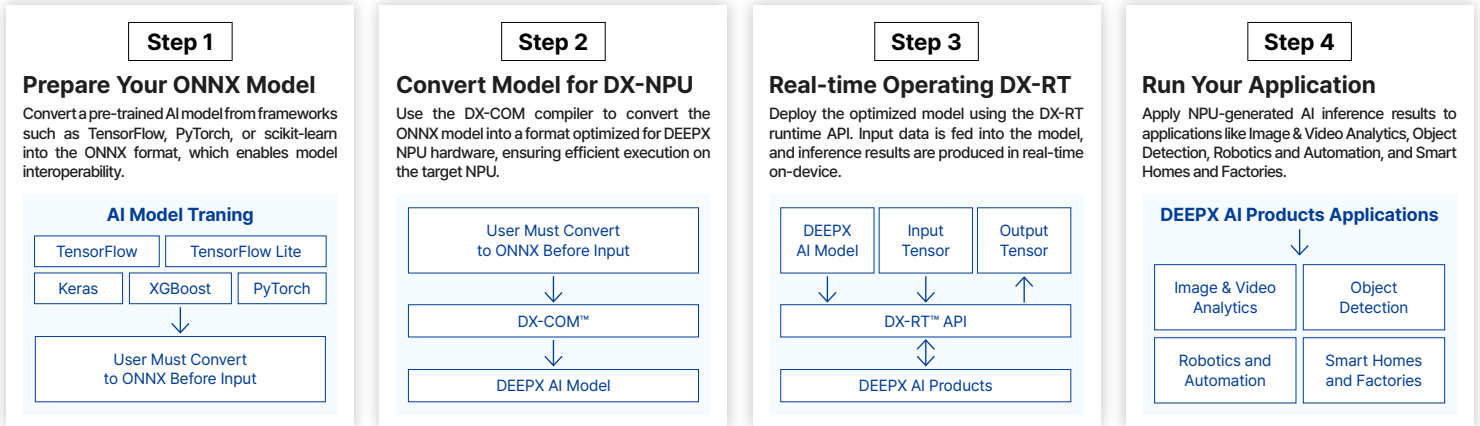# DXNN® SDK

DXNN® (DEEPX Neural Network) SDK streamlines AI deployment on DEEPX NPUs by integrating version-aligned tools for compilation, optimization, simulation, and inference. For **efficient development**, it's offered as the **DX-AS (All Suite), a fully integrated and optimized package.**
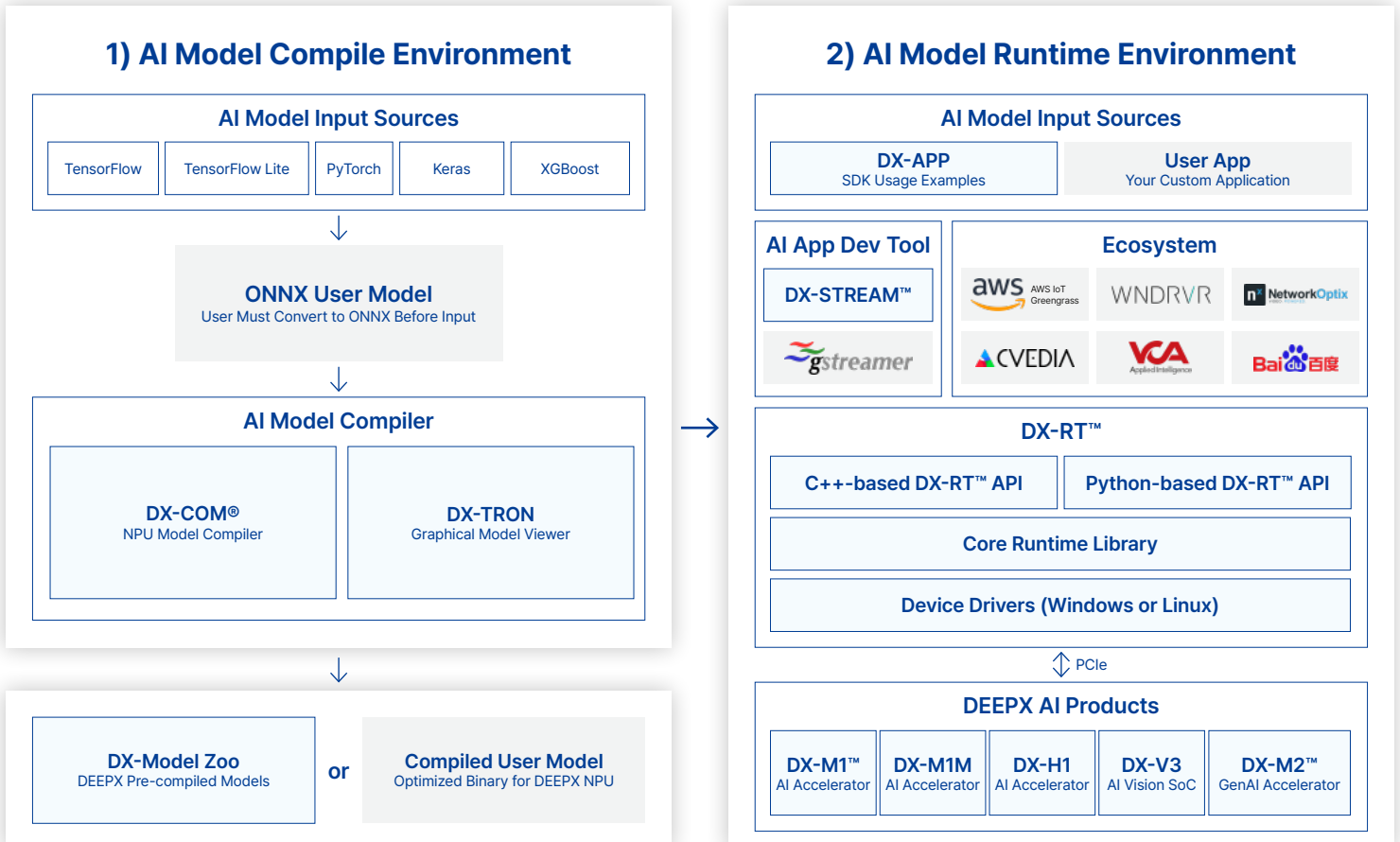
How It Works: 4-Step AI Deployment with DXNN® SDK

Step 1
Prepare Your ONNX Model

Convert a pre-trained AI model from frameworks such as TensorFlow, PyTorch, or scikit-learn into the ONNX format, which enables model interoperability.

AI Model Traning

| TensorFlow | TensorFlow Lite |
| Keras | XGBoost | PyTorch |

↓

User Must Convert to ONNX Before Input

Step 2
Convert Model for DX-NPU

Use the DX-COM compiler to convert the ONNX model into a format optimized for DEEPX NPU hardware, ensuring efficient execution on the target NPU.

User Must Convert to ONNX Before Input

↓

DX-COM™

↓

DEEPX AI Model

Step 3
Real-time Operating DX-RT

Deploy the optimized model using the DX-RT runtime API. Input data is fed into the model, and inference results are produced in real-time on-device.

| DEEPX AI Model | Input Tensor | Output Tensor |

DX-RT™ API

DEEPX AI Products

Step 4
Run Your Application

Apply NPU-generated AI inference results to applications like Image & Video Analytics, Object Detection, Robotics and Automation, and Smart Homes and Factories.

DEEPX AI Products Applications

↓

| Image & Video Analytics | Object Detection |
| Robotics and Automation | Smart Homes and Factories |

DXNN® Full Stack Architecture

DXNN® Full Stack Architecture streamlines AI model deployment onto DEEPX products using its two-stage AI Model Compile and Runtime Environments.

1) AI Model Compile Environment

AI Model Input Sources

| TensorFlow | TensorFlow Lite | PyTorch | Keras | XGBoost |

↓

ONNX User Model
User Must Convert to ONNX Before Input

↓

AI Model Compiler

| **DX-COM®** NPU Model Compiler | **DX-TRON** Graphical Model Viewer |

↓

| **DX-Model Zoo** DEEPX Pre-compiled Models | **or** | **Compiled User Model** Optimized Binary for DEEPX NPU |

→

2) AI Model Runtime Environment

AI Model Input Sources

| **DX-APP** SDK Usage Examples | **User App** Your Custom Application |

AI App Dev Tool

DX-STREAM™

gstreamer

Ecosystem

aws AWS IoT Greengrass | WNDRVR | NetworkOptix

CVEDIA | VCA Applied Intelligence | Bai百度

DX-RT™

| **C++-based DX-RT™ API** | **Python-based DX-RT™ API** |

Core Runtime Library

Device Drivers (Windows or Linux)

↕ PCIe

DEEPX AI Products

| **DX-M1™** AI Accelerator | **DX-M1M** AI Accelerator | **DX-H1** AI Accelerator | **DX-V3** AI Vision SoC | **DX-M2™** GenAI Accelerator |

DEEPX HQ
5F, 20 Pangyoyeok-ro 241beon-gil, Seongnam-si, Gyeonggi-do, South Korea

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

Taiwan
No. 66-10, Yucheng St., Nangang Dist., Taipei City 115012, Taiwan (R.O.C.)

Europe
Tel-Aviv Israel

Buy Now! - DX TechBridge Program

Contact Sales - sales@deepx.ai