tinyML® Foundation

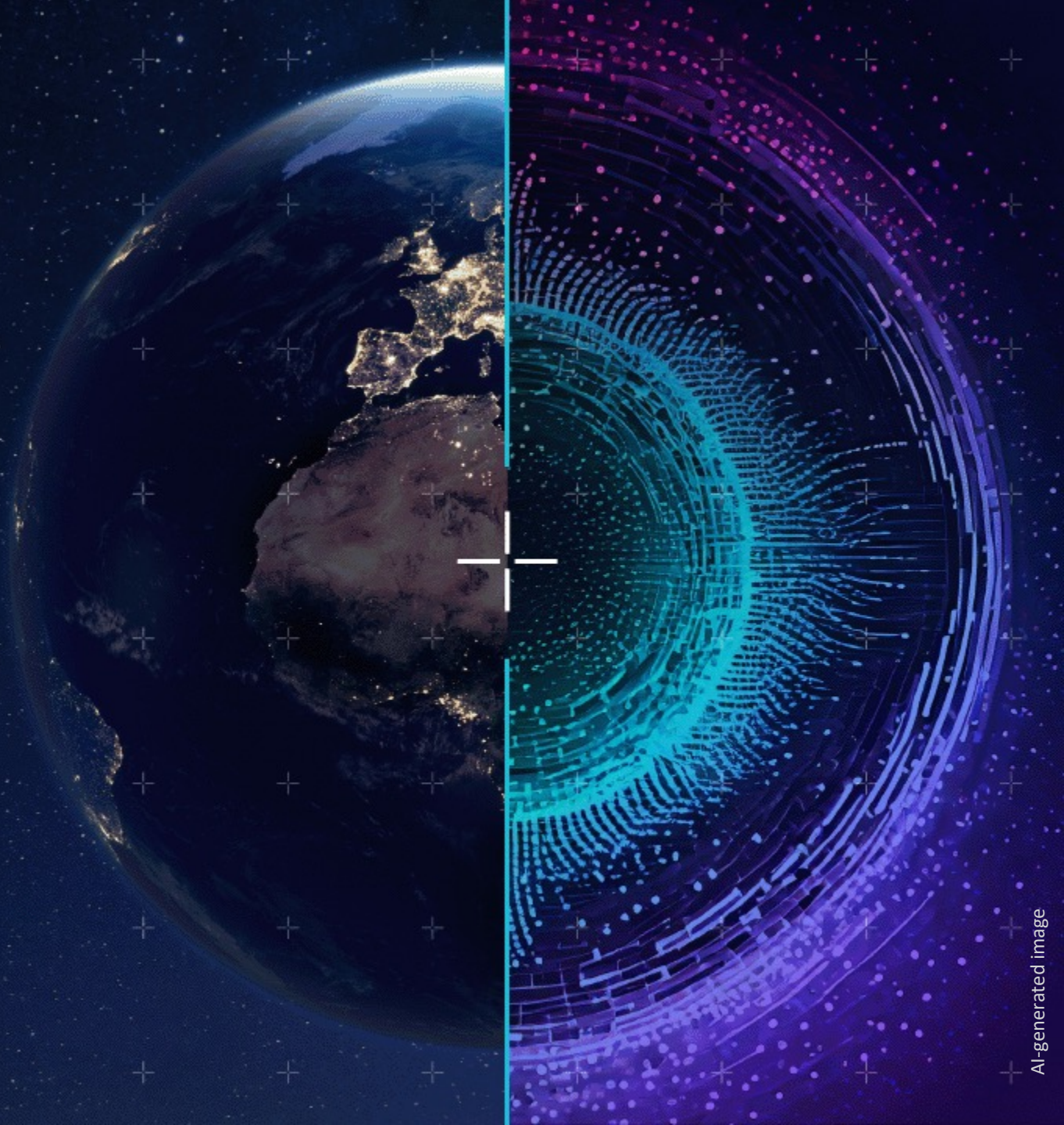
Enabling Ultra-low Power Machine Learning at the Edge

tinyML Summit April 22 - 24, 2024

www.tinyML.org

arm

Accelerating Edge AI Innovation

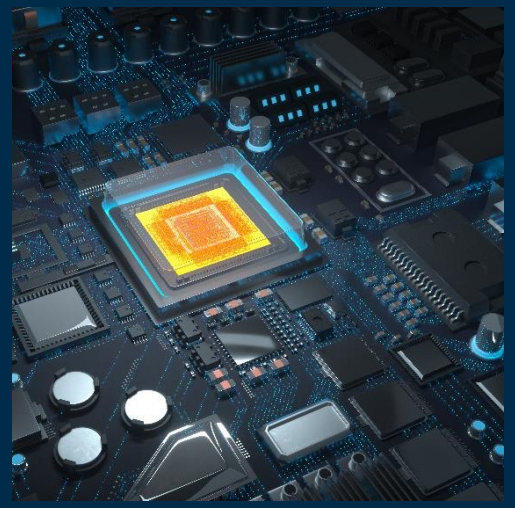


Parag Beeraka

April 2024

AI-generated image

The Evolution of Compute at the Edge



Embedded

Connectivity

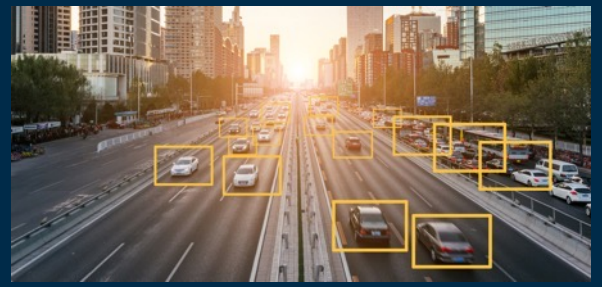
AI

AI is Creating Incredible Opportunities

PAST PRESENT FUTURE

Home

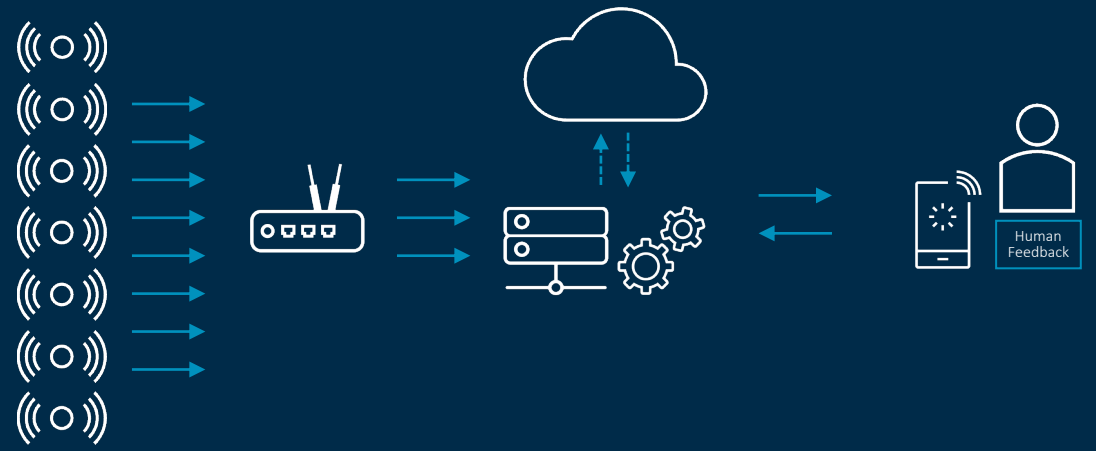
City

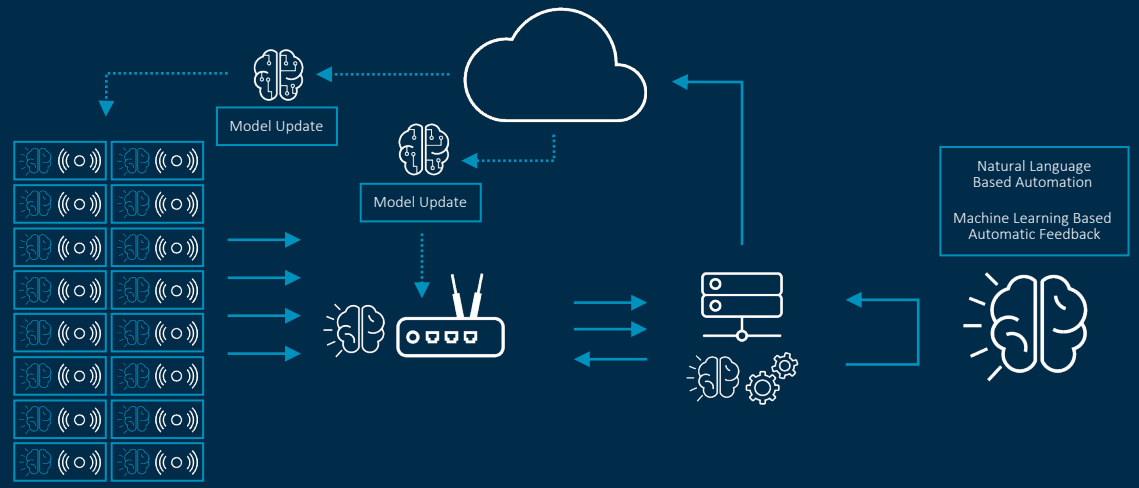


Retail

arm

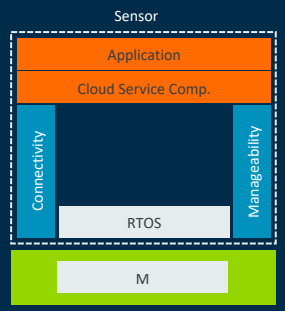
Transformation is Under Way

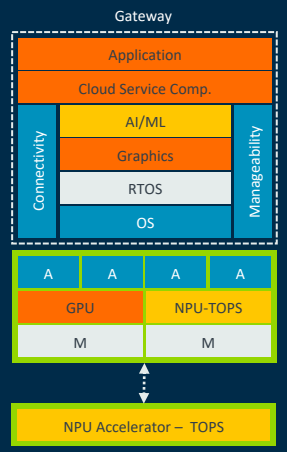


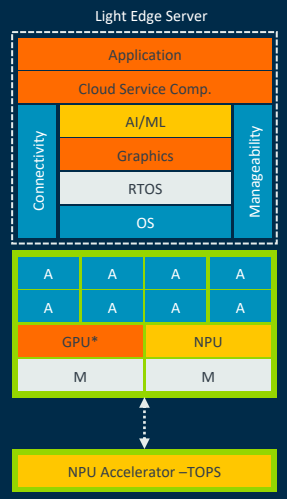


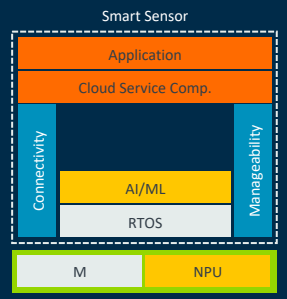
Sensors | Gateway | Cloud Analytics | Manual feedback | Smart Sensors | Smart Gateway | On Prem + Cloud Analytics | Automatic Feedback

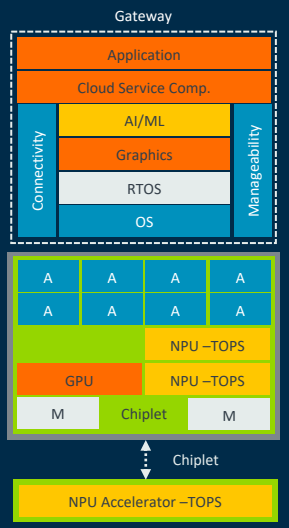
2024 ---------- 2030

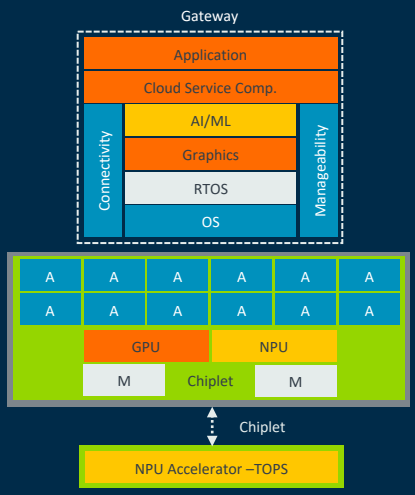








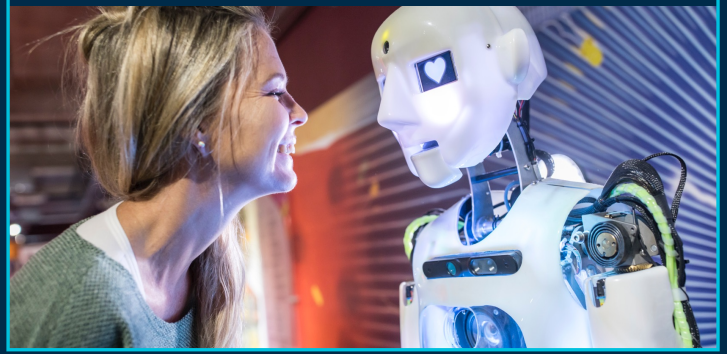




Generative AI will Further Enhance the Opportunity

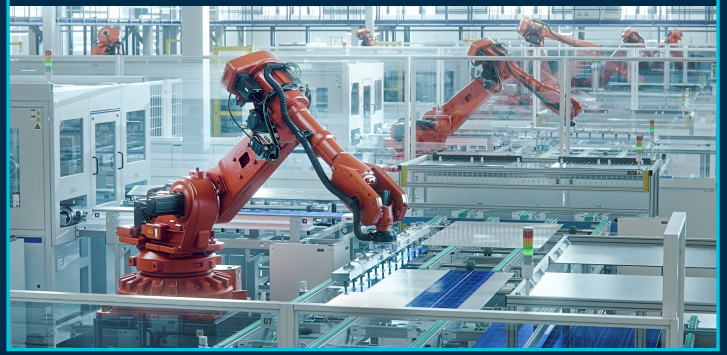
Live Language Translation

Conversational Robots



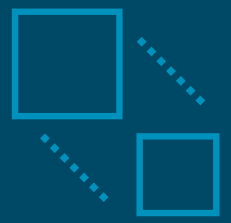
Speech/Music Generation

Code Generation (Industrial)



Meeting the Needs of the Market As It Evolves

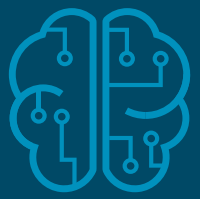
Navigating growing performance demands and increased software complexity



Range of Performance

Time-To-Market

Low-Power Design



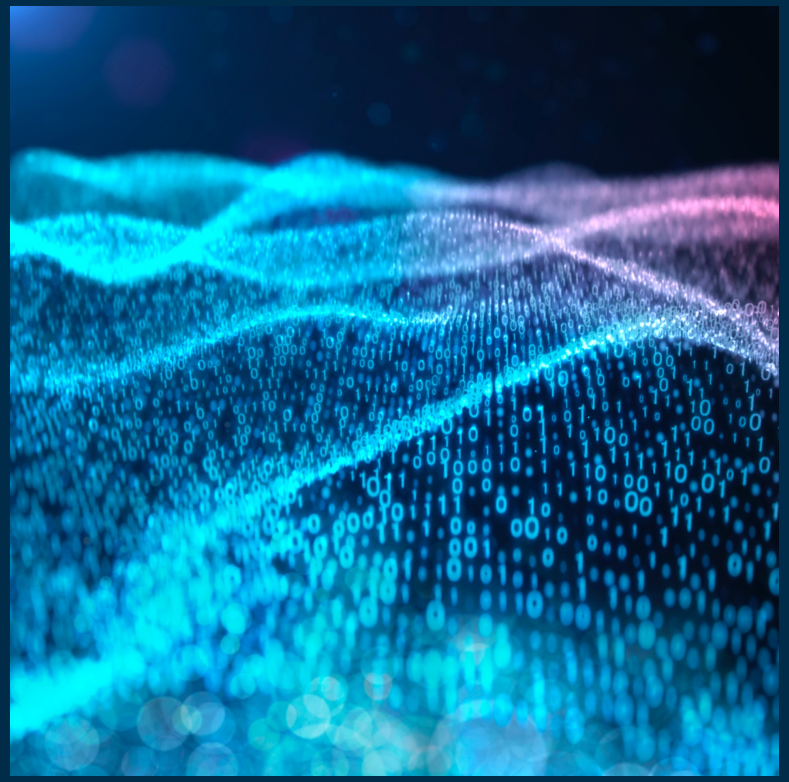
Machine Learning Integration

Software Portability

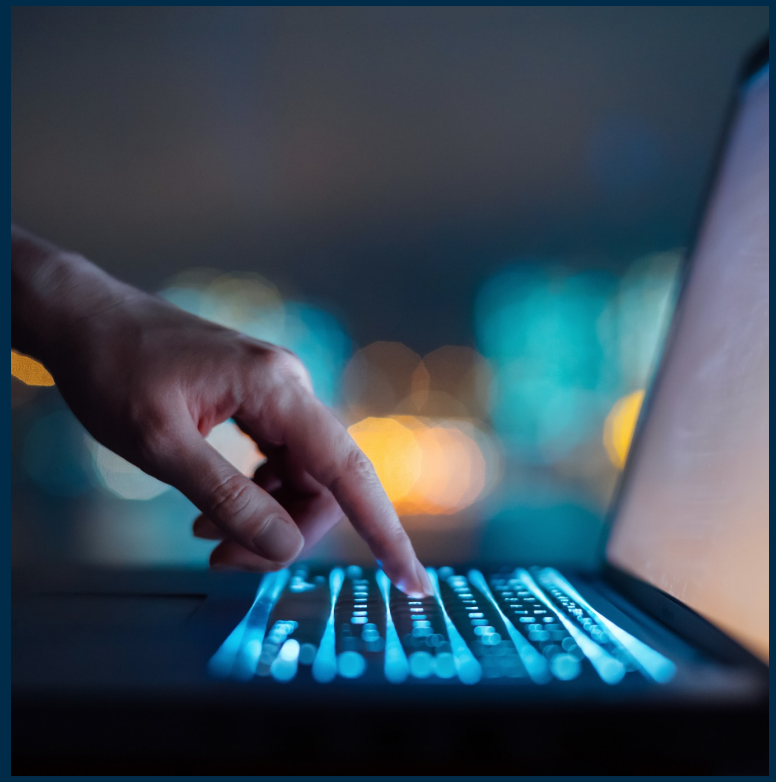
Broad ecosystem support

Arm is Enabling tinyML with Three Focus Areas

Hardware



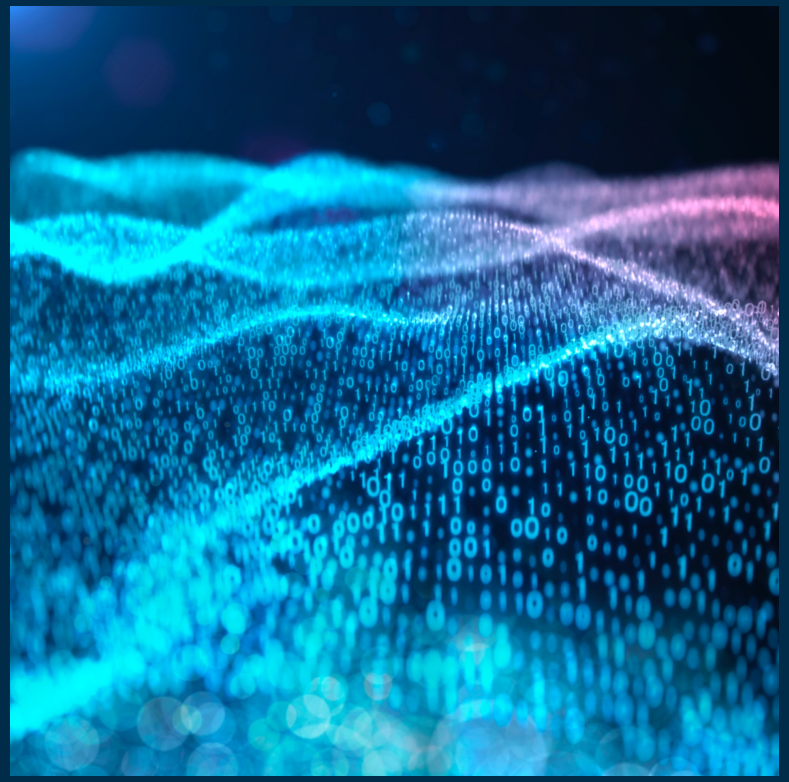
Software



Ecosystem

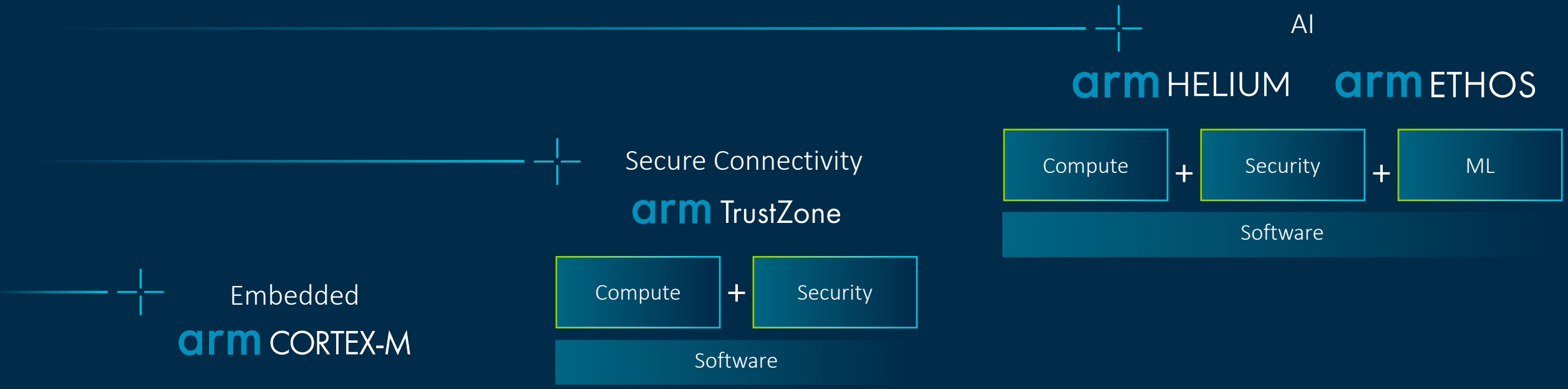
Arm is Enabling tinyML with Three Focus Areas

Hardware



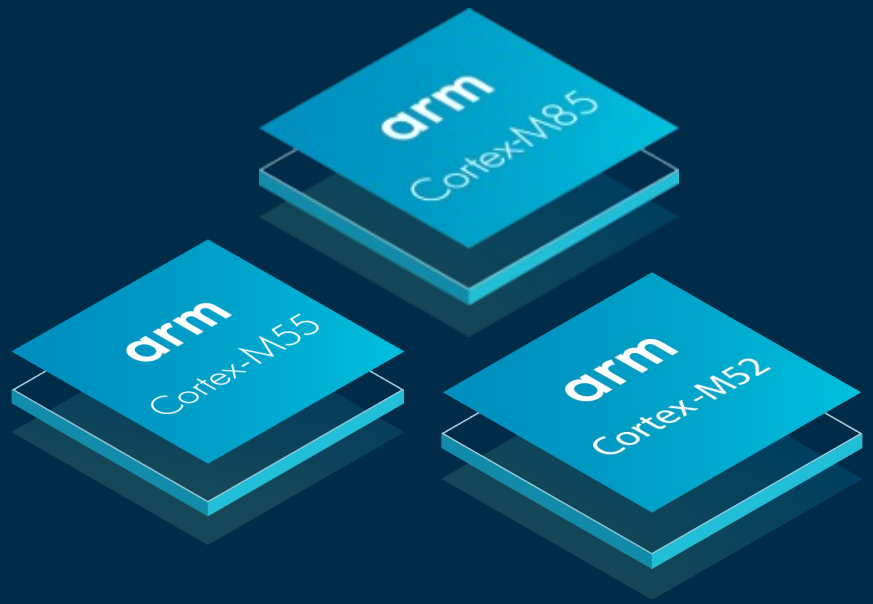
Arm has Delivered Continued Innovation for tinyML

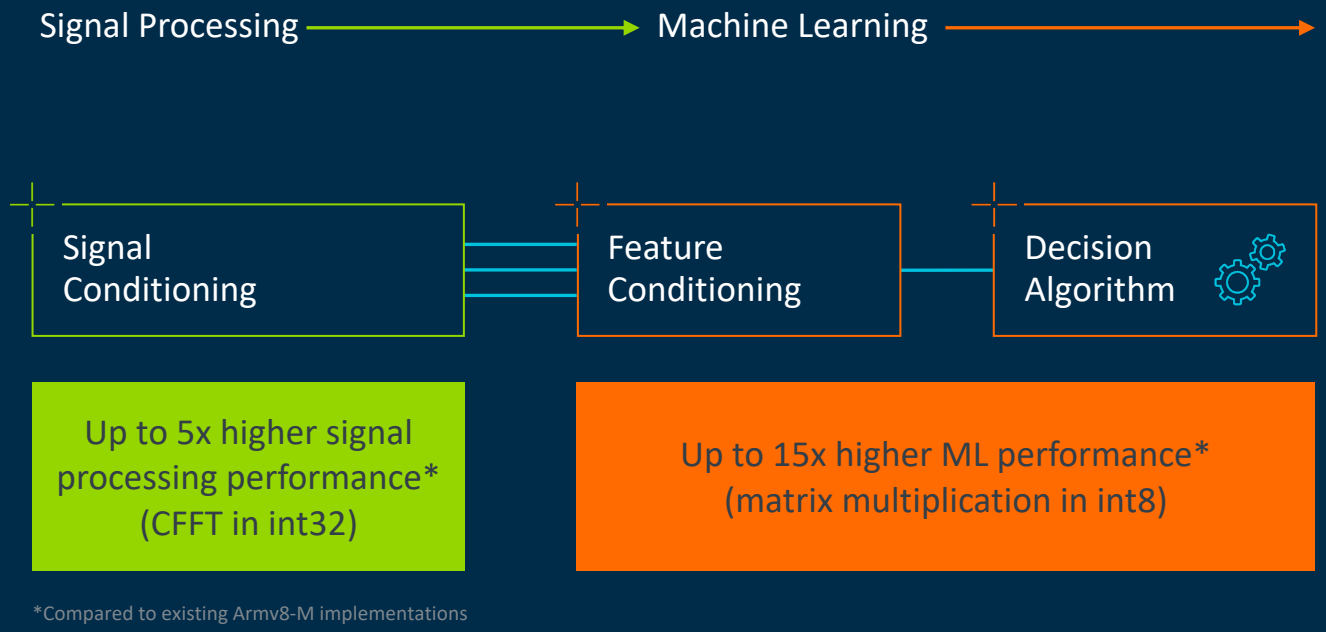
Addressing evolving technology requirements to enable innovation and commercial success



arm

Helium Delivers Significant Uplift for ML on Cortex-M

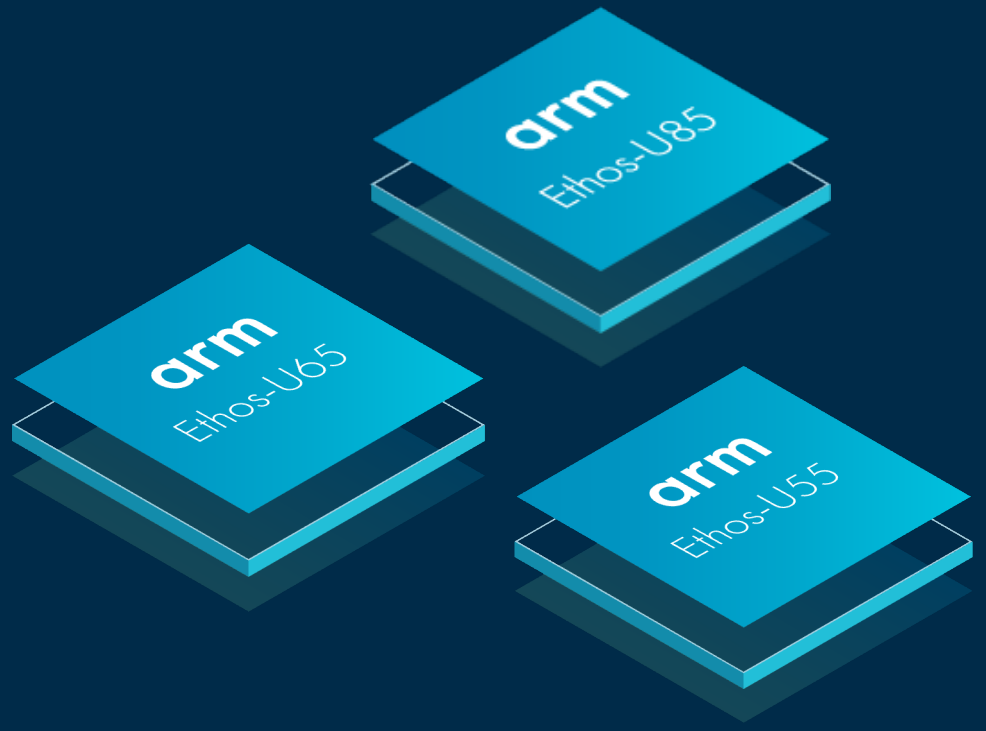




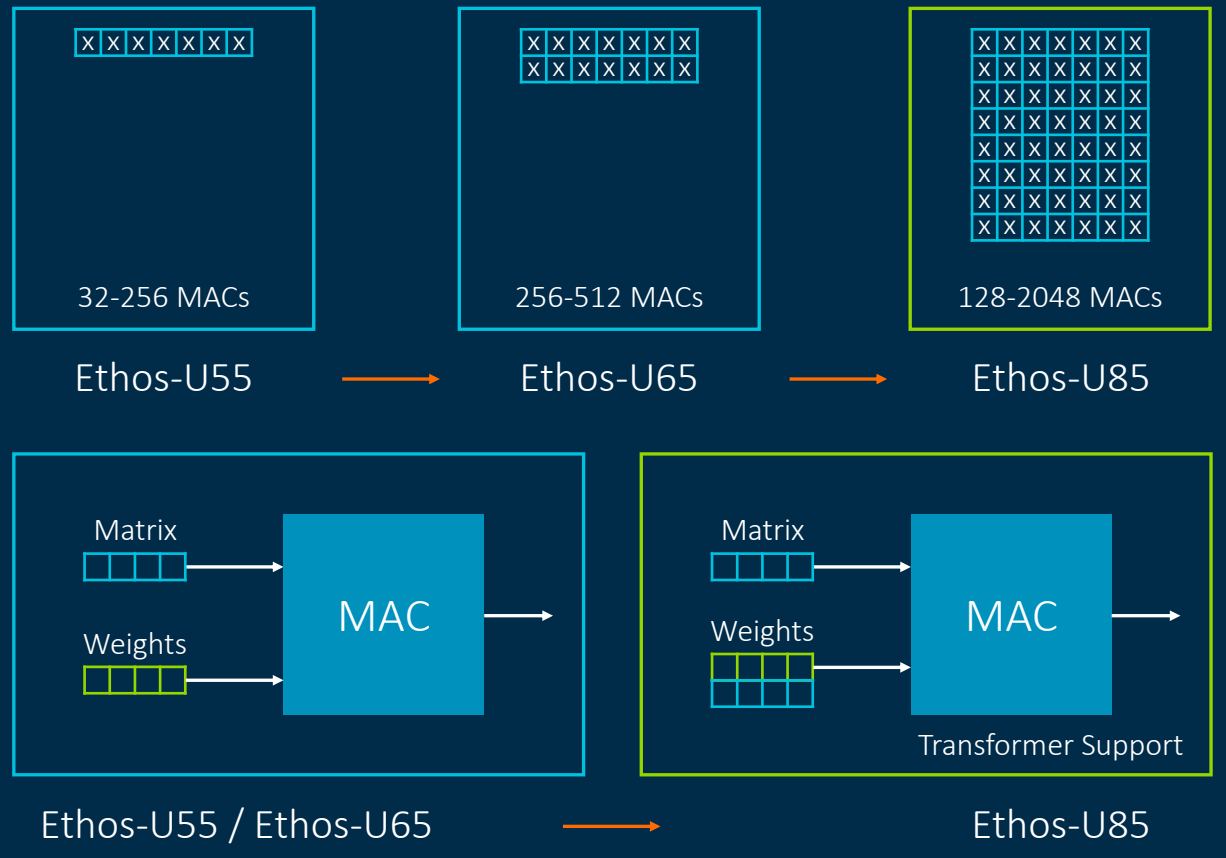
arm

Unlocking the Full Potential of Neural Networks

Accelerating implementation of higher performance AI enabled systems



arm ETHOS-U



arm

Arm Ethos-U85

Offering best in class features for neural network acceleration

Increased Performance

Configurations from 128 MACs to 2048 MACs (4 TOPs)

Higher Power Efficiency

20% more energy efficient than previous generation

Extended Operator Support

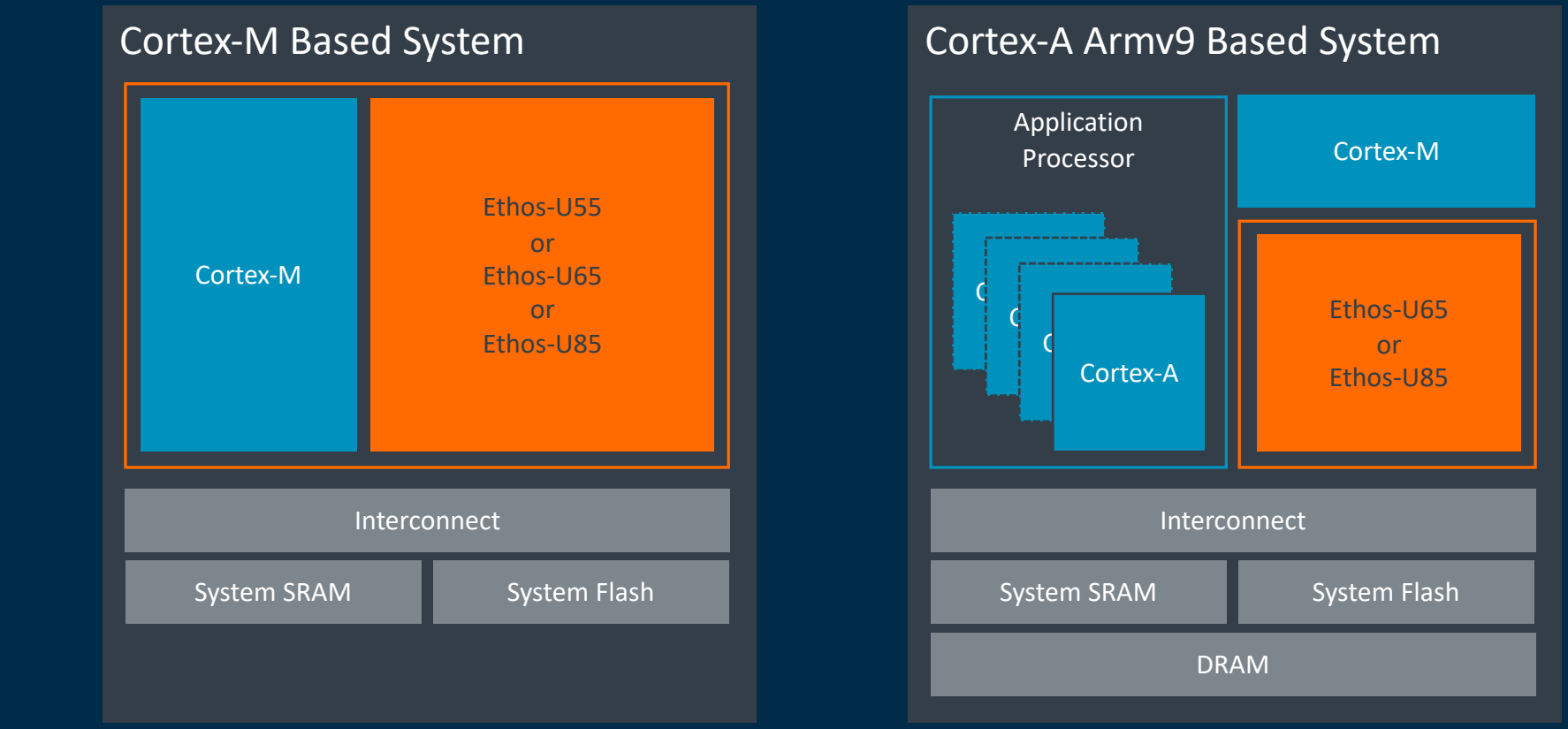
Transformer network support for faster customization

4X

Performance improvement over previous generation

Arm Ethos-U85 Scales to Support tinyML and other Workloads

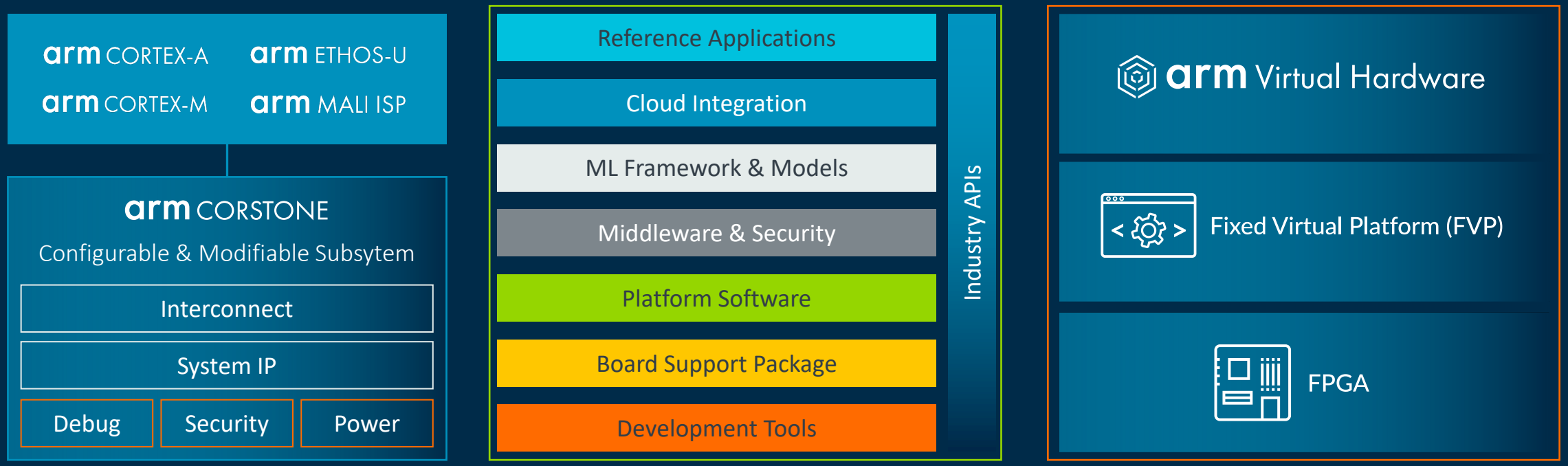
Supports small MCU implementations to high performance Armv9 MPU systems



A Holistic Approach for tinyML

Enabling rapid time to market for edge AI device deployment

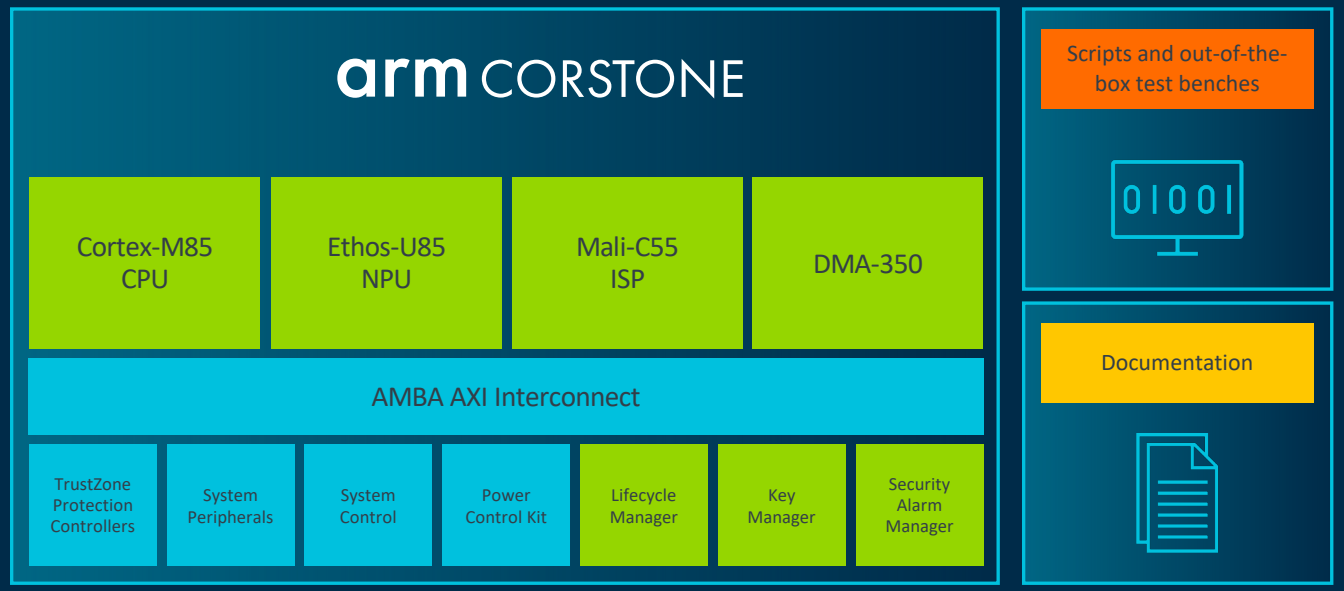
Reference Design | Arm and Ecosystem Software and Tools | Arm Prototyping Platforms



Corstone-320

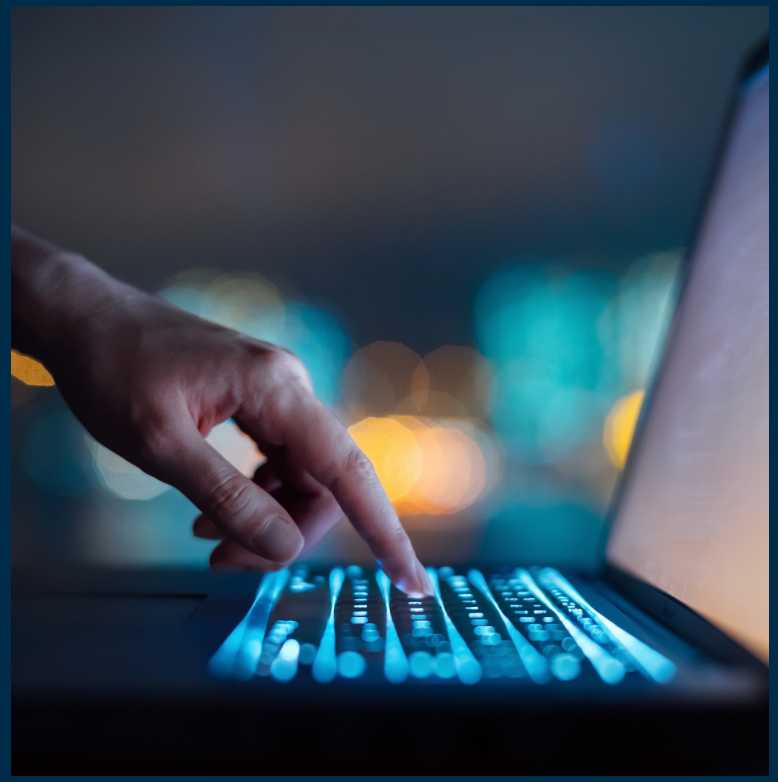
Arm's highest performing IoT reference design

- Bringing together leading-edge embedded IP
  - Cortex-M85, the highest performance MCU processor with Helium technology
  - New Arm Ethos-U85 for up to 4 TOPs AI/ML workload acceleration
  - Mali-C55 image signal processor
- Suitable for use cases including vision, voice, audio and general-purpose Edge AI
- Designed to meet IoT security standards



Arm is Enabling tinyML with Three Focus Areas

Software



Enabling Comprehensive Software and Tools

MACHINE LEARNING

FRAMEWORKS

COMPILERS
RUNTIMES

TensorFlow Lite

PyTorch

Vela

arm COMPILER

TFLM

µtvm

SOFTWARE

LIBRARIES

EVAL KITS

CMSIS-NN

ML Eval Kit

mlia

STANDARDS

IMPLEMENTATION

SPECIFICATIONS

TOSA

TOOLS

ECOSYSTEM TOOLS

ARM DEVELOPMENT TOOLS

arm IP Explorer

arm Virtual Hardware

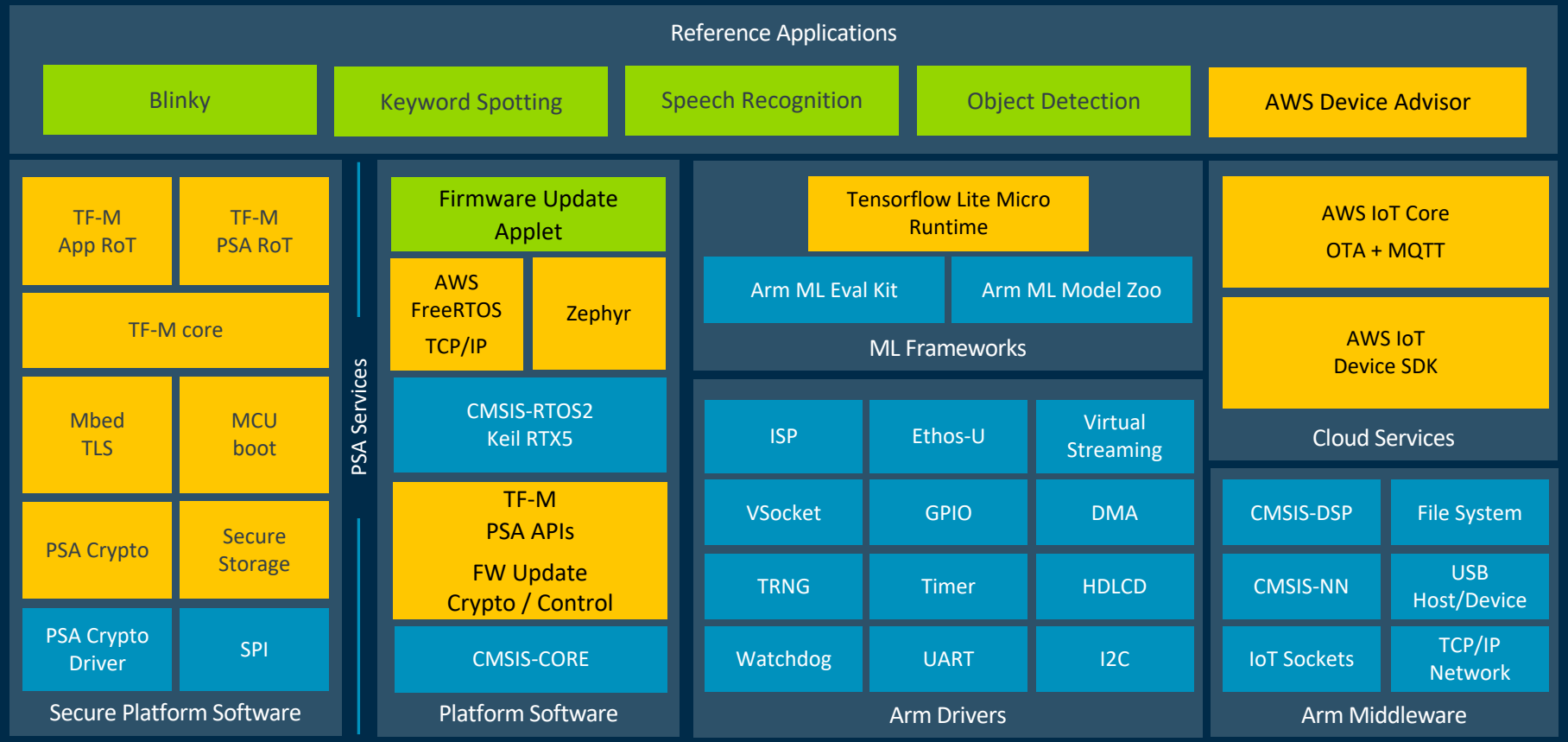
KEIL™ Tools by ARM

arm DS

arm

Enabling Open Source to Accelerate tinyML

https://github.com/FreeRTOS/iot-reference-arm-corstone3xx/releases/tag/v202403.00



Code Development Infrastructure

Unit Tests	Arm MLIA
System Tests	AC6 GCC LLVM
Verification	CMSIS-Toolbox
PSA Secure FW Update	Arm DS Keil MDK v6
PSA Certified Level 2 Ready	Event Logging
AWS Device Qualified	GitHub CI GitLab CI
Standards & Certifications	Build Tools

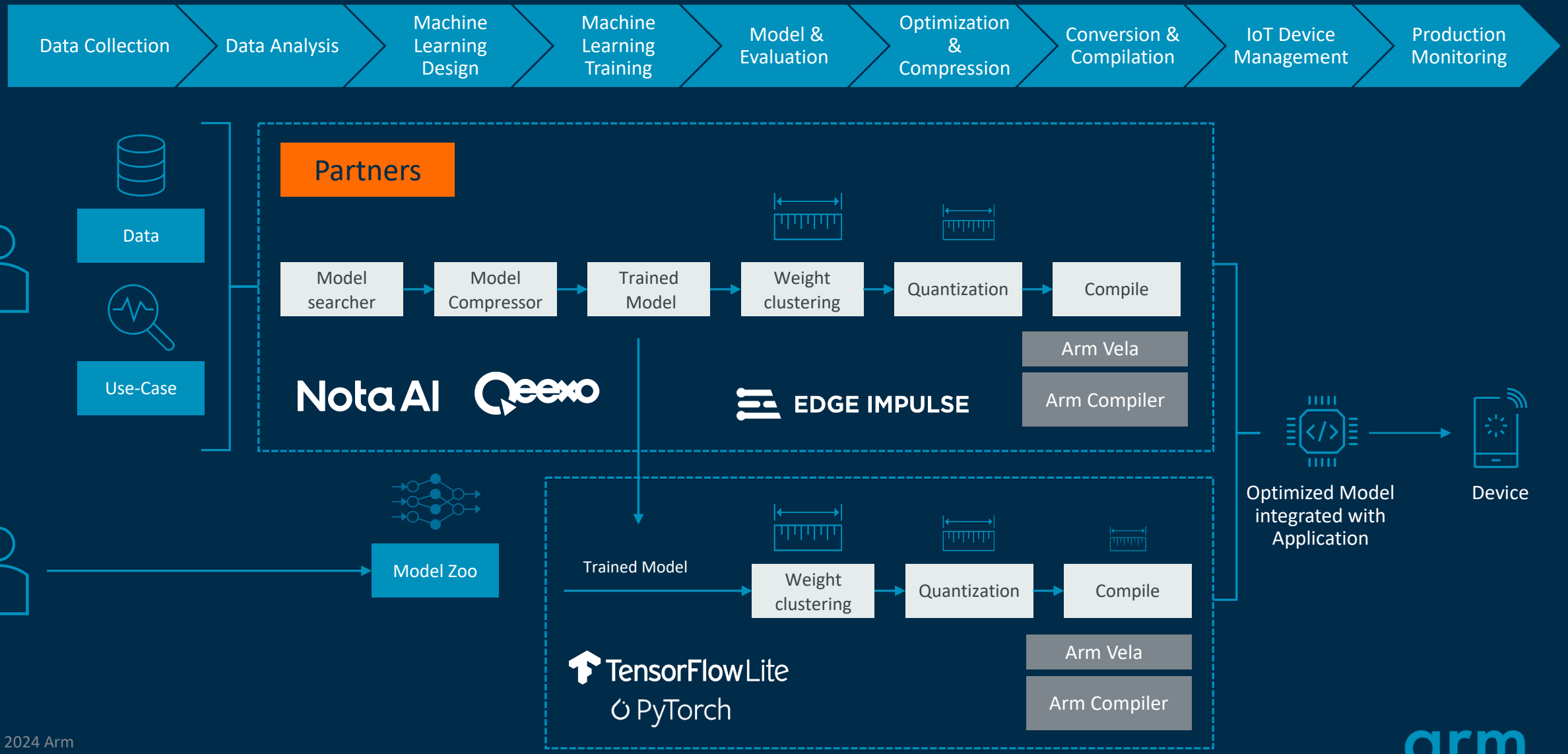
arm Virtual Hardware arm CORSTONE
arm CORTEX-M arm ETHOS

Third party
Arm software
Arm applications

Arm is Enabling tinyML with Three Focus Areas

Ecosystem

Enabling Partners to Have Workflows with Arm Tools



Few specific examples on Ecosystem Enablement for tinyML

Plumerai

ALIF SEMICONDUCTOR

People Identification

seeed studio

Himax

Pose Estimation

AIZIP

RENESAS

Defect Detection

embedUR

synaptics

Image Segmentation

NVIDIA

NXP

People Detection

+ many more

arm

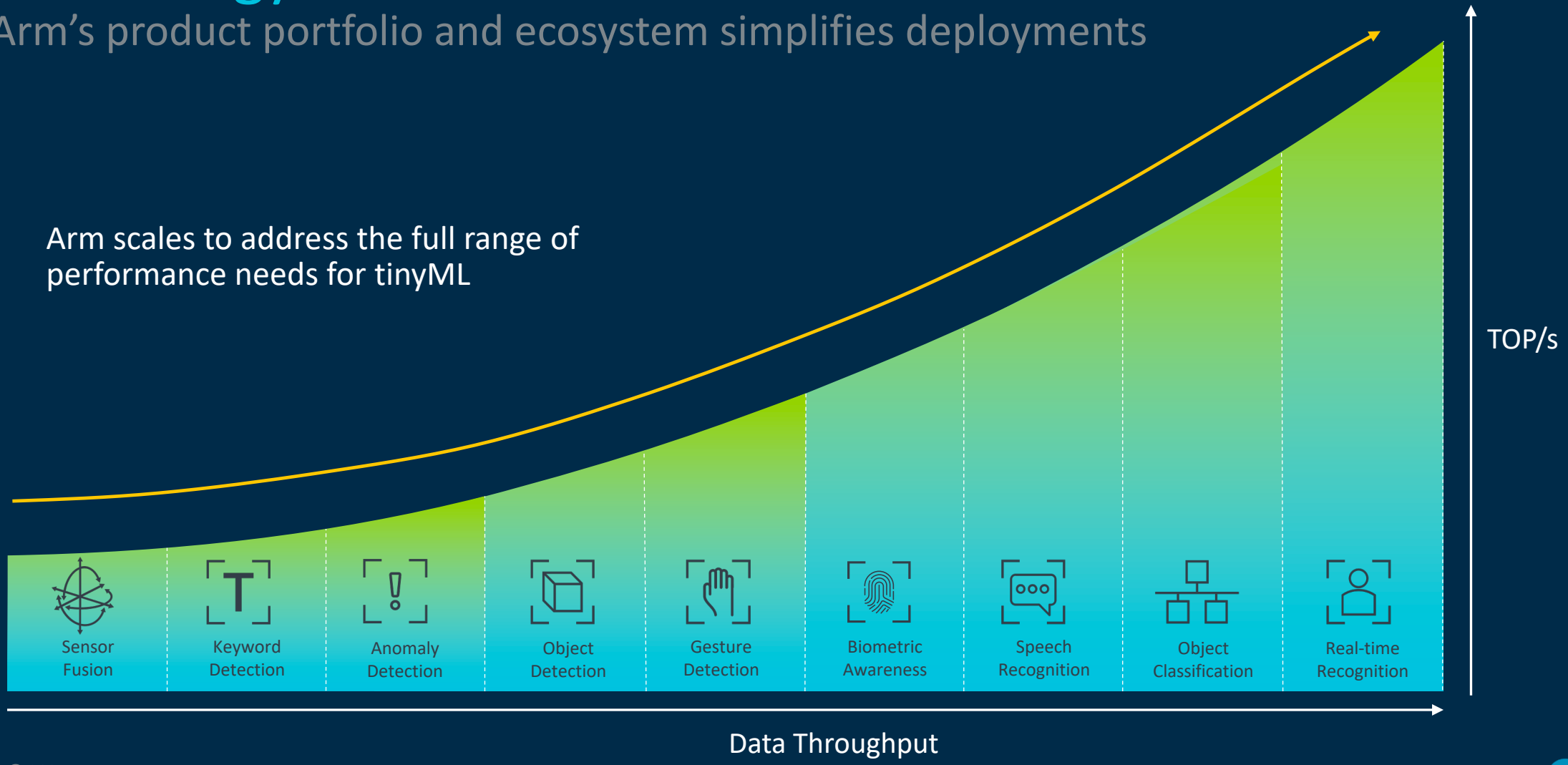
Arm is Enabling a Vibrant Ecosystem for tinyML and Edge AI

Industry leaders are collaborating to realize the potential of tinyML


arm

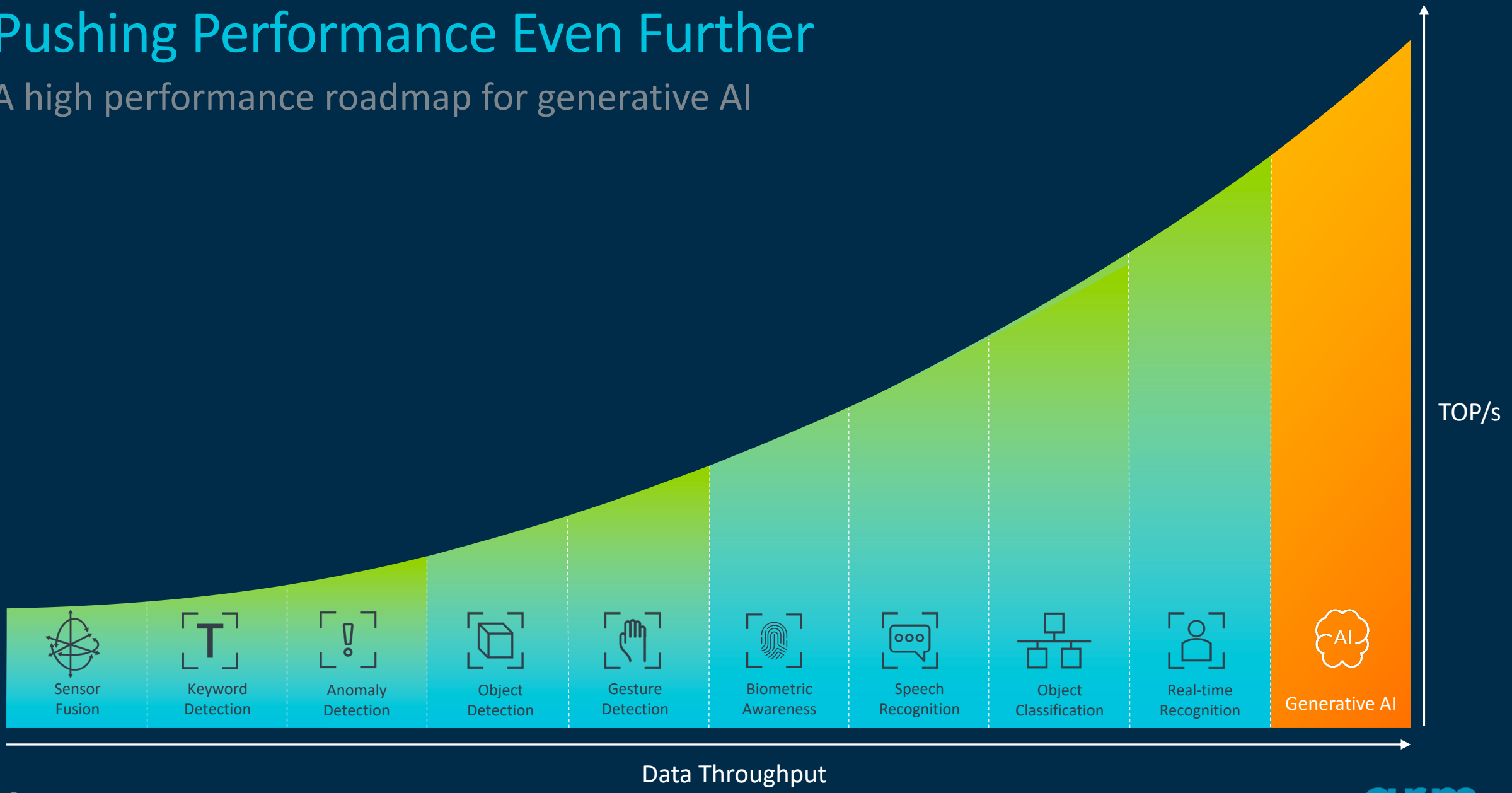
Scaling tinyML and Edge AI with the Broadest Range of Technology

Arm's product portfolio and ecosystem simplifies deployments



Pushing Performance Even Further

A high performance roadmap for generative AI



The Future of tinyML and Edge AI is Being Built on Arm

Enormous opportunity ahead of us to enable tinyML and EdgeAI

Arm is enabling tinyML across multiple vectors - Hardware, Software and Ecosystem

It is important for us to work together to enable a vibrant ecosystem for making tinyML successful

arm

Thank You
Danke
Gracias
Grazie
谢谢
ありがとう
Asante
Merci
감사합니다
धन्यवाद
Kiitos
شكرًا
ধন্যবাদ
תודה
ధన్యవాదములు

Copyright Notice

www.tinyml.org