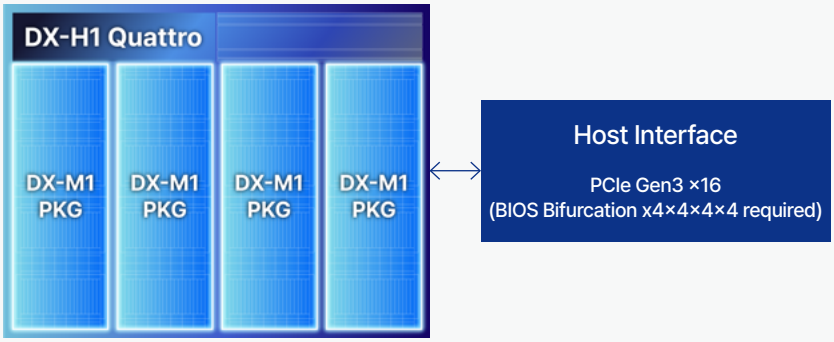
DX-H1 Quattro

Supercharge Your AI Workloads
with 4x Efficiency

Type: AI Accelerator **Performance: 100 TOPS / 20W**

Functional Block Diagram



Specifications

Features	Details
AI Performance	100 TOPS (INT8)
Host Interface	PCIe Gen3 ×16 (BIOS Bifurcation x4×4×4×4 required)
Memory	16GB LPDDR5
Power Consumption	20W
Thermal Solution	Active Cooling (Integrated Fan & Heatsink Module)
Form Factor	Low Profile PCIe Card
Dimensions	166.9 mm × 68.9 mm × 18.30 mm (without Braket)
OS Support	Windows 11/10
	Debian-based Linux (Ubuntu 24.04/22.04/20.04 LTS), Docker
	Yocto Linux
	Android
AI Frameworks	Ultralytics, TensorFlow, PyTorch, ONNX, Keras
System Support	x86, ARM Based Architecture

Overview

Break free from traditional GPU limitations and experience unmatched efficiency.

The **DX-H1 Quattro** delivers a powerful **100 TOPS (INT8)** of performance while maintaining a remarkably low Thermal Design Power (TDP) of just **20W**. This makes it an energy-efficient solution optimized for AI workloads in data center and edge environments.

This card's unique "**Quattro**" design integrates four AI processors to provide powerful parallel processing capabilities. As a standard PCIe card, it's designed to be easily installed in existing servers, making it the most efficient choice for your system upgrade.

Target Applications

- Data Center AI Servers
- Multi-Channel Video Analytics Systems
- Smart Factories and Smart Cities
- Edge AI Servers and Appliances
- Network Video Recorders (NVRs)
- Industrial Robots
- Automation Equipment

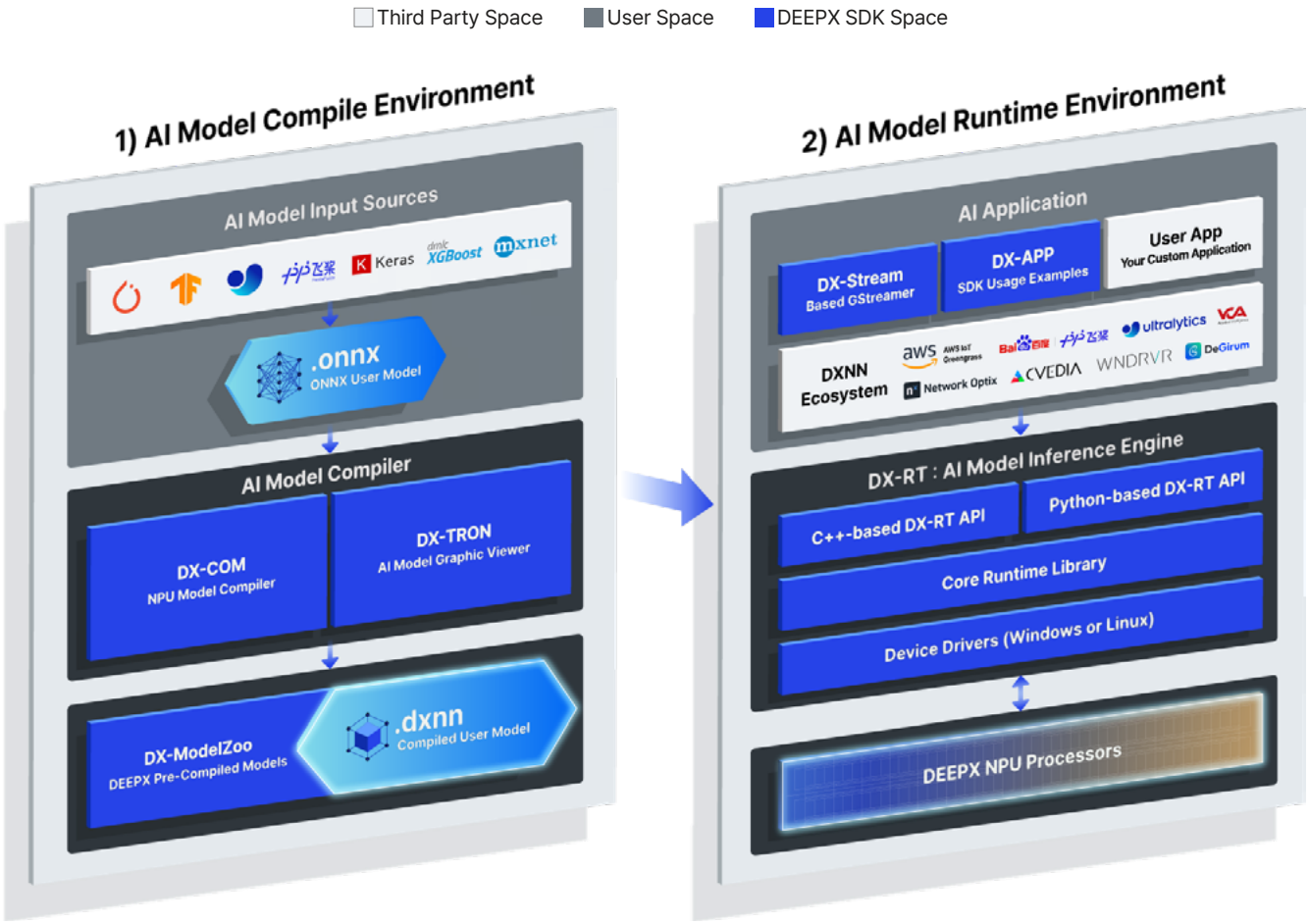
DEEPX

©2026 DEEPX Co., Ltd. All Rights Reserved.

DXNN® - DEEPX SW Development Kit

DXNN® (DEEPX Neural Network) SDK streamlines the AI deployment pipeline on DEEPX NPUs. By integrating essential tools for compilation, optimization, simulation, and inference, it ensures high development efficiency. Experience a ready-to-use development environment with DX-AS (All Suite), a fully integrated and version-aligned package designed for rapid scaling.

DXNN SDK Full Stack Architecture



1) AI Model Compile Environment

Optimizes models from various sources into DEEPX-specific .dxnn formats via the ONNX framework.

2) AI Model Runtime Environment

Deploys optimized models onto hardware via DX-Stream and DX-RT engines using C++ or Python APIs.

DX-AllSuite

DX-AllSuite: The Single Package for Your Complete DXNN Environment. Simplify setup, installation, and updates across local machines and Docker.

Step 1
Supported Environments

- DEEPX Developers — Learn more
- GitHub — Learn more
- aws AWS IoT Greengrass

Step 2
Auto Install & Update

DX-AllSuite

Step 3
Finalize & Start Development

- Docker Install
- Local Install

DEEPX HQ
3F, 20 Pangyoyeok-ro 241beon-gil, Seongnam-si, Gyeonggi-do, South Korea

USA
1735 Technology Drive Suite 740. San Jose, CA U.S.A

China
No 0910, Yangguang Yuehai Building, Nanshan District, Shenzhen, Guangdong, China

Taiwan
Nanjing E. Rd., Zhongshan Dist., Taipei City 104, Taiwan (R.O.C.)

Shop Now!
Contact Sales
sales@deepx.ai

©2026 DEEPX Co., Ltd. All Rights Reserved.