FPT Education
FPT UNIVERSITY

TÀI LIỆU BÀI GIẢNG

Giới thiệu về Trí Tuệ Nhân Tạo (AI)



TRÌNH BÀY

Mr. Ngô Tùng Sơn (Ph.D.)

Trưởng khoa ATTT, Trường Đại học FPT, Cơ sở Hòa Lạc
Giám đốc phòng thí nghiệm ATTT, Đại học FPT

FPT Education
FPT UNIVERSITY

Mục lục

CHƯƠNG 1 — **Giới thiệu về Trí tuệ nhân tạo và tác động của nó** — 3
Vị trí của AI trong cách mạng 4.0, chuyển đổi số, AI trong tổ chức, chính sách AI

CHƯƠNG 2 — **Các khái niệm và đặc tính của Trí tuệ nhân tạo** — 11
Machine Learning, Deep Learning, GenAI và LLM, AI Agent và Agentic AI

CHƯƠNG 3 — **Những khía cạnh ngoài công nghệ khi triển khai AI** — 31
Thách thức, dữ liệu lớn, đạo đức, chi phí đầu tư, tiêu chí đánh giá

CHƯƠNG 4 — **Thực hành với một số công cụ AI** — —
Phần thực hành trên lớp

1

CHƯƠNG 1

Giới thiệu về Trí Tuệ Nhân Tạo và Tác Động của nó

Nội dung chương

1.1 Vị trí của AI trong cuộc cách mạng công nghiệp lần thứ 4 *4*

1.2 Chuyển đổi số - Phương pháp luận của cách mạng 4.0 *5*

1.3 Trí tuệ nhân tạo là gì? *6*

1.4 AI tác động như thế nào tới cuộc sống? *7*

1.5 AI trong hệ thống thông tin quản lý của tổ chức *8*

1.6 Đào tạo AI Literacy theo cấp nhân sự trong tổ chức *9*

1.7 Chính sách AI từ Nghị quyết 57-NQ/TW đến Luật AI *10*

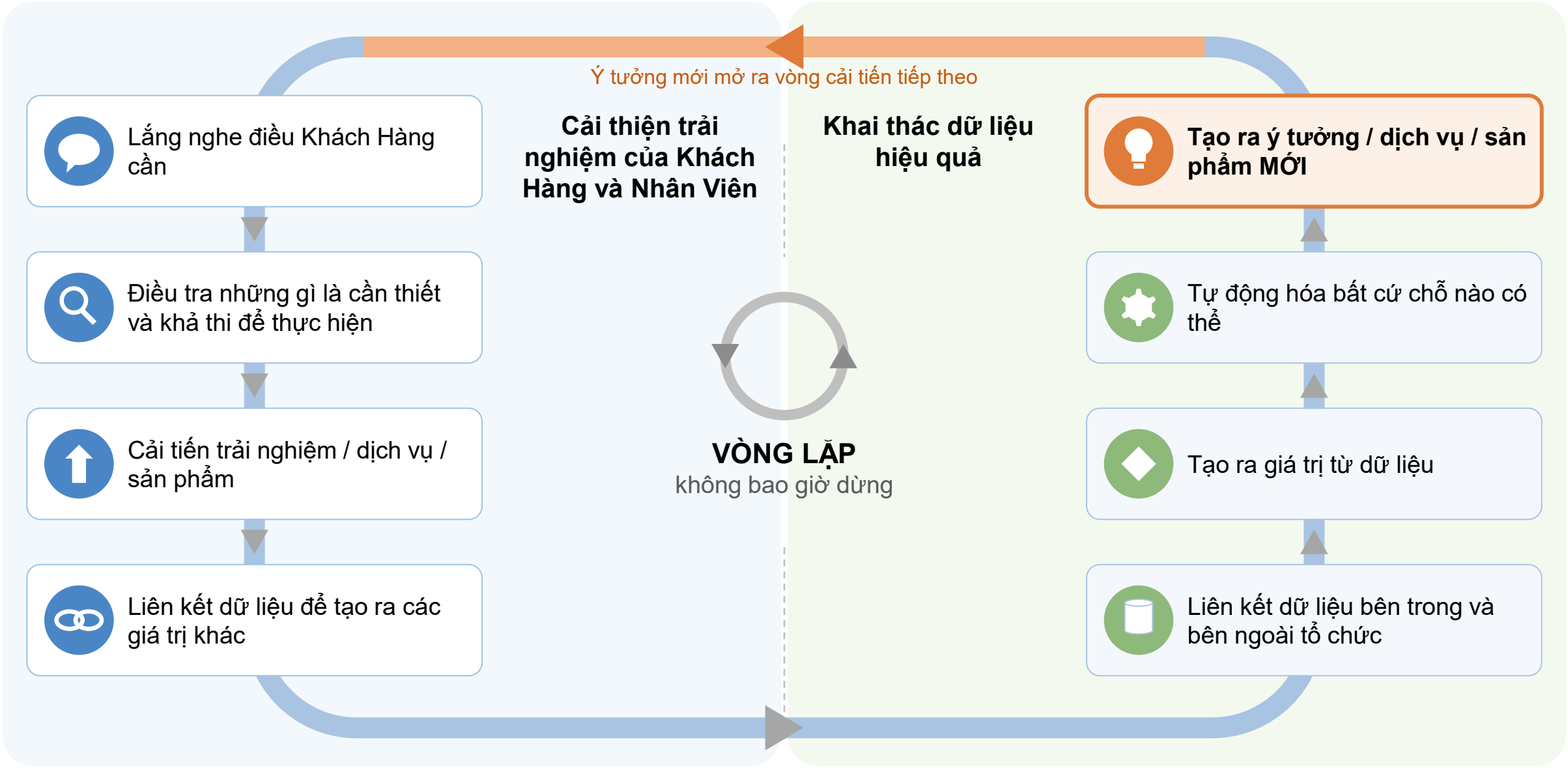
1.1 Vị trí của AI trong cuộc cách mạng công nghiệp lần thứ 4

> *Loài người không bao giờ hài lòng với những gì mình có. Chính sự không hài lòng đó đã thúc đẩy tiến bộ, nhưng cũng gây ra đau khổ.*
>
> **Yuval Noah Harari, Sapiens: Lược sử loài người**

Loài người đã và đang trải qua 4 cuộc cách mạng công nghiệp

1 Lần thứ nhất Sản xuất hàng loạt	2 Lần thứ hai Kiểm soát năng lượng	3 Lần thứ ba Thông tin và tự động hóa	4 Lần thứ tư Vượt giới hạn sinh học
Công nghệ tiêu biểu Máy hơi nước, dệt máy, luyện kim	**Công nghệ tiêu biểu** Điện, dầu mỏ, hóa chất	**Công nghệ tiêu biểu** Máy tính, internet, robot	**Công nghệ tiêu biểu** **AI, công nghệ sinh học, IoT**
Hệ quả tiêu cực chính Bóc lột lao động, ô nhiễm môi trường, đô thị hóa nhanh	**Hệ quả tiêu cực chính** Chủ nghĩa đế quốc, chiến tranh thế giới, khai thác tài nguyên	**Hệ quả tiêu cực chính** Mất việc truyền thống, phân hóa giàu nghèo, khủng hoảng dữ liệu	**Hệ quả tiêu cực chính** Mất quyền riêng tư, khủng hoảng đạo đức, nguy cơ mất kiểm soát

1.2 Chuyển Đổi Số - Phương pháp luận của cách mạng 4.0



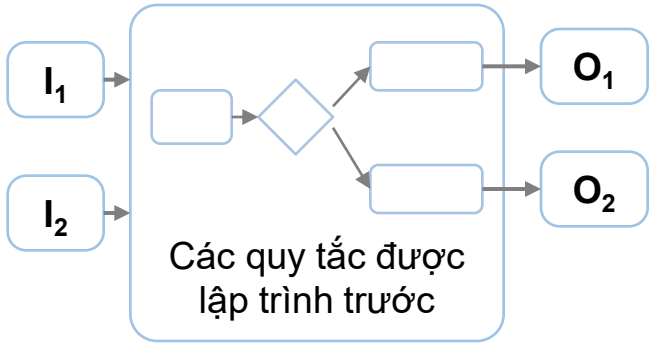
1.3 Trí tuệ nhân tạo là gì?

Phần mềm thông thường

TẤT ĐỊNH (deterministic)

$$y = f(x)$$

f: quy tắc được lập trình trước

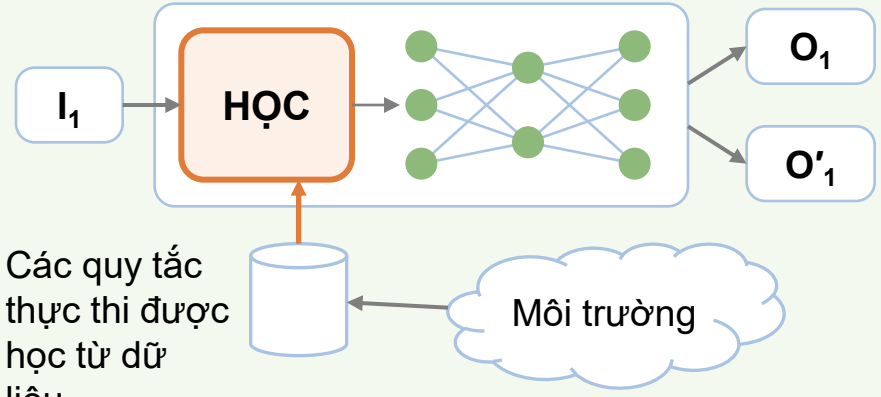


Cùng đầu vào → luôn cùng một đầu ra

Ví dụ: *f(2, 3) = 2 + 3*
→ luôn bằng 5

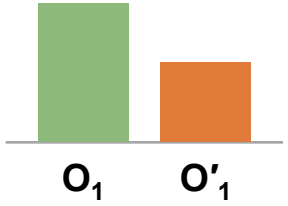
Phần mềm AI

NGẪU NHIÊN (stochastic)



$$y \sim P(y \mid x; \theta)$$

θ: tham số học từ dữ liệu



Cùng đầu vào → nhiều đầu ra có thể, mỗi đầu ra có một xác suất

Phân loại phần mềm AI

AI hẹp (narrow AI)

Đây là loại AI được thiết kế để thực hiện một nhiệm vụ cụ thể. Nó không thể thực hiện được các nhiệm vụ khác ngoài phạm vi đã được lập trình. Ví dụ:

- Theo dõi giao thông
- Hỗ trợ phát âm
- Chơi cờ

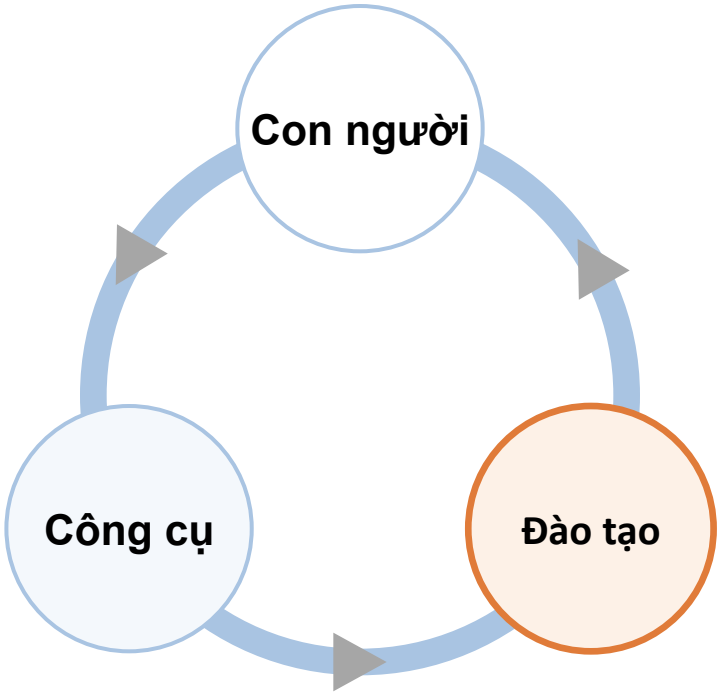
AI tổng quát (AGI)

Đây là loại AI có khả năng thực hiện bất kỳ nhiệm vụ trí tuệ nào mà con người có thể làm, từ suy luận đến sáng tạo.

Loại AI này chưa thực sự được hoàn thiện

1.4 AI Tác động như thế nào tới cuộc sống?





1.5 AI trong Hệ Thống Thông Tin Quản Lý của Tổ Chức

1 Tự động hóa

Giảm thiểu sức lao động của con người

- Robotic process automation
- Xử lý tài liệu tự động (OCR, computer vision, NLP)
- Trợ lý ảo (Chatbot, NLP)
- Nâng cao trải nghiệm người dùng

3 Triển khai Business Intelligence

- Xây dựng các báo cáo thống kê
- Xây dựng mô hình AI (dự đoán, giải thích)
- Định nghĩa mô hình dữ liệu phù hợp với yêu cầu kinh doanh
- Đảm bảo an toàn dữ liệu



2 Tối ưu hóa

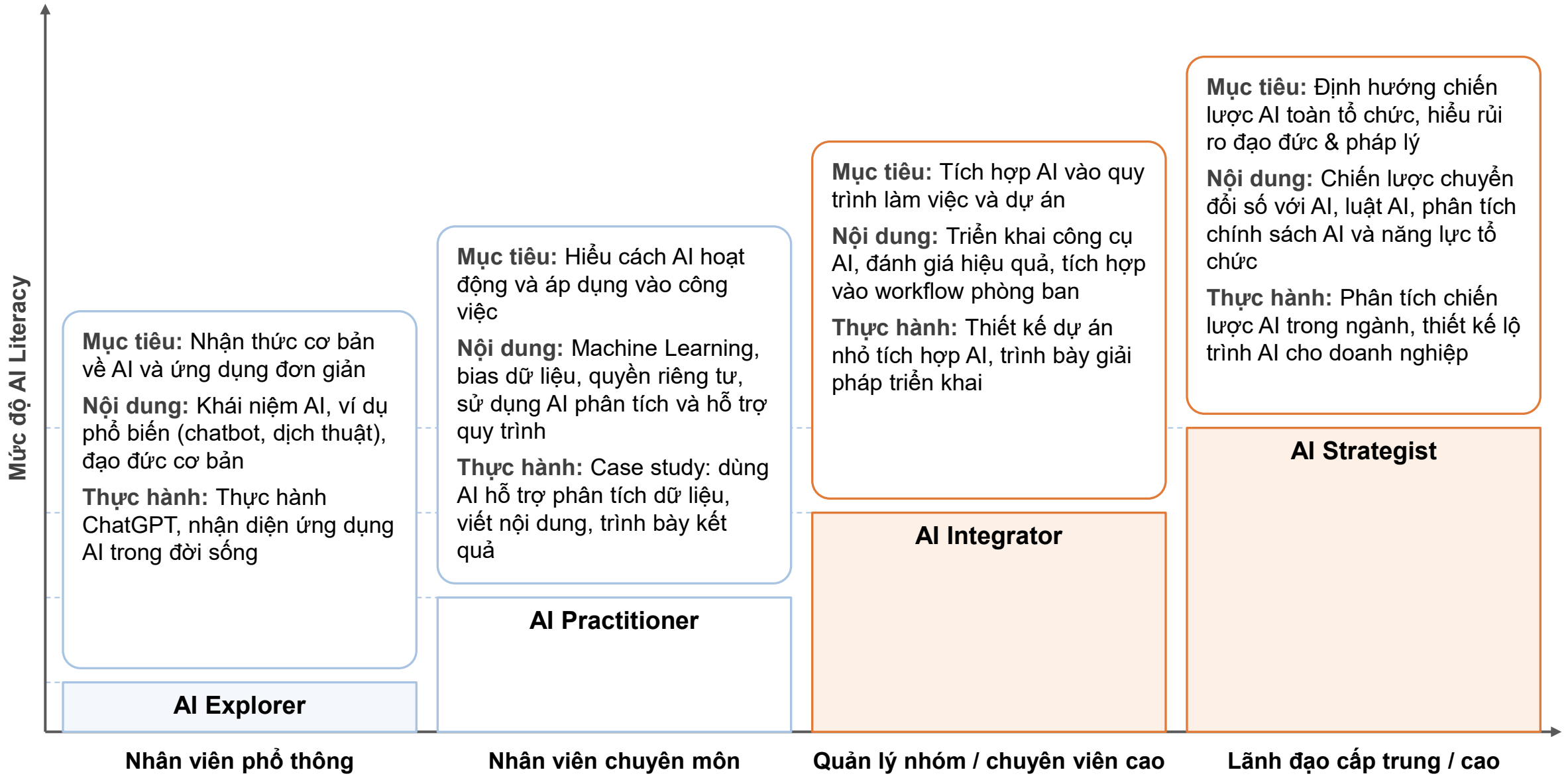
Vận hành hiệu quả, cắt giảm chi phí

- Lập lịch tự động

4 Tự động hóa giám sát vận hành

- Theo dõi lưu lượng mạng và cảnh báo thâm nhập tự động
- Tự động phục hồi sau khi bị tấn công

1.6 Đào tạo AI Literacy theo cấp nhân sự trong tổ chức



1.7 Chính sách AI từ Nghị quyết 57-NQ/TW đến Luật AI

12/2024	02/2025	04/2025	12/2025	08/2026
NQ 57-NQ/TW Bộ Chính trị	**NQ 193/2025/QH15** Quốc hội	**NĐ 88/2025/NĐ-CP** Chính phủ	**Luật Trí tuệ nhân tạo** Luật 134/2025/QH15	**QĐ 1671/QĐ-TTg** Chiến lược AI quốc gia

1 Định hướng của Nghị quyết 57-NQ/TW

Mục tiêu 2030: Việt Nam thuộc top 3 Đông Nam Á về nghiên cứu và phát triển AI

- **Lĩnh vực ưu tiên:** Khuyến khích nghiên cứu, ứng dụng AI trong y tế, giáo dục, nông nghiệp, công nghiệp, hành chính công
- **Hệ sinh thái AI quốc gia:** Trung tâm nghiên cứu, doanh nghiệp khởi nghiệp, quỹ đầu tư, đào tạo chuyên sâu
- **Bảo vệ dữ liệu:** Yêu cầu cao về an ninh thông tin, dữ liệu cá nhân và dữ liệu quốc gia

2 Cơ chế thí điểm đặc biệt

NQ 193/2025/QH15, hướng dẫn bởi NĐ 88/2025/NĐ-CP

- **Doanh nghiệp từ kết quả nghiên cứu:** Viện, trường được thành lập doanh nghiệp; viên chức được tham gia quản lý, điều hành
- **Chấp nhận rủi ro:** Miễn trách nhiệm khi nghiên cứu KH&CN gây thiệt hại nếu đã tuân thủ đúng quy trình, quy định
- **Ưu đãi thuế:** Chi R&D của doanh nghiệp được trừ khi tính thuế TNDN; miễn thuế TNCN cho tiền lương nhiệm vụ KH&CN dùng ngân sách
- **Thu hút:** Đầu tư và nhân lực quốc tế

3 Luật Trí tuệ nhân tạo (134/2025/QH15) · thông qua 10/12/2025, hiệu lực từ 01/3/2026

- Quản lý theo mức độ rủi ro
- Quỹ Phát triển AI quốc gia
- AI Voucher hỗ trợ doanh nghiệp
- Sandbox thử nghiệm

Nguồn: NQ 57-NQ/TW (22/12/2024); NQ 193/2025/QH15 (19/02/2025); NĐ 88/2025/NĐ-CP (13/4/2025); Luật số 134/2025/QH15; QĐ 1671/QĐ-TTg.

2

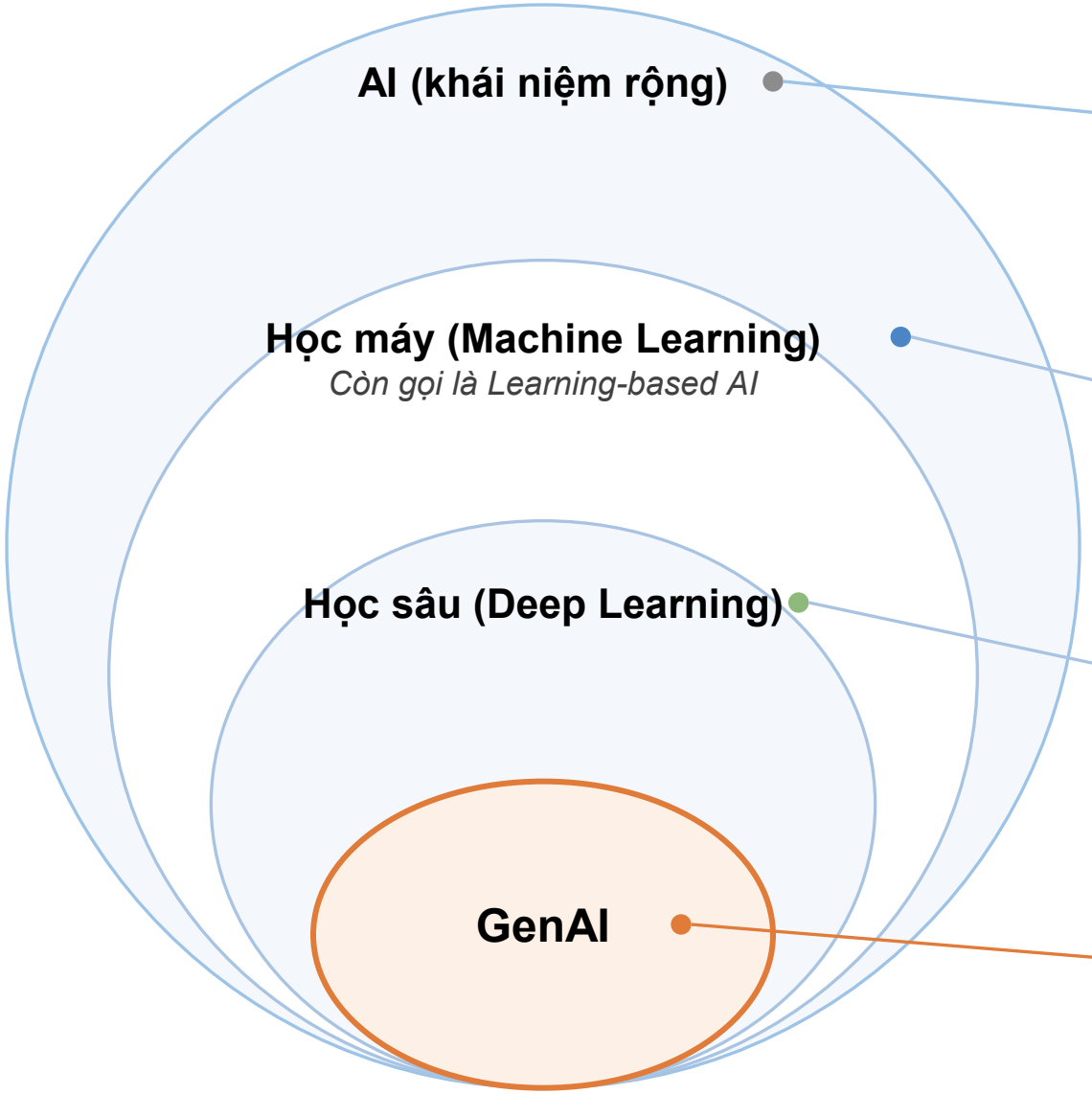
CHƯƠNG 2

Các khái niệm và đặc tính của Trí Tuệ Nhân Tạo

Nội dung chương

2.1	Các khái niệm trong AI và nơi chúng được ứng dụng	*12*
2.2	Một số khái niệm cổ điển trong Trí tuệ nhân tạo	*13*
2.3	Machine Learning – Giai đoạn Training và Prediction	*14*
2.4	Mô hình AI là gì? – Bài toán Hồi quy	*15*
2.5	Một số thuật ngữ liên quan AI Model	*16*
2.6	Mô hình AI là gì? – Bài toán Phân loại	*17*
2.7	Mô hình AI là gì? – Bài toán Phân cụm	*18*
2.8	Điểm yếu và điểm mạnh của Machine Learning	*19*
2.9	Deep Learning là gì?	*20*
2.10	Phân loại các kỹ thuật Deep Learning và ứng dụng	*21*
2.11	Deep Learning và Thị giác máy tính	*22*
2.12	Deep Learning và Xử lý ngôn ngữ tự nhiên	*23*
2.13	Tại sao chương trình AI đòi hỏi nhiều GPU	*24*
2.14	GenAI hay LLM là gì?	*25*
2.15	Các hướng tiếp cận khai thác LLM	*26*
2.16	Nhận thức và phản biện khi sử dụng LLM	*27*
2.17	AI Agent (Tác tử AI)	*28*
2.18	Từ AI Agent đến Agentic AI	*29*
2.19	Các bước triển khai Agentic Systems	*30*

2.1 Các khái niệm trong AI? Nơi chúng có thể được ứng dụng.



AI (khái niệm rộng)

Hệ thống suy luận và lập luận dựa trên quy tắc hoặc thống kê.
Tìm kiếm & tối ưu hóa, kế hoạch và lập lịch, hệ chuyên gia.

Học máy

Dữ liệu vận hành được số hóa và lưu trữ vào các cơ sở dữ liệu có cấu trúc.

Học sâu

Dữ liệu vận hành không có cấu trúc như: ảnh, video, âm thanh, văn bản.

GenAI

Tạo sinh ra các nội dung (không có sẵn) dựa trên một đầu vào (prompt).

(Có vẻ gần nhất với AI Tổng quát)

2.2 Một số khái niệm cổ điển trong lĩnh vực Trí tuệ nhân tạo

AI là lĩnh vực nghiên cứu nhằm tạo ra các hệ thống có khả năng thực hiện các nhiệm vụ đòi hỏi trí thông minh con người: học hỏi, suy luận và ra quyết định.

Các kỹ thuật AI cổ điển

1 Biểu diễn kiến thức *(Knowledge Representation)*

Dùng frame, semantic networks hoặc logic bậc nhất để lưu trữ và thao tác kiến thức; AI suy luận từ các quy tắc định sẵn, ví dụ hệ chuyên gia (expert systems).

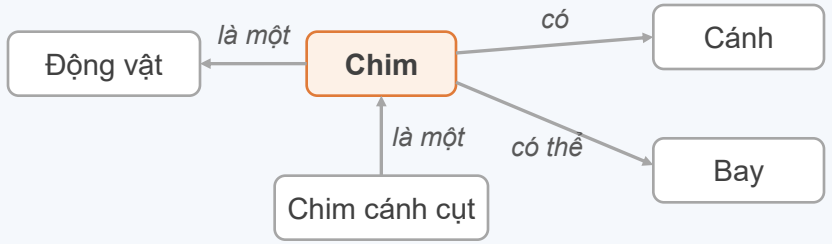
2 Thuật toán tìm kiếm *(Search Algorithms)*

A* search, depth-first search để khám phá không gian vấn đề; áp dụng trong cờ vua hoặc lập kế hoạch.

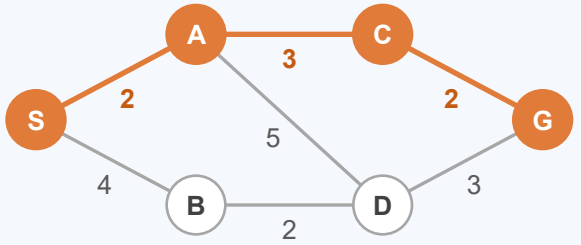
3 Suy luận logic *(Logical Reasoning)*

Lập trình logic (như Prolog) để suy diễn từ các tiền đề, giải quyết vấn đề phức tạp mà không cần học từ dữ liệu lớn.

1 Biểu diễn kiến thức: mạng ngữ nghĩa



2 Tìm kiếm: đường đi ngắn nhất trên đồ thị



Đường ngắn nhất
S→A→C→G
chi phí = 7

3 Suy luận logic: suy diễn từ tiền đề

Quy tắc: **chet(X) :- nguoi(X).**

Sự kiện: **nguoi(socrates).**

→ Kết luận **chet(socrates)**

Truy vấn **?- chet(socrates).** → **true** *(không cần dữ liệu huấn luyện)*

2.3 Machine Learning – Giai đoạn Training và Prediction

1 Giai đoạn huấn luyện AI (Training)

Dữ liệu đã gán nhãn
(Training Data)

→ **Biến đổi dữ liệu**

Chuyển dữ liệu gốc sang không gian véc tơ dựa trên các đặc trưng

→ **Thuật toán Machine Learning**
- Học có giám sát (supervised)
- Học không giám sát (unsupervised)
- Học tăng cường (reinforcement)

Tốn nhiều thời gian thực thi, cần phần cứng chuyên dụng

Dữ liệu đã gán nhãn sinh ra từ:
- Hệ thống vận hành nội bộ
- Sổ tay của chuyên viên
- Mua từ bên ngoài

Có cấu trúc rõ ràng, xác định được các đặc trưng quan trọng

2 Giai đoạn sử dụng AI để dự đoán (Prediction)

Dữ liệu mới
CHƯA gán nhãn

→ **Biến đổi dữ liệu**

→ **Ứng dụng Machine Learning**

Thuật toán dự đoán ← **Mô hình AI**

→ **Nhãn (kết quả)**

Tốn ít thời gian tính toán, phần cứng yêu cầu ít hơn

Dữ liệu CHƯA gán nhãn đến từ:
- Ứng dụng
- Con người, khi thực hiện nghiệp vụ

2.4 Machine Learning - Mô hình AI là gì? – Bài toán Hồi Quy

Ví dụ: AI dự đoán giá nhà từ diện tích, không được cho biết trước giá trung bình mỗi m²

1 Dữ liệu huấn luyện

Diện tích (m²)	Giá nhà (triệu đồng)
106	2628.705
192	4310.099
159	3689.176
139	3081.243
73	1938.033
73	1995.711
58	1807.789
179	4028.173
...	....
...	....

Trích từ file Excel: diện tích và giá đã biết

↓

Thuật toán học máy: **Hồi quy tuyến tính**

↓

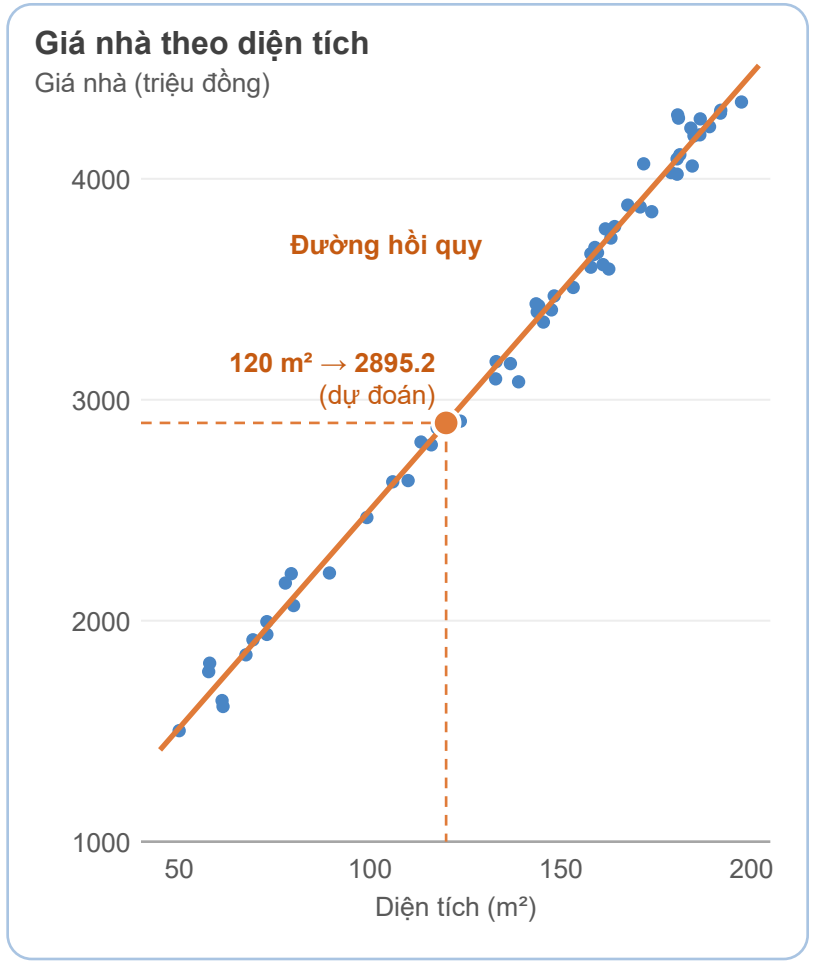
Mô hình AI với 2 tham số

Giá = 19.73 × Diện tích + 527.57

a = 19.73: mỗi m² thêm 19.73 triệu đồng

b = 527.57: phần giá không phụ thuộc diện tích

2 Mô hình học được từ dữ liệu



3 Dự đoán cho dữ liệu mới

Diện tích (m²)	Giá dự đoán (triệu đồng)
120	2895.2
164	3763.3
...	...
...	...

Cách AI tính cho nhà 120 m²

19.73 × 120 + 527.57

≈ 2895.2 triệu đồng (≈ 2.9 tỷ)

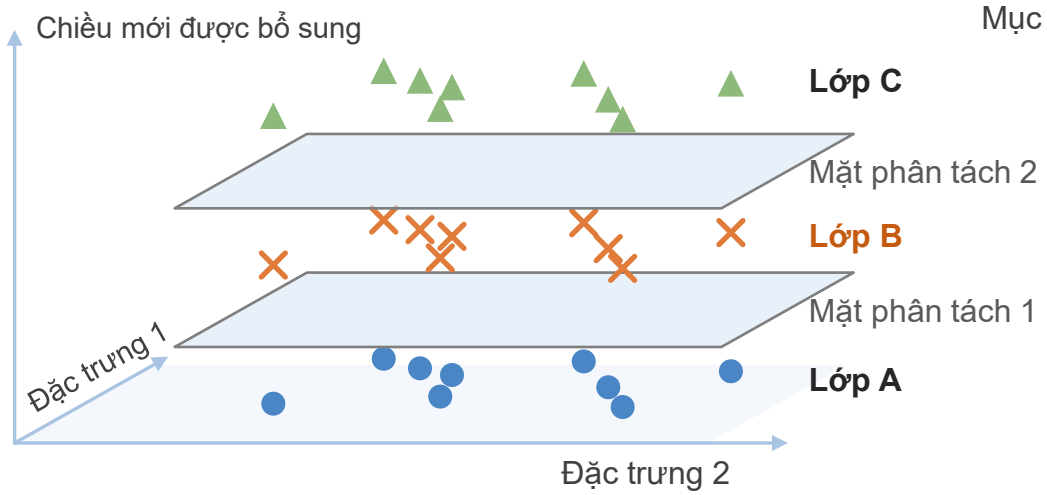
Giới hạn của mô hình

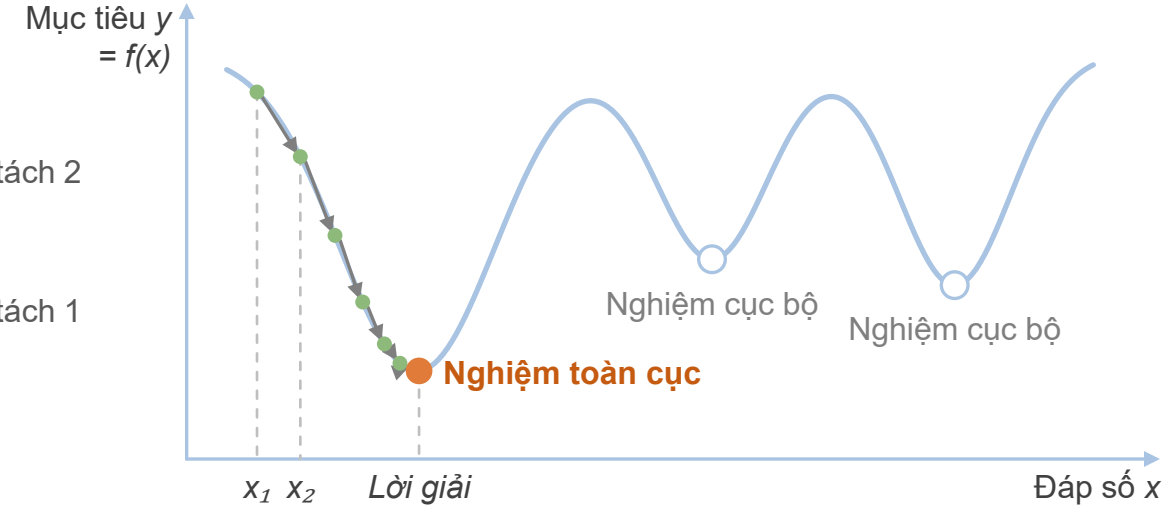
AI chỉ biết những gì có trong dữ liệu huấn luyện. Nó không tự hỏi môi giới hay tra giá thị trường.

Dự đoán đáng tin nhất trong khoảng diện tích đã học (khoảng 50–200 m²).

2.5 Machine Learning – Một số thuật ngữ liên quan AI Model

Parameters	Hyper Parameters
Được tính toán trong quá trình đào tạo với dữ liệu lịch sử	Các giá trị được thiết lập trước bằng tay ☹.
Nó là một phần của mô hình AI.	Không phải là một phần của mô hình AI.
Các giá trị của parameters sẽ được lưu lại trong quá trình đào tạo mô hình AI.	Không phải là một phần của mô hình AI khi được đào tạo và các giá trị không được lưu trữ.
Phụ thuộc vào bộ dữ liệu được sử dụng để huấn luyện AI	Độc lập với bộ dữ liệu huấn luyện.





2.6 Mô hình AI là gì? – Bài toán Phân Loại Dữ Liệu (Classification)

Ví dụ: AI dự đoán chất lượng cà phê từ thời gian rang và nhiệt độ rang

1 Dữ liệu huấn luyện

Thời gian rang (phút)	Nhiệt độ rang (°C)	Chất lượng
15.46119	199.3392	2
9.463318	166.5655	0
10.36882	178.8619	0
9.071868	181.5039	1
12.01535	189.5927	1
7.09744	144.4934	0
8.101292	174.0279	0
15.78594	195.1239	2
6.741174	178.6577	0
11.77232	183.4176	1
...	...	...

Chất lượng: ● 0 = Chưa tới ● 1 = Đạt ● 2 = Quá lửa

↓

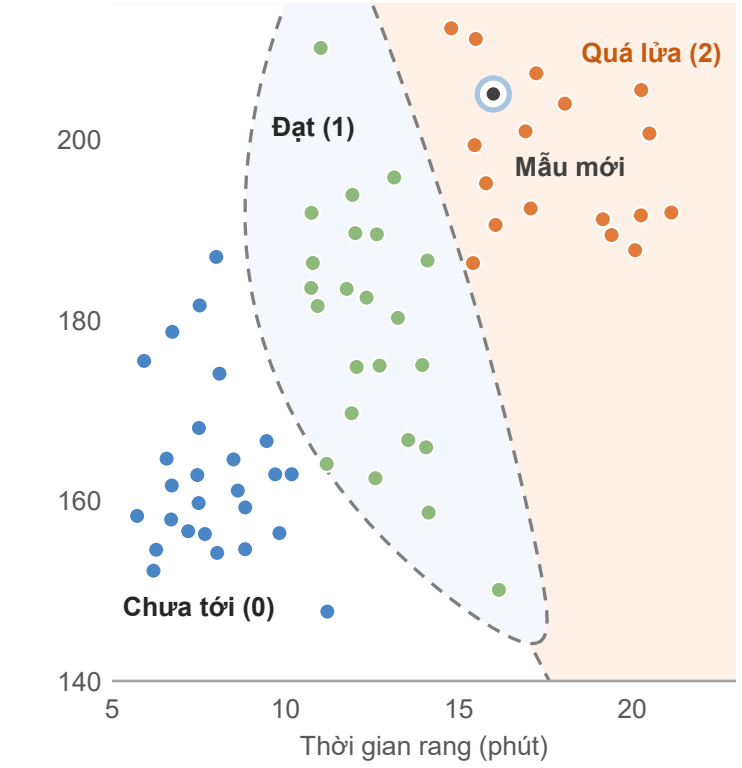
Thuật toán phân loại: **SVM**

↓

Mô hình AI: ranh giới giữa 3 lớp
Mỗi vùng màu ứng với một mức chất lượng.
Mẫu mới rơi vào vùng nào thì được xếp vào lớp đó.

2 Mô hình học ranh giới giữa các lớp

Chất lượng cà phê theo thời gian và nhiệt độ rang



3 Dự đoán cho dữ liệu mới

Mẫu cà phê mới
Thời gian rang: **16 phút**
Nhiệt độ rang: **205 °C**
Dự đoán: Quá lửa (lớp 2)
Điểm này nằm trong vùng cam trên biểu đồ.

Khác với bài toán hồi quy

Hồi quy: dự đoán một con số (giá nhà).

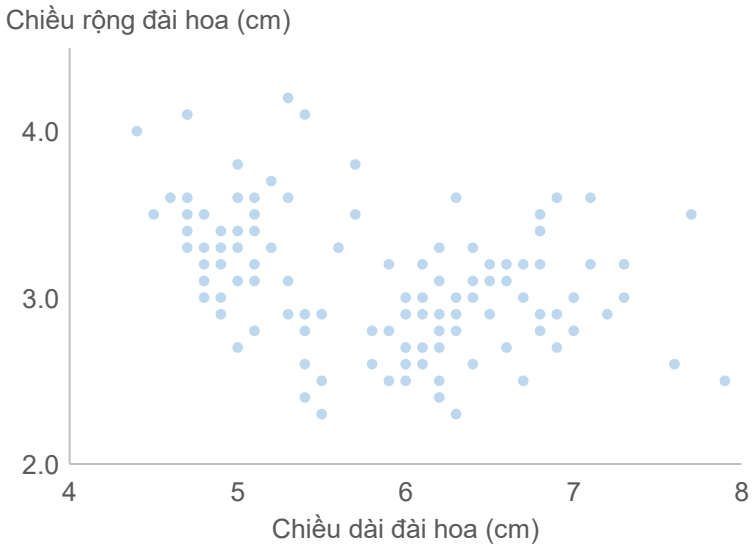
Phân loại: dự đoán một nhóm (lớp 0, 1 hoặc 2).

Ranh giới giữa các lớp càng phức tạp thì mô hình càng cần nhiều tham số.

2.7 Mô hình AI là gì? – Bài toán Phân Cụm Dữ Liệu

Ví dụ: Tìm ra các nhóm tự nhiên của dữ liệu mà không được dạy trước.

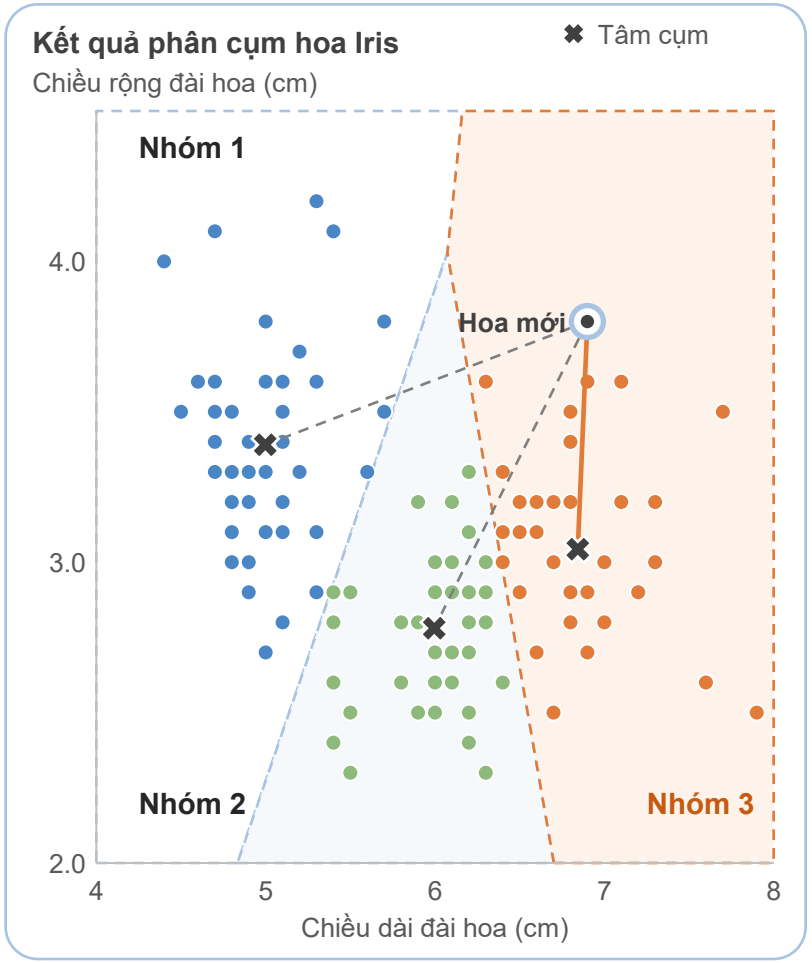
1 Dữ liệu chưa có nhãn



Thuật toán phân cụm: **K-means (k = 3)**

Mô hình AI: 3 tâm cụm
Thuật toán tự gom các điểm gần nhau thành 3 nhóm. Mỗi nhóm có một tâm cụm (×) và một vùng đa giác chứa mọi điểm gần tâm đó nhất.

2 Mô hình tự tìm ra 3 nhóm



3 Xếp nhóm cho dữ liệu mới

Một bông hoa mới
Chiều dài đài hoa: **6.9 cm**
Chiều rộng đài hoa: **3.8 cm**

Khoảng cách tới tâm cụm:
Nhóm 1: 1.9 · Nhóm 2: 1.4 · Nhóm 3: 0.8

Kết quả: Nhóm 3
Hoa được xếp vào nhóm có tâm cụm gần nhất.

Khác với bài toán phân loại

Phân loại: dữ liệu huấn luyện có nhãn sẵn (Chưa tới, Đạt, Quá lửa).

Phân cụm: không có nhãn; AI tự tìm các nhóm.

Ý nghĩa của từng nhóm do con người đặt sau khi xem kết quả.

2.8 Điểm yếu và điểm mạnh của Machine Learning

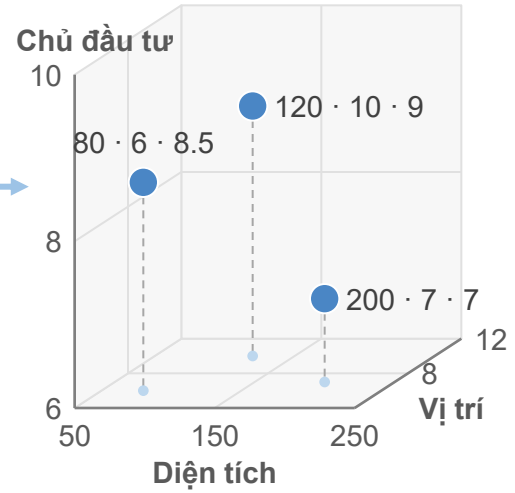
1 Dữ liệu gốc từ hệ thống thông tin

Diện tích	Vị trí	Chủ đầu tư	Đơn giá
120	Quận 1, TPHCM	Vinhomes	18 tỷ
80	Quận 7, TP HCM	Novaland	6.5 tỷ
200	Hà Đông, Hà nội	Hải Phát	9 tỷ
…	…	…	…

2 Đưa mọi thứ về dạng số

Diện tích	Vị trí	Chủ đầu tư
120	10	9
80	6	8.5
200	7	7
…	…	…

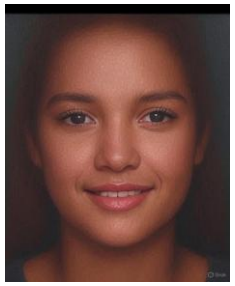
3 Trực quan hóa



Dữ liệu có cấu trúc: đặc trưng thường được xác định dựa trên yếu tố nghiệp vụ. Nhiều kỹ thuật đơn giản nhưng hiệu quả có thể áp dụng: mã hóa one-hot, chuẩn hóa min-max, kết hợp, đánh trọng số…

Nhưng với những dữ liệu phi cấu trúc thì sao?

Hình ảnh



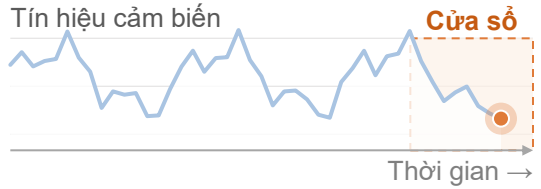
Cô gái này tên là "Linh". Lần tới gặp cô ấy bạn sẽ dễ dàng nhận ra cô ấy.
Nhưng mô tả thế nào? Hay nói cách khác là các đặc trưng gì của Linh giúp máy tính cũng có thể nhận ra cô ấy?

Văn bản

Trăm năm trong cõi người ta,
Chữ tài chữ mệnh khéo là ghét nhau.
Nguyễn Du, Truyện Kiều

Bài thơ có đặc trưng gì?

Dữ liệu dạng luồng

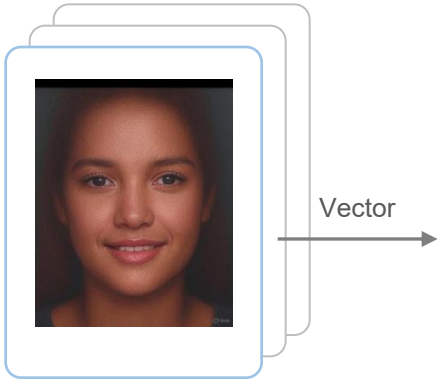


Đặc trưng của dữ liệu dạng luồng là gì?

2.9 Deep Learning là gì?

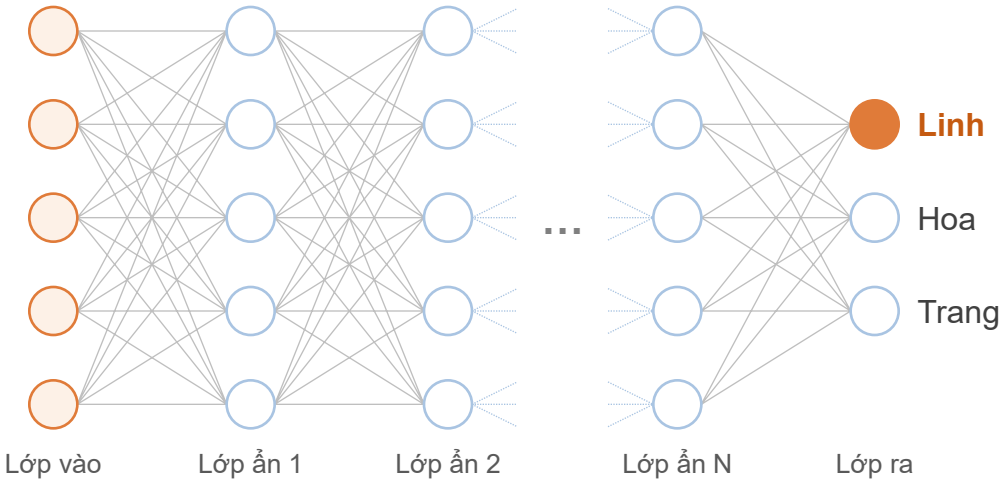
Deep Learning là một nhánh con của Machine Learning, cũng gồm hai giai đoạn huấn luyện (Training) và dự đoán (Prediction). **Điểm khác biệt:** mạng tự trích xuất đặc trưng từ dữ liệu, ít phụ thuộc kiến thức chuyên môn của con người, nên xử lý tốt dữ liệu phi cấu trúc.

1 Dữ liệu đầu vào

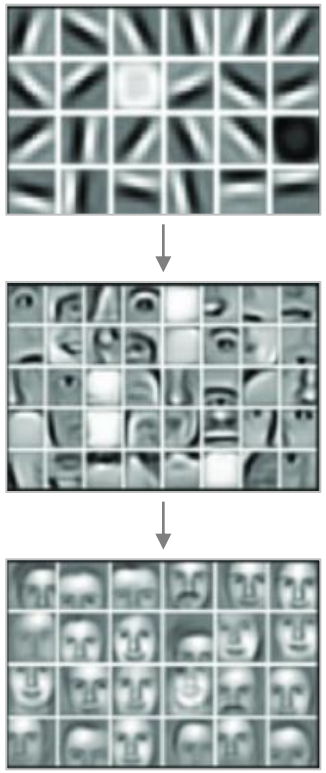


Đa dạng: cùng loại hoặc hỗn hợp (ảnh, văn bản…)

2 Mạng neuron sâu (AI Model)



3 Đặc trưng mạng tự học qua từng lớp



Lớp ẩn 1
Cạnh, nét, góc

Lớp ẩn 2
Bộ phận: mắt, mũi, miệng

Lớp ẩn N
Khuôn mặt hoàn chỉnh

Lớp càng sâu, đặc trưng càng trừu tượng.

Vector đầu vào

Dữ liệu được chuyển thành vector. Trong không gian vector, đặc trưng của dữ liệu trở nên nổi bật.

Mạng neuron sâu

Mạng có rất nhiều lớp, mỗi lớp có nhiều node, nên được gọi là mạng neuron sâu.

Kiến trúc mạng

Các lớp có thể được tổ chức theo nhiều cách khác nhau, gọi là kiến trúc mạng (ví dụ: CNN, RNN, Transformer).

2.10 Phân loại các kỹ thuật Deep Learning và Ứng Dụng

1 Thị giác máy tính
Computer Vision

- **Nhiệm vụ:** phát hiện đối tượng, phân đoạn hình ảnh, nhận diện khuôn mặt, nhận diện hành động
- **Thách thức:** dữ liệu hạn chế, ứng dụng thời gian thực

2 Xử lý ngôn ngữ tự nhiên
Natural Language Processing

- **Nhiệm vụ:** phân tích cảm xúc, nhận diện thực thể tên, tóm tắt văn bản, sinh ngôn ngữ
- **Thách thức:** ngữ cảnh, dữ liệu đa ngôn ngữ, chất lượng phản hồi

3 Học sâu tăng cường
Deep Reinforcement Learning

- **Cách làm:** kết hợp deep learning với học tăng cường; ví dụ chơi game, điều khiển robot
- **Ưu điểm:** học được chính sách tối ưu
- **Thách thức:** hiệu quả dữ liệu, ứng dụng thực tế

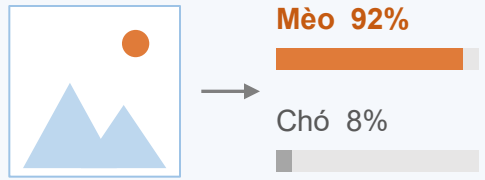
4 Mô hình tạo sinh
Generative Models

- **Cách làm:** dự đoán ra dữ liệu mới (tạo sinh)
- **Đầu ra:** thuộc một loại dữ liệu, hoặc kết hợp nhiều loại dữ liệu khác nhau

2.11 Deep Learning và Thị Giác Máy Tính

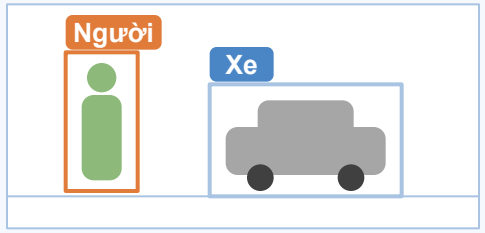
Thị giác máy tính: dạy máy tính hiểu nội dung ảnh và video. Deep Learning là công nghệ chính đứng sau các bài toán dưới đây.

1 Phân loại ảnh



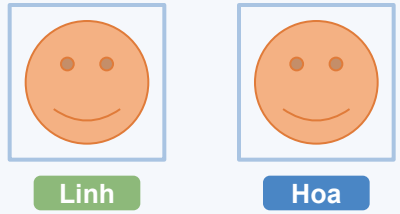
Ảnh → nhãn

2 Phát hiện đối tượng



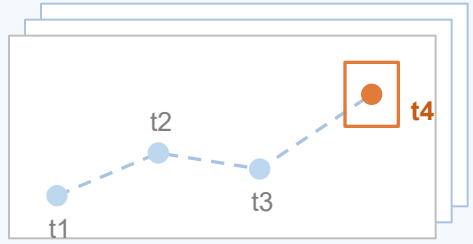
Ảnh → vị trí và nhãn từng đối tượng

3 Nhận diện gương mặt



Ảnh người → tên và vị trí khuôn mặt

4 Theo dõi đối tượng



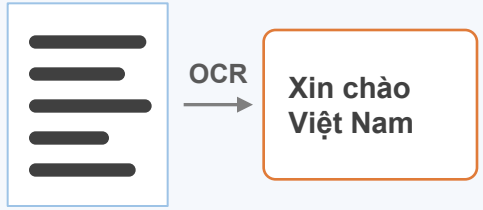
Video → vị trí đối tượng qua từng khung hình

5 Phát hiện hành vi



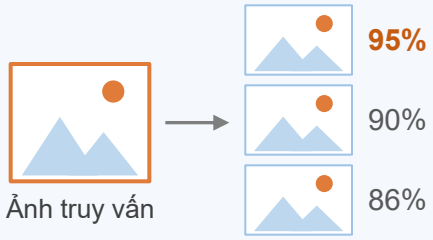
Video → hành vi của người

6 Nhận dạng chữ (OCR)



Ảnh → văn bản

7 Tìm kiếm ảnh



Ảnh → các ảnh tương đồng

Nền tảng chung

Cả 7 bài toán đều dùng mạng neuron sâu, phổ biến nhất là mạng tích chập (CNN).

Các lớp đầu học đặc trưng chung như cạnh, hình dạng. Thay lớp ra là chuyển sang bài toán khác.

2.12 Deep Learning và Xử Lý Ngôn Ngữ Tự Nhiên

3 Ứng dụng

- Phân tích cảm xúc
- Phân tích văn bản
- Dịch máy
- Hỏi đáp tự động
- Trích xuất thông tin
- Tóm tắt văn bản
- Trợ lý ảo
- …

2 Kiến trúc mạng nơ-ron

- CNN
- RNN
- LSTM
- …

1 Xử lý dữ liệu

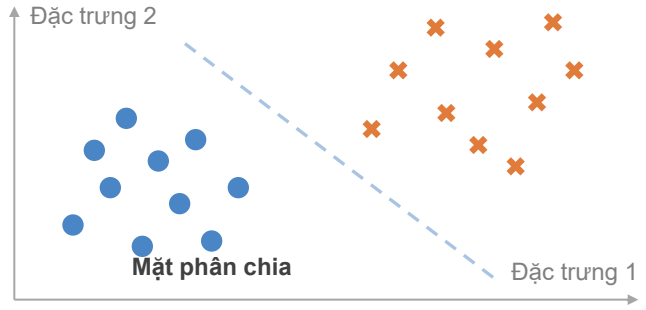
- word2Vec
- Embedding từ con
- GloVe (Global Vector for Word Representation)
- BERT

Đầu vào: dữ liệu dạng văn bản (sách, tài liệu, nội dung web, …)

2.13 Tại sao chương trình AI đòi hỏi nhiều GPU

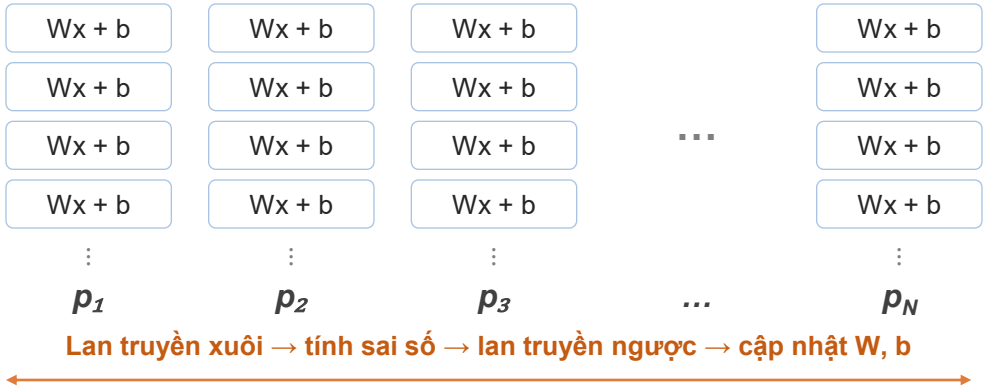
Vì sao AI cần GPU? Huấn luyện mô hình là lặp lại hàng tỷ phép tính nhỏ giống nhau (y = Wx + b). Các phép tính này độc lập với nhau, nên chạy song song sẽ nhanh hơn nhiều so với chạy lần lượt.

1 Bài toán: Phân loại dữ liệu



Thử, đo sai số, chỉnh W và b, lặp lại hàng triệu lần

2 Mỗi lần lặp: hàng triệu phép tính giống nhau



Độc lập với nhau

Không phép tính nào phải chờ phép tính khác, nên có thể chạy cùng một lúc.

Mô hình càng lớn, N càng lớn: tới hàng tỷ phép tính.

3 Chạy N phép tính trên CPU và GPU

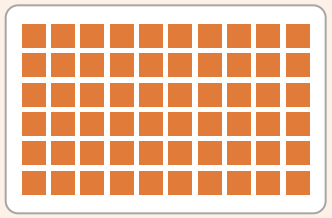
CPU: vài chục lõi mạnh



- Thường **8–64 lõi**, mỗi lõi xử lý phức tạp rất tốt
- Hàng triệu phép tính phải **xếp hàng**, chạy lần lượt từng nhóm nhỏ

Thời gian: Nhiều lượt nối tiếp nhau

GPU: hàng nghìn lõi nhỏ

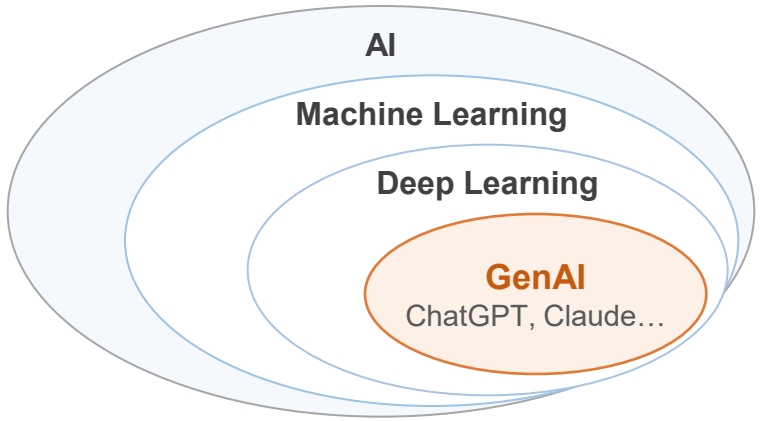


- Mỗi lõi yếu hơn lõi CPU, nhưng có **hàng nghìn lõi**
- Hàng nghìn phép tính **chạy cùng một lúc**, hợp với kiểu tính toán của AI

Thời gian: Ít lượt, nhanh hơn nhiều

2.14 GenAI hay LLM là gì? Tại sao nói về nó nhiều thế?

1 GenAI nằm ở đâu?



2 GPT nghĩa là gì?

G — **Generative: tạo sinh**
Dự đoán từ tiếp theo để tạo ra văn bản

P — **Pre-trained: đào tạo trước**
Học sẵn từ lượng văn bản rất lớn

T — **Transformer**
Kiến trúc mạng neuron, ra đời năm 2017

3 Có tin được LLM không?

LLM học cách tạo văn bản nghe giống người, không học để nói đúng sự thật, nên **có thể bị ảo giác** (bịa ra thông tin).

Cách dùng an toàn: grounding, tức là đưa thêm tài liệu đáng tin vào ngữ cảnh để LLM bám vào khi trả lời.

4 LLM được tạo ra và sử dụng như thế nào?



2.15 Các hướng tiếp cận khai thác LLM

1 Kỹ thuật Prompting
Thiết kế chỉ dẫn đầu vào

- **Mục đích:** hướng dẫn mô hình tạo ra đầu ra mong muốn
- **Kỹ thuật:** chain-of-thought (CoT), few-shot prompting, self-consistency, socratic prompting

2 Tích hợp công cụ và kiến thức bên ngoài
Mở rộng LLM

- **Mục đích:** kết nối LLM với hệ thống bên ngoài hoặc cơ sở kiến thức
- **RAG:** Retrieval-Augmented Generation: truy xuất thông tin từ cơ sở dữ liệu để bổ sung đầu ra

3 Tinh chỉnh mô hình
Fine-tuning

- **Mục đích:** huấn luyện thêm trên dữ liệu có nhãn cho nhiệm vụ cụ thể
- **Kỹ thuật:** Instruction tuning, Supervised fine-tuning (SFT), RLHF (học tăng cường từ phản hồi con người)

4 Tối ưu hóa hiệu suất và bảo trì
Giảm chi phí vận hành

- **Quantization:** giảm độ chính xác tham số để tiết kiệm bộ nhớ
- **Model Editing:** cập nhật kiến thức không cần huấn luyện lại
- **Machine Unlearning:** loại bỏ dữ liệu không mong muốn

2.16 Nhận thức và phản biện khi sử dụng LLM

Khả năng phê phán và đánh giá (phản biện) AI
là yếu tố quan trọng nhất mà người dùng cần đạt được khi sử dụng các mô hình LLM. Bao gồm bốn kỹ năng chính:

1

Đặt câu hỏi và đánh giá độ tin cậy

Nghi vấn và đánh giá mức độ tin cậy của hệ thống AI. Ví dụ: AI detectors không đáng tin cậy, có thể thiên vị chống lại người không nói tiếng Anh bản xứ.

2

Nhận diện thiên kiến

Nhận biết các thiên kiến tiềm ẩn trong tương tác và đầu ra của AI, ví dụ thiên kiến trong ChatGPT hoặc Llama ảnh hưởng đến kết quả.

3

Giảm thiểu GIGO

Hạn chế "rác vào, rác ra" (Garbage In, Garbage Out): bảo đảm đầu vào chất lượng để tránh đầu ra không chính xác.

4

Xem xét vấn đề đạo đức

Cân nhắc các hàm ý đạo đức khi sử dụng AI: công bằng, quyền riêng tư và tác động môi trường.

2.17 AI Agent (Tác tử AI)

AI Agent: chương trình dùng AI để tự nhận biết môi trường, suy luận, ra quyết định và hành động nhằm đạt mục tiêu, với mức độ tự chủ nhất định.

1 Đầu vào

Người dùng
Gửi yêu cầu bằng prompt
Chat | Slack | Email

Sự kiện hệ thống
Agent tự chạy khi có sự kiện
Webhook | Lịch (Cron)

Dữ liệu
Tài liệu, ảnh, giọng nói, cảm biến
PDF | Ảnh | Audio | IoT

Agent

Điều phối: LangChain | LangGraph | AutoGen | CrewAI

2 Nhận biết (Perception)
Chuyển giọng nói, ảnh, tài liệu thành văn bản cho LLM
Whisper (giọng nói) | OCR (ảnh) | Đọc PDF/Excel | Vision model

3 Suy luận (LLM)
Chia mục tiêu thành các bước, chọn công cụ phù hợp
GPT-4o | Claude | Gemini | Llama
ReAct | Chain-of-Thought

4 Tri thức (RAG)
Đưa prompt mẫu và tài liệu nội bộ vào ngữ cảnh trước khi LLM trả lời
System prompt | Tài liệu nội bộ | Embedding

5 Bộ nhớ (Memory)
Ngắn hạn: lịch sử hội thoại trong phiên
Dài hạn: lưu lại để dùng ở lần sau
Pinecone | Chroma | pgvector

6 Hành động (Tools)

Gọi API / công cụ
REST API | MCP

Chạy code
Python | SQL

Tìm kiếm web
Search API | Trình duyệt

Điều khiển thiết bị
Robot | IoT

Gọi Agent khác
A2A | Multi-agent

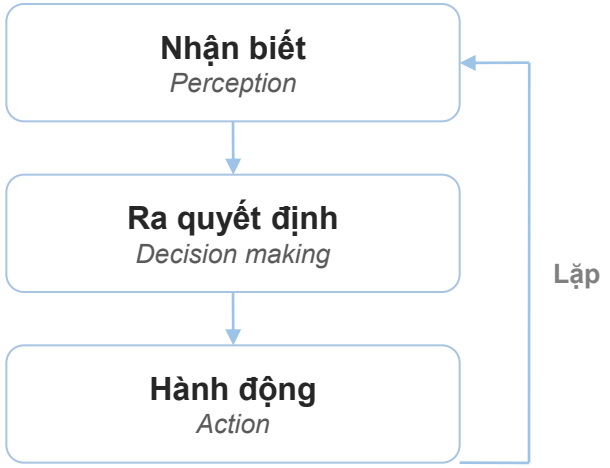
Vòng lặp Agent
lặp lại đến khi đạt mục tiêu

Nhận biết → Suy luận → Hành động → Quan sát kết quả → Trả kết quả cho người dùng

2.18 Từ AI Agent đến Agentic AI

Agentic AI: nhiều agent cùng hợp tác, tự chia nhỏ nhiệm vụ, dùng chung bộ nhớ bền vững và được điều phối để giải quyết mục tiêu lớn.

1 AI Agent: một tác tử



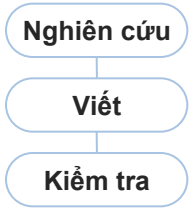
Một agent tự làm trọn một nhiệm vụ, nhớ ngữ cảnh trong phiên làm việc.

Mở rộng

2 Agentic AI: hệ nhiều tác tử phối hợp

1 Đa agent chuyên biệt
Specialized Agents

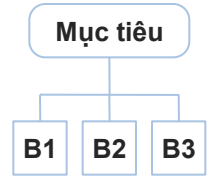
Mỗi agent giỏi một việc và cộng tác với nhau



CrewAI · AutoGen

2 Suy luận và lập kế hoạch
Reasoning & Planning

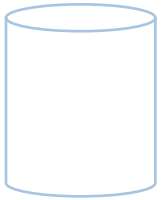
Tự chia mục tiêu lớn thành các bước nhỏ



ReAct · Tree-of-Thought

3 Bộ nhớ bền vững
Persistent Memory

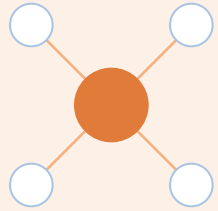
Ngữ cảnh lưu lâu dài, chia sẻ giữa các agent



Vector DB · Knowledge graph

4 Điều phối
Orchestration

Giao việc, theo dõi và gộp kết quả



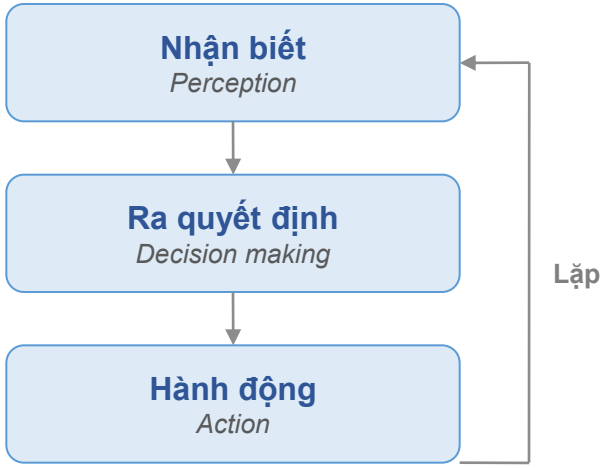
LangGraph · MCP · A2A

AI Agent	Số lượng	Nhiệm vụ	Bộ nhớ
→ **Agentic AI**	1 agent → **nhiều agent phối hợp**	cố định → **tự chia nhỏ, linh hoạt**	trong một phiên → **bền vững, chia sẻ chung**

2.18 Từ AI Agent đến Agentic AI

Agentic AI: nhiều agent cùng hợp tác, tự chia nhỏ nhiệm vụ, dùng chung bộ nhớ bền vững và được điều phối để giải quyết mục tiêu lớn.

1 AI Agent: một tác tử



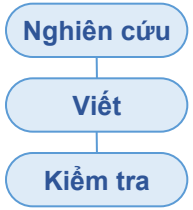
Một agent tự làm trọn một nhiệm vụ, nhớ ngữ cảnh trong phiên làm việc.

Mở rộng

2 Agentic AI: hệ nhiều tác tử phối hợp

1 Đa agent chuyên biệt
Specialized Agents

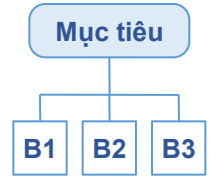
Mỗi agent giỏi một việc và cộng tác với nhau



CrewAI · AutoGen

2 Suy luận và lập kế hoạch
Reasoning & Planning

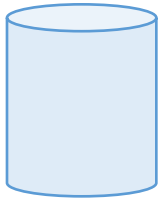
Tự chia mục tiêu lớn thành các bước nhỏ



ReAct · Tree-of-Thought

3 Bộ nhớ bền vững
Persistent Memory

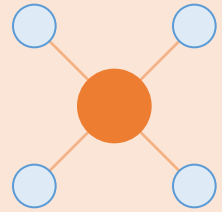
Ngữ cảnh lưu lâu dài, chia sẻ giữa các agent



Vector DB · Knowledge graph

4 Điều phối
Orchestration

Giao việc, theo dõi và gộp kết quả



LangGraph · MCP · A2A

AI Agent → Agentic AI	Số lượng	Nhiệm vụ	Bộ nhớ
	1 agent → **nhiều agent phối hợp**	cố định → **tự chia nhỏ, linh hoạt**	trong một phiên → **bền vững, chia sẻ chung**

2.19 Các Bước Thực Hiện Triển Khai Agentic Systems

1. Định hướng

1 Xác định mục tiêu rõ ràng
Giảm chi phí, tăng doanh thu, cải thiện hiệu quả

2 Phù hợp mục tiêu tổng thể
Gắn với chiến lược lớn để tối đa hóa tác động

3 Hỗ trợ từ lãnh đạo
Phân bổ nguồn lực, quản lý thay đổi tổ chức

2. Chuẩn bị

4 Case-study tác động cao
Ưu tiên dự án mang lại giá trị nhanh

5 Tư vấn chuyên gia
Xây chiến lược, thiết kế Cognitive Skills Module

6 Đánh giá năng lực
Môi trường IT, nền tảng, dữ liệu, nhân lực

3. Thí điểm

7 Bắt đầu nhỏ với pilot
Kiểm tra hiệu quả, human-in-the-loop (AutoGen)

8 Đo lường thành công
KPI, theo dõi hiệu suất, thu thập phản hồi

9 Lặp lại và cải thiện
Tinh chỉnh theo dữ liệu, giảm hallucination qua RAG

4. Mở rộng

10 Mở rộng dần dần
Triển khai rộng hơn, đào tạo và hỗ trợ

11 Giám sát và tối ưu hóa
Theo dõi liên tục, lớp điều phối cho tự trị

12 Mở rộng quy mô
Áp dụng toàn diện, quản lý rủi ro và thay đổi

3

CHƯƠNG 3

Các khía cạnh cần quan tâm bên cạnh yếu tố công nghệ khi triển khai ứng dụng Trí Tuệ Nhân Tạo

Nội dung chương

3.1 Những thách thức cản trở triển khai AI toàn diện 32
3.2 AI phục vụ phân tích dữ liệu và dữ liệu lớn 33
3.3 Các khía cạnh về đạo đức trong AI 34
3.4 Đầu tư từ đầu hay sử dụng dịch vụ AI 35
3.5 Các tiêu chí đánh giá một hệ thống AI 36
3.6 Ba tầng của AI và những ràng buộc khi khai thác 37

3.1 Những Thách Thức cản trở mục tiêu Triển Khai AI toàn diện

1 Thiếu đầu tư, tài chính hạn chế

- Chi phí phần cứng, phần mềm, đào tạo cao khiến AI thành công nghệ "xa xỉ"
- Đầu tư công cho nghiên cứu KH&CN thấp

2 Hạ tầng số và tính toán yếu

- Hạ tầng tính toán thiếu đồng bộ; lưu trữ, chia sẻ dữ liệu hạn chế
- Chiến lược dữ liệu quốc gia, dữ liệu mở chưa phát triển
- Chính phủ điện tử chưa mạnh

3 Thiếu nhân lực chuyên môn

- Thiếu nhân lực AI chuyên sâu; lao động kỹ năng thấp dễ bị thay thế
- Năng lực đào tạo, STEM phổ thông hạn chế
- "Chảy máu chất xám"

4 Khung chính sách chưa hoàn thiện

- Rủi ro về trách nhiệm, bảo vệ dữ liệu, đạo đức AI
- Dễ bị lạm dụng công nghệ từ bên ngoài

5 Rủi ro triển khai và xã hội

- Trở ngại văn hóa làm việc, thiếu nhận thức, mất việc làm
- Vi phạm quyền riêng tư, phân biệt đối xử, phụ thuộc công nghệ nước ngoài

Nguồn: CSIRO Aus4Innovation, Analytical report: National strategy for Artificial Intelligence 2030 (04/2025)

3.2 AI phục vụ Phân tích dữ liệu Thường gắn với Dữ Liệu Lớn

Hệ thống hỗ trợ ra quyết định: gom dữ liệu từ nhiều nguồn vào kho dữ liệu lớn, rồi dùng phân tích và AI để biến dữ liệu thành báo cáo cho người ra quyết định.



Điều kiện để triển khai thành công

Nghiệp vụ và quản trị

- **Mục tiêu rõ ràng:** giảm chi phí, tăng doanh thu, cải thiện hiệu quả
- **Lãnh đạo ủng hộ:** cam kết từ cấp cao (top management support)
- **Văn hóa dữ liệu:** ra quyết định dựa trên dữ liệu, giảm kháng cự từ nhân viên

Hạ tầng công nghệ và con người

- **Hệ thống vận hành** đã số hóa, có dữ liệu để khai thác
- **Đội ngũ kỹ thuật** duy trì và phát triển hệ thống liên tục
- **Đào tạo liên tục** để người dùng khai thác được công cụ

3.3 Các khía cạnh về Đạo Đức trong AI

Đạo đức trong AI: bảy nguyên tắc giúp hệ thống AI công bằng, an toàn và có trách nhiệm với con người.

1 Công bằng, không thiên vị
Fairness

AI không được tạo ra hay củng cố thành kiến (bias) về giới tính, chủng tộc, tôn giáo. Dữ liệu lịch sử thường chứa sẵn thành kiến xã hội, dẫn đến quyết định không công bằng.

2 Minh bạch, giải thích được
Explainability

Người dùng cần hiểu AI ra quyết định thế nào và vì sao. Mô hình "hộp đen" như mạng neuron sâu khó giải thích, gây thiếu tin tưởng.

3 Trách nhiệm giải trình
Accountability

Phải xác định được ai chịu trách nhiệm khi AI gây hậu quả, ví dụ tai nạn xe tự hành hay quyết định sai trong hệ thống tư pháp.

4 Quyền riêng tư
Privacy

AI cần nhiều dữ liệu cá nhân để huấn luyện, nên có nguy cơ vi phạm quyền riêng tư nếu dữ liệu bị lạm dụng hoặc rò rỉ.

5 An toàn và bảo mật
Safety & Security

Hệ thống AI phải tránh gây hại và chống được tấn công. Ví dụ: AI lái xe tự hành không được để bị hack gây tai nạn.

6 Tác động xã hội, việc làm
Social impact

AI có thể làm thay đổi thị trường lao động, gây mất việc làm hoặc tăng bất bình đẳng kinh tế nếu không được quản lý đúng.

7 Đạo đức ngay từ thiết kế
Ethics by design

Tích hợp nguyên tắc đạo đức từ giai đoạn thiết kế để AI phục vụ lợi ích công cộng. Không phát triển AI cho vũ khí tự động hay thao túng tâm lý.

Áp dụng xuyên suốt

mọi giai đoạn của vòng đời AI

Thu thập dữ liệu
↓
Huấn luyện mô hình
↓
Triển khai
↓
Giám sát, đánh giá

3.4 Đầu tư xây dựng từ đầu hay sử dụng các dịch vụ AI

Tiêu chí	AI as a Service (AIaaS)	AI do tổ chức tự phát triển
Chi phí ban đầu	Thấp, dựa trên mô hình pay-as-you-go.	Cao, yêu cầu đầu tư cơ sở hạ tầng và nhân sự.
Thời gian triển khai	Nhanh, thường vài tuần	Lâu hơn, thường nhiều tháng
Khả năng tùy chỉnh	Hạn chế, mô hình "black-box"	Cao, thiết kế riêng theo nhu cầu
Bảo mật dữ liệu	Dữ liệu được gửi đến máy chủ nhà cung cấp, tiềm ẩn nguy cơ rò rỉ hoặc truy cập trái phép. Phụ thuộc vào chính sách bảo mật của nhà cung cấp.	Thấp hơn do dữ liệu được lưu trữ và xử lý nội bộ, giảm nguy cơ rò rỉ bên ngoài. Yêu cầu đội ngũ bảo mật chuyên nghiệp để ngăn chặn tấn công nội bộ hoặc cấu hình sai.
Quyền riêng tư	Dữ liệu nhạy cảm (như dữ liệu y tế, tài chính) có thể bị nhà cung cấp sử dụng để huấn luyện mô hình nếu không có thỏa thuận rõ rang.	Thấp hơn do tổ chức kiểm soát toàn bộ dữ liệu.
Yêu cầu kỹ thuật	Thấp, không cần chuyên gia sâu	Cao, cần đội ngũ chuyên môn về AI và bảo mật
Chi phí dài hạn	Có thể cao nếu sử dụng nhiều API	Cao ban đầu nhưng có thể tối ưu
Bảo trì	Do nhà cung cấp đảm nhiệm	Tổ chức tự chịu trách nhiệm
Phù hợp với	Doanh nghiệp vừa và nhỏ, triển khai nhanh, ít yêu cầu bảo mật nghiêm ngặt.	Tổ chức lớn, ngành nhạy cảm (chính phủ, y tế, tài chính), cần kiểm soát và tùy chỉnh cao.

3.5 Các Tiêu chí Đánh giá một Hệ Thống AI

Cost

Chi phí cho một lần thực thi đơn lẻ (C) phụ thuộc vào số lượng token đầu vào (T_{in}) và đầu ra (T_{out}) cũng như đơn giá tương ứng cho mỗi loại token (P_{in}, P_{out}). Ta có: $C = T_{in} \times P_{in} + P_{out} \times T_{out}$

Latency

Thời gian phản hồi: Là thời gian trung bình, tính bằng giây, cần thiết để hệ thống hoàn thành một quy trình công việc (L_i). Giá trị trung bình được tính trên N lần thực thi như sau: $\bar{L} = \frac{1}{N} \sum_{i=1}^{N} L_i$

Accuracy

Độ chính xác: Phản ánh tỷ lệ phần trăm các công việc được hoàn thành thành công ($N_{correct}$) so với tổng số công việc được thực hiện (N_{total}) được biểu diễn bằng: $A = \frac{N_{correct}}{N_{total}} \times 100\%$

Stability

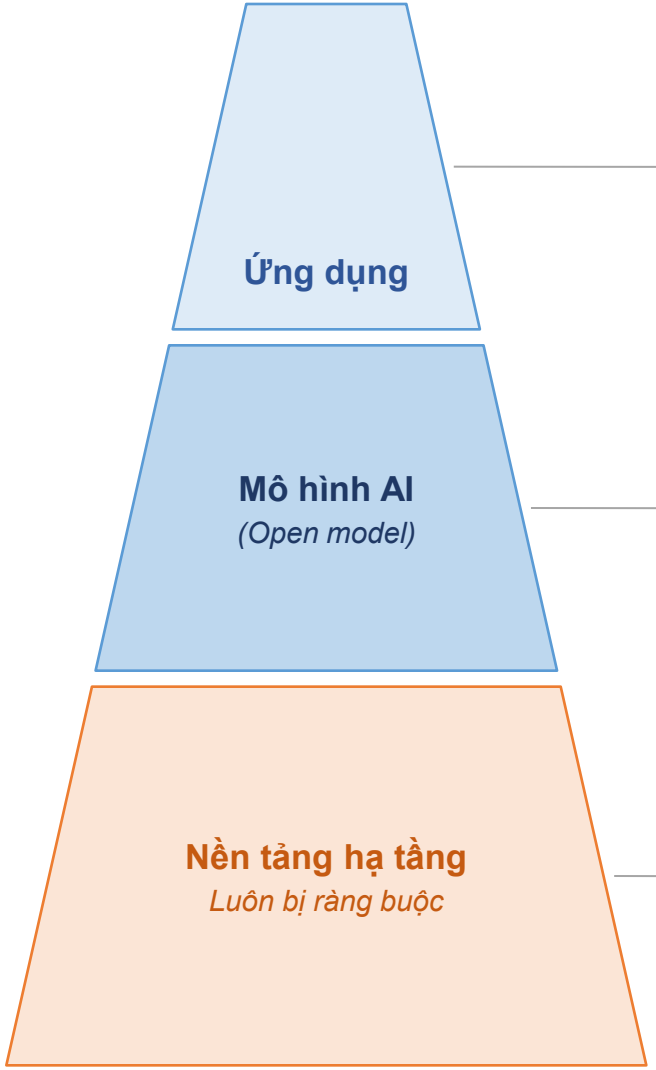
Độ ổn định: Đo lường mức độ nhất quán về hiệu suất hoặc kết quả đầu ra (M) qua nhiều lần (k) thực thi lặp lại với cùng một dữ liệu đầu vào và cấu hình quy trình như sau: $Stability \propto \frac{1}{\sigma(M)}$

Security

An toàn: Chỉ số này được tính bằng tỷ lệ phần trăm số phản hồi được đánh giá là không gây hại (N_{safe}) trên tổng số N_{inject} lời nhắc tiêm nhiễm được sử dụng để thử nghiệm: $S = \frac{N_{safe}}{N_{inject}} \times 100\%$

3.6 Ba tầng của AI và những ràng buộc khi khai thác

1 Các tầng của AI và ràng buộc ở mỗi tầng



Đa dạng, nhưng nhiều rào cản

- AI tạo sinh mã độc → **phục vụ giảng dạy**
- AI tạo sinh kịch bản lừa đảo → **phục vụ điều tra**
- Nhiều loại AI chưa có sẵn hoặc chưa phù hợp để kinh doanh do rào cản kinh tế, đạo đức

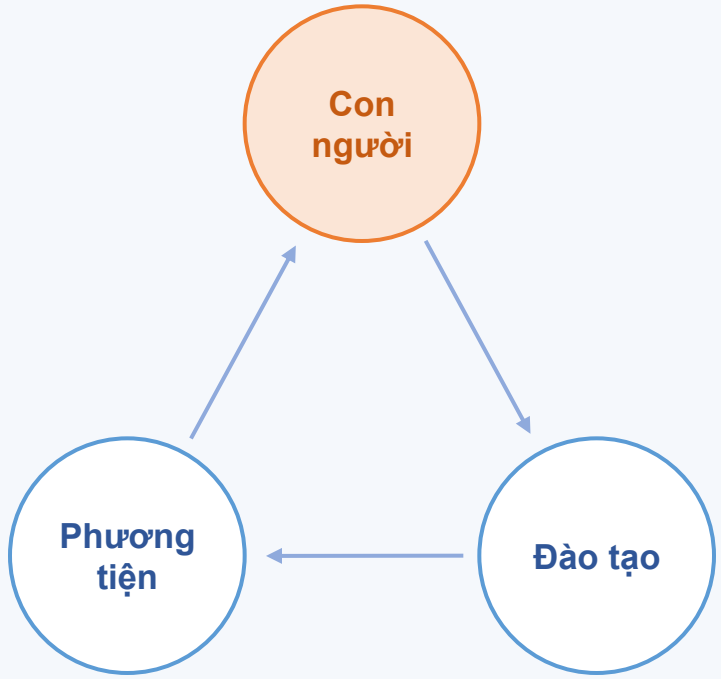
Mục đích rộng, ràng buộc chặt

Mô hình AI được tạo ra với mục đích rất rộng và phải tuân thủ các ràng buộc, nên trong nhiều trường hợp không khai thác được.

Kinh tế và địa chính trị

- **Đầu tư lớn, hàm lượng R&D cao:** rủi ro ở các tầng trên khiến việc quyết định đầu tư khó hơn
- **Ràng buộc địa chính trị:** không cho phép tiếp cận những hạ tầng trọng yếu

2 Nguồn lực cần có



Ba nguồn lực gắn với nhau theo vòng tròn; thiếu một yếu tố, cả hệ thống khó vận hành.