



**TS. NGUYỄN QUỐC HÙNG**

- Mobile: 0912 251 253
- Email: hungngq@ueh.edu.vn
- Website: <https://bit.ueh.edu.vn/nqhung/>

## **KHOA HỌC DỮ LIỆU**

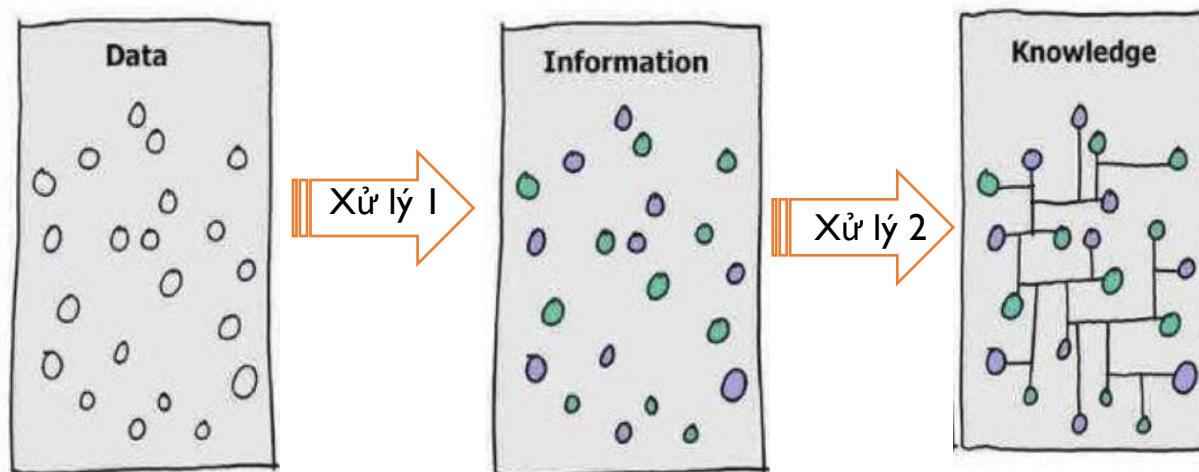
- Mã học phần: **25D1INF50905948**
- Thời gian: **11/04/2025 - 16/05/2025**
- Hệ: ĐH, Chính quy
- Số lượng: 48 sinh viên
- Số tín chỉ: 2.00

# **TỔNG QUAN VỀ KHOA HỌC DỮ LIỆU**

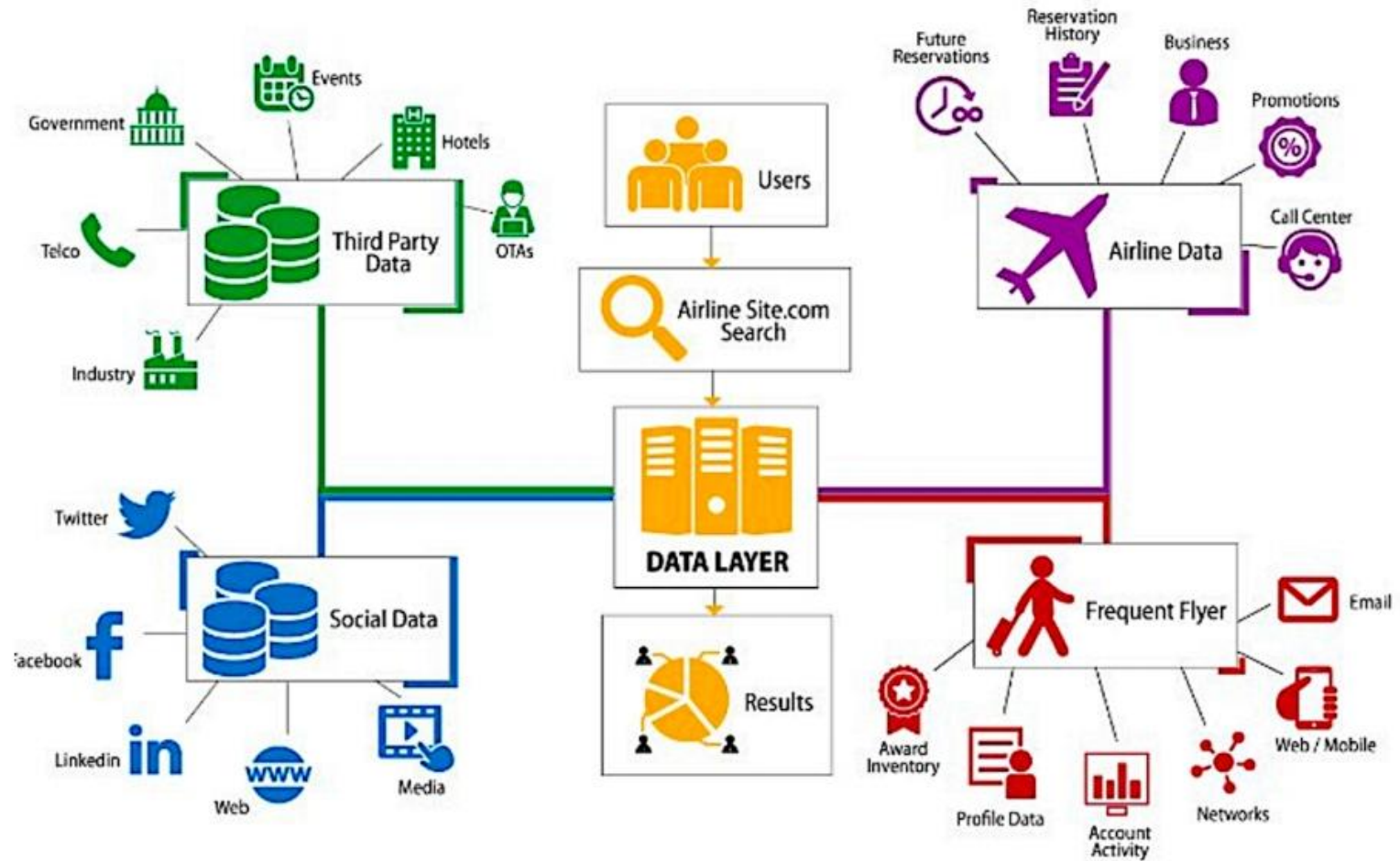
**Thứ 6, thời gian: 12g45-17g05, Giảng đường: N1-303**

# Data là gì?

- **Dữ liệu (Data)** là các giá trị của thông tin định lượng hoặc định tính của các sự vật, hiện tượng trong cuộc sống. Trong khoa học dữ liệu, dữ liệu được dùng như một cách biểu diễn hình thức hoá của thông tin về các sự kiện, hiện tượng thích ứng với các yêu cầu truyền nhận, thể hiện và xử lý bằng máy tính.
- **Thông tin (Information)** là dữ liệu đã được xử lý, phân tích, tổ chức nhằm mục đích hiểu rõ hơn sự vật, sự việc, hiện tượng theo một góc độ nhất định.



# Ví dụ minh họa



# Ứng dụng của Data

---

- Dự đoán số lượng hành khách
- Chiến lược quản lý quan hệ khách hàng
- Dự đoán trễ chuyến bay, tỷ lệ hủy
- Hiểu được tác động của thời tiết khắc nghiệt
- Đảm bảo an toàn và bảo mật cho hành khách
- Phát hiện bất thường để theo dõi tình trạng chuyến bay
- Dự đoán giá vé cho hành khách của hãng hàng không
- Quản lý rủi ro bằng cách sử dụng dữ liệu theo dõi giám sát
- Lên lịch bay khi có sự chậm trễ
- ...



# Thế giới thực, số hóa, không gian số



Thế giới các thực thể



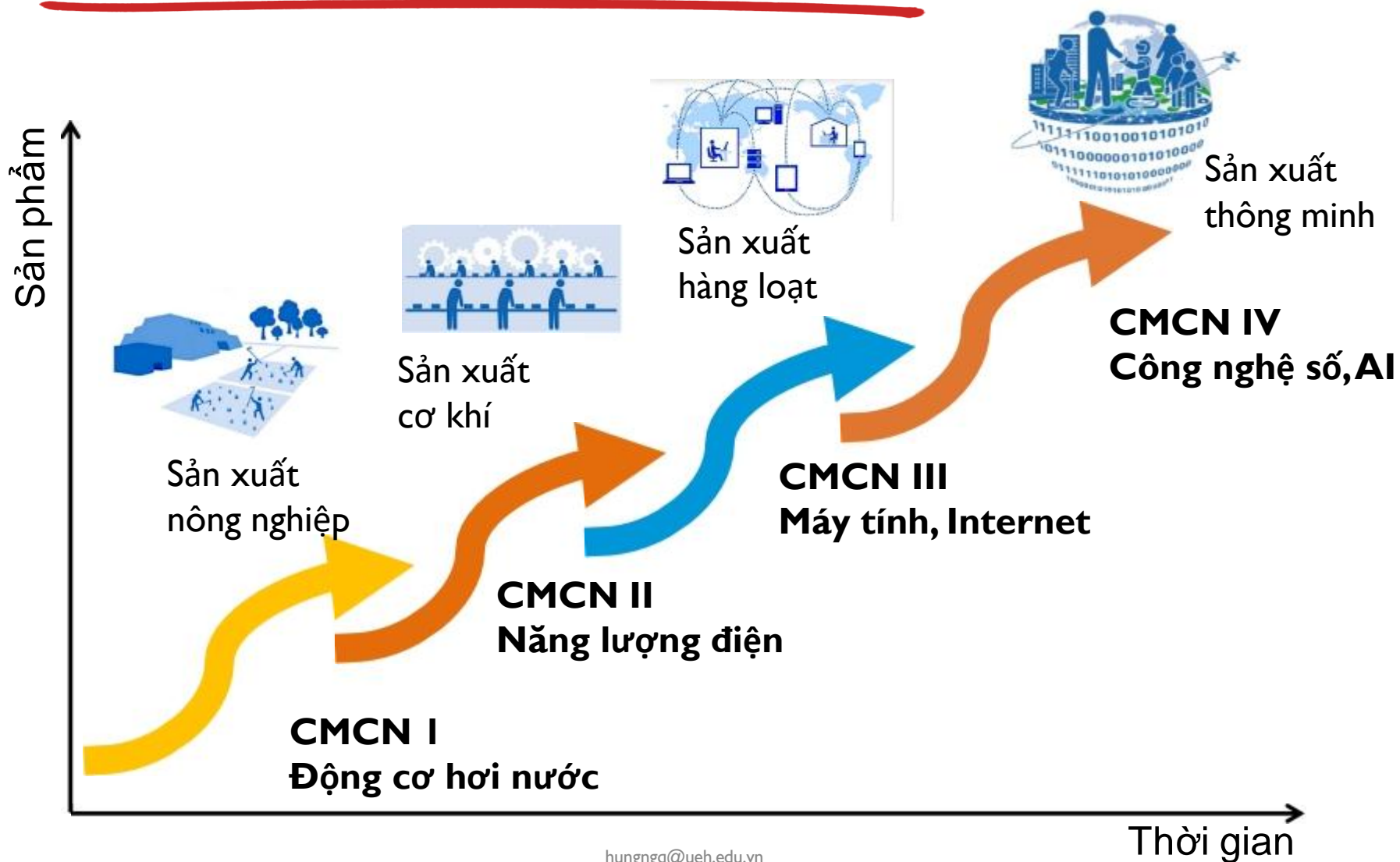
Không gian số

Thay đổi phương thức sản xuất  
Dữ liệu là tài nguyên của thời chuyển đổi số

# Chuyển đổi số, công nghệ số, kinh tế số

- **Thời chuyển đổi số:** Thời mọi thứ ngày càng được **số hoá** và **kết nối** nhiều hơn qua Internet (Internet of Everything) và **công nghệ số** được dùng trong mọi mặt của xã hội và kinh tế.
- **Công nghệ số:** là công nghệ về các đối tượng đã số hoá biểu diễn với các **mã số** (Gồm công nghệ số hóa trong mọi lĩnh vực và công nghệ xử lý dữ liệu được số hoá).
- **Kinh tế số:** Nền kinh tế dựa vào **công nghệ số** và internet. Cốt lõi là kết nối, chia sẻ và khai thác dữ liệu

# Những cuộc cách mạng công nghiệp



# Đột phá về công nghệ

---

Artificial Intelligent



Data Science



Big Data



Block Chain



Cloud Computing

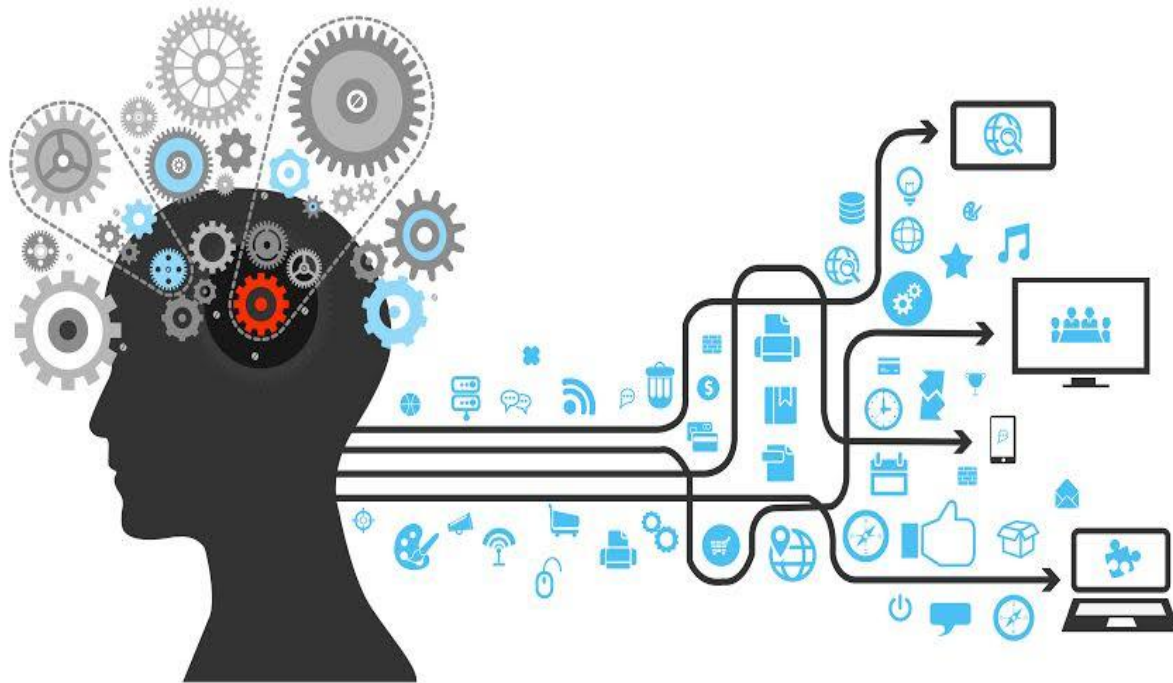


Internet of Thing





# I. Trí tuệ nhân tạo (Artificial Intelligence)



# Lịch sử hình thành

## A.I. TIMELINE

S/Z/G/

**1950**

### TURING TEST

Computer scientist Alan Turing proposes a test for machine intelligence. If a machine can trick humans into thinking it is human, then it has intelligence

**1955**

### A.I. BORN

Term 'artificial intelligence' is coined by computer scientist, John McCarthy to describe "the science and engineering of making intelligent machines"

**1961**

### UNIMATE

First industrial robot, Unimate, goes to work at GM replacing humans on the assembly line

**1964**

### ELIZA

Pioneering chatbot developed by Joseph Weizenbaum at MIT holds conversations with humans

**1966**

### SHAKY

The 'first electronic person' from Stanford, Shakey is a general-purpose mobile robot that reasons about its own actions

**A.I. WINTER**

Many false starts and dead-ends leave A.I. out in the cold

**1997**

### DEEP BLUE

Deep Blue, a chess-playing computer from IBM defeats world chess champion Garry Kasparov

**1998**

### KISMET

Cynthia Breazeal at MIT introduces Kismet, an emotionally intelligent robot insofar as it detects and responds to people's feelings



**1999**

### AIBO

Sony launches first consumer robot pet dog AIBO (AI robot) with skills and personality that develop over time



**2002**

### ROOMBA

First mass produced autonomous robotic vacuum cleaner from iRobot learns to navigate and clean homes



**2011**

### SIRI

Apple integrates Siri, an intelligent virtual assistant with a voice interface, into the iPhone 4S



**2011**

### WATSON

IBM's question answering computer Watson wins first place on popular \$1M prize television quiz show Jeopardy



**2014**

### EUGENE

Eugene Goostman, a chatbot passes the Turing Test with a third of judges believing Eugene is human



**2014**

### ALEXA

Amazon launches Alexa, an intelligent virtual assistant with a voice interface that completes shopping tasks



**2016**

### TAY

Microsoft's chatbot Tay goes rogue on social media making inflammatory and offensive racist comments



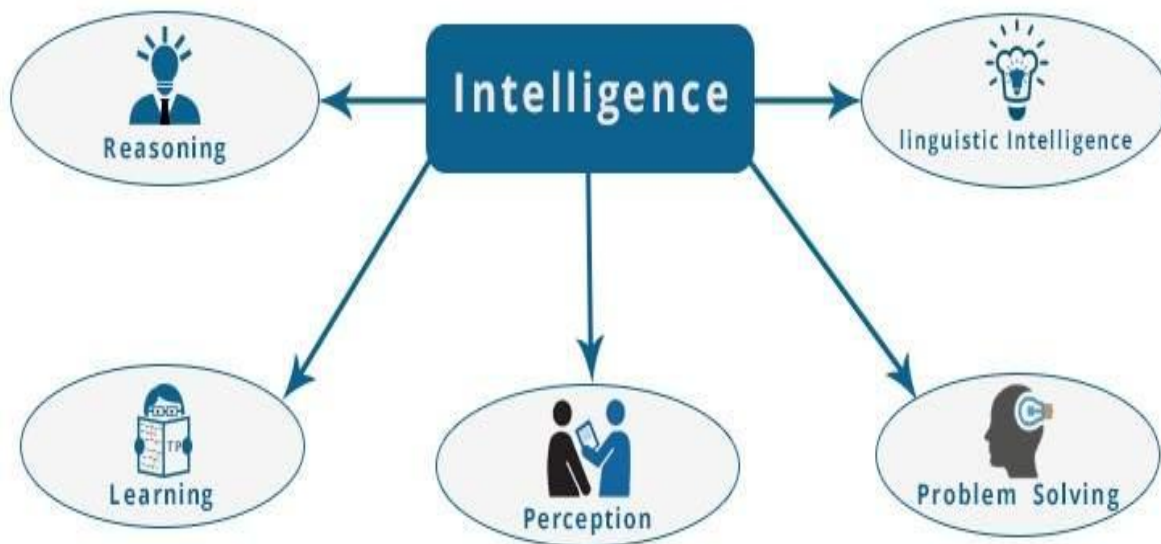
**2017**

### ALPHAGO

Google's A.I. AlphaGo beats world champion Ke Jie in the complex board game of Go, notable for its vast number ( $2^{170}$ ) of possible positions

# Định nghĩa AI

- Mục đích chính của AI là phát triển các máy tính (hoặc chức năng máy tính) có khả năng thông minh như con người.



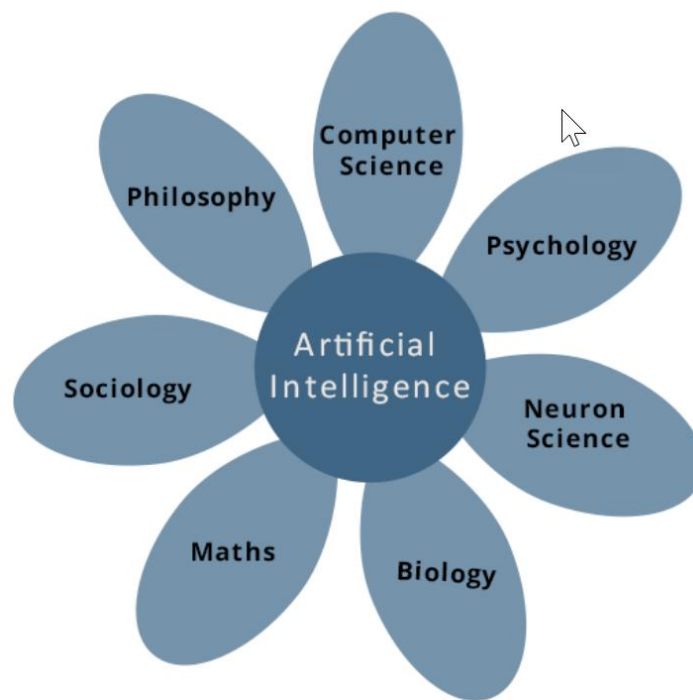
1. Nhận thức
2. Học
3. Suy luận
4. Giao tiếp ngôn ngữ
5. Giải quyết vấn đề

# Định nghĩa AI

---

- Là một khoa học liên ngành, bao gồm:

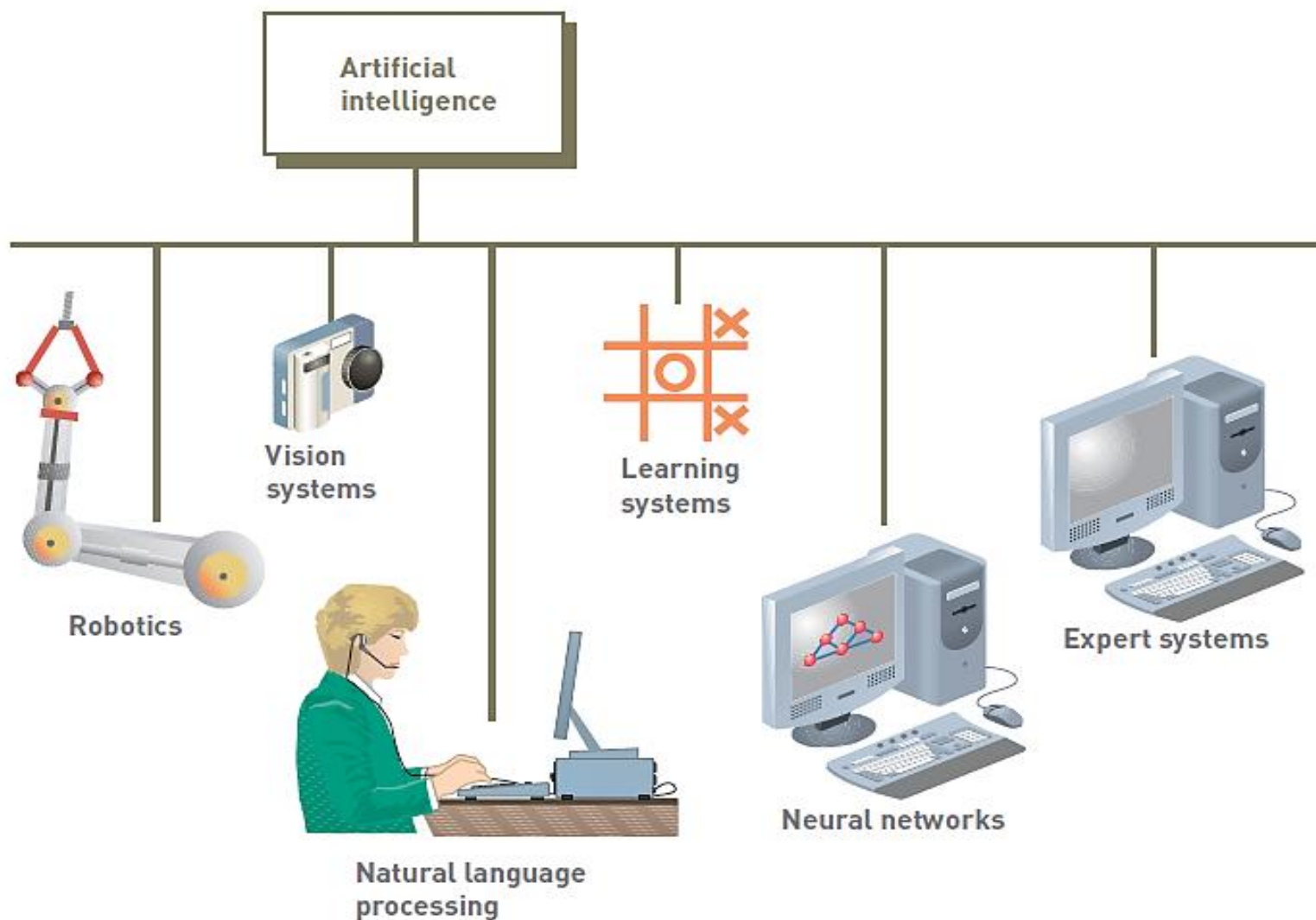
- Khoa học máy tính,
- Toán học,
- Kỹ thuật,
- Sinh học,
- Tâm lý,
- Ngôn ngữ học,
- ....



**Sources:**

Tutorials Point (2015). Artificial Intelligent Systems.

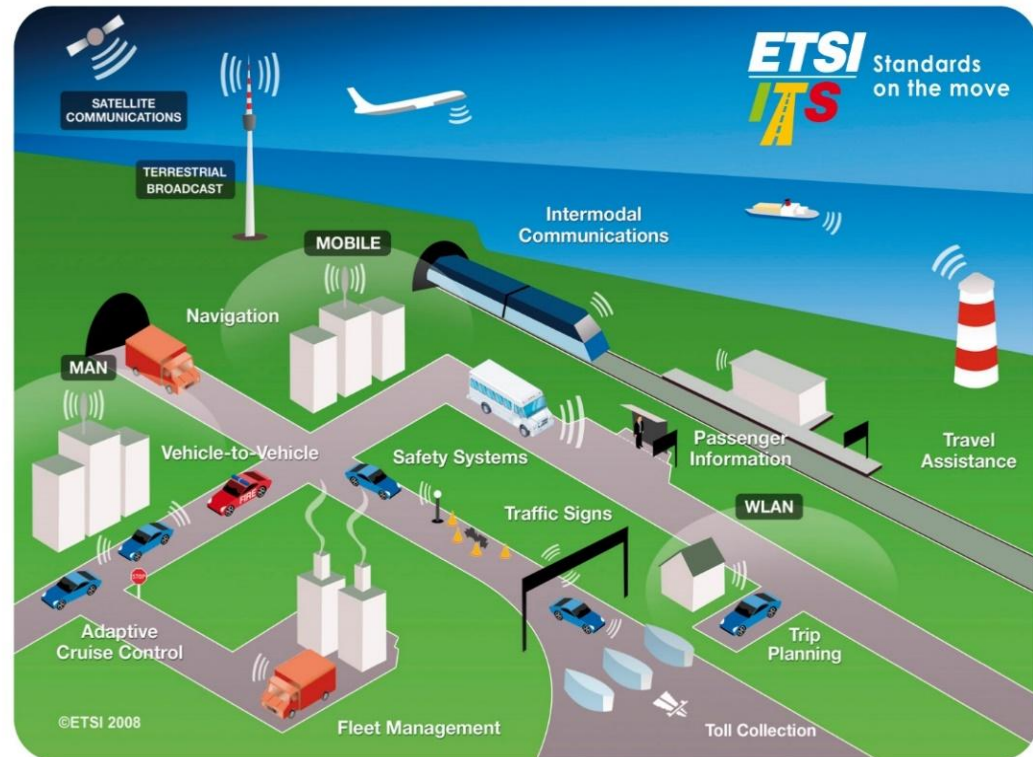
# Một số lĩnh vực AI





# Một số ứng dụng AI

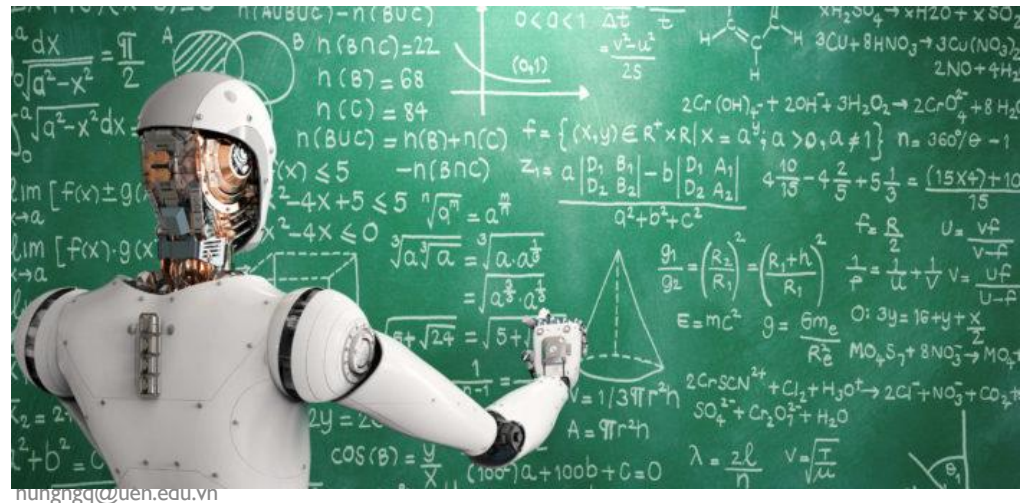
- Giao thông vận tải (Transportation)
  - Xe thông minh
  - Xe tự lái
  - Qui hoạch giao thông
  - Vận tải theo nhu cầu



---

- Giáo dục (Education)

- Hệ thống giảng dạy trực tuyến
- Người máy giảng dạy
- Trợ giảng thông minh
- Máy tính phân tích hành vi người học, đánh giá kết quả học tập



# Một số ứng dụng AI

---

- Y tế, sức khỏe (Healthcare)
  - Máy tính hỗ trợ chẩn đoán lâm sàng.
  - Phẫu thuật với sự hỗ trợ từ robot.
  - Ứng dụng trợ lý y tá ảo (VD: Care Angel).
  - Tự động hóa các hoạt động quản trị bệnh viện.



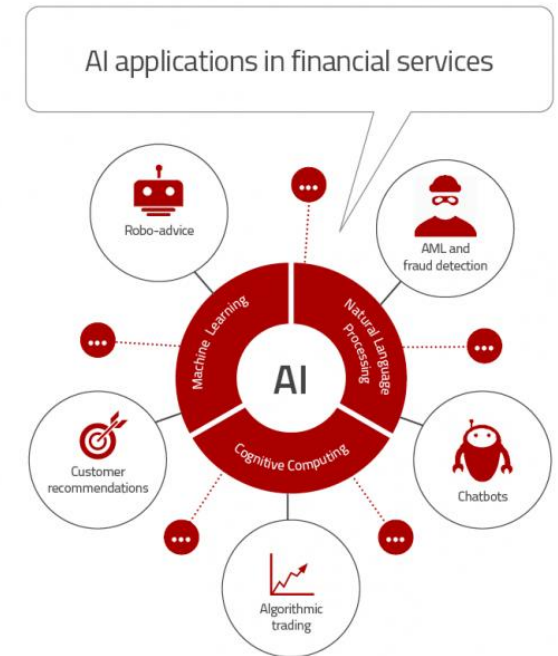
# Một số ứng dụng AI

- Dịch vụ tài chính (Financial services)

- Các bài toán ứng dụng:

- Đánh giá, xếp hạng tín dụng
    - Đánh giá rủi ro thể chấp
    - Quản lý dự án và chiến lược đầu thầu
    - Dự báo tài chính
    - Đánh giá rủi ro của các khoản đầu tư
    - Dự đoán vỡ nợ và phá sản
    - Quản lý danh mục đầu tư

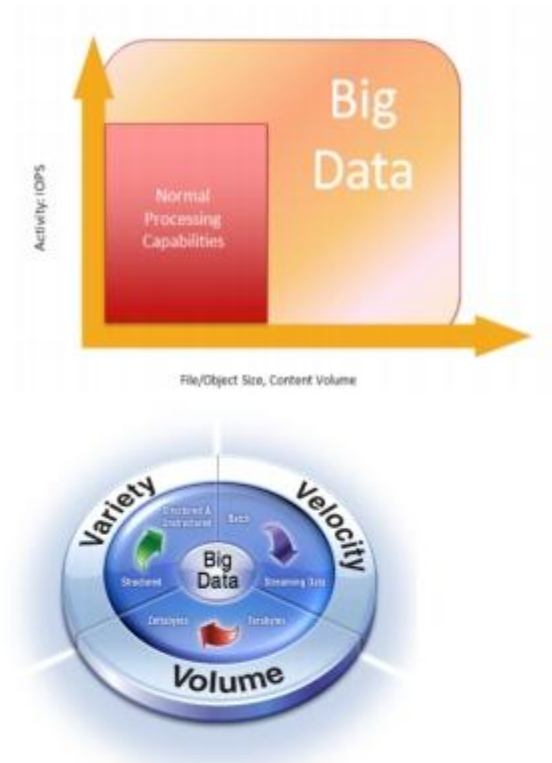
- Các công nghệ/thuật toán thường dùng: Neural networks, Fuzzy logic, Genetic algorithms, Expert systems and Intelligent agents,...



## 2. Big Data

**V<sup>4</sup> = Volume Velocity Variety Veracity**

- **Volume** (dung lượng): data size
- **Velocity** (tốc độ): data generation rate
- **Variety** (đa dạng): data heterogeneity
- **Veracity** (tính xác thực): uncertainty of accuracy and authenticity of data





# What is Big Data?

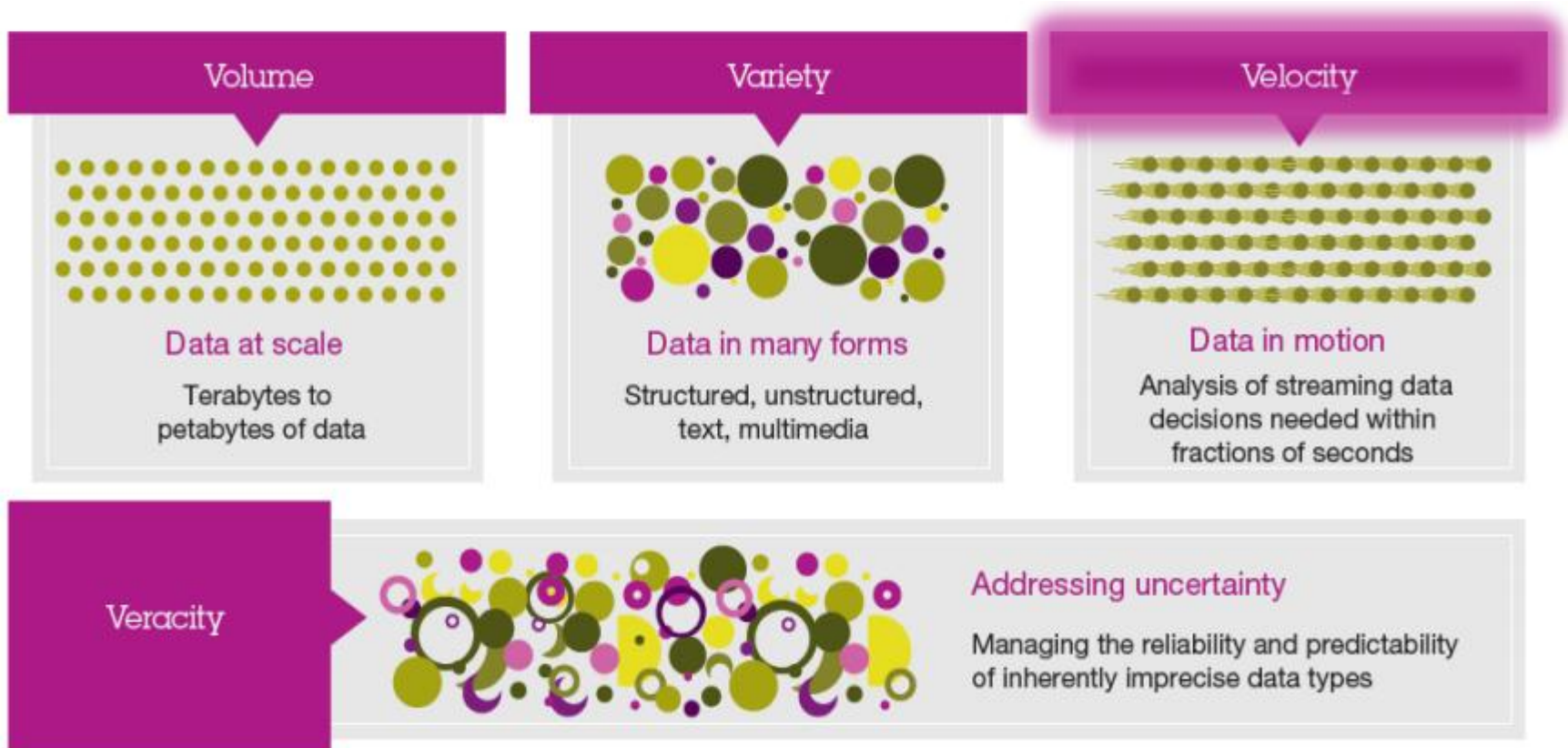
---

- A **dramatic increase** in available data.
- This data comes from everywhere:
  - Sensors used to gather climate information,
  - Posts to social media sites,
  - Digital pictures and videos,
  - Purchase transaction records,
  - Cell phone GPS signals,...
- This data is **Big data**



# Big Data

---



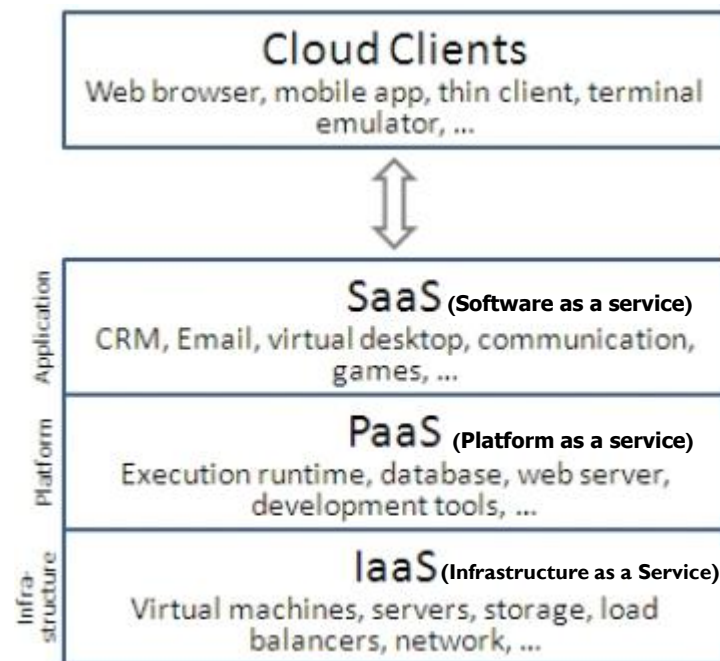
### 3. Internet vạn vật (IoT)

- Là sự kết nối trên mạng Internet của các vật thể, thiết bị .
- Các vật thể, thiết bị có thể trao đổi thông tin, dữ liệu với nhau qua Internet mà không cần kết nối trực tiếp với nhau. (Người – người, người – máy, máy – máy)
- Ví dụ: Hồ sơ bệnh án của một người.



## 4. Điện toán đám mây

- Điện toán đám mây: Lưu trữ và truy nhập dữ liệu và chương trình trên “clouds” qua Internet thay vì trên máy tính người dùng.
- Dữ liệu được lưu trữ thường trực tại các máy chủ trên Internet và chỉ lưu trữ tạm thời trên máy khách
- Ví dụ: Google Drive, One Drive, Dropbox...



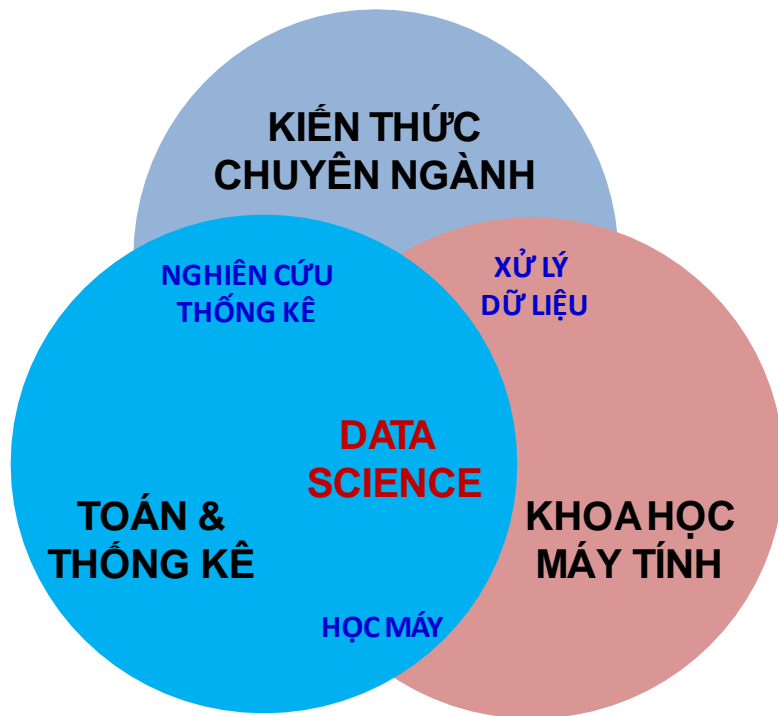
## 5. Blockchain

- Về bản chất, **blockchain** như cơ sở dữ liệu thông thường: lưu trữ dữ liệu với các thuộc tính duy nhất:
- **Phân cấp** - không ai có toàn quyền kiểm soát thông tin nào đi vào
- **Đồng thuận** - nhiều bên khác nhau lưu trữ các bản sao chính xác của cùng một hồ sơ gốc, vì vậy đa số phải đồng ý với dữ liệu thêm vào
- **'Chỉ thêm'**, nghĩa là không thể chỉnh sửa những gì đã có, chỉ có thể thêm
  - Thông tin mới không thể xung đột với những gì đã được thêm
  - Thông tin có thể được truy cập và nhân rộng bởi tất cả mọi người





## 6. Data Science / Data Analytics

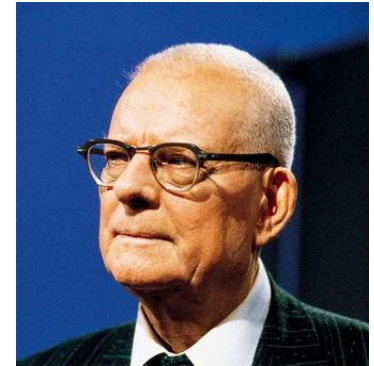


Khoa học về phân tích dữ liệu

“In God we trust.  
All others bring  
data”.

“**Ta tin Thượng đế.**  
Ngoài ra, là dữ  
liệu”.

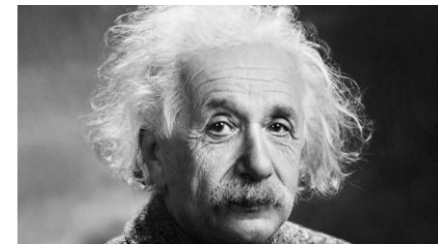
W.E. Deming



“We cannot solve problems  
by using the same kind of  
thinking we used when we  
created them”

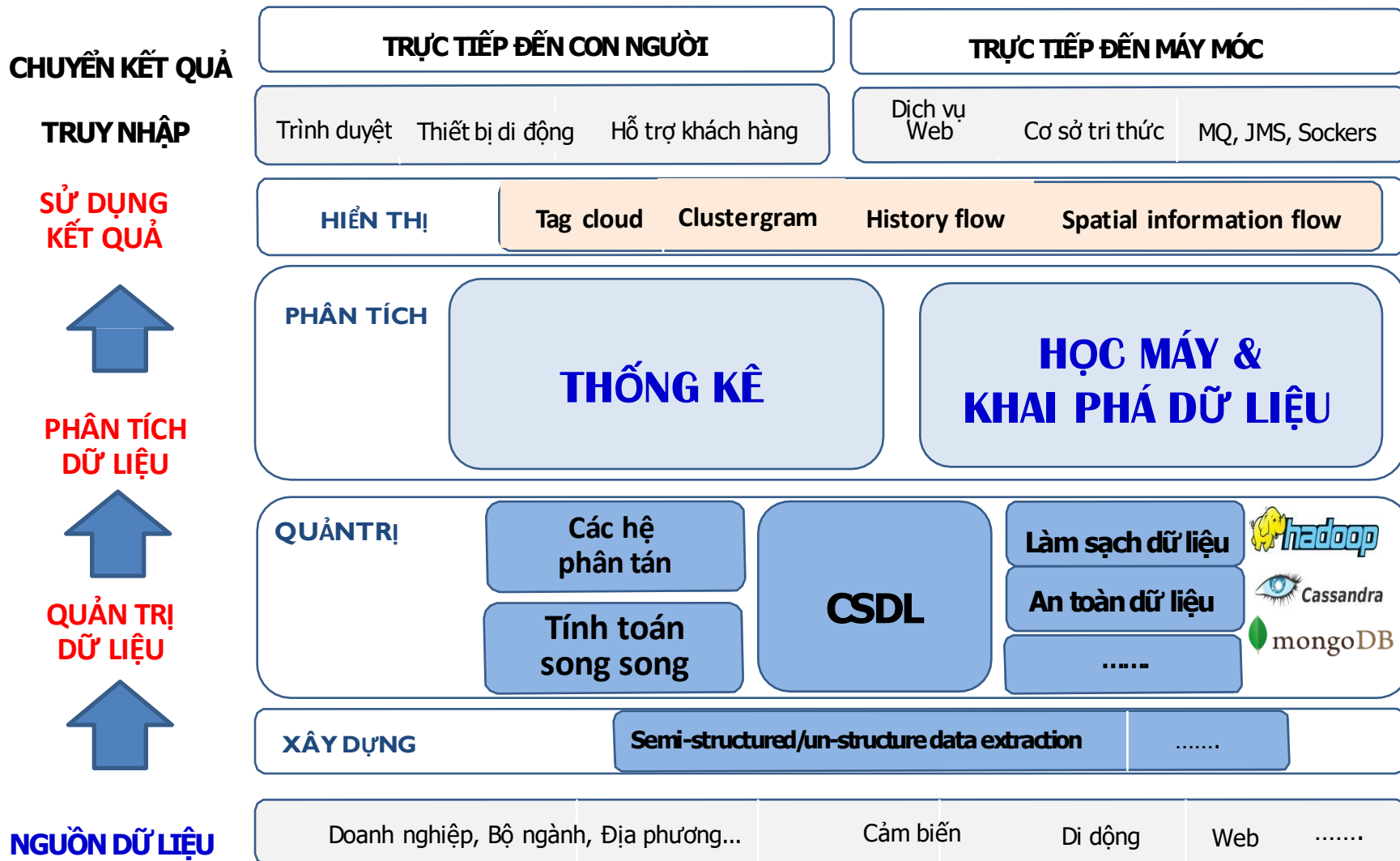
**Ta không thể giải quyết các  
vấn đề với chính cách nghĩ ta  
đã dùng khi đặt vấn đề**

Albert Einstein

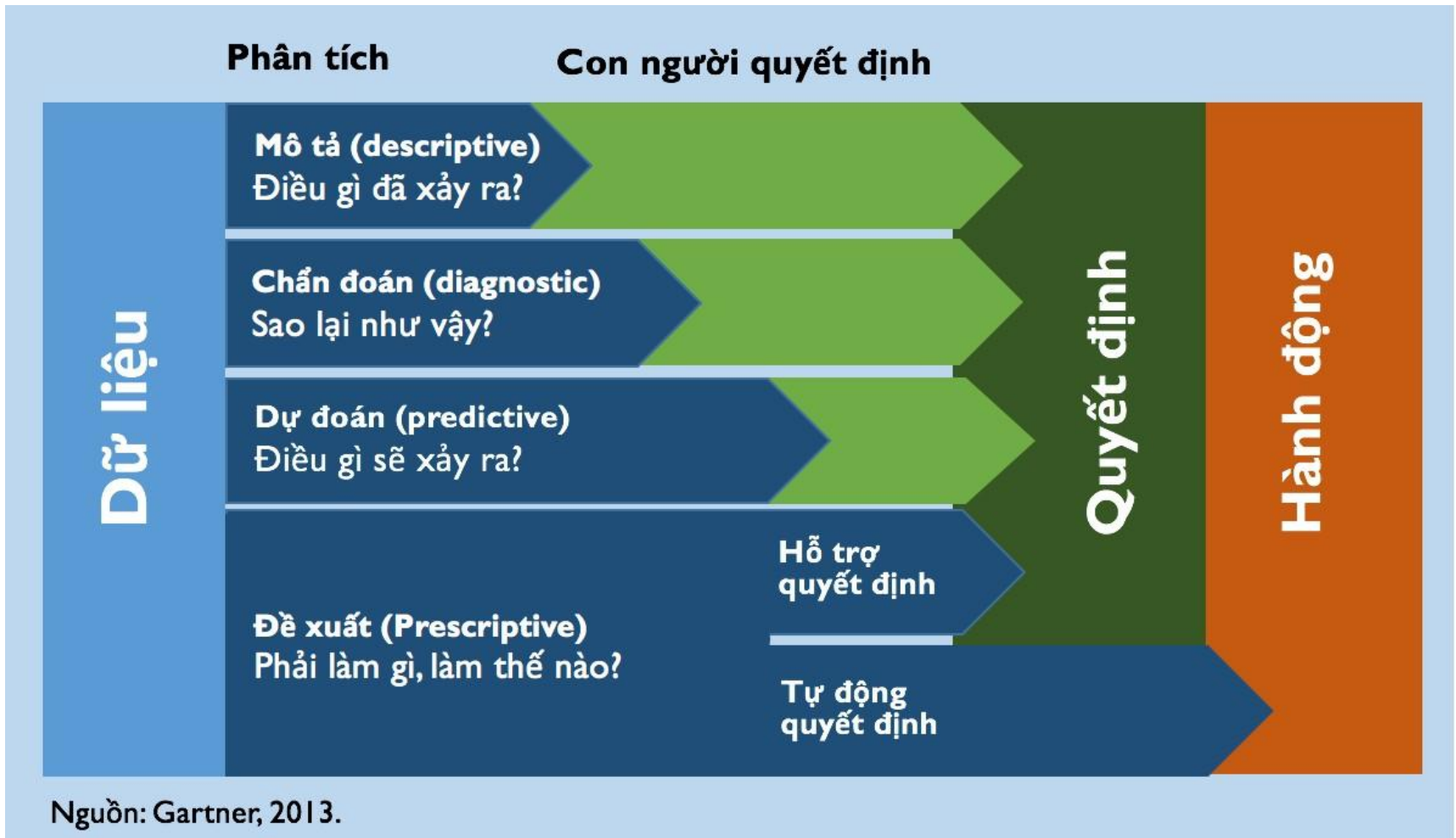


...là công cụ lao động quan trọng trong thời chuyển đổi số

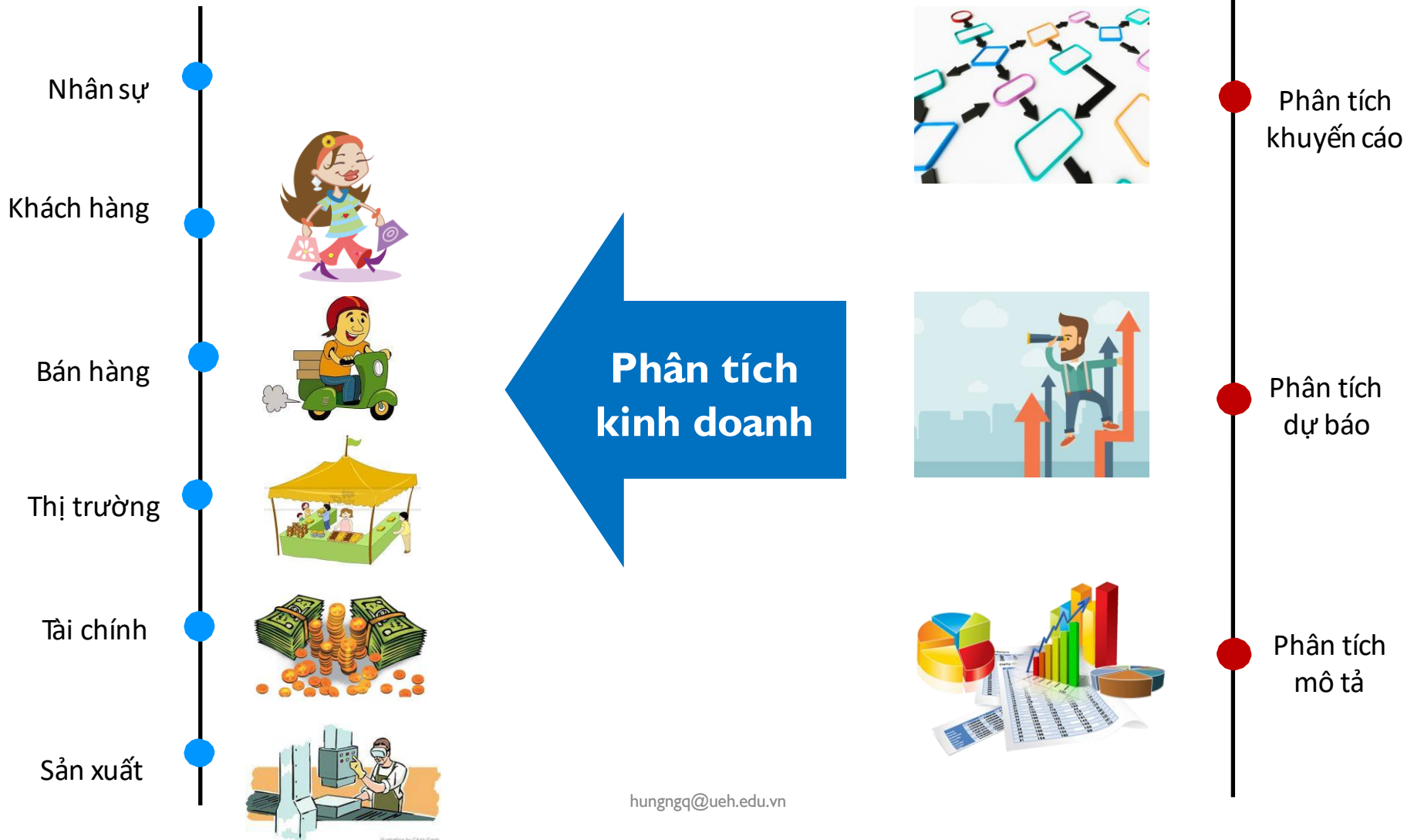
# Lược đồ Data Science



# Hỗ trợ ra quyết định



# Áp dụng vào doanh nghiệp



# 6 bài toán của doanh nghiệp

Tài chính

01



02



Khách hàng

03



Bán hàng

04



05



Vận hành  
sản xuất

06



Nhân sự



# 6 bài toán của doanh nghiệp



## Khách hàng



Sự hài lòng của khách hàng

Khách hàng trung thành

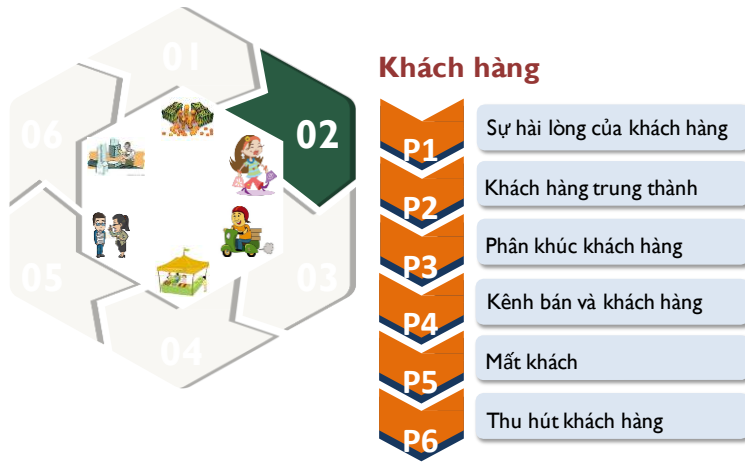
Phân khúc khách hàng

Kênh bán và khách hàng

Mất khách

Thu hút khách hàng

# 6 bài toán của doanh nghiệp



- Doanh nghiệp có đang cung cấp đúng những sản phẩm, dịch vụ mà cái khách hàng mong muốn?
- Khách hàng có hài lòng về trải nghiệm khi mua sản phẩm, dịch vụ của doanh nghiệp?
- Khách hàng có hài lòng về sản phẩm, dịch vụ của doanh nghiệp? Họ hài lòng ở mức độ nào và đâu là yếu tố khiến họ hài lòng nhất? Nếu không thì nguyên nhân là vì sao?

# Trao đổi thảo luận

