

ARA-1 EDGE AI PROCESSOR

BREAKTHROUGH AI PERFORMANCE



The Kinara Ara-1 Edge AI processor delivers an optimal combination of performance, power, and value for AI inferring applications that demand low latency, flexibility, and scalability. The Ara-1 processor fits into a variety of implementations, ranging from cameras to multi-chip edge servers.

EASILY DEPLOY AI AT THE EDGE

Optimized AI at the Edge

The Kinara Ara-1 AI processor, with its patented Polymorphic Dataflow Architecture, enables applications to run multiple AI models with zero overhead context-switching while minimizing data traffic and significantly reduce power consumption. The processor delivers a significant advantage in performance/\$ and performance/Watt over general-purpose GPU's, making it ideal for applications in like Smart Cities, Smart Retail, Automotive, and Industry 4.0.

Unrivaled Software Tools

The Kinara Software Development Kit (SDK) combined with Ara-1 provides unmatched ease of use, flexibility, and insights to deploy AI at the edge. Deep Vision provides Linux and Windows drivers to support runtime communication between most Linux- or Windows-based host systems and Ara-1.

Increase Performance by Offloading Inference

While the host system performs all pre- and post-processing functions, the Ara-1 Edge AI Processor handles the application's end-to-end inferencing requirements. It is optimized for a batch size of '1' and delivers both real-time responsiveness and low latency. For simultaneous multi-model support, the Ara-1 Edge AI Processor provides the ability to run multiple models simultaneously without sacrificing performance. Ara-1 processor scales from endpoints to edge servers - connect multiple Ara-1 processors to a host for a linear increase the AI performance.



**Integrate with PCIe
Plug and Play**



**Exceptional
Performance/Watt**



**Common Frameworks
and the Kinara SDK**



**No Cloud Compute
Needed**

SPECIFICATIONS		
AI Model Frameworks Supported	TensorFlow, PyTorch, MxNet, ONNX, Caffe	
Performance (batch =1)¹	Resnet50-v1: 100 inferences/sec. MobileNet-v1: 554 inferences/sec	
Latency¹	Resnet50-v1: 10 msec. MobileNet-v1: 1.8 msec	
Operating System Support (Compiler)	CentOS, Ubuntu	
Operating System Support (Runtime)	Linux, Android, Windows	
Interface	PCIe Gen3 x4, USB 3.2 Gen1 x1	
Power Consumption (Typical)	1.7W @ 600MHz	
Temperature	0°C to +70°C (Commercial)	-40°C to +85°C (Industrial)
Part Numbers	DVI-2020BA-CA0T-B (600MHz core frequency) DVI-2020BA-CA0T-A (800MHz core frequency)	DVI-2020BA-IA0T-B (600MHz core frequency) DVI-2020BA-IA0T-A (800MHz core frequency)
Packaging	15 mm x 15mm EHS-FCBGA	

¹Maximum performance based on peak computational throughput of Ara-1 (800MHz) and Host System. Specification subject to change without notice. Performance may vary depending on system configuration.

Kinara AI and the Ara-1 are trademarks or registered trademarks of Deep Vision, Inc. in the United States and other countries
© 2022 Deep Vision, Inc.

KINARA | LEADING EDGE AI

Kinara delivers unrivaled edge AI solutions to accelerate and optimize real-time decision making. Our AI accelerators power smart edge devices and gateways that demand responsive AI computing at high energy efficiency. The Kinara AI team, based in Silicon Valley as well as Hyderabad, India, includes Silicon Valley innovators, technology experts from Stanford University, and a world-class hardware and software development group. The company derives its name from the Hindi word for 'edge' and reflects the commitment we make to our customers to build extremely innovative edge devices for retail, smart cities, industry 4.0, and automotive.

Kinara and the Ara-1 are trademarks or registered trademarks of Kinara, Inc. in the US and other countries
© 2022 Kinara, Inc.



KINARA | LEADING EDGE AI

www.kinara.ai