# AIBOX-1684X

## 32T AI Box with High Computing Power


V1.0 2024-4-7

T-CHIP INTELLIGENCE TECHNOLOGY

# Product features


## 32TOPS AI processor with ultra-high computing power

SOPHON AI Processor BM1684X, features an octa-core ARM Cortex-A53 with up to 2.3 GHz of frequency and a 12nm lithography process. With up to 32Tops (INT8) computing power, 16TFLOPS (FP16/BF16), or 2Tops (FP32) high-precision computing power.

## Powerful multi-channel video AI performance

The AI box supports up to 32 channels of 1080P H.264/H.265 video decoding and 32 channels of 1080P HD video processing (decoding + AI analysis), making it ideal for various AI applications such as face detection and license plate recognition on video streaming.

## The private deployment of large language models

Support the private deployment of ultra-large-scale parameter models under the Transformer architecture, including large language models such as LLaMa2, ChatGLM, and Qwen, as well as large vision models like ViT, Grounding DINO, and SAM.

## Private deployment of AIGC image generation models

Support the private deployment of the Stable Diffusion V1.5 image generation model in the AIGC field and Docker container management technology.

# Product features


## Multiple deep learning frameworks

Support traditional network architectures such as CNN, RNN, and LSTM;
a variety of deep learning frameworks, including TensorFlow, PyTorch, MXNet, PaddlePaddle, and ONNX, as well as custom operator development.

## All-aluminum alloy enclosure for heat dissipation

The industrial-grade all-metal enclosure with aluminum alloy structure for thermal conduction. The side of the top cover features a grille design for efficient heat dissipation. Its top cover is a porous hexagonal design, combining elegance with high efficiency. The compact, exquisite device operates stably and meets the needs of various industrial-grade applications.

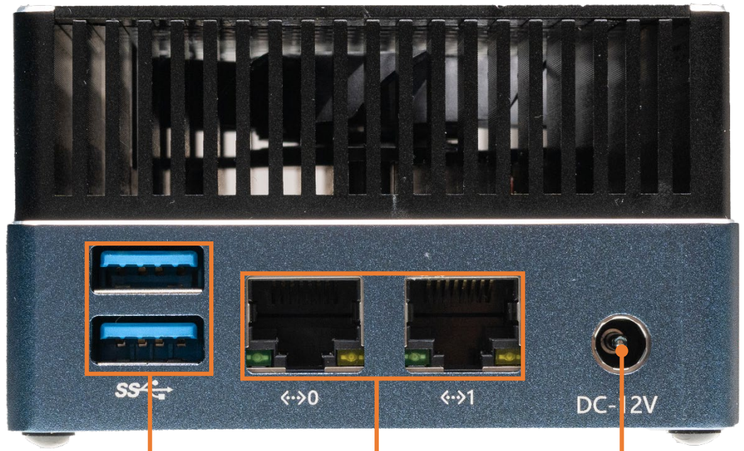
## A wide range of applications

The device is widely used in intelligent surveillance, AI education, services based on computing power, edge computing, private deployment of large models, and data security and privacy protection.

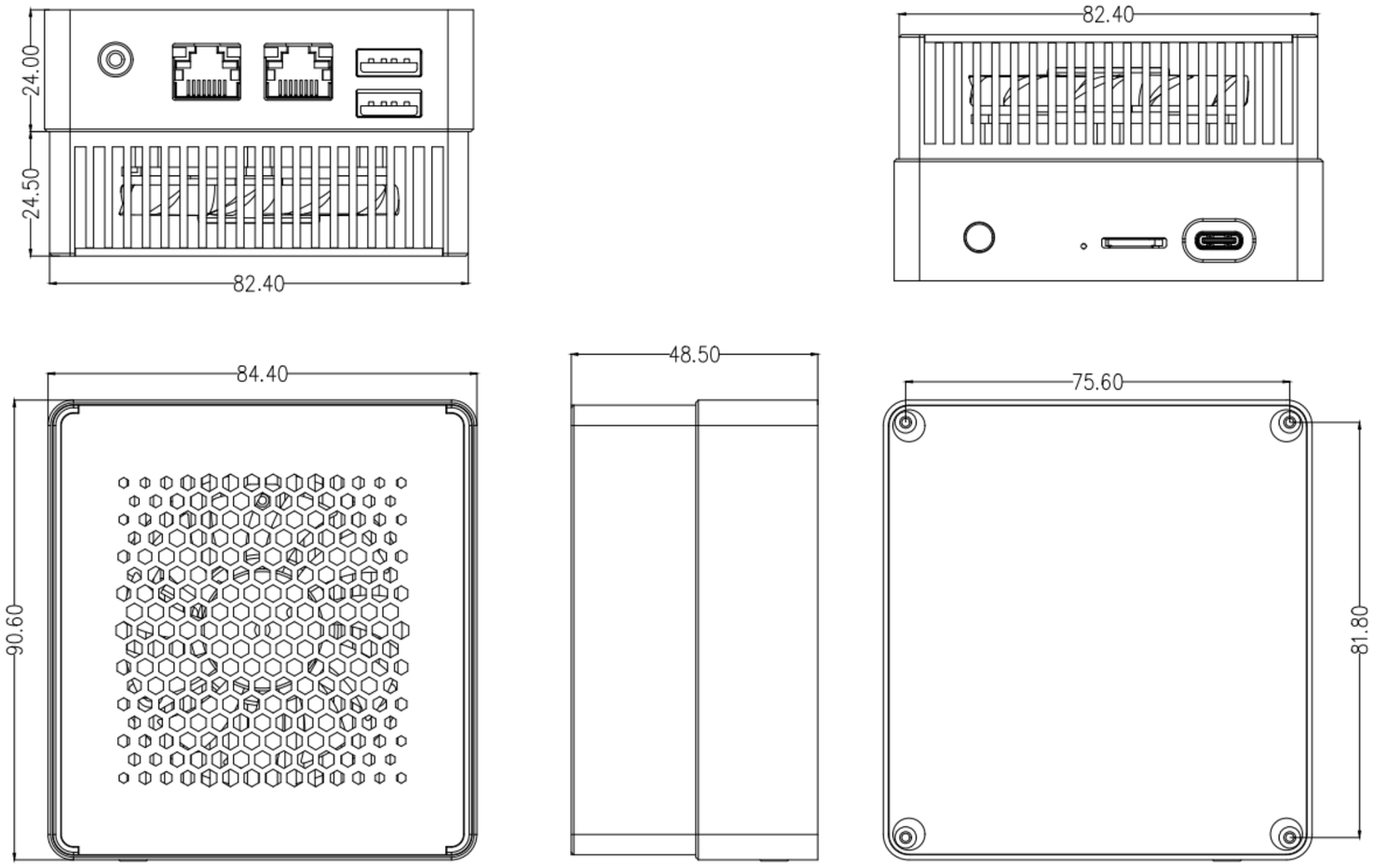
# Specifications

| Specifications | | |
|---|---|---|
| Basic Specifications | SOC | SOPHON BM1684X |
| | CPU | High-performance octa-core ARM A53, 12nm lithography process, frequency up to 2.3 GHz |
| | TPU | Built-in tensor computing module TPU, computing power up to: 32TOPS (INT8), 16TFLOPS (FP16/BF16), 2TFLOPS (FP32) |
| | VPU | 32-channel H.265/H.264 1080p@25fps video decoding, 32-channel 1080P@25fps HD video processing (decoding + AI analysis), 12-channel H.265/H.264 1080p@25fps video encoding |
| | RAM | 8GB/12GB/16GB LPDDR4/LPDDR4X |
| | Storage | 32GB/64GB/128GB eMMC, 1*TF Card |
| | Power | DC 12V/4A (DC 5.5*2.5mm) |
| | OS | Linux |
| | Software Support | • The private deployment of ultra-large-scale parameter models under the Transformer architecture, including large language models such as LLaMa2, ChatGLM, and Qwen, as well as major visual models like ViT, Grounding DINO, and SAM.<br>• The private deployment of the Stable Diffusion V1.5 image generation model in the AIGC field.<br>• Traditional network architectures such as CNN, RNN, and LSTM; a variety of deep learning frameworks, including TensorFlow, PyTorch, MXNet, PaddlePaddle, and ONNX, as well as custom operator development<br>• Docker container management technology |
| | Dimension | 90.6mm * 84.4mm * 48.5mm |
| | Weight | ≈ 420g |
| | Environment | Operating temperature: -20℃ ~ 60℃, Storage temperature: -20℃ ~ 70℃, Storage humidity: 10% ~ 90%RH (non-condensing) |
| Interface Specifications | Ethernet | 2*1000Mbps |
| | USB | 2*USB 3.0 (Current limit: 1A) |
| | Other | 1*power button, 1*Type-C (Debug serial) |

# Interface description

# Dimension



## T-CHIP INTELLIGENCE TECHNOLOGY


Contact Us
(+86)18688117175

E-mail
global@t-firefly.com

Website
https://en.t-firefly.com/

Address
Room 2101, Hongyu Building, #57 Zhongshan 4Rd, East District, Zhongshan, Guangdong, China.