AI nào cho Việt Nam?

Hồ Tú Bảo
Viện Nghiên cứu Cao cấp về Toán
Viện John von Neumann, ĐHQG HCM
Japan Advanced Institute of Science and Technology
bao@viasm.edu.vn, bao.ho@jvn.edu.vn, bao@jaist.ac.jp

JAIST
JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY
1990

JVN Institute
Vietnam National University HCM

VIASM
Vietnam Institute for Advanced Study in Mathematics

Content

1. **AI?**
2. AI nào cho Việt Nam?

Thời chuyển đổi số
Digital transformation time

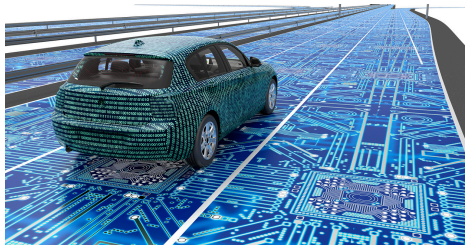
- **Số hoá** (digitization/digitalization): *Tạo ra phiên bản số* của các thực thể, từ dạng liên tục hoặc vật lý.
- **Công nghệ số** (CNS) Công nghệ tạo ra các thiết bị, phương pháp, hệ thống... để khai thác các thực thể số hoá.
- **Thời chuyển đổi số:** mọi vật dần được số hoá và rất nhiều công nghệ số được phát triển và sử dụng.

Chuyển đổi số = Số hoá + Công nghệ số

https://www.zdnet.com/article/what-is-digital-transformation-everything-you-need-to-know-about-how-technology-is-reshaping/

Thời chuyển đổi số
Digital transformation time

Thế giới các thực thể

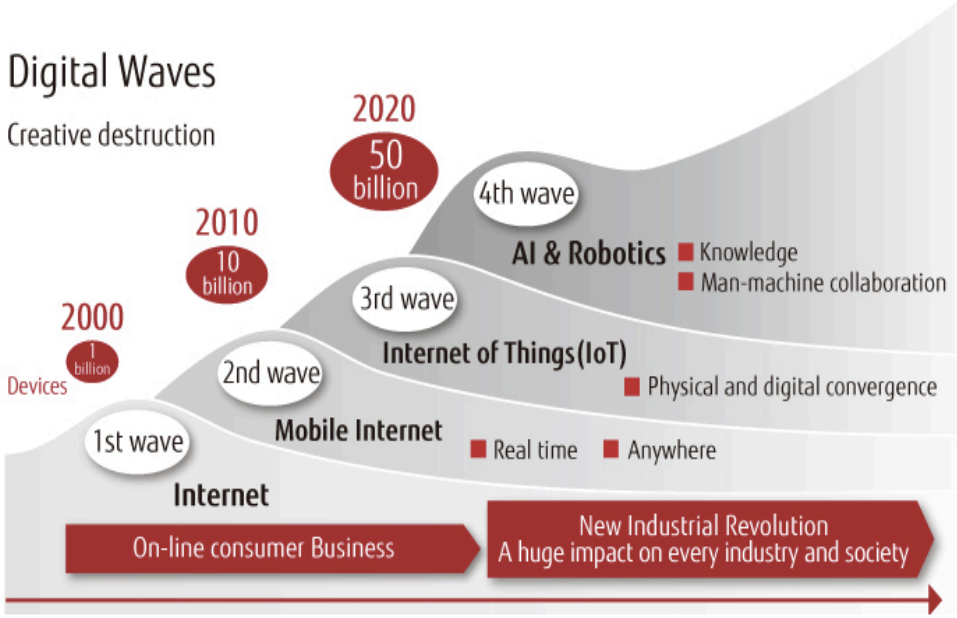


Thí dụ: 1 triệu xe kinh doanh trên toàn quốc đã gắn GPS, ~500GB/day.

Mọi điều xảy ra trên thế giới các thực thể ⟷ Tính toán, điều khiển trên không gian số

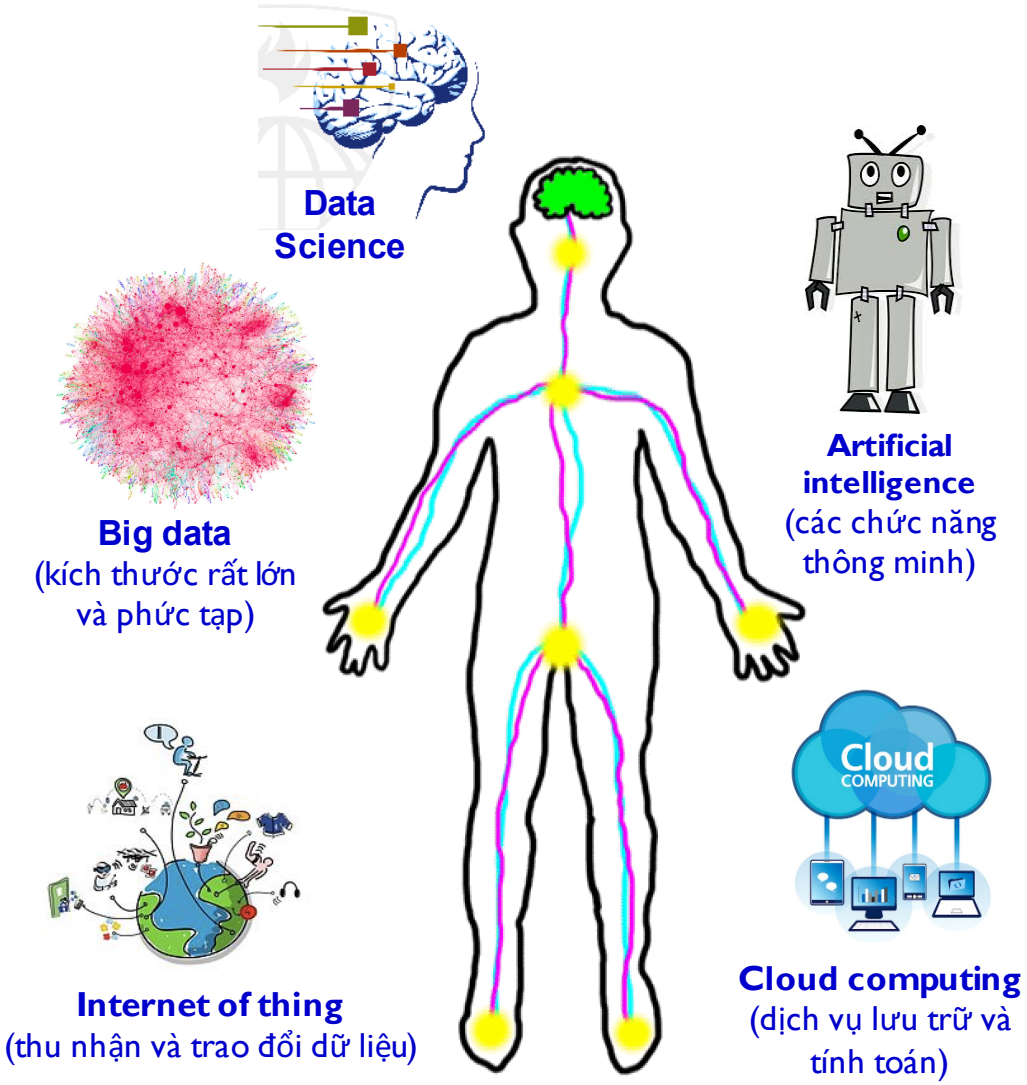
Cách mạng số hoá
Thế giới thực thể-không gian số

- '**Phiên bản số**' các thực thể: Biểu diễn các thực thể bằng '0' và '1' trên máy tính (digital version)
- **Hệ thống không gian số-thế giới thực thể (cyber-physical system): kết nối** các thực thể qua 'phiên bản số' của chúng.



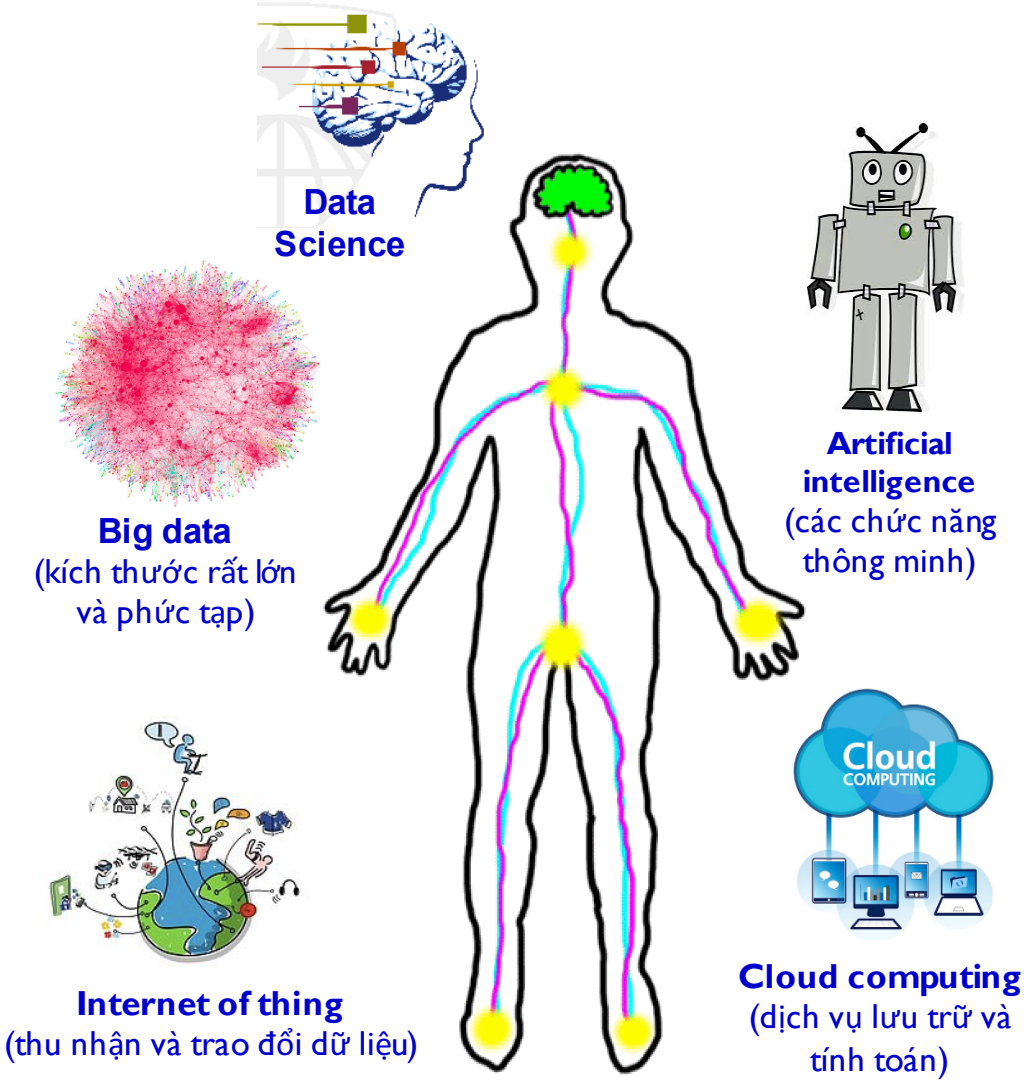
How digital technology will transform the world, Fujitsu Journal, 1.2016

Nhiều công nghệ số đột phá đồng thời

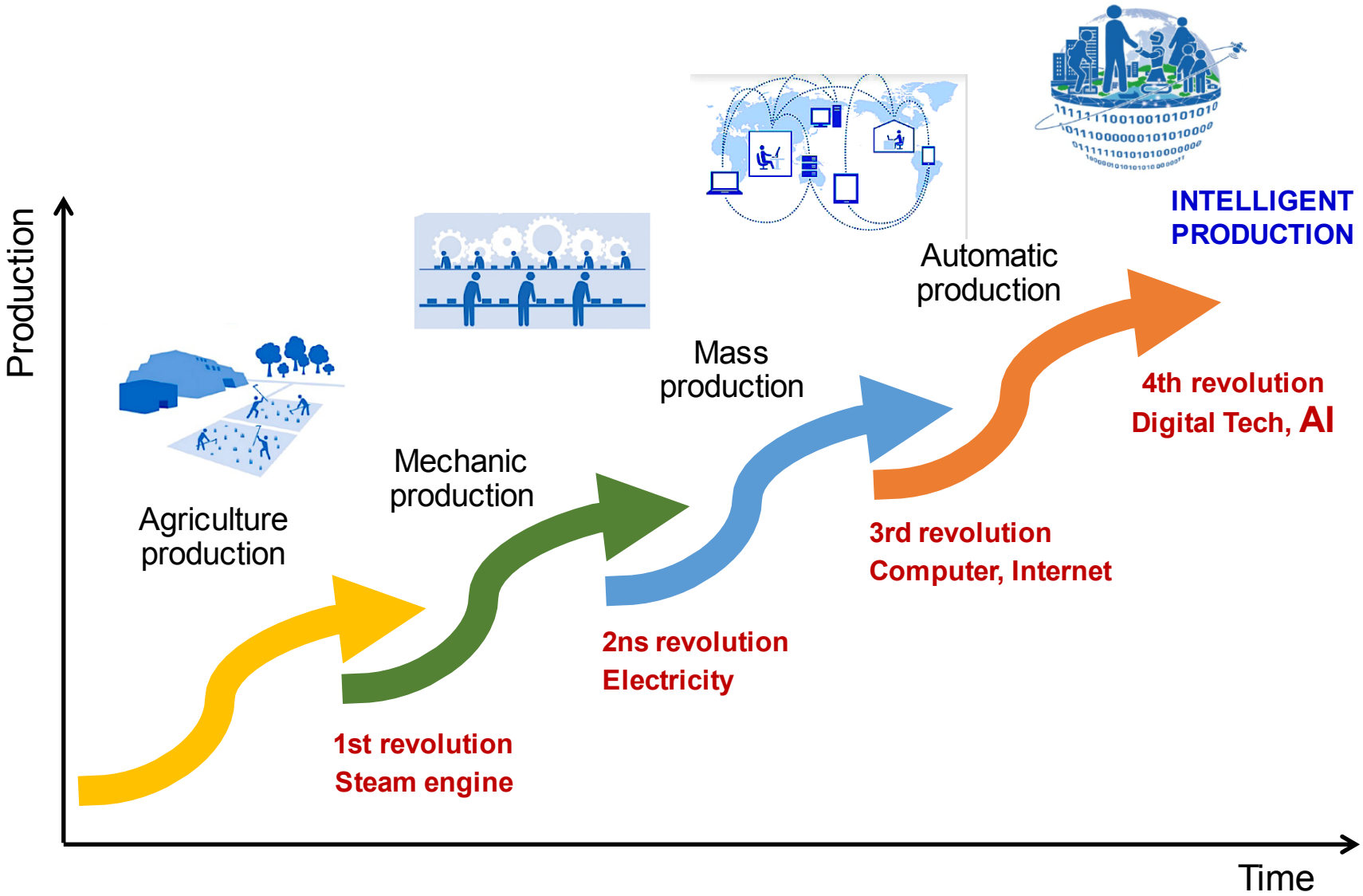


Nhiều công nghệ số đột phá đồng thời

- **Điện toán đám mây**: Môi trường
- **Dữ liệu lớn**: Năng lượng
- **Internet vạn vật**: Huyết mạch
- **Trí tuệ nhân tạo**: Chức năng thông minh
- **Khoa học dữ liệu**: "Bộ não" phân tích dữ liệu để hỗ trợ đưa ra quyết định.



Những cuộc cách mạng công nghiệp



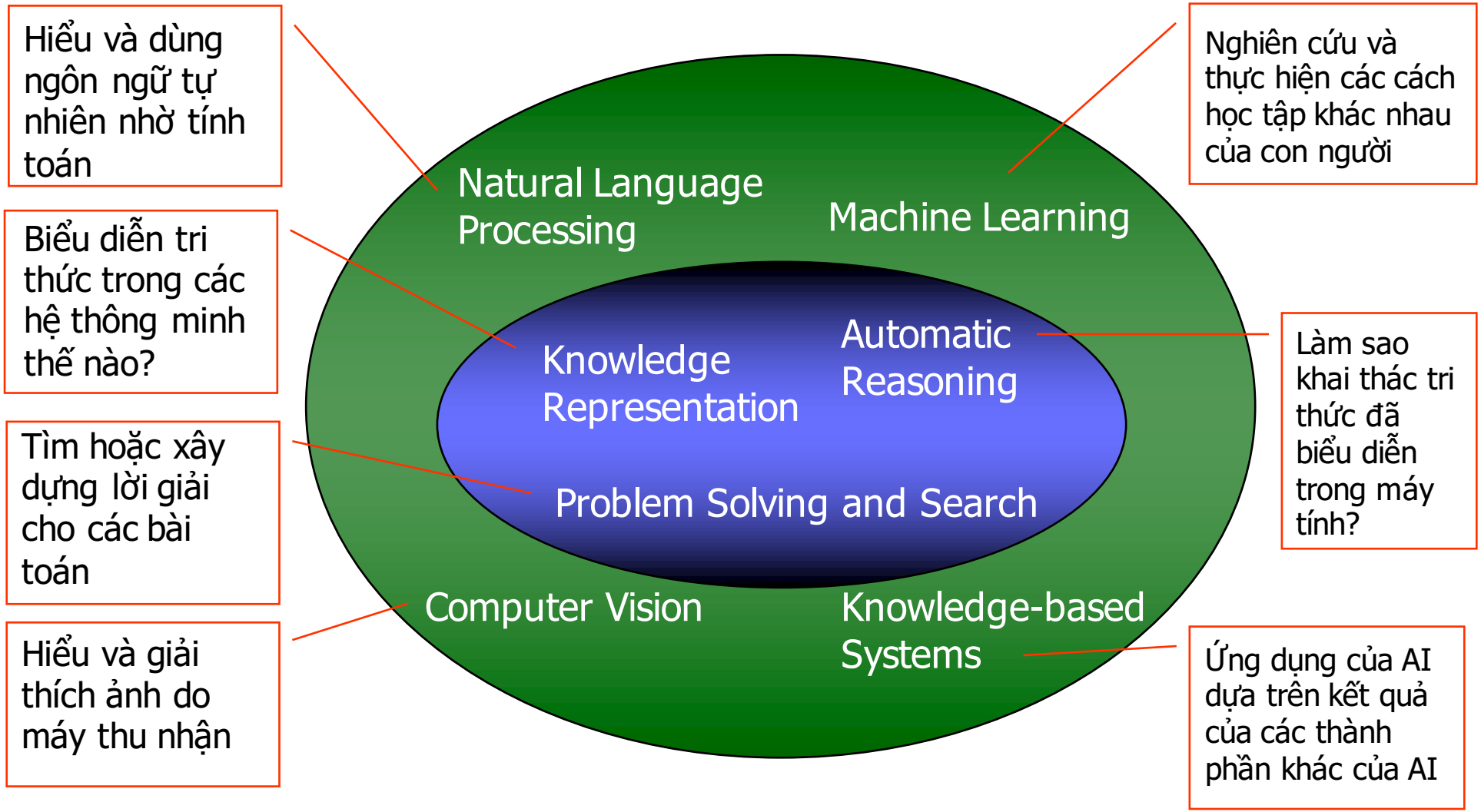
What is intelligence?

- no agreed-upon scientific **meaning** of intelligence
- some dictionary **definition** of intelligence

"the capacity to meet situations, especially if new or unforeseen, by a rapid and effective adjustment of behavior" (Webster) → khả năng đáp ứng các tình huống, đặc biệt là tình huống mới hoặc không lường trước được, bằng cách điều chỉnh hành vi nhanh chóng và hiệu quả

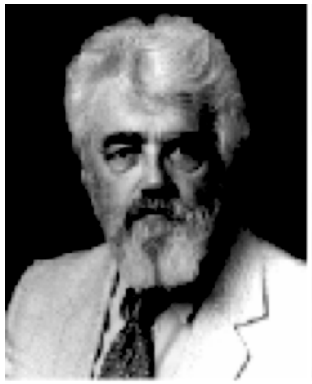
"the ability to learn, understand, and think about things" (Longman) → khả năng học hỏi, hiểu và suy nghĩ về mọi thứ

Elements of artificial intelligence

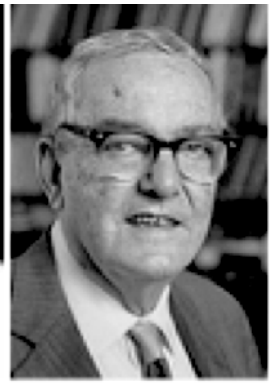


Lịch sử ngành AI

- Hội nghị mùa hè1956 tại trường Dartmouth ở Mỹ, đặt tên AI.
- AI "fathers": Ba trong bốn người 29 tuổi, cả bốn nhận giải Turing của ACM, Simon nhận cả giải Nobel kinh tế.



John McCarthy



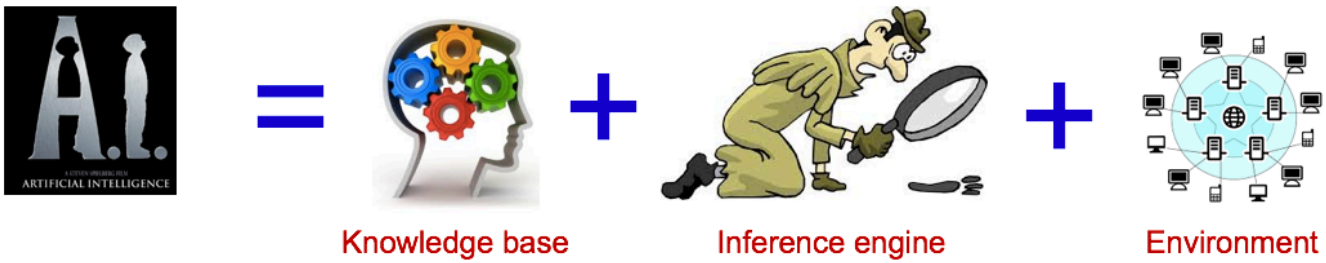
Herbert Simon

Donald Michie

Gặp Minsky ở PRICAI 96, Cairns

Artificial Intelligence – Trí tuệ nhân tạo

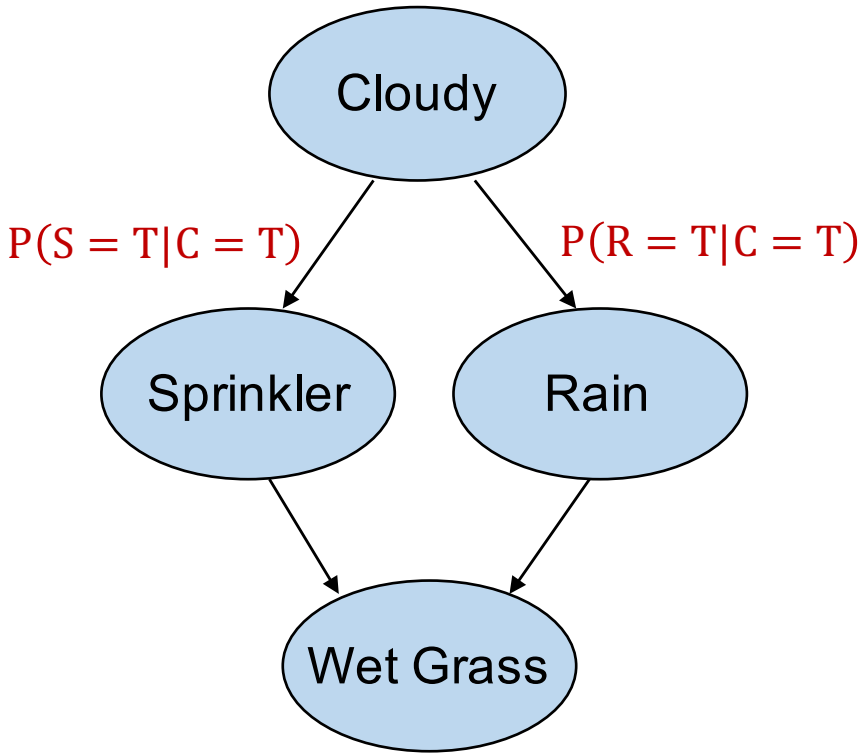
- Lĩnh vực làm cho máy (máy tính) hoạt động như có trí thông minh của con người (suy luận, giải quyết vấn đề, nhận thức, hiểu ngôn ngữ, học tập...).
- AlphaGo, hiểu ngôn ngữ, nhận dạng tiếng nói, chẩn đoán ung thư, ô-tô tự lái...



Sherlock Holmes lập luận

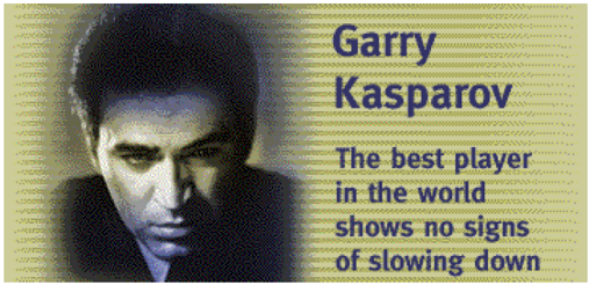
Tại sao cỏ ướt ngoài sân?

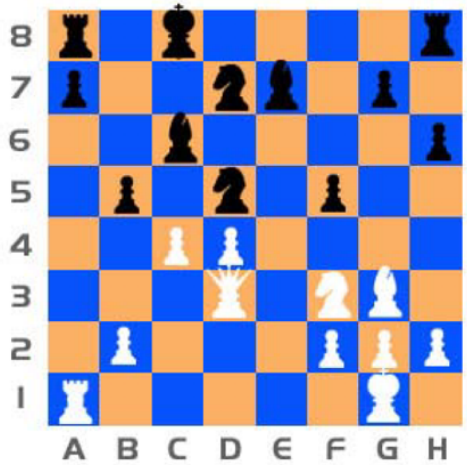
- Thám tử Holmes rời nhà
  - Thấy cỏ ướt trước sân nhà.
  - Có hai lý do: có thể trời mưa hoặc vòi phun nước nhà ông Holmes đã phun suốt đêm.
- Thám tử Holmes nhìn lên trời và thấy nhiều mây xám
  - Khi trời nhiều mây và thường thì vòi phun nước luôn tắt nên khả năng trời mưa nhiều hơn.
  - Holmes kết luận rằng có nhiều khả năng là mưa gây ra cỏ ướt.

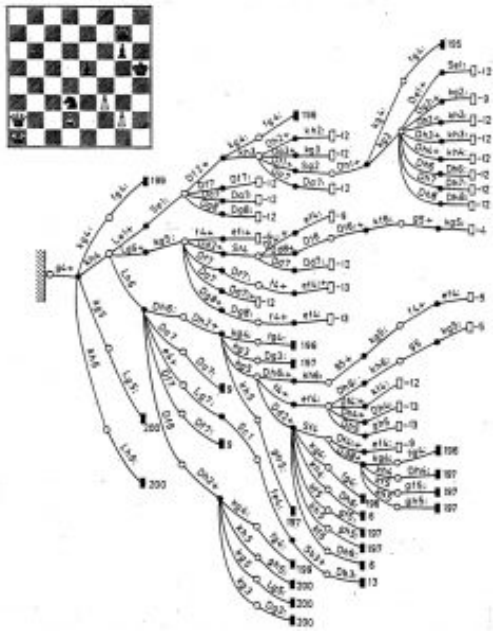


Máy tính đánh cờ

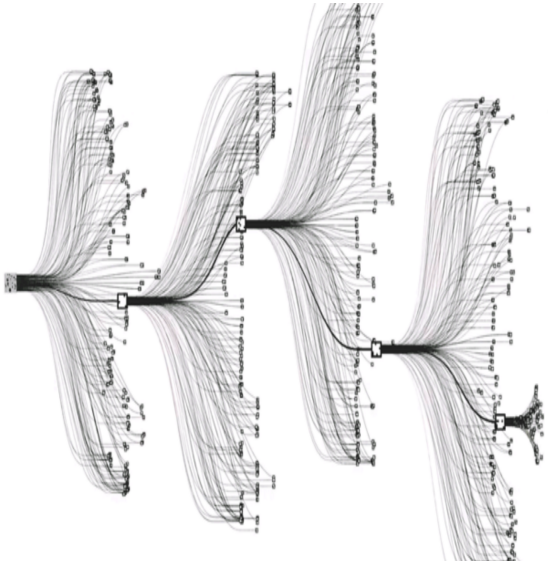
Ngày 1 tháng năm 1997, IBM Deep Blue thắng Kasparov ở ván thứ sáu, đánh dấu một bước ngoặt trong lịch sử cờ vua, một bước ngoặt của công nghệ trước ngưỡng cửa thế kỷ mới.







Developed by Google DeepMind, played in Seoul, South Korea between 9 and 15 March 2016. Won all.

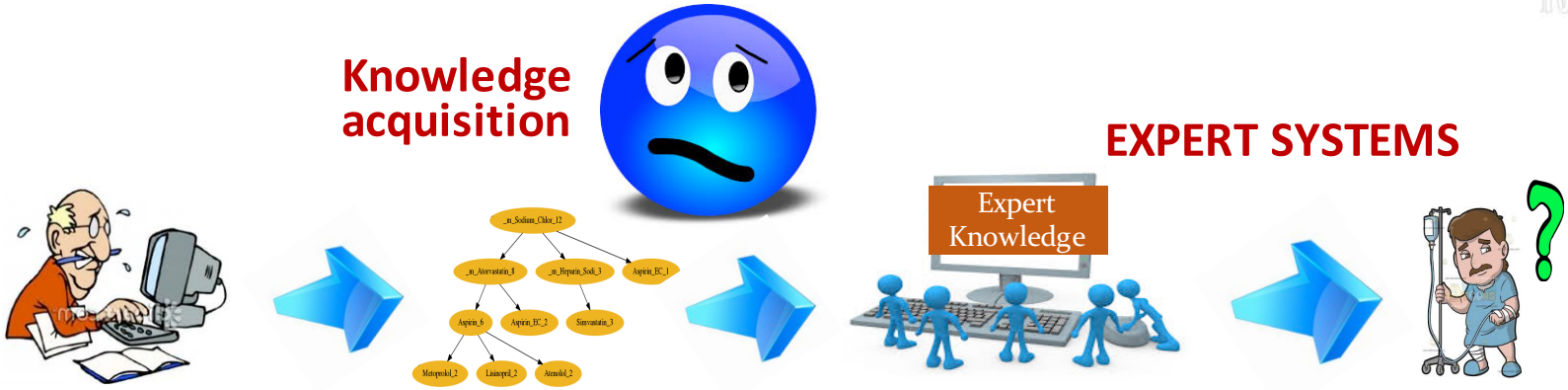


19 October 2017, AlphaGo Zero, a version created without using data from human games

Expert systems in early days of AI

the deduction approach

Một hệ chuyên gia là một hệ máy tính có thể đưa ra quyết định và hành động như một chuyên gia trong một lĩnh vực cụ thể.

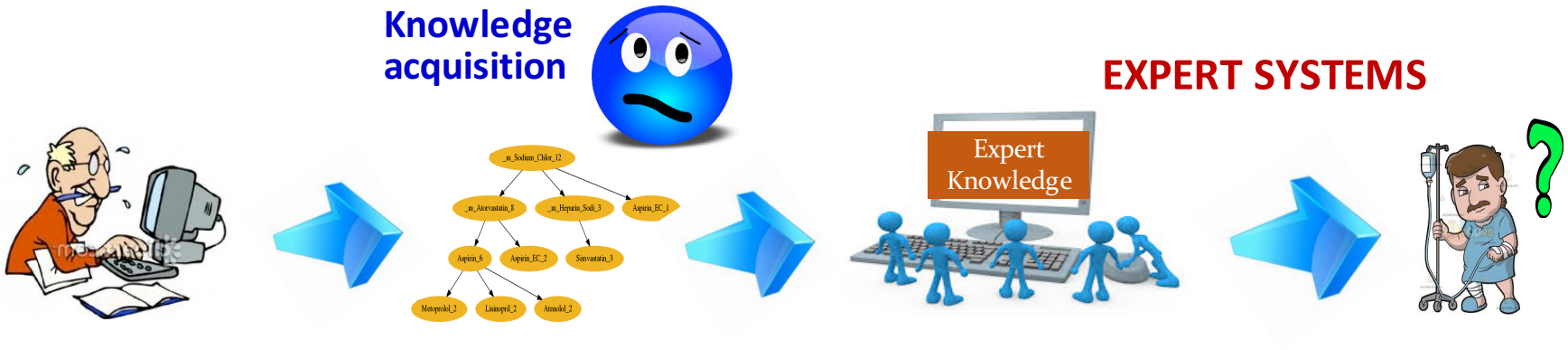


MYCIN (Shortliffe, Feigenbaum, 1979): Infection Diagnosis.

IF
1. the infection is primary bacteremia, and
2. the site of the culture is one of the sterile sites, and
3. the suspected portal of entry of the organism is the gastro intestinal tract

THEN there is suggestive evidence (0.7) that the identity of the organism is bacteroides.

Fifth Generation of Computer Systems (FGCS)

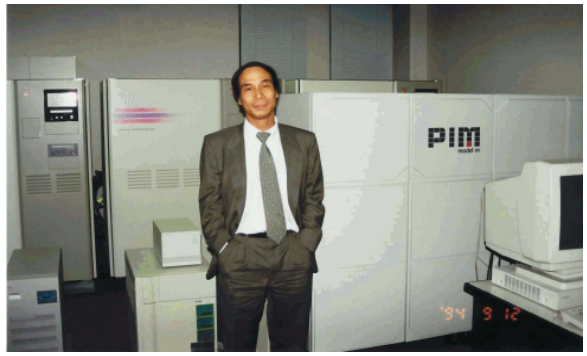


General rules or Principles → Deduction → Specific Cases

FGCS (1982-1992)

Program = Set of axioms.
Computation = Proof of a statement from axioms.
(logic programming)

PIM (Parallel and Inference Machine)



JAIST's PIM (IS center, 1992-1997)

EMRs in medicine

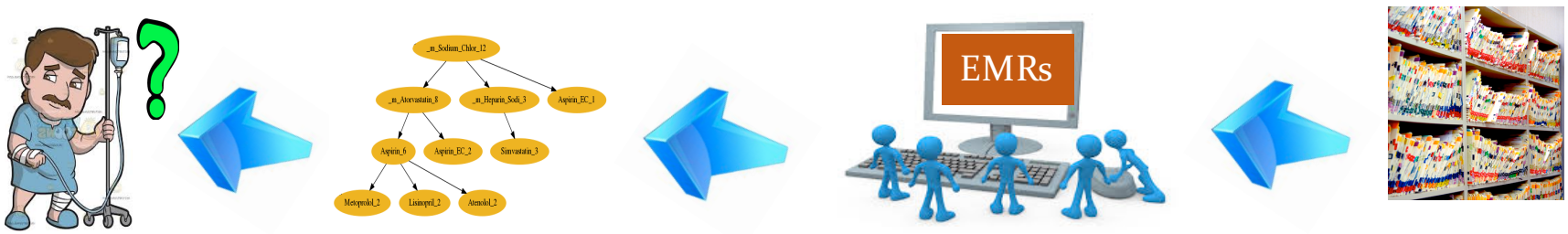
the induction approach (data-driven approach)

General rules or Principles

Deduction →

← Induction

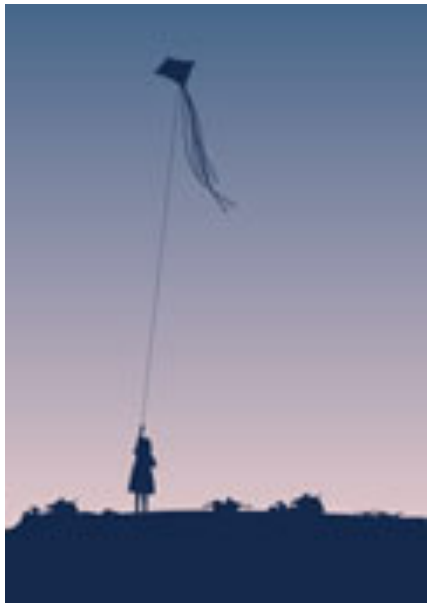
Specific Cases



Digitalization

Natural language processing

- Goal: automated language understanding
  - this isn't possible
  - instead, go for subgoals of text analysis, e.g.,
    - word sense disambiguation
    - phrase recognition
    - semantic associations
- Common current approach (trend):
  - statistical analyses over very large text collections



Marvin Minsky: "Consider a word like "string" or "rope." No computer today has any way to understand what those things mean. For example, you can pull something with a string, but you cannot push anything. You can tie a package with string, or fly a kite, but you cannot eat a string or make it into a balloon. In a few minutes, any young child could tell you a hundred ways to use a string – or not to use a string – but no computer knows any of this."

Natural language processing

Lexical / Morphological Analysis

Word segmentation

Tagging

Chunking

Syntactic Analysis

Grammatical Relation Finding

Named Entity Recognition

Word Sense Disambiguation

Semantic Analysis

Reference Resolution

Discourse Analysis

text

meaning

Shallow parsing

The woman will give Mary a book

POS tagging

The/Det woman/NN will/MD give/VB
Mary/NNP a/Det book/NN

chunking

[The/Det woman/NN]$_{NP}$ [will/MD give/VB]$_{VP}$
[Mary/NNP]$_{NP}$ [a/Det book/NN]$_{NP}$

relation finding

subject

[The woman] [will give] [Mary] [a book]

i-object

object

Sự nhập nhằng của ngôn ngữ

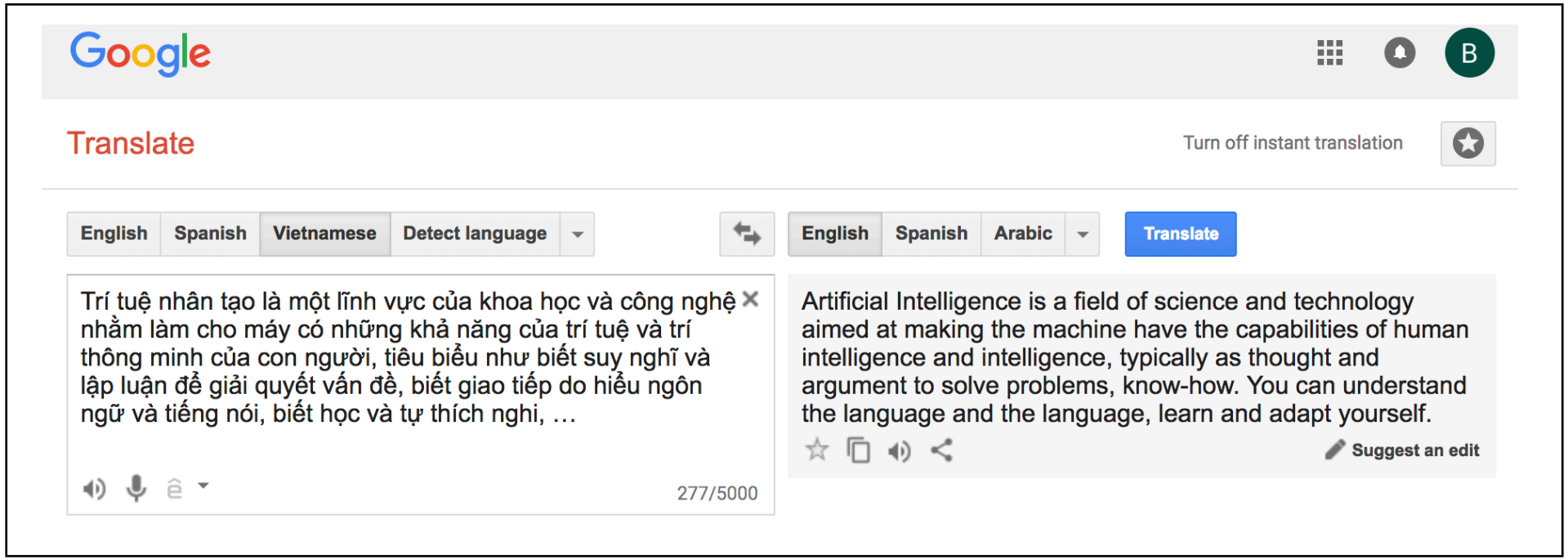
- Câu này có bao nhiêu nghĩa?

Ông già đi nhanh quá

- Tuỳ theo cách phân cụm từ và nghĩa của từ

1. [Ông già] [đi] [nhanh quá] → The old man walks too fast
   → My father walks too fast
2. [Ông già] [đi] [nhanh quá] → The old man died too fast
   → My father died too fast
3. [Ông] [già đi] [nhanh quá] → You get old too fast
   → Grandfather gets old too fast

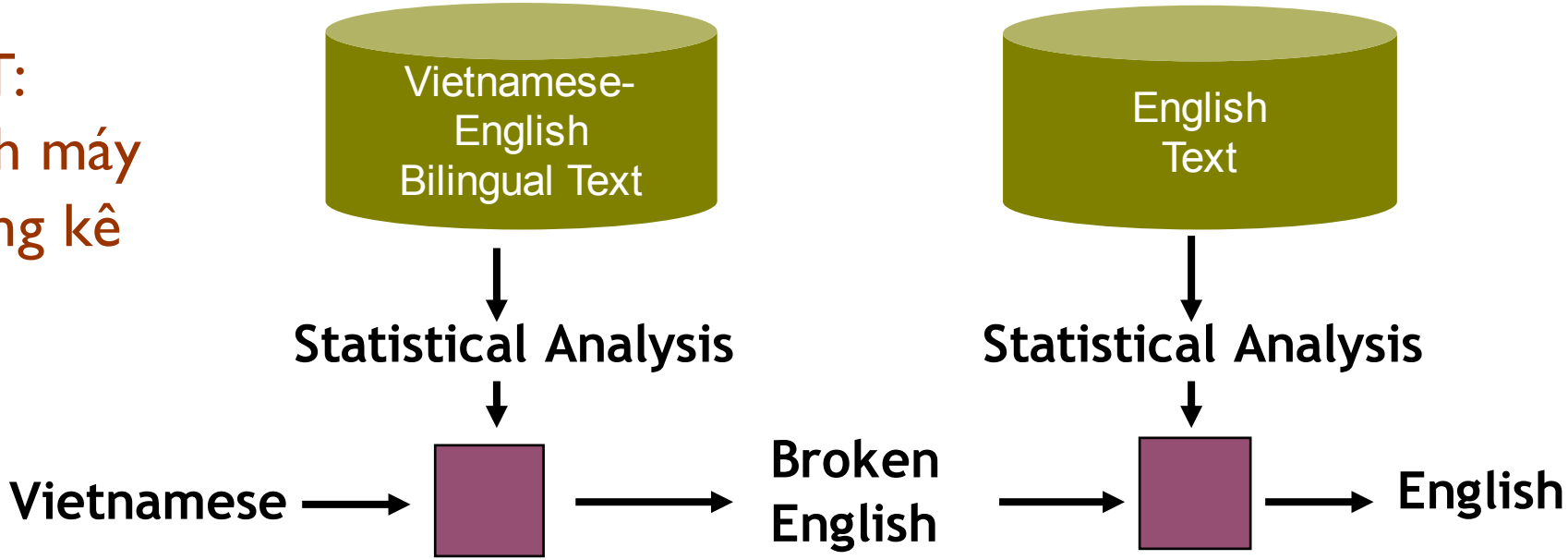
Dịch máy – Machine translation



https://translate.google.com

Machine translation

SMT:
Dịch máy
thống kê

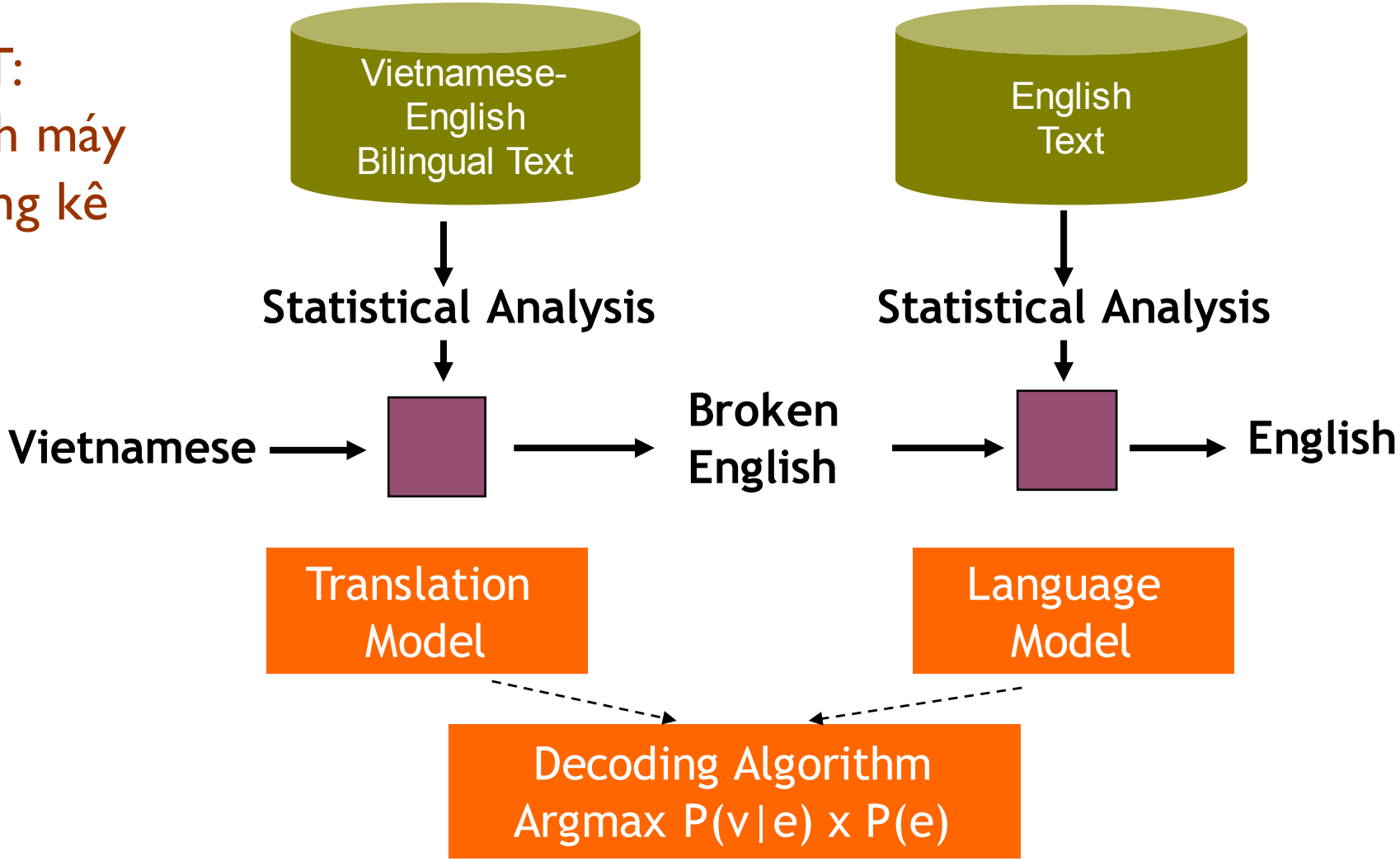


Ông già đi nhanh quá → Died the old man too fast
The old man too fast died
The old man died too fast
Old man died the too fast → The old man
died too fast

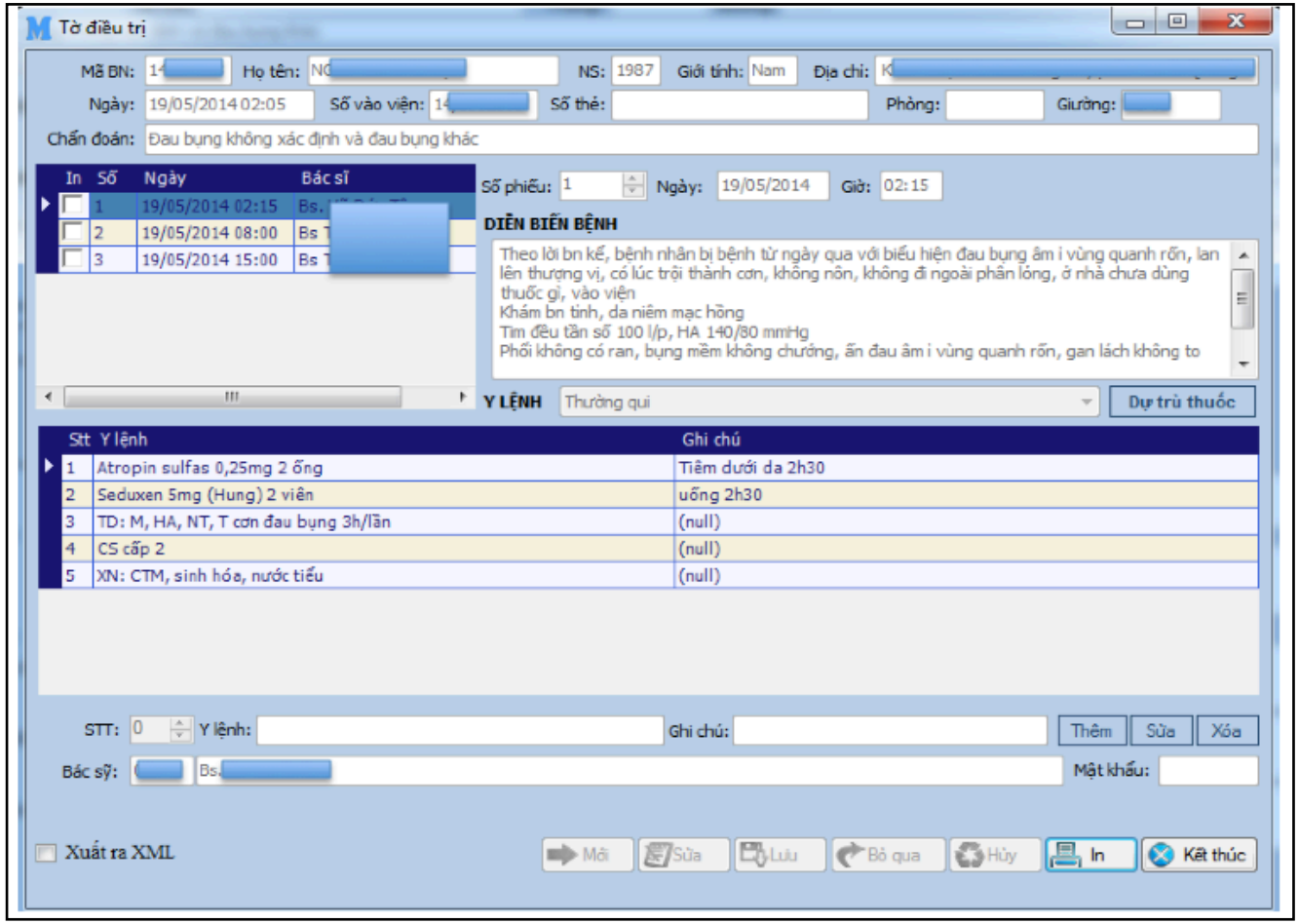
Adapted from tutorial on SMT, K. Knight and P. Koehn

Machine translation

SMT:
Dịch máy
thống kê



Nhận dạng tiếng nói

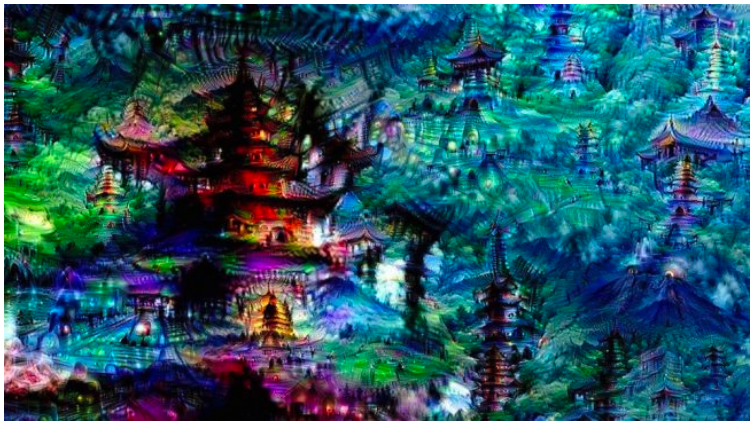


https://voris.skdd.org/

Từ đột phá học máy đến đột phá AI

"Rất nhiều người làm các hệ AI nay đã nhận ra rằng, đối với rất nhiều ứng dụng, việc huấn luyện một hệ thống từ các thí dụ đầu vào-đầu ra [machine learning] để có quyết định hành động là dễ hơn rất nhiều việc soạn sẵn các quyết định mong muốn cho mọi tình huống có thể xảy ra [expert systems].

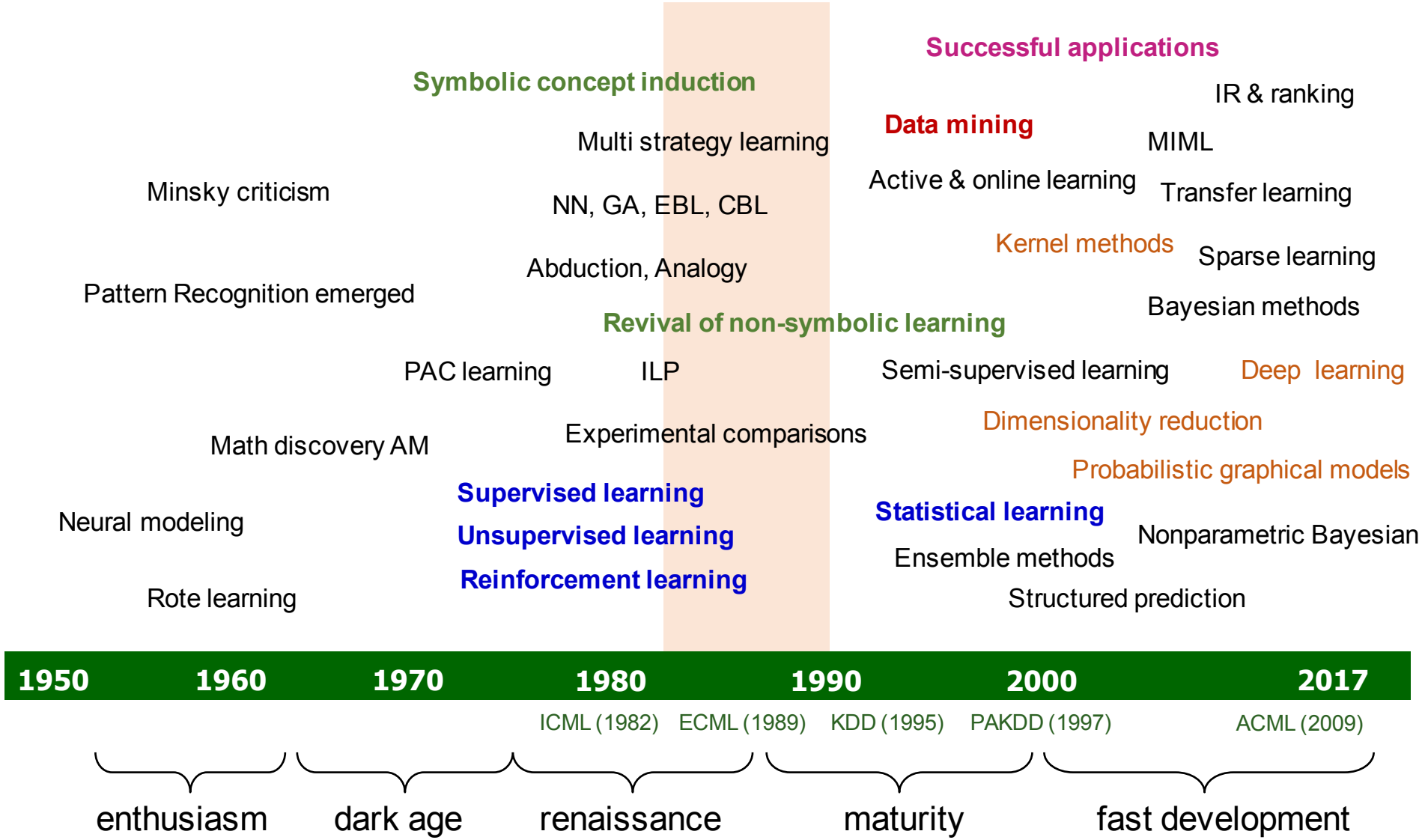
Reinforcement learning led to AlphaGo's stunning victory over a human Go champion last year.



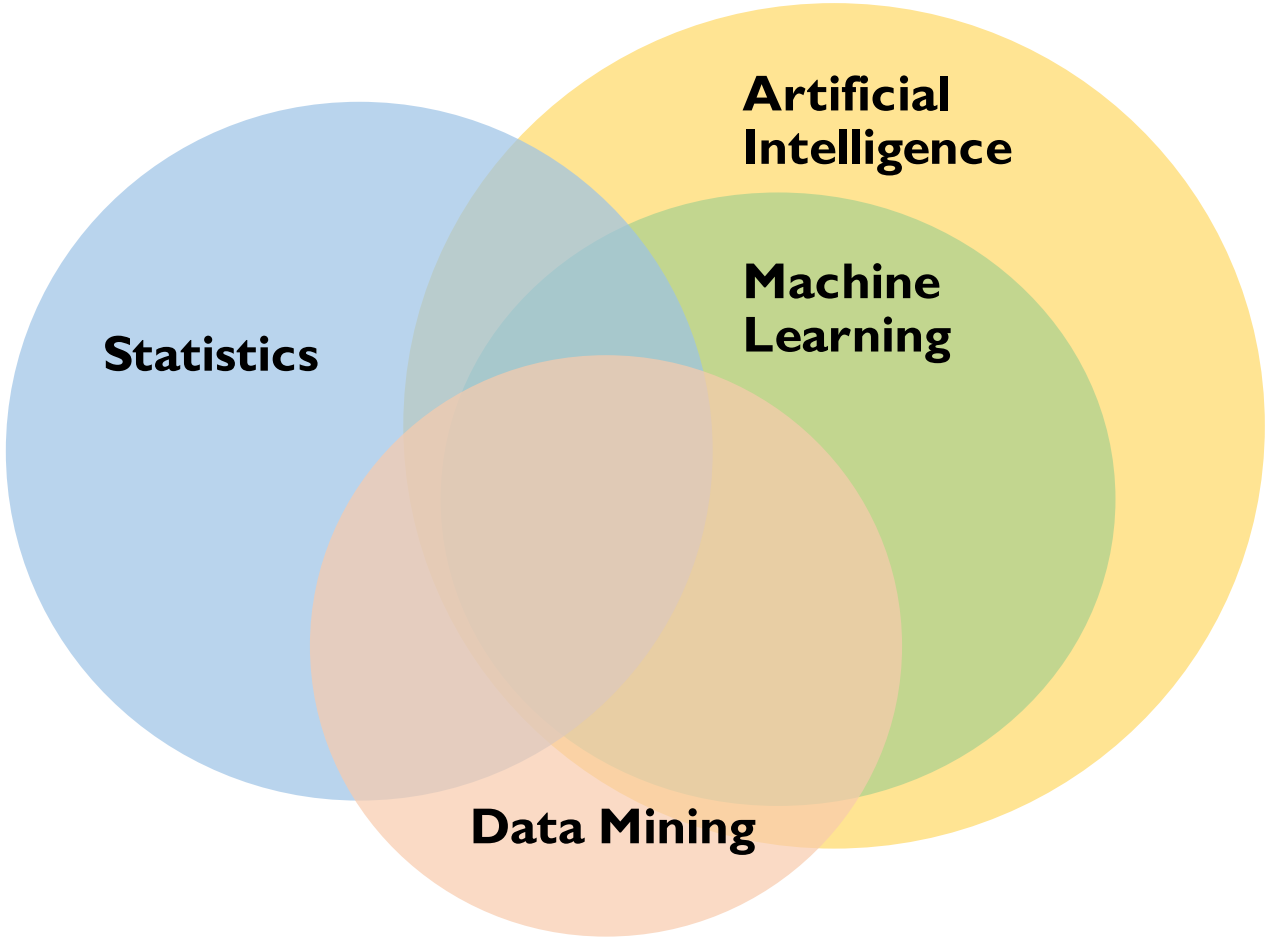
An image created by a Google Neural Network

- M.I. Jordan, T. Mitchell. Machine Learning: Trends, perspectives, and prospects. *Science*, 349 (6245), 255–260, 2015.
- AI4LIFE program: 17-18/33 talks are about or strongly related to machine learning

Development of machine learning

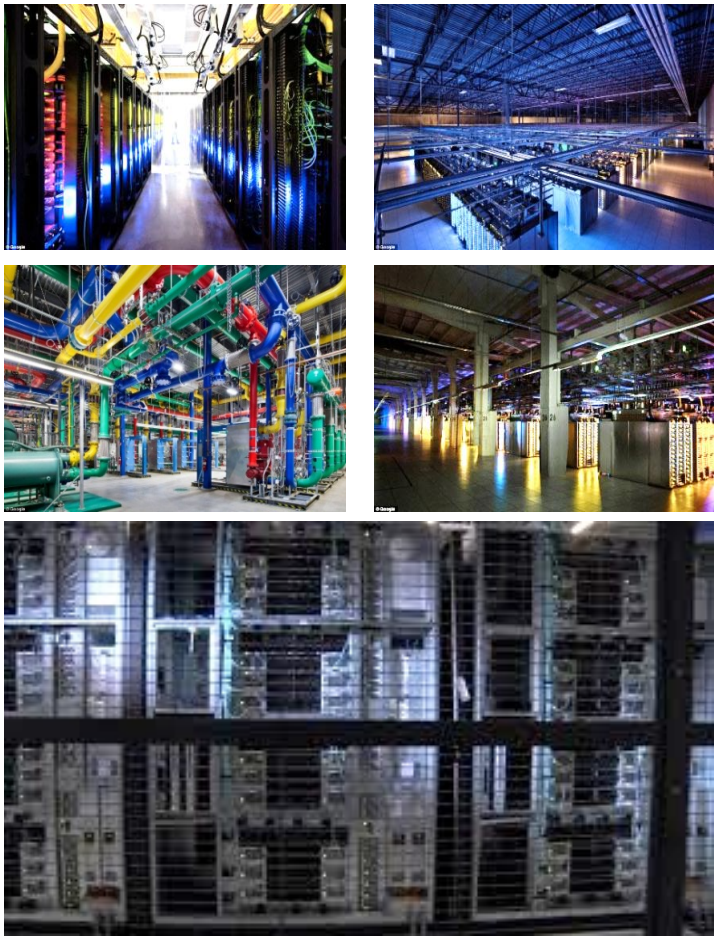


Machine learning, data mining, statistics, and artificial intelligence



Cloud Storage và BigQuery của Google

Google Data Center

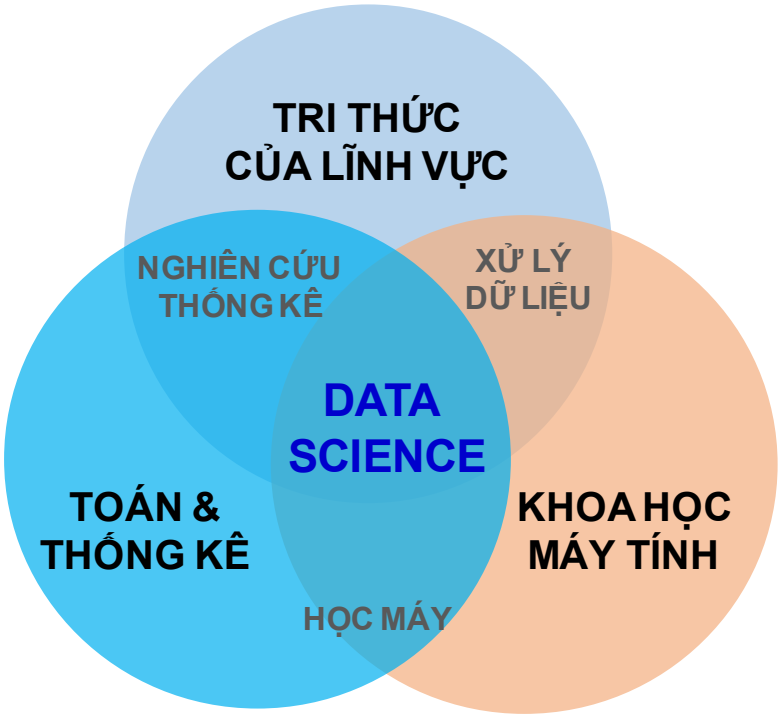


- Công nghệ: BigQuery (Tableau), Cloud Storage.
- Machine learning core
  - Logistic & linear regression, general convex losses
  - Infusion of L1 and L2 regularization
  - On-the-fly curvature estimation
- System infrastructure
  - MapReduce for parallelism
  - Multiple cores and threads per computer
  - Data stored in compressed column-based form

Problem	Number of raw features (M)	Non-zero weights (M)	Fraction of non-zero weights
A	868	20	2.3%
B	333	8	2.4%
C	1762	252	14.3%
D	2172	372	17.1%

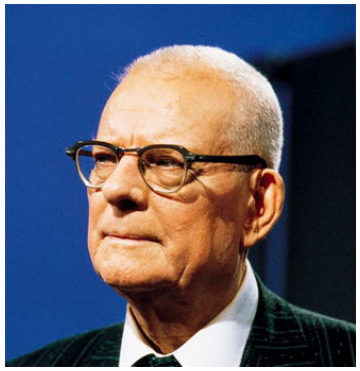
Singer Yoram, keynote at ACML'14

Khoa học Dữ liệu

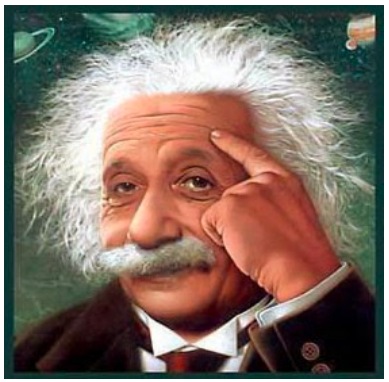


Khoa học về phân tích dữ liệu

"In God we trust. All others bring data".
"Ta tin Thượng đế. Ngoài ra, là dữ liệu".
W.E. Deming

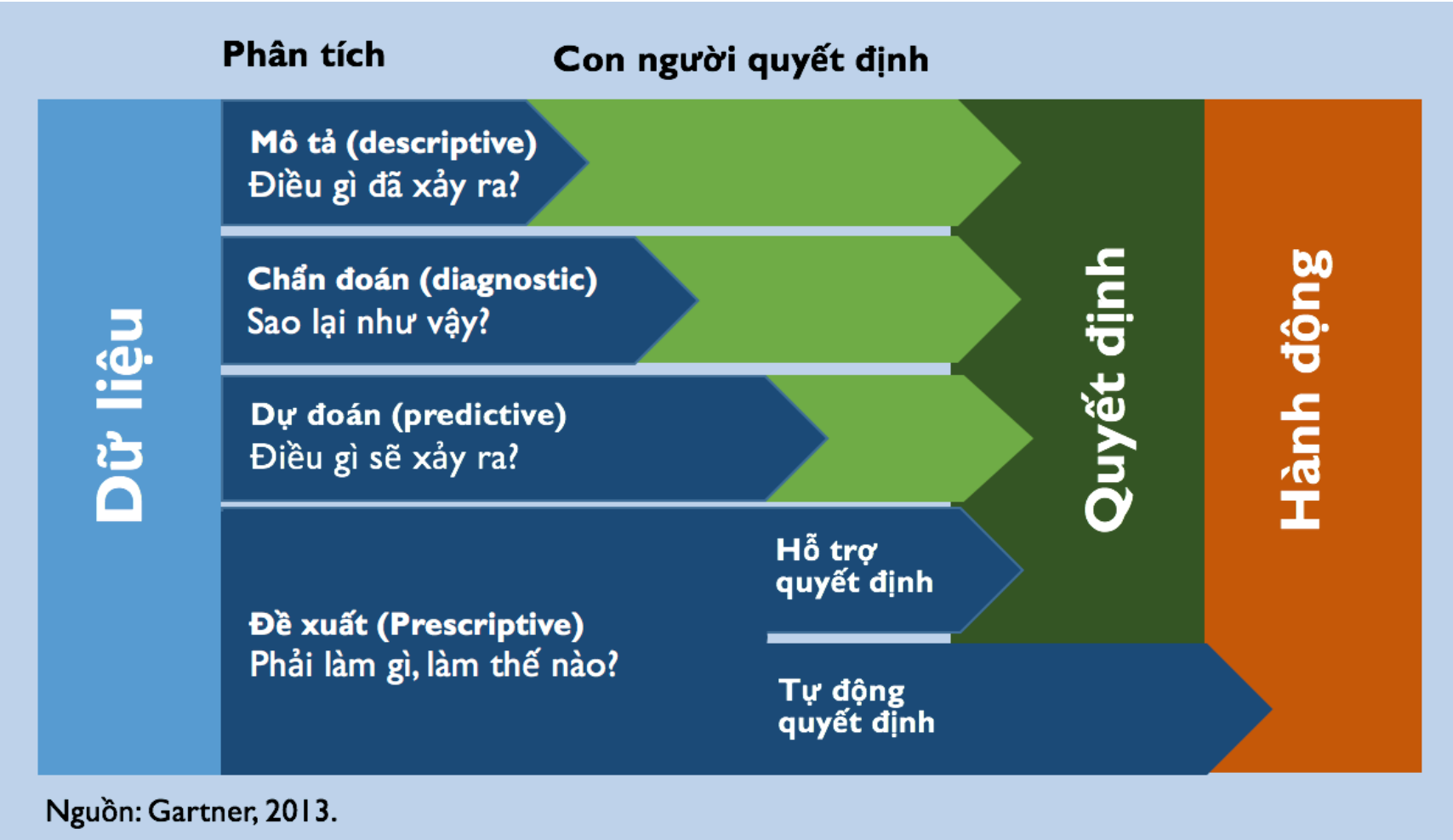


"We cannot solve problems by using the same kind of thinking we used when we created them"
Ta không thể giải quyết các vấn đề với chính cách nghĩ ta đã dùng khi đặt vấn đề
Albert Einstein



Kết hợp của Toán học và Tin học là cốt lõi của khai thác dữ liệu

Khoa học dữ liệu: Cơ sở của quyết định



Nguồn: Gartner, 2013.

Data science is **the essence of using data**

Content

1. AI?
2. **AI nào cho Việt Nam?**

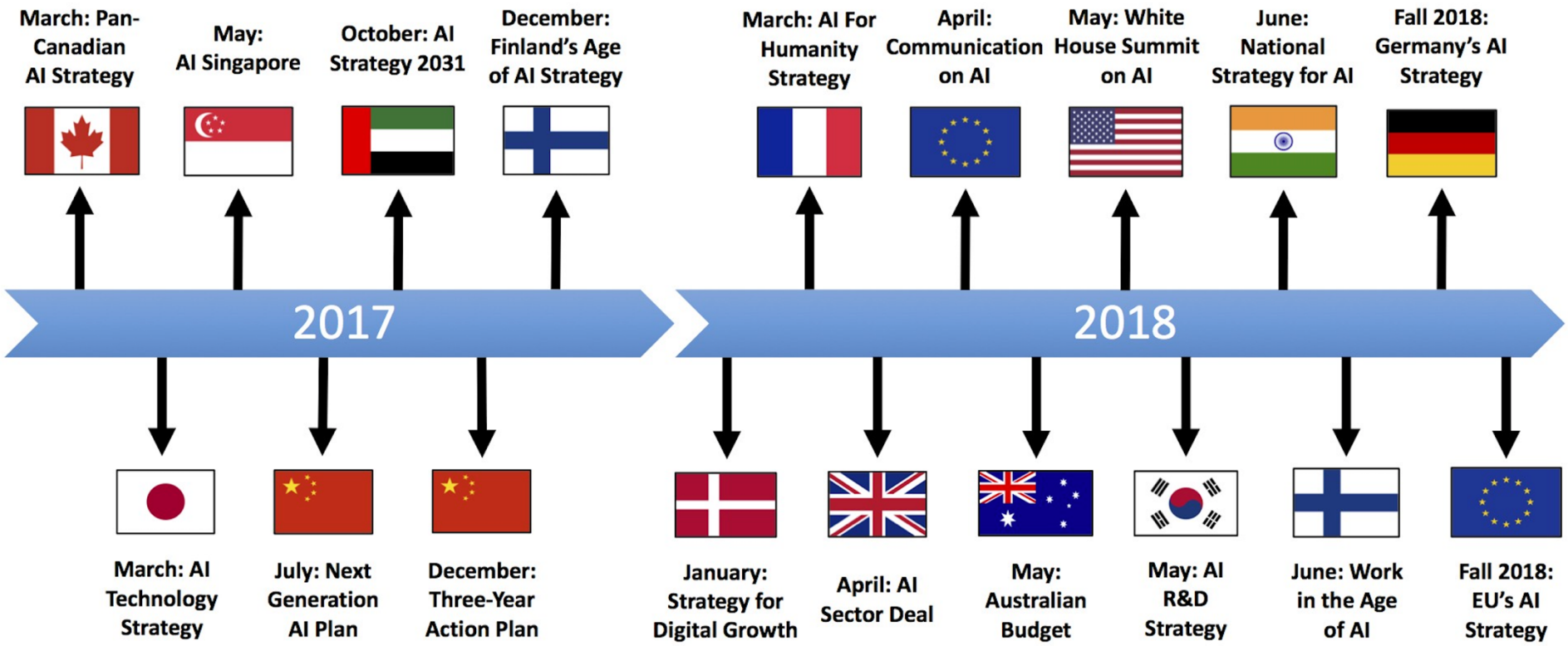
Asiad 2018

A small number of winners take all?

- "**Mối đe doạ là chỉ một số ít quốc gia thắng cuộc sẽ lấy tất**" ("Threat: A small number of winners takes all")*.
- Ai là người thắng cuộc trong thời chuyển đổi số? Vì sao họ thắng? Việt Nam có nằm trong số các nước sẽ thua cuộc? Nếu có thì làm sao thoát ra?
- Nền tảng của thắng cuộc là công nghệ số: có dữ liệu, làm chủ và khai thác dữ liệ.
- Mỗi lĩnh vực khoa học làm gì thời chuyển đổi số?

* Prof. Sang Kyun Cha, Director of Korea Institute of Big Data, PAKDD 2017

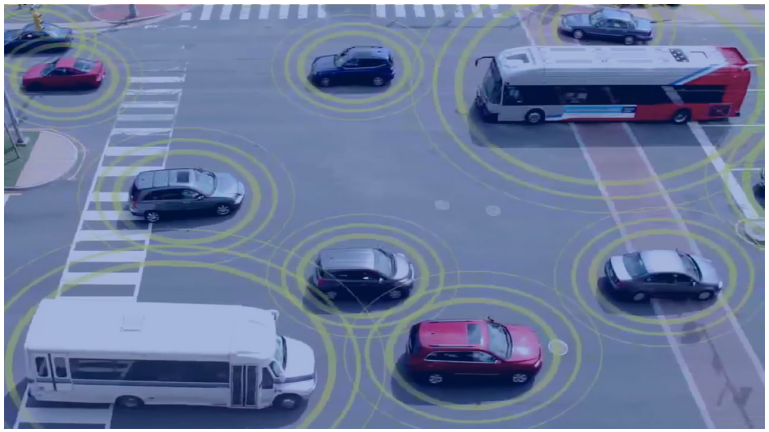
AI strategies



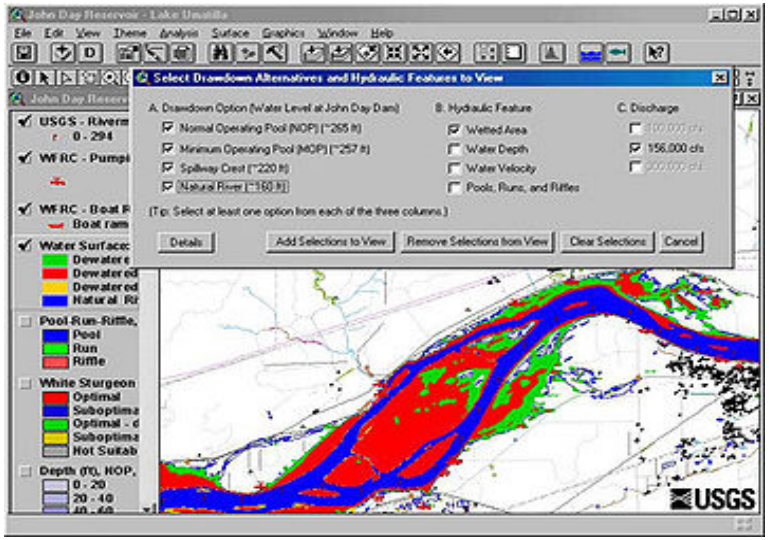
Chiến lược về AI của một số quốc gia (https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd)

AI nào cho Việt Nam?

- Hệ thống tự động ("AI robot")



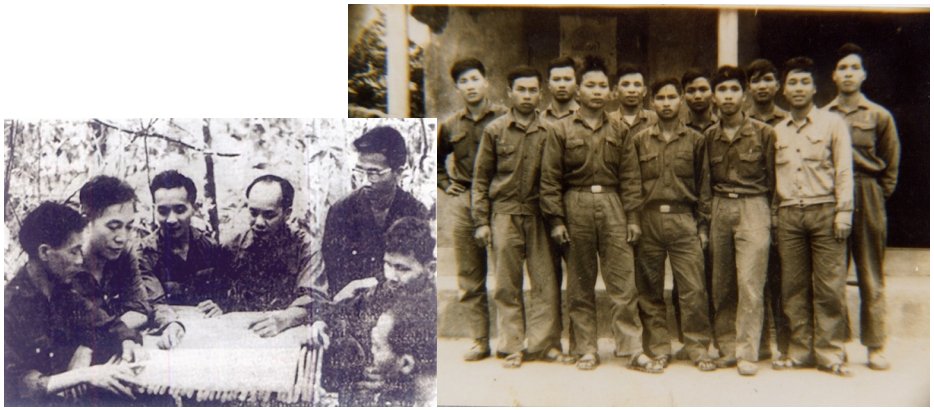
- Hỗ trợ quyết định, đưa ra lời giải ("AI data")



AI nào cho Việt Nam?

- Nông nghiệp thông minh?
- Du lịch thông minh?
- Môi trường an toàn?
- Y tế tiên tiến?
- …

Luôn phải quyết định & hành động



Chỉ huy ra quyết định với tin tức từ lính trinh sát

Đội hình nào?

Chọn nước cờ nào?

Quyết định và kinh doanh

Thảo luận và quyết định

Chuyện nước nhà

Quyết định thường được tạo ra thế nào?

1

Đặt vấn đề

2

Thu thập dữ liệu

3

Có các khả năng nào?

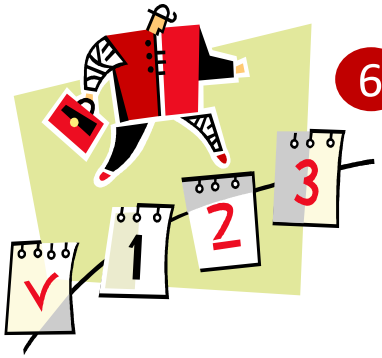
4

Cân nhắc (what-if?)

5

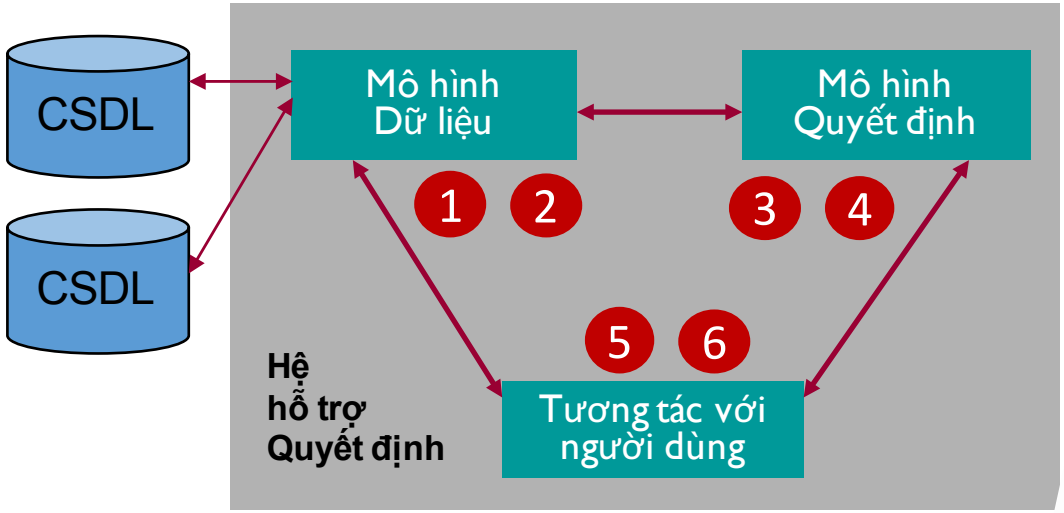
Quyết định

6



Kế hoạch hành động

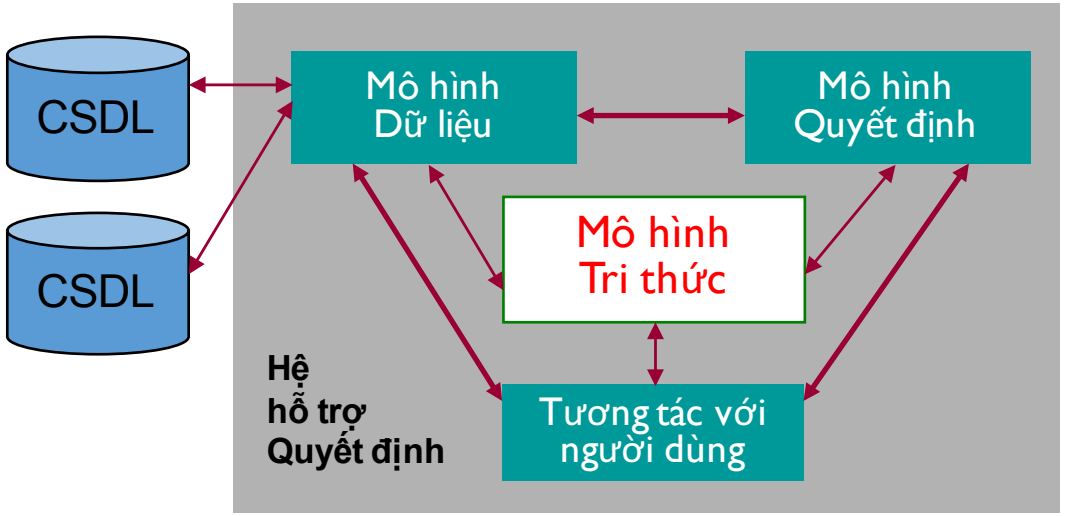
Hệ hỗ trợ quyết định (DSS)



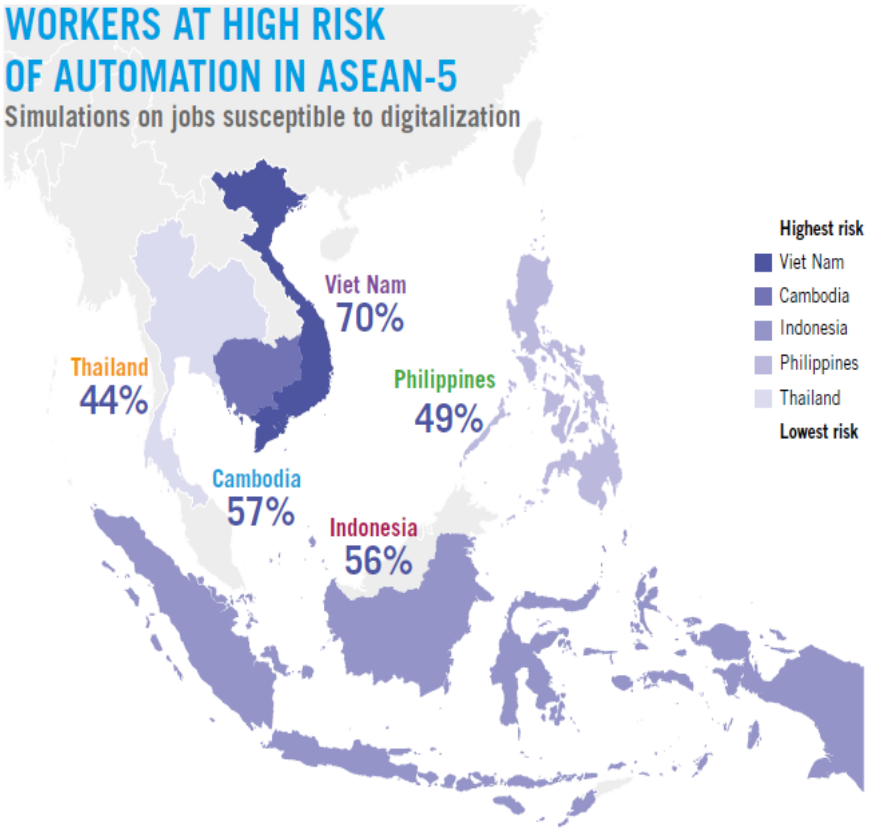
Ba thành phần truyền thống của các hệ hỗ trợ quyết định

Những thập kỷ qua DSS đã gắn với việc dùng tri thức của AI

Thời chuyển đổi số: Thế giới liên kết, các luồng thông tin dày đặc, KH&CN đột phá → có cả thách thức và cơ hội cho quyết định và hành động.



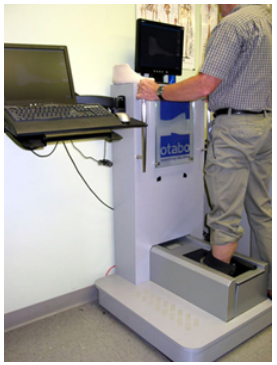
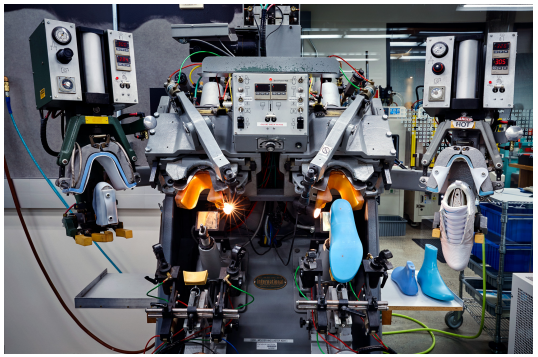
Việc dễ tổn thương thời chuyển đổi số



ILO: ASEAN in transformation: the future of jobs at risk of automation (2016)

May mặc, đóng giày sẽ tiến đến đặc chế

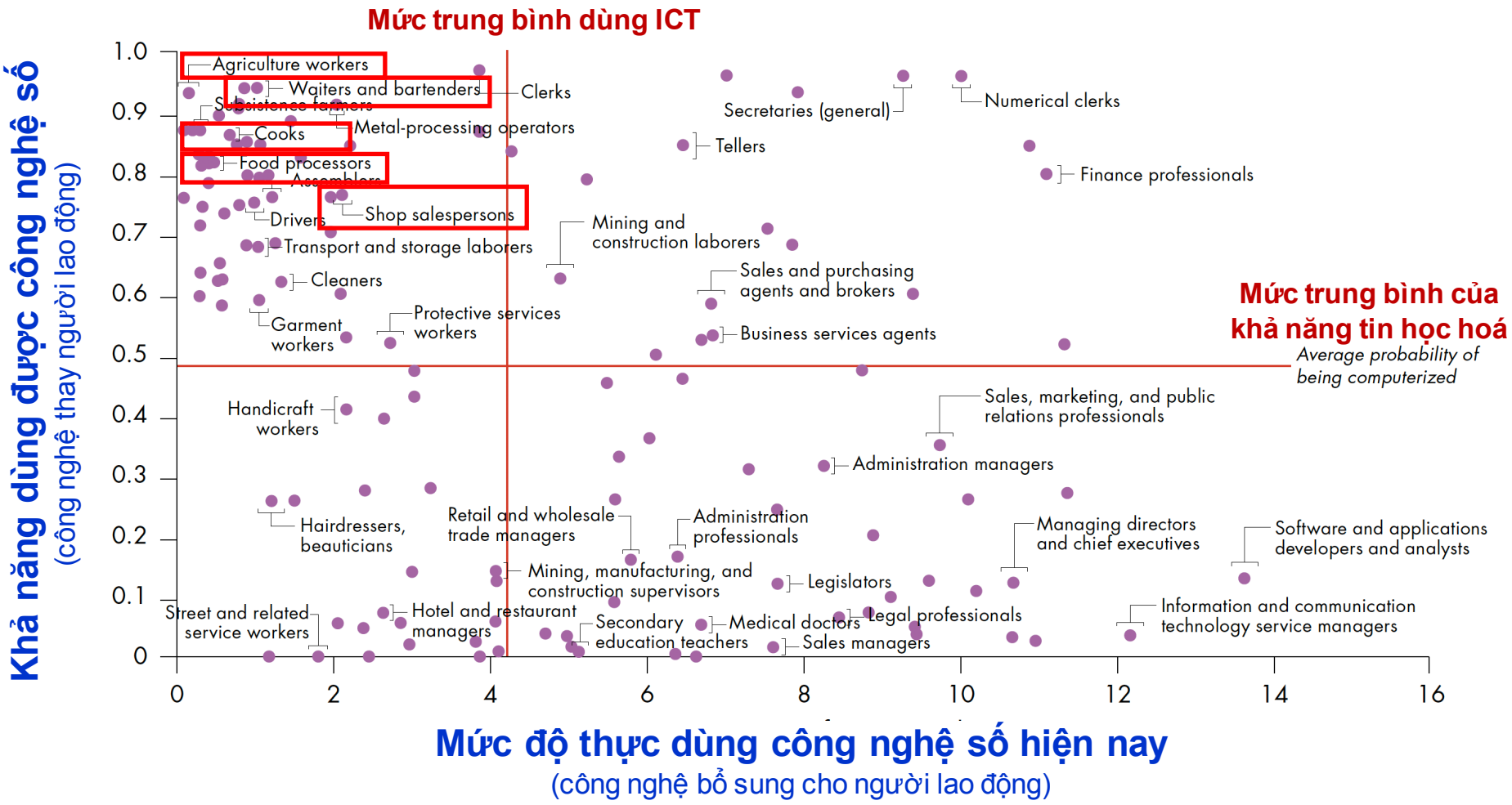




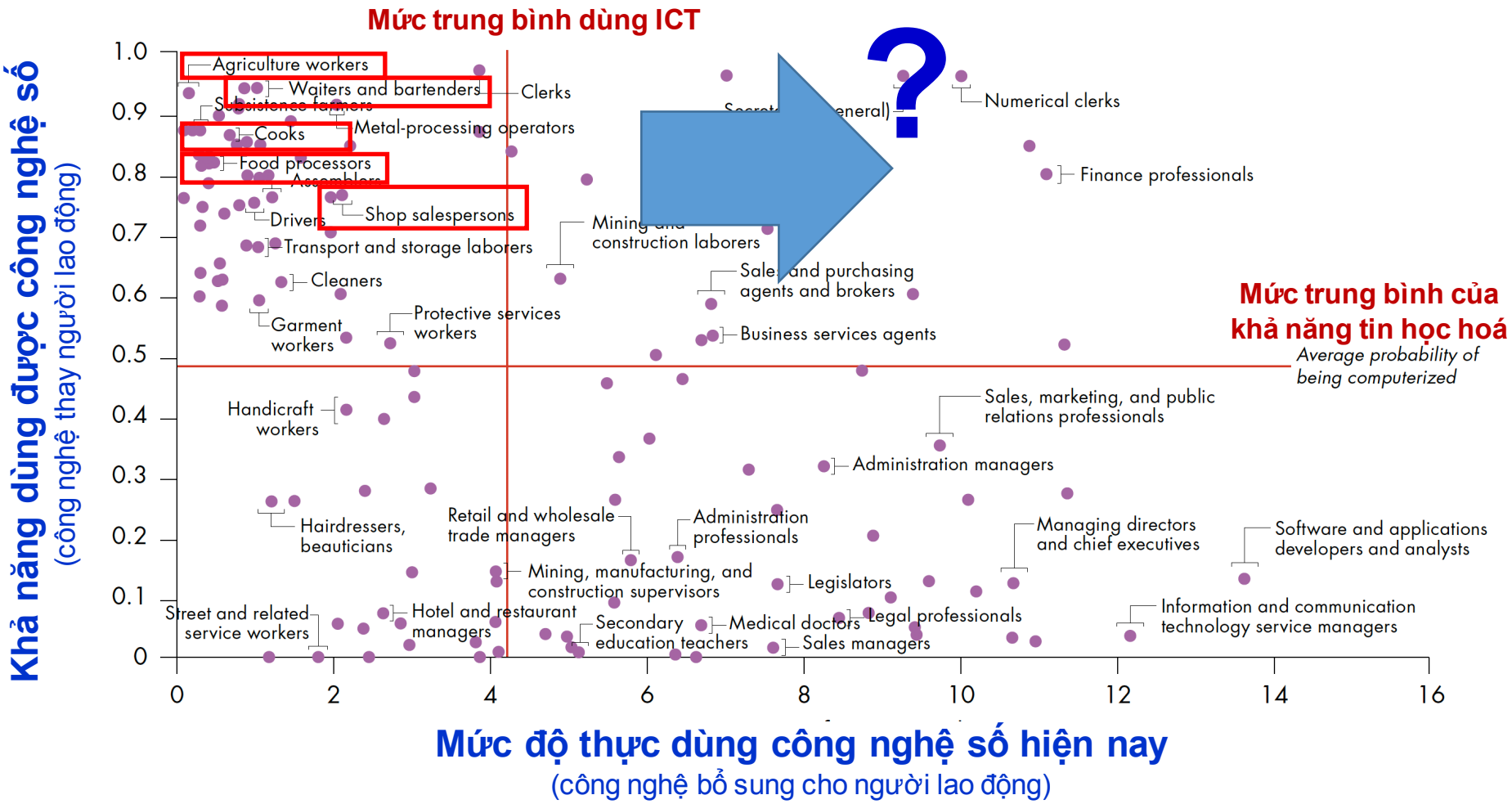
Điều gì đang xảy ra? Sao đặc chế tự động được?

Khả năng và thực tế dùng công nghệ số



World development report 2016: Digital Dividends, World bank group.

Khả năng và thực tế dùng công nghệ số



World development report 2016: Digital Dividends, World bank group.

Một số bài toán của e-health

- Xây dựng các hệ thông tin bệnh viện tiên tiến.
- Từng bước tạo bệnh án điện tử (EMR) ở các bệnh viện.
- CSDL hồ sơ sức khoẻ điện tử (EHR) của mọi công dân.
- Hồ sơ sức khoẻ cá nhân (PHR)
- **Làm sao rút ngắn được thời gian khám bệnh?**

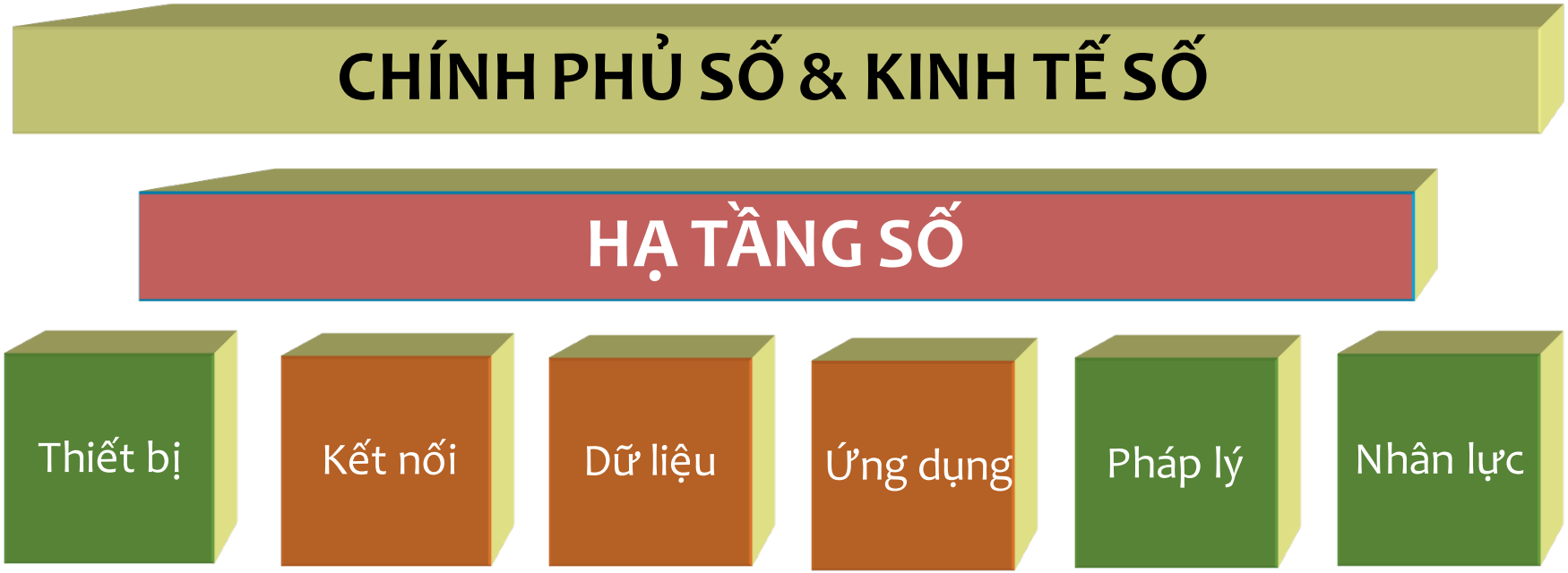
Thành phần trọng yếu của hạ tầng số?

CHÍNH PHỦ SỐ & KINH TẾ SỐ

HẠ TẦNG SỐ

- Xác định và xây dựng các thành phần chính của hạ tầng số
- Lộ trình xây dựng phải hợp lý, không phải phá làm lại

Thành phần trọng yếu của hạ tầng số?



- Xác định và xây dựng các thành phần chính của hạ tầng số.
- Lộ trình xây dựng phải hợp lý, không phải phá làm lại.

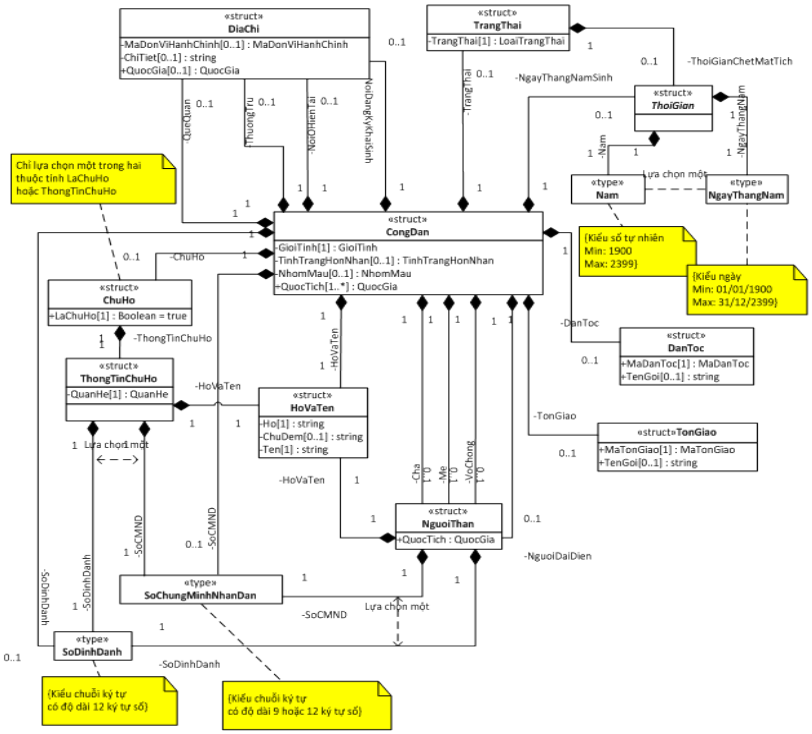
Nhận thức về hạ tầng số của ta

- **Hạ tầng số** chưa được **nhận thức** ở mọi cấp như các hạ tầng khác (đường xá, cầu cống, năng lượng ...)
- **Dữ liệu** quốc gia, địa phương, toàn cầu đều *thiếu*, chất lượng cũng như việc *chia sẻ và kết nối* còn hạn chế.
- Còn **thiếu hành lang pháp lý** cho xã hội số và kinh tế số.
- Thiếu **nhân lực số** ở cả *diện rộng* và ***tinh nhuệ***.
- Hạ tầng số của ta **chưa an toàn**?

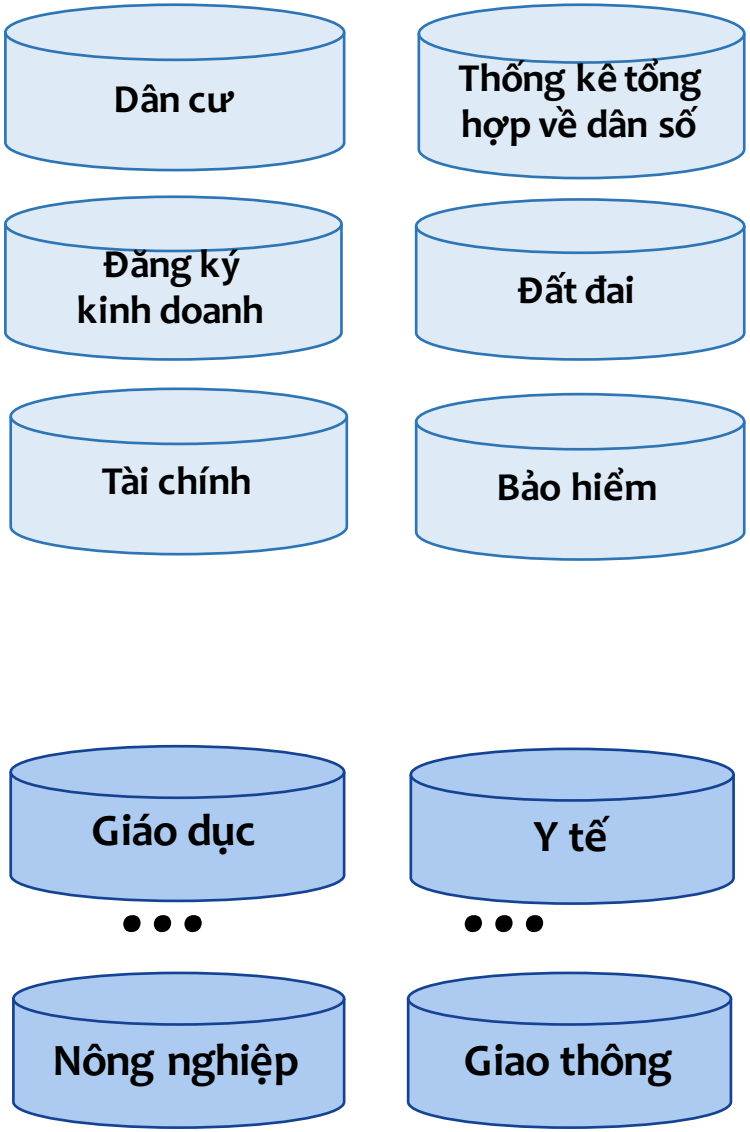
Nhận thức và tập trung xây dựng hạ tầng số, coi đây là **nhiệm vụ hàng đầu** trong thời chuyển đổi số.

CSDL ưu tiên và CSDL các bộ ngành

QĐ 714, 22/5/2015, 6 CSDL quốc gia được ưu tiên triển khai

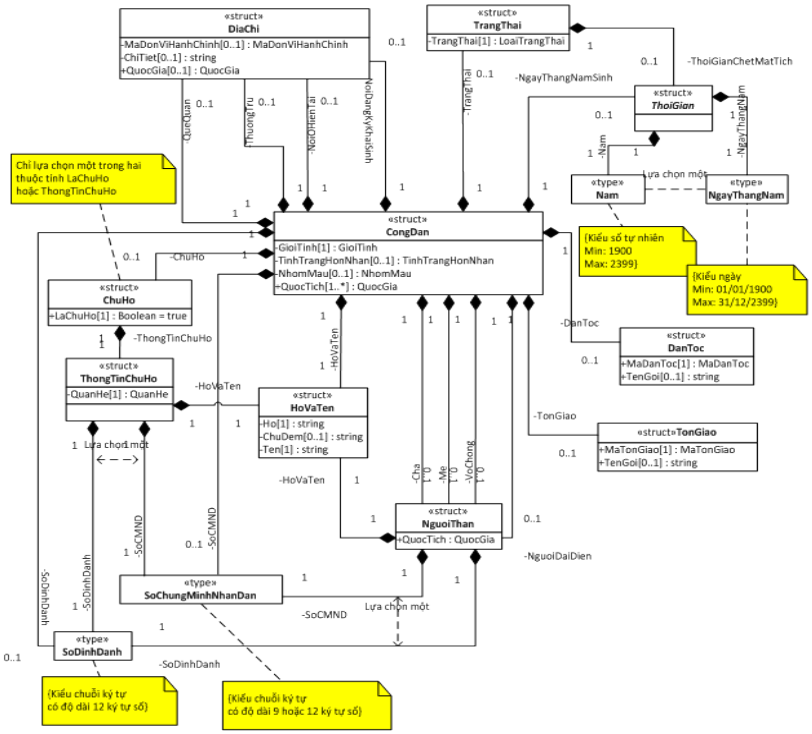


Số định danh cá nhân gồm **12 số** chứa mã **giới tính**, mã **năm sinh**, mã **quê quán**.



CSDL ưu tiên và CSDL các bộ ngành

QĐ 714, 22/5/2015, 6 CSDL quốc gia được ưu tiên triển khai



Số định danh cá nhân gồm **12 số** chứa mã **giới tính**, mã **năm sinh**, mã **quê quán**.

Cơ sở dữ liệu tập trung

Giúp ta **biết thiên hạ** để phát triển trong quan hệ toàn cầu.

Xây dựng các **cơ sở dữ liệu** như **tài sản chính của quốc gia** và **doanh nghiệp,** hạ tầng của chính phủ số và kinh tế số.

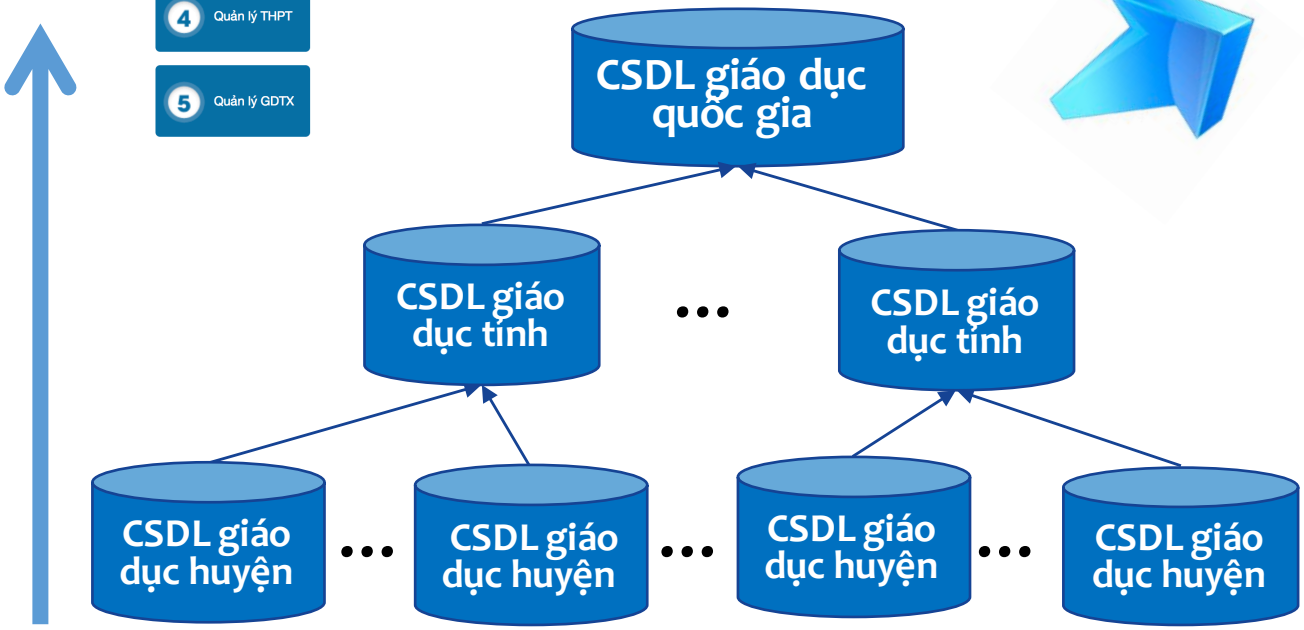
Hạ tầng dữ liệu toàn cầu

- Kết nối với **chuẩn** và **kiến trúc** tốt.
- Xây dựng hạ tầng dữ liệu với các loại dữ liệu (**đóng, chia sẻ, mở**)
- Ba mức: **địa phương, quốc gia, toàn cầu.**

Cơ sở dữ liệu giáo dục quốc gia

www.csdl.moet.gov.vn

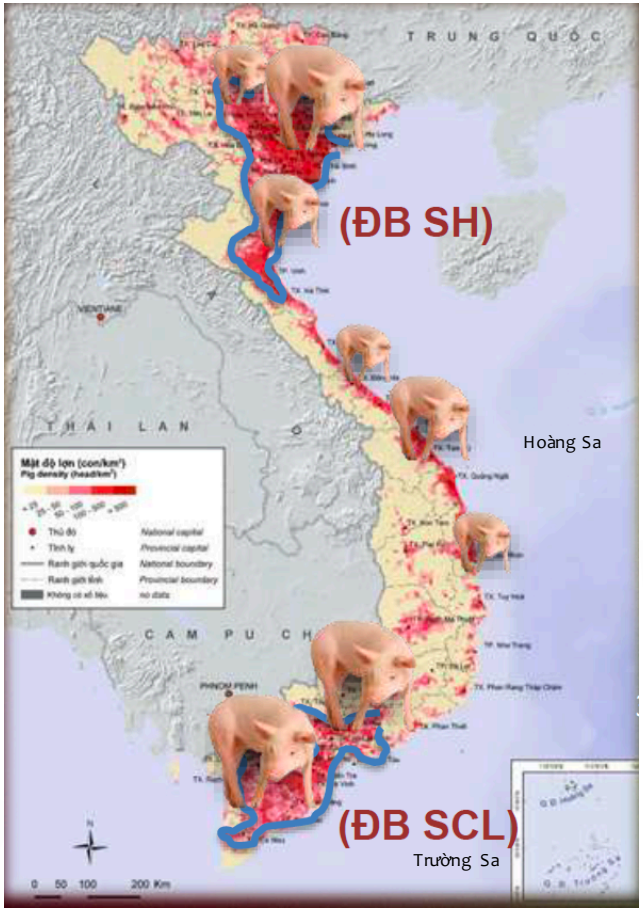
Số định danh cá nhân ngành giáo dục gồm **10 số**, thống nhất và duy nhất.



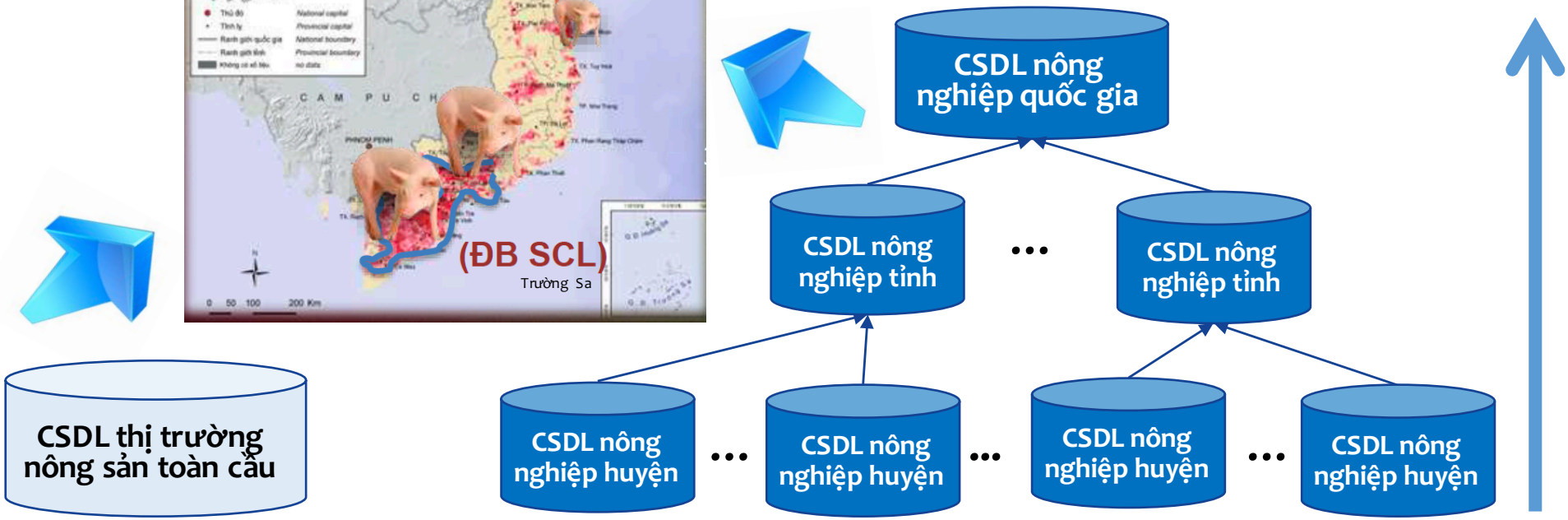
CSDL lõi, gồm

- Trường (44 nghìn)
- Lớp (760 nghìn)
- Giáo viên (>1 triệu)
- Học sinh (23 triệu)

Cơ sở dữ liệu nông nghiệp quốc gia?



Nông nghiệp thông minh không chỉ làm trang trại thông minh, mà phải tính được cung-cầu, chọn sản phẩm "át chủ bài", dự đoán thị trường… tất cả dựa vào **cơ sở dữ liệu nông nghiệp** và **phân tích dữ liệu.**



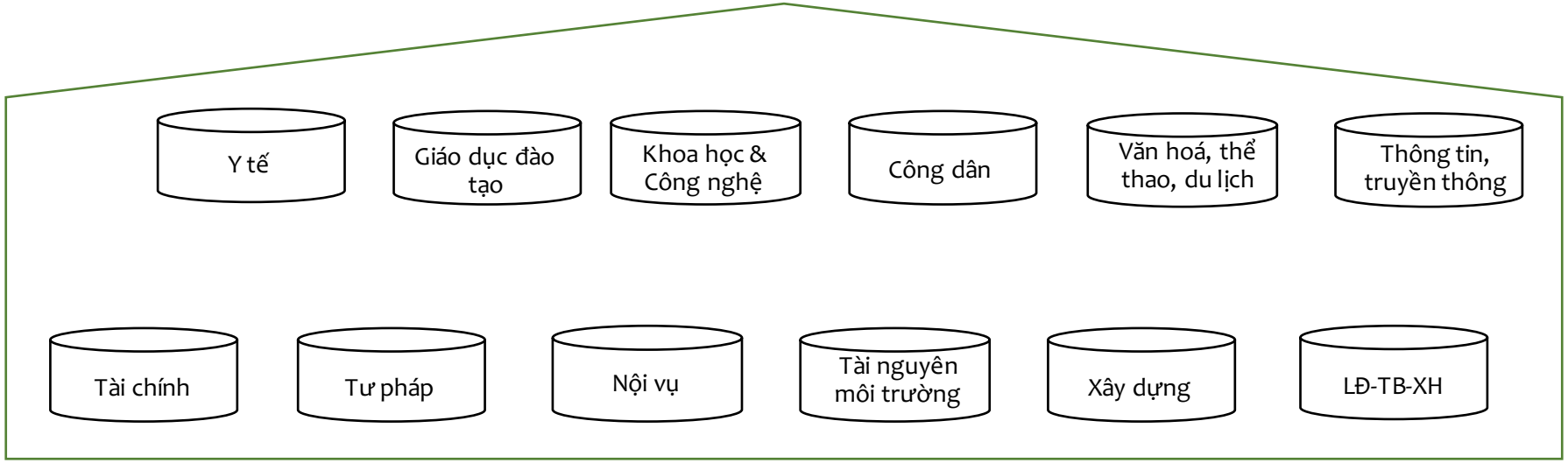
Kho dữ liệu dùng chung: kết nối & chia sẻ

Phải đào tạo ngành nghề gì và đào tạo thế nào để đất nước có đủ nhân lực thời chuyển đổi số?

Làm sao tăng giá trị gia tăng của sản phẩm chứ không chỉ xuất khẩu thô?

Làm sao để 70% dân số sống ở nông thôn, chiếm trên 43% lao động, đóng góp nhiều hơn 18% GDP như hiện nay?

Làm sao giảm nợ xấu và sở hữu chéo để hệ thống tài chính quốc gia tránh rủi ro?



KHO DỮ LIỆU DÙNG CHUNG (KẾT NỐI VÀ CHIA SẺ)

Kho dữ liệu dùng chung: kết nối & chia sẻ

Phải đào tạo ngành nghề gì và đào tạo thế nào để đất nước có đủ nhân lực thời chuyển đổi số?

Làm sao tăng giá trị gia tăng của sản phẩm chứ không chỉ xuất khẩu thô?

3 loại dữ liệu cần làm:

1. **CSDL chuyên ngành** (nâng chất lượng các CSDL đang có)
2. **CSDL dùng chung** (kết nối và **chia sẻ**)
3. **CSDL mở & liên kết của chính quyền** (cho dân)

KHO DỮ LIỆU DÙNG CHUNG (KẾT NỐI VÀ CHIA SẺ)

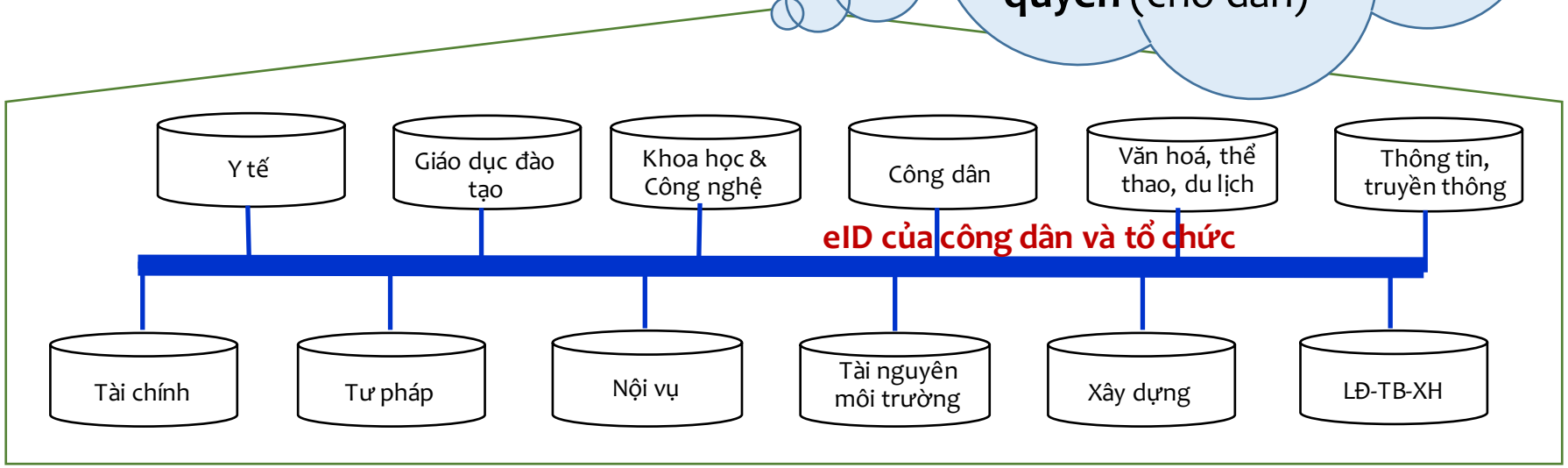
Kho dữ liệu dùng chung: kết nối & chia sẻ

Phải đào tạo ngành nghề gì và đào tạo thế nào để đất nước có đủ nhân lực thời chuyển đổi số?

Làm sao tăng giá trị gia tăng của sản phẩm chứ không chỉ xuất khẩu thô?

3 loại dữ liệu cần làm:

1. **CSDL chuyên ngành** (nâng chất lượng các CSDL đang có)
2. **CSDL dùng chung** (kết nối và **chia sẻ**)
3. **CSDL mở & liên kết của chính quyền** (cho dân)

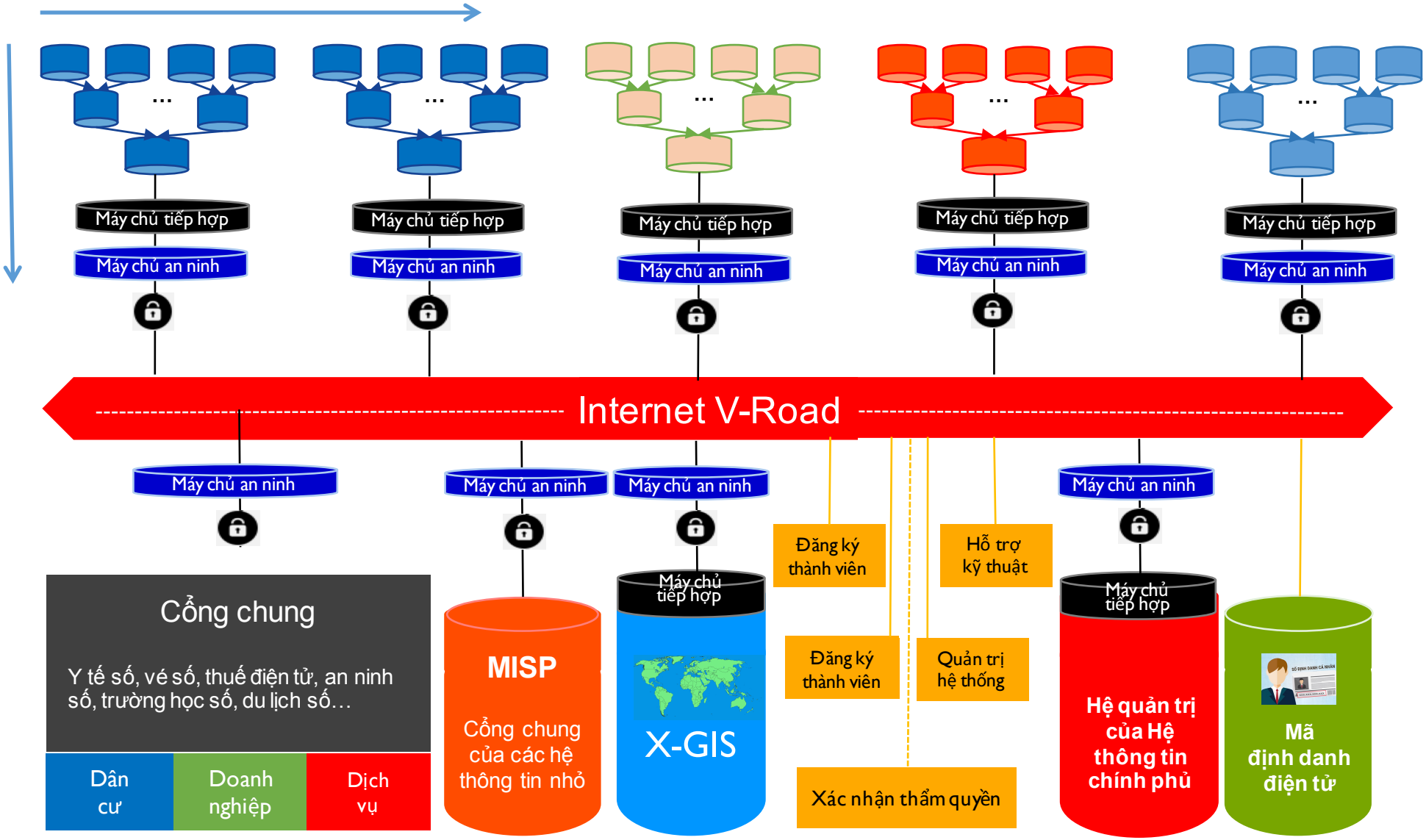


KHO DỮ LIỆU DÙNG CHUNG (KẾT NỐI VÀ CHIA SẺ)

Dữ liệu dùng chung: kết nối & chia sẻ

- Rủi ro lớn đối với Nhà nước: tập trung dữ liệu không kiểm soát được vào một cơ sở dữ liệu duy nhất → nên tránh CSDL tập trung.
- **Quy tắc Anderson**: Hệ thống lớn được thiết kế để dễ truy cập thì không an toàn; nếu khó truy nhập lại không dùng được hiệu quả
  - Các **CSDL không nên quá lớn, mỗi tổ chức có và quản trị CSDL riêng.**
  - Các **CSDL riêng được nối và chia sẻ với CSDL** của các thành viên khác theo **cam kết** trong một môi trường công nghệ (thí dụ mô hình X-Road của Estonia).

Đề xuất mô hình kết nối và chia sẻ dữ liệu



Nhân lực số và kỹ năng lao động mới

năm 2020	năm 2015
1. Giải quyết vấn đề phức tạp	1. Giải quyết vấn đề phức tạp
2. Tư duy phê phán	2. Hợp tác với người khác
3. Tính sáng tạo	3. Quản lý nhân sự
4. Quản lý nhân sự	4. Tư duy phê phán
5. Hợp tác với người khác	5. Đàm phán
6. Trí tuệ cảm xúc	6. Quản lý chất lượng
7. Phán quyết và ra quyết định	7. Định hướng dịch vụ
8. Định hướng dịch vụ	8. Phán quyết và ra quyết định
9. Đàm phán	9. Lắng nghe tích cực
10. Linh hoạt trong nhận thức	10. Tính sáng tạo

Nhân lực số phụ thuộc vào dòng dịch chuyển người lao động, kết quả đào tạo, tính linh hoạt và thích nghi. Giáo dục phải thay đổi từ cấp phổ thông (đã có kế hoạch) đến học nghề và đại học (chậm) → **Kỹ năng lao động mới.**

Thông điệp cùng chia sẻ

- AI là chìa khoá của thời kỳ chuyển đổi số.
- AI cần cho ta là AI gắn với các mục tiêu phát triển của đất nước.
- Học máy có vai trò đột phá trong AI.
- Trước khi đến một chiến lược AI của đất nước, nhìn rõ những AI nào ta đang cần và ưu tiên là một điều cần thiết.

Cám ơn!