Developing Optimized Systems with Akida

Rob Telson- VP World Wide Sales

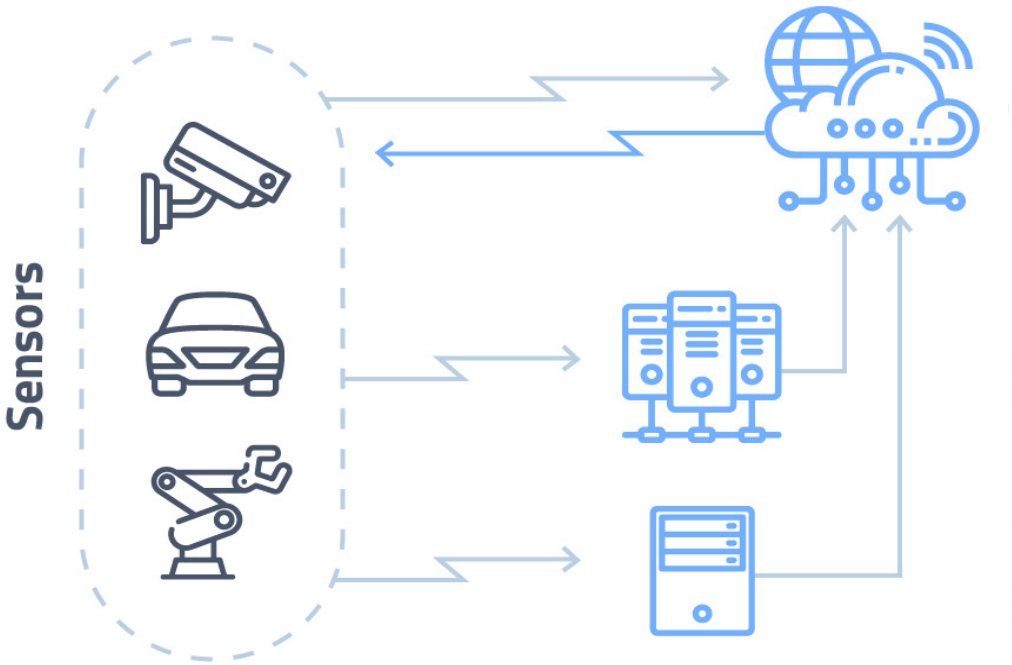
Todd Vierra – Director Customer Engagement

February 24th, 2022

About BrainChip – Founded 2013

- 15+ yrs fundamental AI architecture research & technologies
- 65+ data science, hardware & software engineers
- 21 Patents (pending included)
- Publicly traded:
  - ASX: BRN.AX
  - OTCX: BCHPY
- We make Sensors Smart- Simplicity for Edge AI Devices
  - Audio
  - Gustatory
  - Olfactory
  - Tactile
  - Vision

What is Edge AI Computing?



Conventional AI captures data from sensors and sends it to the cloud for processing.

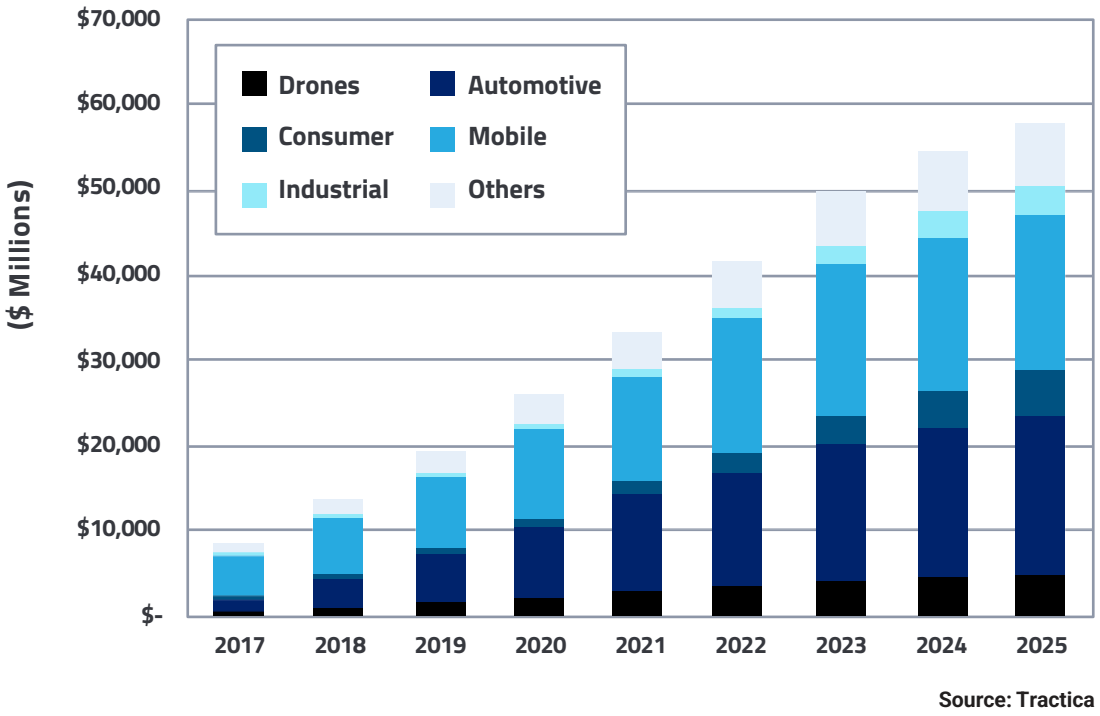
- Large Data Sets
- High Bandwidth
- Computationally Inefficient
- Power Intensive

Edge AI with BrainChip Akida™ brings intelligence to the sensor in the device.

- Computes on-chip, in-device
- Order of magnitude less data & computation
- Fractional power
- Enables one-shot learning
- Device Personalization

Edge Device Market Outlook

Edge Based Devices requiring AI - $60B by 2025



Source: Tractica

Implications

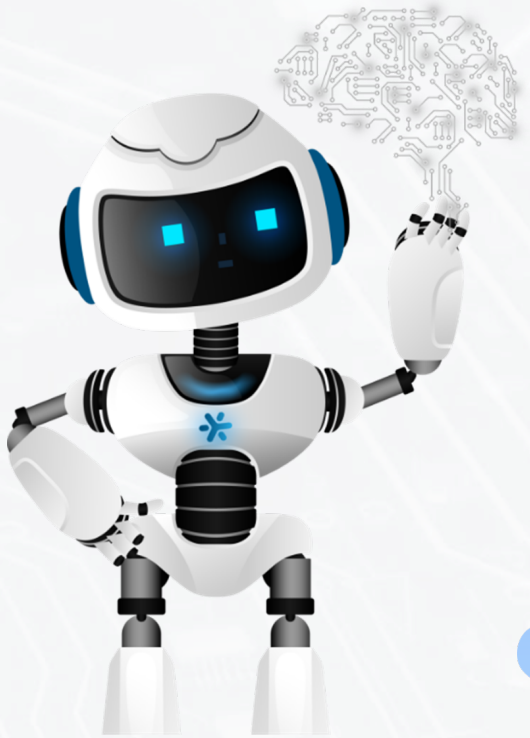
- More capability
- More data
- More compute
- More bandwidth

Call to action

- Current AI compute architecture is unsustainable
- Need to move AI to the Edge…

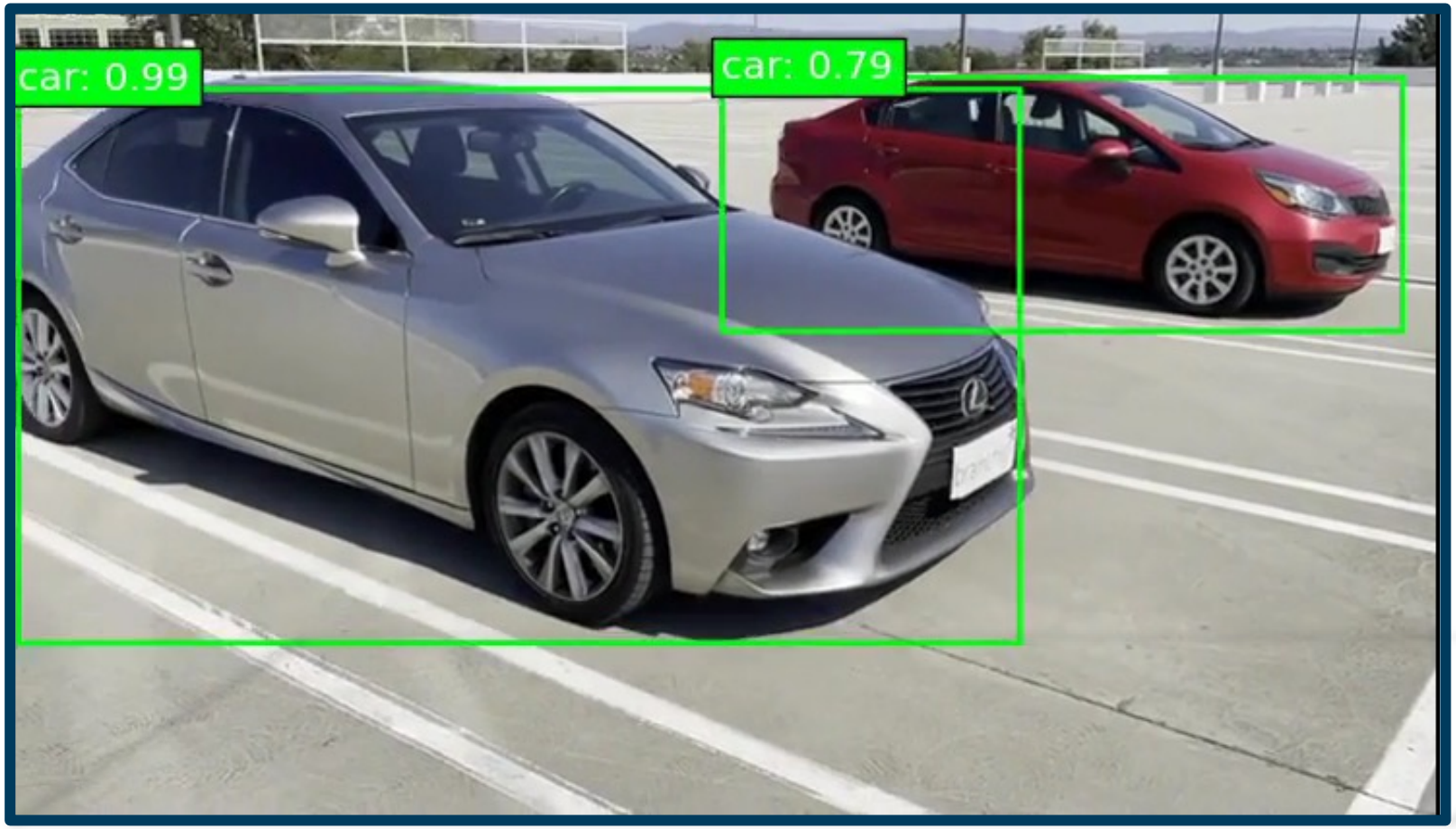
brainchip

Akida™
Object Detection
Person Detection



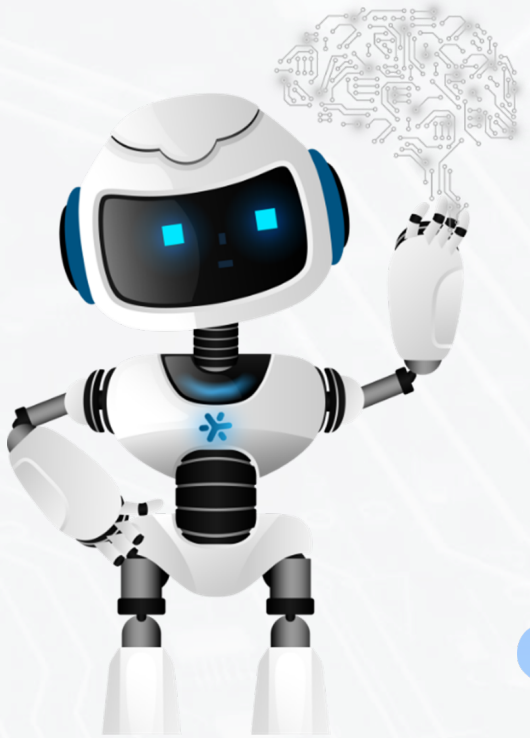
Revolutionizing AI at the Edge

Mobilenet SSD – Cars and People



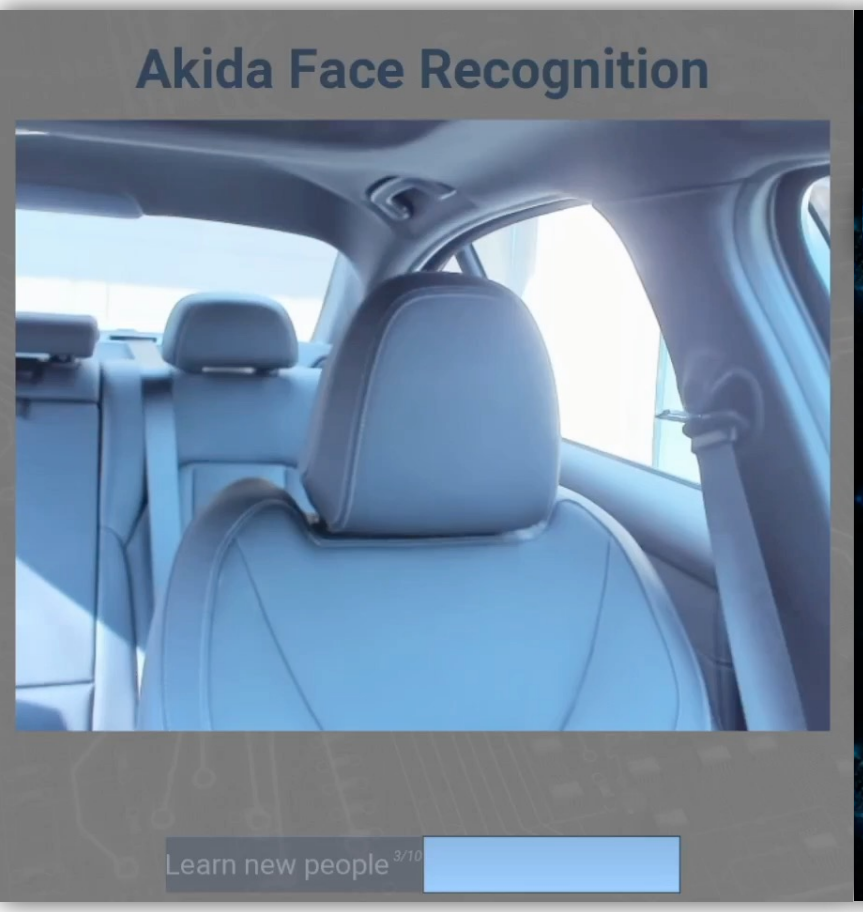
brainchip

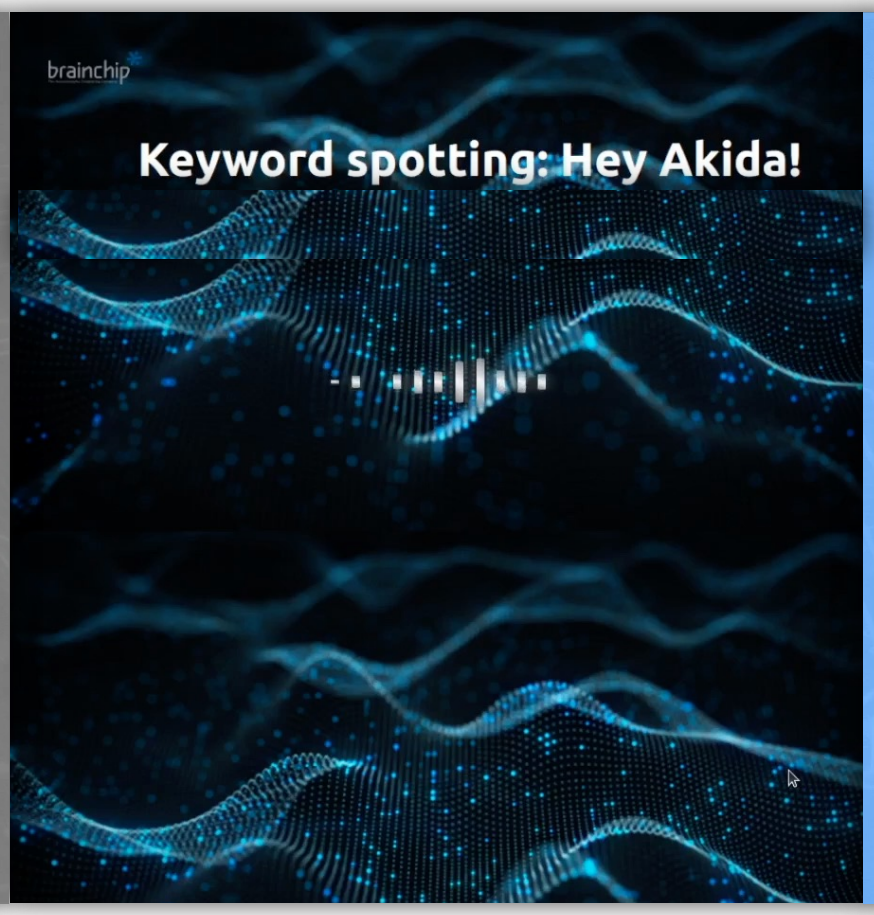
Akida™
Facial Recognition
Keyword Spotting
Visual wake

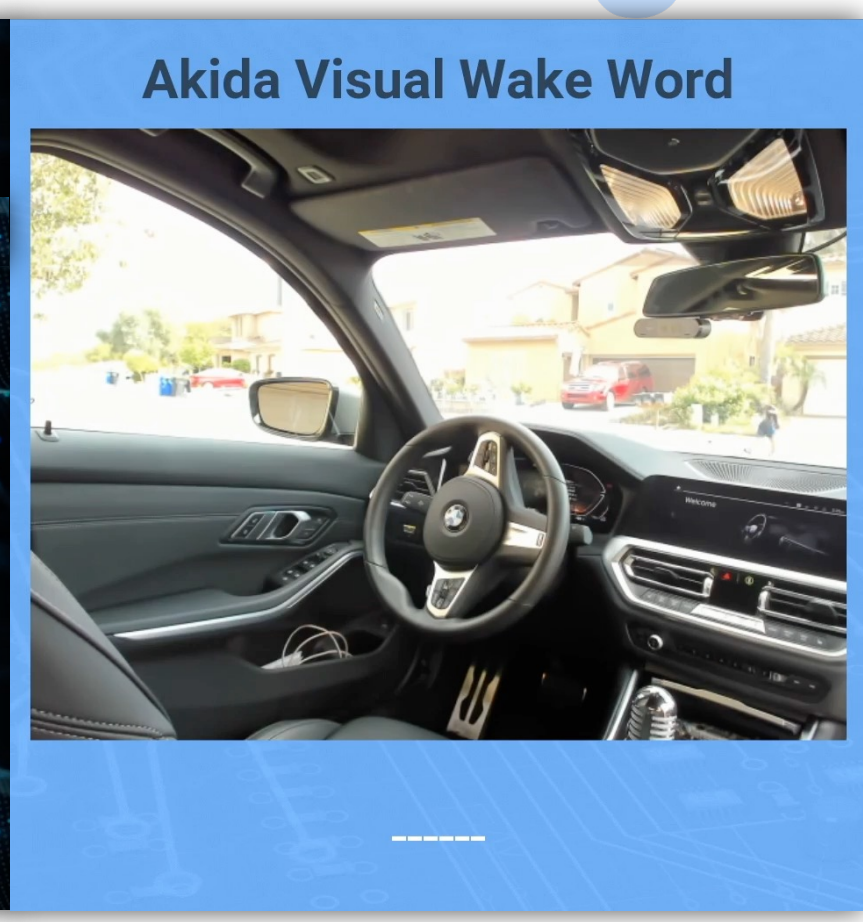


Revolutionizing AI at the Edge

Smart Automotive - In Cabin Experience







Facial Recognition: 22.6mW

Keyword Spotting: 600*u*W

Visual Wake: 6mW

BrainChip's AKIDA Neuromorphic Design Principles

Distributed Computation

- Computation spread across many cores (neural processing units – NPUs)
- Each NPU has its own dedicated computational engine and memory, which reduces data movement

Event-Based Processing

- Non-zero activation map values are represented as multi-bit (1 to 4-bit) events
- NPUs only perform computation on events, not activation maps

Event-Based Communication

- NPUs communicate by sending events to each other over a mesh network without host CPU intervention
- Neural network connectivity is configurable in the field

Event-Based Learning

- AKIDA implements an on-chip, learning algorithm
- No costly communication with cloud required

Akida (AKD1000) Neuromorphic Processor

Data Input Interfaces

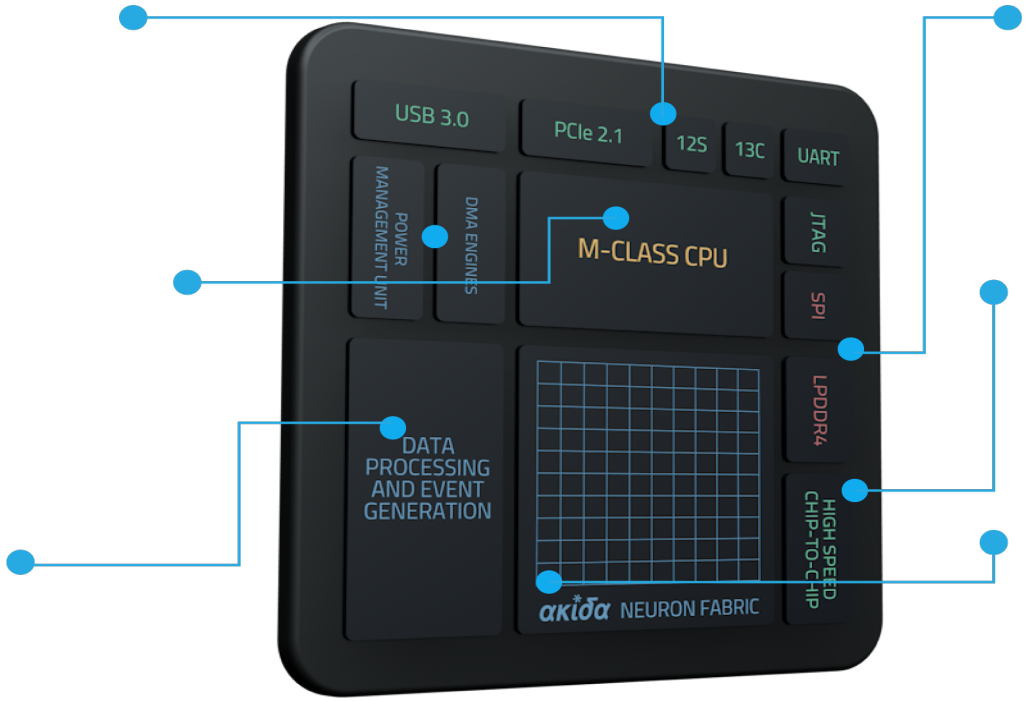
- PCI Express 2.1 x2 Lane Endpoint
- USB 3.0 Endpoint
- I3S, I2C, UART, JTAG

On-Chip Processor

- M-Class CPU with FPU & DSP
- System Management
- Akida Configuration

Data Processing

- Pixel-Event Converter
- SW Data-Event Encoder
- Any multivariable digital data
- Sound, pressure, temp., others



External Memory Interfaces

- SPI FLASH for boot/storage
- LPDDR4 Program/Weights

Multi-Chip Expansion

- PCIe 2.1 2 lane root complex
- Connects up to 64 devices

Flexible Akida Neuron Fabric

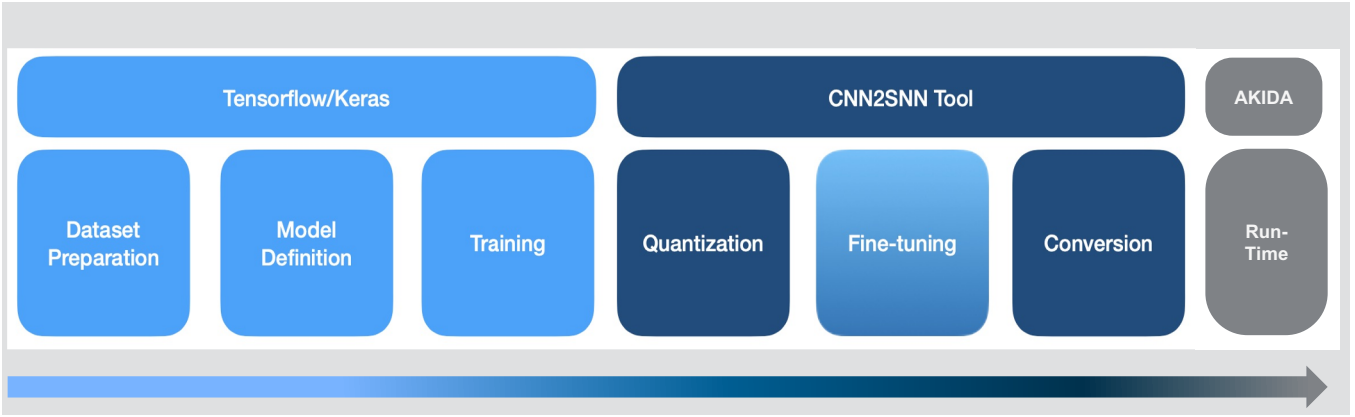
- Implements 80 NPUs
- All Digital logic with SRAM (8MB)
- Also Available as Licensed IP Core
- First Implementation: TSMC 28nm

: Framework Structure and Workflow Saves Time

- Deep-learning professionals do not need to learn a new framework
- Start using **MetaTF** immediately
- Craft models in TensorFlow Keras
- Convert them for deployment on the AKIDA™ neural processor

All this in a few short steps:

1. **CNN design** (TensorFlow)
2. **CNN training** (TensorFlow)
3. **CNN quantization** (cnn2snn/quantize)
4. Optional **CNN tuning** (TensorFlow)
5. **SNN conversion** (cnn2snn/convert)
6. SNN deployment (akida/predict)



- The resulting **SNN model** runs on the AKIDA™ neural processor MetaTF Runtime
- Intuitive and automated network conversion flow = Deep Learning (DL- SNN)

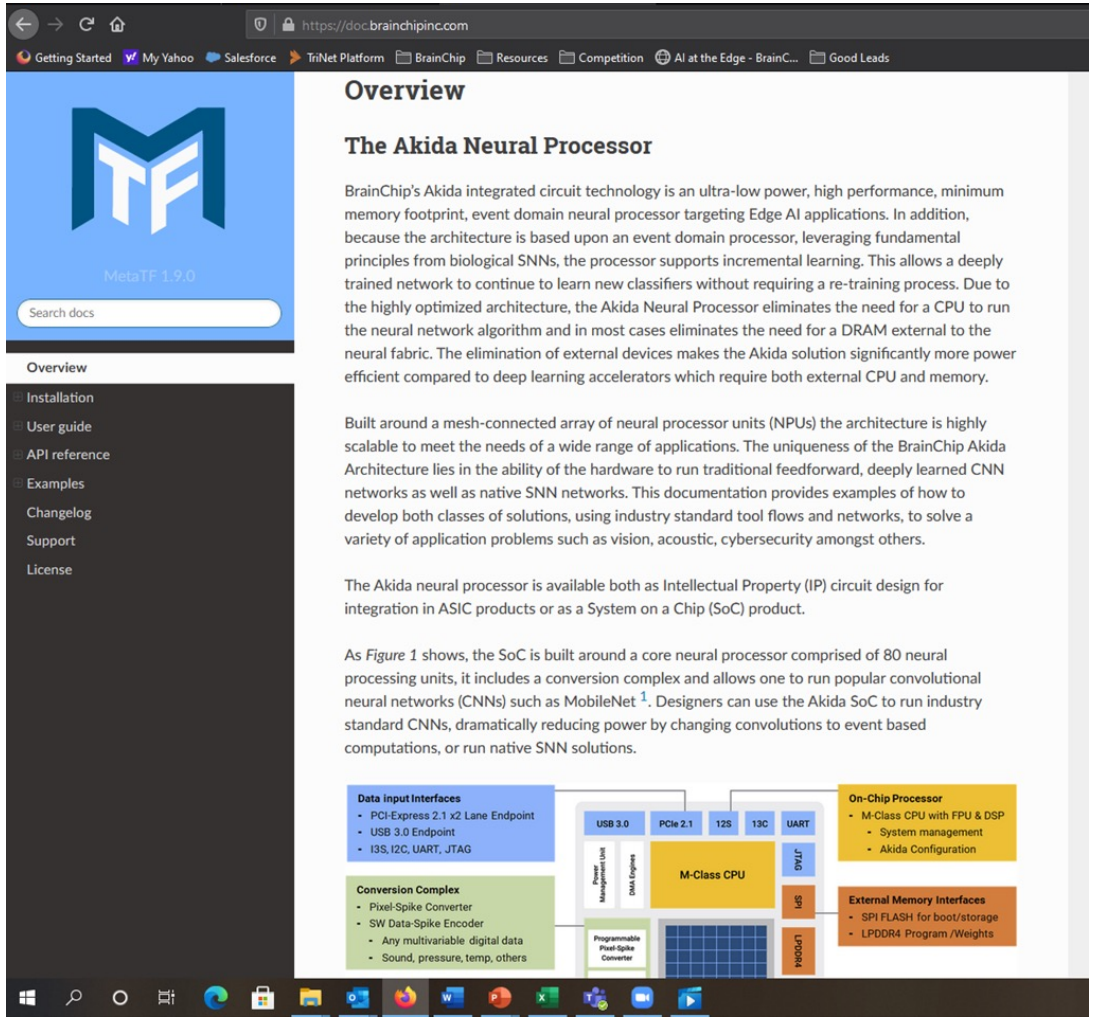
The MetaTF Development Framework: It's a simple environment leveraging standard CNN models design in TensorFlow.

www.brainchip.com/developer

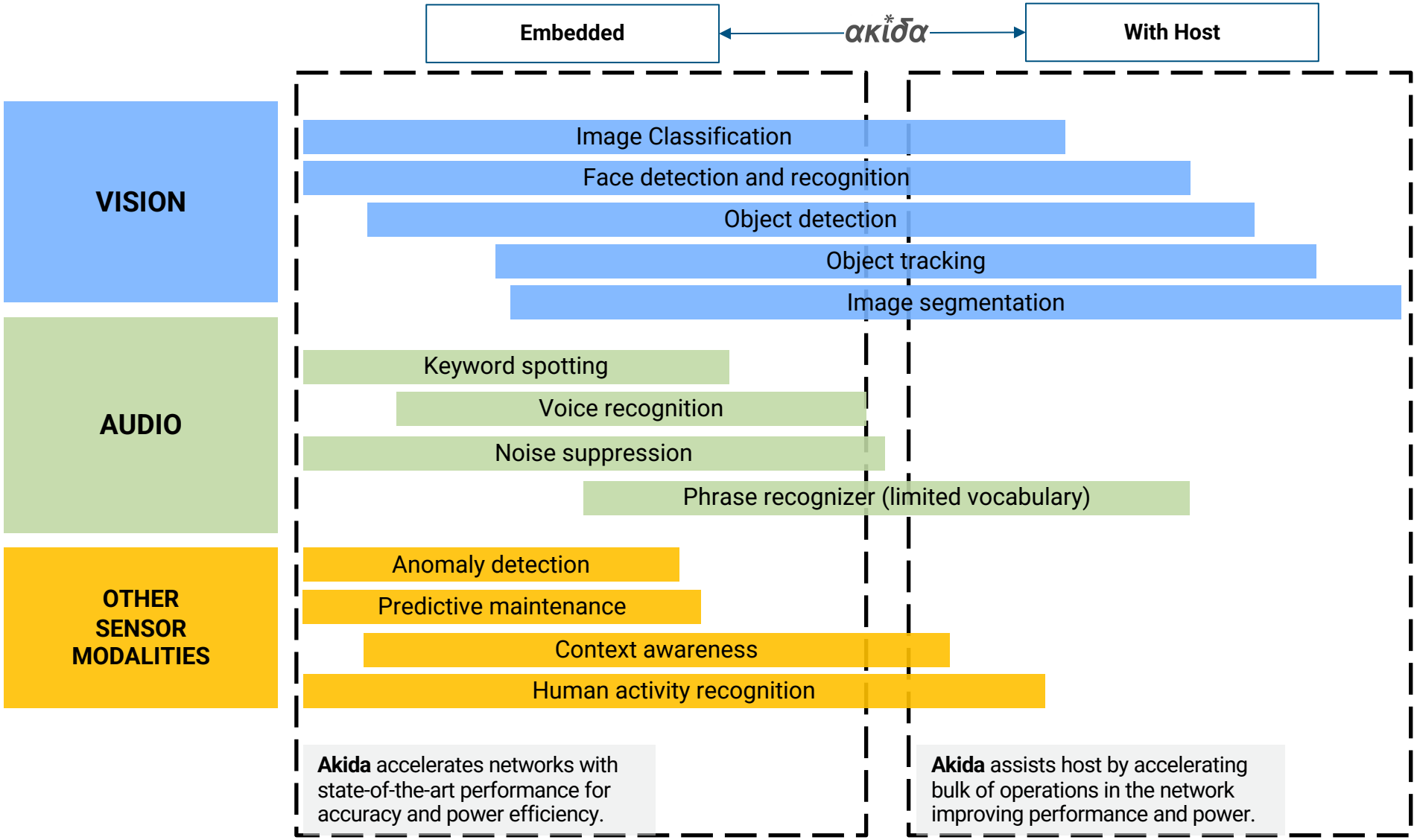
MetaTF Development Environment

- Standard Tensor Flow environment
  - Installation manual
  - User Guide
  - API reference
  - Example/Workflows
- Network and Dataset samples
- Convert normal CNN to SNN
- ~5000 users in 2021

www.brainchip.com/developer

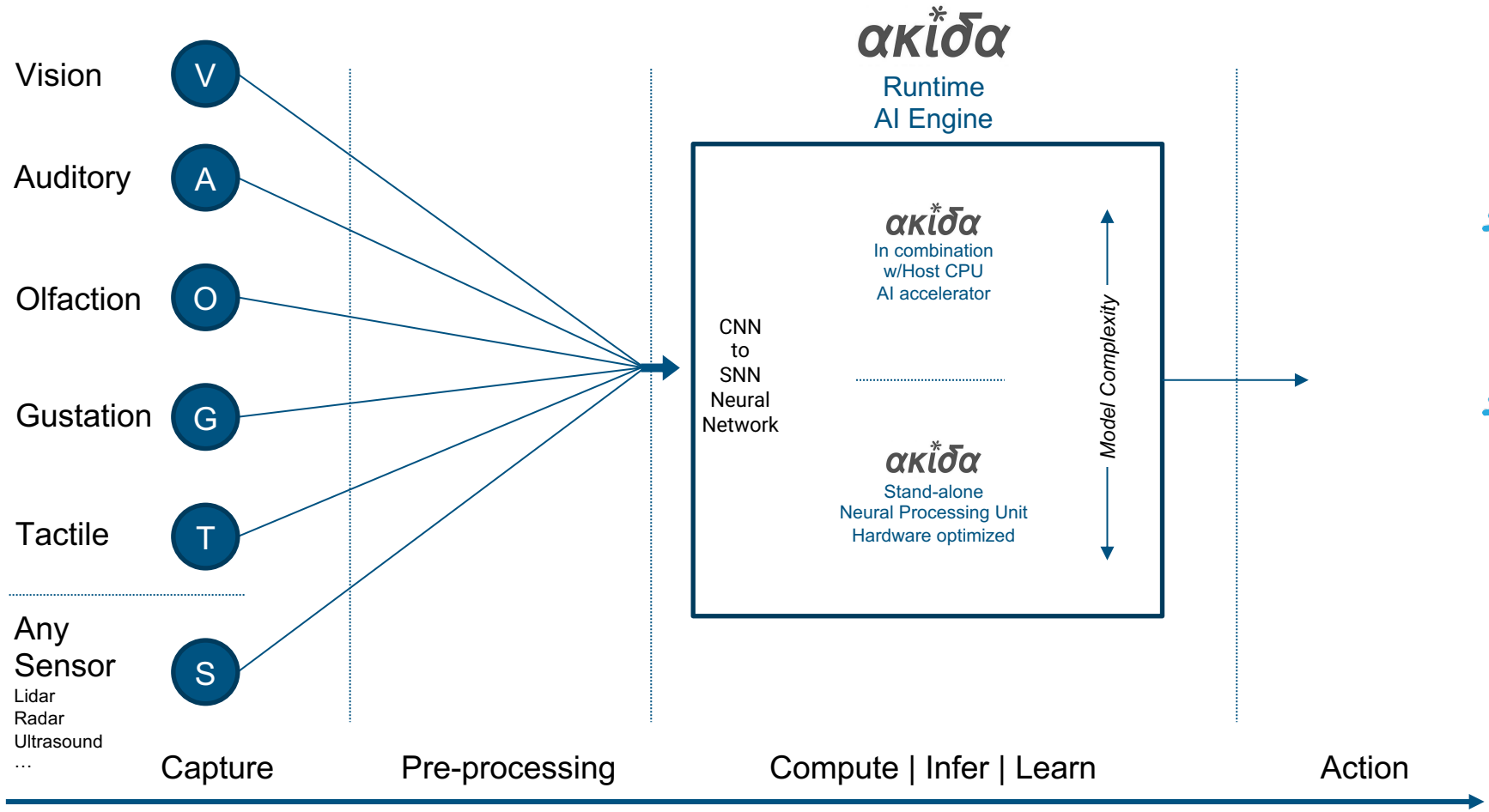


AI Processing in the Edge…



From Sensor to Inference

Performant and Efficient



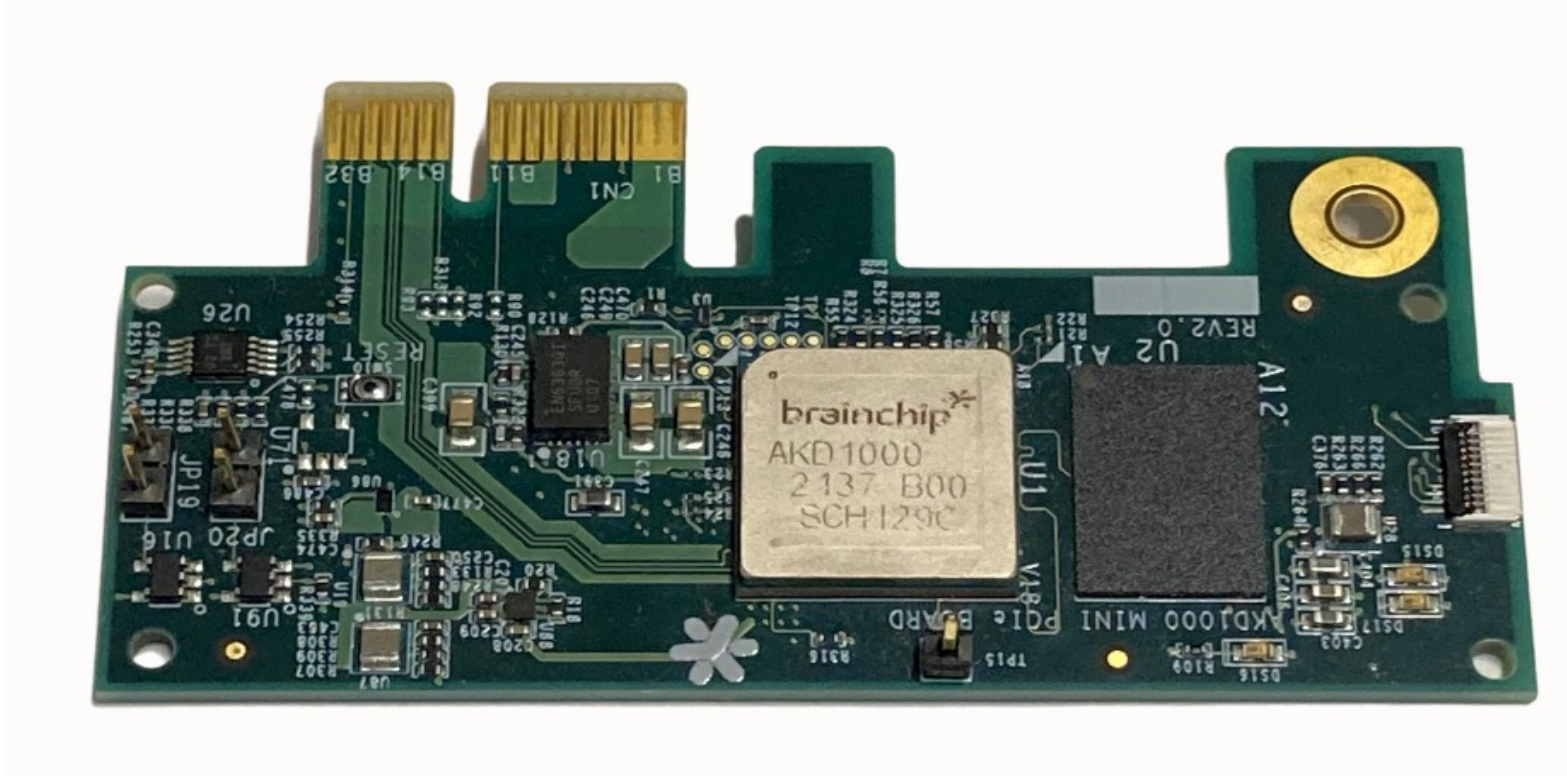
- Akida supports output from any sensor, with its data and network, and applies in-chip neuromorphic AI to efficiently process and infer with radically reduced power consumption.
- Akida can also be deployed as an AI Accelerator to work in combination with a host CPU, still providing power consumption efficiencies.

Optimized compute, radically reduced power…

Akida *AKD1000*

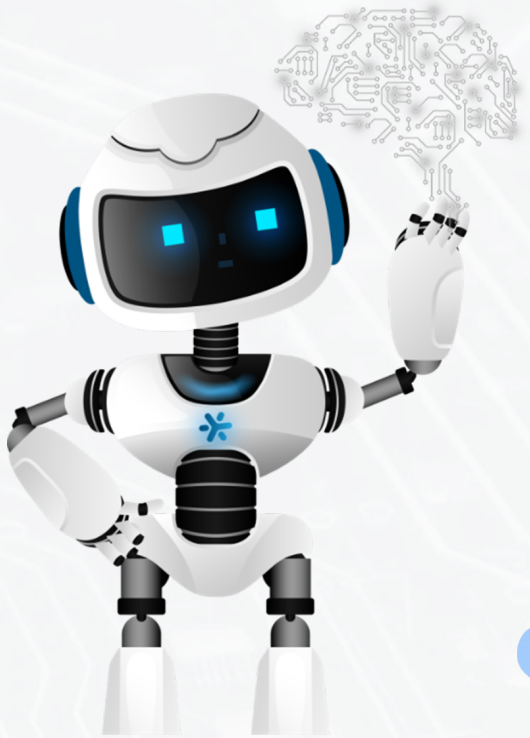
Full Commercialization…

- AKD1000 Chip
- Shuttle PC Development Kit
- Raspberry Pi Development Kit
- AKD PCIe board
- Akida IP



[Order Online](#)
shop.brainchipinc.com

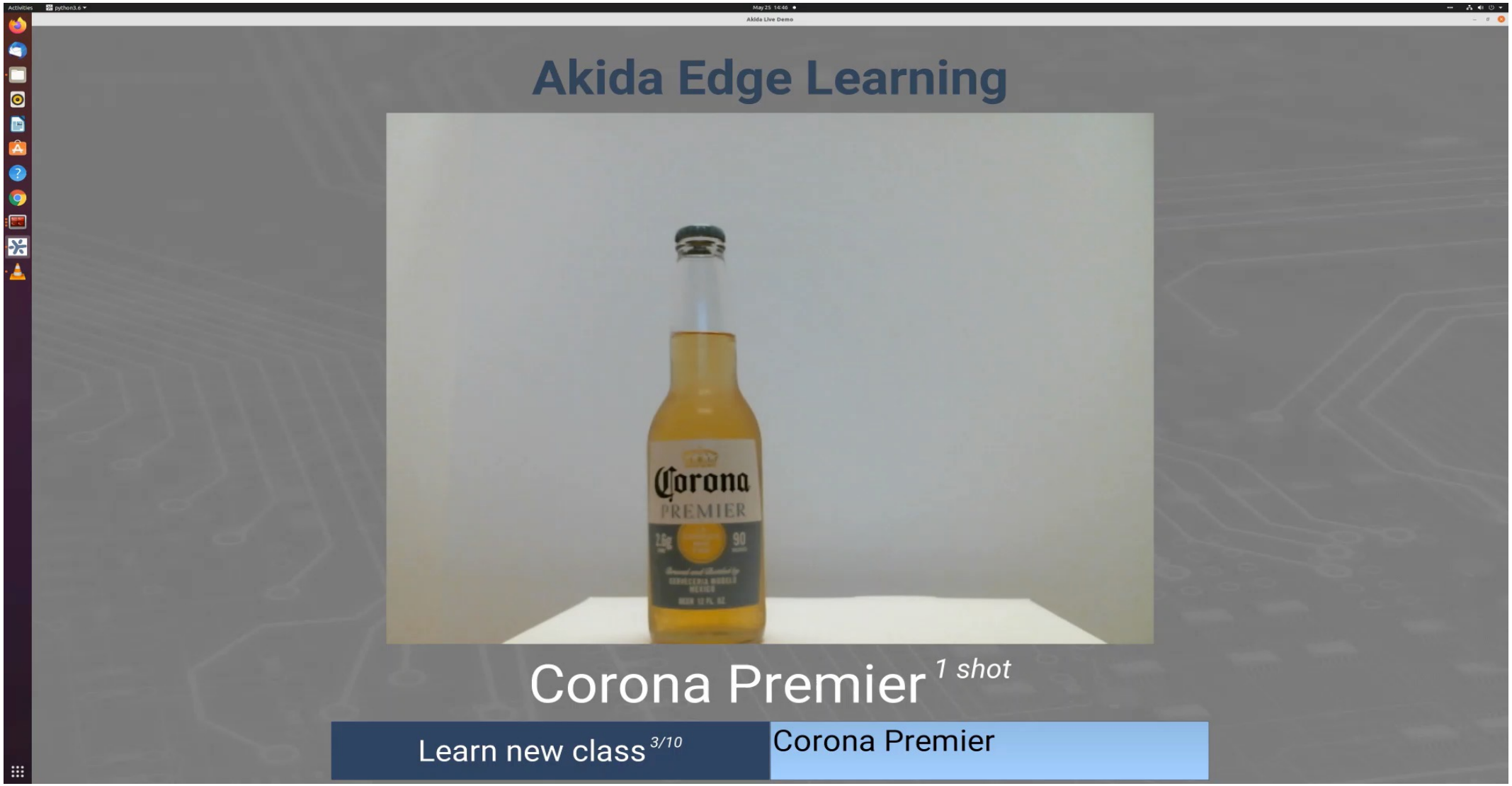
brainchip



Akida™
Edge Based Learning

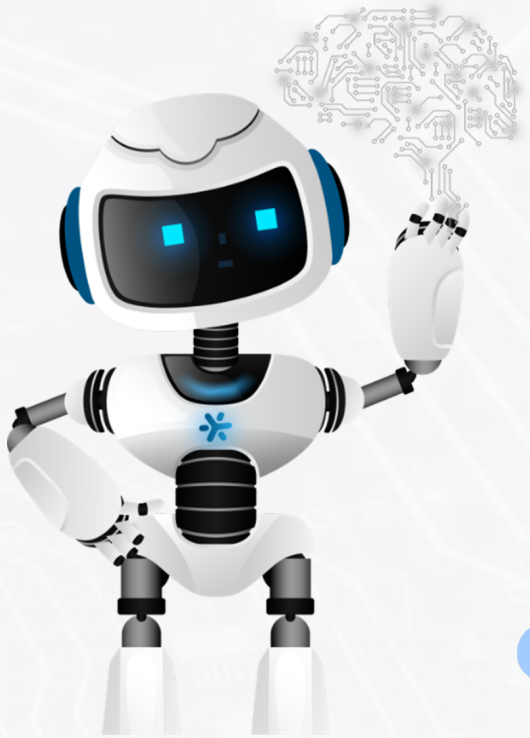
Revolutionizing AI at the Edge

Object Classification: Time for a Beer?



ImageNet Mobilenet V1 Edge learning at High Speed

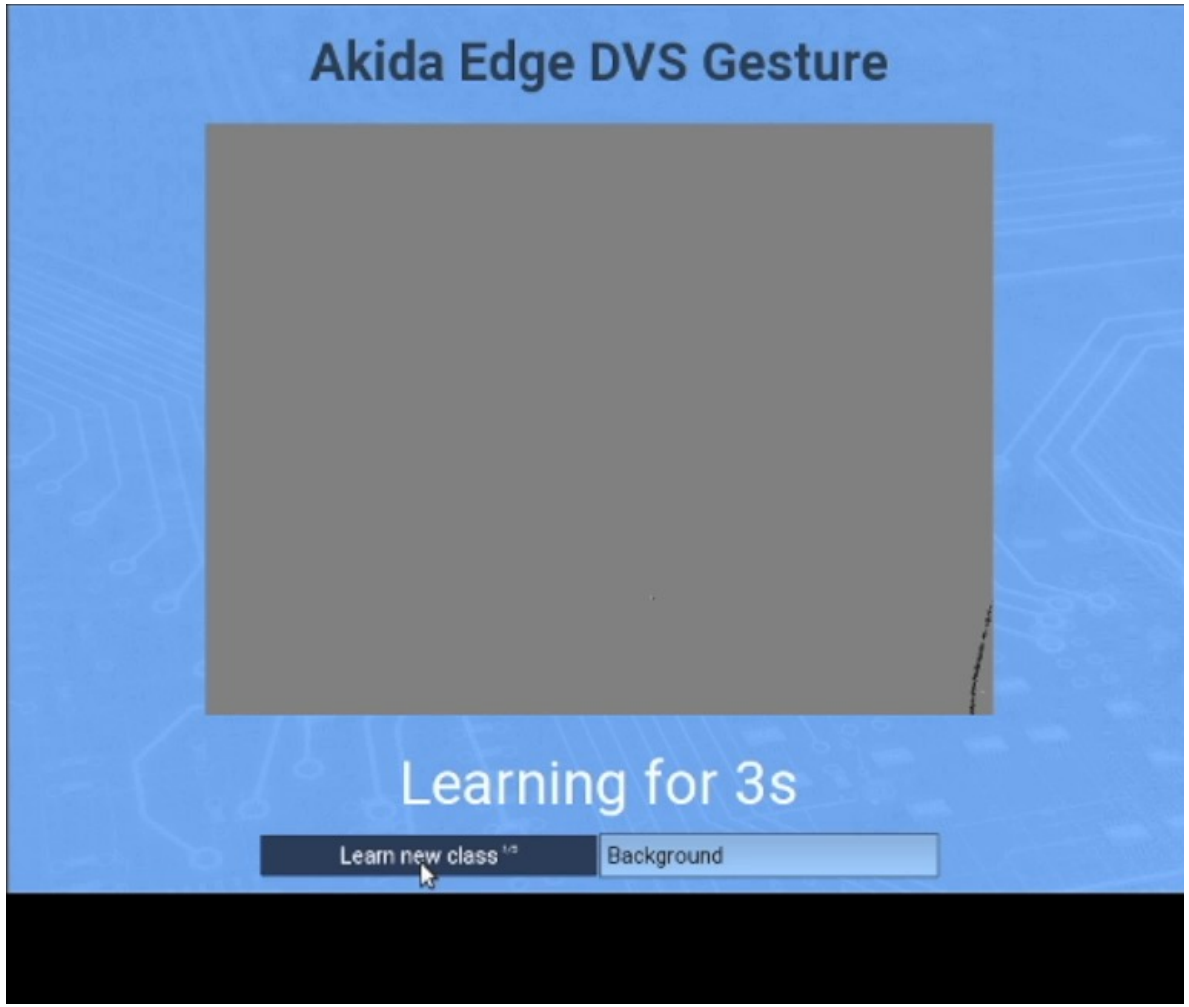
brainchip

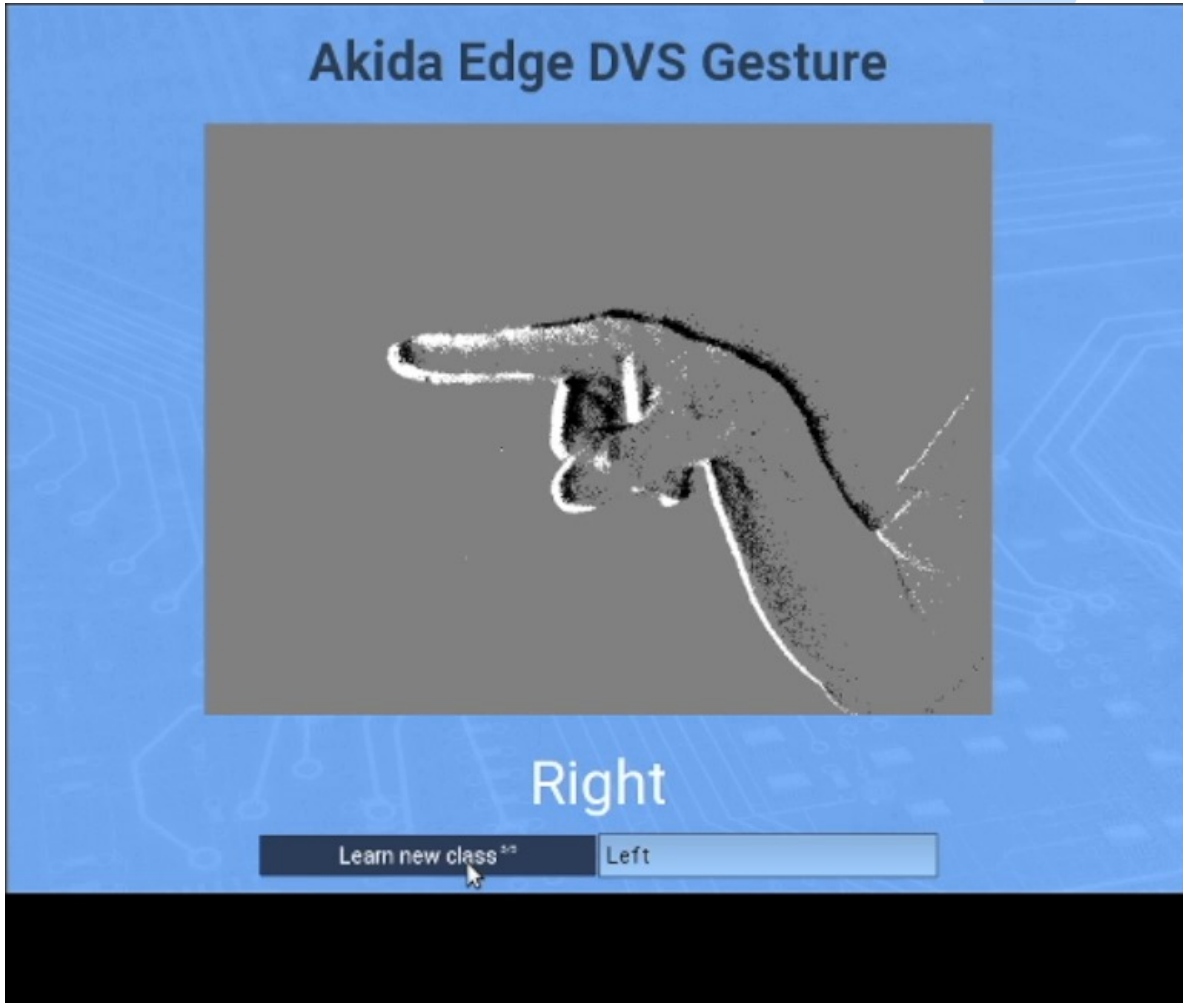


Akida™ Gesture Control

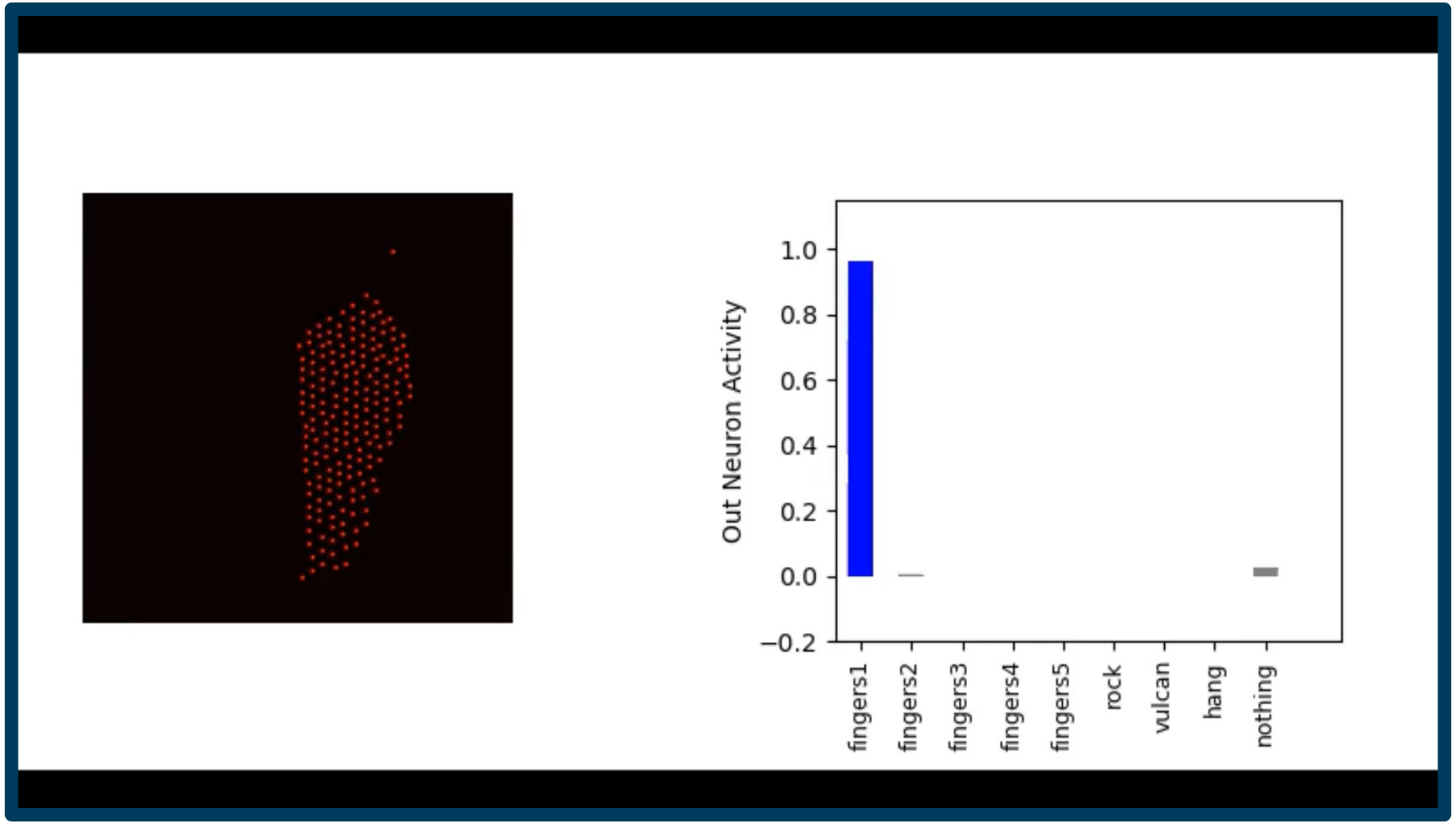
Revolutionizing AI at the Edge

DVS Gesture Learning & Recognition



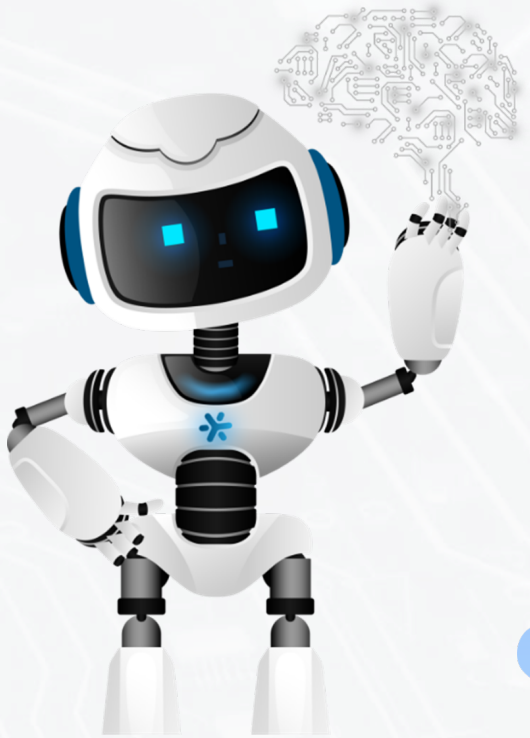


Gesture Classification



Gesture

brainchip



Akida™ Benchmark Data

Revolutionizing AI at the Edge

Keyword Spotting (KWS) Model Benchmarking

- We used a variation of a popular KWS model* that was compiled and optimized for
  - AKD1000 dev. board – 4-bit integer
  - Google Coral dev. board – 8-bit integer
  - Nvidia Jetson Nano – 16-bit floating point
- The KWS model was trained on the Google Speech Commands Data Set**
- We presented 10k inputs (batch size = 1) to each system and measured the idle power, inference power, and dynamic power

DS-CNN

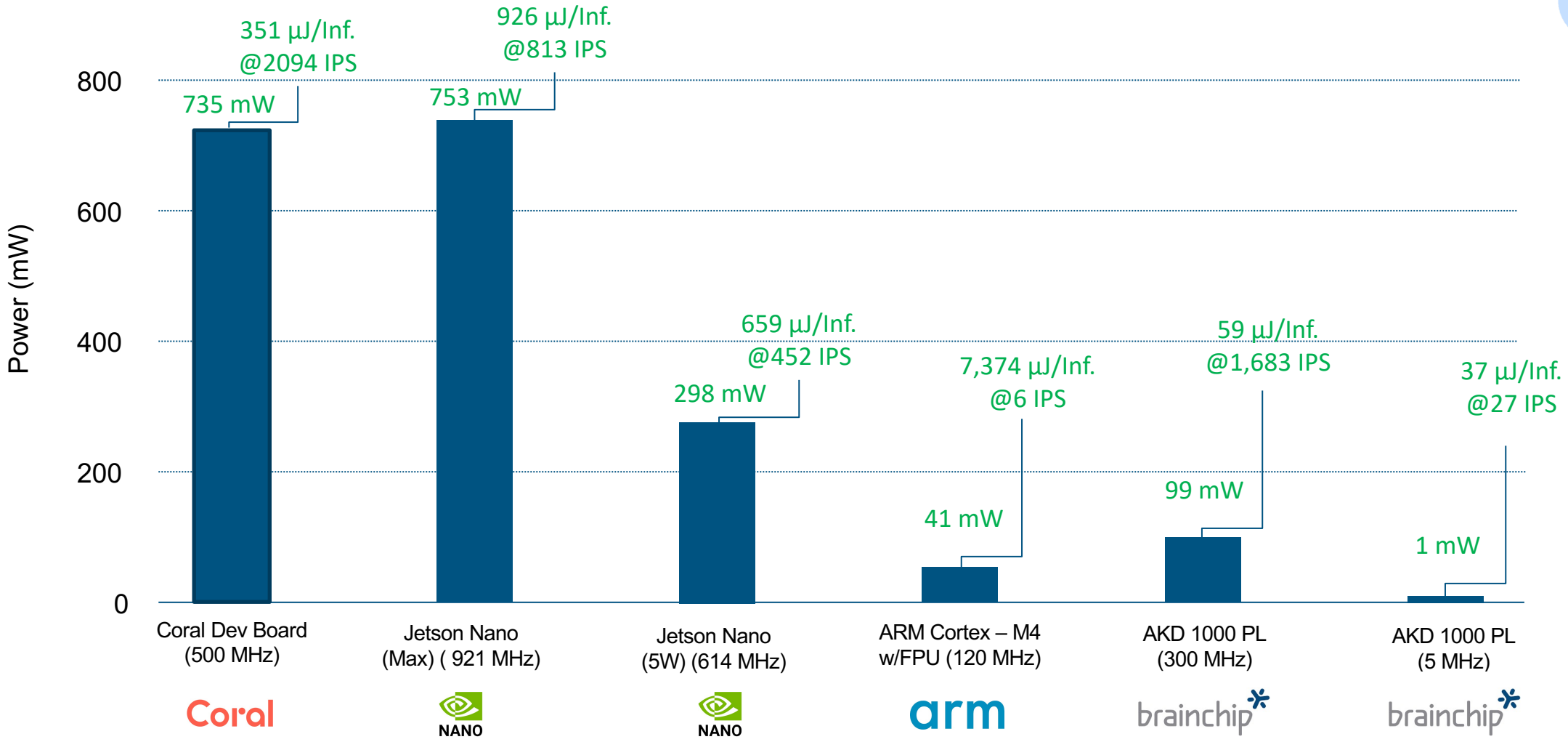
Layers	Output Dim
Input	49x10x1
Conv MP 5x5	25x5x32
DWS Conv 3x3	25x5x64
DWS Conv 3x3	25x5x64
DWS Conv 3x3	25x5x64
DWS Conv 3x3	25x5x64
DWS Conv 3x3	25x5x64
DWS Conv 3x3	25x5x64
GAP	1x1x64
DWS Conv 3x3	1x1x256
Dense	1x1x33

Total Params = 47,232 MACs/Inference = 2,538,112

* Zhang, Yundong, Naveen Suda, Liangzhen Lai, and Vikas Chandra. 2018. ArXiv:1711.07128 [Cs, Eess]. http://arxiv.org/abs/1711.07128.

** Warden, Pete. 2018. ArXiv:1804.03209 [Cs], http://arxiv.org/abs/1804.03209.

Power, Efficiency and IPS for Key Word Spotting (KWS)



We are the New Standard for Edge AI

BrainChip is revolutionizing the future of in-device Artificial Intelligence (AI) and is the worlds first commercial producer of neuromorphic semiconductor chips and IP.

Mercedes-Benz
Media Newsroom USA

Neuromorphic computing – a car that thinks like you

Another key efficiency feature of the VISION EQXX that takes its cue from nature is the way it thinks. It uses an innovative form of information processing called **neuromorphic computing.** The hardware runs spiking neural networks. Information is coded in discrete spikes and energy is only consumed when a spike occurs, which reduces energy consumption by orders of magnitude. **Working with California-based artificial intelligence experts BrainChip, Mercedes-Benz engineers developed systems based on BrainChip's Akida hardware and software.** The example in the VISION EQXX is the "Hey Mercedes" key-word detection. **Structured along neuromorphic principles, it is five to ten times more efficient than conventional voice control.**

Although neuromorphic computing is still in its infancy, systems like these will be available on the market in just a few years. When applied on scale throughout a vehicle, they have the potential to **radically reduce the energy needed to run the latest AI technologies.**

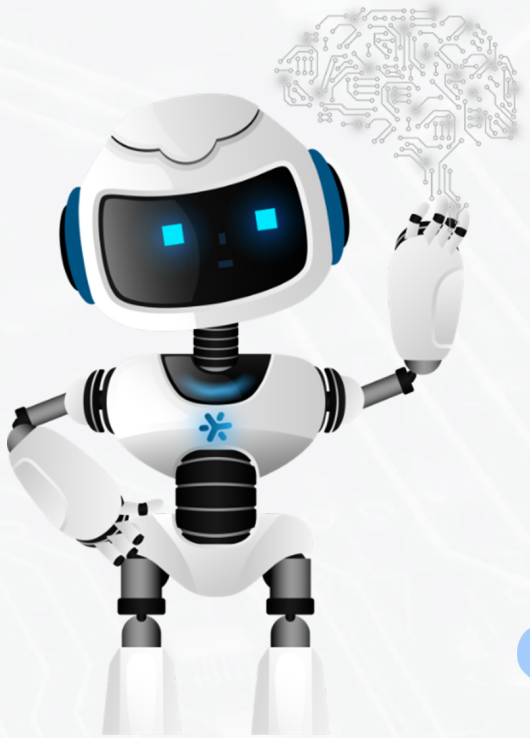
VentureBeat

EE Times

Design & Reuse

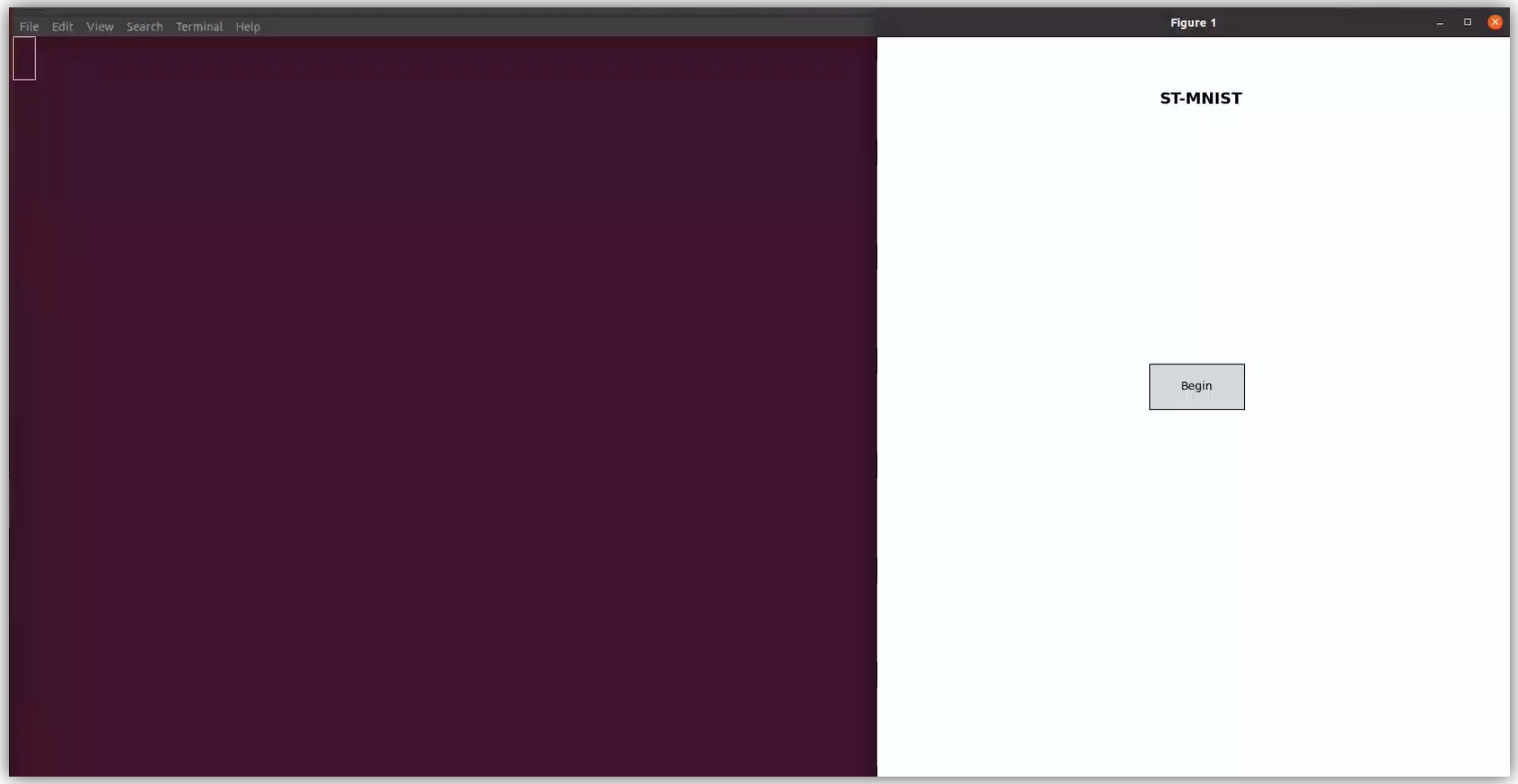
brainchip

Akida™
Vibration Analysis
Tactile Sensing

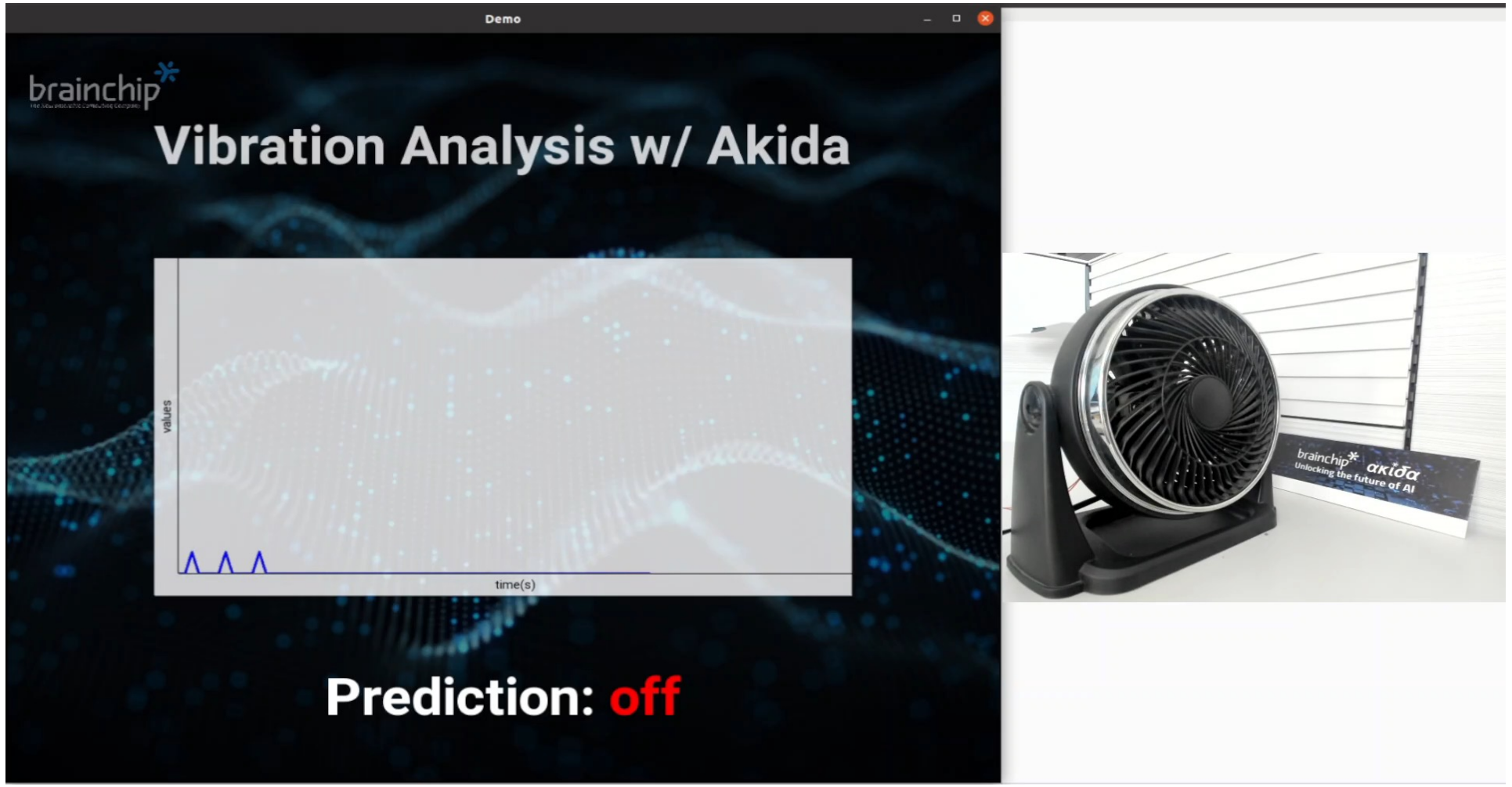


Revolutionizing AI at the Edge

ST-MNIST

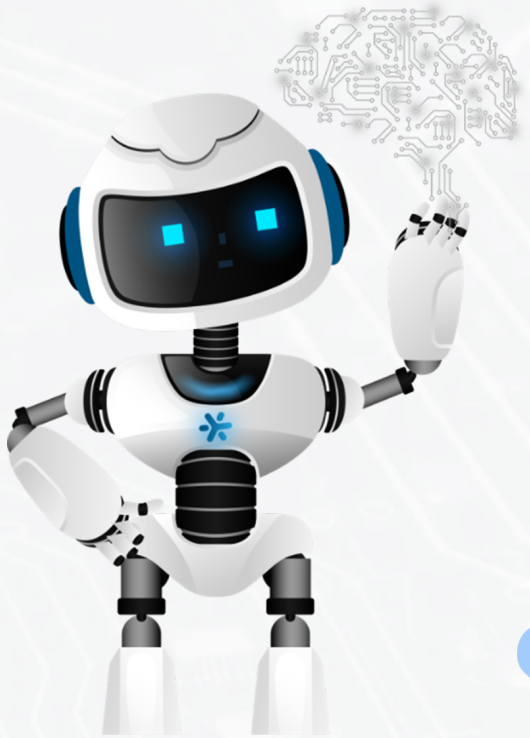


Vibration Analysis



brainchip

Akida™ Gustatory Beer Tasting



Revolutionizing AI at the Edge

Gustatory Sensing Demo

BrainChip - This is Our Mission

- Traditional AI is compute intensive

- Traditional AI solutions do not address the future for Edge AI environments
  - Too much irrelevant data is being processed
  - Consuming too much power
  - Too many dollars are being spent in the wrong areas

- The key is not to process more data faster

- **Process Relevant Data Efficiently with Accuracy**

- Neuromorphic Computing is the next generation of AI
  - Processing the most relevant data: smarter, faster and most economical

- BrainChip is first to market and is leading the way!

This is our Mission

We don’t make the sensors

- We make them smart

We don’t add complexity

- We eliminate it

We don’t waste time

- We save it

We solve the tough *Edge* AI problems

- Others do not or cannot solve

Intelligent AI Everywhere
This is our Mission

www.brainchip.com

Empowering Product Creators to Harness Edge AI and Vision

The Edge AI and Vision Alliance (www.edge-ai-vision.com) is a partnership of 100+ leading edge AI and vision technology and services suppliers, and solutions providers

Mission: To inspire and empower engineers to design products that perceive and understand.

The Alliance provides low-cost, high-quality technical educational resources for product developers

Register for updates at www.edge-ai-vision.com

The Alliance enables edge AI and vision technology providers to grow their businesses through leads, partnerships, and insights

For membership, email us: membership@edge-ai-vision.com

Join us at the Embedded Vision Summit

May 17-19, 2022—Santa Clara, California

The only industry event focused on practical techniques and technologies for system and application creators

- *"Awesome! I was very inspired!"*
- *"Fantastic. Learned a lot and met great people."*
- *"Wonderful speakers and informative exhibits!"*

Embedded Vision Summit 2022 highlights:

- **Inspiring keynotes** by leading innovators
- High-quality, practical **technical, business and product talks**
- Exciting **demos**, **tutorials** and **expert bars** of the latest applications and technologies

Visit www.EmbeddedVisionSummit.com to learn more

BrainChip Demo Links- YouTube @BrainChipInc

Wine Tasting: https://youtu.be/RwWTietRioM?t=1

Edge Based Learning: https://youtu.be/912hYD1qJuk

Keyword Spotting: https://youtu.be/wDWHQqi69qo

Visual Wake & Facial Recognition: https://youtu.be/EoG0_AUv3Zo

Smart Automotive In Cabin Experience: https://youtu.be/tin55B83r_M?t=107

Edge Based Learning (High Speed Environment): https://youtu.be/GvfDKILIbOk

Gesture Control: https://youtu.be/Rin5RGTHiOk