Những khái niệm cơ bản của AI, Khoa học dữ liệu và Chuyển đổi số (AI, DS và DX)

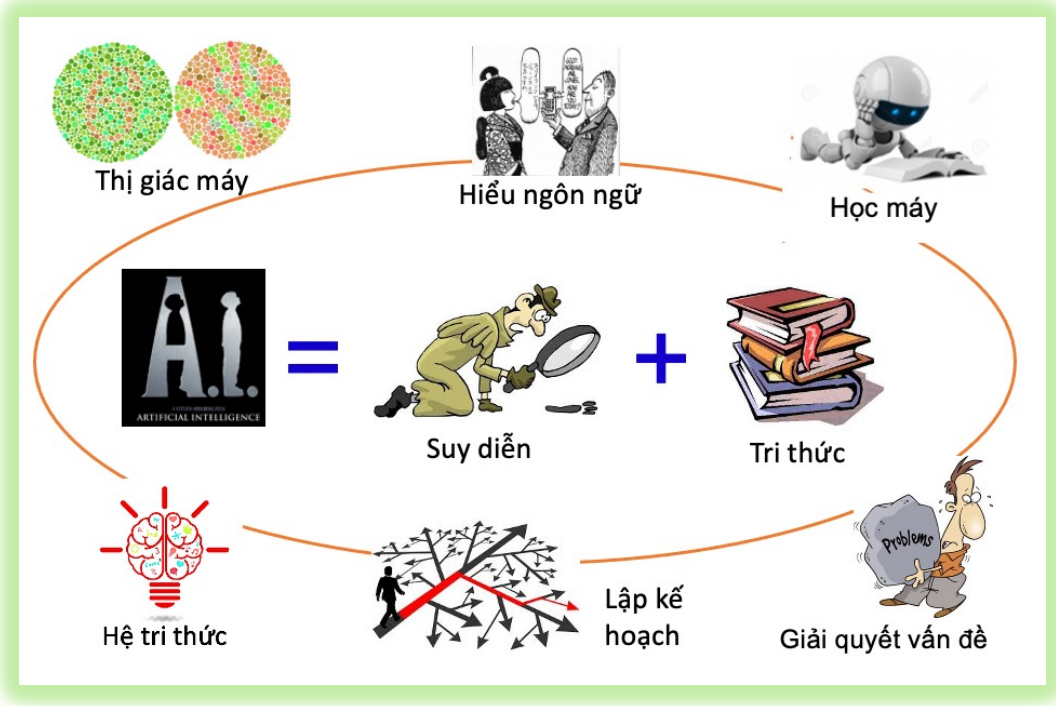
Hồ Tú Bảo
Viện Nghiên cứu cao cấp về Toán/BKAI

Viện Khoa học và Công nghệ Tiên tiến Nhật Bản

Viện Nghiên cứu cao cấp về Toán

- Toán ứng dụng ĐHBKHN (1978); Trí tuệ nhân tạo ĐH Paris 6 (ThS 1984, TS 1987), TS Habil (1998)
- Viện Hàn Lâm KH&CN Việt Nam (1979-1993), PGS (1991)
- GS, Japan Advanced Institute of Science and Technology (1993-2018), GS danh dự (từ 2018-)
- GĐ Lab Khoa học dữ liệu VIASM (4.2018-), Giám đốc khoa học BKAI.

Từ thông minh tự nhiên đến thông minh nhân tạo

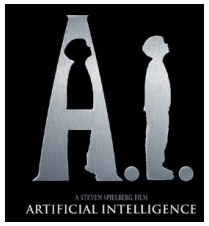


"Nếu có thể Bảo nên chuyển qua nghiên cứu về Trí tuệ nhân tạo, vì đấy là tương lai của Tin học"
GS Phan Đình Diệu, 1984

ROBOCUP
‘Robots must defeat human world champions by 2050’

Tích hợp nhiều công nghệ: robot di chuyển như người, tự điều khiển, hợp tác với robot khác, hoạt động theo chiến thuật, quan sát và quyết định theo thời gian thực … (general AI). https://www.youtube.com/watch?v=zXyxvgWSAdQ
https://www.youtube.com/watch?v=Ub1Z02dVKXM

Cốt lõi của thông minh: Lập luận và Tri thức



=

+

LẬP LUẬN

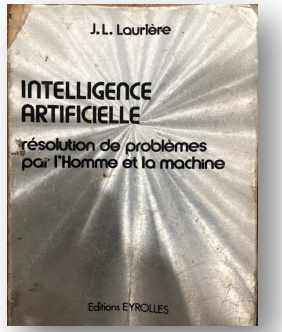
TRI THỨC

tam đoạn luận

Nếu A thì B
A đúng
———————
B đúng

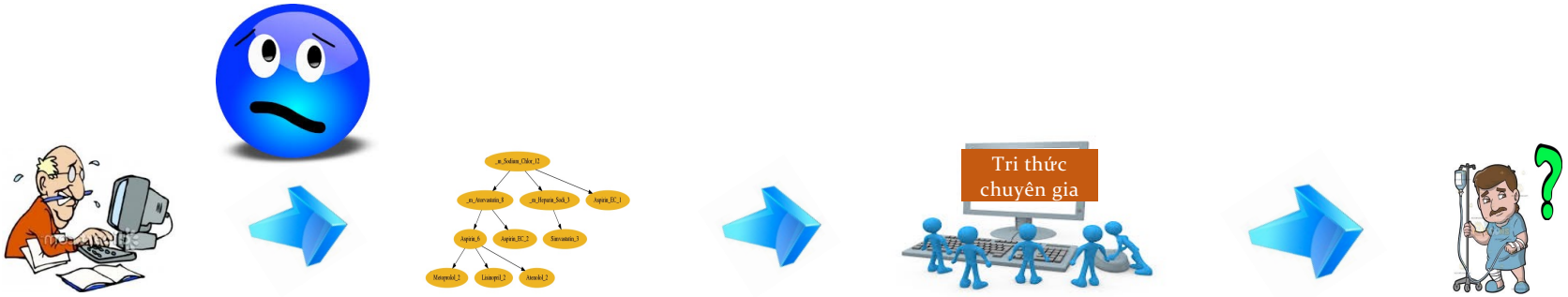
Chuồn chuồn bay thấp thì mưa
Mọi người ai cũng phải chết
Mắt vàng da vàng thì mắc bệnh gan
Công Phượng vào là làm bàn
...

A = chuồn chuồn bay thấp B = mưa



My first AI book
(Paris, 1984)

AI suy diễn (deduction-based AI)



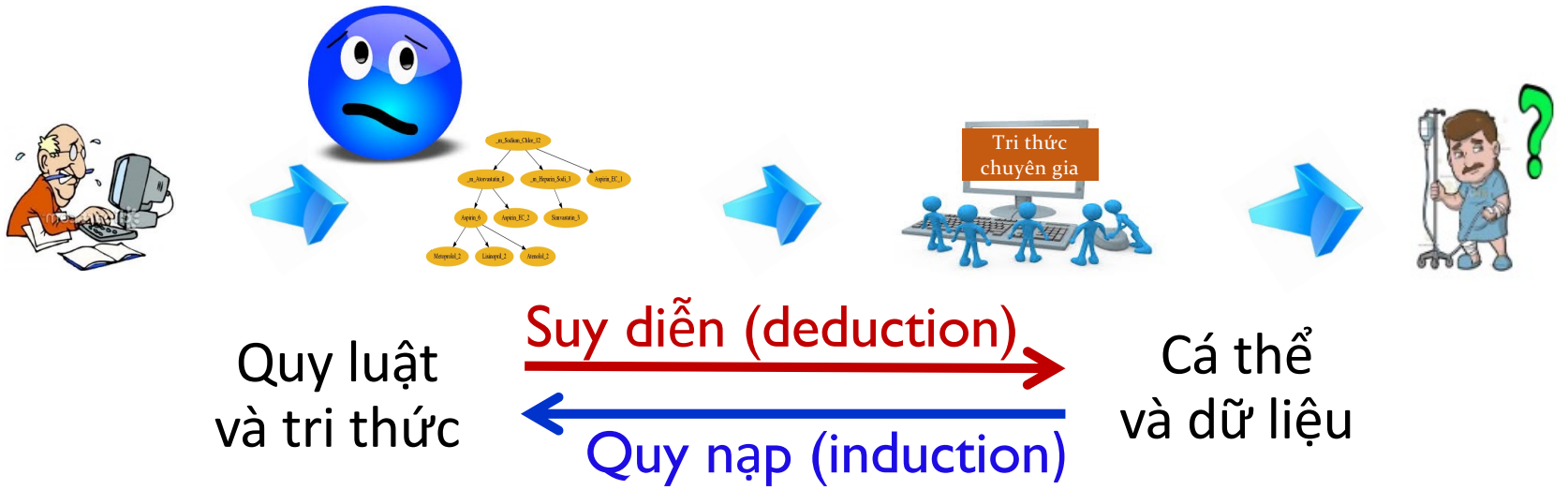
Quy luật
và tri thức

Suy diễn (deduction)

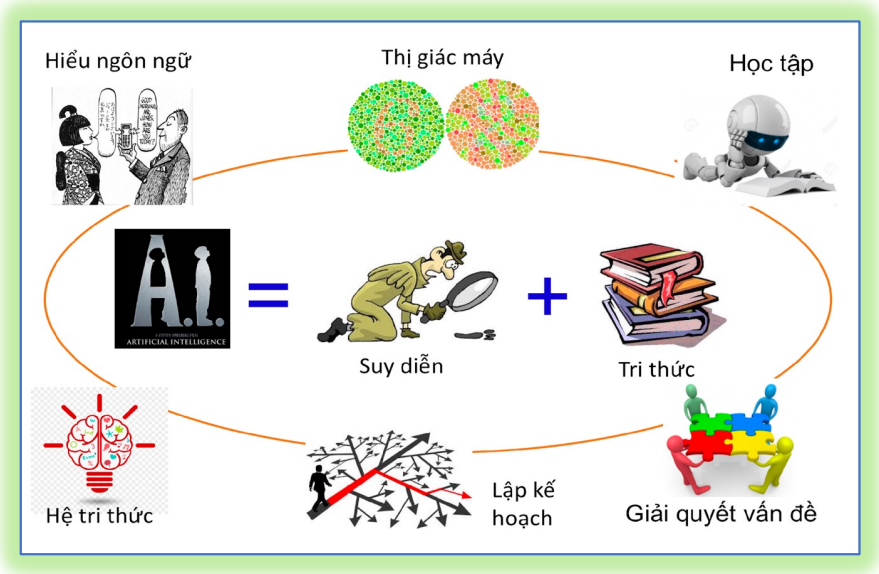
Cá thể
và dữ liệu

"Mắt vàng da vàng thì có
khả năng cao mắc bệnh về gan"

AI quy nạp (induction-based AI)



AI tổng quát và AI chuyên dụng



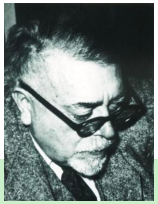
General AI or **Strong AI**
(**AI tổng quát** hoặc **AI mạnh**)

Nhằm làm cho máy biết "lập luận" và "suy nghĩ" ở mức cao của trí tuệ con người.

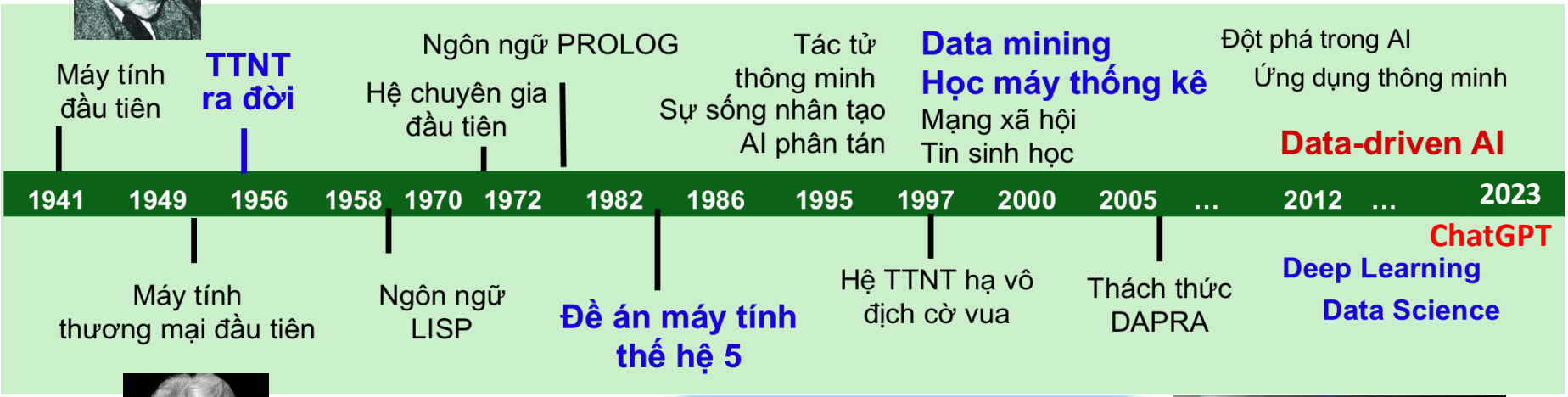
Narrow AI or **Weak AI**
(**AI chuyên dụng** hoặc **AI hẹp**)

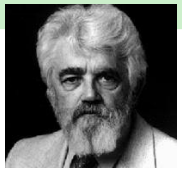
Nhằm làm máy hỗ trợ hoạt động trí óc của con người (intelligence augmentation), giải quyết các việc cụ thể nhờ khai thác, sử dụng dữ liệu với toán học và khoa học máy tính.

AI dựa vào dữ liệu – toán và điều khiển học

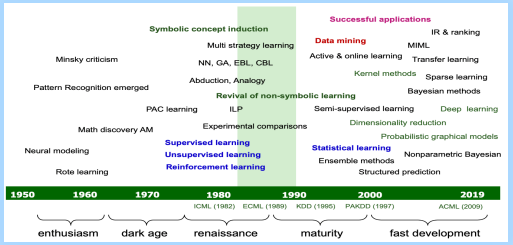


Norbert Wiener, MIT, "Cybernetics" (1948) = "khoa học về điều khiển và giao tiếp ở động vật và máy móc", dựa trên toán tối ưu, thống kê, nhận dạng, lý thuyết thông tin…

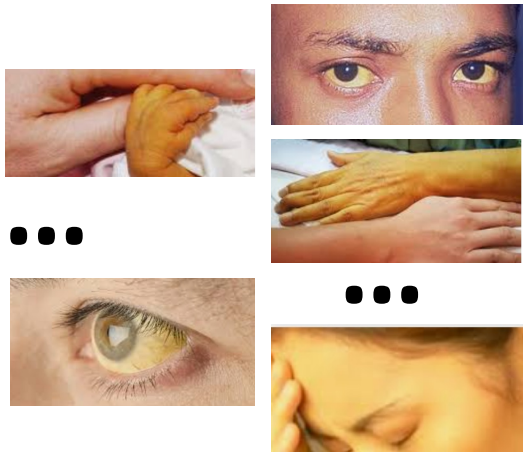




John McCarthy, Dartmouth, "Artificial Intelligence" (1955)
chủ yếu dựa trên logic toán



Học máy: Từ dữ liệu đến tri thức

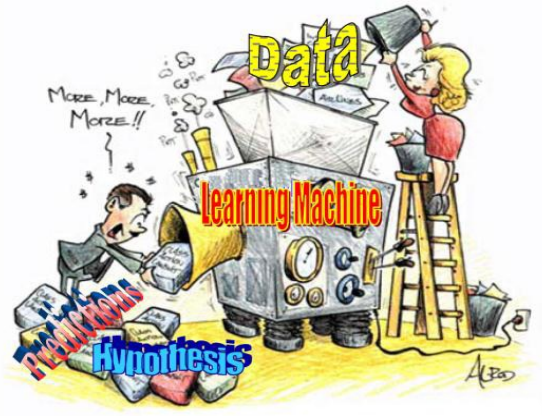


"Mắt vàng da vàng thì có khả năng cao mắc bệnh về gan"

MACHINE LEARNING

Machine learning

- Field of study that gives computers the ability to learn without being explicitly programmed (Arthur Samuel, 1959).
- *Có*
  - Một nguồn dữ liệu gồm n đối tượng X = $\{(\boldsymbol{x}_i, y_i)\}$, $i = 1, \ldots, n$, với $\boldsymbol{x}_i$ là mô tả của đối tượng (data) và y_i là nhãn of $\boldsymbol{x}_i$ (tính chất đặc biệt) (y_i có hữu hạn hoặc vô giá trị, đã biết hoặc không biết).
  - Tập hợp các bệnh án điện tử.
- *Tìm*
  - Hàm $p(y|\boldsymbol{x})$ để dự đoán nhãn của các đối tượng mới hoặc hàm $p(\boldsymbol{x})$ đặc trưng cho các tính chất của các đối tượng.
  - Tri thức chẩn đoán bệnh hoặc gợi ý phác đồ điều trị.



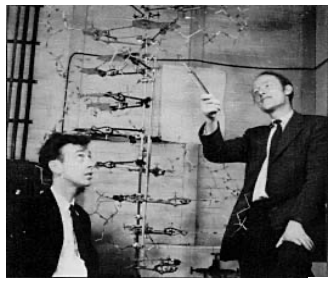
(Source: Eric Xing lecture)

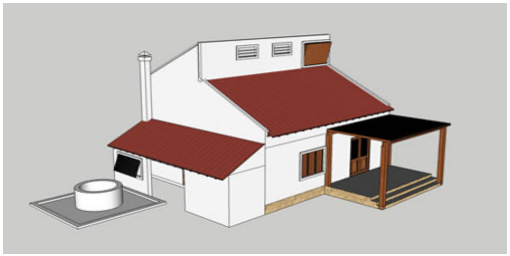
$$\mathbf{y} = \mathbf{f}(\mathbf{x}; \mathbf{W})$$

Machine learning: Máy học hay Học máy?

Model selection: Cốt lõi của học máy

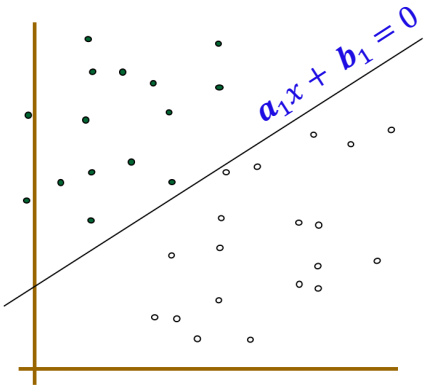
Model - **Mô hình**
Abstract description or representation of a reality.
Mô tả hoặc biểu diễn khái quát của một hiện thực.

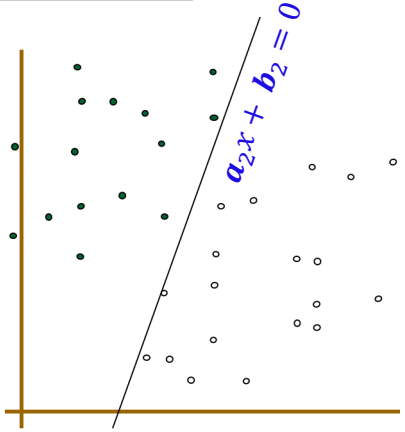


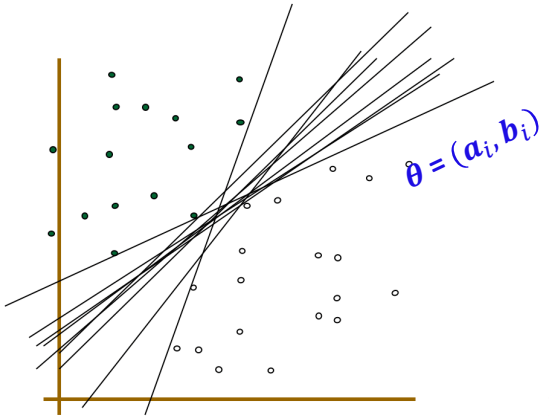


Một mô hình trong học máy được định nghĩa là một bộ các tham số của các phân bố xác suất, biểu diễn bởi các tham số của mô hình

$$\boldsymbol{M} = \{\boldsymbol{f}(\boldsymbol{y}|\boldsymbol{\theta})|\boldsymbol{\theta} \in \Omega\}$$





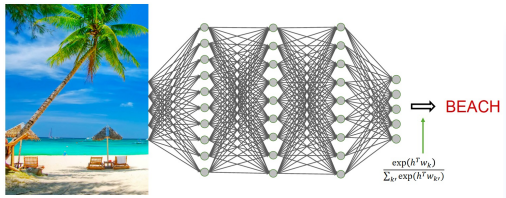


Machine learning: Các trường phái lớn

Tribes	Origins	Master Algorithm
Symbolists	Logic, philosophy	*Inverse deduction*
Evolutionary	Evolutionary biology	*Genetic programming*
Connectionists	Neuroscience	*Backpropagation*
Probabilistic	Statistics	*Bayes' theorem*
Analogizers	Psychology	*Support Vector Machines*

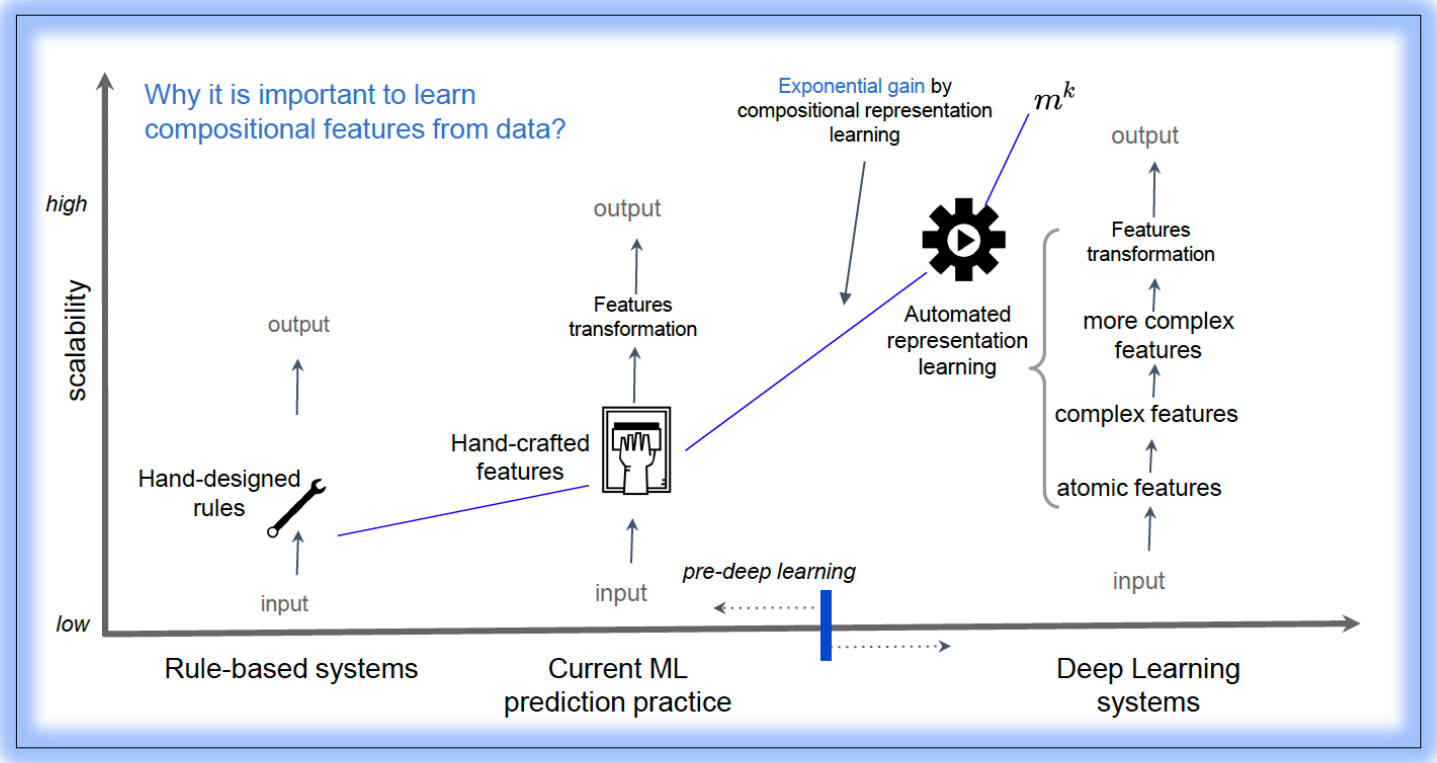
Petro Domingos, The Five Tribes of Machine Learning

Deep learning

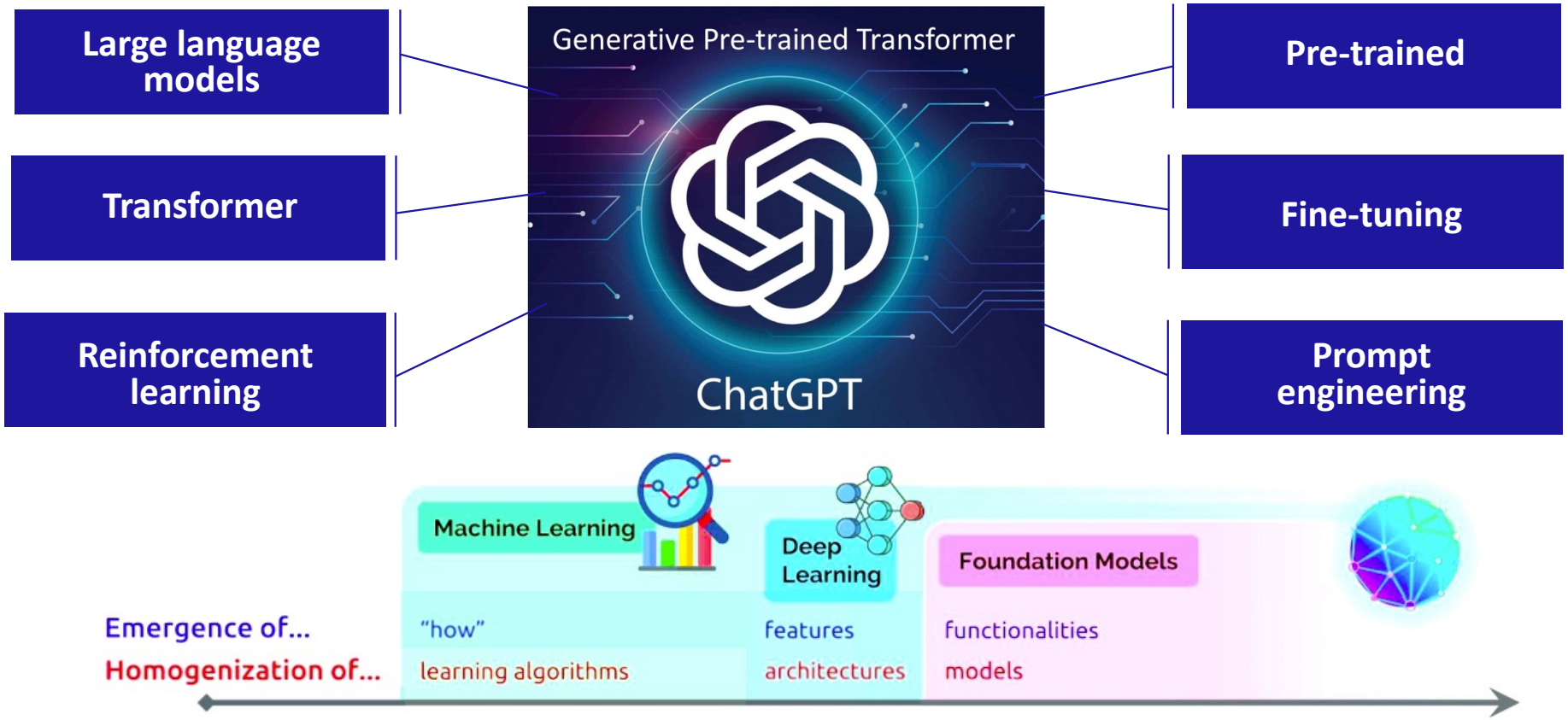


Phương pháp học nhiều tầng để biểu diễn và mô tả dữ liệu

Methods with multi-level learning for representing and describing the data.

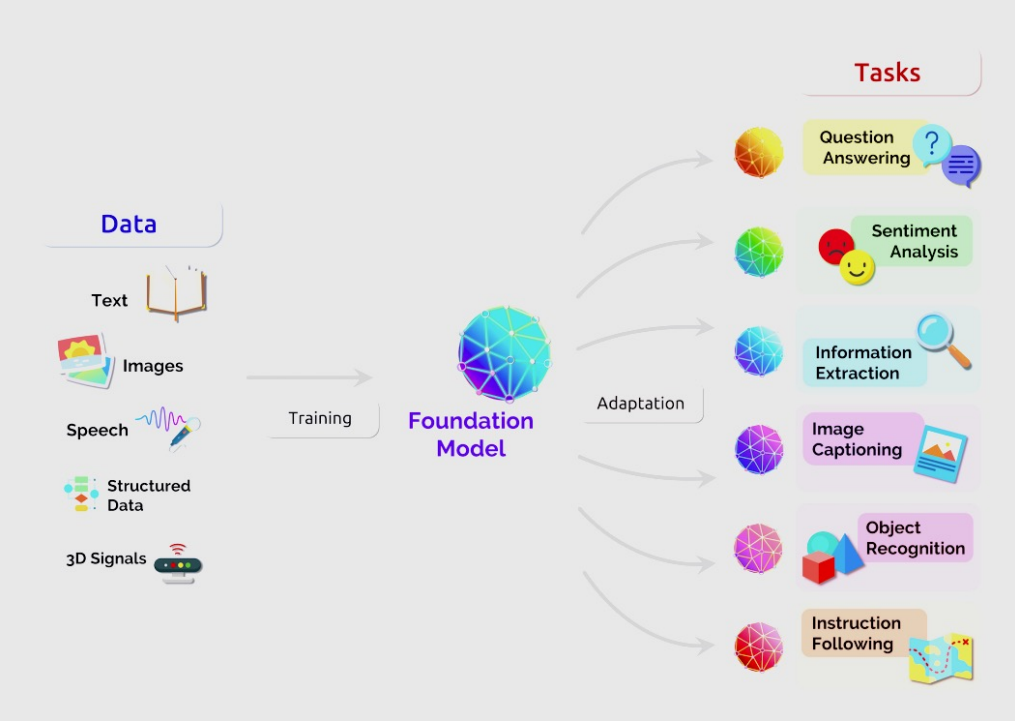


ChatGPT và những bước tiến tiếp theo



Bommasani et al. (2021). On the opportunities and ríks of foundation models

Foundation models

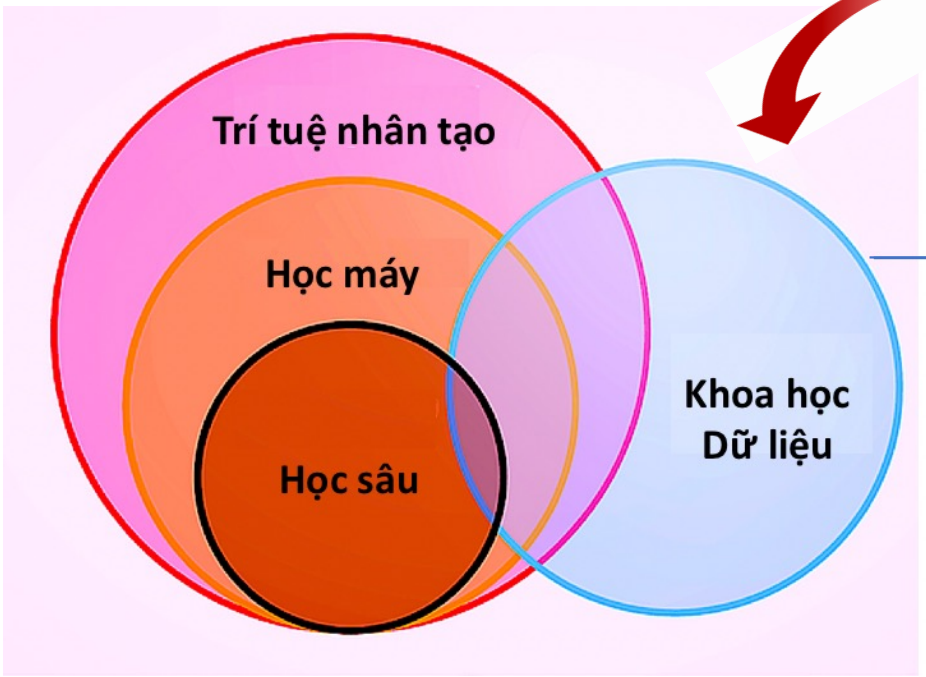


A foundation model is a model trained at broad scale that can adapted to a wide range of downstream tasks.

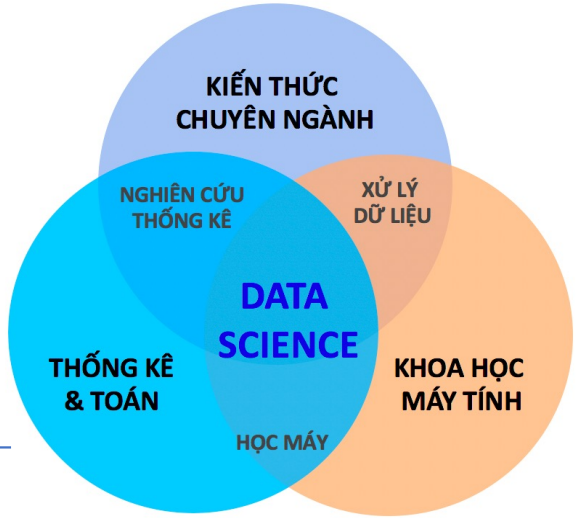
Mô hình nền tảng là mô hình được huấn luyện ở quy mô rộng và có thể thích ứng với nhiều việc khác nhau (máy cái).

Bommasani et al. (2021). On the opportunities and ríks of foundation models

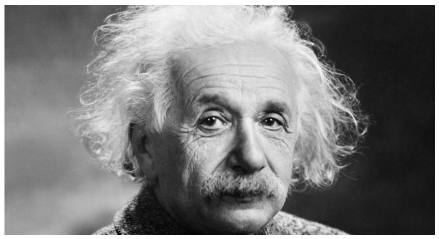
AI và Khoa học dữ liệu



DATA MINING



"We cannot solve problems by using the same kind of thinking we used when we created them"
Albert Einstein



Source: Data Science for the C-Suite, New York: Digital Living Press, 2015

Spotlight

ARTWORK Tamar Cohen, *Andrew J Buboltz*
2011, silk screen on a page from a high school yearbook, 8.5" x 12"

Data Scientist:

The Sexiest Job of the 21st Century

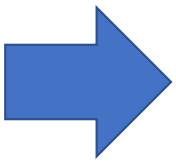
Meet the people who can coax treasure out of messy, unstructured data.

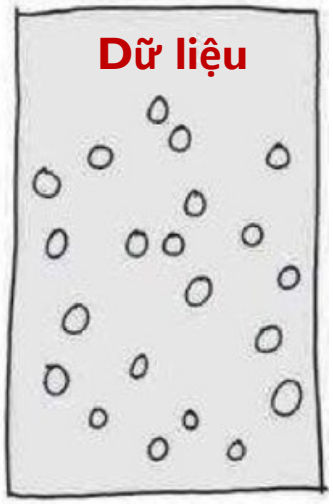
by Thomas H. Davenport and D.J. Patil

Harvard Business Journal, Oct. 2012

Data, information, and knowledge

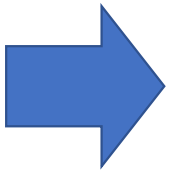
- Quan sát
- Đo đạc
- Thu thập

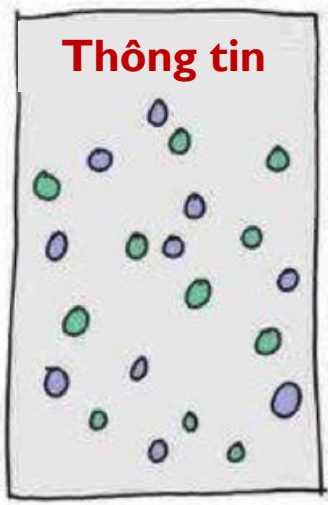




Tín hiệu thô
chưa giải nghĩa

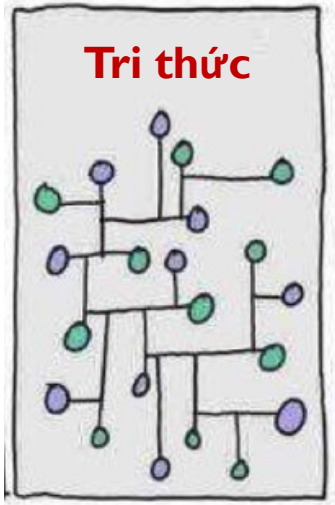
- Xử lý
- Tính toán





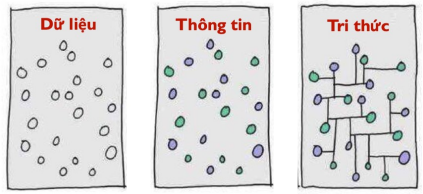
Nghĩa của dữ liệu

- Nhận thức
- Học tập
- Khám phá



Thông tin tích hợp
và quan hệ của chúng

Ví dụ về dữ liệu, thông tin, và tri thức



DỮ LIỆU	THÔNG TIN	TRI THỨC
dãy số 2, 3, 5, 7, 11	là những số nguyên tố	2 là số nguyên tố chẵn duy nhất, 3 là số nguyên tố lẻ bé nhất…
các hợp đồng được ký kết bởi một công ty.	một số đang đến hạn, cần đàm phán lại để đảm bảo các điều khoản và giá cả.	quản lý hợp đồng là quan trọng, đảm bảo thành công và lợi nhuận
độ pH của đất trong một khu vực nông nghiệp	đất trong khu vực có độ pH thấp và là đất sét	hợp cho trồng lúa và cần thêm phân bón để cải thiện chất lượng đất.
mức độ ô nhiễm nước sông tại một trạm giám sát.	mức độ ô nhiễm nước trên mức cho phép, có dấu hiệu ô nhiễm hóa chất	hại cho hệ sinh thái và sức khỏe, tìm nguồn gốc của ô nhiễm và xử lý nước
Điểm số của nhóm bệnh nhân khi xét nghiệm chức năng thận.	hầu hết có điểm số trong khoảng bình thường, một số có điểm thấp.	điểm thấp là chức năng thận có thể có vấn đề, cần kiểm tra bổ sung để chẩn đoán và điều trị

Tại sao dữ liệu ngày càng nhiều và càng quan trọng?

THU ĐƯỢC DỮ LIỆU VỚI TỐC ĐỘ CAO

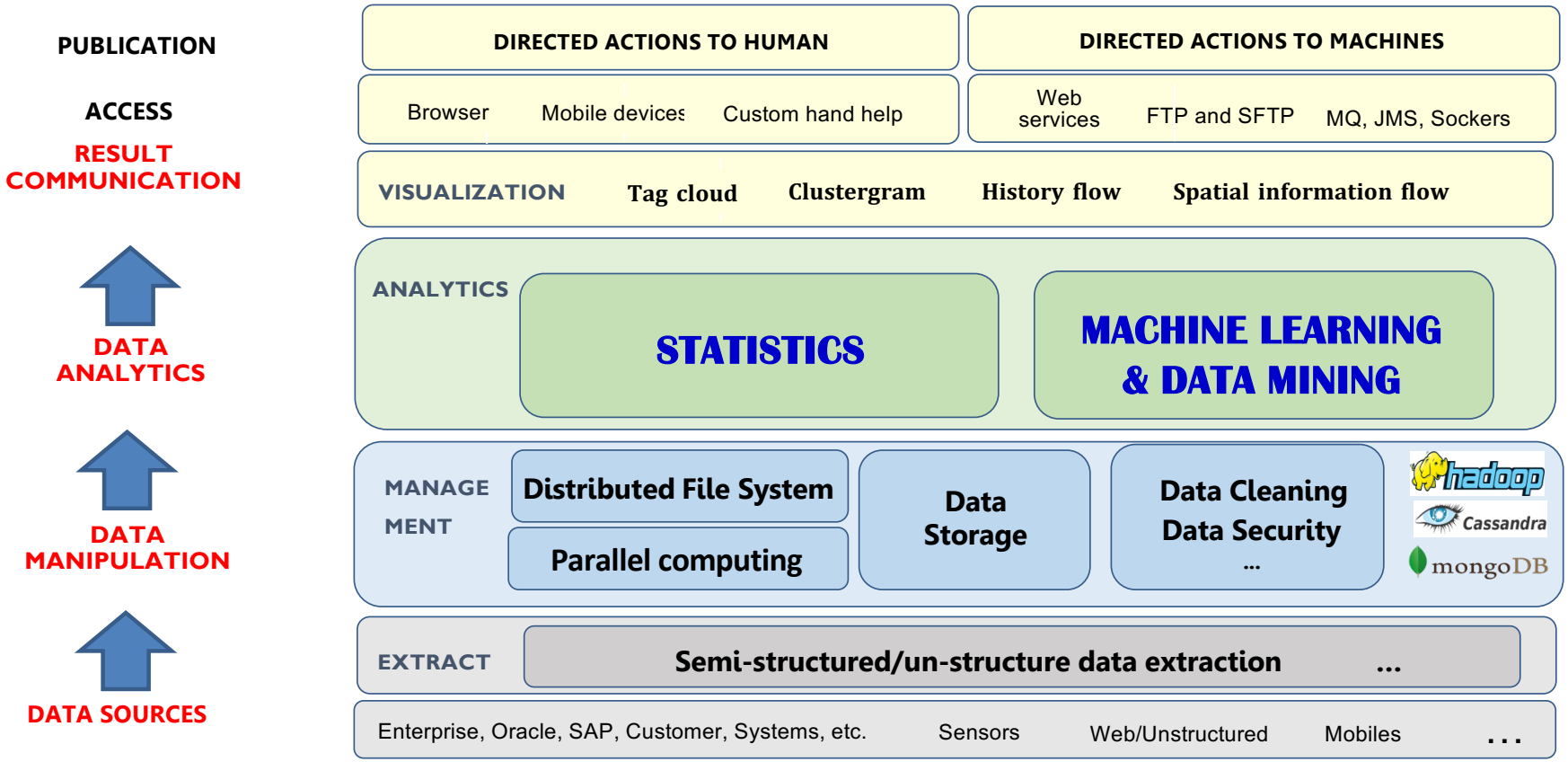
Khảo sát, điều tra, quan sát, ghi chép, thu thập, đo đạc

PHÂN TÍCH, XỬ LÝ DỮ LIỆU HIỆU QUẢ

Phân tích mô tả, dự báo, khuyến cáo… với statistics, AI, big data, business analytics

- Thông tin và tri thức ẩn chứa trong dữ liệu thu từ các hoạt động.
- Dữ liệu cho ta thấu hiểu, tìm ra thông tin, tri thức, để mỗi hoạt động đều có thể thông minh.

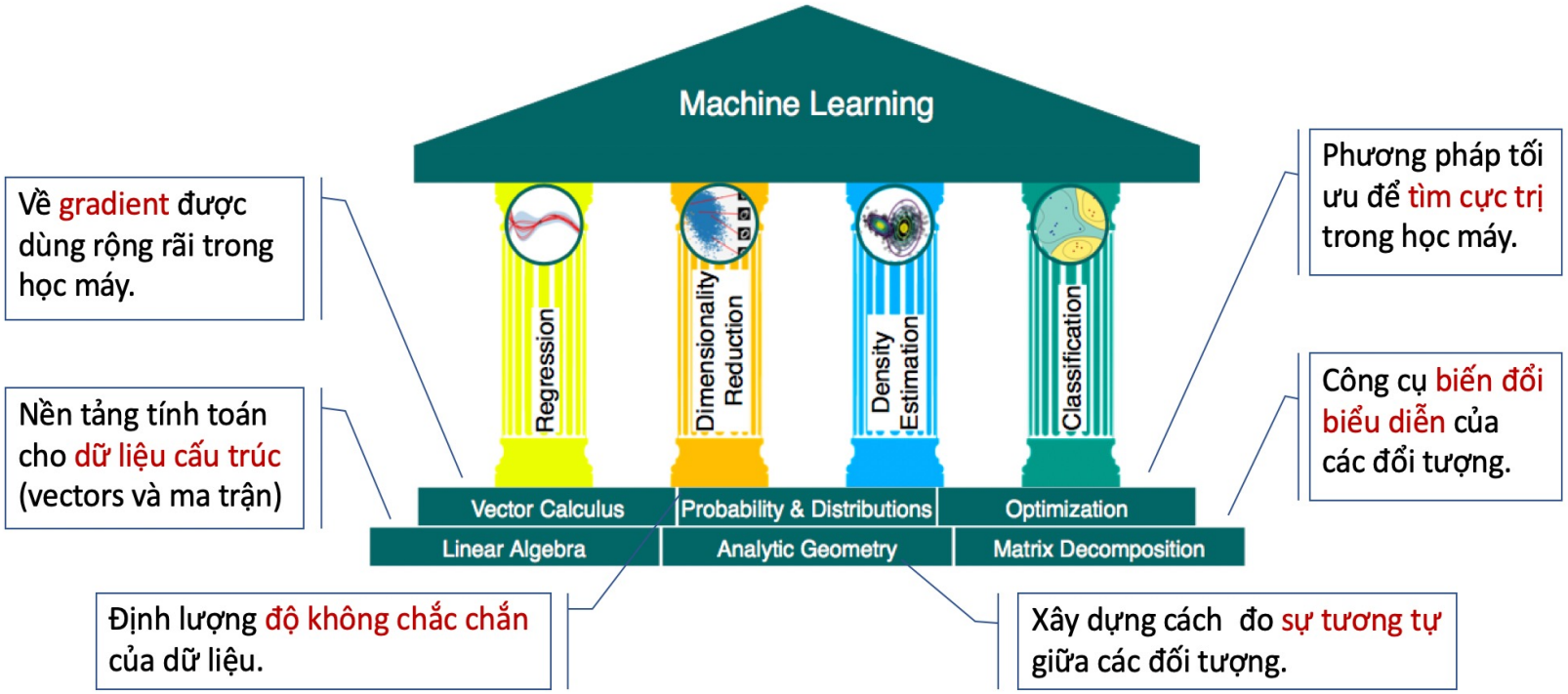
Một lược đồ của khoa học dữ liệu



Source: WAMDM, Web group

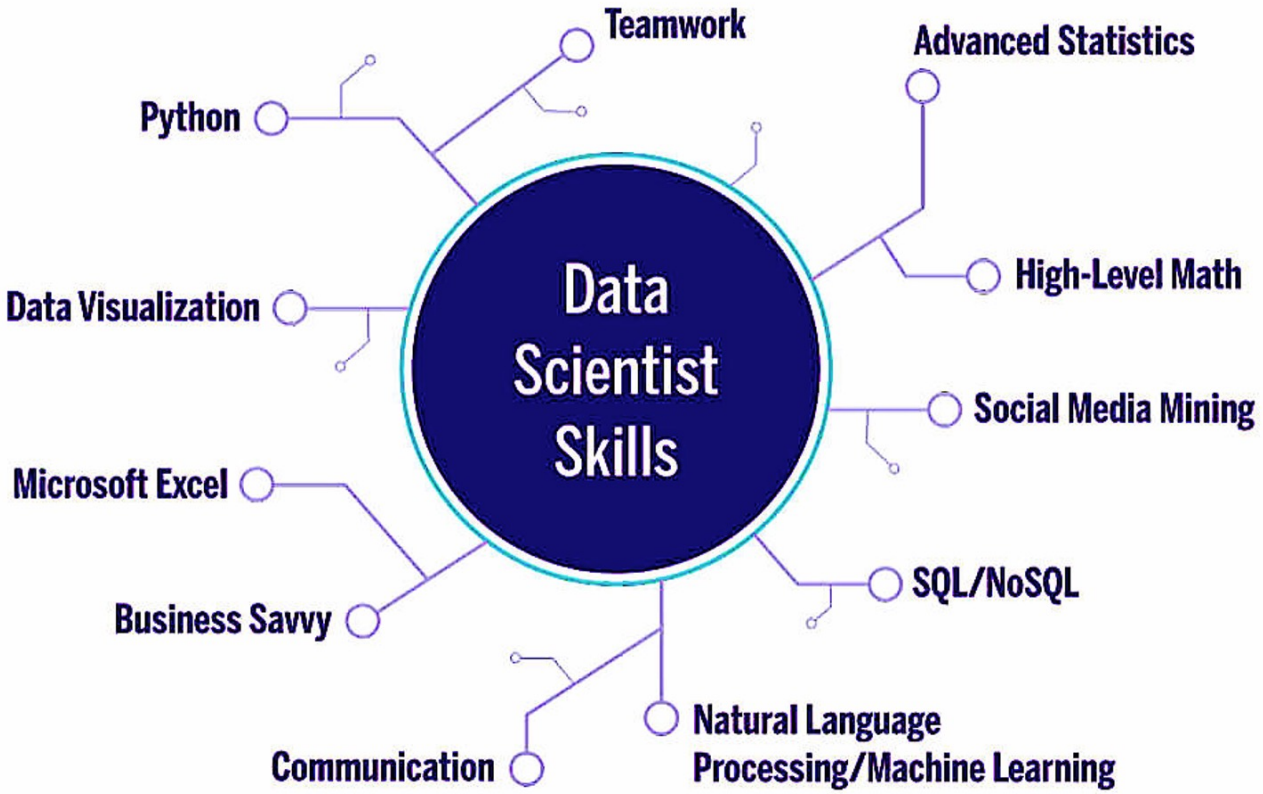
Tại sao khoa học dữ liệu cần cả Toán và Tin?

Toán học đóng vai trò phân tích dữ liệu, xây dựng mô hình dự đoán và cho hiểu biết sâu hơn về dữ liệu.



Tại sao khoa học dữ liệu cần cả Toán và Tin?

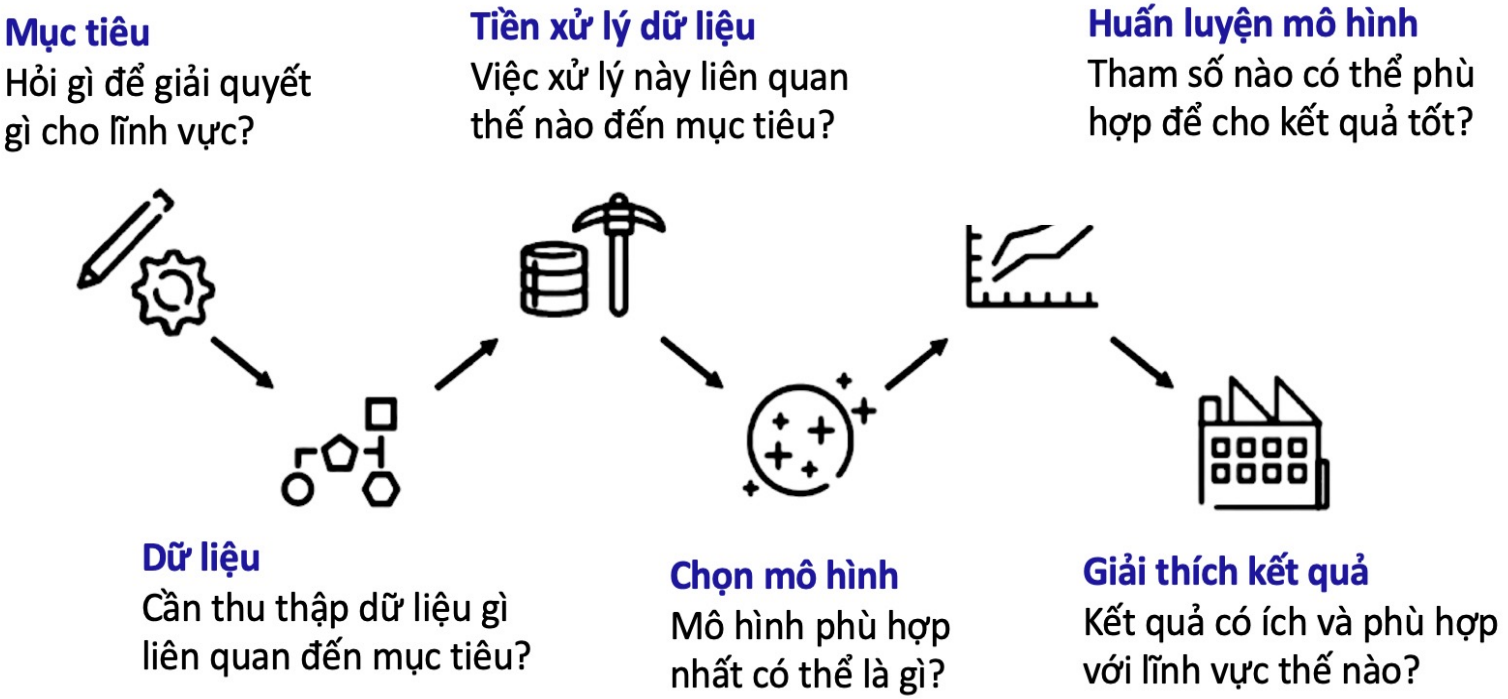
Khoa học máy tính cung cấp phương pháp, kỹ thuật và công cụ để xây dựng và triển khai các mô hình dự đoán và phân tích dữ liệu trong khoa học dữ liệu.



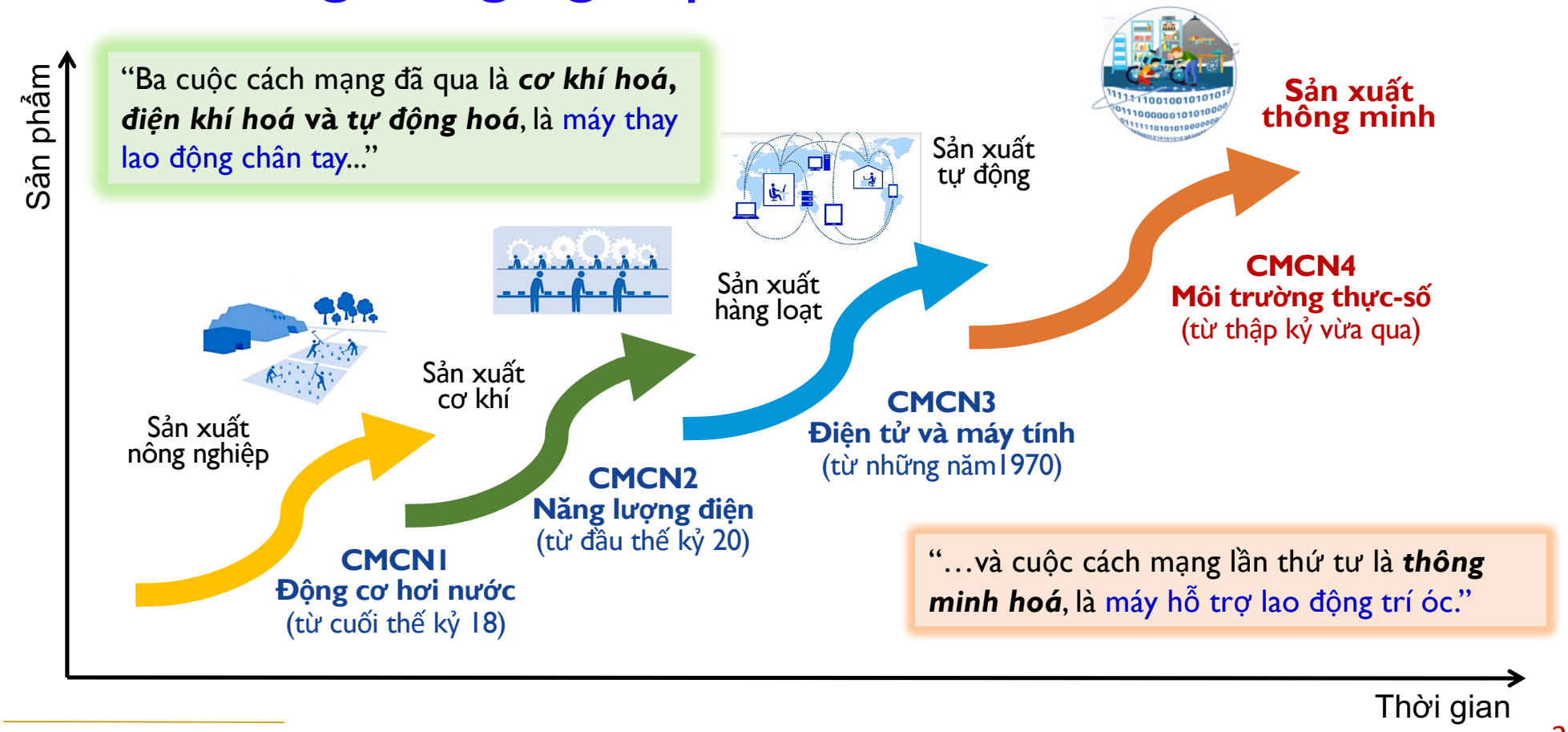
https://bootcamp.berkeley.edu/blog/data-scientist-skills/

Tại sao khoa học dữ liệu cần tri thức của lĩnh vực?

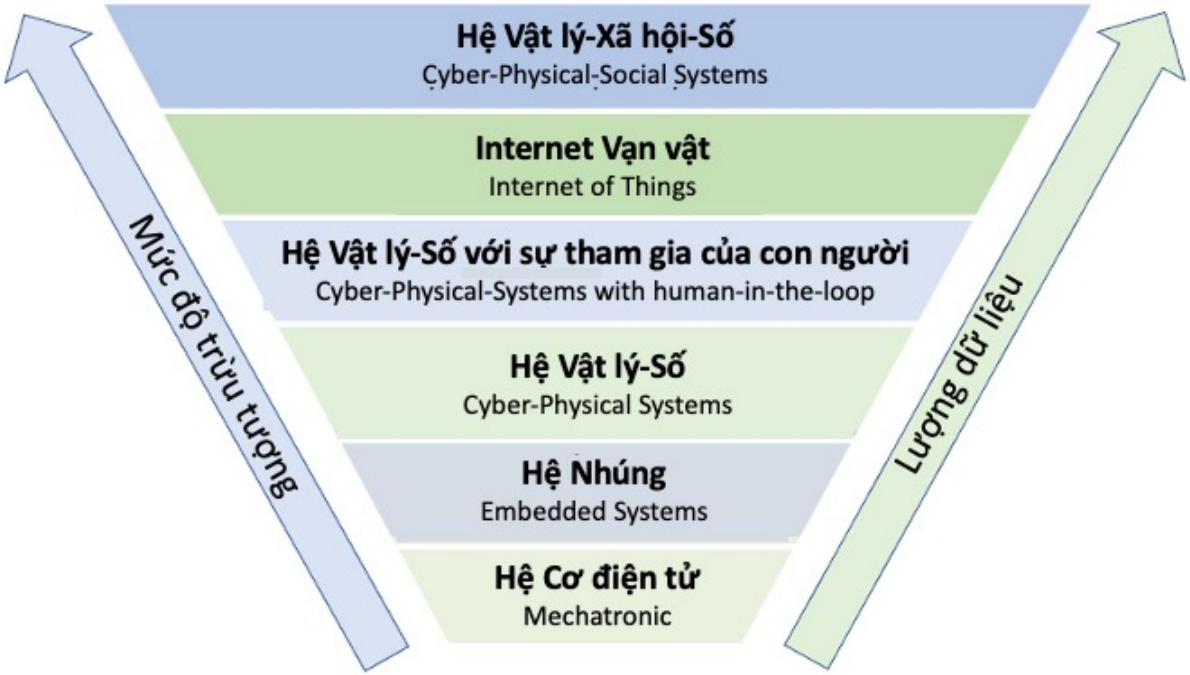
Cho hiểu biết về dữ liệu, chọn biến quan trọng, xây mô hình hợp lý, tối ưu quá trình, và cho hiểu biết sâu hơn kết quả phân tích.

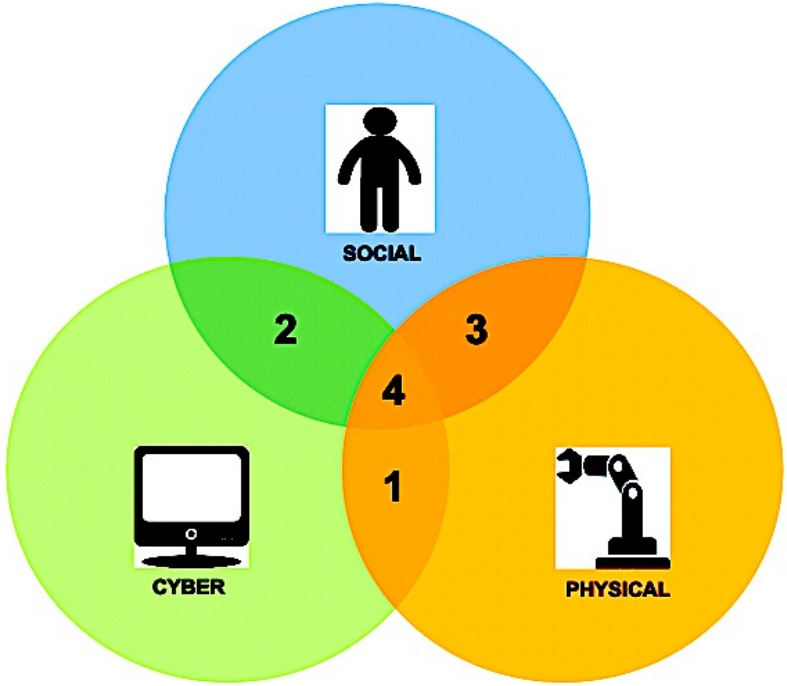


Cách mạng công nghiệp lần thứ tư



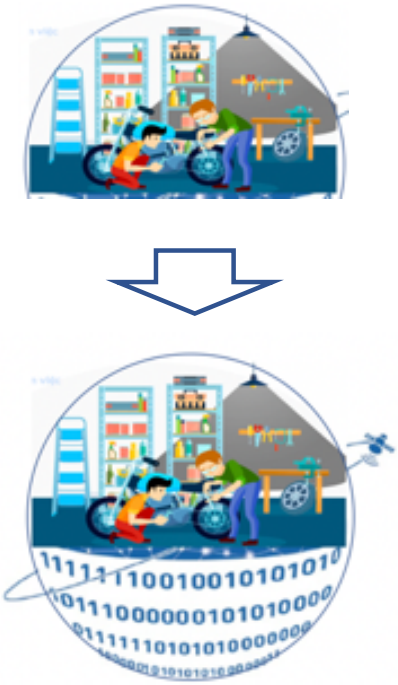
Con người và máy móc hoà nhập trên không gian mạng



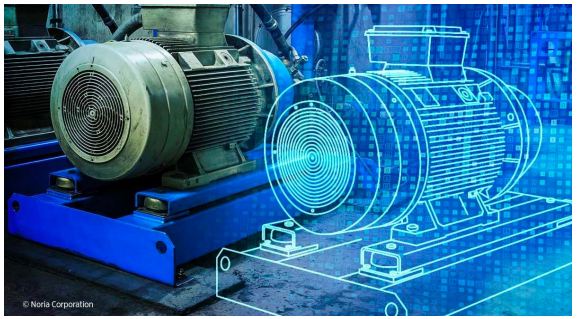


Cyber-Physical—Cyber-Social—Physical-Social → Cyber-Physical-Social

Môi trường thực-số



Môi trường thực-số (môi trường số) là môi trường ta đang sống được nối với không gian mạng, ở đó các thực thể được số hoá tạo thành các phiên bản số (**dữ liệu**) và có thể **kết nối** được với nhau.



Thực thể và phiên bản số (digital twin)

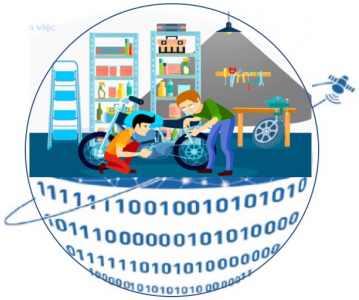
Thực thể kết nối được qua phiên bản số

Chuyển đổi số là gì?

Chuyển đổi số là quá trình thay đổi cách sống và cách làm việc trên môi trường thực-số.

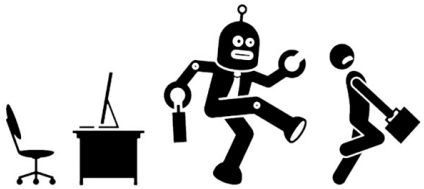
Môi trường thực thể

CHUYỂN ĐỔI SỐ
Dữ liệu + Kết nối



Môi trường thực-số

CƠ HỘI SỐ



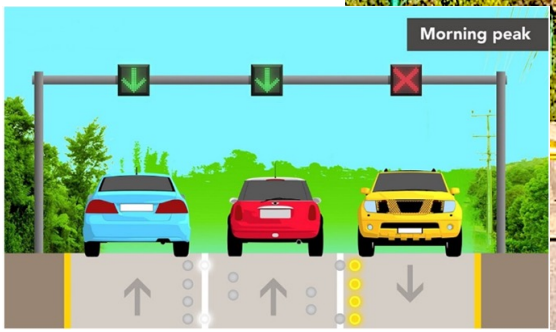
Phá huỷ sáng tạo disruption

THÁCH THỨC SỐ

Thông minh là năng lực nắm bắt và phản ứng hiệu quả đối với sự thay đổi của môi trường.

Thông minh để thích nghi?

Dynamic road
lanes management

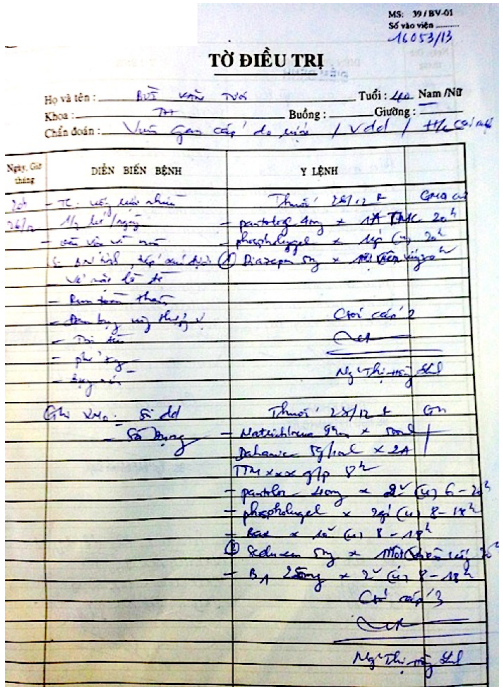


Thông minh = **Hợp lý** và **hiệu quả** trong **điều kiện cụ thể** dựa trên sự **thấu hiểu** (nhờ **dữ liệu** & **kết nối**)

thể chế **đổi mới sáng tạo** — **CON NGƯỜI**

công nghệ số và AI — **CÔNG NGHỆ**

Từ bệnh án giấy đến bệnh án điện tử (EMR)



TỜ ĐIỀU TRỊ

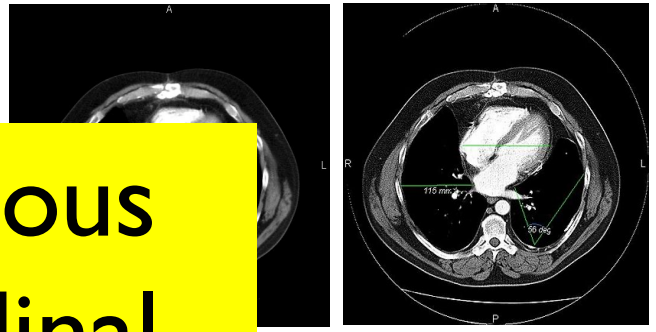
Ngày, Giờ tháng	DIỄN BIẾN BỆNH	Y LỆNH

Bệnh nhân được nhập viện vào Khoa Hô Hấp với triệu chứng ho kèm theo khó thở kéo dài từ 2 tuần trước khi đến viện. Bệnh nhân miêu tả cơn ho nl tă cá K tỉ kh lần/phút, thường, không có nhịp tim bất thường. Phổi: Hơi thở nhanh, có tiếng kéo, giảm dải thở.

Nhận xét của bác sĩ:Tình trạng của bệnh nhân thể hiện một sự gia tăng của triệu chứng hen suyễn, đồng thời có sự kết hợp với viêm phổi cấp.

Heterogeneous and longitudinal

CLINICAL DATA (clinical text)



ay, CT scan, MRI, ... in PACS

			280 - 360	280 - 360	06/10/2016 14:5
MCV	81.2	fL	83.0 - 98.0	83.0 - 98.0	06/10/2016 14:5
MPV	9.6	fL	6.0 - 13.0	6.0 - 13.0	06/10/2016 14:5
Mid#	1.5	GPL	0.2-0.8	0.2-0.8	06/10/2016 14:5
Mid%	21.9	%	5 - 8	5 - 8	06/10/2016 14:5
P - LCR	22.2	%			06/10/2016 14:5
PDW	11.3	fL	6.0 - 10.0	6.0 - 10.0	06/10/2016 14:5
RBC(Hồng cầu)	4.67	/mm^3	4.0 - 5.9	4.0 - 5.9	06/10/2016 14:5
RDW	40.1	%	8.0 - 12.0	8.0 - 12.0	06/10/2016 14:5
THR(Tiểu cầu)	238	/mm^3	150 - 450	150 - 450	06/10/2016 14:5
WBC(Bạch cầu)	6.9	/mm^3	4.0 - 10.0	4.0 - 10.0	06/10/2016 14:5
Tổng phân tích nước tiểu (B					
pH	7.0		4.8-7.4	4.8-7.4	06/10/2016 14:5
BIL (Bilirubin)	Âm tính	umol/L	<3.4	<3.4	06/10/2016 14:5
BLO (Hồng cầu)	VẾT	/µ	<5	<5	06/10/2016 14:5
GLU (Glucose nước tiểu)	Âm tính	mmol/L	3.7 - 6.2	3.7 - 6.2	06/10/2016 14:5
KET (Ketone)	Âm tính	mmol/L	<5	<5	06/10/2016 14:5
LEU (Bạch cầu)	-	/µ	<10	<10	06/10/2016 14:5

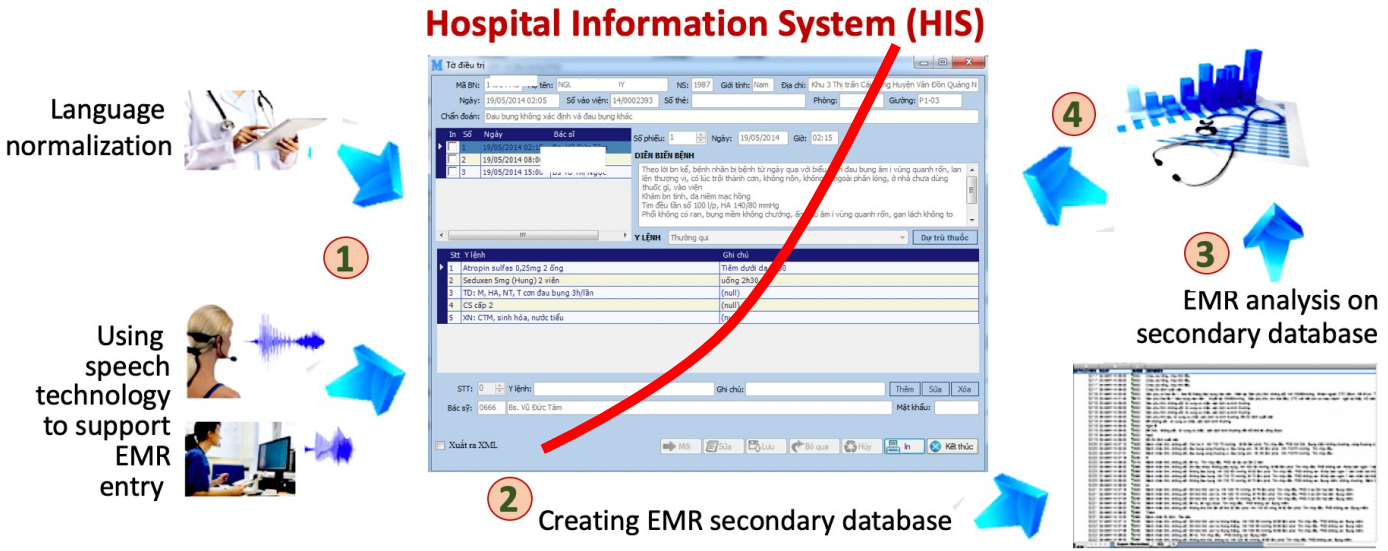
PARA-CLINICAL DATA (numbers)

Mục tiêu nghiên cứu

Đề tài B2016-42-01 và B2019-42-01

"Phát triển các công nghệ lõi để khai thác bệnh án điện tử tiếng Việt"

ĐHQG HCM 2016-2021



1. Phát triển và sử dụng các phương pháp và công cụ để hỗ trợ việc tạo ra bệnh án điện tử.

2. Phát triển các phương pháp để khai thác bệnh án điện tử, đặc biệt với văn bản lâm sàng tiếng Việt.

Project members and collaboraters

Pr. HT Bao (PI)

Pr. CH Tru

Pr. VTN Chau

Dr NV Binh

Pr. Le Ly

Dr. MTM Huyen

Pr. NDCong

Pr. NV Thanh

Pr.Takabayashi.

Pr. Suzuki

Pr. Kanda

16 master students

NH Dang

TA Sang

HH Quan

BQ Minh

PT Dat

P. Ouankhamchan

NT Tùng

NN Quang

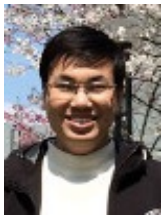
PT Dung

Vo Kiet

ND Phuong

6 PhD students

Hoang K. Hung


Dang T Thai

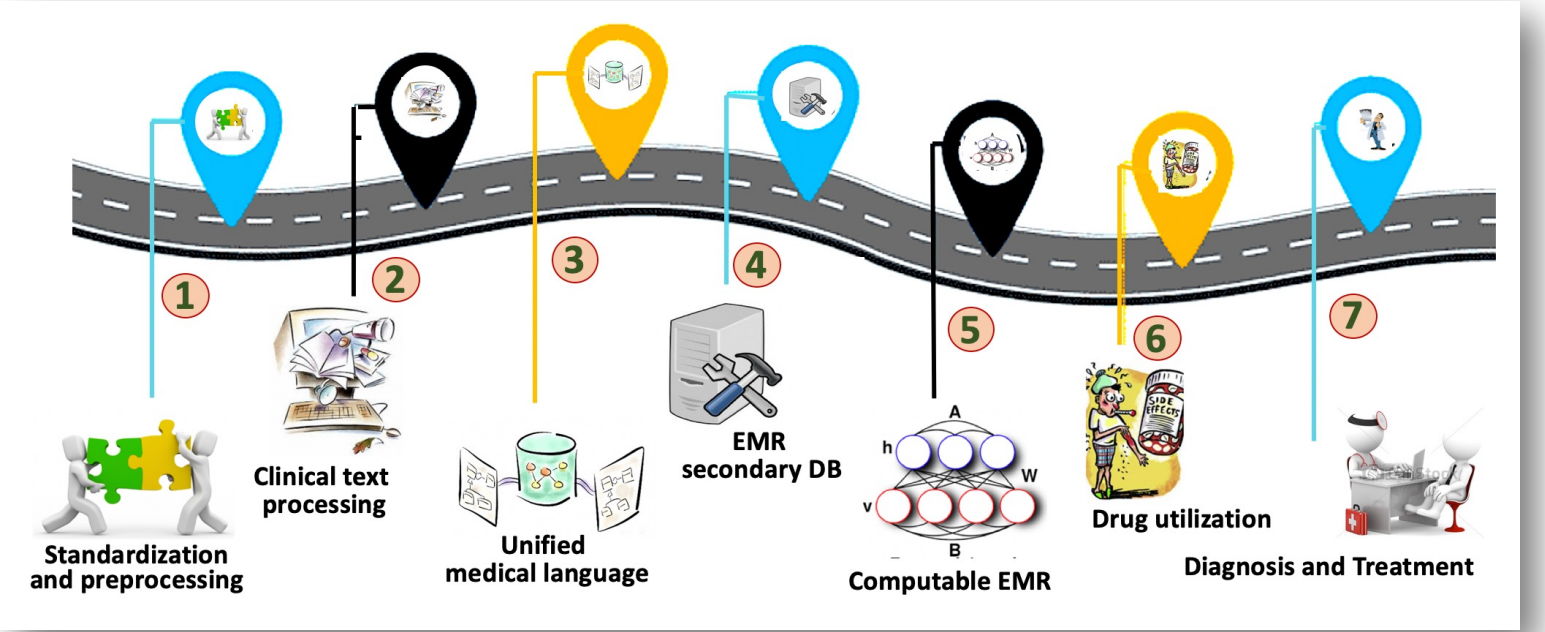
G. Moharasan.

S. Taewijit

S. Nuttapong

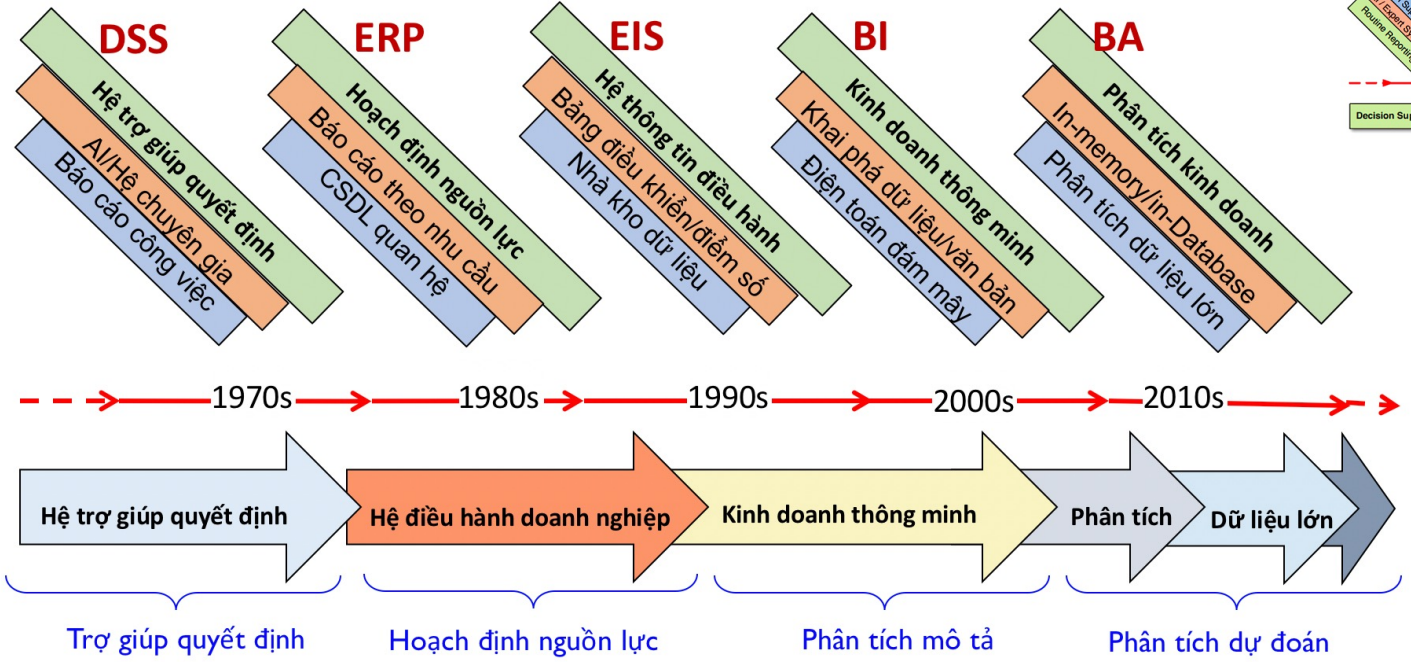
M. Matsuo

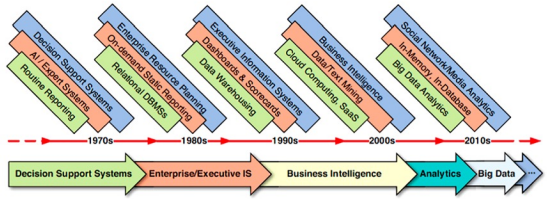
Những bài toán lớn trên lộ trình khai thác BAĐT



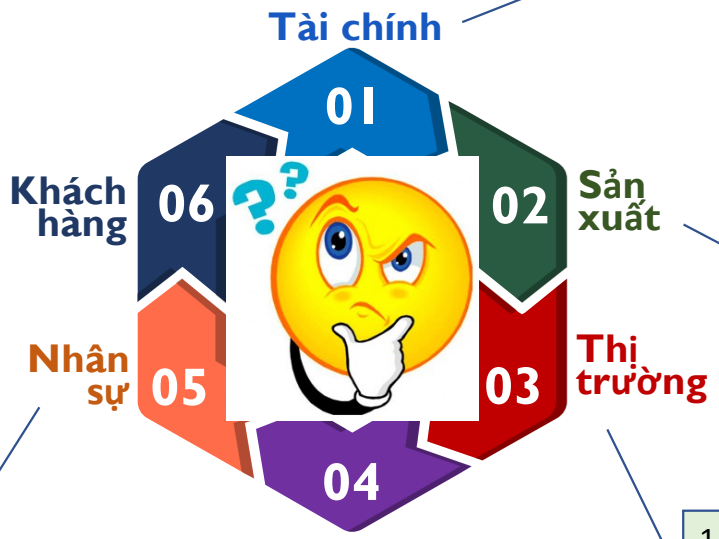
Để tài xác định một lộ trình với những bài toán quan trọng và đưa ra những lời giải cơ bản ban đầu cho chặng đường dài.

Từ hệ trợ giúp quyết định đến phân tích kinh doanh





Sáu bài toán kinh doanh cơ bản



1. Dự đoán doanh thu bán hàng
2. Phân tích lợi nhuận từ khách hàng
3. Phân tích lợi nhuận từ các sản phẩm
4. Phân tích dòng tiền
5. Phân tích các yếu tố tạo giá trị tài chính
6. Phân tích giá trị đem lại cho cổ đông

1. Lập kế hoạch dự án
2. Lập kế hoạch hoạt động
3. Bố trí cơ sở sản xuất
4. Quản lý chất lượng sản phẩm
5. Quản lý chuỗi cung ứng
6. Quản lý hàng tồn kho
7. Quản lý hoạt động dịch vụ

1. Tư vấn sản phẩm và dịch vụ
2. Phân khúc khách hàng
3. Phân tích phương tiện truyền thông xã hội
4. Phân tích khách hàng rời bỏ sản phẩm và dịch vụ

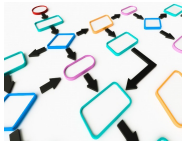
1. Phân tích khả năng
2. Phân tích năng lực
3. Phân tích kênh tuyển dụng
4. Phân tích hiệu suất của nhân viên
5. Phân tích lãnh đạo
6. Phân tích nhân viên rời bỏ công ty

1. Phân tích sự hài lòng của khách hàng
2. Phân tích vòng đời của khách hàng
3. Phân tích phân khúc khách hàng
4. Phân tích mức độ tương tác của khách hàng
5. Phân tích sự thu hút khách hàng

© Bao Ho. AI, DS và DX

Doanh nghiệp và kinh tế số: Phân tích kinh doanh

Phân tích khuyến cáo
(cần kinh doanh thế nào?)



Phân tích dự báo
(kinh doanh sẽ thế nào?)

Phân tích mô tả
(kinh doanh đã và đang thế nào?)

Phân tích kinh doanh

Dữ liệu mở

Phân tích kinh doanh dựa trên phân tích dữ liệu (**AI và khoa học dữ liệu**) là con đường kinh tế số của các doanhnghiệp.

Luật cho kinh tế số

?

Nhân sự

Khách hàng

Bán hàng

Thị trường

Tài chính

Sản xuất

Kinh doanh

Take home message

- Việt Nam ba lần 'lỡ tàu' các cuộc cách mạng công nghiệp.
- Cơ hội vô giá: Vượt lên nhờ thay đổi chính mình.
- Cơ hội cuối cùng: Nhiều chục năm mới có một lần.
- "The winner takes it all"
- "Từ AK đến AI"?