



**TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH**  
HO CHI MINH CITY UNIVERSITY OF FOREIGN LANGUAGES - INFORMATION TECHNOLOGY

# Giới thiệu về Máy học

Biên soạn: **ThS. Vũ Đình Ái** ([aivd@huflit.edu.vn](mailto:aivd@huflit.edu.vn))

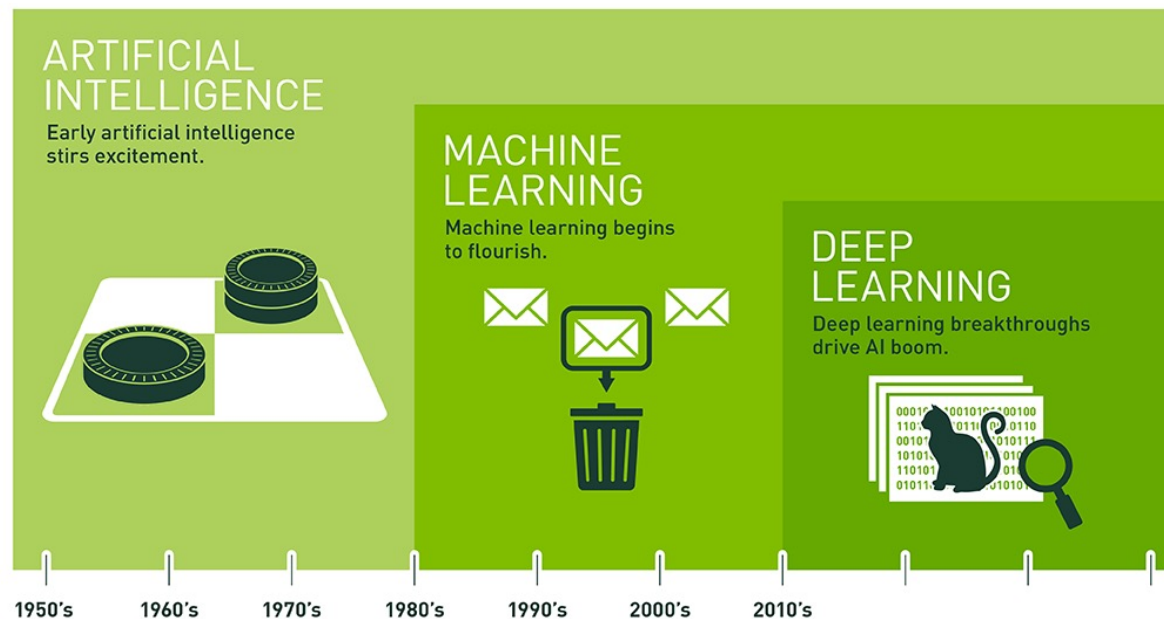
*Cập nhật: tháng 09/2023*

[www.huflit.edu.vn](http://www.huflit.edu.vn)

- Giới thiệu máy học
- Một số định nghĩa về máy học
- Các loại bài toán trong máy học
- Thảo luận – Phân công đề tài

# Giới thiệu máy học

- Machine Learning (học máy) là một lĩnh vực của trí tuệ nhân tạo (AI).
- Mục tiêu của Machine Learning là hiểu được cấu trúc của dữ liệu và tìm ra được các mô hình mà con người có thể hiểu và sử dụng.



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# Giới thiệu máy học

- Ví dụ:



```
if (speed < 4):  
    status = WALKING  
else if (speed < 12):  
    status = RUNNING  
else:  
    status = BIKING
```



# Giới thiệu máy học

## ▪ Làm thế nào Machine Learning có thể giải quyết vấn đề ?

- Trong các phương pháp lập trình truyền thống, các thuật toán được lập trình rõ ràng cho từng vấn đề.
- Bài toán lập trình theo cách truyền thống, chúng ta sử dụng những quy tắc (Rules) cùng với dữ liệu (Data) để đưa ra đáp án (Answers):



# Giới thiệu máy học

- Xét bài toán học tập có giám sát (supervised learning) trong Machine Learning, chúng ta sẽ thu thập dữ liệu cùng với nhãn (label) của từng dữ liệu tương ứng (hay còn gọi là đáp án của dữ liệu đó).
- Từ những dữ liệu này, Machine Learning sẽ đưa ra các quy tắc thông qua việc biểu diễn sự liên hệ giữa Data và Answers



# Giới thiệu máy học

- Với vấn đề cần giải quyết bên trên, chúng ta sẽ thực hiện thu thập một lượng lớn dữ liệu về các hành động Walking, Running, Biking, Golfing và gán nhãn (label) cho từng dữ liệu.
- Chỉ ra với dữ liệu này là hành động đi bộ (Walking) hay với dữ liệu kia là hành động chơi Golf (Golfing).



```
0101001010100101010  
1001010101001011101  
0100101010010101001  
0101001010100101010
```

Label = WALKING



```
1010100101001010101  
0101010010010010001  
0010011111010101111  
1010100100111101011
```

Label = RUNNING



```
1001010011111010101  
1101010111010101110  
1010101111010101011  
1111110001111010101
```

Label = BIKING

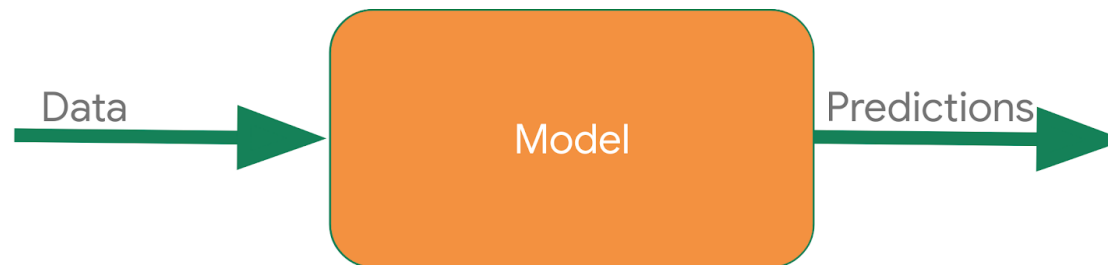


```
1111111111010011101  
0011111010111110101  
0101110101010101110  
101010101001111110
```

Label = GOLFING  
(Sort of)

# Giới thiệu máy học

- Từ những dữ liệu có nhãn này, Machine Learning sẽ đưa ra cho chúng ta quy luật các hành động mà không cần phải lập trình cụ thể.
- Với quy luật được học này, với bất kỳ một dữ liệu mới nào, hệ thống sẽ có thể dễ dàng đưa ra được câu trả lời dự đoán. Kết quả dự đoán sẽ tùy thuộc vào từng bài toán và lượng dữ liệu thu thập được.





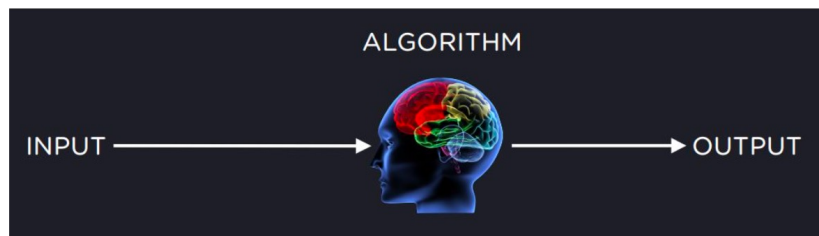
# Một số định nghĩa về máy học

- Định nghĩa 1 bởi Arthur Samuel mô tả (trích nguyên văn):

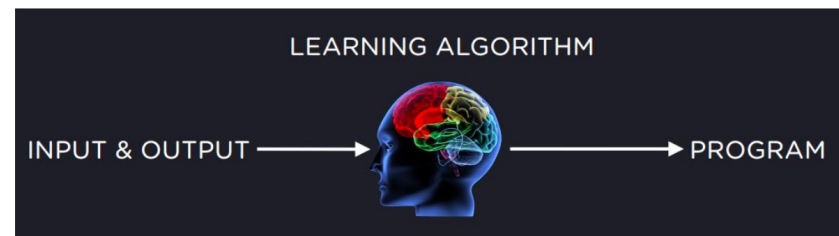
**"The field of study that gives computers the ability to learn without being explicitly programmed."**

- Tạm dịch: Học máy là một lĩnh vực nghiên cứu cung cấp cho máy tính khả năng học hỏi mà không cần phải lập trình rõ ràng.

Traditional programming



Machine Learning



# Một số định nghĩa về máy học

- Định nghĩa 2 bởi Tom Mitchell cung cấp một định nghĩa hiện đại hơn (trích nguyên văn):

**"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."**

- Tạm dịch: Một chương trình máy tính được gọi là học tập từ kinh nghiệm E để hoàn thành nhiệm vụ T và được đo bởi hiệu suất P, nếu hiệu suất tại nhiệm vụ T, được đo bằng P, cải thiện theo kinh nghiệm E.

# Một số định nghĩa về máy học

- T (task): nhiệm vụ/ bài toán cần giải quyết
- E (Experience): dữ liệu cần thu thập
- Performance: cách đánh giá kết quả

Question/  
Hypothesis

(Task)



Image

$x^{(i)}$

$[0,255]^{600 \times 400 \times 3}$



Cat?  
Not cat?

Prediction

$\hat{y}^{(i)}$

True / False  
 $\{1, 0\}$  hoặc  $\{1, -1\}$

Assessment

(Performance)

$$\text{Accuracy} = \frac{1}{M} \sum_m \mathbb{I}(\hat{y}^{(m)} = y^{(m)})$$

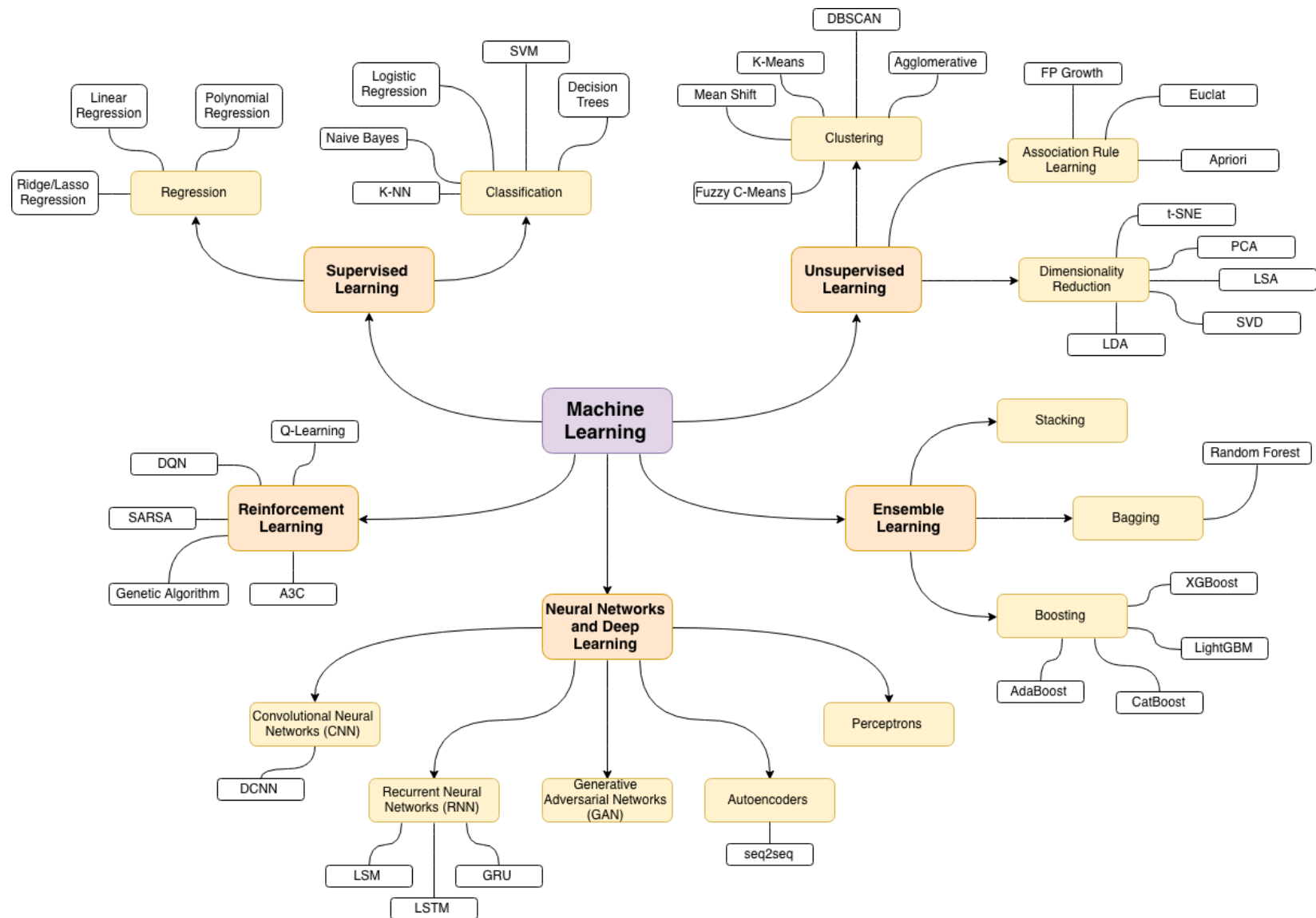
Data  
collection

(Experience)



$$\mathcal{D} = \{x^{(n)}, y^{(n)}\}_1^N$$

# Các loại bài toán trong máy học



# Các loại bài toán trong máy học

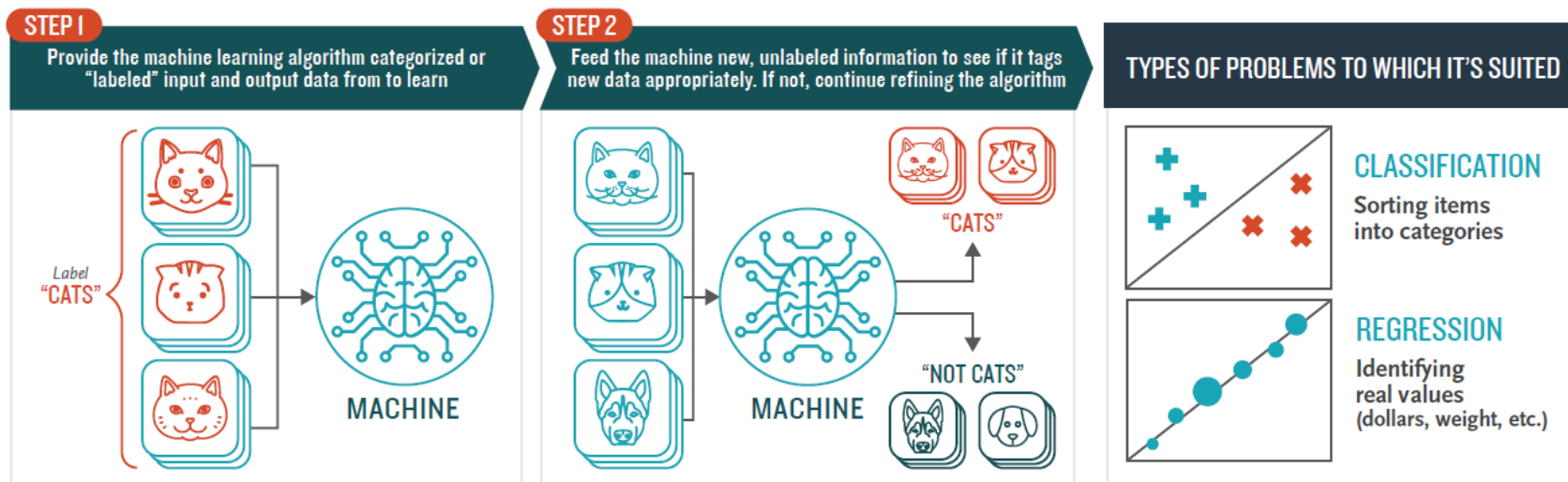
- Dựa trên phương thức học mà có 4 nhóm thuật toán chính trong Machine Learning như sau:
  - Học có giám sát (Supervised learning)
  - Học không giám sát (Unsupervised learning)
  - Học bán giám sát (Semi-supervised learning)
  - Học củng cố (Reinforcement learning)

## ▪ Học có giám sát (Supervised Learning)

- Học có giám sát (supervised learning) là phương pháp học máy mà máy tính được cung cấp đầu vào có gắn nhãn (label). Có nghĩa là: dữ liệu huấn luyện bao gồm các cặp (data, label).
- Đầu ra của bài toán có thể là một giá trị liên tục (gọi là hồi qui - Regression), hay có thể là dự đoán một nhãn phân loại cho một đối tượng đầu vào (gọi là phân loại - Classification).
  - ✓ Bài toán hồi quy có thể sử dụng để dự đoán giá nhà, cân nặng,...
  - ✓ Bài toán phân loại có thể sử dụng để phân loại email (có phải là Spam hay không), phân loại hình ảnh chó / mèo ...

# Các loại bài toán trong máy học

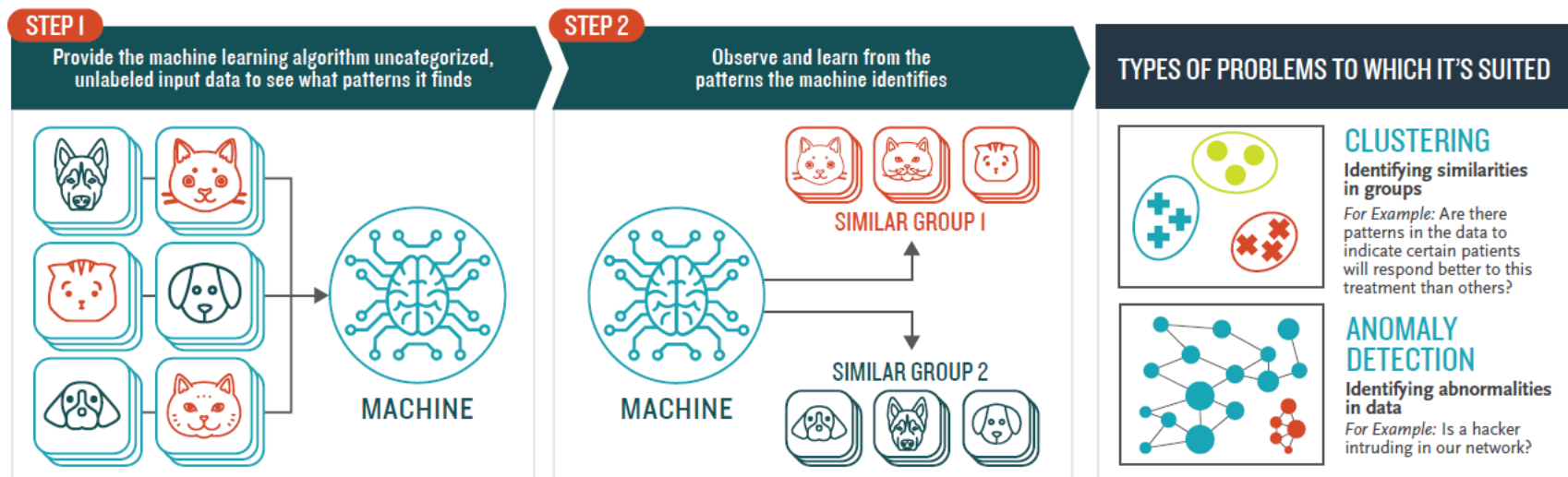
- Ví dụ trong bài toán phân loại hình ảnh chó hoặc mèo. Dữ liệu đầu vào bao gồm hình ảnh chó / mèo cùng với nhãn dán tương ứng cho hình ảnh. Thông qua quá trình đào tạo (train) mà hệ thống sẽ có thể phân loại hình ảnh mới có phải là chó hay mèo không.



# Các loại bài toán trong máy học

## ▪ Học có giám sát (Supervised Learning)

- Trong bài toán học không giám sát (Unsupervised learning) dữ liệu đầu vào sẽ **không** được gắn nhãn. Vì vậy mà thuật toán này sẽ cố gắng tìm các điểm chung của dữ liệu. Trong thực tế, các dữ liệu không có gắn nhãn thường nhiều hơn dữ liệu được gắn nhãn. Chính vì vậy mà phương pháp học không giám sát này luôn mang giá trị đặc biệt.





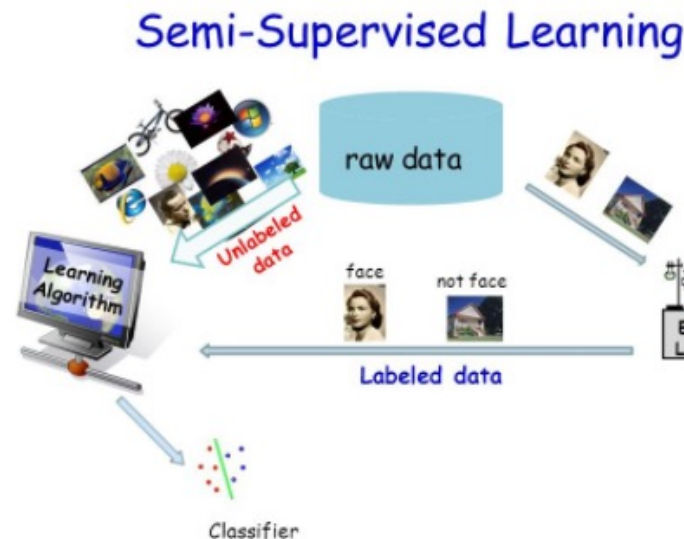
# Các loại bài toán trong máy học

- Các bài toán học không giám sát (Unsupervised learning) được phân thành 2 loại chính:
  - **Clustering (Phân nhóm):** Với bộ dữ liệu đầu vào, hệ thống sẽ dựa trên mức độ liên quan của các dữ liệu mà phân chia thành các nhóm khác nhau. Ví dụ trong bài toán có đầu vào là các hình ảnh chó và mèo (nhưng không được gắn nhãn), sau đó qua quá trình học tập sẽ phân chia được thành 2 nhóm: 1 nhóm hình ảnh bao gồm các con mèo và 1 nhóm hình ảnh bao gồm các con chó.
  - **Anomaly Detection (Phát hiện bất thường):** Trong bài toán này, hệ thống sẽ tìm ra các quy luật của các dữ liệu cho trước. Nếu có bất kỳ dữ liệu mới nào mà lệch khỏi quy luật này sẽ được tính vào là một hành vi bất thường. Ví dụ như trong giao dịch ngân hàng, đột nhiên có một tài khoản quá 12 giờ đêm mà giao dịch hàng tỷ đồng.

# Các loại bài toán trong máy học

## ▪ Học bán giám sát (Semi-Supervised Learning)

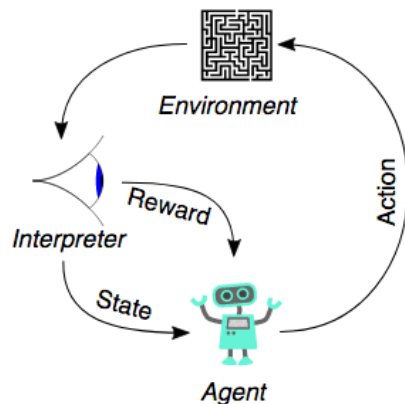
- Trong học bán giám sát (Semi-supervised learning), dữ liệu đầu vào một phần sẽ được gắn nhãn - là sự kết hợp của 2 phương pháp học có giám sát (Supervised learning) và học không giám sát (Unsupervised learning).
- Ví dụ bạn có một kho hình ảnh khổng lồ, nhưng chỉ có dữ liệu về người được gắn nhãn. Mà bạn lại muốn phân loại thành nhiều nhóm khác nhau như nhóm xe đạp, nhóm ảnh thiên nhiên, ...



# Các loại bài toán trong máy học

## ▪ Học củng cố (Reinforcement Learning)

- Học củng cố hay học tăng cường (Reinforcement Learning) thường được áp dụng vào "lý thuyết trò chơi" (Game Theory), thuật toán sẽ cần phải xác định các bước đi tiếp theo để đạt được điểm số cao nhất. Trong học củng cố sẽ có phần thưởng và phạt dựa trên các hành vi - chẳng hạn: trong bước tiếp theo máy đạt được điểm cao hơn thì sẽ nhận được phần thưởng và ngược lại nếu điểm số hạ xuống sẽ bị phạt.
- Ứng dụng điển hình của học củng cố là AlphaGo - chơi cờ vây có thể thắng được con người.



## Thảo luận – Phân công đề tài

- Mỗi đề tài cần 2-3SV
- Thời gian: 12 tuần
- Link đăng ký
- Các bước triển khai
  - Xác định loại đề tài (Dự đoán, phân lớp (nhị phân, đa lớp), phân cụm...)
  - Thu thập dữ liệu → Xử lý dữ liệu → Lựa chọn mô hình (huấn luyện mô hình) → Đánh giá và tinh chỉnh mô hình → Xây dựng ứng dụng demo
  - Viết báo cáo