

INTERA Eagle

Product Brief

Near-Memory AI Acceleration Platform
Delivering Maximum Performance per Watt

INTERA provides a family of ultra-low power AI accelerators, from low-latency RISC-V custom instructions to near-memory computing engines, enabling TinyML systems to achieve maximum performance-per-watt with minimal silicon and memory footprint.

Key features

-  Ultra-low-power matrix-vector accelerator
-  Always-on availability
-  Smallest practical memory-mapped AI engine
-  Excellent cost/performance ratio

Key features

-  Sovereignty
 -  Data privacy
 -  Low latency
 -  Ideal for battery-powered applications
- Optimized memory mapped core



INTERA Cognit IP Selection Guide

AI Accelerator	Architecture	Operation accelerated	Compute size (*PE) and scalability
Falcon	Monolithic, streaming processor	MxV	Single PE, 2 to 32-bit implementation
Hawk Advance	RISC-V custom instruction	MxV / MxM ⁽²⁾	Single PE, 2 to 32-bit implementation Entry-level version, scales to Hawk Pro
Hawk Pro	RISC-V custom instruction	MxV / MxM ⁽²⁾	1, 2, or 4 PEs Configurable 2 to 32-bit
Eagle	Near-memory computing	MxV⁽²⁾	4, 16 PEs; tested up to 32 banks/PEs Configurable 2 to 32-bits
Harrier	Monolithic, streaming processor	MxM	Scalable matrix array; 8x8 = 64 PEs; configurable as 32x32, 32x64, 64x64, ...

*PE = processing element

⁽²⁾ Indirect MxM multiplication by software arrangement

Typical target applications



Health monitoring wearables



Wake-word detection



Acoustic event detection (AED)



Neuromorphic vision



Occupancy sensing



Predictive maintenance



Fall detection



Radar perception



Smart security sensors



BMS (state of charge)

Software Suite for Classical AI

(CNN, LSTM, MLP/FCNN)

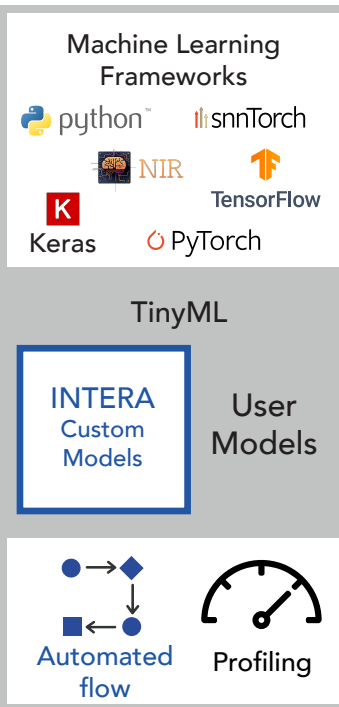
INTERA provides an open-source tool suite which seamlessly interoperates with existing deep learning development frameworks, enabling fast and efficient deployment of AI models on INTERA accelerators.



Development Workflow

-  Seamless integration with existing deep learning frameworks
-  Pre-trained benchmark models
-  Maximizing AI compute performance by efficient use of NN core resources

Model Build Environment



Runtime Environment



-  Ultra-low-power applications
-  Near sensor AI computing
-  SNN applications (spiking neural networks)
-  INTERA SW component
-  Third party SW component

Neural Network Models & Application Support

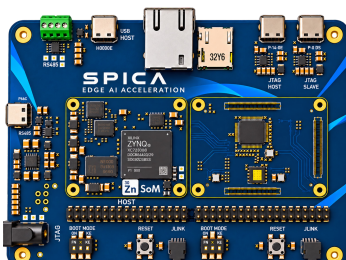
AI Frameworks

TensorFlow, TensorFlow Lite, Keras
PyTorch & ONNX, snnTorch, NIR

Comprehensive dataflow compiler & open frameworks enable customers to easily port their models (CNN, LSTM, MLP/FCNN, SNN) or use INTERA custom models

Starter development kit

Hardware^(*)



Documentation

User guide
Datasheet
Examples

Software

Plug & Play software SDK
Classical AI (CNN, LSTM, MLP/FCNN)
SNN (spiking neural networks)