

Latam Regional Workshop
on SciTinyML:
Scientific Use of
Machine Learning on
Low-Power Devices



11-15 July 2022
Online

Further information:
<https://tinyMLedu.org/SciTinyML>
edu@tinyML.org

The Future of Machine Learning is Tiny and Bright

Brian Plancher
Barnard College, Columbia University
Harvard John A. Paulson School of Engineering and Applied Sciences
brianplancher.com



Quick Disclaimer:

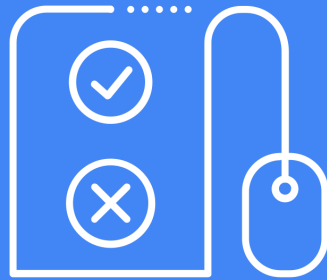
Today will be **both too fast**
and **too slow!**

What Is Embedded ML? (TinyML)

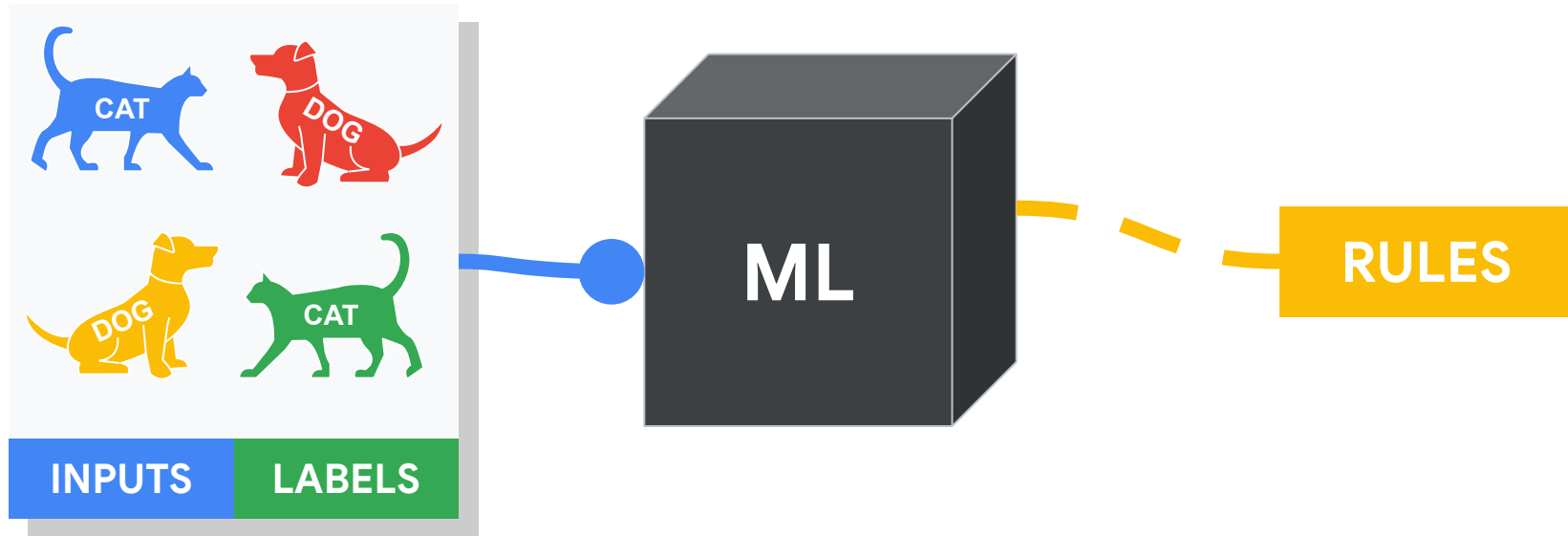
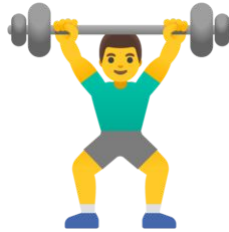
What Is ML?

Artificial Intelligence

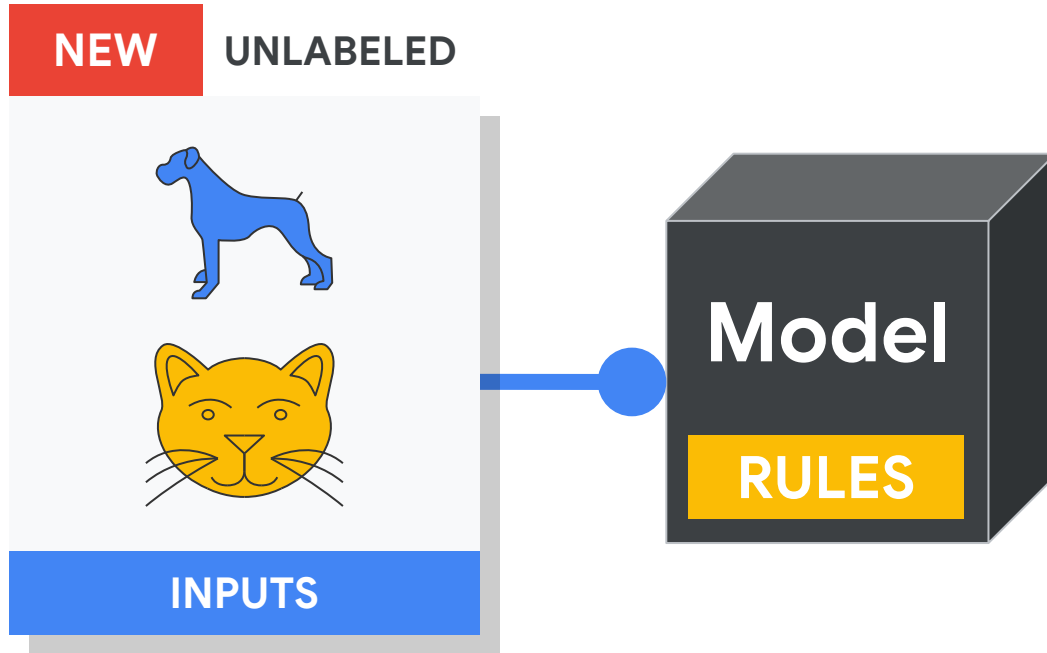
Machine Learning



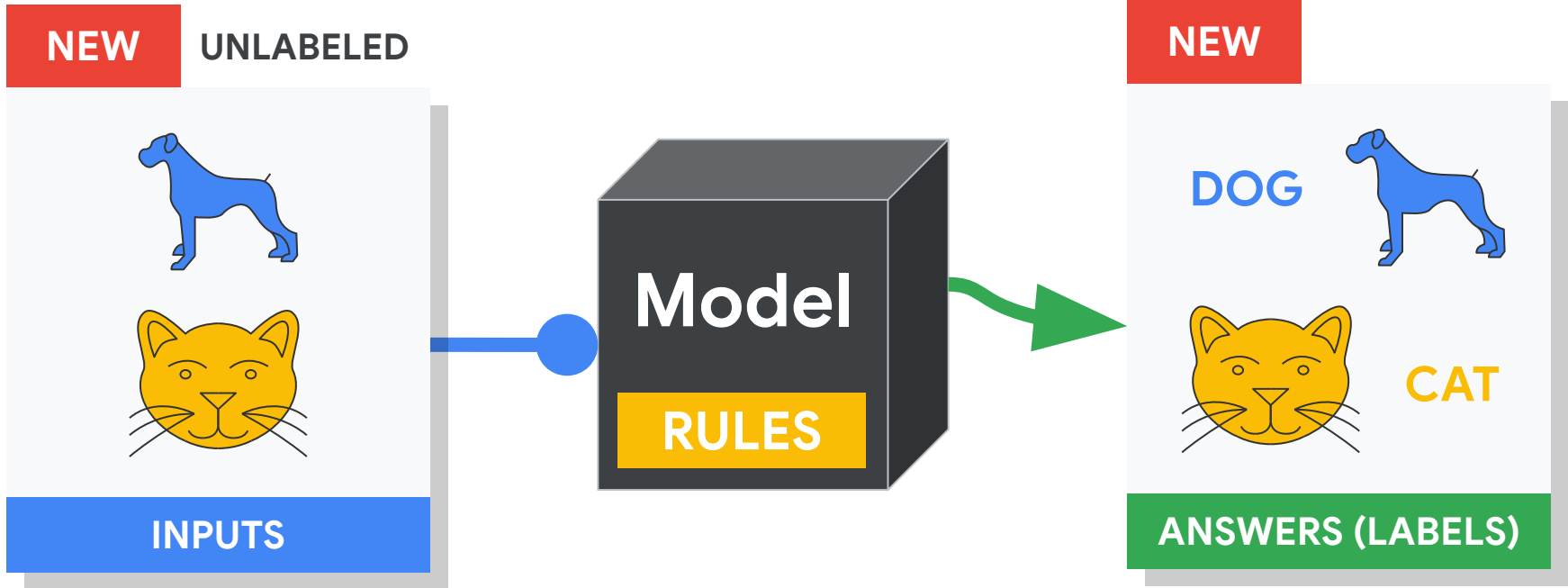
ML works by Training a model



After it's learned...



After it's learned you can make predictions:



What Is Embedded ML? (TinyML)

What is Tiny Machine Learning (**TinyML**)?

TinyML



Fastest-growing field of **ML**



What is Tiny Machine Learning (**TinyML**)?

TinyML

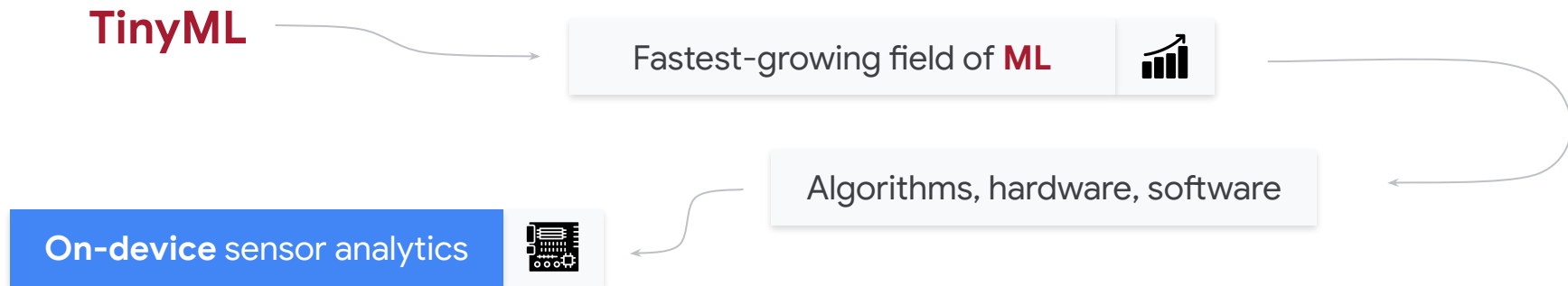


Fastest-growing field of **ML**

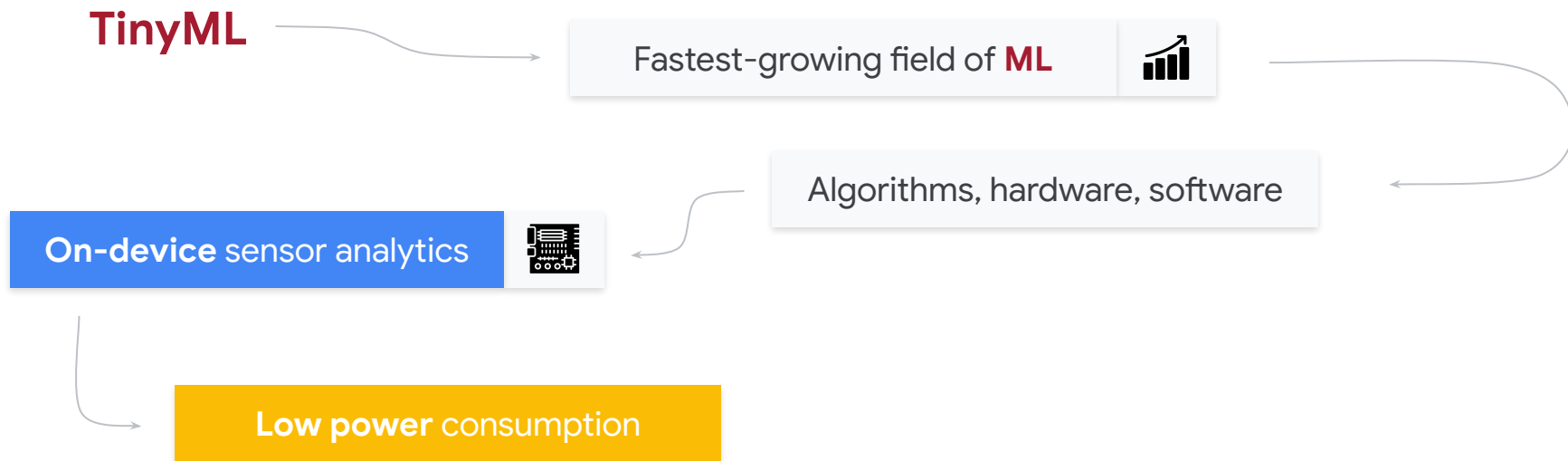


Algorithms, hardware, software

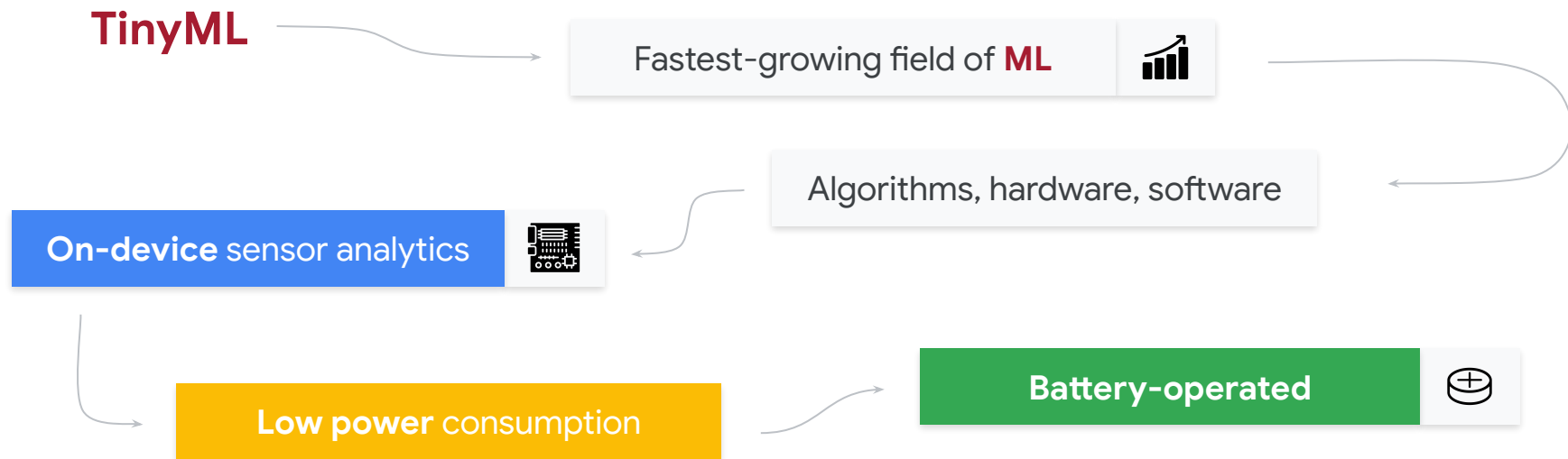
What is Tiny Machine Learning (**TinyML**)?



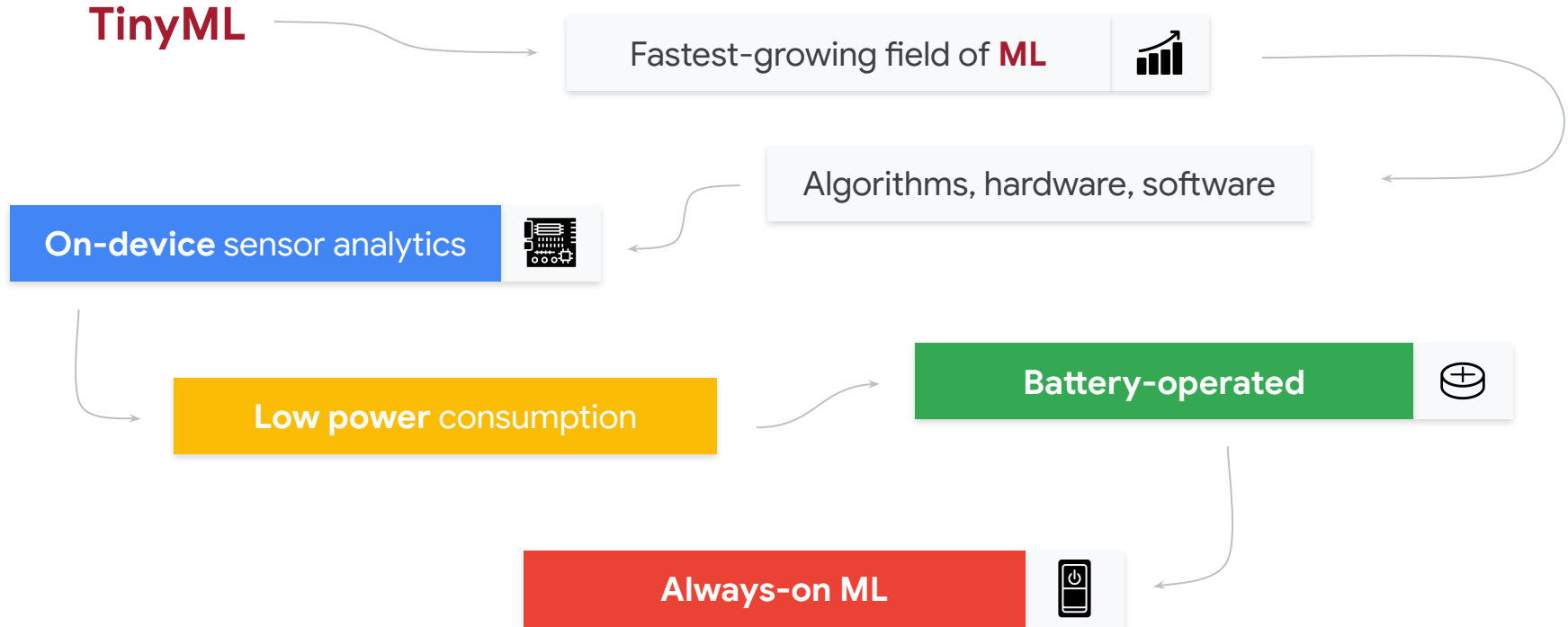
What is Tiny Machine Learning (**TinyML**)?



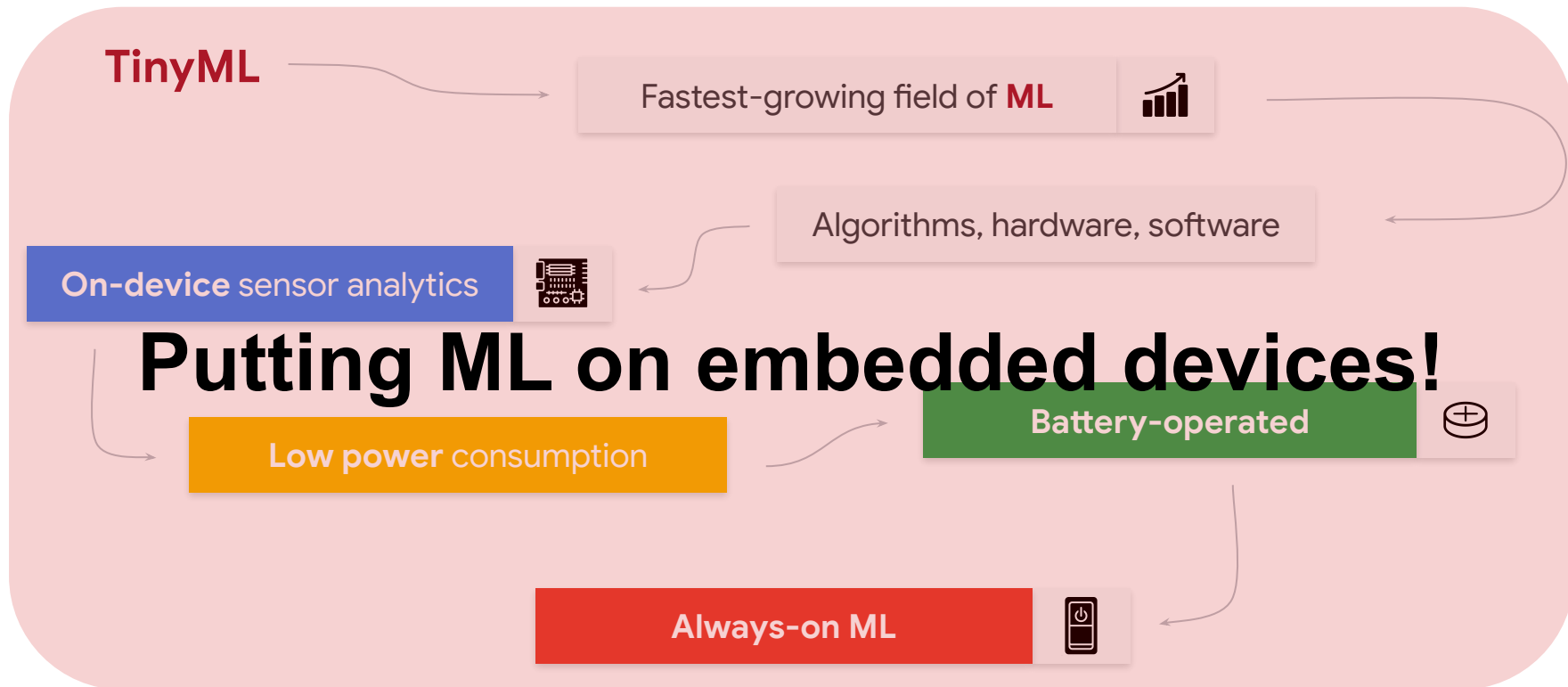
What is Tiny Machine Learning (**TinyML**)?



What is Tiny Machine Learning (**TinyML**)?



What is Tiny Machine Learning (**TinyML**)?





Kicking

Penalty kicking

Passing

Dribbling

...



Promising Social Applications of TinyML

Wildlife conservation

ElephantEdge

Building The World's Most Advanced
Wildlife Tracker.



Agriculture

May be able to reduce agrichemical use to 0.1%
of conventional blanket spraying

Technology: The Future of Agriculture

[Anthony King](#)

[Nature](#) 544, S21–S23 (2017) | [Cite this article](#)

161k Accesses | 132 Citations | 209 Altmetric | [Metrics](#)

And many more!



Tiny Robot Learning: Challenges and Directions for Machine Learning in Resource-Constrained Robots

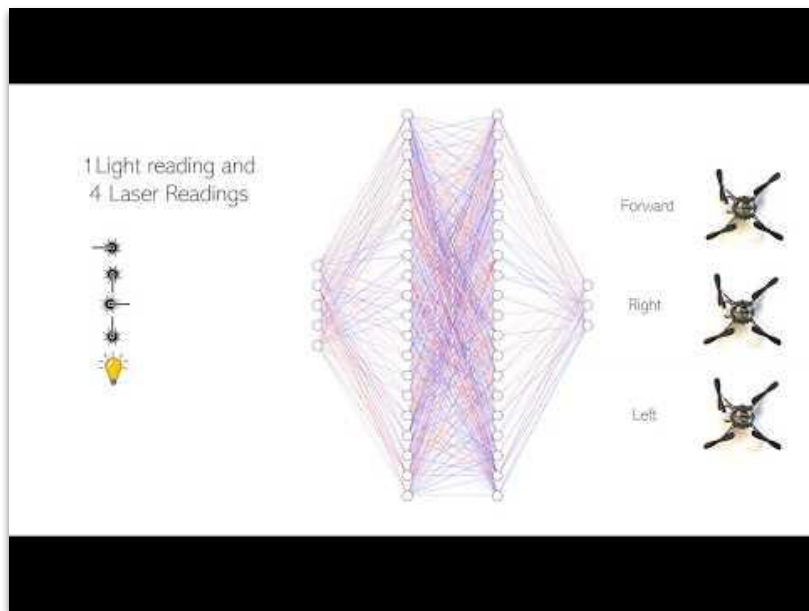
Sabrina M. Neuman¹, Brian Plancher¹, Bardienus P. Duisterhof², Srivatsan Krishnan¹, Colby Banbury¹,
Mark Mazumder¹, Shvetank Prakash¹, Jason Jabbour³, Aleksandra Faust⁴,
Guido C.H.E. de Croon⁵, and Vijay Janapa Reddi¹

Harvard University¹, CMU², University of Virginia³, Google Brain⁴, Delft University of Technology⁵

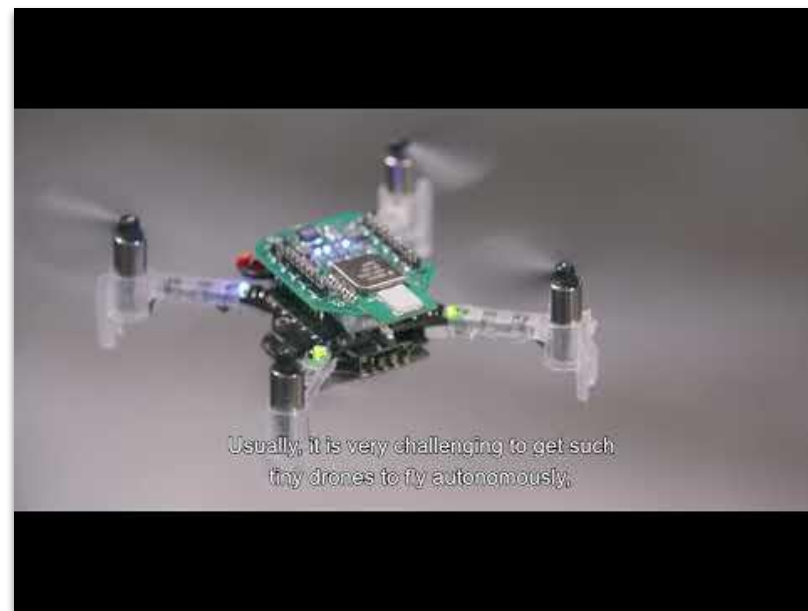
{sneuman@seas, brian_plancher@g, srivatsan@g, cbanbury@g, markmazumder@g, sprakash@g, vj@eecs}.harvard.edu,
bduister@andrew.cmu.edu, jjj4se@virginia.edu, aleksandra.faust@gmail.com, g.c.h.e.decroon@tudelft.nl

<https://arxiv.org/pdf/2205.05748.pdf>

TinyRL: Autonomous Navigation on Nano Drone



[ICRA'21]



[IROS'21]

Why Tiny?

IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things



No Good Data Left Behind

5 Quintillion

bytes of data produced
every day by IoT

<1%

of unstructured data is
analyzed or used at all

IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things

B

L

E

R

P

IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things

Bandwidth

Latency

Energy

**Side Note: I've also seen
the E represent Economics**

Reliability

Privacy

IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things

Energy



**Battery Life is
only O(months)
and only sends
GPS signal**

IoT 1.0:
Internet
of Things



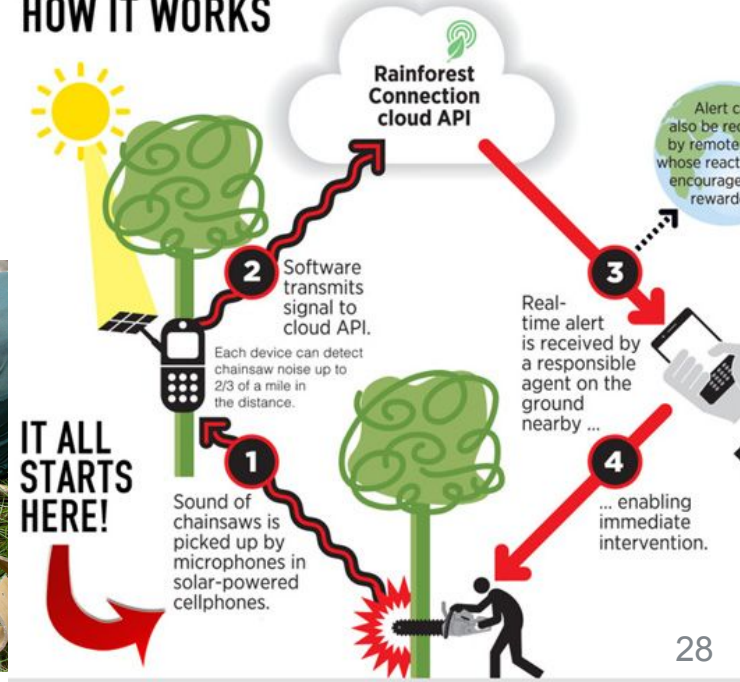
IoT 2.0:
Intelligence
on Things

Bandwidth

Energy



RAINFOREST CONNECTION: HOW IT WORKS



IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things

Bandwidth

Energy

The OpenCollar
initiative



IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things

Privacy



Google Assistant



IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things



Latency



Google Assistant



Privacy

IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things

Energy
Reliability
Privacy



IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things

Bandwidth

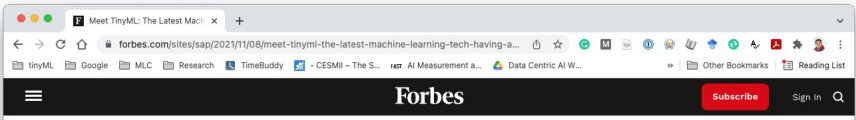
Latency

Energy

Reliability

Privacy

**TinyML to
the rescue!**



Meet TinyML: The Latest Machine Learning Tech Having An Outsize Business Impact

Dr. Nicholas Nicoloudis Brand Contributor
SAP BRANDVOICE | Paid Program
Innovation

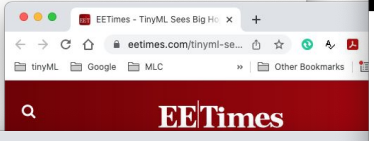
As device sensors proliferate across product development through insurmountable data, there are sound economic reasons why researchers predict IoT will have a trillion by 2025, identifying manufacturing (trillion).



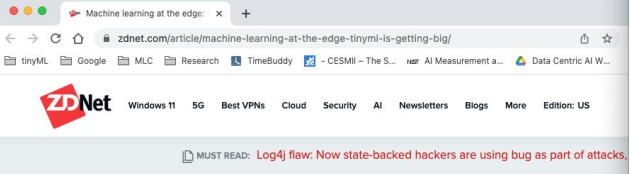
The rise of tinyML to collect data from edge devices is exploding the number of sensors in pretty much every industry.

The tinyML community was established to share learning architectures, techniques, and on-device analytics for a variety of applications (chemical, and others) at low power devices. One of the tinyML founders

"...we are in the midst of the digital transformation. The ultimate benefits of extreme energy intelligence and analytics at low cost are just beginning to be realized."



Machine learning at the edge: TinyML is getting big



Machine learning at the edge: TinyML is getting big

Being able to deploy machine learning applications at the edge is the key to unlocking the full potential of IoT. TinyML is the art and science of producing machine learning models frugal enough to run on low-power devices.

Written by **George Anadiotis**, Contributing Writer
Posted in Big on Data on June 7, 2021 | Topic: Big Data

Is it **\$61 billion and 38.4% CAGR by 2028** or **\$43 billion and 37.4% CAGR by 2027**? Depends on which report outlining the growth of **edge computing** you choose to go by, but in the end it's not that different.

What matters is that **edge computing is booming**. There is growing interest by vendors, and ample coverage, for good reason. Although the definition of **what constitutes edge computing** is a bit fuzzy, the idea is simple. It's about taking compute out of the data center, and bringing it as close to where the action is as possible.

Whether it's stand-alone IoT sensors, devices of all kinds, **drones**, or **autonomous vehicles**, there's one thing in common. Increasingly, data generated at the edge are used to feed applications powered by machine learning models. There's just one problem: machine learning models were never designed to be deployed at the edge. Not until now, at least. Enter **TinyML**.



What is machine learning? Everything you need to know

Tiny machine learning (TinyML) is broadly defined as a fast growing



Home



SPONSORED

How TinyML is powering big ideas across critical industries

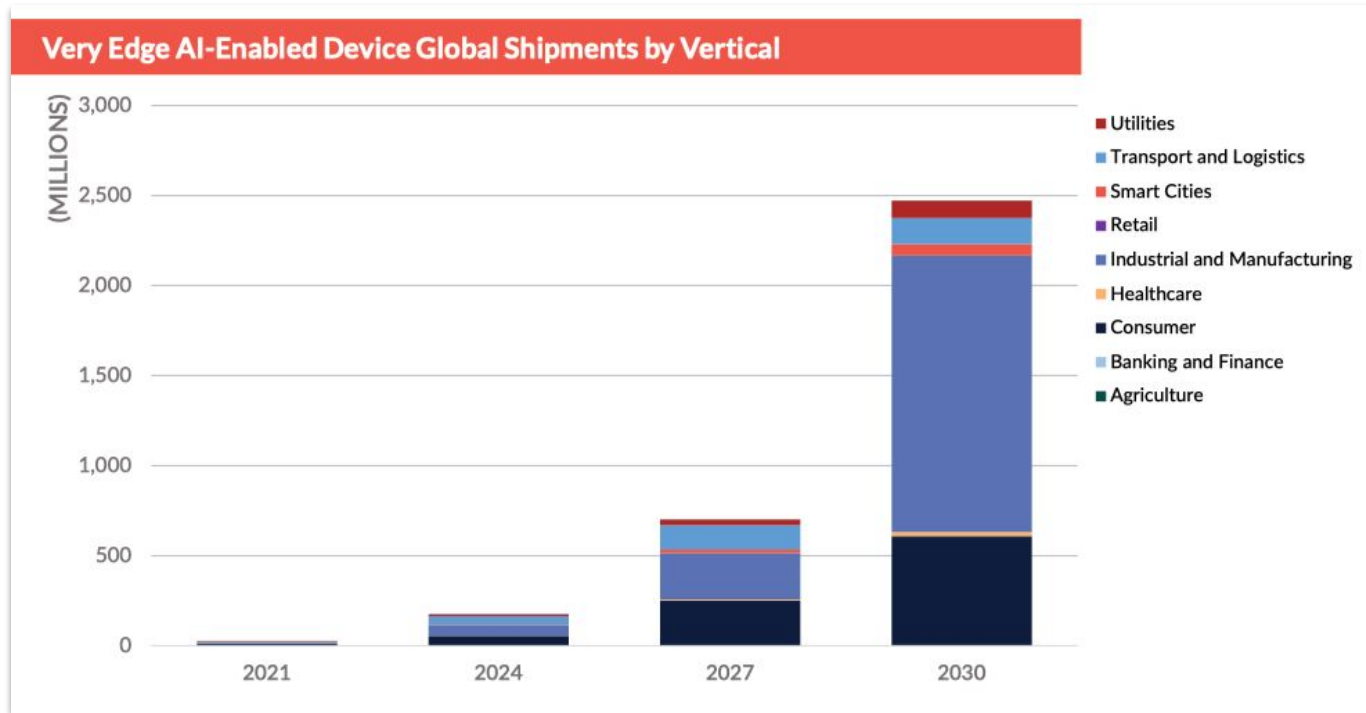
BrandPost Sponsored by SAP | Learn More | JUL 18, 2021 4:31 PM PDT



From cars and TVs to lightbulbs and doorbells. So many of the objects in everyday life have 'smart' functionality because the manufacturers have built chips into them.

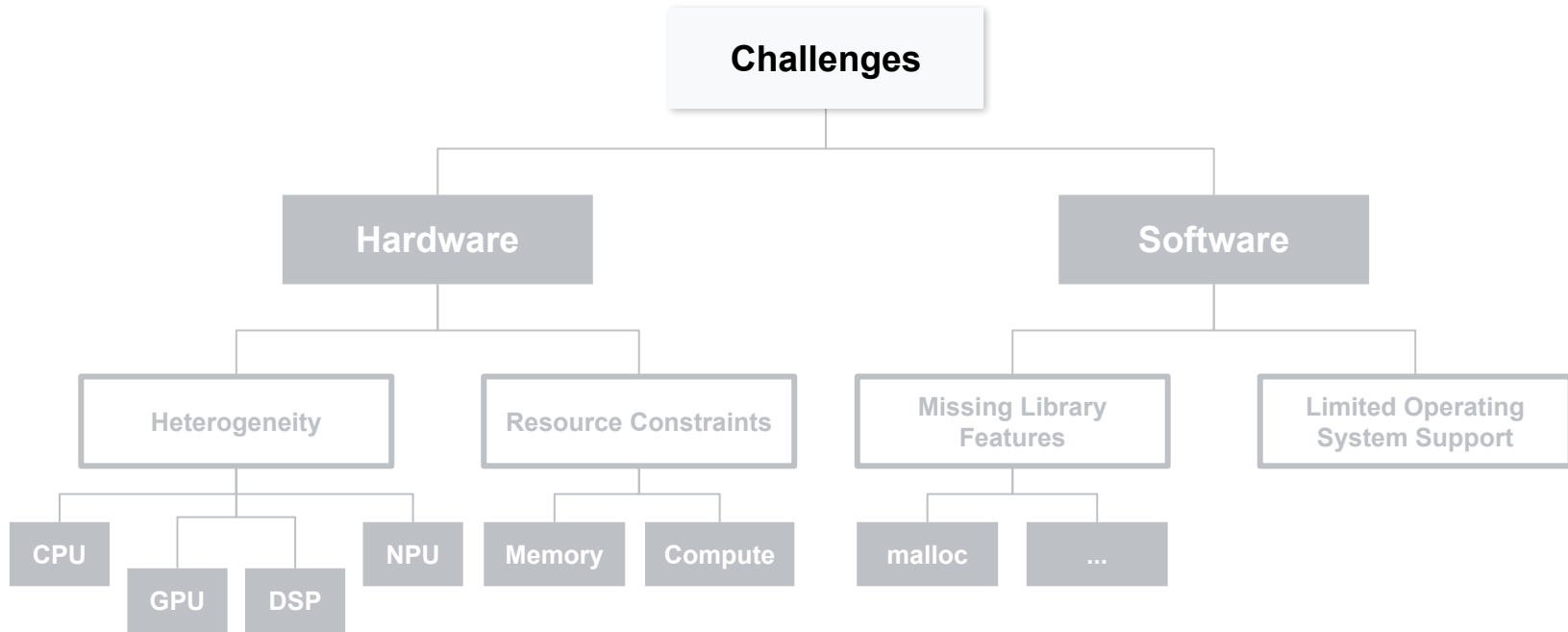
But what if you could also run machine learning models in something as small as a **golf ball dimple**? That's the reality that's being enabled by TinyML, a **broad movement** to run tiny machine learning algorithms on embedded devices, or those with

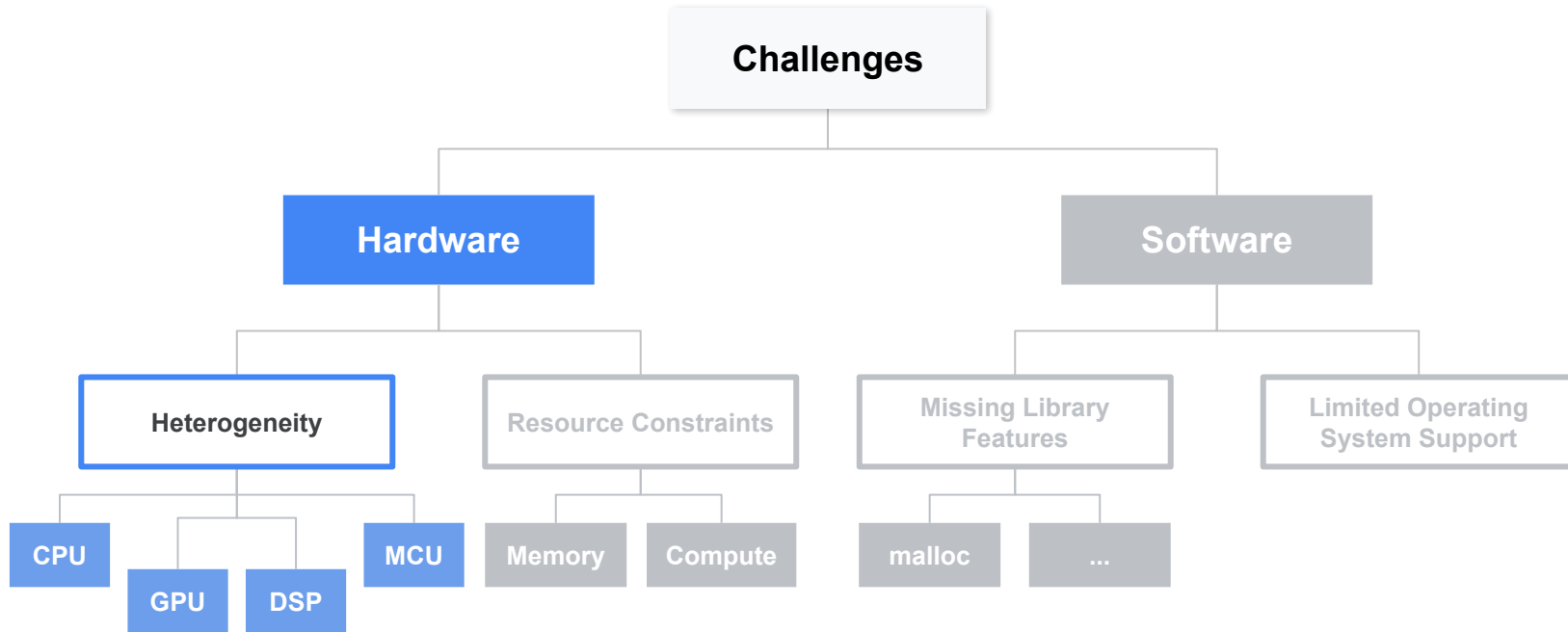
Market Forecast



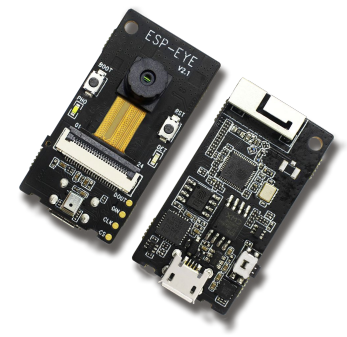
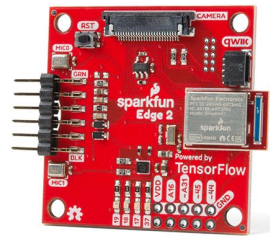
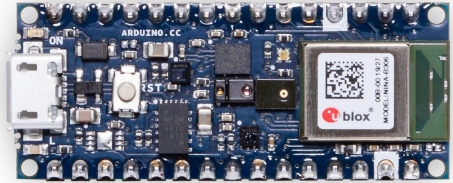
Source: ABI Research: TinyML





TinyML Challenges

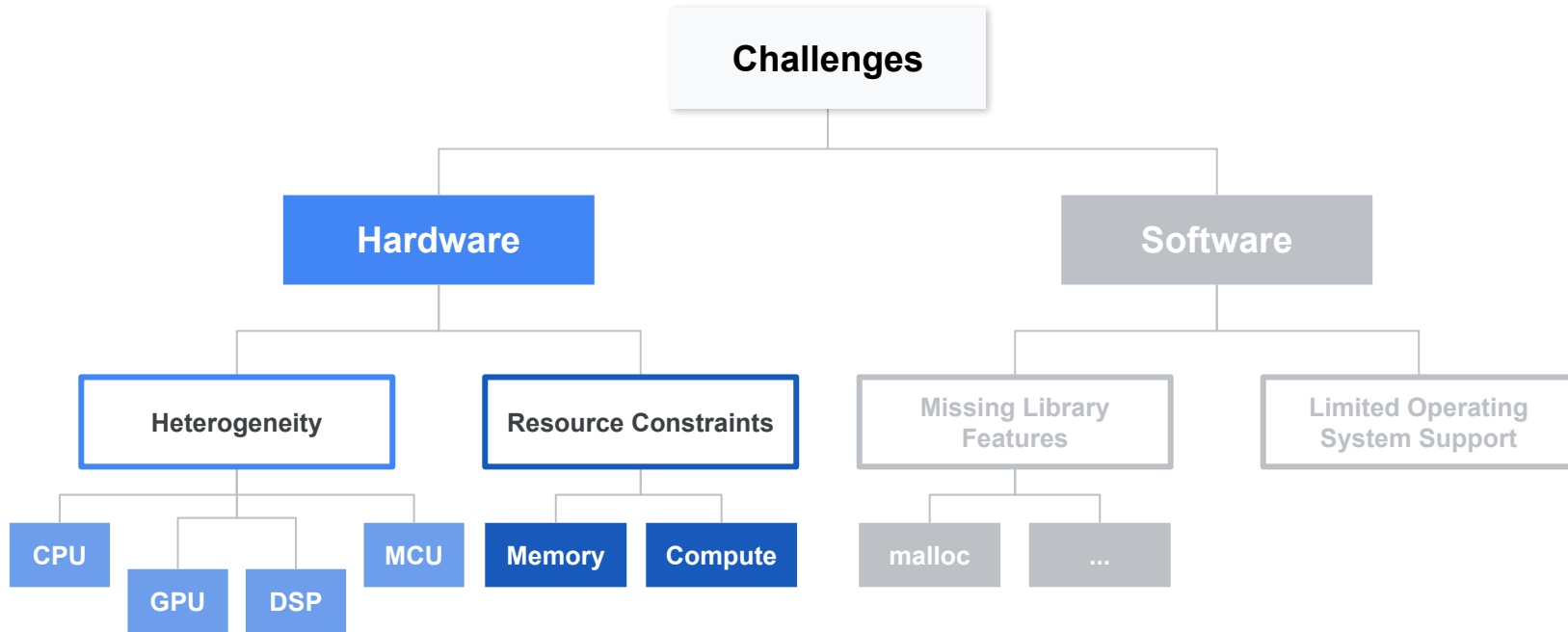




250 Billion
MCUs today

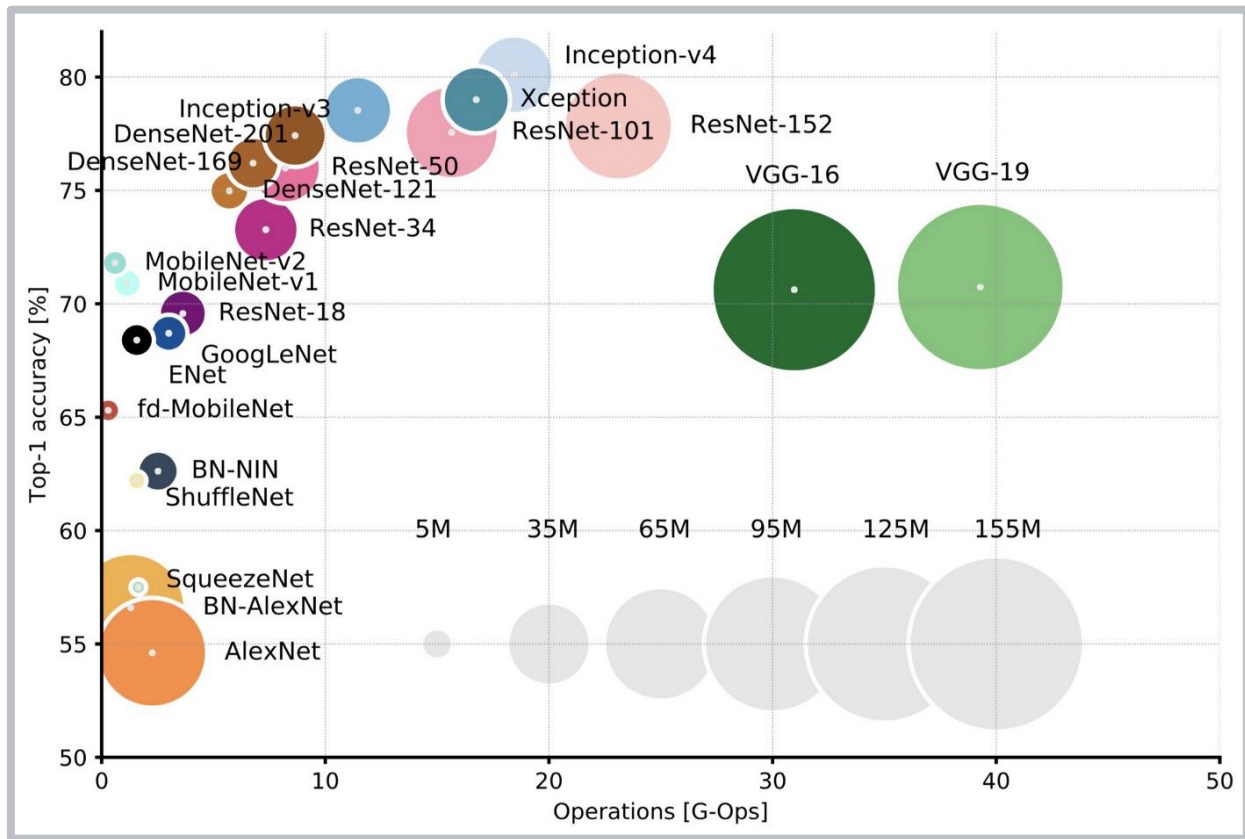


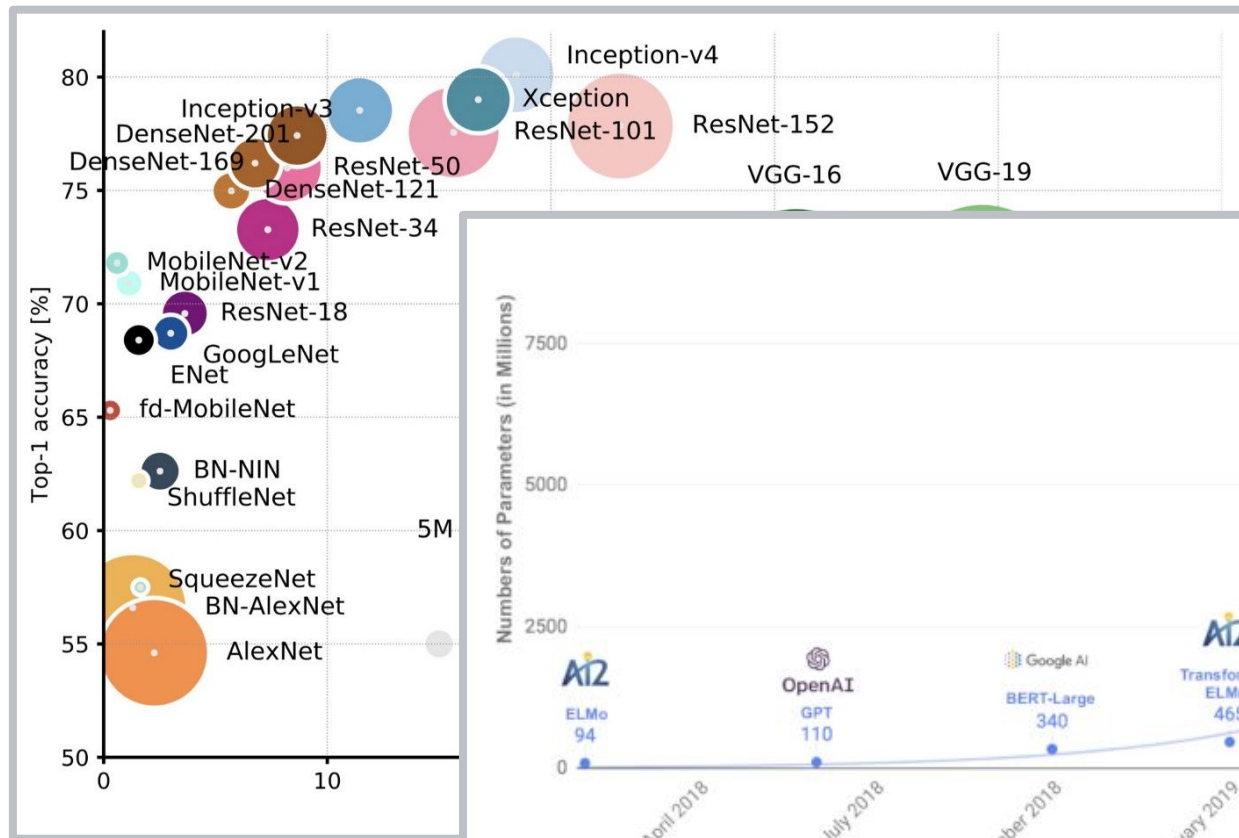
	Board	MCU / ASIC	Clock	Memory	Sensors	Radio
	Himax WE-I Plus EVB	HX6537-A 32-bit EM9D DSP	400 MHz	2MB flash 2MB RAM	Accelerometer, Mic, Camera	None
	Arduino Nano 33 BLE Sense	32-bit nRF52840	64 MHz	1MB flash 256kB RAM	Mic, IMU, Temp, Humidity, Gesture, Pressure, Proximity, Brightness, Color	BLE
	SparkFun Edge 2	32-bit ArtemisV1	48 MHz	1MB flash 384kB RAM	Accelerometer, Mic, Camera	BLE
	Espressif EYE	32-bit ESP32-D0WD	240 MHz	4MB flash 520kB RAM	Mic, Camera	WiFi, BLE

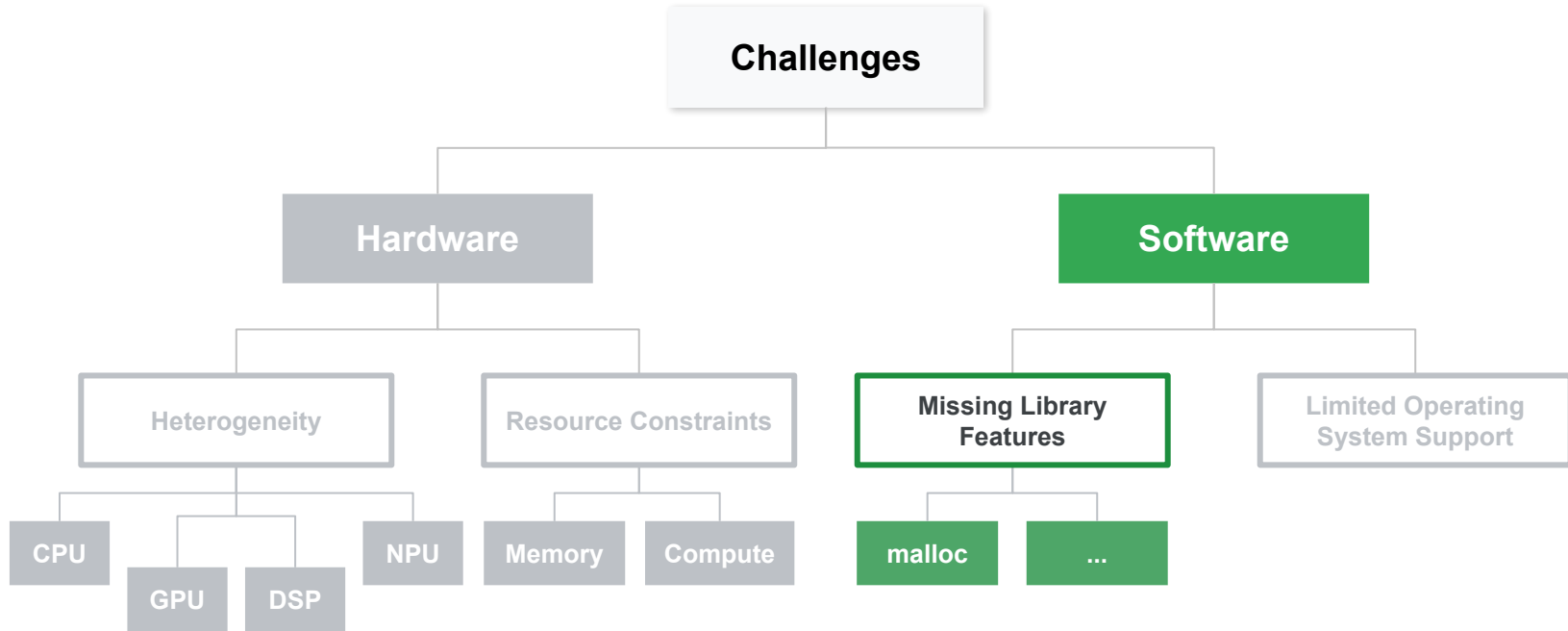


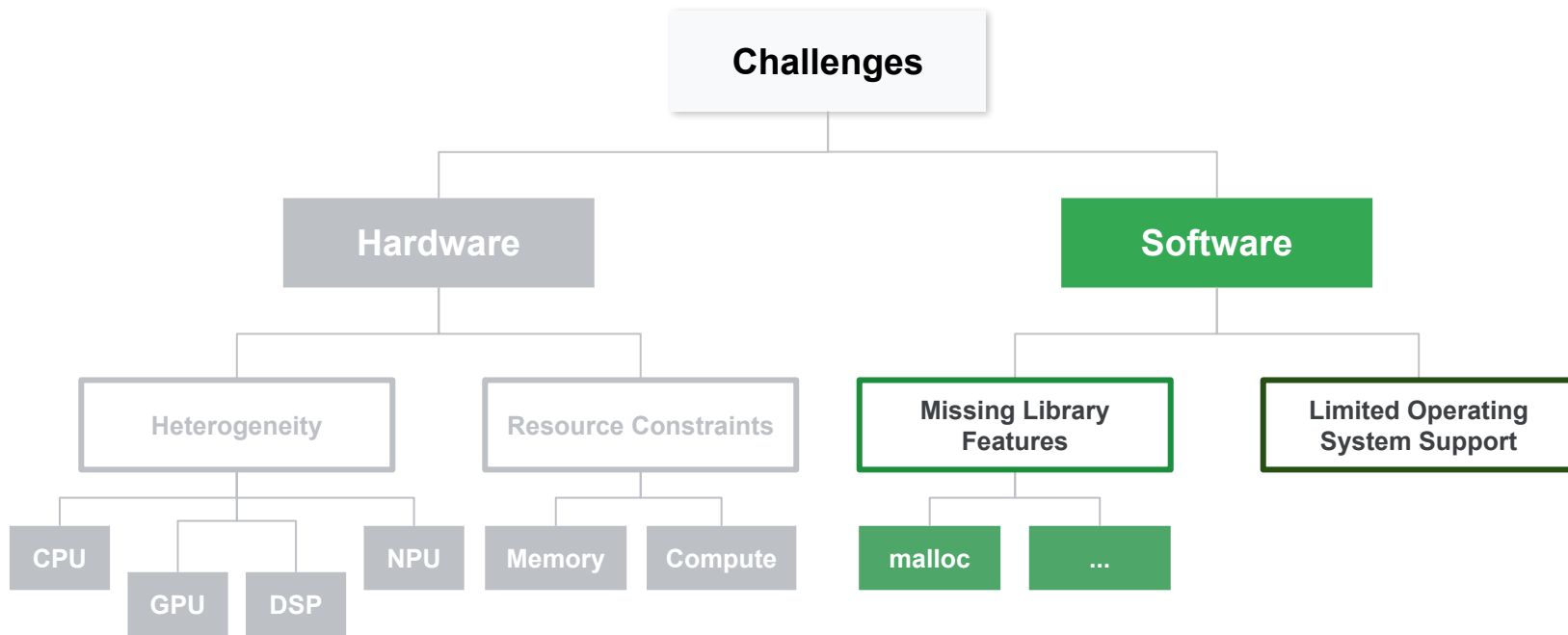


Board	MCU / ASIC	Clock	Memory	Sensors	Radio
Himax WE-I Plus EVB	HX6537-A 32-bit EM9D DSP	400 MHz	2MB flash 2MB RAM	Accelerometer, Mic, Camera	None
Arduino Nano 33 BLE Sense	32-bit nRF52840	64 MHz	1MB flash 256kB RAM	Mic, IMU, Temp, Humidity, Gesture, Pressure, Proximity, Brightness, Color	BLE
SparkFun Edge 2	32-bit ArtemisV1	48 MHz	1MB flash 384kB RAM	Accelerometer, Mic, Camera	BLE
Espressif EYE	32-bit ESP32-D0WD	240 MHz	4MB flash 520kB RAM	Mic, Camera	WiFi, BLE



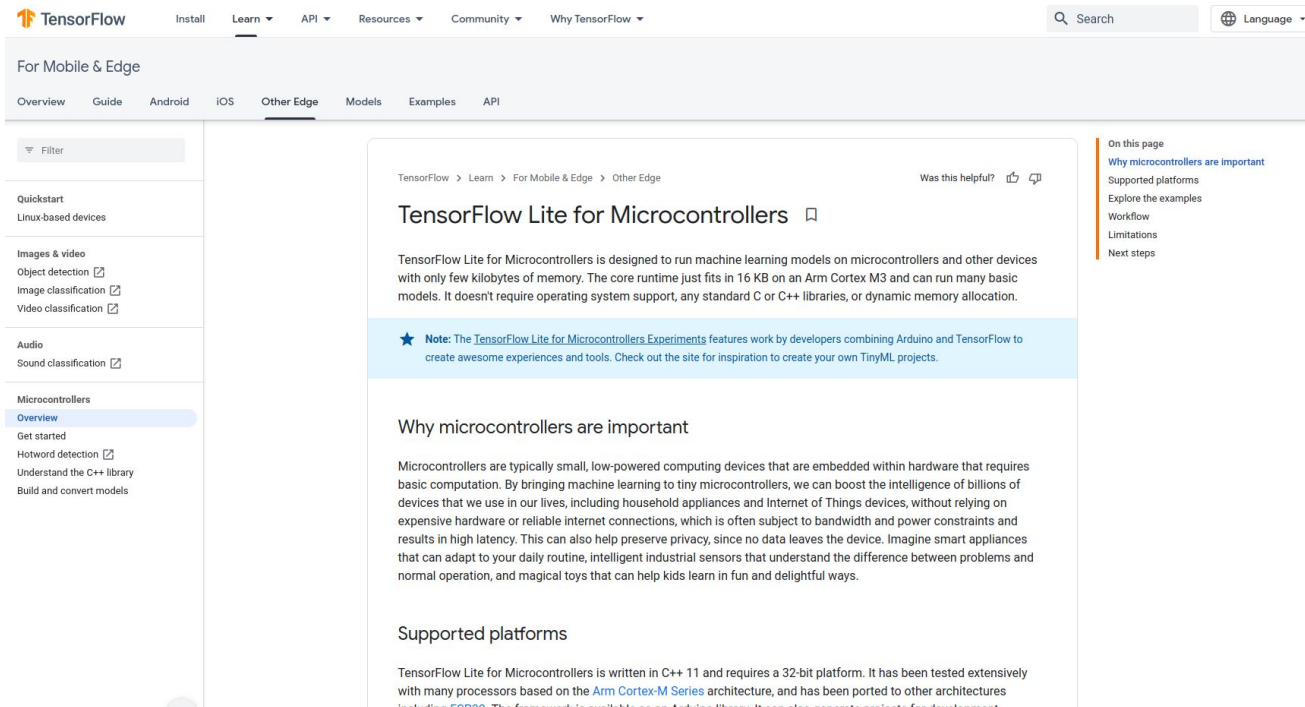








TensorFlow Lite Micro



The screenshot shows the TensorFlow Lite Micro documentation page. At the top, there is a navigation bar with the TensorFlow logo, 'Install', 'Learn', 'API', 'Resources', 'Community', and 'Why TensorFlow' menus. A search bar and a language selector are also present. Below the navigation bar, the page is titled 'For Mobile & Edge' and has sub-navigation for 'Overview', 'Guide', 'Android', 'iOS', 'Other Edge', 'Models', 'Examples', and 'API'. The 'Other Edge' section is active. On the left, there is a sidebar with a 'Filter' dropdown and categories: 'Quickstart' (Linux-based devices), 'Images & video' (Object detection, Image classification, Video classification), 'Audio' (Sound classification), and 'Microcontrollers' (Overview, Get started, Hotword detection, Understand the C++ library, Build and convert models). The main content area has a breadcrumb 'TensorFlow > Learn > For Mobile & Edge > Other Edge' and a 'Was this helpful?' feedback link. The main heading is 'TensorFlow Lite for Microcontrollers'. The text describes that TensorFlow Lite for Microcontrollers is designed to run machine learning models on microcontrollers with only a few kilobytes of memory. A blue callout box contains a note about the TensorFlow Lite for Microcontrollers Experiments. Below this, there are sections for 'Why microcontrollers are important' and 'Supported platforms'. On the right side, there is a 'On this page' sidebar with links to 'Why microcontrollers are important', 'Supported platforms', 'Explore the examples', 'Workflow', 'Limitations', and 'Next steps'.

TensorFlow > Learn > For Mobile & Edge > Other Edge

Was this helpful? [👍](#) [👎](#)

TensorFlow Lite for Microcontrollers

TensorFlow Lite for Microcontrollers is designed to run machine learning models on microcontrollers and other devices with only few kilobytes of memory. The core runtime just fits in 16 KB on an Arm Cortex M3 and can run many basic models. It doesn't require operating system support, any standard C or C++ libraries, or dynamic memory allocation.

★ **Note:** The [TensorFlow Lite for Microcontrollers Experiments](#) features work by developers combining Arduino and TensorFlow to create awesome experiences and tools. Check out the site for inspiration to create your own TinyML projects.

Why microcontrollers are important

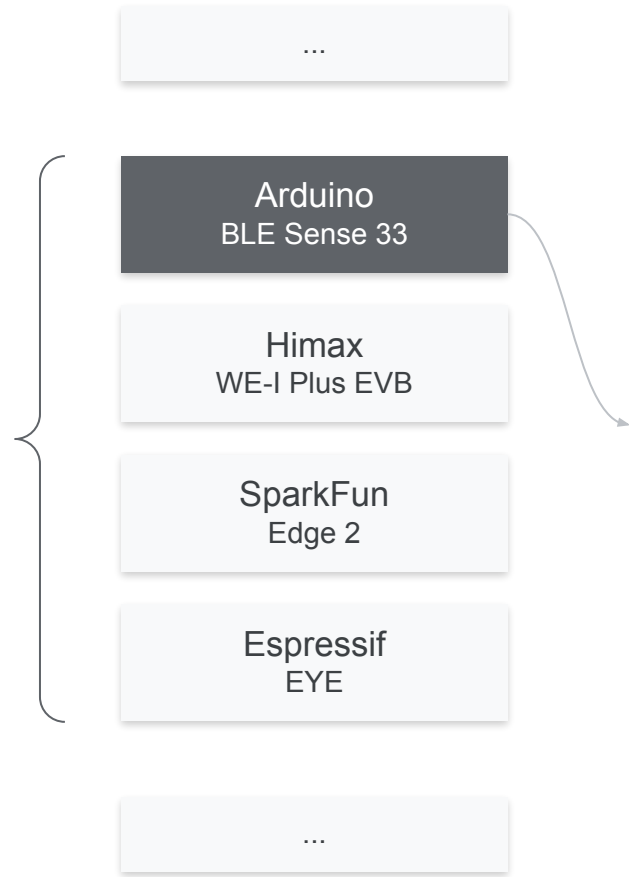
Microcontrollers are typically small, low-powered computing devices that are embedded within hardware that requires basic computation. By bringing machine learning to tiny microcontrollers, we can boost the intelligence of billions of devices that we use in our lives, including household appliances and Internet of Things devices, without relying on expensive hardware or reliable internet connections, which is often subject to bandwidth and power constraints and results in high latency. This can also help preserve privacy, since no data leaves the device. Imagine smart appliances that can adapt to your daily routine, intelligent industrial sensors that understand the difference between problems and normal operation, and magical toys that can help kids learn in fun and delightful ways.

Supported platforms

TensorFlow Lite for Microcontrollers is written in C++ 11 and requires a 32-bit platform. It has been tested extensively with many processors based on the [Arm Cortex-M Series](#) architecture, and has been ported to other architectures including [ESP32](#). The framework is available as an Arduino library. It can also generate projects for development

On this page

- [Why microcontrollers are important](#)
- [Supported platforms](#)
- [Explore the examples](#)
- [Workflow](#)
- [Limitations](#)
- [Next steps](#)










EDGE IMPULSE



Create library

Turn your impulse into optimized source code that you can run on any device.

 <p>C++ library</p>	 <p>Arduino library</p>	 <p>Cube.MX CMSIS- PACK</p>
 <p>WebAssembly</p>	 <p>TensorRT library</p>	

Run your impulse directly

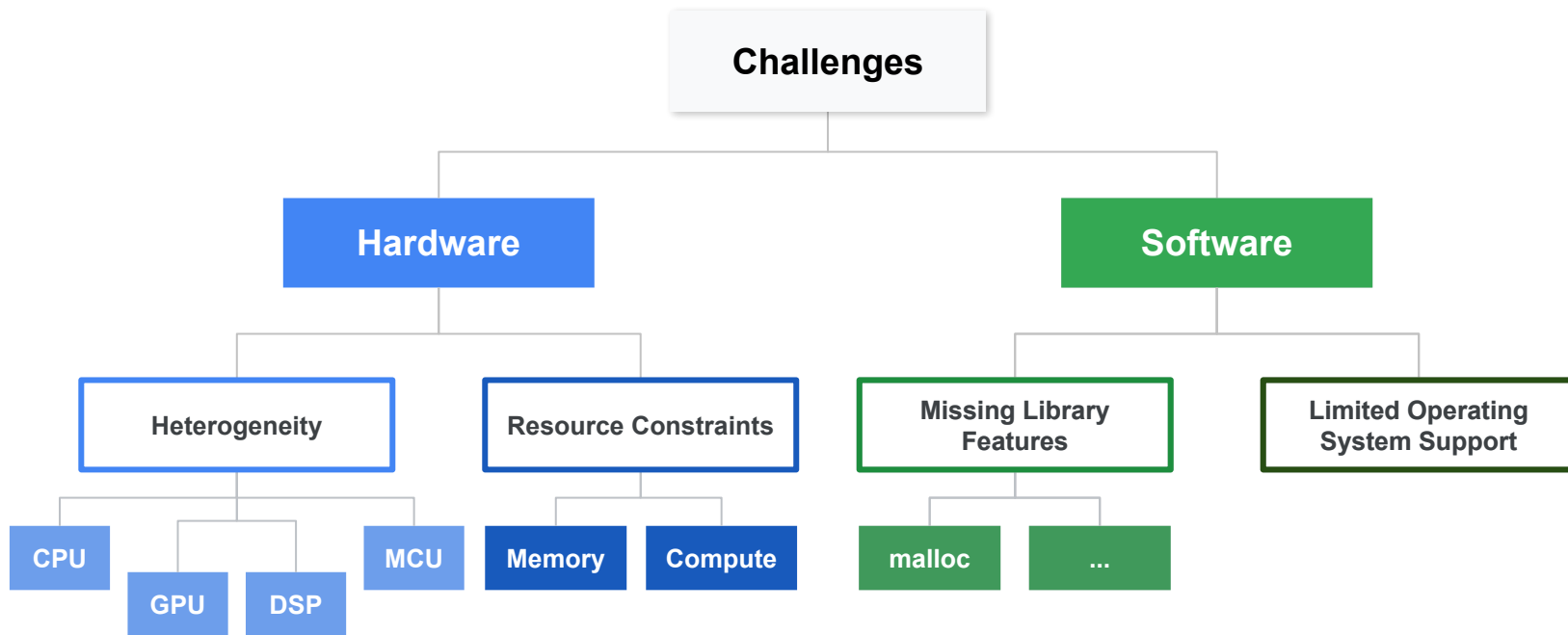
Run this impulse directly on your mobile phone or computer, no app required.

 <p>Computer</p>	 <p>Mobile phone</p>
---	---

Build firmware

Get a ready-to-go binary for your development board that includes your impulse.

 <p>ST IoT Discovery Kit</p>	 <p>Arduino Nano 33 BLE Sense</p>	 <p>Espressif ESP-EYE (ESP32)</p>
 <p>Raspberry Pi RP2040</p>	 <p>SiLabs Thunderboard Sense 2</p>	 <p>SiLabs xG24 Dev Kit</p>
 <p>Himax WE-I Plus</p>	 <p>Nordic nRF52840 DK + IKS02A1</p>	 <p>Nordic nRF5340 DK + IKS02A1</p>



Scaling TinyML

Why do 87% of data science projects never make it into production?

Transform 2019
San Francisco, July 10 & 11, 2019
#VBTRANSFORM

Build and scale with up to \$100,000 in AWS Activate credits

AWS Activate offers free tools, training, and more for startups to help you quickly build and scale quickly – plus, you can receive up to \$100,000 Activate credits.

[Apply here!](#)

“If your competitors are applying AI, and they’re finding insight that allow them to accelerate, they’re going to peel away really, really quickly.” Deborah Lef, CTO for data science and AI at IBM, said on stage at [Transform 2019](#).

On their panel, “What the heck does it even mean to ‘Do AI’? Lef and Chris Chapo, SVP of data and analytics at Gap, dug deep into the reason so many companies are still either kicking their heels or simply failing to get AI strategies off the ground, despite the fact that the inherent advantage large companies had over small companies is gone now, and the paradigm has changed completely. With AI, the fast companies are outperforming the slow companies, regardless of their size. And tiny, no-name companies are actually stealing market share from the giants.

But if this is a universal understanding, that AI empirically provides a competitive edge, why do only 13% of data science projects, or just one out of

Let's quantify this a bit. In 2019 alone, approximately **USD 40 billions** were invested into privately held AI companies. If we extrapolate this and throw the approximated success rate of AI projects into these figures (and completely exclude intracompany ML investments), we reach the conclusion that in 2019, around **USD 38 billions were wasted due to unsuccessful Machine Learning projects.**

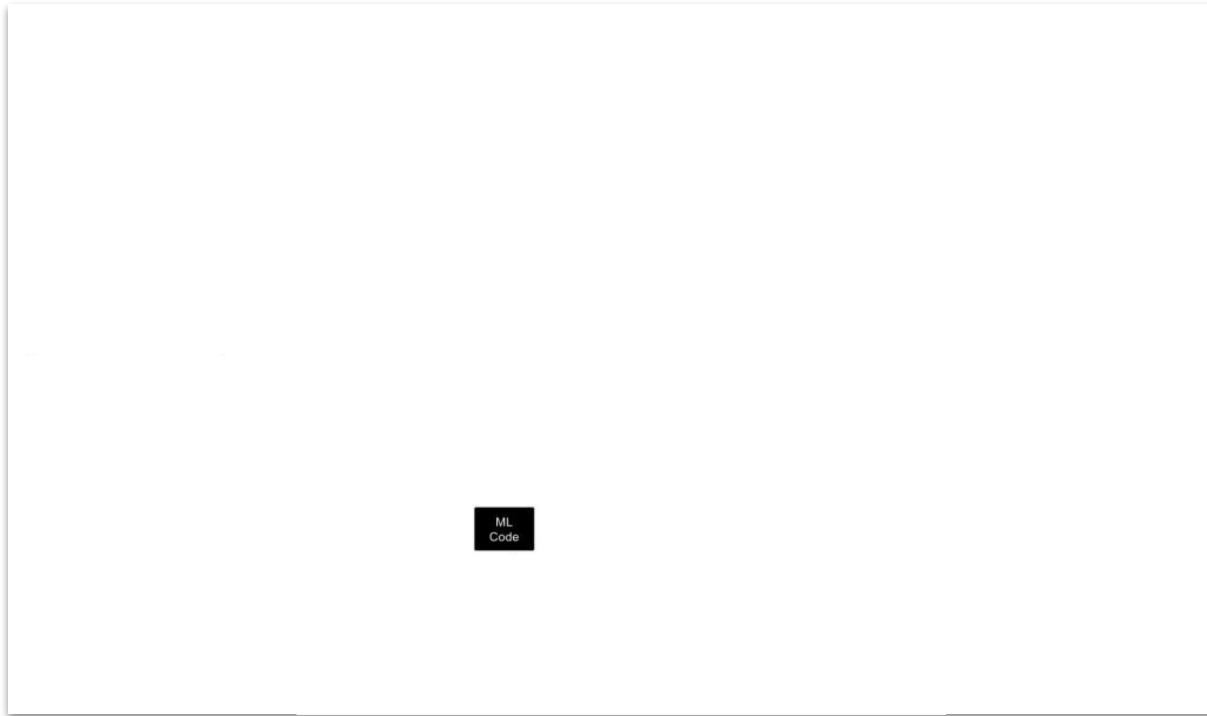


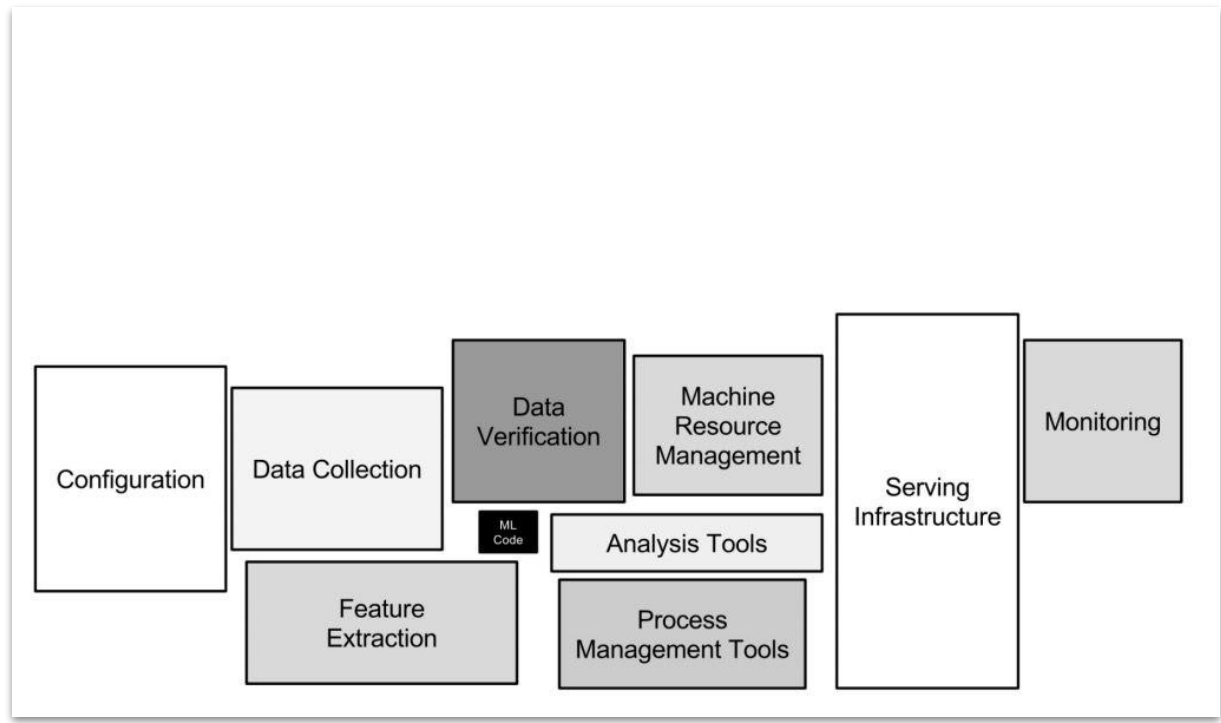
Predicts 2019: Analytics and BI Solutions



- Through 2020, 80% of AI projects will remain alchemy, run by wizards whose talents will not scale in the organization.
- Through 2022, only 20% of analytic insights will deliver business outcomes.
- By 2021, proof-of-concept analytic projects using quantum computing infrastructure will have outperformed traditional analytic approaches in multiple domains by at least a factor of 10

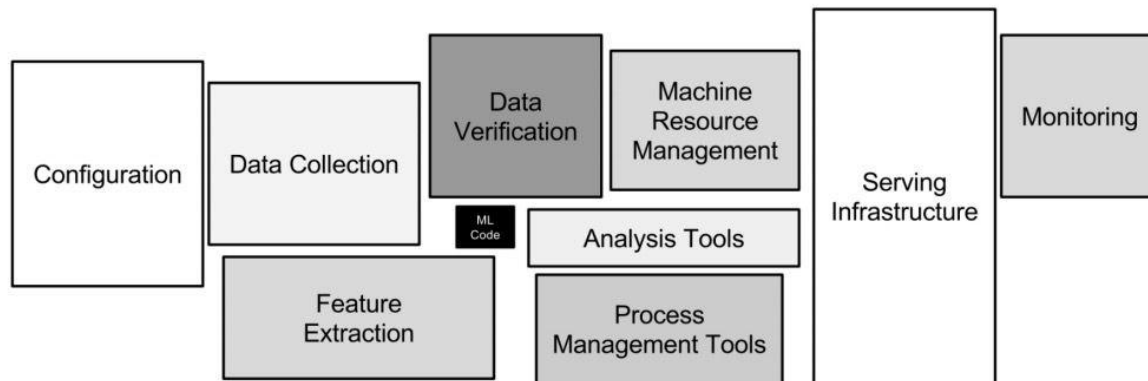
Source: https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/



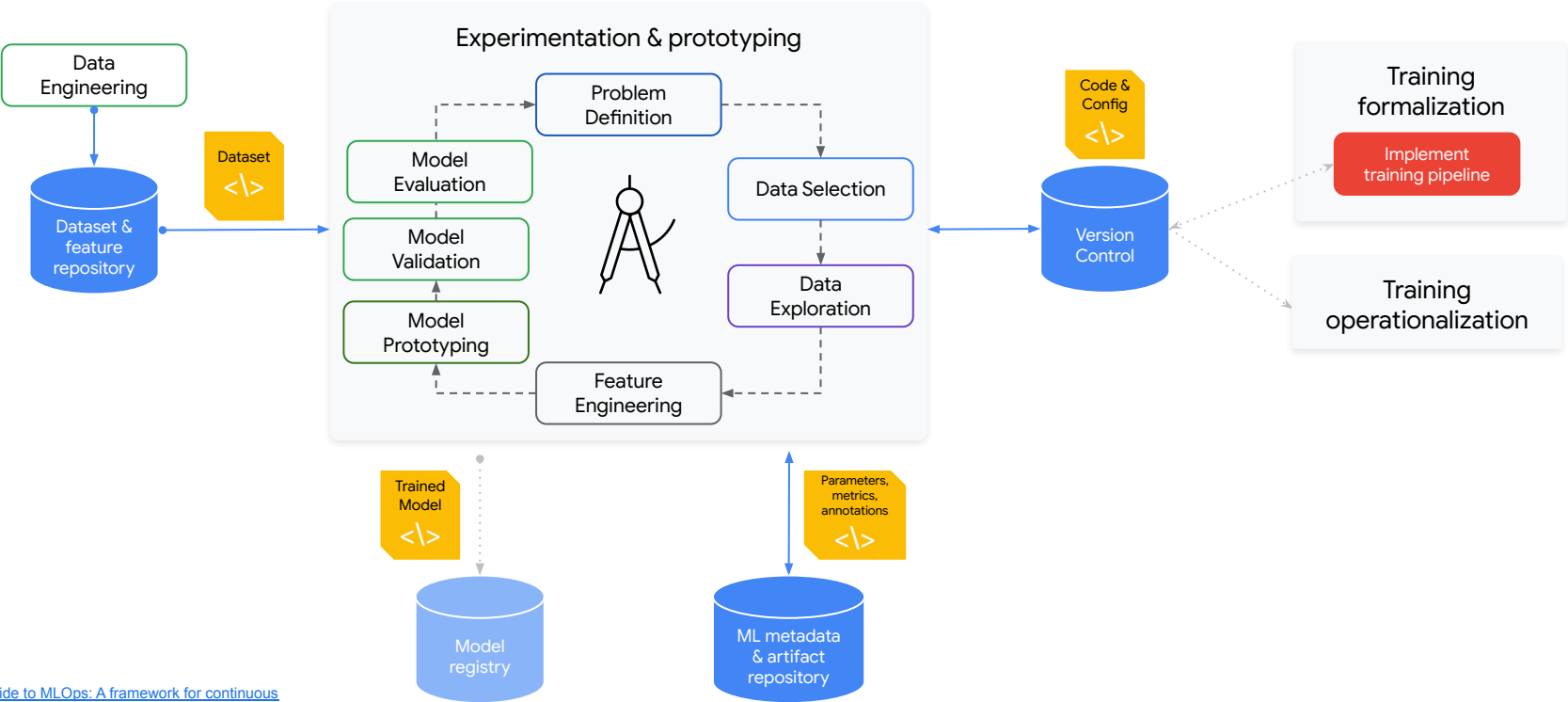


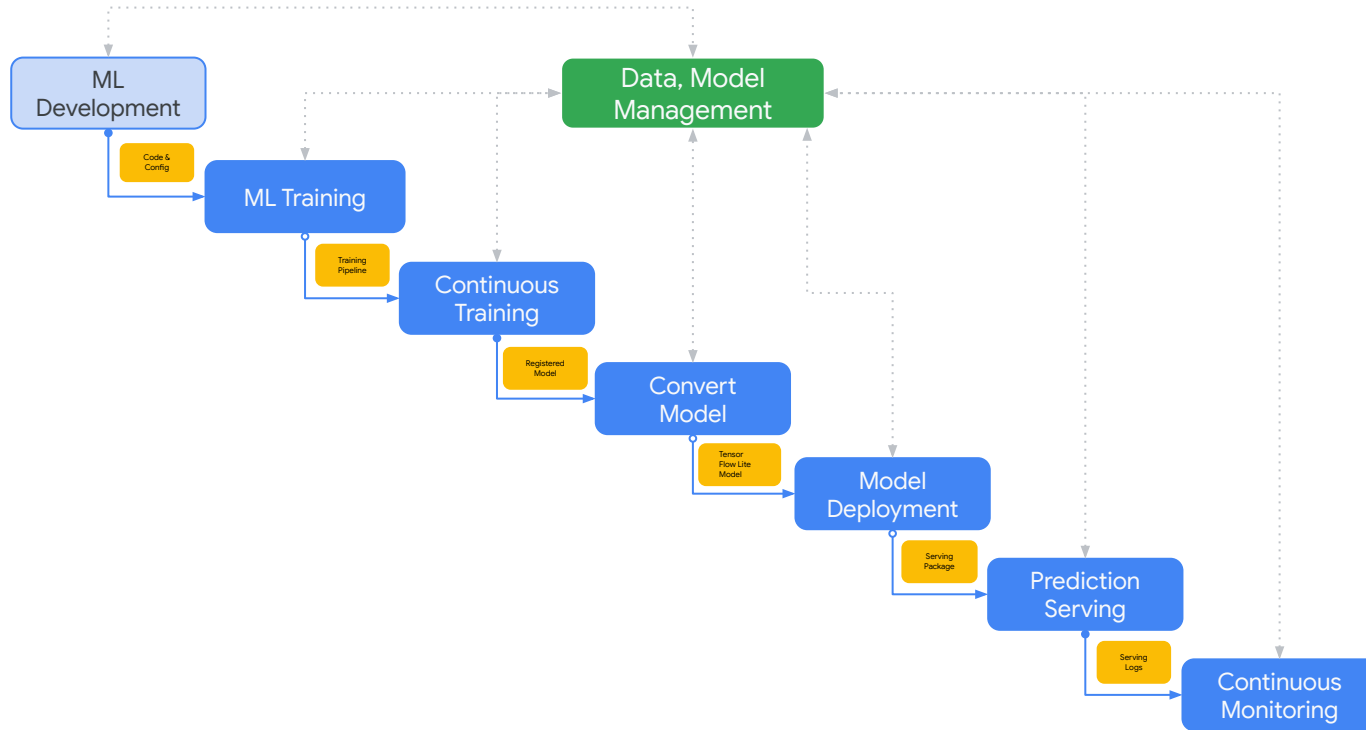
Hidden Technical Debt in Machine Learning Systems

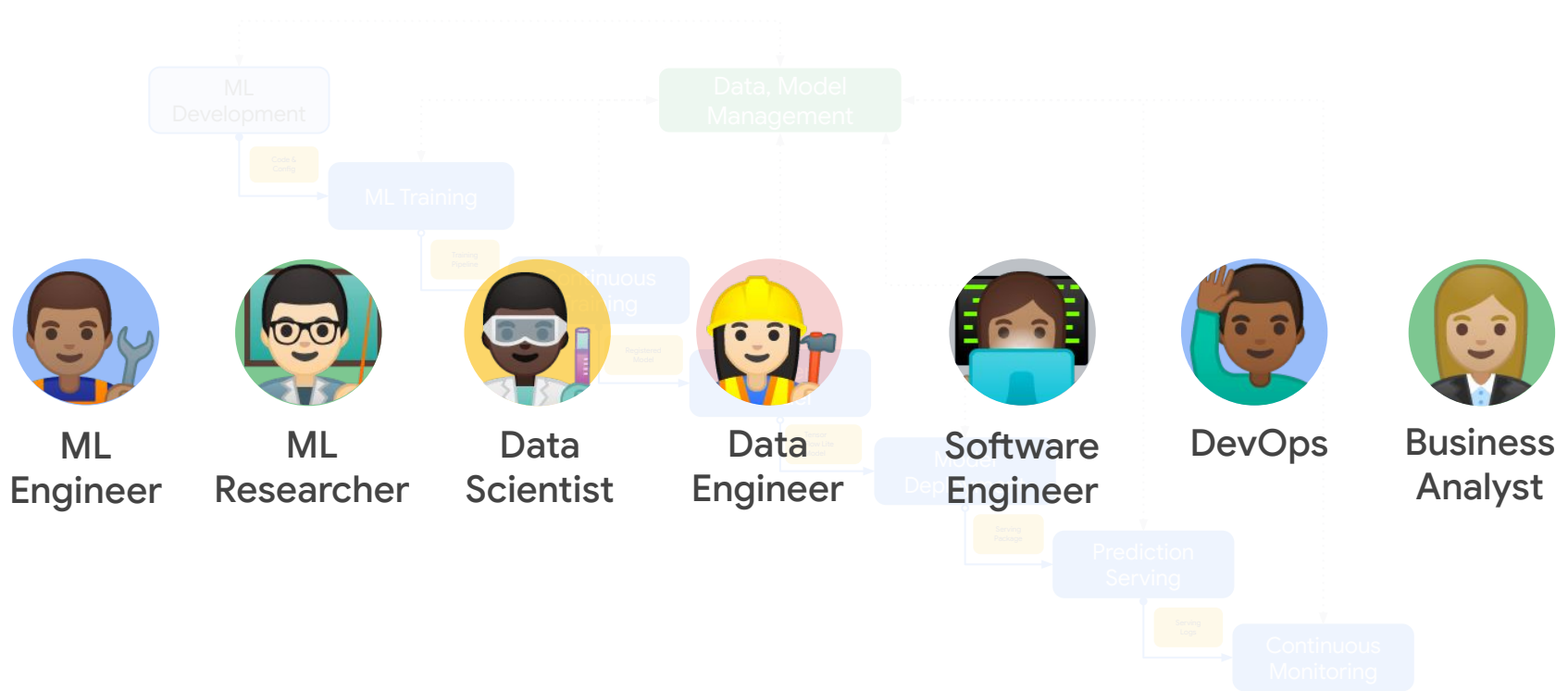
D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley, gholt, dgg, edavydov, toddphillips}@google.com
Google, Inc.



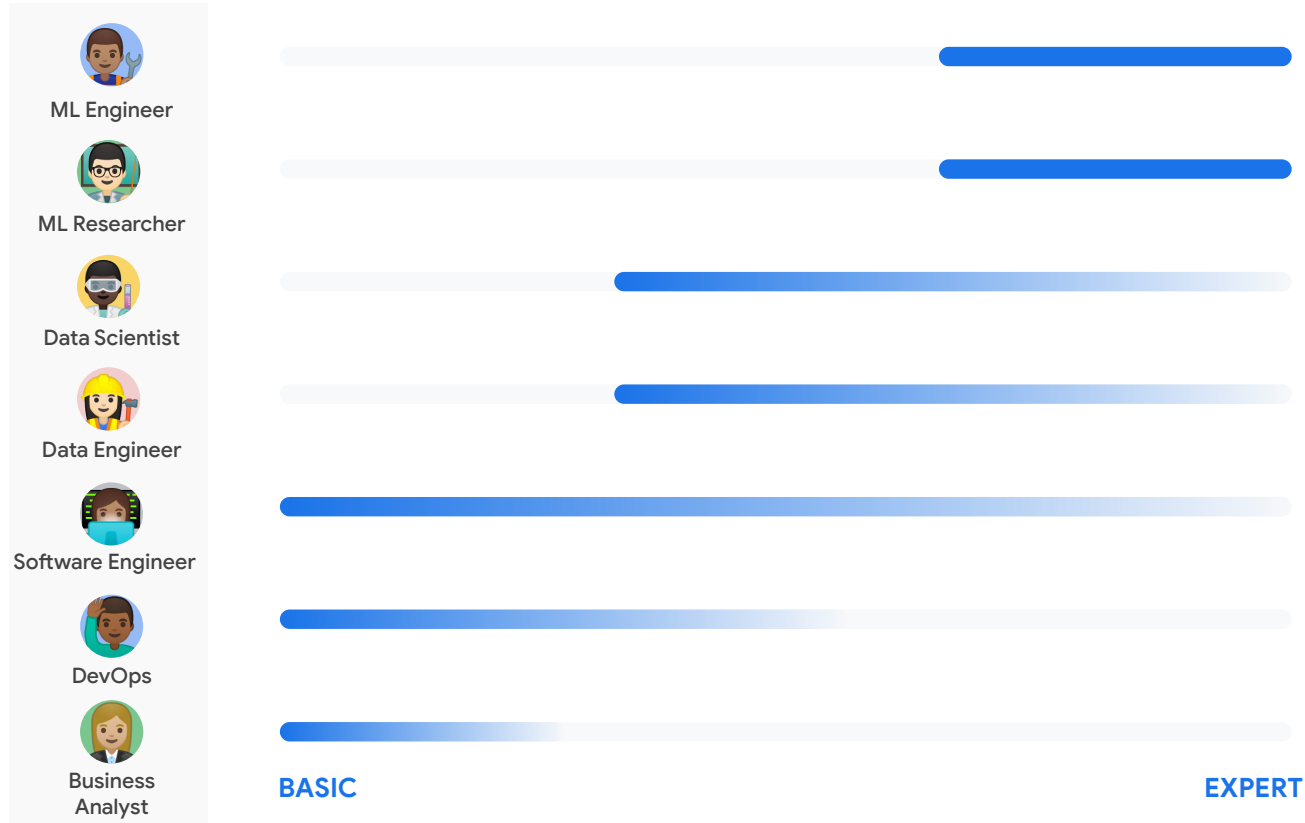
MLCode



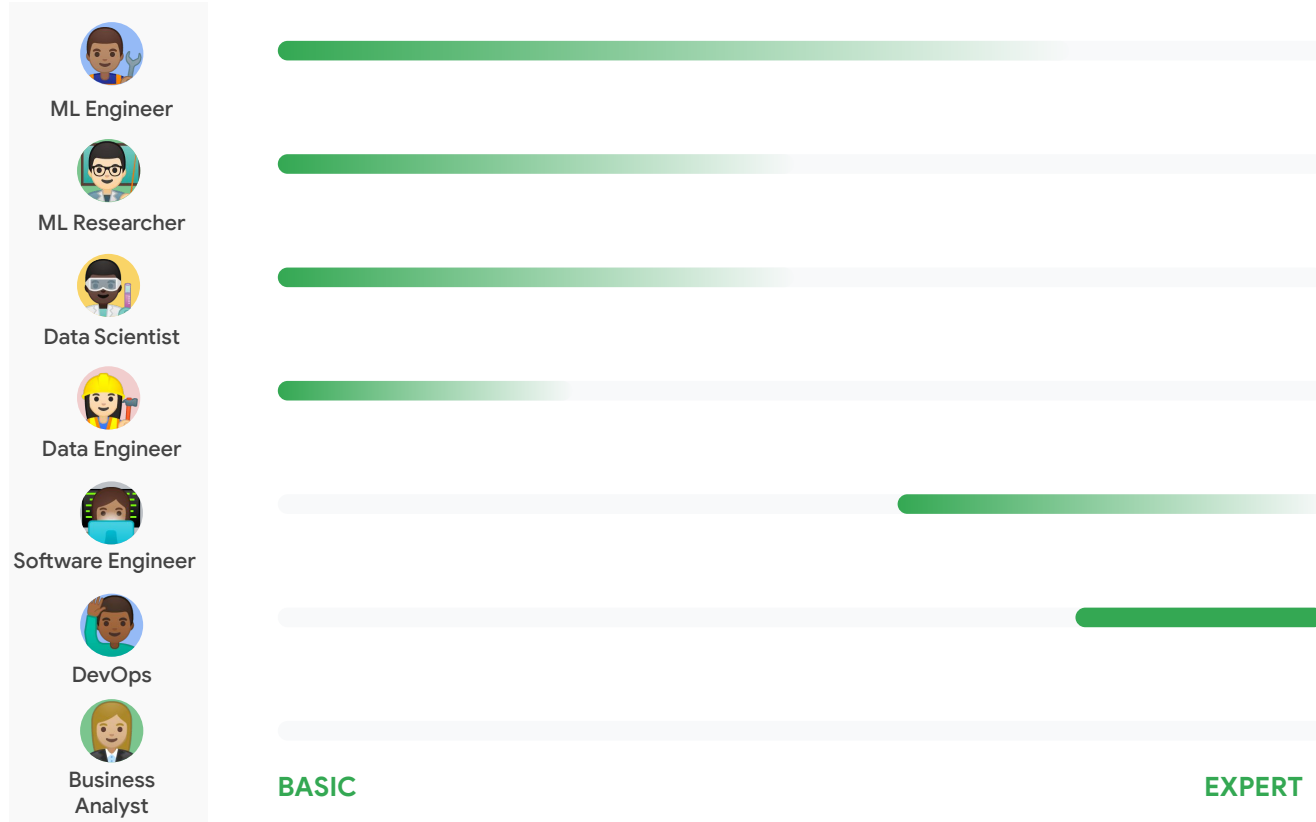


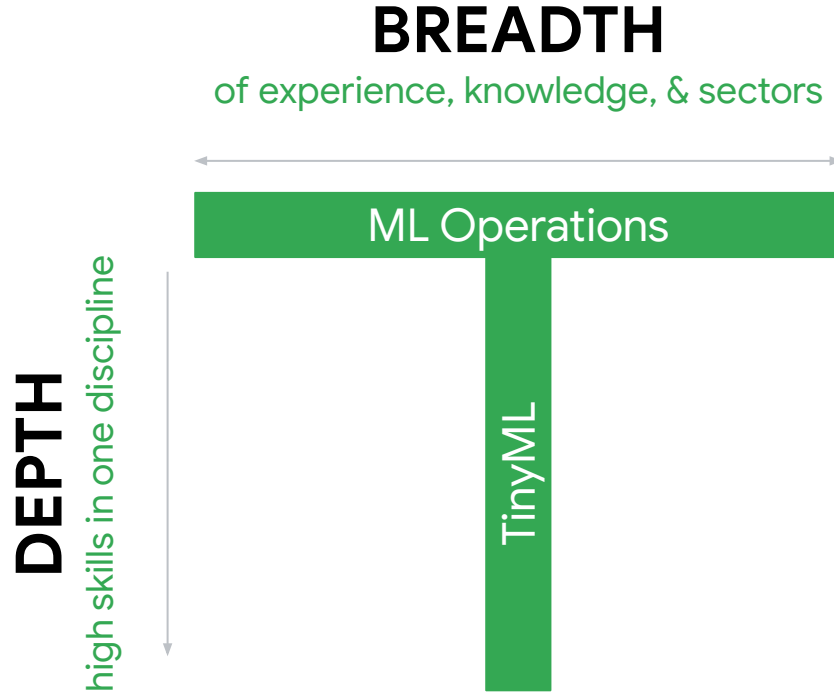


ML Expertise



Deployment Expertise





MLOps for Scaling TinyML | edX

edX.org/course/mlops-for-scaling-tinyml

TinyML Harvard MLC Research Seed CS141 TimeBuddy Jasper.AI VJs Funding Other Bookmarks

edX Courses Programs & Degrees Schools & Partners What do you want to learn? edX for Business Sign In Register


edX is [part of 2U](#): the next era of online learning begins today! Visit our Help Center to [read more](#) about changes at edX

Catalog > Computer Science Courses

HARVARD UNIVERSITY

MLOps for Scaling TinyML

This course introduces learners to Machine Learning Operations (MLOps) through the lens of TinyML (Tiny Machine Learning). Learners explore best practices to deploy, monitor, and maintain (tiny) Machine Learning models in production at scale.

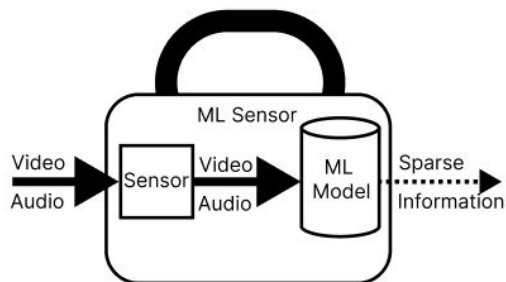


Estimated 7 weeks
2-4 hours per week

Self-paced
Progress at your own speed

Free
Optional upgrade available

ML Sensors



MACHINE LEARNING SENSORS

**Pete Warden¹ Matthew Stewart² Brian Plancher² Colby Banbury² Shvetank Prakash² Emma Chen²
Zain Asgar¹ Sachin Katti¹ Vijay Janapa Reddi²**

¹Stanford University ²Harvard University

ABSTRACT

Machine learning sensors represent a paradigm shift for the future of embedded machine learning applications. Current instantiations of embedded machine learning (ML) suffer from complex integration, lack of modularity, and privacy and security concerns from data movement. This article proposes a more data-centric paradigm for embedding sensor intelligence on edge devices to combat these challenges. Our vision for “sensor 2.0” entails segregating sensor input data and ML processing from the wider system at the hardware level and providing a thin interface that mimics traditional sensors in functionality. This separation leads to a modular and easy-to-use ML sensor device. We discuss challenges presented by the standard approach of building ML processing into the software stack of the controlling microprocessor on an embedded system and how the modularity of ML sensors alleviates these problems. ML sensors increase privacy and accuracy while making it easier for system builders to integrate ML into their products as a simple component. We provide examples of prospective ML sensors and an illustrative datasheet as a demonstration and hope that this will build a dialogue to progress us towards sensor 2.0.

IoT 1.0:
Internet
of Things



IoT 2.0:
Intelligence
on Things

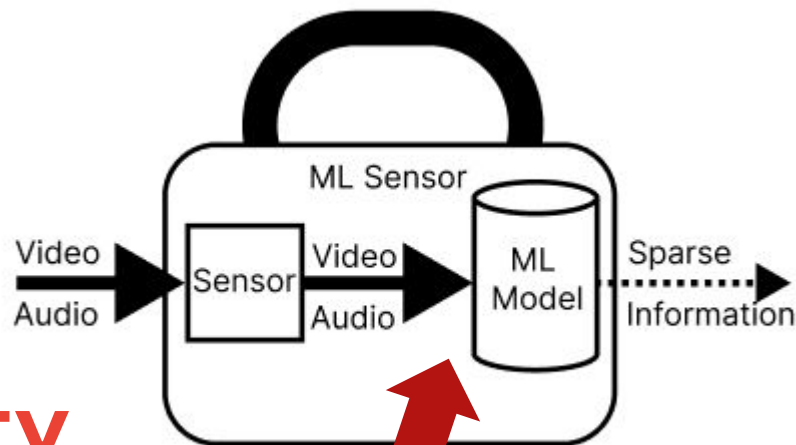
Bandwidth

Latency

Energy

Reliability

Privacy



TinyML

PA1 Person Detection Module

Description: The PA1 Person Detection Module enables you to quickly and easily add smarts to your IoT deployment to monitor and detect for humans. You can use this module indoors and outdoors to understand where and when humans arrive at your deployment site.

Features:

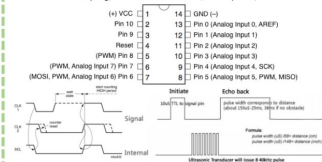
- Real-time Person Detection with On-Device ML
- Indoor and Outdoor use
- Finds a person at a maximum distance of 10 meters to a minimum distance of 5 centimeters
- Operates in low and high light environments (1-20000 Lux) across a wide temperature range (0 to 50 °C)
- Features Color and Black-and-White Detection Modules

Use Cases:

- Smart business and home security systems
- Multi-modal key word spotting for virtual assistants
- Occupancy sensors and other infrastructure sensors

Description, Features, and Use Cases

Sources: fabacademy.org, electroshematics.com, and nxp.com/docs



Communication Specification and Pinout

Source: datanutrition.org

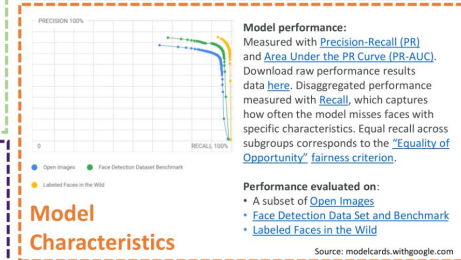
Source: internetofprivacy.com

Source: docs.luxonis.com

Diagrams and Form Factor

Camera Specs	Color camera	Stereo pair					
Sensor	IMX214	OV7251	SYMBOL	RATING	MIN	MAX	UNIT
EVFov / HFOV / VFOV	81° / 61° / 54°	86° / 73° / 58°	V _{max}	Recommended	4.75	5.25	V
Resolution	1.3MP (1280x800)	480P (640x480)	V _{min}	Minimum	3.5	5.5	V
Frame Rate	30 FPS → DR FV: 50cm →	300 FPS	I _{max}	Maximum	0	1.5	A
Max. Estimate	60 FPS	300 FPS	I _{min}	Minimum	0	1.5	A
T-number	2.2 ± 5%	2.2	F	Focal	4	6	W
Lens size	3.73 inch	1.75 inch	F _{max}	Max. Focal	2.4	2.6	W
Effective Focal Length	3.37mm	1.3mm	T _{max}	Max. Temp.	45	55	°C
Distortion	< 1%	< 1.5%					
Pixel size	1.12µm × 1.12µm	3µm × 3µm					

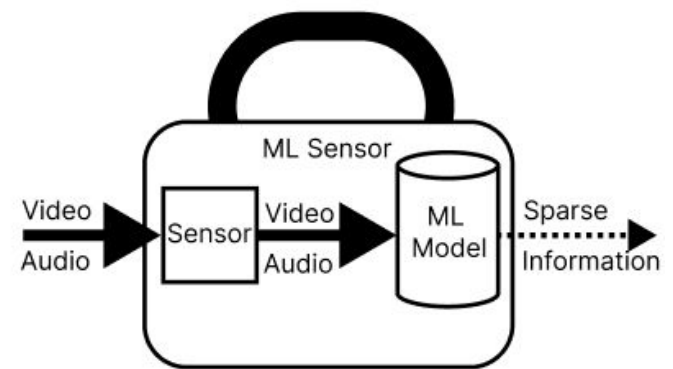
Source: docs.luxonis.com



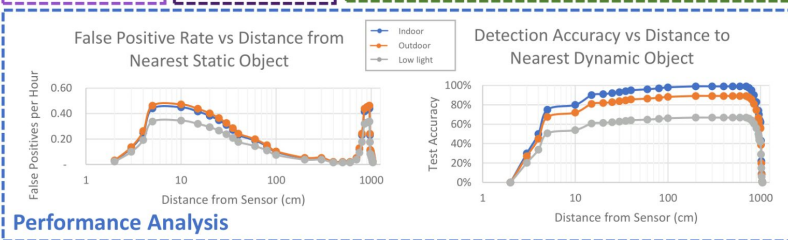
Environmental Impact: Full report can be found [here](#).

390g CO₂-eq, 23L Water

Source: st.com



<https://mlsensors.org/>



How can you learn more?

The Future of ML is Tiny and Bright



Professional Certificate in
Tiny Machine Learning (TinyML)

I'm interested

What you will

- Fundamentals of machine learning
- How to gather data and preprocess it
- How to use Python to build machine learning models
- How to optimize machine learning models
- How to conceive and design machine learning systems
- How to program in TensorFlow

MLOps for Scaling TinyML | edX

edX Courses Programs & Degrees Schools & Partners

edX is part of 2U: the next era of online learning begins today! Visit our Help Center to read more about changes at edX

Catalog > Computer Science Courses

Introduction to Embedded Machine Learning

Offered by **EDGE IMP**

★★★★★ 4.8 306 ratings | 96%

Shawn Hymel +1 more instructor

Enroll for Free
Starts Apr 25

Financial aid available

16,642 already enrolled

About Instructors Syllabus Reviews Enrollment Options FAQ

Welcome to the Tiny Machine Learning Open Education Initiative (TinyMLedu)

We are an international group of academics and industry professionals working to improve global access to educational materials for the cutting-edge field of TinyML. TinyML brings the transformative power of machine learning (ML) to the performance- and power-constrained domain of embedded systems. Successful deployment in this field requires knowledge of applications, algorithms, hardware, and software. TinyMLedu is hosted by the Harvard John A. Paulson School of Engineering and Applied Sciences in collaboration with the tinyML

Home TinyML Courses Teach TinyML Working Groups Discourse

tinyML 7.23K subscribers

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT

Uploads

Video Title	Duration	Views	Time Ago
tinyML Auto ML Tutorial with Notia AI	59:48	130 views	1 day ago
tinyML Auto ML Tutorial with Neuton	58:09	97 views	1 day ago
tinyML Challenge 2022: Smart weather station	1:02:30	208 views	5 days ago
tinyML Talks South Africa - What is tinyML?	34:31	171 views	12 days ago
tinyML Talks: The new Neuromorphic Analog Signal Processing	1:00:30	304 views	12 days ago
tinyML Talks Shenzhen: 分享主题：机器学习应用于K12...	1:06:44	80 views	12 days ago
tinyML Auto ML Forum - Depois	1:53:07		
tinyML Auto ML Forum - Keynote - State of the...	42:13		
tinyML Auto ML Forum - Keynote - State of the...	33:29		
tinyML Trailblazers Success Stories with Mouna Elkhatib	1:00:16		
tinyML Talks Taiwan in Mandarin and English...	1:10:53		
tinyML Talks: Meetup Italy with small-medium industries	1:27:13		

Latam Regional Workshop
on SciTinyML:
Scientific Use of
Machine Learning on
Low-Power Devices



11-15 July 2022
Online

Further information:
<https://tinyMLedu.org/SciTinyML>
edu@tinyML.org

The Future of Machine Learning is Tiny and Bright

Brian Plancher
Barnard College, Columbia University
Harvard John A. Paulson School of Engineering and Applied Sciences
brianplancher.com

