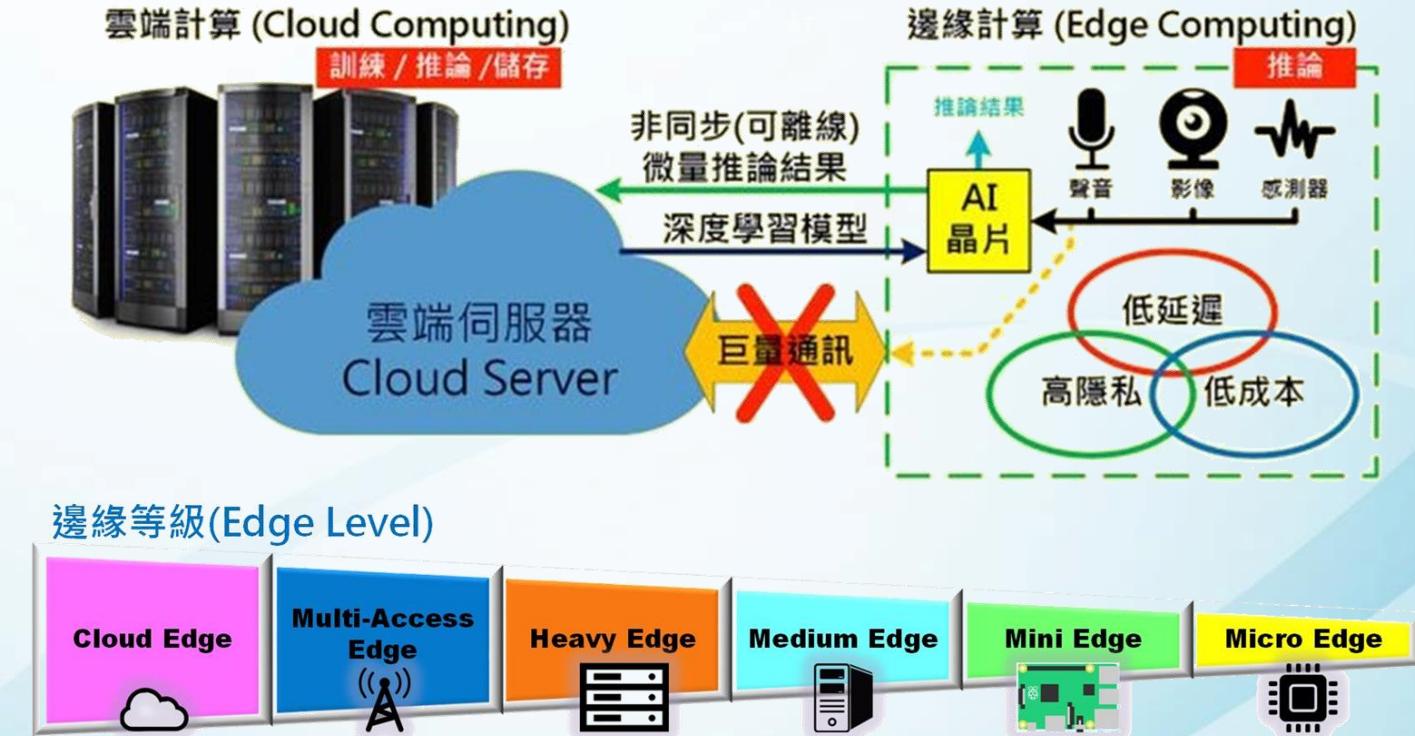


# OmniXRI's Edge AI & TinyML 小學堂



只有更邊



## 【第1講】邊緣人工智慧簡介



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

# 簡報大綱



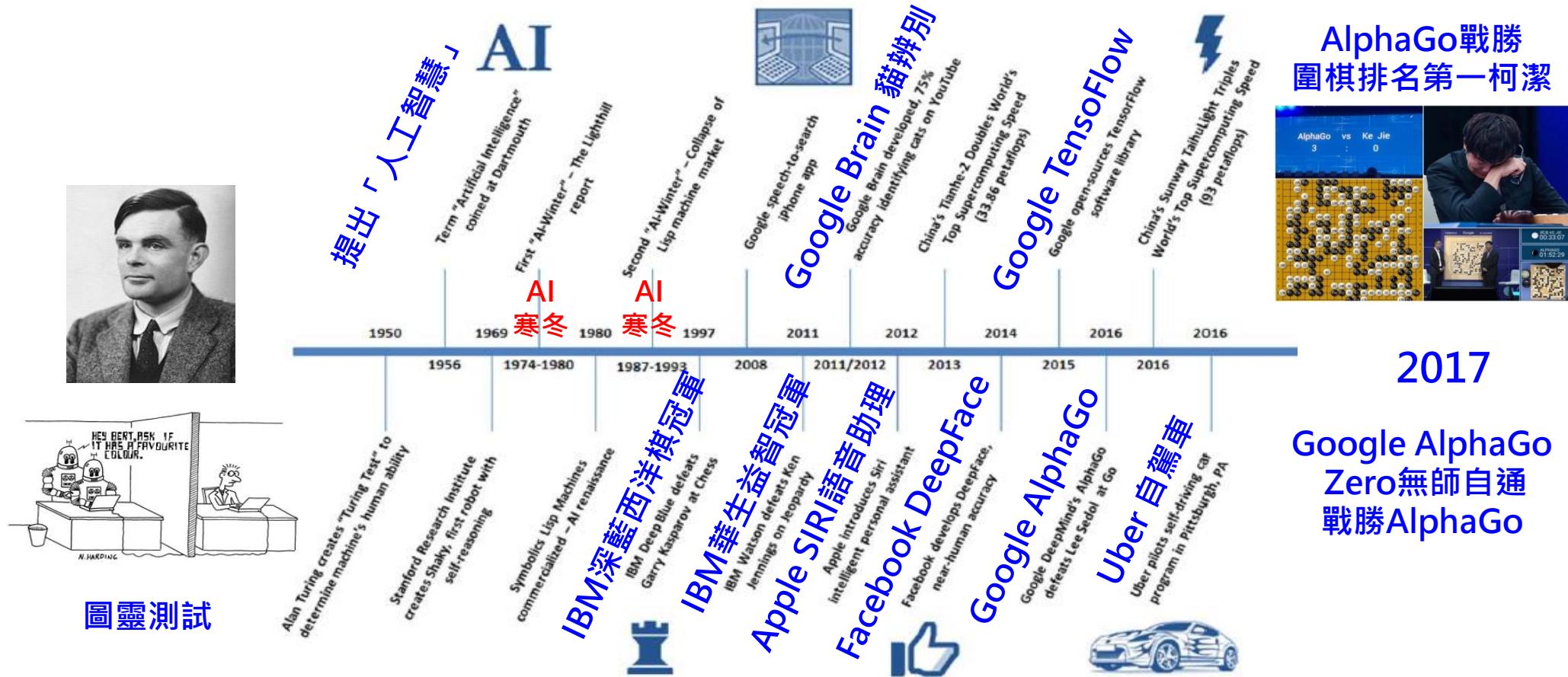
- 1.1. 人工智慧
- 1.2. 機器學習
- 1.3. 深度學習
- 1.4. 雲端與邊緣運算

## 1.1. 人工智慧



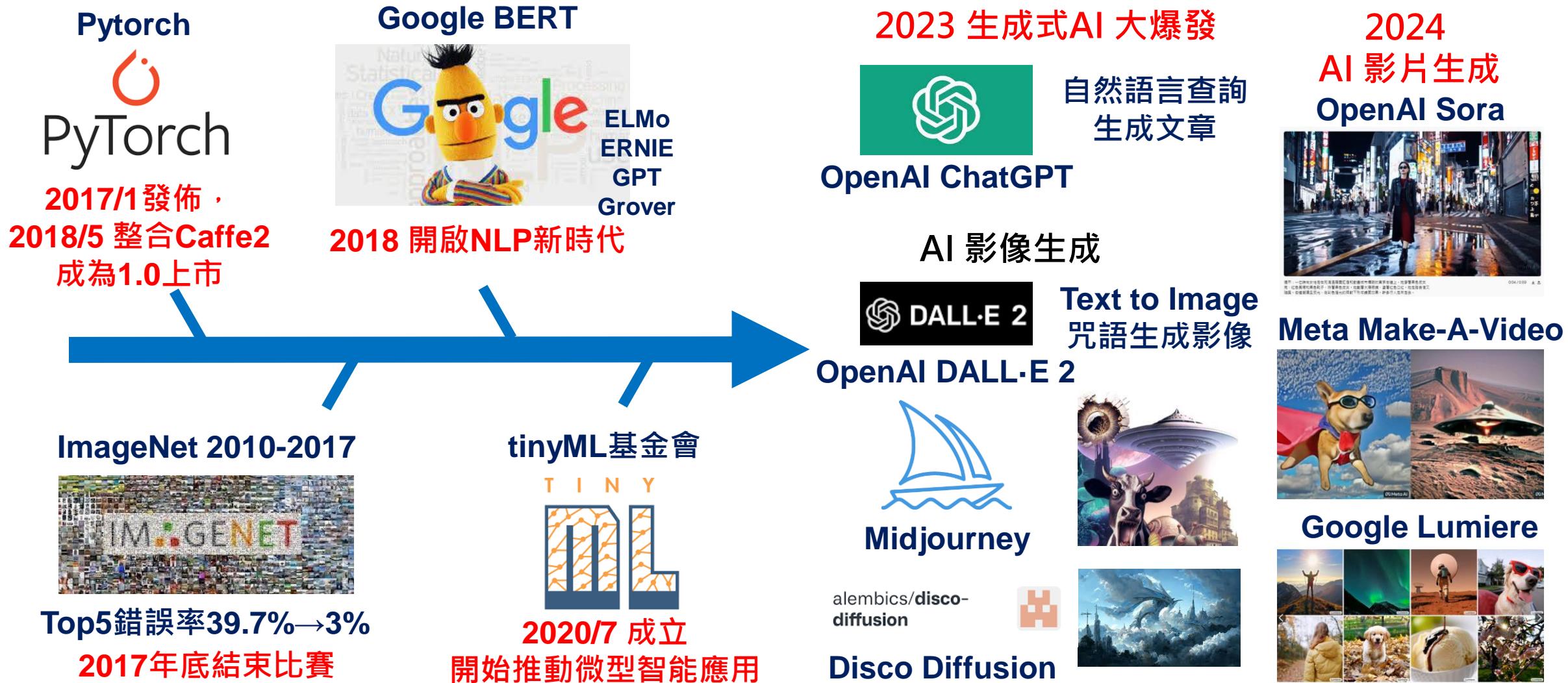
- 發展簡史
- 應用情境
- 產業發展
- 主要基石
- 未來發展

# 發展簡史—重要里程碑 (1/2)



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

# 發展簡史—重要里程碑 (2/2)



# 發展簡史—技術衍進



# 應用情境(1/2)

影像分類、物件偵測、  
影像分割、人臉表情、  
姿態偵測、超解析度、  
人臉特徵、風格轉換、  
**AIGC, MultiModal ...**

**2023 AIGC**  
文字（對話）、  
影音生成、  
多模態邊緣化

喚醒偵測、語音命令、  
異音偵測、鳥叫辨識、  
**NLP, NLU, LLM, AIGC...**

影片解讀  
與生成

電腦視覺

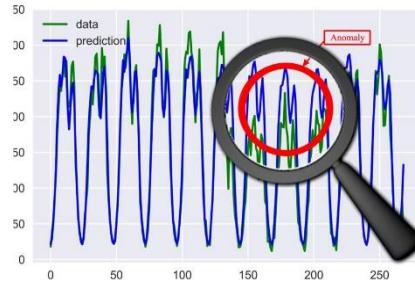
智慧感測

自然語言



運動感測、環境偵測、  
照度色溫、重量變化、  
距離偵測、電壓電流、  
數據預測、異常偵測 ...

# 應用情境(2/2)



異常偵測  
(Anomaly Detection)



推薦系統  
(Recommended System)



預測維修系統  
(Predictive Maintenance System)



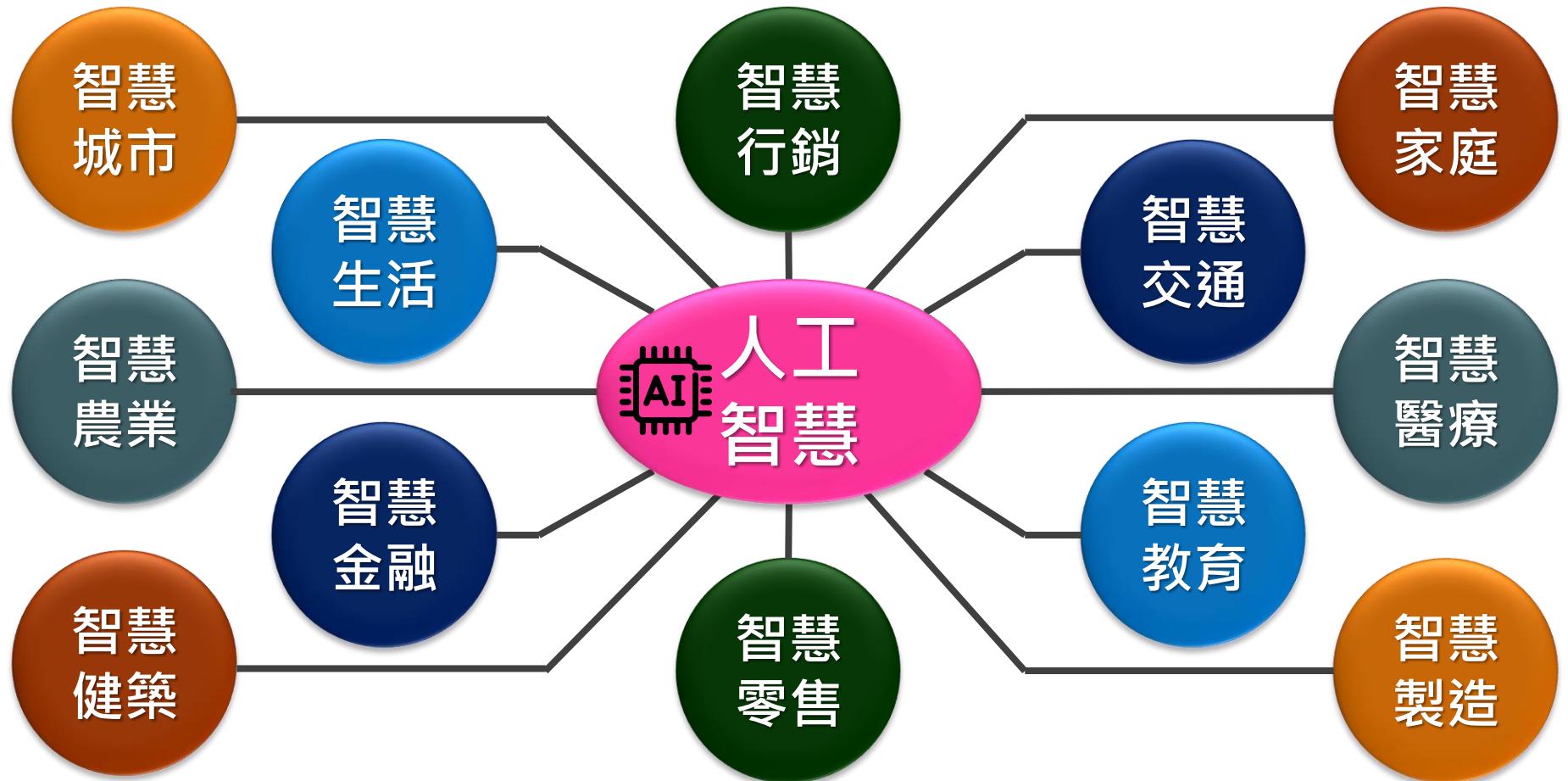
消費行為  
(Consumer Behavior)



資訊安全  
(Information Security)



# 產業發展



# 產業發展—CES 2024 邊緣智慧相關產品

## CES 2024 陪伴機器人相關產品



**ELLI-Q**



**LG Smart Home AI Agent**



**Rabbit R1**



**Samsung Ballie**

OmniXRI 整理製作, 2024/02/15

## CES 2024 智慧家電相關產品



**Samsung Bespoke Flex Refrigerator**



**Seer Grills Perfecta**



**Kohler PureWash E930**



**Flappie Catflap**

OmniXRI 整理製作, 2024/02/15

## CES 2024 智慧生活相關產品



**Philips Palm Smart Deadbolt**



**Evolve MVMT**



**Movono Health Evie Ring**



**Swarovski Optik AX Visio**

OmniXRI 整理製作, 2024/02/15

## CES 2024 AI Chip 相關產品



**Ambarella N1**



**Aondevices AON1120**



**Bosch Sensortec**



**Ceva NeuPro-M NPU + Visionary.ai**



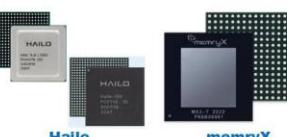
**Brainchip AKD1000**



**DeepX DX-M1/DX-H1/DX-V1/DX-V2**



**EY3D(Etron) eCV4**



**Halio Hallo-8L/Hallo-15**



**memryX MX3**

OmniXRI 整理製作, 2024/02/15

資料來源：<https://omnixri.blogspot.com/2024/02/vmaker-edge-ai-14-ces-2024-edge-aitinyml.html>

# 產業發展一相關簡報

台北科技大學車輛所/自動所/能源所111學年度第1學期聯合學術研討會




從智慧製造邁向工業元宇宙

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

<https://omnixri.blogspot.com/2022/11/20221130.html>

資訊與決策科學研究所 專題演講




深度學習應用於智慧零售影像分析之趨勢

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

<https://omnixri.blogspot.com/2022/11/20221128.html>

關南大學 Kainan University



健康資料處理與分析

08 健康資料收集與處理—穿戴式及健康照護感測器

協同計畫主持人 許哲豪 博士



<https://omnixri.blogspot.com/2022/11/20221031week08.html>

中華大學資訊工程學系 專題演講



AIoT與tinyML生態系國際發展趨勢與國產IC未來方向

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

<https://omnixri.blogspot.com/2022/12/20221219aiottinymlc.html>

# 主要基石



# 未來發展

## AI產業化

- 建構完整產業鏈及人才培訓

## 產業AI化

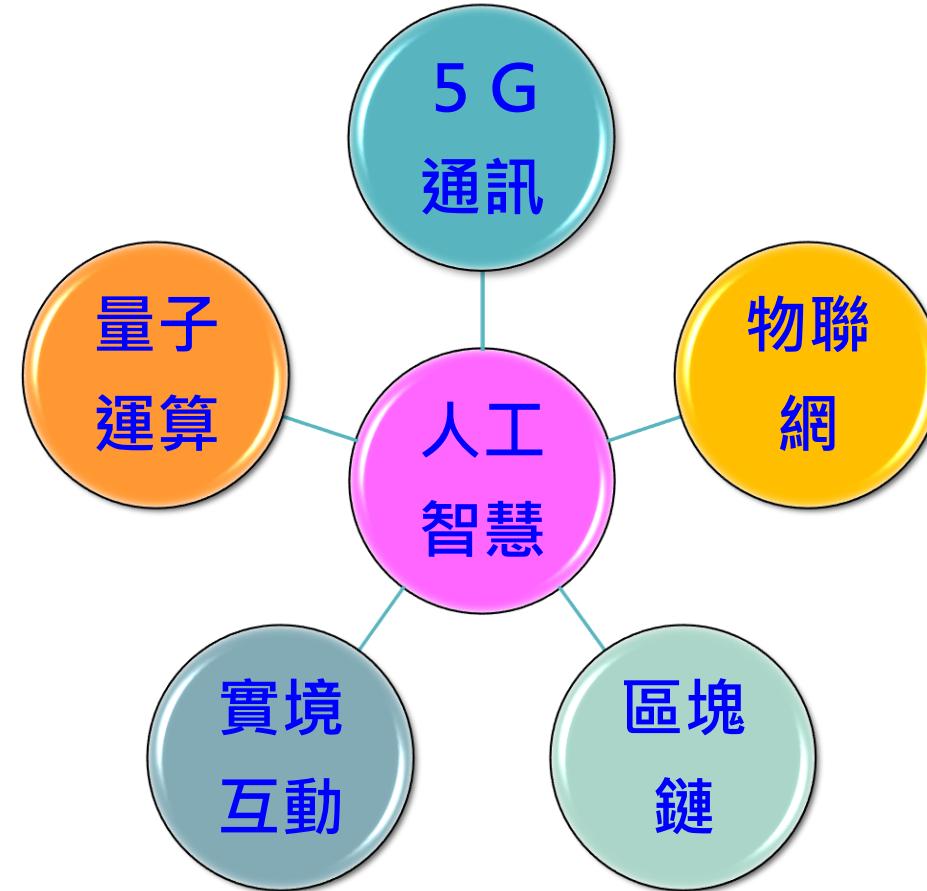
- 產業導入AI提升營運績效

## 算力普及化

- 雲端、邊緣算力隨手可得

## 資料集公開化

- 站在巨人肩膀看世界
- 我為人人、人人為我

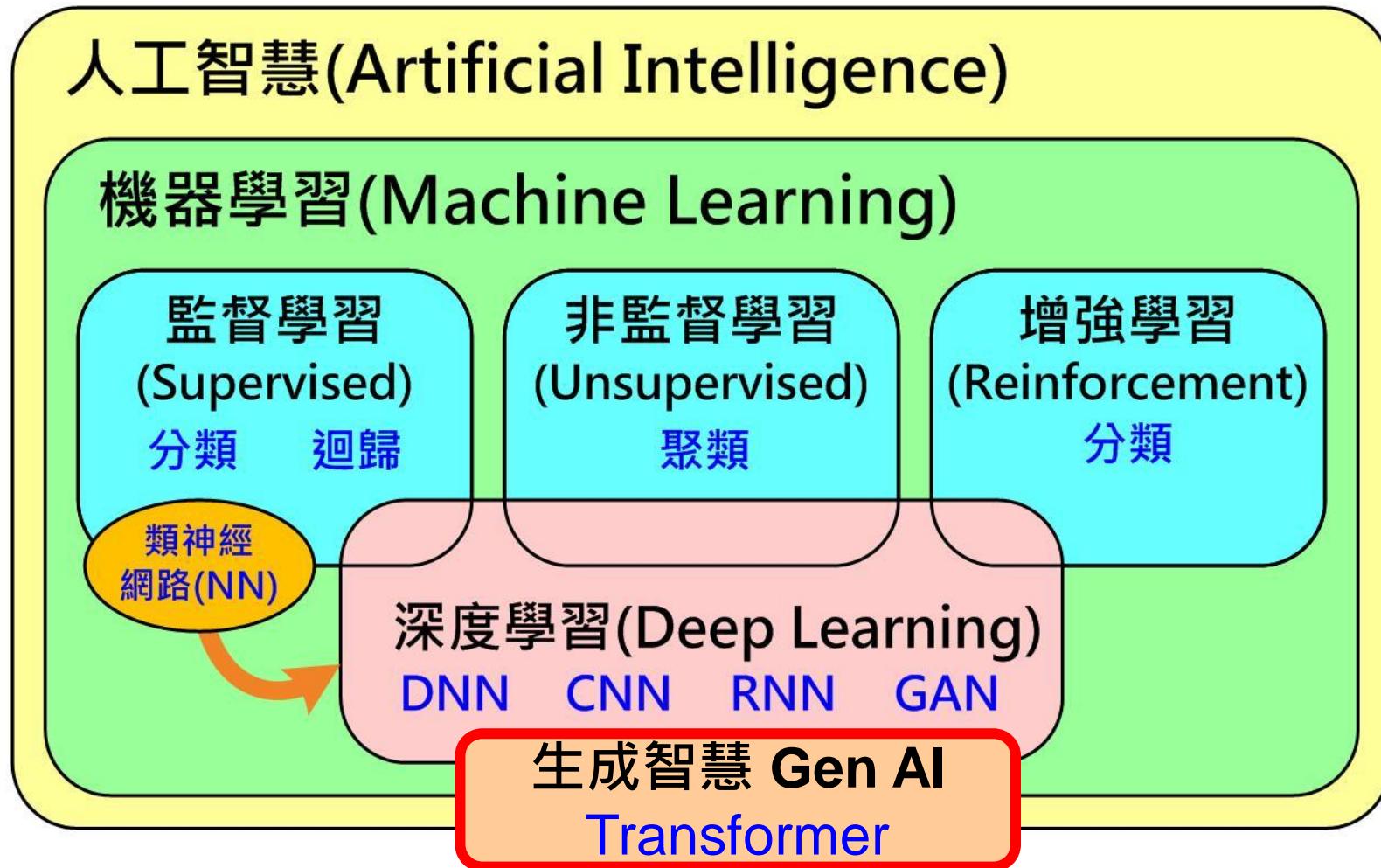


## 1.2. 機器學習



- 主要分類
- 監督式學習
- 非監督式學習
- 增強式學習

# 機器學習—主要分類

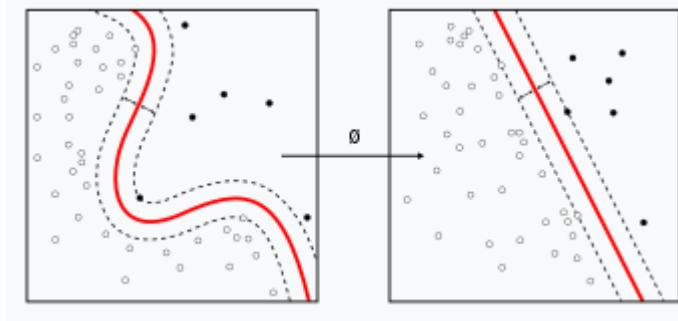


資料來源：<http://omnixri.blogspot.com/2018/05/aimaker.html>

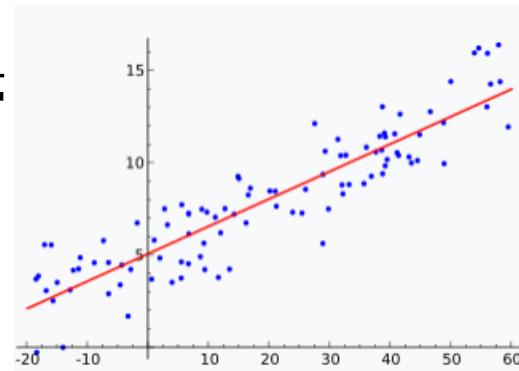
# 監督式學習

## 主要目的

- 分類(Classification)



- 回歸

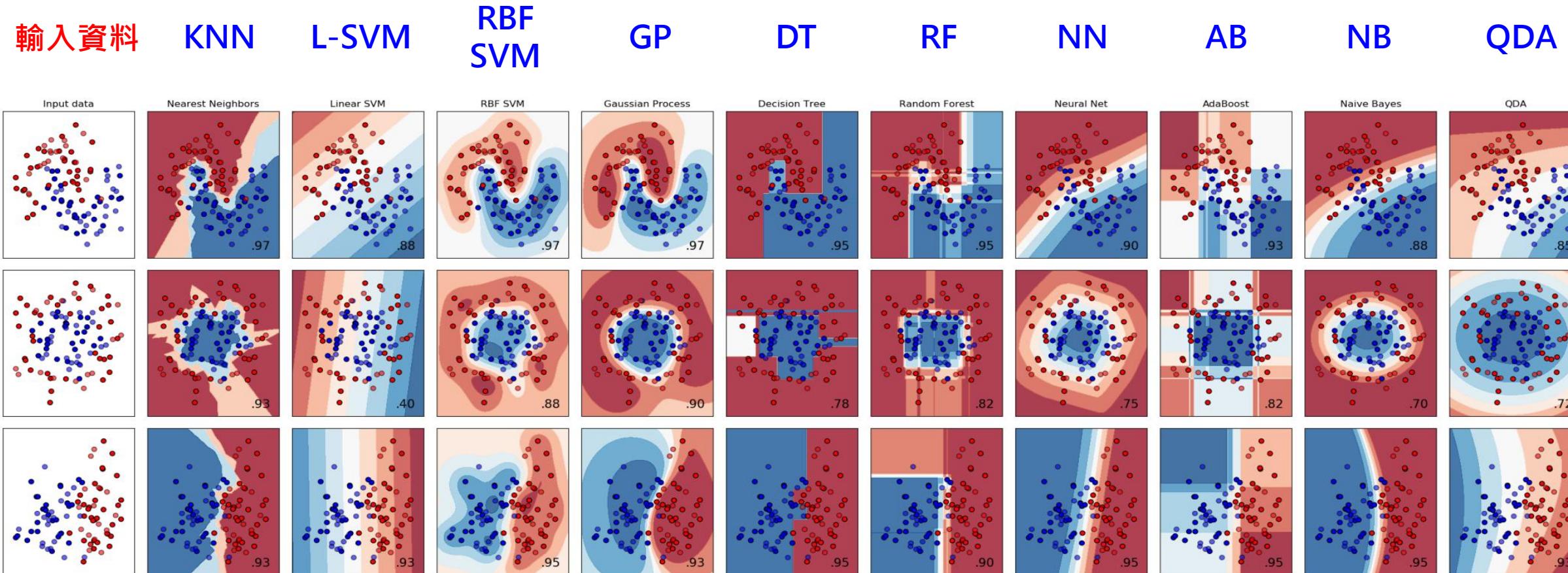


## 常見算法

- 決策樹整合 ( 裝袋 , 提升 , 隨機森林 )
- k-NN
- 線性回歸
- 樸素貝葉斯神經網路
- 邏輯回歸
  - 感知器
- 支持向量機 ( SVM )
- 相關向量機 ( RVM )

資料來源：<https://zh.wikipedia.org/wiki/%E7%9B%91%E7%9D%A3%E5%AD%A6%E4%B9%A0>

# 監督式學習一分類比較

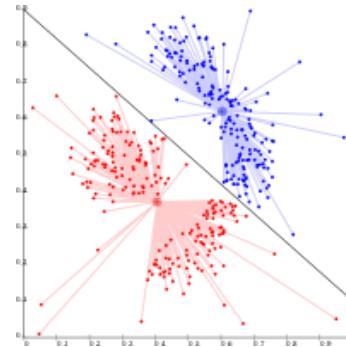
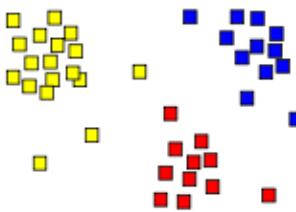


資料來源：[http://scikit-learn.org/stable/auto\\_examples/classification/plot\\_classifier\\_comparison.html](http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html)

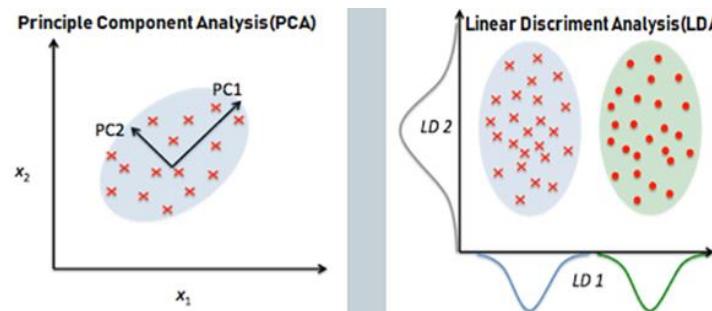
# 非監督式學習

## 主要目的

- 聚類 (Cluster)



- 降維(Dimensionality reduction)

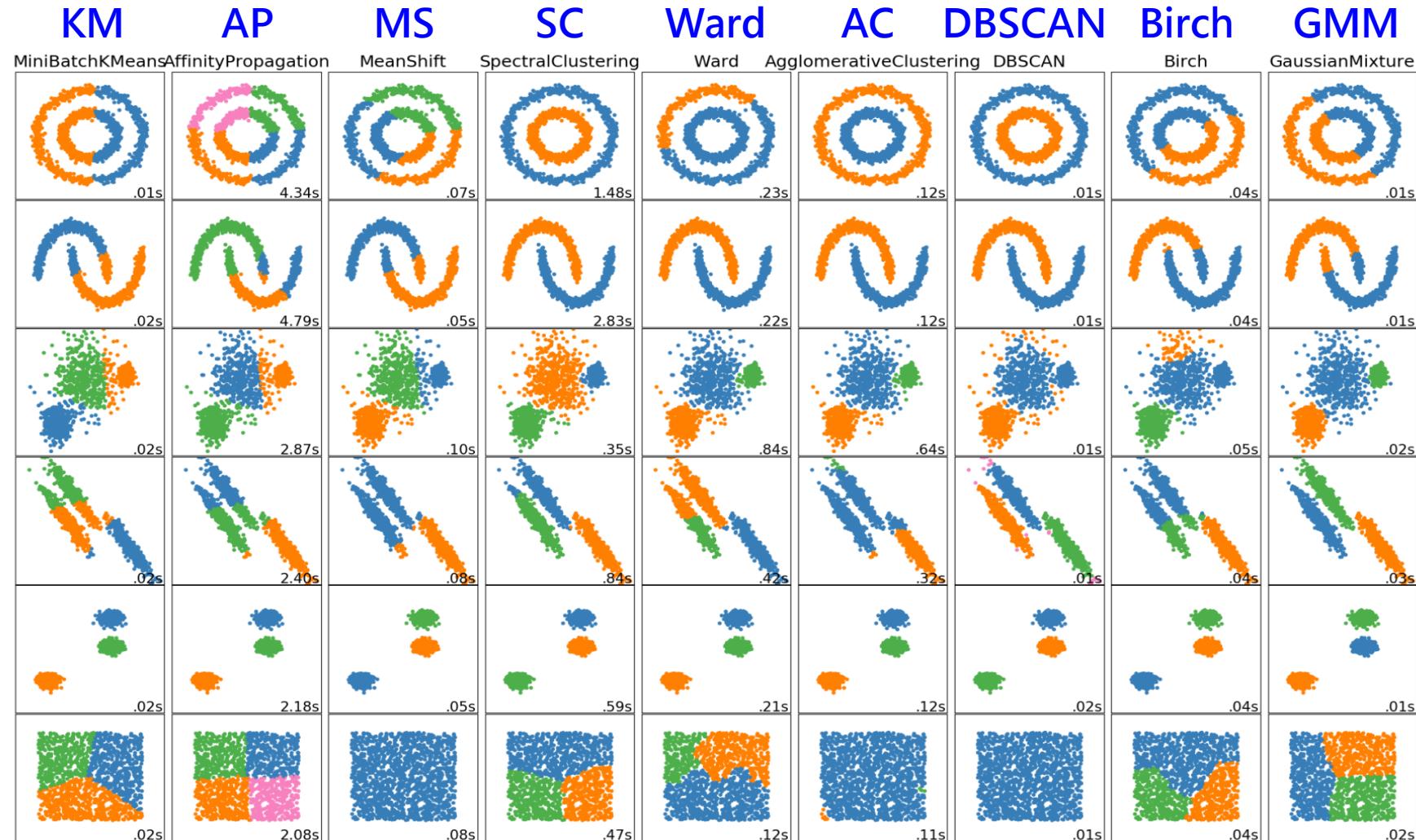


## 常見算法

- |             |         |
|-------------|---------|
| ➤ 聚類        | ➤ 降維    |
| ➤ BIRCH層次   | ➤ 因素分析  |
| ➤ k-平均      | ➤ CCA   |
| ➤ 期望最大化(EM) | ➤ ICA   |
| ➤ DBSCAN    | ➤ LDA   |
| ➤ OPTICS    | ➤ NMF   |
| ➤ 均值飄移      | ➤ PCA   |
|             | ➤ LASSO |
|             | ➤ t-SNE |

資料來源：<https://zh.wikipedia.org/wiki/%E7%84%A1%E7%9B%A3%E7%9D%A3%E5%AD%B8%E7%BF%92>

# 非監督式學習—聚類比較

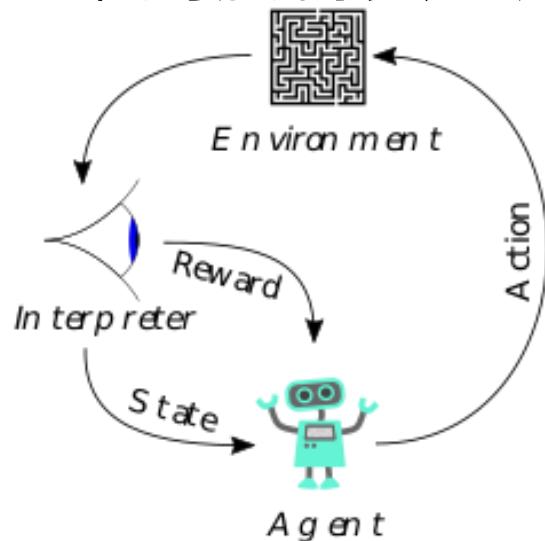


資料來源：[http://scikit-learn.org/stable/auto\\_examples/cluster/plot\\_cluster\\_comparison.html](http://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html)

# 增強式學習

## 主要目的

➤ 利用獎懲機制提高學習能力，無需輸入資料集，只需定義決策條件，經常用於博奕、遊戲。



## 常見算法

- 蒙特卡洛學習 Monte-Carlo Learning
- Temporal-Difference Learning
- SARSA
- Q-Learning

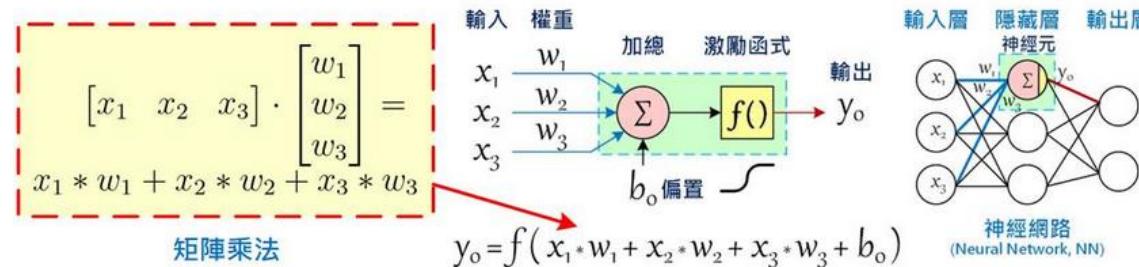
資料來源：<https://zh.wikipedia.org/wiki/%E5%BC%BA%E5%8C%96%E5%AD%A6%E4%B9%A0>

## 1.3. 深度學習

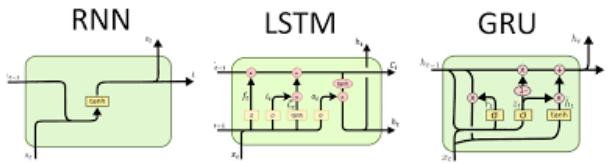


- 神經網路技術發展
- 神經網路簡介
  - 神經網路(NN)
  - 深度神經網路(DNN)
  - 卷積神經網路(CNN)
  - 循環神經網路(RNN)
  - 生成對抗網路(GAN)
  - 自動編碼網路(AE)
  - 圖神經網路(GNN)
  - 脈衝神經網路(SNN)
  - BERT / GPT / Transformer

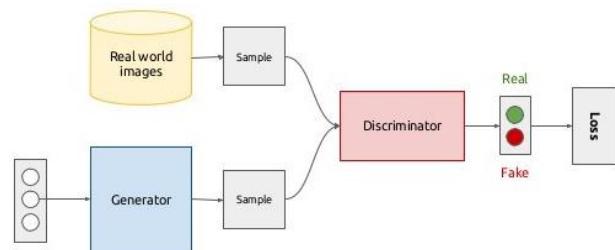
# 神經網路技術發展



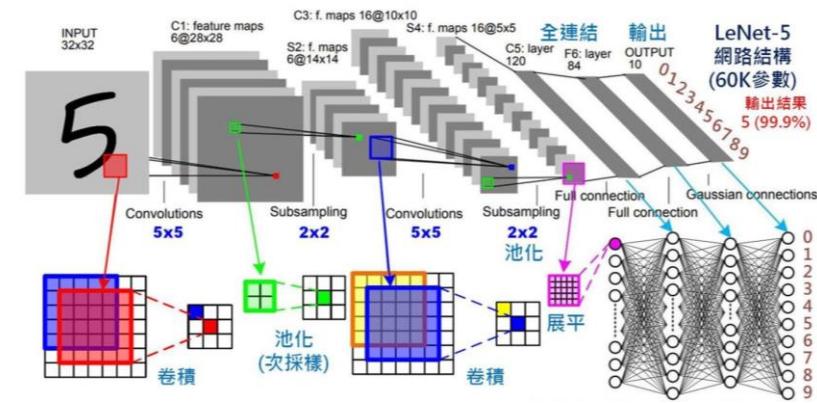
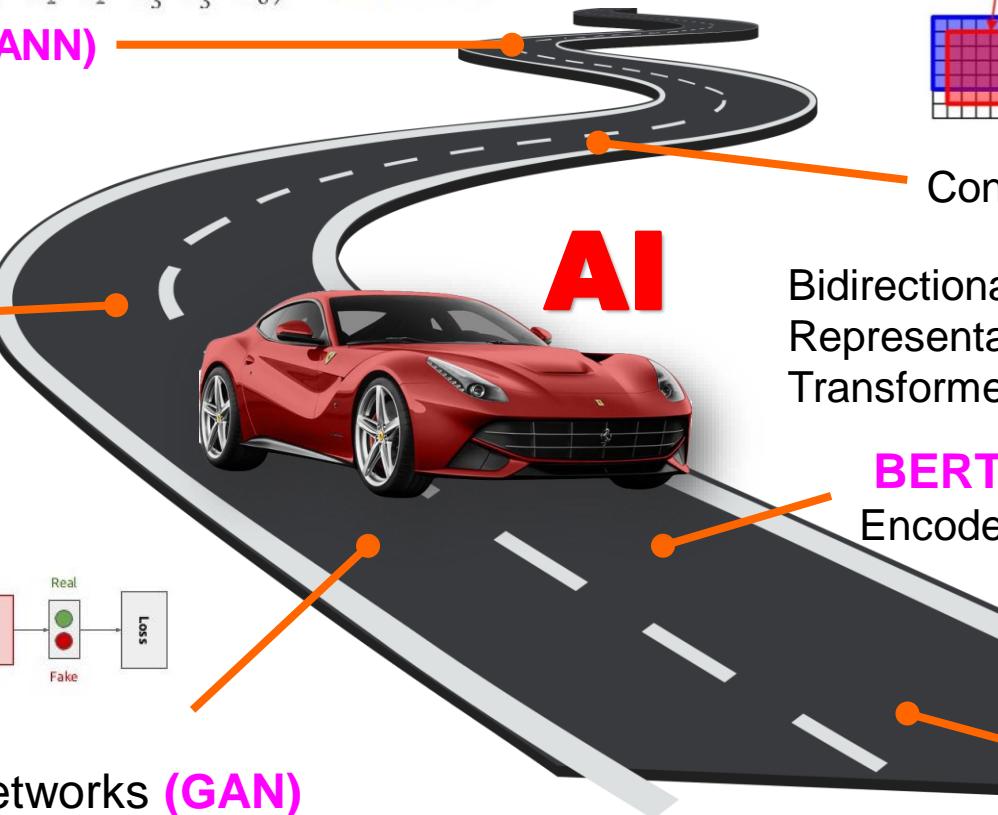
Artificial Neural Network (ANN)



Recurrent Neural Network (RNN)  
Long-Short Term Memory (LSTM)  
Gated Recurrent Unit (GRU)



Generative Adversarial Networks (GAN)



Convolutional Neural Network (CNN)

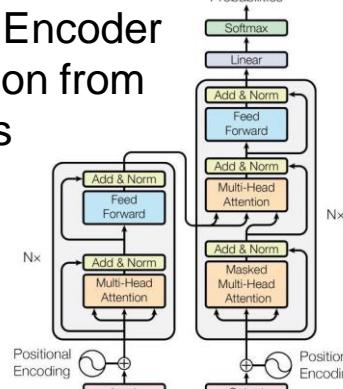
Bidirectional Encoder Representation from Transformers

BERT Encoder

Transformer

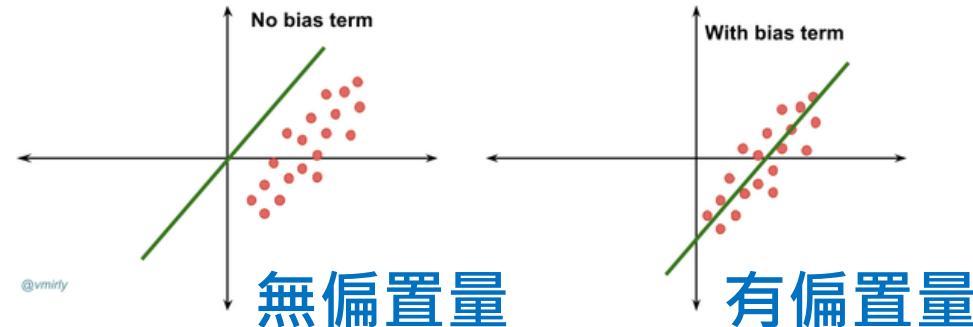
Generative Pre-trained Transformer

GPT Decoder

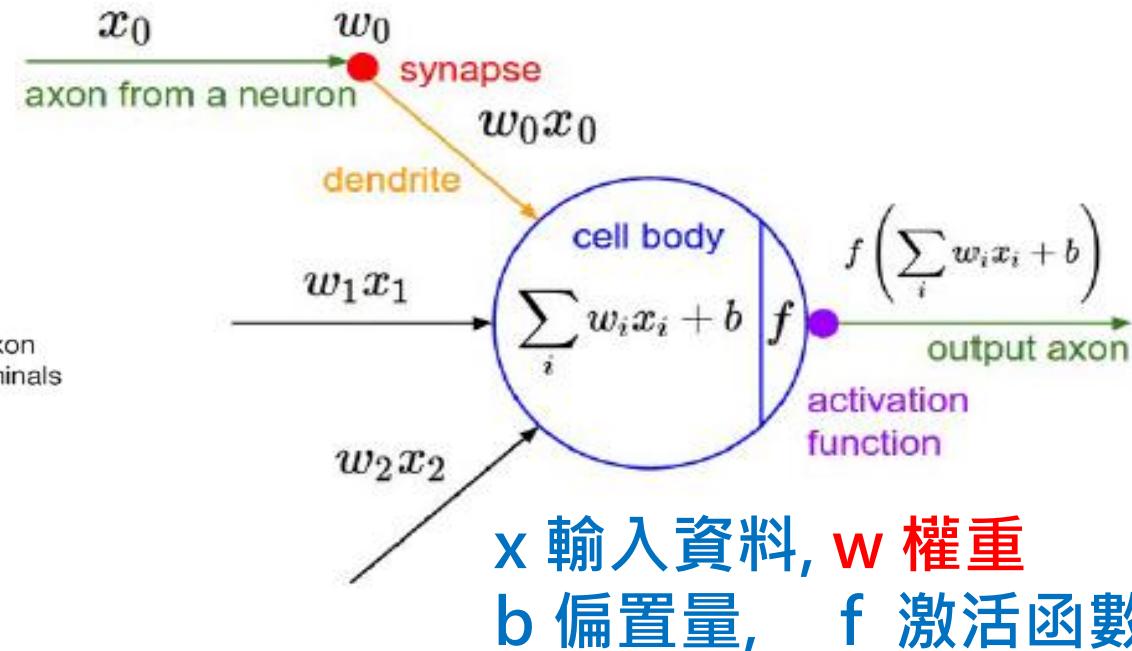
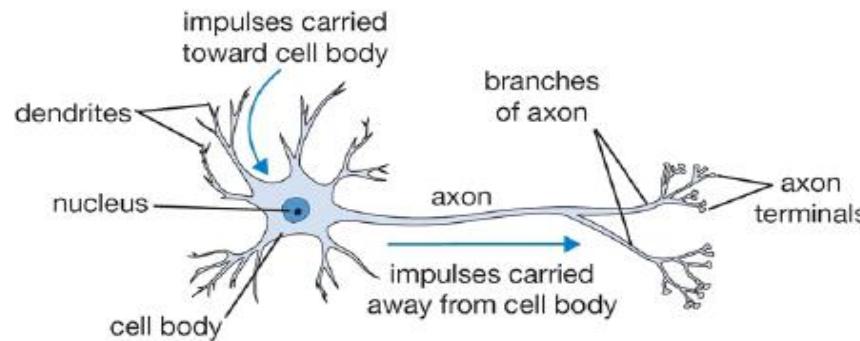


ChatGPT

# 神經網路概念—神經元



## 生物神經元



資料來源：[http://blog.csdn.net/cyh\\_24/article/details/50593400](http://blog.csdn.net/cyh_24/article/details/50593400)

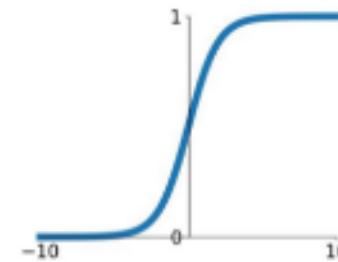
# 神經網路概念—激活函數

## Activation Functions

(一定要為非線性函式)

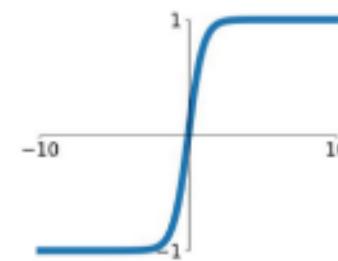
### Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



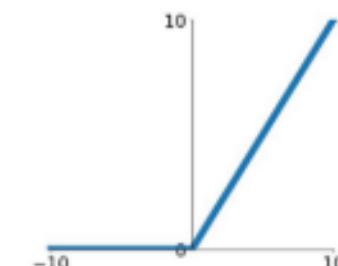
### tanh

$$\tanh(x)$$



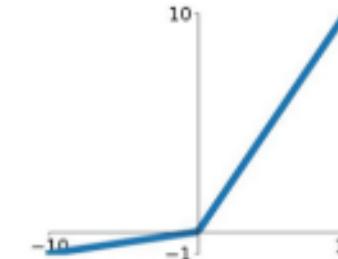
### ReLU

$$\max(0, x)$$



### Leaky ReLU

$$\max(0.1x, x)$$

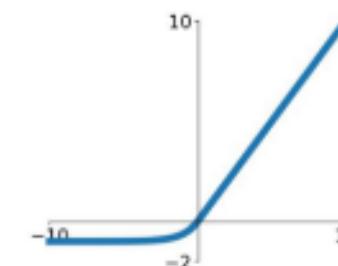


### Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

### ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

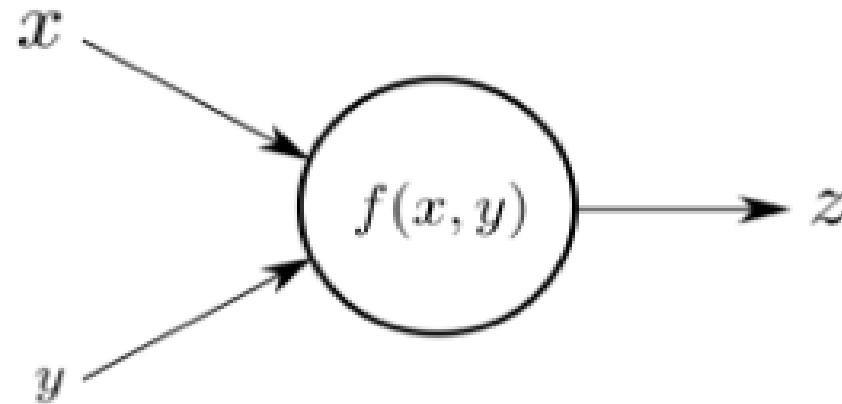


資料來源：[http://blog.csdn.net/cyh\\_24/article/details/50593400](http://blog.csdn.net/cyh_24/article/details/50593400)

# 神經網路概念—傳播

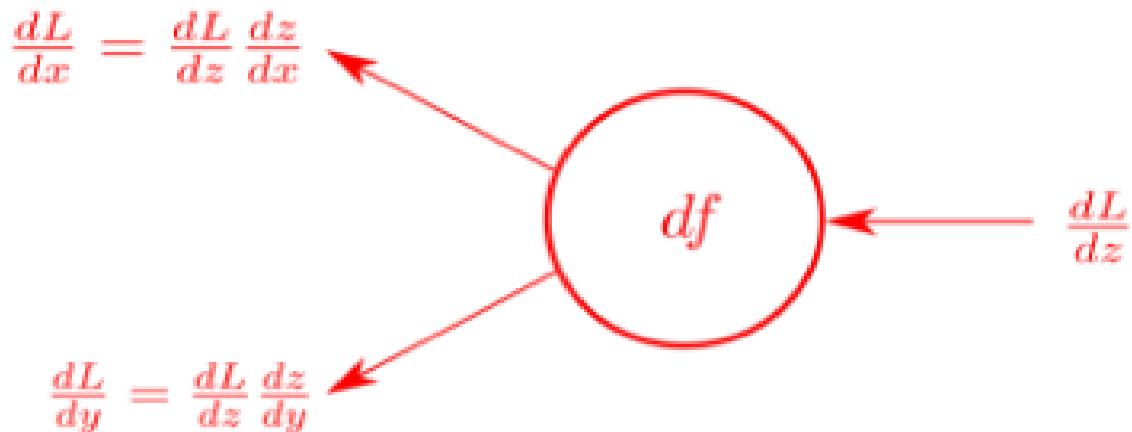
正向傳播(推論)

Forwardpass



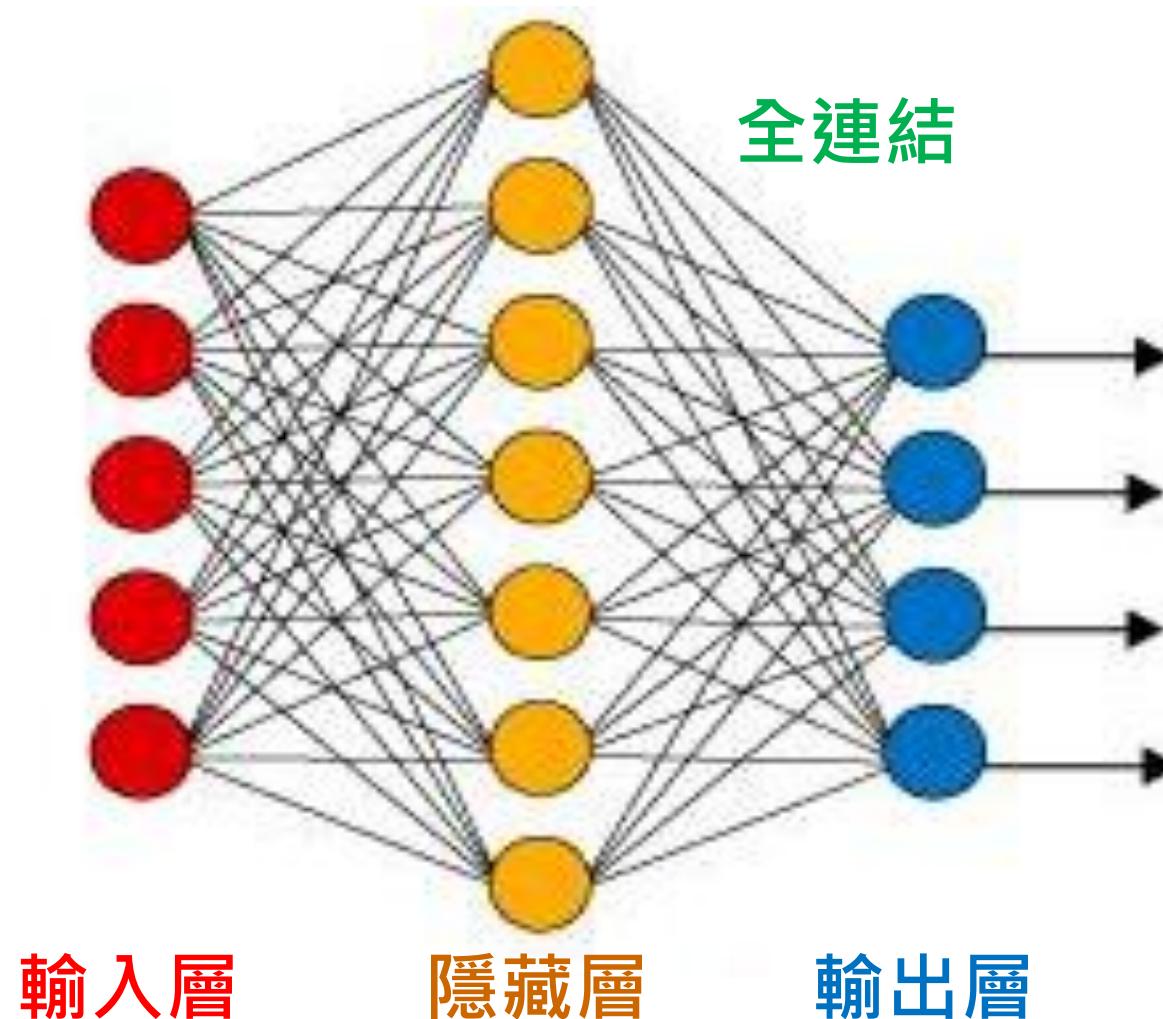
反向傳播(訓練)

Backwardpass

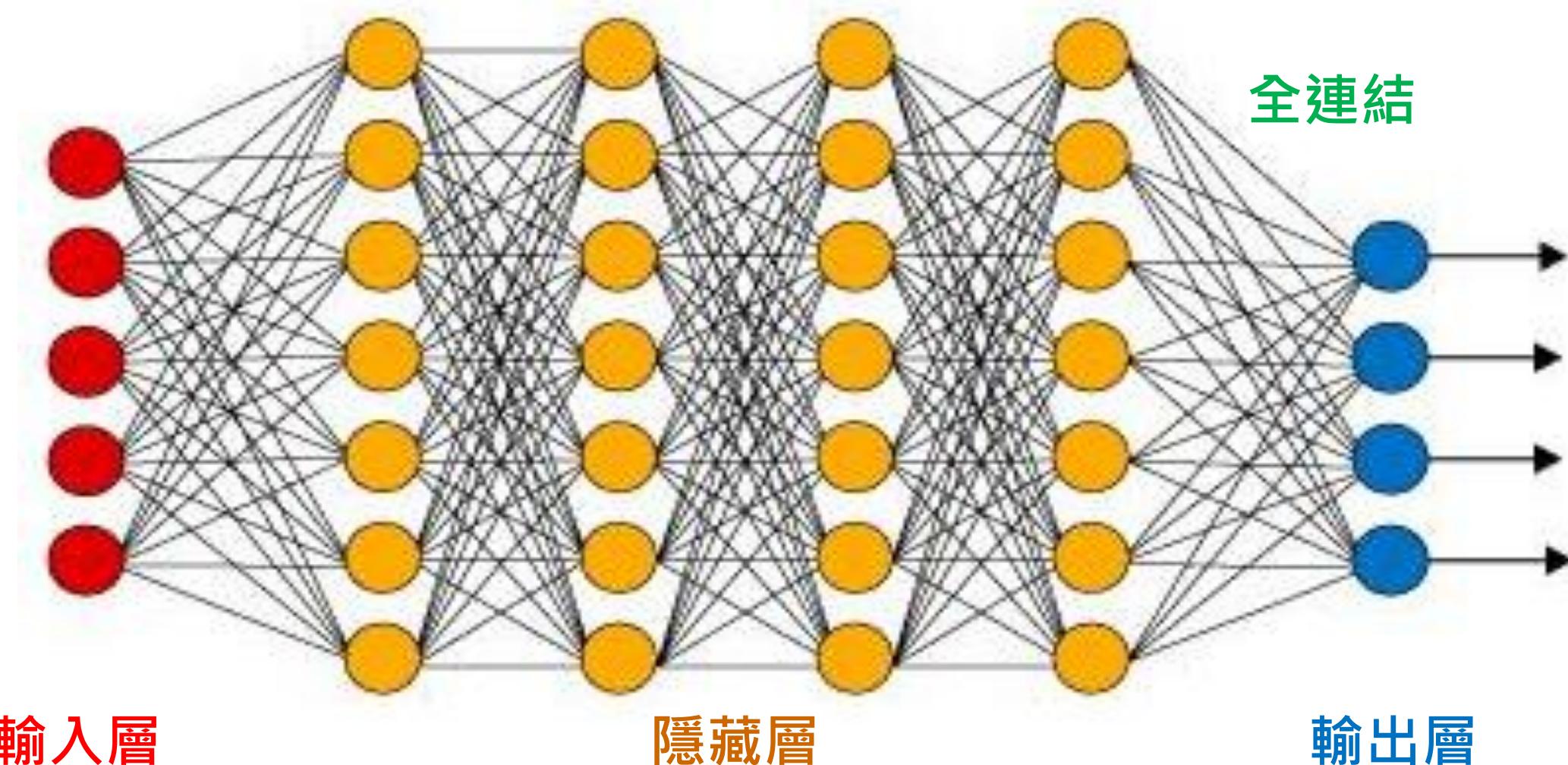


資料來源：[http://blog.csdn.net/cyh\\_24/article/details/50593400](http://blog.csdn.net/cyh_24/article/details/50593400)

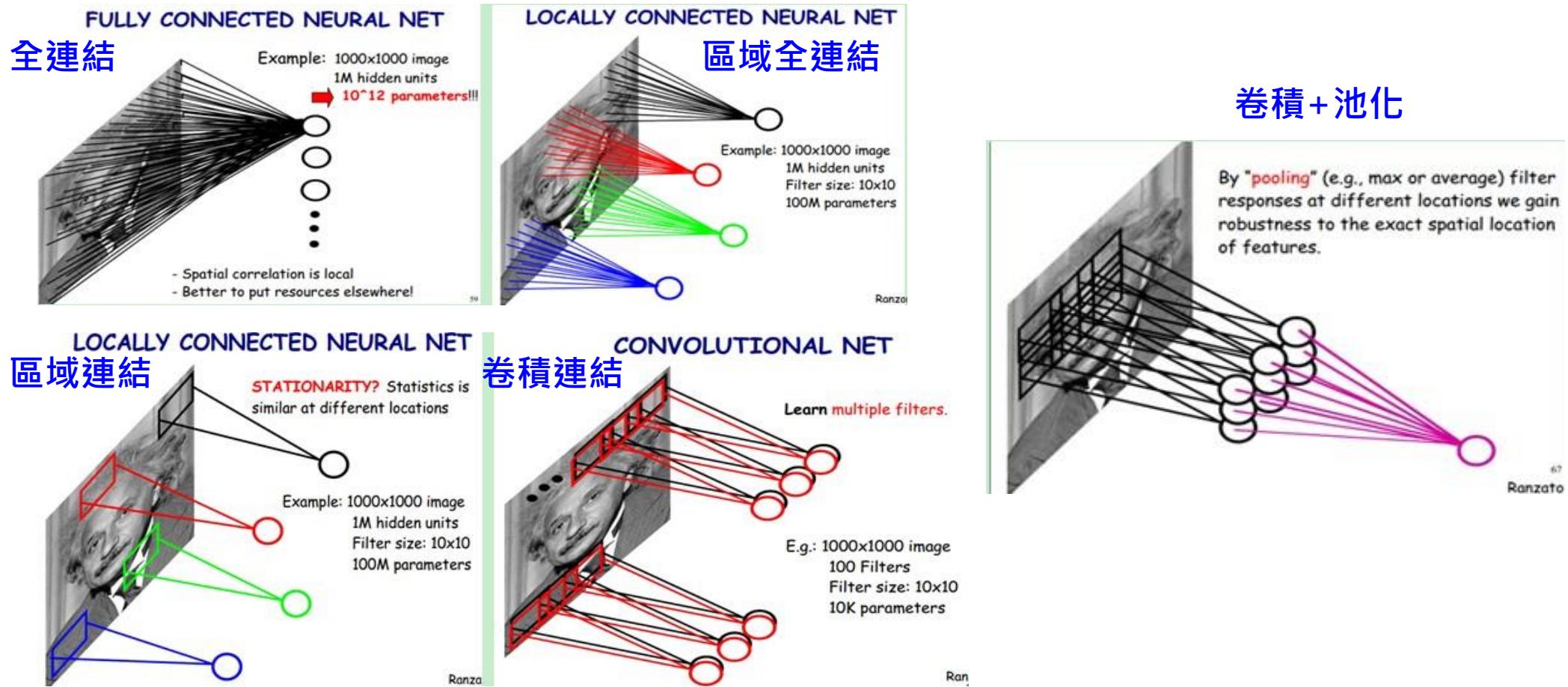
# 神經網路概念一神經網路(NN)



# 深度神經網路(DNN)



# 卷積神經網路—演化



資料來源：<http://blog.csdn.net/zouxy09/article/details/8781543>

# 卷積神經網路—卷積、池化

5x5 原始影像

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Image

經3x3卷積核  
卷積後影像

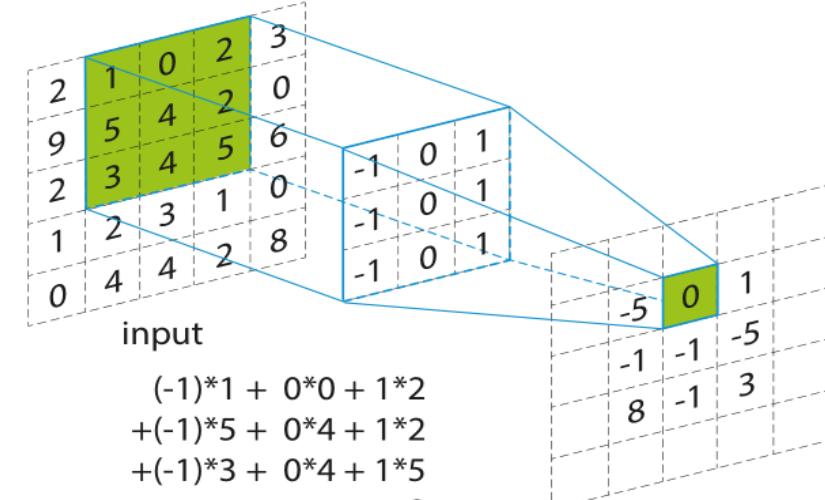
4		

Convolved  
Feature

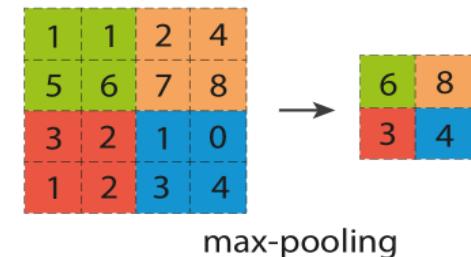
3x3卷積核

1	0	1
0	1	0
1	0	1

3x3 卷積  
(Convolution)



最大池化  
(Max Pooling)

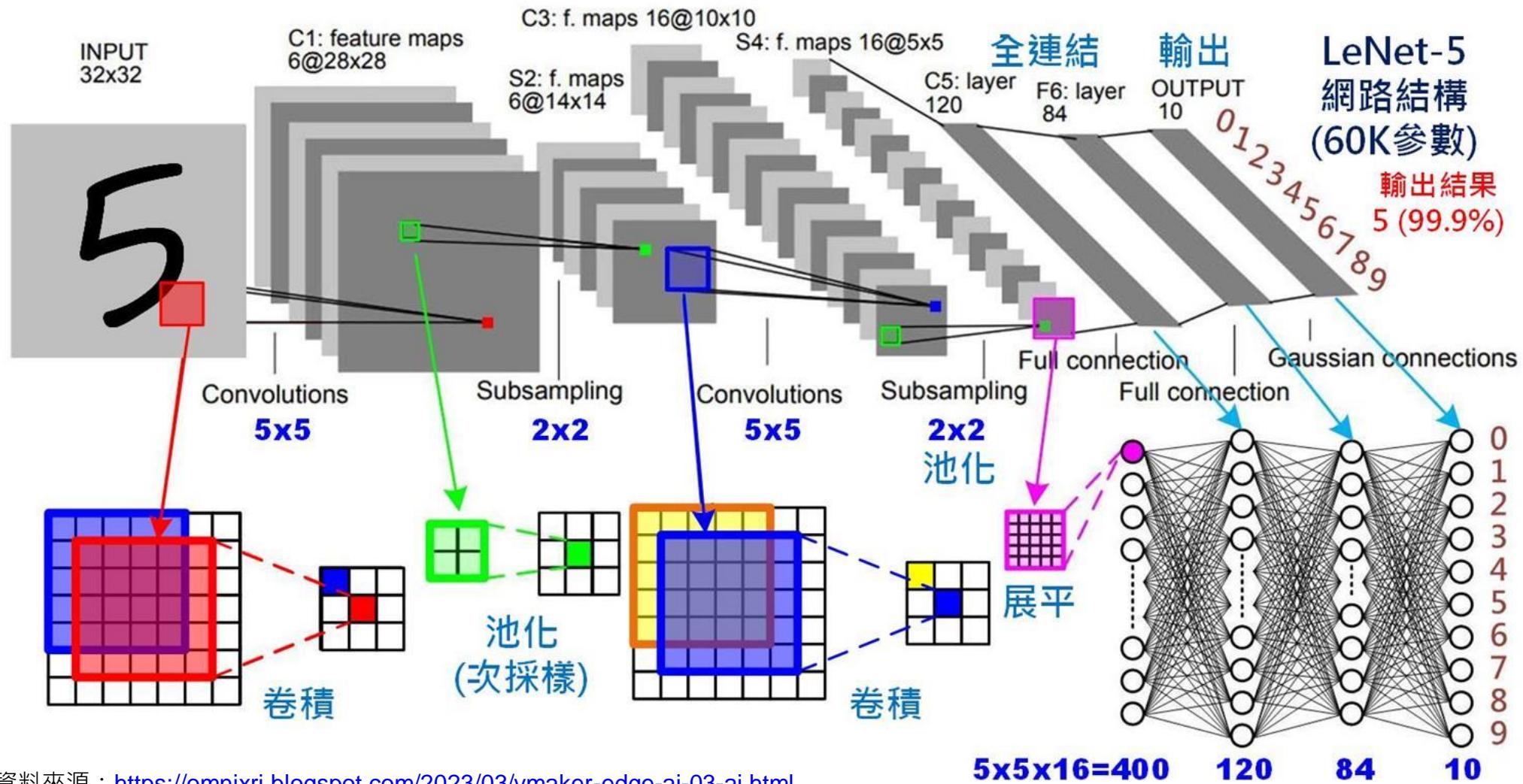


云栖社区 (b).aliyun.com

資料來源：<https://ifun01.com/BTUTFP8.html>

資料來源：<https://yq.aliyun.com/articles/64813>

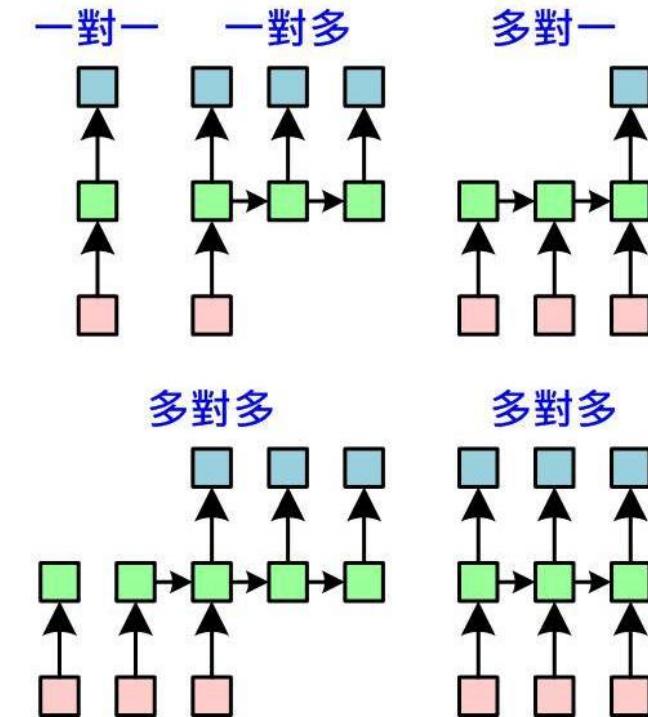
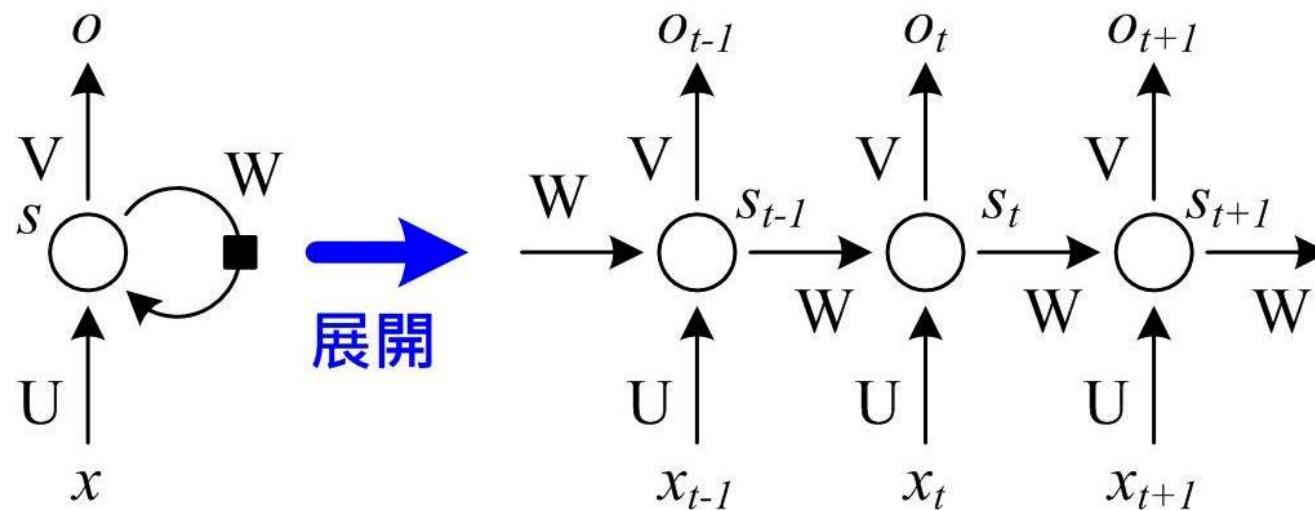
# 卷積神經網路—LeNet5



資料來源：<https://omnixri.blogspot.com/2023/03/vmaker-edge-ai-03-ai.html>

# 循環神經網路(RNN)

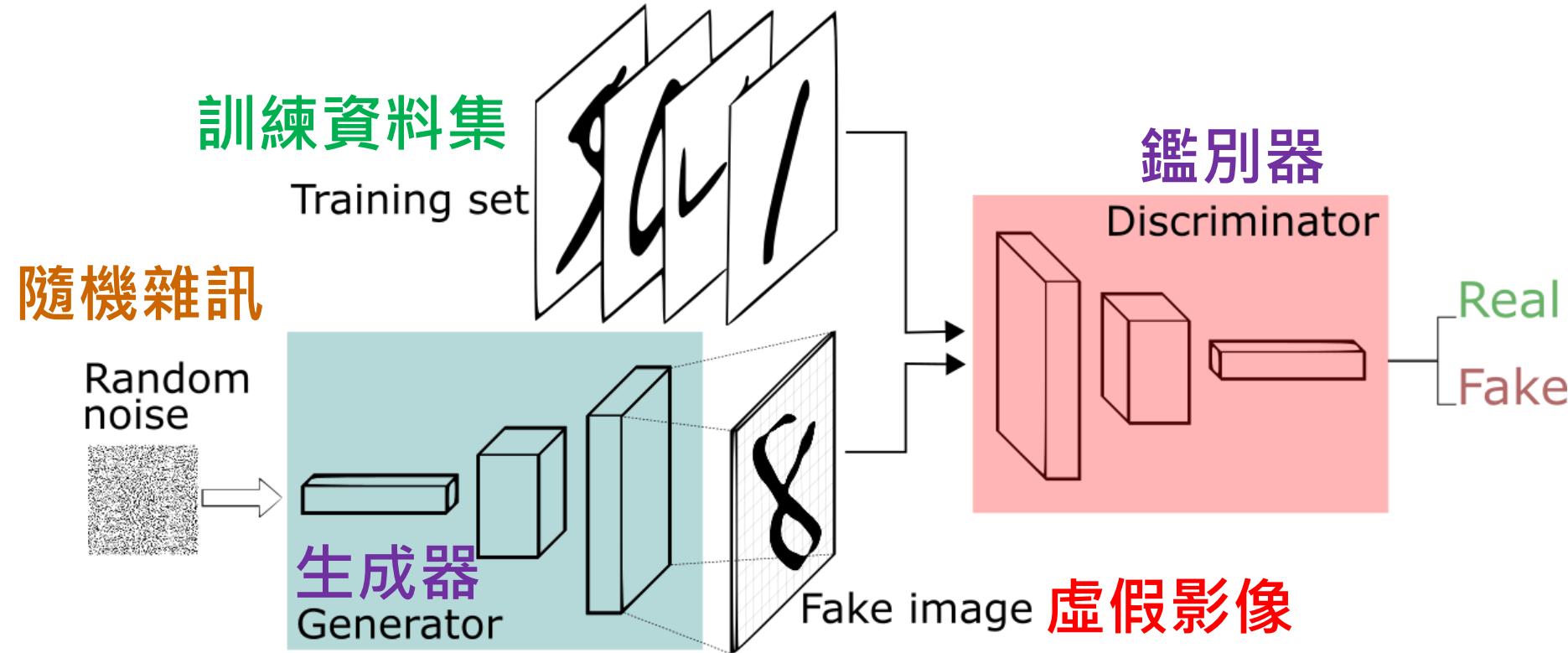
## Recurrent Neural Networks, RNN ( 時間序列型網路 )



資料來源：<http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>

# 生成對抗網路(GAN)

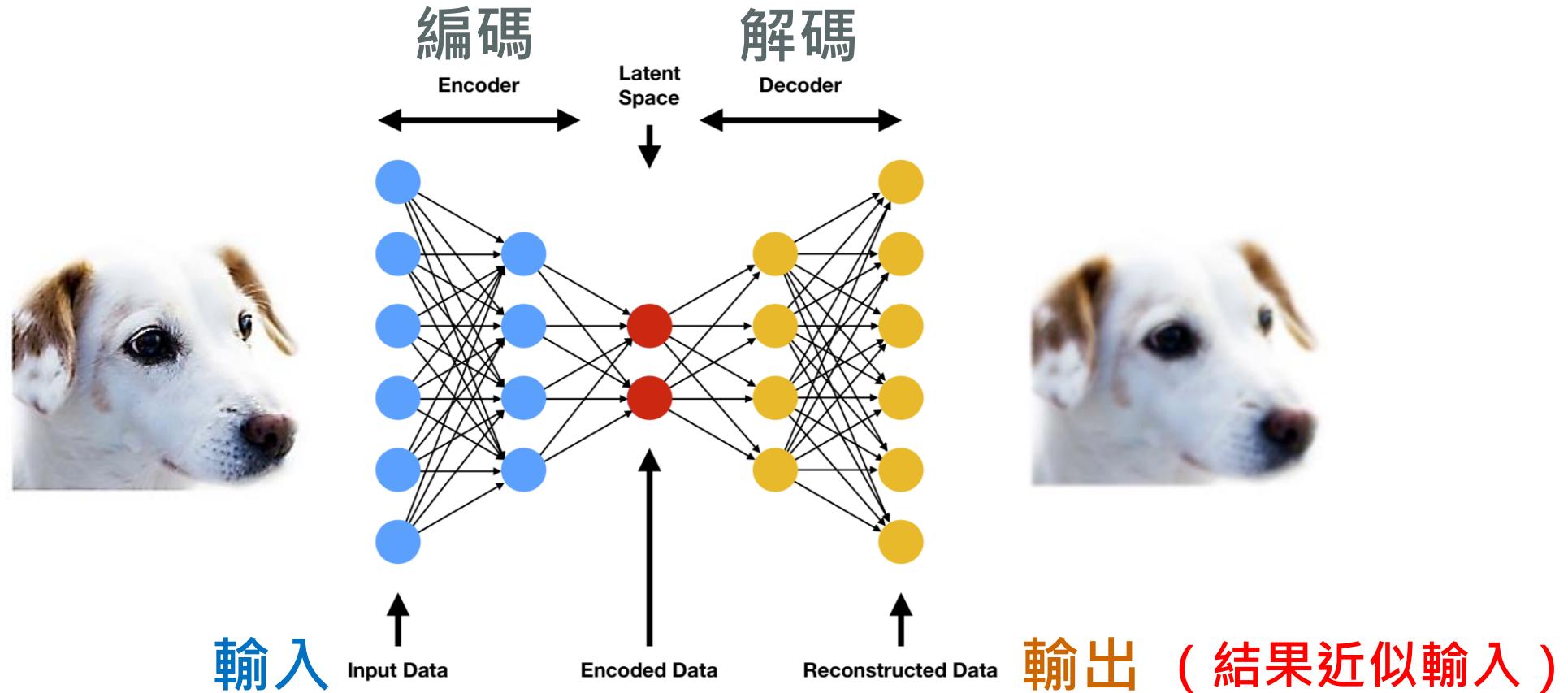
## Generative Adversarial Networks, GAN (增強式學習 )



資料來源：<https://deeplearning4j.org/generative-adversarial-network>

# 自動編碼網路(AE)

## Auto-Encoder Networks, AE(自動編/解碼式網路 )

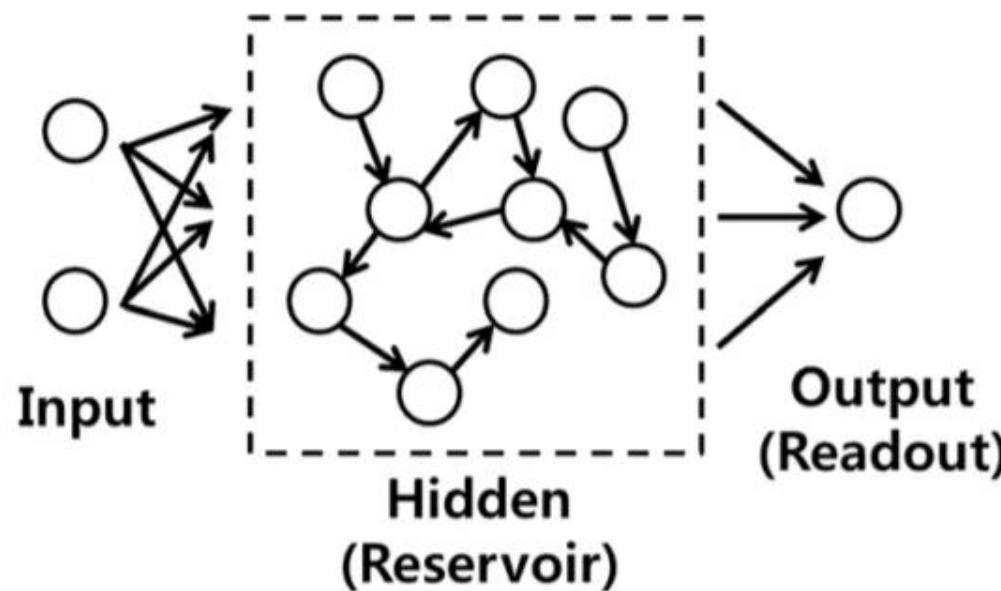


資料來源：<https://www.compthree.com/blog/autoencoder/>

# 圖神經網路(GNN) / 脈衝神經網路(SNN)

圖神經網路

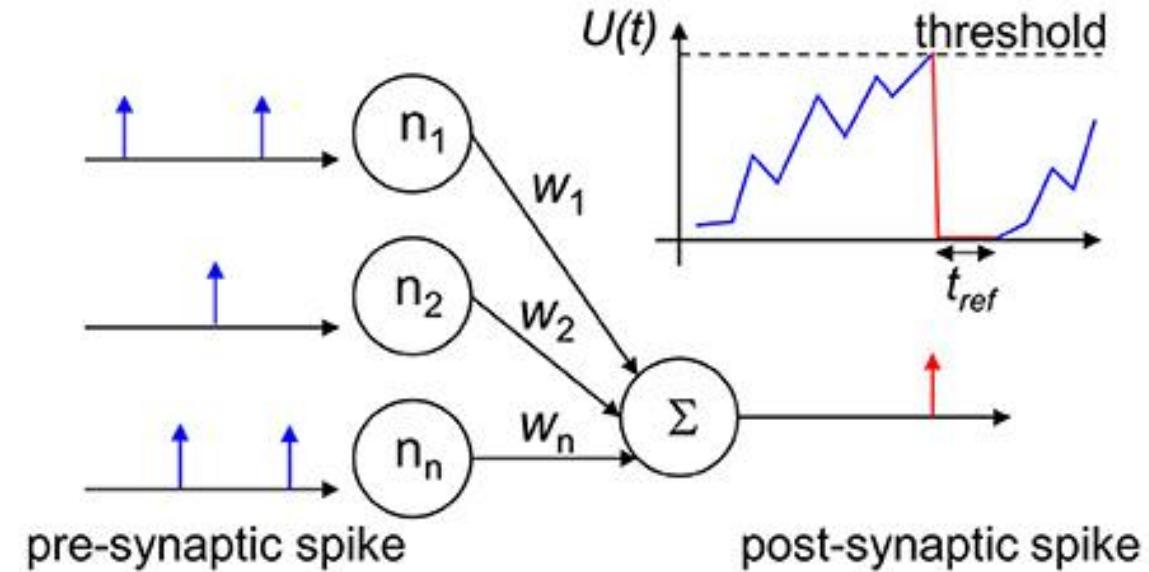
Graph Neural Network(GNN)



圖片來源：<https://pse.is/4sq9ft>

脈衝神經網路

Spiking Neural Network(SNN)



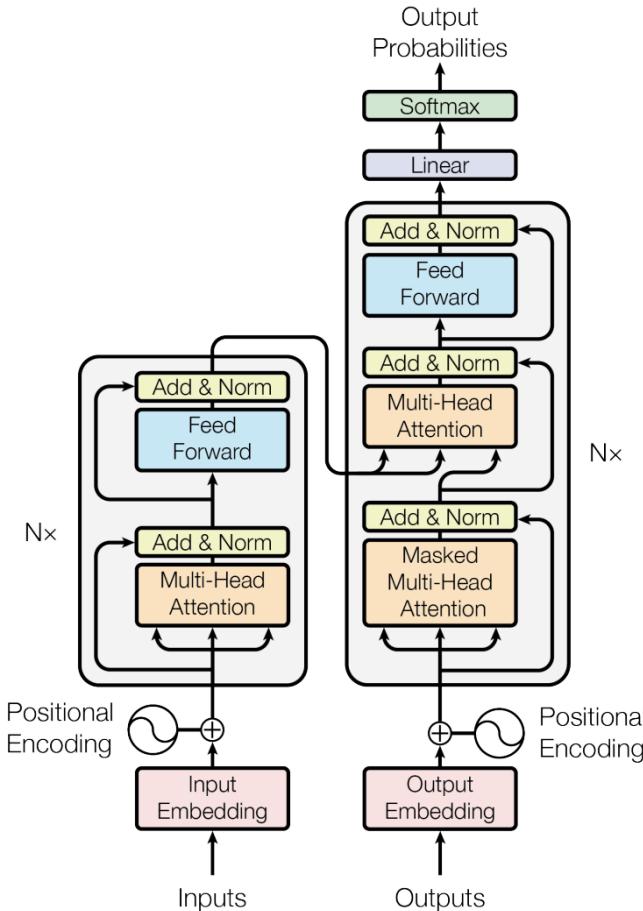
圖片來源：<https://www.frontiersin.org/articles/10.3389/fncom.2021.646125/full>

# BERT / GPT / Transformer

## BERT

### Encoder

基於變換器的雙向編碼器表示技術 ( **Bidirectional Encoder Representations from Transformers** , **BERT** )



## GPT

### Decoder

基於轉換器的生成式預訓練模型 ( **Generative pre-trained transformer**, **GPT** )

## Transformer

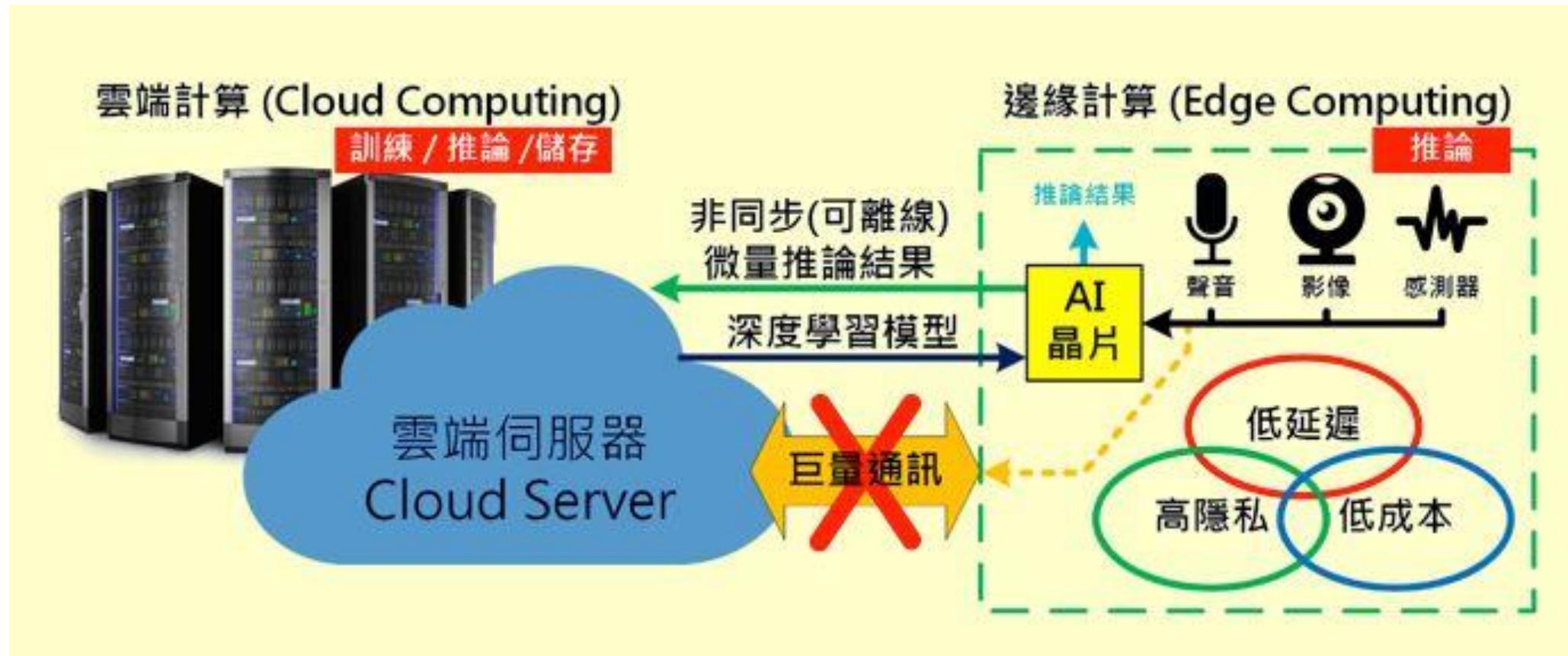
資料來源：<https://heidloff.net/article/foundation-models-transformers-bert-and-gpt/>

## 1.4. 雲端與邊緣運算



- 雲端與邊緣運算架構
- 雲端虛擬服務
- 邊緣智慧等級
- 邊緣智慧裝置
- 硬體模型選用考量

# 雲端與邊緣運算架構



OmniXRI Jul. 2018 整理繪製

圖片來源：<https://www.facebook.com/groups/edgeaitw/>

# 雲端虛擬服務—主機

## 傳統虛擬主機服務

- CPU / RAM / HDD
- 快速啟用指定作業系統
- 自行安裝所有工作軟體
- 所有運算由自己開發程式
- 資料集須自費儲存及管理



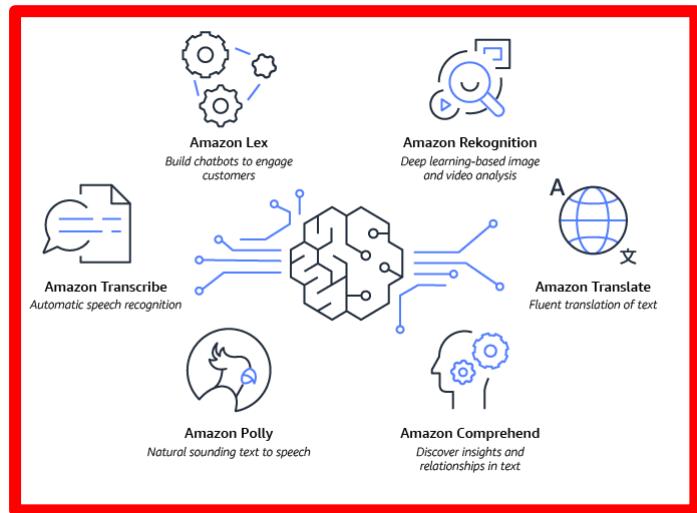
## AI專用型虛擬主機

- 支援容器(Docker, K8S)快速安裝工作環境含AI相關開發套件包
- 可動態配置GPU算力加速訓練及推論時間
- 提供專用API完成影像 / 語音 / 數據等智慧分析服務
- 共享常用資料集

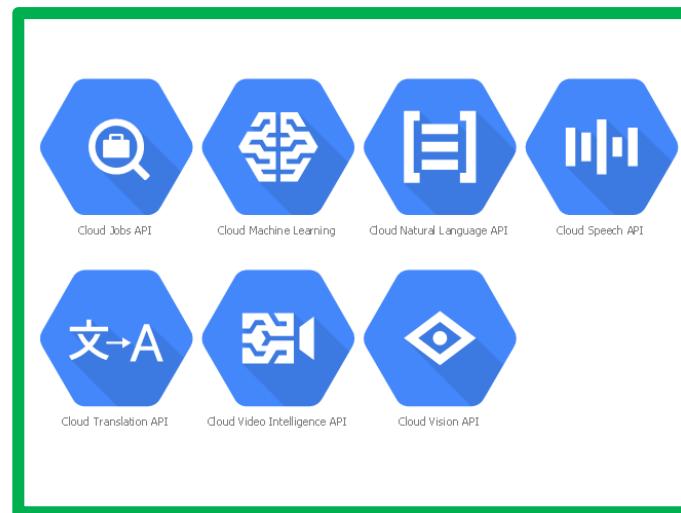
# 雲端虛擬服務—AI API



**Amazon  
AWS**

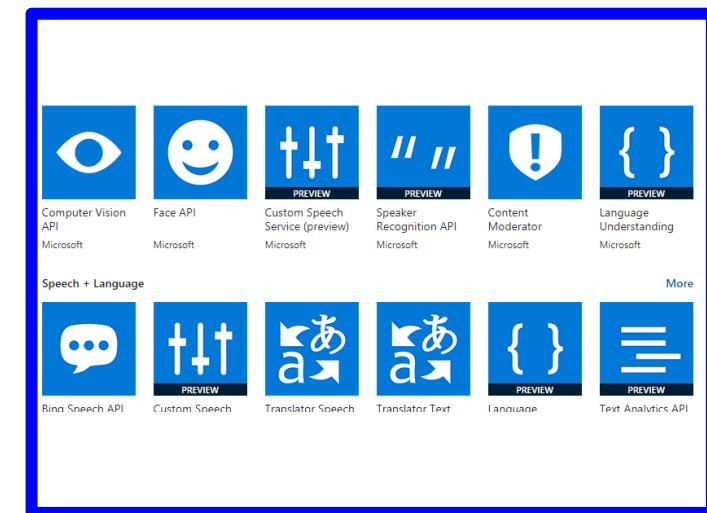


**Microsoft  
Azure**



Google Cloud Platform

**Google  
GCP**



**GenAI  
API**



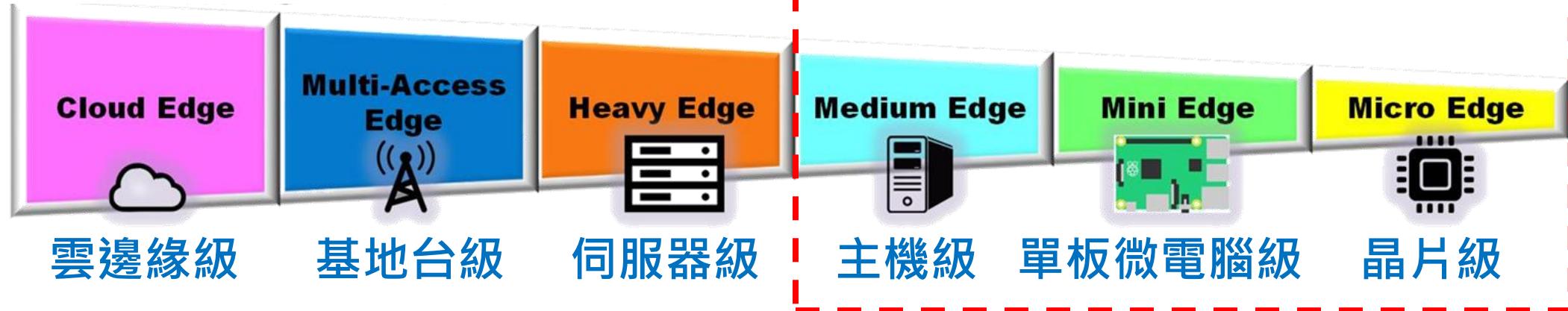
**OpenAI  
Gemini**



# 邊緣智慧等級

## 邊緣等級(Edge Level)

本課程重點範圍



**廣義邊緣智慧裝置：**不上網就具備AI模型推論能力之裝置

**狹義邊緣智慧裝置：**低功耗、高性能、可移動、電池推動之具備AI模型推論能力之裝置

# 邊緣智慧裝置（電池供電等級）

## AI PC (NB)



## MPU (SBC) AI



NPU

## Mobile AI



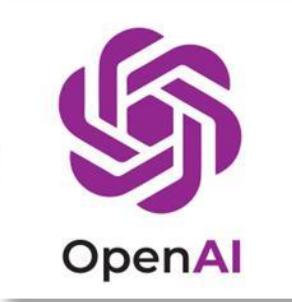
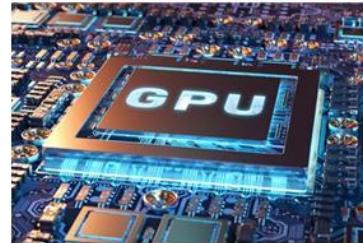
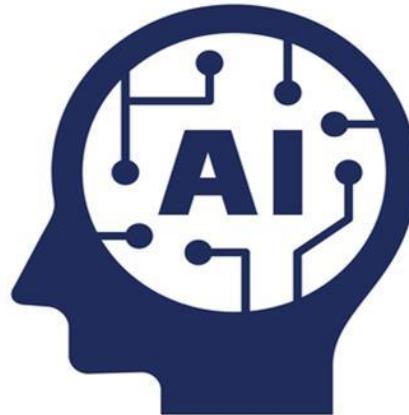
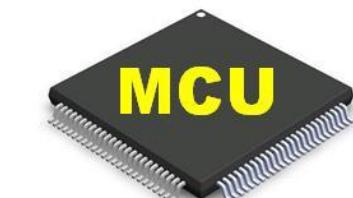
## MCU AI (TinyML)



# 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>

# 邊緣智慧裝置硬體及模型選用考量

## 硬體選用考量

- 權重容量(FLASH, HDD)
- 記憶體(SRAM, DRAM)
- 耗電量(電池/插電)
- 單價
- 更新模型方式
  - 直接本地燒錄
  - 更換記憶卡
  - 遠端OTA（額外FLASH）
- 加密及資料安全

## 模型選用考量

- 原始開發框架
- 轉換/優化工具
- 推論精度
- 推論速度
- 單獨推論時間
- 載入輸入內容時間
- 載入模型參數及轉換時間
- 前後處理時間
- 訓練設備及時間成本

# 小結

---

- 「邊緣智慧」從單晶片到近雲伺服器跨度太大，所以本課程將著重在使用電池供電的智慧裝置上如何開發及部署AI應用程式，包括單晶片（MCU）、單板微電腦（MPU）、AI PC（NB），而Mobile AI（AI手機）部份則暫時略過。
- 深度學習及常見神經網路架構已初步介紹過，後續各章節展開時會再深入探討。
- 建構「邊緣智慧」應用較大型AI應用所需基礎知識更為繁多，不僅要了解模型訓練，還有更多建立資料集及部署問題要了解，後續將陸續介紹。

# 參考文獻

---

[1] 許哲豪，NTUST Edge AI 人工智慧與邊緣運算實務(2021~2023)

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

[2] 許哲豪，【vMaker Edge AI專欄 #14】從CES 2024 看Edge AI及TinyML最新發展趨勢

<https://omnixri.blogspot.com/2024/02/vmaker-edge-ai-14-ces-2024-edge-aitinyml.html>

[3] 許哲豪，【vMaker EDGE AI專欄 #02】要玩AI前，先來認識數字系統

<https://omnixri.blogspot.com/2023/02/vmaker-edge-ai-02-ai.html>

# 延伸閱讀

---

[A] 許哲豪，【課程簡報】20221031\_開南健管系\_健康資料處理與分析\_week08\_穿戴式及健康照護感測器

<https://omnixri.blogspot.com/2022/11/20221031week08.html>

[B] 許哲豪，【課程簡報】20221130\_北科自動化所\_從智慧製造邁向元宇宙

<https://omnixri.blogspot.com/2022/11/20221130.html>

[C] 許哲豪，【課程簡報】20221128\_臺北商大\_資訊與決策科學研究所\_深度學習應用於智慧零售影像分析之趨勢

<https://omnixri.blogspot.com/2022/11/20221128.html>

[D] 許哲豪，【課程簡報】20221219\_中華資工\_AIoT與tinyML生態系國際發展趨勢與國產IC未來方向

<https://omnixri.blogspot.com/2022/12/20221219aiottinyml.html>

沒有最邊



只有更邊



歡迎加入  
邊緣人俱樂部



**YOUTUBE 直播 :** <https://www.youtube.com/@omnixri1784streams>



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](#)

[電子郵件 : omnixri@gmail.com](#)

[部落格 : https://omnidri.blogspot.tw](#)

[開 源 : https://github.com/OmniXRI](#)