导入需要一定的适配周期，公司需要投入研发资源进行适配和优化。

若公司无法在短期内有效拓展其他头部客户，可能会导致客户结构多元化进度不及预期，一定程度上将影响公司的抗风险能力及市场布局，对公司长期经营业绩可能产生不利影响。”

中介机构核查程序及核查意见：

一、核查程序

保荐人、申报会计师履行了以下核查程序：

1、访谈发行人管理层，了解公司与腾讯的合作历史和合作模式；获取并查阅发行人与腾讯签署的合作协议、销售合同等文件，核查主要合同条款的具体约定；核查报告期内发行人对主要客户的销售情况，分析在手订单的变化趋势，判

断发行人与主要客户合作的稳定性、可持续性；

2、访谈发行人研发及产品负责人，了解公司 DPU 产品设计开发等内容，了解向腾讯提供服务过程中积累的技术经验情况及可复用性；

3、获取发行人销售明细表，对新老客户贡献情况进行汇总分析，分析报告期内主要客户数量、销售金额及占比变化情况；核查发行人与非关联客户的合作协议、产品认证进展及订单签署情况；

4、对发行人主要客户进行走访和函证，了解其与发行人的合作背景、当前合作情况和后续合作意愿，是否存在关联关系；

5、查阅行业研究报告等公开信息，了解行业发展现状、下游集中度情况、市场空间及未来发展趋势；查阅同行业可比公司的公开披露信息，与同行业可比公司的客户集中度情况进行对比分析；

6、对发行人的资产、人员、财务、机构等方面独立性进行核查，检查是否存在与客户混同的情形，分析发行人是否具备独立面向市场获取业务的能力。

二、核查意见

经核查，保荐人、申报会计师认为：

1、发行人与腾讯的业务合作具有稳定性，DPU 产品针对腾讯存在一定程度的适配性开发，但该等开发主要集中于场景及软硬件生态适配层面，而非芯片核心架构与底层技术，产品本身具有通用可编程性，且在腾讯数据中心应用过程中积累的技术经验与工程能力具备充分的可复用性。凭借前述产品特性，报告期内发行人已成功开拓中国移动、南方电网、欣诺通信、新浪等其他非关联客户。基于双方长期稳定的合作基础及当前产品合作与验证进展，预计发行人与腾讯及相关方的交易金额未来将继续增长，交易占比将呈现短期保持高位态势，长期下降的趋势；

2、发行人对腾讯的重大依赖符合 DPU 行业下游客户集中的客观规律及公司所处发展阶段的特征，不会对正常经营发展和持续经营能力产生重大不利影响。发行人与腾讯的关联交易是基于行业特性和双方市场地位形成的正常商业合作，具有必要性、合理性和公允性，不影响发行人的独立性。

问题 4.关于实际控制人与股东情况

申报材料显示：发行人实际控制人萧启阳、张学利合计控制发行人约 30% 的表决权，各方约定，意见不一致时，以萧启阳意见为准提起提案和进行表决。腾讯通过相关主体合计持有发行人约 20% 股权。

请发行人披露：

（1）结合云豹创芯、云豹创智两家持股平台的合伙协议以及关于普通合伙人、表决权行使的约定情况，分析实际控制人是否可以有效控制持股平台。

（2）结合发行人股权结构、创立及历史经营管理，与其他主要股东的协调机制或公司治理安排，各平台表决机制等情况，分析发行人实际控制权的稳定性。

请保荐人、发行人律师说明核查依据、过程，并发表明确意见。

发行人披露：

一、结合云豹创芯、云豹创智两家持股平台的合伙协议以及关于普通合伙人、表决权行使的约定情况，分析实际控制人是否可以有效控制持股平台。

（一）云豹创芯、云豹创智两家持股平台的合伙协议以及关于普通合伙人、表决权行使的约定情况

云豹创芯、云豹创智的合伙协议均约定由普通合伙人执行合伙事务，对外代表合伙企业。其他合伙人不再执行合伙事务，不得对外代表合伙企业。

云豹创芯、云豹创智由执行事务合伙人对合伙企业对外投资等事务进行决策代表合伙企业向被投资企业行使董事提名权、表决权等股东权利，该等安排的具体约定如下：执行事务合伙人负责合伙企业日常运营，对外代表合伙企业，拥有全权负责合伙企业及投资业务以及其他合伙事务之管理、运营、控制、决策的全部职权，包括但不限于代表合伙企业对股权投资项目进行管理，全权决定合伙企业股权投资项目的重大事项，代表合伙企业参加合伙企业股权投资项目的股东（大）会，行使表决权、董事提名权等股东权利，决定有限合伙人入伙、退伙、

当然退伙、合伙人财产份额转让、变更合伙企业的名称、地点、经营范围、经营期限等事宜。

（二）实际控制人是否可以有效控制持股平台

自 2023 年 1 月 1 日至本回复出具之日，云豹创芯的普通合伙人及执行事务合伙人均为张学利，云豹创智的普通合伙人执行事务合伙人均为萧启阳。报告期内，公司历次股东会均由张学利代表云豹创芯行使股东权利，由萧启阳代表云豹创智行使股东权利，相关安排持续稳定执行。

综上，依据执行事务合伙人的产生、职权范围及报告期内的实际运作情况，萧启阳、张学利拥有按合伙协议约定全权负责两家持股平台、投资业务以及其他合伙事务的管理、运营、控制、决策的全部职权，能够有效支配云豹创智、云豹创芯所持公司股份对应的股东权利，二人可以有效控制上述两家持股平台。

二、结合发行人股权结构、创立及历史经营管理，与其他主要股东的协调机制或公司治理安排，各平台表决机制等情况，分析发行人实际控制权的稳定性。

截至本回复出具之日，公司实际控制人为萧启阳、张学利，最近二年未发生变更，依据公司股权结构、创立及历史经营管理，与其他主要股东的协调机制或公司治理安排，各平台表决机制等情况，公司实际控制权稳定，具体分析如下：

（一）发行人股权结构

截至本回复出具之日，萧启阳直接持有公司 5.8195%的股权，通过担任云豹创智执行事务合伙人控制公司 11.0341%的表决权，合计控制公司 16.8536%的表决权；张学利直接持有公司 1.3324%的股权，通过担任云豹创芯执行事务合伙人控制公司 11.6247%的表决权，合计控制公司 12.9572%的表决权；萧启阳、张学利、云豹创智、云豹创芯签署了《一致行动协议》，明确了各方之间的一致行动关系和萧启阳、张学利共同控制公司的相关安排。萧启阳、张学利两人合计控制公司 29.8108%的表决权。

截至本回复出具之日，公司共有 113 名股东，股权结构分散。除深圳腾讯、云豹创芯、云豹创智、萧启阳外，公司其他股东持有的公司股份比例均未达到 5% 以上。腾讯通过深圳腾讯及其关联方广西腾讯、苏州湃益合计持有公司 19.7792% 股份，与实际控制人萧启阳、张学利共同控制的表决权比例差异为 10.0316%，差异较大。

深圳腾讯及其关联方广西腾讯、苏州湃益已出具《关于不谋求公司控制权的承诺函》，确认自腾讯相关主体入股公司以来，腾讯尊重萧启阳、张学利为公司共同实际控制人，并承诺自公司上市之日起 36 个月内不会单独或联合其他方以任何方式（包括但不限于增持、协议、合作、委托表决、关联方关系、一致行动关系、征集投票权等）扩大持股比例或表决权比例从而谋求公司的控制权。

综上，萧启阳、张学利直接和间接合计控制公司表决权比例远高于其他股东的表决权控制比例，萧启阳、张学利可以对股东会实现有效控制。

（二）创立及历史经营管理

公司由萧启阳、张学利于 2020 年 8 月共同设立，深圳腾讯、苏州湃益、广西腾讯分别于 2021 年 4 月、2021 年 9 月、2022 年 5 月以外部投资人身份通过增资成为公司股东。

报告期内，萧启阳一直担任云豹智能有限及公司的执行董事/董事长和经理，张学利一直担任云豹智能有限及公司的董事，且自股份公司设立以来一直担任股份公司副经理，萧启阳及张学利系公司日常生产经营的核心领导，二人主持并制定了公司整体发展规划和战略，主导了公司各项重大决策，对公司的重大经营方针、决策和经营管理层的提名及任免等拥有决定性的影响。公司高级管理人员、财务人员均无外部投资人股东委派的情形，且公司核心技术人员均系公司内部培养，亦不存在腾讯相关经历或背景，实际控制人可以控制公司日常经营管理。

（三）股东之间的协调机制

（1）一致行动安排

萧启阳、张学利、云豹创智、云豹创芯签署了《一致行动协议》，明确了各方之间的一致行动关系和萧启阳、张学利共同控制公司的相关安排。具体情况如下：

①《一致行动协议》明确约定了各方发生意见分歧或纠纷时的解决机制：萧启阳、张学利、云豹创智及/或云豹创芯作为公司股东及董事提名人，在向公司股东会或董事会提出议案或表决前，应事先沟通；意见一致时，以共同名义或一致结果提案和表决；若沟通后无法达成一致，则以萧启阳和张学利的意见为准，若二人意见不一致，则以萧启阳的意见为准提起提案和进行表决。

②《一致行动协议》对公司的实际控制人做了确认：萧启阳及张学利自公司设立至今一直为公司的实际控制人，共同对公司实施控制。

③《一致行动协议》对协议有效期做了约定：协议自各方签署之日起生效，至公司股票上市之日起满三十六（36）个月或者相关证券监管机构要求的更长期限届满之日终止，上述期限届满后，经各方协商一致可延长本协议的有效期。

（2）腾讯已出具关于不谋求公司控制权的承诺函

公司其他持股 5%以上的股东深圳腾讯及其关联方广西腾讯、苏州湃益均出具了《关于不谋求公司控制权的承诺函》，承诺自发行人上市之日起 36 个月内不会谋求公司的控制权。

（3）“一票否决权”的实际执行情况

深圳腾讯提名的董事虽享有董事会一票否决权，但该条款系投资人保护性权利安排，且仅限于特定事项。依据公司的《公司章程》《股东会议事规则》《董事会议事规则》等文件的约定，深圳腾讯无法据此单方面对发行人实现控制。报告期内，深圳腾讯从未在董事会、股东会实际行使“一票否决权”，未曾通过行使“一票否决权”等方式干预公司治理、日常经营管理。

（四）公司治理情况

公司实际控制人始终在董事会中占多数席位。自 2023 年 1 月 1 日至 2025 年 12 月 25 日，公司董事会由 3 名董事组成，萧启阳及张学利均占公司 2 名董事席位，腾讯占 1 名董事席位；自 2025 年 12 月 25 日至本回复出具之日，公司董事会由 7 名董事组成，其中独立董事 3 名，职工董事 1 名，剔除独立董事及职工董事，萧启阳、张学利合计控制的董事会席位已经超过相关董事席位的半数，两人对董事会的运作有着实际控制性影响。

报告期内，董事会相关决议均由全体董事一致表决通过，未出现其他董事提出反对意见的情形。腾讯及其提名董事均在股东会、董事会表决中投赞成票促成相关决议通过，未曾出现与实际控制人萧启阳、张学利决策不一致的情形。

公司 2026 年第三次临时股东会决议，公司股东一致确认萧启阳、张学利为公司实际控制人。

（五）各平台表决机制

如“问题 4/一、结合云豹创芯、云豹创智两家持股平台的合伙协议以及关于普通合伙人、表决权行使的约定情况，分析实际控制人是否可以有效控制持股平台。”部分所述，依据云豹创芯、云豹创智合伙协议约定及报告期内实际运行情况，萧启阳、张学利能够有效支配云豹创智、云豹创芯所持公司股权对应的股东表决权，二人可以有效控制上述两家持股平台。

综上，依据公司股权结构、创立及历史经营管理，与其他主要股东的协调机制或公司治理安排，各平台表决机制等情况，公司的实际控制人为萧启阳及张学利，且最近两年未发生变更，公司控制权稳定。

中介机构核查程序及核查意见：

一、核查程序

保荐人及发行人律师履行了以下核查程序：

1、取得并查阅了云豹创芯、云豹创智的《合伙协议》，了解其关于普通合伙人、表决权行使的约定情况；

2、取得并查阅发行人的股东名册，了解发行人股权结构；取得并查阅萧启阳、张学利、云豹创智、云豹创芯签署的《一致行动协议》，了解一致行动安排相关的约定及纠纷解决机制；

3、访谈萧启阳、张学利，了解公司创立及历史经营、公司治理情况；取得并查阅发行人报告期内的董事会、股东会文件，了解报告期内董事会、股东会的运行情况；取得了腾讯出具的《关于不谋求公司控制权的承诺函》；

4、取得发行人的说明及确认文件。

二、核查意见

经核查，保荐人、发行人律师认为：

1、依据云豹创芯、云豹创智执行事务合伙人的产生、职权范围及报告期内的实际运作情况，萧启阳、张学利拥有按合伙协议约定全权负责两家持股平台、投资业务以及其他合伙事务的管理、运营、控制、决策的全部职权，能够有效支配云豹创智、云豹创芯所持公司股份对应的股东权利，二人可以有效控制上述两家持股平台。

2、依据公司股权结构、创立及历史经营管理，与其他主要股东的协调机制或公司治理安排，各平台表决机制等情况，公司的实际控制人为萧启阳及张学利，且最近两年未发生变更，公司控制权稳定。

问题 5.关于收入与客户

申报材料显示：2025 年，发行人 DPU 板卡价格有所下降；在手订单金额由 18,706.96 万元下降为 5,748.29 万元。

请发行人披露：

（1）板卡产品的定价模式，DPU 板卡降价原因，在 DDR 及 SSD 等原材料价格上涨的情况下，板卡价格下降的原因、对毛利率的影响，分析原材料价格波动风险能否有效向下游传导。

（2）2025 年末在手订单大幅下滑的原因，结合期后最新销售收入及在手订单情况，除腾讯外其他新开拓客户期后收入贡献情况，分析收入增长的可持续性。

请保荐人、申报会计师说明核查依据、过程，并发表明确意见。

发行人披露：

一、板卡产品的定价模式，DPU 板卡降价原因，在 DDR 及 SSD 等原材料价格上涨的情况下，板卡价格下降的原因、对毛利率的影响，分析原材料价格波动风险能否有效向下游传导。

（一）板卡产品的定价模式，DPU 板卡降价原因。在 DDR 及 SSD 等原材料价格上涨的情况下，板卡价格下降的原因、对毛利率的影响

公司 DPU 板卡产品采用客户差异化、市场化定价机制，实行友好协商、动态调整的议价原则。定价综合考虑市场目标、需求规模，以成本为基础，叠加合理预期毛利综合确定，该定价机制公允且具备商业合理性。

2025 年度公司 DPU 板卡平均单价低于 2024 年度，主要系：2024 年，公司量产初期销售的 DPU 板卡系向中国移动、南方电网等客户的小批量送样，单卡价格较高；2025 年，公司与主要客户腾讯的业务合作调整为板卡交付的行业标准模式，当期向腾讯销售板卡 43,079 张，腾讯作为公司战略性种子客户，公司综合考虑其采购规模和公司自身的市场拓展策略给予略低售价。此外，公司 DPU

板卡需根据不同客户的服务器型号、接口规格、性能配置及网络生态需求进行设计和生产，该等软硬件解决方案具有差异化，因此产品单价不具备直接可比性。

在 DDR 及 SSD 等原材料价格上涨的情况下，公司 DPU 板卡单价金额保持上升态势，毛利率有所波动主要是受公司存储器件备货策略实施情况以及同客户商议调价节奏的影响，具有合理性，具体情况参见本题回复之“（二）分析原材料价格波动风险能否有效向下游传导”。

（二）分析原材料价格波动风险能否有效向下游传导。

公司自行购买 DDR、SSD 等原材料，能够控制原材料并承担原材料的价格波动风险。公司 DPU 板卡产品采用市场化定价机制，遵循友好协商、动态调整的议价原则，以成本为基础叠加合理预期毛利综合确定。基于上述定价模式，公司可通过与下游客户的动态议价，将原材料价格波动风险向下游传导，但传导时效性取决于公司与客户商务谈判的节奏；同时，公司结合对上下游市场行情的预判制定动态备货策略，以有效管控原材料成本。

二、2025 年末在手订单大幅下滑的原因，结合期后最新销售收入及在手订单情况，除腾讯外其他新开拓客户期后收入贡献情况，分析收入增长的可持续性。

（一）2025 年末在手订单大幅下滑的原因

公司于 2024 年 12 月 31 日在手订单金额为 18,706.96 万元，2025 年 12 月 31 日在手订单金额为 5,748.29 万元，下降原因系公司与腾讯的交易模式发生变动所致。

在公司首款芯片量产初期，受限于所处发展阶段及客观资源约束，公司曾短暂采取直接销售芯片的过渡性业务模式。后随着业务规模扩大、供应链体系成熟及资源储备完善，公司已回归行业主流的板卡销售业务模式。公司与腾讯在产品商业化后的合作模式可简要划分为两个阶段，具体情况如下：

时间	产品/服务	交易模式
2024 年 7 月 -2025 年 5 月	DPU 芯片	1.直接交易：腾讯直接向公司采购 DPU 芯片，提供给板卡加工商用于生产 DPU 板卡，后经服务器厂商集成为整机服务器；2.间接交易：板卡加工商指定采购代理商，向公司采购 DPU 芯片作为原材料并加工成板卡，销售给下游服务器厂商
2025 年 6 月 开始至今	DPU 板卡	间接交易：公司向服务器厂商（如客户 A、新华三信息技术有限公司、超聚变数字技术有限公司等）销售 DPU 板卡，后经服务器厂商集成为服务器

在直接销售芯片的过渡性业务模式下，由腾讯直接向公司下达一定周期内的芯片集中采购订单；而在销售板卡的模式下，腾讯向各服务器厂商动态下达采购订单及需求预测，各服务器厂商结合腾讯的需求与自身备货策略，再向公司下达短期采购订单。受上述业务模式转换影响，截至 2025 年 12 月 31 日，公司在手订单金额较上年末有所下降。

（二）结合期后最新销售收入及在手订单情况，除腾讯外其他新开拓客户期后收入贡献情况，分析收入增长的可持续性

1、期后最新销售收入及在手订单情况

公司 2023-2025 年度及 2026 年 1-6 月收入金额如下：

单位：万元

项目	2026 年 1-6 月 (未经审计)	2025 年度	2024 年度	2023 年度
收入	93,988.86	36,989.17	3,635.57	17.32

由上表可见，2026 年 1-6 月公司收入仍保持高速增长态势，截至 2026 年 6 月 30 日在手订单金额为 49,443.30 万元，在手订单充裕，为公司未来收入的持续增长提供了有力支撑。

2、除腾讯外其他新开拓客户期后收入贡献情况

经过数年深耕，公司已构建良好的应用生态和品牌影响力。公司 DPU 产品已在头部云服务商腾讯实现规模化导入、多场景商用；同时，公司也已成为中国移动、中国联通等国内主要电信运营商以及国产知名 AI 大模型厂商、互联网企

业、通信设备商等客户的重要合作伙伴，并在金融、电力等行业的关键客户拓展方面已取得进展、完成项目落地。通过与前述各领域头部客户的深度合作，公司已在产品适配性、生态构建及解决方案落地方面积累了经验，形成了相应的先发优势与竞争壁垒，有助于公司新客户开拓及导入。

DPU 作为数据中心基础设施的关键网络硬件，下游客户（尤其是头部云服务商、电信运营商等）对产品的技术验证、兼容性测试及场景适配要求严苛，验证周期通常较长。具体的验证周期根据客户行业、技术生态及产品具体需求有所不同，云服务商、电信运营商及大型金融机构的验证周期通常为 6-12 个月，AI 大模型公司、通信设备商的验证周期通常为 6-9 个月；完成验证至量产出货的周期通常为 3-6 个月。

目前，公司 DPU 产品已在头部云服务商腾讯完成大规模部署与多场景商用，并在电信运营商中国移动、国产知名 AI 大模型厂商、新浪、欣诺通信、锐捷网络、南方电网、工商银行、招商银行等多领域头部客户实现销售出货，同时积极推动公司全功能 DPU、AI DPU 两款产品在其他头部云服务、电信运营商、垂类行业客户的测试验证、量产导入，加速拓展下游市场与丰富客户结构，逐步构建“客户开拓-订单增长”的良性循环。

针对除腾讯外的其他客户，公司 2026 年 1-6 月收入已超过 2025 年度全年收入，收入规模稳步上升。

综上，2025 年末在手订单金额较上年末下降，主要系公司与腾讯的合作模式变化所致。从期后情况看，2026 年 1-6 月公司收入仍保持高速增长态势，截至 2026 年 6 月 30 日在手订单充裕。同时，公司积极拓展非腾讯客户并效果显著，其他领域客户订单贡献逐步提升，公司收入增长具备良好的可持续性。

中介机构核查程序及核查意见：

一、核查程序

保荐人、申报会计师履行了以下核查程序：

1、获取公司 2023-2025 年度及 2026 年 1-6 月收入成本表进行分客户、分季度、分产品的单价及毛利率对比分析；

2、获取公司 2023-2025 年度及 2026 年 1-6 月采购表，分析公司对存储器件等核心原材料的采购价格变化；

3、获取公司截至 2024 年末、2025 年末及 2026 年 6 月 30 日在手订单，分析在手订单变化情况及收入增长的可持续性。

二、核查意见

经核查，保荐人、申报会计师认为：

1、公司 DPU 板卡产品采用客户差异化、市场化定价机制，实行友好协商、动态调整的议价原则，定价机制公允且具备商业合理性；2025 年度公司 DPU 板卡平均单价低于 2024 年度主要系采购规模、公司自身的市场拓展策略及产品配置差异等因素综合影响；在 DDR 及 SSD 等原材料价格上涨的情况下，公司 DPU 板卡单价金额保持上升态势，毛利率有所波动主要是受公司存储器件备货策略实施情况以及同客户商议调价节奏的影响，公司可通过与腾讯等客户的商务谈判，将原材料价格波动风险在一定程度向下游客户传导；

2、公司 2025 年末在手订单金额较上年末下降，主要系公司与腾讯的合作模式变化所致。从期后情况看，2026 年 1-6 月公司收入仍保持高速增长态势，截至 2026 年 6 月 30 日在手订单充裕；同时，公司积极拓展非腾讯客户并取得积极进展，其他领域客户订单贡献逐步提升，公司收入增长具备良好的可持续性。

问题 6.关于毛利率

申报材料显示，2024 年产品量产初期 DPU 板卡主要向客户小批量送样，单价较高。2025 年发行人 DPU 板卡产品毛利率同比下滑 21.91 个百分点。

请发行人披露：

（1）2025 年 DPU 板卡毛利率大幅下滑的原因，结合 DPU 产品供求情况、期后销售毛利率情况，分析未来毛利率是否存在持续下滑风险。

请保荐人、申报会计师说明核查依据、过程，并发表明确意见。

发行人披露：

一、2025 年 DPU 板卡毛利率大幅下滑的原因，结合 DPU 产品供求情况、期后销售毛利率情况，分析未来毛利率是否存在持续下滑风险。

（一）2025 年 DPU 板卡毛利率大幅下滑的原因

2024 年及 2025 年，公司 DPU 板卡毛利率分别为 46.29%和 24.38%，变动较大，主要原因系：

1、2025 年和 2024 年度客户结构不同

2024 年，公司量产初期销售的 DPU 板卡主要向中国移动、南方电网等客户的小批量送样，单卡价格较高，整卡毛利率较高；2025 年，公司与主要客户腾讯的业务合作调整为板卡交付的行业标准模式，当期向腾讯销售板卡 43,079 张，腾讯作为公司战略性种子客户，公司综合考虑其采购规模和公司自身的市场拓展策略给予其较低的售价，致使 2025 年度公司的 DPU 板卡毛利率大幅下滑。

2、2025 年下半年上游存储原材料价格上行，抬升单位生产成本

2025 年下半年开始年全球存储器件进入持续涨价周期，存储类原材料为 DPU 板卡核心成本构成，原材料采购单价同比上涨直接增加单卡单位生产成本，但向下游客户的成本传导存在错期滞后，进而压缩 DPU 板卡产品盈利空间，加剧毛利率同比下滑幅度。

（二）结合 DPU 产品供求情况、期后销售毛利率情况，分析未来毛利率是否存在持续下滑风险。

2026 年 1-6 月公司 DPU 板卡毛利率为 23.70%（未经审计），较 2025 年略微下降，主要系 DDR、SSD 等存储配件价格进一步上涨所致。公司未来毛利率可能因原材料涨价仍然存在波动的风险，但是综合下游 DPU 市场供需发展情况、公司动态报价及备货策略、产品结构拓展等因素，公司长期毛利率水平具备提升基础。

1、DPU 产品市场需求增长、国产替代加速推动公司产业链地位提升

从需求端来看，DPU 作为释放 CPU 及 GPU 算力、加速网络互联、提升计算集群效能的数据中心基础设施关键组件，已广泛应用于 AI 智算、云计算、存储、通信、网络安全及边缘计算等关键战略领域。未来，随着全球新一代算力基础设施的规模化建设、人工智能与工业互联网等新兴领域需求增长，DPU 市场规模预计将持续扩大。根据弗若斯特沙利文数据，全球 DPU 市场规模由 2021 年的人民币 649.93 亿元增长至 2025 年的人民币 1,964.91 亿元，期间年复合增长率为 31.86%；预计到 2030 年将增长至 4,362.39 亿元，2026 年至 2030 年间的年复合增长率达 17.07%。中国 DPU 市场规模由 2021 年的 150.44 亿元增长至 2025 年的 497.58 亿元，期间年复合增长率达 34.86%；预计到 2030 年将增长至 1,290.91 亿元，是 AI 智算与云计算基础设施领域中增长最为迅速、成长空间最为广阔的细分市场之一。

从供给端来看，目前国内 DPU 市场主要被英伟达等海外大厂占据。从国产替代、供应链安全以及算力基础设施自主可控角度，国产 DPU 厂商迎来重要发展机遇，国内数据中心、电信运营商、云服务商均逐步加大对自研国产 DPU 产品的导入验证与批量采购力度，国产替代空间广阔。公司作为国内 DPU 行业头部企业，具备产品领先性与商业化卡位优势，已实现对头部云服务商、电信运营商、AI 大模型公司、电力及金融等多行业标杆客户批量供货，市场竞争力较强，依托下游需求持续增长与供应链国产化浪潮，公司于产业链的地位将持续增强，上下游议价能力稳步提升，进而提升产品毛利率。

2、动态备货与成本传导机制保障毛利空间

2025 年下半年以来，全球存储器件进入价格上行周期，导致公司 DPU 板卡产品单位材料成本有所上升。针对原材料价格波动，公司主要采取以下应对措施：

一是市场化定价机制下具备一定成本传导能力。公司 DPU 板卡产品采用市场化议价机制，以成本为基础叠加合理毛利综合定价，可结合主要原材料价格变动情况与客户协商动态调整阶段性价格，具备一定的向下游传导成本压力的能力。

二是实施动态备货策略平抑短期价格波动。公司结合对上下游市场供需格局的研判，实施灵活的库存管理与动态备货策略，合理控制原材料采购时点与规模，缓释存储器件价格短期波动对毛利率的冲击。

通过上述措施，公司能够在一定程度上减小原材料价格上涨的影响，降低毛利率持续下行的风险。

3、产品结构拓展成为驱动公司盈利能力持续提升的重要动力

公司第二款 AI DPU 芯片已流片成功并加速推向市场，预计 2026 年下半年形成规模化收入。相较第一款全功能 DPU 产品，公司第二款 AI DPU 定位于当前需求增长更快的 AI 智算集群后端网络市场。基于芯片架构简化带来的生产良率和产出率提升，以及 AI DPU 板卡无需集成较多高价值元器件，该款产品的定价弹性更高、毛利率有望增长，将成为驱动公司盈利能力持续提升的重要动力。

综上，短期内公司毛利率仍可能因存储元器件价格波动而存在阶段性波动风险；但长期来看，受益于下游 DPU 市场需求持续增长及国产替代加速推动公司产业链地位提升、公司市场化定价与动态备货策略的有效实施，以及第二款高毛利 AI DPU 芯片逐步量产落地等多重积极因素支撑，公司未来毛利率水平具备稳步提升的坚实基础。

中介机构核查程序及核查意见：

一、核查程序

保荐人、申报会计师履行了以下核查程序：

1、获取公司 2023-2025 年度及 2026 年 1-6 月收入成本表进行分客户、分产品毛利率对比分析；

2、获取公司 2023-2025 年度及 2026 年 1-6 月采购表，分析公司对存储器件等核心原材料的采购价格变化；

3、获取 DPU 产品相关的行业研究报告，分析 DPU 产品的市场供求情况。

二、核查意见

经核查，保荐人、申报会计师认为：

2025 年 DPU 板卡毛利率大幅下滑主要系客户结构变化与存储原材料成本上涨双重影响。短期内公司毛利率仍可能因存储元器件价格波动而存在阶段性波动风险；但长期来看，受益于下游 DPU 市场需求持续增长及国产替代加速推动公司产业链地位提升、公司市场化定价与动态备货策略的有效实施，以及第二款高毛利 AI DPU 芯片逐步量产落地等多重积极因素支撑，公司未来毛利率水平具备稳步提升的坚实基础。

问题 7.关于存货

申报材料显示：2025 年末，发行人存货账面余额同比增长，而在手订单同比下降。

请发行人披露：

（1）各类存货的具体构成与库龄情况，并结合各存货变动情况、原材料采购周期、在手订单及期后销售情况等情况，分析 2025 年末存货余额增长原因，期后销售情况，在手订单下降的情况下存货余额增长的原因。

（2）结合各报告期末主要产品销售价格、原材料采购价格、存货库龄情况等，分析存货跌价计提的充分性。

请保荐人、申报会计师说明核查依据、过程，并发表明确意见。

发行人披露：

一、各类存货的具体构成与库龄情况，并结合各存货变动情况、原材料采购周期、在手订单及期后销售情况等情况，分析 2025 年末存货余额增长原因，期后销售情况，在手订单下降的情况下存货余额增长的原因

（一）各类存货的具体构成与库龄情况

报告期各期末，公司存货构成如下：

单位：万元

存货类型	2025/12/31		2024/12/31		2023/12/31	
	金额	占比	金额	占比	金额	占比
原材料	3,652.12	21.60%	191.77	6.62%	38.82	3.73%
半成品	3,392.18	20.07%	628.92	21.73%		
委托加工物资	4,414.62	26.11%	145.91	5.04%	-	-
库存商品	4,703.73	27.82%	1,827.85	63.15%	188.26	18.07%
发出商品	466.81	2.76%	72.21	2.49%	-	-
合同履约成本	275.80	1.63%	27.99	0.97%	814.49	78.20%
合计	16,905.25	100.00%	2,894.66	100.00%	1,041.57	100.00%

报告期各期末，公司存货库龄如下：

单位：万元

库龄	2025/12/31		2024/12/31		2023/12/31	
	金额	占比	金额	占比	金额	占比
1 年以内 (含 1 年)	16,312.36	96.49%	2,780.12	96.04%	1,041.57	100.00%
1-2 年	478.35	2.83%	114.54	3.96%		
2-3 年	114.54	0.68%				
合计	16,905.25	100.00%	2,894.66	100.00%	1,041.57	100.00%

公司存货主要由原材料、半成品、委托加工物资和库存商品构成。报告期各期末，公司存货库龄主要集中在一年以内。

（二）2025 年末存货余额增长原因

1、2025 年末公司存货变动情况

2025 年末，公司存货余额为 16,905.25 万元，相较于 2024 年末增加 14,010.60 万元，主要系原材料、半成品、委托加工物资、库存商品合计增加 13,368.20 万元所致。2025 年末存货余额增长主要系腾讯 2025 年第四季度开始采购需求快速释放，公司加紧生产排期及备货所致。

2、原材料采购周期

公司的 DPU 芯片采购周期（从下单到交付）通常为 6 个月，SSD、DDR 内存条和 DDR 颗粒的采购周期（从下单到交付）一般为 1 个月左右，公司委托板卡加工厂将 DPU 芯片加工为 DPU 板卡的周期一般为 1-2 个月。

3、在手订单情况

在手订单情况请参见本回复之“问题 5/二/（一）2025 年末在手订单大幅下滑的原因”。

4、存货期后销售情况

2025 年末，公司存货的结转、销售情况如下：

单位：万元

存货类型	账面余额	结转/销售金额	结转/销售比例
原材料	3,652.12	3,103.33	84.97%
半成品	3,392.18	2,980.83	87.87%
委托加工物资	4,414.62	4,371.68	99.03%
库存商品	4,703.73	3,850.34	81.86%
发出商品	466.81	399.50	85.58%
合同履约成本	275.80	-	-
合计	16,905.25	14,705.67	86.99%

注：2025 年末存货的期后结转数据均截至 2026 年 6 月 30 日。

2025 年末公司整体存货余额期后结转/销售率为 86.99%，存货期后结转/销售情况良好。其中，2025 年末公司库存商品期后销售比例为 81.86%，主要系库存商品中存在部分因产品型号停售等原因的呆滞品 838.62 万元所致，公司已针对该部分存货全额计提了跌价准备；2025 年末公司发出商品期后销售比例为 85.58%，主要系工商银行向公司采购 DPU 板卡，由于工商银行原合作的服务器厂商中途退出，导致项目整体进度较预期延后所致。

5、2025 年末存货余额增长的主要原因

公司 2025 年末存货余额增长，主要系公司主要客户腾讯自 2025 年第四季度开始采购需求快速释放，为保障交付安全，公司加紧生产排期及备货，致使期末原材料、半成品、委托加工物资与库存商品余额均明显提高。

截至 2026 年 6 月 30 日，2025 年末公司存货余额期后结转/销售率为 86.99%，存货期后结转、销售情况良好，2025 年末存货余额增长具有合理性。

（三）在手订单下降的情况下存货余额增长的原因

1、2025 年末在手订单下降的情况

公司于 2024 年 12 月 31 日在手订单金额为 18,706.96 万元，2025 年 12 月 31 日在手订单金额为 5,748.29 万元，下降原因系公司与腾讯的交易模式发生变动所致。

在公司首款芯片量产初期，受限处于所处发展阶段及客观资源约束，公司曾短暂采取直接销售芯片的过渡性业务模式。后随着业务规模扩大、供应链体系成熟及资源储备完善，公司已回归行业主流的板卡销售业务模式。公司与腾讯在产品商业化后的合作模式可简要划分为两个阶段，具体情况如下：

时间	产品/服务	交易模式
2024 年 7 月 -2025 年 5 月	DPU 芯片	1.直接交易：腾讯直接向公司采购 DPU 芯片，提供给板卡加工商用于生产 DPU 板卡，后经服务器厂商集成为整机服务器；2.间接交易：板卡加工商指定采购代理商，向公司采购 DPU 芯片作为原材料并加工成板卡，销售给下游服务器厂商
2025 年 6 月 开始至今	DPU 板卡	间接交易：公司向服务器厂商（如客户 A、新华三信息技术有限公司、超聚变数字技术有限公司等）销售 DPU 板卡，后经服务器厂商集成为服务器

在直接销售芯片的过渡性业务模式下，由腾讯直接向公司下达一定周期内的芯片集中采购订单；而在销售板卡的模式下，腾讯向各服务器厂商动态下达采购订单及需求预测，各服务器厂商结合腾讯的需求与自身备货策略，再向公司下达短期采购订单。受上述业务模式转换影响，截至 2025 年 12 月 31 日，公司在手订单金额较上年末有所下降。

2、在手订单下降的情况下存货余额增长的原因

2025 年 6 月开始，公司主要采取以板卡形态出货的行业标准模式，公司下游主要客户为服务器厂商，由腾讯向各服务器厂商动态下达采购订单及需求预测，各服务器厂商结合腾讯的需求与自身备货策略，再向公司下达短期采购订单。同时，由于公司 DPU 板卡所需的 DPU 芯片整体生产周期较长，且公司需要向供应

商提前排产备货，公司备货量通常大于在手订单。

2025 年末，公司结合销售规模快速增长的预期，为保障交付安全，公司加紧生产排期及备货，致使期末原材料、半成品、委托加工物资与库存商品余额均明显提高。截至 2026 年 6 月 30 日，2025 年末公司存货余额期后结转/销售率为 86.99%，存货期后结转、销售情况良好。综上，公司在 2025 年末在手订单下降的情况下存货余额增长具有合理性。

二、结合各报告期末主要产品销售价格、原材料采购价格、存货库龄情况等，分析存货跌价计提的充分性

（一）存货跌价准备计提政策

公司存货跌价准备计提政策为：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。

（二）存货库龄情况

报告期各期末，公司按库龄区分的存货跌价准备计提情况如下：

单位：万元

年度	库龄	账面余额	存货跌价准备	计提比例
2025/12/31	1 年以内	16,312.36	887.53	5.44%
	1-2 年	478.35	313.94	65.63%
	2 年以上	114.54	114.54	100.00%
	合计	16,905.25	1,316.01	7.78%

年度	库龄	账面余额	存货跌价准备	计提比例
2024/12/31	1 年以内	2,780.12	192.25	6.92%
	1-2 年	114.54	114.54	100.00%
	2 年以上	-	-	-
	合计	2,894.66	306.79	10.60%
2023/12/31	1 年以内	1,041.57	-	-
	1-2 年	-	-	-
	2 年以上	-	-	-
	合计	1,041.57	-	-

报告期各期末，公司存货账面余额分别为 1,041.57 万元、2,894.66 万元和 16,905.25 万元，其中库龄 1 年以上的存货余额分别为 0.00 万元，114.54 万元和 592.89 万元，占比分别为 0.00%、3.96%和 3.51%，库龄 1 年以上的存货规模较小。报告期内，公司对库龄 1 年以上的存货计提的存货跌价准备金额为 0.00 万元，114.54 万元和 428.48 万元，占库龄 1 年以上存货的账面余额比例分别为 0.00%、100.00%、72.27%。

（三）存货跌价计提的充分性

1、原材料

报告期内，公司主要原材料平均采购单价总体呈波动上升趋势，期后 2026 年 1-6 月主要原材料平均采购单价有进一步较大幅度的提升，主要受存储市场供求关系影响。报告期各期末，公司原材料库龄主要集中在一年以内。公司的原材料主要为板卡配件材料，系为板卡加工而采购，一般不会对外直接销售。故对于原材料，公司系以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，与原材料对应的成本进行比较，确定存货跌价准备。

报告期各期末，公司针对原材料计提的存货跌价准备分别为 0.00 万元、7.28 万元、9.67 万元，主要系少量原材料因对应生产的板卡型号停售、借出物料无转

销机会而计提存货跌价准备。其余各项原材料可被领用至生产与发行人主营业务相关的各类产品，相关原材料加工为产成品后的估计售价减去预计将要发生的完工成本、估计税费、销售费用等后的可变现净值显著高于存货成本，无需计提存货跌价准备。

2、半成品、委托加工物资、库存商品、发出商品

报告期内，公司 DPU 芯片各期平均单价基本保持稳定。2025 年 DPU 板卡平均单价较 2024 年有所下降，主要受产品与客户结构变动影响，同时系公司 DPU 板卡业务由小批量送样阶段进入规模化量产阶段所致。2026 年 1-6 月，DPU 板卡销售单价增长主要系公司因存储相关原材料价格上涨同步上调产品售价；DPU 芯片销售单价下降主要系向客户欣诺通信销售 DPU 芯片用于 NPB 网络安全分流设备，公司基于该特定场景简化了部分 DPU 功能配置，因此产品定价相对略低。报告期各期末，公司半成品、委托加工物资、库存商品、发出商品等存货库龄主要集中在一年以内。

报告期各期末，公司针对半成品、委托加工物资、库存商品、发出商品计提的存货跌价准备分别为 0.00 万元、299.52 万元、1,306.34 万元，主要系对部分停售板卡产品、不良品计提存货跌价准备。其余半成品、委托加工物资加工为产成品后的估计售价减去预计将要发生的完工成本、估计税费、销售费用等后的可变现净值显著高于存货成本，其余库存商品、发出商品的估计售价减去估计的销售费用和相关税费后的可变现净值显著高于存货成本，无需计提存货跌价准备。

3、合同履约成本

报告期各期末，公司的合同履约成本按照合同售价减去预计将要发生的完工成本、估计税费、销售费用等后的可变现净值显著高于存货成本，无需计提减值准备。

综上所述，公司已根据《企业会计准则》的要求，充分评估了相关存货的减值风险，并计提了相应的存货跌价准备，存货跌价准备计提充分。

中介机构核查程序及核查意见：

一、核查程序

保荐人、申报会计师履行了以下核查程序：

1、获取发行人存货及库龄明细表、发行人原材料采购周期、在手订单情况，检查期后销售情况，分析各存货变动情况、2025 年末存货余额增长原因、在手订单下降的情况下存货余额增长的原因；

2、获取报告期各期主要产品销售价格、原材料采购价格、存货库龄情况，分析存货跌价计提的充分性；

3、获取发行人期后订单价格及原材料市场价格，了解发行人存货减值的测试方法及存货可变现净值的测算过程，检查存货跌价准备计提的金额及依据是否准确合理。

二、核查意见

经核查，保荐人、申报会计师认为：

1、公司存货主要由原材料、半成品、委托加工物资和库存商品构成，报告期各期末存货库龄主要集中在一年以内。2025 年末，公司存货余额增长主要系腾讯 2025 年第四季度开始采购需求快速释放，公司加紧生产排期及备货所致，期后结转/销售情况良好。公司在 2025 年末在手订单下降的情况下存货余额增长，主要系结合销售规模快速增长的预期，为保障交付安全，公司加紧生产排期及备货，具有合理性；

2、报告期各期末，公司存货库龄主要集中在一年以内，公司已根据《企业会计准则》的要求，结合期后订单价格及原材料市场价格充分评估了相关存货的减值风险，相关存货跌价准备计提充分、准确。

(本页无正文，为深圳云豹智能股份有限公司《关于深圳云豹智能股份有限公司首次公开发行股票并在创业板上市申请文件的审核问询函之回复报告》之盖章页)



深圳云豹智能股份有限公司

2026 年 8 月 6 日

发行人董事长声明

本人已认真阅读深圳云豹智能股份有限公司本次审核问询函回复全部内容，确认本回复不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担相应的法律责任。

董事长：


萧启阳


深圳云豹智能股份有限公司
2026年8月6日

（本页无正文，为中信证券股份有限公司《关于深圳云豹智能股份有限公司首次公开发行股票并在创业板上市申请文件的审核问询函之回复报告》之签章页）

保荐代表人：

彭捷

彭捷

王纯然

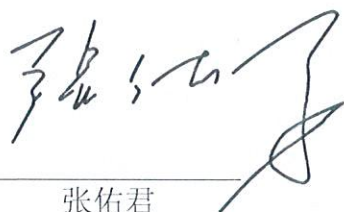
王纯然



保荐人董事长、法定代表人声明

本人已认真阅读深圳云豹智能股份有限公司本次审核问询函回复的全部内容，了解文件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，本次审核问询函的回复不存在虚假记载、误导性陈述或重大遗漏，并对上述文件的真实性、准确性、完整性承担相应的法律责任。

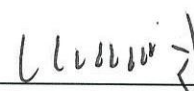
董事长：


张佑君

本所及经办律师已阅读《关于深圳云豹智能股份有限公司首次公开发行股票并在创业板上市申请文件的审核问询函之回复报告》（下称“回复报告”），确认回复报告与本所出具的法律意见书无矛盾之处。本所及经办律师对发行人在回复报告中引用的法律意见书的内容无异议，确认回复报告不致因上述内容而出现虚假记载、误导性陈述或重大遗漏。

北京市中伦律师事务所（盖章）
负责人： 

张学兵

经办律师： 

郭晓丹

经办律师： 
石 璁

2026 年 8 月 6 日

(此页无正文，为天健会计师事务所(特殊普通合伙)关于《关于深圳云豹智能股份有限公司首次公开发行股票并在创业板上市申请文件的审核问询函之回复报告》之会计师签章页，我们仅对审核问询函中需要会计师进行核查的事项发表核查意见)

签字注册会计师：

龙琦 之龙
印琦

龙 琦

李凤 之李
印凤

李 凤

天健会计师事务所(特殊普通合伙)

二〇二六年八月六日

