

# 深度學習實作與應用

# Deep learning and its applications

## 1. Course introduction

IM5062, Spring 2024

黃意婷



加簽表單

<https://reurl.cc/QemMyO>

# 黃意婷 Yi-Ting Huang

- Email: [yitingh@ntu.edu.tw](mailto:yitingh@ntu.edu.tw)
- Assistant Professor,  
National Taiwan University of Science and Technology
- TA:施瑋翔
  - Email: [r12725019@ntu.edu.tw](mailto:r12725019@ntu.edu.tw)
  - 任何問題請先Email與助教聯繫，並副本(cc)給老師。
  - 請記得留下清楚的主旨、問題、個人資訊（學號、姓名）

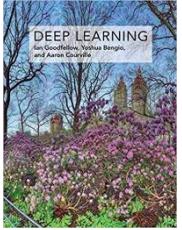
# Course logistics

- Time: Thursday, 9:10~12:00
  - 9:10~10:00
  - 10:10~11:00
  - 11:10~12:00
- Location: 管一405
- Course website: <https://cool.ntu.edu.tw/courses/37549>
  - Slides will be uploaded before or after each lecture.

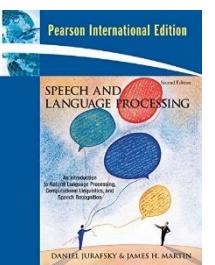
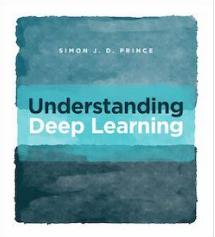
# Goals

- Introduction to Deep Learning
  - Basic MLP, convolutions, sequences
- Theory
  - Capacity control (dropout, batch norm)
  - Optimization, models, overfitting, objective functions
- Practice
  - Write code in Python
  - Solve realistic problems
- Talks from Domain experts
  - Computer vision, NLP, etc.
- Project
  - Ability to start solve a real-world problem in deep learning in a team

# Reference Books



- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Retrieved from <http://www.deeplearningbook.org/>
- Quinn, Joanne, Joanne McEachen, Michael Fullan, Mag Gardner, and Max Drummy. *Dive into deep learning: Tools for engagement*. Corwin Press, 2019. Retrieved from <https://d2l.ai/>
- Simon J.D. Prince. (2023). *Understanding Deep Learning*. MIT Press. Retrieved from <https://udlbook.github.io/udlbook/>
- Dan Jurafsky and James H. Martin. (2024). *Speech and Language Processing* (3rd ed. draft). Retrieved from <https://web.stanford.edu/~jurafsky/slp3/>
- Some slides used in this course are from the book publishers.



# Pre-requisites

- Programming language: Python
  - Python notebooks in CoLab
  - You will need a Google account (make one before next Tuesday)
    - Alternatively, set up Jupyter Notebooks yourself
      - Numpy
      - Matplotlib
      - PyTorch / Tensorflow/...
- Basic knowledge of Probability, Calculus and Linear Algebra

# Syllabus

Week	Date	Topic	備忘
1	2/20	Course introduction	
2	2/27	Basic Neural Network (I): from regression to neural networks. Regression, perceptron, forward propagation, activation functions	
3	3/5	Basic Neural Network (II): backward propagation. Loss functions, gradient descent, optimizers, normalization	
4	3/12	Basic Neural Network (III): regularization, evaluation metrics.	HW1 announce
5	3/19	Static Embedding Techniques	
6	3/26	Convolutional Networks: Architectures, convolution / pooling layers	HW2 announce
7	4/2	Guest Lecturer (1): 中研院資訊所 王建堯博士	
8	4/9	Recurrent Neural Networks: RNN, GRU, LSTM	
9	4/16	Midterm	

# Syllabus

Week	Date	Topic	備忘
10	4/23	Project Proposal	
11	4/30	Sequence to sequence learning: encoder-decoder, attention mechanism	HW3 announce
12	5/7	Guest Lecturer (2): TBD	
13	5/14	Transformer: Attention is all you need, BERT, GPT	
14	5/21	Guest Lecturer (3):八維智能 陳珮華 營運長	
15	5/28	Project presentation	
16	6/4	Project presentation	

# Grading policy

- 3 assignments + 3 (or 4) reflection reports: 40%
- NOT late submission will be acceptable.
  - 3 assignments:  $12\% * 3 = 36\%$ 
    - Kaggle (<https://www.kaggle.com/>)
      - Any problem, such as network lag, is your responsibility.
      - Discussion is allowed, but make it on your own.
  - 3 (or 4) reflection reports for talks: 4%
    - Ask a question: list your question(s) on your report sheet
    - Otherwise: write down your reflection (more than 400 words)

# Grading policy (cont.)

- Midterm exam: 30%
  - paper and pencil test
- Final project: 30%
  - In a team (2~3)
  - Novel problem: 5%
  - Data collection and validation: 5%
  - Novel approach: 5%
  - Results w/o comparisons: 5% (w/ 5%)
  - Proposal / Presentation / Report: 5%

# 加選與旁聽

- 加選：

- 若有加簽需求，在**2/24 11:59PM**以前至Google表單填寫加簽意願，  
會在**2/26 5PM**前，將授權碼寄給大家。
- 優先順序：
  - 資管所
  - 資管系(含雙主修)，以年級為優先順序
  - 管院學生，以年級為優先順序
  - 其他

- 旁聽：

- 在**2/26 5PM**以前寄信給助教
- 可使用**NTU COOL**，參加非同步的課程與線上作業。



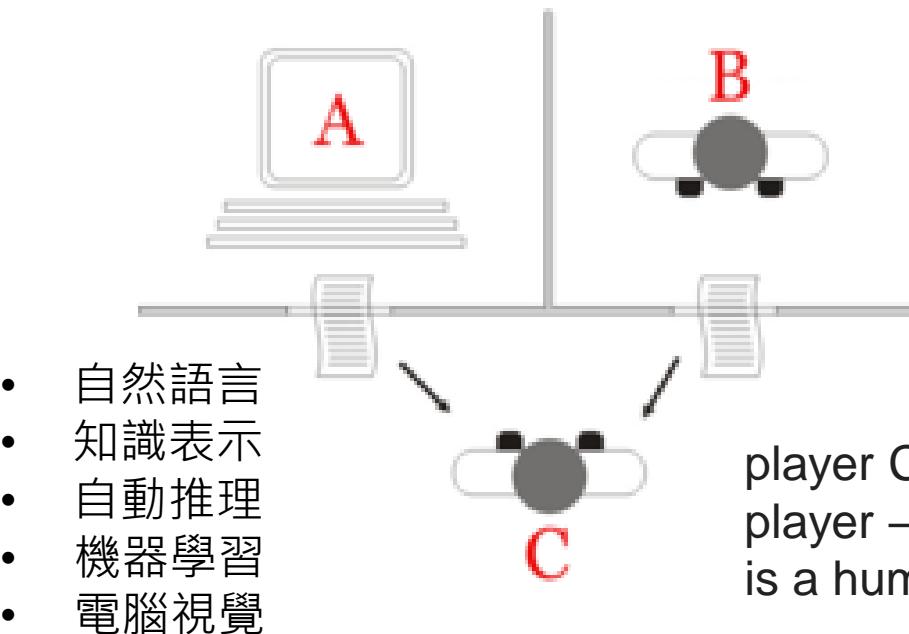
加簽表單

<https://reurl.cc/QemMyO>

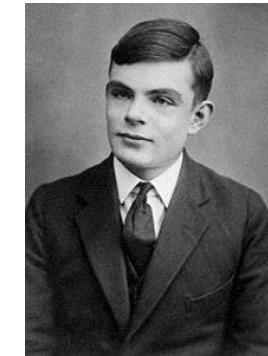
# History of AI and Deep Learning

# History (1948~1974)

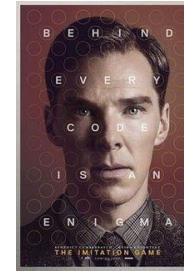
- In 1948, Alan Turing: “Intelligent Machinery”
  - An early discussion of neural networks
  - *Can Machines Think?*
- Turing Test



Source: [Wikipedia](#)



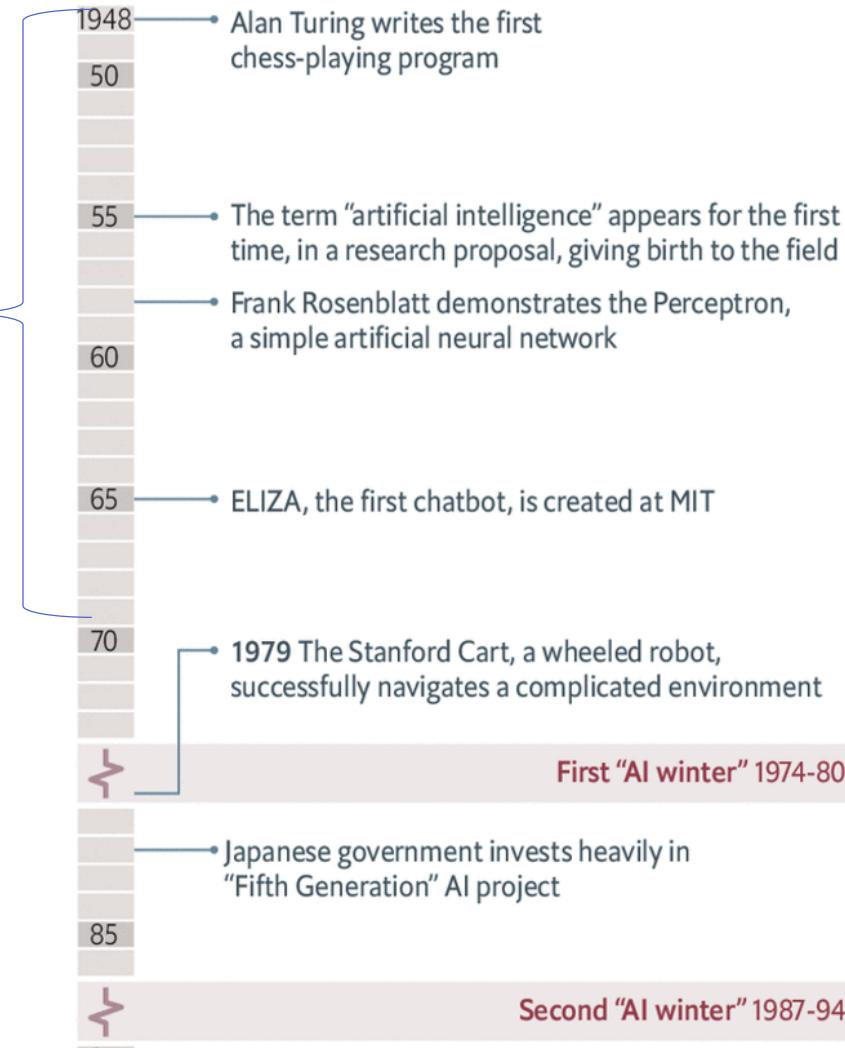
Alan Turing 模仿遊戲



深度學習實作與應用

## The new revolution

Artificial intelligence, selected events



Source: *The Economist*

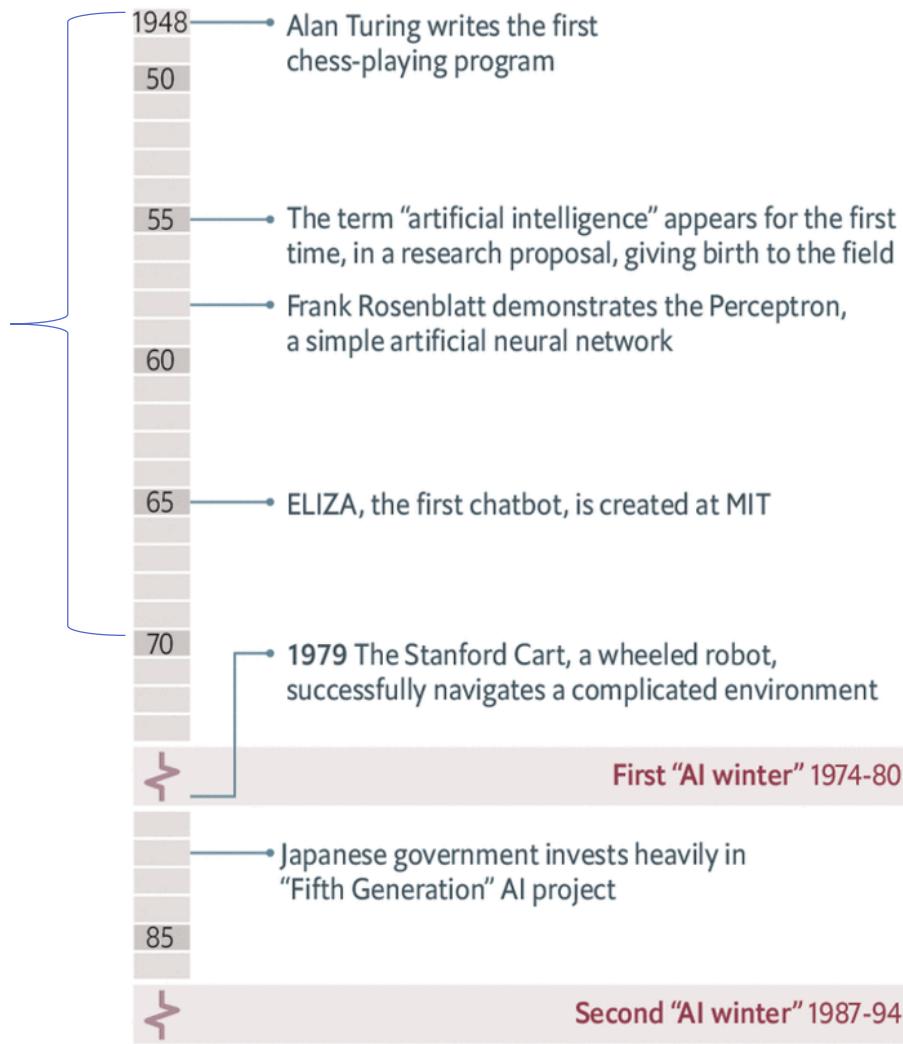
The Economist

# History (1948~1974)

- In 1948, Alan Turing: “Intelligent Machinery”
- In 1958, Frank Rosenblatt demonstrates the Perceptron, a simple **artificial neural network**.
- In 1965, ELIZA, the first **chatbot**, is created at MIT.

## The new revolution

Artificial intelligence, selected events



# History (1974~1994)

- First AI winter 1974-80

- Limited computing/data resources

1) 電腦的運算能力非常有限，不管是記憶體或是CPU的運算都不足以解決任何實際的AI問題。

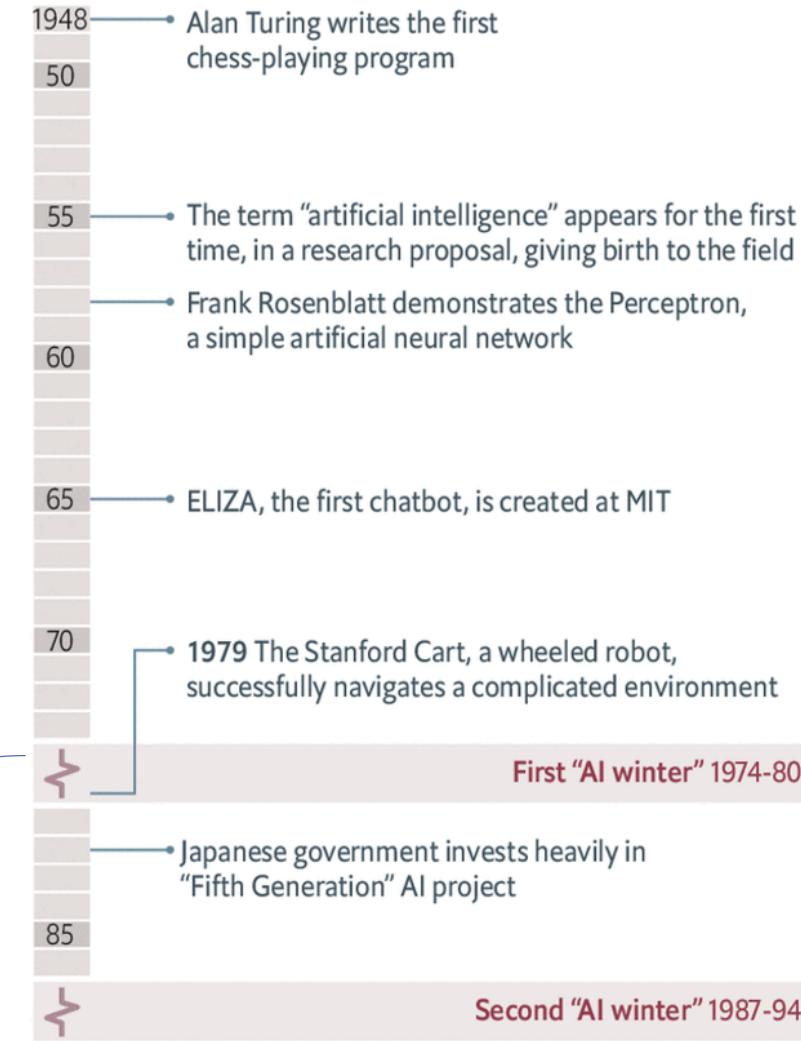
例如，在自然語言方面的研究結果，只能處理二十個字彙量。

2) 許多問題，計算所需要的問題，與輸入的規模，存在指數時間的關係。

3) 以及有很多AI的應用，像是CV和NLP，都是需要大量的world knowledge，如Wikipedia。

## The new revolution

Artificial intelligence, selected events



# History (1974~1994)

- In 1980, Expert system

- 專家系統是一種包含很多規則(rule)，透過在知識中推演，能夠像專家一樣的，回答某一個特定領域的問題。

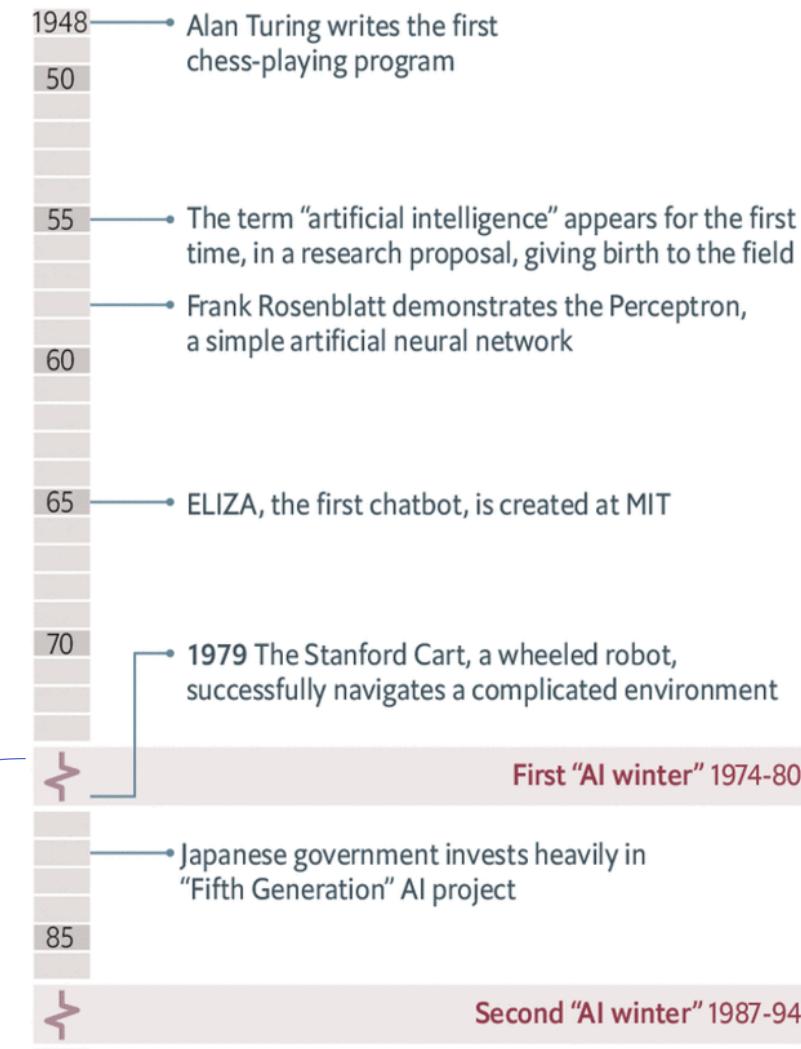
- In 1981, Japanese government invests heavily in “Fifth Generation” AI project.

- 大家發現專家系統需要維護，而且他的實用性僅限於某些特定的情景

- Second AI winter 1987-94

## The new revolution

Artificial intelligence, selected events

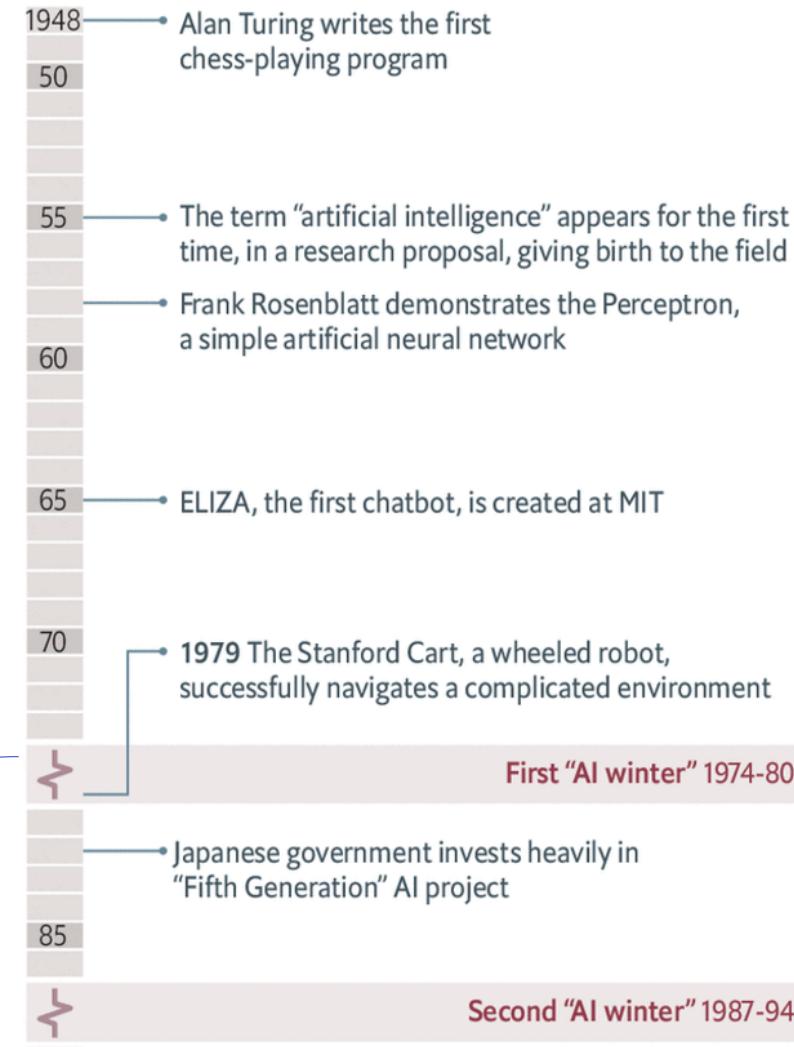


# History (1974~1994)

- In 1980, Expert system
  - 專家系統是一種包含很多規則(rule)，透過在知識中推演，能夠像專家一樣的，回答某一個特定領域的問題。
  - In 1981, Japanese government invests heavily in "Fifth Generation" AI project.
  - 大家發現專家系統需要維護，而且他的實用性僅限於某些特定的情景
- 1986 - Back propagation
  - Rumelhart, D. E., G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol.323, no.6088, pp.533-536, 1986.
- Second AI winter 1987-94

## The new revolution

Artificial intelligence, selected events



# History (1996~1997)

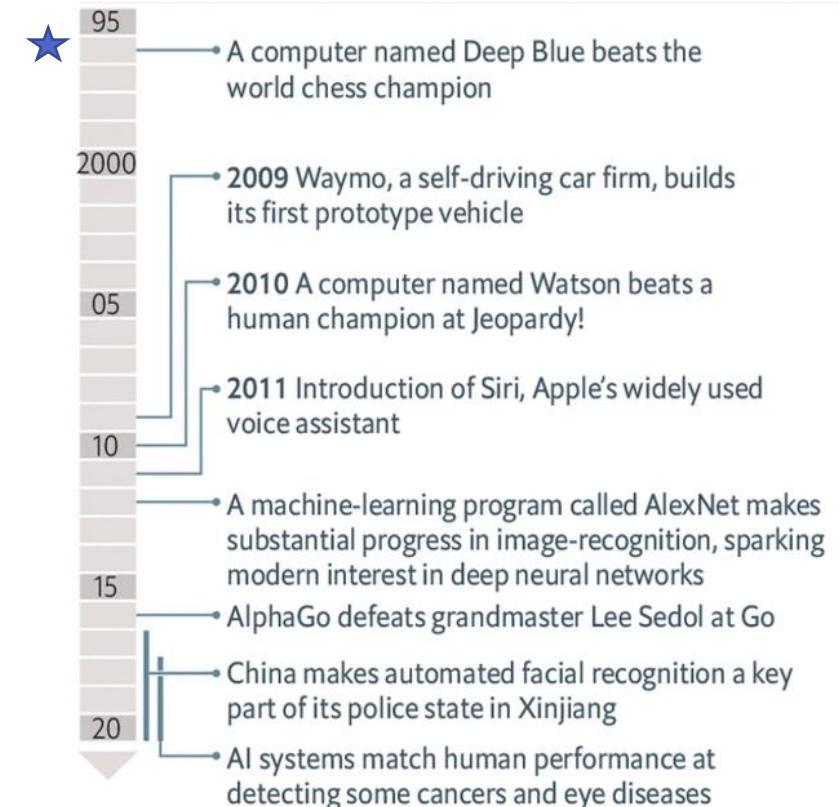
- In 1997, Deep Blue beats the world chess champion.
  - 1997年的深藍可搜尋及估計隨後的12步棋，而一名人類棋手大約可估計隨後的10步棋。



Source: [Wikipedia](#)

## The new revolution

Artificial intelligence, selected events



Source: *The Economist*

The Economist

# History (2006)

- 2006 - Deep learning (Theory) Canadian Institute for Advanced Research
  - Hinton, G. E., S. Osindero, and Y. W. Teh, "A fast learning algorithm for **deep belief nets**," Neural Computation, vol.18, no.7, pp.1527-1554, 2006.
  - Hinton, G. E. and R. R. Salakhutdinov, "Reducing the dimensionality of data with **neural networks**," Science, vol.313, no.5786, pp.504-507, 2006.

# History (2009)

- In 2009, Waymo, a self-driving car firm, builds its first prototype vehicle.



Source: [Wikipedia](#)

這些車輛使用照相機、雷達感應器和雷射測距機來「看」其他的交通狀況，並且使用詳細地圖來為前方的道路導航。

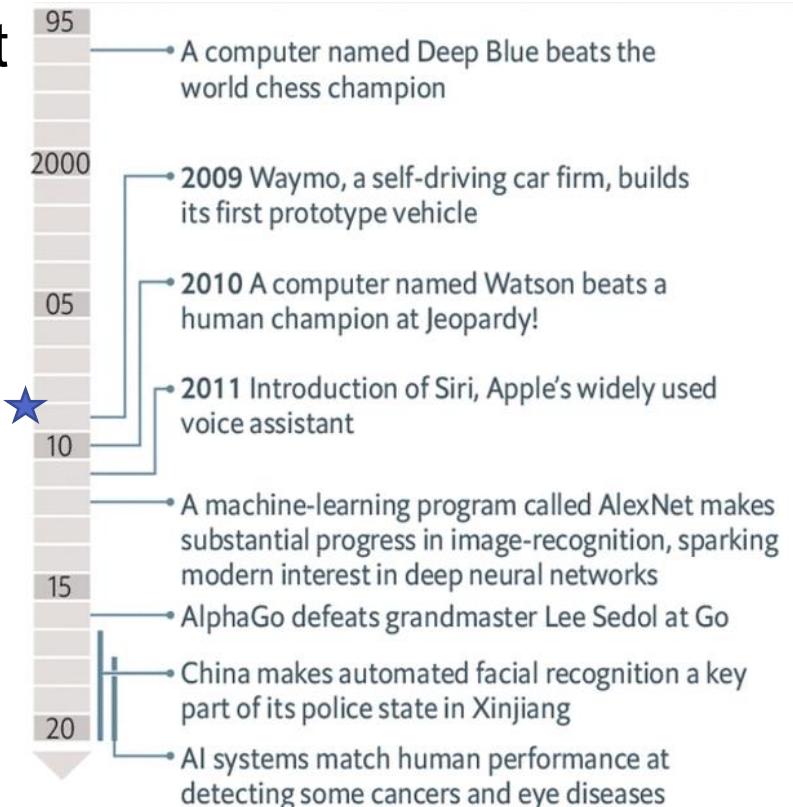


舊金山灣區的九曲花街

深度學習實作與應用

## The new revolution

Artificial intelligence, selected events



Source: *The Economist*

The Economist

Source: [Wikipedia](#)

# History (2011)

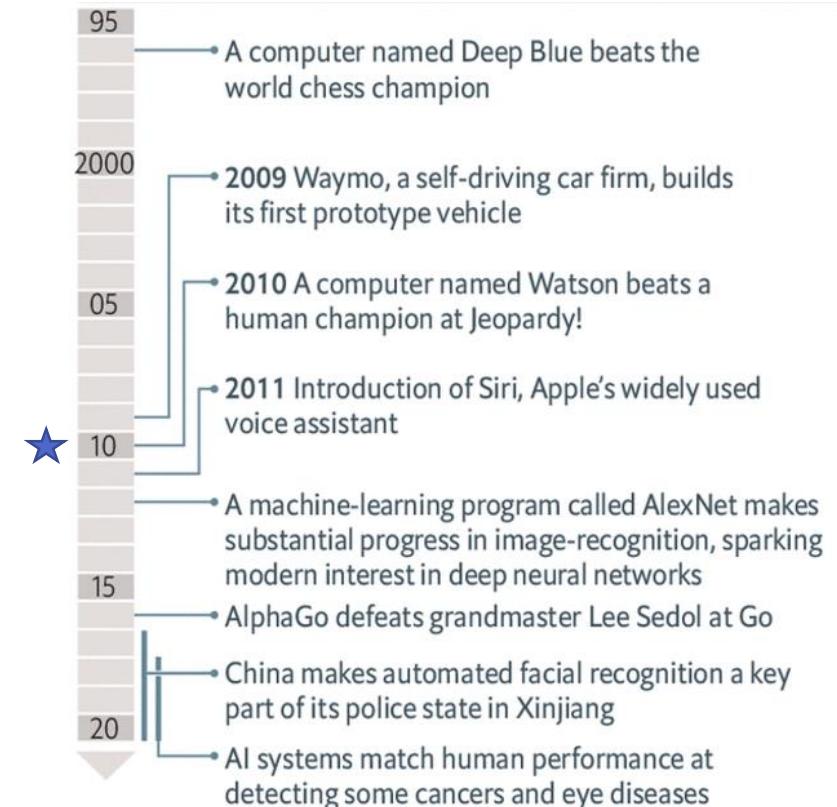
- In 2011, Introduction of Siri (Speech Interpretation and Recognition Interface), Apple's widely used voice assistant.



Source: [Wikipedia](#)

## The new revolution

Artificial intelligence, selected events



Source: *The Economist*

The Economist

# History (2012)

- In 2012, Convolutional neural networks (Image classification)
- Krizhevsky, A., I. Sutskever, and **G. E. Hinton**, "ImageNet classification with deep convolutional neural networks," in Proc. 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, NV, 2012, pp.1097-1105

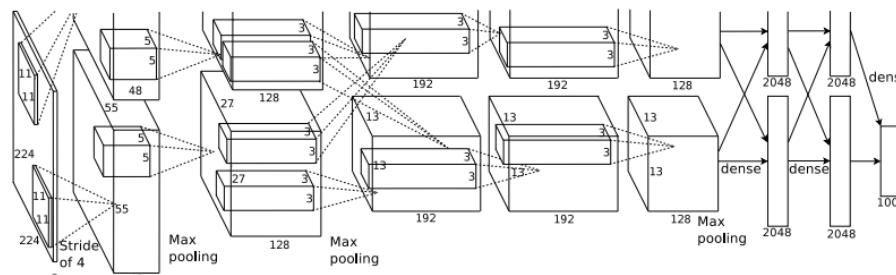
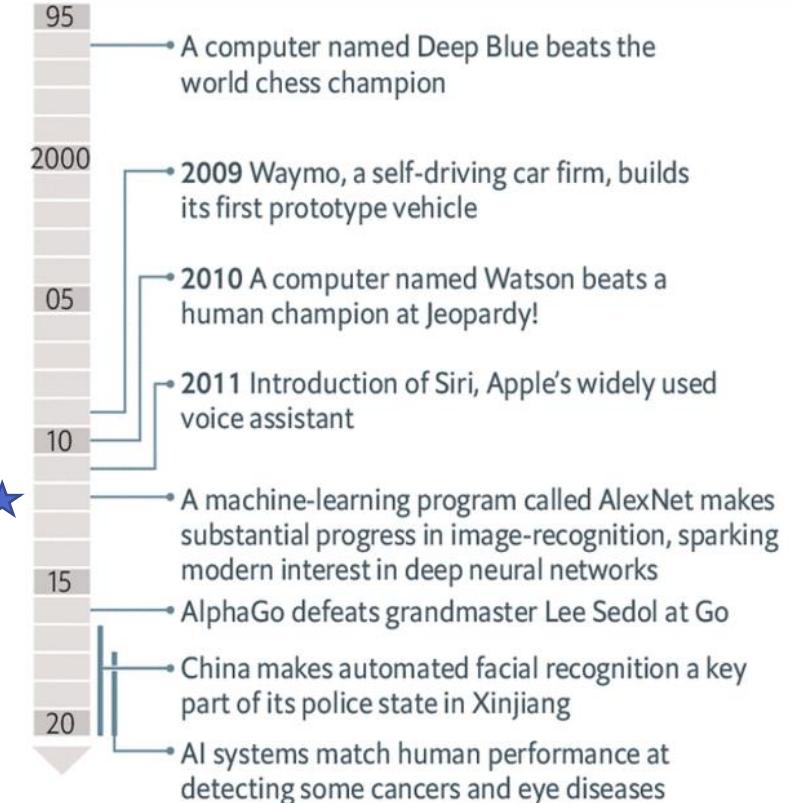


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

深度學習實作與應用

## The new revolution

Artificial intelligence, selected events



Source: *The Economist*

The Economist

# History (2016)

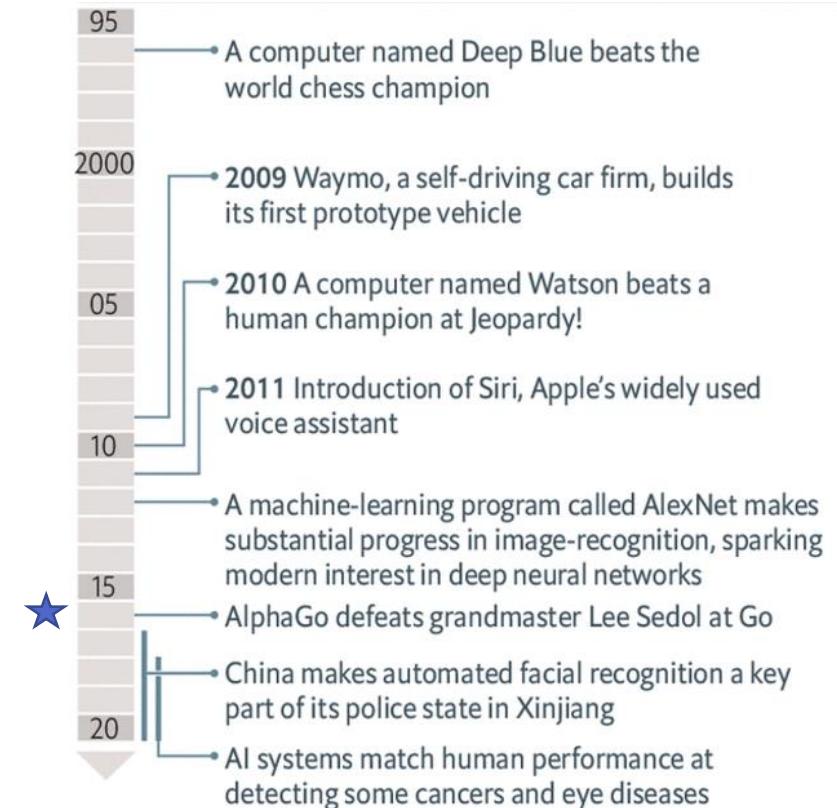
- In 2016, AlphaGo defeats grandmaster Lee Sedol at Go.



Source: CNN

## The new revolution

Artificial intelligence, selected events

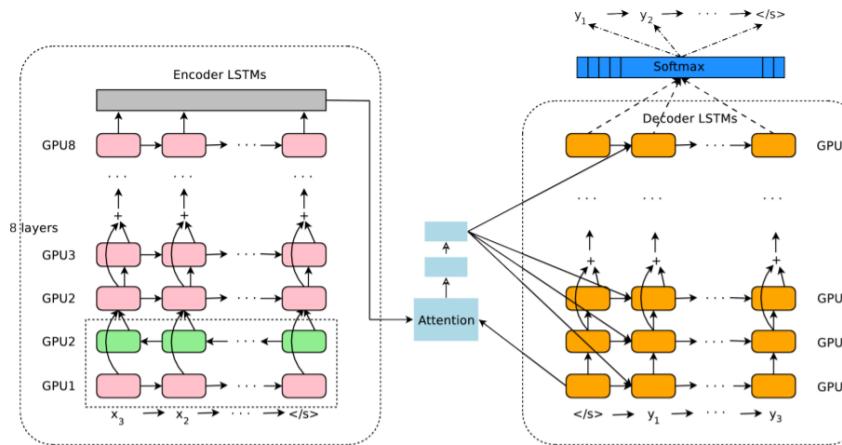
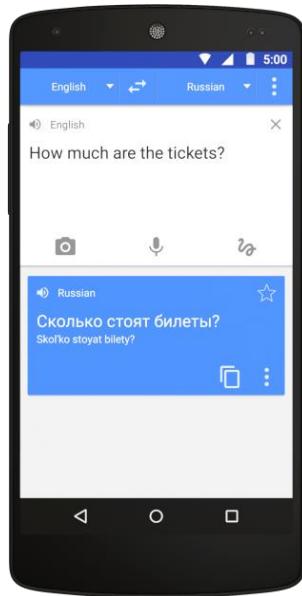


Source: *The Economist*

The Economist

# History (2016)

- In late 2016, Google announce the Google Neural Machine Translation system (GNMT), which utilizes state-of-the-art training techniques to achieve the largest improvements to date for machine translation quality.



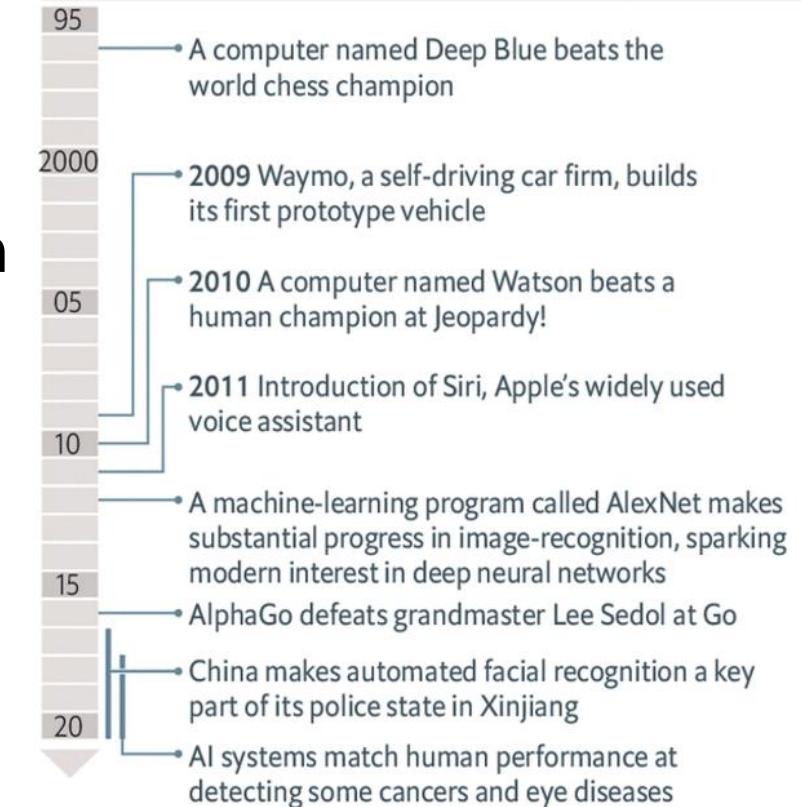
Source: Google NMT (Wu et al., 2016)

Source: Google Translate

深度學習實作與應用

## The new revolution

Artificial intelligence, selected events



Source: *The Economist*

The Economist

# History (2017~)

- Facial recognition



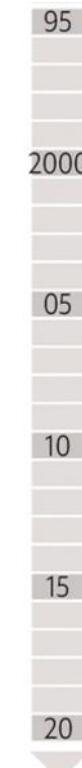
Source: [Washington Post](#)

A screenshot of a National Geographic article titled "Face-mask recognition has arrived—for better or worse". The article discusses how new algorithms can police mask-wearing. It includes a photograph of a woman wearing a mask with a bounding box around her face, and a sidebar showing statistics: "Wearing mask: 100% (1/1)" and "Wearable mask: 100% (1/1)". The sidebar also includes a "Currently wearing mask" section with a small thumbnail of a person wearing a mask.

Source: [National Geographic](#)

## The new revolution

Artificial intelligence, selected events



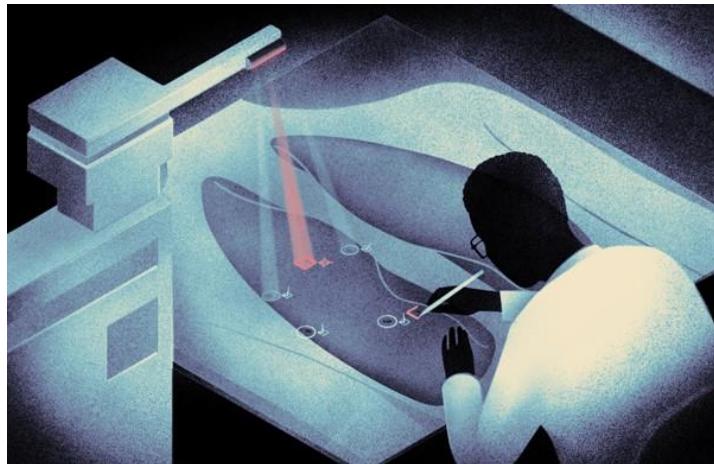
- A computer named Deep Blue beats the world chess champion
- 2009 Waymo, a self-driving car firm, builds its first prototype vehicle
- 2010 A computer named Watson beats a human champion at Jeopardy!
- 2011 Introduction of Siri, Apple's widely used voice assistant
- A machine-learning program called AlexNet makes substantial progress in image-recognition, sparking modern interest in deep neural networks
- AlphaGo defeats grandmaster Lee Sedol at Go
- China makes automated facial recognition a key part of its police state in Xinjiang
- AI systems match human performance at detecting some cancers and eye diseases

Source: *The Economist*

The Economist

# History (2018~)

- AI systems match human performance at detecting some cancers and eye diseases.



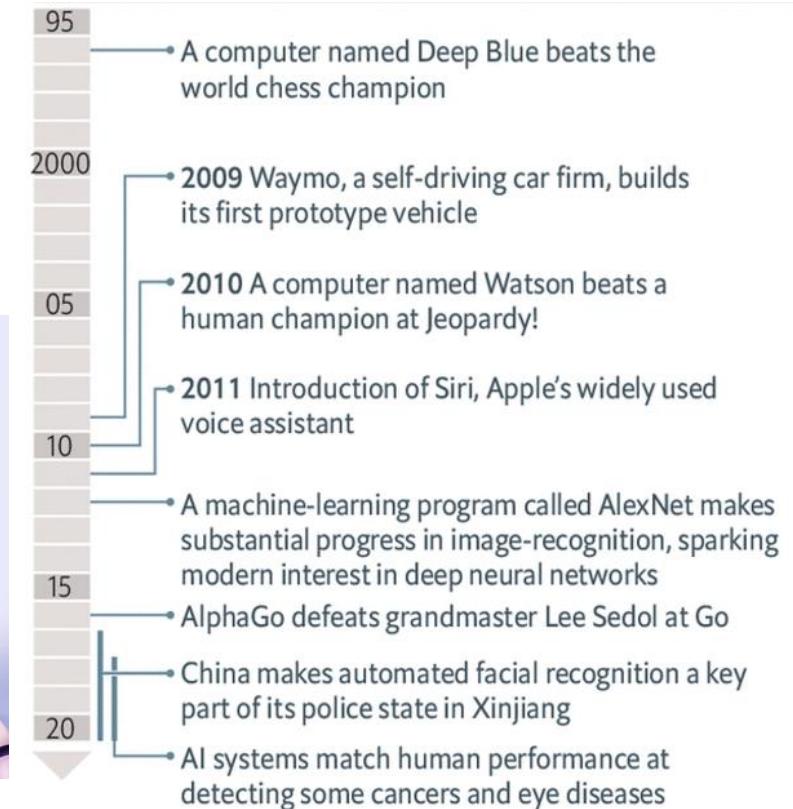
Source: [Nature](#)



Source: [Nature](#)

## The new revolution

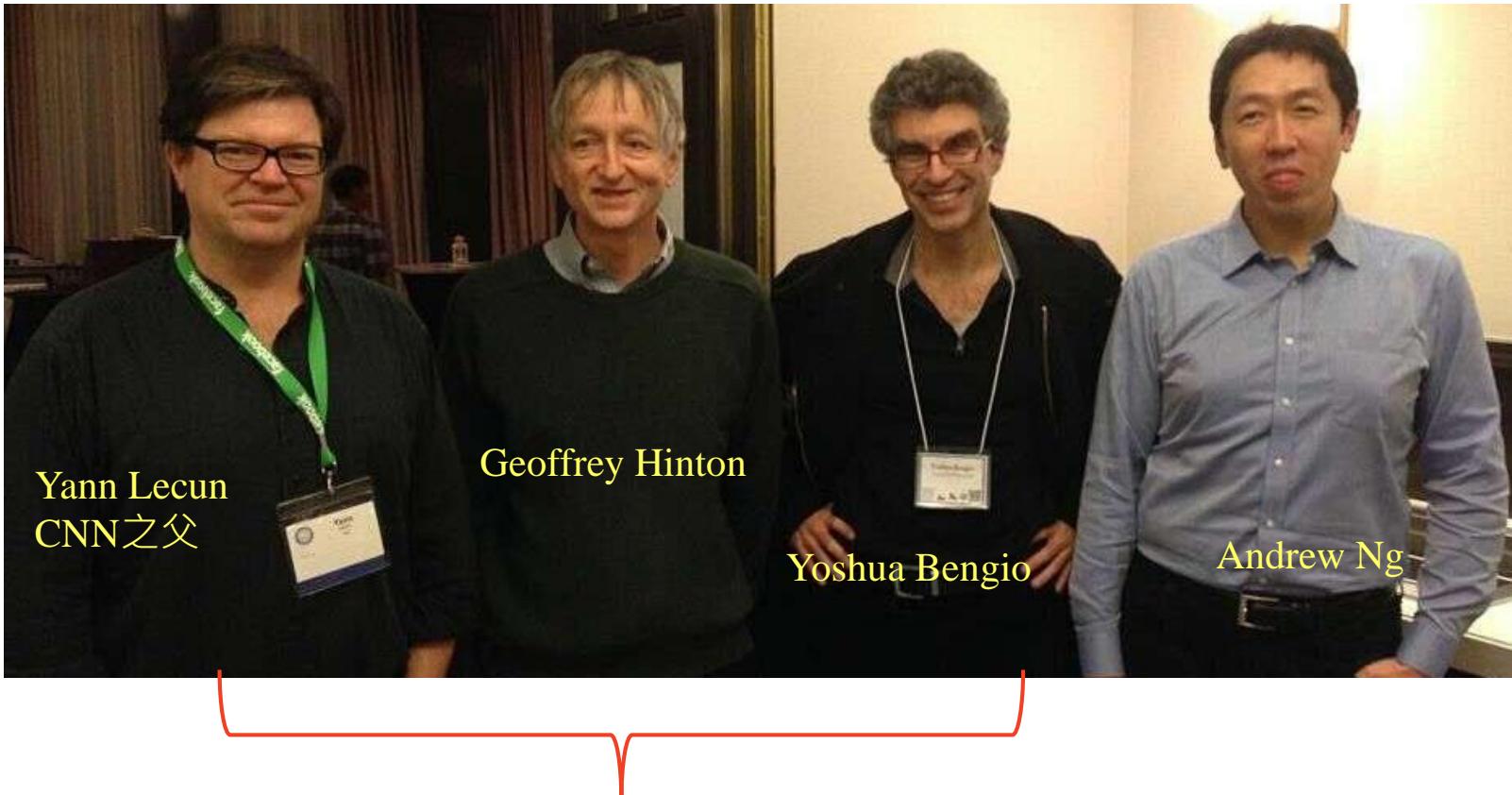
Artificial intelligence, selected events



Source: *The Economist*

The Economist

# Godfathers of AI/Deep Learning



# Recently ...



ChatGPT  
Nov., 2022

9:41

ChatGPT >

Me I am hosting my friends for brunch on Sunday and they are asking what to bring. Could you give me a list of suggestions? I'll be making a frittata and mimosas.

ChatGPT Certainly! Here's a list of brunch items your friends could bring:

1. Fresh fruit salad or a fruit platter
2. Pastries or croissants
3. Bagels with cream cheese and smoked salmon
4. Yogurt with granola and berries
5. Breakfast potatoes or hash browns
6. Sausages or bacon
7. Quiches or savory tarts
8. Pancakes or waffles with syrup
9. Coffee or tea
10. Freshly squeezed orange juice

These options can complement your frittata and mimosas nicely and provide a variety of flavors for your brunch gathering. ●

+ Send a message... 🎧



Stable Diffusion  
2022



a photograph of an astronaut riding a horse

深度學習實作與應用



Sora  
2024

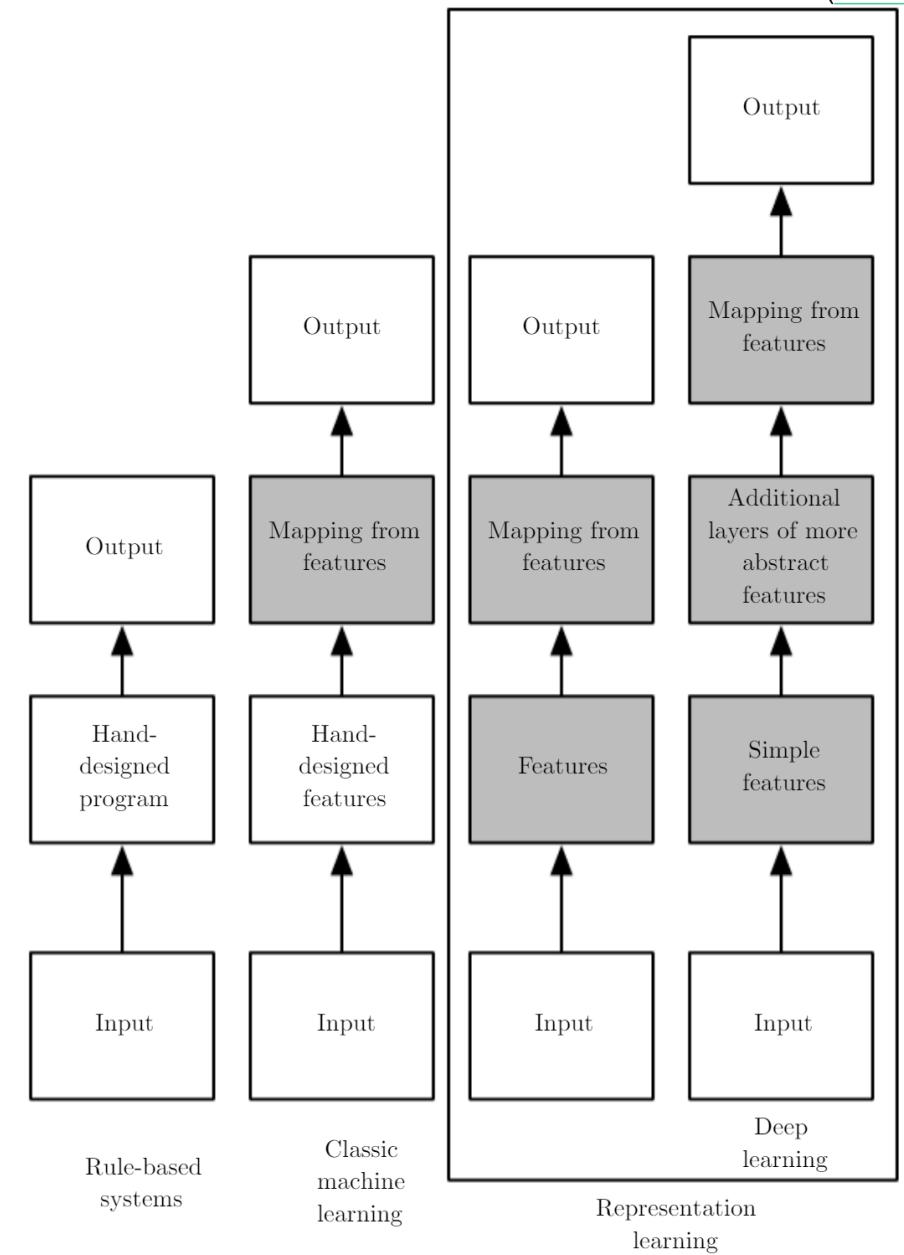
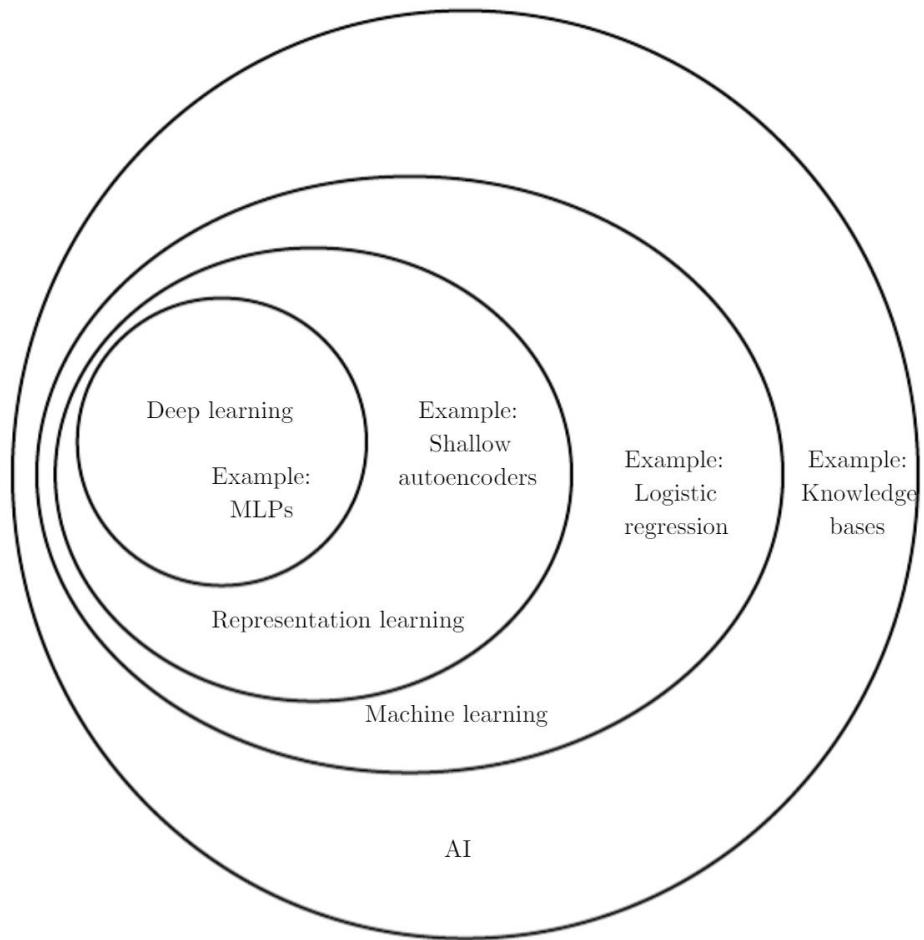


A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colorful lights. Many pedestrians walk about.

# Reasons for exploring deep learning

- In 2006, deep learning techniques started outperforming other machine learning techniques. But, why now?
  - DL techniques benefit more from **a lot of data**.
  - **Faster machines** and multicore CPU/**GPU** help DL.
  - New models, algorithms, and ideas.
- Improved performance (first in speech and vision, then NLP)
- Deep learning can learn
  - **supervised** (with specific labels such as positive/negative) and
  - **unsupervised** (from raw data).

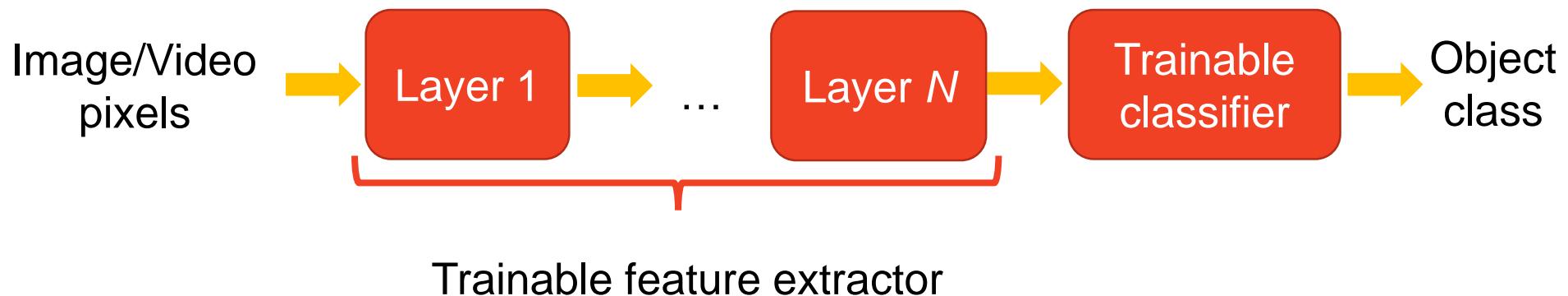
# AI / ML / RL / DL



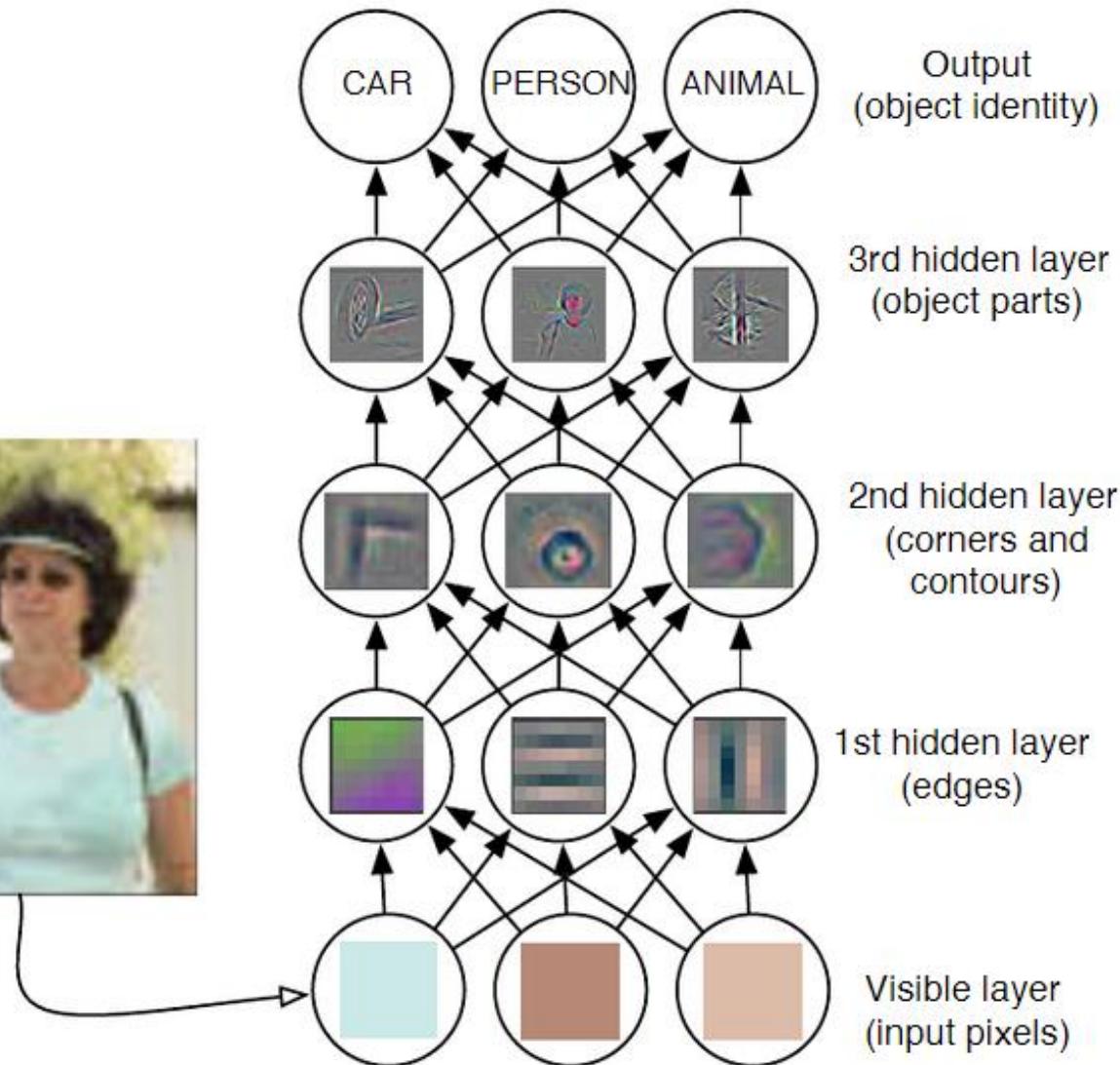
- Traditional recognition

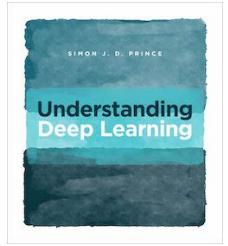


- Deep learning



# Deep learning





# Understanding Deep Learning

## CH 1. Introduction

Simon J.D. Prince. <https://udlbook.github.io/udlbook/>

# Artificial intelligence

# Artificial intelligence

Machine learning

## Artificial intelligence

- Machine learning is an area of artificial intelligence that fits mathematical models to observed data.
- It can coarsely be divided into supervised learning, unsupervised learning, and reinforcement learning.

### Machine learning

Supervised  
learning

Unsupervised  
learning

Reinforcement  
learning

# Artificial intelligence

Deep neural networks contribute to each of these areas.

Machine learning

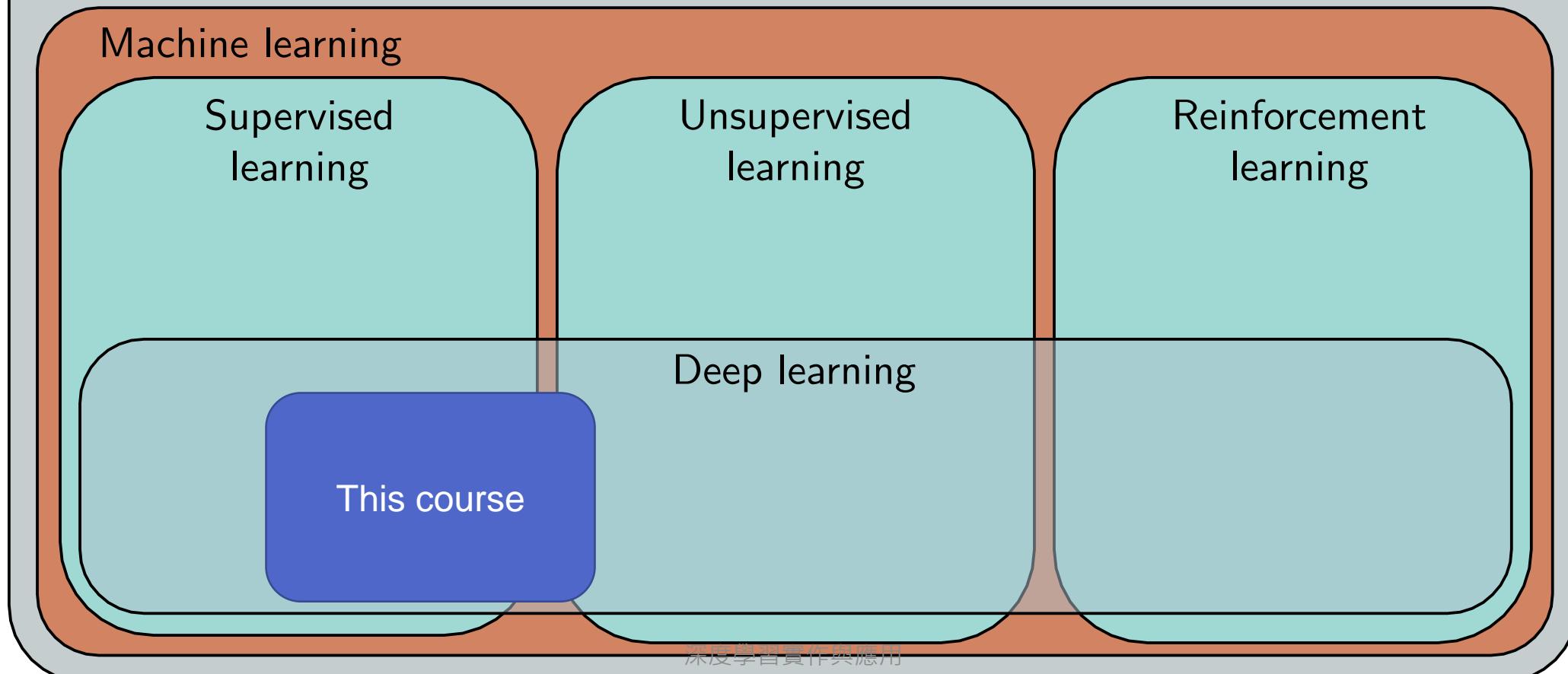
Supervised  
learning

Unsupervised  
learning

Reinforcement  
learning

Deep learning

# Artificial intelligence



# Artificial intelligence

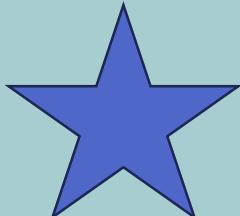
Machine learning

Supervised  
learning

Unsupervised  
learning

Reinforcement  
learning

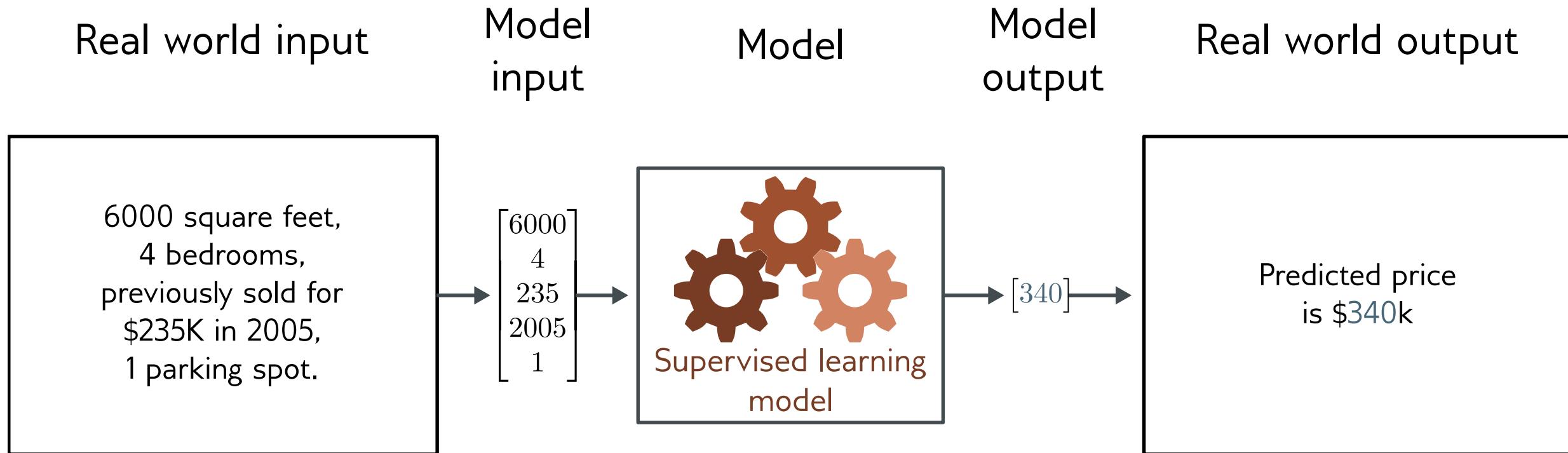
Deep learning



# Supervised learning

- Define a mapping from input to output
- Learn this mapping from paired input/output data examples

# Regression: a predicts the price of a house

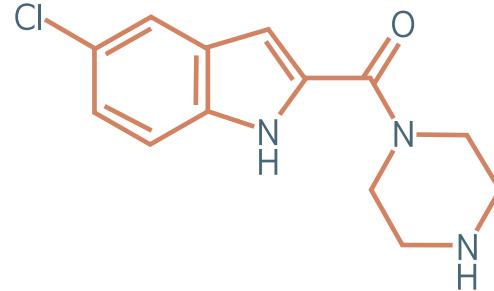


- Univariate **regression** problem (one output, **real value**)
- Fully connected network

# Graph regression: the chemical structure of a molecule as an input and predicts both the melting and boiling points

(分子的化學結構作為輸入並預測熔點和沸點)

Real world input



Model  
input

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ \vdots \\ 17 \\ 1 \\ 1 \\ \vdots \end{bmatrix}$$

Model



Model  
output

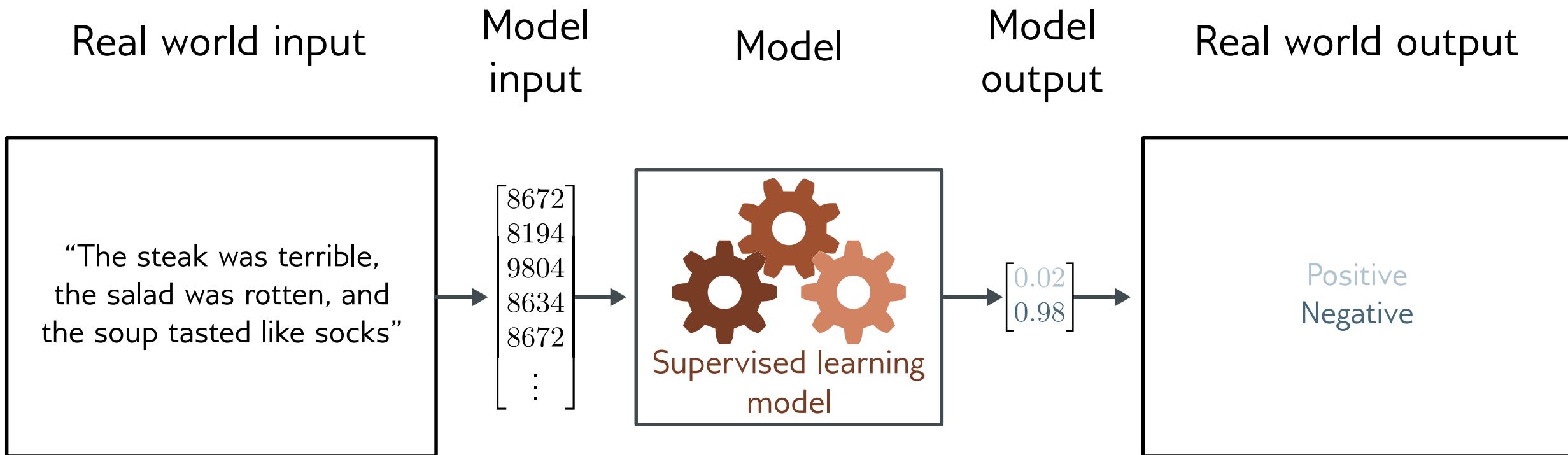
$$\begin{bmatrix} -12.9 \\ 56.4 \end{bmatrix}$$

Real world output

Freezing point  
is  $-12.9^{\circ}\text{C}$   
Boiling point  
is  $56.4^{\circ}\text{C}$

- Multivariate regression problem ( $>1$  output, real value)
- Graph neural network

Text classification: receives a text string containing a restaurant review as input and predicts whether the review is positive or negative



- **Binary classification** problem (**two discrete classes**)
  - The model attempts to assign the input to one of two categories
  - The output vector contains the probabilities that the input belongs to each category
- Transformer network

# Music genre classification: the input is an audio file, and the model predicts which genre of music it contains

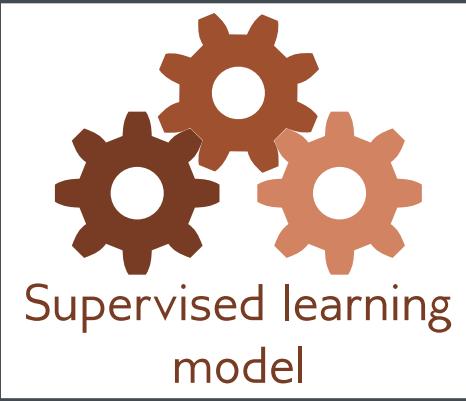
Real world input



Model  
input

$$\begin{bmatrix} 125 \\ 12054 \\ 1253 \\ 6178 \\ 24 \\ 4447 \\ \vdots \end{bmatrix}$$

Model



Model  
output

$$\begin{bmatrix} 0.03 \\ 0.52 \\ 0.18 \\ 0.07 \\ 0.12 \\ 0.08 \\ \vdots \\ 0.01 \end{bmatrix}$$

Real world output

Classical  
Electronica  
Hip Hop  
Jazz  
Pop  
Metal  
Punk

- Multiclass classification problem (discrete classes, >2 possible values)
- Recurrent neural network (RNN)

# Image classification: the input is an image, and the model predicts which object it contains

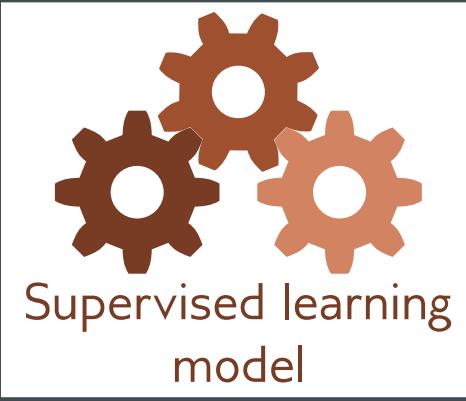
Real world input



Model  
input

$$\begin{bmatrix} 124 \\ 140 \\ 156 \\ 128 \\ 142 \\ 157 \\ \vdots \end{bmatrix}$$

Model



Model  
output

$$\begin{bmatrix} 0.00 \\ 0.00 \\ 0.01 \\ 0.89 \\ 0.05 \\ 0.00 \\ \vdots \\ 0.01 \end{bmatrix}$$

Real world output

Aardvark  
Apple  
Bee  
Bicycle  
Bridge  
Clown  
⋮

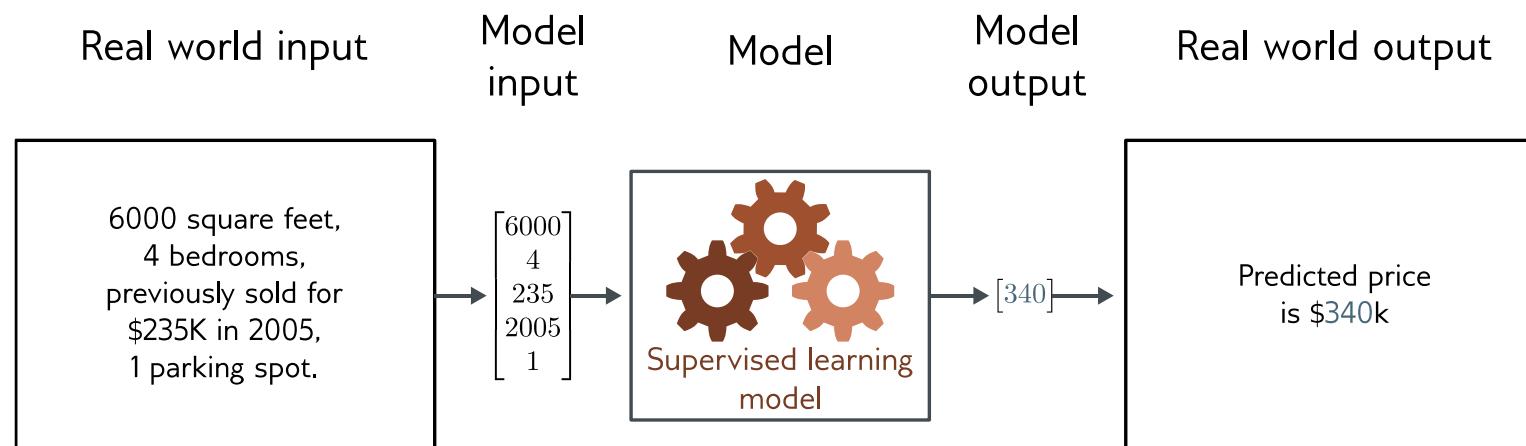
- Multiclass classification problem (discrete classes, >2 possible classes)
- Convolutional network

# Terms

- Regression = continuous numbers as output
- Classification = discrete classes as output
- Two class (binary) and multiclass classification treated differently
- Univariate = one output
- Multivariate = more than one output (multi-label)

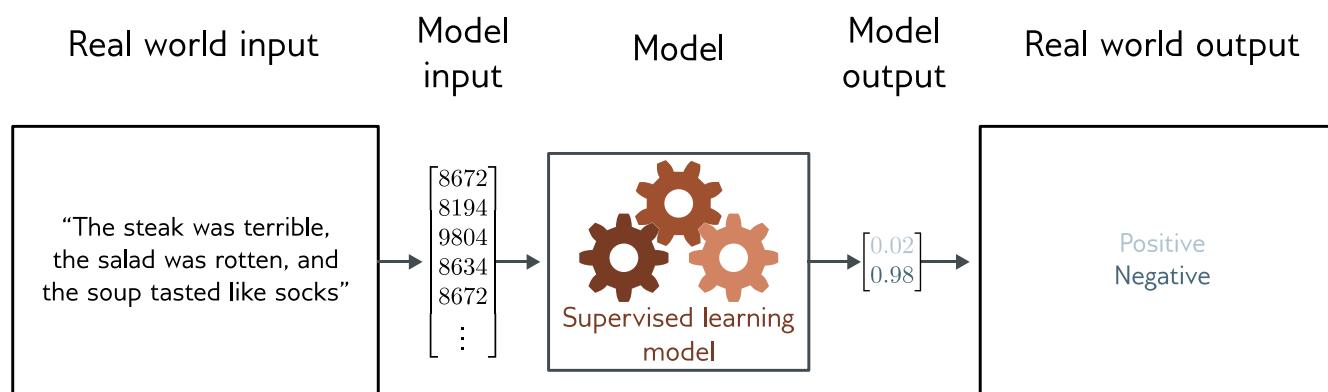
# Inputs

- In the house pricing example, the input is a fixed-length vector containing values that characterize the property.
- This is an example of tabular data because it has no internal structure;
  - if we change the order of the inputs and build a new model, then we expect the model prediction to remain the same.



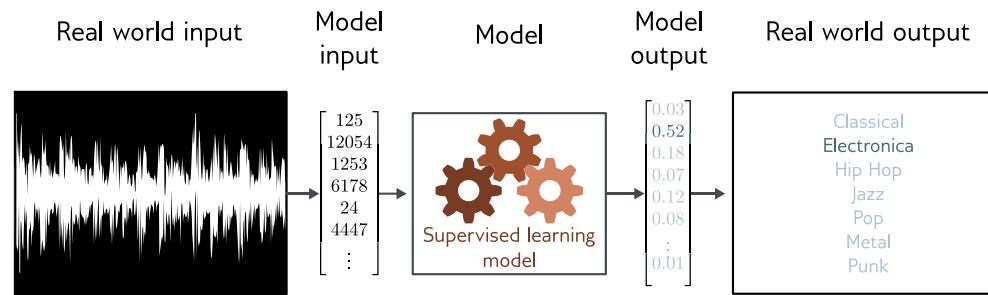
# Inputs (cont.)

- The input in the restaurant review example is a body of text.
  - This may be of **variable length** depending on the number of words in the review, and here input order is important
    - my wife ate the chicken
    - the chicken ate my wife
  - The text must be encoded into numerical form before passing it to the model.
  - Here, we use a fixed vocabulary of size 10,000 and simply concatenate the word indices.

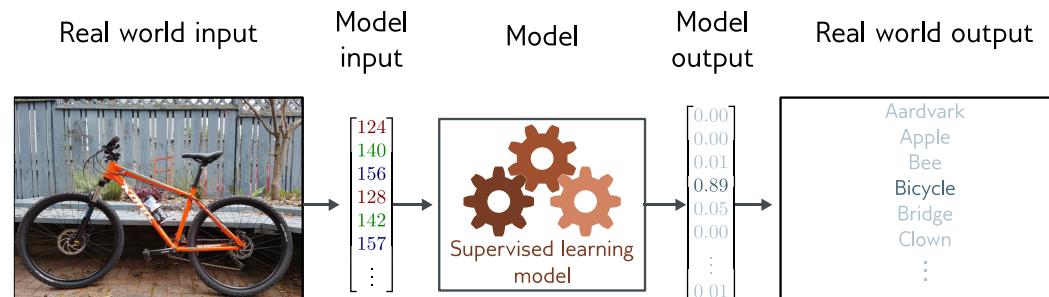


# Inputs (cont.)

- For the music classification example, the input vector might be of fixed size (perhaps a 10-second clip) but is very high-dimensional.



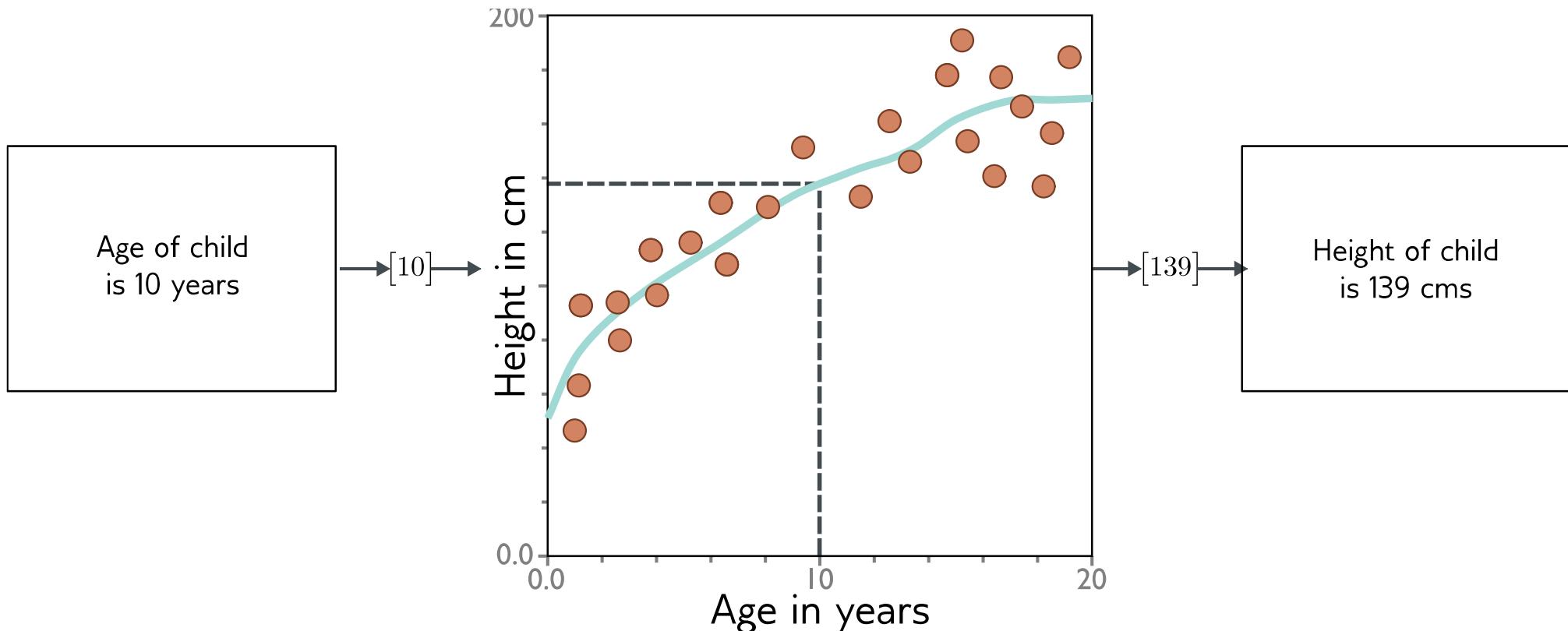
- The input in the image classification example (which consists of the concatenated RGB values at every pixel) is also enormous.
- Moreover, its structure is naturally two-dimensional; two pixels above and below one another are closely related, even if they are not adjacent in the input vector.



# What is a supervised learning model?

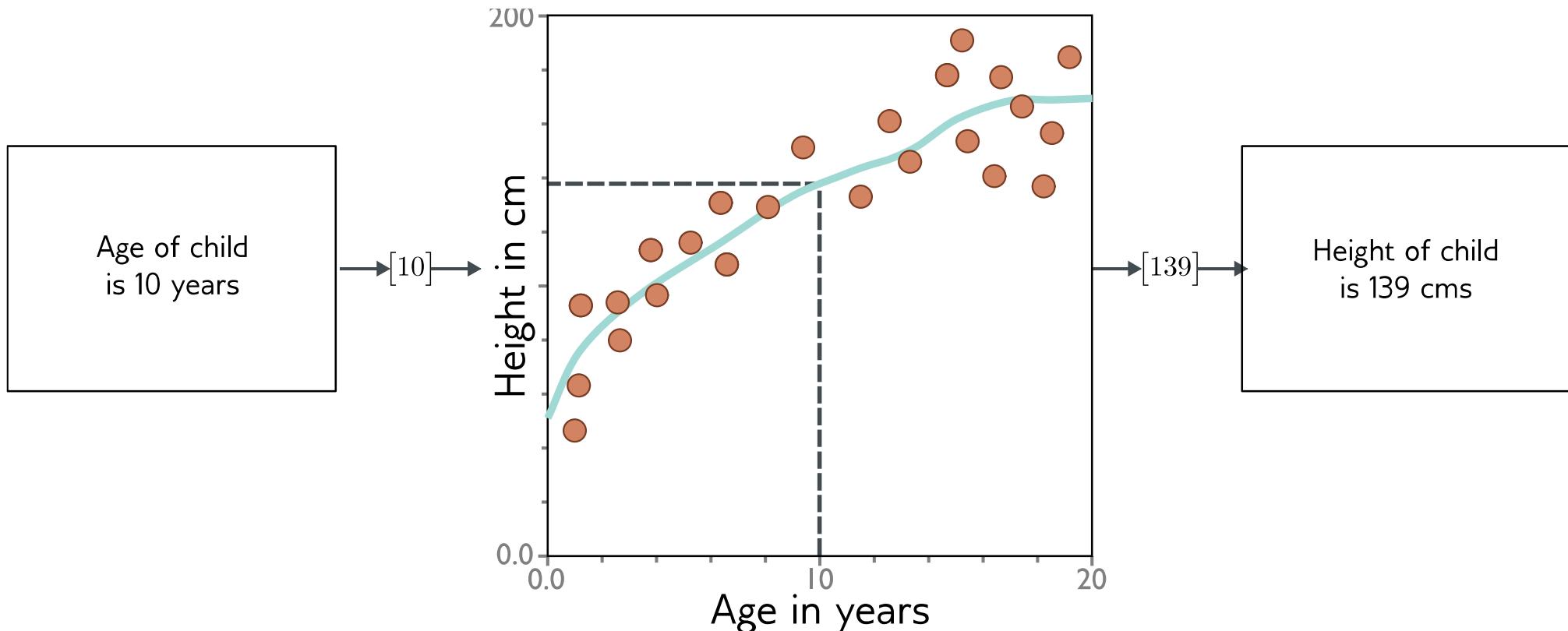
- Until now, we have treated the machine learning model as a black box that takes an input vector and returns an output vector
- But what exactly is in this black box?

# To predict the height of a child from their age



- The machine learning model is a mathematical equation that describes how the average height (input) varies as a function of age (output)
- The model represents a family of equations mapping the input to the output (i.e., a family of different cyan curves).

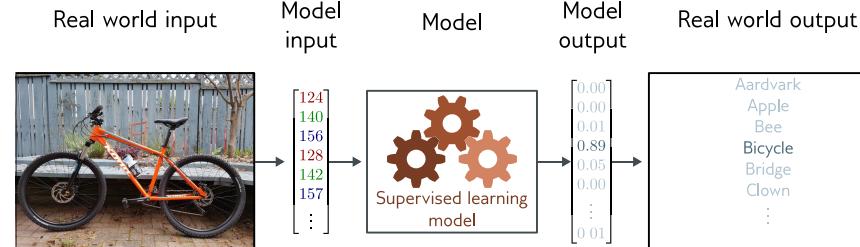
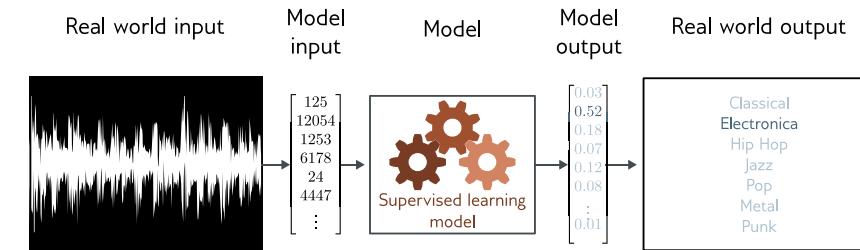
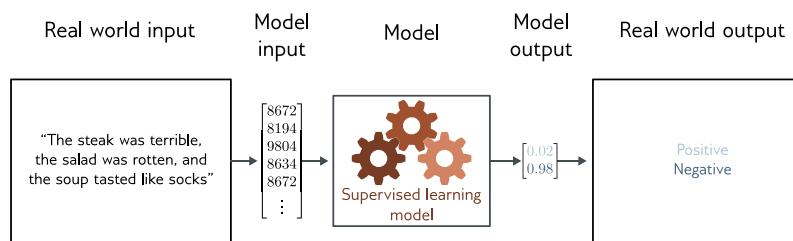
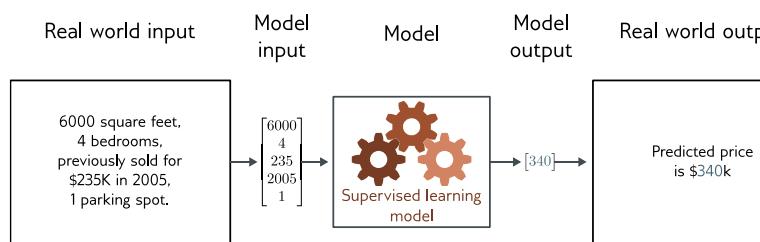
# To predict the height of a child from their age



- The particular equation (curve) is chosen using training data (examples of input/output pairs).
- These pairs are represented by the orange points, and we can see that the model (cyan line) describes these data reasonably.
- When we talk about **training** or **fitting** a model, we mean that we search through the family of **possible equations (possible cyan curves)** relating input to output to find the one that describes the training data most accurately.

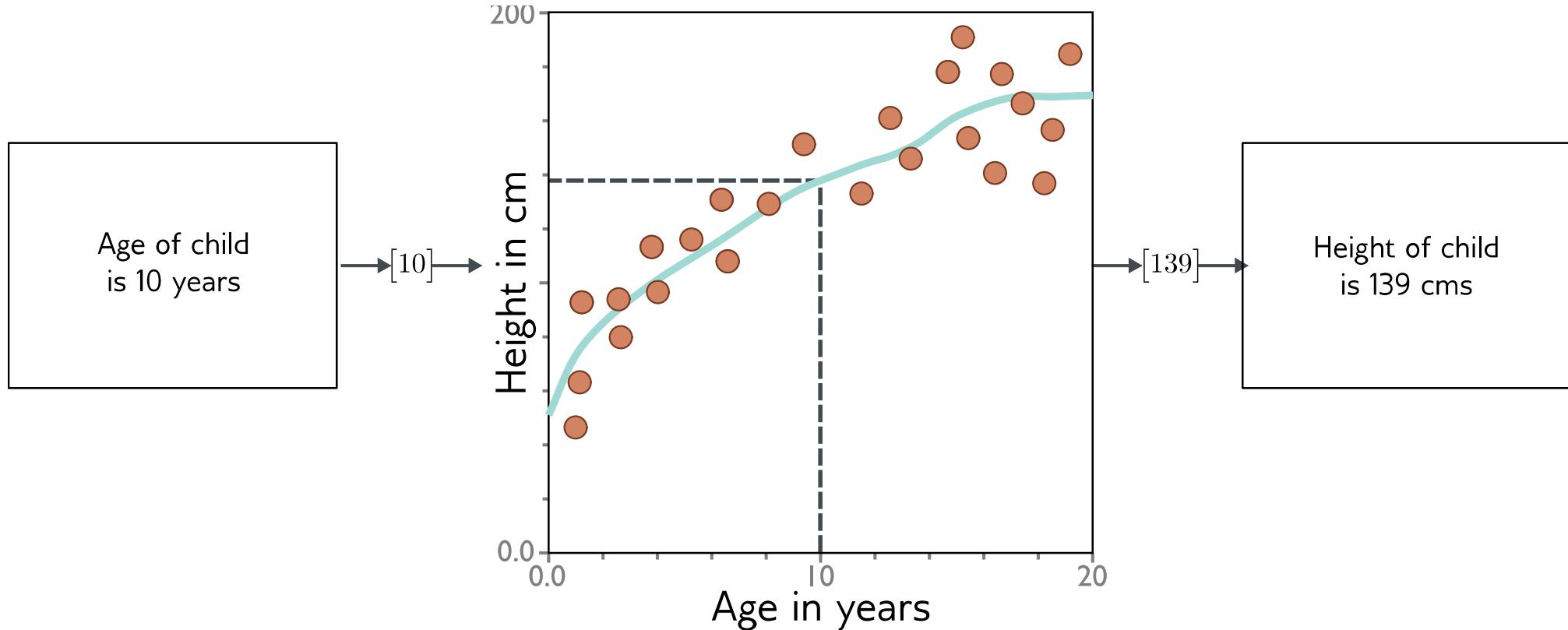
# What is a supervised learning model?

- It follows that the models in the previous figures require **labeled input/output pairs** for training.



- For example, the music classification model would require a large number of audio clips where a human expert had identified the genre of each.
- These input/output pairs take the role of a teacher or supervisor for the training process, and this gives rise to the term **supervised learning**.

# What is a supervised learning model in deep learning?

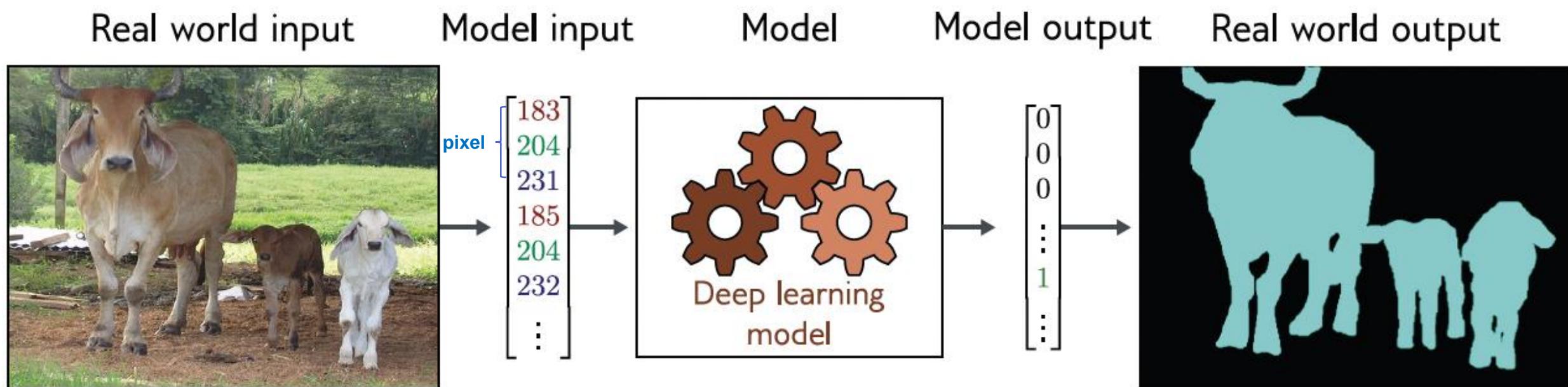


- Deep neural networks are just a very flexible family of equations
- Fitting deep neural networks = “Deep Learning”

# Deep neural networks

- Deep neural networks can process **inputs** that are **very large, of variable length**, and **contain various kinds of internal structures**.
- They can **output**
  - single real numbers (regression),
  - multiple numbers (multivariate regression), or
  - probabilities over two or more classes  
(binary and multiclass classification, respectively).
- Let's discuss the following **structured outputs**.

Image segmentation: every pixel of an input image is assigned a binary label that indicates whether it belongs to a cow or the background



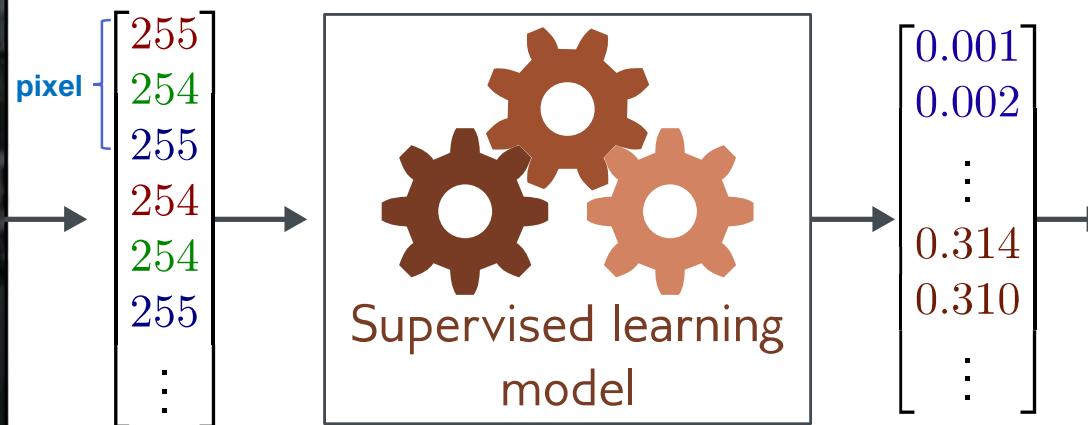
- Multivariate binary classification problem (many outputs, two discrete classes)
- Convolutional encoder-decoder network

Depth estimation: the input is an image of a street scene and the output is the depth at each pixel (distance from camera)

Real world input



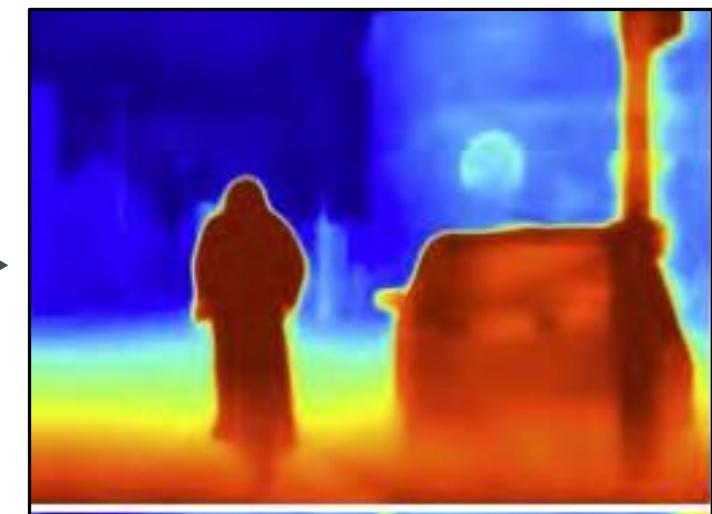
Model  
input



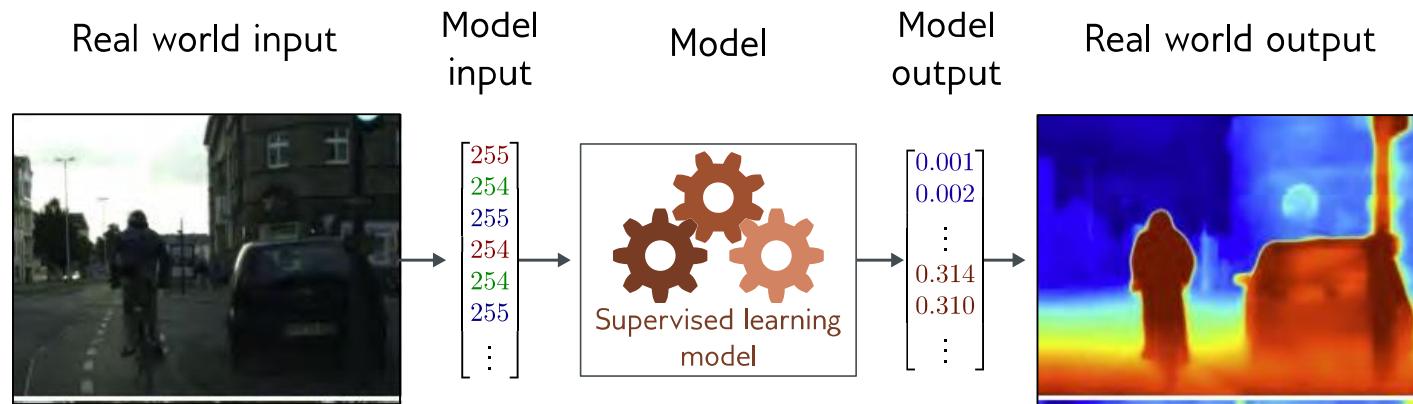
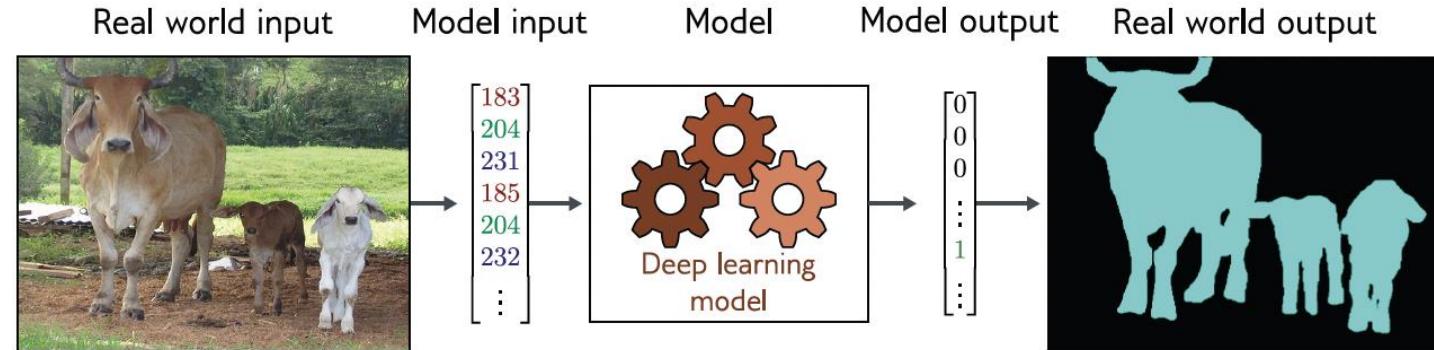
Model

Model  
output

Real world output

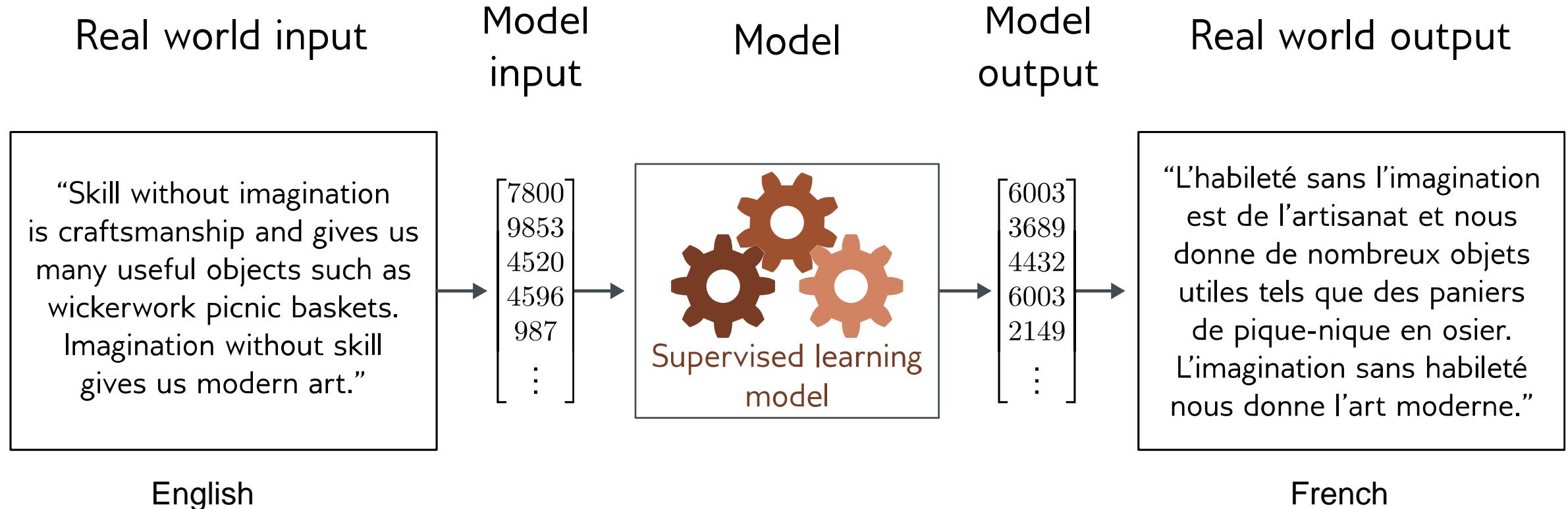


- Multivariate regression problem (many outputs, continuous)
- Convolutional encoder-decoder network

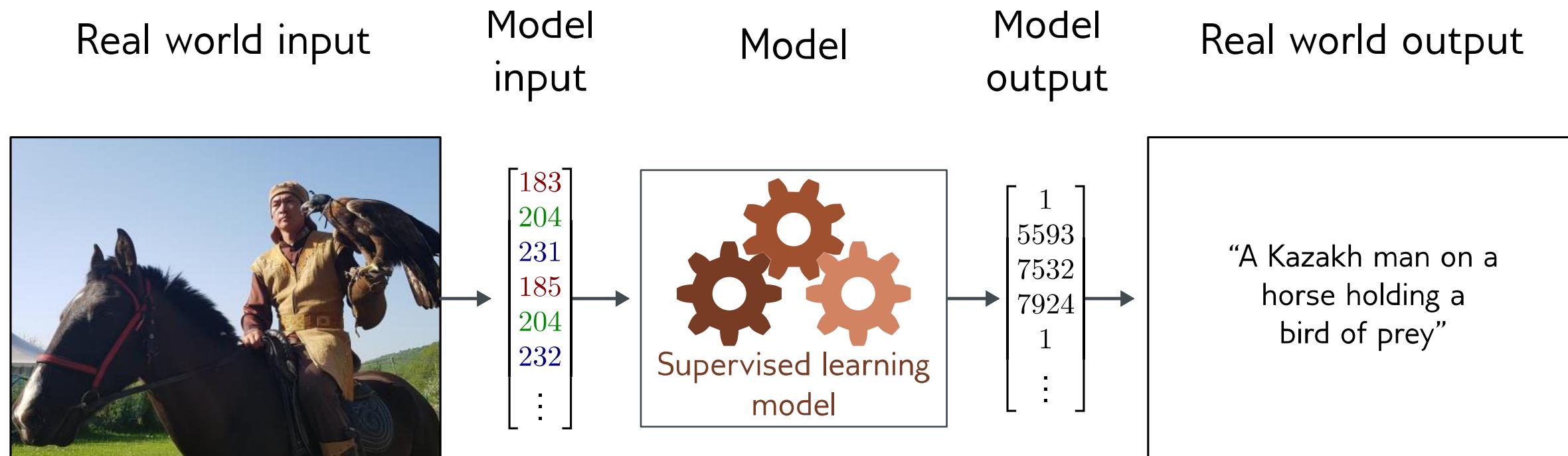


- In both cases, the output is high-dimensional and structured.
- This structure is closely tied to the input, and this can be exploited.
- If a pixel is labeled as “cow,” then a neighbor with a similar RGB value probably has the same label.

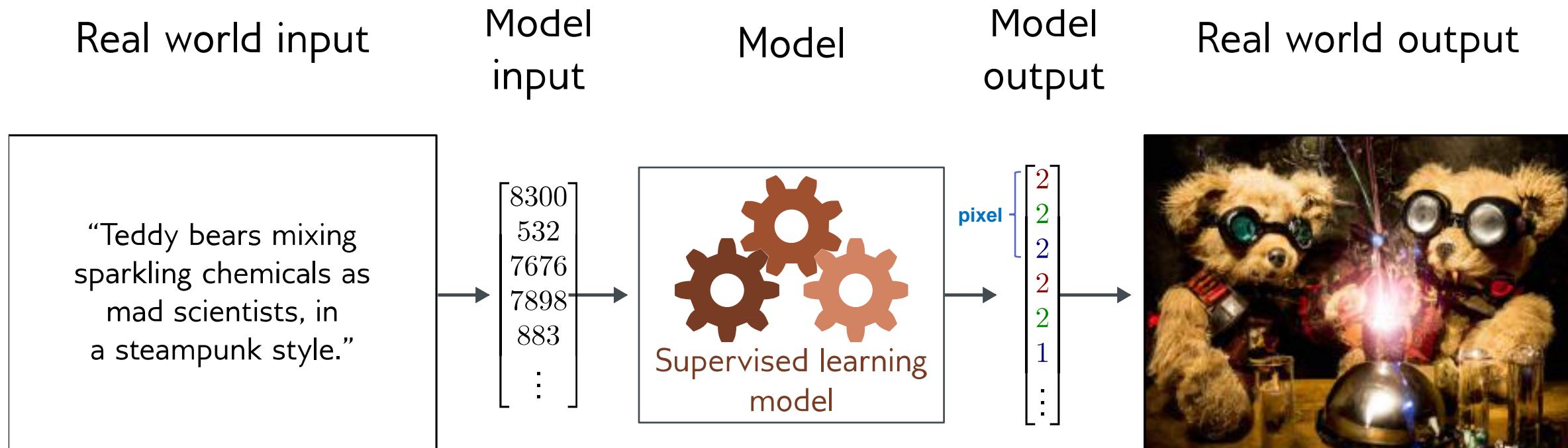
# Translation



# Image captioning



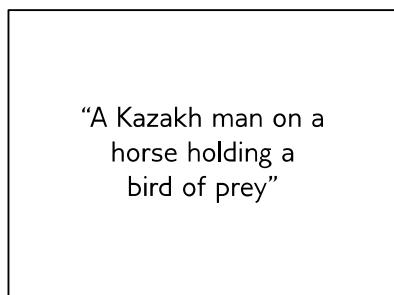
# Image generation from text: image synthesis model maps a caption to an image



<https://openai.com/dall-e-2/>

# What do these examples have in common?

- In principle, the latter three tasks can be tackled in the standard supervised learning framework, but they are more difficult for reasons.
  - the output may genuinely be ambiguous
  - the output contains considerable structure
- Very complex relationship between input and output
- But outputs (and sometimes inputs) obey rules



Language obeys grammatical rules



Natural images also have “rules”

# Idea

- Learn the “grammar” of the data from unlabeled examples
- Can use a gargantuan amount of data to do this (as unlabeled)
- Make the supervised learning task easier by having a lot of knowledge of possible outputs

# Artificial intelligence

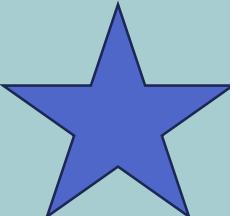
Machine learning

Supervised  
learning

Unsupervised  
learning

Reinforcement  
learning

Deep learning

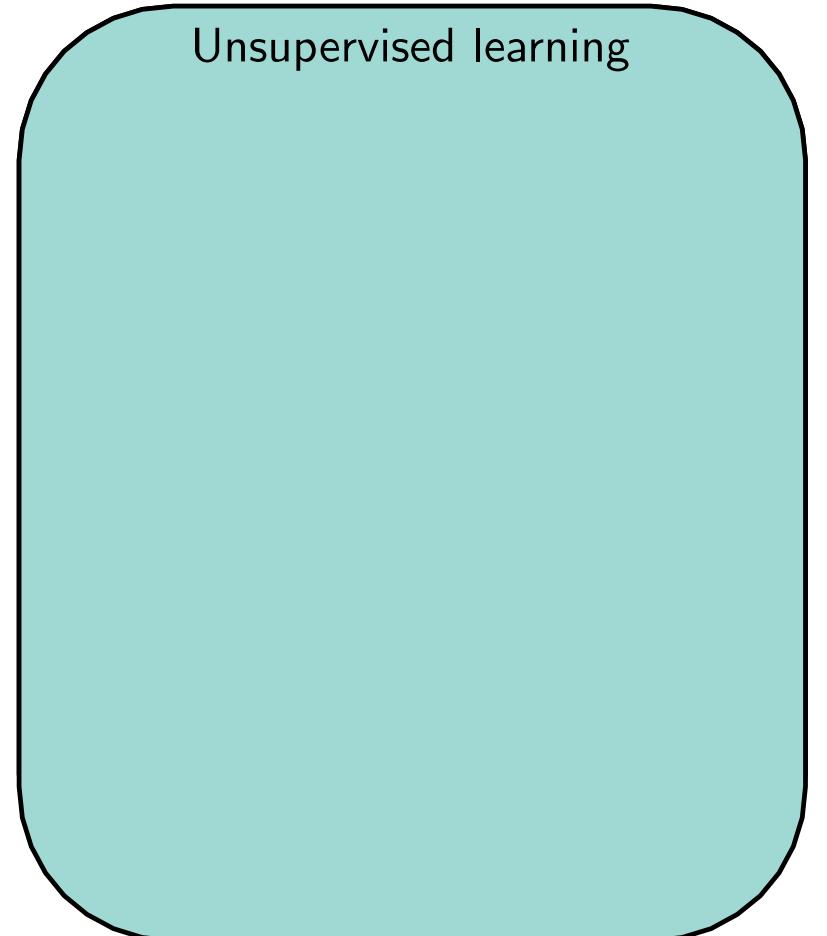


# Unsupervised Learning

- Constructing a model from input data without corresponding output labels is termed unsupervised learning;
- The absence of output labels means there can be no “supervision.”
- Rather than learning a mapping from input to output, **the goal is to describe or understand the structure of the data.**
- As was the case for supervised learning, the data may have very different characteristics; it may be discrete or continuous, low-dimensional or high-dimensional, and of constant or variable length.

# Unsupervised Learning

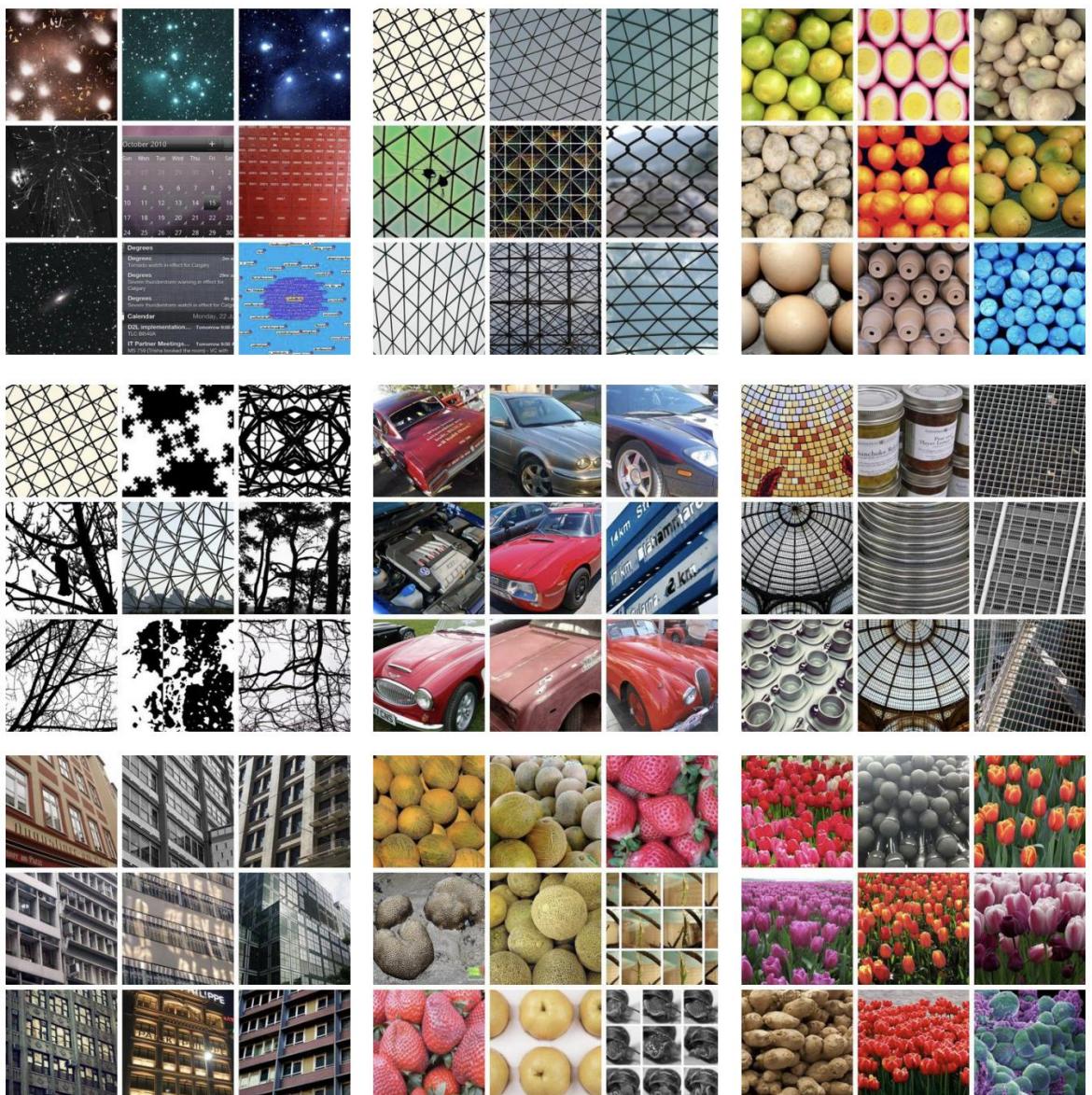
- Learning about a dataset without labels
  - Clustering
  - Finding outliers
  - Generating new examples (this book focuses)
  - Filling in missing data



Unsupervised learning



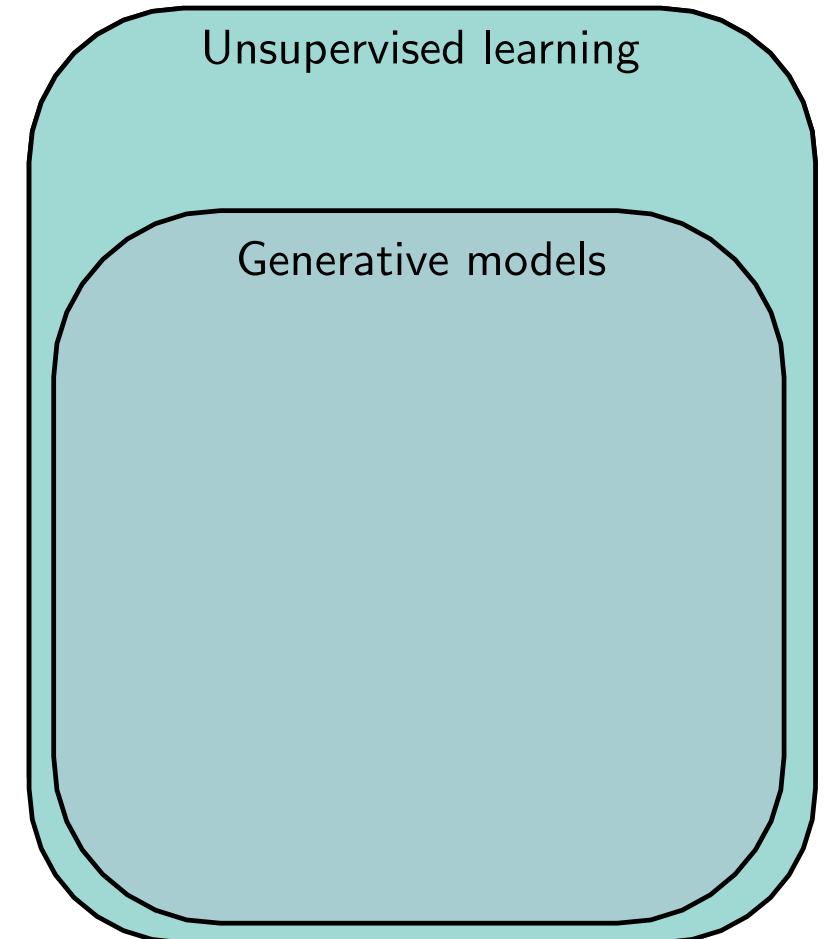
DeepCluster: Deep Clustering for Unsupervised Learning of Visual Features (Caron et al., 2018)



DeepCluster: Deep Clustering for Unsupervised Learning of Visual Features (Caron et al., 2018)

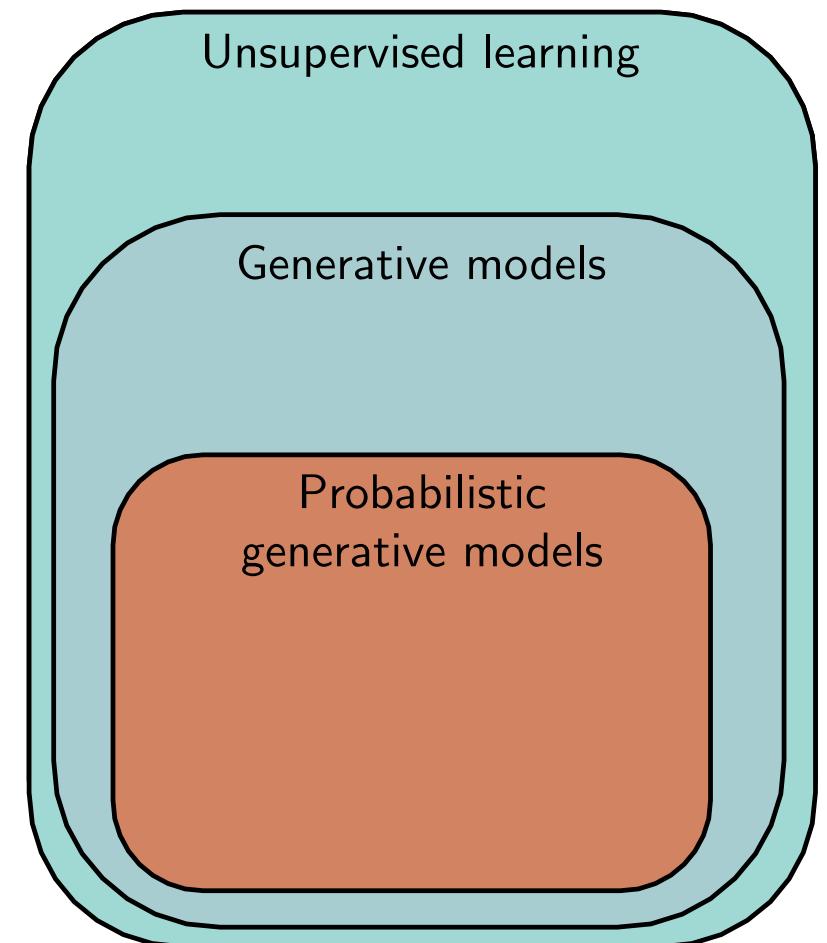
# Unsupervised Learning

- Learning about a dataset without labels
  - e.g., clustering
- Generative models can create examples
  - e.g., generative adversarial networks



# Unsupervised Learning

- Learning about a dataset without labels
  - e.g., clustering
- Generative models can create examples
  - e.g., generative adversarial networks (GAN)
- PGMs learn distribution over data
  - e.g., variational autoencoders (VAE)
  - e.g., diffusion models



# Generative models

- Learn to synthesize new data examples that are statistically indistinguishable from the training data.
- Some generative models explicitly describe the probability distribution over the input data and here **new examples are generated by sampling from this distribution**, e.g. VAE.
- Others merely learn a mechanism to generate new examples **without** explicitly describing their distribution, e.g. GAN
- State-of-the-art generative models can synthesize examples that **are extremely plausible but distinct from the training examples**, e.g. diffusion models

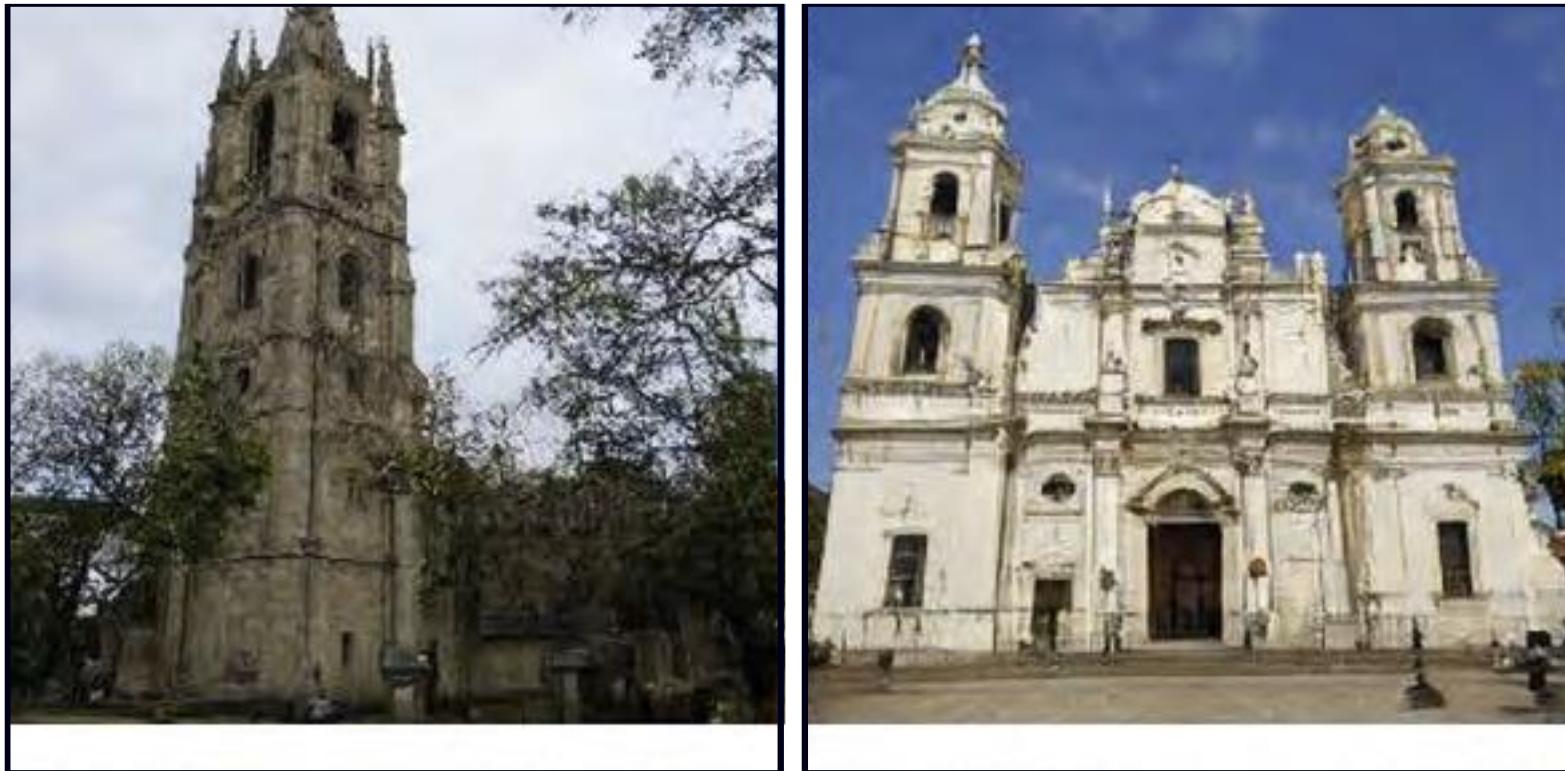
# Generative model: cat

- Generated from a model trained on pictures of cats.
- These are not real cats, but samples from a probability model.



# Generative models

- Two images generated from a model trained on images of buildings



Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020b). Analyzing and improving the image quality of StyleGAN. IEEE/CVF Computer Vision & Pattern Recognition, 8110–8119

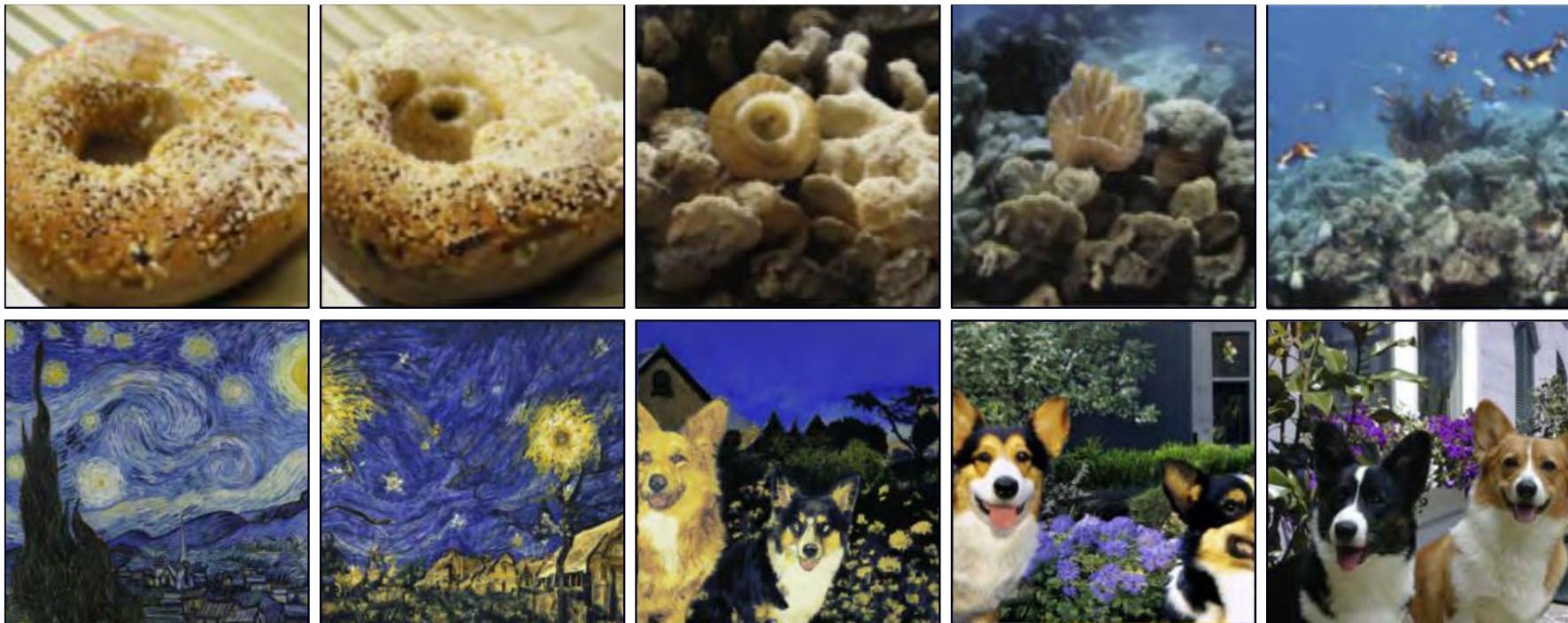
# Conditional synthesis



Inpainting. In the original image (left), the boy is obscured by metal cables. These undesirable regions (center) are removed and the generative model synthesizes a new image (right) under the constraint that the remaining pixels must stay the same.

Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., & Norouzi, M. (2022a). Palette: Image-to-image diffusion models. ACM SIGGRAPH

# Interpolation



In each row the left and right images are real and the three images in between represent a sequence of interpolations created by a generative model. The generative models that underpin these interpolations have learned that all images can be created by a set of underlying latent variables. By finding these variables for the two real images, interpolating their values, and then using these intermediate variables to create new images, we can generate intermediate results that are both visually plausible and mix the characteristics of the two original images.

# Artificial intelligence

Machine learning

Supervised  
learning

Unsupervised  
learning

Reinforcement  
learning

Deep learning



# Reinforcement learning

- A set of **states (input)**: observation from environment
- A set of **actions (output)**: model's action on environment
- A set of **rewards (goal)**
- Goal: take actions to change the state so that you receive rewards
- You don't receive any data – you have to explore the environment yourself to gather data as you go

# Example: chess

- States are valid states of the chess board (the current chess on the current position)
- Actions at a given time are valid possible moves
- Positive rewards for taking pieces, negative rewards for losing them

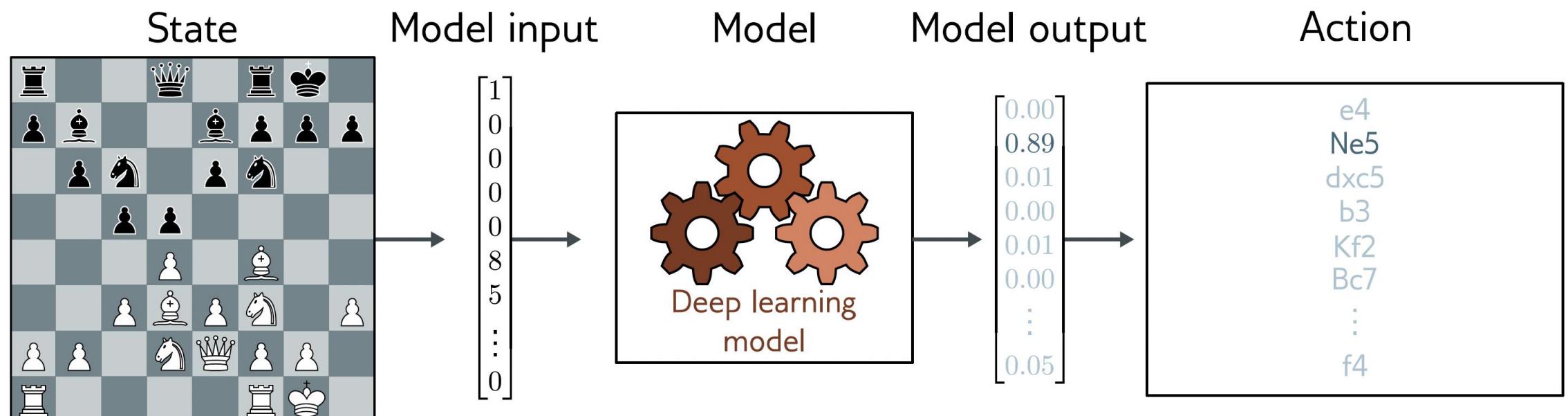


Action

e4  
Ne5  
dxc5  
b3  
Kf2  
Bc7  
:  
f4

# Example: chess

- States are valid states of the chess board
- Actions at a given time are valid possible moves
- Positive rewards for taking pieces, negative rewards for losing them



# Why is this difficult?

- Stochastic (隨機性)
  - Make the same move twice, the opponent might not do the same thing
  - Rewards also stochastic (opponent does or doesn't take your piece)
- Temporal credit assignment problem
  - Did we get the reward because of this move? Or because we made good tactical decisions somewhere in the past?
- Exploration-exploitation trade-off
  - If we found a good opening, should we use this?
  - Or should we try other things, hoping for something better?

# Framework

- PyTorch: <https://pytorch.org/>
- TensorFlow: <https://www.tensorflow.org/>
- Keras: <https://keras.io/>
- MXNet: <https://mxnet.apache.org/>
- Colab: <https://colab.research.google.com/>



TensorFlow



# Related Conferences

- Deep Learning (Theory) :
  - NeurIPS
  - ICLR
  - ICML
- Deep Learning (Application) :
  - AAAI
  - IJCAI
- Computer Vision:
  - CVPR
  - ICCV
  - ECCV
- Natural Language Processing:
  - ACL
  - NAACL
  - EMNLP
- Speech:
  - ICASSP
  - INTERSPEECH

## Scientist vs Engineer

Scientists use systematic methods  
to solve long-term problems

Engineers use existing methods to  
solve current problems

