DX-M1 Chip

Intelligence liberated — powerful, efficient, and universal at the edge

Type: AI Accelerator **Performance: 25 TOPS**

Overview

The DX-M1 AI chip is more than silicon; it is the declaration that intelligence belongs everywhere, not just in the cloud. Driven by a breakthrough efficiency design, it delivers the force of 200 GPU-class TOPS, transforming the edge from a passive receiver into an active decision-maker.

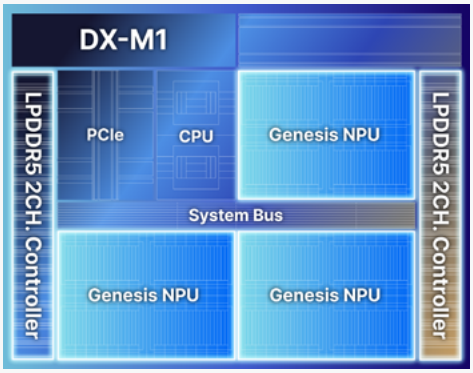
Through its innovative memory design, DX-M1 allows **multiple AI models to work side** by side, not as a luxury but as a new standard. What once demanded costly, massive systems is now within reach of **every device, every factory, every city.**

This is not a step forward in chips. It is a step forward in civilization, where **real-time intelligence becomes universal, affordable, and inevitable.**

Specifications

Features	Details
AI Performance	25 TOPS (INT8)
Host Interface	PCIe Gen3 ×4 (Supports Gen 1/2/3 & x1/x2/x4)
Memory Interface	LPDDR4x or LPDDR5
	Memory Capacity : Up to 8GB
Operating Temperature	0 ~ 70°C (Commercial)
	-40 ~ 85°C (Industrial)
	-40 ~ 85°C (AEC-Q100 Grade3)
	Supports AEC-Q100 Grade upon request
Package	FC-BGA, 17 × 17mm 625-ball, 0.65mm Pitch
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

Functional Block Diagram



Target Applications

- Edge Cameras Systems
- Smart Mobility
- Smart Factory
- Smart Cities
- Robotics
- Drones
- Edge Computing
- Smart Homes
- Smart Retail

DEEPX

Key Benefits

Superior Thermal Management

Maintains high performance while staying safely within industrial temperature range (-40°C to 85°C), making it uniquely suitable for real-world applications where competing solutions fail due to overheating.

Exceptional Performance Efficiency

Delivers around 20x power performance efficiency compared to GPGPUs and outperforms competitors by more than 2x, ensuring on-device systems can maintain high-level AI capabilities without compromise.

GPU-Level AI Accuracy

Achieves GPU-level AI accuracy using quantized Int-8 precision instead of power-hungry FP32, ensuring reliable on-device intelligence that meets industry standards of less than 1% accuracy drop.

Industry-Leading TCO

Even if competitors offered their chips for free, the DX-M1 would still be more economical. Lower power consumption translates to dramatically reduced operational costs, making advanced AI economically feasible at scale.

Performance

Real-world AI inference FPS(Frames Per Second) benchmarks comparing DEEPX with leading competitors

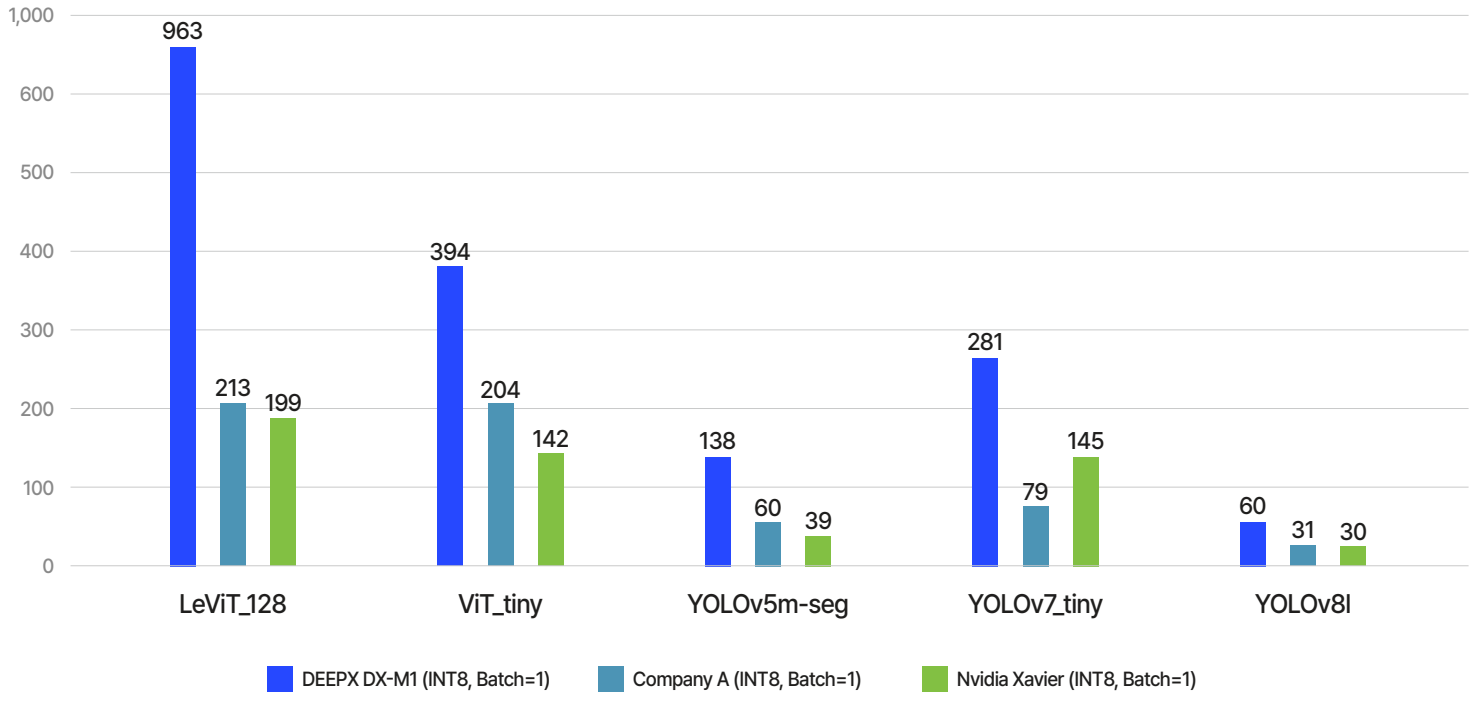
*This graph shows the benchmark results for Batch=1, all tested under the same conditions.

*Xavier measured on Jetson AGX kit (room temp) using INT8 and 'no batch' (BS=1), per NVIDIA recommendations..

*Company A results used the optimize option of "random-calib-set" with default values for the parser and compiler.

*DEEPX DX-M1 numbers are based on the SDK (September 2025), DX-Compiler v.2.0.0, and DX-RT v.3.0.0.

*Company A and DX-M1 were measured at room temperature on a single device through a PCIe interface on an x86 Host PC (Intel® Core™ i7-14700K CPU @ 5.6GHz).



ultralytics

DEEPX x Ultralytics YOLO26: Powering Physical AI Together

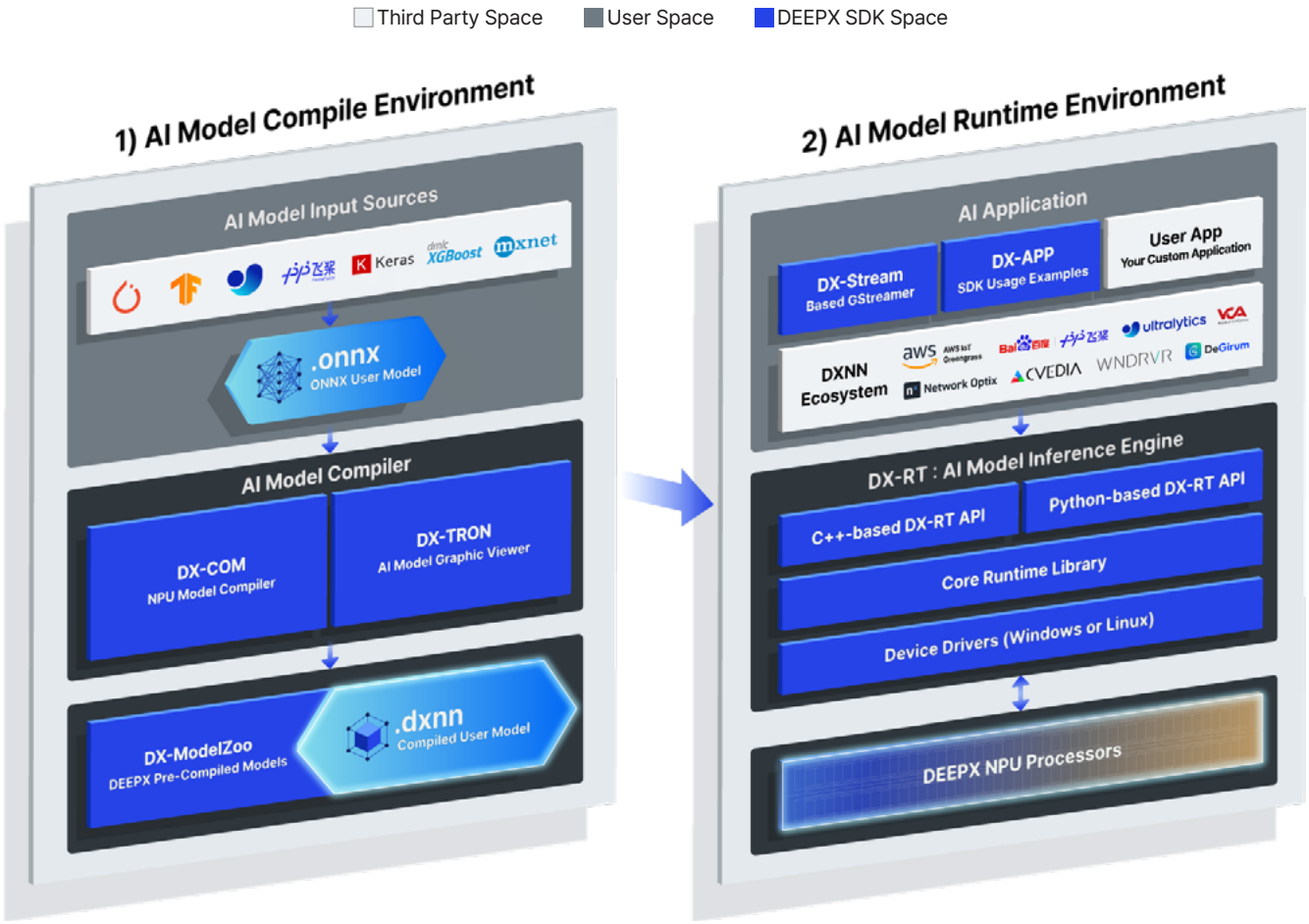
"As a key partner of the Open-Source Physical AI Alliance, DEEPX provides first-class support for YOLO26, enabling real-time intelligence for robots and smart cities."

Learn more

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