



# Effortless ML for Embedded Edge

Gopal Hegde

SVP, Engineering & Operations

AI Hardware & Edge AI Summit

September 2023

# Who we are

Expertise

**160+**

**of the industry's  
best ML minds**

Funding

**\$200M**

**Fidelity, Dell,  
MSD, Adage..**

Create Tomorrow, Today

**June 2022**

**Shipping to  
customers**

# Our focus: Embedded edge market



Smart Vision



Robotics &  
Industrial 4.0



Advanced Driver  
Assistance Systems



Government  
Sector



Industrial Drones



Healthcare



Smart Agriculture



Smart  
Construction

We are a **software**  
company that is *building*  
*our own silicon*

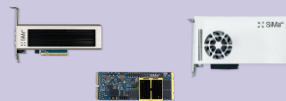
Goal: **No code** ML  
environment for 1000+  
customers

Scale ML at the Embedded Edge

Effortless ML Customer Experience

ANY. 10X. Pushbutton

 **Palette**™ Software



Boards



MLSoC™ Silicon

# MLPerf: First startup to be better than NVIDIA in performance and power for embedded edge category



## Software Solves a Silicon Problem, How a SiMa.ai Chip Outperformed the Industry Leader

Randa Scott  
Apr 12, 2023 · 4 min

Collection: Our Portfolio

PROFILES



“Silicon Valley has been living up to its name too much with the belief that every problem can be solved only by improving the semiconductor.”  
— Krishna Rangasayee, CEO SiMa.ai



Despite rapid advancements in artificial intelligence (AI), only a small fraction of today's total addressable market is benefitting from these innovations, according to Krishna Rangasayee, founder and CEO of AI chipmaker SiMa.ai. Applications where running AI processes at the edge could spur significant innovation are being left out. “Automotive, medical, robotics, industrial, government, all these smart-vision systems, they're vastly being ignored (by chipmakers),” he says. Yet the financial opportunity for successfully developing a chip that can efficiently run AI at the embedded edge for any one of those individual vehicles is enormous and obvious. How is it possible there's such a gap in solutions?

“Silicon Valley has been living up to its name too much with the belief that every problem can be solved only by improving the semiconductor,” Rangasayee says. “But if you really talk to customers and understand their pain points, they visualize their problem through the lens of software.”

## Start-up SiMa trumps Nvidia's chips

The start-up's performance data exceeds that of the established competition. The expectations of SiMa and founder Krishna Rangasayee are therefore immense - and even TSMC seems convinced.

Martin W. Szeszani | Joachim Huber

04/11/2023 · 09:08 am · 3 min read



Krishna Rangasayee  
Chips from the SiMa founder have attracted industry leaders in terms of performance. (Photo: SiMa.ai)

**Stuttgart.** For the fact that Krishna Rangasayee builds what is probably the most powerful chips for artificial intelligence (AI), he speaks calmly about his success. “We set out to be a leader in the machine learning industry. Now we have the technical confirmation for this,” says the founder of the start-up SiMa.ai. The MLPerf (Machine Learning performance) brought confirmation - a kind of annual chip competition.

Just over four years ago, the now 53-year-old Stanford graduate started SiMa.ai. In the most recent performance comparison, his chip performed 50 percent better than Nvidia's in terms of speed and power.

## SiMa.ai leads in its debut MLPerf Benchmark performance in the Closed Edge Power category

SiMa.ai is the first startup to participate and achieve winning results in the industry's most popular MLPerf image benchmark: ResNet-50, enabling scaling of ML at the embedded edge.

ETAuto  
Updated On: Apr 6, 2023 at 03:42 PM IST

Based By: 53 Industry Professionals



Krishna Rangasayee, CEO and Founder, SiMa.ai and Harald Kreyer, President, Automotive Business, SiMa.

### Startup with Indian roots beats Nvidia at ML Olympics

ARUN GEORGE | THE / Apr 9, 2023, 10:15 IST

**STARTING WITH INDIAN ROOTS BEATS NVIDIA AT ML OLYMPICS**

YOU'RE READING

It's almost impossible to beat Nvidia in AI enhancing. Yet, that's exactly what Krishna Rangasayee, a year-old SiMa.ai founder

SiMa.ai is a 4-year-old semiconductor startup founded by Krishna Rangasayee in San Jose, California, with half of its 140 employees based out of Bengaluru. This little David has just beaten the Goliath in the semiconductor space, Nvidia, in a crucial category in what some call the machine learning Olympics. And it did it in its maiden attempt.

The competition, called MLPerf, pits the latest chips against each other in benchmarking trials in different categories to see who comes out on top. It's conducted by an initiative called MLCommons, founded by the who's who of the tech world, and whose objective is to accelerate machine learning innovation to benefit everyone.

Winning the MLPerf competition is a big deal. It's a benchmarking trial in different categories to see who comes out on top. It's conducted by an initiative called MLCommons, founded by the who's who of the tech world, and whose objective is to accelerate machine learning innovation to benefit everyone.

With the world's best competing companies on the planet, it's hard to win. It's a bit like the Olympics, but instead of athletes, it's AI chips. The winner is the one who can do the most work with the least power. It's a bit like the Olympics, but instead of athletes, it's AI chips. The winner is the one who can do the most work with the least power.

We announced the MLPerf competition in 2019 and we've had over 100 people and 100+ companies participating in it. It's a bit like the Olympics, but instead of athletes, it's AI chips. The winner is the one who can do the most work with the least power.

With all the announcements in the past few months, it's hard to win. It's a bit like the Olympics, but instead of athletes, it's AI chips. The winner is the one who can do the most work with the least power.

160 engineers | Pushbutton results | 16nm MLSoC



# Deploying performant ML at embedded Edge is hard

Accelerating models isn't enough | **Accelerate entire application**

**Cloud: Only accelerate ML models**

INT8 required for higher performance + power | **Quantize with acceptable accuracy**

**Cloud: FP16 provides better accuracy**

Wide variety of models, customer proprietary data | **Pushbutton tools are a MUST**

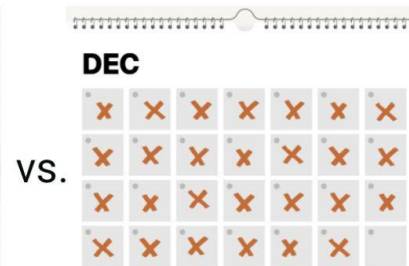
**Cloud: Experienced engineers can use complex tools to get best possible performance**

---

# Our Vision: Effortless machine learning for the embedded edge

## Our strategy

Run **any** computer vision application, **any** network, **any** model, **any** framework, **any** sensor, **any** resolution.



Any

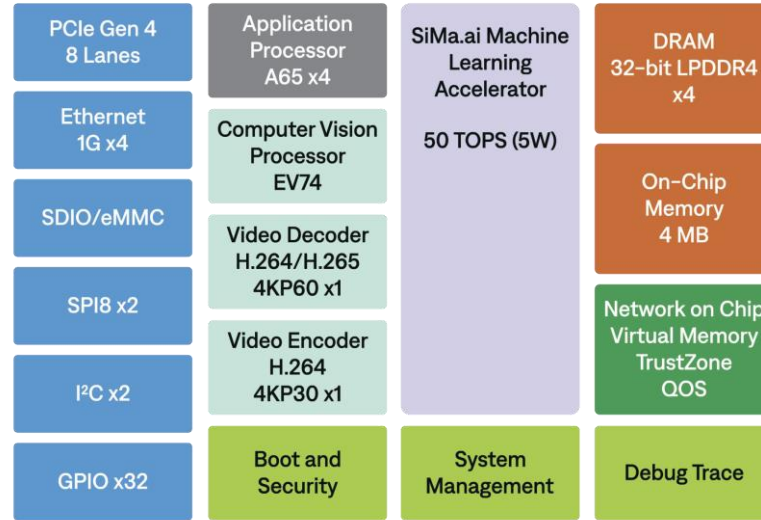
10x

Pushbutton

# MLSoC™: A new product category purpose-built for the Edge



## SiMa.ai MLSoC™



First **software-centric** purpose-built MLSoC platform with **push-button** performance



# Key innovations

## ANY and 10x

### Highly flexible ML accelerator

**Fully programmable MLA**



Supports full complement of CV applications at lowest power

**SW controlled data movement, scheduling & synchronization**



SW control of cache/memory hierarchy, data movement: minimal data movement, small cache, high compute efficiency and lowest power

### Secure, self-contained SoC

**Seamless heterogeneous compute**



Enables ML capabilities for legacy apps; future proofs applications

**Optimized end-to-end Customer app**



Highly optimized building blocks enable best in class end-to-end performance

## PUSHBUTTON

**Effortless evaluation, Development & deployment**

**Effortless customer integration**



**Low code ML**  
Customers can develop & deploy applications without having to understand details of HW

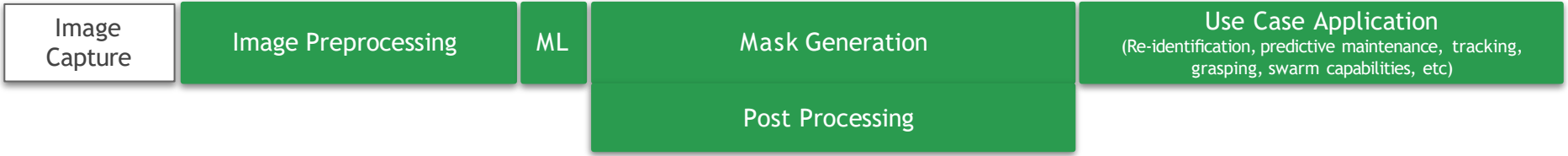
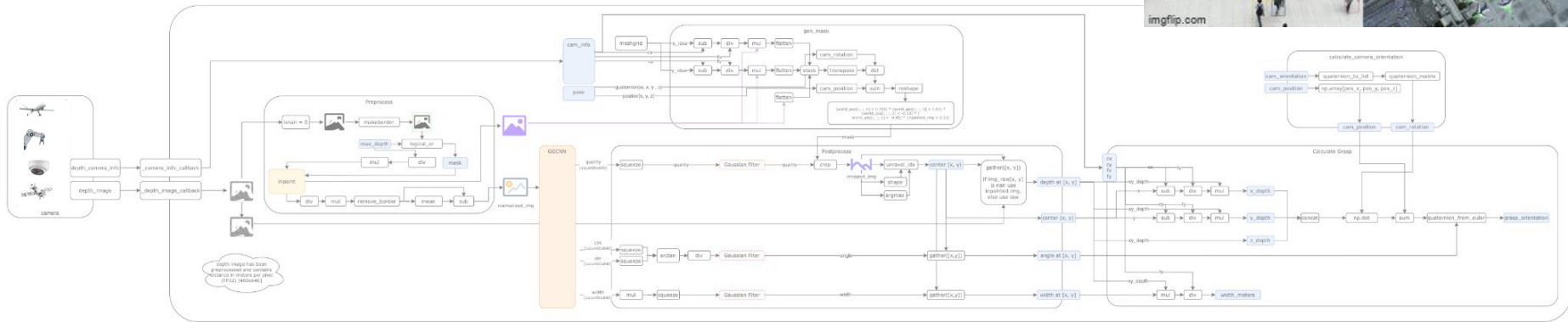
**Low code customer evaluation**



**Palette - Edgematic**  
Allows customers to build apps & evaluate without writing code

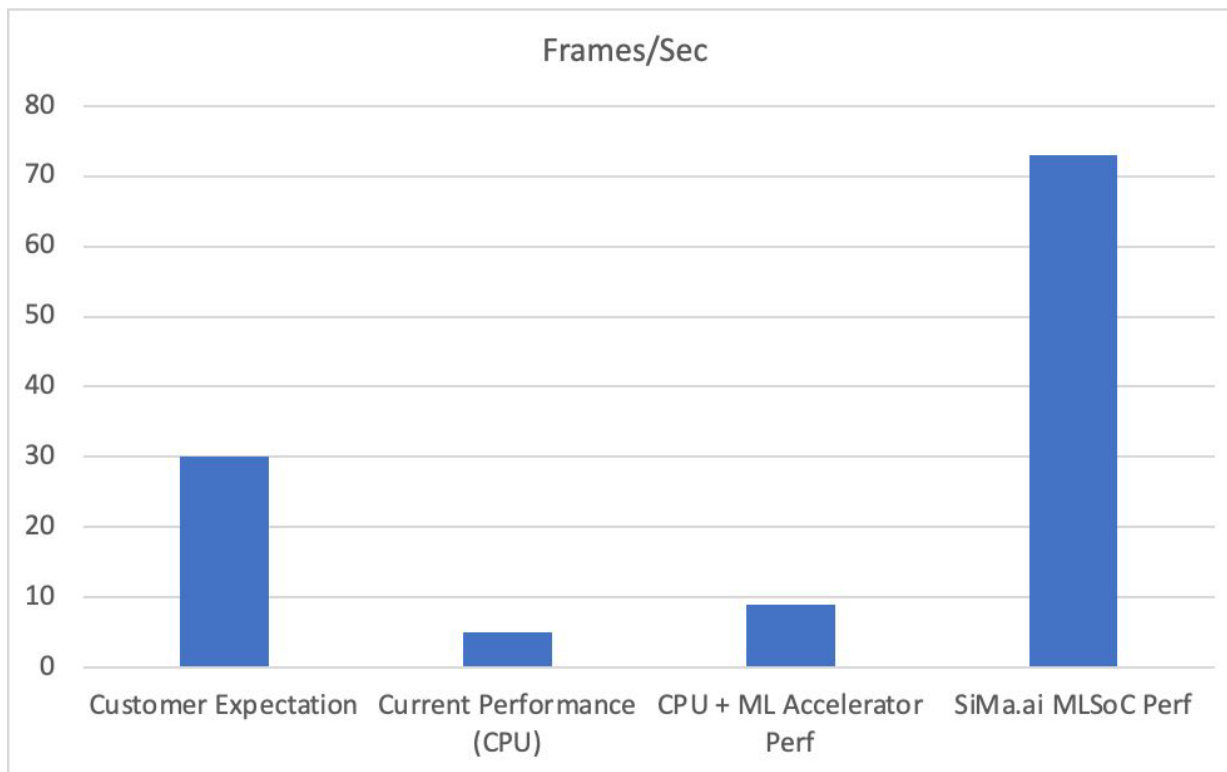
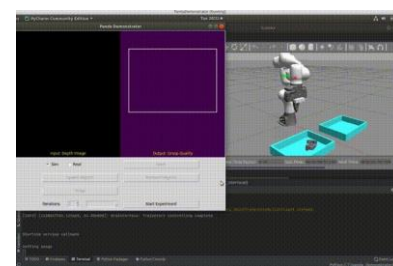
# MLSoC accelerates end2end customer apps

## Delivers unparalleled end2end performance



# Accelerating end2end applications

## Real-world Robotics customer example



# SiMa.ai offerings - In production today

Software



Production silicon



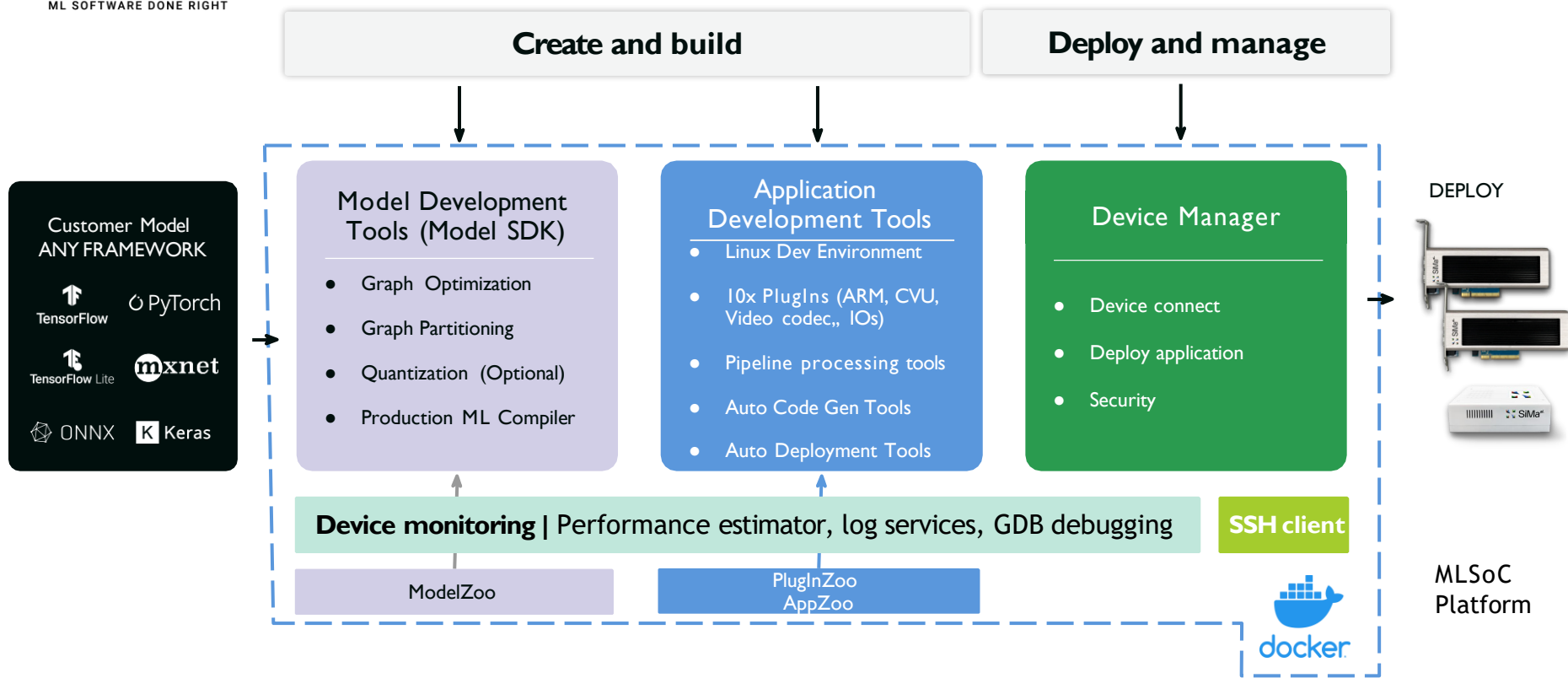
Production boards



DevKit

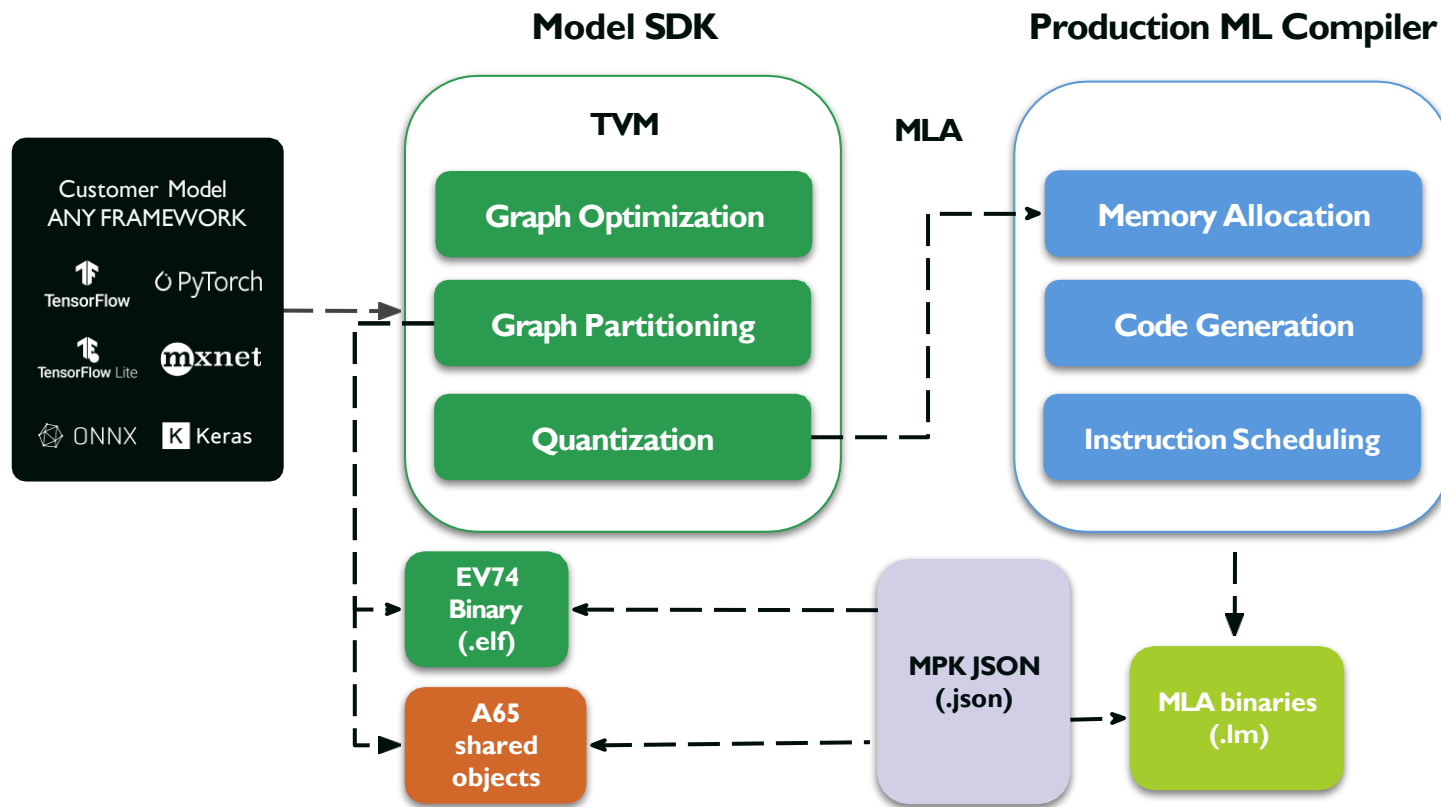


# SDK Overview - PushButton Tools



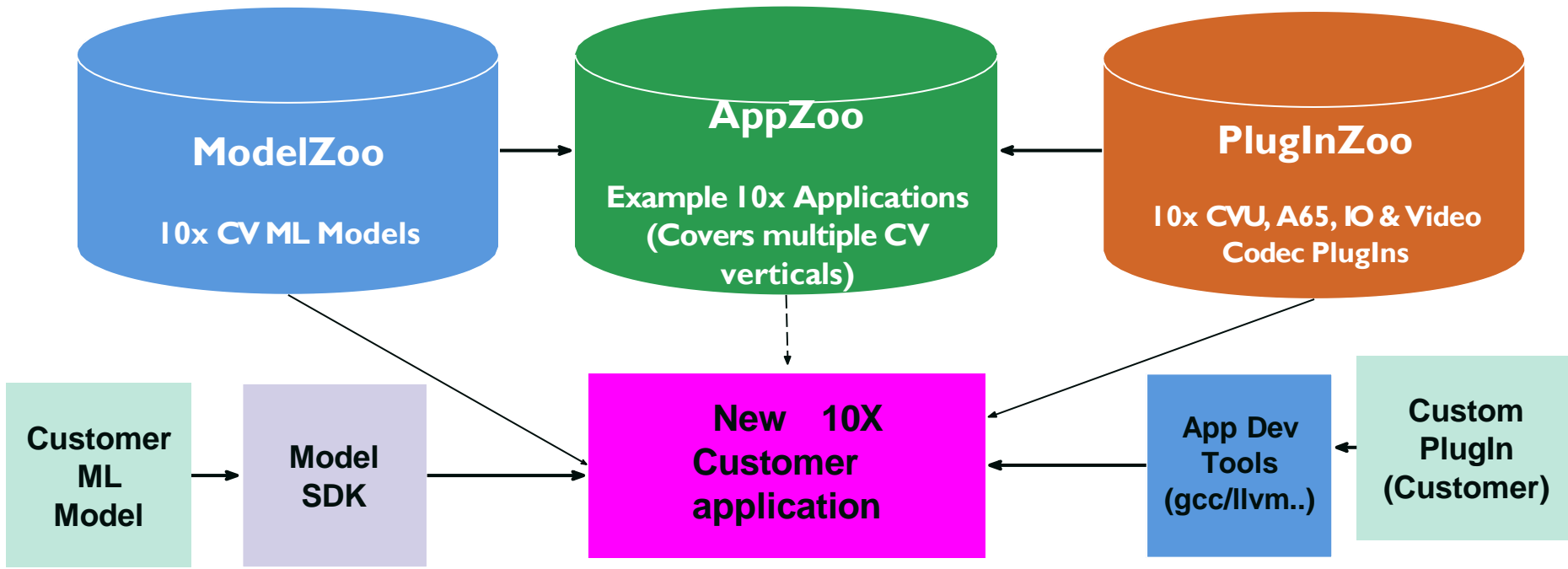
**Development tools run on any host machine**

# Enabling “Any” CV ML model



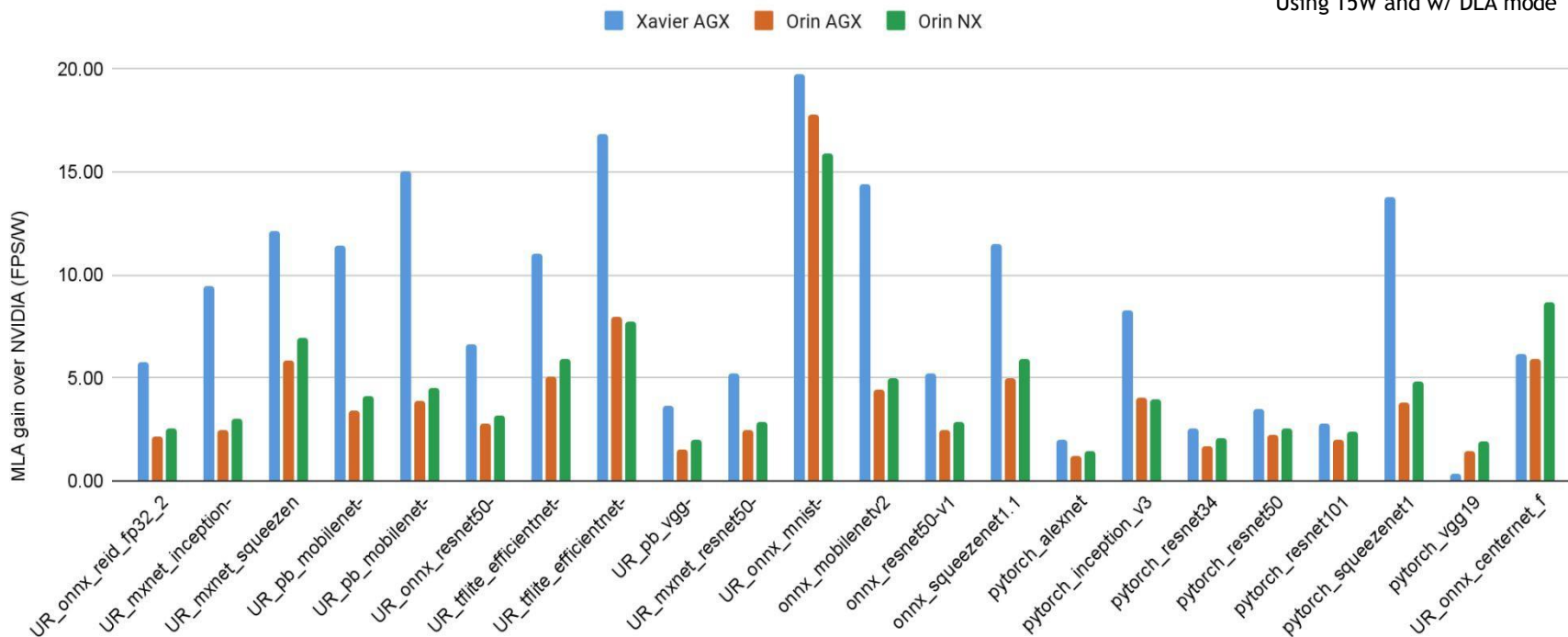
# Enabling 10x CV End2End Applications

Leverage prebuilt optimized building blocks or build your own using PushButton tools



# SiMa.ai beats NVIDIA across wide variety of Edge ML networks

Using 15W and w/ DLA mode



MLSoC is 10x Xavier and 4x Orin on embedded edge ML network performance



# Summary

- SiMa.ai MLSoC is industry's **FIRST** purpose built ML inference solution for embedded edge
- 1st edge machine learning SoC solution to demonstrate performance superiority over nVidia in MLPerf benchmarking
- Palette (™) provides highly scalable, Effortless software experience to our customers
- In production today