

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

第二轮审核问询函之回复报告

保荐人（主承销商）



中信证券股份有限公司
CITIC Securities Company Limited

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 8 月 20 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第二轮审核问询函》（审核函〔2026〕010155 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 DPU 产品与数据中心建设更新	4
问题 2. 关于云服务商等客户的 DPU 产品需求	21
问题 3. 关于客户集中	45
问题 4. 关于尚未盈利与前瞻性信息	61

问题 1. 关于 DPU 产品与数据中心建设更新

申报材料及首轮问询回复显示：目前服务器仍是数据中心建设中最核心的投资环节,全功能 DPU 及 AI DPU 在相关客户资本支出中的占比可达 4.5%左右。

请发行人披露：

（1）云服务商等客户每年新增相关资本支出、存量服务器改造等因素对 DPU 市场渗透率的影响，如资本支出缩减对 DPU 行业及发行人业绩的影响，并进行风险提示。

（2）预期存量服务器及近年新建数据中心批量更新迭代的周期，发行人 DPU 替换进入存量算力中心的可行性，DPU 产品替换特性、服务器批量更新预期等情况对发行人未来业绩的影响，并进行风险提示。

请保荐人说明核查依据、过程，并发表明确意见。

【回复】

一、云服务商等客户每年新增相关资本支出、存量服务器改造等因素对 DPU 市场渗透率的影响，如资本支出缩减对 DPU 行业及发行人业绩的影响，并进行风险提示

（一）云服务商等客户每年新增相关资本支出、存量服务器改造等因素对 DPU 市场渗透率的影响

1、云服务商等客户资本支出规模与增长趋势

DPU 作为数据中心基础设施的核心芯片，其市场需求与云服务商、电信运营商及 AI 大模型公司等下游客户的资本支出规模高度相关。近年来，在全球 AI 产业浪潮驱动下，下游客户持续加大数据中心等算力基础设施建设投入，资本支出规模显著增长，为 DPU 市场扩张提供了强劲的需求支撑。

（1）全球云服务商、大模型厂商资本支出规模持续扩大

根据 TrendForce 于 2026 年 8 月发布的最新数据，2026 年全球九大云服务提供商（谷歌、亚马逊、Meta、微软、甲骨文，及字节跳动、腾讯、阿里、百度）合计资本支出规模将突破 8,867 亿美元，同比增长接近 90%。其中，北美五大超大规模云厂商合计占比近 90%，持续引领全球 AI 基础设施投资。展望 2027 年，

TrendForce 预估九大云服务商总资本支出规模将扩大至约 1.3 万亿美元，AI 基础设施投资依然保持强劲态势。TrendForce 同时上调 2026 年 AI 服务器出货年增长率至近 31%（此前预期为 28%），反映出 AI 算力基础设施建设的加速态势。

除上述传统云服务商外，头部 AI 大模型公司亦在大规模自建或共建算力基础设施，成为全球 AI 资本支出的重要增量来源，具体情况如下：

OpenAI: 2025 年 1 月，OpenAI 联合软银、甲骨文等宣布拟在未来四年内投资约 5,000 亿美元用于 AI 基础设施建设，其中 2025 年首批部署规模 1,000 亿美元。2026 年，OpenAI 完成 1,220 亿美元融资，资金重点投向定制芯片、全球数据中心建设及模型研发。

Anthropic: 2026 年 2 月，Anthropic 完成 300 亿美元 G 轮融资，此前曾宣布投资 500 亿美元在美国建设自有 AI 数据中心。2026 年 5 月，Anthropic 完成 650 亿美元 H 轮融资。据报道，2026 年其训练与推理基础设施投入预计合计约 190 亿美元。

xAI(Grok): 其 Colossus 超级计算机于 2024 年 9 月建成 10 万卡集群，2024 年 12 月扩容至 20 万卡，2026 年初进一步扩建至约 55.5 万块英伟达 GPU。根据 SpaceX 美股上市文件 Form S-1 披露，xAI 的 2025 年度资本支出达 127 亿美元，2026 年第一季度达 77 亿美元，据此年化测算 xAI 的 2026 年度全年资本支出约 308 亿美元。

头部 AI 大模型公司已从单纯租用云算力转向大规模自建、共建算力基础设施，其资本支出规模已进入百亿至千亿美元量级，且大部分投向 AI 训练与推理集群，构成了独立于传统云服务商之外的又一重要 AI 基础设施投资主体。

（2）中国云服务商、电信运营商、AI 大模型厂商资本支出增长显著

中国云服务商及电信运营商的资本支出增长尤为突出。根据 TrendForce 数据，2026 年字节跳动、腾讯、阿里巴巴、百度的合计资本支出将同比增长逾 80%。根据伯恩斯坦 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%。主要客户的具体资本支出情况如下：

云服务商方面：1）根据美银证券报告，腾讯 2026-2028 年资本支出预计将分别达到 1,850 亿元、2,250 亿元和 2,500 亿元，呈持续增长态势。腾讯 2026 年

全年资本支出同比显著上行，下半年加速超级算力节点布局，加大国产 AI 芯片适配与规模化应用。2) 根据 TrendForce 资讯，字节跳动 2026 年资本支出将超过 2,000 亿元，年内多次上调资本支出计划，重点新建大规模 AI 智算中心、扩容 GPU 集群。3) 根据大和证券研究，百度 2026 年资本支出预计将达到 230 亿元，持续加码自有智算中心建设。4) 阿里巴巴公告计划未来三年投入超过 3,800 亿元用于建设云和 AI 硬件基础设施，2025 财年资本支出已达 1,261 亿元，同比增长显著。

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 年算力网络资本支出预计 378 亿元；中国联通 2026 年算力相关资本支出预计约为 175 亿元。

国产 AI 大模型公司亦在加速算力基础设施建设，持续扩大训练与推理算力集群规模，为公司 DPU 产品提供了新的增长空间。根据 OFweek 人工智能网资讯，DeepSeek 于 2026 年 6 月完成 500 亿元融资，本次融资主要用于算力基础设施建设支出。据其官方招聘信息，DeepSeek 正从租用云算力转向自建/共建万卡级智算中心；智谱于 2026 年完成 314 亿港元配售融资，投入基座模型研发和算力基础设施建设，境内 AI 大模型公司整体处于算力投入高速扩张期。

(3) DPU 在客户资本支出中的价值量占比

全功能 DPU 及 AI DPU 主要用于数据中心建设中的服务器及相关网络设备。根据美银证券、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 AI DPU 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 AI DPU 在相关客户资本支出中的占比可达 4.5%左右。综上，估算公司合作的几家云服务商、电信运营商、AI 大模型公司 2026 年 DPU 采购需求合计超过 210 亿元，为公司 DPU 产品销售提供了广阔的市场空间。

2、DPU 市场渗透率现状及提升路径

(1) DPU 市场渗透率现状

DPU 作为数据中心“第三颗主力芯片”，其市场渗透率仍处于快速提升阶段。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，是算力基础设施领域中增长最为迅速的细分市场之一。2025 年，中国存量服务器中 DPU 渗透率超过 20%，预计 2030 年将突破 50%。从新增服务器的角度看，2025 年新增服务器中 DPU 预装率已达 40%，预计 2030 年将突破 80%。

分服务器类型看，国内通算服务器侧 DPU 渗透率预计由 2025 年约 25%提升至 2030 年 70%，而 AI 服务器侧 DPU 渗透率预计由 2025 年约 98%提升至 2030 年约 100%，英伟达也在 2026 年 CES 大会上提出“DPU 是每个计算节点的标准配置”。从细分场景看，DPU 渗透率呈现分化特征，头部互联网及 AI 智算服务器 DPU 渗透率提升速度更快，通用企业私有云、传统存储服务器渗透率提升速度相对平滑，增量空间巨大。

(2) DPU 市场渗透率提升的驱动因素

国内 DPU 市场渗透率的提升主要受以下因素驱动：

第一，AI 智算集群建设加速。AI 大模型训练与推理对网络互联性能的要求持续提升，DPU 作为 GPU 集群网络核心枢纽，在万卡乃至十万卡集群中已成为标配组件。随着 AI 推理及 Agent 应用场景的快速扩展，推理侧算力基础设施对 DPU 的需求进一步释放，驱动 DPU 渗透率在 AI 智算场景中快速提升。

第二，云原生架构演进。云计算从传统架构向“CPU+GPU+DPU”三大主力芯片协同的新型架构演进，DPU 作为基础设施卸载层的价值在多租户、虚拟化及云原生场景中不可替代。随着云服务商加速数据中心架构升级，DPU 在新建设施中的配置率持续提升。

第三，技术标准成熟与生态完善。DPU 产品形态自 2020 年前后逐步成熟，PCIe Gen5、CXL 3.0、UCIe 等互联标准的量产验证推动 DPU 与主流 CPU、GPU 硬件的适配日趋完善，降低了客户导入门槛，加速渗透率提升。

第四，国产替代加速。当前中国 DPU 市场中，国外厂商占据较高市场份额，在国产替代政策推动及供应链安全需求驱动下，国产 DPU 渗透率有望在行业整体渗透率提升的基础上加速提升，为公司提供了超越行业平均增速的增长机遇。

3、资本支出对 DPU 渗透率的驱动机制分析

云服务商等客户的资本支出与 DPU 市场渗透率之间存在多重传导机制，具体分析如下：

（1）资本支出的同一性：存量、增量需求统一体现为资本支出投入

“存量数据中心更新”与“新数据中心建设”并非两类割裂市场。虽然理论上 DPU 硬件支持在终端客户已有服务器上热插拔和单独更换，但这在商业实践中并不具备普适性，旨在引入 DPU 以提升算力效率的举措通常依托于全新的服务器采购。因此，存量改造的需求并不会独立于新建需求存在，而是直接转化为新增服务器的采购订单。无论是为了扩大产能而建设新数据中心，还是为了技术升级而替换老旧设备，在财务计量上，其资金投入最终均体现为当年数据中心基础设施的新增资本支出投入。二者共同构成了云服务商年度硬件采购的总预算池。

（2）资本支出总量决定 DPU 市场需求

当前，新增服务器 DPU 预装率呈增长趋势，且与未来 DPU 市场渗透率正相关。未来 DPU 市场渗透率提升高度依赖于云服务商及互联网巨头、大模型厂商每年新增资本支出的景气度。资本支出总量直接决定了当年新增服务器采购规模，在 DPU 渗透率仍处于快速提升阶段的背景下，资本支出总量的增长不仅扩大了 DPU 市场的绝对规模，同时为渗透率的持续提升提供了更多的配置窗口。

根据弗若斯特沙利文报告，中国 DPU 市场的下游需求主要来自云服务商、电信运营商及其他行业，其中云服务商亦是最核心、规模最大的下游需求来源，2025 年需求占比约 57%，电信运营商需求占比约 21%。具体而言，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 2025 年贡献的 DPU 市场规模约 104.49 亿元，预计到 2030 年将以 19.83% 的年复合增长率增长至 258.18 亿元。

（3）资本支出结构影响 DPU 渗透速度

除资本支出总量外，资本支出的结构亦是影响 DPU 渗透率的重要因素。当资本支出更多投向 AI 智算基础设施时，由于 AI 服务器 DPU 配置率高于通算服务器，如前文所述，2025 年 AI 服务器 DPU 渗透率超 90%，通算服务器约 25%，DPU 渗透率将加速提升。根据 TrendForce 数据，2026 年全球九大云服务商资本

支出同比增长接近 90%，其中 AI 相关投资占比持续提升，这一结构性特征有利于 DPU 渗透率的加速提升。反之，若资本支出结构向非 AI 领域倾斜，如传统 IT 基础设施维护、非算力类资本支出等，则 DPU 渗透率提升速度可能放缓。但从当前行业趋势看，AI 算力投资在云厂商资本支出中的占比持续上升，DPU 受益于结构性配置需求的趋势明确。

4、存量服务器改造对 DPU 市场渗透率的影响

存量服务器更新改造将推动 DPU 市场渗透率进一步提升。由于 DPU 主要于近年逐步实现规模化应用，较早部署的服务器中 DPU 预装率相对较低，尤其在通用计算领域，较多存量服务器仍采用普通网卡等传统网络方案。

数据中心服务器通常具有约 3 至 5 年的使用周期，随着已运行较长时间的存量服务器陆续进入更新换代周期，客户通常会根据当前网络架构及业务升级需求重新配置服务器网络方案，并更多采用 DPU 产品。同时，随着数据中心网络带宽持续向更高速率演进，以及 AI 计算、云计算等场景对高带宽、低时延及数据处理能力的要求不断提升，新一代服务器对 DPU 的配置需求进一步增加。因此，随着原有 DPU 渗透率较低的存量服务器逐步更新并更多采用 DPU 方案，DPU 在整体服务器市场中的配置比例将持续提升，从而推动 DPU 市场渗透率进一步提高。

需要特别说明的是，数据中心服务器及配套网络设备并非短期一次性替换。虽然理论上 DPU 硬件支持在终端客户已有服务器上热插拔和单独更换，存量服务器升级 DPU 网卡在技术上可行，但在商业实践中并不具备普适性。大型数据中心通常提前半年以上进行服务器硬件选型和备货，新 DPU 部件的导入窗口通常锁定在新一轮服务器代际更替的规划节点上下。下游客户（尤其是大型云服务商）通常遵循严格的硬件生命周期管理策略。除非 DPU 本身发生物理故障，否则客户通常不会仅仅为了追求算力效率提升而在服务器生命周期中途单独更换 DPU 板卡。因此，存量服务器改造需求并不会独立于新建需求存在，而是直接转化为新增服务器的采购订单。

存量服务器改造对 DPU 渗透率的影响主要体现在存量服务器到期更替时新一代服务器普遍配置 DPU（反映为新增服务器中的 DPU 预装率），直接提升

DPU 整体渗透率。该等改造通常与服务器代际更替同步进行，其传导路径最终统一体现为下游客户年度资本支出投入。DPU 渗透率的提升速度在根本上取决于下游客户前述资本支出的总量及结构。

（二）如资本支出缩减对 DPU 行业的潜在影响

当前，中国云服务商、电信运营商、大模型厂商的资本支出仍处于加速增长阶段。根据 TrendForce 数据，2026 年字节跳动、腾讯、阿里巴巴、百度的合计资本支出同比增长逾 80%。若极端情况下，下游客户资本支出出现缩减，对 DPU 行业的影响需从总量效应和结构效应两个维度分析：

1、DPU 市场规模增速可能放缓

由于 DPU 市场规模的增量直接来源于下游客户资本支出中用于服务器及网络设备采购的部分，若资本支出总量缩减，DPU 市场的绝对增量空间将相应收窄、可能面临增量需求收缩的压力。然而，由于 DPU 整体渗透率仍处于低位快速提升阶段，渗透率提升本身可在一定程度上对冲资本支出总量增速放缓的影响。即使资本支出总量零增长，DPU 渗透率提升，仍将带来增量市场空间。短期内 DPU 市场的增长韧性可能高于资本支出总量的波动幅度。

2、DPU 相对配置优先级可能提升

在资本支出总量受限的情境下，云厂商倾向于优化投资结构，将有限预算投向投资回报率更高的环节。DPU 作为提升 CPU、GPU 利用率和集群综合算力效率的关键组件，其投入产出比优于单纯增加 CPU、GPU 数量。根据英伟达与 F5 的 2026 年联合测试数据，通过英伟达 BlueField-3 DPU 卸载可实现 CPU 开销降低 80%；根据中国电信天翼云数据，以 1,000 台服务器规模为例，部署 DPU 后硬件采购成本降低约 1.2 亿元、年电费减少约 400 万元。

据此，在资本支出总量收紧时，云厂商可能削减低端服务器或非 AI 基础设施的采购预算，但倾向于保持甚至增加 DPU 配置比例，以在不增加 CPU、GPU 采购的前提下提升集群整体算力效率。该等结构效应使得 DPU 在资本支出缩减环境下的抗风险能力可能优于高价值量芯片。

3、行业竞争格局可能加速分化

资本支出增速放缓可能加速 DPU 行业竞争格局分化。在市场增量收窄的环境下，具备技术领先性、产品性价比优势及已实现规模化商用的厂商将获得更高的市场份额，而处于早期阶段或缺乏差异化竞争力的厂商面临更大的生存压力。对于公司而言，作为已率先实现 400Gbps 全功能 DPU 规模化商用、在头部客户完成大规模部署的国产独立厂商，在行业竞争加剧的环境中具备先发优势和规模效应。

（三）如资本支出缩减对公司业绩的潜在影响及有关风险提示

虽然 DPU 在 AI 智算场景中的刚需属性使其配置优先级较高，DPU 整体渗透率仍处于快速提升阶段，可部分对冲资本支出总量下降的影响，公司产品在国产 DPU 市场具备领先地位和先发优势，在国产替代加速趋势下，公司市场份额有望持续提升，但是在极端情况下，云服务商等客户的资本支出增速不及预期甚至缩减资本支出导致该等客户降低其 DPU 采购规模，将对公司业绩造成不利影响，使得公司收入和毛利率低于预期。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，上述预测性信息为公司管理层基于公司经营规划及市场情况做出预测，不构成公司的盈利预测或业绩承诺

公司已在招股说明书“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“二、与行业相关的风险”中分析并补充披露了“下游 AI 智算与云计算基础设施投资周期波动的风险”，具体如下：

“公司产品主要面向 AI 智算与云计算等领域。下游云服务商、电信运营商等客户的资本开支、服务器升级换代、算力部署与网络架构调整受宏观经济、行业景气度、政策导向、技术路线与竞争格局等多因素影响，存在周期性波动。当

前全球 AI 智算及云计算基础设施建设正处于高速扩张期，然而云服务商资本支出呈现周期性波动特征，AI 驱动的资金支出高速增长存在周期性回调风险。

若全球或国内云服务商、电信运营商资本支出增速放缓或进入调整期，公司产品下游订单可能面临增速下降甚至绝对额缩减；若云服务商、电信运营商在资本支出收缩期加强对供应商的成本管控和议价压力，可能导致公司产品单价下降、毛利率收窄；若下游需求突然放缓，公司基于景气期需求预期预订的产能和建立的库存可能面临利用不足或减值风险，将进一步影响盈利能力。”

（四）公司的抗风险能力与极端情况下应对措施

即便考虑上述行业波动，公司仍具有一定的抗风险能力。截至 2026 年 6 月 30 日，公司货币资金、交易性金融资产、一年内到期的非流动资产及大额存单本息合计达 22.49 亿元，资金储备相对充裕，能够充分满足当前研发投入、业务拓展及日常运营的资金需求；公司 2026 年 1-8 月实现营业收入 13.08 亿元（未经审计），截至 2026 年 8 月末的在手订单金额为 6.26 亿元，为短期业绩提供了一定保障。与此同时，公司产品矩阵持续丰富，AI DPU 产品风驰 2.0 系列已成功流片，正推动客户量产导入，有望在 2027 年实现放量销售，为公司打开新的增长曲线；公司亦积极拓展客户结构，以降低单一客户资本支出波动带来的潜在影响，进一步增强了经营的稳定性与抗风险能力。

针对前述下游客户资本支出波动可能带来的业绩风险，公司还将采取以下应对措施：

1、持续拓展客户结构，分散单一客户波动风险。公司客户正逐渐从腾讯拓展至其他头部云服务商、电信运营商、互联网公司、国产知名 AI 大模型厂商等，并在金融、电力等行业头部客户完成项目落地。多元化的客户结构有助于分散单一客户资本支出波动的风险；

2、不断丰富产品矩阵，拓展应用场景。公司全功能 DPU 云霄 1.0 系列已在腾讯、中国移动等客户实现规模化商用；AI DPU 风驰 2.0 系列已成功流片，正推动客户量产导入。公司下一代 800Gbps 速率的云霄 2.0 全功能 DPU 及最高支持 800Gbps/1.6Tbps 的风驰 3.0 系列 AI DPU 产品已经启动研发。产品矩阵的丰富使公司能够覆盖云计算、AI 智算、存储、网络安全等多个下游场景，降低单

一场景需求波动的影响；

3、坚持强化产品性能，提升配置优先级。在资本支出总量受限的环境下，更高性能、更具价格优势的国产 DPU 产品更易获得客户配置优先级，有望在行业增速放缓时实现份额提升；

4、保持充足研发投入，巩固技术领先性。公司报告期内累计研发费用投入超 21 亿元，持续维持高强度的研发投入。截至 2026 年 6 月 30 日，公司货币资金、交易性金融资产、一年内到期的非流动资产及大额存单本息金额合计 22.49 亿元，资金储备相对充足，能够保障未来一定期间内的研发投入；

5、深化国产替代战略，把握政策红利。在国家算力自主可控战略推动下，国产 DPU 渗透率有望加速提升。公司作为全功能 DPU 市场国产独立厂商第一名，在国产替代加速趋势中具备先发优势和稀缺性价值，有望在资本支出增速放缓的不利环境下仍保持平稳经营。

二、预期存量服务器及近年新建数据中心批量更新迭代的周期，发行人 DPU 替换进入存量算力中心的可行性，DPU 产品替换特性、服务器批量更新预期等情况对发行人未来业绩的影响，并进行风险提示

（一）预期存量服务器及近年新建数据中心批量更新迭代的周期

1、当前云服务商服务器更新周期

亚马逊 2023 年委托标普全球市场财智（S&P Global Market Intelligence）开展专题研究并发布《企业上云：构建 IT 实力和减碳》报告，调研超过 300 家中国企业、涵盖多个行业，研究发现：平均而言，云服务商服务器的使用时间仅为 3 到 4 年。近 40%的受访企业通常会等待 3-6 个月才部署最新一代服务器。相比之下，云服务商通常会在产品在市场上普及之前就采用最新的技术，因为他们更愿意自己设计、测试并投资新芯片和服务器技术。云服务提供商的服务器通常更新更快、效率更高，能够确保实现最佳的性能。

根据 IDC 于 2025 年发布的《过时的服务器基础设施如何支撑创新？有效掌控技术债务，实现超前管理》白皮书，经调研约 200 家英特尔、戴尔服务器使用企业，约 44%的企业每 3 年或更短时间更新一次本地服务器，约 98.5%的企业 5 年以内更新一次本地服务器。同时超过 40%的企业表示，2020 年以来更新或升

级服务器设备频率正在提高，主要原因系过去五年的创新和技术变革速度变快，企业更新、升级周期相应缩短。

根据云服务商的年度报告等公开信息，当前主要云服务商服务器等固定资产的会计核算折旧年限主要区间为 3 至 5 年。但有关资产折旧年限是管理层会计估计，不等于服务器实际使用寿命。头部云服务商承载关键业务的服务器实际更新周期受当前技术加速迭代影响，可能快于上述周期。

云服务商	服务器折旧年限
腾讯	电脑及其他运营设备 2 至 10 年
阿里	计算设备和软件 3 至 5 年，CFO 在 2026 年 8 月 20 日业绩会提及：“集团服务器的使用周期为 5 年，AI 服务器约 3 年即覆盖硬件成本，第 4、5 年转化为自由现金流”
百度	计算机设备 3 至 6 年
华为	电子及其他设备 2 至 5 年
亚马逊	自 2025 年 1 月 1 日起更改部分服务器及网络设备的使用寿命估计，由 6 年调整为 5 年。使用寿命缩短的原因在于技术发展速度加快，尤其是在人工智能和机器学习领域
微软	服务器及网络设备 2 至 6 年
Meta	服务器及网络资产 4 至 5 年，2025 年 1 月起部分延长至 5.5 年
谷歌	服务器及网络设备一般在 6 年内折旧
甲骨文	计算机、网络、机器设备 1 至 6 年，服务器及网络设备预计使用寿命为 6 年

数据来源：公司年度报告、业绩说明会

2、未来 AI 服务器更新周期呈缩短趋势

AI 技术的爆发式发展正在大幅压缩服务器更新迭代周期。德国 PCB 设备制造商 SCHMID 集团（NASDAQ: SHMD）首席销售官表示，典型的人工智能服务器板的生命周期估计只有 18 至 24 个月，新架构就会推动对下一代设计的需求。美国计算机协会通讯（CACM）2026 年 3 月文章指出，AI 数据中心可能需要在百万芯片规模下，每 18 到 36 个月更换一次 GPU。根据上述行业数据，AI 服务器未来更新周期可能缩短至 18 至 36 个月。主要驱动因素包括：

（1）算力芯片架构迭代提速

英伟达硬件更新周期已从 2 年缩短至 1 年。Blackwell 架构集群落地不足半年，新一代 Vera Rubin 平台即全面铺开。2024 年前后集中部署的 A100/H100 服务器在算力、能效方面预计无法满足未来大模型和 AI 推理业务需求，运维电费

居高不下，大量云服务商和互联网企业批量下线老旧算力集群。

（2）功耗与散热的物理极限

2024 年前后建造的数据中心大多基于单机柜 10-40kW、风冷为核心的设计。英伟达在售的新一代 AI 架构（Blackwell Ultra、Rubin 等），单机柜功耗已暴涨至 100-240kW 以上，必须采用液冷散热。老旧数据中心的承重、层高和管道设计无法原地改造，面临提前退役或整体重建。

（3）网络互联的木桶效应

AI 集群对网络吞吐量的需求以每 1-2 年翻倍的节奏升级。交换机端口速率从 2024 年的 200G+400G 组合，到 2025 年切换为 400G，2026 年已有厂商提前规划部署 800G。同一 AI 集群同一网络平面内不宜混速部署交换机和光模块，理想状况下网络设备应跟随 GPU 代际同步更新。DPU 作为网络互联的核心组件，其更新与服务器和网络设备的迭代紧密绑定。

（4）大模型参数规模与业务场景变化

当前，大模型参数规模呈指数级增长，驱动算力需求急剧攀升。以单机部署能力为例，在 H200 时代，8 卡即可承载 7,000 亿参数级模型，而在 B200 时代，需 64 卡集群方可支撑 2.4 万亿参数级模型，GPU 集群规模随参数规模同步倍增。此外，Agent 和推理场景正在改变 CPU 与 GPU 的配比关系，传统架构中一个 CPU 配 4-8 个 GPU，而 Agent 时代 CPU 配比需求有所提升，服务器算力架构变化推动集群整体更新升级。

（二）发行人 DPU 替换进入存量算力中心的可行性

1、DPU 的 PCIe 插卡式部署特性

DPU 产品在物理形态上以 PCIe 插卡形式存在，可直接插入标准服务器的 PCIe 插槽内。公司 DPU 产品采用标准 PCIe 接口，符合 PCIe 及以太网协议规范，可兼容主流 x86 及 Arm 服务器平台。从物理部署角度看，只要存量服务器具备可用的 PCIe 插槽并满足 DPU 的功耗要求，即可部署 DPU 卡实现功能升级。

行业标准服务器通常配置多个 PCIe 插槽。以 2U 机架式服务器为例，典型通算服务器配置提供 6-10 个 PCIe 插槽，在安装 RAID 卡、NVMe 卡等设备后，

通常仍有空闲 PCIe 插槽可用于 DPU, AI 服务器的 PCIe 插槽数则根据 GPU 配置需要, 更为灵活。根据服务器平台技术规格, AMD EPYC 9004/9005 系列处理器提供 128 条 PCIe 通道, Intel Xeon 6 系列处理器提供 88~136 条 PCIe 通道, 足以支撑 GPU+DPU 的并行部署。

2、公司产品具备技术兼容性与平滑迁移能力

公司 DPU 产品在技术设计上具备以下关键特性, DPU 替换进入存量算力中心具备可行性:

(1) Virtio 虚拟化兼容实现平滑迁移

公司 DPU 通过硬件加速 Virtio-net/Virtio-blk 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。Virtio-net/Virtio-blk 是当前数据中心主流的网络和存储软件虚拟化技术, 公司 DPU 产品对其提供全面兼容和硬件加速, 客户原有基于 Virtio 架构的应用无需修改即可在 DPU 上运行, 保障业务连续性。这一特性使存量算力中心在不改动现有软件架构的前提下, 即可实现从传统网卡到 DPU 的渐进式替换。

(2) 热插拔、热迁移与热升级

公司 DPU 产品支持热插拔、热迁移和热升级三项关键特性: 1) 热插拔: 可在不关闭服务器电源的情况下插入或移除 DPU 卡; 2) 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; 3) 热升级: 支持在不中断业务的情况下对 DPU 固件进行在线升级和迭代, 确保基础设施持续优化不影响业务连续性。上述特性使 DPU 替换可在不停机或极短停机窗口内完成, 满足数据中心改造的业务连续性要求。

(3) 统一架构降低改造成本

公司 DPU 产品为数据中心提供弹性裸金属、虚拟机和容器云服务统一架构和运维解决方案, 全面解决三种不同场景的资源虚拟化和统一管理。存量数据中心在改造过程中往往同时运行裸金属、虚拟机和容器等多种业务形态, 公司 DPU 的统一架构可避免为不同业务形态分别部署不同的网络加速设备, 有效降低改造成本和运维复杂度。

（4）标准 Linux 生态兼容

公司 DPU 产品支持标准 Linux 操作系统，提供软硬件卸载和加速以及灵活可编程能力。客户原有基于 x86 服务器开发的 DPDK 应用，仅需简单编译即可移植到 DPU 上运行，无需重新开发，大幅降低了存量算力中心改造的技术门槛和迁移成本。

3、存量服务器加装 DPU 具备物理可行性，但存量算力中心更新主要通过整机升级更换

从存量算力中心的产品形态看，较早期服务器侧主要配置普通网卡，相关产品主要承担基础网络连接及数据转发功能；由于 DPU 作为独立数据面处理单元的产品概念主要在 2020 年前后开始被市场广泛认知，例如英伟达 BlueField-2 于 2020 年正式推出，因此在此之前建设的算力中心服务器通常尚未大规模配置全功能 DPU。随着 DPU 产品形态逐步成熟，存量算力中心更新需求主要包括两类：一类是早期普通网卡服务器在更新时升级为全功能 DPU 方案，另一类是已采用 FPGA DPU 过渡方案的客户逐步向 DPU SoC 方案切换。

算力中心服务器及配套网络设备并非短期一次性替换。虽然理论上 DPU 硬件支持在终端客户已有服务器上热插拔和单独更换，但这在商业实践中并不具备普适性。大型算力中心通常提前半年以上进行服务器硬件选型和备货，新 DPU 部件的导入窗口通常锁定在新一轮服务器代际更替的规划节点附近。下游客户（尤其是大型云服务商）通常遵循严格的硬件生命周期管理策略。除非 DPU 本身发生物理故障，否则客户通常不会仅仅为了追求算力效率提升而中断业务、单独升级 DPU 硬件。DPU 产品的生命周期因此常被约束在整台服务器的折旧年限内，随服务器一同淘汰。

从更新周期看，部分存量服务器在实际运行中可维持约 3 至 5 年的使用周期。因此，2019 年及以后部署的服务器在当前及未来几年仍构成存量算力中心的重要组成部分。随着 AI 应用、云原生架构及高密度数据中心建设持续推进，该等较早部署的服务器在网络带宽、能效水平、稳定性及基础设施卸载能力方面将逐步产生更新换代需求，存量服务器替换及存量算力中心架构升级预计成为公司全功能 DPU 市场的重要需求来源。

（三）DPU 产品替换特性、服务器批量更新预期等情况对发行人未来业绩的影响

“存量数据中心更新”与“新数据中心建设”并非两类割裂市场，导致下游客户对 DPU 需求均表现为增量需求，详细分析参见“问题 1/一/（一）/4、存量服务器改造对 DPU 市场渗透率的影响”。DPU SoC 作为领先方案，其在整体 DPU 市场渗透率提升，并不依赖于对海量存量服务器的碎片化改造，而是高度依赖于云服务商及互联网巨头每年新增资本支出的景气度。

当前，DPU 下游 AI 智算及云服务市场正处于高速扩张期，下游客户大幅增加有关资本支出。根据 TrendForce 数据，2026 年字节跳动、腾讯、阿里巴巴、百度的合计资本支出将同比增长逾 80%。随着采用传统网卡及 FPGA DPU 方案的旧服务器生命周期陆续期满，DPU SoC 作为当前主流技术路线，已逐渐成为包括华为、腾讯、字节跳动在内云服务厂商的主流选择，有望在未来数据中心建设中占据主要份额、持续获取新订单。

在 DPU 渗透率持续提升、AI 服务器退役高峰的驱动下，根据弗若斯特沙利文数据，公司面临的国内 DPU 市场需求从 2026 年约 620 亿元增长至 2028 年约 912 亿元。若仅按维持 5%不变的市场份额测算，对应潜在收入规模从 2026 年约 31 亿元增长至 2028 年约 46 亿元，为公司未来业绩增长提供了坚实的市场基础。

（四）有关风险提示

公司已在招股说明书“第三节 风险因素”之“二、与行业相关的风险”中补充披露了“存量数据中心更新周期延长及 DPU 渗透率提升不及预期的风险”，具体如下：

“（五）存量数据中心更新周期延长及 DPU 渗透率提升不及预期的风险

尽管 AI 服务器迭代周期正加速缩短，但下游厂商可能通过第三方维保服务、延长固定资产折旧年限等方式推迟旧服务器的更新换代。若下游厂商存量服务器更新周期延长，将导致 DPU 增量需求的释放节奏相应放缓，进而对公司 DPU 产品销售产生不利影响。此外，受资本支出波动、存量服务器更新周期变动等因素影响，存量数据中心更新中的 DPU 渗透率提升可能不及预期，将限制公司产品在存量市场的拓展空间，亦将对公司未来业绩造成不利影响。”

中介机构核查程序及核查意见：

一、核查程序

保荐人履行了以下核查程序：

1、查阅弗若斯特沙利文、IDC、TrendForce 等研究机构发布的相关行业研究报告和公开披露数据，了解全球及中国 DPU 市场规模、渗透率现状及增长趋势，了解下游客户资本支出规模及增长预测；

2、查询腾讯、百度、阿里等下游客户的年度报告、业绩发布会及公开披露信息，了解其资本支出计划及 AI 基础设施投资规划，分析全功能 DPU 及 AI DPU 在资本支出中的价值量占比；分析资本支出缩减对 DPU 行业及发行人业绩的潜在影响；

3、获取发行人盈利预测资料及在手订单情况，了解发行人对下游客户资本支出波动的应对措施及抗风险能力，分析收入敏感性及时点影响；

4、查阅发行人招股说明书风险因素章节，核查发行人是否已充分披露下游投资波动相关风险；

5、查阅发行人说明，了解发行人客户拓展进展、产品矩阵丰富计划及国产替代战略推进情况，分析发行人应对行业波动的措施有效性。

二、核查意见

经核查，保荐人认为：

1、云服务商等客户每年新增资本支出是 DPU 市场渗透率提升的核心驱动因素。国内整体 DPU 渗透率仍处于低位快速提升阶段，资本支出总量增长与渗透率提升共同驱动 DPU 市场规模扩张，存量服务器改造需求最终统一体现为下游客户年度资本支出投入。发行人关于资本支出、存量服务器改造等因素对 DPU 市场渗透率影响的分析具有合理性。

2、若下游客户资本支出增速放缓或出现缩减，DPU 行业增速可能受到一定影响，但 DPU 渗透率提升的增量空间、AI 智算场景中的刚需属性以及在资本支出受限环境下的相对配置优先级提升，使得 DPU 具有一定抗风险能力。发行人关于资本支出缩减对 DPU 行业及发行人业绩影响的分析具有合理性。

3、发行人已针对下游投资波动风险采取了客户结构多元化、产品矩阵丰富化、保持充足研发投入及深化国产替代战略等应对措施，具备一定的抗风险能力。发行人已在招股说明书中补充披露“下游 AI 智算与云计算基础设施投资周期波动的风险”“存量数据中心更新周期延长及 DPU 渗透率不及预期的风险”，相关风险提示充分、完整。

问题 2. 关于云服务商等客户的 DPU 产品需求

申报材料及首轮问询回复显示：多数国内云服务商自研的全功能 DPU 现阶段以基于商业 FPGA 解决方案的过渡方案为主。

请发行人披露：

(1) 在全功能 DPU 领域，下游云服务商等客户自研 FPGA 方案目前情况及未来一段时间的预期使用情况，FPGA 方案性能上限能否满足行业主流数据中心当前及未来需求，发行人预计 FPGA 方案受限于带宽与功耗难题、DPU SoC 演进路线将成为主流路线的依据。

(2) 发行人全功能 DPU 产品与客户自研 FPGA 方案相比的竞争优势，自研 FPGA 方案客户转向发行人全功能 DPU 的考虑因素及对发行人业绩的预期影响，并进行相关风险提示。

(3) 在 AI DPU 领域，客户选择自研或外采的主要考虑，发行人在该领域的竞争优势，发行人 AI DPU 产品后续批量出货的主要影响因素。

请保荐人说明核查依据、过程，并发表明确意见。

【回复】

一、在全功能 DPU 领域，下游云服务商等客户自研 FPGA 方案目前情况及未来一段时间的预期使用情况，FPGA 方案性能上限能否满足行业主流数据中心当前及未来需求，发行人预计 FPGA 方案受限于带宽与功耗难题、DPU SoC 演进路线将成为主流路线的依据

(一) 在全功能 DPU 领域，下游云服务商等客户自研 FPGA 方案目前情况及未来一段时间的预期使用情况

1、云服务商自研 FPGA 方案目前情况

根据弗若斯特沙利文数据，中国全功能 DPU 市场中 FPGA DPU 方案占比约 25%。目前，FPGA 方案主要留存于低网络带宽、定制化功能开发等特定场景，在主流数据中心高性能网络场景中的使用比例下降。

国内云服务商全功能 DPU 方案呈现“自供 FPGA DPU+外采 DPU SoC”并

行的特征。根据弗若斯特沙利文截至 2025 年的调研数据及公开查询信息，云服务商在全功能 DPU 领域自研 FPGA 方案在其业务中的具体应用情况如下：

华为、字节跳动均以 DPU SoC 为主。字节跳动自研火山引擎 DPU 1.0 于 2021 年公测上线，采用 FPGA 方案，2.0 于 2023 年发布，3.0 于 2026 年发布，最高支持 400Gbps 网络带宽。但从部署规模看，字节跳动自研产品占比较少，外采 DPU SoC 仍占据主流。

腾讯处于自研 FPGA DPU 方案及外采 DPU SoC 并行阶段，正由 FPGA 方案转向 DPU SoC。腾讯早期自研的水杉、银杉网卡分别于 2020 年和 2021 年上线，采用 FPGA 方案，银杉 1.0 网卡最高支持 200Gbps 网络带宽。由于 FPGA 路线逐渐逼近性能和功耗的瓶颈，2024 年起腾讯与公司合作的玄灵网卡（基于公司 DPU SoC 方案）开始大规模部署，应用于第九代云服务器及 AI 智算服务器。根据公司对其销售规模，2026 年 DPU SoC 方案占比预计大幅提升。

阿里、百度以自研 FPGA DPU 方案为主。阿里为自身云服务定制的 CIPU 2.0 于 2024 年 9 月云栖大会发布，最高支持 400Gbps 网络带宽。此前，阿里神龙 MOC 1.0（2017 年）、1.5（2018 年）、2.0（2019 年）、2.5/CIPU 1.0（2022 年）均基于 FPGA 方案开发，CIPU 2.0 在此基础上开始引入 ASIC 技术（eFPGA）以解决 FPGA 的应用问题，历经 7 年从 FPGA 走向 ASIC 化，也侧面反映了 400Gbps 高性能场景下 FPGA 方案的自身局限性。百度自研太行 DPU 1.0 于 2020 年发布，采用 FPGA 方案，2.0 于 2023 年发布，最高支持 200Gbps 网络带宽。但目前百度数据中心网络需求已演进到 400Gbps 带宽，太行 DPU 2.0 最高 200Gbps 的网络带宽能力与 AI 智算集群 400Gbps 的网络接入需求存在代际差距，在支撑 AI 场景下的高性能网络需求时存在限制。

2、未来一段时间的预期使用情况

综合上述产业实践及行业发展趋势，公司预计云服务商自研 FPGA 方案未来一段时间的预期使用情况如下：

在 400Gbps 以上高性能全功能 DPU 领域，FPGA 方案将逐渐退出客户大规模部署的主流选择范围。当前行业主流数据中心网络已进入 400Gbps 规模化部署、800Gbps 方案规划阶段，AI 智算场景对网络带宽、时延和集群效率的要求持续

提升，FPGA 方案受限于功耗、成本、性能与规模部署可靠性的固有约束，难以满足上述场景需求。阿里在 400Gbps 高性能场景开始引入 ASIC 技术，目前国内外未有成功商业化 800Gbps FPGA 方案问世，而国际头部厂商英伟达支持 800Gbps 的 DPU SoC 产品已于 2026 年发布，因此，客户未来若在大规模商用中选择 FPGA 方案，可能面临性能或功耗瓶颈。

在 400Gbps 以下中低网络带宽、定制化或小规模技术方案验证等场景，FPGA 方案仍存在一定应用空间。FPGA 方案在部分对功耗不敏感、功能定制化程度高、部署规模相对有限场景中可满足特定需求，该部分场景预计将继续维持一段时间的存量应用。

综上所述，从长期趋势看，在高性能、大规模商用场景，云服务商等客户自研 FPGA 方案将逐步被 DPU SoC 方案替代。随着 DPU SoC 产品在功耗、性能、成本方面的综合优势持续显现，以及行业主流数据中心网络带宽持续升级，FPGA 方案的应用场景将愈发局限，DPU SoC 演进路线预计将成为行业主流路线。

（二）FPGA 方案性能上限能否满足行业主流数据中心当前及未来需求

1、FPGA 方案的技术架构与性能特征

全功能 DPU 领域的 FPGA 方案通常采用“FPGA+外置 CPU”的双芯片架构：由 FPGA 芯片承载报文解析、流表匹配、加解密等数据面处理功能，由外置 CPU 芯片承载控制面处理、虚拟化及存储管理和上层协议栈等功能。该架构下，FPGA 芯片与 CPU 芯片之间需通过高速串行接口（如 PCIe、SerDes 等）进行数据交互，数据面处理性能受限于 FPGA 可编程逻辑资源的规模与工作频率。

FPGA 方案的核心优势在于可编程性与快速迭代能力：由于功能由可编程逻辑实现，客户无需经过芯片流片即可调整硬件功能，适用于网络协议尚未定型、功能需求频繁变化的验证及应用阶段。其核心劣势则在于性能与效率：受限于可编程逻辑阵列的架构特征，FPGA 方案在功耗、芯片面积效率、单位成本等方面劣于 DPU SoC 芯片。

2、FPGA 方案的主要性能瓶颈

FPGA 方案的性能瓶颈主要包括如下几方面：

（1）功耗瓶颈

FPGA 芯片的可编程逻辑资源由查找表（LUT）及可编程布线资源构成，无论资源实际利用率高低，配置存储器与闲置布线始终存在，持续产生静态与动态功耗，功耗难以随有效算力线性增长。根据电子设计工程领域权威媒体《Electronic Design》文章，在实现同等功能条件下，采用 ASIC 方案可将信号处理的逻辑功耗降低约 4 至 10 倍；基于 16nm 工艺 FPGA 搭建的子系统整机功耗约为 3 瓦，而同规格功能的 12nm/7nm 工艺 ASIC 方案整体功耗可控制在 1 至 1.25 瓦。根据 Arm、中国半导体行业协会信息，基于 Arm 架构的 DPU SoC 的板卡功耗只有基于 FPGA 架构 DPU 卡的约三分之一。

DPU 板卡功耗直接挤占服务器散热与供电预算、抬升数据中心 PUE。FPGA 方案在 400Gbps 带宽下整卡功耗已逼近或达到系统整体预算上限，向更高速率演进时功耗问题将更加突出；而 DPU SoC 通过专用加速引擎与多核 CPU 的单芯片集成，在能效比上具备结构性优势。因此，在数据中心单卡功耗预算约束下，FPGA 方案在实现 400Gbps 以上带宽的 DPU 功能时，功耗将触及单卡功耗上限，难以满足大规模部署的散热与能耗要求。

（2）芯片面积瓶颈

由于可编程逻辑与布线资源占用大量芯片面积，FPGA 器件在单位面积上的晶体管集成密度低于 ASIC 方案。根据 Ian Kuon 与 Jonathan Rose 发表于 IEEE 期刊研究结果，在相同 90nm 工艺下实测，实现同等逻辑功能，FPGA 所需的硅片面积约为标准单元 ASIC 的 18~35 倍。腾讯云智能网卡研发负责人 2021 年在中国智能网卡研讨会上公开表示，200Gbps 网络带宽 FPGA 方案已触及芯片面积物理上限。在 400Gbps 以上网络带宽，FPGA 方案需采用更大尺寸器件或增加器件数量，将导致芯片面积、功耗和成本同步大幅上升，在物理与经济层面均难以满足要求。

（3）成本瓶颈

当前国内高端 FPGA 依赖海外厂商赛灵思和 Altera，高端型号产能受限、交期紧张、单价较高。同时，FPGA 方案需额外配置外置 CPU 芯片、高速接口器件等，且 FPGA 与 CPU 芯片均需独立 DDR 内存，单卡物料成本高于单芯片 SoC

方案。此外，FPGA 方案在规模部署下的功耗成本亦更高，综合持有成本（TCO）劣势明显。由字节跳动 Haoran Chen 与北京大学 Yuemeng Xu 等人在 2025 年发表的《Gryphon: Scaling Hyperscale Multi-Tenant Gateways Beyond the Petabit-Era via DPU-Augmented Hierarchical Co-Offloading》研究指出，在相同的 2U 机箱下进行的基准测试显示，与商用 FPGA 方案相比，基于 DPU SoC 的网卡将功耗降低了 21%，硬件成本降低了 14%，降低了大规模部署时的总运营成本。

（4）运维瓶颈

在规模部署运维层面，FPGA 方案的核心瓶颈在于故障率与故障处置能力。FPGA 依赖大规模可重配置存储，对单粒子翻转（SEU）的敏感性高于固化逻辑的 ASIC 芯片，且随制程缩小加剧，同时 SEU 引起的静默数据损坏定位难度大。该等故障率在单机部署下尚可接受，但在数万至数十万节点的数据中心规模下，故障率被部署基数线性放大。叠加国内高端 FPGA 依赖海外厂商赛灵思和 Altera 的因素，故障备件供应与海外技术支持的响应速度亦不及本土厂商。而 DPU SoC 芯片层面采用固化逻辑架构，不存在大规模可重配置存储阵列，规模化部署下的故障率水平与故障处置复杂度低于 FPGA 方案，运维可依托成熟的 Linux 诊断工具链完成。

3、新一代 FPGA 芯片技术进展及其适用边界

近年来，FPGA 厂商通过集成硬核 IP 持续提升芯片性能，例如赛灵思（AMD）Versal Premium 系列支持 400Gbps/800Gbps 以太网接口、112G PAM4 SerDes；Altera Agilex 7 系列支持 400Gbps MAC 等。海外高端 FPGA 在接口速率方面或可支持 400Gbps 及以上网络带宽。

但是，FPGA 芯片仅靠接口速率能力的提升并不能解决 FPGA 方案在功耗、面积与成本方面的固有瓶颈：其一，接口速率提升的同时，可编程逻辑部分处理 400Gbps 级数据面功能所需的资源规模、功耗随之急剧上升，整卡功耗将超出数据中心单卡功耗预算；其二，FPGA 方案仍需外置 CPU 承载控制面功能，双芯片架构的功耗、面积、成本劣势不因器件接口速率提升而消除；其三，采用海外高端 FPGA 芯片的整卡方案在量产成本上高于同规格 SoC 单芯片方案，难以满足云服务商、运营商等客户大规模部署的成本要求。

因此，新一代 FPGA 在接口速率上的进展，主要用于交换机、网络测试、原型验证等场景，预计无法支撑 FPGA 方案在 400Gbps 以上主流数据中心高性能全功能 DPU 场景中实现规模化应用。

4、行业主流数据中心网络需求演进

从行业主流数据中心网络需求看，当前数据中心网络已进入 400Gbps 规模化部署、800Gbps 方案规划阶段；AI 智算场景下，GPU 集群规模从千卡向万卡、十万卡演进，对网络带宽、时延、拥塞控制、多路径传输等能力提出更高要求。国际头部厂商英伟达 BlueField-4 为 800Gbps，并进一步向 1.6Tbps 演进。腾讯已大规模部署公司 400Gbps 云霄 1.0。行业主流数据中心当前及未来需求已明确指向 400Gbps 及以上高性能全功能 DPU，FPGA 方案在 400Gbps 档位即已触及面积与功耗上限，在 800Gbps 档位可能无法满足未来需求。

综上所述，FPGA 方案在中低带宽网络、定制化及小规模技术验证场景下尚可满足客户特定需求，但因功耗、芯片面积、性能、成本及运维等的固有瓶颈，其性能上限难以支撑 400Gbps 以上高性能网络规模化部署，预计无法满足行业主流数据中心未来需求。

（三）发行人预计 FPGA 方案受限于带宽与功耗难题、DPU SoC 演进路线将成为主流路线的依据

1、市场数据依据

DPU 市场规模的快速增长主要来源于云计算基础设施持续扩张与 AI 智算需求爆发，该等需求对网络带宽、时延、集群效率的要求持续提升，进一步推动市场向高性能 DPU SoC 产品集中，为 DPU SoC 路线主流化提供了坚实的市场基础。根据弗若斯特沙利文数据，中国全功能 DPU 市场中 SoC/ASIC 方案占比约 75%，FPGA 方案占比约 25%。DPU SoC 方案对 FPGA 方案的市场替代正在实际发生。

2、技术演进依据

从技术演进规律看，DPU SoC 方案将 FPGA 方案中分立的数据面处理引擎、控制面处理器以及网络、存储、安全等专用加速引擎集成于单颗芯片，具有以下优势：一是单芯片集成大幅降低整卡功耗与物料成本，可满足数据中心单卡功耗预算约束，并支撑 400Gbps 及以上速率的规模化部署；二是芯片内部高速互联替

代芯片间高速串行接口，降低数据交互时延、提升处理确定性；三是可通过多核通用处理器（如 Arm 架构或 RISC-V 架构核心）承载丰富软件栈，结合 P4 等可编程数据面语言，在保持硬件加速性能的同时提供软件可编程能力，兼顾性能与灵活性。上述技术特征决定了 DPU SoC 方案是兼顾性能、功耗、成本与可编程性的最优路径，具备成为行业主流路线的技术基础。

3、产业实践依据

从产业实践看，全球头部 DPU 厂商与头部云服务商采用路线均以 SoC/ASIC 方案为主：头部 DPU 厂商方面，英伟达自 2020 年推广 DPU 概念以来，BlueField 系列产品持续采用 SoC 路线，BlueField-3 实现 400Gbps、BlueField-4 实现 800Gbps，后续还将向 1.6Tbps 演进，超威半导体、博通、美满电子等国际头部厂商的 DPU 产品亦均采用 SoC/ASIC 路线；头部云服务商方面，亚马逊、微软自研 DPU SoC 产品，谷歌与英特尔合作开发的 Mount 系列 IPU 采用 SoC/ASIC 路线，Meta、甲骨文的全功能 DPU 大规模采用英伟达、超威半导体等成熟方案。

国内头部云服务商方面，华为自研 DPU 为 SoC/ASIC 方案；腾讯由 FPGA 方案转向基于 DPU SoC 的玄灵网卡；阿里 CIPU 2.0 开始引入 ASIC 技术以解决 FPGA 的失效问题，历经 7 年从 FPGA 走向 ASIC 化；字节跳动自研火山引擎 DPU 采用 FPGA 方案，但从部署规模看，外采 DPU SoC 仍占据主流。上述头部厂商的产业实践充分表明，DPU SoC 演进路线已成为全球及国内产业界的主流选择。

2020 年以来，业界关于 FPGA 方案受限于性能与功耗难题、DPU SoC 演进路线将成为主流路线的有关表述具体如下：

美国 FPGA 厂商 Achronix 最早于 2020 年发布的白皮书《FPGAs Enable the Next Generation of Communication and Networking Solutions》及其官网指出：“在传统 FPGA 架构内支撑单路 400G 以太网流，需要在可编程逻辑中运行 1024 位总线@724MHz 或 2048 位总线@642MHz，这两种组合在当今最先进的 FPGA 中均不可能实现。如此宽的总线难以布线，会消耗 FPGA 结构内大量布线资源，所需时钟频率下的时序收敛几乎不可能。”Achronix 提出 eFPGA 技术，即在 FPGA 内嵌入硬化 ASIC 形态的 400G 以太网 MAC 和类 ASIC 的二维片上网络，将 400G 流拆分为 4 路 100G 流再交给可编程逻辑，反映了 FPGA 向 ASIC 硬化电路融合

的技术趋势。

腾讯云智能网卡研发负责人 2021 年在中国智能网卡研讨会上公开表示，FPGA 方案在 200Gbps 时将碰到芯片面积的物理上限和单卡功耗上限，长期演进风险较高，值得警惕。

阿里云智能集团弹性计算资深技术专家在阿里云开发者社区文章中及 2024 年云栖大会明确表示：“FPGA 可重配置逻辑实现转发技术，time to market 能力有很大优势，但是对于 400Gbps/800Gbps 转发业务，挑战很大。”、“（相较于 FPGA）ASIC 技术显著降低芯片失效率，ECS 数据平面稳定性提升 20%”。

字节跳动 Haoran Chen 与北京大学 Yuemeng Xu 等人在 2025 年发表的《Gryphon: Scaling Hyperscale Multi-Tenant Gateways Beyond the Petabit-Era via DPU-Augmented Hierarchical Co-Offloading》研究表明，与商用 FPGA 方案相比，基于 DPU SoC 的网卡将功耗降低了 21%，硬件成本降低了 14%，降低了大规模部署时的总运营成本。

中国移动等联合编制的《云计算通用可编程 DPU 发展白皮书（2023）》指出，随着网络带宽的快速增长，基于 FPGA 硬件架构的 DPU 在性能上将难以满足，功耗方面也会存在很大挑战，ASIC 成为 DPU 产品的发展趋势。

中国联通于 2023 年发布的《中国联通 DPU 网络域场景应用白皮书》指出，基于 FPGA+CPU 硬件架构的 DPU 有功耗等方面限制，且 FPGA 芯片价格相对较高，在产品规模商用后也不利于控制成本。DPU SoC 是当前国内外头部 DPU 厂商和云服务提供商选择的主流技术路线之一。

中国半导体行业协会援引 Arm 业务拓展副总裁 2023 年接受《中国电子报》采访时发言，基于 Arm 架构的 DPU SoC 的板卡功耗只有基于 FPGA 架构 DPU 卡的约三分之一，DPU SoC 的产品在顾及可编程、高灵活性等特性之余，还同时具备高性能、低功耗、低成本等优势，必然成为未来 DPU 发展的主流方向。

DPU 行业主管部门工信部直属科研单位中国信息通信研究院发布的《DPU 发展分析报告（2026 年）》明确指出：“SoC 凭借高集成度、性能优化以及软件协同设计等优势，已成为当前 DPU 发展的主流路线之一。”报告还指出，SoC 将处理器、存储器、接口等功能模块集成到单个芯片上，降低系统复杂性，提高

整体性能和能效比，满足智算中心等场景对高性能计算的需求。

4、公司实践验证

公司是国内较早选择并率先实现 DPU SoC 方案规模化商用的芯片企业，公司首款全功能 DPU 云霄 1.0 于 2023 年一次性流片成功，可实现单芯片 400Gbps 高性能网络和通用可编程能力，是国内率先实现技术突破的厂商。公司 DPU SoC 产品已获得头部客户规模化验证：

腾讯基于公司 DPU 芯片的玄灵智能网卡自 2024 年第四季度起批量交付并商用，相较腾讯上一代基于海外商业 FPGA 方案芯片的网卡，基于公司 DPU SoC 芯片的第九代云服务实例在网络吞吐、存储性能等方面均有提升，从客户实际部署层面验证了 DPU SoC 方案相较 FPGA 方案的综合优势。

中国移动已发布基于公司 DPU SoC 芯片的磐石、智算琢光、灵耀三款 DPU 产品，其中磐石已落地商用，相比上一代基于 FPGA 方案的磐石 DPU 3.0，磐石 DPU 4.0 产品功耗降低 50%，成本也降低 50%，进一步提升数据中心基础设施的部署效率和使用效益。

截至 2026 年 8 月末，公司 DPU SoC 产品累计出货量已超过 13 万颗，并已启动 800Gbps 新一代产品研发。

综上所述，FPGA 方案受限于带宽与功耗等固有难题，其性能上限无法满足行业主流数据中心当前及未来需求；随着网络带宽持续升级、AI 智算需求快速释放，DPU SoC 方案在性能、功耗、成本等方面的综合优势将持续凸显，DPU SoC 演进路线将成为全功能 DPU 领域的主流技术路线，上述判断具有技术规律、产业实践、市场数据及公司产品验证等多方面依据支撑。

二、发行人全功能 DPU 产品与客户自研 FPGA 方案相比的竞争优劣势，自研 FPGA 方案客户转向发行人全功能 DPU 的考虑因素及对发行人业绩的预期影响，并进行相关风险提示

（一）发行人全功能 DPU 产品与客户自研 FPGA 方案相比的竞争优劣势

全功能 DPU 进一步按照技术路线划分为 FPGA DPU 和 DPU SoC，两类方案对比如下：

对比维度	客户自研 FPGA 方案	公司 DPU SoC 方案
芯片架构	FPGA+外置 CPU 双芯片方案，板卡上需集成两颗以上芯片	单芯片集成多核 CPU+网络/存储/安全加速引擎，如公司云霄 1.0 集成 16 核 Arm N2+64 核 RISC-V，集成度更高
制程工艺	受 FPGA 供应商制约，相对落后制程	相对先进制程，集成度更高，性能优势更强
功耗水平	板卡功耗较高，基于 Arm 架构 DPU SoC 的板卡功耗相较 FPGA 方案低	功耗低，单芯片集成设计避免了多芯片间的能耗损耗
单卡成本	成本昂贵，高端 FPGA 芯片价格高，且需额外配置 CPU 芯片	成本具竞争力，专用 SoC 芯片大规模量产单价低于 FPGA 方案
开发门槛	需专业 HDL 开发团队，调试困难，人力投入大；硬件直接使用外购芯片，可以节省芯片设计、流片时间	支持 C/C++ 等高级语言编程，软件开发效率高，生态更广泛；芯片设计、流片周期长、存在流片失败风险
灵活性	极高，可重构数据通路，适合协议快速迭代场景	可编程（通过 P4/RISC-V 等），兼顾灵活性与性能，灵活性略低于 FPGA 但满足绝大多数场景需求
带宽支持能力	在向 400Gbps 以上演进时，受限于芯片面积上限、单卡功耗等物理天花板	可支持 400Gbps 及更高带宽，公司云霄 1.0 已达到 400Gbps，云霄 2.0 将支持 800Gbps
供应链安全	高端 FPGA 芯片供应商集中在海外（赛灵思/Altera），存在缺货、受限等供应链风险	公司自研 SoC 拥有自主知识产权，供应有保障
长期演进路径	长期演进风险较高，目前暂未有成功商业化的 800Gbps FPGA 方案	技术路线清晰，可持续迭代升级，与行业带宽升级趋势一致

综合上述对比，公司全功能 DPU 产品与客户自研 FPGA 方案各有适用场景：在面向行业主流数据中心高性能需求的全功能 DPU 领域，公司产品的综合竞争优势明显；客户自研 FPGA 方案的优势主要体现在中低带宽网络、定制化程度高的特定场景，两者在应用场景上存在一定分层。

从长期成本角度看，全功能 DPU 领域，FPGA DPU 和 DPU SoC 的差异主要体现在初始研发投入、使用成本、量产成本、长期演进四方面。

在初始研发投入方面，DPU SoC 由于采用 ASIC 化设计，需要经历芯片架构设计、流片、验证及生态适配等环节，前期研发和导入成本通常高于 FPGA DPU；而 FPGA DPU 基于通用 FPGA 配合外挂 CPU 实现，早期硬件开发门槛相对较低，更适合功能验证和方案迭代。但 DPU SoC 一旦进入规模化量产阶段，前期硬件研发投入可通过大规模出货逐步摊薄，长期单位成本具备下降空间。

在使用成本方面，FPGA DPU 采用多芯片实现方式，系统集成度相对较低，功耗控制和运行稳定性存在一定限制；DPU SoC 则在单芯片内集成 CPU 核、数据面处理单元及相关加速模块，硬件集成度更高，能够更高效地完成网络、存储、

安全、虚拟化及管控等基础设施任务卸载，从而降低服务器长期运行中的能耗成本和运维复杂度。

在量产成本方面，FPGA DPU 依赖高端 FPGA 芯片及外挂 CPU，芯片采购成本较高，且芯片供应主要集中于少数海外厂商，长期成本弹性相对有限；DPU SoC 虽然前期流片及验证投入较高，但采用 ASIC 方案后，规模化量产下单位芯片成本通常低于 FPGA 方案，成本可控性更强。

从长期演进来看，随着网络带宽向 400Gbps 及以上演进，FPGA 方案在功耗和成本方面的约束逐步显现，长期演进存在风险，DPU SoC 成为全功能 DPU 的主流演进方向。国际头部厂商英伟达支持 800Gbps 的 DPU SoC 产品已于 2026 年发布，而目前未有成功商业化的 800Gbps FPGA 方案问世。因此，虽然 DPU SoC 导入仍需完成客户网络、存储、虚拟化及云平台软件体系适配，存在一定验证周期和切换成本，但考虑到其在性能、能效、集成度和长期 TCO 方面的优势，DPU SoC 技术路线替代 FPGA DPU 具备明确的技术及商业合理性。

（二）自研 FPGA 方案客户转向发行人全功能 DPU 的考虑因素

腾讯于 2020 年 9 月上线第一代基于 FPGA 的自研产品“水杉”，支持 2x50G 网络接口；2021 年 10 月上线第二代“银杉”，支持 2x100G 网络接口。两代产品均采用 FPGA 方案，实现了软硬协同热迁移、弹性网卡/云盘密度和网络性能等技术突破。但腾讯云智能网卡研发负责人 2021 年在中国智能网卡研讨会上公开表示，FPGA 解决方案在 200Gbps 时将碰到芯片面积的物理上限和单卡功耗上限，长期演进风险较高。目前，腾讯已从 FPGA 路线方案逐步转向公司的 DPU SoC 路线。

根据公司客户实际部署情况、客户公开技术分享及行业公开信息，自研 FPGA 方案客户选择转向公司全功能 DPU 产品，主要基于以下考虑因素：

1、网络性能升级需求是客户切换方案的直接触发因素

在 200G 网络时代，客户采用自研 FPGA 方案尚处于方案可承载范围内。随着云服务商、运营商业务规模扩大与 AI 智算需求爆发，客户对数据中心网络带宽、吞吐能力、时延确定性的要求持续提升，其自研 FPGA 方案的潜在性能瓶颈难以满足未来需求。腾讯云智能网卡研发负责人曾公开指出 FPGA 路线逐渐逼近

性能和功耗的瓶颈，在数据中心网络普遍升级到 400G 的时代，DPU SoC 路线成为头部云服务商满足其 AI 基础设施持续升级需求的现实选择。因此，网络性能升级是促使客户从自研 FPGA 方案转向 SoC 方案的直接触发因素。

腾讯第九代云服务器配置基于公司 DPU SoC 的玄灵智能网卡，相较配置海外商业 FPGA 芯片方案网卡的上一代服务器，实例提供更强大的吞吐能力、原生支持 IPv6、降低网络延迟，为金融高频交易、短视频、直播等场景提供微秒级确定性时延，并在连接数、吞吐性能、存储性能、磁盘密度、数据加解密性能等方面均有所提升，可承载百万级实时并发请求。

2、功耗与成本是客户在大规模部署场景下的刚性约束

数据中心大规模部署场景下，单卡功耗与综合成本是客户选型必须满足的刚性约束。FPGA 方案在高性能数据处理时功耗可能超出数据中心单卡功耗预算，且高端 FPGA 器件单价高、叠加外置 CPU 的物料成本高，规模部署的综合持有成本劣势明显。单芯片 SoC 方案在满足同等网络性能要求的同时，可在功耗与成本上更好地适配客户的数据中心运营约束。客户在同等性能满足前提下，会将功耗与综合成本作为选型的重要约束条件纳入评估，这也是促使客户转向 DPU SoC 产品的重要现实考量。

3、生态兼容性与切换成本是客户决定是否选择 SoC 方案的关键考量

对已深度使用自研 FPGA 方案的客户而言，切换技术路线的首要顾虑在于新方案能否与既有计算生态平滑兼容、迁移与切换成本是否可控——若生态兼容性不足，即使新方案性能或成本更优，客户亦难以实施切换。公司产品基于开放标准协议，可兼容绝大多数主流 CPU 与 GPU 服务器平台，客户无需对存量计算平台进行大规模改造即可实现方案切换，生态适配成本较低。

同时，客户坚持自研 FPGA 方案需持续投入大量研发资源进行功能开发、软件栈建设和运维支持，且随速率升级需不断迭代重构，研发投入与时间成本高昂；外采成熟的全功能 DPU SoC 产品可缩短产品开发与上市周期，客户可将有限研发资源集中于 AI 算力、云计算等核心业务能力建设。生态兼容性与切换成本是否可控，直接决定客户是否愿意且能够选择 SoC 方案。

4、结合未来集群发展趋势提前布局 SoC 方案是客户的主动规划

头部云服务商在确定网络技术路线时，不仅考虑当前业务需求，更会结合未来 2-3 年 AI 集群规模扩张、算力代际升级及网络架构演进趋势进行前瞻性布局。全功能 DPU SoC 方案具备架构可扩展性与代际延续性，能够支撑客户后续集群向更高带宽、更大规模方向演进，避免因技术路线频繁切换产生重复投入，客户在规划新一代 AI 基础设施时倾向于将 SoC 方案纳入其长期技术架构，提前完成产品验证与生态建设，以匹配未来集群建设节奏。

同时，全功能 DPU 落地部署需要多年的持续投入，算力底层生态适配成本高，客户倾向于与具备成熟产品能力的独立 DPU 厂商建立长期合作关系；公司产品已实现规模化量产与稳定交付，供应保障能力经过头部客户检验，亦系客户选择长期合作对象的重要基础。

5、实际客户转化情况

从公司实际客户转化情况看，已有多家原采用 FPGA 方案的客户转向采用公司全功能 DPU 产品：

（1）腾讯上一代智能网卡基于海外商业 FPGA 芯片方案，现基于公司 DPU 芯片的玄灵智能网卡自 2024 年第四季度起批量交付并商用，2026 年 8 月底累计交付超 12 万颗；

（2）中国移动曾于 2019 年联合联想、赛灵思、Napatech 等合作伙伴展示 5G 接入网云化方案，其中加速硬件设备包括基于赛灵思 FPGA 芯片的 Napatech 智能网卡；2023 年 4 月发布基于 FPGA 方案的磐石 DPU 3.0，同年 10 月磐石 DPU 4.0 的技术概念首次公开。2024 年起中国移动开始采购公司 DPU SoC 产品，2024 年 4 月起先后发布了基于公司 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，大云磐石 DPU 4.0 已于 2025 年落地商用，实现了高达 5000 万 PPS 的网络转发性能，超 100 万 IOPS 存储性能，相比上一代 FPGA 方案产品功耗降低 50%，成本也降低 50%，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；

（3）公司已与其他自研 FPGA 的云服务商对接业务需求且正在导入测试公司产品。此外，公司已与 AI 大模型厂商、新浪、招商银行、工商银行、南方电

网等金融、电力行业客户建立合作，客户结构持续多元化。

（三）对发行人业绩的预期影响及相关风险提示

随着腾讯等曾自研 FPGA 方案客户逐步转向公司全功能 DPU 产品，公司报告期内经营业绩已实现快速增长，2025 年公司营业收入增长至 3.70 亿元，同比增长超过 900%，2026 年上半年营业收入已加速增长至 9.40 亿元。鉴于公司 DPU SoC 产品已在存量客户腾讯、中国移动完成产品验证及规模化部署，未来公司与二者的合作预计持续深化，采购规模有望保持增长。

自研 FPGA 方案新客户拓展方面，公司已与除腾讯外的头部云服务商长期保持业务与技术交流，公司已完成与云服务商 B 关于全功能 DPU 产品的需求对接与技术规格对齐，计划在 2026 年第四季度送测导入产品，并积极与云服务商 A 探讨下一代全功能 DPU 产品的合作。目前，云服务商 A 虽有自研 FPGA 方案，但主要依靠外采 DPU SoC 产品实现大规模部署。公司相关新客户从自研 FPGA 方案转向公司全功能 DPU 的时间节奏与采购规模，受自身资本支出计划、产品验证进度、市场竞争格局、技术路线演进等多方面因素综合影响，存在一定不确定性。若公司有关新客户的技术方案转换节奏不及预期，考虑极端情形下，公司全功能 DPU 产品无法如期对其实现销售，根据公司预测数据，该等情形对公司预测收入的影响总体较小。

公司已在招股说明书“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”中补充披露了“自研 FPGA 方案客户转向公司全功能 DPU 产品不及预期的风险”，具体如下：

“报告期内，部分自研 FPGA 方案的客户已逐步转向采购公司全功能 DPU 产品，该等客户的业务转向系公司未来收入增长的重要来源。自研 FPGA 方案客户转向公司全功能 DPU 的业务拓展进度，受客户自身资本支出计划、产品验证进度、市场竞争格局、技术路线演进等多方面因素的综合影响，公司新拓展替代 FPGA 方案客户形成业绩增长的时间节奏与规模存在不确定性。若该等客户转向不及预期，公司可能面临收入增长放缓等风险，并对公司经营业绩实现产生不利影响。”

三、在 AI DPU 领域，客户选择自研或外采的主要考虑，发行人在该领域的竞争优劣势，发行人 AI DPU 产品后续批量出货的主要影响因素

（一）AI DPU 领域云服务商选择自研或外采的主要考虑

AI DPU 是面向 AI 智算场景设计的高性能网络数据处理器芯片，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，是影响 GPU 集群利用率和 AI 基础设施整体效率的关键组件。

1、云服务商选择自研的主要考虑

截至本回复出具之日，国内头部云服务商仅华为、阿里于 2026 年第二季度开始陆续公布 AI DPU 自研进展，自研 AI DPU 对厂商的技术积累、资金实力与需求规模均有较高门槛，目前仅少数云服务商同时具备上述条件，主要考虑因素预计如下：

（1）具备算力网络技术与高端芯片的研发能力，系自研 AI DPU 的重要技术前提

AI DPU 是高性能网络芯片，在大模型训练与推理场景中，AI DPU 与 GPU 之间需建立协同的数据通路，实现网络数据面与 AI 框架、调度系统、拥塞控制算法、网络协议的耦合。公司凭借在网络芯片领域的持续深耕、独立厂商的中立定位，以及与头部云服务商长期合作中沉淀的需求理解和产品迭代能力，已在 AI DPU 与多品牌 GPU 协同方面形成了技术优势，云服务商若自研亦需要对算力网络技术及高端芯片研发具备深厚积淀。

华为长期投入昇腾系列 AI 处理器研发并配套 CANN 软件栈，同时具备自研网络芯片（含 DPU、交换芯片）的长期积累；阿里平头哥已相继推出含光 800 AI 芯片、倚天 710 通用 CPU 并于 2026 年发布真武 M890 训推一体 AI 芯片及配套的自研 ICN Switch 互联交换芯片，具备 AI 算力芯片、服务器 CPU 及 AI 互联交换芯片的研发能力。对于已经形成自有 AI 算力芯片、互联交换芯片及配套软件生态的华为、阿里而言，自研 AI DPU 可与自有 AI 芯片及软件栈耦合，实现协同效应。

相较而言，其他国内云厂商尚未形成成熟的算力网络技术与 AI 算力芯片研

发能力闭环，即便不考虑 AI DPU 自身较高的研发门槛，该等厂商额外大规模投入自研 AI DPU 实现与外采 GPU 协同，预期收益有限。同时，DPU 行业技术迭代较快，专业 DPU 厂商在产品迭代速度、生态兼容性及规模化交付方面已形成成熟体系，通过外购专业 DPU 厂商产品满足当前 AI 算力大规模部署的紧迫需求，系更具确定性、经济性与效率的选择。

因此，是否自研 AI DPU 取决于云服务商是否具备前述硬件研发基因与扎实技术积累，此系华为、阿里区别于其他云服务商的关键基础。尽管如此，采取自研 AI DPU 的云服务商出于供应链多元化保障考量，也会外采市场成熟 DPU 产品。

(2) 战略投入与人才资源可支撑长周期、高投入的芯片研发

AI DPU 属复杂的系统级芯片，从架构设计、多次流片、软件栈建设到生态适配，研发周期长达数年，前期投入规模大且存在流片失败风险，产品成功需要公司战略层面的长期坚定投入以及相关领域人才资源的积累。华为、阿里基于各自历史渊源长期在硬件领域战略投入，并基于过往自研芯片经验形成了覆盖芯片设计、软件开发、系统协同等建制化的技术团队，对高端网络芯片等高难度、长周期研发项目提供了充足的资源支持。

国内多数云服务商现阶段更倾向于将研发资源聚焦于 AI 应用层高效算力基础设施建设，而非聚焦于网络硬件自研，通过外采成熟 DPU 产品可以满足算力基础设施需求，规避自研芯片的长周期投入风险。

(3) 自身系 AI 算力核心供给方与 AI DPU 大需求方，具备规模化落地场景

自研芯片的研发成本需通过足够规模的部署予以摊薄，自身需求确定性系自研决策的重要前提。华为的昇腾 AI 服务器及超节点方案系在当前国内 AI 算力自主供给占据重要份额；阿里持续大规模建设自有 AI 基础设施并对外提供 AI 算力服务。两家自研 AI DPU 可随其自有 AI 算力体系实现规模化部署，需求体量大、确定性高，能够形成“自研—自用—反馈迭代”的闭环，自研具备经济性基础。

(4) 差异化竞争与自主可控等战略需要

在具备上述能力与条件的基础上，差异化竞争与自主可控等需求侧因素进一

步强化了华为、阿里的自研动机：一方面，在 GPU 算力竞争背景下，更低时延、更高集群效率的网络数据面能力可形成生态技术壁垒；另一方面，自研方案可保障核心网络芯片的技术、供应自主可控。

2、云服务商选择外采的主要考虑

多数头部云服务商、电信运营商、大模型厂商仍选择外采 AI DPU，主要考虑因素预计如下：

（1）技术门槛与研发投入考量

AI DPU 属高性能网络芯片，涉及 RDMA、高速接口、数据平面卸载架构、网络可编程、高级网络特性（如逐包负载均衡、拥塞控制）等复杂技术，产品需在支持大规模集群高速互联的场景下保持稳定性与可靠性，研发门槛较高，且芯片设计、软件栈开发、系统验证的研发投入巨大、周期长。如云服务商自身不具备深厚的算力网络技术积累、高端芯片研发能力与量产工程化经验，自研可能存在流片失败或量产芯片良率低、性能不及预期等风险。而与专业芯片厂商合作可降低资本支出与技术风险。对于多数算力基础设施建设者，自研 AI DPU 在技术、资金、人才资源等方面均不具备成熟条件，预计更倾向于外采第三方产品。

（2）需求紧迫性考量

头部云服务商对 AI 算力基础设施的投资紧迫、项目建设周期短，对应网络芯片等核心硬件的交付时效要求高；AI DPU 与 GPU 主流配比关系 1:1，万卡集群规模部署用量大，外采成熟 DPU 产品可缩短网络方案开发与部署周期，快速支撑 AI 集群上线。

（3）生态兼容性、供应安全性考量

AI 集群组网正走向开放以太网路线（如超以太网联盟 UEC，成员涵盖 AMD、博通、微软、Meta 等主流厂商），与封闭专有网络体系并行竞争。国内智算建设普遍采用多元化算力采购策略，同一集群中多品牌 GPU 并存的部署形态日益普遍，就要求网络层的 AI DPU 中立地兼容多品牌 GPU 的高速互联与集合通信加速。华为、阿里采用自研 DPU 面向自有体系定制，无法满足多数云服务商“多元化 GPU 与算力芯片互联”的开放需求。而独立 DPU 厂商天然中立，与下游客户不存在云计算、AI 智算业务的直接竞争关系，客户网络基础设施不因 GPU 选

型而与特定芯片体系单点绑定，可随多元化算力采购策略灵活组合、分散供应风险，是开放生态下云服务商的更优选择。

（4）性价比与经济性考量

自研路径下，高昂的芯片研发、流片与量产工程化投入由单一使用方分摊，单卡综合成本通常高于外采；且流片失败、良率爬坡风险均需自行承担，同时大规模投入自研 AI DPU 与外采 GPU 协同的模式收益有限。相较而言，而独立 DPU 厂商的产品经多客户规模化验证，成熟度与可靠性高，产品迭代与升级较快，且通过规模化量产摊薄成本，在综合性价比方面对外采客户更具吸引力。

3、当前市场格局及未来趋势

从当前市场格局看，国内下游客户以外采英伟达等国际厂商的 AI DPU 产品为主，仅个别头部云服务商于 2026 年第二季度开始公布自研进展。阿里旗下平头哥于 2026 年 4 月发布首款自研 AI DPU 磐脉 920，最高支持 400Gbps 网络带宽。华为于 2026 年 5 月鲲鹏昇腾开发者大会 2026 公开发布 SP560 系列 800G AI DPU 产品，支撑十万级超大规模集群互联，依托集合通信卸载、高速 RDMA 技术，满足 AI 训练与推理场景下的海量数据传输需求。

总体来看，AI DPU 领域自研与外采的选择取决于客户自身技术积累与高端芯片研发能力、战略规划与人才资源情况、生态建设需要与需求紧迫性等因素。鉴于海外厂商代表英伟达的 ConnectX 系列 AI DPU 的先发优势、在国内占据较高市场份额，预计未来一段时期内，多数国内厂商仍会以外采为主。对于自研 AI DPU 的云服务商，出于供应链多元化保障考量，也会外采市场成熟 DPU 产品。

在算力领域国产替代的大趋势下，公司作为国产独立 DPU 厂商代表，已流片正量产导入的风驰 2.0 AI DPU 具备对标 ConnectX-7 的基础性能（支持单口 400Gbps 速率、低通信时延等）、并重点优化高级智算网络特性，同时在生态兼容性方面，既可以适配英伟达、超威半导体的 CPU/GPU，还将广泛适配昆仑芯、天数智芯、燧原科技、沐曦股份等国产 AI 加速卡，未来有望大规模导入国产算力集群，因而具备广阔市场空间。

当前除云服务商外，头部大模型公司亦在大规模自建或共建算力基础设施，有关模型厂商受限于芯片研发周期与需求紧迫性，外采 AI DPU 或将成为优先选

择，未来有望成为公司重要的增量收入来源。

（二）公司在 AI DPU 领域的竞争优势

1、公司 AI DPU 产品布局情况

公司在高性能全功能 DPU SoC 技术积累的基础上，已面向 AI 智算场景设计并成功流片了支持 400Gbps 网络带宽的风驰 2.0 AI DPU，并已启动下一代支持 800Gbps/1.6Tbps 网络带宽的 AI DPU 研发，以适配下一代 AI 数据中心需求。公司 AI DPU 产品支持 GDR、IBGDA 等技术以及多路径拥塞控制、逐包负载均衡等高级网络特性，可降低 GPU 集群通信时延、提升集群综合算力效率，可覆盖通算、智算服务器前后端网络需求。

2、公司竞争优势

（1）AI DPU 产品性能优势

公司风驰 2.0 AI DPU 在设计规格和性能设定上对标英伟达 ConnectX-7，在 RoCEv2 消息速率指标达到英伟达 ConnectX-7 同等水平，保证了公司产品在关键数据路径上的包处理能力足够适应数据中心/智算中心的各种应用场景。公司风驰 2.0 AI DPU 还重点优化了面向智算大规模网络的高级特性，如端侧逐包负载均衡、乱序重排、基于端到端信用流量控制等。

公司风驰系列 AI DPU 产品基于 RISC-V 多核以及硬件可编程流水线的混合架构设计，具有低功耗优势，能够保证客户大规模、长时间部署的稳定性，有助于客户优化集群 TCO。公司自研指令集的 RISC-V 核，采用创新的多线程技术，具有较强的开放性和通用可编程性，能够灵活适配各类客户自定义网络和存储协议，针对性地满足 AI 智算训练及推理场景对高能效、低延迟和灵活客户自定义数据处理的核心需求。公司 AI DPU 针对 RISC-V 处理器自研了大量专有指令，在复杂业务场景处理数据包时的“动作”相较采用标准指令集明显缩减，可节省约 85%的工作量，并且可以组合支持各种不同的业务级加速，能够有效提升数据转发速度和网络吞吐能力。

（2）全功能 DPU 的技术积累与 AI 网络关键技术能力优势

公司系国内率先实现 400Gbps 高性能 DPU 技术突破的独立厂商，公司云霄

1.0 系列全功能 DPU 产品关键性能对标英伟达在售主力产品 BlueField-3。风驰系列 AI DPU 产品系在公司成熟的全功能 DPU SoC 平台上针对性扩展 AI 网络能力而来，技术基础扎实、产品成熟度高。公司掌握 GDR、IBGDA 及多路径拥塞控制等 AI 网络关键技术，在国产 AI 网络协议生态建设中具有先发优势。

（3）开放生态与供应保障优势

作为采用开放生态的独立 DPU 厂商，公司风驰系列 AI DPU 产品广泛与国内外主流的 CPU 平台、GPU 平台以及网络交换机适配，并对主流操作系统和软件平台进行兼容性设计，有助于降低大型云服务商规模部署的适配、开发难度。公司首款全功能 DPU 云霄 1.0 已实现规模化量产，供应保障能力与规模部署可靠性已经过头部客户检验。

（4）在客户智算场景的验证积累与国产替代优势

公司 AI DPU 尚未规模销售但全功能 DPU 已先行在头部客户 AI 智算场景中实际部署验证、积累了丰富的可复用经验。云霄 1.0 全功能 DPU 已量产部署于腾讯 AI 服务器，用于云端调度的智算业务；与中国移动联合发布的智算琢光 DPU 已完成智算网络及主流交换芯片的适配验证，磐石 DPU 已应用于国产 GPU 超节点方案前端网络；招商银行方面，公司已完成云霄 1.0 全功能 DPU 在其智算服务器的适配测试及小批量出货。此外，在国产化需求与政策支持下，公司作为领先的国产 DPU 独立厂商，在 AI 智算国产替代进程中具有先发优势。

3、公司竞争劣势

（1）AI DPU 规模化验证与应用时间相对较短

公司 AI DPU 产品完成流片及市场推广时间较短，在大规模 AI 集群中的规模化部署验证数据积累相对有限，相较英伟达等国际巨头在超大规模 AI 集群中的长期验证存在差距，客户对新产品完成充分验证需要一定周期。

（2）头部厂商生态绑定程度较高

英伟达构建了“GPU+DPU+InfiniBand 专用网络技术”的完整生态，其 AI 网络方案与 GPU 平台深度耦合，公司作为独立 DPU 厂商，在英伟达 GPU 生态中的适配与协同深度弱于英伟达 ConnectX 系列产品，且英伟达 ConnectX-9 已支

持 2*800Gbps 网络带宽，公司新一代产品尚在研发中，与国际头部厂商在代际上存在差距。

同时，华为亦依托昇腾 AI 处理器、自研 DPU、数据中心网络设备及 CANN 软件栈，构建了覆盖“算力芯片+网络芯片+网络设备+软件生态”的全栈自主技术体系。其自研 DPU 与昇腾服务器深度耦合，公司作为独立 DPU 厂商深度参与该全栈生态内部的方案协同具有一定的门槛，在华为生态所覆盖的 AI DPU 客户群体中面临一定生态进入挑战。但华为对外销售服务器的 DPU 选型独立于服务器采购决策，公司仍可通过华为整机客户配套及整机厂商合作进入华为生态市场。

（三）公司 AI DPU 产品后续批量出货的主要影响因素

1、产品技术成熟度与验证进度

AI DPU 产品需在真实 AI 集群环境中完成大规模、长时间的场景测试与性能打磨，包括与主流 GPU 平台的互联适配、与主流网络协议的互通验证、大规模分布式训练场景下的稳定性验证等。公司 AI DPU 产品正处于客户验证与逐步放量阶段，产品在万卡级集群中的规模化验证进度将直接影响后续批量出货的时间与规模。当前，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

2、客户验证与导入周期

AI DPU 属系统级芯片，客户导入需经历产品测试、平台适配、小批量验证、规模部署等阶段，验证与导入周期较长；客户在 AI 网络核心环节对产品的可靠性、确定性要求高，通常需在真实业务场景中完成充分验证后方可批量采购。公司已与运营商、AI 大模型厂商、金融等行业客户建立合作并开展适配验证，相关客户验证与导入进度将影响 AI DPU 产品批量出货节奏。

目前，公司风驰 2.0 产品已在中国移动研究院完成智算网络协议的适配验证测试，预计将作为支持其智算网络协议的 AI DPU 进入中国移动智算服务器集采标准配置清单。同时，风驰 2.0 产品已完成国产知名大模型厂商导入测试，产品符合客户应用要求，正在积极推进灰度部署订单，预计 2027 年进入规模化采购阶段。此外，风驰 2.0 产品在腾讯等头部云服务商、其他电信运营商、网络通信设备商的导入测试工作亦在积极推进中。

3、下游 AI 算力基础设施建设与资本支出节奏

当前全球 AI 智算及云计算基础设施建设正处于高速扩张期。AI DPU 产品的批量出货与下游 AI 智算基础设施建设进度高度相关。国内 AI 算力中心、智算集群建设受政策引导、产业投资周期影响较大，若下游 AI 智算基础设施建设进度加快，将直接拉动 AI DPU 产品需求；反之，若下游资本支出收缩或建设进度放缓，公司 AI DPU 产品的批量出货将相应延后。

4、供应保障能力

AI DPU 产品有关晶圆产能、封装测试资源的供应保障直接影响产品的量产交付能力。公司需确保晶圆产能的稳定供应，并建立多源供应保障体系，以满足客户批量出货需求；若供应链出现产能紧张或供应中断情形，将影响 AI DPU 产品的批量出货进度。

5、政策与国产化需求

国家政策持续支持算力基础设施自主可控，《算力基础设施高质量发展行动计划》明确提出开展 DPU、无损网络等技术升级与试点应用。在 AI 算力国产化进程中，头部云服务商、运营商、金融、电力等行业客户对国产 DPU 产品的采购需求持续增长。政策支持力度与国产化推进节奏将影响公司 AI DPU 产品在运营商及行业客户中的批量出货进度。

6、市场竞争与生态格局

AI DPU 领域竞争格局对公司批量出货具有重要影响。英伟达通过生态绑定占据主导地位，其新一代产品迭代将持续强化生态壁垒；华为、阿里近期公布的自研方案将分流部分国内市场空间；国内其他 DPU 厂商亦在 AI 网络领域加快布局。公司 AI DPU 产品的规模化出货，需在开放生态兼容、国产协议标准、性价

比等方面持续构筑竞争力。

综上所述，公司 AI DPU 产品后续批量出货受产品技术成熟度、客户验证导入周期、下游 AI 算力基础设施建设与资本支出节奏、供应保障能力、政策与国产化需求及市场竞争格局等多方面因素综合影响。公司将持续推进 AI DPU 产品在头部客户 AI 智算场景中的验证与部署，加快新一代产品研发，积极把握 AI 基础设施建设的市场机遇。

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报告及相关行业研究报告，核查全球及中国 DPU 市场规模、增速及技术路线结构数据及发行人市场地位数据，AI DPU 领域自研与外采格局及主要竞争方案情况；

2、查阅腾讯、字节跳动、阿里等客户及行业厂商的公开技术分享、产品发布会信息，了解头部云服务商自研 FPGA 方案及 SoC/ASIC 方案演进情况，客户由 FPGA 方案转向发行人 SoC 方案的具体情况与性能对比；

3、查阅英伟达官方发布的技术资料，核查其 DPU 产品路线及 SoC 架构情况；查阅 FPGA 器件厂商公开产品资料及行业功耗对比技术数据，核查 FPGA 方案性能上限相关论据；

4、查阅发行人说明，了解发行人 DPU SoC 产品技术指标、客户验证及量产情况，了解发行人技术路线选择依据及行业判断逻辑，了解客户选型考虑因素、客户选择自研或外采的考虑因素、发行人竞争优势、AI DPU 产品批量出货的影响因素以及发行人 AI DPU 产品布局、技术指标及研发进度；

5、查阅发行人招股说明书，核查有关风险提示的充分性与准确性；

6、取得发行人销售明细及主要客户合同/订单，了解发行人 AI DPU 产品客户验证与导入进度，核查发行人主要客户销售情况、收入确认及客户转化情况。

二、核查意见

经核查，保荐人认为：

1、发行人已充分披露下游云服务商等客户自研 FPGA 方案情况；在全功能 DPU 领域，FPGA 方案性能上限预计无法完全满足行业主流数据中心当前及未来需求；发行人预计 FPGA 方案受限于带宽与功耗难题、DPU SoC 演进路线将成为主流路线具有合理依据；

2、发行人全功能 DPU 产品与客户自研 FPGA 方案相比具有技术、成本等方面的优势；发行人已充分披露自研 FPGA 方案客户转向发行人全功能 DPU 的考虑因素及对发行人业绩的预期影响，并已在招股说明书提示有关风险；

3、当前国内仅少数同时具备算力网络技术、高端芯片研发能力及规模化自用需求的头部云服务商满足自研 AI DPU 门槛，多数云服务商、运营商及大模型厂商基于技术门槛、研发投入、需求紧迫性、生态兼容性及性价比等因素仍将以外采为主；发行人在 AI DPU 领域具备产品性能对标英伟达 ConnectX-7、RISC-V 低功耗架构、全功能 DPU 技术积累、开放生态兼容及国产替代先发等竞争优势，但亦存在规模化验证时间较短、头部厂商生态绑定较强等竞争劣势；发行人 AI DPU 产品后续批量出货主要受产品技术成熟度与验证进度、客户验证与导入周期、下游 AI 算力基础设施建设与资本支出节奏、供应保障能力、政策与国产化需求以及市场竞争与生态格局等多重因素综合影响。

问题 3. 关于客户集中

申报材料及首轮问询回复显示：因下游云服务厂商集中度较高、且腾讯急需高性能国产 DPU 产品以优化算力基础设施，发行人对头部云服务厂商客户腾讯的收入占比较高。发行人在腾讯数据中心应用过程中积累了可复用的技术经验与工程能力。

请发行人披露：

(1) 结合下游行业需求分布情况、DPU 产品批量应用的主要场景需求等，进一步分析发行人因下游行业集中度较高导致客户集中的合理性，相关依据充分性。

(2) 结合竞品情况、腾讯选择其他 DPU 方案或产品的可行性及商业考虑、DPU 产品及国产化供应链对腾讯的必要性，进一步分析发行人是否存在单方面依赖腾讯的风险，发行人对腾讯收入占比较高是否构成重大不利影响。

(3) 发行人在腾讯数据中心应用过程中积累的可复用技术经验在其他客户场景中的具体体现。结合发行人其他头部客户开拓及客户认可情况，分析发行人以腾讯为种子客户形成品牌背书、市场影响力策略的有效性，并进一步分析发行人未来降低关联交易的可行性。

请保荐人、申报会计师、发行人律师说明核查依据、过程，并发表明确意见。

【回复】

一、结合下游行业需求分布情况、DPU 产品批量应用的主要场景需求等，进一步分析发行人因下游行业集中度较高导致客户集中的合理性，相关依据充分性。

(一) DPU 产品批量应用的主要场景和下游行业需求分布集中

DPU 产品批量应用集中于云计算与 AI 智算两大场景：

1、云计算场景

云计算场景是 DPU 产品最核心的下游应用领域之一。在数据中心的云服

务商通过虚拟化技术将物理服务器的计算、网络和存储资源池化，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。全功能 DPU 主要负责在服务器中卸载 CPU 任务（包括网络、安全、存储、虚拟化等功能），释放 CPU 性能，降低数据中心总体运营成本。

2、AI 智算场景

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。

上述两大场景的核心参与者，主要集中于云服务商与电信运营商。云服务商深耕大规模算力基础设施与云服务输出，向云计算、AI 智算延伸具备天然的业务协同优势；电信运营商则依托广泛的网络资源与机房基础设施，以云网融合为战略支点，深度嵌入算力基建与运营。根据弗若斯特沙利文的数据，2025 年中国 DPU 市场下游需求中，二者合计占比约为 78.0%，需求结构相对集中；预计至 2030 年，该占比将进一步提升至约 90.0%，市场集中趋势显著。

（二）云服务商和电信运营商市场高度集中

1、云服务商

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。随着云计算基础设施规模不断扩大以及算力密度持续提升，云服务商通过在服务器前端网络 and 后端网络侧广泛引入 DPU，以提升整体资源利用效率和平台稳定性。同时，云服务商亦通过对外提供云基础设施与云原生服务，将自身数据中心能力转化为面向企业与开发者的服务输出。内部基础设施建设与对外云服务供给的双重需求，使云服务商成为推动中国 DPU 市场发展的核心力量。

我国云服务市场高度集中，头部云服务商凭借在基础设施规模、技术能力及生态体系方面的综合优势，占据了市场的主要份额，从下游竞争格局看，我国云服务市场呈现“巨头主导”特征，根据弗若斯特沙利文数据，按收入统计 2025 年中国云服务 TOP5 厂商及电信运营商合计占据公有云市场约 82.15% 的市场份额。

2、电信运营商

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。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。在这一过程中，DPU 被广泛用于支撑电信级云平台的低时延网络、虚拟化、安全隔离及存储加速等关键数据面功能。同时，电信运营商在推进自身业务系统云化与智能化升级的过程中，也不断提升对数据中心效率与系统稳定性的要求。面向外部客户的云服务需求与内部系统升级需求相互叠加，使电信运营商成为 DPU 市场中具备持续配置能力的重要需求方。

我国电信运营商市场高度集中，主要由中国移动、中国联通、中国电信三大电信运营商主导。

（三）公司客户集中具有合理性

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。

从 DPU 应用的主要场景和下游行业需求分布来看，DPU 产品作为高性能服务器的关键组件，云计算和 AI 智算是最主要的两个应用场景，在我国云服务市场和电信运营商高度集中的竞争格局下，国产 DPU 厂商优先服务头部云服务及电信运营商客户，符合行业客观发展规律。

头部云服务及电信运营商客户对 DPU 产品的性能指标、稳定性、软硬件生态兼容性、供货保障能力均设置严苛准入标准，产品从初步送测导入到实现规模量产，需要历经样品测试、软硬件适配、灰度测试、大规模量产等多轮环节，项目周期长、资源投入消耗大。公司因早期发展阶段研发人员、资金等资源相对有限，难以同步铺开对多个头部客户的并行导入，为保证首款产品的成功，公司选择集中资源优先攻坚头部“种子客户”，导致阶段性客户集中度较高，具有客观合理性。公司选择腾讯作为“种子客户”主要因为：

一是与 GPU 侧重单芯片计算性能有所不同，DPU 作为服务器及算力集群的系统级“管家”，需要支持客户多元化、复杂业务场景下的基础设施任务卸载和数据处理需求，且具备软硬件深度耦合的技术属性。因此，DPU 产品成熟一方面需要在设计阶段基于对行业及市场需求的充分理解以完善功能定义与规格标准，另一方面则需要在终端真实、全面的业务场景中持续迭代打磨。腾讯拥有国内规模领先、场景复杂度极高的云服务和 AI 智算业务，具备体量庞大、品类丰富、高并发与高时效的业务流量与数据处理要求。公司与腾讯开展业务合作，有利于首款全功能 DPU 完善功能定义、优化产品性能，并在高规格、高复杂度的真实业务场景下完成产品打磨，加速商业化进程。

二是腾讯作为国际领先的科技企业，业务覆盖云服务、多模态大模型、AI 智能体、数字支付、音视频、社交、游戏等多元场景，其业务规模、并发复杂度均处于行业第一梯队。海量用户访问、高频金融交易、大模型高吞吐推理、实时消息流转等业务特征，对 DPU 的性能及长期运行可靠性提出了严苛要求。公司选择与腾讯开展业务合作，是对自身 DPU 产品性能与可靠性的充分检验，能够

快速打磨商业化能力，形成品牌背书、市场影响力。

伴随首款全功能 DPU 在腾讯规模化量产及多场景商用，公司完成产品打磨、经验积累与标杆案例建设，并向国内其余头部云服务商、电信运营商领域的客户推广产品，目前公司在国内其余头部云服务商、电信运营商领域的客户开拓工作已取得实质进展，部分项目已实现落地，预计未来公司客户集中度水平将有所下降。

二、结合竞品情况、腾讯选择其他 DPU 方案或产品的可行性及商业考虑、DPU 产品及国产化供应链对腾讯的必要性，进一步分析发行人是否存在单方面依赖腾讯的风险，发行人对腾讯收入占比较高是否构成重大不利影响。

（一）DPU 产品对腾讯的必要性

DPU 是人工智能时代与 CPU、GPU 并列的数据中心“三大主力芯片”，也是支撑国家“十五五”期间构建核心“六张网”之一算力网络的关键一环。DPU 作为提升算力网络运载能力、降低算力综合成本的关键底座，能够为 AI 智算、云计算打造一条“信息高速公路”，推动人工智能与数字经济产业的快速发展。

腾讯作为国内业务版图最为多元的头部科技企业之一，业务覆盖云计算、AI 大模型、社交与通讯、网络游戏、网络广告、金融科技及企业服务等多个领域，各业务板块均依托自建的大型数据中心及算力基础设施运行，对 DPU 产品的需求贯穿云计算、AI 智算等多个核心应用场景。作为国内领先的云服务商及 AI 大模型厂商，腾讯面临着数据中心大规模扩容及算力效率提升的迫切需求。近年来，腾讯算力基础设施建设已实现从千卡到万卡再到十万卡的快速跃迁，2026 年腾讯云宣布高性能智算集群 HCC 完成新一轮升级，已全面支持十万卡规模 AI 大模型训练，可支持万亿级参数规模的大模型训练，网络有效带宽利用率和稳定性均达到行业领先水平。根据美银证券最新研究报告，腾讯 2026-2028 年资本支出预计达到 1,850 亿元、2,250 亿元及 2,500 亿元。按照服务器价值占比 60%，全功能 DPU 及 AI DPU 合计价值量占比 7.5%，估算腾讯 2026-2028 年 DPU 采购规模将分别达到 83.25 亿元、101.25 亿元和 112.50 亿元，需求持续提升。

DPU 是腾讯算力基础设施升级的关键硬件，在云计算场景下，能够将网络虚拟化、存储、安全等基础设施负载从主机 CPU 卸载，释放 CPU 算力用于业务

处理，提升云主机可售卖算力，优化数据中心整体 TCO，改善网络吞吐与业务时延，支撑自身高并发业务运行；在 AI 智算场景中，DPU 依托 RDMA 高速互联、可编程拥塞控制能力，缓解大规模 GPU 集群通信瓶颈，提升 GPU 利用率，更好满足腾讯的混元大模型训练、推理的算力网络需求。

全功能 DPU 方面，腾讯当前供应包括自研 FPGA 方案（银杉系列）及外采独立厂商 DPU SoC 产品，公司云霄 1.0 全功能 DPU SoC（玄灵系列）2026 年占据较大份额。AI DPU 方面，腾讯当前主要采购海外产品，公司风驰 2.0 系列 AI DPU 正在测试导入。

（二）DPU 国产化对腾讯的必要性

长期以来，国内高端 DPU 市场主要被英伟达等海外厂商占据。公司自成立以来坚持行业主流方向的 SoC 技术路线，通过自主研发与技术创新，围绕 DPU 核心技术、应用场景构建了软硬件一体化的全栈产品体系。2023 年，公司率先推出面向多场景、最高支持 400Gbps 网络带宽的国产全功能 DPU SoC 芯片，关键性能指标对标英伟达主流产品，成功打破海外国际大厂在国内市场的垄断，补齐算力网络“卡脖子”环节，对夯实我国数字产业根基、保障算力供应链安全、推动智算基础设施国产替代，具备重大产业与战略价值。

在实现国家算力基础设施自主可控的产业大背景下，国产 DPU 的技术突破同时也为腾讯等国内头部云服务商和电信运营商提供了可靠的本土供应链备选方案。具体而言，推进 DPU 供应国产化对腾讯具有重要意义：

一是供应链稳定与及时性。受益于人工智能产业的突破发展，全球算力基础设施建设进入快速扩张周期，带动云计算、AI 智算等场景对 DPU 产品需求呈现爆发式增长。受全球旺盛需求影响，海外头部厂商通常优先服务北美市场，国内市场部分 DPU 产品出现交付周期拉长、供货紧张、售后维护不及时的情况，一定程度上影响包括腾讯在内的国内云服务厂商大规模算力集群的建设节奏。相较于海外供应商，国内 DPU 本土厂商在供应链保障、交付排期及运维支持等方面具备明显优势，能够更好匹配国内云厂商快速扩容的建设节奏，降低供应链交付和售后维护不确定的风险。

二是供应链安全。近年来，受地缘政治影响，美国政府陆续出台多项法令以

限制向国内出口先进芯片和相关技术，当前主要围绕支持人工智能产业的高端算力芯片领域。尽管 DPU 是数据中心三大主力芯片之一，但因其本身属于网络芯片，尚未在当前制裁的聚焦范围。在中美科技博弈持续深化的背景下，境外 DPU 采购渠道可能受阻，因此推动 DPU 国产化对腾讯的供应链安全至关重要。

（三）综合竞品情况、其他 DPU 方案或产品的可行性及商业考虑，公司的 DPU 产品是腾讯的最优选择

结合技术方案、竞争格局及腾讯自身需求，预期腾讯关于 DPU 技术方案有以下选择：

序号	方案		可供选择的供应商
1	FPGA 方案		腾讯自研 FPGA 方案（银杉）
2	SoC/ASIC	海外厂商	英伟达（BlueField 和 ConnectX）、博通（Thor 系列）DPU 等
3		国内其他厂商	华为或国内其他独立厂商 DPU
4		公司	云霄和风驰系列 DPU

基于产品的可行性及商业考虑，腾讯采购公司 DPU 产品是最优选择，具体分析如下：

（1）从 DPU 方案或产品可行性角度分析，传统 FPGA DPU 方案在功耗和成本方面的约束逐步显现，采用 DPU SoC/ASIC 成为腾讯的更优选择

FPGA DPU 通常作为全功能 DPU 的过渡方案，采用 FPGA 配合外挂 CPU 的实现方式，由 FPGA 承担网络、存储与安全等数据面功能，CPU 负责控制与管理逻辑；DPU SoC 则采用 ASIC 实现路径，在单芯片内集成高性能 CPU 核和专用处理单元，将数据面处理与控制面管理功能高度融合，在性能、能效及系统稳定性方面具备优势，更适合规模化部署及长期稳定运行。

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 FPGA 方案网卡难以满足自身云基础设施持续升级需要，采用符合行业主流发展趋势的 DPU SoC 方案成为必由之路，腾讯未来回到 FPGA 方案为主的可能性较低，具体分析如下：

一是 FPGA 在网络带宽 400G 时已碰到物理和功耗瓶颈，带宽达到 800G 时的瓶颈将更加明显。在 800G 带宽下，FPGA 的性能、功耗和成本预计难以承受，

而 DPU SoC 可通过先进制程和单芯片集成有效解决这些问题；

二是腾讯与公司的合作已取得显著成效。截至 2026 年 8 月末，玄灵网卡已部署超 12 万卡，双方技术互信与生态合作较深。腾讯已实现了从 FPGA 到 SoC 的顺利转型，2026 年基于公司首款 400G DPU SoC 的玄灵智能网卡供应预计占同类产品采购较高比例；

三是腾讯自研的 FPGA 方案网卡在 400G 网络带宽阶段已遇到性能及功耗瓶颈，英伟达 DPU 一直采取 SoC 方案，最新一代 BlueField-4 系列全功能 DPU 已支持 800G，目前行业内尚无成熟商业化 800G FPGA 方案发布，说明 FPGA 方案在高性能场景竞争中处于劣势。腾讯选择与公司在下一代 800G 产品合作，是基于技术路线演进和成本效率的理性选择。对于下一代 DPU 产品规划，腾讯计划继续与公司产品合作迭代玄灵，预计不会在规模部署中转回 FPGA 方案。

（2）从商业考虑和供应链安全角度分析，腾讯将优先选择公司产品

与 GPU 不同，以公司产品为代表的国产 DPU 与国际在售主力产品 Bluefield3 等在综合性能方面不存在明显代差，可适配主流服务架构的同时更具性价比：

①全功能 DPU 领域

在全功能 DPU 领域，腾讯当前主要采用公司的云霄 1.0 全功能 DPU SoC，采购境外厂商的全功能 DPU 较少。

公司云霄 1.0 全功能 DPU 在综合性能、处理器性能、网络带宽及可编程性等方面已达到英伟达主力在售 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，同时价格相对较低，综合性价比优势明显。

②AI DPU 领域

在 AI DPU 领域，腾讯当前主要采购海外厂商产品，公司风驰 2.0 系列 AI DPU 当前正在测试导入。

公司风驰 2.0 系列 AI DPU 与英伟达 ConnectX-7 在通信延迟、网络带宽等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，同时重点优化针对智算领域的高级网络特性（如逐包负载均衡、CBFC 拥塞控制等），且价格

更具市场竞争力。

此外，全球 AI 算力扩张带动 DPU 市场需求大幅增长，海外头部厂商产品在国内市场存在交期拉长、供货短缺及售后维护不及时等问题，本土 DPU 厂商在供应链保障、交付排期、本地化运维支持方面具备突出优势。同时，在中美科技博弈持续深化的背景下，境外高端 DPU 芯片未来存在被纳入出口管制的政策风险。因此，在产品性能对齐、成本优势显著的情况下，采购国产高性能 DPU 以替代海外厂商产品，是腾讯降低国际供应商依赖、保障供应链安全的关键战略举措。

综上所述，在同等产品性能条件下，公司 DPU 具备突出的性价比优势，同时基于供应链自主可控需求，腾讯将优先选择公司产品。

(3) 从国内市场竞争格局及竞品情况分析，腾讯将优先选择公司产品

国内 DPU 行业发展尚处于初期阶段，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。公司自成立以来深耕 DPU SoC 技术路线，依托具备深厚大芯片研发积累和丰富行业经验的顶尖技术团队，以及持续的高强度研发投入，于 2023 年率先推出国内领先的基于 SoC 方案、支持 400Gbps 高速网络的全功能 DPU 芯片，核心指标对标国际大厂主流产品，并已实现在头部云厂商的大规模量产与多场景商用，有望填补国内高性能 DPU 的市场空白。

在全功能 DPU 领域，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。根据公开信息查询，国内其他第三方厂商在全功能 DPU 领域主要采用 FPGA 方案或为最高支持 200Gbps 的 SoC 产品，与公司产品在技术路线和产品定位上存在差异。因此，相较公司云霄 1.0 全功能 DPU SoC，国内其他厂商暂不具备提供同等性能、符合腾讯实际需求的产品。

在 AI DPU 领域，华为相关产品偏向封闭生态体系，基本采用私有协议实现网络加速；阿里旗下平头哥于 2026 年 4 月 28 日发布首款自研 AI DPU 磐脉 920，主要面向自供。相较同为云服务商的阿里、华为，公司作为芯片厂商更具中立性，产品灵活性与生态兼容性也更具优势。国内当前暂无已实现规模化部署、支持 1*400Gbps 带宽规格的独立厂商成熟 AI DPU 产品，因此现阶段国产 GPU 集群

组网仍大量依赖英伟达等海外厂商的 AI DPU。对于国产 AI DPU，腾讯正优先导入公司风驰 2.0 产品，当前已完成产品性能测试、正推动在智算服务器整机厂商的机型导入。

综上，无论是在全功能 DPU 领域还是在 AI DPU 领域，公司均为腾讯在国内独立 DPU 供应商的最佳选择。此外，公司与腾讯合作基础扎实，技术粘性与产品迭代共同保障双方合作长期稳定。公司与腾讯自 2021 年起即开展合作，双方彼此熟悉对方的技术能力与业务需求。若腾讯更换 DPU 供应商，需重新经历产品定义、功能开发、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。据此分析，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，公司与腾讯持续推进多款新一代 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，充分考虑大型云服务商、运营商基于 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，与腾讯继续在该 DPU 产品开展合作。

因此，结合国内 DPU 市场竞争格局及竞品供给情况，腾讯优先选用公司产品具有合理性。双方合作基础扎实、具有较高的技术粘性，预计未来业务合作将保持稳定。

（四）双方是基于业务发展的双向选择而非公司单方面依赖腾讯，公司对腾讯收入占比较高不构成重大不利影响

从腾讯角度来看，高性能 DPU 是布局下一代数据中心架构、建设高效算力基础设施的必要选择，而在可选方案中，自研 FPGA DPU 在性能、功耗及成本方面的约束逐步显现，难以满足自身云基础设施的持续升级需要；海外产品的性价比较低，且在全球 DPU 市场需求大幅增长、中美科技博弈持续深化的背景下，过度依赖国际厂商会面临潜在 DPU 产品交付不确定、售后维护不及时、供应链安全等风险；国内其他主流 DPU 第三方厂商暂不具备提供满足腾讯性能要求、生态兼容性需要的 400G DPU SoC 产品。

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%。前述服务器性能的显著提升帮助腾讯增强云业务领域的市场竞争力，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、保障供应链安全的最佳选择。

从公司角度来看，高端芯片企业在发展初期优先选择头部“种子客户”完成产品打磨、落地标杆案例，依托头部客户形成市场示范效应，再逐步向全行业拓展、实现规模化商业化，是国内高端芯片行业较为通行的商业化路径。海光信息、燧原科技等算力芯片企业在商业化起步阶段，均存在对单一或少数种子客户依赖度较高的特征，客户集中度偏高是高端芯片企业发展初期阶段较为普遍的行业现象。腾讯作为国内云服务与 AI 应用领域的领军企业，同时也是国内算力基础设施的核心投资主体，是国产高端 DPU 产品落地验证的优质优先合作对象。在此行业背景下，公司针对首款 DPU 产品采取“以点带面、重点突破”的市场化拓展策略，将腾讯定位为云服务商领域的核心种子客户，在商业化早期阶段集中内部研发、技术支持、交付保障等资源向该种子客户倾斜，一方面有利于推动公司首款 DPU 产品在头部云厂商环境下快速实现规模上量，在真实业务环境中持续打磨产品性能、完善商业化交付能力；另一方面借助头部客户的大规模商用落地形成品牌背书与市场影响力，为后续进一步推进国内 DPU 领域国产替代、持续抢占市场份额筑牢基础。

综上，双方合作是基于业务发展的双向选择，而非公司单方面依赖腾讯，公司对腾讯收入占比较高不构成重大不利影响。

三、发行人在腾讯数据中心应用过程中积累的可复用技术经验在其他客户场景中的具体体现。结合发行人其他头部客户开拓及客户认可情况，分析发行人以腾讯为种子客户形成品牌背书、市场影响力策略的有效性，并进一步分析发行人未来降低关联交易的可行性。

（一）发行人在腾讯数据中心应用过程中积累的相关技术经验的可复用性，及在其他客户场景中的具体体现

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，商业化初期与腾讯开展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：

1、优化产品定义及功能特性

DPU 芯片设计阶段需要对大量可选功能做取舍权衡，若功能配置不足，难以覆盖头部客户的主流业务诉求；若集成过多功能，会增加芯片面积、提高功耗，影响整机部署效果。腾讯作为国内业务版图最为多元的头部科技企业之一，业务覆盖云计算、AI 大模型、社交与通讯、网络游戏、网络广告、金融科技及企业服务等多个领域，各业务板块主要依托自建的大型数据中心及算力基础设施运行，对 DPU 产品的需求贯穿云计算、AI 智算、存储、网络安全等多个核心应用场景。在与腾讯的合作过程中，公司充分理解头部云服务厂的核心业务需求，进而优化 DPU 芯片的核心功能定义、规格要求，避免出现功能不足或者特性缺失等问题。公司将提炼后的下游客户通用核心需求落实到芯片架构、硬件规格以及软件栈的研发设计当中，使得公司产品能够充分适配其他头部云服务商、电信运营商的共性业务场景，减少面向新客户的二次适配开发工作，降低客户导入难度与验证成本，推动同款通用可编程 DPU 在不同客户的应用落地。

2、产品性能优化

腾讯作为大型云服务商，对公司 DPU 产品的各项性能、稳定性均提出严苛标准。双方合作过程中，腾讯要求公司针对复杂业务场景下的网络时延痛点开展针对性的产品性能优化。例如，公司基于腾讯云对数据包转发处理方面的高标准、多元化业务需求，在首款云霄 1.0 全功能 DPU 芯片集成 64 个自研指令集的 RISC-V CPU 核，并采用创新的多线程技术，实现网络数据的可编程与加速处理。

具体而言，公司结合腾讯业务需求与优化方向，针对 RISC-V 处理器自研多条专有指令，使其在复杂业务场景处理数据包时的“动作”较采用标准指令集时的步骤明显缩减、节省约 85%的工作量，并且可以组合支持各种不同的业务级加速，提升处理数据转发速度。前述基于自研 RISC-V 指令集的网络加速能力，一方面充分满足腾讯云的高标准要求，另一方面复用在其他客户的云业务场景也取得良好效果，有利于公司持续开拓云服务商、电信运营商客户。例如，中国移动基于公司云霄 1.0 芯片的磐石 DPU，实现了每秒百万级数据包处理能力、5 微秒 RDMA 低时延，显著提升数据中心基础设施的部署效率和使用效益。

3、高可靠性量产部署能力

公司基于腾讯运营超大规模数据中心对稳定性、故障率的严苛标准，一方面将相关要求在产品开发时即纳入 DPU 软硬一体化设计 RAS（可靠性、可用性、可维护性）特性内，另一方面在量产阶段持续打磨 DPU 软件以进一步提升产品部署的稳定性。具体而言，公司针对腾讯要求而优化的 DPU 软硬件方案，可以提升大规模数据中心的管控运维效率，包括实现无感上下电、故障一键收集、丢包溯源诊断等，同时降低设备故障率、增强自动修复能力，包括对 PCIe 总线、内存及接口等关键器件的 RAS 异常处理与故障率优化，尤其是降低规模化 DDR5 的异常宕机率。前述可靠性量产与部署能力对公司拓展与服务下游大型客户、进入超大规模数据中心起到重要作用。

（二）结合发行人其他头部客户开拓及客户认可情况，分析发行人以腾讯为种子客户形成品牌背书、市场影响力策略的有效性，并进一步分析发行人未来降低关联交易的可行性

目前，公司以腾讯为种子客户形成的品牌背书和市场影响力策略已取得显著成效。经过数年深耕，公司已成为中国移动、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。

下游头部客户在对外采购 DPU 产品、遴选供应商时，核心考量主要包括：

一是产品功能适配性。客户需确认 DPU 功能规格能否精准匹配复杂业务场景需求，切实解决虚拟化卸载、网络加速、存储卸载等核心痛点，实现算力效率

提升与整体成本、功耗的优化；

二是产品性能与稳定性。头部客户对 DPU 在实际运行中的吞吐量、时延、IOPS 等关键性能指标及长期稳定性要求严苛，须确保高负载、复杂场景下的持续可靠表现；

三是供应商的量产部署能力。DPU 采购涉及供应链安全与业务连续性，加之 AI 算力基础设施建设窗口期有限，客户对供应商的量产一致性、产能爬坡及交付时效、量产部署后稳定性与运维成本要求极高。

腾讯作为国内业务版图最为多元的头部科技企业之一，业务覆盖云计算、人工智能、社交与通讯、网络游戏、网络广告、金融科技及企业服务等多个领域，对 DPU 产品的需求贯穿多个核心应用场景。同时，腾讯的数据中心在服务器规模、业务并发量、可靠性要求等方面均处于行业标杆水平，对 DPU 产品的性能指标、稳定性、可维护性及供货保障能力提出了严苛的要求。因此，针对上述核心考量，公司与腾讯的成功合作案例已形成充分品牌背书，有效降低新客户信任建立成本与导入门槛：

一是腾讯数据中心的部署案例成为公司技术实力、产品功能适配性、产品性能及可靠性的有力证明，显著降低了其他客户对公司产品的验证成本与决策顾虑；

二是在服务腾讯大规模落地的过程中，公司沉淀形成了完整的商业落地的全流程体系和量产管控能力，构建起标准化的客户导入工具链与问题排查知识库。面对其他客户需求时，可复用已经过大规模现网验证的基础软件组件、调优方法论以及运维处置经验，减少重复开发与从零试错的工作量，进而缩短量产周期。

三是腾讯作为数据中心产业链的链主企业，对上下游供应商具有较强的辐射带动作用，其供应链认证与采用行为在产业链上下游具有一定的公信力。公司在服务腾讯的过程中，与国内多家头部服务器厂商开展产品适配、技术磨合，为服务其他客户奠定基础。

综上，公司与腾讯的成功合作案例已形成充分品牌背书，有效降低新客户信任建立成本与导入门槛，并借助链主供应链传导与腾讯多元场景沉淀的全场景方案能力，成功开拓电信运营商、AI 大模型、互联网、金融、电力等多行业客户，公司以腾讯为种子客户形成的品牌背书与市场影响力策略已取得成效。

此外，在产品拓展方面，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 产品已完成流片及测试、正推动客户量产导入，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。

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、发行人律师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》及公司说明文件，了解公司 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，分析公司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、查阅公司销售明细表及在手订单，获取公司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、实地访谈公司主要客户，了解公司产品销售背景、应用场景及用户评价；

6、取得公司产品测试报告，评估公司产品综合性能；

7、访谈公司研发及产品负责人，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；

8、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。

二、核查意见

经核查，保荐人、申报会计师、发行人律师认为：

1、从 DPU 应用的主要场景和下游行业需求分布来看，DPU 产品作为高性能服务器的关键组件，云计算和 AI 智算是最主要的两个应用场景，在我国云服务市场和电信运营商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，公司客户集中具有合理性；

2、从商业价值角度和腾讯选择其他 DPU 方案或产品分析，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；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、保障供应链安全的最佳选择，公司不存在单方面依赖腾讯的风险，公司对腾讯收入占比较高不构成重大不利影响；

3、公司在腾讯数据中心应用过程中积累的经验可复用，公司与腾讯交易不影响对其他非关联客户的开拓。公司与腾讯的合作降低了新客户信任成本与导入门槛，并借助链主供应链传导与腾讯多元场景沉淀的全场景方案能力，成功开拓电信运营商、AI 大模型、互联网、金融、电力等多行业客户，公司以腾讯为种子客户形成的品牌背书与市场影响力策略已取得显著成效。同时，随着多代际、多场景产品矩阵不断完善，公司未来将降低与腾讯关联交易收入的占比。

问题 4. 关于尚未盈利与前瞻性信息

申报材料显示：发行人风驰 2.0 系列 AI DPU 已成功流片，预计在 2026 年下半年量产销售。

请发行人披露：

（1）风驰 2.0 系列 AI DPU 产品的进展情况，能否如期在 2026 年下半年实现销售。

（2）结合 2026 年上半年业绩情况及后续期后业绩情况，进一步分析发行人盈利时点预测的准确性，期后是否存在影响预测的重大不利变化。

请保荐人、申报会计师说明核查依据、过程，并发表明确意见。

【回复】

一、风驰 2.0 系列 AI DPU 产品的进展情况，能否如期在 2026 年下半年实现销售

（一）风驰 2.0 系列 AI DPU 产品的研发及流片情况

公司风驰 2.0 系列 AI DPU 于 2025 年完成芯片设计，2026 年上半年回片测试良好，正推动客户量产导入并实现了部分样品销售，2026 年 1-6 月收入金额为 137.97 万元。

该产品重点针对 AI 智算场景进行优化，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，同时也满足私有云、边缘计算、存储等场景对高性能 RDMA 网卡的需求。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、网络带宽等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，同时重点优化针对智算领域的高级网络特性（如逐包负载均衡、CBFC 拥塞控制等），且价格更具市场竞争力，亦符合当前国产化及自主可控趋势，具备进一步推向市场的基础条件。

（二）客户验证导入进展顺利，预计可如期在 2026 年下半年实现批量销售

基于第一款全功能 DPU 云霄 1.0 系列的成功商业化落地，公司已构建良好的应用生态和品牌影响力。通过与头部云服务商、电信运营商等客户的长期合作，公司已在产品适配性、生态构建及解决方案落地等方面积累了经验，为公司风驰 2.0 系列 AI DPU 产品的市场开拓打下了良好的基础，有助于该产品更高效快速地推向市场。

2026 年上半年风驰 2.0 系列 AI DPU 流片成功后，公司积极推进各主要客户的测试验证及导入工作，目前已实现对锐捷网络的销售收入，在腾讯、中国移动、国产知名 AI 大模型厂商等重要客户以及多个头部服务器厂商已启动测试导入工作，其中部分已完成，前述客户预计在 2026 年四季度进入灰度部署、小批量采购，在 2027 年实现放量；此外，公司风驰 2.0 系列 AI DPU 还将在 2026 年四季度送测导入其他头部云服务商和通信行业客户，并推进后续批量采购。

截至 2026 年 8 月末，风驰 2.0 系列 AI DPU 产品在手订单金额 650.41 万元。基于产品竞争力以及客户导入进展，公司风驰 2.0 系列 AI DPU 产品预计可如期在 2026 年下半年实现批量销售。

二、结合 2026 年上半年业绩情况及后续期后业绩情况，进一步分析发行人盈利时点预测的准确性，期后是否存在影响预测的重大不利变化

（一）2026 年上半年业绩情况及后续期后业绩情况

随着云霄 1.0 系列全功能 DPU 产品在主要客户侧持续放量，公司 2026 年 1-6 月实现营业收入 9.40 亿元，同比大幅增长。2026 年 7-8 月，公司实现营业收入 3.69 亿元（未经审计），保持快速增长态势。截至 2026 年 8 月末，公司在手订单金额为 6.26 亿元，在手订单充足。

因此，公司 2026 年上半年业绩及期后业绩均保持快速增长趋势，且在手订单充足，为公司全年业绩增长提供了有力支撑。

（二）发行人盈利时点预测的准确性

公司 2026 年 1-8 月实现营业收入 13.08 亿元（未经审计），截至 2026 年 8 月末的在手订单金额为 6.26 亿元，2026 年全年预测收入具备可实现性。

若公司产品开发、客户导入和验证进度，以及行业市场空间等因素未发生重大不利变化，公司预计最早可在 2028 年实现盈利，2027 年及 2028 年收入将随着 AI DPU 产品的逐步放量及全功能 DPU 新产品的推出而稳步增长。2026 年上半年及期后业绩情况与前瞻性信息预测情况总体相符，具体如下：

1、营业收入

2026 年 1-8 月公司实现营业收入 13.08 亿元（未经审计），截至 2026 年 8 月末，公司在手订单金额为 6.26 亿元，在手订单充足。随着全功能 DPU 产品进一步放量，AI DPU 产品逐步实现批量销售，公司预计 2026 年全年营业收入具备可实现性。

2、销售单价

2026 年 1-6 月公司全功能 DPU 平均销售单价及在手订单单价与 2026 年预测平均单价相匹配，2027-2028 年预测单价逐步回落，具有谨慎性；2026-2028 年预测 AI DPU 平均销售单价低于 2026 年 1-6 月单价，系公司结合样品出货价格、竞品价格等情况及规模效应带动成本下降等因素，谨慎预计 AI DPU 在预测期内平均单价逐年小幅下降，具有合理性。

3、毛利率

2026 年 1-6 月，公司全功能 DPU 产品毛利率为 23.68%，高于其 2026 年全年的预测毛利率，具有谨慎性和可实现性；2027-2028 年，随着存储配件价格预期回落及公司规模效应进一步显现，预计全功能 DPU 产品毛利率将有所回升。2026 年 1-6 月，公司 AI DPU 产品毛利率与 2026-2028 年的预测毛利率基本持平，具有合理性。

4、期间费用

2026 年 1-6 月公司期间费用为 4.65 亿元，期间费用较 2025 年上半年同比增长主要系风驰 2.0 系列 AI DPU 在上半年确认流片相关费用，公司实际发生的期间费用与预测水平总体相符。2027-2028 年，随着公司整体规模扩张和研发持续投入，预计期间费用呈增长态势。2027 年期间费用较高，主要系公司 2027 年计划完成两款芯片的流片，相关光罩费用较大所致。

因此，公司 2026 年上半年及期后业绩保持增长，盈利时点预测的相关指标总体符合预期，具有合理性，盈利时点预测具有准确性。

（三）期后不存在影响预测的重大不利变化

公司 2026 年上半年营业收入保持快速增长，期后业绩持续向好。公司 2026 年 1-6 月实现营业收入 9.40 亿元，同比增长超过 1,500%。2026 年 7-8 月，公司实现营业收入 3.69 亿元（未经审计），保持快速增长态势。截至 2026 年 8 月末，公司在手订单金额为 6.26 亿元，在手订单充足。

公司在市场需求、产品研发、客户开拓及产品测试导入、上游供应链等方面总体保持良好态势，不存在影响预测的重大不利变化，具体如下：

1、市场需求

算力需求的持续高速增长带动大规模 AI 智算集群与云计算基础设施建设的持续投入。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，资本支出规模仍在不断提升。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%。根据上市公司财报及业绩发布会，腾讯 2026 年上半年资本支出达到 847 亿元，同比增长 82%；中国移动 2026 年上半年算力网络资本支出达到 156 亿元，同比增长 71%，均保持高速增长。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，为公司业绩持续增长提供了良好的市场需求环境。

2、产品研发

公司产品研发进展良好。风驰 2.0 系列 AI DPU 已成功流片，正推动客户量产导入并实现了样品销售，2026 年三季度已完成了部分客户的验证测试，预计在 2026 年下半年实现批量销售，在 2027-2028 年进一步放量。该产品具有与国际巨头竞品对标的基本技术实力，同时重点优化针对智算领域的高级网络特性，价格更具竞争力且更符合当前国产化及自主可控趋势。AI DPU 产品将成为公司新的业务增长点，进一步拓宽公司收入来源，增强业绩增长可持续性。

3、客户开拓与产品测试导入

公司持续开拓云服务商、电信运营商以及人工智能等重要领域的头部客户，积极推动产品测试导入，近期主要进展如下：

在云服务商领域，公司云霄 1.0 全功能 DPU 产品在腾讯持续放量增长，计划于 2026 年四季度在另一家知名云服务商推进测试导入。公司风驰 2.0 产品已在腾讯完成产品性能测试，正在其服务器供应商推动机型导入，预计四季度取得小批量订单、完成灰度上线并推动量产；此外，公司已与三家知名云服务商就风驰 2.0 产品进行对接，计划 2026 年四季度送测导入。

在电信运营商领域，中国移动基于公司云霄 1.0 产品发布的大云磐石 DPU 已入选其“2026 年至 2027 年 PC 服务器产品集中采购项目”中标服务器厂商的配置计划，并被列入其智算服务器集采的配置方案，目前已在相关智算服务器中完成适配与部署；公司风驰 2.0 产品已在中国移动智算服务器集采项目中标服务器供应商完成测试及导入，同时也在中国移动研究院完成了适配智算网络的产品验证测试。此外，风驰 2.0 产品在其他电信运营商的智算服务器供应商加快测试导入，预计 2026 年四季度取得小批量订单，后续会进一步加强在智算等方面的合作。

在 AI 大模型领域，国产知名 AI 大模型厂商已完成风驰 2.0 AI DPU 产品的导入测试，预计 2026 年四季度下达小批量订单，2027 年开始规模化采购用于其推理集群和存储集群。

4、上游供应链

上游供应链方面，公司与主要供应商保持稳定合作。针对芯片代工生产，公司通过预付货款锁定产能，并已与主要芯片代工供应商建立稳定排产协作机制，供应商按双方确认的排产计划组织生产，目前不存在因产能紧张导致交付延迟进而影响公司经营的情形；针对全球存储器件价格上涨的情况，公司已结合原材料价格变动与主要客户协商动态调价，目前存储价格上涨未对公司经营造成重大不利影响。公司上游供应链总体保持稳定，期后不存在影响预测的重大不利变化。

综上所述，报告期后，公司营业收入保持快速增长，在手订单充足，在市场需求、产品研发、客户开拓与产品测试导入、上游供应链等方面总体保持良好态

势，不存在影响预测的重大不利变化。

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、查阅发行人说明，了解发行人风驰 2.0 系列 AI DPU 产品的研发及流片相关情况，以及客户开拓及验证导入进展、预期实现批量销售的时间等情况；

2、获取发行人报告期内收入成本明细，期后财务报表和在手订单等，了解公司产品的销售情况、2026 年上半年业绩情况和期后业绩情况；

3、查询头部云服务商、电信运营商等主要客户的定期报告、业绩发布会及公开披露信息，了解其资本支出规模增长情况；

4、获取发行人盈利预测资料，查阅发行人审计报告及财务报表，查阅相关行业研究报告及发行人说明，了解发行人盈利时点预测的准确性，分析期后是否存在影响预测的重大不利变化。

二、核查意见

经核查，保荐人、申报会计师认为：

1、发行人风驰 2.0 系列 AI DPU 已于 2026 年上半年成功回片，产品的进展情况良好，预计可如期在 2026 年下半年实现批量销售；

2、发行人 2026 年上半年业绩及后续期后业绩保持快速增长，发行人盈利时点预测具有准确性，期后不存在影响预测的重大不利变化。

（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第二轮审核问询函之回复报告》之签章页）


深圳云豹智能股份有限公司
2026年9月16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9月16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第二轮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

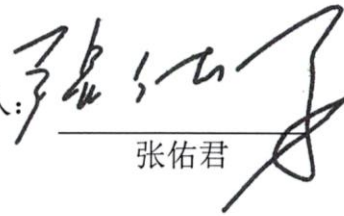
王纯然



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、法定代表人：


张佑君

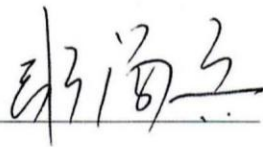


本所及经办律师已阅读《关于深圳云豹智能股份有限公司首次公开发行股票并在创业板上市申请文件的第二轮审核问询函之回复报告》(下称“回复报告”), 确认回复报告与本所出具的法律意见书无矛盾之处。本所及经办律师对发行人在回复报告中引用的法律意见书的内容无异议, 确认回复报告不致因上述内容而出现虚假记载、误导性陈述或重大遗漏。

北京市中伦律师事务所



负责人:


张学兵

经办律师:


郭晓丹


石 璁

2026年 9 月 16 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




李 凤



天健会计师事务所(特殊普通合伙)

二〇二六年九月十六日

