EDGE AI:

Artificial Intelligence at the Edge

Hamed Farbeh

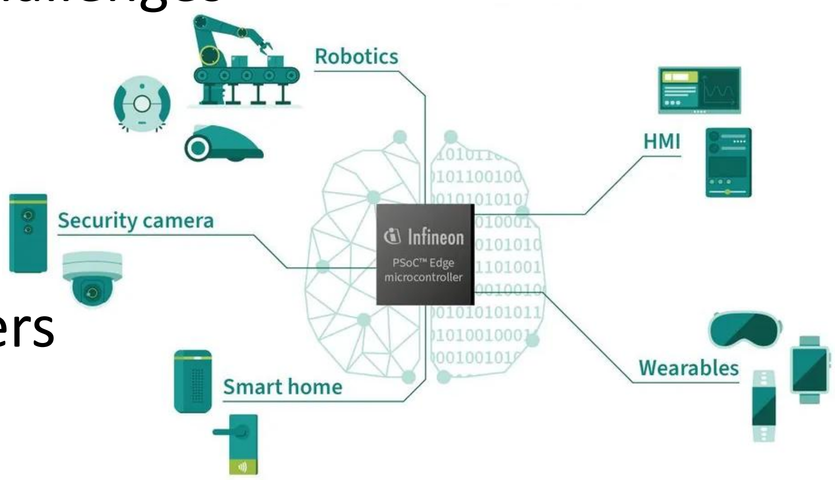
farbeh@aut.ac.ir

Department of Computer Engineering

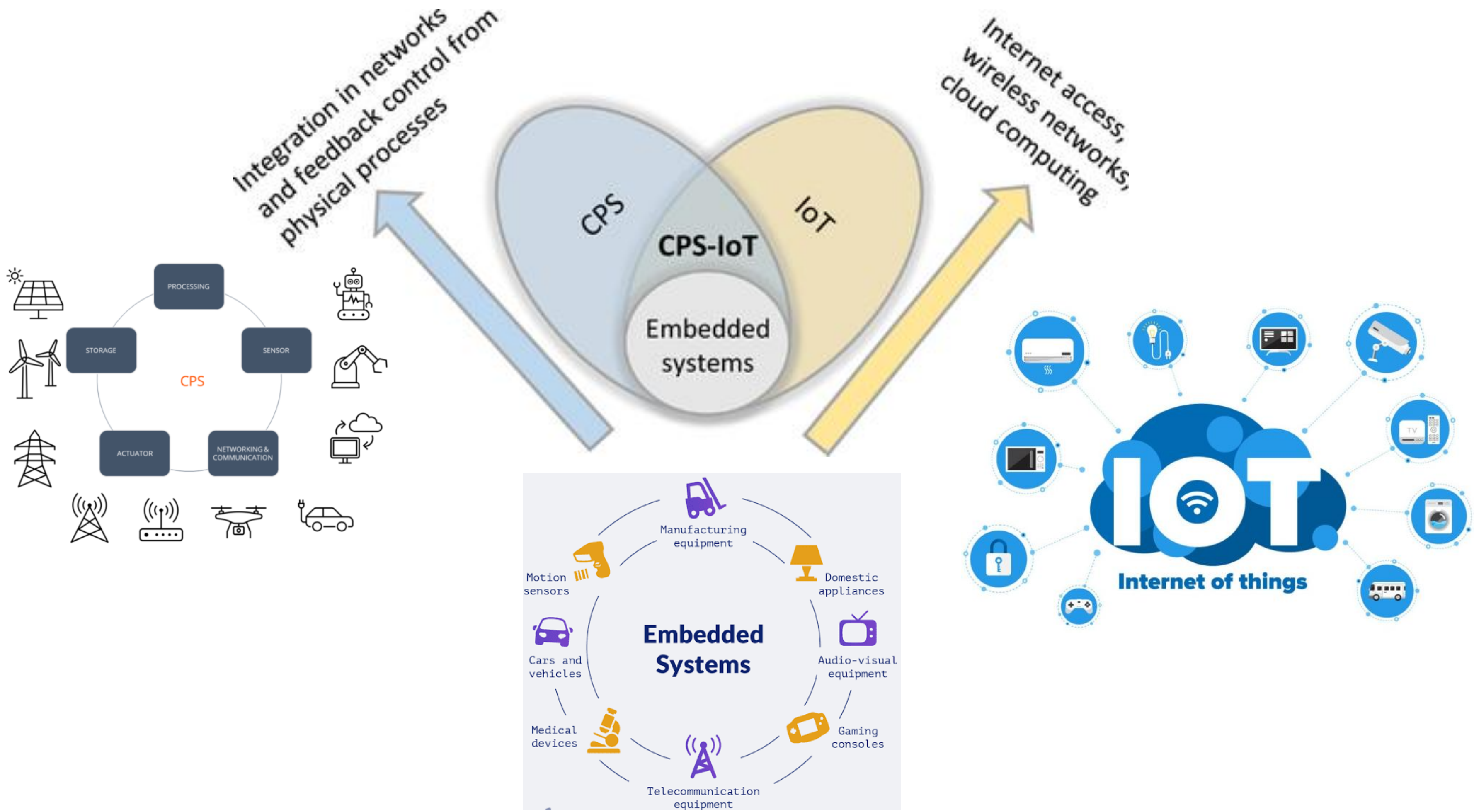
Amirkabir University of Technology

Course Outline

- Internet of Things
- Technical Development and Challenges
- Edge Computing
- Edge AI and TinyML
- TinyML Example and Key Players
- Neuromorphic Computing
- Challenges and Future perspectives
- Conclusion

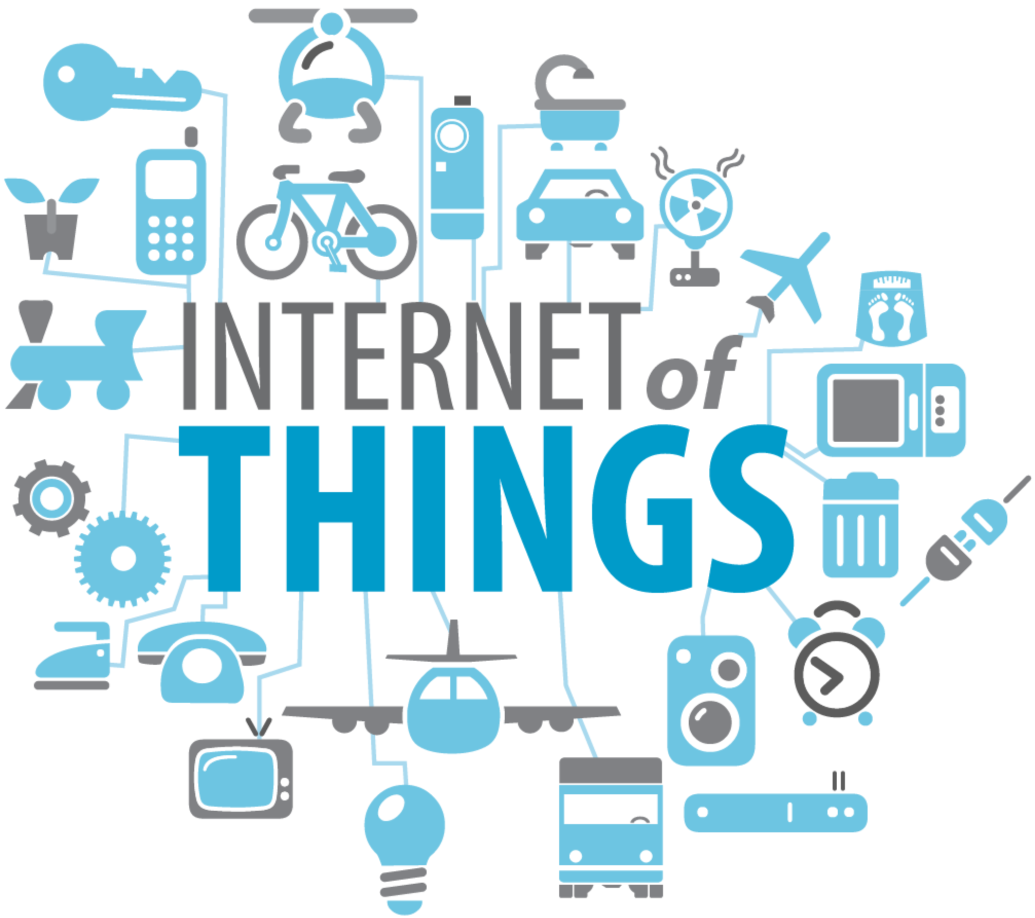


Embedded Systems, CPS, and IoT



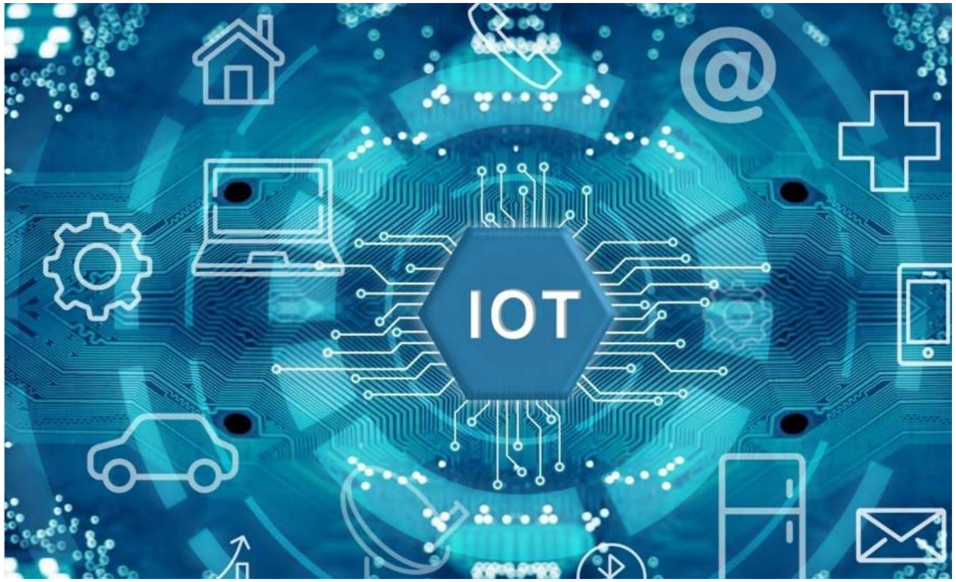
What is IoT

The Internet of things is the network of physical objects or "things" embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data

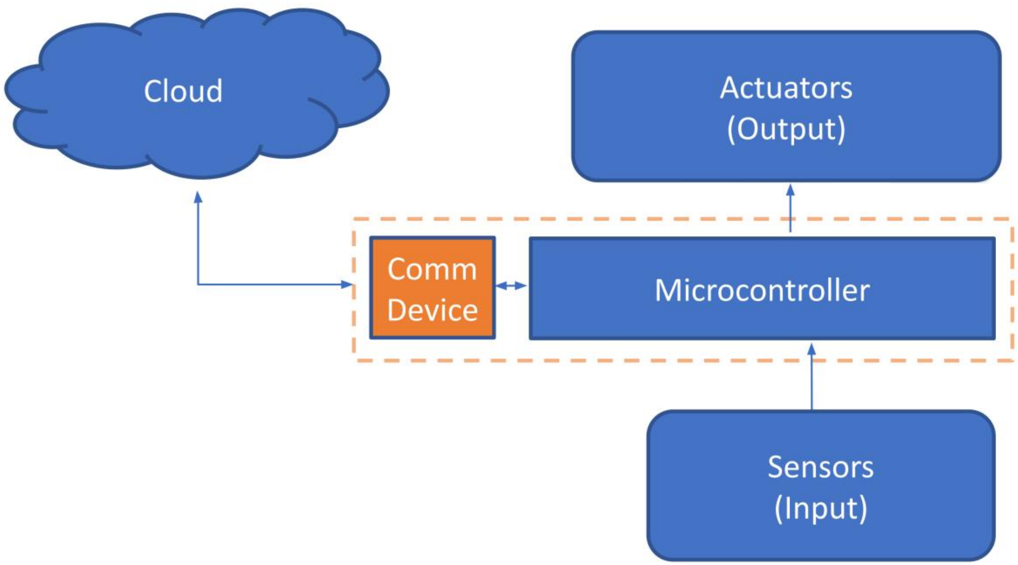


Technical developments for IoT realization

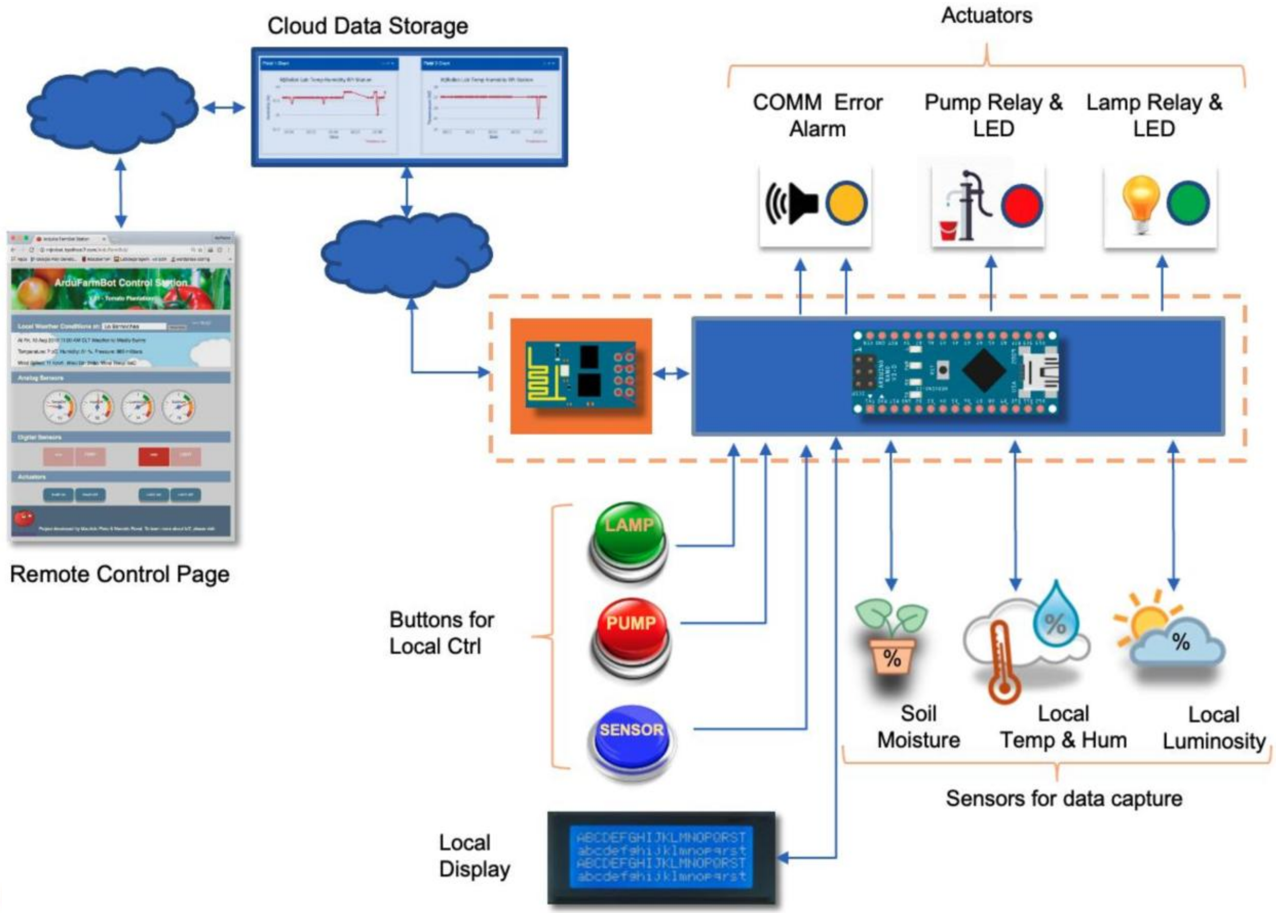
- Communication and cooperation
- Addressability
- Identification
- Sensing
- Actuation
- Embedded information processing
- Localization
- User interfaces



Typical IoT Project

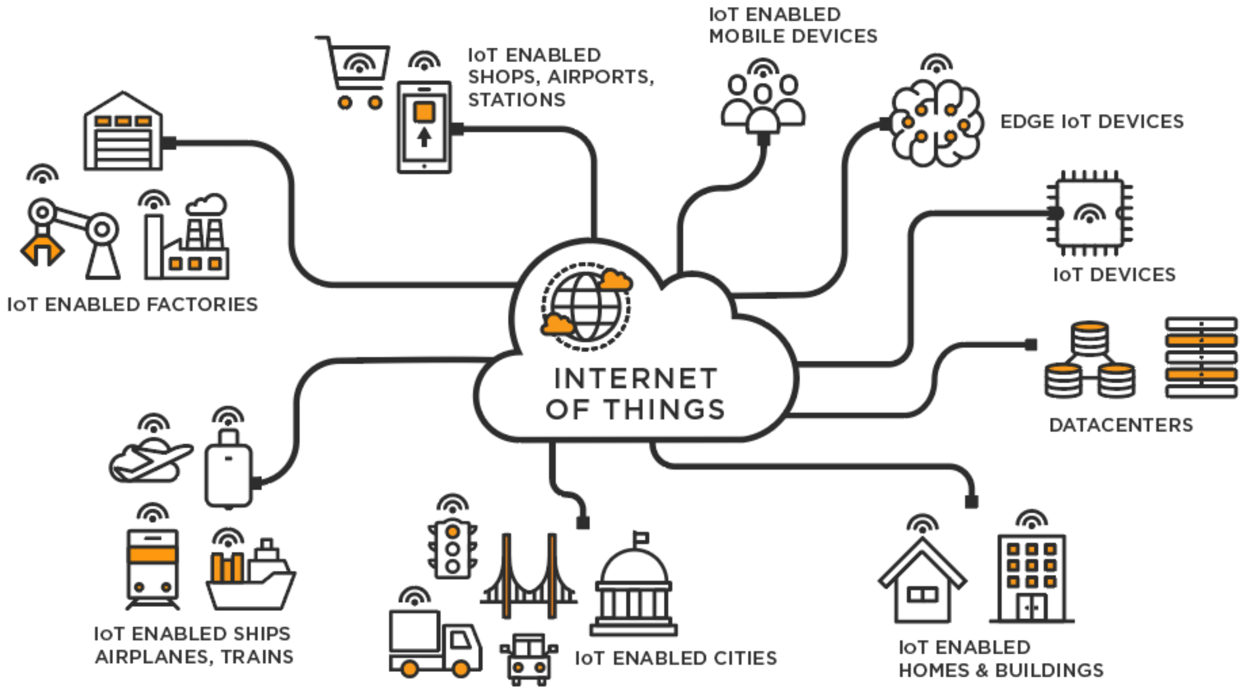


Typical IoT Project



Main IoT Challenges

- Energy Consumption
- Restricted Resources
- Latency and Bandwidth
- Timeliness
- Security
- Privacy

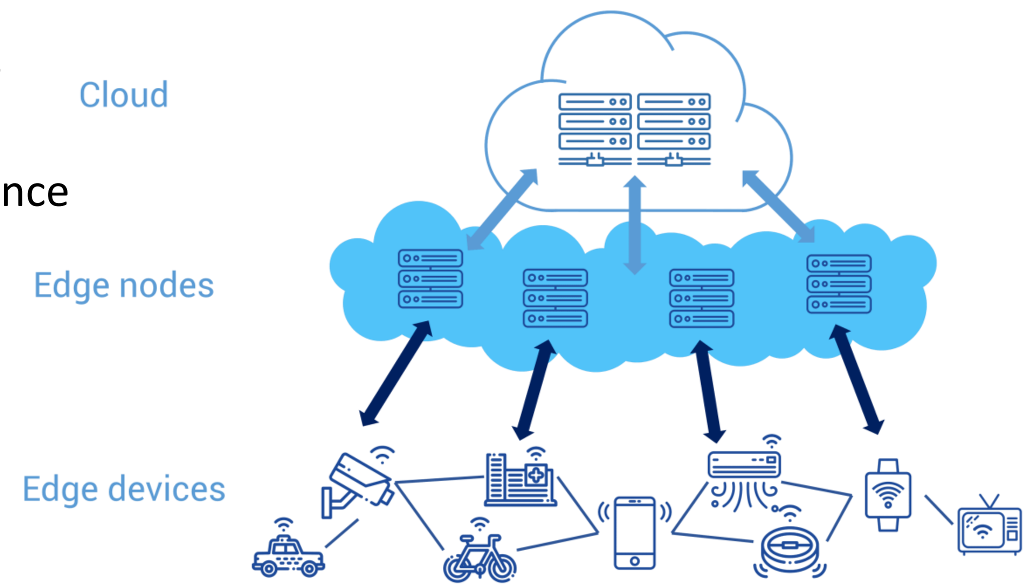


Edge Computing

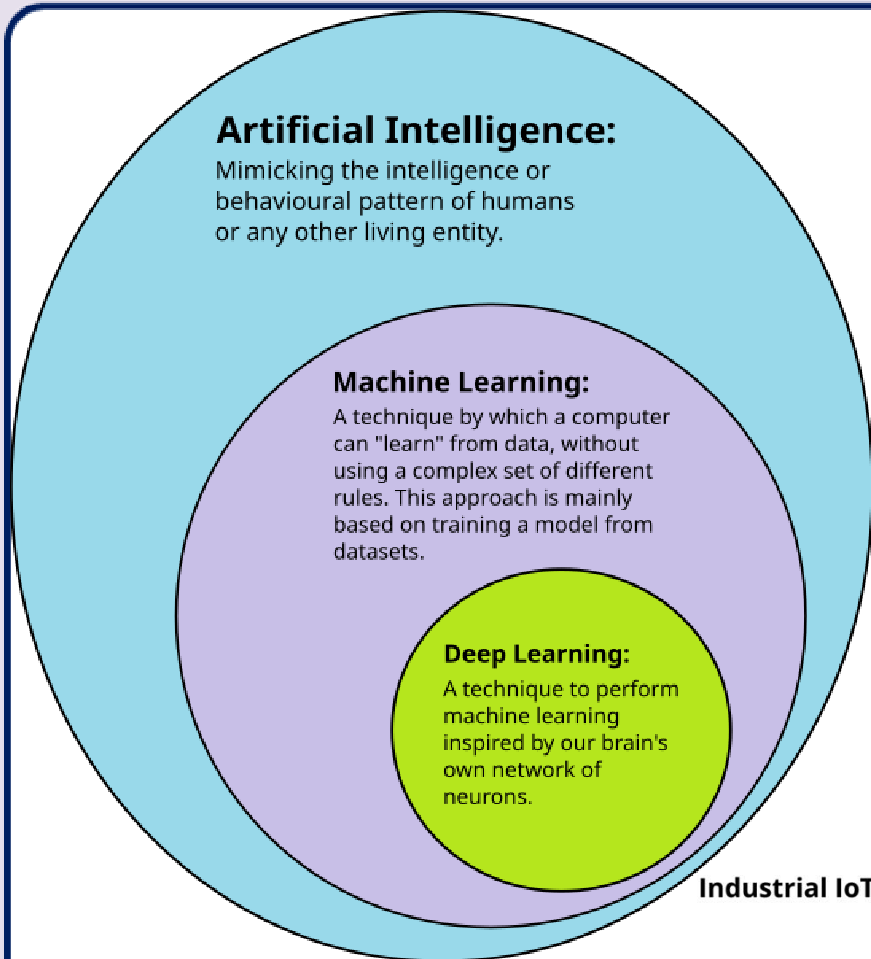
- Decentralized computing model: processes data closer to the source
- Reduces latency, bandwidth usage, and dependency on cloud servers
- Enables real-time processing for IoT applications

Key Applications

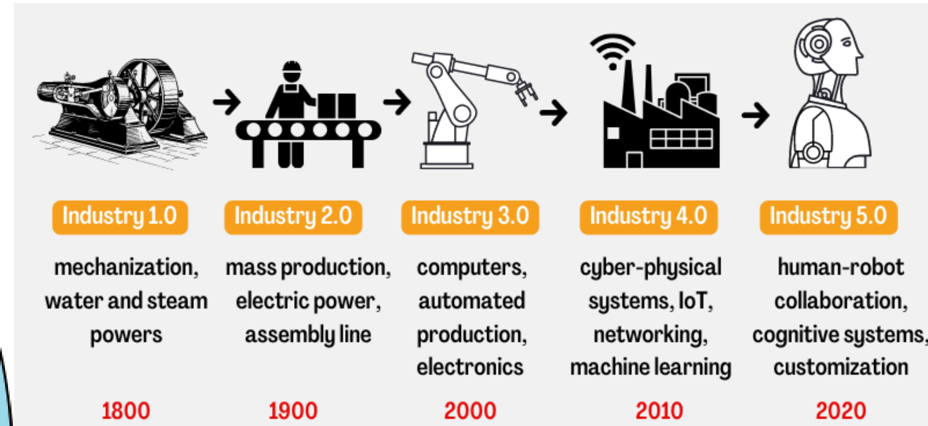
- **Smart Cities** – Real-time traffic management & public safety
- **Healthcare** – Wearable sensors for continuous monitoring
- **Industry 4.0** – Predictive maintenance in factories
- **Autonomous Vehicles** – Instant decision-making without cloud dependency



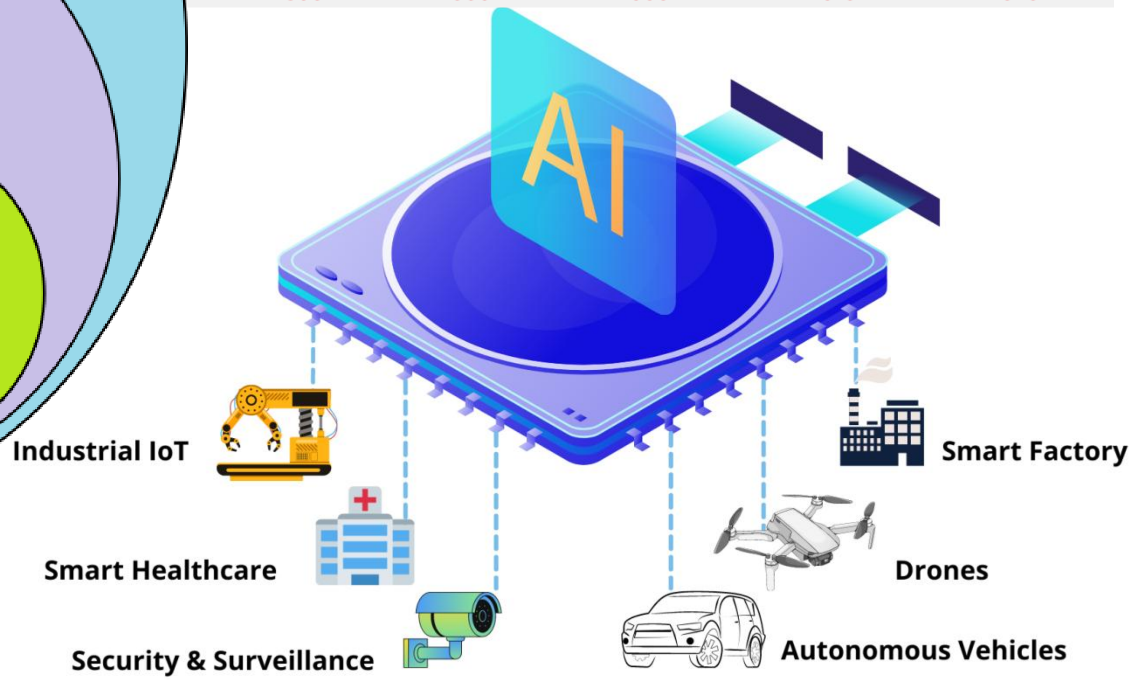
Edge AI



Artificial Intelligence:
Mimicking the intelligence or behavioural pattern of humans or any other living entity.
Machine Learning:
A technique by which a computer can "learn" from data, without using a complex set of different rules. This approach is mainly based on training a model from datasets.
Deep Learning:
A technique to perform machine learning inspired by our brain's own network of neurons.

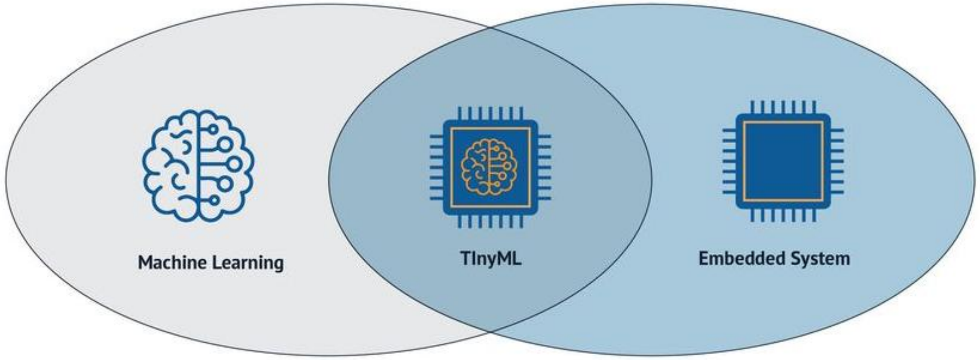


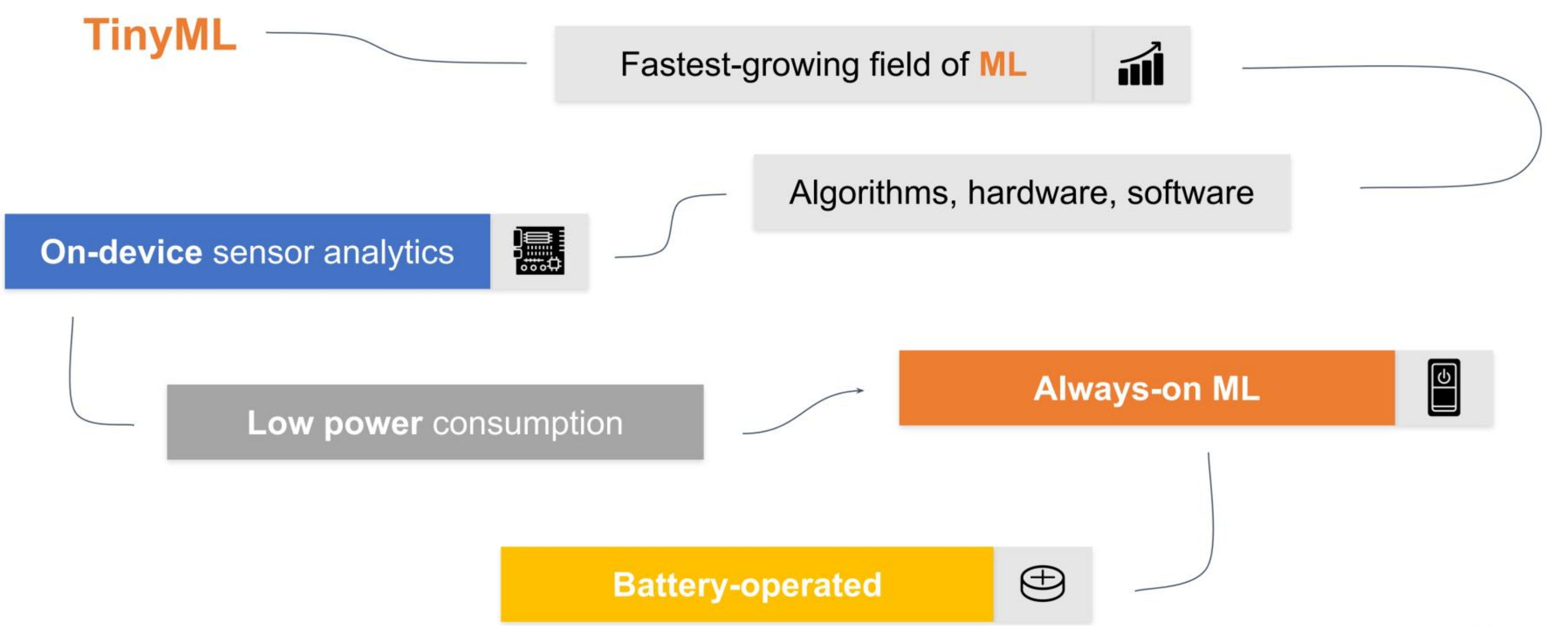
Industry 1.0
mechanization, water and steam powers
1800
Industry 2.0
mass production, electric power, assembly line
1900
Industry 3.0
computers, automated production, electronics
2000
Industry 4.0
cyber-physical systems, IoT, networking, machine learning
2010
Industry 5.0
human-robot collaboration, cognitive systems, customization
2020



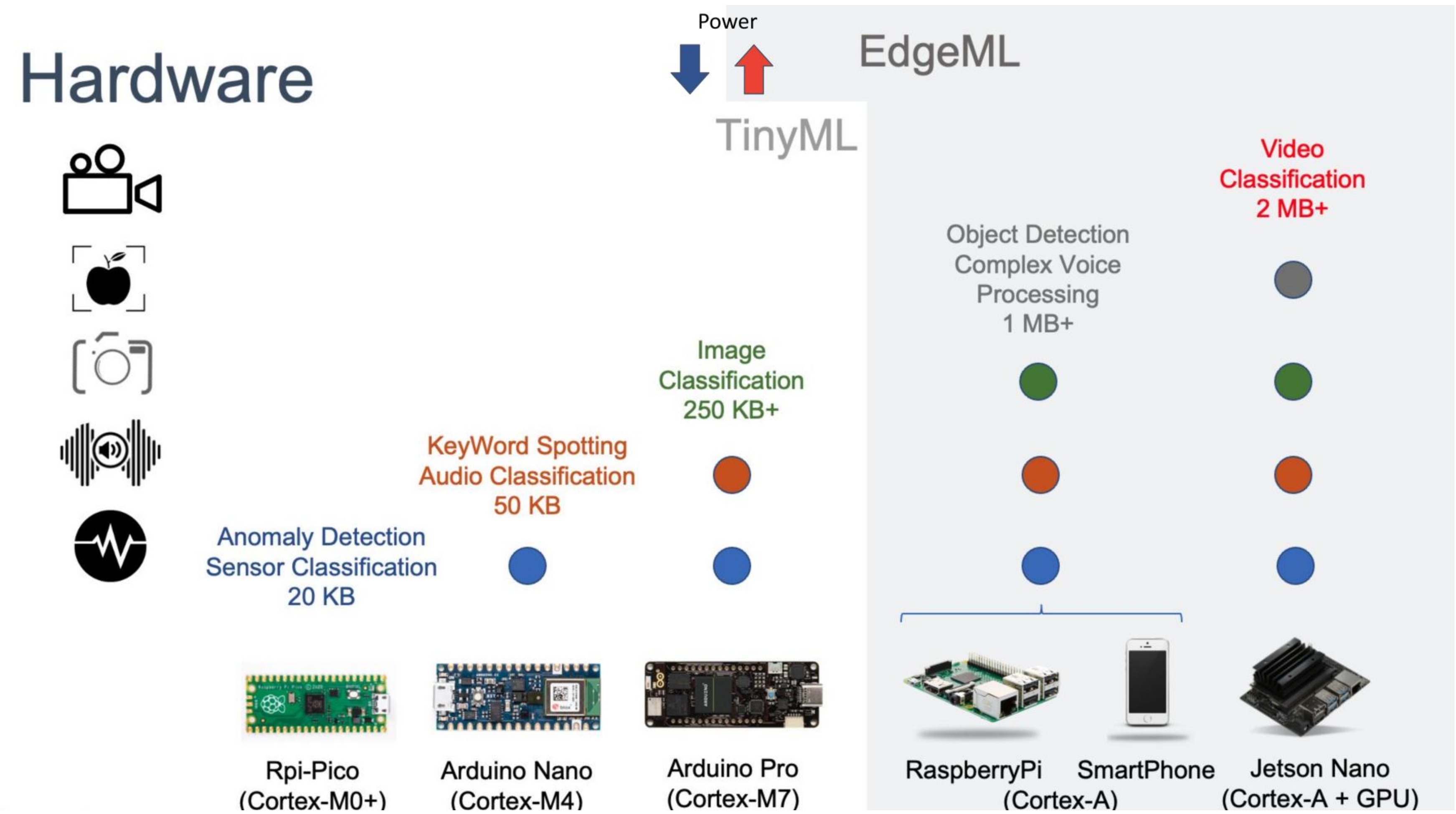
AI
Industrial IoT
Smart Factory
Smart Healthcare
Drones
Security & Surveillance
Autonomous Vehicles

TinyML


Machine Learning
TInyML
Embedded System

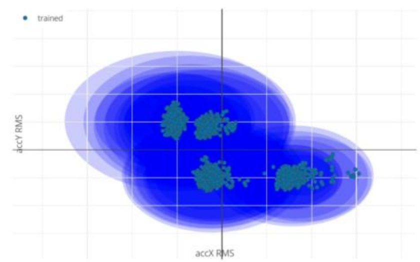

TinyML
Fastest-growing field of ML
Algorithms, hardware, software
On-device sensor analytics
Low power consumption
Always-on ML
Battery-operated

TinyML vs EdgeML



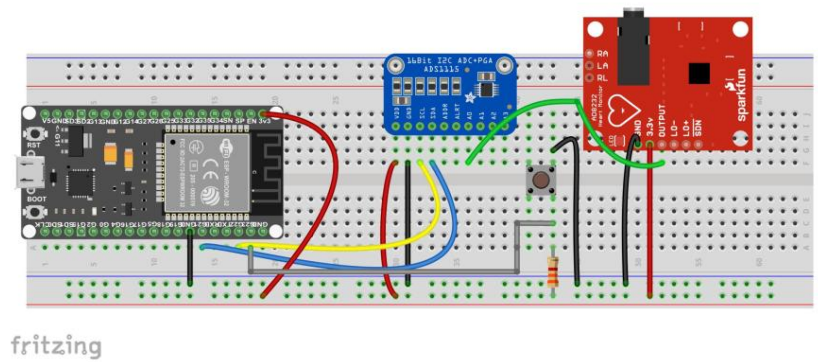
TinyML Examples

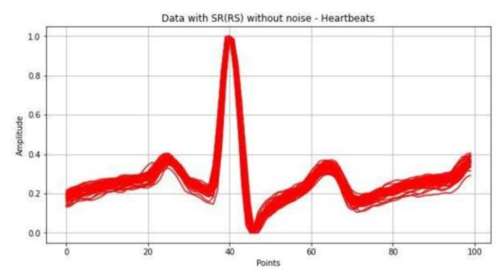
Predictive Maintenance



TinyML Examples

Health- Human Sensing

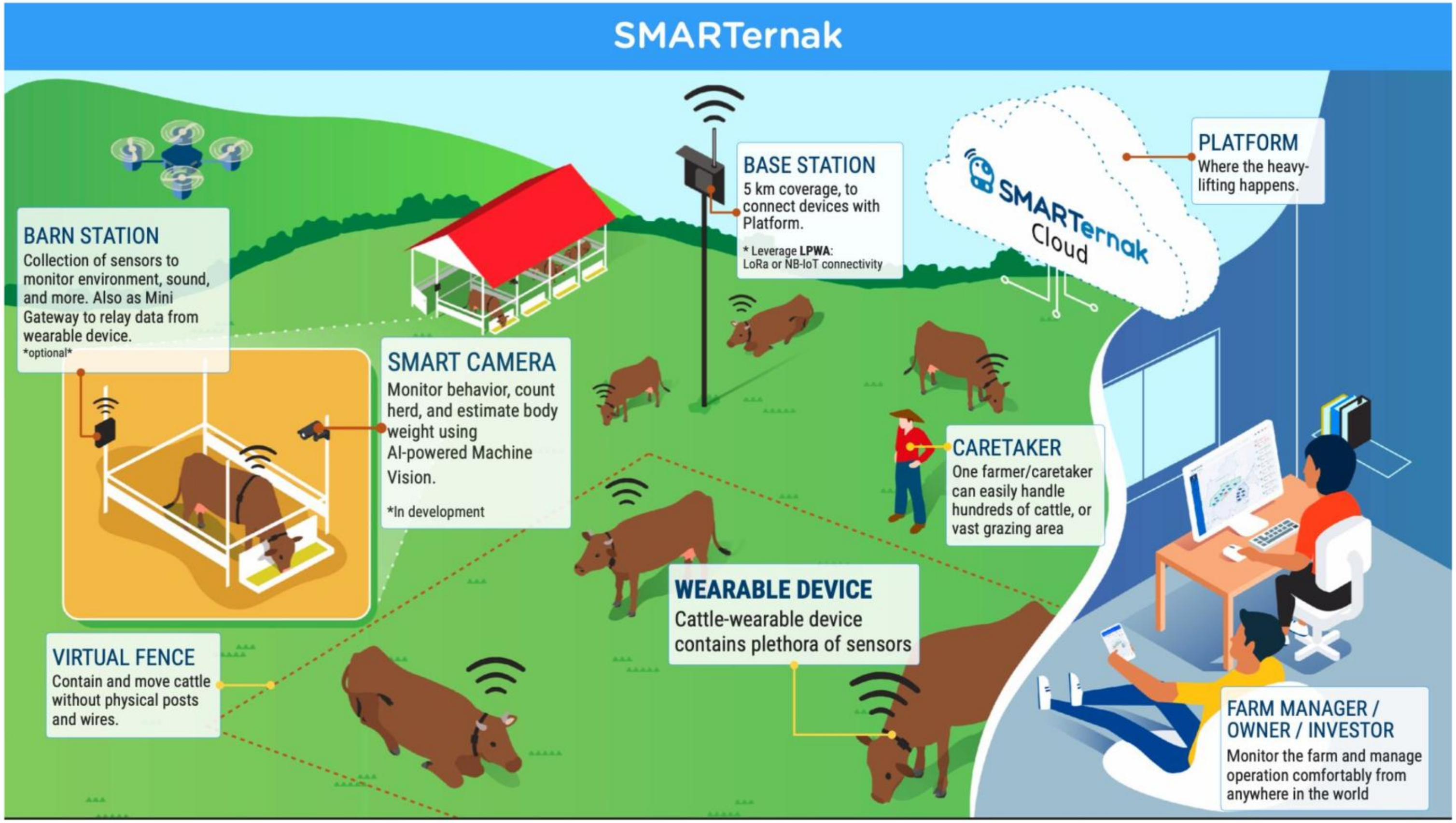




Guilherme Silva
Engenheiro - UNIFEI

TinyML Examples

Agriculture– Smart Farm



TinyML Examples

More than just voice

- Security (Broken Glass / Keyboard)
- Industry (Anomaly Detection)
- Medical (Snore, Toss)
- Nature (Bee*, Mosquito sound)

* Smart Beehive monitoring systems

TinyML Examples

Cow Monitoring

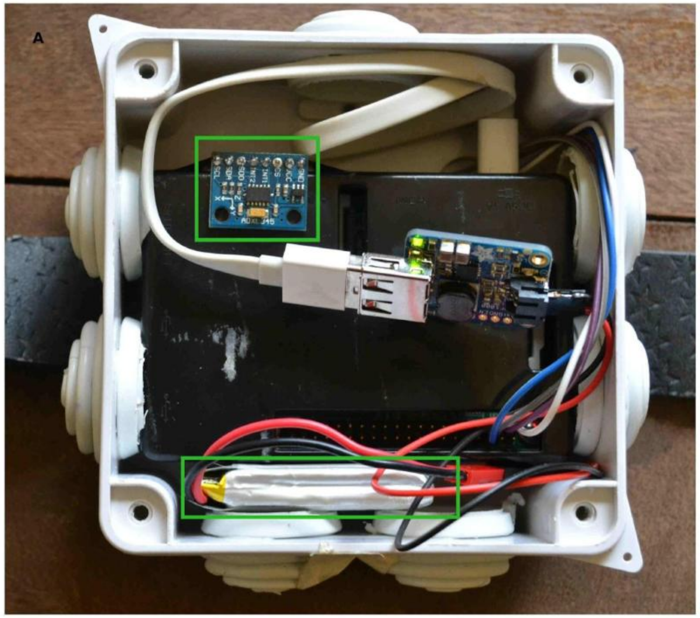
Using the Internet of Things for Agricultural Monitoring

"We aim to deploy a variety of sensors for agricultural monitoring. One of the projects involves using accelerometer sensors to monitor activity levels in dairy cows with a view to determining when the cows are on heat or when they are sick."

Ciira wa Maina, Ph.D.

Senior Lecturer
Department of Electrical and Electronic Engineering
Dedan Kimathi University of Technology
Nyeri Kenya
Email: ciira.maina@dkut.ac.ke

Kenia



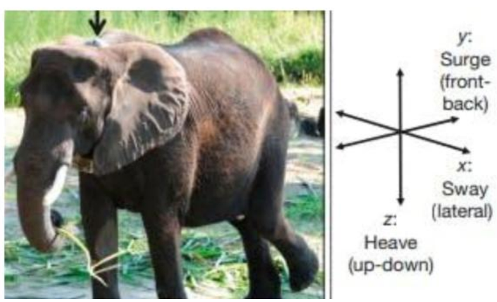
TinyML Examples

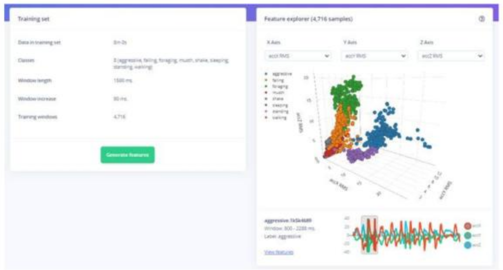
Predict and classify common Elephant behavior

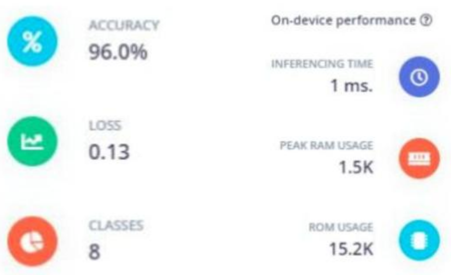
Aggressive

Standing

Sleeping

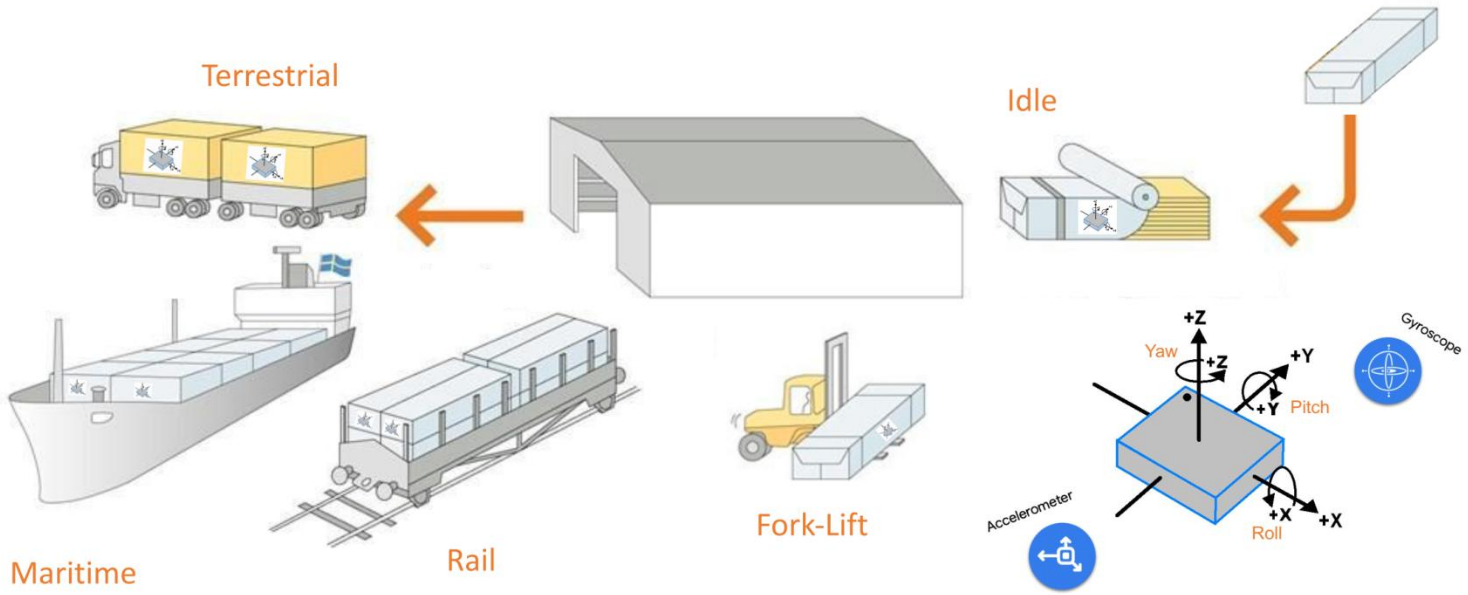






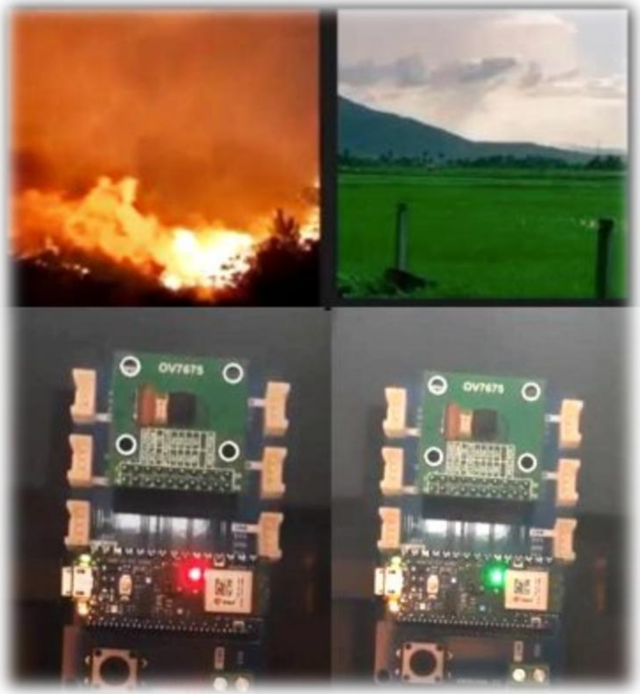
TinyML Examples

Mechanical Stresses in Transport



TinyML Examples

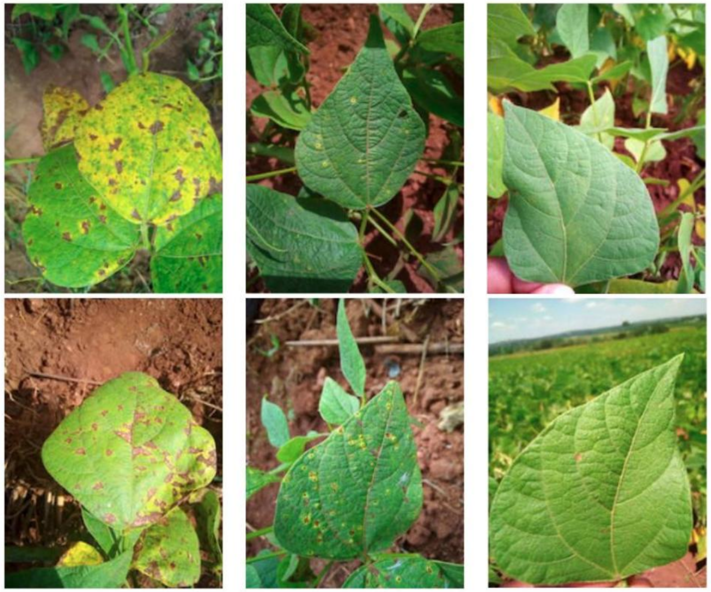
Forest Fire Detection



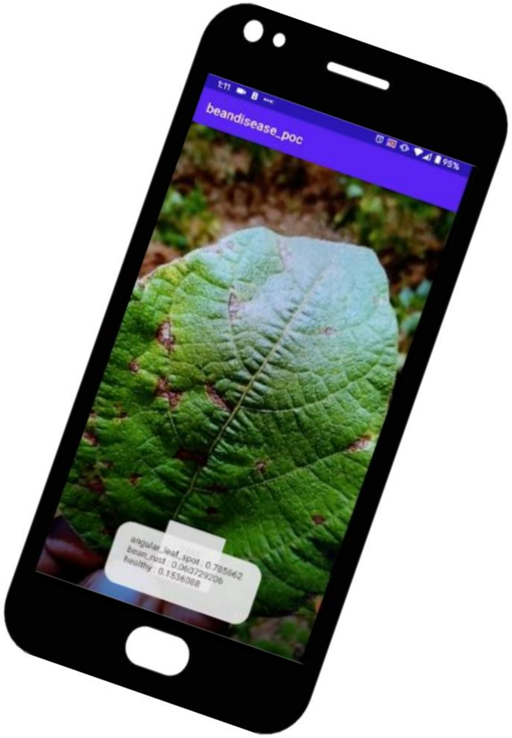
TinyML Examples

Detecting Diseases in the Bean plants

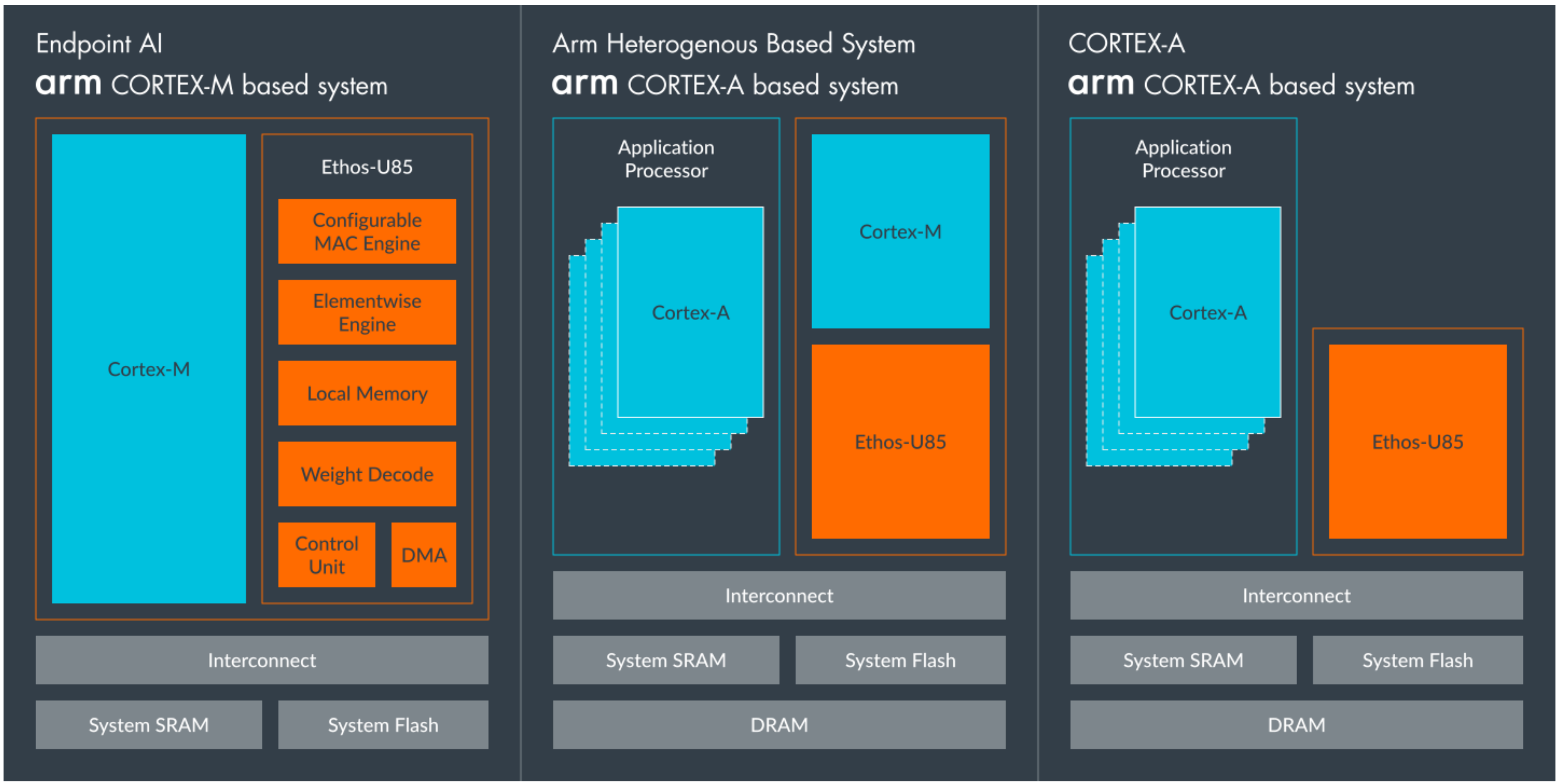
AIR Lab Makerere University
UGANDA



Angular Leaf Spot | Bean Rust | Healthy



TinyML Key Players



TinyML Key Players

Arm Kleidi - AI Libraries for ML and CV (May 2024)

- Microkernels optimized for ARM processor cores
- Ease of adoption into C or C++ machine learning (ML) and AI frameworks
- No dependencies on
  - External libraries
  - Dynamic memory allocation
  - Memory management

TinyML Key Players

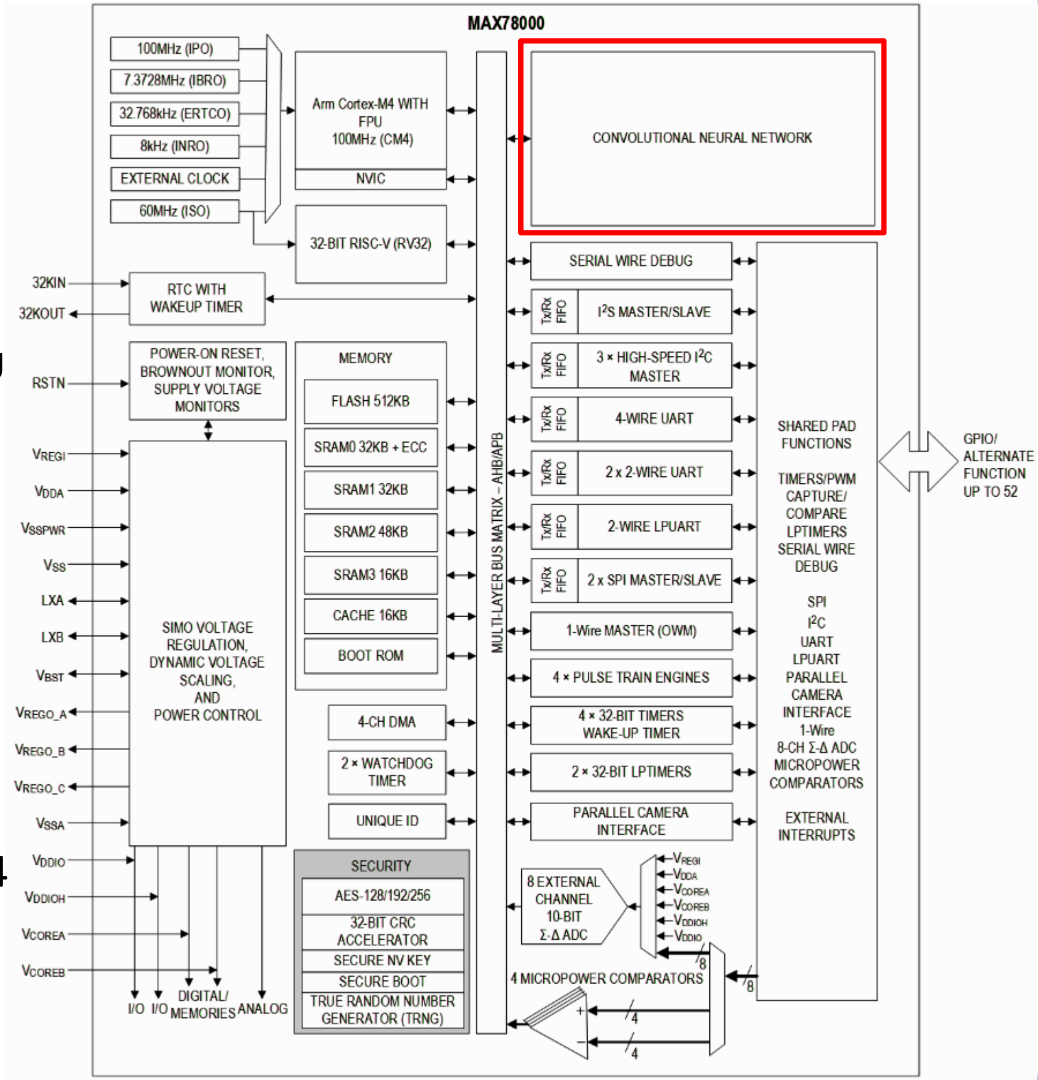
RISC-V is revolutionizing AI development

- Flexible and open Instruction Set Architecture (ISA)
- Seamlessly integrating software and hardware
- Standard Instruction Profiles
- Custom Instruction Capabilities

TinyML Key Players

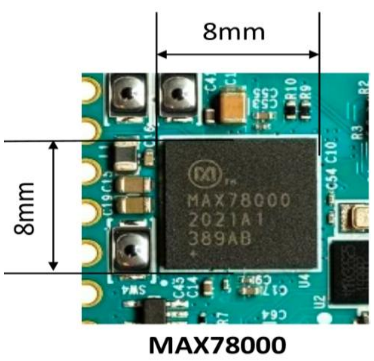
ANALOG DEVICES

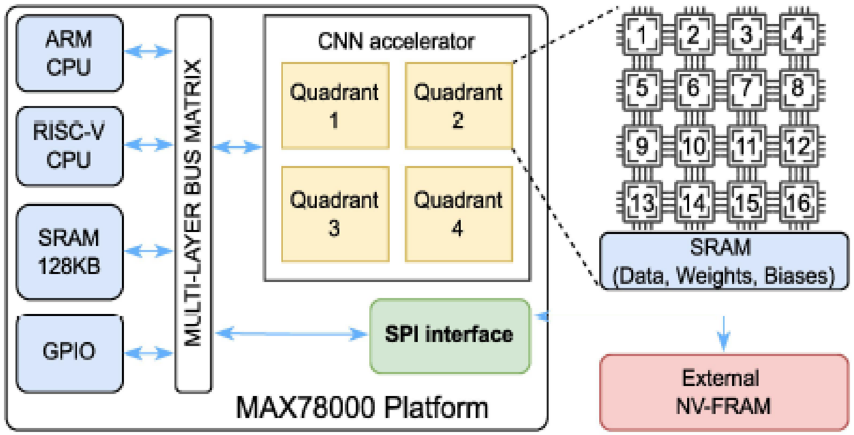
- Dual-Core Ultra-Low-Power MCU
  - Arm Cortex-M4 Processor with FPU
  - 32-Bit RISC-V Coprocessor
- Neural Network Accelerator
- 442k 8-Bit Weight Capacity
- Image Size up to 1024 x 1024 Pixels
- Network Depth up to 64 Layers
- Network Channel Widths up to 1024



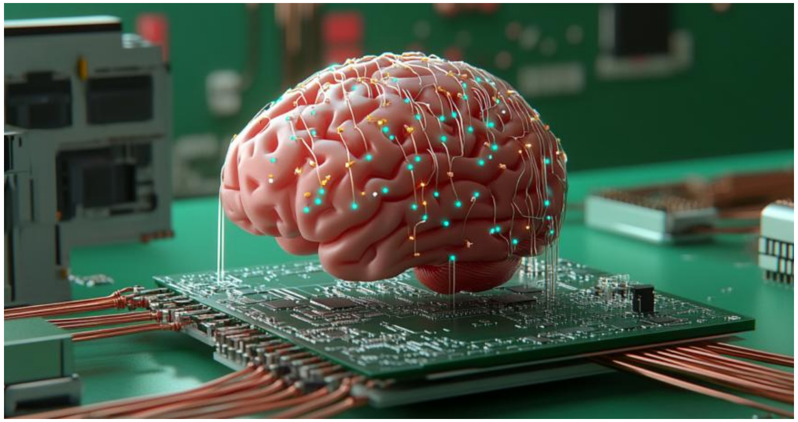
TinyML Key Players

ANALOG DEVICES





Neuromorphic Computing



AI past trend

- Feeding models more data
- Increasing processing speeds

AI systems rely on

- Conventional computing architectures
- Sequential data processing
- Consuming massive amounts of energy

Neuromorphic computing

- Inspired by biological neurons
- Enables AI to learn dynamically
- Process data using event-driven architectures

Neuromorphic Computing

Event-driven computation

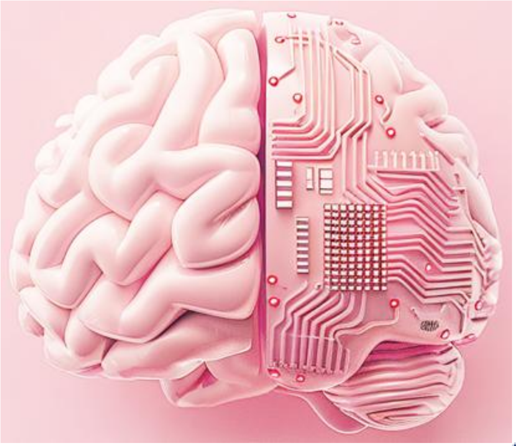
- Spiking neural networks (SNNs)
- Mimics how biological neurons communicate
- Discrete electrical spikes

Integrated Memory and Processing

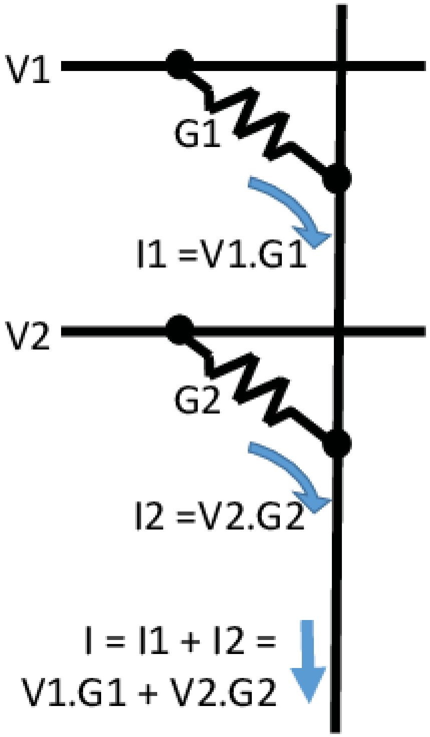
- Merge computation and memory within the same architecture
- Eliminating the von Neumann bottleneck
- Minimizes data transfer delays and power consumption

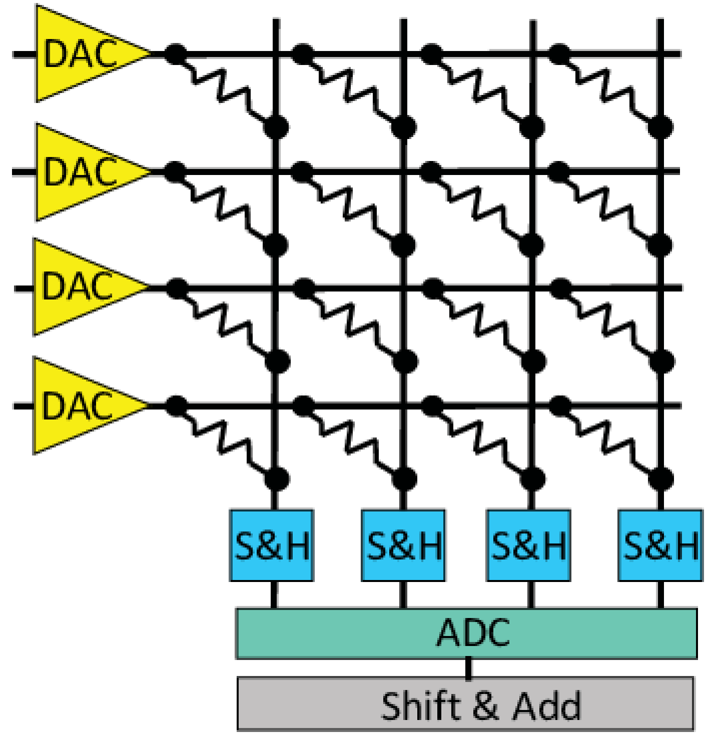
Analog Computing

- Stanford's Neurogrid system
- Energy-efficient drone navigation
- 1/10,000th the power of traditional GPUs



Neuromorphic Computing





$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \end{pmatrix} \times \begin{pmatrix} y_{11} & y_{12} & y_{13} & y_{14} \\ y_{21} & y_{22} & y_{23} & y_{24} \\ y_{31} & y_{32} & y_{33} & y_{34} \\ y_{41} & y_{42} & y_{43} & y_{44} \end{pmatrix}$$

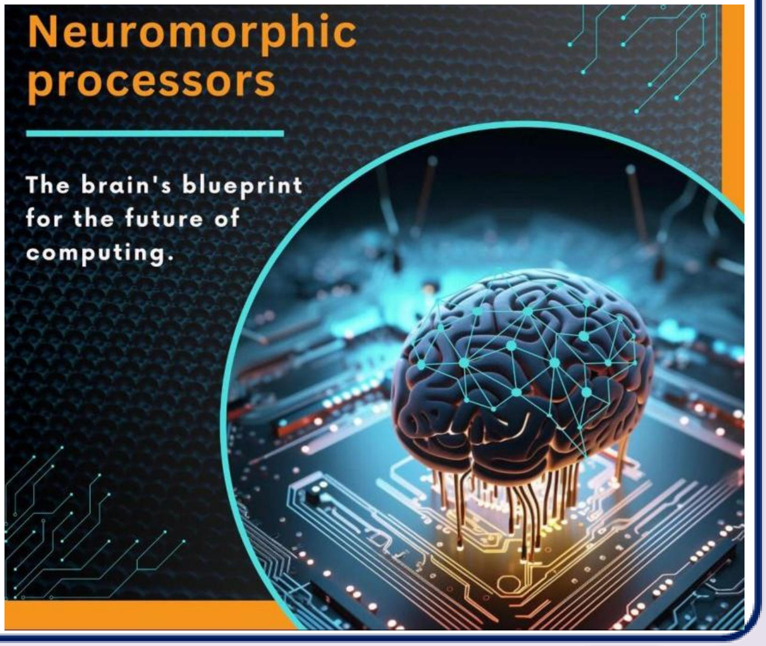
Neuromorphic Chips Advantages

Energy Efficiency

- Consume only 1% to 10% of the power
- Due to
  - Event-Driven Processing
  - Integrated Memory and Compute
- IBM’s TrueNorth: 98% energy consumption reduction

Real-Time Processing

- Asynchronous Operations
- Sony’s neuromorphic chips detect pedestrians 20ms faster



Scalability

- Compact Design
- Optimized for Sparse Data
- Adaptive Learning

Industries Benefiting from Neuromorphic Edge AI

Smarter Consumer Gadgets

- **Smartphones:** Neuromorphic vision sensors
- **Wearables**
  - BrainChip's Akida: fitness trackers and medical wearables
    - Real-time analysis of ECG, glucose levels, and sleep patterns
    - Extending battery life by 10–100x
- **Smart Home Devices**
  - Local processing of voice commands and gestures
  - Reducing cloud dependency and latency
  - Increasing Privacy

Industries Benefiting from Neuromorphic Edge AI

Industrial IoT (IIoT)

- **Predictive Maintenance:** Accenture Labs' neuromorphic systems
  - Vibration and thermal data to detect machinery anomalies

- **Robotic Automation:** SynSense's Speck chip
  - Robots to mimic human movements with sub-millisecond latency
- **Anomaly Detection**
  - Analog neuromorphic circuits
    Sparse sensor data processing
    in noisy industrial environments

Industries Benefiting from Neuromorphic Edge AI

Autonomous Systems

- **Autonomous Vehicles:** Prophesee's Event-Based Metavision sensors
  - Process LIDAR and camera inputs at 0.1ms latency, enabling collision avoidance without cloud reliance
- **Drones:** SynSense's neuromorphic processors
  - Enable drones to navigate complex environments autonomously
- **Vision-Based Robotics:** Intel's Loihi chip powers warehouse robots
  - Reducing energy consumption by 40%

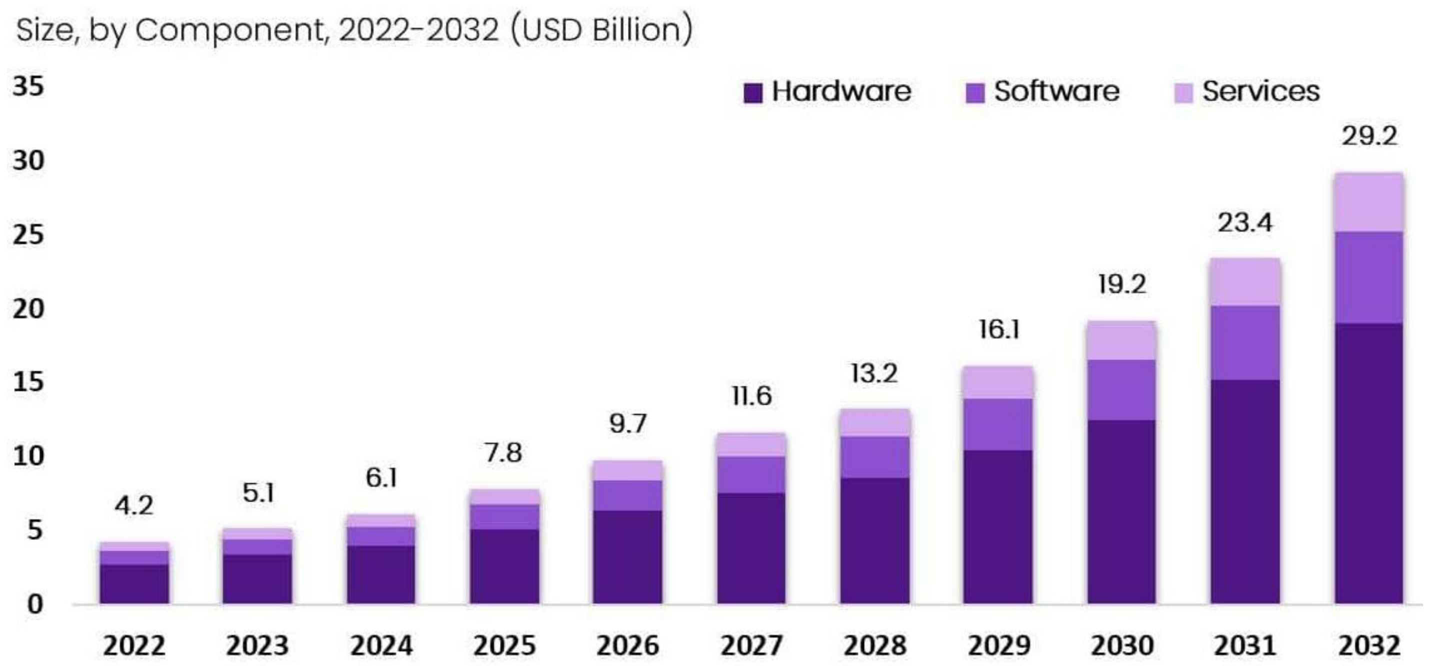
Industries Benefiting from Neuromorphic Edge AI

Healthcare AI

- **AI-Powered Diagnostics:** Neuromorphica's medical devices
  - detecting seizures and arrhythmias with 99% accuracy

- **Medical Wearables:** Stanford's Neurogrid
  - Continuous glucose monitoring with a 30-day battery life

Key Players and Market Availability Timeline



SAMSUNG

IBM

intel

Qualcomm

SONY

Neuromorphic Chips Examples

IBM’s TrueNorth chip (2014)

- 4096 neurosynaptic cores
- 1 million neurons and 256 million synapses
- 5.4 billion transistor
- 46 billion synaptic operations per second
- mimics how biological neurons communicate
- discrete electrical spikes

IBM’s NorthPole chip (2023)

- roughly 4,000 times faster than TrueNorth
- 22 billion transistors in 800 square millimeters
- 256 cores and can perform 2,048 operations per core per cycle
- 25 times more energy efficient than 12-nm GPUs and 14-nm CPUs
- process data using event-driven architectures

Neuromorphic Chips Examples

Intel’s Loihi 2 chip

Resources/Features	Loihi	Loihi 2
Process	Intel 14nm	Intel 4
Die Area	60 mm^2	31 mm^2
Core Area	0.41 mm^2	0.21 mm^2
Transistors	2.1 billion	2.3 billion
Max # Neuron Cores/Chip	128	128
Max # Processors/Chip	3	6
Max # Neurons/Chip	128,000	1 million
Max # Synapses/Chip	128 million	120 million
Memory/Neuron Core	208 KB, fixed allocation	192 KB, flexible allocation
Neuron Models	Generalized LIF	Fully programmable
Neuron State Allocation	Fixed at 24 bytes per neuron	Variable from 0 to 4096 per neuron depending on neuron model requirements

Challenges in Adopting Neuromorphic AI

Lack of Industry-Wide Standardization

- Programming Interfaces
    - PyNN and Intel’s Lava

- **Communication Protocols**
    - Vendor-specific implementations hinder seamless interoperability

- **Host Dependency**
    - Reliance on conventional computers for pre- and post-processing

Challenges in Adopting Neuromorphic AI

Immature Software Ecosystem

- **Framework Compatibility**
  - TensorFlow and PyTorch lack native support for SNNs
- **Algorithm Development**
  - SNN training methods lack standardization
- **Toolchain Limitations**
  - Neuromorphic-specific compilers and debuggers

Challenges in Adopting Neuromorphic AI

Scalability Concerns

- **Physical Scaling**
  - Loihi and TrueNorth, support only ~1 million neurons
    - Far below the 86 billion in biological brains
    - Advancements in 3D integration and high-density memristor technology
- **Material Challenges**
  - Trade-offs between endurance and energy efficiency
- **Thermal Management**
  - significant heat under load
  - limiting device density and long-term reliability

Conclusion

- **Edge AI** – Enabling **real-time intelligence**
- **Decentralized intelligence** and **efficient AI** beyond traditional cloud computing
- **TinyML**– Bringing **machine learning** to ultra-low-power devices
- **Revolutionizing** computing with **brain-inspired architectures**
  - **Neuromorphic Processors**
- **Final Thought:** *"The future belongs to intelligent systems that think, learn, and adapt—right at the edge!"*

Thanks for your attention