Acknowledgements & Copyright Notice

Credits: This presentation includes slides and images from various books, research articles, and online sources, including:

- NVIDIA's Edge AI and Robotics Teaching Kit
- Harvard - TinyML Courseware and Machine Learning Systems
- MIT - TinyML and Efficient Deep Learning Computing
- Dive into Deep Learning and Luis Serrano's Academy and his books and videos
- IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things, 2017 by David Hanes
- Introduction to IoT, 2021 by Anandarup Mukherjee; *AI at the Edge* by Daniel Situnayake, Jenny Plunkett - O'Reilly

The author acknowledges and thanks the creators for their contributions. Images and content are used under fair use for educational purposes.

CP 330:

Edge AI

Lecture 1: Introduction

Pandarasamy Arjunan (Samy)
Assistant Professor
Robert Bosch Centre for Cyber Physical Systems
Indian Institute of Science, Bangalore
samy@iisc.ac.in
https://www.samy101.com/edge-ai-25/

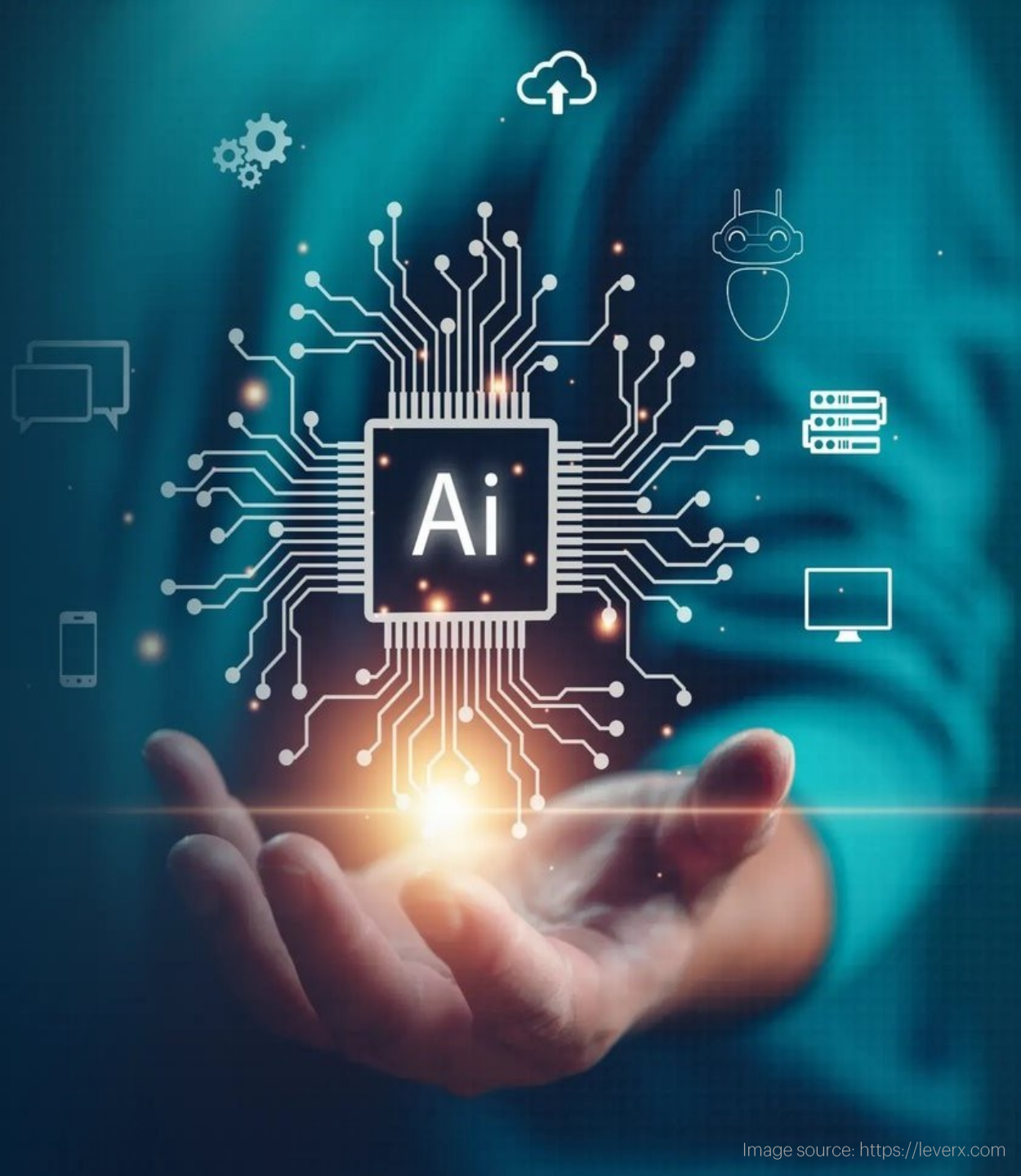


Image source: https://leverx.com

About me

Career (Academic and Industry):

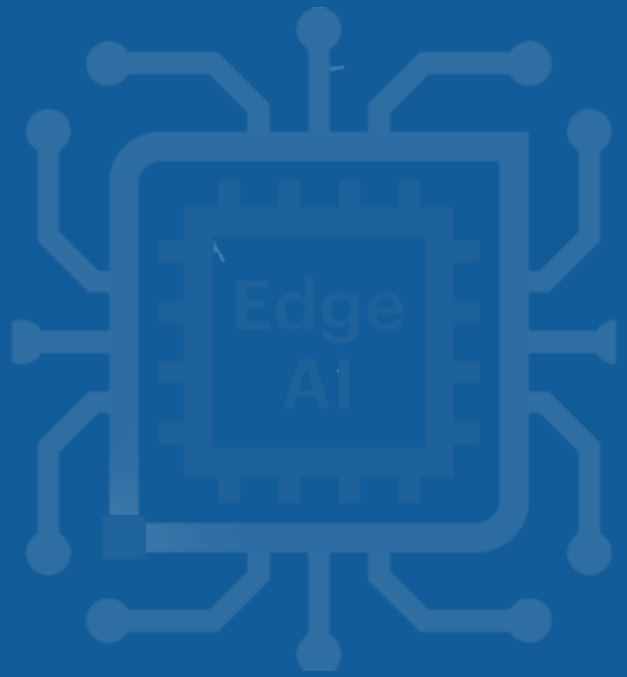
- Assistant Professor, Indian Institute of Science, Bangalore (Since July 2023)
- Postdoc Scholar, BEARS, UC Berkely's Lab in Singapore
- PhD Scholar, IIIT-Delhi
- Software Engineer, IBM India Private Limited, Bangalore
- Software Engineer, HCL Technologies Limited, Chennai

Research Interests and Expertise:

- Internet of Things and Cyber Physical Systems
- AI/ML and Edge AI
- Application Domains: Smart Buildings, Smart Cities, Smart Energy, Smart Mobility, and Smart Agriculture

More details at https://www.samy101.com/

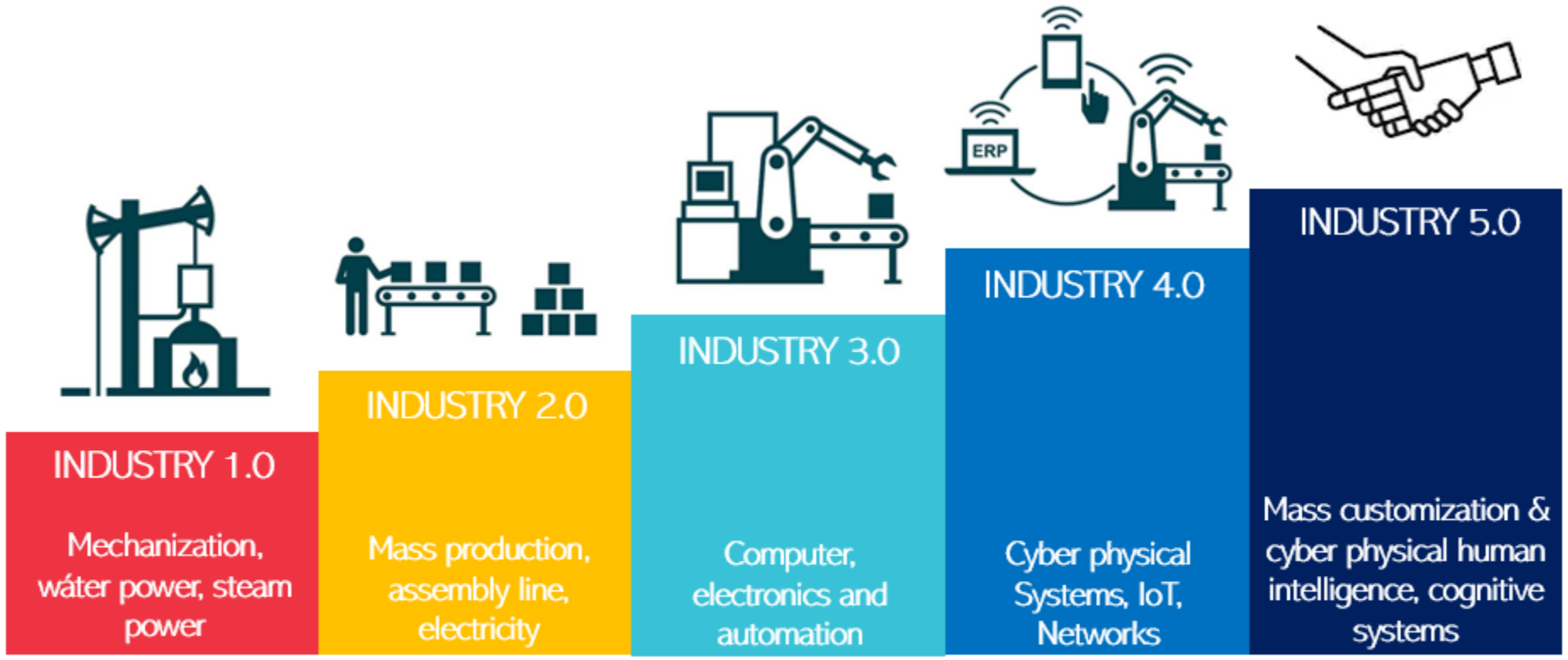
Course Logistics



Class timings

- Lecture: Tuesdays 11:30 AM to 1:00PM
- Lab session: Thursdays 11:30 AM to 1:00PM

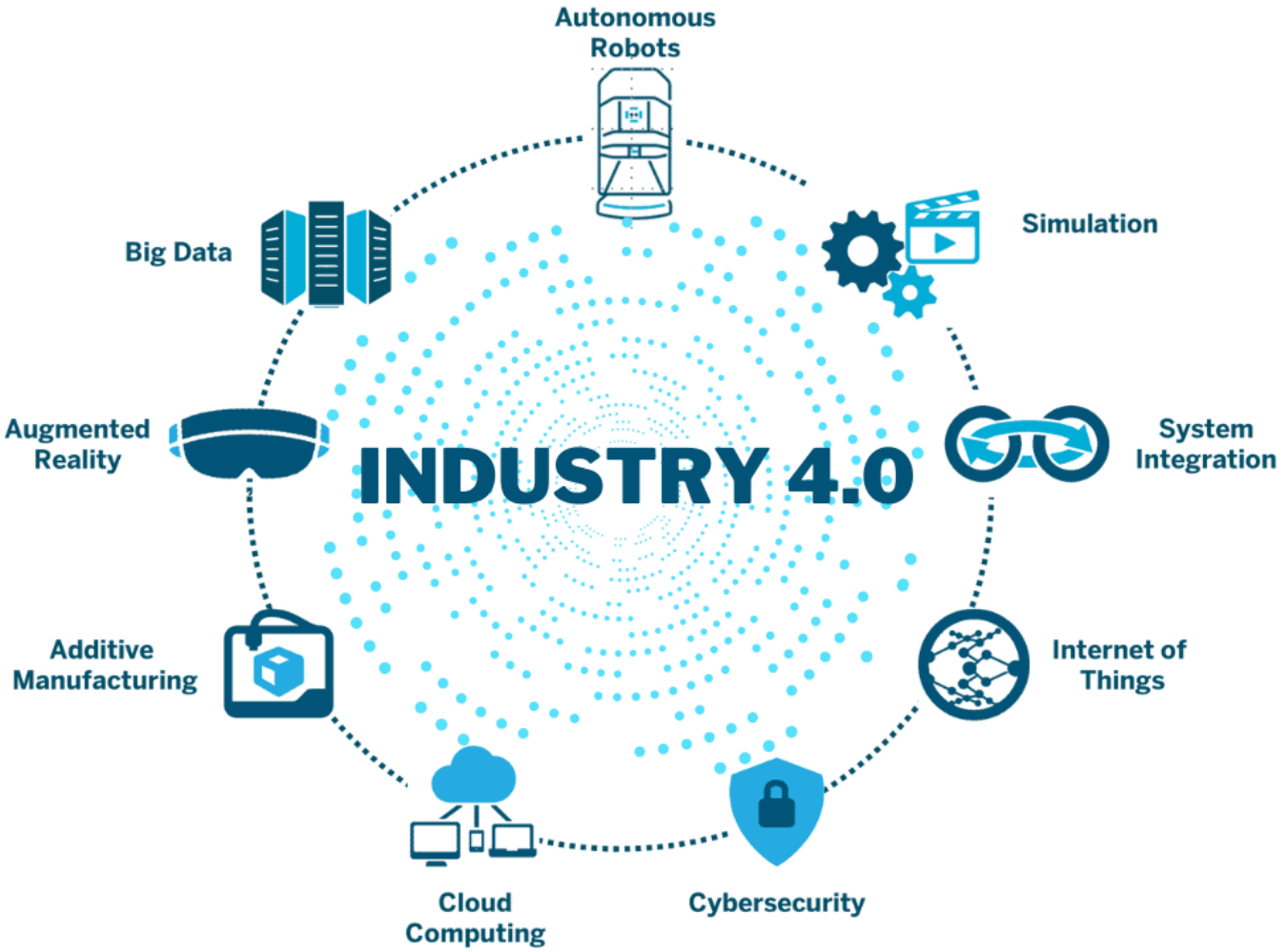
Industrial Revolution



*Years are estimates

Source: Internet

Industry 4.0



Source: https://www.calsoft.com/what-is-industry-4-0/

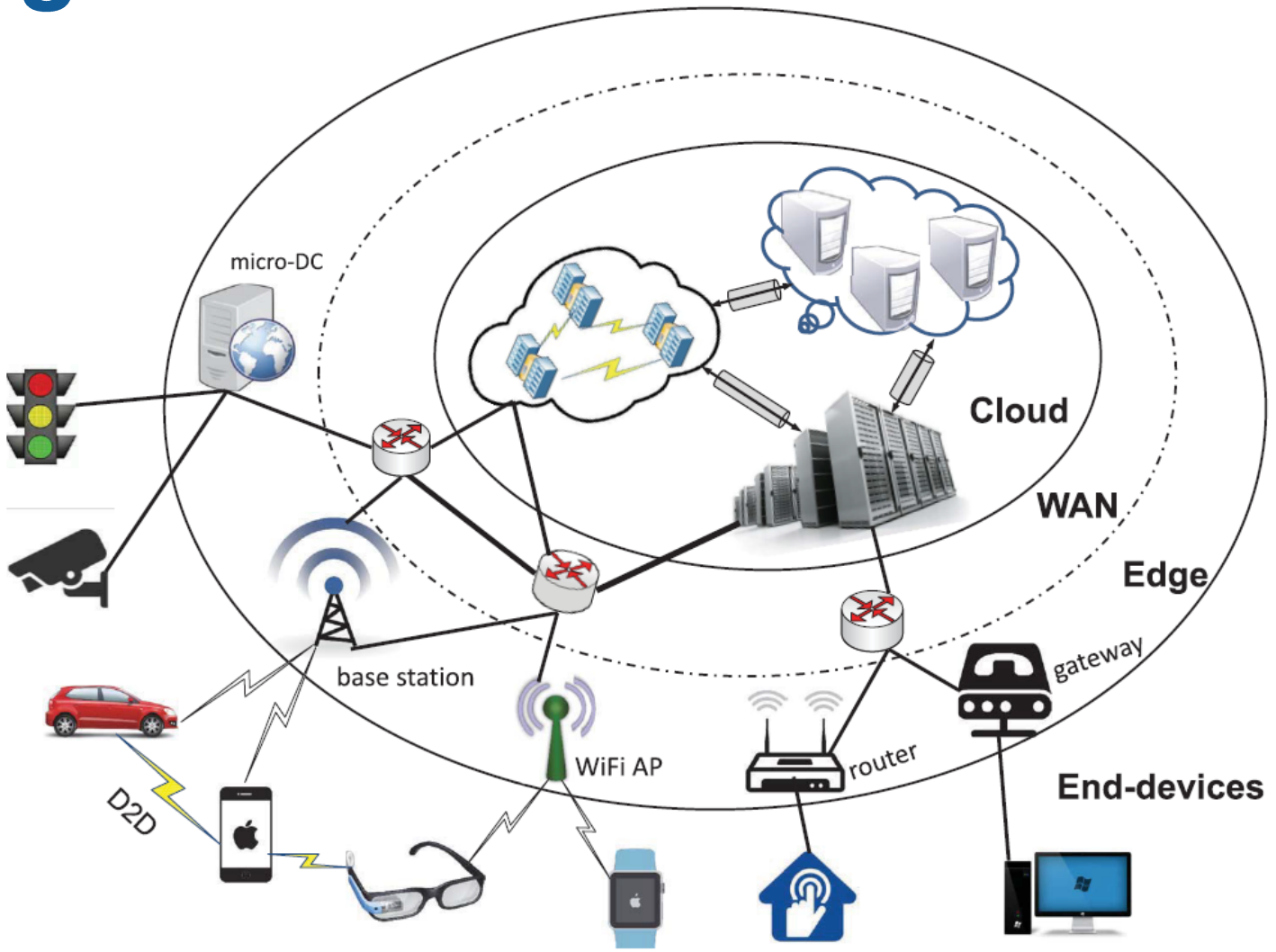
Industry 5.0

Source: https://www.momenta.one/

What is Internet of Things (IoT)?

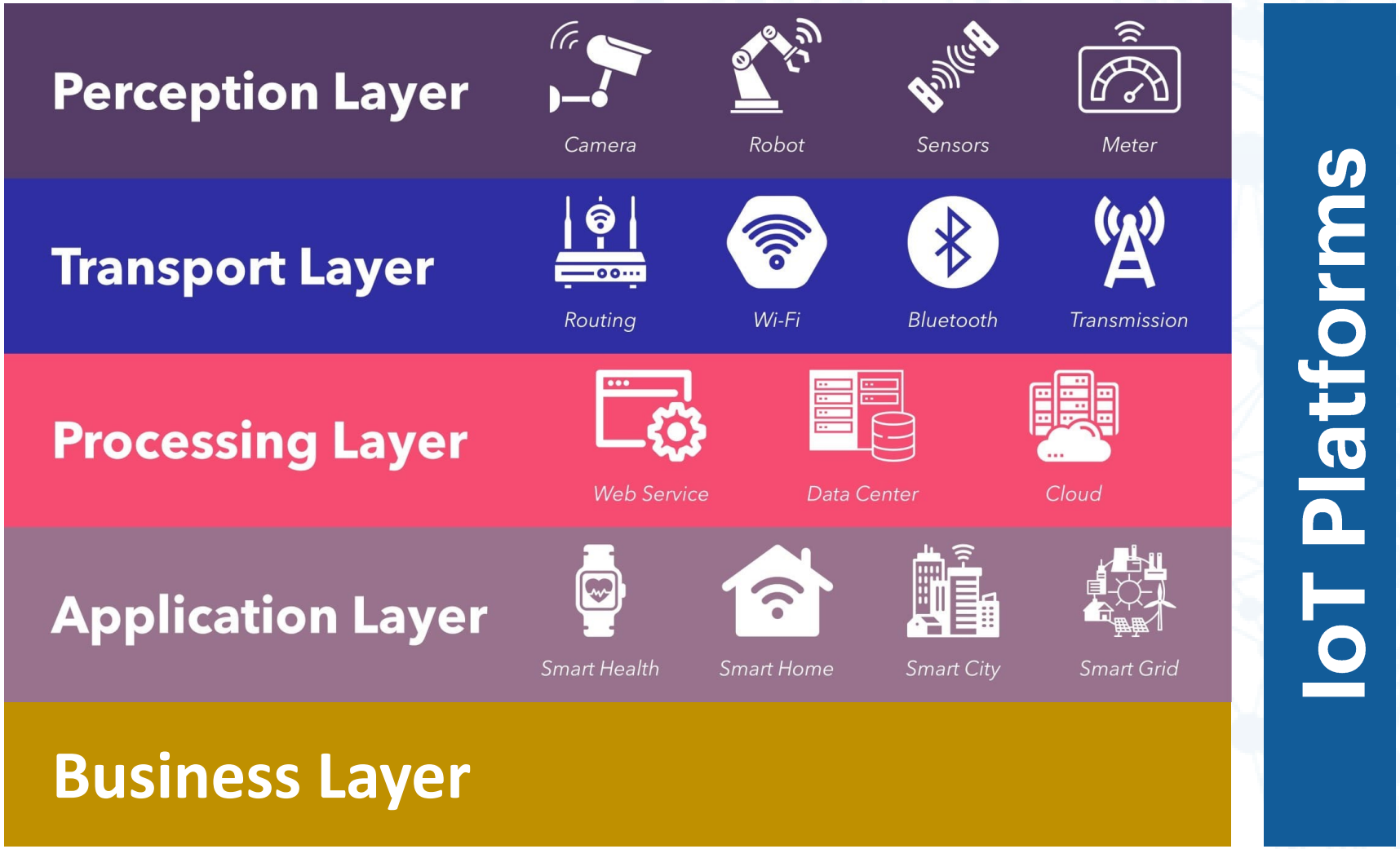
IoT is a network of physical objects that are embedded with sensors, software, and network connectivity, allowing them to collect, exchange, and act on data over the internet.

IoT and Edge Devices



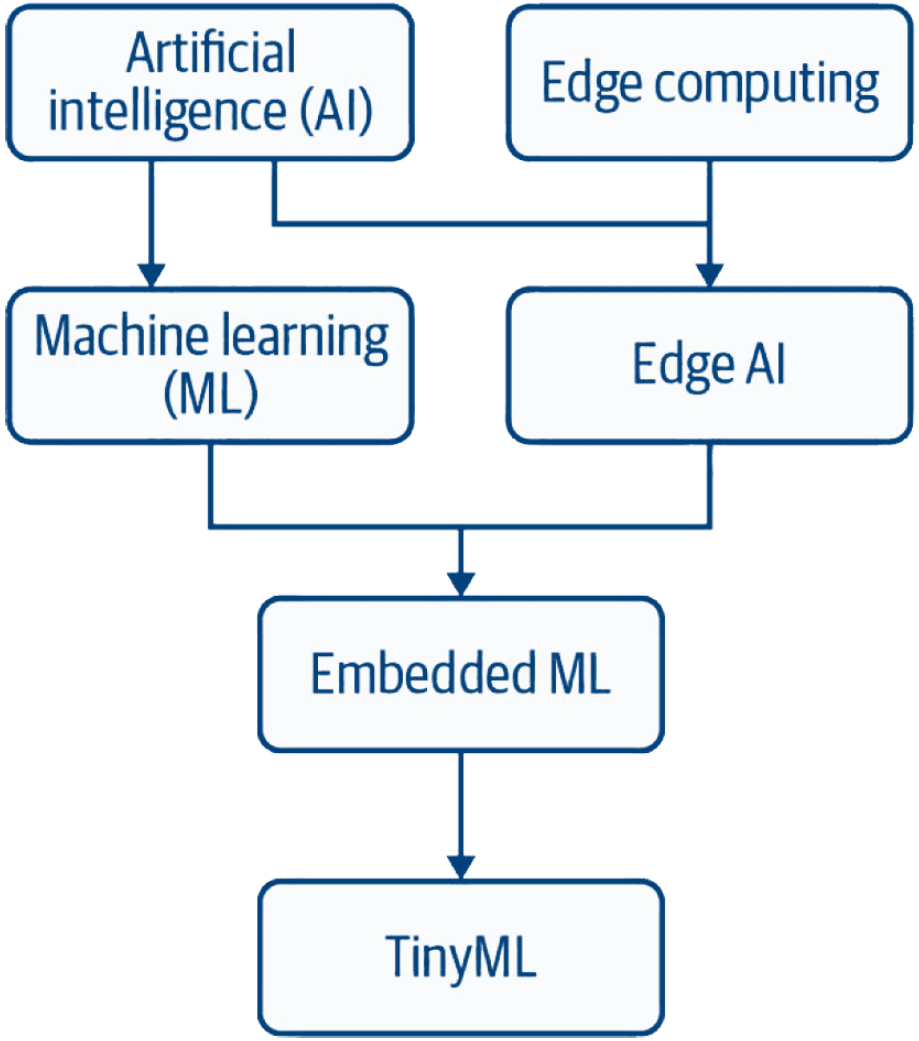
Source: Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing (Zhou et al. 2019)

IoT Architecture and Layers



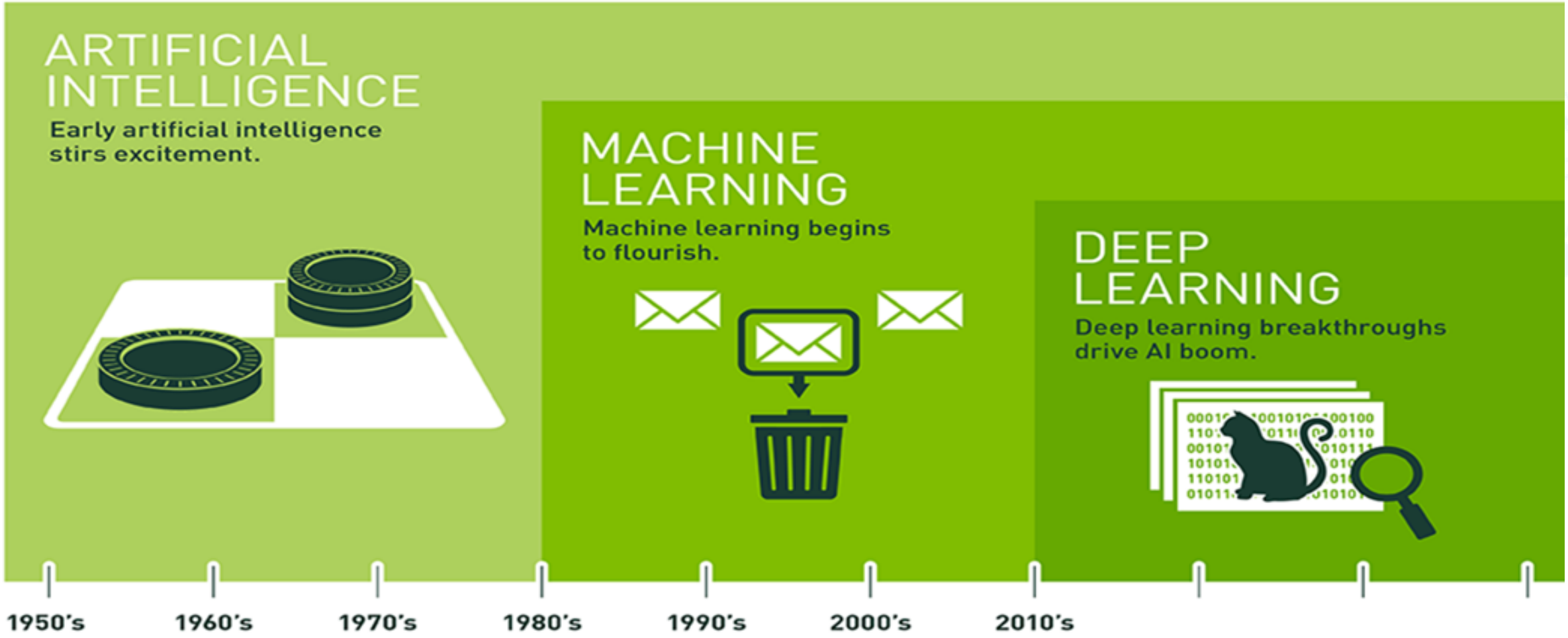
Source: https://dgtlinfra.com/

The BIG picture of Edge AI



Source: [3]

Evolution of AI



Source: [1]

Why AI at the Edge?

Bandwidth

1 billion+ cameras worldwide
10's of petabytes per day

Latency

Safety-critical services
Realtime decisions

Privacy

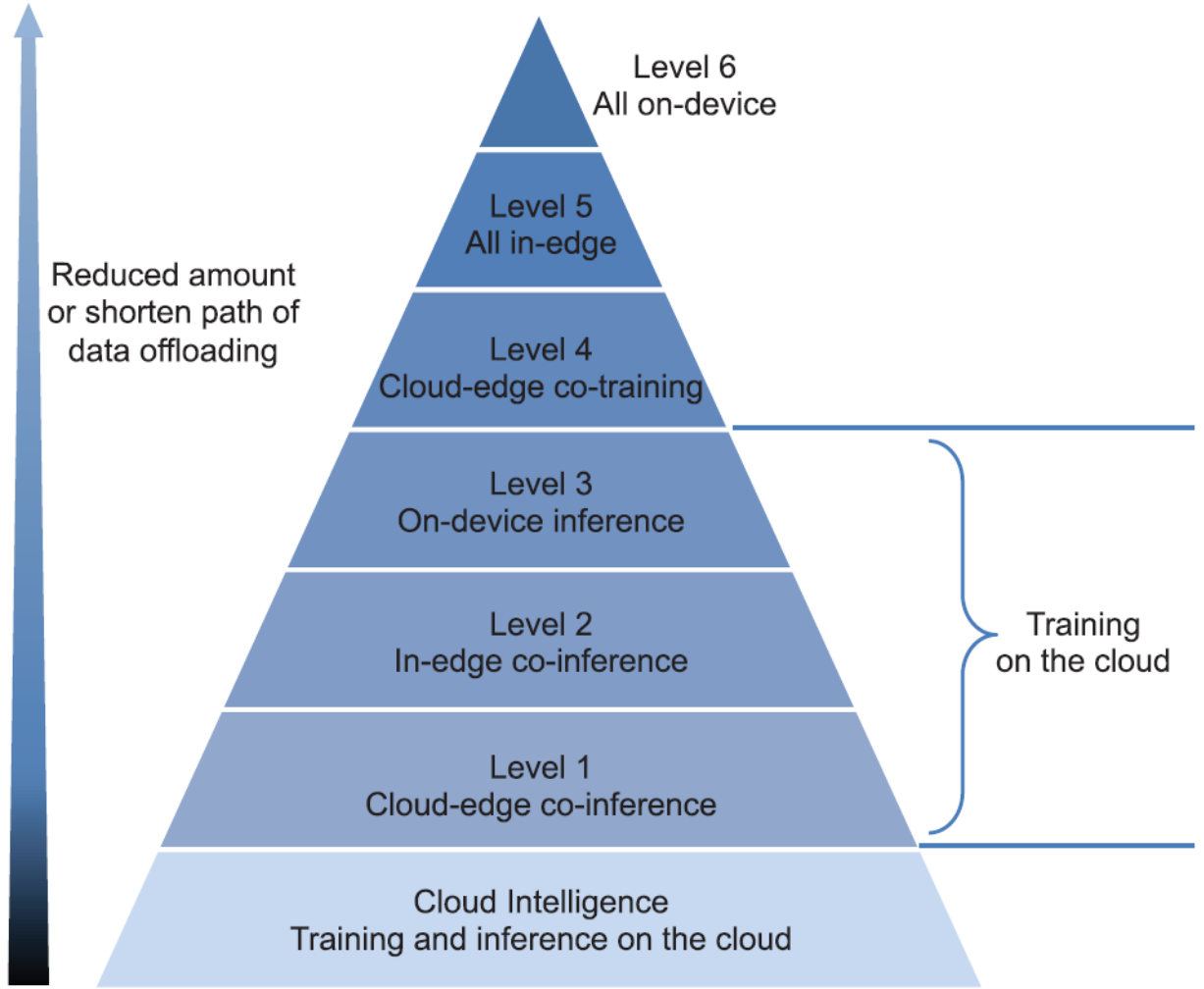
Data Redaction + Confidentiality
Private cloud or on-premises storage

Connectivity

50% of populated world < 8mbps
Bulk of uninhabited world no 3G+

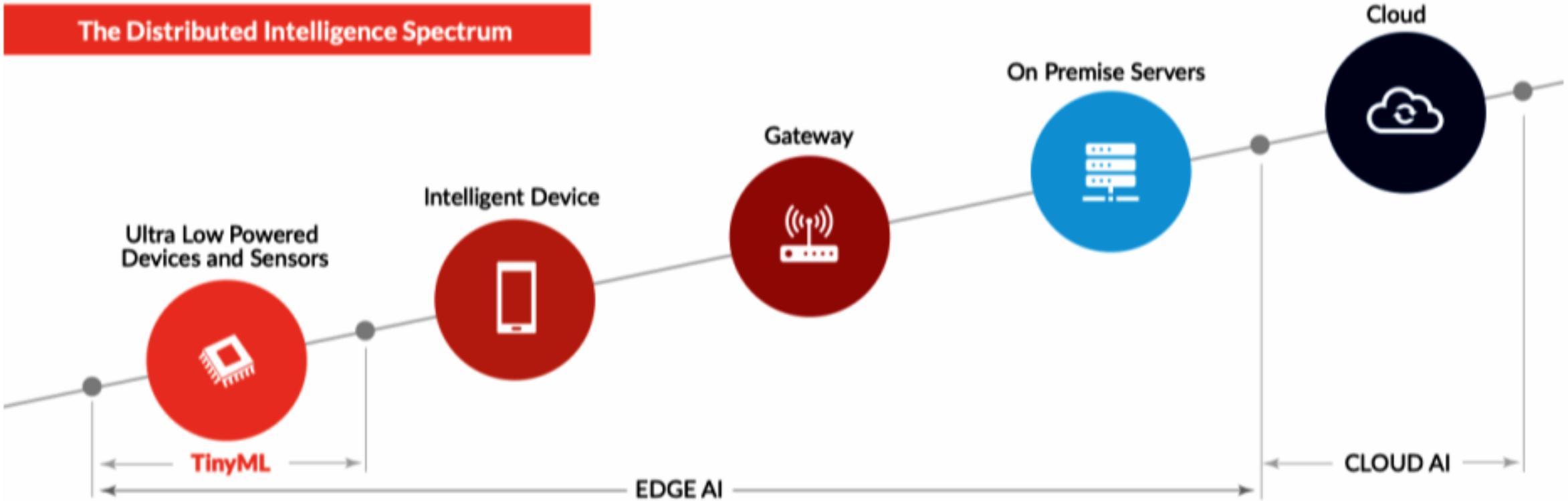
Source: [1]

Six-level ratings of Edge Intelligence



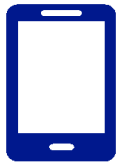
Source: Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing (Zhou et al. 2019)

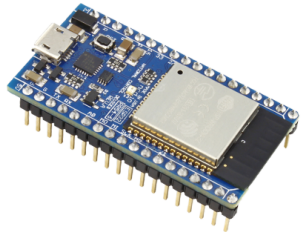
Cloud vs. Edge vs. Mobile vs. Tiny AI



Source: ABI Research: TinyML

Cloud vs. Mobile vs. Tiny AI





Cloud ML (~2006)

- DNN
- Large Models (16-32GB)
- X Millions of Paremeters
- TFLOPs
- Focus on Accuracy
- Hardware: GPU, TPU, FPGA
- AlexNet, Inception, ResNet, VGGnet
- Data: Storage, Sharing (1%)

Mobile ML (~2016)

- CNN (light)
- Constrained resources: memory 8GB RAM, Application size limitation
- GFLOPs
- Focus on Accuracy-efficiency trade-off
- Hardware: SoC, NPU
- AlexNet, Inception, ResNet, VGGnet
- MobileNet_v1, ShuffleNet, SqueezeNet
- Data: Pics, Audio, Clicks, GPS(5%)

TinyML (2019)

- CNN-micro
- Severly Constrained resources
- ~100KB RAM
- MCU with HW accelerators
- Sensors: CMOS Cameras, IR Cameras, Audio, IMU, Temp, Chemical, Accelerometers
- Data: Sensing the physical world (95%)

Source: TinyML for Ultra-Low Power AI and Large Scale IoT Deployments: A Systematic Review (Schizas et al. 2022)

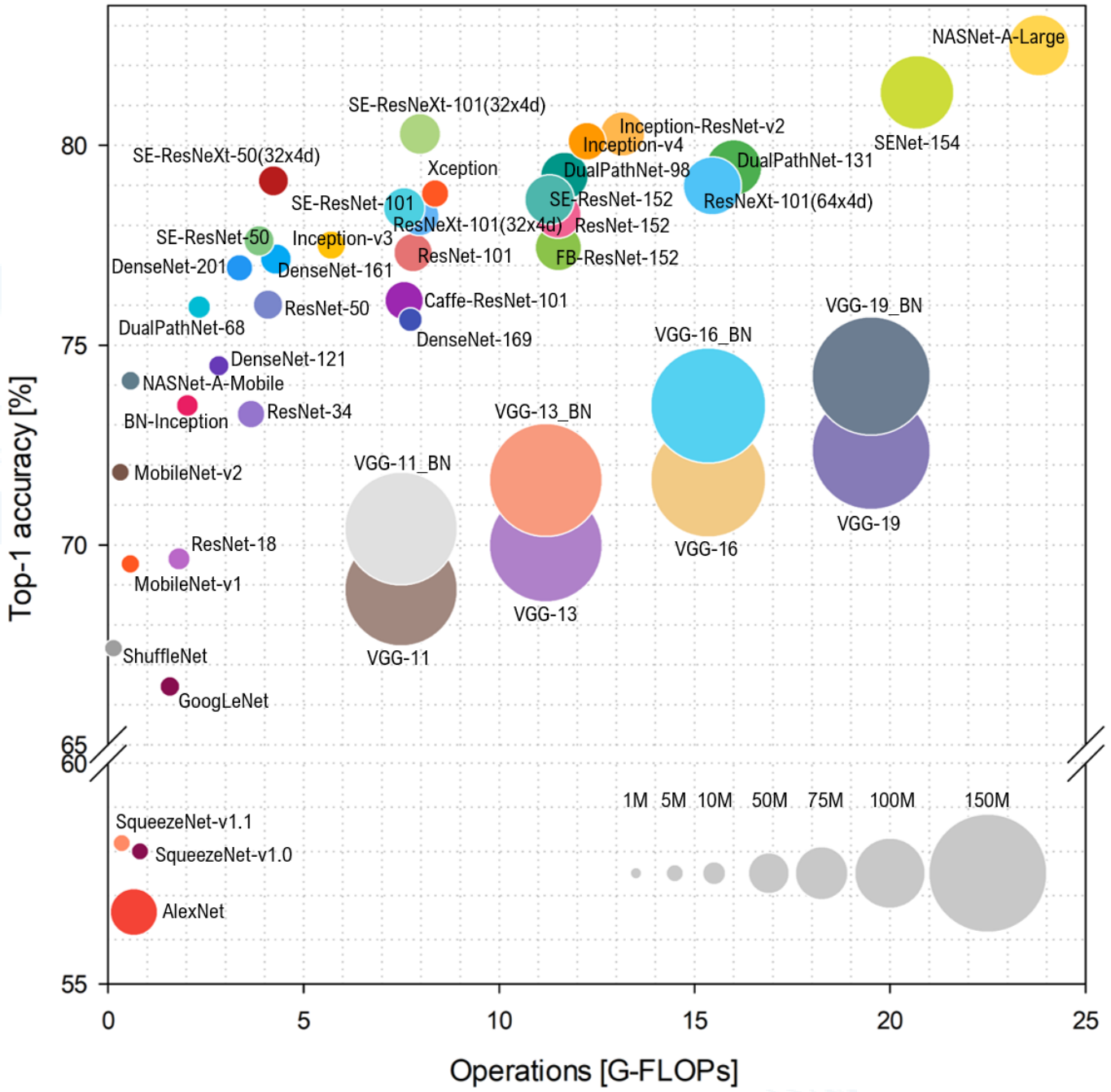
Cloud vs. Mobile vs. Tiny AI

	Cloud AI (NVIDIA V100)		Mobile AI (iPhone 11)		Tiny AI (STM32F746)		ResNet-50	MobileNetV2	MobileNetV2 (int8)
Memory	16 GB	4× →	4 GB	3100× →	**320 kB**	← gap →	7.2 MB	6.8 MB	1.7 MB
Storage	TB~PB	1000× →	>64 GB	64000× →	**1 MB**	← gap →	102MB	13.6 MB	3.4 MB

Source: Tiny Machine Learning: Progress and Future (Lin et al. 2023)

AI Model Size

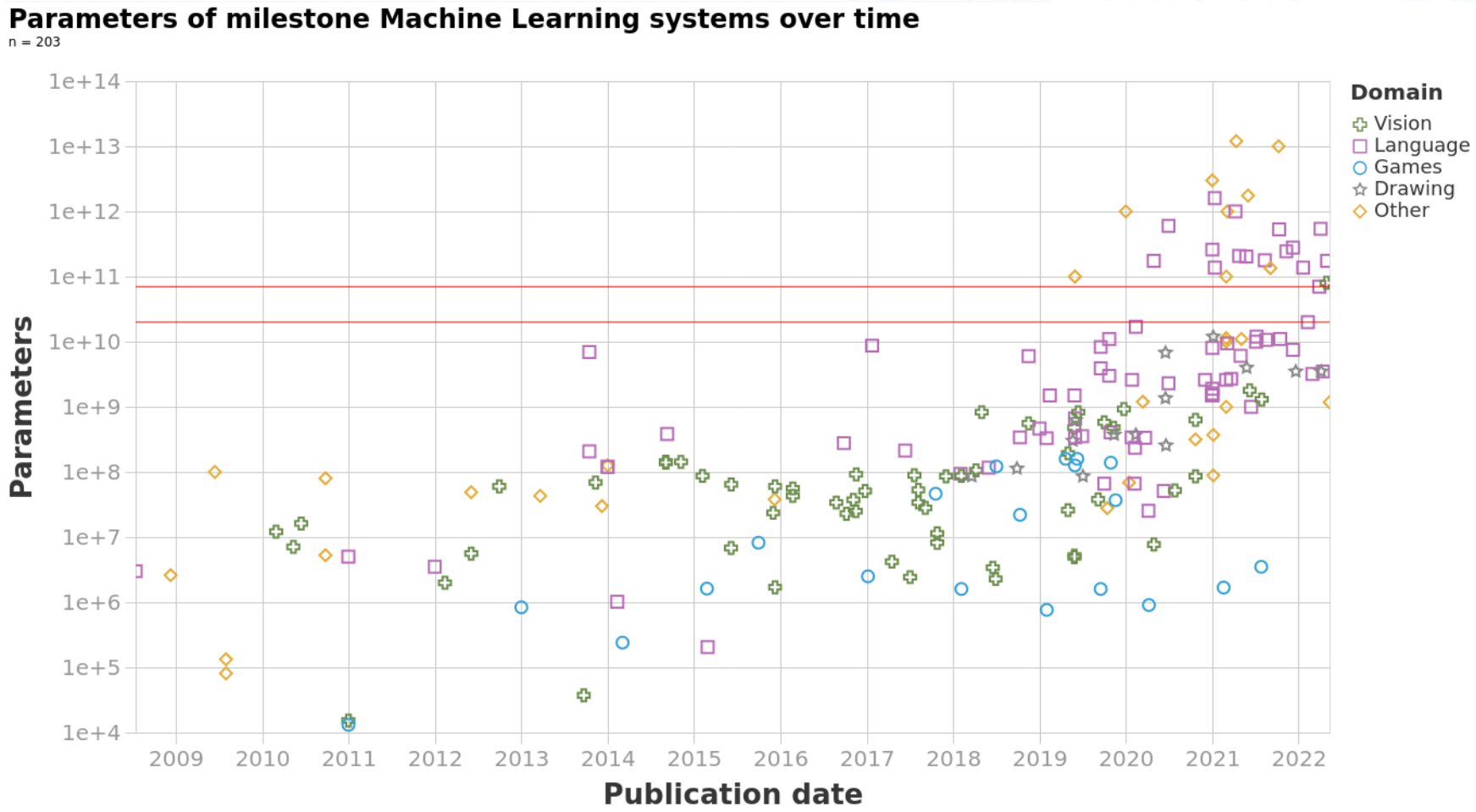
Top-1 accuracy vs. computational complexity on ImageNet



Source: https://github.com/CeLuigi/models-comparison.pytorch

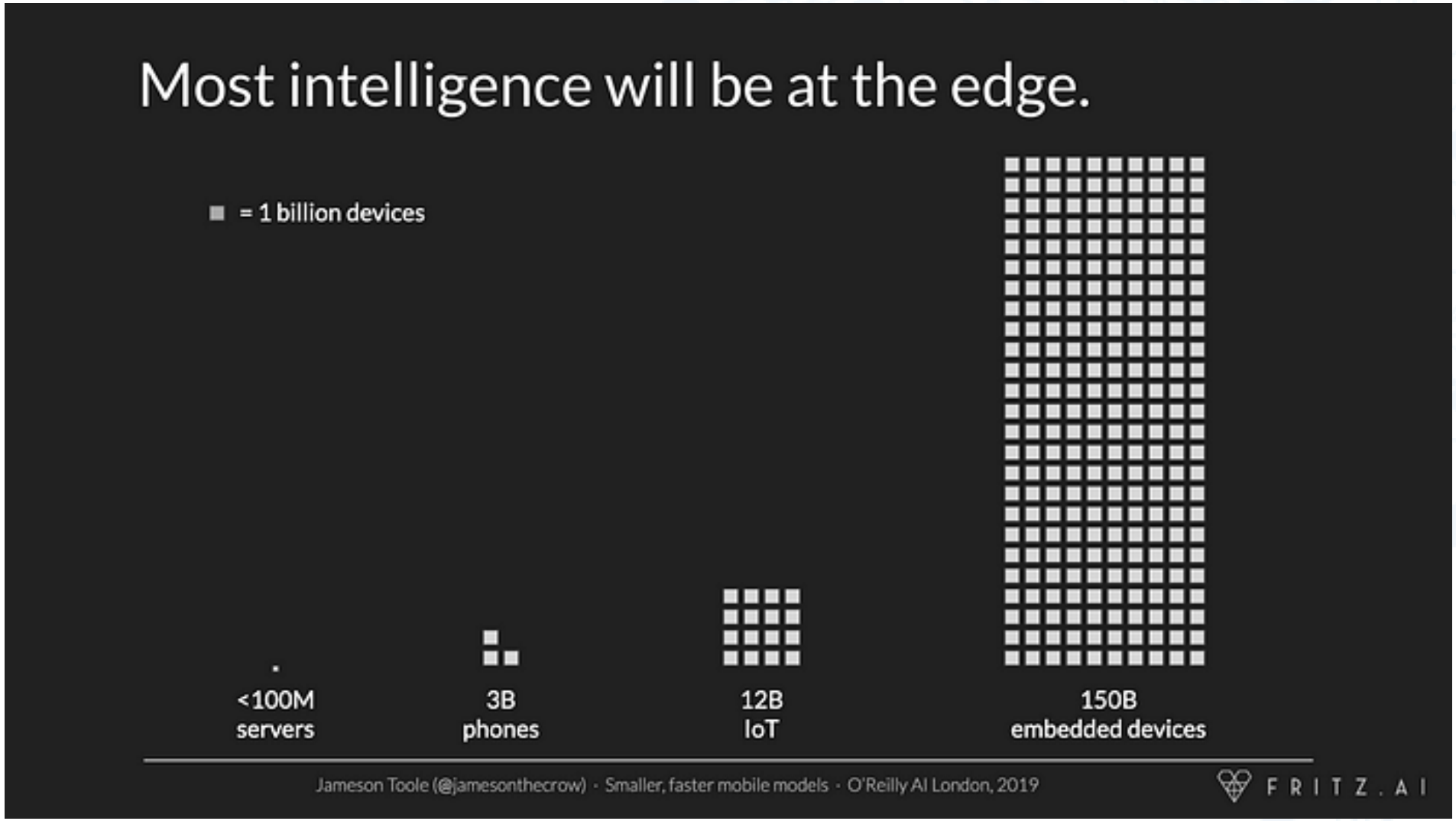
AI Model Size

ML Model size has grown 10x faster than before since 2018

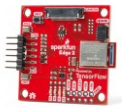


Source: https://epoch.ai/ and Machine Learning Model Sizes and the Parameter Gap (Villalobos et al., 2022)

Deep learning has a size problem!!!

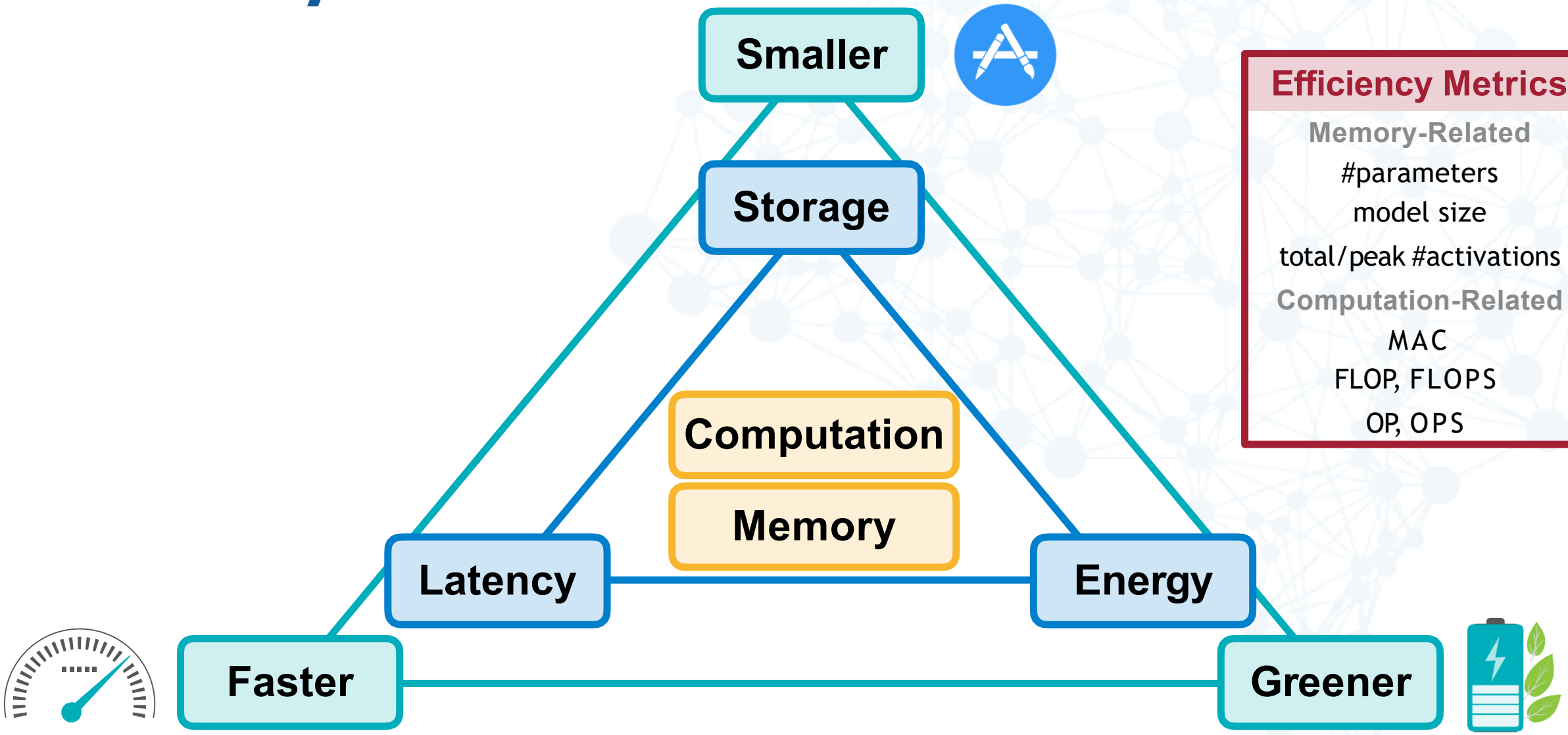


Embedded Systems

	Board	MCU / ASIC	Clock	Memory	Sensors	Radio
	Himax WE-I Plus EVB	HX6537-A 32-bit EM9D DSP	400 MHz	2MB flash 2MB RAM	Accelerometer, Mic, Camera	None
	Arduino Nano 33 BLE Sense	32-bit nRF52840	64 MHz	1MB flash 256kB RAM	Mic, IMU, Temp, Humidity, Gesture, Pressure, Proximity, Brightness, Color	BLE
	SparkFun Edge 2	32-bit ArtemisV1	48 MHz	1MB flash 384kB RAM	Accelerometer, Mic, Camera	BLE
	Espressif EYE	32-bit ESP32-DOWD	240 MHz	4MB flash 520kB RAM	Mic, Camera	WiFi, BLE

Source: [4]

Efficiency of Neural Networks



Efficiency Metrics

Memory-Related

#parameters

model size

total/peak #activations

Computation-Related

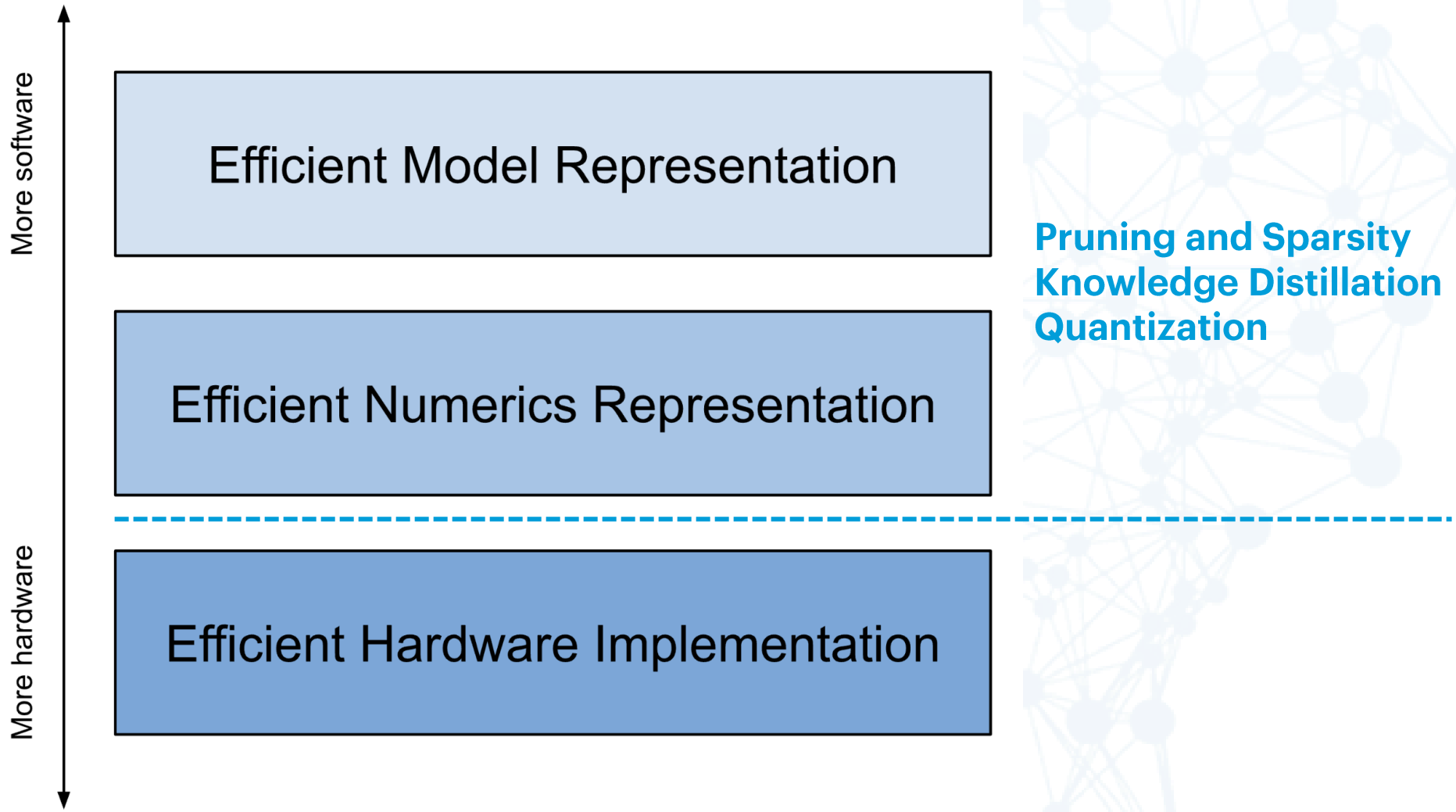
MAC

FLOP, FLOPS

OP, OPS

Image source: 1

Efficient AI models



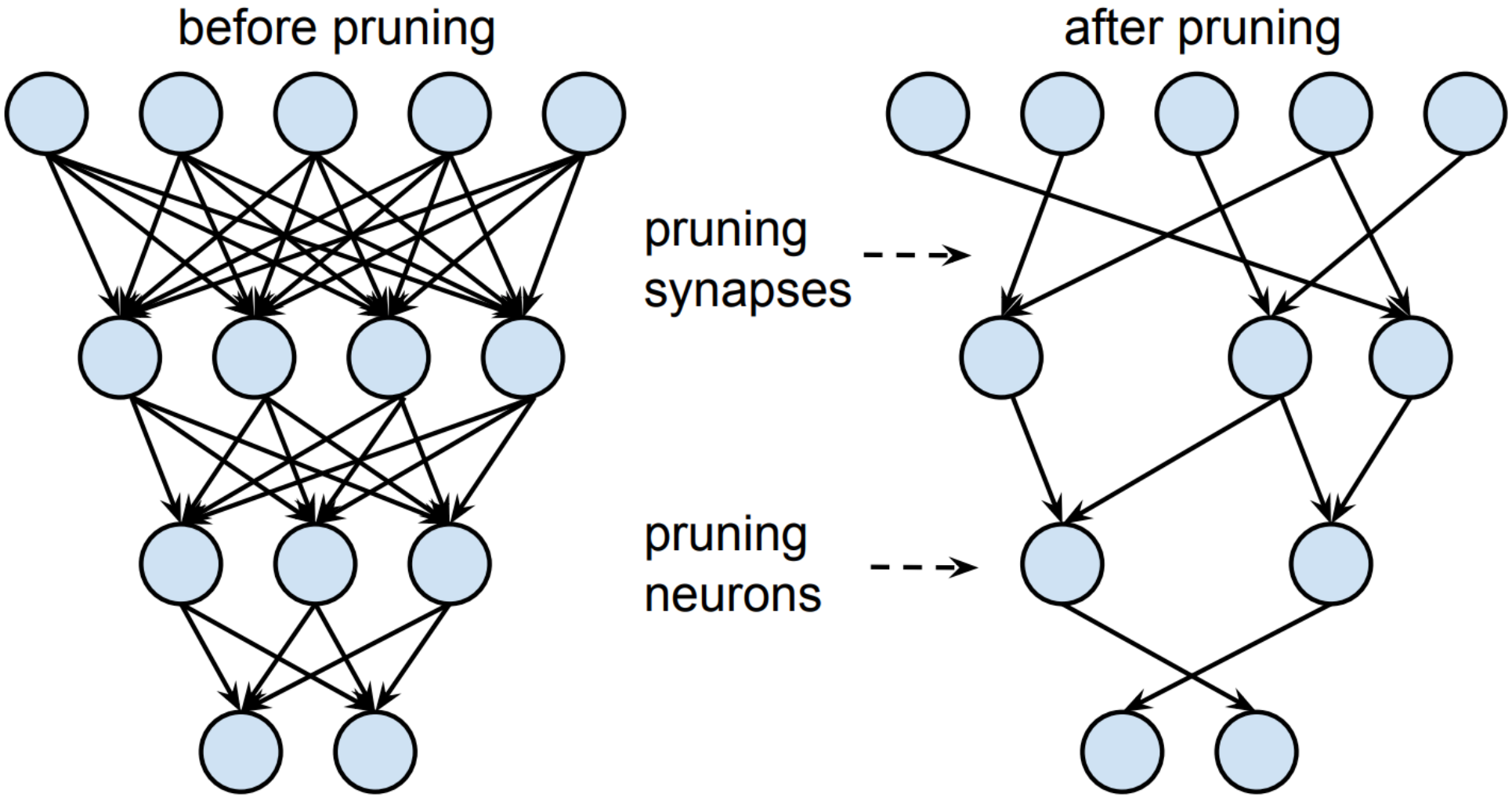
Source: [2]

Efficient AI models

Make AI models run faster and efficiently on low-power hardware

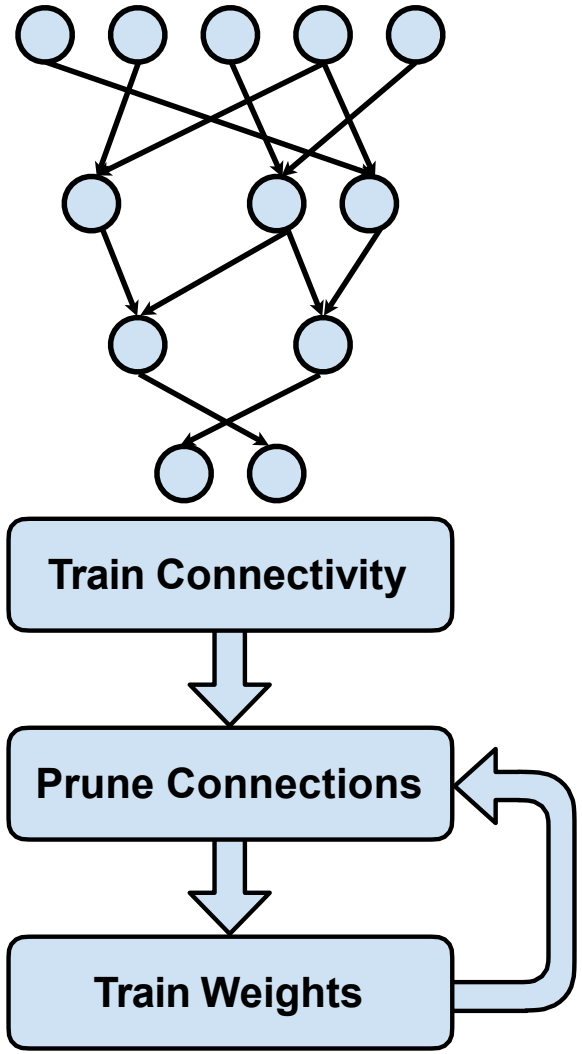
Source: [4]

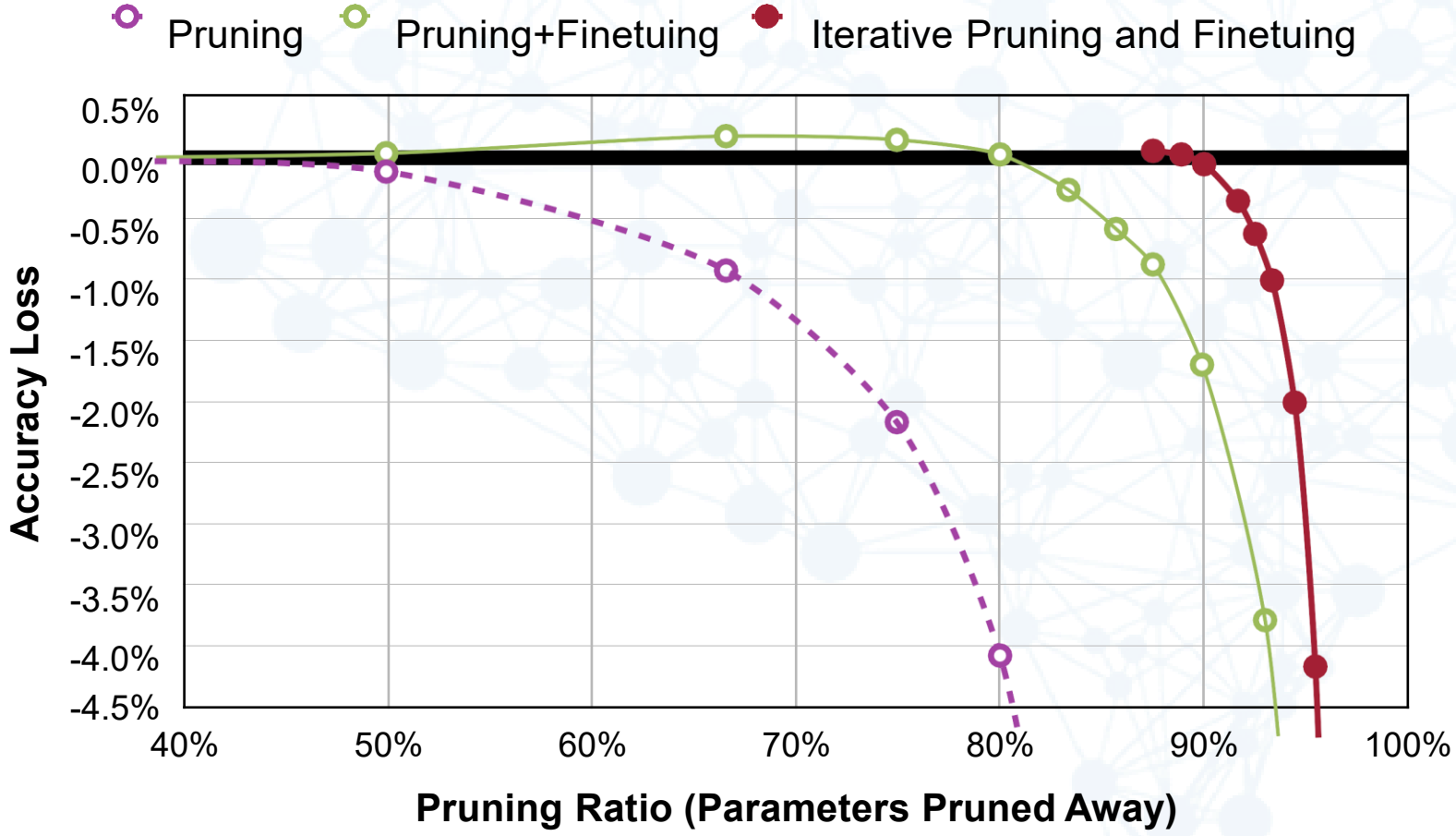
Pruning in Neural Networks



Source: Learning both Weights and Connections for Efficient Neural Networks (Han et al. 2015)

Pruning in Neural Networks





Source: Learning both Weights and Connections for Efficient Neural Networks (Han et al. 2015)

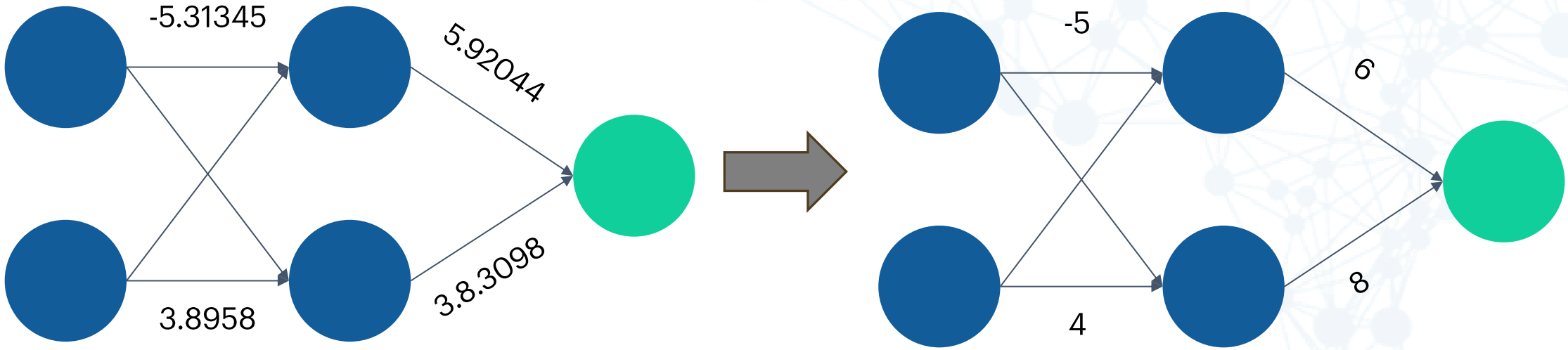
Pruning in Neural Networks

Neural Network	#Parameters			MACs
	Before Pruning	After Pruning	Reduction	Reduction
AlexNet	61 M	6.7 M	9 ×	3 ×
VGG-16	138 M	10.3 M	12 ×	5 ×
GoogleNet	7 M	2.0 M	3.5 ×	5 ×
ResNet50	26 M	7.47 M	3.4 ×	6.3 ×
SqueezeNet	1 M	0.38 M	3.2 ×	3.5 ×

Source: Efficient Methods and Hardware for Deep Learning (Han 2017)

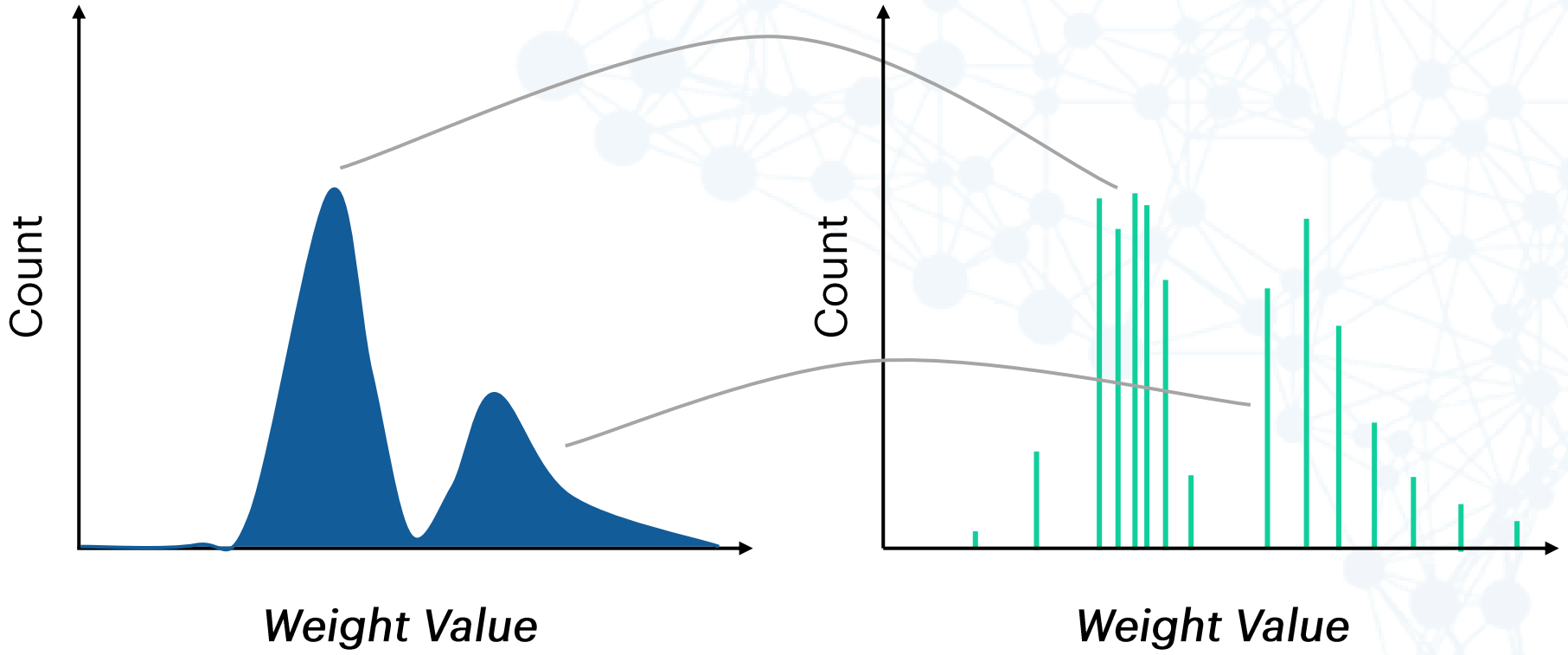
Quantization

Quantization is an optimization that works by reducing the precision of the numbers used to represent a model's parameters, which by default are 32-bit floating point numbers. This results in a smaller model size, better portability and faster computation.



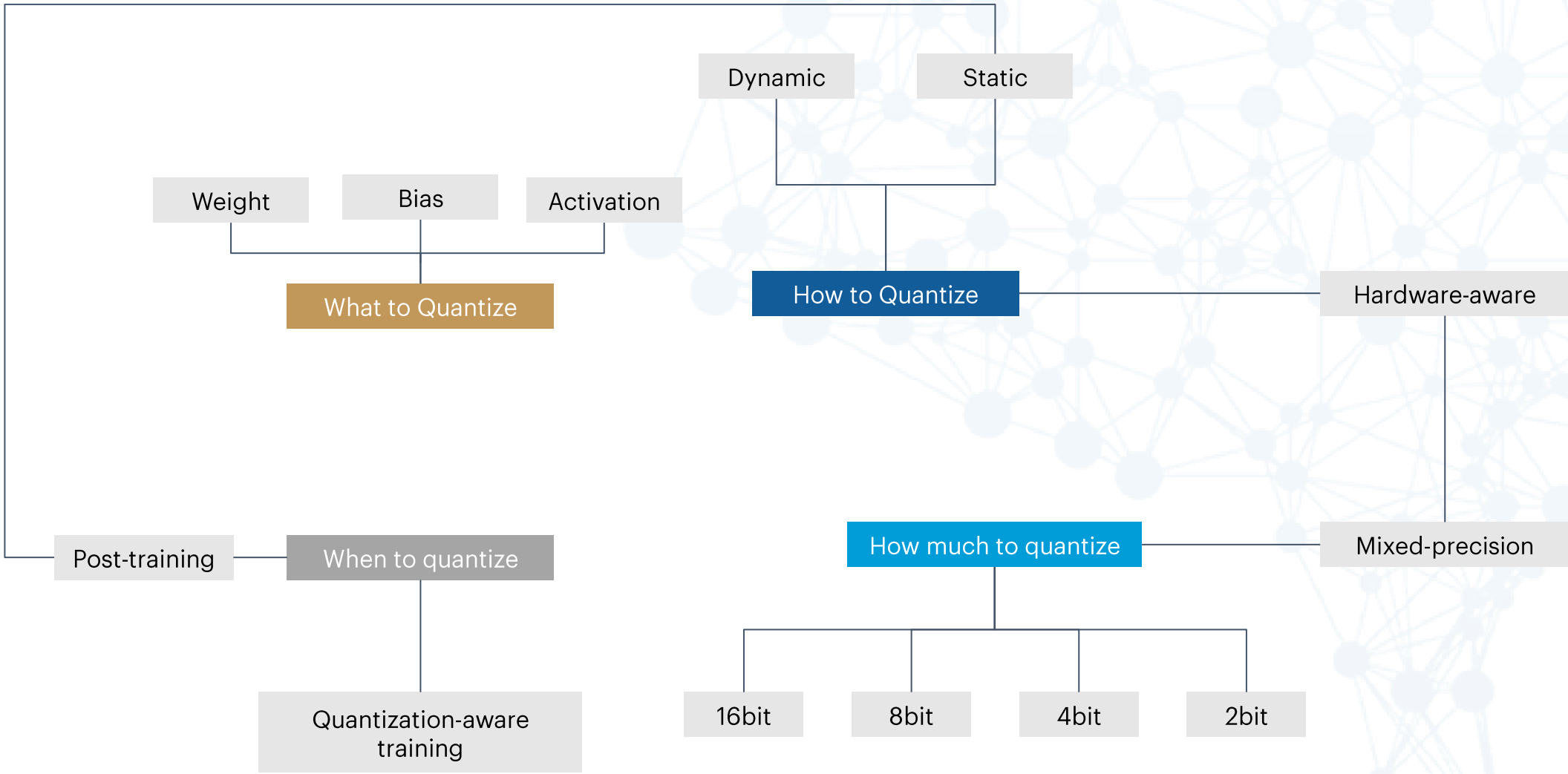
Source: [4]

Reduce Precision (Discretize)



Source: [4]

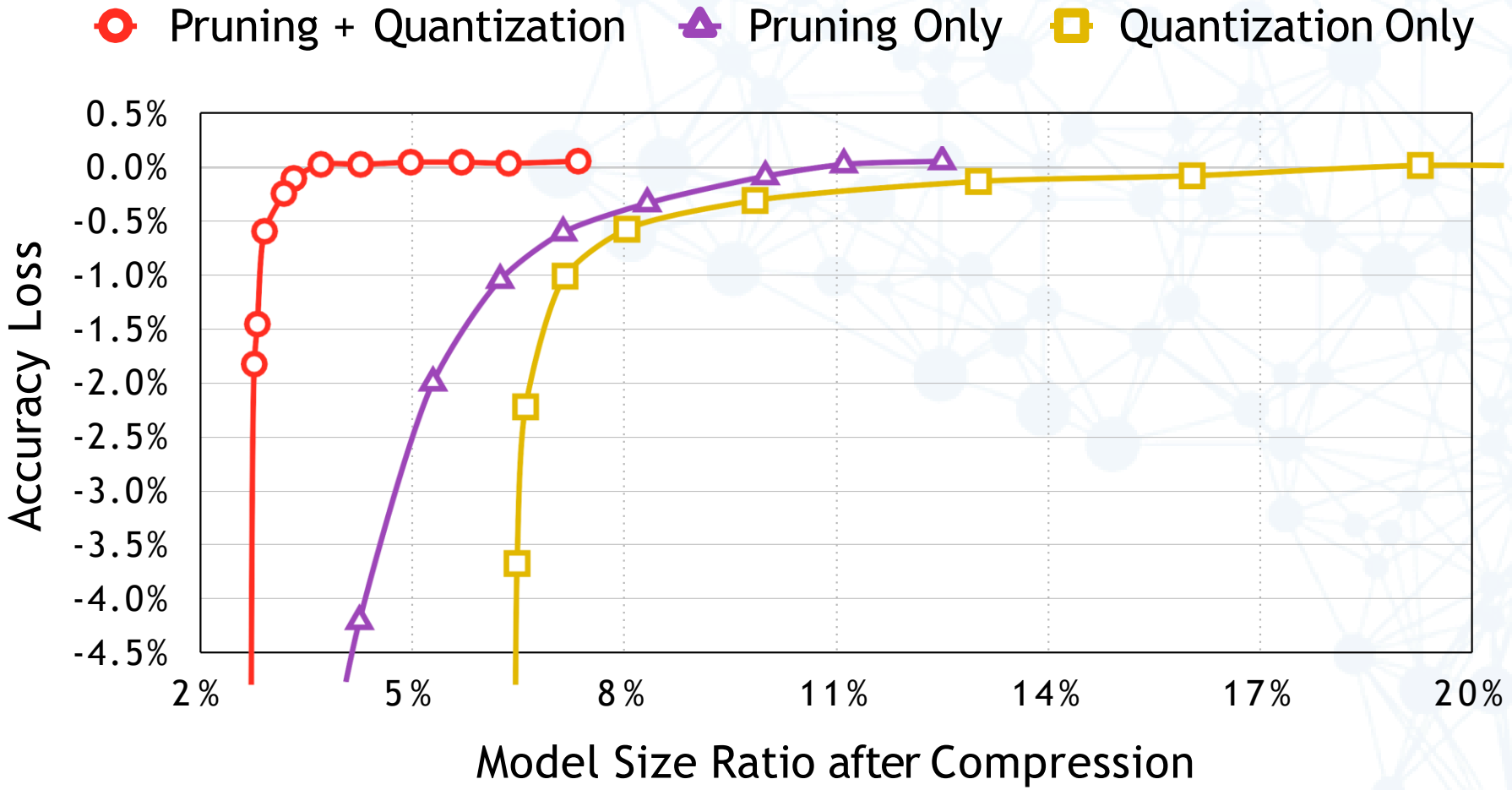
Quantization



Source: [4]

Clustering-based Weight Quantization

Accuracy vs. compression rate for AlexNet on ImageNet dataset



Source: Deep Compression (Han et al., ICLR 2016)

Google AI Edge

Deploy AI across mobile, web, and embedded applications

On device
Reduce latency. Work offline. Keep your data local & private.

Cross-platform
Run the same model across Android, iOS, web, and embedded.

Multi-framework
Compatible with JAX, Keras, PyTorch, and TensorFlow models.

Full AI edge stack
Flexible frameworks, turnkey solutions, hardware accelerators

Google's LiteRT

LiteRT (short for Lite Runtime), formerly known as TensorFlow Lite, is Google's high-performance runtime for on-device AI. You can find ready-to-run LiteRT models for a wide range of ML/AI tasks, or convert and run TensorFlow, PyTorch, and JAX models to the TFLite format using the AI Edge conversion and optimization tools.

Key features

- **Optimized for on-device machine learning**: LiteRT addresses five key ODML constraints: latency (there's no round-trip to a server), privacy (no personal data leaves the device), connectivity (internet connectivity is not required), size (reduced model and binary size) and power consumption (efficient inference and a lack of network connections).
- **Multi-platform support**: Compatible with Android and iOS devices, embedded Linux, and microcontrollers.
- **Multi-framework model options**: AI Edge provides tools to convert models from TensorFlow, PyTorch, and JAX models into the FlatBuffers format (`.tflite`), enabling you to use a wide range of state-of-the-art models on LiteRT. You also have access to model optimization tools that can handle quantization and metadata.
- **Diverse language support**: Includes SDKs for Java/Kotlin, Swift, Objective-C, C++, and Python.
- **High performance**: Hardware acceleration through specialized delegates like GPU and iOS Core ML.

Source: https://ai.google.dev/edge/litert

TensorFlow Model Optimization Toolkit

The *TensorFlow Model Optimization Toolkit* is a suite of tools for optimizing ML models for deployment and execution. Among many uses, the toolkit supports techniques used to:

- Reduce latency and inference cost for cloud and edge devices (e.g. mobile, IoT).
- Deploy models to edge devices with restrictions on processing, memory, power-consumption, network usage, and model storage space.
- Enable execution on and optimize for existing hardware or new special purpose accelerators.

Choose the model and optimization tool depending on your task:

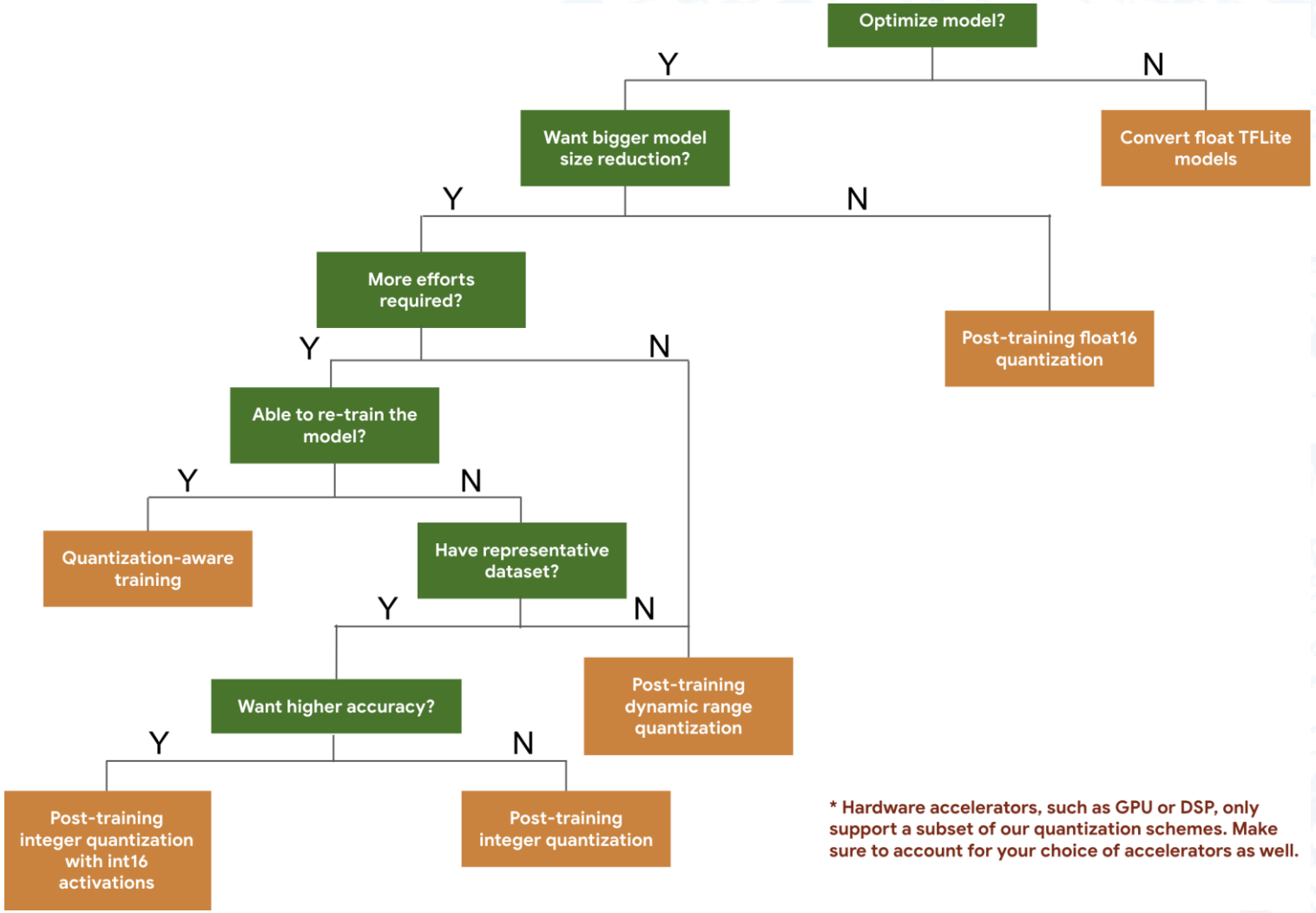
Improve performance with off-the-shelf models

In many cases, pre-optimized models can improve the efficiency of your application.

```
import tensorflow as tf
import tensorflow_model_optimization as tfmot
import tf_keras as keras

model = keras.Sequential([...])

pruning_schedule = tfmot.sparsity.keras.PolynomialDecay(
                      initial_sparsity=0.0, final_sparsity=0
                      begin_step=2000, end_step=4000)

model_for_pruning = tfmot.sparsity.keras.prune_low_magnitude
    model, pruning_schedule=pruning_schedule)
...

model_for_pruning.fit(...)
```

Source: https://www.tensorflow.org/model_optimization

TensorFlow Model Optimization Toolkit



* Hardware accelerators, such as GPU or DSP, only support a subset of our quantization schemes. Make sure to account for your choice of accelerators as well.

Source: https://ai.google.dev/edge/litert/models/model_optimization

TensorFlow Model Optimization Toolkit

Quantized models are 4x smaller and 1.4x faster!

Model	Top-1 Accuracy (Original)	Top-1 Accuracy (Post Training Quantized)	Top-1 Accuracy (Quantization Aware Training)	Latency (Original) (ms)	Latency (Post Training Quantized) (ms)	Latency (Quantization Aware Training) (ms)	Size (Original) (MB)	Size (Optimized) (MB)
Mobilenet-v1-1-224	0.709	0.657	0.70	124	112	64	16.9	4.3
Mobilenet-v2-1-224	0.719	0.637	0.709	89	98	54	14	3.6
Inception_v3	0.78	0.772	0.775	1130	845	543	95.7	23.9
Resnet_v2_101	0.770	0.768	N/A	3973	2868	N/A	178.3	44.9

Source: https://ai.google.dev/edge/litert/models/model_optimization

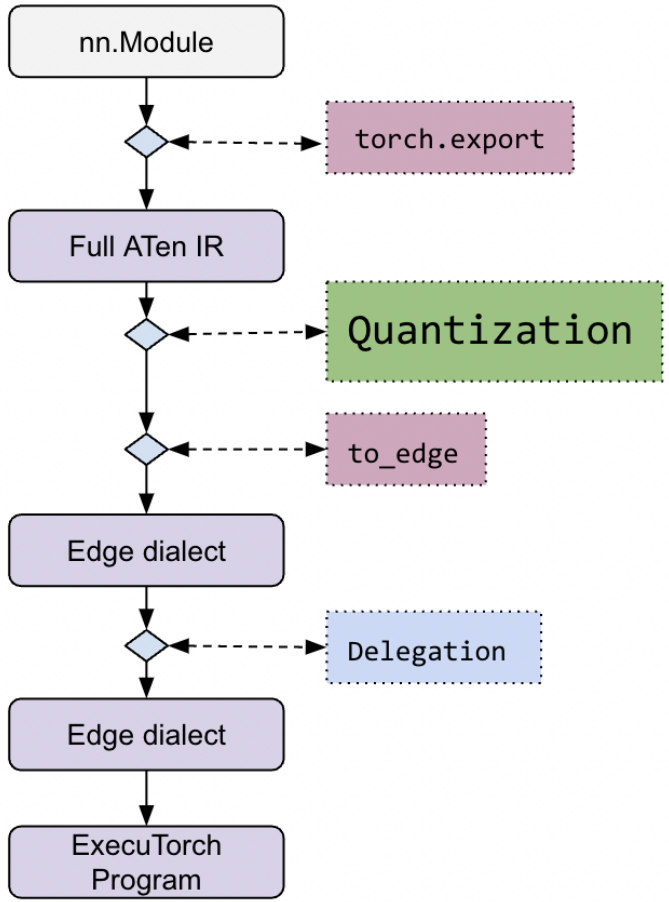
PyTorch Edge

ExecuTorch: A powerful on-device AI Framework

CONTRIBUTORS 219 | STARS 2.6K | DISCORD JOIN US | DOCUMENTATION

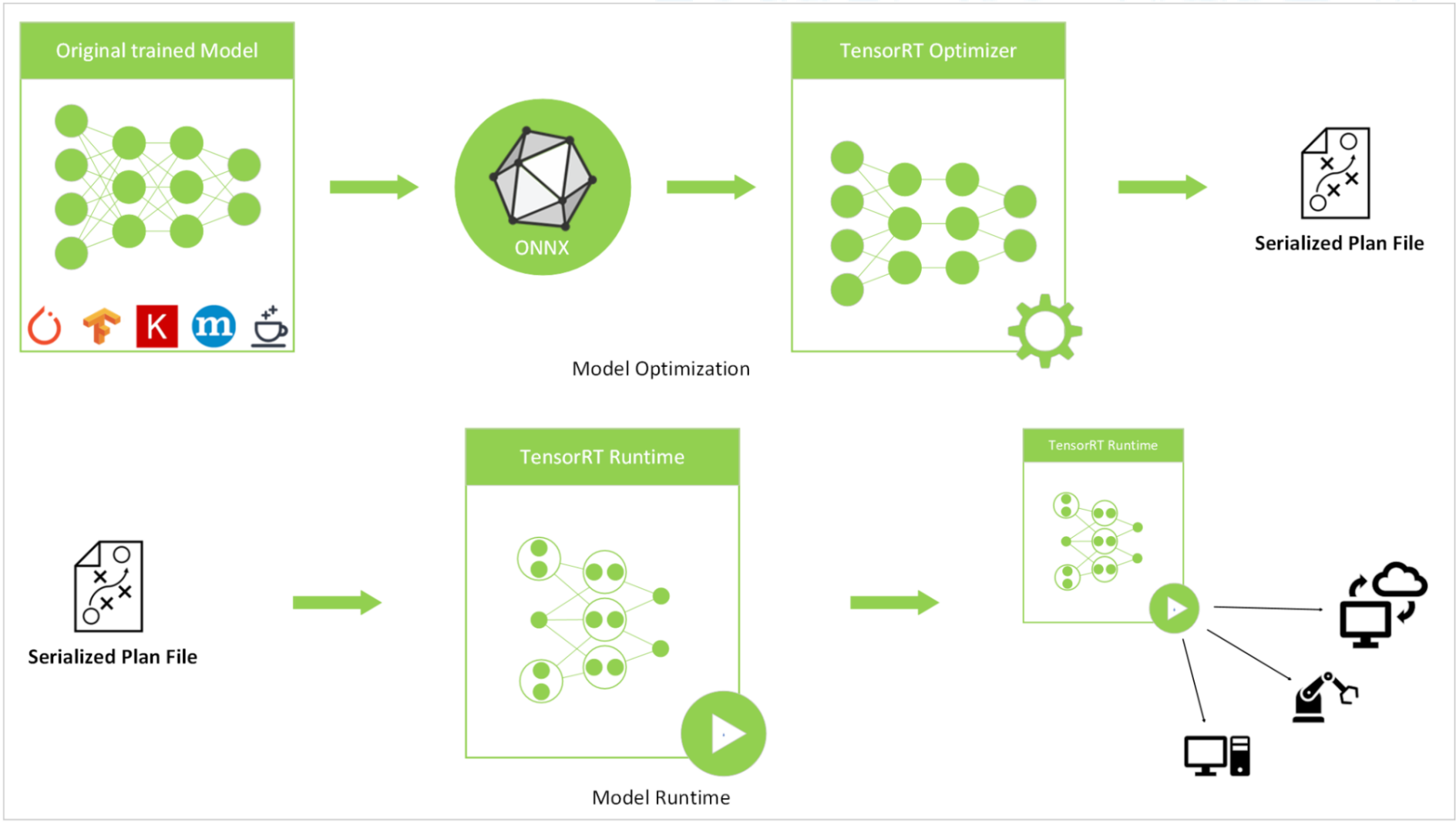
ExecuTorch is an end-to-end solution for on-device inference and training. It powers much of Meta's on-device AI experiences across Facebook, Instagram, Meta Quest, Ray-Ban Meta Smart Glasses, WhatsApp, and more.

It supports a wide range of models including LLMs (Large Language Models), CV (Computer Vision), ASR (Automatic Speech Recognition), and TTS (Text to Speech).



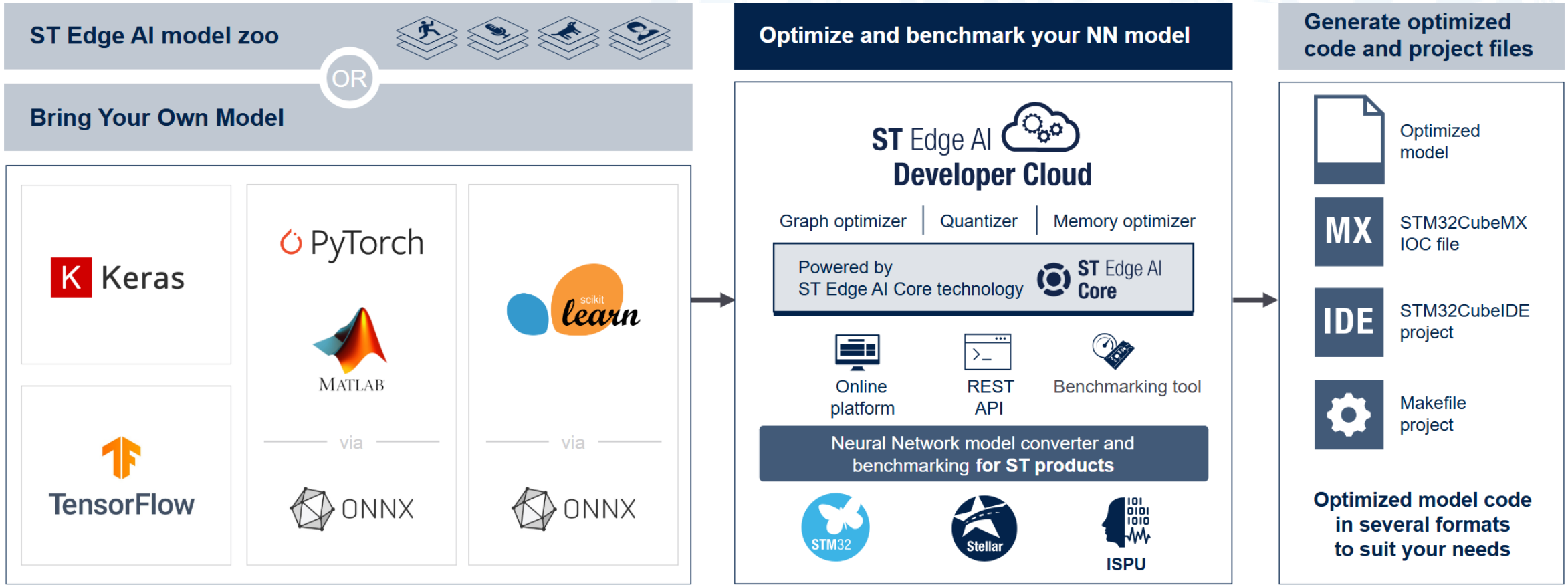
Source: https://pytorch.org/executorch-overview

NVIDIA TensorRT Model Optimizer



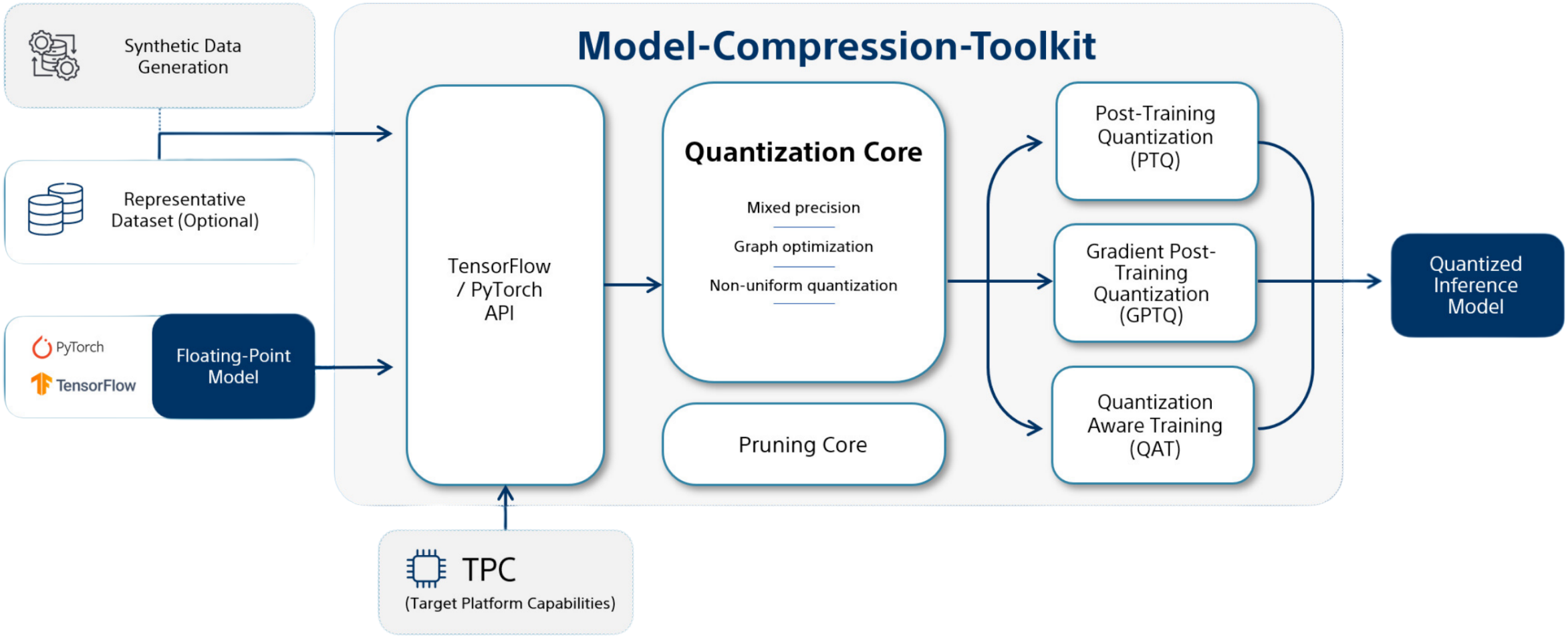
Source: https://developer.nvidia.com/tensorrt https://github.com/NVIDIA/TensorRT-Model-Optimizer

STMicroelectronics's Edge AI Platform



Source: https://stm32ai-cs.st.com/home

SONY's Model Compression Toolkit (MCT)



Source: https://github.com/sony/model_optimization

Intel OpenVINO

1 | Model → 2 | Optimize → 3 | Deploy

PyTorch · TensorFlow · TensorFlow Lite · PaddlePaddle · ONNX · Keras

Industry Model Zoos

Use open source models from industry model zoos such as Hugging Face, TensorFlow, Keras, PyTorch, PaddlePaddle, and ONNXpretrained models.

Optimum* for Intel

Use OpenVINO™ toolkit as an extension in Hugging Face* transformer models and gain model compression and performance benefits.

intel GETi

Build computer vision models in a fraction of the time and with less data.

Model Converter for OpenVINO™

Convert a trained model from supported frameworks.

IR Data → **Read, load, infer**

OpenVINO format (intermediate representation file, .pb, .tflite, .onnx)

Direct Model Conversion for TensorFlow* and PyTorch*

An option to skip steps to get to deployment faster.

Model Compression with NNCF

Neural Network Compression Framework (NNCF) provides quantization aware-training, model pruning, and sparsity along with posttraining optimization.

Jupyter* Notebook

Get sample code on the latest models to help get your application into production faster.

OpenVINO Model Server

Serve models over gRPC, REST, or C API endpoints.

OpenVINO Runtime

Common Python*, C, and C++ APIs that abstract low-level programming for each of the following devices.

intel Xeon · intel Core · intel Atom · intel Core Ultra · intel Arc Graphics · intel Iris Xe MAX Graphics · intel Data Center GPU · arm · intel FPGA AI Suite

Source: https://docs.openvino.ai/2025/index.html

Intel OpenVINO

Neural Network Compression Framework (NNCF)

Key Features • Installation • Documentation • Usage • Tutorials and Samples • Third-party integration • Model Zoo

release v2.15.0 | website docs | license Apache 2.0 | downloads 3M

python 3.9+ | backends openvino | pytorch | onnx | tensorflow | OS Linux | Windows | MacOS

Neural Network Compression Framework (NNCF) provides a suite of post-training and training-time algorithms for optimizing inference of neural networks in OpenVINO™ with a minimal accuracy drop.

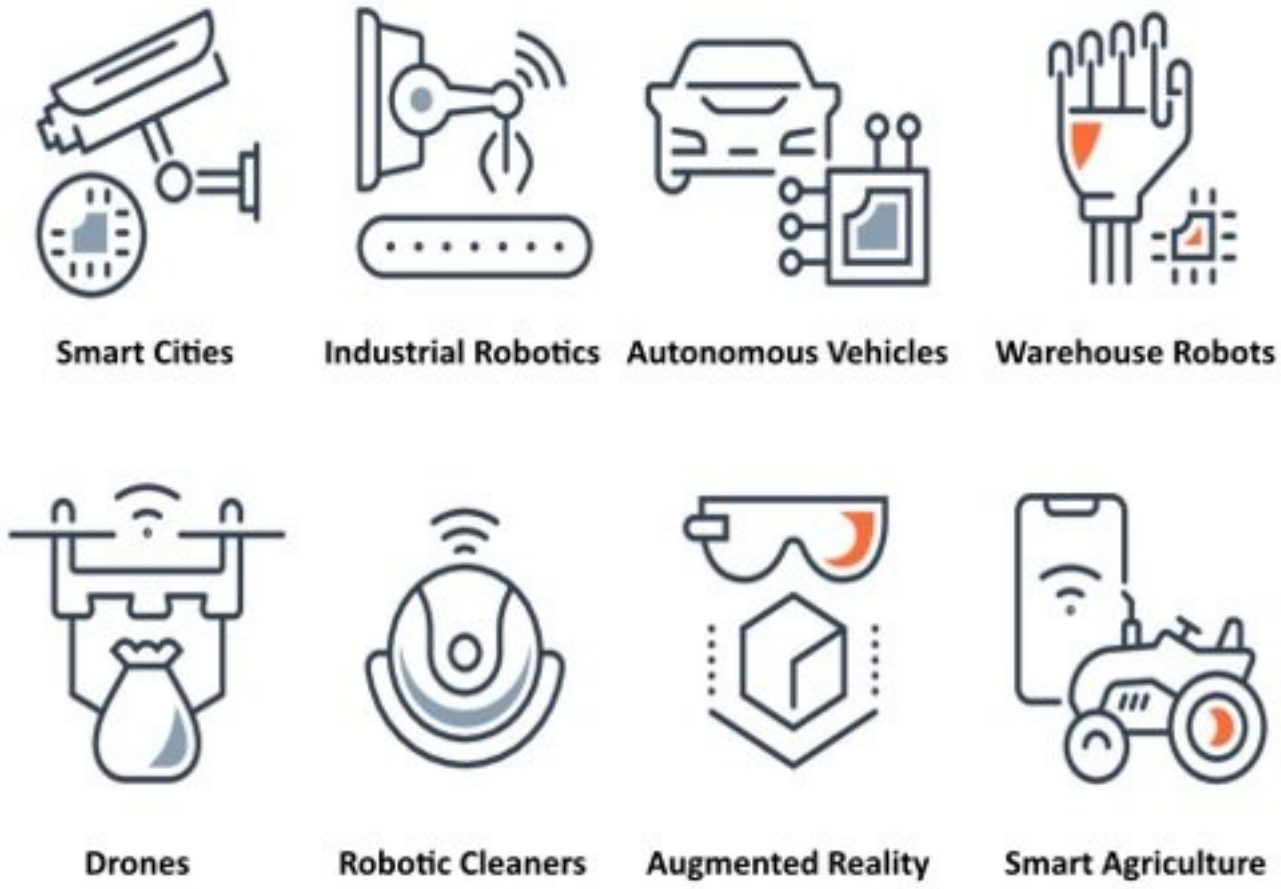
NNCF is designed to work with models from PyTorch, TorchFX, TensorFlow, ONNX and OpenVINO™.

NNCF provides samples that demonstrate the usage of compression algorithms for different use cases and models. See compression results achievable with the NNCF-powered samples on the NNCF Model Zoo page.

The framework is organized as a Python* package that can be built and used in a standalone mode. The framework architecture is unified to make it easy to add different compression algorithms for both PyTorch and TensorFlow deep learning frameworks.

Source: https://github.com/openvinotoolkit/nncf

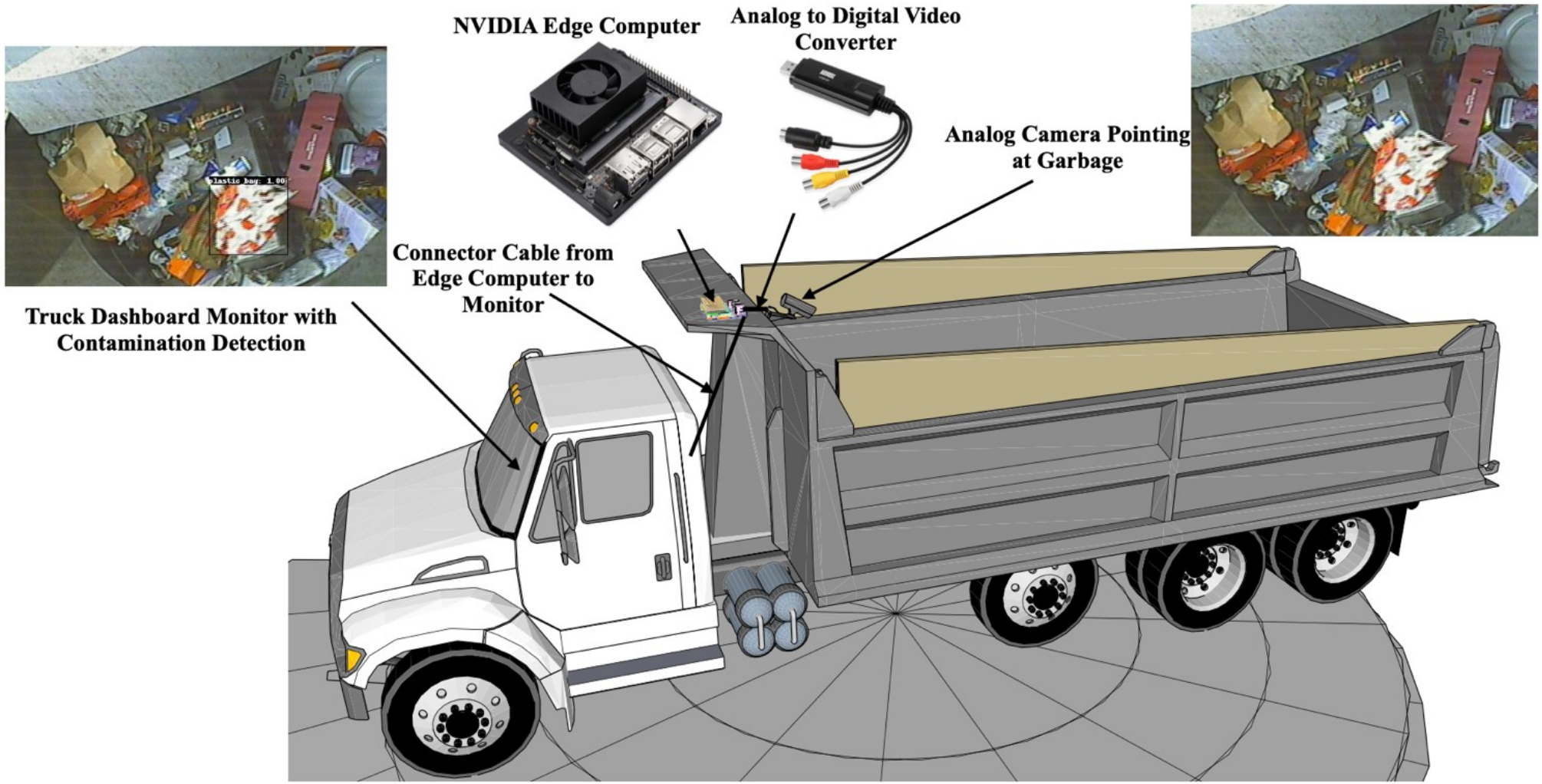
Edge AI applications



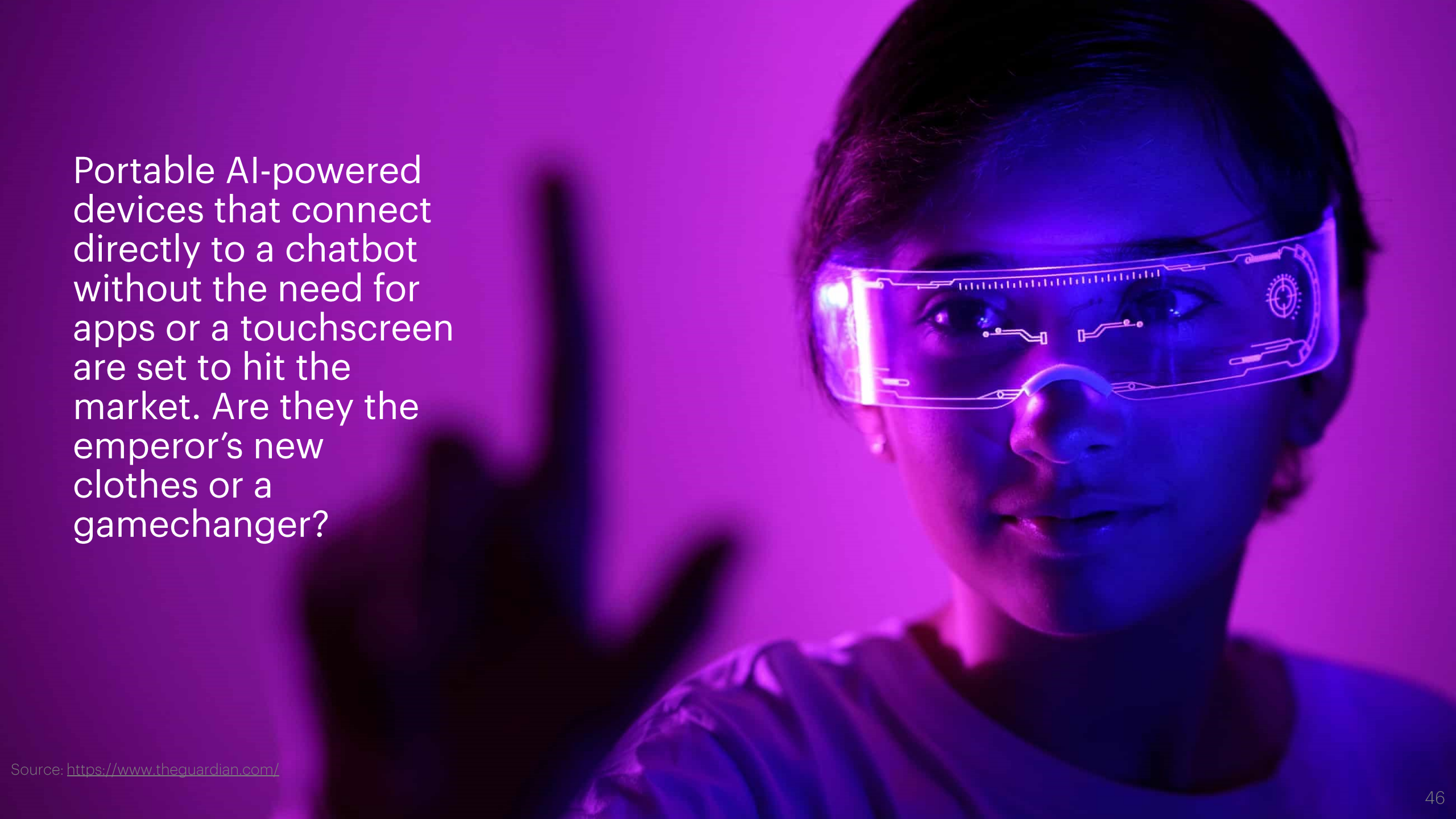
Source: Internet

Source: Internet

Detecting Real-Time Waste Contamination Using Edge Computing and Video Analytics



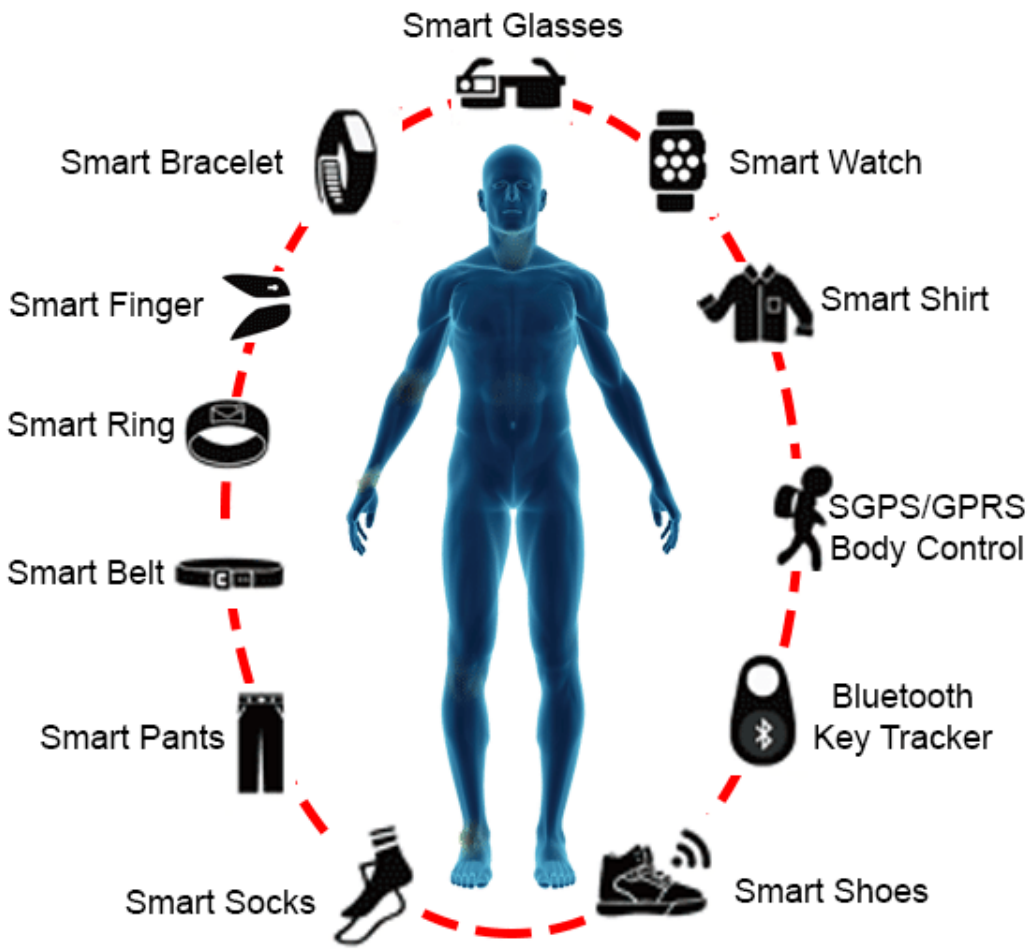
Source: https://developer.nvidia.com/blog/detecting-real-time-waste-contamination-using-edge-computing-and-video-analytics/



Portable AI-powered devices that connect directly to a chatbot without the need for apps or a touchscreen are set to hit the market. Are they the emperor’s new clothes or a gamechanger?

Source: https://www.theguardian.com/

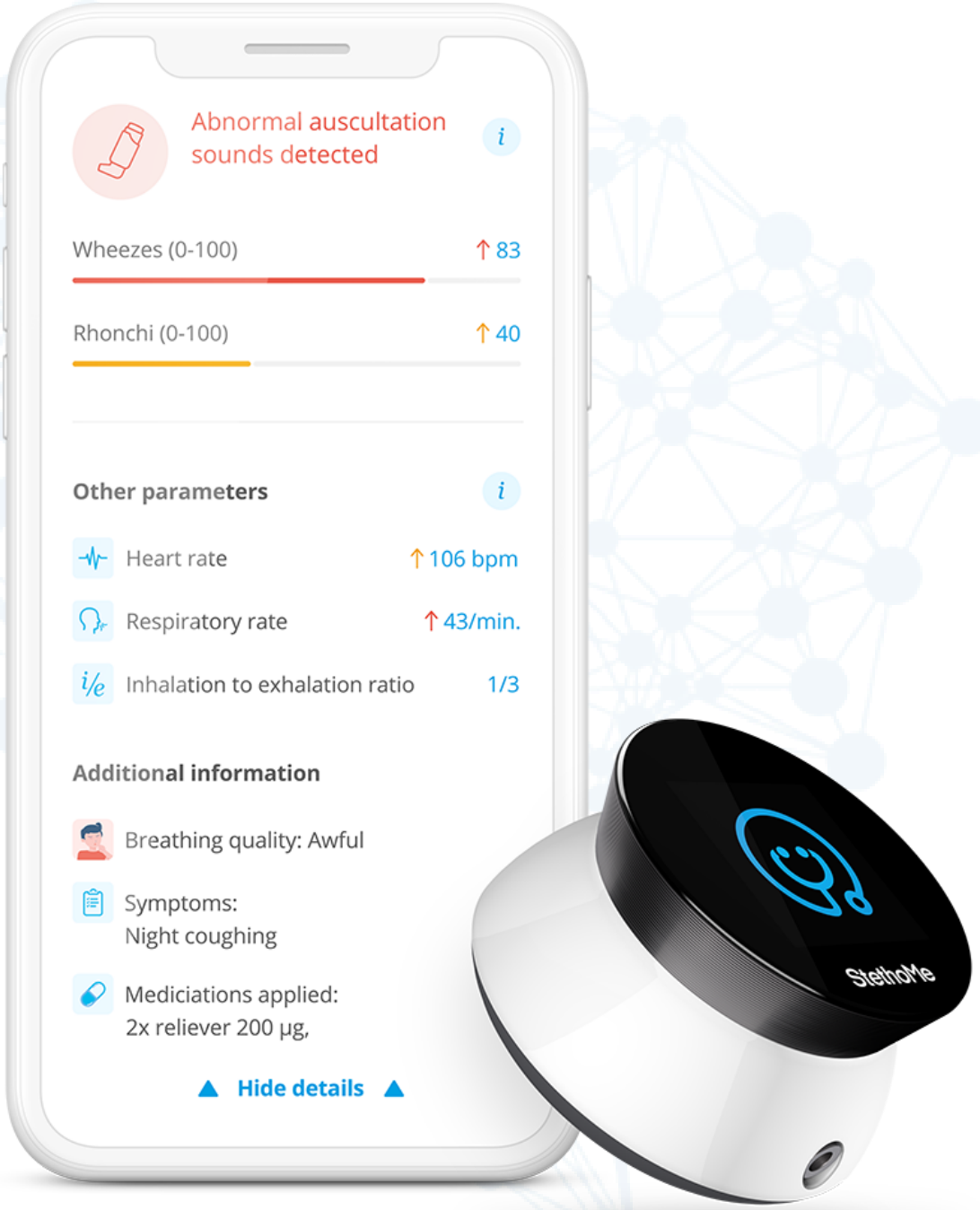
Wearable AI



Source: Internet

Healthcare AI

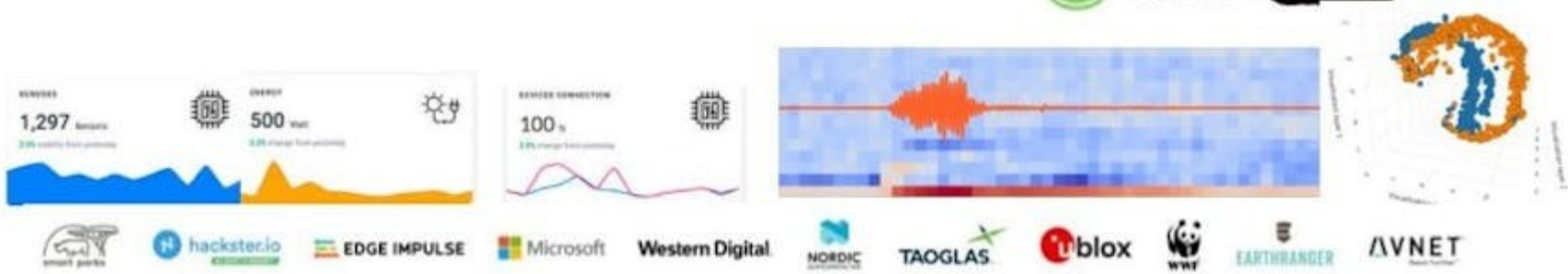
StethoMe® AI detects abnormal sounds in the respiratory system!!!



Source: https://www.stethome.com/

EleTect – EDGE IMPULSE

and /IOTCONNECT based Smart Wildlife Tracker Technology

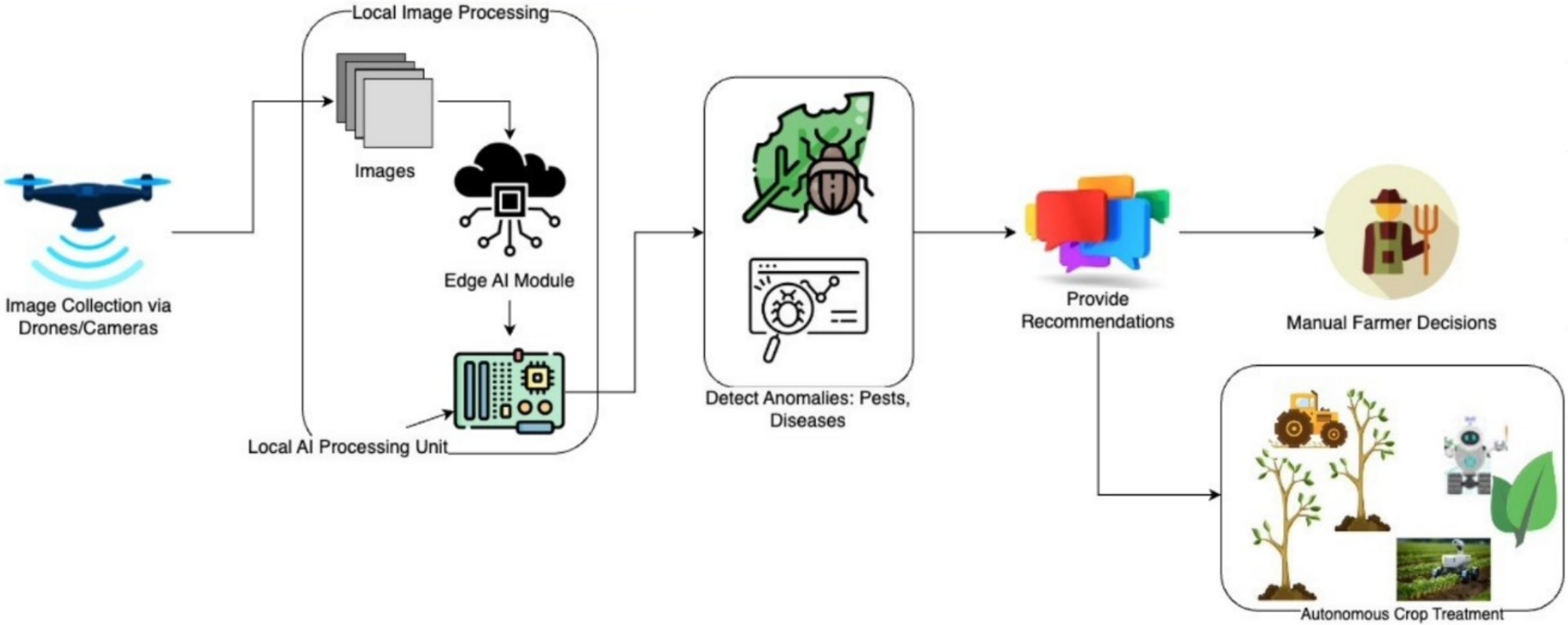


Source: https://www.hackster.io/dhruvsheth_/eletect-tinyml-and-iot-based-smart-wildlife-tracker-c03e5a

Smart Farming

Source: https://www.libramli.ai/

Smart Farming



Source: https://www.xenonstack.com/blog/edge-ai-agriculture

AI Camera

Edge-analytics powered
No dependence on server

On-premise solution
No need to pull live video over poor network

Source: https://www.honeywellbuildings.in

Edge AI Based Object & Face Detection Cameras VM-72B5AIVE

VMukti's smart product family, 4k products are equipped with Edge AI features such as Facial Recognition, Object Detection, Intrusion Detection and more. VMukti FHD, H.265+ PTZ Dome & Bullet Camera delivers up to FHD resolution (2592×1944, effectively four times that of Full HD) at 30 frames per second (fps), providing users with ultra-high-definition video viewing experience.

Source: https://www.vmukti.com/products/object-detection-camera/

Top stories

CES 2025 kicks off in Las Vegas

In these articles AMD ↑3.33%

Yahoo Finance

AMD announces new AI PC, gaming chips at CES 2025

7 hours ago

Firstpost

CES 2025: AMD announces new Ryzen gaming and AI PC processors, to offer…

2 hours ago

Interesting Engineering

AMD's new Ryzen 9 9950X3D CPU boosts gaming by 8%

10 hours ago

Wccftech

AMD Ryzen AI Max "Strix Halo" APUs Are The Ultimate AI PC APUs: 16 "Zen 5" & …

10 hours ago

Investor's Business Daily

AMD Reveals Next-Generation AI PC Chips, Gets Dell Endorsement

10 hours ago

Also in the news

Forbes

Nvidia Rocks CES With Grace-Blackwell AI PC Platform

1 hour ago

NVIDIA Newsroom

NVIDIA Launches AI Foundation Models for RTX AI PCs

2 hours ago

Source: Internet

What is an AI PC?

AI PCs use artificial intelligence technologies to elevate productivity, creativity, gaming, entertainment, security, and more. They have a CPU, GPU, and NPU to handle AI tasks locally and more efficiently.
-Intel

Source: https://www.scmp.com/ and https://www.intel.com/

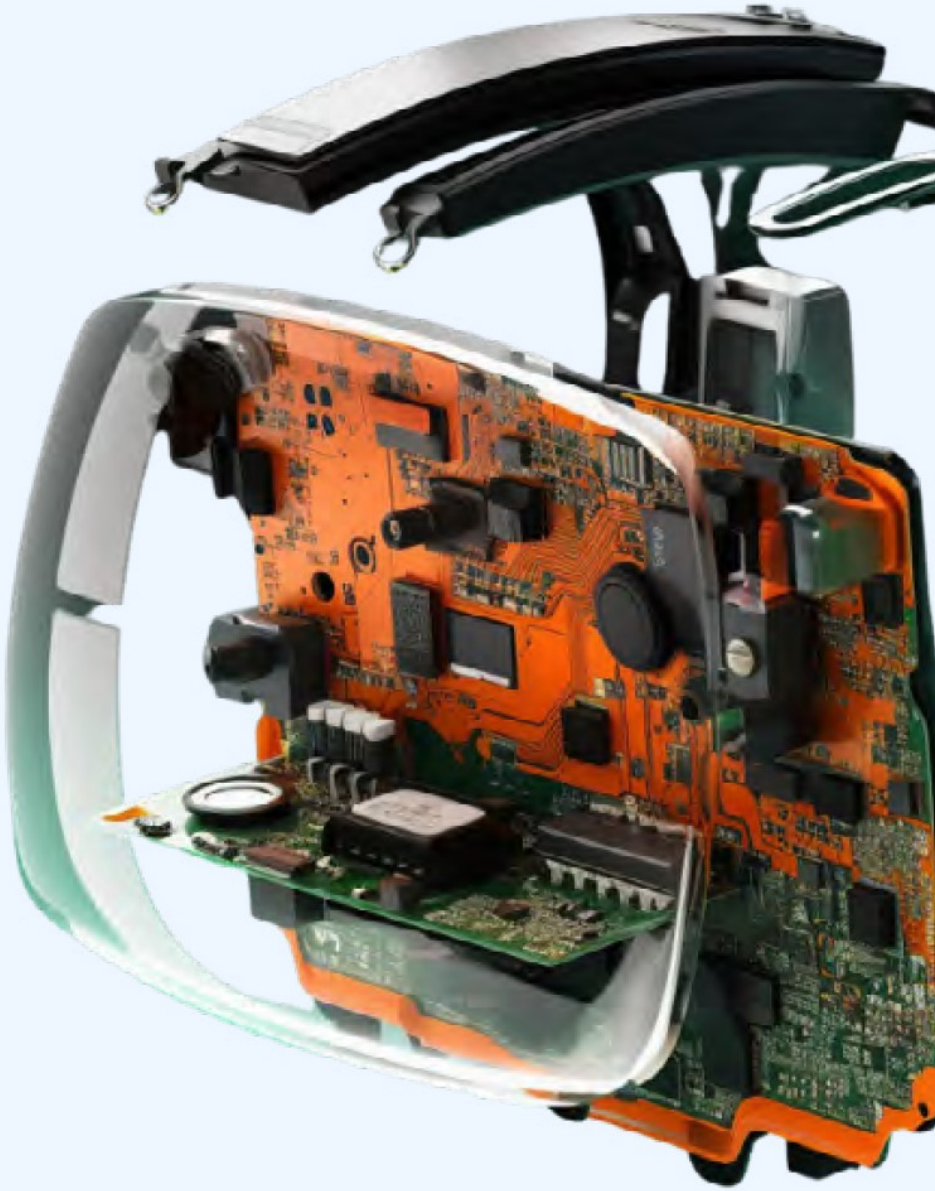
What is an AI PC?

An "AI-PC" is a PC with **new NPU silicon** that brings new AI experiences in productivity, creativity, and security through a combination of the CPU, GPU, and the new NPU.

Source: https://robquickenden.blog/2024/03/what-is-an-ai-pc/

2024 STATE OF EDGE AI REPORT

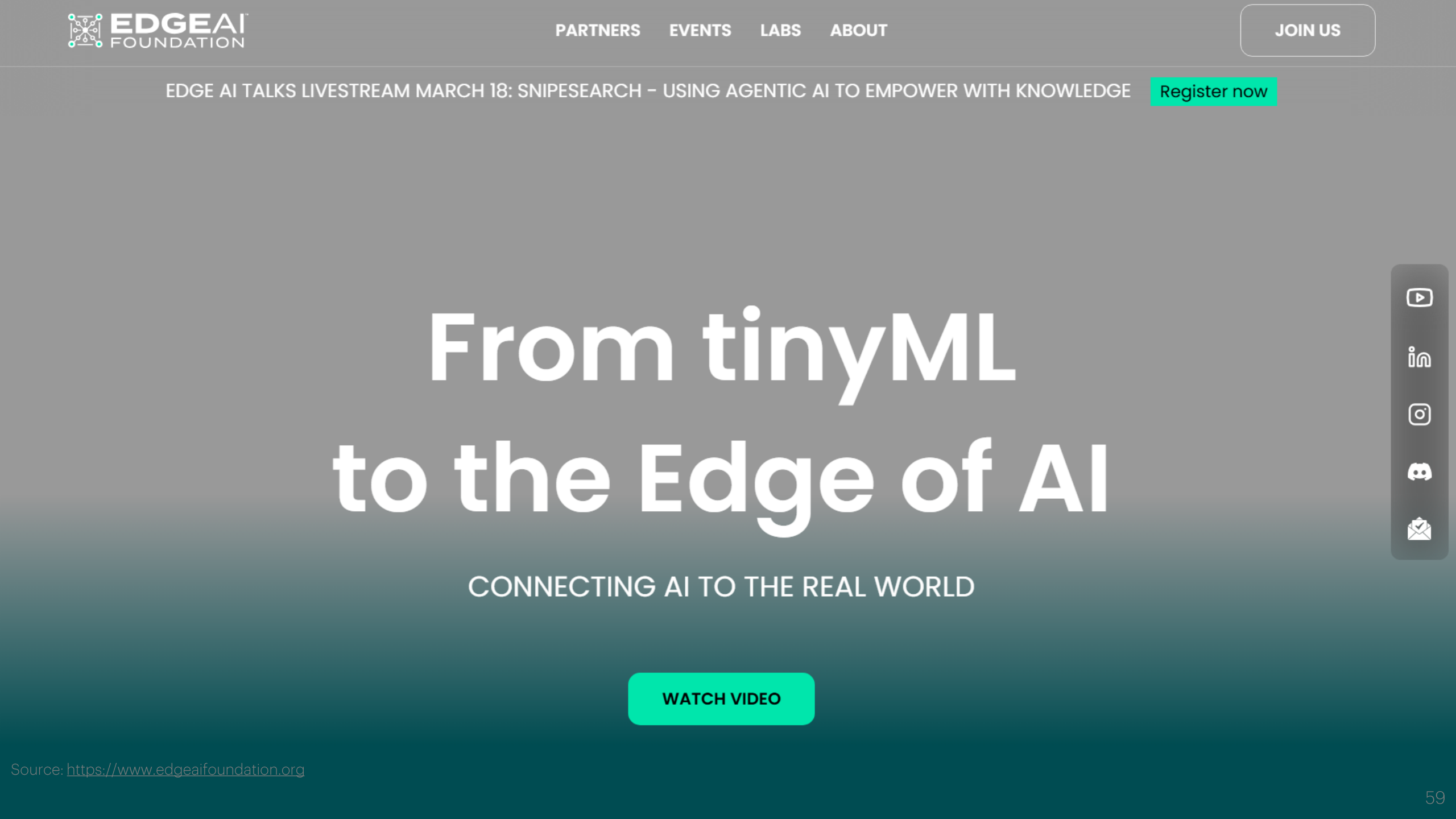
Exploring the dynamic world of Edge AI applications across industries



Source: https://www.wevolver.com/article/2024-state-of-edge-ai-report

The Edge AI Market is expected to reach USD 143.6 Billion by 2032

Source: https://market.us/



Source: https://www.edgeaifoundation.org

Source: https://www.edge-ai-vision.com

edge ai + vision ALLIANCE
Formerly the Embedded Vision Alliance

Home | The Alliance | Resources | The Summit | Members | News | Registration | Log In

If you're building AI or vision-enabled products, you've come to the right place.

The Alliance

The Edge AI and Vision Alliance is a worldwide industry partnership bringing together technology providers and end-product companies to accelerate the adoption of edge AI and vision in products.

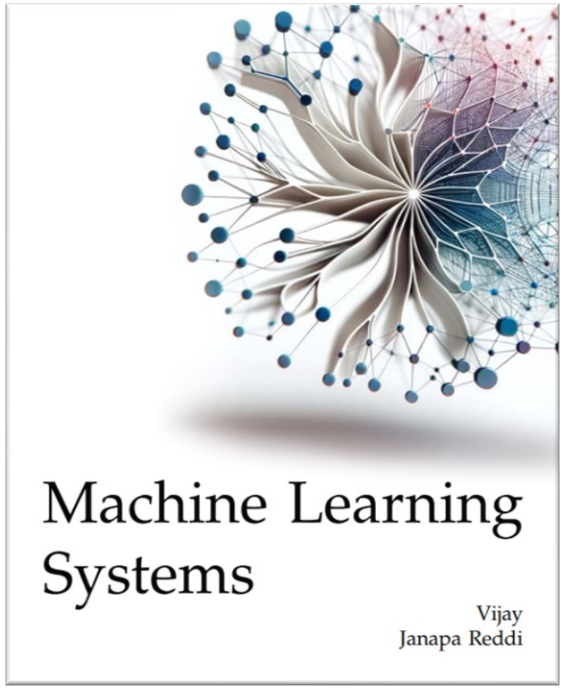
Subscribe to the Edge AI and Vision Insights Newsletter and stay up to date on the latest technology, applications, markets and trends in computer vision and edge AI.

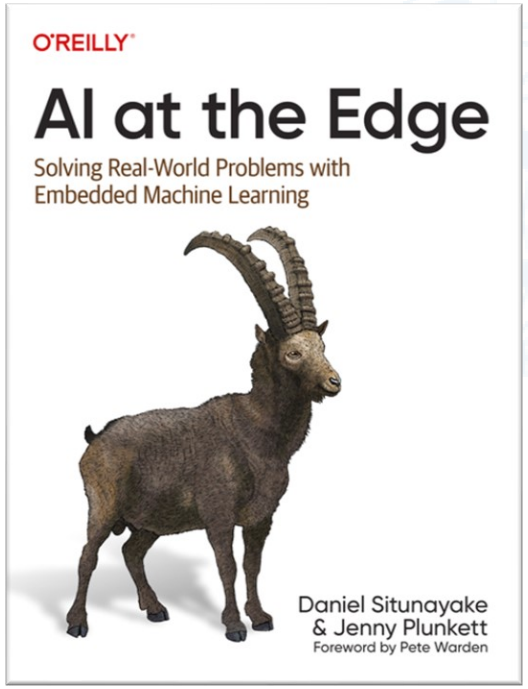
Course Topics

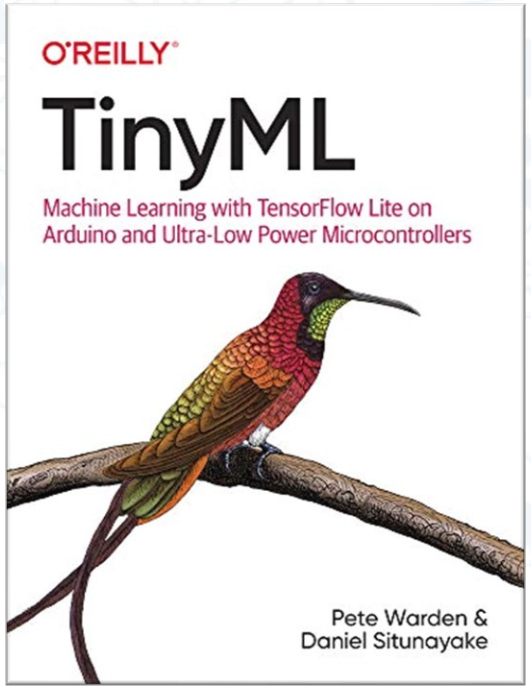
What will you learn in this course?

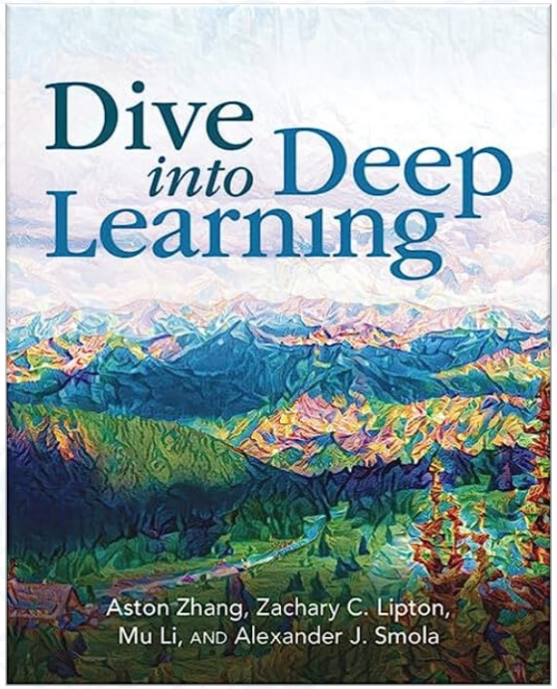
- Introduction to IoT, AI, and Edge AI
- IoT Devices and Embedded Systems
- Basics of AI, Machine Learning, Deep Learning, and Computer Vision
- Edge AI Hardware Platforms and Accelerators
- Edge AI Software Frameworks and Tools
- Model Compression and Optimization Techniques (Pruning, Quantization, and Knowledge Distillation).
- Deployment of efficient AI models and end-to-end applications
- Federated Learning
- Sustainable AI
- Applications: Human Activity Detection, Magic Wand, (Visual) Wake Word Detection
- Case Studies: Smart Cities, Smart Agriculture, Smart Healthcare

Books and reference materials









Plus, hands-on tutorials and research papers

Grading

- Five programming assignments (50%)
  - Implementation of an Embedded ML components and end-to-end applications
- One research seminar (10%)
  - Reading and presenting a research article
- Course project (40%)
  - Working Demo/Poster presentation

Prerequisites:

- Proficiency in Python and C programming
- Optional: Basics of AI/ML, programming microcontrollers and embedded systems, and IoT

Hardware Platforms

Arduino Tiny Machine Learning Kit

Arduino Nano 33 BLE Sense

OV7675 Camera

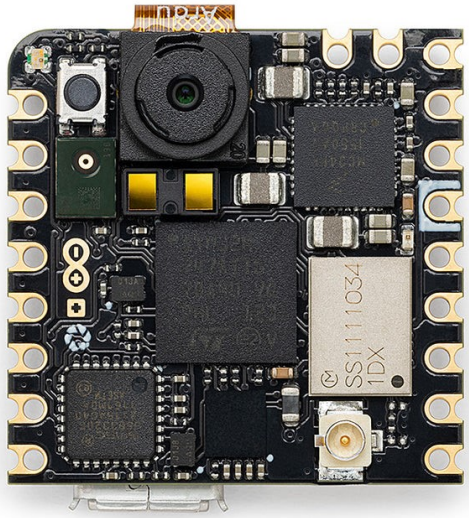
Arduino Tiny Machine Learning Shield

USB A to Micro USB Cable

Source: https://store.arduino.cc/products/arduino-tiny-machine-learning-kit

Hardware Platforms

Arduino Nicla Vision

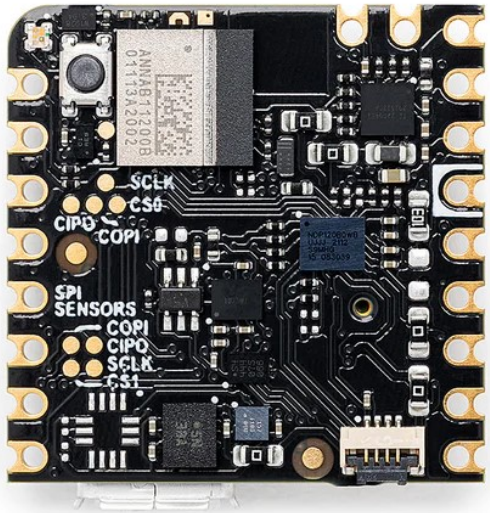


Microcontroller	•STM32H747AII6 Dual Arm® Cortex® M7/M4 IC:1x Arm® Cortex® M7 core up to 480 MHz •1x Arm® Cortex® M4 core up to 240 MHz
Sensors	•2 MP Color Camera •6-Axis IMU (LSM6DSOX) •Distance / Time Of Flight sensor (VL53L1CBV0FY/1) •Microphone (MP34DT05)
Power	•3.7V Li-po battery with Integrated battery charger and fuel gauge (MAX17262REWL)
Memory	2MB Flash / 1MB RAM 16MB QSPI Flash for storage
Connectivity	Wi-Fi / Bluetooth® Low Energy 4.2 (Murata 1DX - LBEE5KL1DX-883)

Source: https://docs.arduino.cc/hardware/nicla-vision/

Hardware Platforms

Arduino Nicla Voice

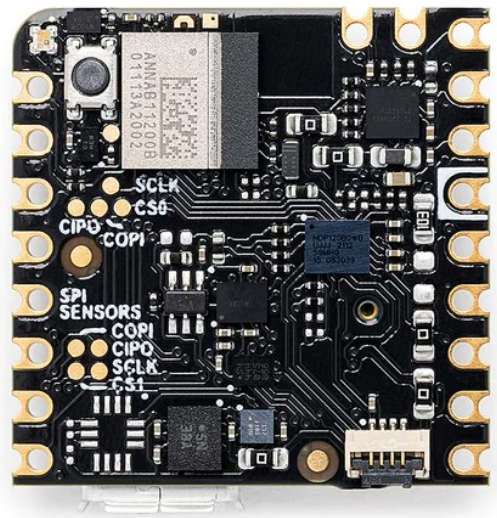


Microprocessor	•Syntiant® NDP120 Neural Decision Processor™ (NDP):1x Syntiant Core 2™ ultra-low-power deep neural network inference engine •1x HiFi 3 Audio DSP •1x Arm® Cortex® M0 core up to 48 MHz
Microcontroller	•Nordic Semiconductor nRF52832:64 MHz Arm® Cortex M4
Sensors	•High performance microphone (IM69D130) •6-Axis IMU (BMI270) •3-axis magnetometer (BMM150)
Memory	•512KB Flash / 64KB SRAM •16MB SPI Flash for storage •48KB SRAM dedicated for NDP120
Power	•3.7V Li-po battery with Integrated battery charger and fuel gauge (BQ25120AYFPR)
Connectivity	•Bluetooth® Low Energy (ANNA-B112)

Source: https://docs.arduino.cc/hardware/nicla-voice/

Hardware Resources Available

Arduino Nicla Sense ME



Microcontroller	64 MHz Arm® Cortex M4 (nRF52832)
Sensors	BHI260AP - **Self-learning AI smart sensor** with integrated accelerometer and gyroscope BMP390 - Digital pressure sensor, BMM150 - Geomagnetic sensor, BME688 - Digital low power gas, pressure, temperature & humidity sensor with AI
Connectivity	Bluetooth® 4.2
Power	Micro USB (USB-B), Pin Header, 3.7V Li-po battery with Integrated battery charger
Memory	512KB Flash / 64KB RAM, 2MB SPI Flash for storage, 2MB QSPI dedicated for BHI260AP

Source: https://docs.arduino.cc/hardware/nicla-sense-me/

Hardware Resources Available

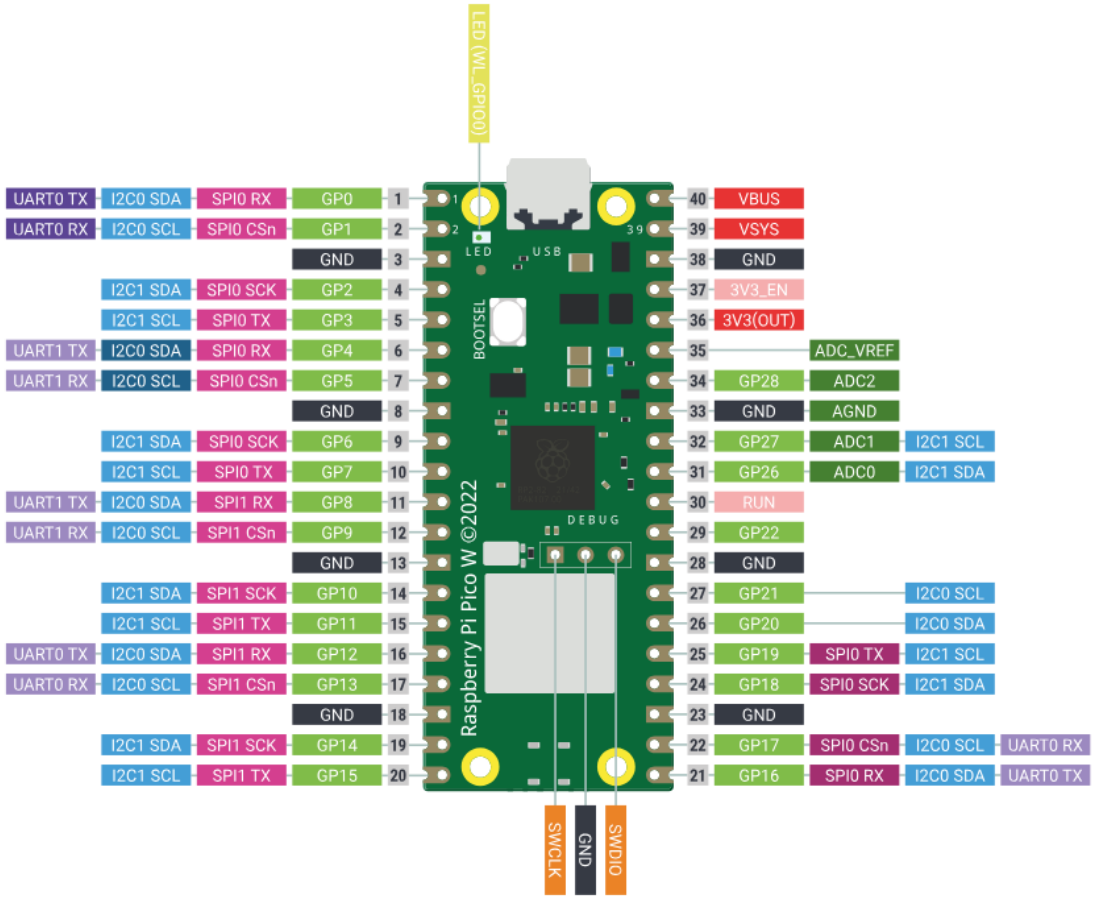
WIO Terminal

Main Chip	Core Processor	ARM® Cortex®-M4F running at 120MHz
	Memory	4 MB External Flash, 192 KB RAM
	Maximum Speed	200MHz
	External Flash	4MBytes
LCD Screen	Resolution	320x240
	Display Size	2.4inch
	Drive IC	ILI9341
Wireless	Wi-Fi	802.11 a/b/g/n 1x1, Dual Band 2.4GHz & 5GHz
	Bluetooth	Support BLE5.0
Built-in Modules	Accelerometer	LIS3DHTR
	Microphone	1.0V-10V -42dB
	Speaker	≥78dB @10cm 4000Hz
	Light Sensor	400-1050nm
	Infrared Emitter	940nm
Interface	MicroSD Card Slot	Maximum 16GB
	GPIO	40-PIN (Raspberry Pi Compatible)
	Supported Protocol	SPI, I2C, I2S, ADC, DAC, PWM, UART(Serial)
Operation Interface	5-Way Switch	
	Power/Reset Switch	

Source: https://www.seeedstudio.com/Wio-Terminal-p-4509.html

Hardware Resources Available

Raspberry Pi Pico W

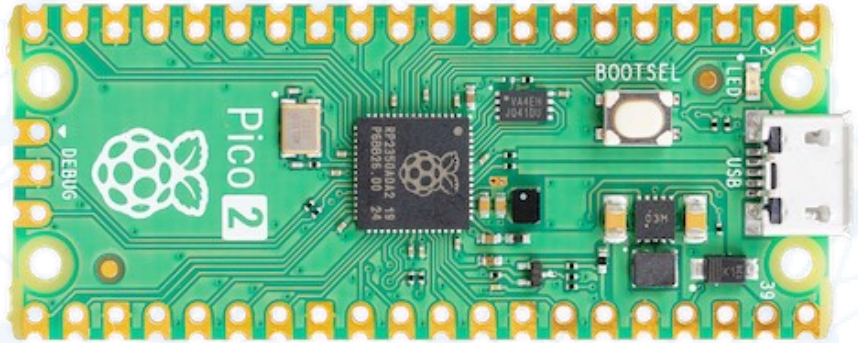


RP2040

- Power
- Ground
- UART / UART (default)
- GPIO, PIO, and PWM
- ADC
- SPI / SPI (default)
- I2C / I2C (default)
- System Control
- Debugging

Infineon 43439

- GPIO



- Dual-core Arm Cortex M0+ processor, flexible clock running up to 133 MHz
- 264kB of SRAM, and 2MB of on-board flash memory
- Wireless (802.11n), single-band (2.4 GHz)
- Bluetooth 5.2
- **Ideal board for connecting external sensors!**

Source: https://www.raspberrypi.com/documentation/microcontrollers/pico-series.html

Hardware Platforms

- reComputer J1010 Edge AI Device with Jetson Nano
- JetBot AI Kit

1. Powered by the quad-core ARM® Cortex®-A57 MPCore processor.
2. 128-core NVIDIA Maxwell™ GPU with 128 NVIDIA CUDA® cores delivers 0.5 TFLOPs (FP16).
3. RTC Connector
4. M.2E Connector
5. Module: JetSon Nano
6. USB Type: C power Connector
7. Rich peripherals including Gigabit Ethernet port, USB 3.0 and USB 2.0 Type-A ports, HDMI port.
8. Pre-installed NVIDIA official JetPack software, ready for cloud native application.

Source: https://robu.in/product/diy-nvidia-jetsonnano-development-learning-kit-jetbot/

Similar Courses

- edX tinyML Specialization (Harvard University) [Link]
- CS249r: Tiny Machine Learning (Harvard University) [Link]
- ESE 3600: Tiny ML (University of Pennsylvania) [Link]
- TinyML and Efficient Deep Learning Computing (MIT) [Link]
- Embedded Deep Learning & TinyML (Carnegie Mellon University) [Link]
- IoT and Tiny Machine Learning (Marquette University) [Link]
- Machine Learning for Embedding Devices (UNIFEI, Brazil)[Link]
- Introduction to Embedded Machine Learning (Edge Impulse) [Link]
- Computer Vision with Embedded Machine Learning (Edge Impulse) [Link]
- Edge AI and Robotics (NVIDIA) [Link]
- AI on the Edge with Computer Vision (Intel) [Link]
- Machine Learning at the Edge on Arm: A Practical Introduction (Arm)[Link]

Intel® Edge AI Certification

Advance your career and get recognized for your new, marketable skills with Intel® Edge AI Certification. The self-paced, online training course includes virtual classroom instruction and hands-on projects as you learn to use the latest Intel® developer tools and platforms to create your own portfolio of edge AI solutions. Register today to start earning your certificate and qualify for great career opportunities.

Register

Sign In

Supercharge Your Career

Intel Edge AI Certification training courses are a testament to your competency of building edge AI solutions using Intel developer tools and platforms. The courses, hands-on exer
and completed at no charge. At the end of successful completion, you will be given a course completion certificate. Annual follow up is recommended to access course u

Sign Up

Certification Benefits

Get access to exciting career opportunities and strong professional networks. Your Intel Edge AI Certification tells employers that you are a solid candidate with the skills and expertise they need on their development team.

Update Your Skills

Technology moves fast. Keep up with the latest tools and methods, and gain practical and hands-on solution-based skills to build your portfolio of

Earn Your Certificate

The course completion certificate attests to your hands-on experience, knowledge, and skills in the high-growth, high-demand field of edge AI

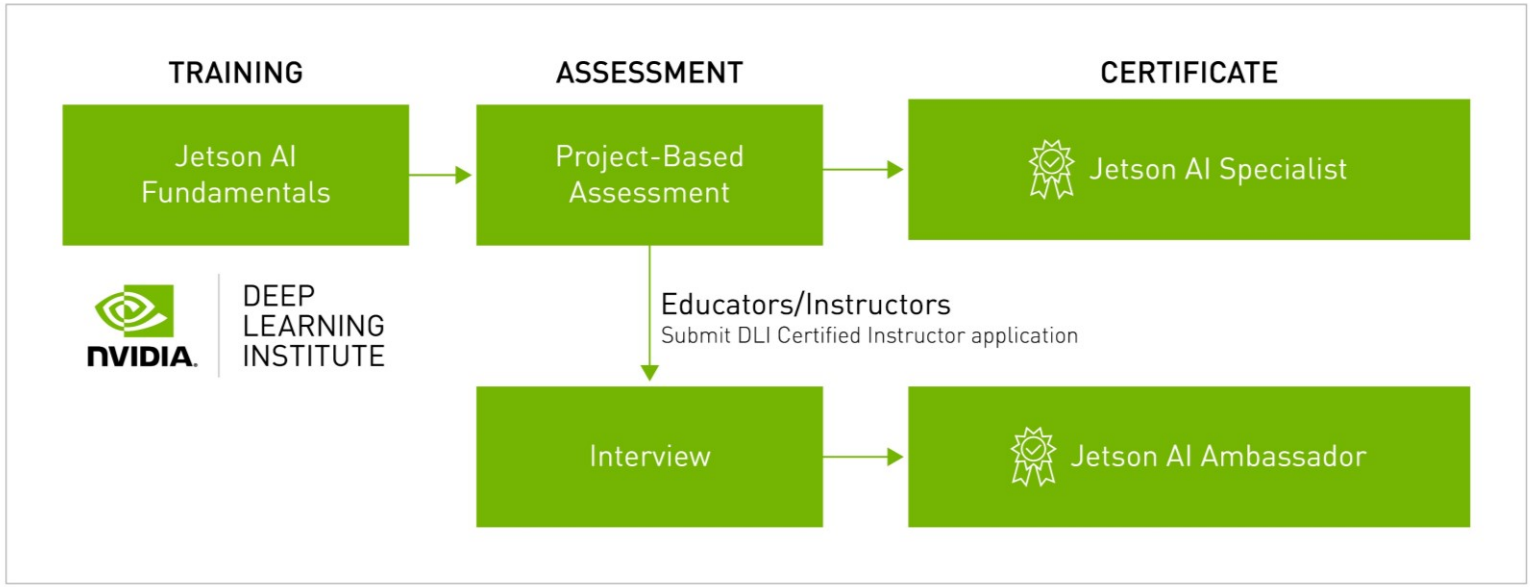
Boost Your Career

Use your newly acquired skills to advance in your career or find your next opportunity.

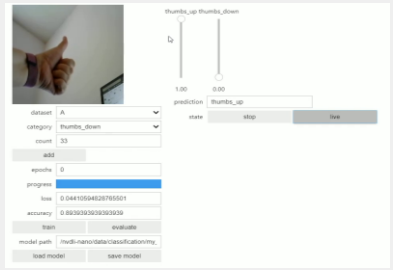
Grow Your Network

Expand your network by sharing your work and certificate with others in the AI community.

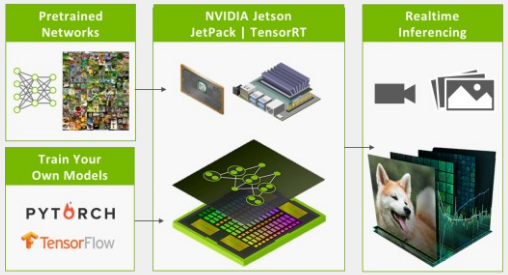
Jetson AI Certifications



Jetson AI Fundamentals



Getting Started with AI on Jetson Nano



Hello AI World

JetBot

nvidia.com/jetson-ai-certification

References

1. NVIDIA Edge AI and Robotics Teaching Kit [Link]
2. Machine Learning Systems [Link]
3. AI at the Edge Book [Link]
4. TinyML Courseware [Link]
5. Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing, Zhou et al. 2019 [Link]
6. Tiny Machine Learning: Progress and Future, Lin et al. 2023 [Link]

EDGE AI

THE FUTURE OF ARTIFICIAL INTELLIGENCE

Source: https://premioinc.com