

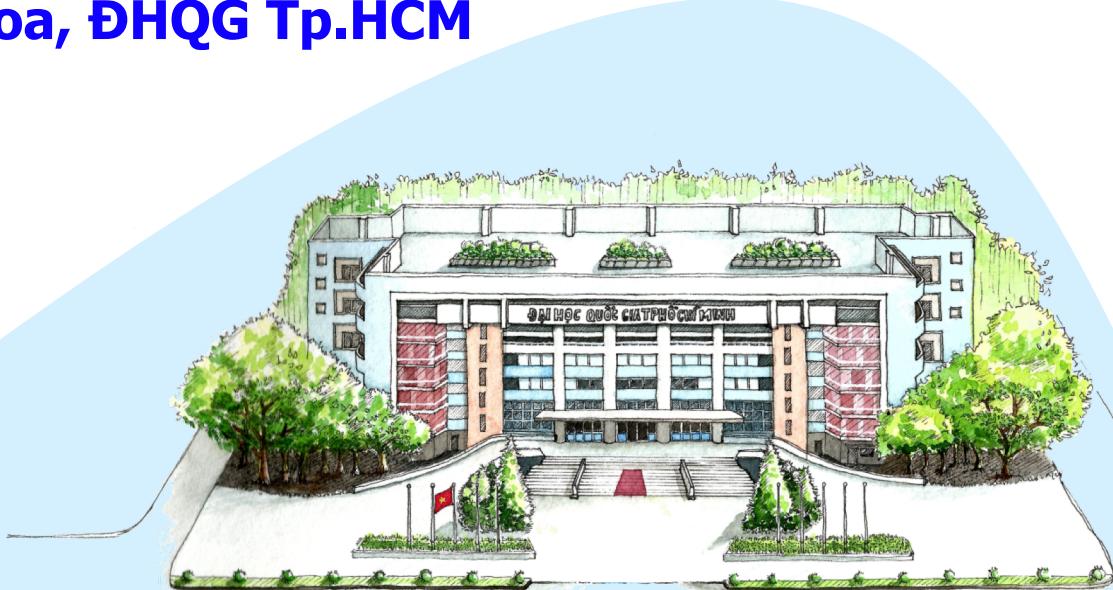
Trí tuệ nhân tạo

Tổng quan, ứng dụng, nguy cơ, và thách thức

TS. Lê Thành Sách (LTSACH@hcmut.edu.vn)

Khoa Khoa học & Kỹ thuật Máy tính

Trường Đại học Bách Khoa, ĐHQG Tp.HCM



NỘI DUNG



Trí tuệ nhân tạo là gì?



AI: Tại sao sử dụng?



AI: Ứng dụng và tiềm năng



AI: Bằng cách nào?



AI: Nguy cơ tiềm ẩn



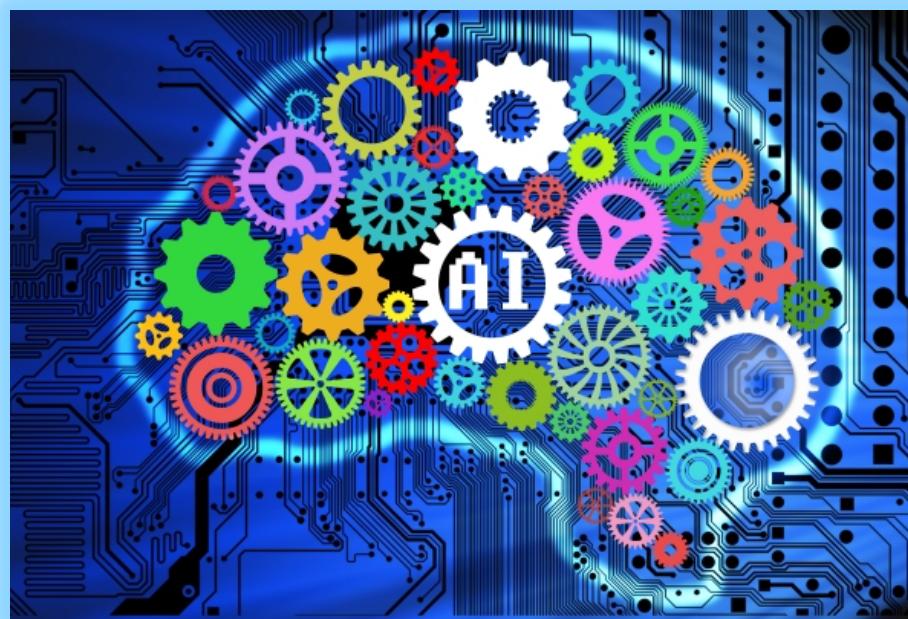
AI: Các thách thức



Kết luận

Trí tuệ nhân tạo là gì?

- **Trí tuệ nhân tạo (Artificial Intelligence, AI)**
 - Là một lĩnh vực nghiên cứu giúp cho máy móc^(*) có các khả năng của con người.



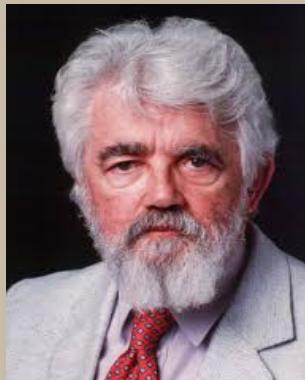
(*) Các thực thể nhân tạo

Trí tuệ nhân tạo là gì?

- **Trí tuệ nhân tạo (Artificial Intelligence, AI)**
 - Là một lĩnh vực nghiên cứu giúp cho máy móc có các khả năng của con người.

Đề xuất tại: Hội nghị Dartmounth (1956)

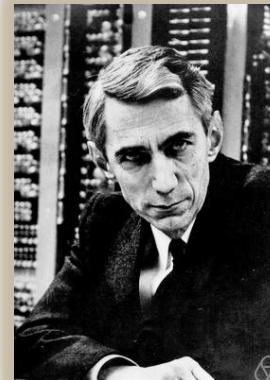
(Dartmouth Summer Research Project on Artificial Intelligence)



John McCarthy



Marvin Minsky



Claude Shannon

...



Allen Newell

Trí tuệ nhân tạo là gì?

Con người có những khả năng đặc biệt nào?

Trí tuệ nhân tạo là gì?

Tâm nhìn từ ~1950

1

Khả năng sử dụng ngôn ngữ

Hiểu văn bản, tóm tắt văn bản, viết văn

2

Khả năng nhìn - hiểu (thị giác)

Hiểu ảnh và video

3

Khả năng giao tiếp qua nghe-nói

Nhận dạng tiếng nói và phát âm

Trí tuệ nhân tạo là gì?

Tâm nhìn từ ~1950

4

Khả năng biểu diễn tri thức và thực hiện suy diễn

Thực hiện chuỗi biến đổi:

Dữ liệu → thông tin → tri thức → thông minh.

Suy diễn dựa vào tri thức

*Suy diễn là khả năng yếu nhất hiện nay của AI.
Vui hay không?*

AI (hiện nay): Khả năng phân tích dữ liệu thông minh và đưa ra thông tin có giá trị

Trí tuệ nhân tạo là gì?

Tâm nhìn từ ~1950

5

Khả năng thể hiện hành động

Thực hiện hành động thông minh:
biểu cảm,
đi, đứng, chạy, nhảy, phối hợp,
điều khiển thiết bị, v.v.

6

Khả năng học

Học tập để nâng cao tri thức và kỹ năng.

Học tập, giáo dục, đào tạo: được quan tâm bởi mọi quốc gia

Trí tuệ nhân tạo là gì?

Tâm nhìn từ ~1950

Con người

Học,
Học nữa,
Học mãi.

Trí tuệ nhân tạo



NỘI DUNG



Trí tuệ nhân tạo là gì?



AI: Tại sao sử dụng?



AI: Ứng dụng và tiềm năng



AI: Bằng cách nào?



AI: Nguy cơ tiềm ẩn



AI: Các thách thức

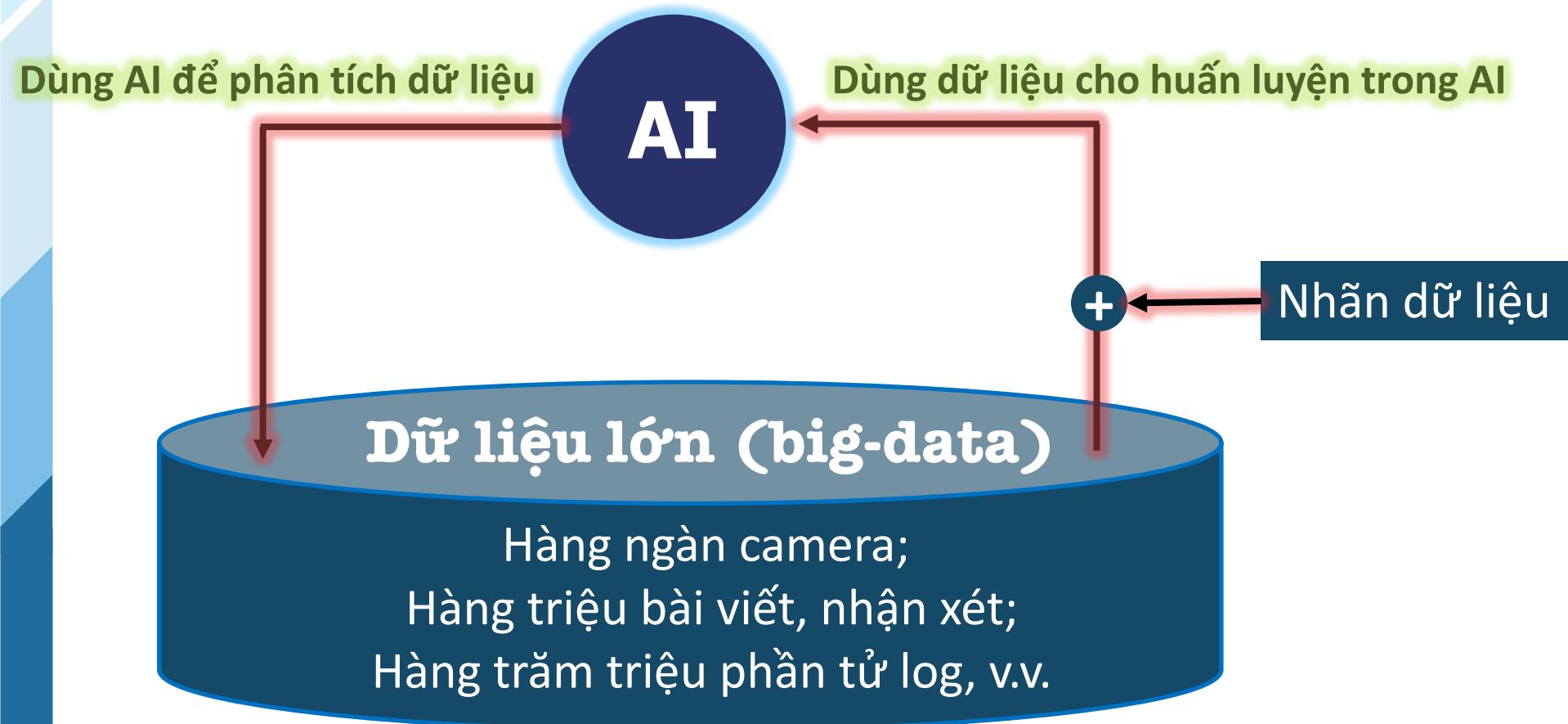


Kết luận

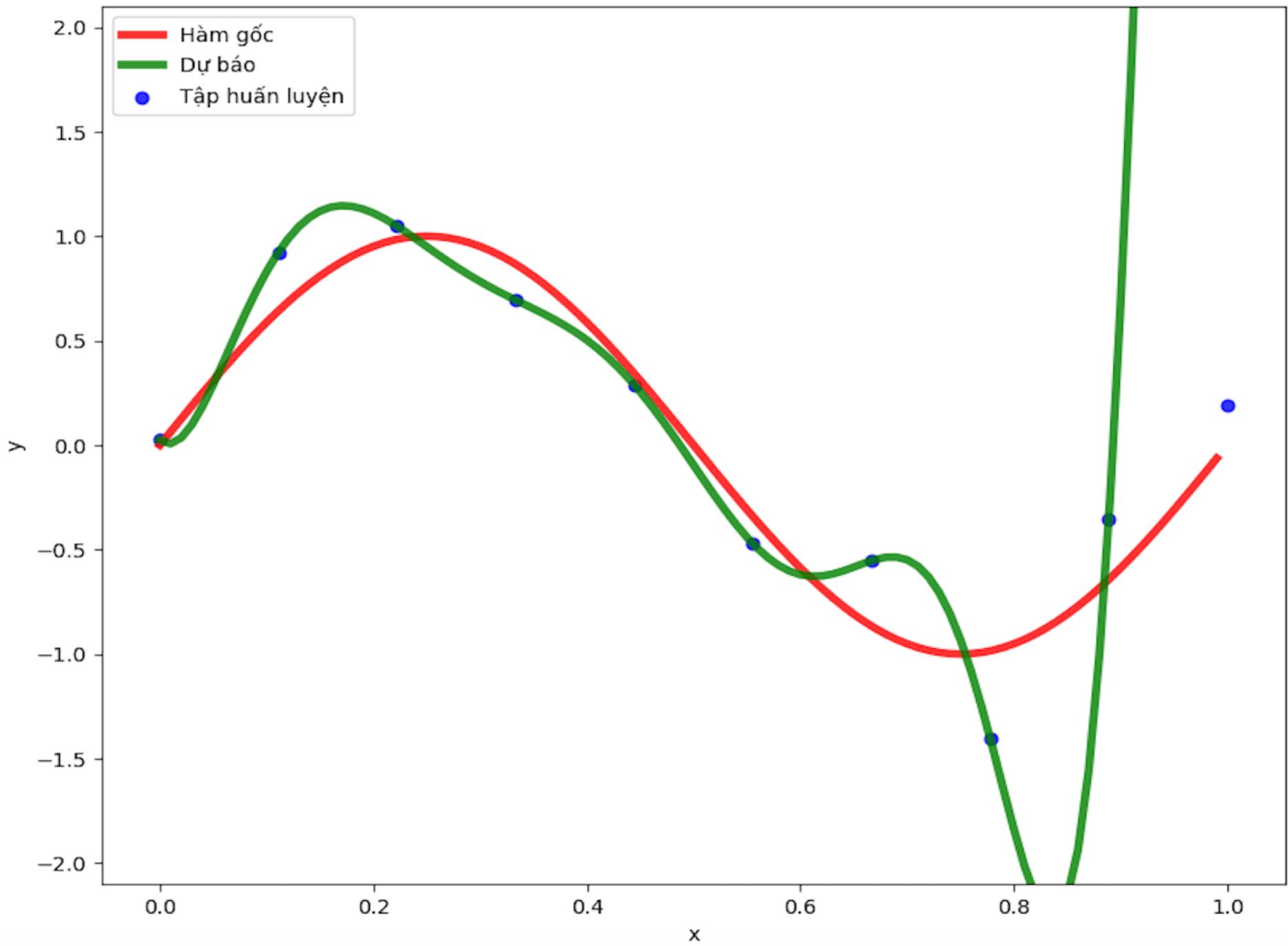
AI: Tại sao sử dụng?

1 Dữ liệu lớn

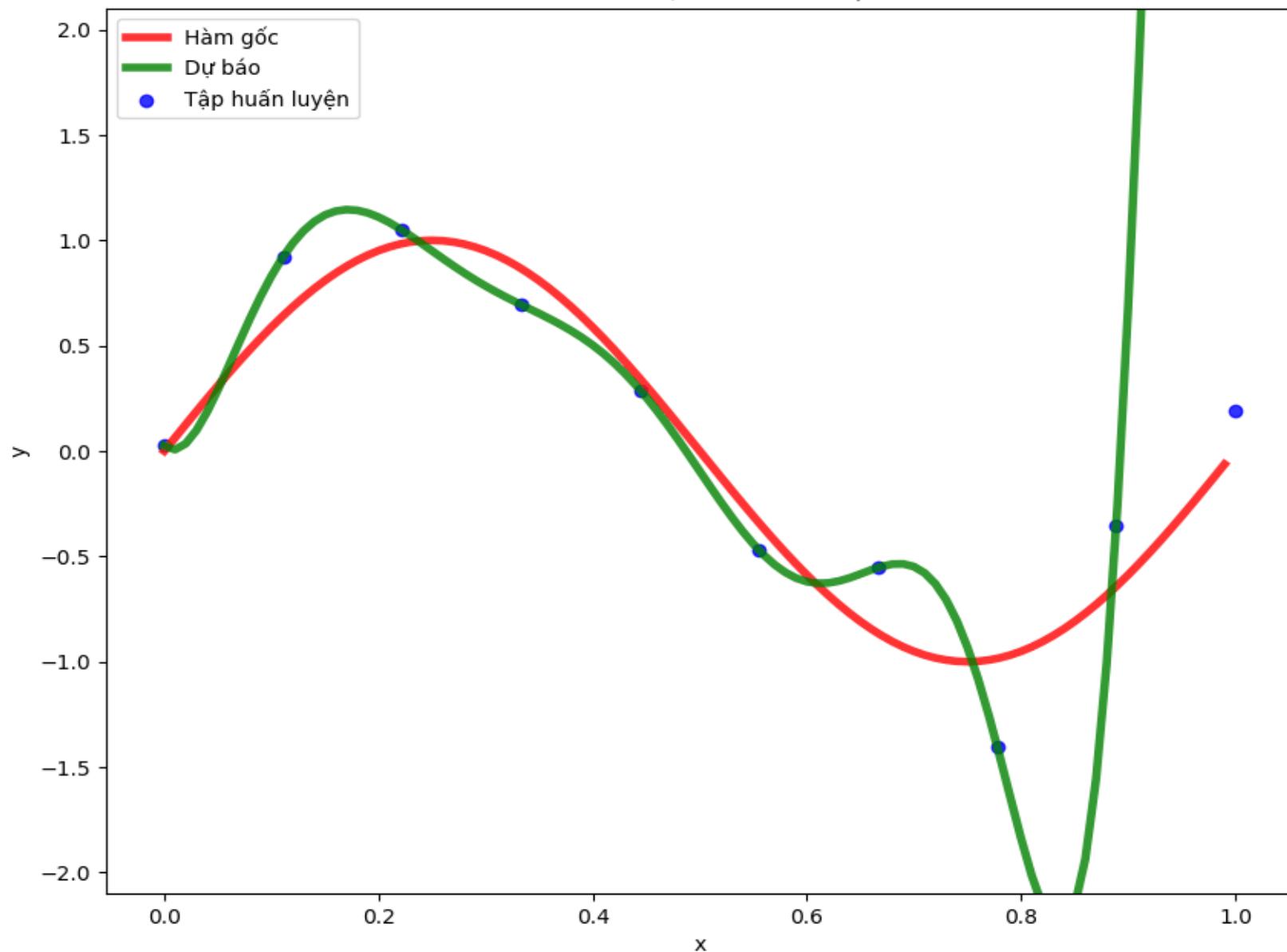
Sự thúc đẩy từ dữ liệu



Số tham số: 16; số điểm dữ liệu:10



Số tham số: 16; số điểm dữ liệu: 10



AI: Tại sao sử dụng?

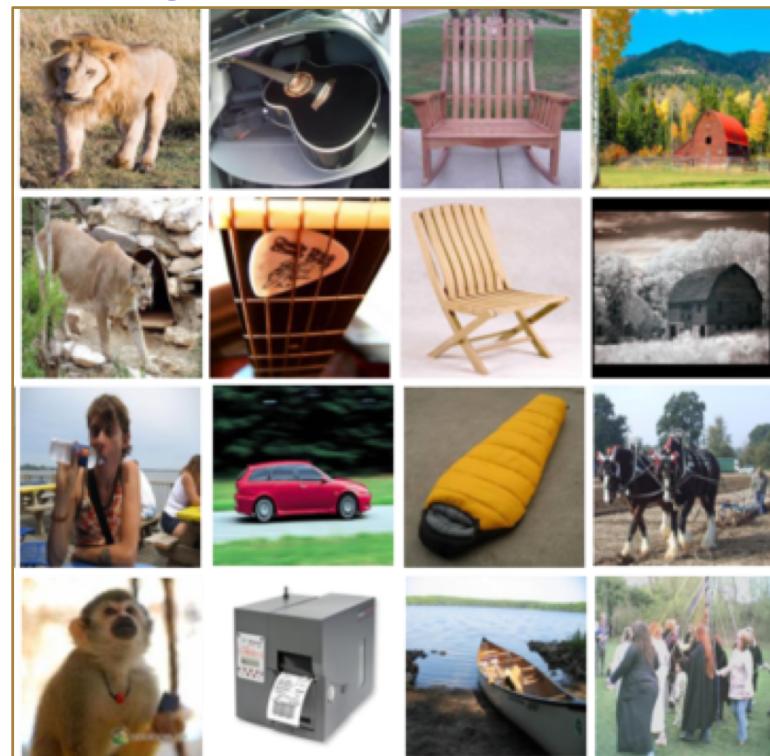
1 Dữ liệu lớn

Sự thúc đẩy từ dữ liệu

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

<http://www.image-net.org>

- Số lớp: 1000
 - ✓ Các giống mèo,
 - ✓ Các giống chó,
 - ✓ Các giống cá, ...
- Số ảnh huấn luyện (**có nhăn**)
 - ILSVRC-2010: 1.2 triệu ảnh
 - ILSVRC-2017: > 14 triệu ảnh
- Số ảnh kiểm tra (**có nhăn**)
 - ILSVRC-2010: 100.000 ảnh
 - ILSVRC-2017: vài trăm ngàn ảnh



AI: Tại sao sử dụng?

② Công nghệ tính toán mới

Sự hỗ trợ từ máy móc

Rút ngắn thời gian học



NVIDIA V100



NVIDIA T4



Jetson TX2

Xử lý dữ liệu nhanh

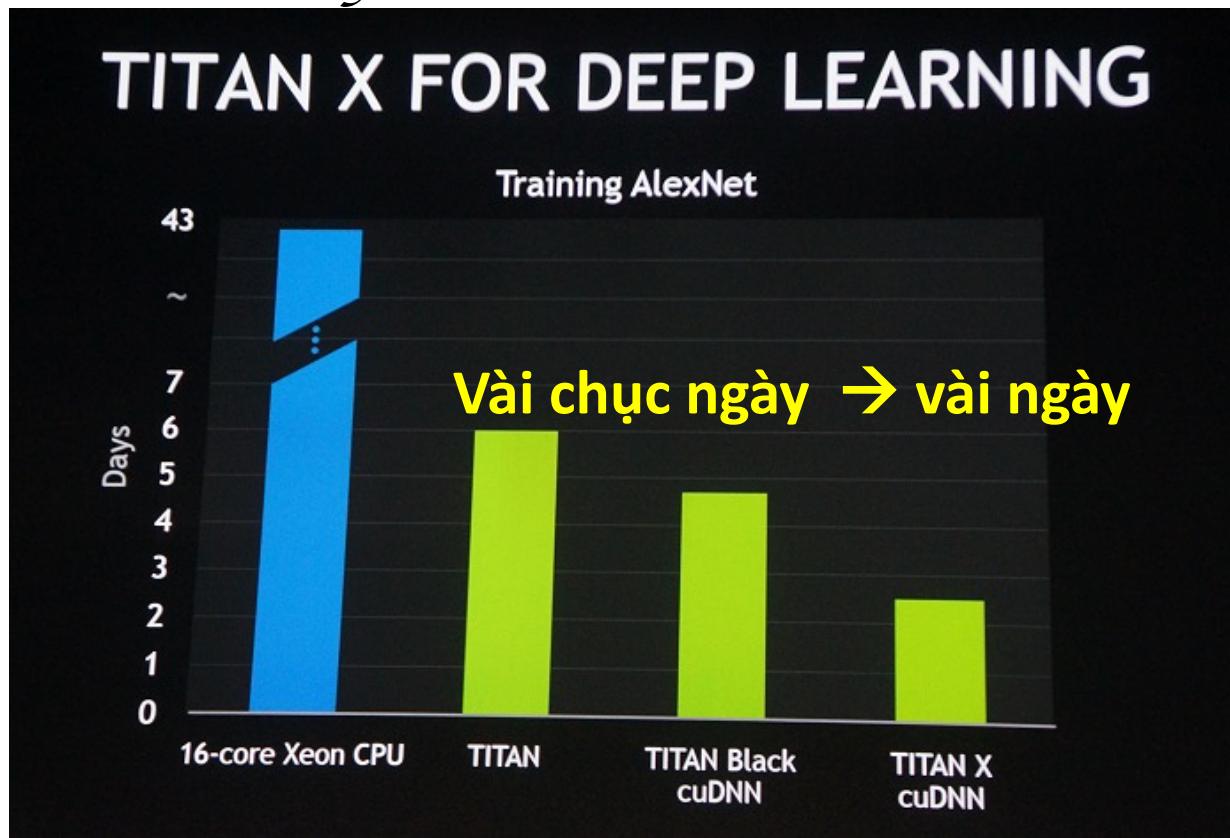
Hỗ trợ nhiều dạng ứng dụng

Tính toán tại sensor, tính toán tập trung

AI: Tại sao sử dụng?

② Công nghệ tính toán mới

Sự hỗ trợ từ máy móc

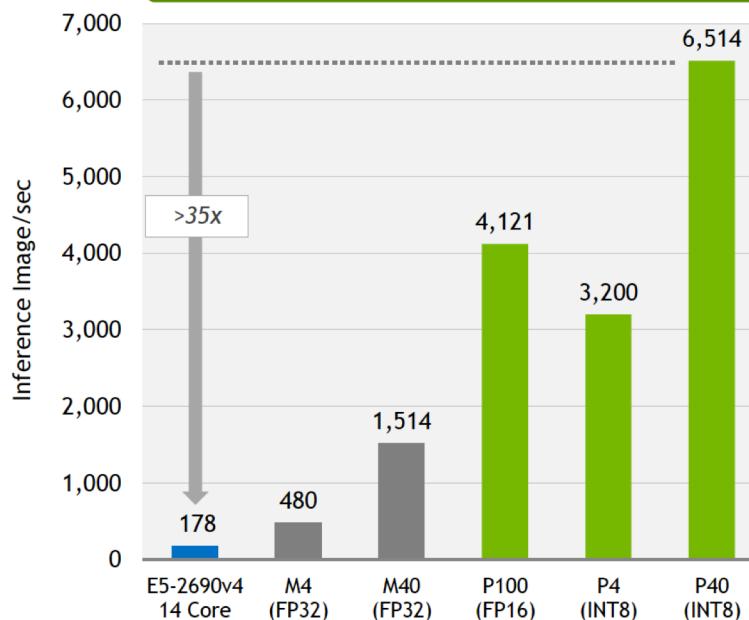


AI: Tại sao sử dụng?

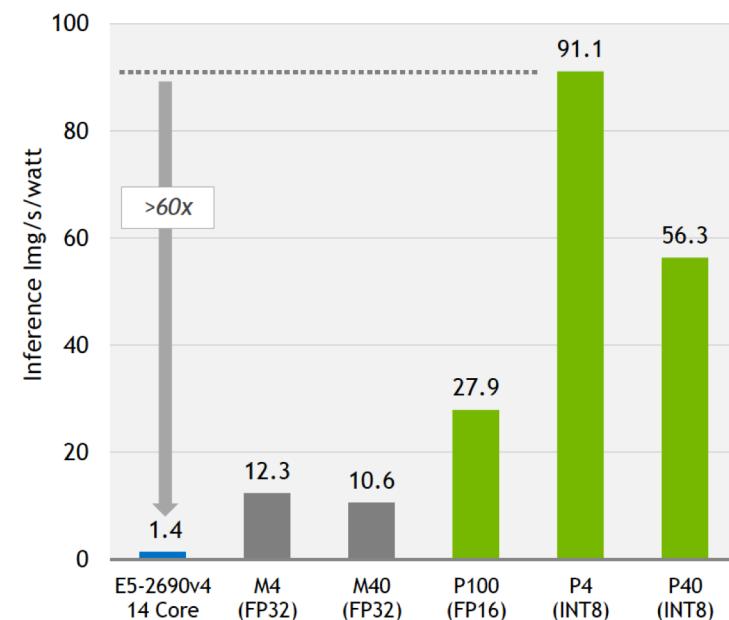
2 Công nghệ tính toán mới

Sự hỗ trợ từ máy móc

P40 For Max Inference Throughput



P4 For Max Inference Efficiency



All results are measured, based on GoogLeNet with batch size 128
Xeon uses MKL 2017 GOLD with FP32, GPU uses TensorRT internal development ver.

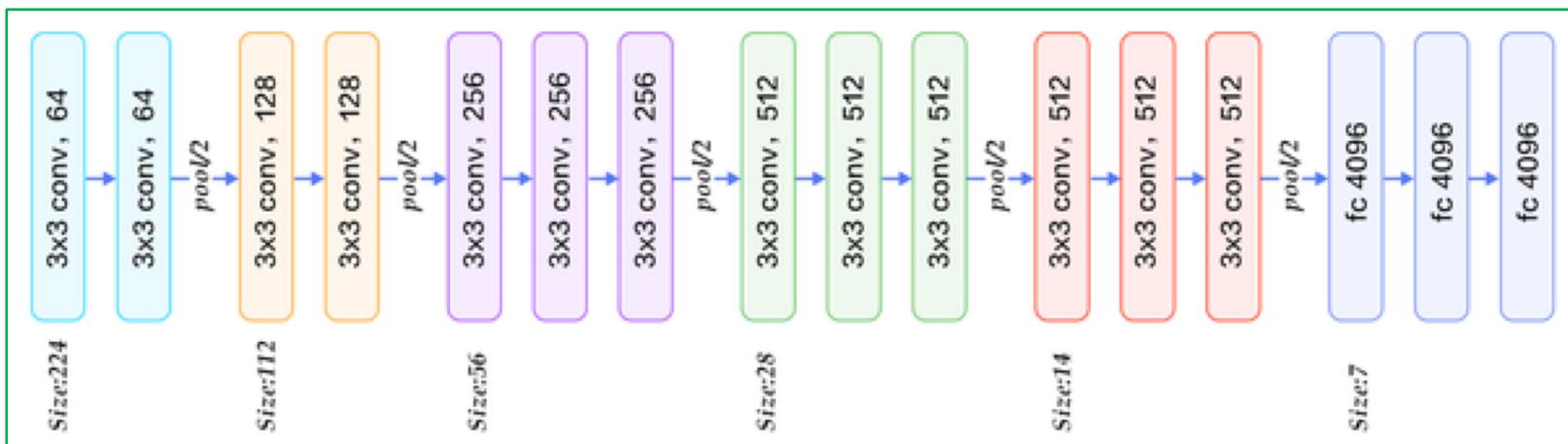
NVIDIA CONFIDENTIAL. DO NOT DISTRIBUTE.

23 NVIDIA

AI: Tại sao sử dụng?

3 Kỹ thuật học mới – Học sâu (deep learning)

Sự hỗ trợ từ khoa học & kỹ thuật



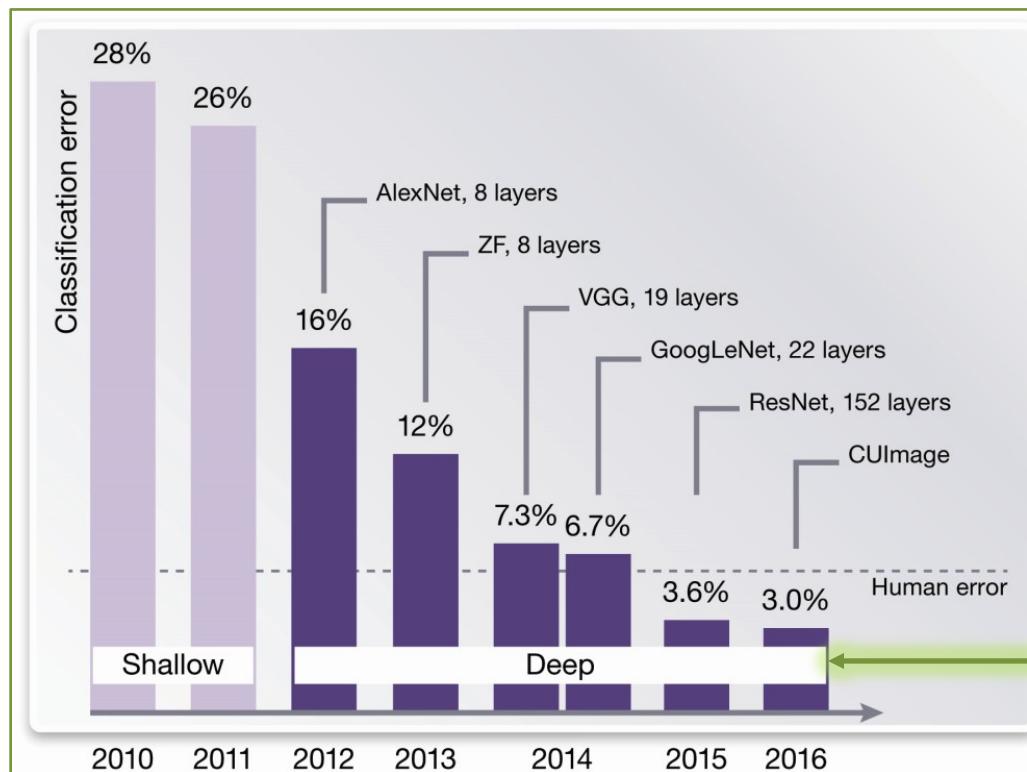
Học sâu:

- Tăng độ chính xác
- Thích nghi tốt với dữ liệu và tình huống chưa gặp (học bổ sung)
- Dễ sử dụng (end-to-end)
- Thống nhất cách tiếp cận cho các nhánh của AI (All-in-one)

AI: Tại sao sử dụng?

3 Kỹ thuật học mới – Học sâu (deep learning)

Kết quả cuộc thi về nhận dạng ảnh (ILSVRC)



AI đã vượt qua con người

AI: Tại sao sử dụng?

Độ chính xác cao

Tính toán nhanh

Thích nghi tốt với dữ liệu

Dễ sử dụng

Tiềm năng ứng dụng rộng rãi

Một kỹ thuật cho nhiều loại khả năng

AI
là chủ đề nóng

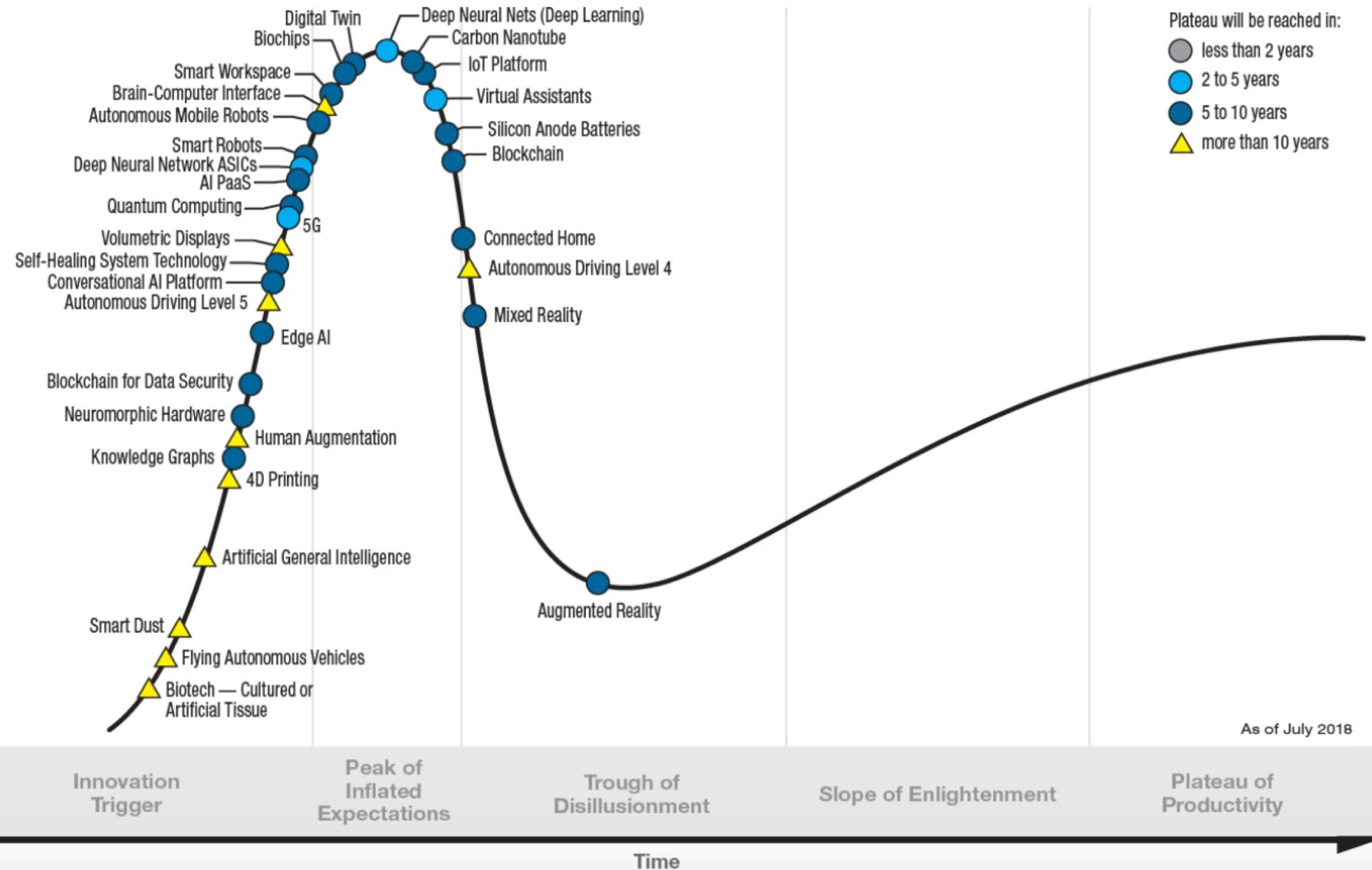
Nghe,
nói

Nhìn

Ngôn
ngữ

Hành
động

Phân
tích
dữ liệu



(Gartner Hype Cycle for 2018)

NỘI DUNG



Trí tuệ nhân tạo là gì?



AI: Tại sao sử dụng?



AI: Ứng dụng và tiềm năng



AI: Bằng cách nào?



AI: Nguy cơ tiềm ẩn

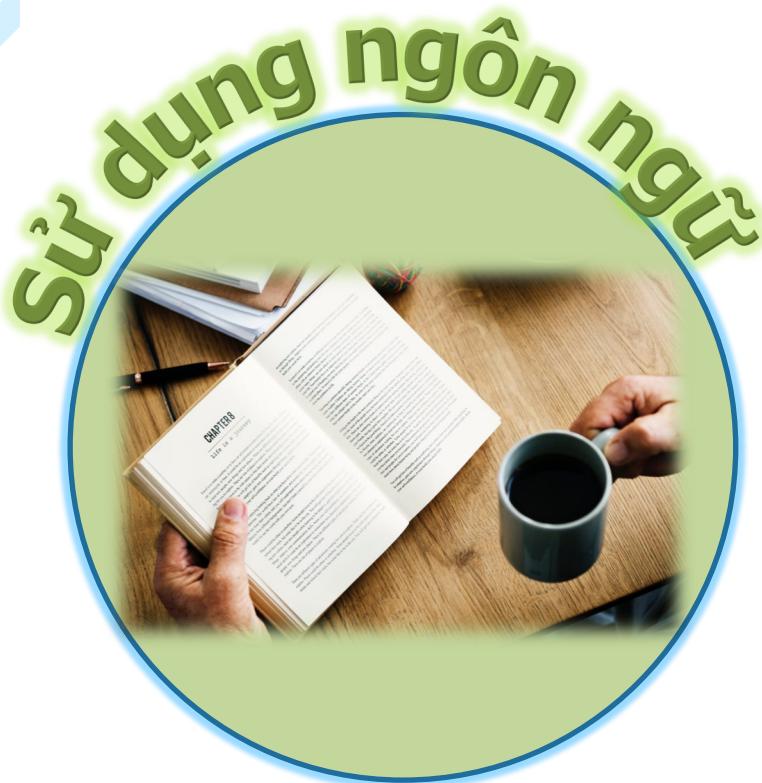


AI: Các thách thức



Kết luận

AI: Ứng dụng và tiềm năng



Dịch máy

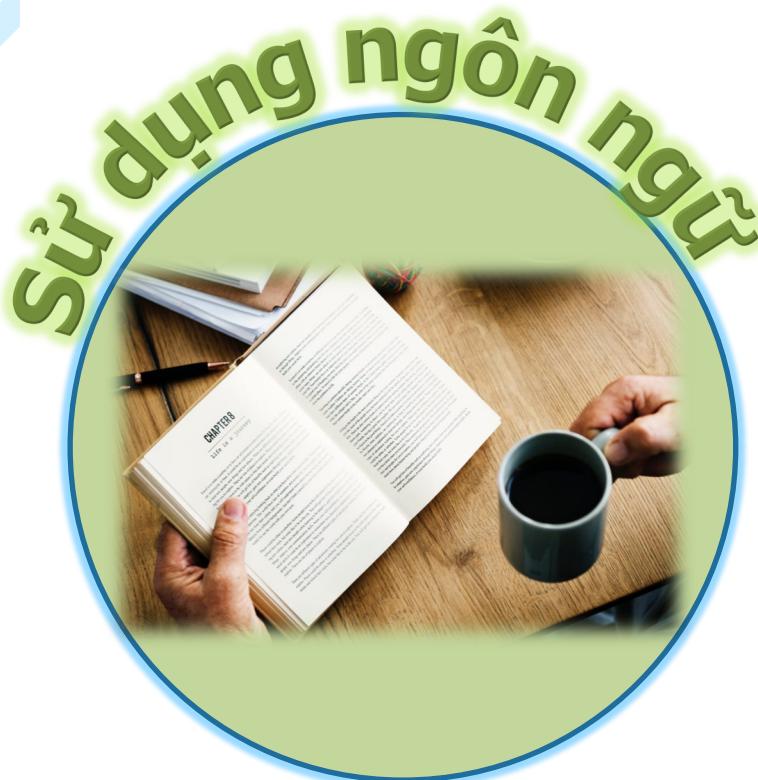
[Google Translate](#)

Hội thoại

Tóm tắt,
phân loại nội dung

Phân tích cảm xúc

AI: Ứng dụng và tiềm năng



Dịch máy

Google Translate

Giao tiếp, du lịch

Hội thoại

Trợ lý ảo

Tóm tắt,
phân loại nội dung

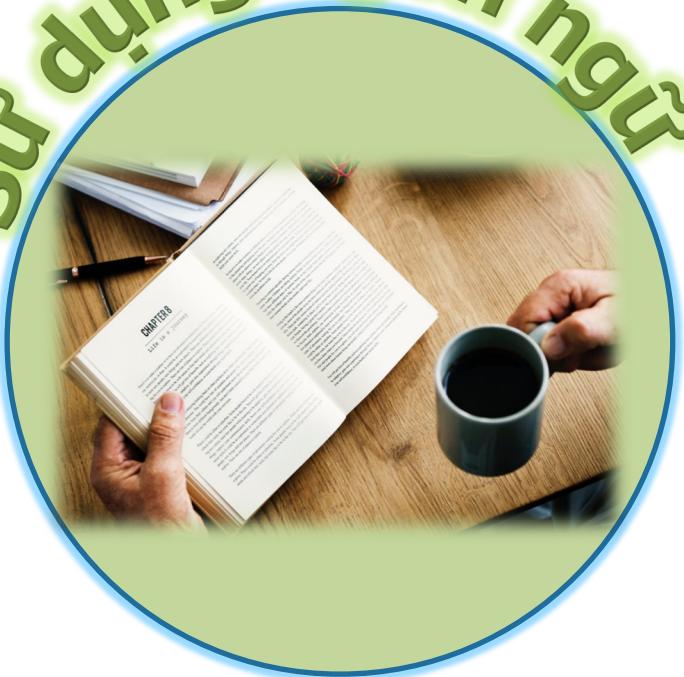
An ninh, tuyên giáo

Phân tích cảm xúc

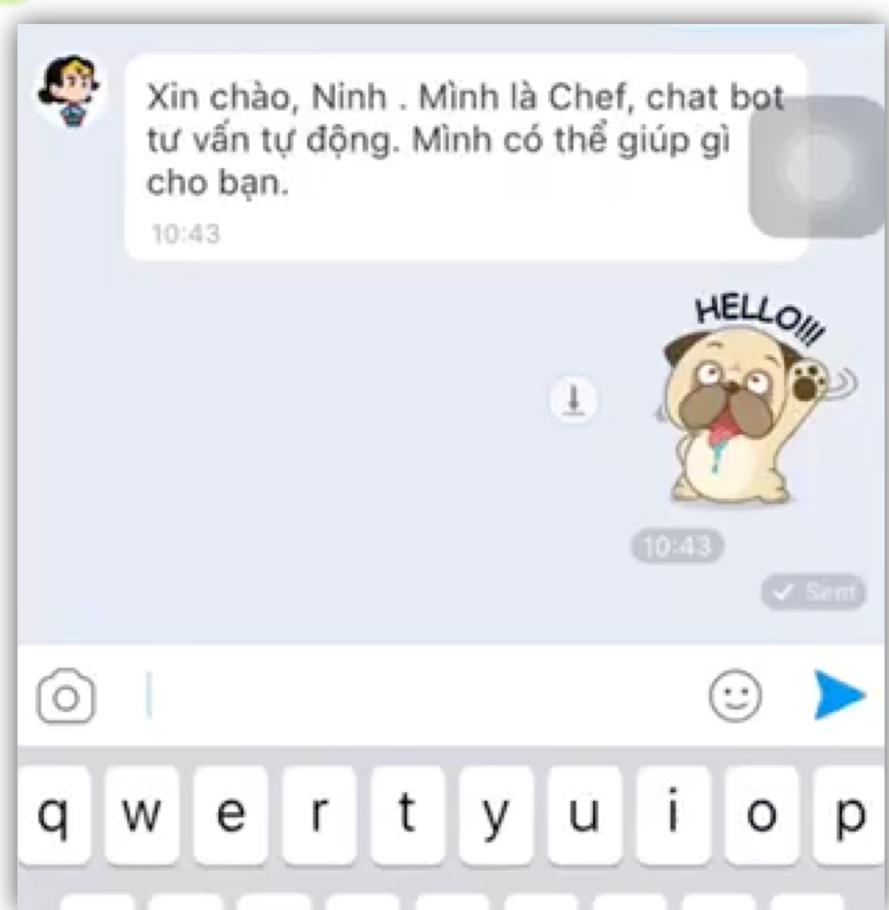
Kinh doanh, thương mại

AI: Ứng dụng và tiềm năng

Sử dụng ngôn ngữ



Hội thoại

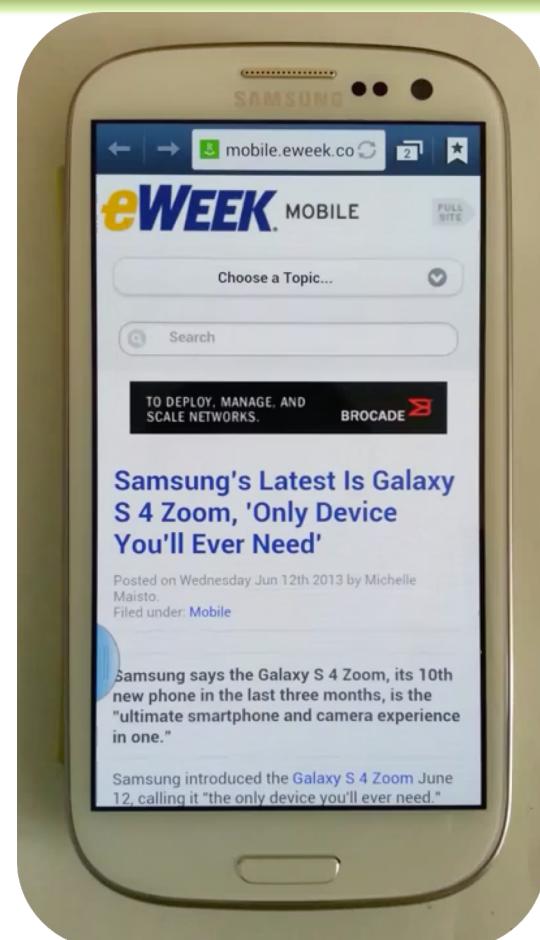


AI: Ứng dụng và tiềm năng

Sử dụng ngôn ngữ



Tóm tắt nội dung



AI: Ứng dụng và tiềm năng

Thể hiện hành động



Robot tựa con người

Di chuyển, phối hợp, thể hiện cảm xúc, v.v

Xe tự hành

Máy bay tự điều khiển

AI: Ứng dụng và tiềm năng

Thể hiện hành động



Cách tiếp cận hiện đại trong chế tạo Robots:

- ➡ **Học để di chuyển, chạy, nhảy;**
- ➡ **Học để phối hợp;**
- ➡ **Học là nền tảng cho hành động thông minh.**

Cơ khí chính xác;
Cơ điện tử;
Chương trình cài đặt sẵn

AI: Ứng dụng và tiềm năng

Thể hiện hành động

Học tránh chướng ngại vật

Học để điều khiển
(lựa chọn hành động)

AI: Ứng dụng và tiềm năng



Nhận dạng tiếng nói

Phổ biến: Google, Apple Siri

Tổng hợp tiếng nói

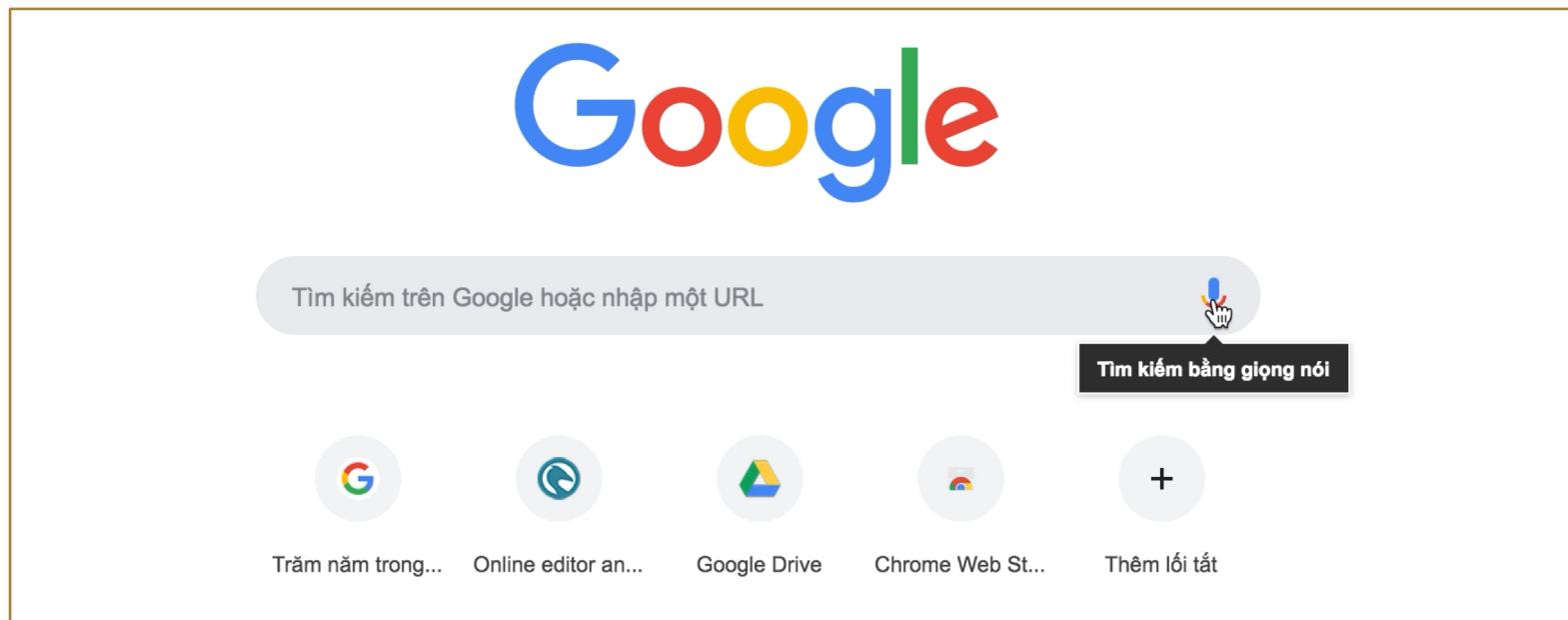
Tin tức báo chí: VOA, Newyork times, v.v

Ứng dụng tổng hợp

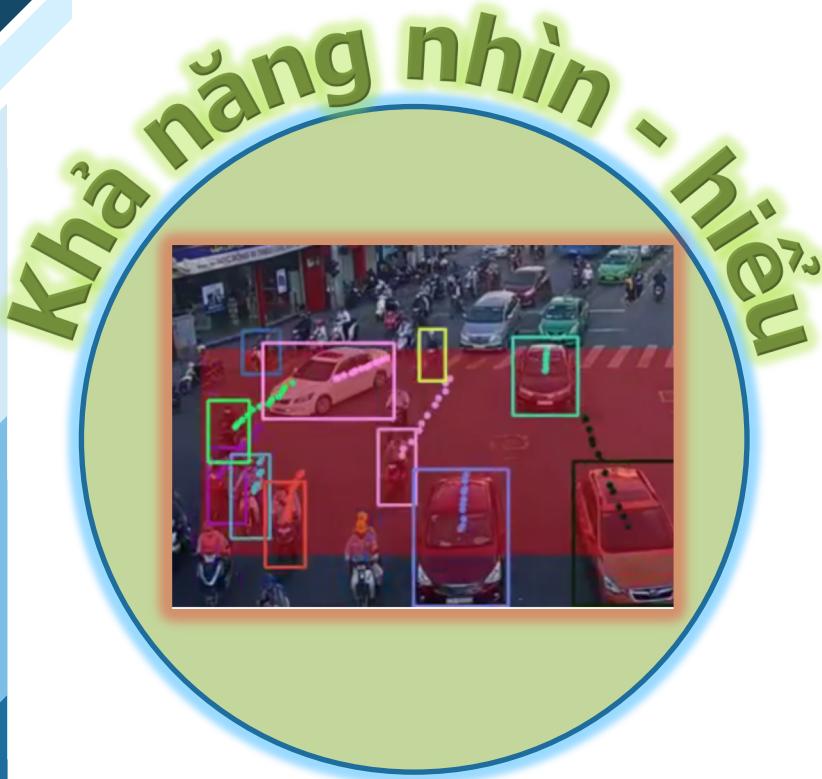
Tạo lý ảo (Siri), Biên tập viên TH, v.v.

AI: Ứng dụng và tiềm năng

Giao tiếp qua nghe và nói



AI: Ứng dụng và tiềm năng



Định danh qua khuôn mặt

Khuôn mặt, hình thể, liên kết qua dây camera

Phân tích ảnh giao thông

Mật độ, tốc độ, loại xe, tai nạn giao thông, v.v.

Phân tích ảnh y khoa

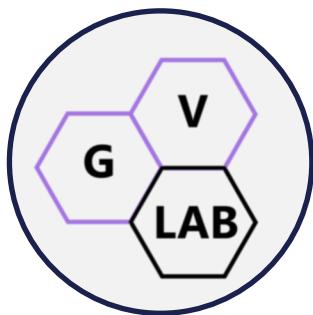
Phát hiện tổn thương trên ảnh gan, phổi da, v.v.

Phân tích hình ảnh/video khác

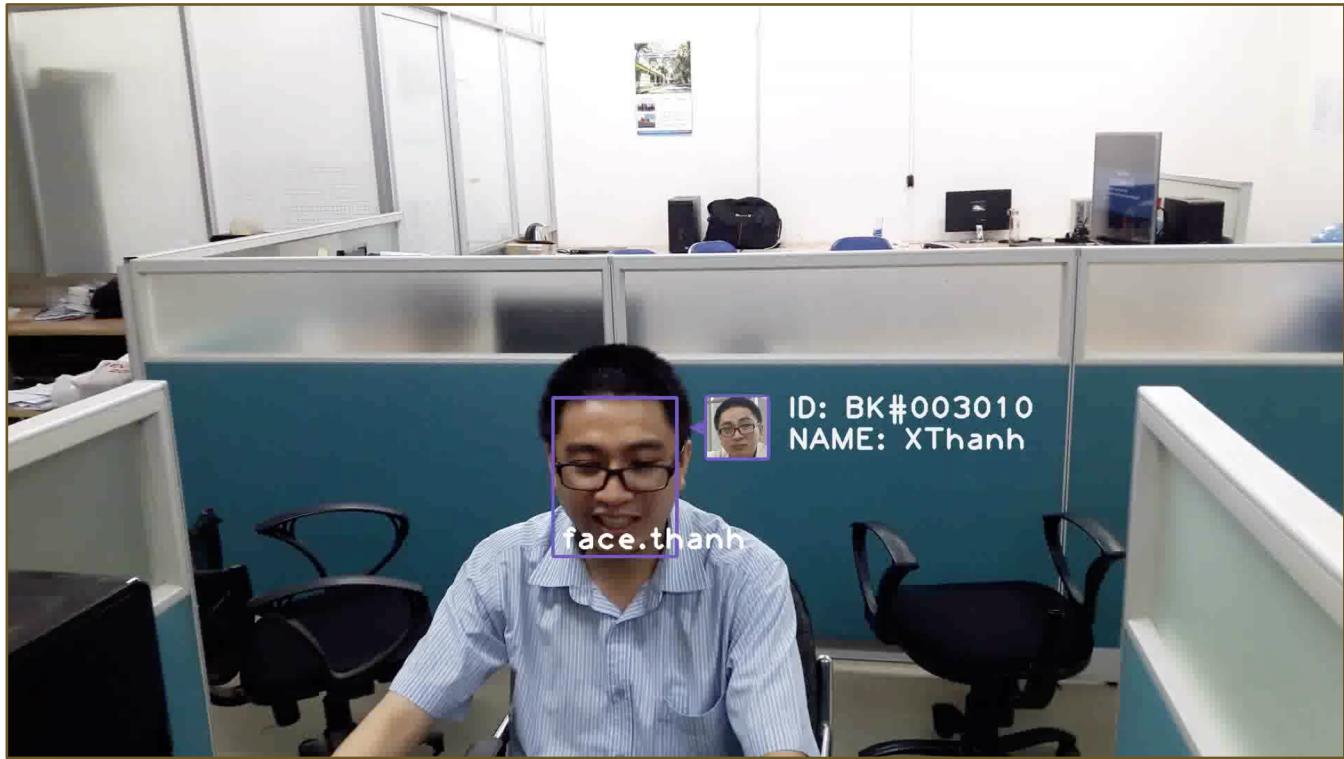
Phân tích cảm xúc, phát hiện cháy, xả rác, v.v.

AI: Ứng dụng và tiềm năng

Khả năng nhìn - hiểu



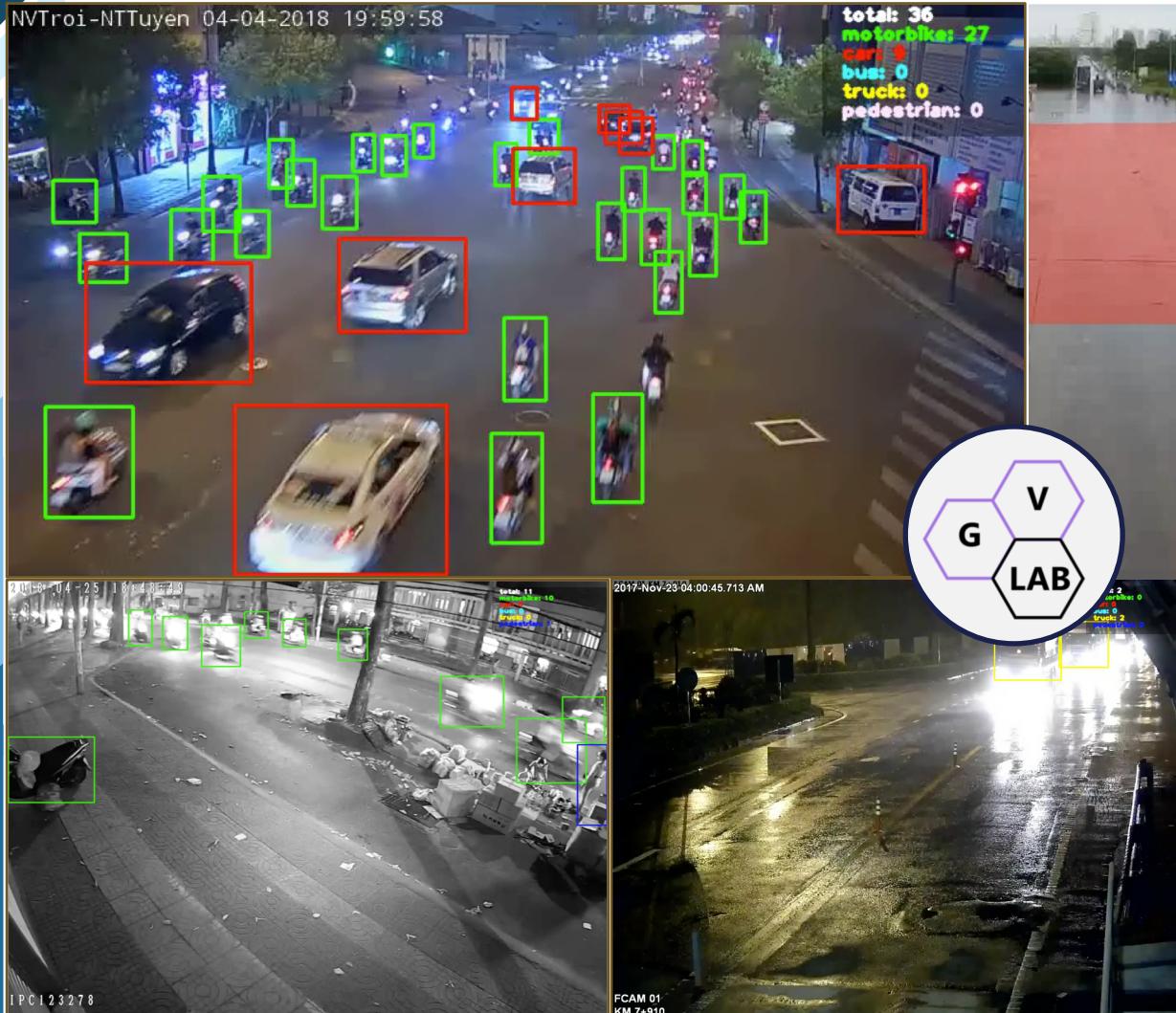
gvEyes



Hỗ trợ các ứng dụng: điểm danh PTN, lớp học, an ninh, v.v.

AI: Ứng dụng và tiềm năng

Khả năng nhìn - hiểu



Học để phát hiện,
phân loại,
ước lượng, v.v.

AI: Ứng dụng và tiềm năng

Sử dụng ngôn ngữ viết

1

Nhìn - hiểu (thị giác)

2

Nghe-nói

3

Suy diễn với tri thức

4

Thể hiện hành động

5

Học tập

6

Đủ
tốt

Tự động hóa cao,
Giảm sức người,
Giảm giá thành,
Tăng lợi nhuận, ...

NỘI DUNG



Trí tuệ nhân tạo là gì?



AI: Tại sao sử dụng?



AI: Ứng dụng và tiềm năng



AI: Bằng cách nào?



AI: Nguy cơ tiềm ẩn



AI: Các thách thức



Kết luận

AI: Bằng cách nào?

(AI theo xu hướng học sâu)

5 công việc

- Chuẩn bị dữ liệu
Thống kê, làm nhãn, v.v.
- Thiết kế mạng nơron
- Thiết kế hàm tổn thất
- Huấn luyện kiến trúc nơron
- Triển khai ứng dụng

3 yêu cầu

- 1 Dữ liệu của cho bài toán
- 2 Nhân sự giỏi
 - Vững Toán
 - Giới lập trình
 - Có tư duy phản biện
- 3 Hạ tầng tính toán:
 - Mạnh cho huấn luyện
 - Đủ mạnh cho triển khai

AI: Bằng cách nào?

(AI theo xu hướng học sâu)

Trí tuệ nhân tạo

Con người

Chuẩn bị dữ liệu **1**

Thống kê, làm nhãn, v.v.

Thiết kế mạng nơron **2**

Chuẩn bị tư liệu học tập:
sách, bài tập, v.v.

tương đương ?
Hỗ trợ: chăm sóc sức khỏe

Thiết kế hàm tổn thất **3**

Phương pháp đánh giá.
Chính sách thưởng, phạt

Huấn luyện kiến trúc nơron **4**

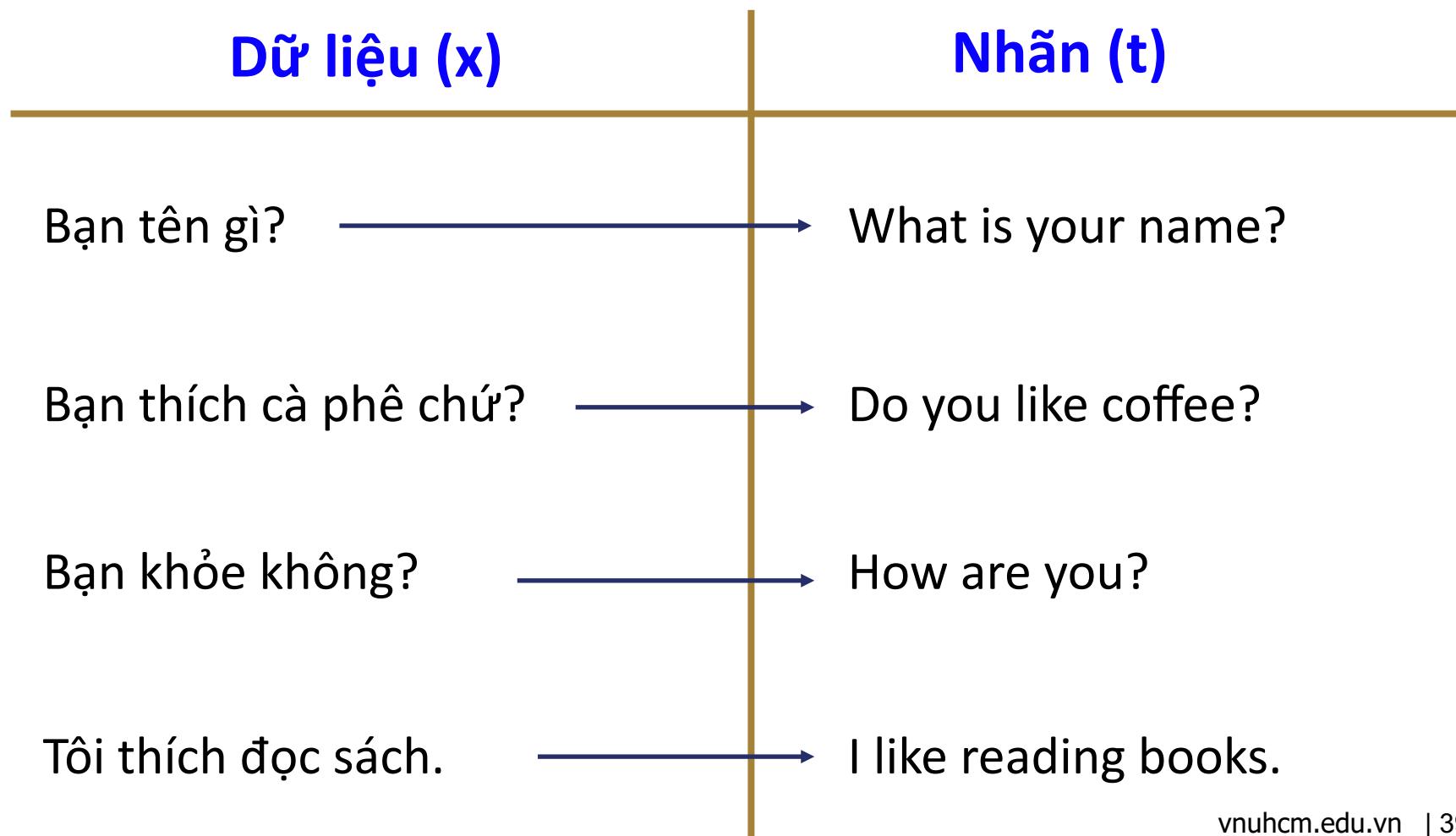
Tổ chức học tập:
giáo dục + đào tạo

Triển khai ứng dụng **5**

Tuyển dụng, sử dụng lao động

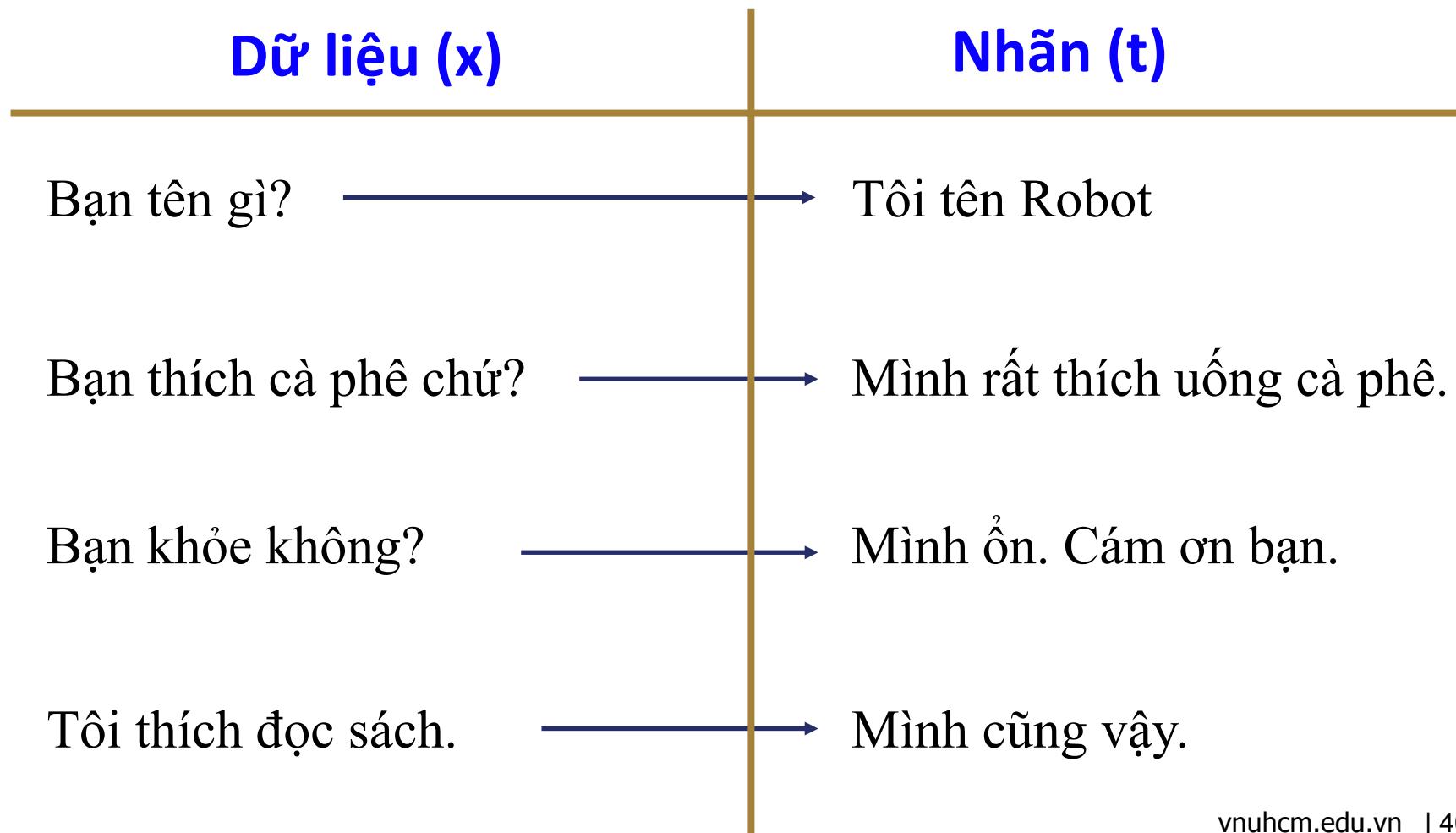
AI: Bằng cách nào?

Làm nhãn dữ liệu: Dịch máy



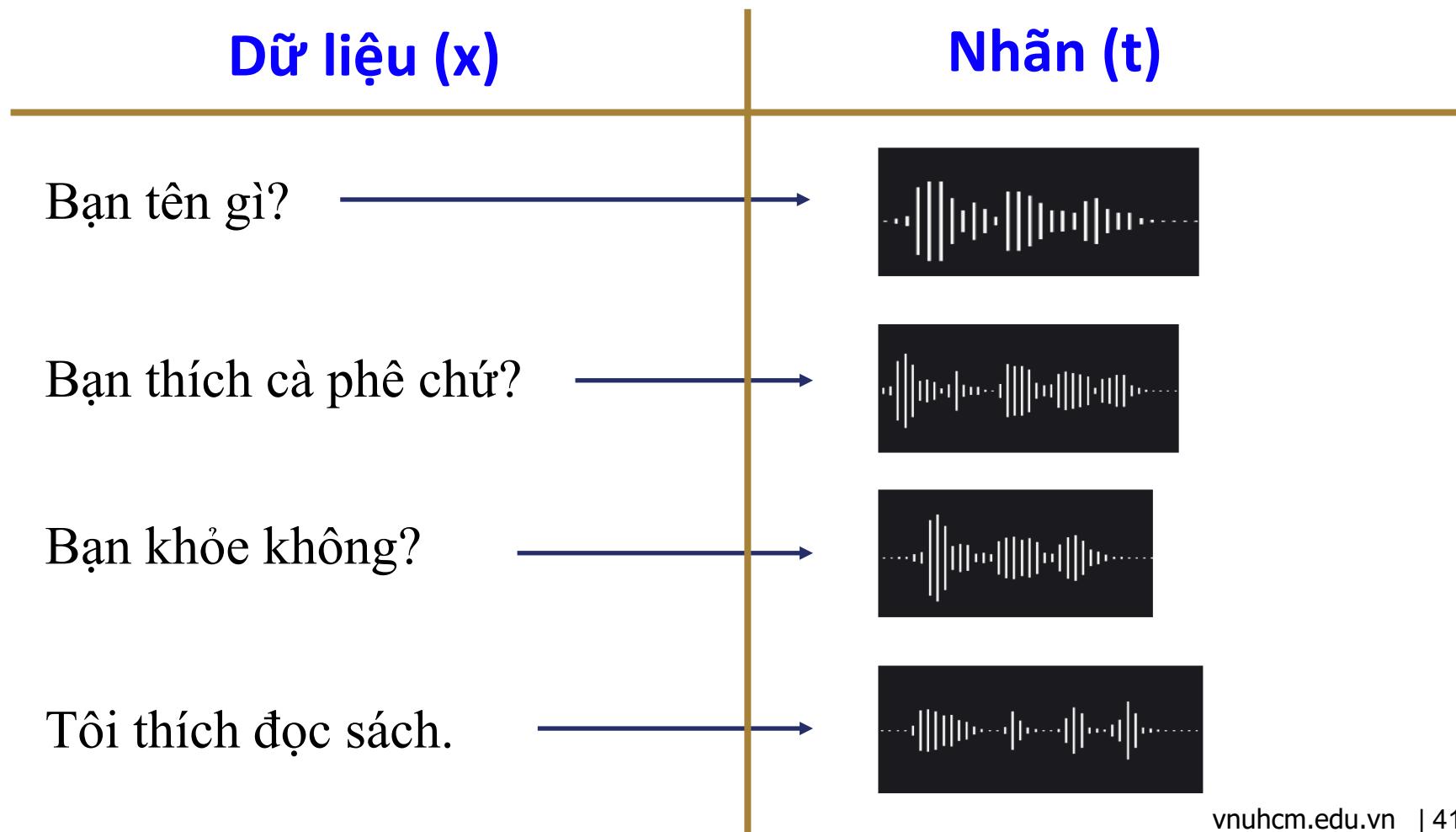
AI: Bằng cách nào?

Làm nhãn dữ liệu: Hỏi - đáp



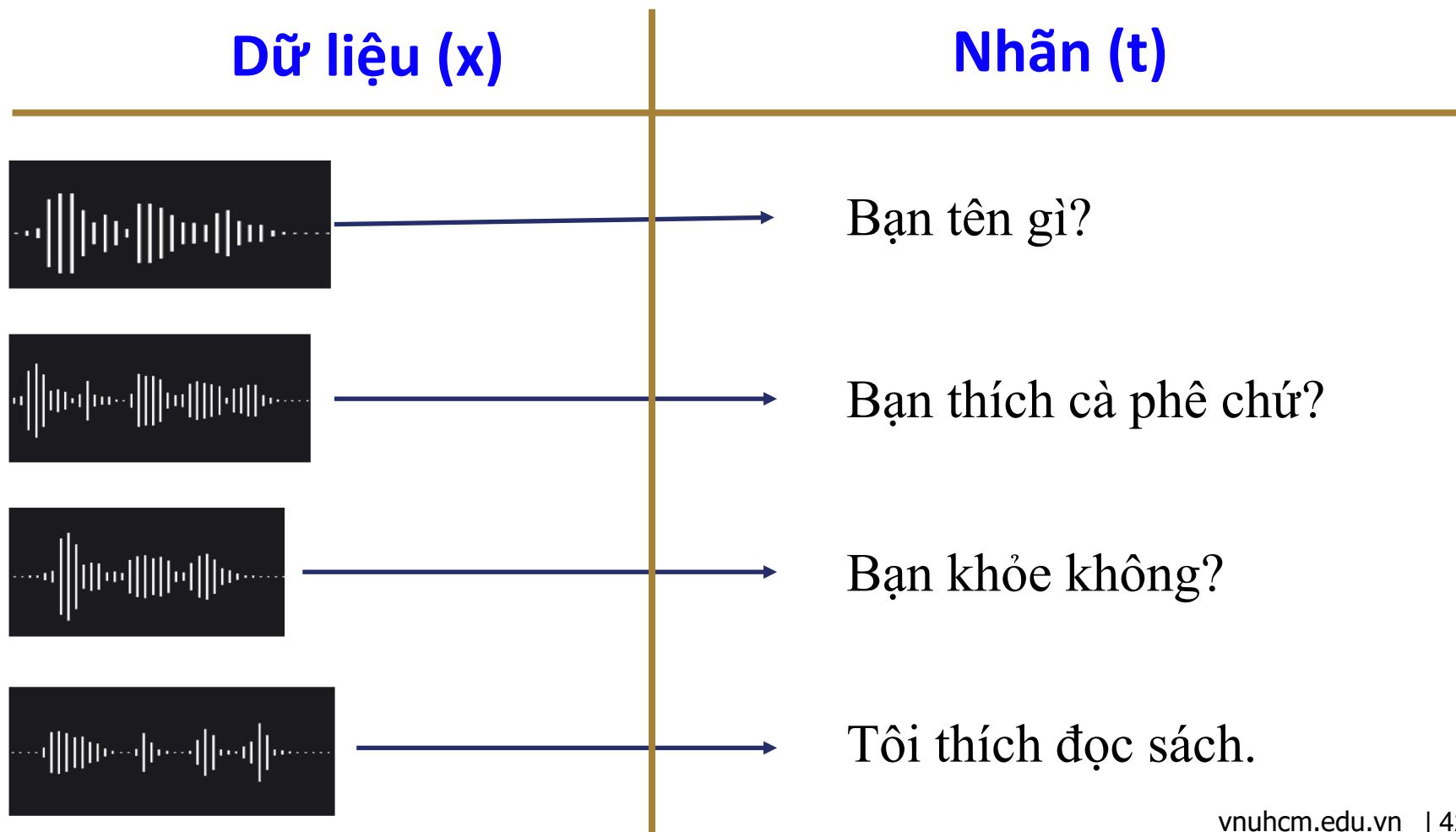
AI: Bằng cách nào?

Làm nhãn dữ liệu: Tổng hợp tiếng nói



AI: Bằng cách nào?

Làm nhãn dữ liệu: Nhận dạng tiếng nói



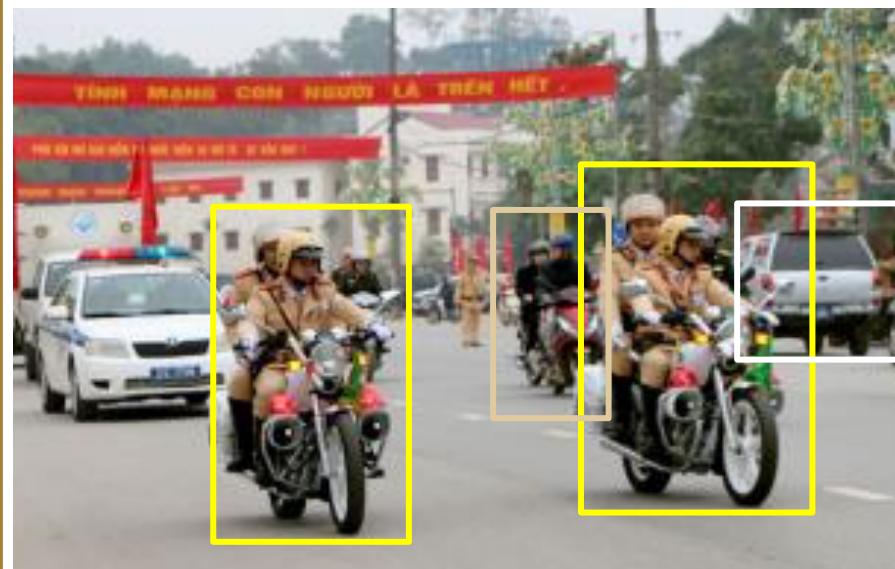
AI: Bằng cách nào?

Làm nhãn dữ liệu: Phát hiện đối tượng

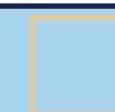
Dữ liệu (x)



Nhãn (t)



Xe hơi



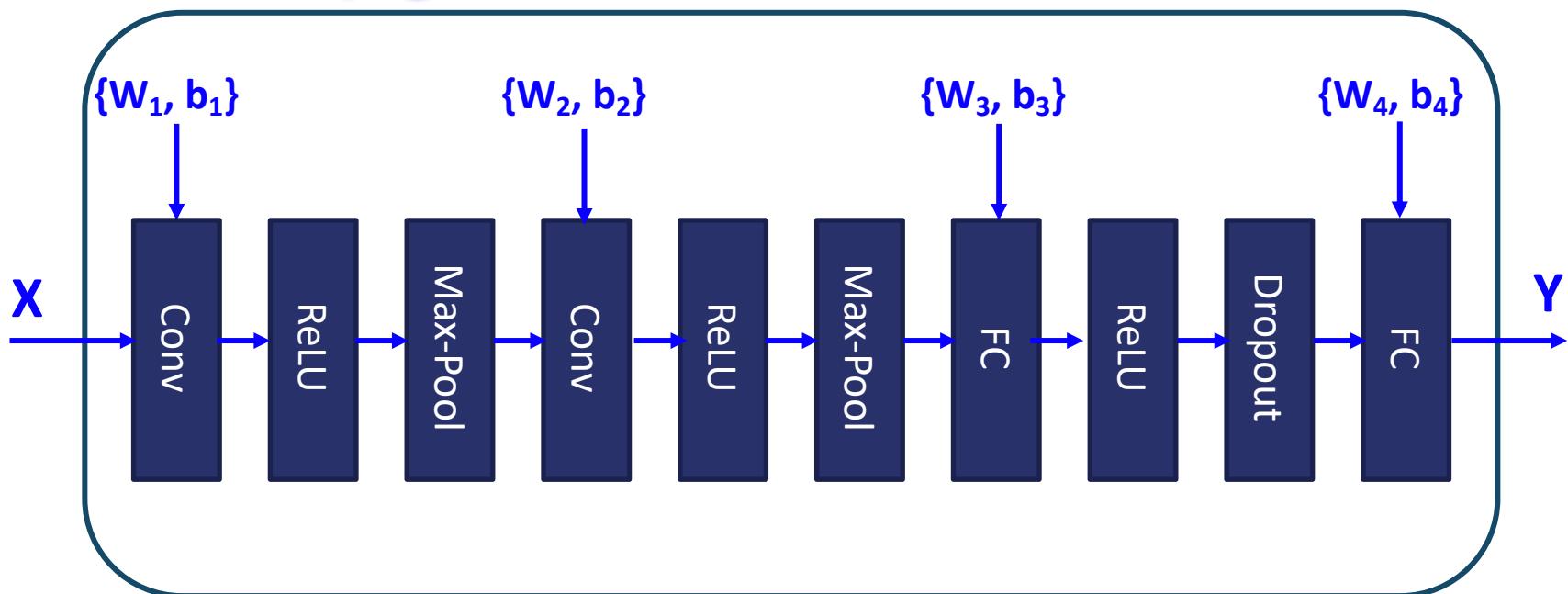
Xe máy



Môtô cảnh sát

AI: Bằng cách nào?

Thiết kế mạng nơron



Mạng nơron học sâu ≡ Đồ thị tính toán



: nút tính toán

(chỉ có chừng hơn 10 loại nút tính toán cơ bản)

→ : cạnh nối

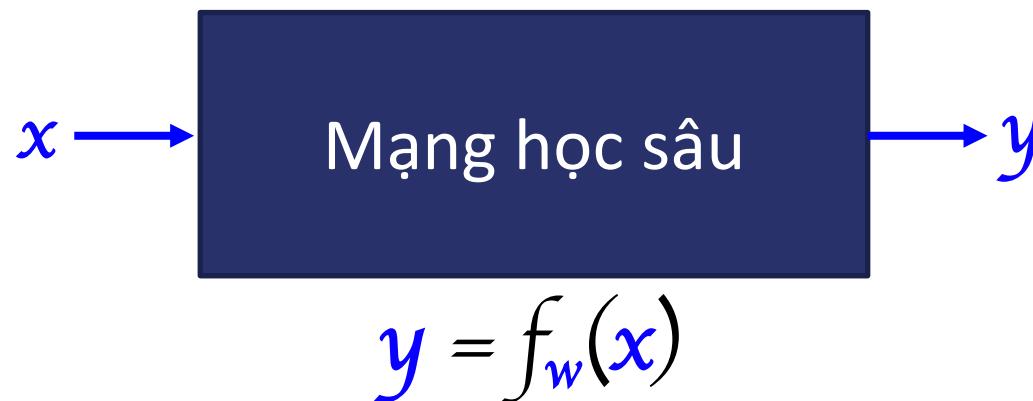
AI: Bằng cách nào?

Thiết kế mạng nơron

Mạng nơron học sâu

≡ Đồ thị tính toán

≡ Hàm toán học phụ thuộc vào thông số học



w : là tham số của mạng, được xác định lúc huấn luyện

AI: Bằng cách nào?

Thiết kế mạng nơron

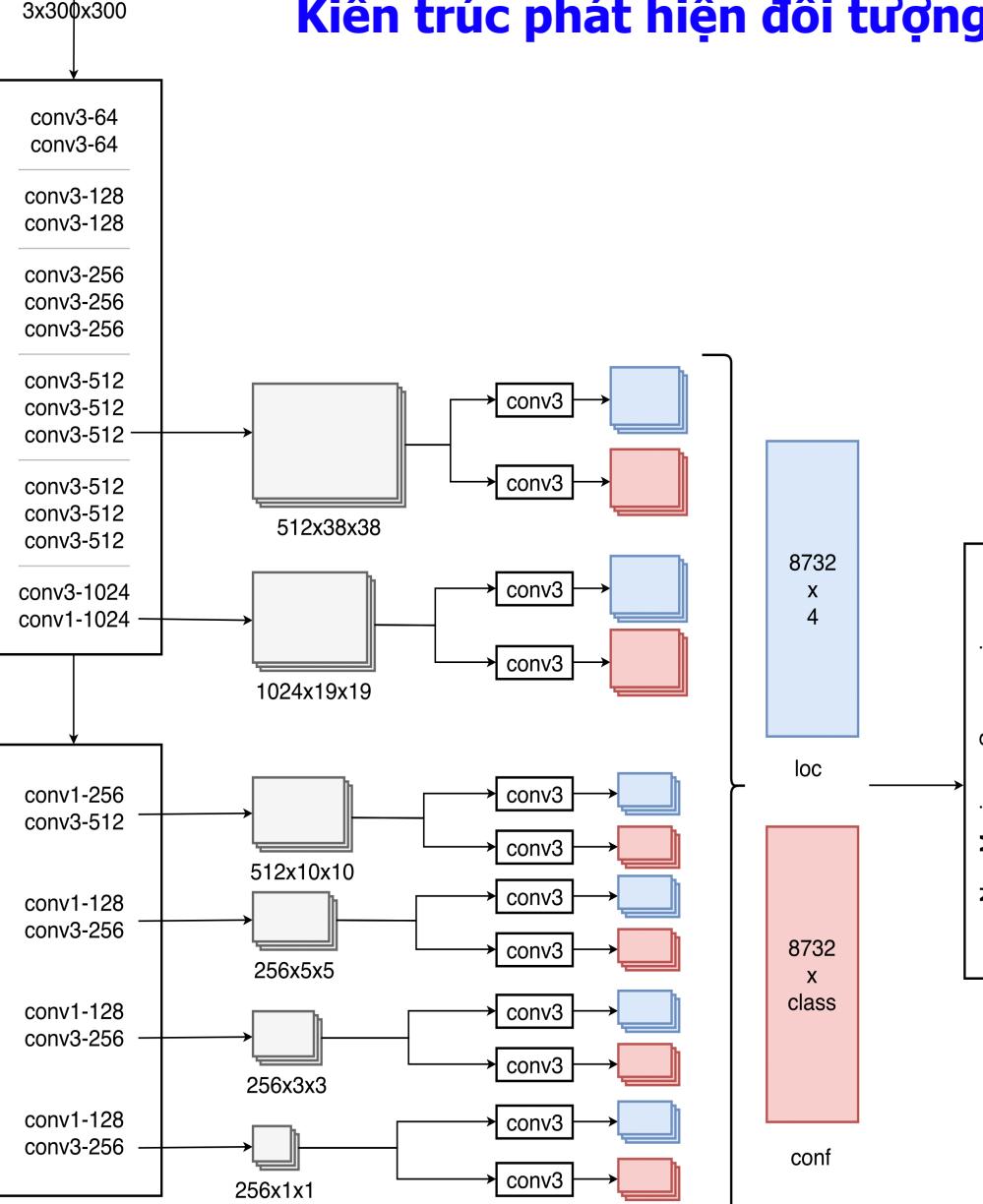
Thiết kế mạng

- ≡ Tạo ra kiến trúc tính toán
- ≡ Tạo ra kiến trúc nơron

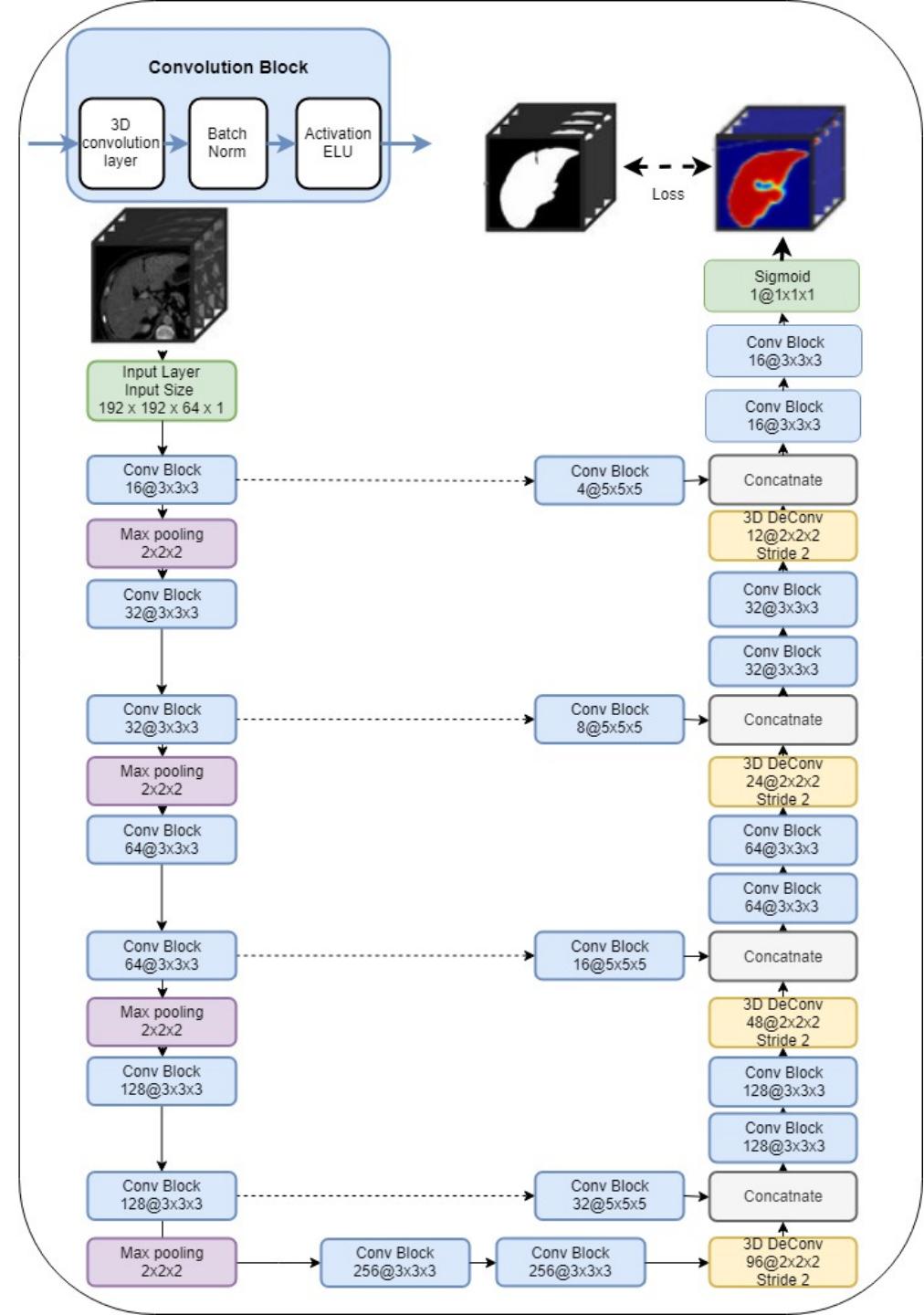
Hiện nay đang là thời kỳ bùng nổ các kiến trúc nơron

Image

Kiến trúc phát hiện đối tượng

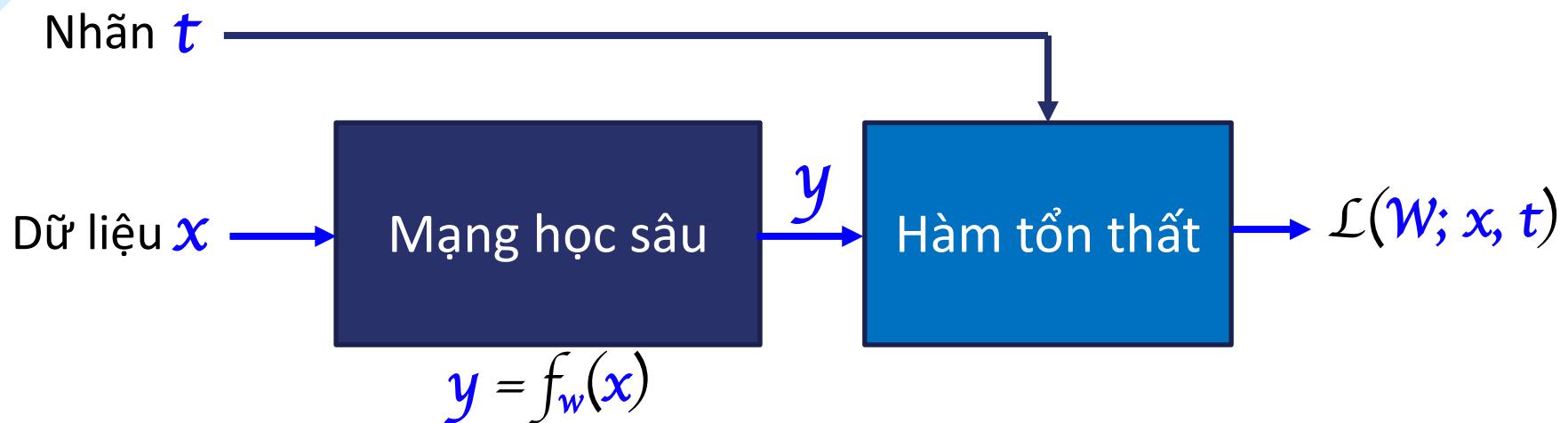


Kiến trúc phân đoạn đối tượng



AI: Bằng cách nào?

Thiết kế hàm tổn thất

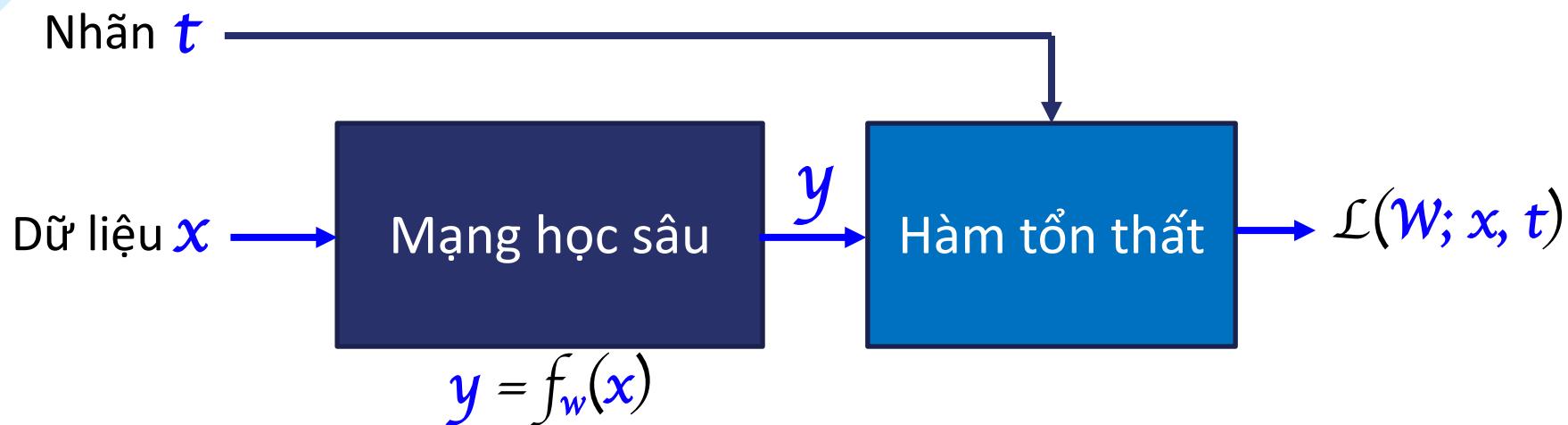


Hàm tổn thất đo lường sự khác biệt giữa
dự báo y của mạng và giá trị nhãn t

Giá trị w tối ưu \equiv Giá trị làm cho hàm $\mathcal{L}(w; x, t)$ nhỏ nhất

AI: Bằng cách nào?

Huấn luyện kiến trúc nơron



Áp dụng nguyên lý toán học (đạo hàm) để tìm giá trị w^* làm cho hàm $\mathcal{L}(w^*; x, t)$ **đủ nhỏ** trên tập dữ liệu dùng để **kiểm thử**.

AI: Bằng cách nào?

(AI theo xu hướng học sâu)

Con người

Khiếm khuyết không thể học

Học qua loa, không sâu

Học vẹt

Cần học tập chủ động

Bài khó cần học nhiều lần hơn

Trí tuệ nhân tạo

Hiện tượng không thể học
(*Vanishing gradients*)

Hiện tượng không hội tụ về điểm tối ưu
(*học nhanh, hệ số học lớn*)

Hiện tượng quá khớp
(*overfitting*)

Cần học tập chủ động

Mẫu dữ liệu khó học nhiều lần hơn

AI: Bằng cách nào?

(AI theo xu hướng học sâu)

Con người

Cần lựa chọn tư liệu học tập (sách, bài tập) hiệu quả

Phương pháp: học tập chủ động
cần tăng dần mức độ từ nhỏ → lớn

Phương pháp đánh giá
ảnh hưởng lớn chất lượng dạy/học

Khoảng cách nội dung học tập (tổ chức đào tạo) vs công việc thực tiễn
(công nghiệp)

Trí tuệ nhân tạo

Cần lựa chọn mẫu dữ liệu để huấn luyện, kiểm thử và kiểm tra

Phương pháp: học tập chủ động
cần tăng dần mức độ theo thời gian

Cần sử dụng hàm tổn thất và thưởng
phạt phù hợp

Phân phối của các tập huấn luyện, kiểm thử và kiểm tra

NỘI DUNG



Trí tuệ nhân tạo là gì?



AI: Tại sao sử dụng?



AI: Ứng dụng và tiềm năng



AI: Bằng cách nào?



AI: Nguy cơ tiềm ẩn



AI: Các thách thức



Kết luận

AI: Nguy cơ tiềm ẩn

Hệ thống AI dễ bị đánh lừa

One pixel attack for fooling deep neural networks

Jiawei Su*

Kyushu University
Japan

jiawei.su@inf.kyushu-u.ac.jp

Danilo Vasconcellos Vargas*

Kyushu University
Japan

vargas@inf.kyushu-u.ac.jp

Kouichi Sakurai

Kyushu University
Japan

sakurai@csce.kyushu-u.ac.jp

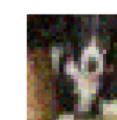
Abstract

Recent research has revealed that the output of Deep Neural Networks (DNN) can be easily altered by adding relatively small perturbations to the input vector. In this paper, we analyze an attack in an extremely limited scenario where only one pixel can be modified. For that we propose a novel method for generating one-pixel adversarial perturbations based on differential evolution. It requires less adversarial information and can fool more types of networks. The results show that 68.36% of the natural images in CIFAR-10 test dataset and 41.22% of the ImageNet (ILSVRC 2012) validation images can be perturbed to at least one target class by modifying just one pixel with 73.22% and 5.52%

AllConv



NiN



VGG



AI: Nguy cơ tiềm ẩn

Hệ thống AI dễ bị đánh lừa

One pixel attack for fooling deep neural networks

Chỉ sửa 1 pixel: độ chính xác thay đổi đáng kể
→ Nguy cơ gây tai nạn cho các hệ thống tự hành

jiawei.su@inf.kyushu-u.ac.jp

vargas@inf.kyushu-u.ac.jp

sakurai@csce.kyushu-u.ac.jp

Abstract

Recent research has revealed that the output of Deep Neural Networks (DNN) can be easily altered by adding relatively small perturbations to the input vector. In this paper, we analyze an attack in an extremely limited scenario where only one pixel can be modified. For that we propose a novel method for generating one-pixel adversarial perturbations based on differential evolution. It requires less adversarial information and can fool more types of networks. The results show that 68.36% of the natural images in CIFAR-10 test dataset and 41.22% of the ImageNet (ILSVRC 2012) validation images can be perturbed to at least one target class by modifying just one pixel with 73.22% and 5.52%

AllConv



SHIP

CAR(99.7%)

NiN



HORSE

FROG(99.9%)

VGG



DEER

AIRPLANE(85.3%)



HORSE

DOG(70.7%)



DOG

CAT(75.5%)



BIRD

FROG(86.5%)

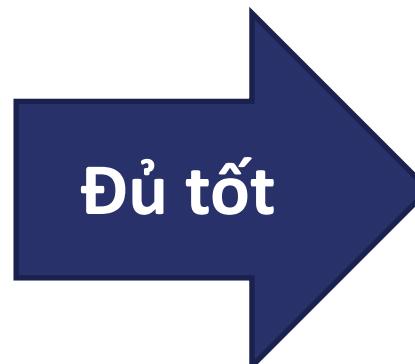
AI: Nguy cơ tiềm ẩn

Các tin tức giả mạo



AI: Nguy cơ tiềm ẩn

Tỉ lệ thất nghiệp cao



- Tự động hóa cao
- Tỉ lệ thất nghiệp cao

AI: Nguy cơ tiềm ẩn

Xung độ giữa con người vs Robots

Nếu “A.I. agent” vượt qua phép thử Turing cho tất cả các khả năng?



Vượt qua
phép thử Turing



Nguy cơ rất lớn
(ở thì tương lai)

AI: Nguy cơ tiềm ẩn

Xung độ giữa con người vs Robots



Bao giờ Ms. Sophia vượt qua được phép thử Turing?

NỘI DUNG



Trí tuệ nhân tạo là gì?



AI: Tại sao sử dụng?



AI: Ứng dụng và tiềm năng



AI: Bằng cách nào?



AI: Nguy cơ tiềm ẩn



AI: Các thách thức



Kết luận

AI: Các thách thức

1

Cần lượng lớn dữ liệu và gán nhãn dữ liệu

- Dữ liệu khó thu thập: y khoa, tài chính, an ninh, v.v.
- Gán nhãn cần chuyên gia

Đề xuất:

- Cần sự điều phối của nhà nước;
- Cần cộng tác giữa các đơn vị để chia sẻ dữ liệu;
- Cần hợp tác với các chuyên gia để gán nhãn số liệu (y khoa);
- Đầu tư xây dựng cơ sở dữ liệu có nhãn chung cho một số bài toán quan trọng;

AI: Các thách thức

2

Cần hạ tầng tính toán mạnh

- Cần hệ thống tính toán mạnh để huấn luyện mạng, đặc biệt là để chạy chương trình huấn luyện với nhiều bộ siêu tham số khác nhau.
- Cần hệ thống tính toán cho bước triển khai để nghiên cứu và làm chủ toàn bộ quy trình từ huấn luyện đến triển khai kiến trúc nơron.

Đề xuất:

- Cần đầu tư của nhà nước và các doanh nghiệp;
- Cần phối hợp để chia sẻ tài nguyên tính toán

AI: Các thách thức

3

Cần nguồn nhân lực có chuyên môn cao về AI

- Để xây dựng các nền tảng phục vụ cho chính AI, làm nhãn, chia sẻ dữ liệu, chia sẻ kiến trúc nơron, v.v.
- Để thiết kế và thử nghiệm các kiến trúc nơron

Đề xuất (các đại học):

- Tái cấu trúc chương trình và nội dung đào tạo
- Đẩy mạnh việc sử dụng AI trong nhiều chuyên ngành/bài toán khác hơn là chỉ ngành Khoa học Máy tính, Khoa học Dữ liệu và AI.

AI: Các thách thức

4

Cần sự chấp nhận của xã hội về AI

- Người dân có sẵn sàng bị AI giám sát (ví dụ, FaceID) và phân tích hành vi, và phân tích dữ liệu cá nhân?
- Doanh nghiệp có sẵn sàng đầu tư để số hóa và áp dụng công nghệ AI?
- Nhà nước có sẵn sàng đầu tư để nghiên cứu và triển khai AI, có sẵn sàng chịu rủi ro do AI mang lại?

Đề xuất:

- Nhà nước nên có chiến lược phù hợp về AI;
- Đẩy mạnh tuyên truyền một cách đúng đắn về AI.

NỘI DUNG



Trí tuệ nhân tạo là gì?



AI: Tại sao sử dụng?



AI: Ứng dụng và tiềm năng



AI: Bằng cách nào?



AI: Nguy cơ tiềm ẩn



AI: Các thách thức



Kết luận

Kết luận

Thành công hiện tại

Dữ liệu lớn
Bigdata

Công nghệ tính toán
Computing Infra-structure

Học máy
Machine Learning

Kết luận

Thành công hiện tại

Dữ liệu lớn
Bigdata

Công nghệ tính toán
Computing Infra-structure

Học máy
Machine Learning

(GPU)

(Deep learning)

Tri thức
Knowledge

Công nghệ tính toán
Computing Infra-structure

Suy luận máy
Machine Reasoning

Thành công tương lai
(kỳ vọng)

Kết luận

- Trí tuệ nhân tạo là chủ đề:
 - Có phạm vi ứng dụng rộng;
 - Có tiềm năng và có thể gây ảnh hưởng lớn trong xã hội
 - Hiện đã rất thành công trong nhiều lĩnh vực xử lý
 - Video
 - Âm thanh
 - Ngôn ngữ tự nhiên
- Nhà nước cần đầu tư và chiến lược phát triển AI phù hợp
- Tính liên ngành và tính cộng tác là điểm quan trọng để phát triển AI



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
TP. HỒ CHÍ MINH



Xin cảm ơn sự lắng nghe

Mong nhận được sự chia sẻ và góp ý!

Lê Thành Sách
LTSACH@hcmut.edu.vn

