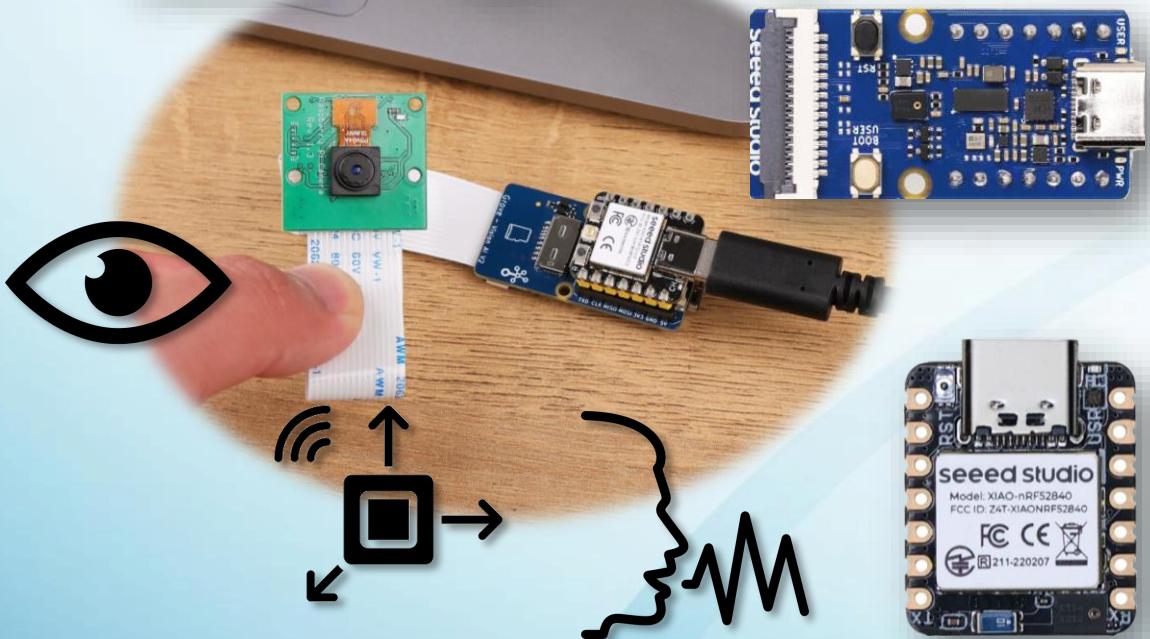


OmniXRI TinyML 小學堂 2025



Himax



歡迎加入
邊緣人俱樂部



只有更邊



Cortex-M
Processor



Ethos-U
MicroNPU

【第 1 講】 微型人工智慧簡介

簡報大綱



- 1.1. 智慧物聯網與微型人工智慧
- 1.2. TinyML 常見應用
- 1.3. TinyML 常見開發板
- 1.4. TinyML 常見開發工具

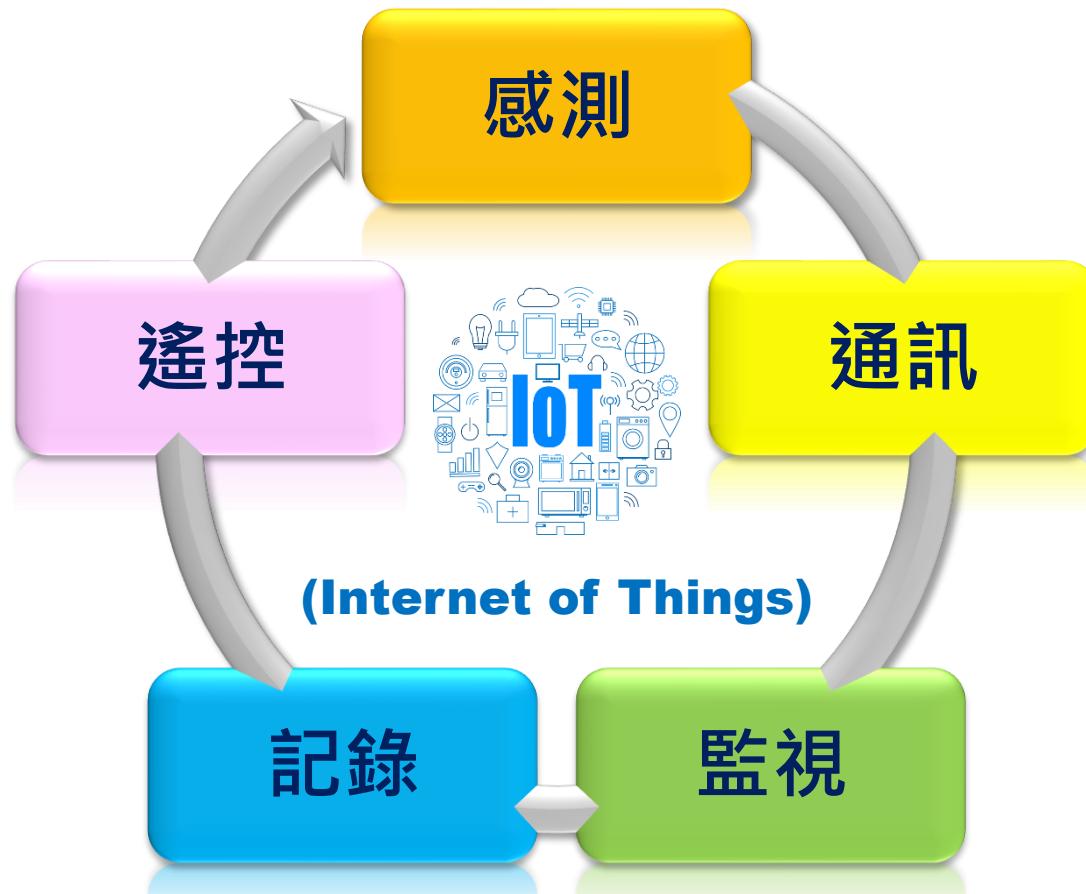
本課程完全免費，請勿移作商業用途！
歡迎留言、訂閱、點讚、轉發，讓更多需要的朋友也能一起學習。

完整課程大綱：<https://omnixri.blogspot.com/2025/03/omnixri-tinyml-2025-0.html>
課程直播清單：<https://www.youtube.com/@omnixri1784streams>



1.1.智慧物聯網與微型人工智慧 AIoT & TinyML

傳統物聯網(IoT)工作流程



智慧農業、工業、城市、
交通、醫療、家庭...

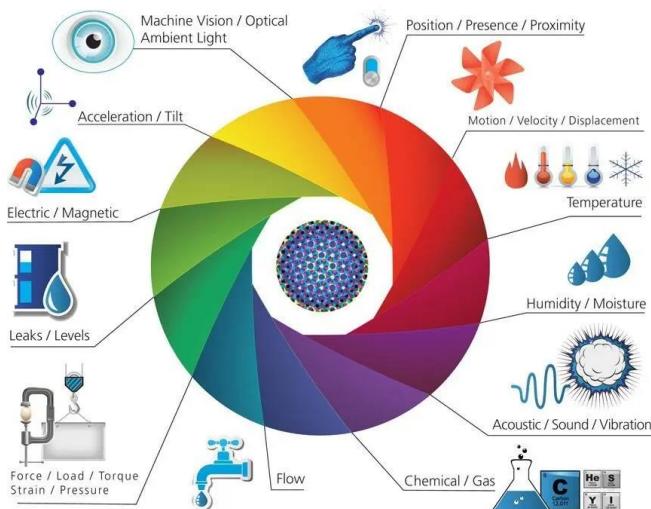


工人智慧(HI) vs. 人工智慧(AI)

工業物聯網 (IIoT) 架構

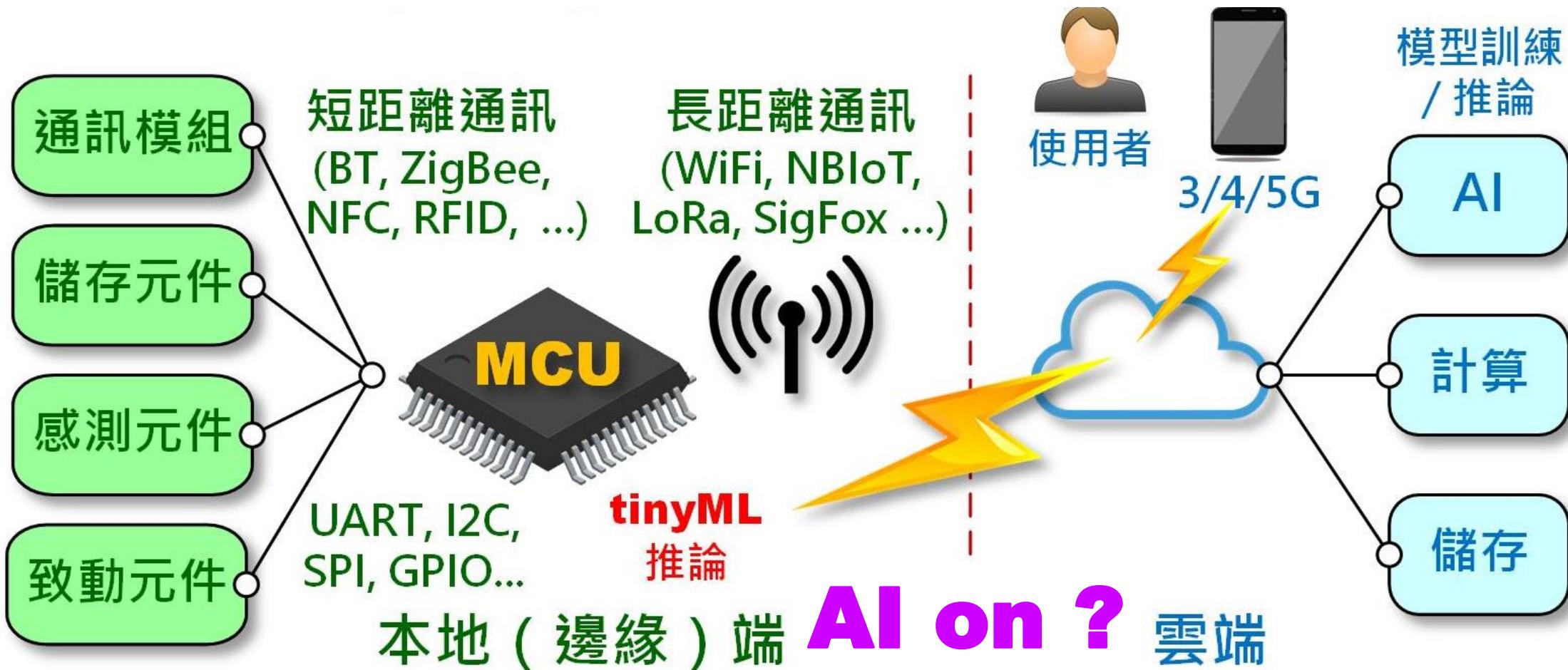
常見感測器類型：

- A. 電阻、電容、電感式
- B. 光電式
- C. 壓電式
- D. 電聲式
- E. 微機電式
- F. 電磁波式
- G. 影像式



影像來源：<https://www.edntaiwan.com/20190201ta31-designers-guide-to-iiot-sensor-systems/> <https://3smarket-info.blogspot.com/2019/06/iiot.html>

智慧物聯網 (AIoT) 架構



圖片來源：<https://omnixri.blogspot.com/2021/09/aiottinymlmcu.html>

OmniXRI 整理繪製, 2021/9/9

智慧健康（穿戴式）物聯網



➤ 感測器

- ◆ 麥克風
- ◆ 加速度計
- ◆ 陀螺儀
- ◆ 血氧計
- ◆ 心率計
- ◆ 體溫計
- ◆ 車禍偵測
- ◆ 跌倒偵測
- ◆ 運動狀態
- ◆ 睡眠品質
- ◆ 心臟健康
- ◆ 體溫監測

➤ 主要功能

➤ 主要功能

- ◆ 行為分析
- ◆ 影像辨識
- ◆ 語音辨識
- ◆ 記錄回溯
- ◆ GPS定位
- ◆ 影音通訊

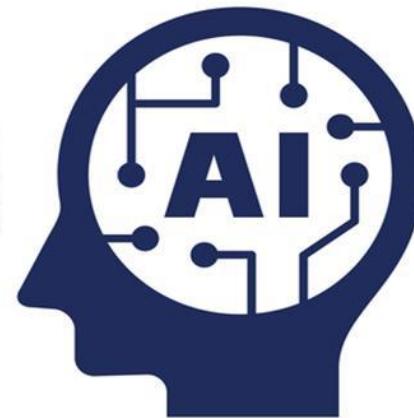
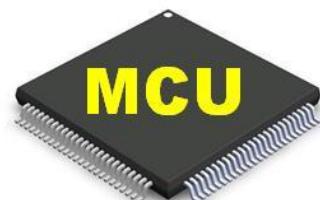
➤ 主要功能

- ◆ 長期資料儲存
- ◆ 資料統計分析
- ◆ 記錄可視處理
- ◆ 跨越裝置通訊
- ◆ 模型訓練推論
- ◆ 模型更新部署

邊緣智慧 vs. 生成智慧

- 邊緣端極少資源
儲存(Flash, RAM)、計算
- 智慧感測
聲音、影像、運動...
- 超小模型、超低功耗

極小化



極大化

- 雲端無限資源
儲存、計算、頻寬、功耗
- AI生成
對話、影像、程式...
- 超巨大資料集、模型

資料來源：<https://omnixri.blogspot.com/2023/04/20230420aiexpo-aiedge-ai.html>

Edge AI 是什麼？

Edge AI 就像可以獨立移動的**交通工具**，不同的公司會給你不同答案。



邊緣智慧 (Edge AI) 誰說了算？

Edge AI : 不同的公司會給你不同答案，從幾塊到幾十萬美金都有可能。
廣義：不上網能完成所有AI推論，**狹義**：用電池供電完成所有AI推論。

Edge Server



Nvidia EGX



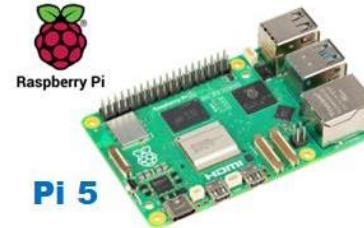
NB AI (AI PC)



Mobile AI (SoC)



SBC AI (MPU / NPU / GPU)



Wearable AI (MCU / TinyML)



何謂微型人工智慧 (TinyML)



2019年3月 **tinyML** 基金會舉辦首次峰會，有超過90家公司共同參與。



2024年11月更名為 **Edge AI** 基金會

➤ **微型機器學習(Tiny Machine Learning)**被廣泛定義為一個快速發展的機器學習技術和應用領域，包括**硬體**（專用積體電路）、**算法**和能夠執行設備上感測器（視覺、聲音、運動IMU、生物醫學等）**數據分析**的軟體極低的功耗，通常在 **mW** 範圍內及以下，因此可實現各種始終在線的用例並針對電池供電設備。**(非強制定義)**

➤ **TinyML**同義字，**MCU AI**, **Tiny AI**, **Micro AI**, **On Device AI**, **Embedded ML**, **Embedded AI**, **Smart Sensor**, **Intelligence Sensor** ...

TinyML 相關論文

2019年之後論文，229筆。



The screenshot shows the homepage of the NDLTD (National Digital Library of Theses and Dissertations in Taiwan) with a search bar at the top. Below the search bar, there is a large banner for 'TinyML 微型機器學習 嵌入式'. The main content area displays a search result table titled '臺灣博碩士論文熱門排行榜' (Top Ranking of Taiwan Theses and Dissertations). The table lists the top 5 schools with their respective citation counts and download counts.

名次	學校名稱	已授權全文	書目
1	國立清華大學	672	765
2	國立成功大學	569	582
3	國立臺灣大學	539	899
4	國立臺灣師範大學	532	579
5	國立政治大學	512	621

<https://omnixri.blogspot.com/2024/06/tinyml.html>

2019年之後論文，614筆。

TinyML相關學術論文

這裡主要搜集單晶片 (MCU) 等級的機器學習、人工智慧、深度學習等相關研究及論文。而單板微電腦(SBC)、行動裝置或小型工業電腦等級之相關研究及論文請參考另一篇「[Edge AI相關學術論文](#)」。

註：相關論文連結不一定有提供PDF可供下載，或者必須有學術網路帳號才能下載，請自行點擊查閱。以下論文清單依發表時間(相同月份)由新到舊月份排序。目前小計614篇。

最後更新日期：2025/02/18

上一次更新日期：2023/11/29

https://hackmd.io/@OmniXRI-Jack/tinyML_papers

MCU等級TinyML常見應用及限制

ARM MCU等級晶片 智慧運算能力與適用情境

非影像類智慧感測應用



MCU等級跨度大
以Arm Cortex-M為例, M0+, M3, M4, M7, M55, M85

指令速度從數十MHz到數百MHz
程式碼儲存空間從數KB到數MB
SRAM從數KM到數MB

圖片來源：<https://www.arm.com/blogs/blueprint/ai-for-iot-devices>

運算量(TOP/S)

MCU等級TinyML 優點

- 低單價、功耗
- 低延時(反應快)
- 高隱私(免上網)
- 易連接各式感測器及通訊模組

缺點

- 速度、算力不足
- 記憶體不足
- 儲存能力小
- 難以在線訓練

智慧物聯網 vs. 微型人工智慧

AIoT

視覺/語音/數據
雲端計算

長距大量通訊
雲端巨量儲存

通用IDE, LIB
專用IDE, SDK

MCU, SBC, PC



智慧應用

通訊儲存

軟體開發

硬體設計

tinyML

智慧感測
邊緣裝置運算

短距小量通訊
邊緣微量儲存

整合型平台
通用IDE, LIB

MCU (+NPU)

資料來源：<https://omnixri.blogspot.com/2022/11/20221110icaiottinyml-ic.html>



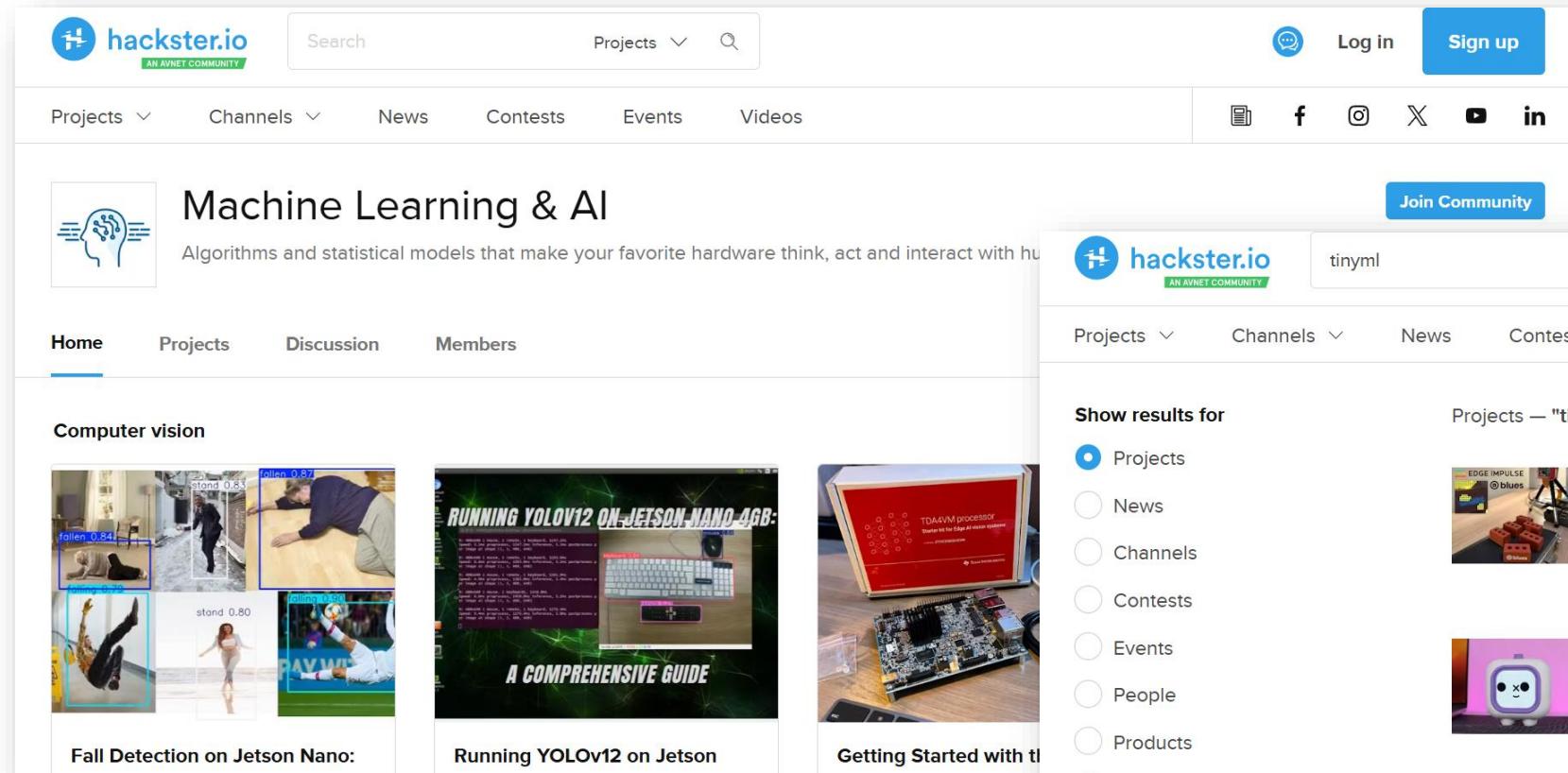
1.2. TinyML 常見應用

TinyML案例分享（技術分類）



TinyML應用大全 (30組案例) https://hackmd.io/@OmniXRI-Jack/tinyML_30_projects

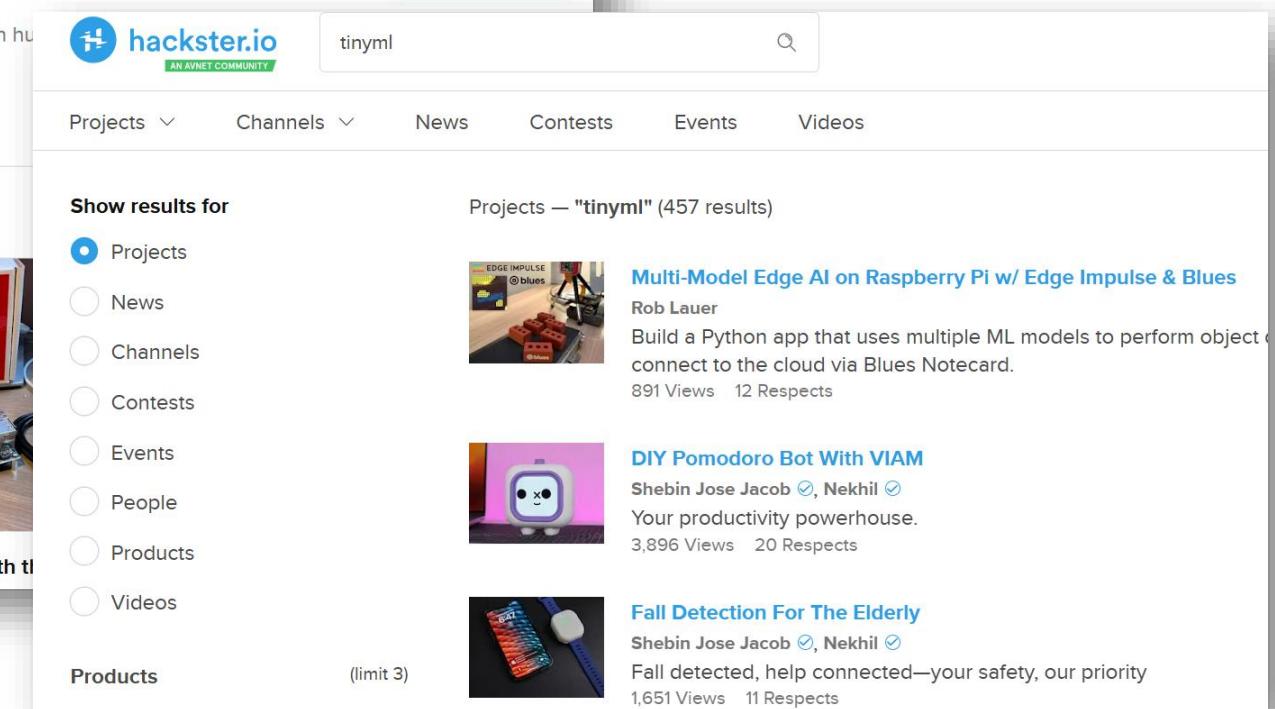
TinyML案例分享 – Hackster.io



The screenshot shows the Hackster.io homepage with the "Machine Learning & AI" category selected. The main content area displays projects related to computer vision, such as "Fall Detection on Jetson Nano:" and "Running YOLOv12 on Jetson". A sidebar on the right provides navigation options for the community.

<https://www.hackster.io/ML>

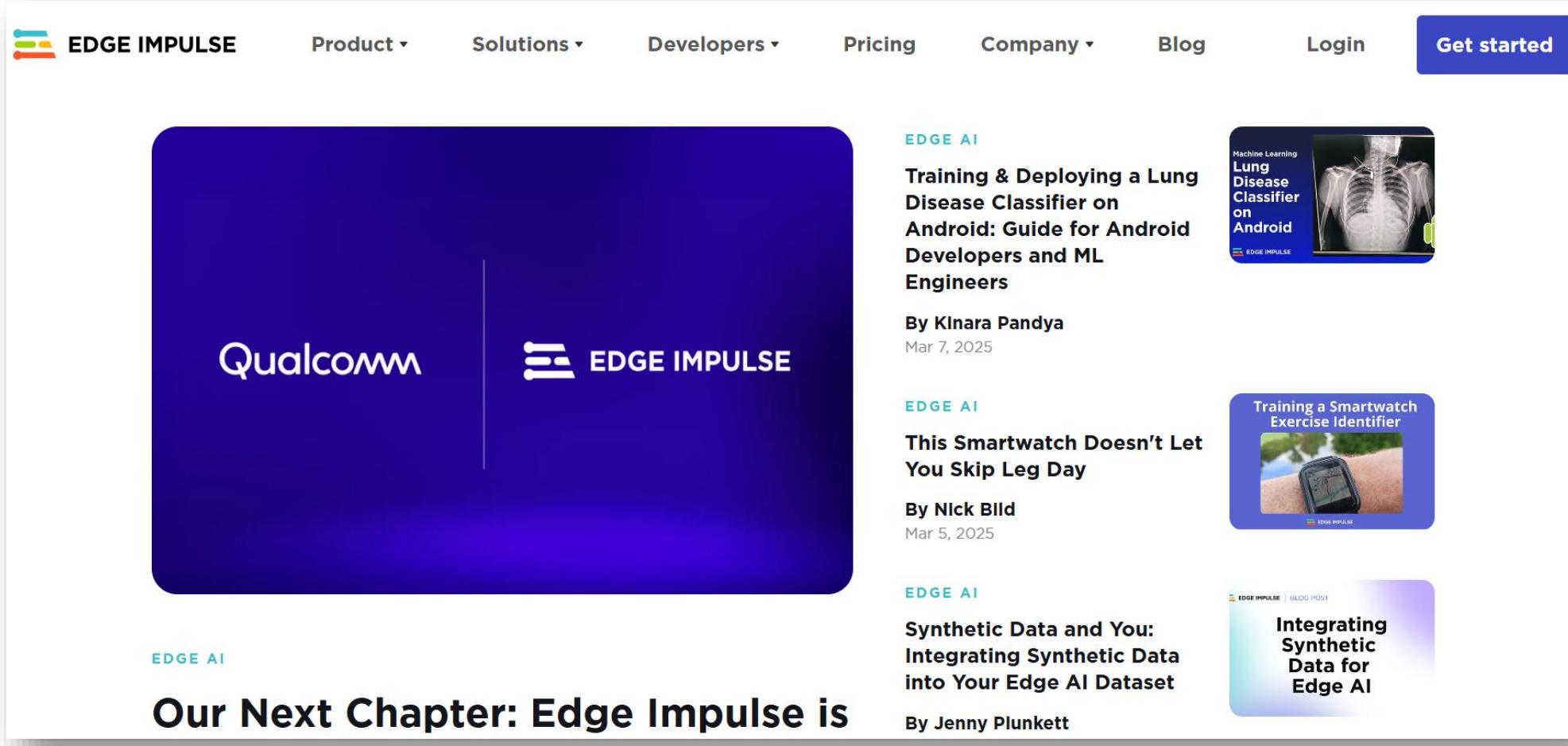
TinyML Projects
(指定開發板查詢會更精準)



The screenshot shows the search results for "tinyml" on Hackster.io. The search bar at the top has "tinyml" entered. The results page lists 457 projects. The "Projects" option is selected in the "Show results for" dropdown. The first few results include "Multi-Model Edge AI on Raspberry Pi w/ Edge Impulse & Blues", "DIY Pomodoro Bot With VIAM", and "Fall Detection For The Elderly".

<https://www.hackster.io/search?q=tinyml&i=projects>

TinyML案例分享 – Edge Impulse Blog

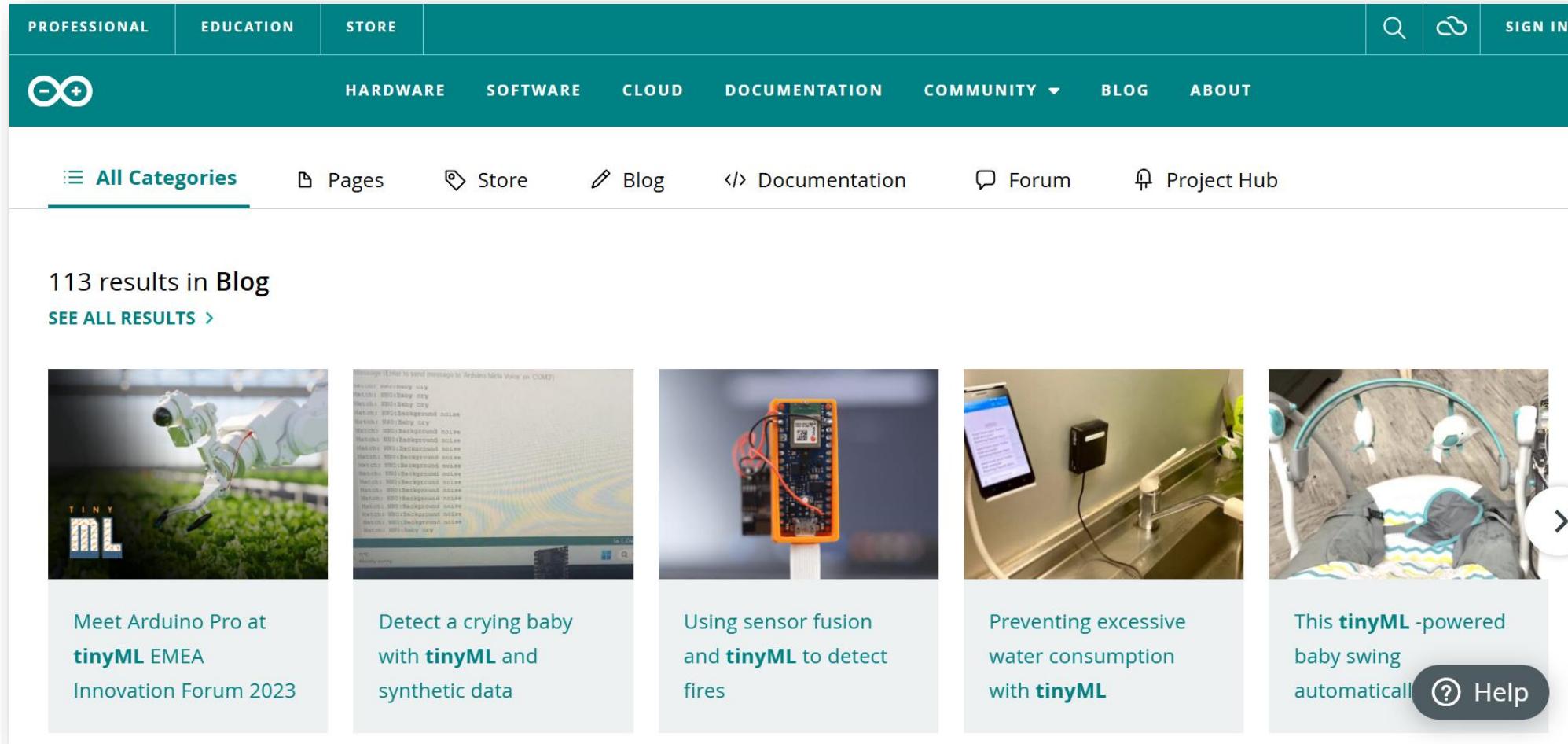


The screenshot shows the Edge Impulse blog homepage. At the top, there is a navigation bar with links for Product, Solutions, Developers, Pricing, Company, Blog, Login, and a prominent blue "Get started" button. Below the navigation bar, there is a large, semi-transparent purple overlay containing the Qualcomm logo and the Edge Impulse logo. The main content area displays three blog posts under the "EDGE AI" category:

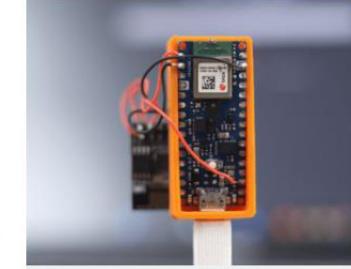
- Training & Deploying a Lung Disease Classifier on Android: Guide for Android Developers and ML Engineers**
By Kinara Pandya | Mar 7, 2025 | 
- This Smartwatch Doesn't Let You Skip Leg Day**
By Nick Bld | Mar 5, 2025 | 
- Synthetic Data and You: Integrating Synthetic Data into Your Edge AI Dataset**
By Jenny Plunkett | 

<https://www.edgeimpulse.com/blog/>

TinyML案例分享 – Arduino TinyML Blog

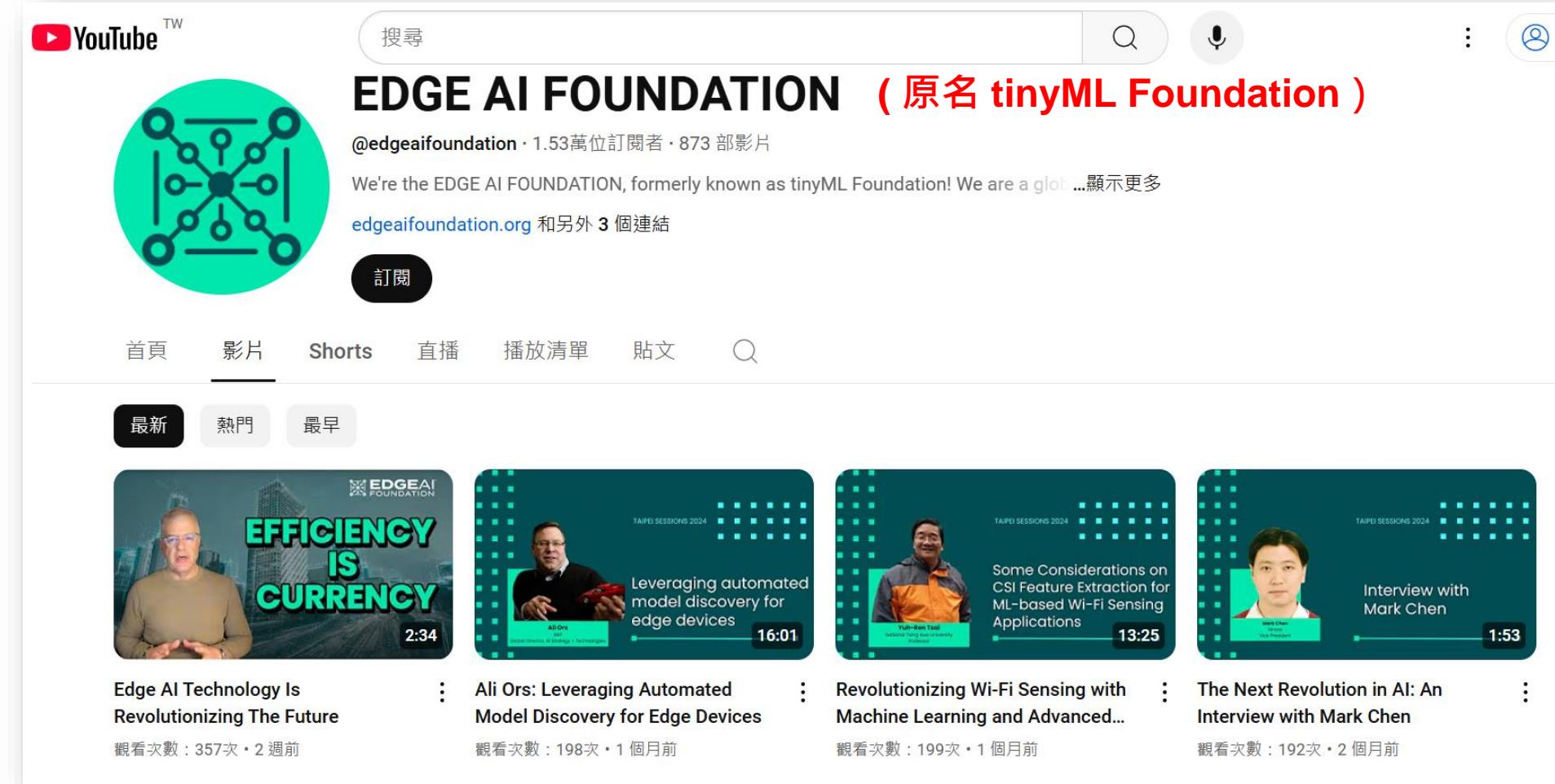


The screenshot shows the Arduino TinyML Blog homepage. At the top, there's a navigation bar with links for PROFESSIONAL, EDUCATION, STORE, and SIGN IN. Below that is a secondary navigation bar with links for HARDWARE, SOFTWARE, CLOUD, DOCUMENTATION, COMMUNITY, BLOG, and ABOUT. A search bar is located at the top right. The main content area displays 113 results in the BLOG category. A "SEE ALL RESULTS >" link is visible. Five featured projects are shown in cards:

-  Meet Arduino Pro at tinyML EMEA Innovation Forum 2023
-  Detect a crying baby with tinyML and synthetic data
-  Using sensor fusion and tinyML to detect fires
-  Preventing excessive water consumption with tinyML
-  This tinyML-powered baby swing automatically [Help]

<https://search.arduino.cc/search?tab=&q=tinyML>

TinyML案例分享 – Edge AI基金會(Youtube)



The screenshot shows the YouTube channel page for 'EDGE AI FOUNDATION (原名 tinyML Foundation)'. The channel has 1.53 million subscribers and 873 videos. The main video thumbnail features a man speaking with the text 'EFFICIENCY IS CURRENCY' and a duration of 2:34. Below it, four other video thumbnails are displayed, each with a different speaker and title related to edge AI technology and machine learning applications.

EDGE AI FOUNDATION (原名 tinyML Foundation)

@edgeaifoundation · 1.53萬位訂閱者 · 873 部影片

We're the EDGE AI FOUNDATION, formerly known as tinyML Foundation! We are a global community of... [顯示更多](#)

[edgeaifoundation.org](#) 和另外 3 個連結

訂閱

首頁 影片 Shorts 直播 播放清單 貼文

最新 熱門 最早

Edge AI Technology Is Revolutionizing The Future

觀看次數：357次 · 2 週前

Ali Ors: Leveraging Automated Model Discovery for Edge Devices

觀看次數：198次 · 1 個月前

Revolutionizing Wi-Fi Sensing with Machine Learning and Advanced...

觀看次數：199次 · 1 個月前

Interview with Mark Chen

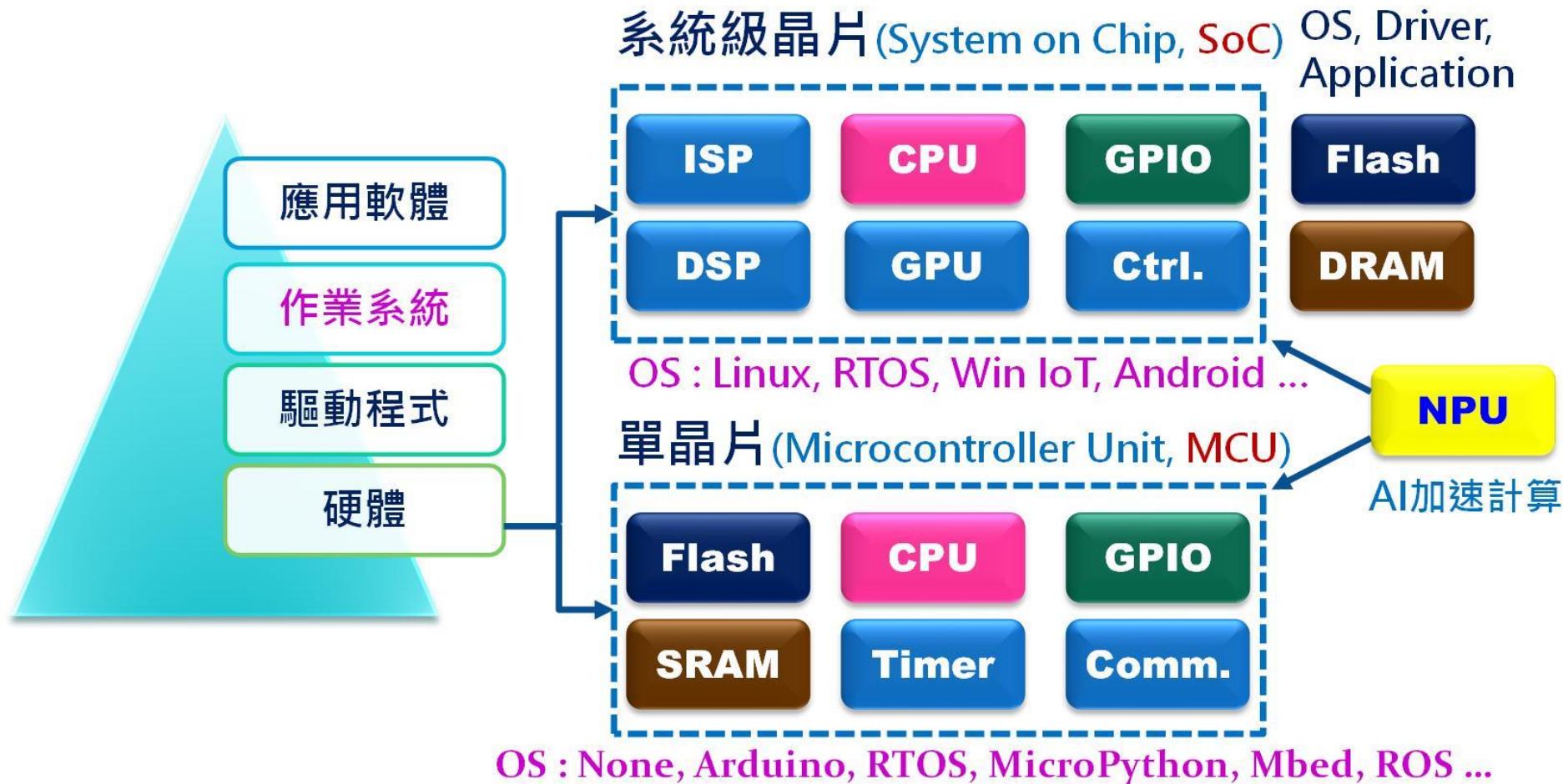
觀看次數：192次 · 2 個月前

<https://www.youtube.com/@edgeaifoundation/videos>



1.3. TinyML 常見開發板

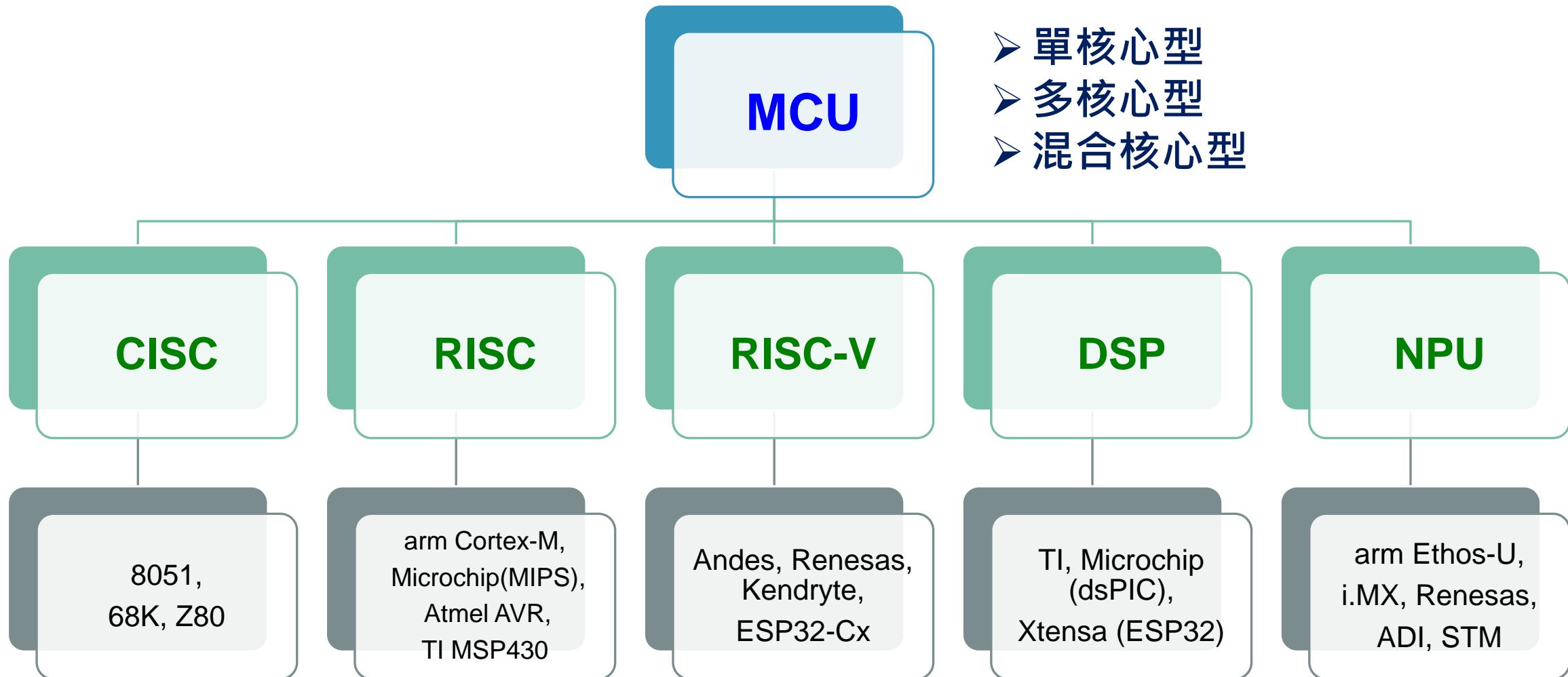
MCU vs. MPU / SoC



資料來源：<https://omnixri.blogspot.com/2021/09/aiottinymlmcu.html>

OmniXRI 整理繪製, 2021/8/14

MCU等級主要核心晶片分類



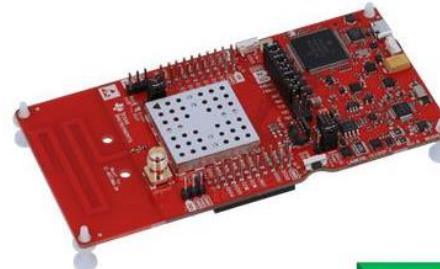
常見tinyML開發板(智能感測)



Raspberry
Pi Pico (RP2040)



Arduino
Nano 33 IoT /
BLE / BLE Sense



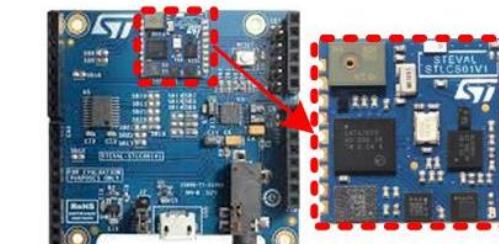
TI
CC1352P Launchpad



Seeed
Xiao BLE / BLE Sense



Silicon Labs
xG24-DK2601B



ST
SensorTile
(STEVAL-STLKT01V1)



Coretronic MEMS
CoreMaker-01



Syntiant
Tiny ML Board

M0+ x2

M4

M4

M4

M33

M4

M4

資料來源：<https://omnixri.blogspot.com/2022/07/20220731coscupaimcutinyml.html>

OmniXRI整理製作, 2022/7/31

常見tinyML開發板(自帶攝影機)



Arducam
Pico4ML (BT)

M0+



OpenMV
Cam H7 Plus

M7



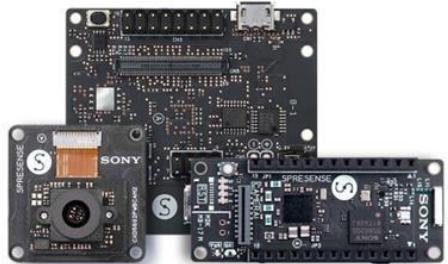
Arduino
Nicla Vision

M7



Arduino
Portenta+Vision Shield

M7



Sony
Spresense

M4 x6



M5Stack
UNITV2-AI-CAMERA-
GC2145

A7 x2



Himax
WE-I Plus

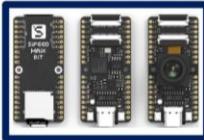
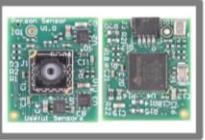


Espressif
ESP-EYE(ESP32)

資料來源：<https://omnixri.blogspot.com/2022/07/20220731coscupaimcutinyml.html>

OmniXRI整理製作, 2022/7/31

常見具視覺TinyML開發板

							
Arducam Pico4ML-BLE	Arduino Nano 33 BLE Kit	Sony Spresense	Silicon Labs xG24 Dev Kit	Realtek AMB82-Mini	Sipeed Maix Bit	Espressif ESP-EYE (ESP32)	Lilygo T-Camera S3
Pi RP2040 MCU : (Cortex-M0+ @133MHz x2) Flash : 2MB SRAM : 264KB Camera : Himax M01B0 (320x240) Others : IMU, BLE, LCD	nRF52840 MCU : (Cortex-M4 @64MHz) Flash : 1MB SRAM : 256KB Camera : *IO OV7675 (640x480)	CXD5602 MCU : (Cortex-M4F @156MHz x6) Flash : 8MB SRAM : 1.5MB Camera : CAM1(2608x1960) CAM2W(1280x960)	EFR32MG24 MCU : (Cortex-M33 @78MHz) Flash : 1536KB SRAM : 256KB Camera : *SPI Arducam Mini 2MP Plus (1632x1232) Mic, IMU, T&H, BP...	RTL8735B MCU : (Armv8-M @500MHz) NPU : 0.4 TOPS Flash : 768KB, 64MB(ext) RAM : DDR 128MB Camera : *(1920x1080) Sensors : Mic	Kendryte K210 MCU : (RISC-V 64bit @400MHz) NPU : 1TOPS Flash : 16MB SRAM : 8MB Camera : OV2640 (1600x1200) Sensors : None	ESP32 MCU : (Xtensa LX6 x2 @240 MHz + ULP FSM) Flash : 448K, 4MB(ext) RAM : S:536KB, PS:8MB Camera : OV2640 (1600x1200) Others : Mic, WiFi, BLE	ESP32-S3 MCU : (Xtensa LX7 x2 @240MHz + ULP RISC-V+ ULP FSM) Flash : 384KB, 16MB(ext) RAM : S:512KB, PS:8MB Camera : OV2640 (1600x1200) Others : Mic, WiFi, BLE
							
OpenMV Cam H7 Plus	Arduino Protenta H7+Vision	Arduino Nicla Vision	Alif Ensemble E7 DK	Eta Compute ECM3532-AVBK	Himax WE-I Plus EVB	Seeed Studio Grove- Vision AI	Useful Sensors Person Sensor
STM32H743II MCU : (Cortex-M7 @480MHz) Flash : 2MB, 32MB(ext) RAM : S:1MB, SD:32MB Camera : OV5640 (2592x1944) Sensors : None	STM32H747XI MCU : (CM7@480MHz+ CM4@240MHz) Flash : 2MB+16MB(ext) RAM : S:1MB+SD:8MB Camera : Himax HM-01B0 (320x240) Others : Mic, WiFi/BT, EtherNet	STM32H747AI6 MCU : (CM7@480MHz+ CM4@240MHz) Flash : 2MB+16MB(ext) SRAM : 1MB Camera : GC2145 (1616x1232) Sensors : IMU, Mic, ToF	Cortex-M55 x2 @160M / 400 MHz NPU : Ethos-U55 x2 128 / 256 MAC MPU : Cortex-A32 x2 @800 MHz MRAM : 5.5MB SRAM : 13.5MB Camera : * MIPI Interface	ECM3532 (Cortex-M3) DSP : CoolFlux DSP16 2MAC@100MHz Flash : 512KB, 8MB(ext) SRAM : 352KB Camera : Himax HM0360 (640x480) Sensors : IMU, Mic, ALS	HX6537-A MCU : (ARC EM9D DSP @400MHz) Flash : 2MB SRAM : 2MB Camera : Himax HM0360 (640x480) Sensors : IMU, Mic	HX6537-A MCU : (ARC EM9D DSP @400MHz) Flash : 2MB SRAM : 2MB Camera : OV2640 (1600x1200) Others : Mic, IMU, BLE	HX6537-A MCU : (ARC EM9D DSP @400MHz) Flash : 2MB SRAM : 2MB Camera : Himax HM0360 (640x480) Sensors : None

 Cortex-M0+ Cortex-M3 Cortex-M4 Cortex-M7 Cortex-M33/M35 Cortex-M55 RISC-V DSP

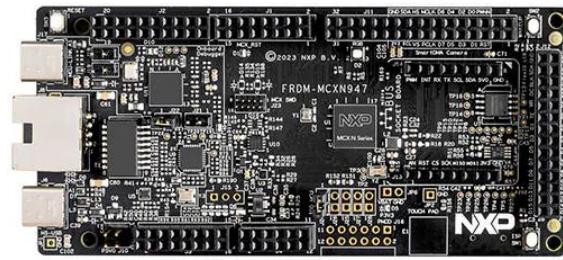
OmniXRI整理製作 · 2022/12/29

資料來源：<https://omnixri.blogspot.com/2022/12/tinymlmcu.html>

TinyML MCU 等級相關開發板

NXP MCX-N FRDM-MCXN947

- Cortex-M33 x2
- eIQ NPU x1



Nordic nRF9160 SiP Thingy:91

- Cortex-M33 x1



Raspberry Pi Pico 2W

- Cortex-M33 x2
- RISC-V Hazard3 x2



Seeed Xiao ESP32-S3 Sense

- Xtensa LX7 (DSP) x2



Seeed Wio Terminal

- Cortex-M4F x1



Seeed SenseCap Watcher

- ESP32-S3 x1
- Xtensa LX7 (DSP) x2
- Himax WE2 x 1
- arm Cortex-M5 x2
- arm Ethos-U55 x1



Arm Cortex-M55/M85 + Ethos-U55/U65

ALIF Ensemble E7 AI/ML AppKit

- Cortex-M55 x2
- Ethos-U55 x2
- Cortex-A32 x2



新唐(Nuvoton) NuMaker-M55M1

- Cortex-M55 x1
- Ethos-U55 x1



Seeed Grove Vision AI v2 Kit (奇景 Himax WiseEye2 HX6538)

- Cortex-M55 x2
- Ethos-U55 x1



STM STM32-N6 STM32N6570-DK

- Cortex-M55 x1
- Neural-ART(NPU) x1



Infineon PSoC Edge Development Kit

- Cortex-M55 x1
- Ethos-U55 x1
- Cortex-M33 x1



NXP i.MX93 Evaluation Kit

- Cortex-A55 x2
- Cortex-M33 x1
- Ethos-U65 x1



Renesas EK-RA8M1 Evaluation Kit

- Cortex-M85 x1

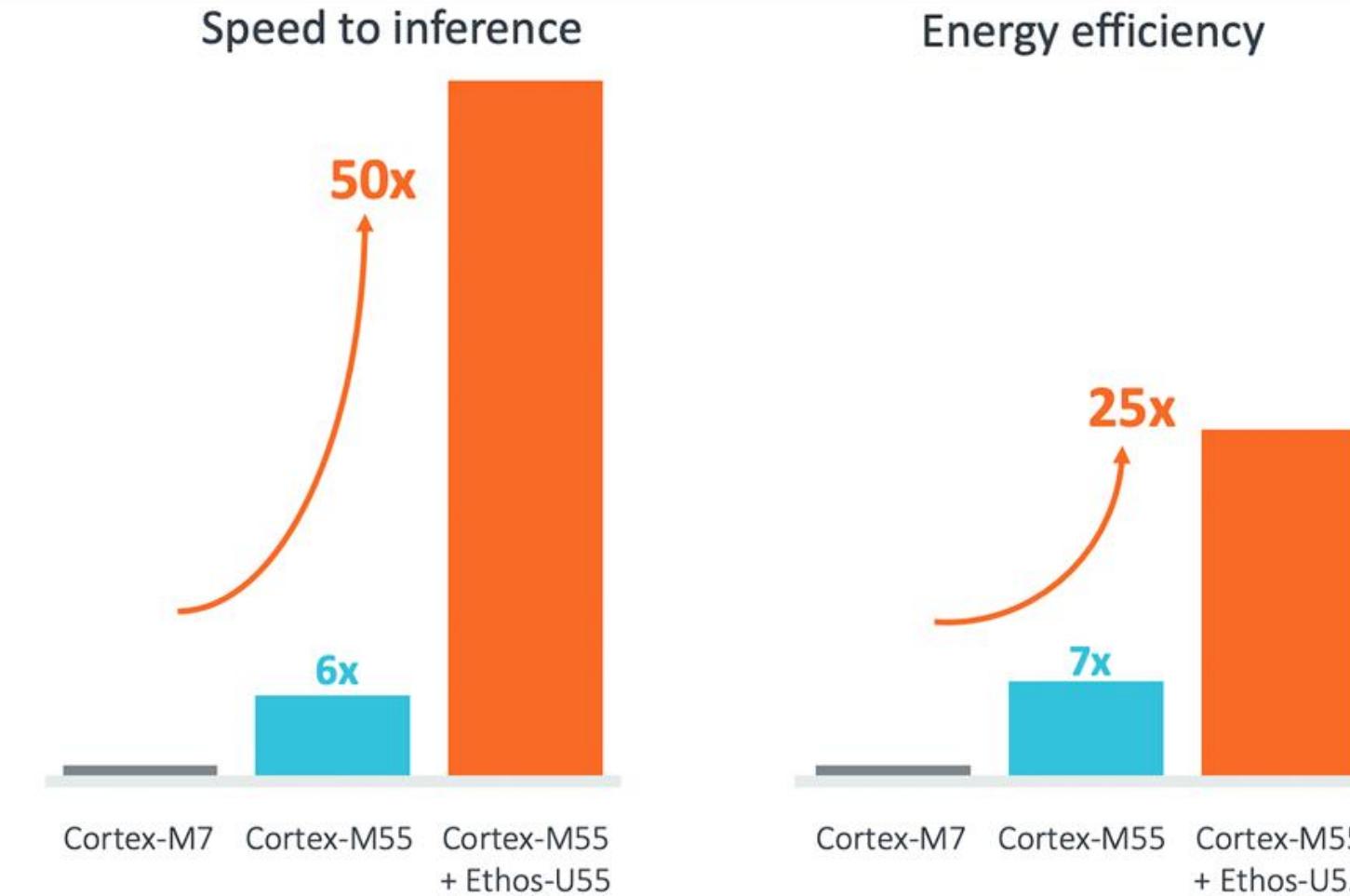


Ambiq Apollo 510

- Cortex-M55 x1



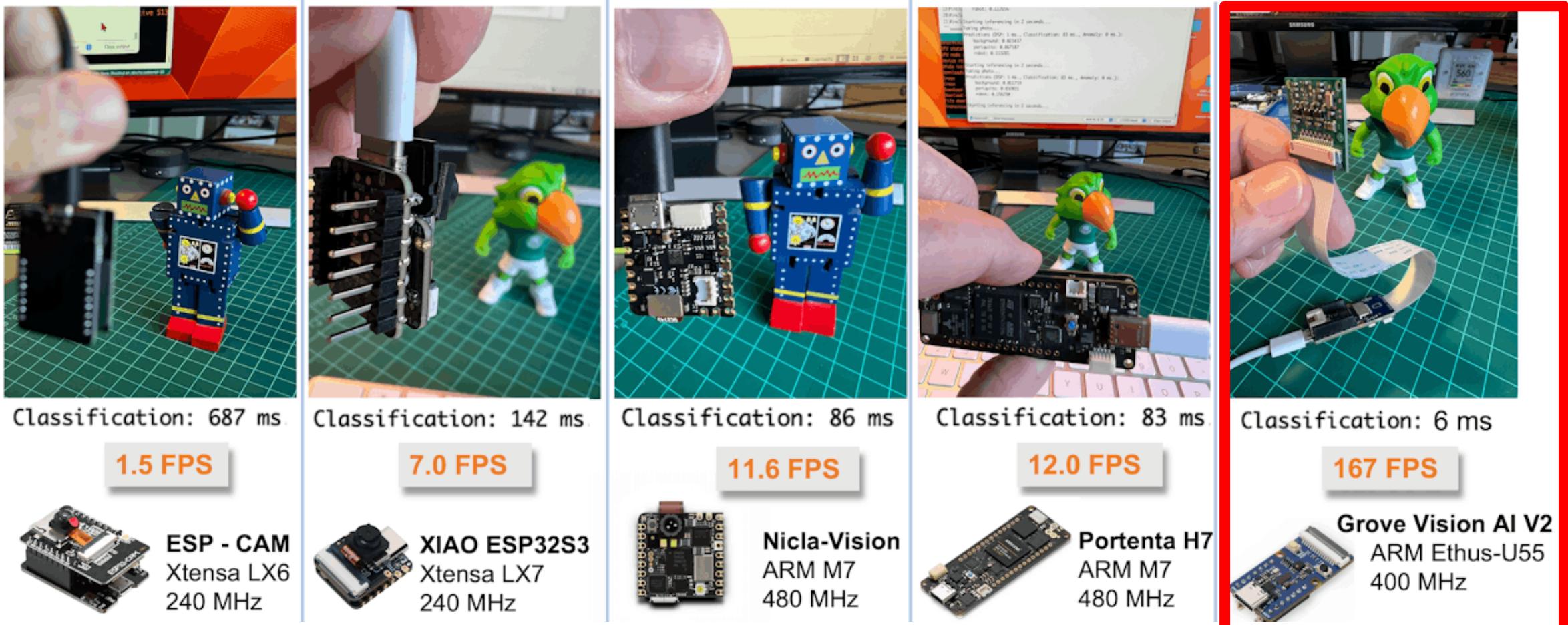
單晶片推論速度與耗能比較



資料來源：<https://community.arm.com/arm-community-blogs/b/architectures-and-processors-blog/posts/cortex-m55-and-ethos-u55-processors-extending-the-performance-of-arm-ml-portfolio-for-endpoint-devices>

各式TinyML視覺開發板效能比較

評比方式：**MobilenetV2 96x96**



資料來源：<https://www.hackster.io/mjrobot/computer-vision-at-the-edge-with-grove-vision-ai-module-v2-0003c7>



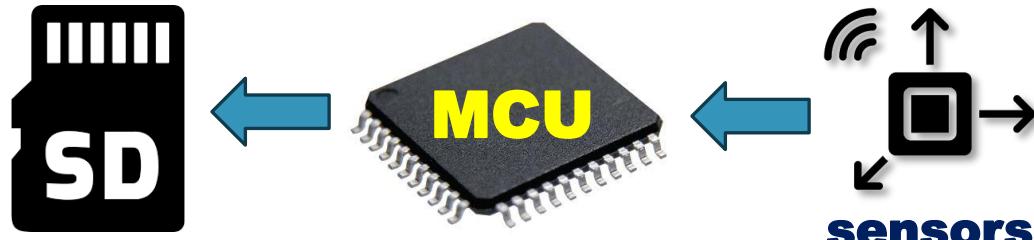
1.4. TinyML 常見開發工具

TinyML 應用開發流程



資料集建置工具

感測器資料集建置工具



Data Forwarder



Data Capture Lab



影像資料集建置工具



CVAT
<https://www.cvat.ai/>



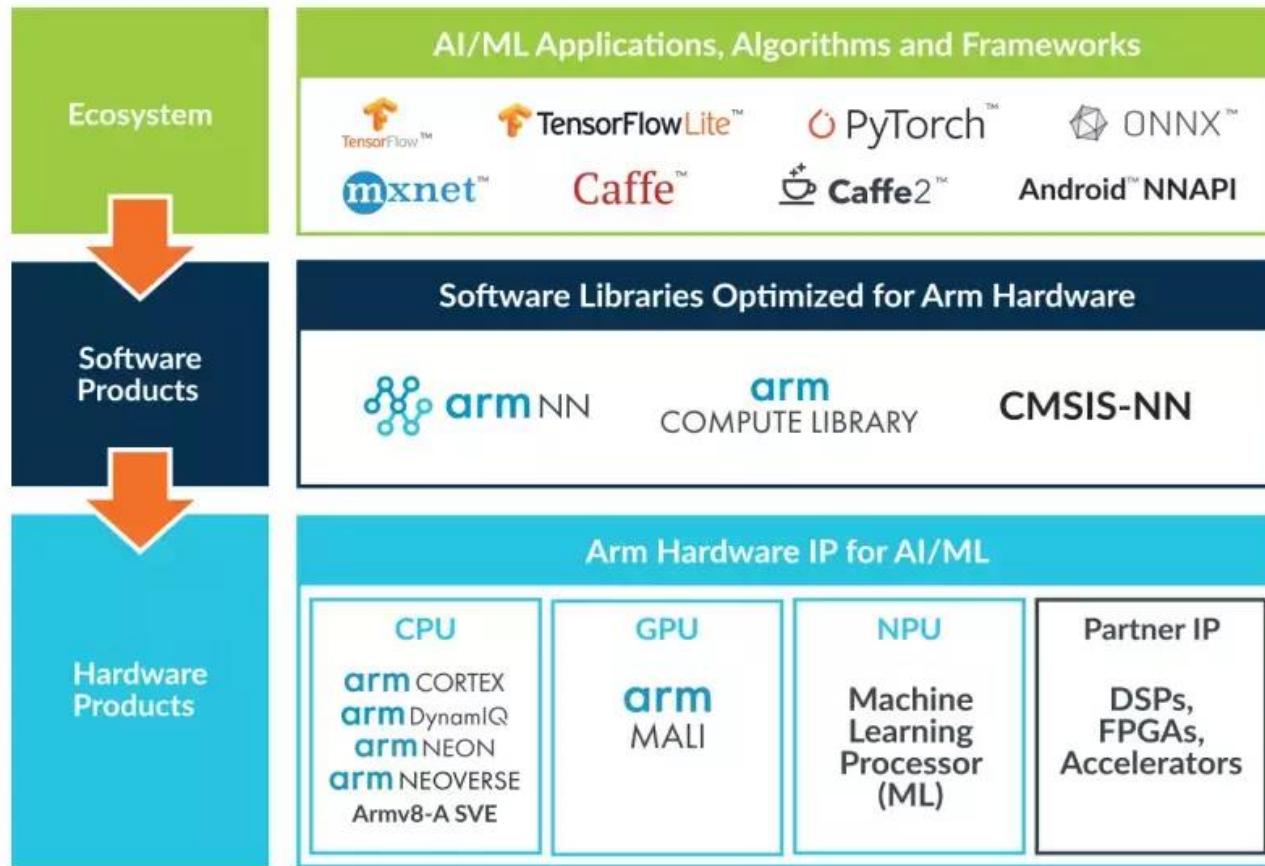
Roboflow
<https://roboflow.com/>



SenseCraft AI
<https://sensecraft.seeed.cc/ai/>

開發框架與工具 (MCU)

Arm Based MCU / MPU / GPU / NPU 通用型



MCU 專用 (半通用) 型

- ST : CubeAI, Nano Edge
- NXP : eIQ
- Renesas : Reality AI
- Nordic : Atlazo
- TDK : Qeexo
- Nuvoton : NuEdgeWise
- QuickLogic : SensiML
- Infineon : Imagimob

TinyML 開發平台與技術供應商



開發框架 (CPU / GPU / NPU / MCU)



TensorFlow

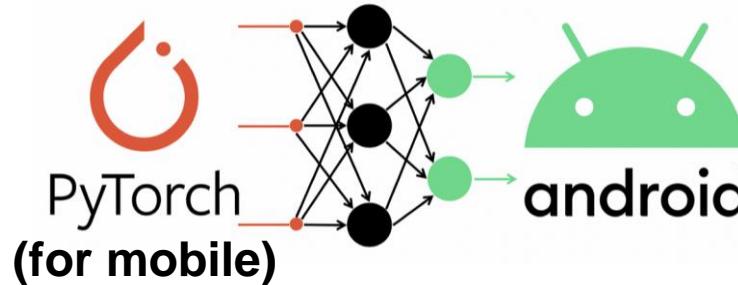


TensorFlow Lite



TensorFlow Lite
(for microcontroller)

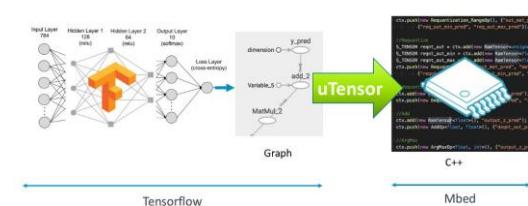
PyTorch



ExecuTorch

獨立工具

uTensor



uTensor

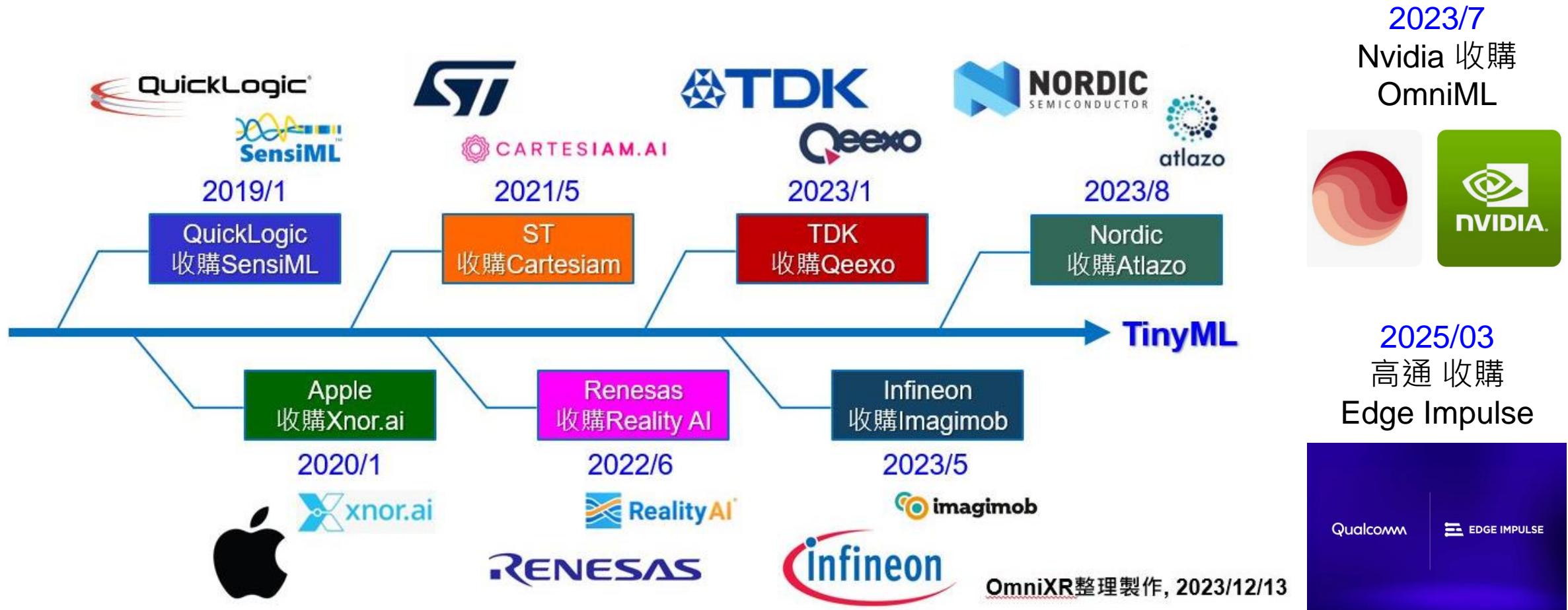


microTVM



AlfES

MCU硬體廠商收購TinyML軟體開發商



ML Commons (MLPerf) Benchmarks:Tiny

評測情境

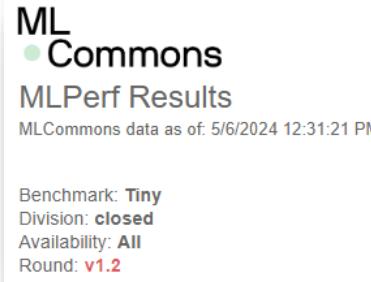


ML
• Commons

Scenario	Query Generation	Duration	Samples/query	Latency Constraint	Tail Latency	Performance Metric
Single stream	LoadGen sends next query as soon as SUT completes the previous query	1024 queries and 60 seconds	1	None	90%	90%-ile measured latency
Multiple stream (1.1 and earlier)	LoadGen sends a new query every <i>latency constraint</i> if the SUT has completed the prior query, otherwise the new query is dropped and is counted as one overtime query	270,336 queries and 60 seconds	Variable, see metric	Benchmark specific	99%	Maximum number of inferences per query supported
Multiple stream (2.0 and later)	LoadGen sends next query, as soon as SUT completes the previous query	270,336 queries and 600 seconds	8	None	99%	99%-ile measured latency
Server	LoadGen sends new queries to the SUT according to a Poisson distribution	270,336 queries and 60 seconds	1	Benchmark specific	99%	Maximum Poisson throughput parameter supported
Offline	LoadGen sends all queries to the SUT at start	1 query and 60 seconds	At least 24,576	None	N/A	Measured throughput

<https://mlcommons.org/benchmarks/inference-tiny/>

ML Commons (MLPerf) Benchmarks: Tiny



評測結果範例 (Open / Close)

Version	Division	Availability	Host Processor Frequency	Model MLC	Accelerator	Organization	Processor
v1.2	closed	(All)	(All)	(All)	(All)	(All)	(All)

Public ID	Organization	Availability	System Name (Click + f..)	Board Name	Host Processor ..	Software	Processor	Accelerator
1.2-0001	Qualcomm	Available	Snapdragon_8_Gen_3_..	Snapdragon 8 Gen 3 Mobile HDK	Prime Core, up to 3.4Ghz, 5 Performance Cor..	Qualcomm AI Stack	Qualcomm Kryo CPU	Qualcomm Sensing Hub

Benchmark / Model MLC / Units							
Tiny							
Deep AutoEncod..		DSCNN		MobileNetV1 0.25x		ResNet-V1	
Energy..	Latenc..	Energy..	Latenc..	Energy..	Latenc..	Energy..	Latenc..
		0.07		0.09		0.19	

<https://mlcommons.org/benchmarks/inference-tiny/>

評測項目

Task	Dataset	Model	Mode	Quality	Latest Version Available
Keyword Spotting	Google Speech Commands	DS-CNN	Single-stream, Offline	90% (Top 1)	v1.1
Visual Wake Words	Visual Wake Words Dataset	MobileNetV1 0.25x	Single-stream	80% (Top 1)	v1.1
Image classification	CIFAR10	ResNet-8	Single-stream	85% (Top 1)	v1.1
Anomaly Detection	ToyADMONS	Deep AutoEncoder	Single-stream	0.85 (AUC)	v1.1

參考文獻

- 許哲豪，臺灣科技大學資訊工程系「人工智慧與邊緣運算實務」(2021~2023)
<https://omnixri.blogspot.com/p/ntust-edge-ai.html>
- 許哲豪，OmniXRI's Edge AI & TinyML 小學堂 Youtube 直播課程總結
<https://omnixri.blogspot.com/2024/06/omnixris-edge-ai-tinyml-youtube.html>
- 許哲豪，歐尼克斯實境互動工作室系列發文—TinyML(MCU AI)系列
<https://hackmd.io/1PK1URhIQ7GutcWgpgsWbg#TinyMLMCU-AI%E7%B3%BB%E5%88%97>



只有更邊



Q&A



歐尼克斯實境互動工作室
(OmniXRI Studio)
許哲豪 (Jack Hsu)

Facebook : Jack Omnidri
FB社團 : Edge AI Taiwan 邊緣智能交流區
電子信箱 : omnixri@gmail.com
部落格 : <https://omnidri.blogspot.tw>
開 源 : <https://github.com/OmniXRI>

YOUTUBE 直播 : <https://www.youtube.com/@omnidri1784streams>