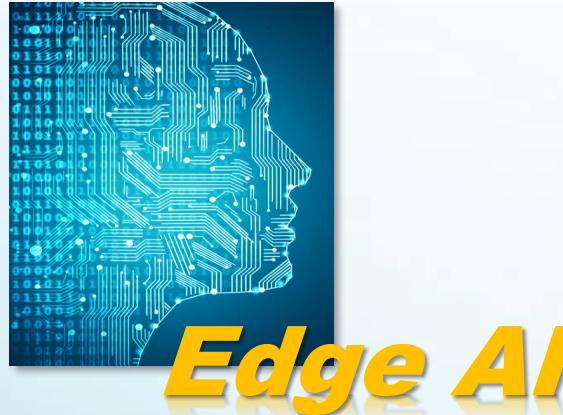
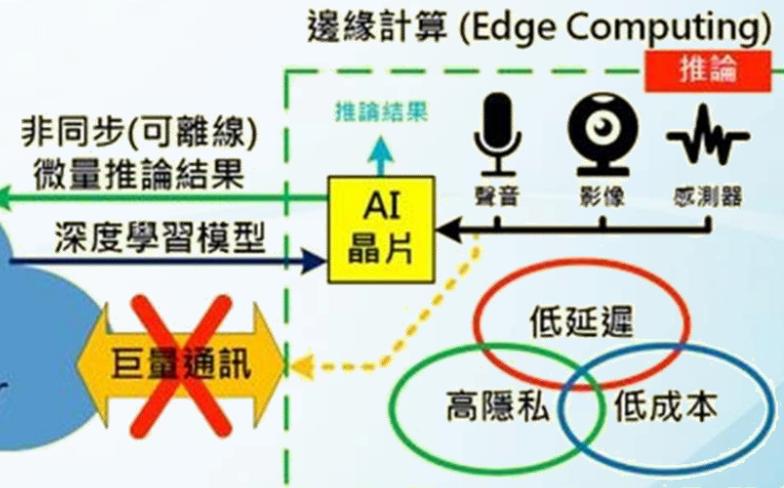
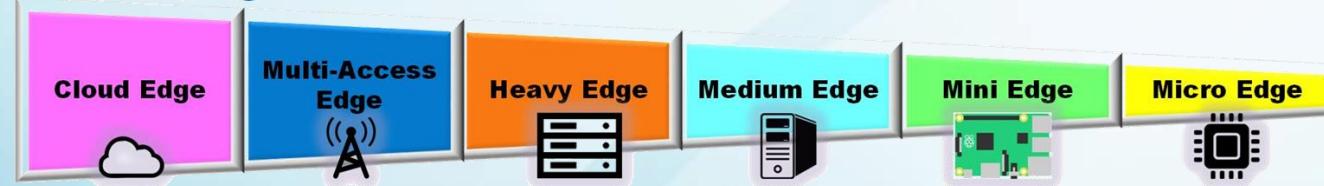


邊緣人工智能實務 (EE5354701)



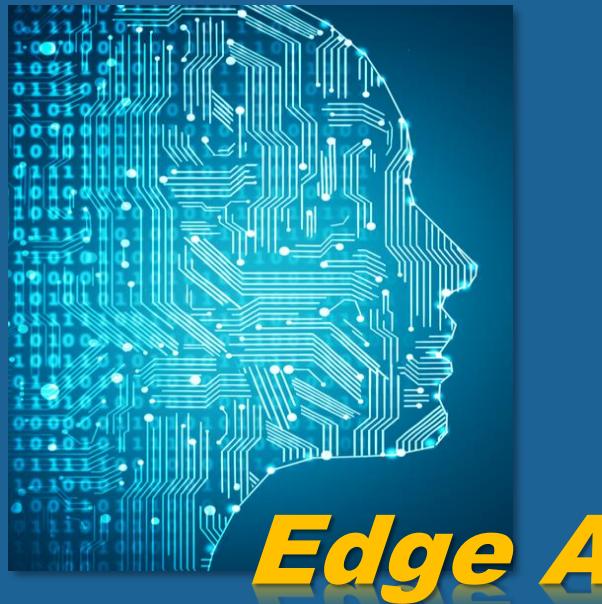
邊緣等級(Edge Level)



1. 邊緣人工智能簡介

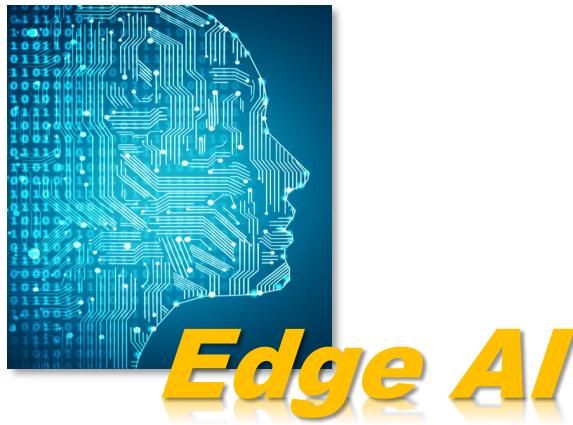
電機工程系 人工智慧應用產碩專班 許哲豪 兼任助理教授

簡報大綱

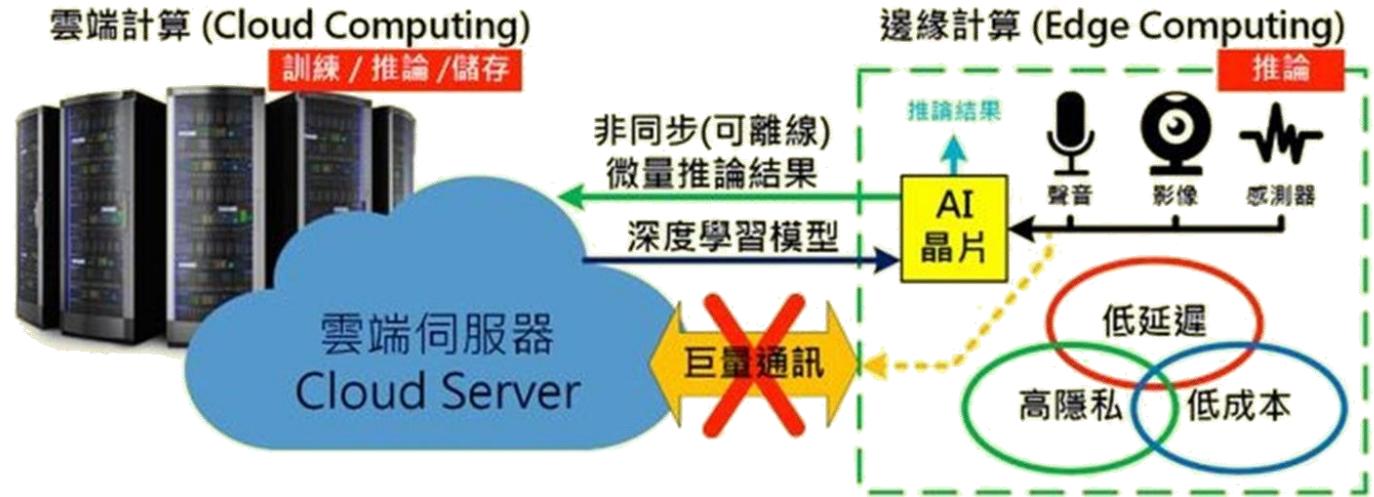
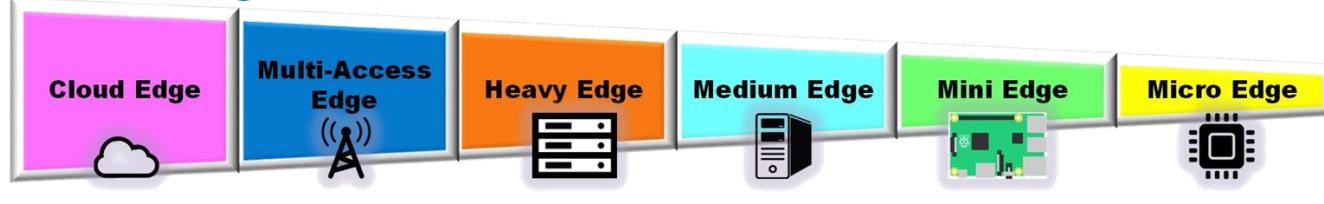


- 1.1. 邊緣智慧發展歷史
- 1.2. 邊緣智慧基本原理
- 1.3. 邊緣智慧應用情境

https://hackmd.io/@OmniXRI-Jack/NTUST_EdgeAI_2026

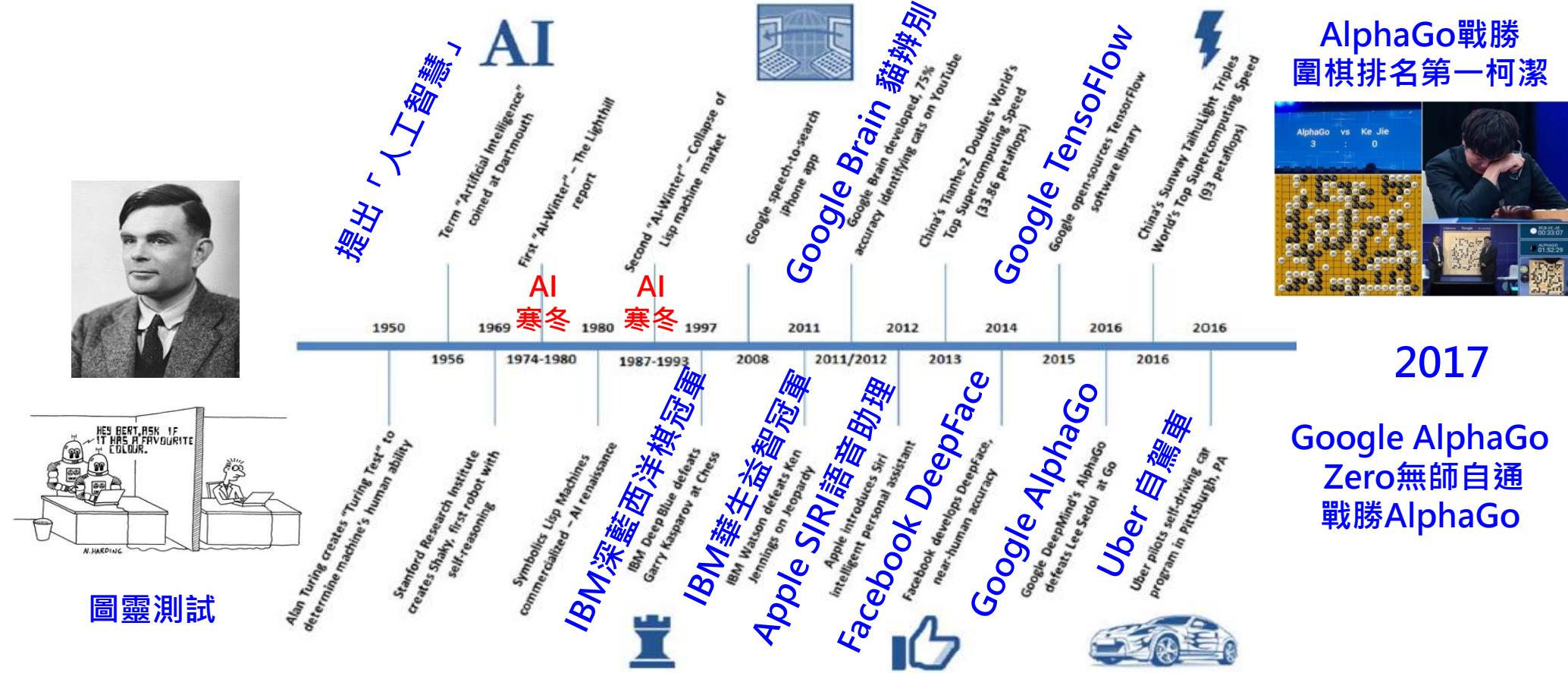


邊緣等級(Edge Level)



1.1. 邊緣智慧發展歷史

AI 發展重要里程碑 (1/2)



資料來源：<http://www.smallake.kr/wp-content/uploads/2017/05/P020161223538320477062.pdf>

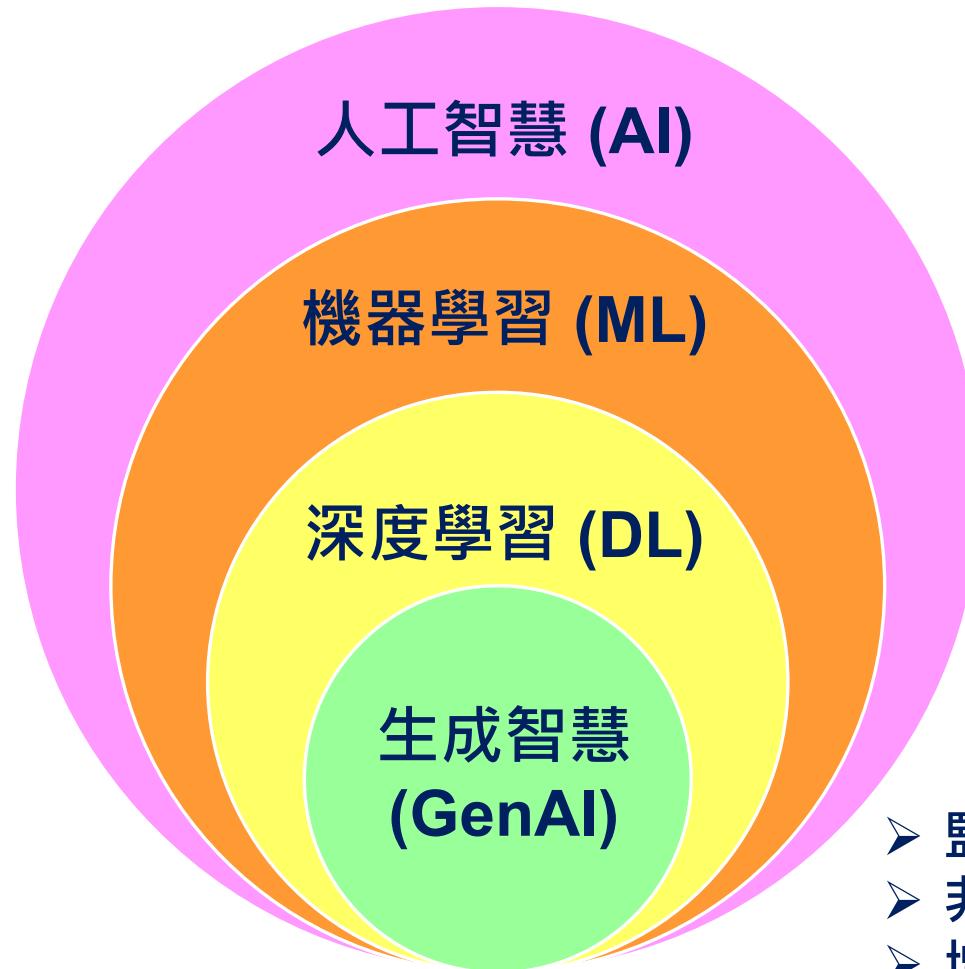
AI 發展重要里程碑 (2/2)



AI 發展寒冬與爆發



人工智慧與主要基石



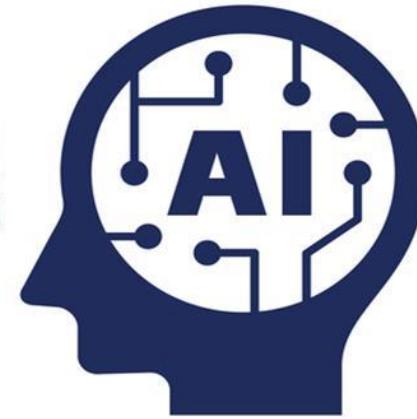
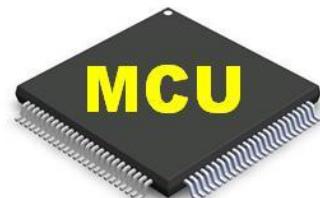
缺一不可

- 監督式學習 (分類、迴歸)
- 非監督式學習 (聚類)
- 增強學習 (分類)

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

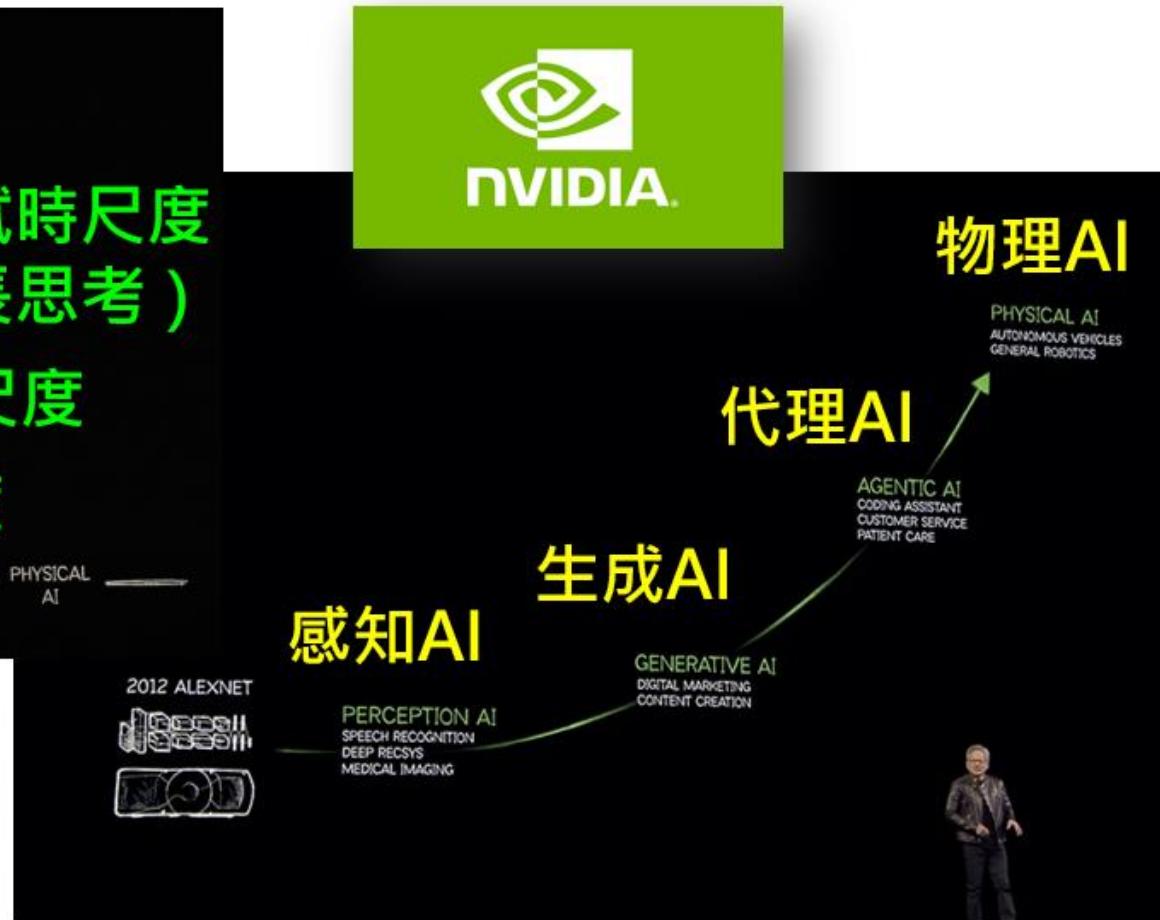
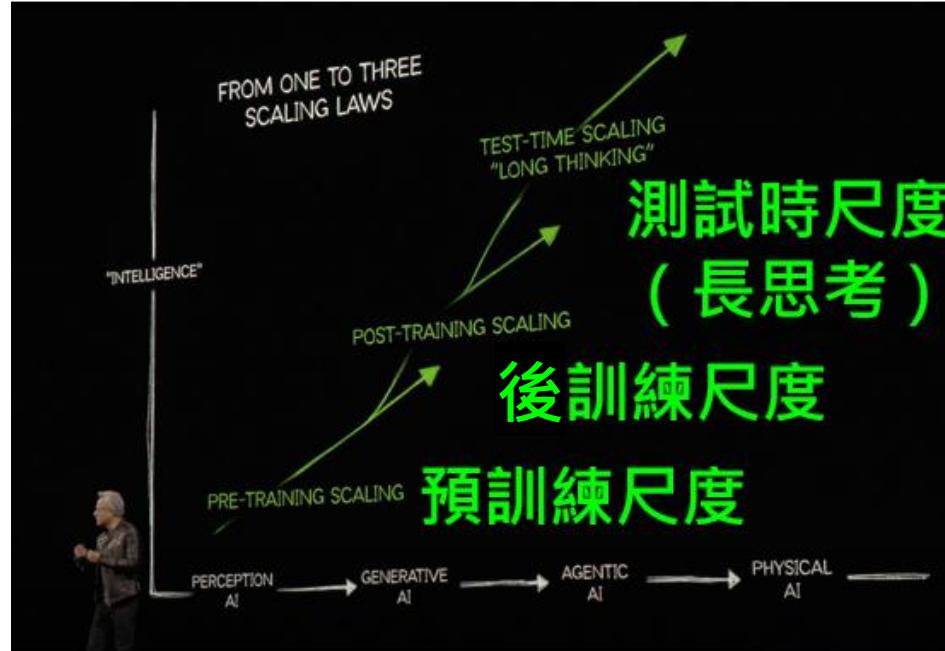


極大化

- 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>

AI 發展趨勢



- AI 核心技術
- 半導體技術
- 計算能力
- 開發工具
- 基礎模型
- 多模態技術

資料來源：<https://omnixri.blogspot.com/2025/03/20250327.html>

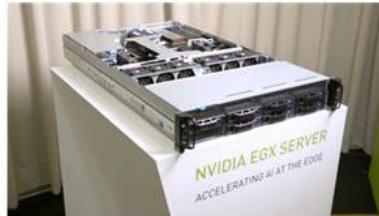
Edge AI 誰說了算？

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。

Edge Server



Nvidia EGX



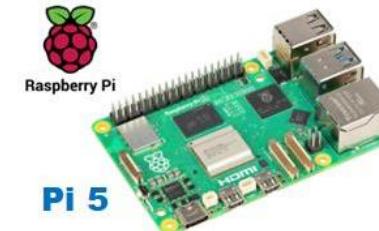
NB AI (AI PC)



Mobile AI (SoC)



SBC AI (MPU / NPU / GPU)



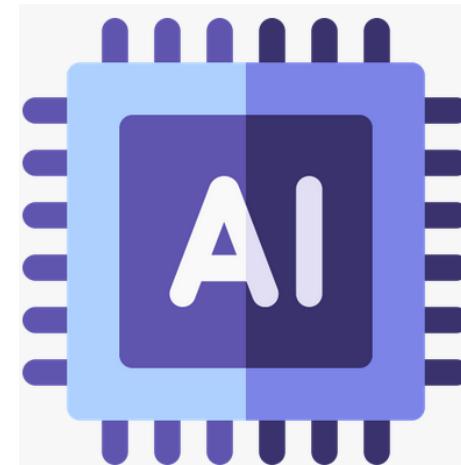
Wearable AI (MCU / TinyML)



Edge AI 如何定義？

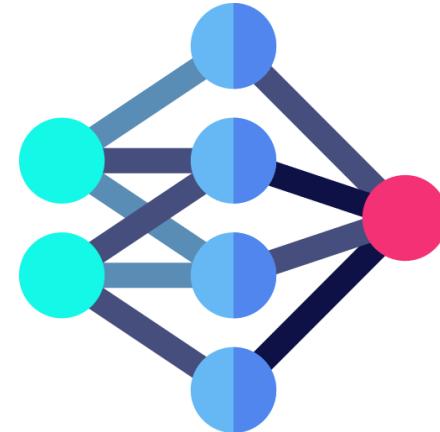
廣義定義：

不上網就能完成所有**AI推論**



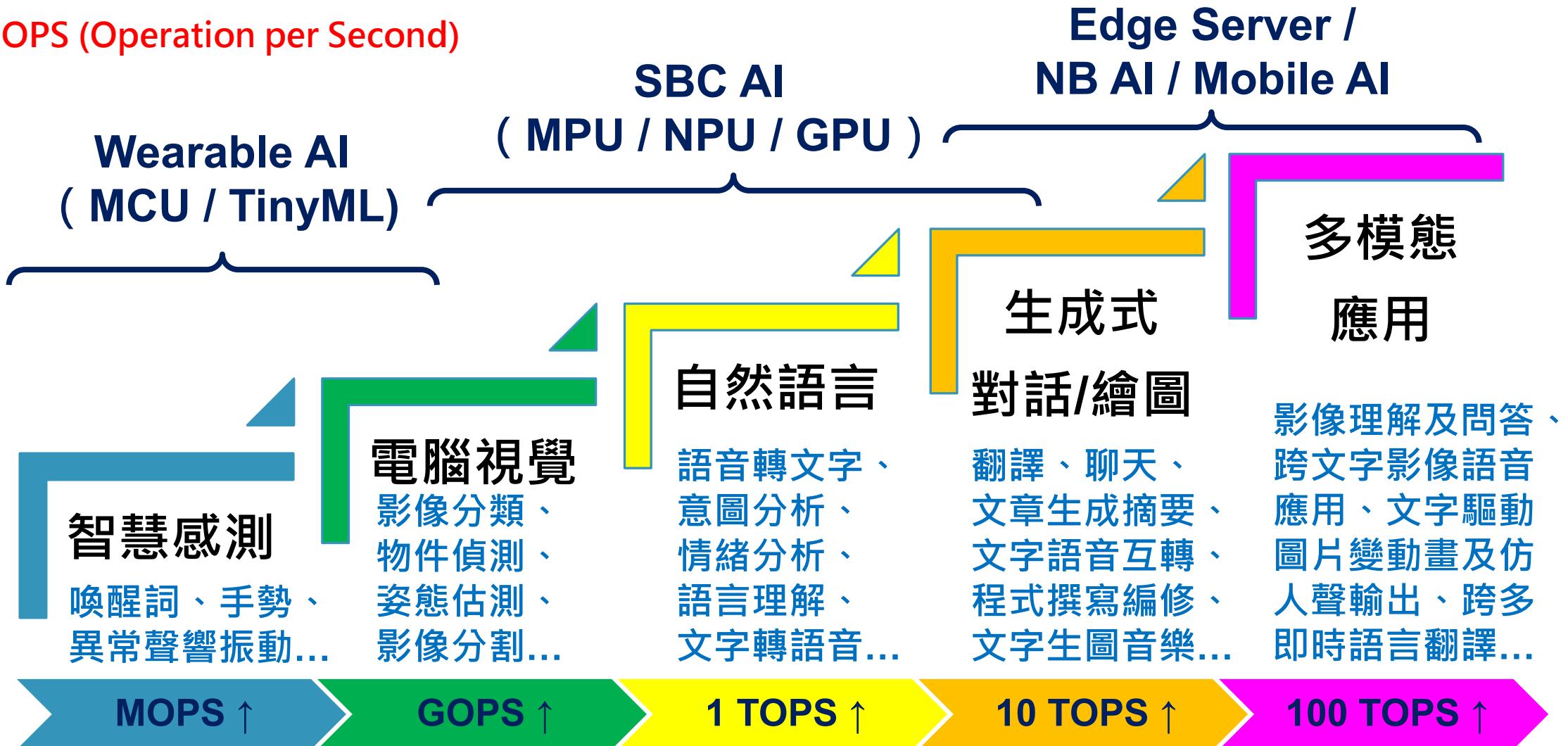
狹義定義：

使用**電池供電**完成所有**AI推論**
的**獨立裝置（Endpoint Device）**

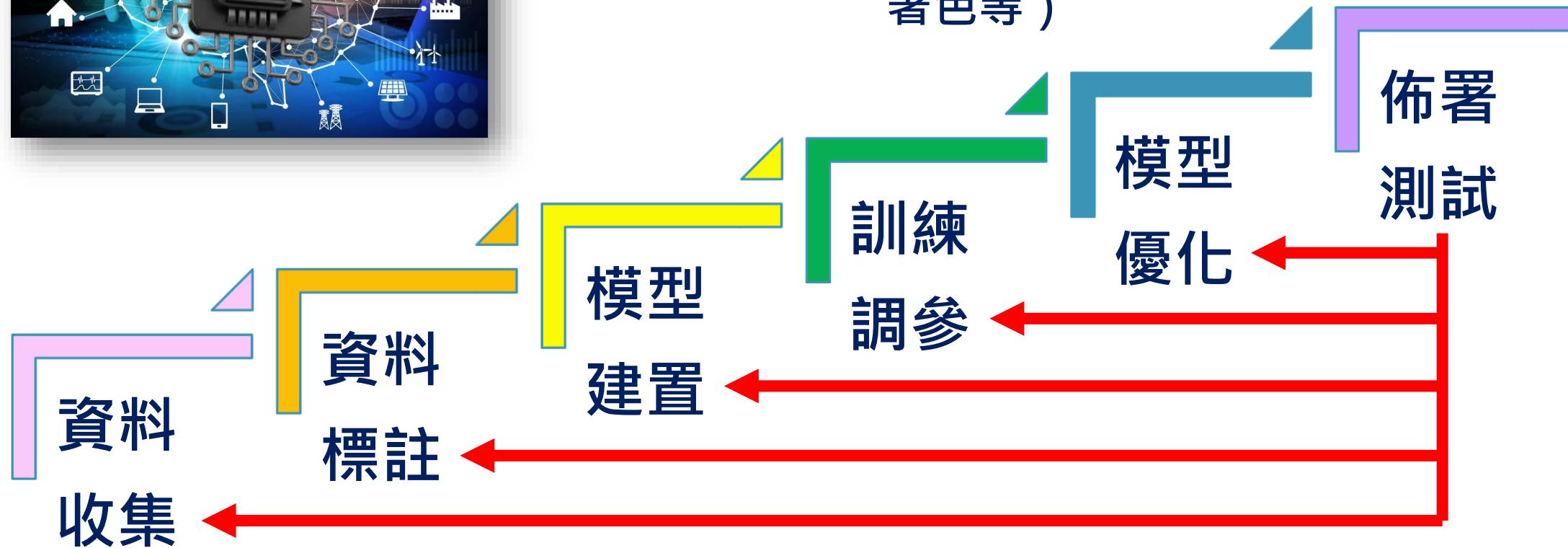


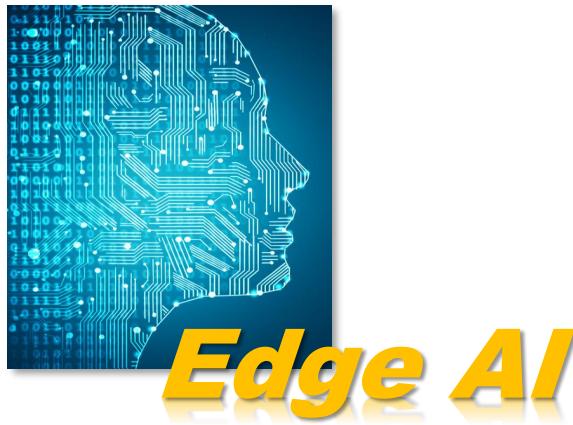
應用與算力需求

* OPS (Operation per Second)

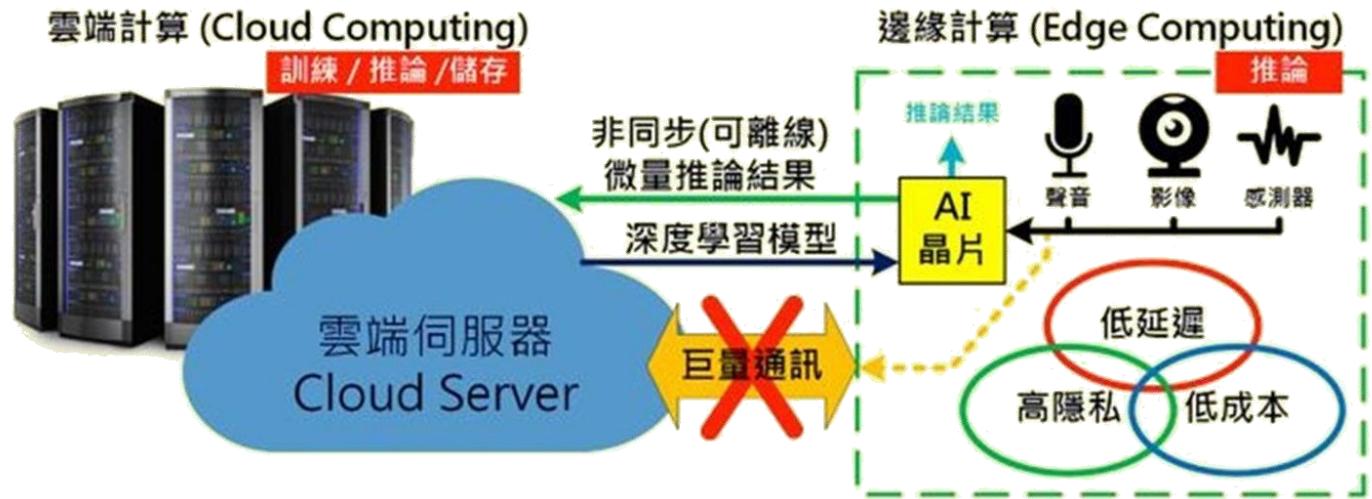
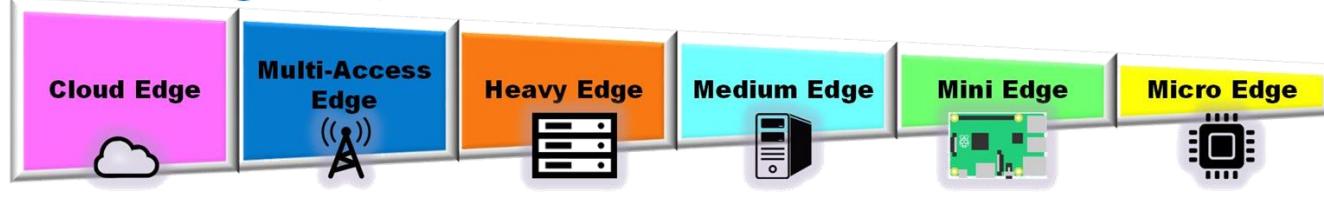


Edge AI 開發流程



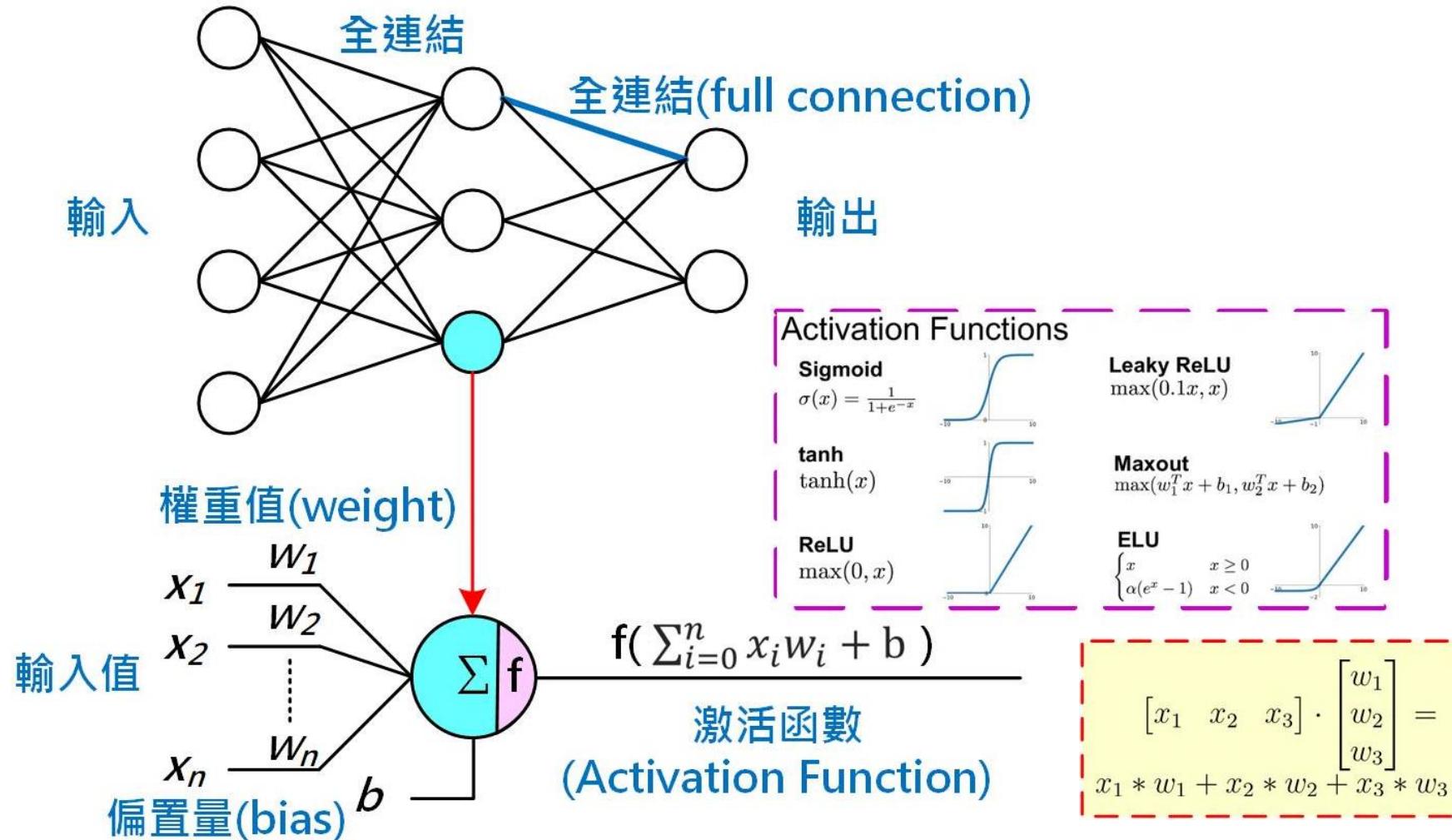


邊緣等級(Edge Level)



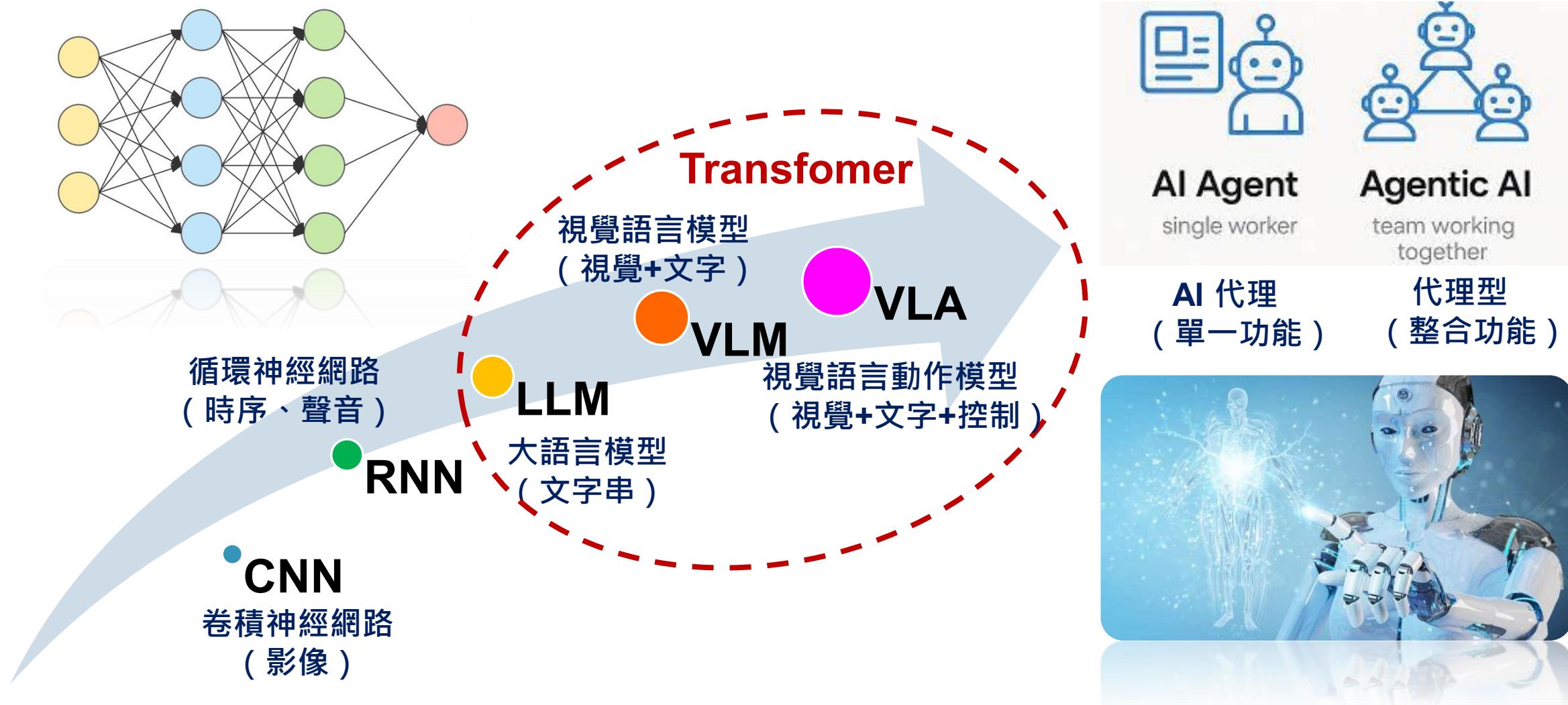
1.2. 邊緣智慧基本原理

神經元與全連結神經網路



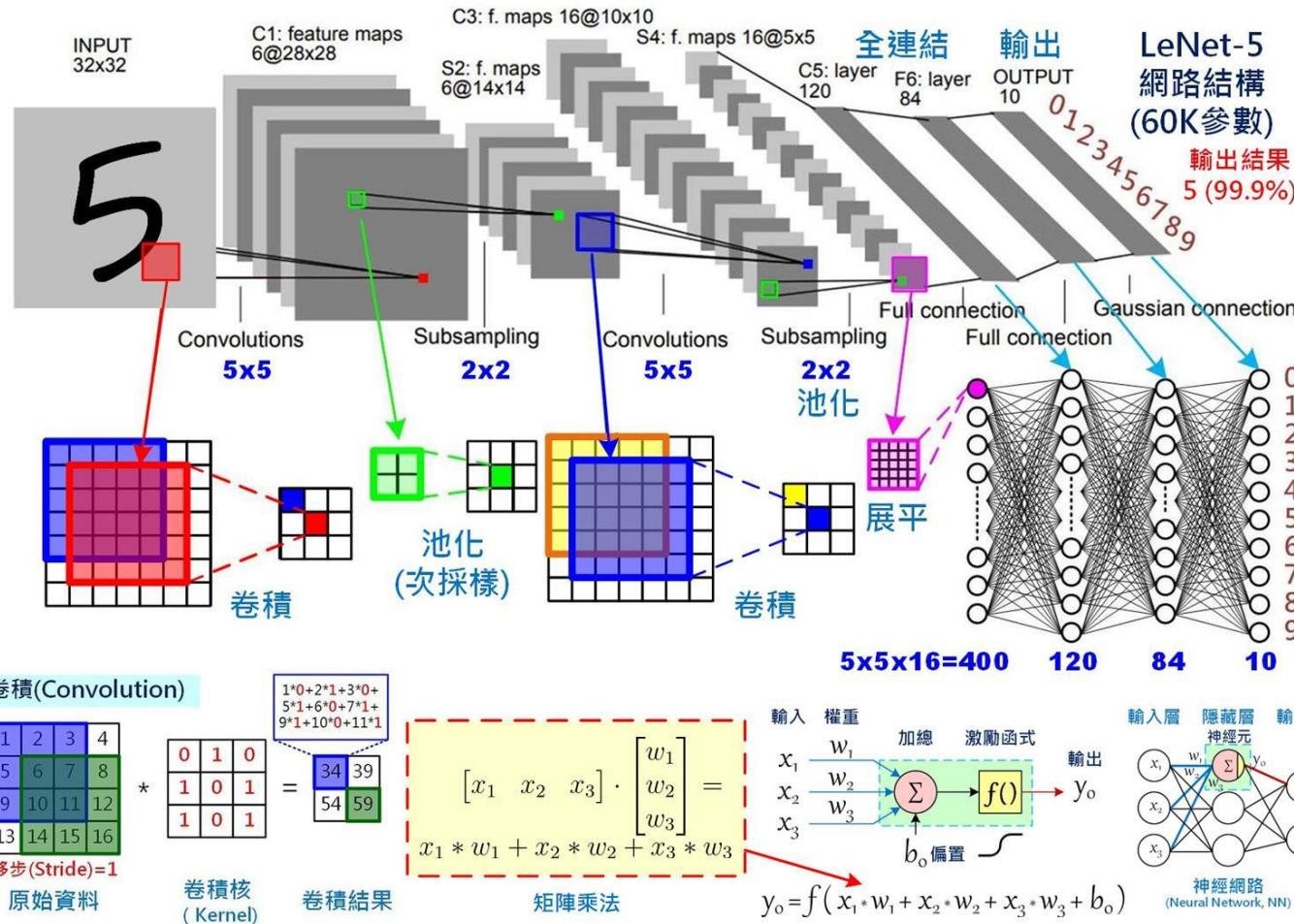
* 相當於Python numpy 的矩陣點乘 `a.dot(b)` 而非矩陣各元素相乘的 `a*b`

AI（神經網路）技術演進



卷積神經網路 (CNN)

Convolution Neural Network (CNN)



$$Y = A * X + B$$

乘積累加運算 MAC
(Multiply Accumulate)

硬體加速方式 (算力)

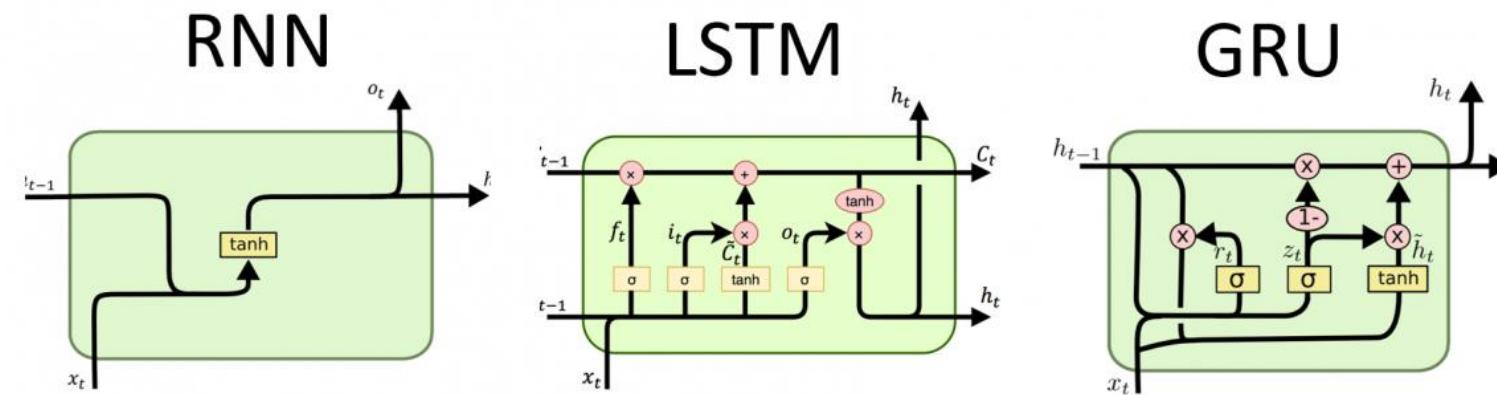
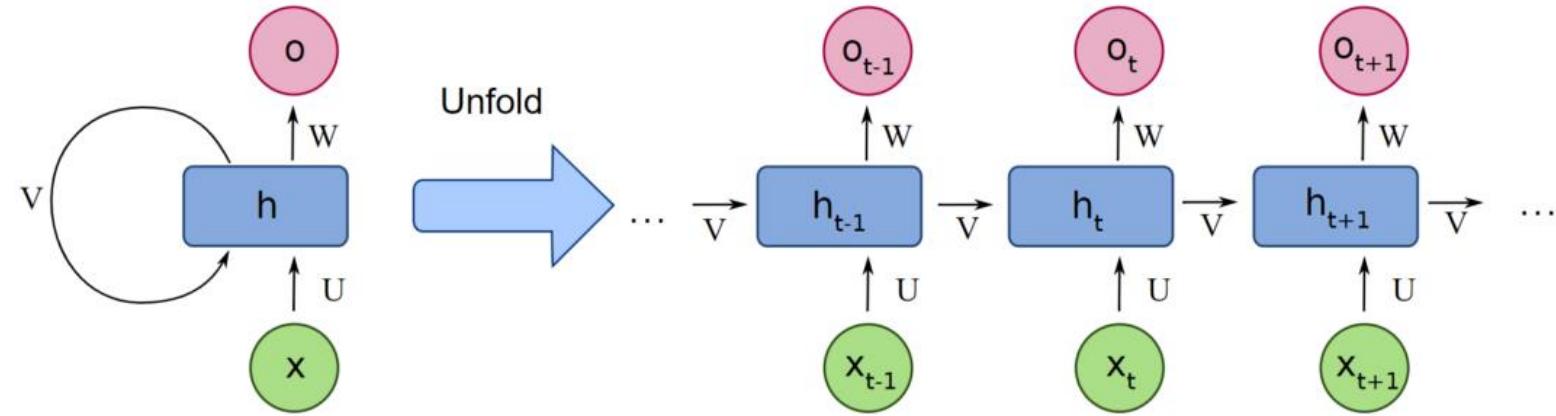
- 提高工作時脈速度
- 平行 / 向量指令集
- 多核(同質、大小核、異質) 加速
- 增加MAC數量

算力單位(注意數值格式)

- TOPS (每秒 10^{12} 次計算)
- GOPS (每秒 10^9 次計算)

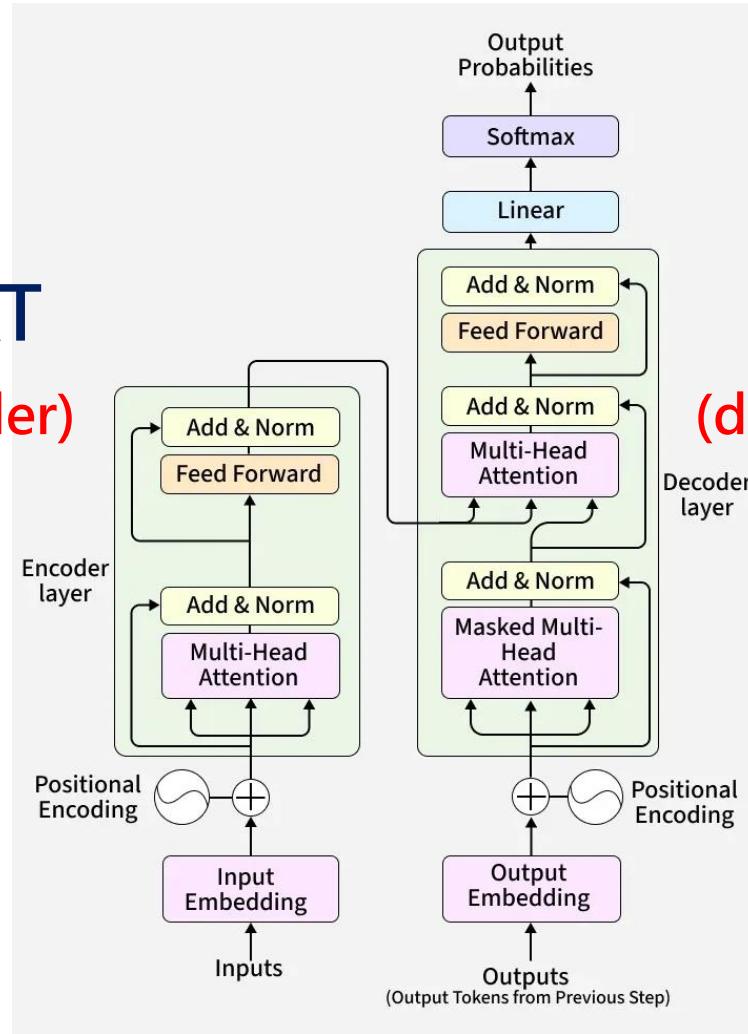
循環神經網路 (RNN)

Recurrent Neural Network (RNN)

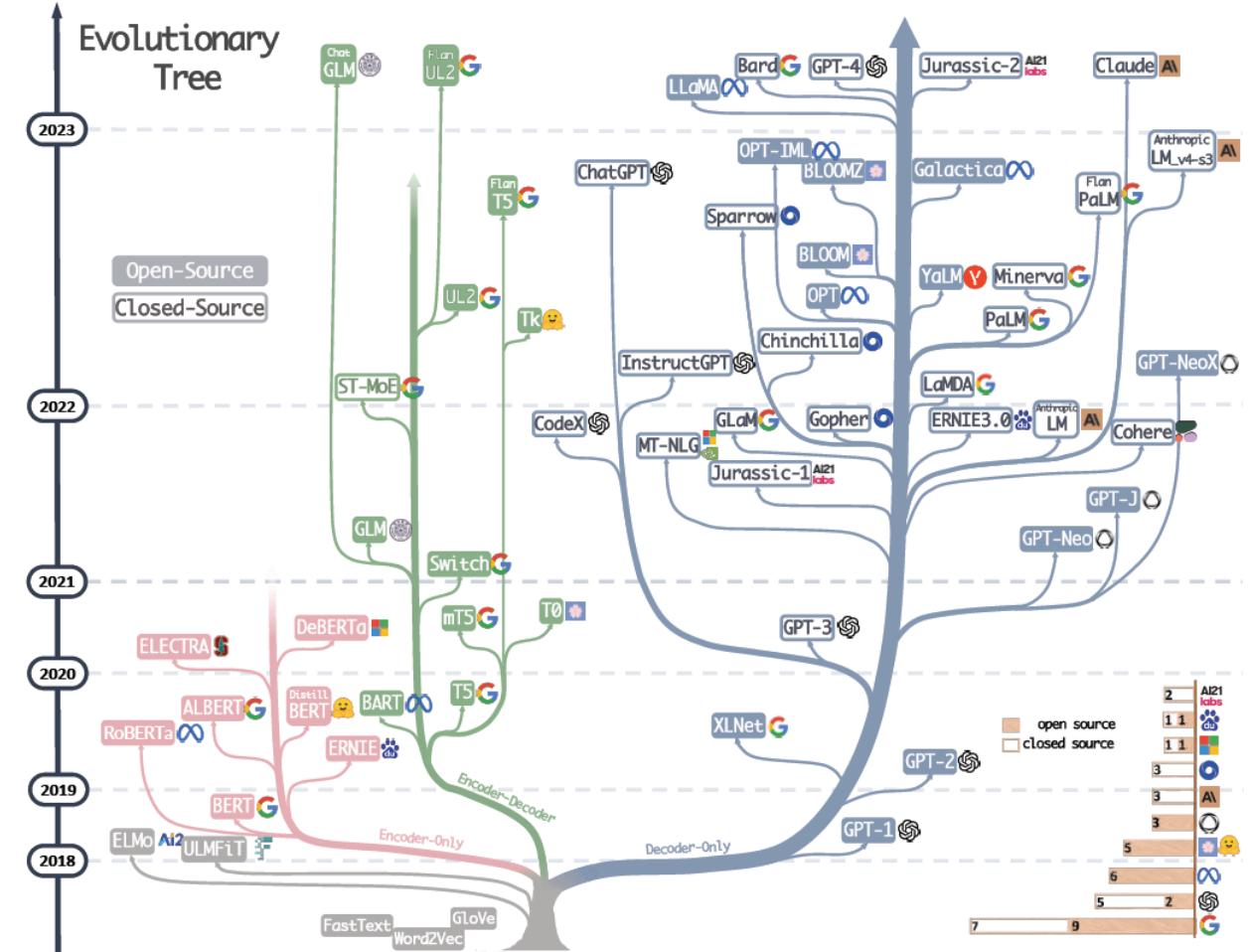


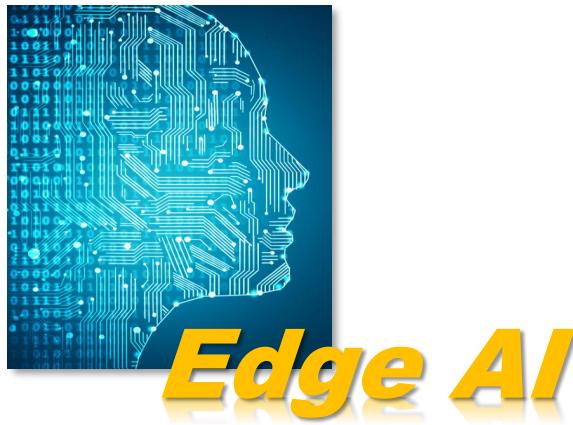
變換器 (Transformer)

BERT
(Encoder)

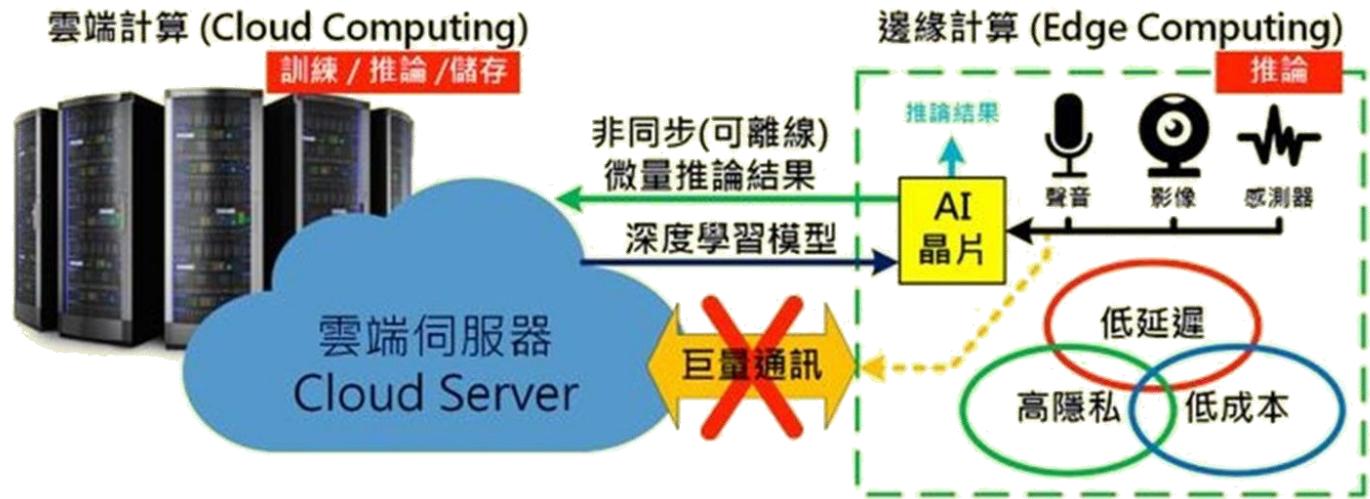
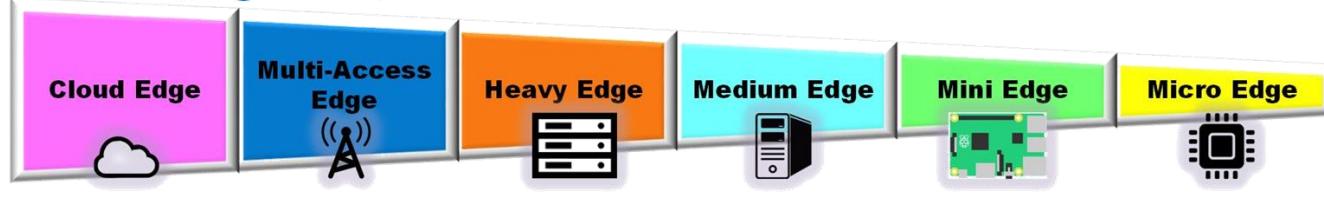


GPT
(decoder)





邊緣等級(Edge Level)



1.3. 邊緣智慧應用情境

不同等級邊緣硬體應用情境

單晶片 單板微電腦 手機/平板 桌機/伺服器



運算速度 /
平均功耗 /
推論精度 /
單位成本

感測分析
語音命令
運動偵測

影像分類
物件偵測
語音辨識

影像辨識
文字生成
影音生成

模型訓練
鑑別智慧
生成智慧

TinyML案例分享（技術分類）



TinyML30組案例 https://hackmd.io/@OmniXRI-Jack/tinyML_30_projects

Edge Impulse 案例研究

EDGE IMPULSE Product Solutions Developers Pricing Company Blog Login Get started

Embedded ML solutions, with Edge Impulse

Edge Impulse puts the power of ML into real products, across every industry.

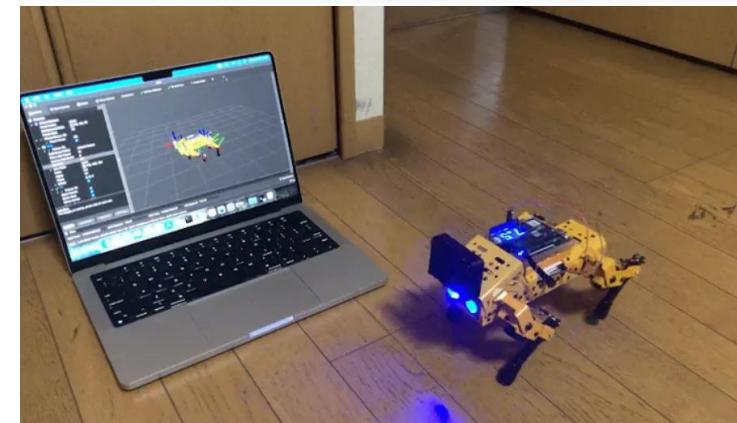
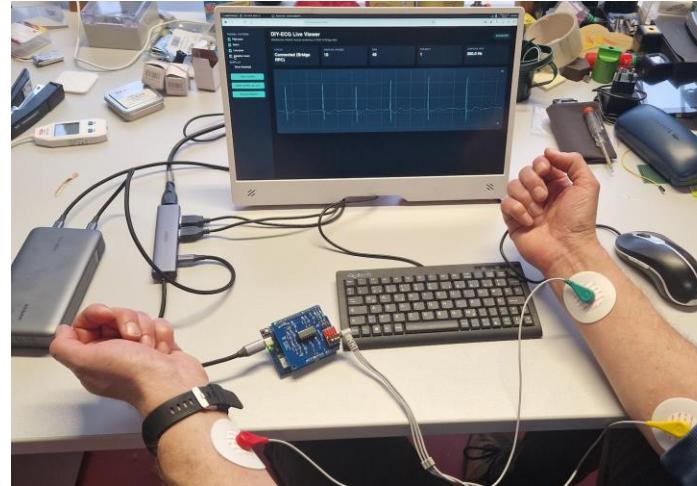
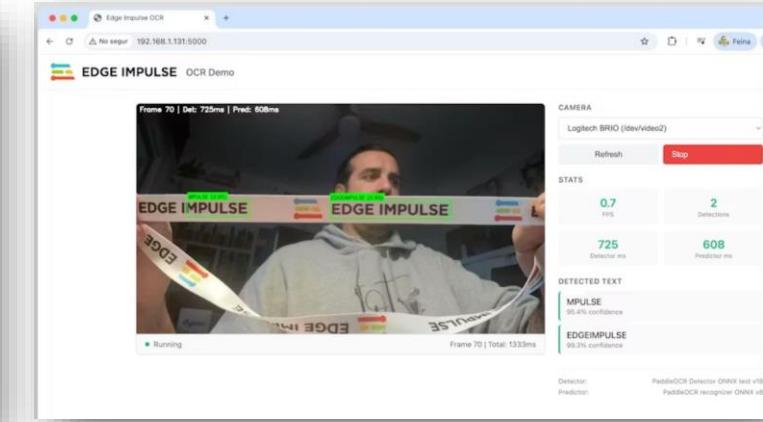
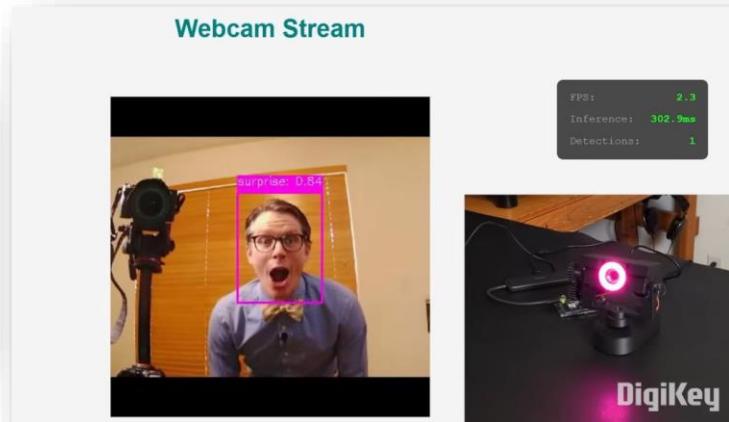
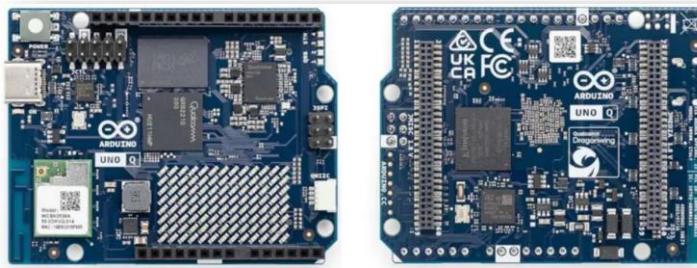


Health	Consumer devices	Automotive
Revolutionizing Fall Detection with Edge AI <small>The prevalence of falls among the elderly and those with limited mobility is a significant factor.</small>	HP: Bringing Voice Control to Earbuds & Headsets <small>When one of the best-known brands in business and personal productivity hardware set out to...</small>	GlobalSense Drives New Standards in Automotive Diagnostics with Edge AI

- 健康
- 車輛
- 資產追蹤
- 預測維護
- 電腦視覺
- 人機介面
- 農業
- 工業
- 建築
- 穿戴式

<https://www.edgeimpulse.com/case-studies>

Arduino UNO Q 教學案例整理



<https://omnixri.blogspot.com/2026/02/arduino-uno-q.html>

Hackster.io 人工智慧及機器學習專案

The screenshot shows the Hackster.io website interface. At the top, there is a navigation bar with the logo, search bar, and user authentication buttons (Log in, Sign up). Below the navigation bar, there are links for Projects, Channels, News, Contests, Events, and Videos, along with social media sharing icons. A prominent banner for the 'AI & Machine Learning' community is displayed, featuring a brain icon and a brief description: 'Algorithms and statistical models that make your favorite hardware think, act and interact with humans and machines alike.' A 'Join Community' button is also present. Below the banner, there are tabs for Home, Projects, Discussion, and Members, with 'Home' being the active tab. A section titled 'Computer vision' displays four project cards:

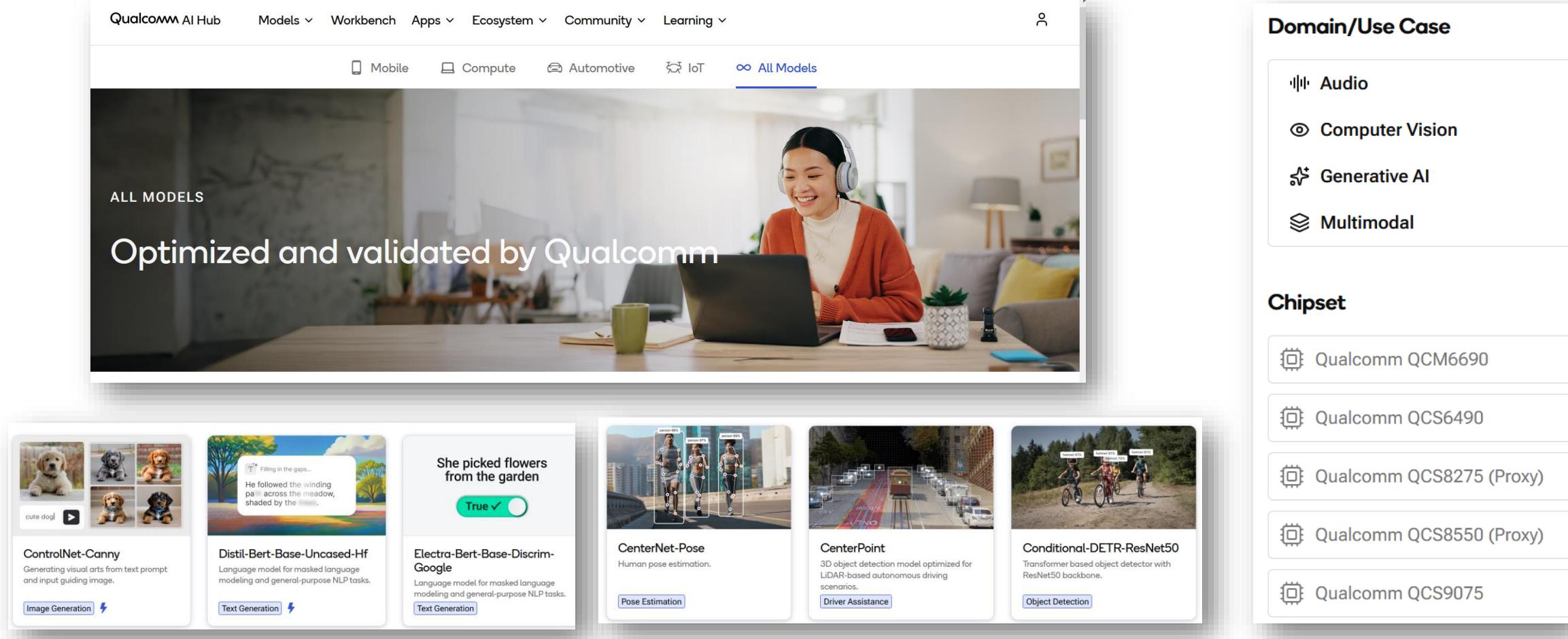
- Managing AI-Agent Context**: VISION-DRIVEN AI AGENT CONTEXT MANAGEMENT, showing a laptop screen with AI agent context management.
- AMD Ryzen AI NPU: Image Transpose using Multi-Dim DMA**: YantraVision AMD Partner, showing a diagram of the NPU architecture.
- Anti-Sleep Camera-Based Alarm System Using Arduino & OpenCV**: ANTI-SLEEP CAMERA ALARM, showing a camera connected to an Arduino board.
- AMD Ryzen AI NPU: Multi-Dimensional Shim DMA**: YantraVision AMD Partner, showing another diagram of the NPU architecture.

Each project card includes a 'View all' link at the top right.

建議以主晶片、
開發板名稱或
應用主題搜尋

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

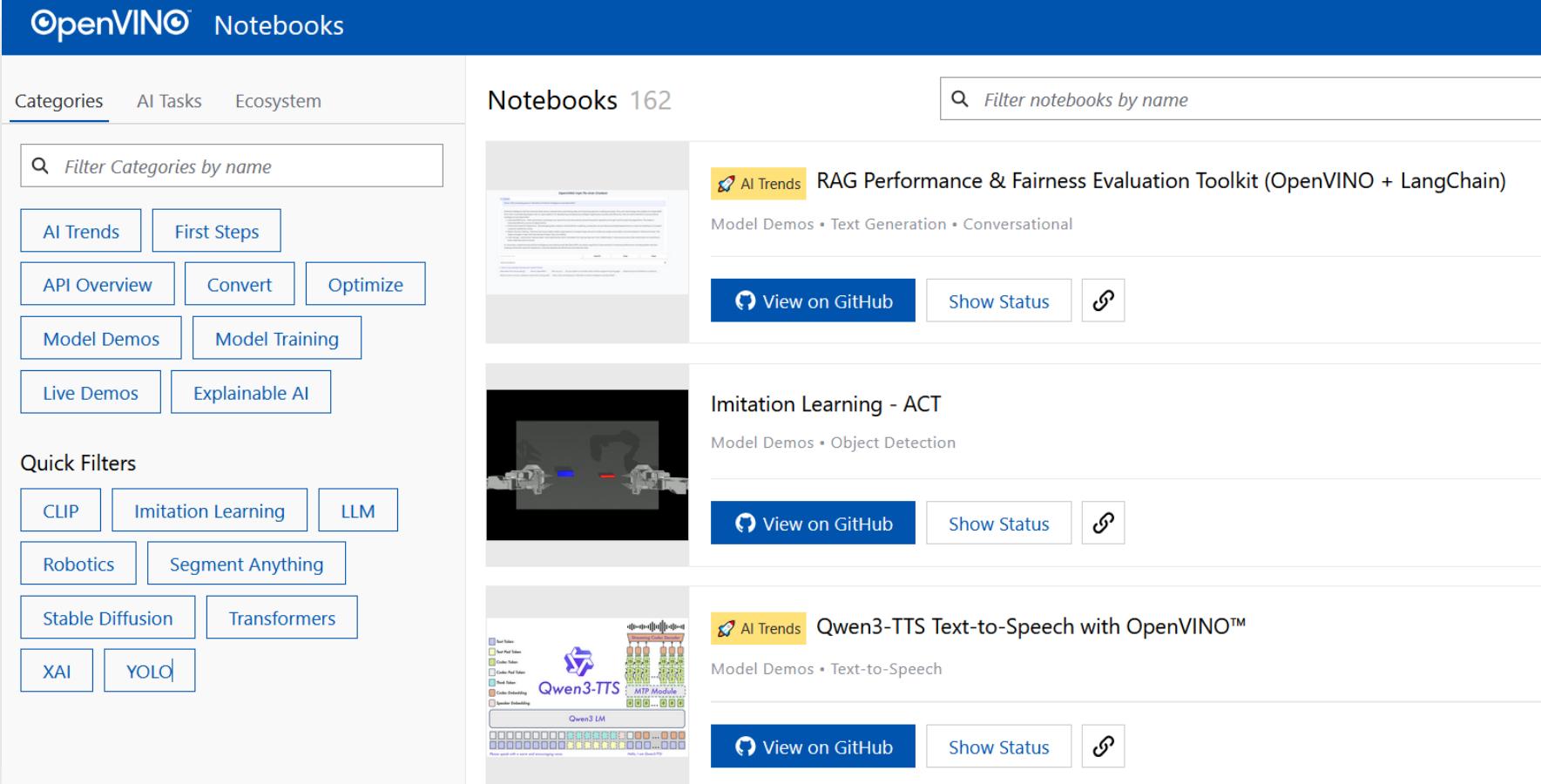
Qualcomm AI Hub



The screenshot shows the Qualcomm AI Hub website. At the top, there's a navigation bar with links: Qualcomm AI Hub, Models, Workbench, Apps, Ecosystem, Community, and Learning. Below the navigation, there are category tabs: Mobile, Compute, Automotive, IoT, and All Models (which is underlined). A large banner image features a woman wearing headphones and working on a laptop at a desk. The text "ALL MODELS" and "Optimized and validated by Qualcomm" is overlaid on the banner. Below the banner, there are several model cards. From left to right: 1. ControlNet-Canny: Generates visual arts from text prompt and input guiding image. 2. Distil-Bert-Base-Uncased-Hf: Language model for masked language modeling and general-purpose NLP tasks. 3. Electra-Bert-Base-Discrim-Google: Language model for masked language modeling and general-purpose NLP tasks. 4. CenterNet-Pose: Human pose estimation. 5. CenterPoint: 3D object detection model optimized for LiDAR-based autonomous driving scenarios. 6. Conditional-DETR-ResNet50: Transformer based object detector with ResNet50 backbone. On the right side of the page, there's a sidebar titled "Domain/Use Case" with icons for Audio, Computer Vision, Generative AI, and Multimodal. Another sidebar titled "Chipset" lists Qualcomm QCM6690, QCS6490, QCS8275 (Proxy), QCS8550 (Proxy), and QCS9075.

<https://aihub.qualcomm.com/models>

Intel OpenVINO Notebooks



The screenshot shows the Intel OpenVINO Notebooks interface. On the left, there's a sidebar with categories like AI Trends, First Steps, API Overview, Convert, Optimize, Model Demos, Model Training, Live Demos, and Explainable AI. Below that are quick filters for CLIP, Imitation Learning, LLM, Robotics, Segment Anything, Stable Diffusion, Transformers, XAI, and YOLO. The main area displays a list of 162 notebooks. The first item is "RAG Performance & Fairness Evaluation Toolkit (OpenVINO + LangChain)" under the "AI Trends" category. The second item is "Imitation Learning - ACT" under "Model Demos" and "Object Detection". The third item is "Qwen3-TTS Text-to-Speech with OpenVINO™" under "Model Demos" and "Text-to-Speech". Each notebook entry includes a thumbnail, a title, a category tag (e.g., AI Trends), a brief description, and buttons for "View on GitHub", "Show Status", and a link icon.

https://github.com/openvinotoolkit/openvino_notebooks

https://openvinotoolkit.github.io/openvino_notebooks/

Intel Edge AI Suite 使用者案例

[Retail AI Suite](#)[Manufacturing AI Suite](#)[Metro AI Suite](#)[Robotics AI Suite](#)

Accelerate Hardware Decisions for Retail AI Workloads at the Edge

The Retail AI Suite is an open-source software framework designed to accelerate AI workload evaluation and hardware selection for point-of-sale use cases at the edge. It helps retail solution builders assess device configurations across Intel® product generations to enhance decision-making and reduce the total cost of ownership. Key use cases include:

- **Checkout/Self-Checkout:** Product recognition (detection, classification, and tracking), full pipeline (product, weight, text, and barcode), age verification
- **Loss Prevention:** Fake scans, items in basket, multi-product identification, product switching, shopper behavior (obscuring/hiding an item), event video summation
- **Order Accuracy:** Order validation (product recognition), packing video summation



- 智慧零售
- 智慧製造
- 智慧交通
- 智慧機器人

<https://www.intel.com/content/www/us/en/software/edge-platform-industry.html#tab-blade-1-0>

Roboflow 部落格案例研究

roboflow Products Solutions Resources Pricing Docs Blog

Case Studies

LATEST POSTS CASE STUDIES PRODUCT UPDATES LOGISTICS MANUFACTURING

Build New Consumer Experiences



roboflow

5 Nov 2025 • 4 min read

How to Deliver Better Consumer Experiences with Vision AI

Build Factory Robots with Vision AI



roboflow

4 Nov 2025 • 4 min read

How to Build Factory Robots with Vision AI

Learn how to build adaptive factory robots with Vision AI

Build and Deploy with Roboflow for Free

Use Roboflow to manage datasets, train models in one-click, and deploy to web, mobile, or the edge.

Try It Now

<https://blog.roboflow.com/tag/case-studies/>

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 (CS5149701)

<https://omnixri.blogspot.com/p/ntust-edge-ai.html>

➤ 許哲豪，歐尼克斯實境互動工作室【部落格】

<https://omnixri.blogspot.com>

➤ 許哲豪，歐尼克斯實境互動工作室【YOUTUBE】

<https://www.youtube.com/@omnixri1784>

➤ 許哲豪，歐尼克斯實境互動工作室【GITHUB】

<https://github.com/OmniXRI>

➤ 許哲豪，Edge AI Taiwan 邊緣智能交流區

<https://www.facebook.com/groups/edgeaitw>