

致也科技

AI 轻量化推理专业交付服务

专注数据清洗、高性能推理与标准化交付，打造极致效率的边缘计算护城河。

轻量化推理：AI 规模化的关键

随着 AI 应用的井喷，大模型在边缘侧的高效落地成为核心痛点。致也科技通过轻量化流程，助力平台实现：

- **响应毫秒化**：实现高并发场景下的极速推理。
- **算力低耗化**：大幅降低设备资源占用，提升部署密度。
- **交付标准化**：满足各大内容平台对模型的高质量要求。



全链路推理交付方案

 **专业数据清洗**：针对 b 站/字节等海量素材进行本地清洗，去除噪声与冗余数据。

 **高算力推理计算**：利用自有集群进行模型本地推理，确保结果精准。

 **标准化打包交付**：统一的规格化上传与接口适配，对接平台标准流程。

主流平台深度交付经验

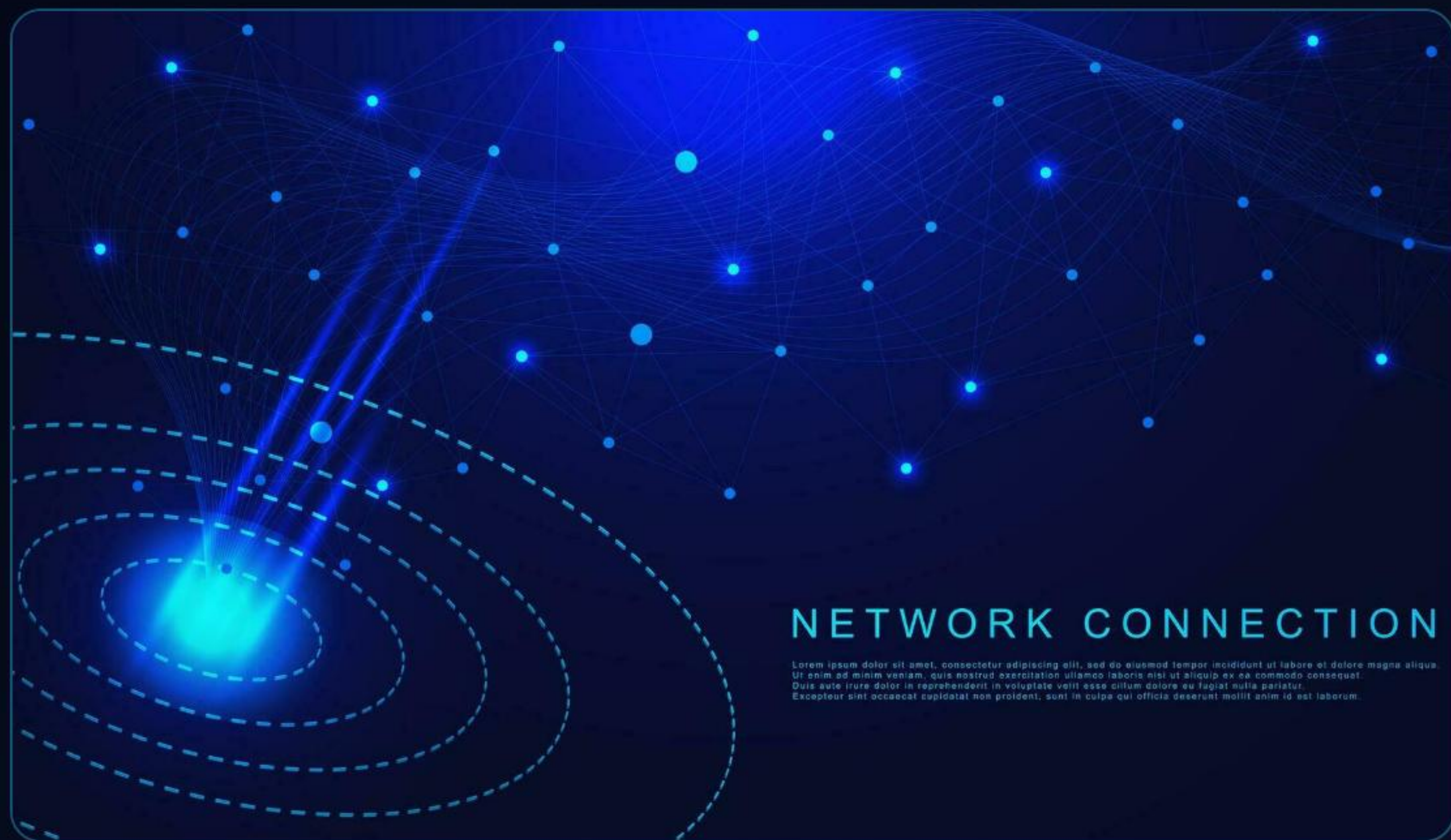
Bilibili 合作生态

承接社区高频互动数据推理任务，优化内容理解与分发策略，提升平台运营效率。

字节跳动生态对接

针对视频流数据的轻量化清洗与推理，确保大规模视频内容的自动识别与处理准确度。

高性能边缘推理架构



极致 Throughput：针对特定模型优化算子，吞吐效率显著高于行业水平。

稳定本地专线：确保数据清洗与传输过程中的极低延迟与高安全性。

本地化数据与隐私保障

我们承诺所有数据清洗与计算环节均在**物理隔离的内部环境**执行。

通过本地化推理方案，彻底消除大模型调用过程中的数据外泄风险，完全符合平台方的合规要求。



专业交付交付流程

阶段	操作内容	输出目标
数据接收	加密采集、分类汇总	原始数据完整
清洗推理	去噪、加权、本地推理	高可用处理结果
适配上传	规格压缩、协议适配	平台级交付标准

降本增效的量化表现

+45%

特定模型推理吞吐性能提升

致也科技通过深度的模型轻量化与本地数据预处理，助力合作伙伴节省约 **40% 的云端计算开销**，同时确保交付质量满足平台严格的审核标准。

致也科技，赋能 AI 轻量化落地

数据精处理，推理高效率，助力平台生态更智联。

Image Sources



https://media.istockphoto.com/id/2273664489/vector/abstract-data-server-center-room-with-ai-cloud-internet-network.jpg?s=612x612&w=0&k=20&c=v4bBB-DGsfimGj282tPvUGB0ycB-gbaw5j_I8BjApTM=

Source: www.istockphoto.com



https://static.vecteezy.com/system/resources/previews/033/348/045/non_2x/digital-technology-high-speed-transfer-data-or-internet-of-things-iot-concept-connecting-dots-and-line-for-network-connection-and-communication-background-illustration-vector.jpg

Source: www.vecteezy.com



Thumbnail

https://png.pngtree.com/png-clipart/20250709/original/pngtree-futuristic-digital-shield-protecting-email-and-data-security-png-image_21294702.png

for

Source: pngtree.com

pngtree.com