



# Навчання з підкріпленням

Лекція 0: Вступ

Кочура Юрій Петрович  
[iuriy.kochura@gmail.com](mailto:iuriy.kochura@gmail.com)  
[@y\\_kochura](https://twitter.com/y_kochura)

# Сьогодні

- Огляд основ машинного навчання
- Вступ до навчання з підкріпленням

# Огляд основ машинного навчання

Що таке машинне навчання?

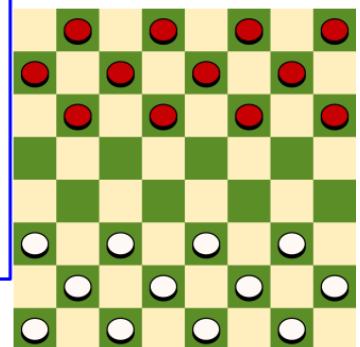
# Визначення за Артур Семюель

Артур Семюель (1959): Машинне навчання - це область навчання, яка надає комп'ютеру можливість вчитися не будучи явно запrogramованим.



A. L. Samuel\*

**Some Studies in Machine Learning  
Using the Game of Checkers. II—Recent Progress**



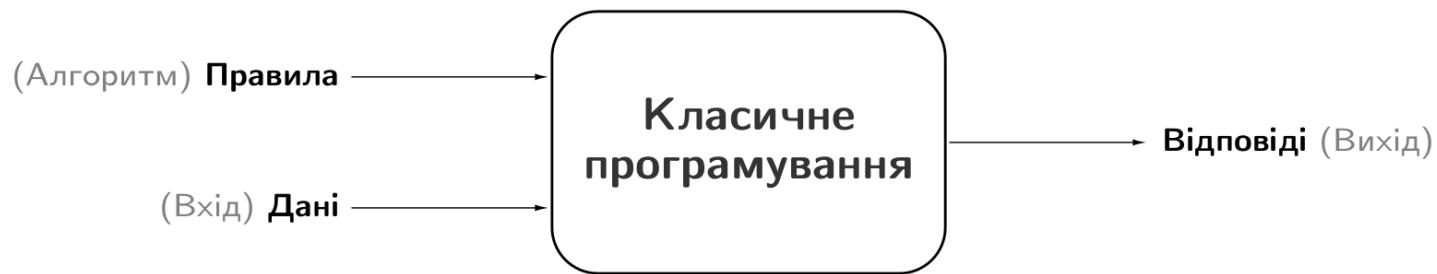
# Визначення за Том Мітчелл

Том Мітчелл (1998): Комп'ютерна програма, яка учається з досвіду  $E$  по відношенню до деякого класу задач  $T$  та міри продуктивності  $P$  називається машинним навчанням, якщо її продуктивність у задачах з  $T$ , що вимірюється за допомогою  $P$ , покращується з досвідом  $E$ .



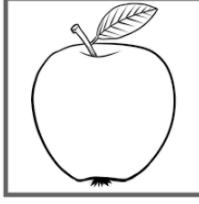
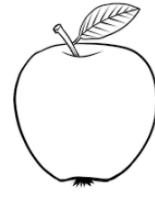
- Досвід (дані): ігри в які грає програма сама з собою
- Вимір продуктивності: коефіцієнт виграшу

# Класичне програмування vs машинне навчання



# Типи навчання

За характером навчальних даних (**досвіду**) машинне навчання поділяють на чотири типи: контролюване (з учителем), напівконтрольоване, неконтрольоване (без учителя) та з підкріплленням.

Контрольоване навчання <b>Supervised learning</b>	Напівконтрольоване навчання <b>Semi-supervised learning</b>	Неконтрольоване навчання <b>Unsupervised learning</b>	Навчання з підкріпллення <b>Reinforcement learning</b>
<b>Дані:</b> $(x, y)$ $x$ – приклад, $y$ – мітка	<b>Дані:</b> $(x, y)$ та $x,  (x, y)  <  x $ $x$ – приклад, $y$ – мітка	<b>Дані:</b> $x$ $x$ – приклад, немає міток!	<b>Дані:</b> пари стан-дія
Мета – знайти функцію відображення $x \rightarrow y$	Мета – знайти функцію відображення або категорію $x \rightarrow y$	Мета – знайти правильну категорію.	Мета – максимізація загальної винагороди, отриманої агентом при взаємодії з навколошнім середовищем.
<b>Приклад</b>  Це є яблуко.	<b>Приклад</b>  Це є яблуко.	<b>Приклад</b>  Цей об'єкт схожий на інший.	<b>Приклад</b>  Їжте це, бо це зробить вас сильнішим.

# Як вчиться людина?

- Ми та інші розумні істоти, вчимось завдяки **взаємодії** із своїм оточенням
- Взаємодії часто бувають **послідовними** - майбутні взаємодії можуть залежати від попередніх
- Ми направлені на **результат**
- Ми можемо вчитися **не маючи прикладів** оптимальної поведінки

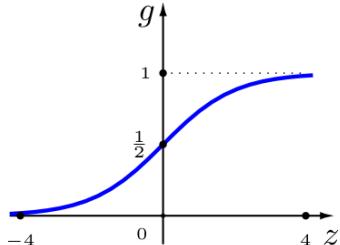
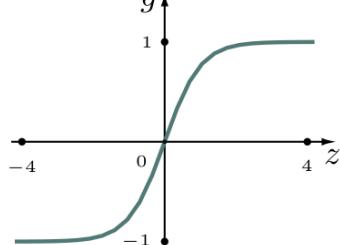
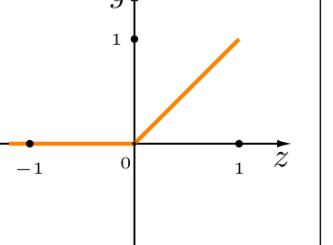
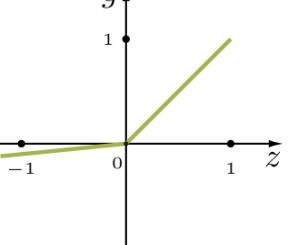
# Мозок людини

Базовою обчислювальною одиницею мозку є нейрон. Мозок дорослої людини складається з 86 мільярдів нейронів, які з'єднані між собою приблизно  $10^{14}$  -  $10^{15}$  синапсами.

# Біологічний та штучний нейрон

Біологічний нейрон	Штучний нейрон
<p>Дендрити Ядро Тіло клітини Аксон Відростки аксону Термінали аксону Імпульс спрямований до тіла клітини Імпульс спрямований від тіла клітини</p>	<p>Імпульс від аксону (вхід) <math>x_0</math>   Синапс Дендрит <math>\omega_0</math>   <math>b = \omega_0 x_0</math> <math>x_0</math>   <math>x_1</math>   <math>x_2</math>   <math>\vdots</math>   <math>x_m</math> <math>\omega_1</math>   <math>\omega_2</math>   <math>\omega_m</math> <math>\omega_1 x_1</math>   <math>\omega_2 x_2</math>   <math>\omega_m x_m</math> <math>z = \sum_{n=1}^m \omega_n x_n + b</math> Тіло клітини Активаційна функція <math>g</math>   <math>g(z)</math> Імпульс на аксоні (вихід)</p>

# Деякі функції активації

Sigmoid	Tanh	ReLU	Leaky ReLU
$g(z) = \frac{1}{1 + e^{-z}}$	$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$g(z) = \max(0, z)$	$g(z) = \max(\epsilon z, z)$ $\epsilon \ll 1$
			

Людина добре сприймати  
візуальну інформацію



Що Ви бачите?



Собака-вівця чи швабра?

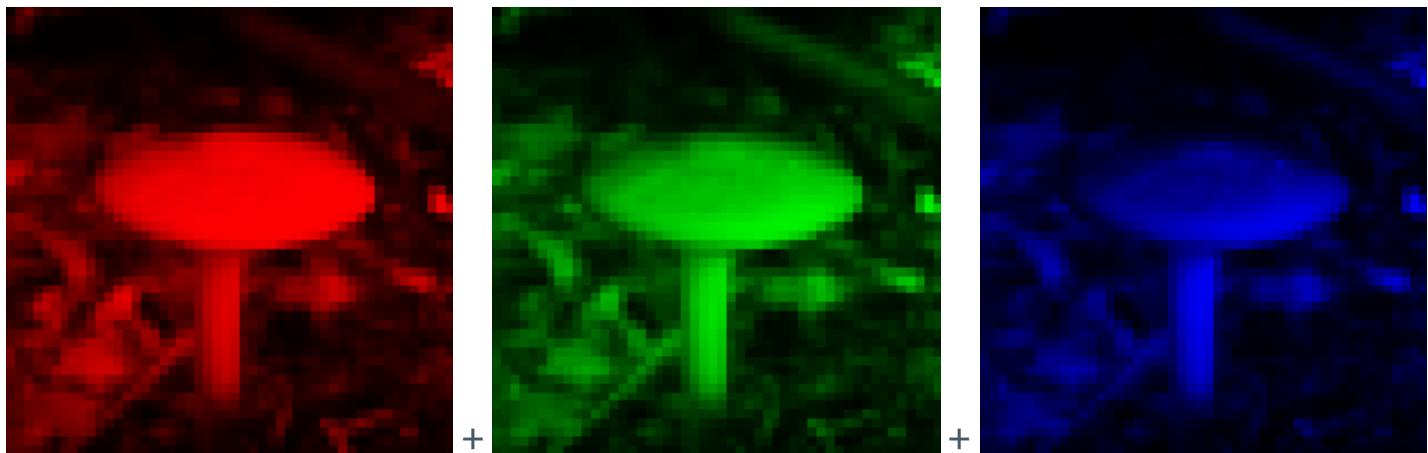
Людський мозок настільки добре інтерпретує візуальну інформацію, що **роздріб** між зображенням та його семантичною інтерпретацією (пікселями) важко оцінити інтуїтивно:



Це мухомор.



Це мухомор.



Це мухомор.

```
array([[[0.03921569, 0.03529412, 0.02352941, 1.          ],
       [0.2509804 , 0.1882353 , 0.20392157, 1.          ],
       [0.4117647 , 0.34117648, 0.37254903, 1.          ],
       ...,
       [0.20392157, 0.23529412, 0.17254902, 1.          ],
       [0.16470589, 0.18039216, 0.12156863, 1.          ],
       [0.18039216, 0.18039216, 0.14117648, 1.          ]],

      [[0.1254902 , 0.11372549, 0.09411765, 1.          ],
       [0.2901961 , 0.2509804 , 0.24705882, 1.          ],
       [0.21176471, 0.2        , 0.20392157, 1.          ],
       ...,
       [0.1764706 , 0.24705882, 0.12156863, 1.          ],
       [0.10980392, 0.15686275, 0.07843138, 1.          ],
       [0.16470589, 0.20784314, 0.11764706, 1.          ]],

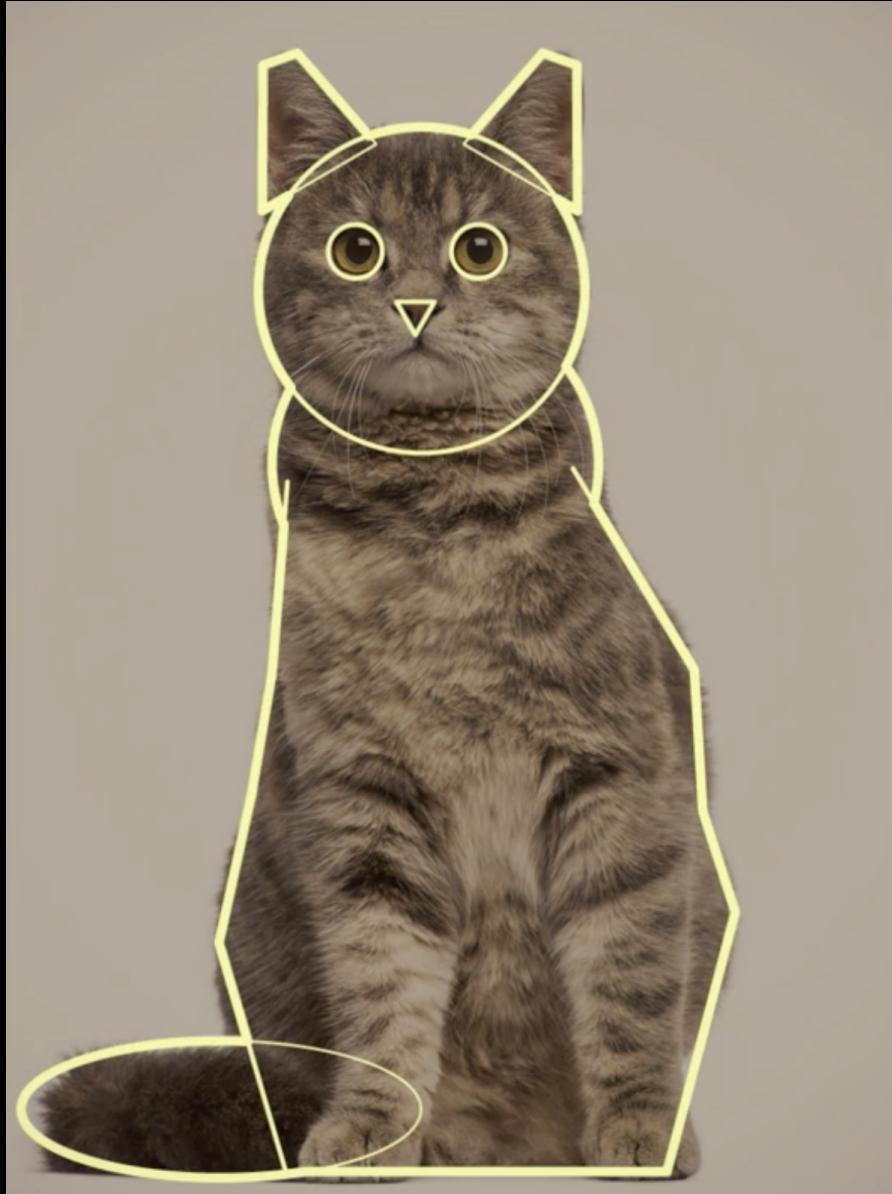
      [[0.14117648, 0.12941177, 0.10980392, 1.          ],
       [0.21176471, 0.1882353 , 0.16862746, 1.          ],
       [0.14117648, 0.13725491, 0.12941177, 1.          ],
       ...,
       [0.10980392, 0.15686275, 0.08627451, 1.          ],
       [0.0627451 , 0.08235294, 0.05098039, 1.          ],
       [0.14117648, 0.2        , 0.09803922, 1.          ]],

      ...,
```

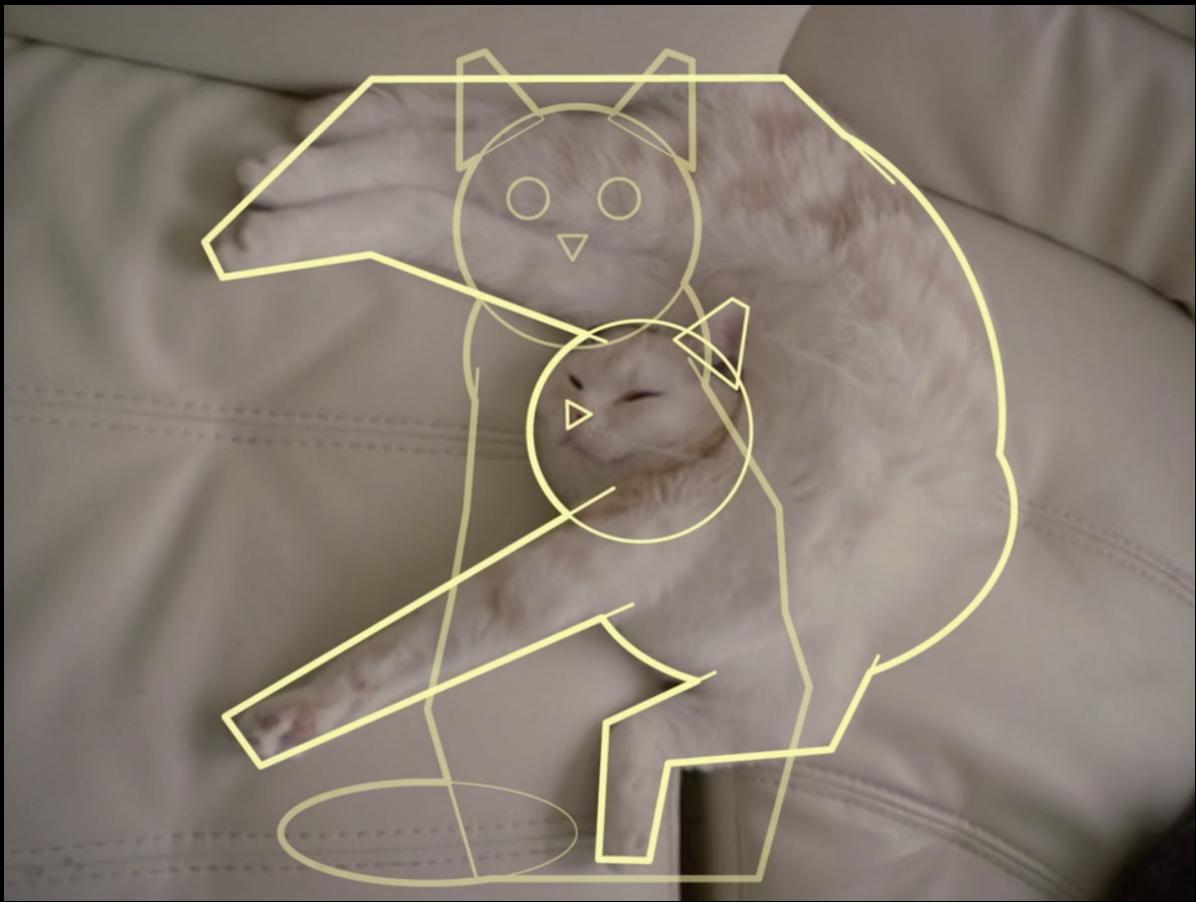
Це мухомор.

Як навчити машин бачити?



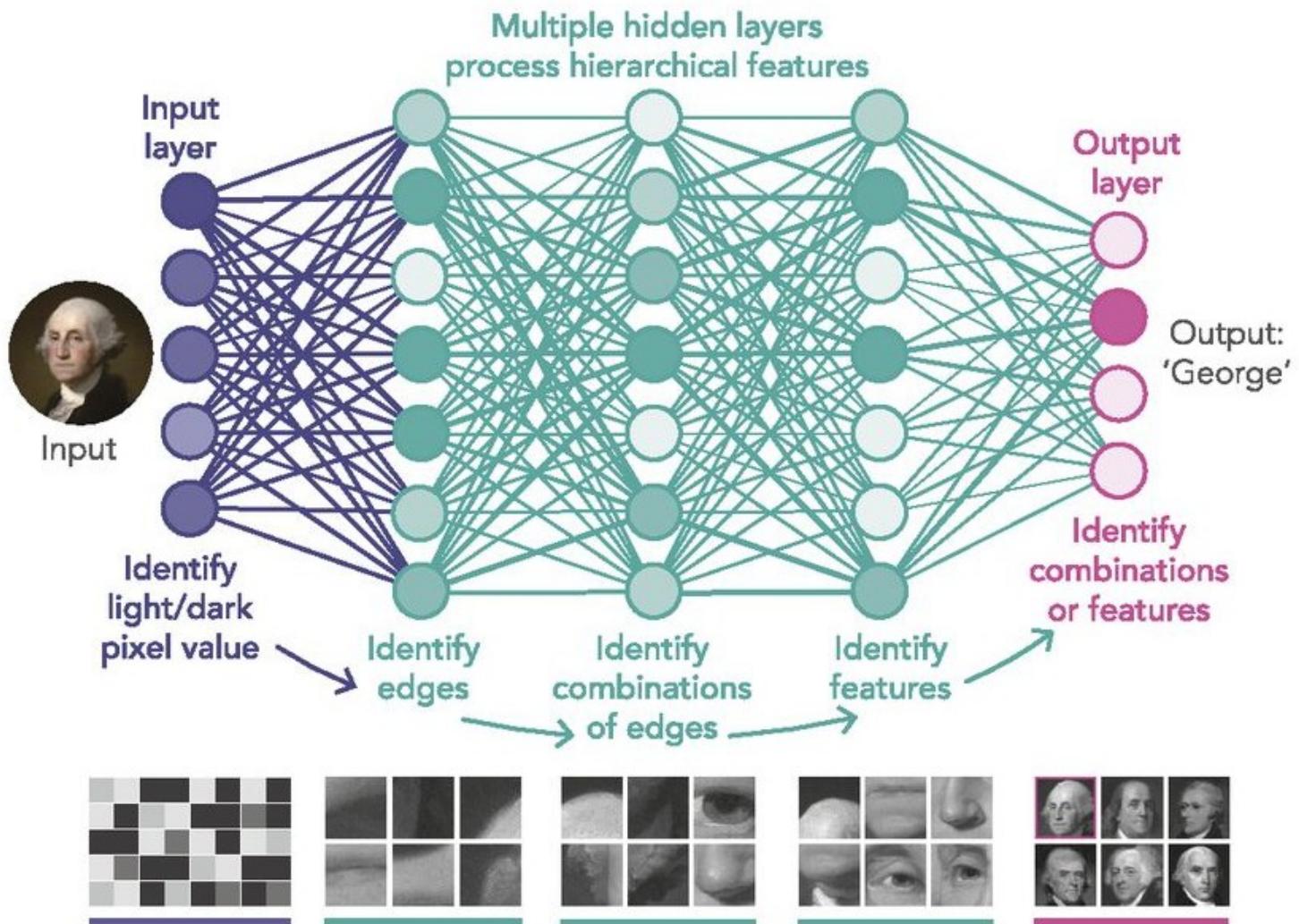






Для пошуку шаблону в даних (витягування семантичної інформації, ознак) потрібна побудова **складних моделей**, які б отримати вручну було б дуже складно.

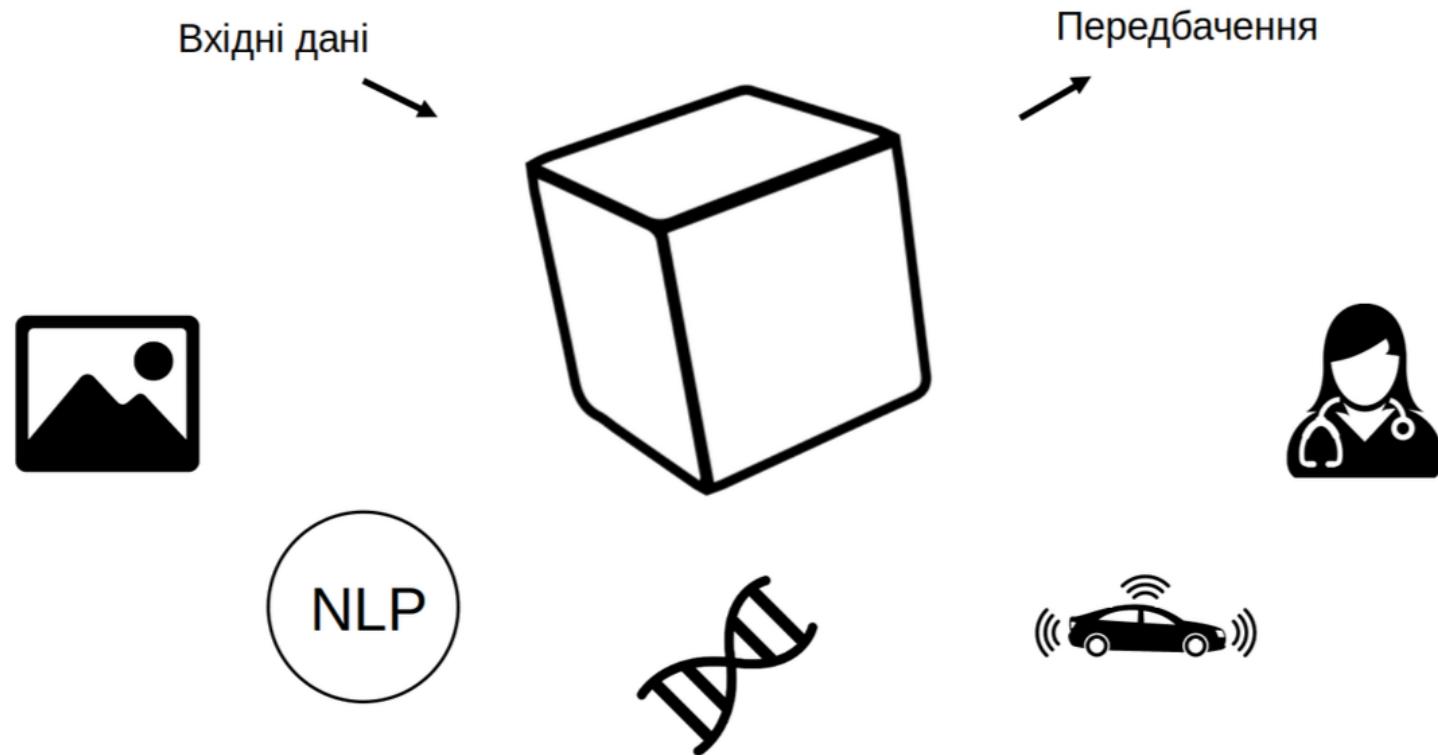
Однак, можна написати програму, яка буде **вчитись** знаходити шаблон в даних самостійно.



# Що входить до задачі машинного навчання?

- Постановка проблеми + дані
- Навчання моделі
- Визначення функції втрат
- Вибір алгоритму оптимізації

# Які дані використовуються?



# Ознаки у машинному навчанні

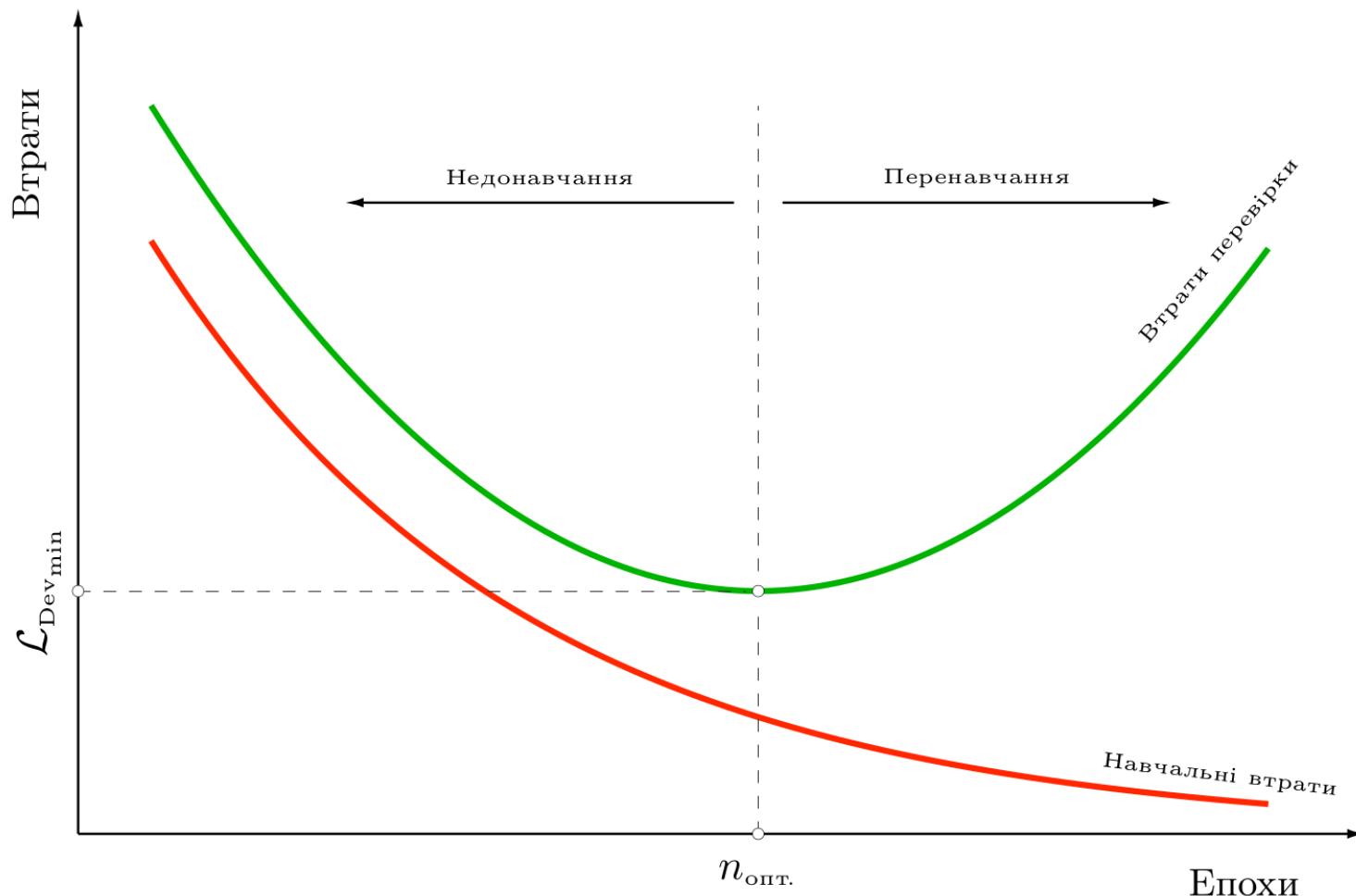
Ознаки - це спостереження, які використовуються для прийняття рішень моделлю.

- Для класифікації зображень  **кожен** піксель є ознакою
- Для розпізнавання голосу, **частота** та **гучність** є ознаками
- Для безпілотних автомобілів дані з **камер, радарів і GPS** є ознаками

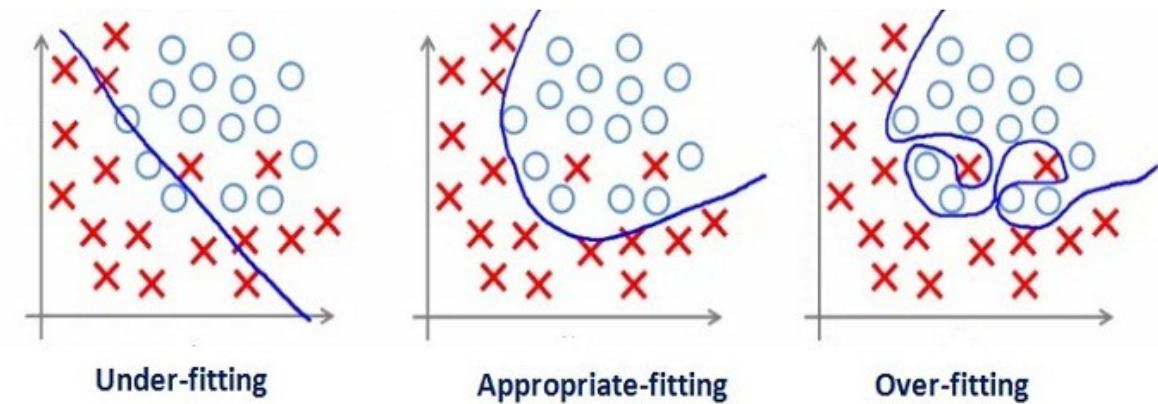
# Типи ознак у робототехніці

- Пікселі (RGB дані)
- Глибина (сонар, лазерні далекоміри)
- Орієнтація або прискорення (гіроскоп, акселерометр, компас)

# Недонавчання vs перенавчання

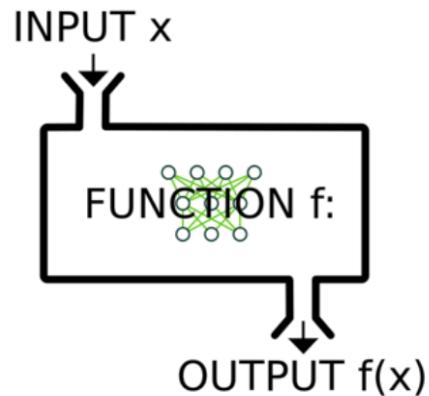


# Недонавчання vs перенавчання



# Що таке модель?

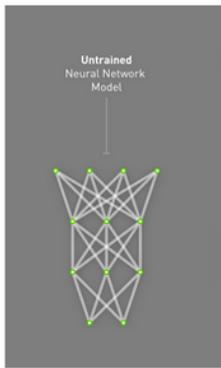
Хоча те, що знаходиться всередині глибинної нейронної мережі, може бути складним, за своєю суттю це просто функції. Вони беруть певні вхідні дані: **INPUT x** і генерують деякі вихідні дані: **OUTPUT f(x)**



# З чого складається модель?

## Компоненти моделі

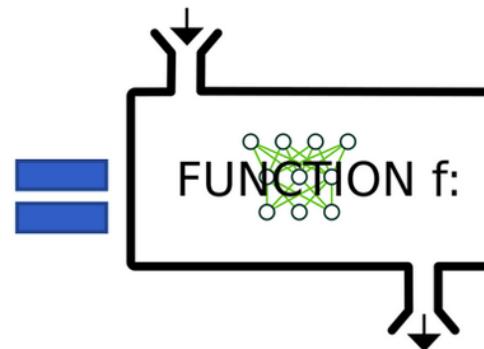
Архітектура мережі = [deploy.prototxt](#)



Навченні ваги = \*\*\*.caffemodel



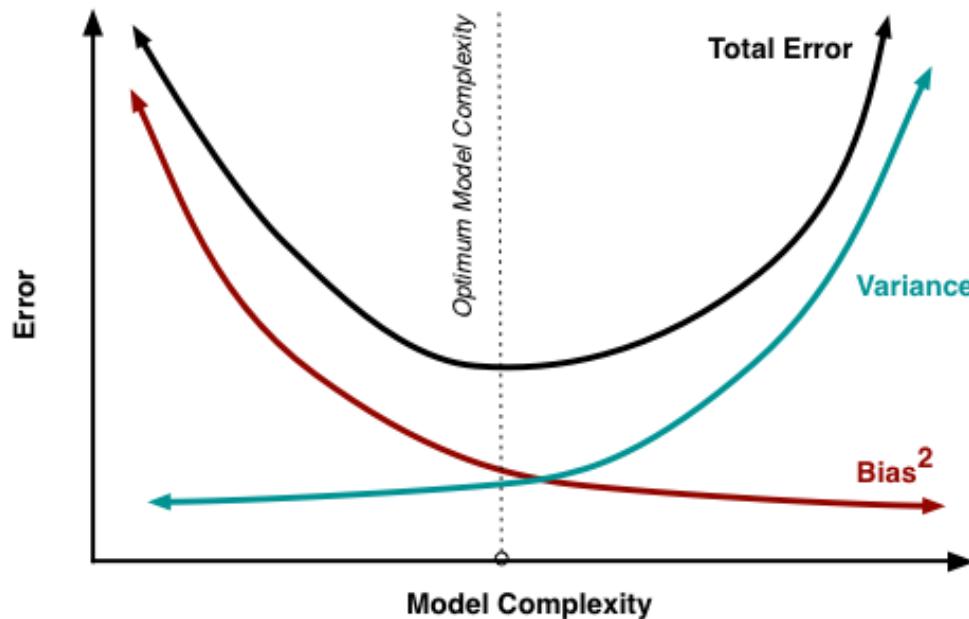
Модель



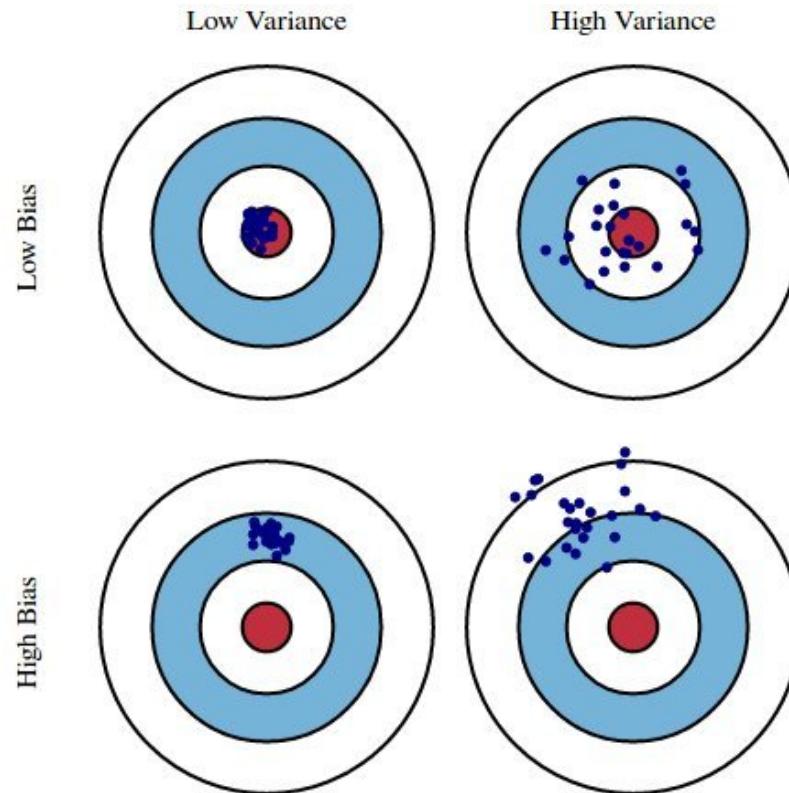
# Джерела помилок моделі

- Зсув (Bias)
- Розкид (Variance)
- Шум (Irreducible error)

$$Err = Bias^2 + Variance + Irreducible\ error$$



# Інтуїція



# **Applications and successes**



Detectron2: A PyTorch-based modular object detection ...



Copy link



Object detection, pose estimation, segmentation (2019)



Google DeepMind's Deep Q-learning playing ...



Watch later



Share



Reinforcement learning (Mnih et al, 2014)



## AlphaStar Agent Visualisation

Watch later Share



Strategy games (Deepmind, 2016-2018)



NVIDIA Autonomous Car



Watch later



Share



Autonomous cars (NVIDIA, 2016)



Full Self-Driving



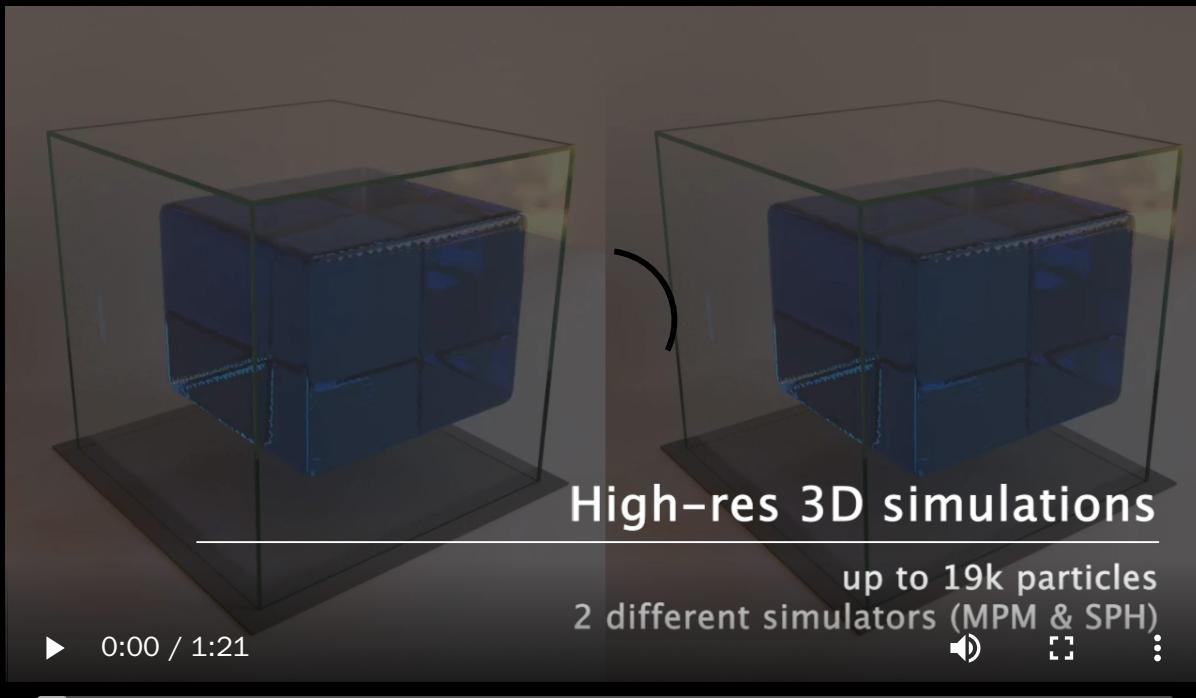
Watch later



Share



Autopilot (Tesla, 2019)



Physics simulation (Sanchez-Gonzalez et al, 2020)



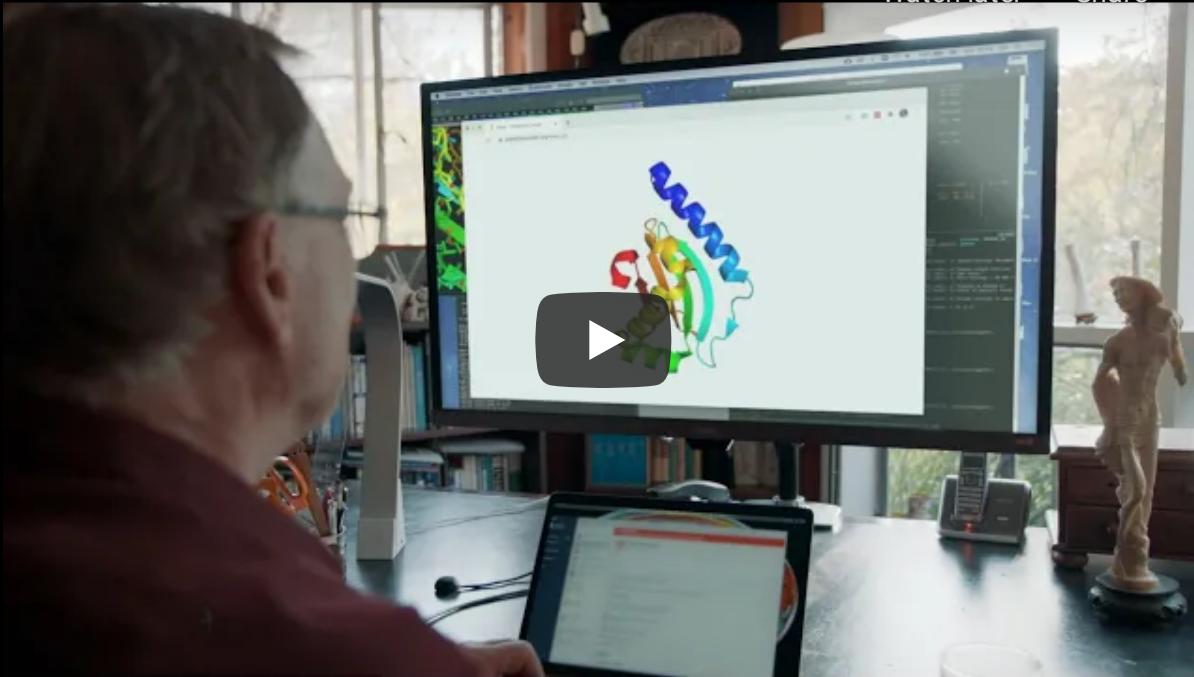
AlphaFold: The making of a scientific breakt...



Watch later



Share



AI for Science (Deepmind, AlphaFold, 2020)



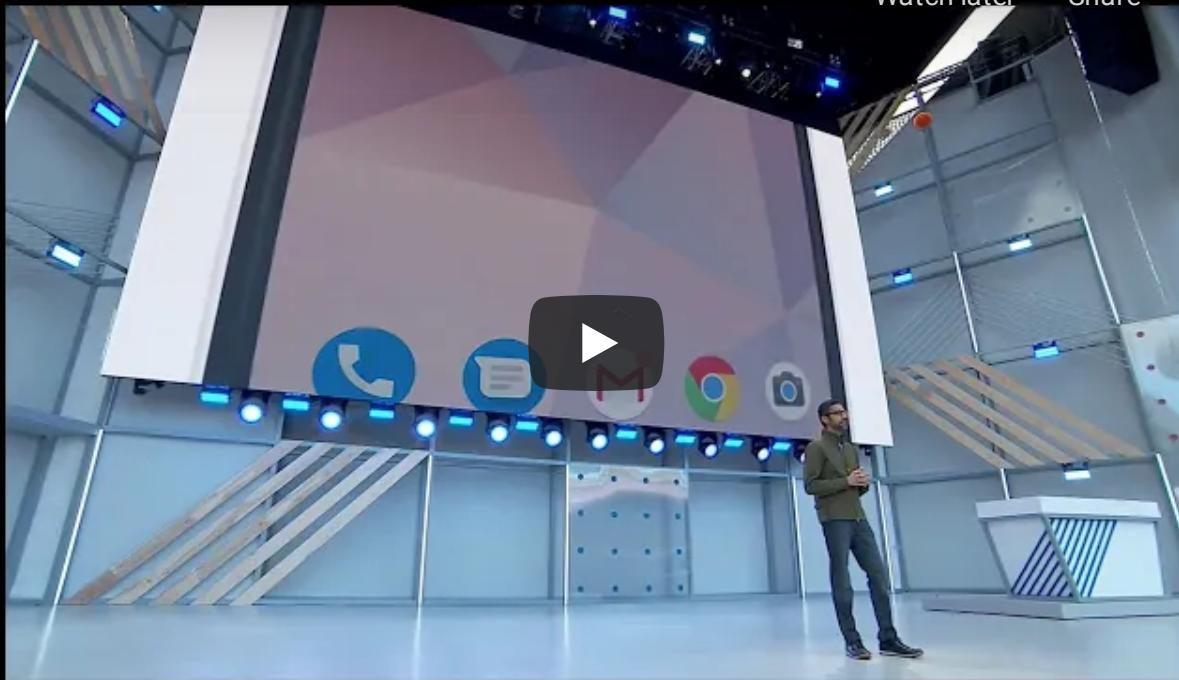
Google Assistant will soon be able to call res...



Watch later



Share



Speech synthesis and question answering (Google, 2018)



Artistic style transfer for videos



Watch later



Share

Sintel movie, III



Artistic style transfer (Ruder et al, 2016)

T

# A Style-Based Generator Architecture for Ge...



Watch later



Share

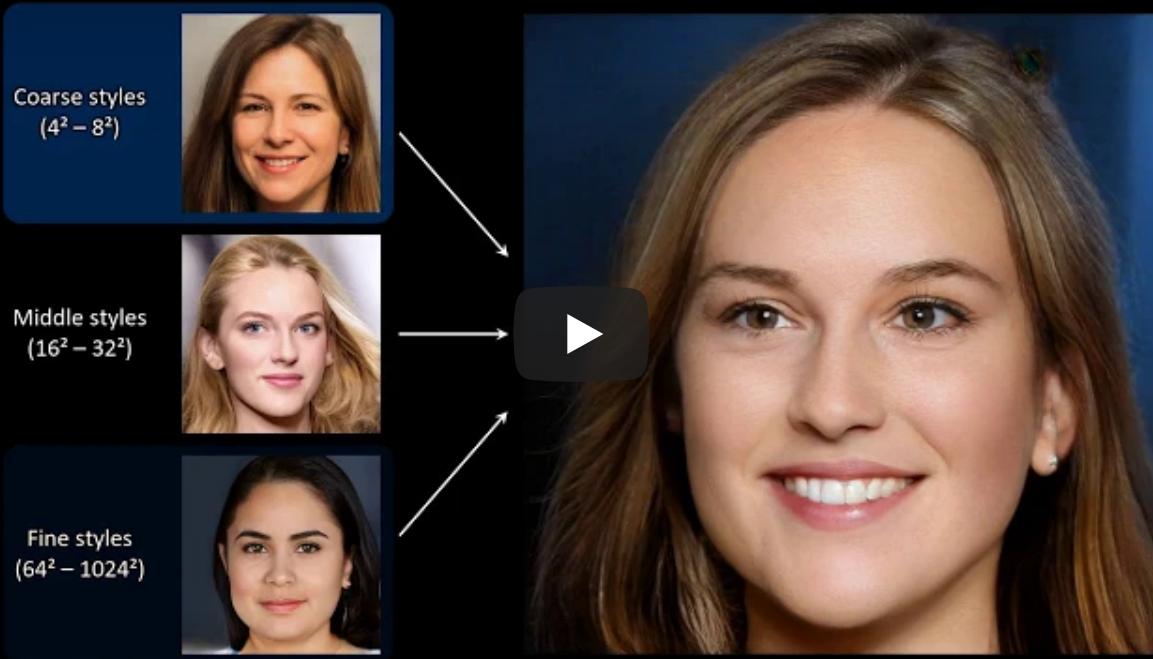


Image generation (Karras et al, 2018)



GTC Japan 2017 Part 9: AI Creates Original ...



Watch later



Share



Music composition (NVIDIA, 2017)



Behind the Scenes: Dali Lives



Watch later



Share



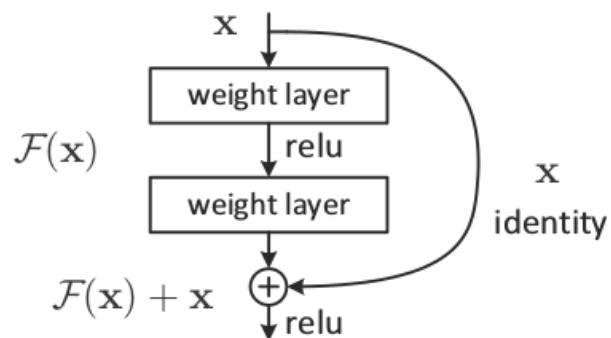
Dali Lives (2019)



Асоціацією обчислювальної техніки (ACM) нагороджено в 2018 році премією Тюрінга таких науковців: **Yann LeCun, Geoffrey Hinton, Yoshua Bengio** за концептуальні та інженерні прориви, які зробили в глибинних нейронних мережах.

# Чому DL працює?

Алгоритми (старі та нові)



Зростає кількість даних



Програмне забезпечення



Більш швидкі обчислювальні машини





*"For the last forty years we have programmed computers; for the next forty years we will train them."*

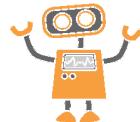
Chris Bishop, 2020.

# Вступ до навчання з підкріленням

Основним викликом штучного інтелекту та машинного навчання є прийняття правильних рішень в умовах **невизначеності**

# Визначення RL

Навчання з підкріленням (reinforcement learning, RL) - сімейство алгоритмів, які вивчають оптимальну стратегію, метою якої є максимізація загальної винагороди, отриманої агентом при взаємодії з навколошнім середовищем.



- Наприклад, кінцевою винагородою більшості ігор є перемога. Система навчання з підкрілення може стати експертом у складних іграх, шляхом оцінювання послідовності попередніх ігрових ходів, які в підсумку привели до перемоги або програшу.

# Визначення RL

RL - наука про те, як приймати рішення на основі взаємодій

- Це вимагає від нас задуматися над:
  - часом
  - (довгостроковими) наслідками спричинені діями
  - збором досвіду
  - передбаченням майбутнього
  - боротьбою з невизначеністю

# Застосування RL

- Ігри ([Atari](#), [AlphaGo](#))
- Робототехніка ([End-to-End Training](#))
- Фінанси
- Взаємодія людини з комп'ютером
- ...

# Причини використання RL

## 1. Пошук раніше невідомих рішень

- Приклад, програма, яка може грати в Go краще, ніж будь-яка людина, будь-коли

## 2. Пошук рішень в режимі реального часу за непередбачених обставин

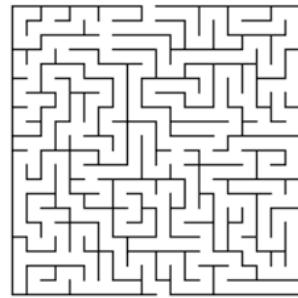
- Приклад, робот, який може орієнтуватися на місцевості, яка значно відрізняється від будь-якої очікуваної місцевості
- Алгоритми навчання з підкріпленням намагаються задовільнити обидва випадки
- Зауважте, що другий пункт стосується не (просто) узагальнення - це більшою мірою про ефективне навчання в режимі реального часу під час взаємодії з середовищем

# Агент (agent)



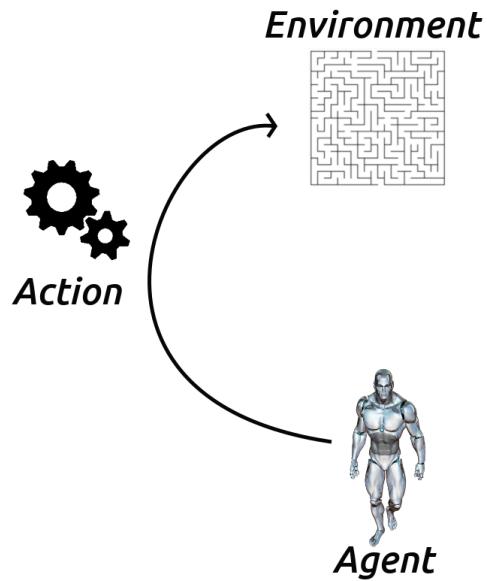
Агент (agent) - це те, що існує окремо від інших речей та використовує певну стратегію (policy) для максимізації очікуваної винагороди (reward), отриманої від переходу між станами середовища (environment).

# Середовище (environment)



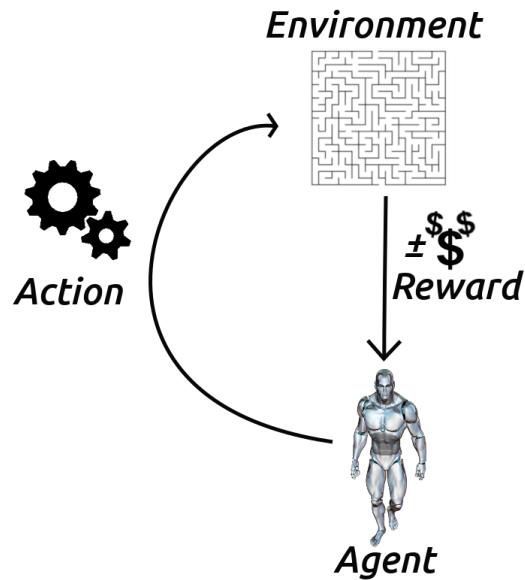
Середовище - це стохастичний та невизначений світ, в якому агент існує та діє. Дія (action)

# Дія (action)



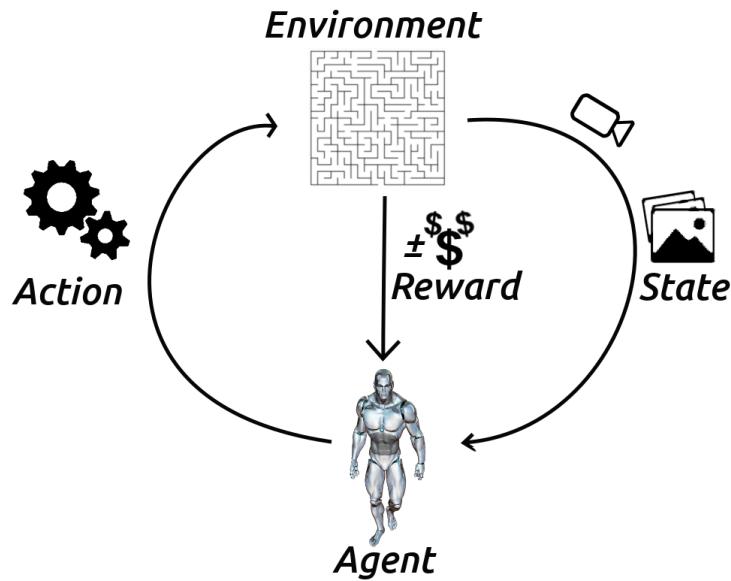
**Дія** - механізм, за допомогою якого агент переходить між дозволеними середовищем станами. Агент обирає дію, використовуючи стратегію.

# Винагорода (reward)



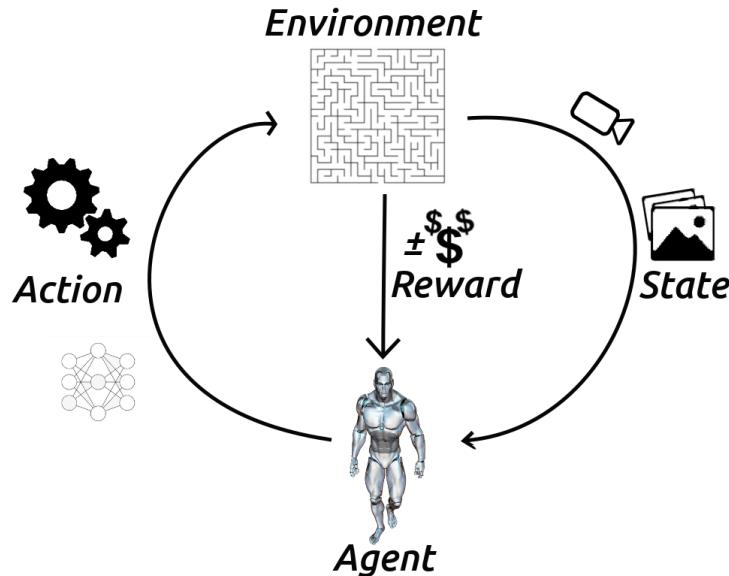
**Винагорода** - числовий результат, отриманий агентом у наслідок переходу між стананами, які визначені середовищем (дії).

# Цикл взаємодії



**Стан** - значення параметрів, що описують поточну конфігурацію середовища. Агент використовує ці параметри для вибору дії.

# Глибинне RL (Deep RL)



При глибинному навчанні з підкріпленням агент зазвичай обробляє 2D-зображення із використанням згорткових нейронних мереж (CNN) - це дає йому можливість навчатись "із побаченого" завдяки **наскрізній мережі**, яка перетворює набір пікселів у дії.

# Характеристика RL

Чим навчання з підкріплення відрізняється від інших парадигм машинного навчання?

- Ніякого контролю, лише сигнал про винагороду
- Зворотній зв'язок може затримуватися, а не миттєво передаватися
- Час має значення
- Більш ранні рішення агента впливають на його наступні дії

# Основні поняття RL

- Середовище (environment)
- Винагорода (reward)
- Агент (agent), який включає:
  - Стан агента (agent state)
  - Стратегію (policy)
  - Q-функцію, яка відома як функція значення стан-дія (state-action value function)
  - Модель (за бажанням)



# Література

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.