

HỌC MÁY CƠ BẢN

AI Academy Vietnam

Nội dung buổi học

- Giới thiệu về khóa học.
- Giới thiệu về Trí Tuệ Nhân Tạo (AI và vai trò của học máy).
- Khái niệm về học máy.
- Các dự án trong khóa học.

GIỚI THIỆU KHÓA HỌC

Học Viện Trí Tuệ Nhân Tạo

- Được thành lập bởi các nhà nghiên cứu, giảng viên, chuyên gia ngành, doanh nghiệp kết hợp lại nhằm tạo dựng môi trường học tập và ứng dụng Trí tuệ nhân tạo để cùng sáng tạo và đổi mới.
- **Các chương trình đào tạo cho cộng đồng:**
 - Full stack Data science Master program
 - Industry4.0 and Smart city for government officers
 - AI for Business leaders
 - AI for All
- **Các chương trình đào tạo In-house cho các tập đoàn:**
 - In-house, specialized AI courses for enterprises

Thời lượng, Phương pháp

- Bao gồm 12 + 1 bài
 - 12 Bài lý thuyết + Thực hành
 - 1 Workshop kết nối công nghiệp
 - Hỗ trợ sau khóa học
- Nằm trong chương trình tổng thể
 - Chương trình Khoa học Dữ liệu
 - Chương trình Trí tuệ nhân tạo
- Với cách tiếp cận dựa trên dự án (project-based), người học có được các kỹ năng:
 - Sử dụng các kỹ thuật và công cụ của học máy
 - Thực hành và giải quyết các vấn đề thực tế

Nội dung cơ bản

- Học viên được tiếp cận các khái niệm, kỹ thuật, và mô hình học máy cơ bản và phổ dụng
 - Học có giám sát (Supervised learning)
 - Học không giám sát (Unsupervised learning)
 - Giới thiệu về Học sâu (Deep learning).
- Học viên được tiếp cận với các ứng dụng của học máy với nhiều ví dụ minh họa, dữ liệu, và dự án từ các bài toán thực tế.
 - Dự đoán/dự báo giá cả
 - Đánh giá tín dụng cá nhân
 - Phân loại phương tiện giao thông/ảnh/tin tức
 - Phát hiện thâm nhập trong an ninh mạng

WORKSHOP

- Trình diễn kết quả
- Giới thiệu việc làm

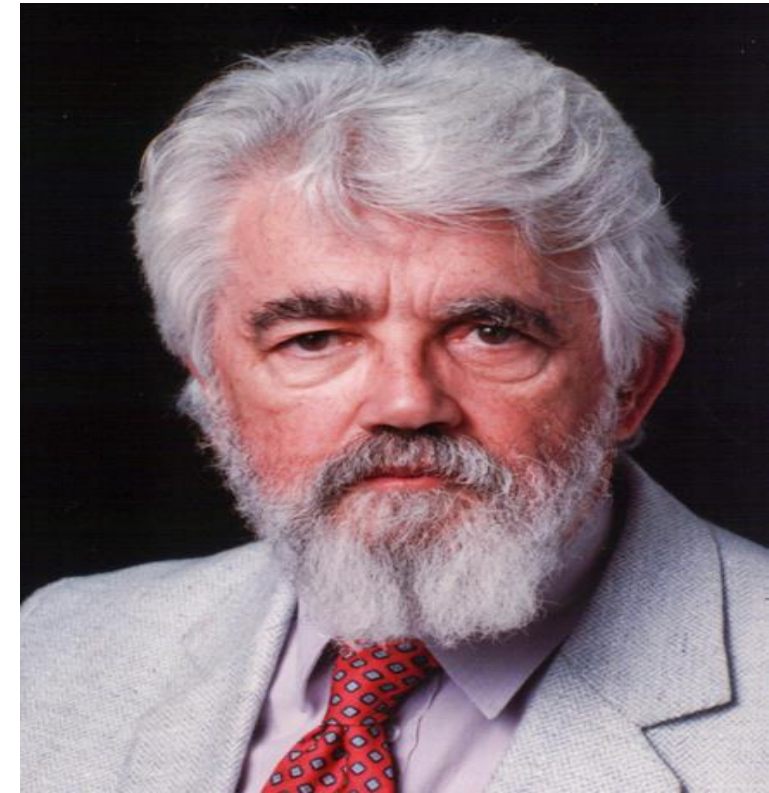
Phụ trợ trong khóa học

- Website quản lý: <http://103.216.114.222/moodle/>
 - Các nội dung, bài giảng
 - Các công cụ tương tác học viên – giáo viên
 - Các dự án, đánh giá và phản hồi
 - Yêu cầu học viên đăng ký qua email
- Dự án trong khóa học:
 - <https://github.com/ai-academy-vn/BasicML>

GIỚI THIỆU VỀ AI

AI là gì?

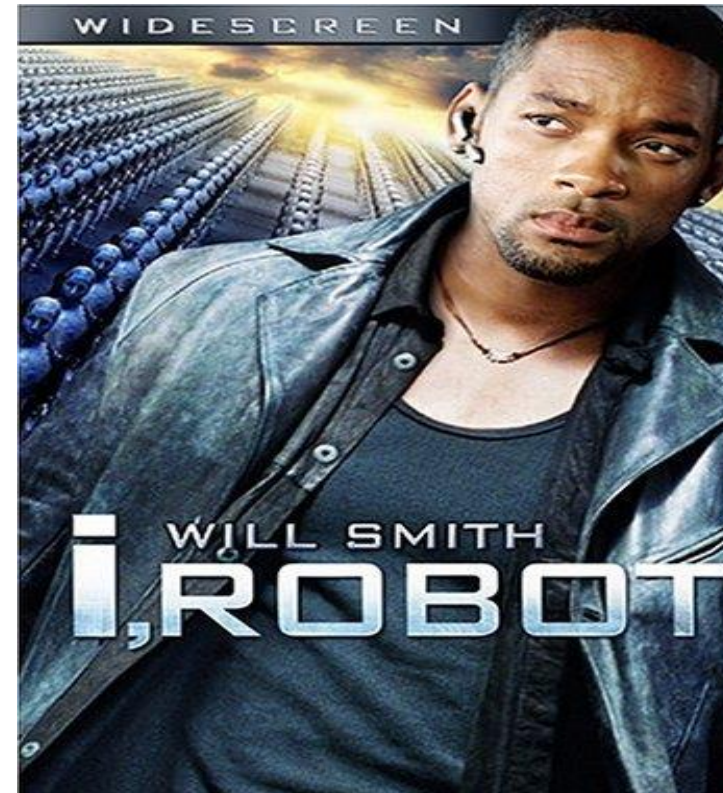
- John McCarthy đưa ra thuật ngữ “Artificial Intelligence” vào khoảng năm 1955.
- Định nghĩa AI không rõ ràng và dễ gây hiểu nhầm.
- 2 quan niệm về AI : **Strong (general) AI** và **Weak (specific) AI**.



Strong AI là gì?

Strong AI: Robot, chương trình AI có thể trở thành 01 giống loài mới (human-being, self-aware).

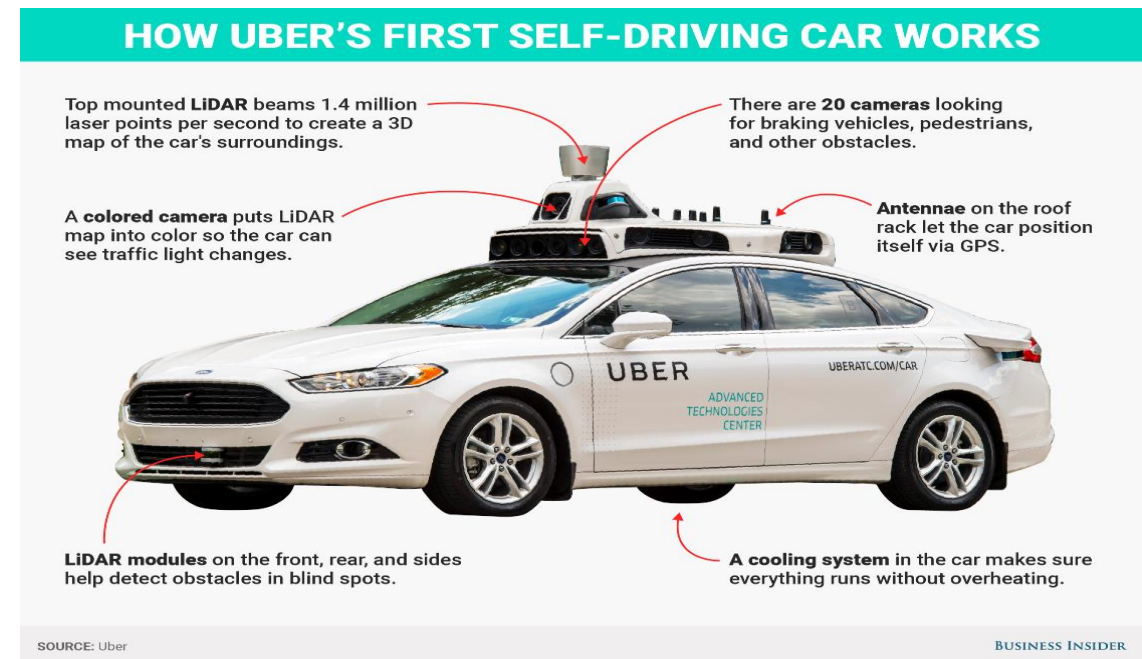
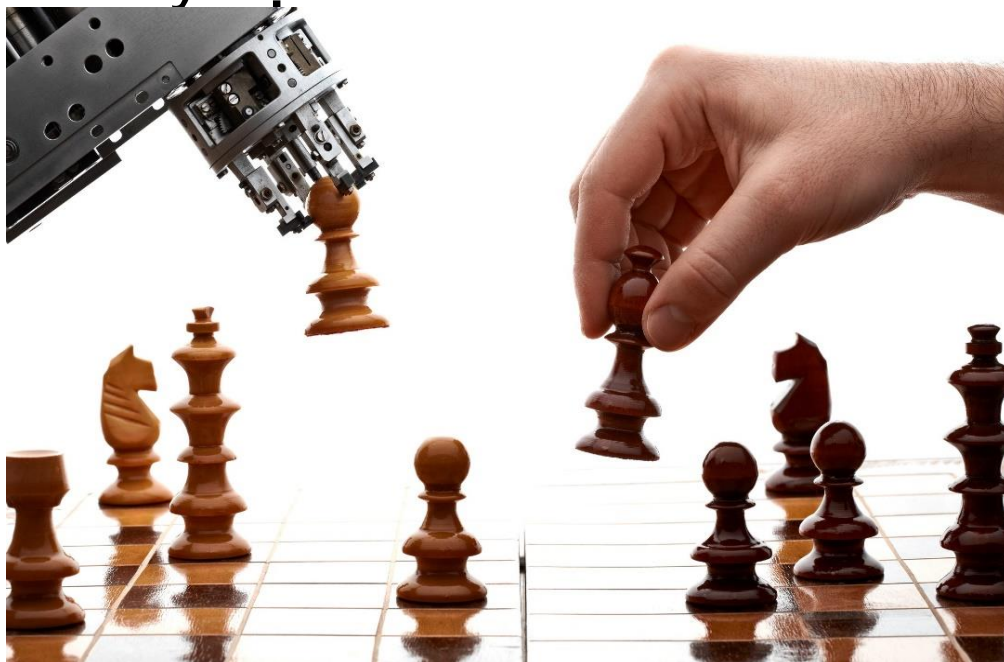
- Máy tính có thể nghĩ, có lý trí, có nhận thức!
- Được kỳ vọng bởi một số nhà khoa học về AI, nhà tương lai học, và Hollywood!



Weak AI là gì?

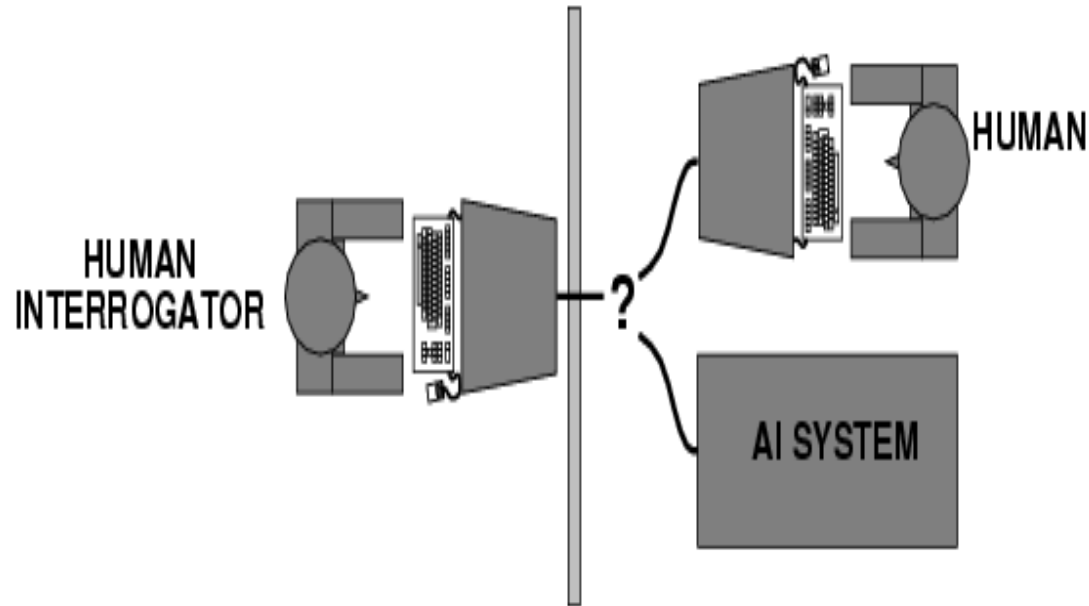
Weak AI: AI chỉ có thể mô phỏng một số hành vi của (trí tuệ) con người.

- Được phần lớn các nhà nghiên cứu AI chấp nhận.
- Hay bị hiểu nhầm thành Strong AI.



Turing Test

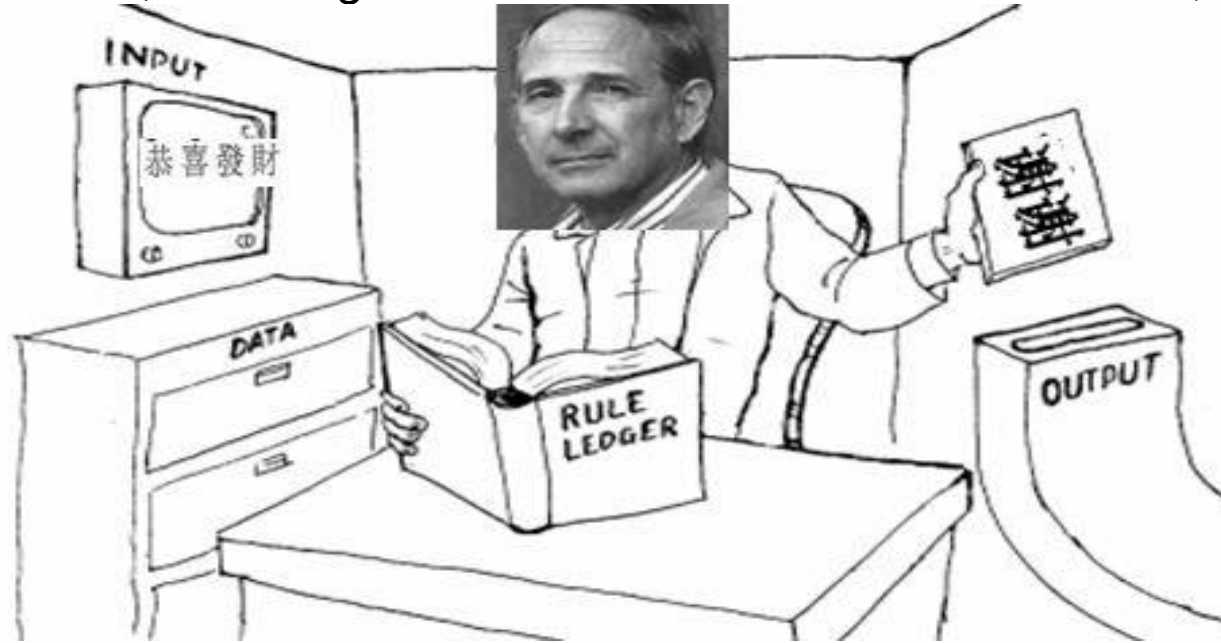
- A. M. Turing (1950) *Computing Machinery and Intelligence*, Mind 49: 433-460.



- Từng có dự đoán vào năm 2000, máy tính AI có 30% cơ hội “đánh lừa” người dùng trong 5 phút.
- Làm cơ sở cho nhiều lập luận trống lại AI trong 50 năm sau đó.
- Gợi ý các thành phần chính mà hệ thống AI phải có: tri thức (knowledge), lập luận (reasoning), xử lý ngôn ngữ tự nhiên (language understanding), học(learning).

Chinese Argument

Searle, John. *Minds, Brains, and Programs*. Behavioral and Brain Sciences 3, 1980, pp. 417-424.



- Các máy tính AI được lập trình không thể thực sự suy nghĩ hay có lý trí mà chỉ có thể mô phỏng con người thậm chí ngay cả khi chúng vượt qua Turing test!

Cách tiếp cận để xây dựng AI

- Top-down: Khoa học nhận thức (Cognitive science) → Symbolism (Allen Newell and Herbert A. Simon) Symbolic representation, Symbolic Computation.
 - General Problem Solver, Planner, Knowledge Based Systems, ...

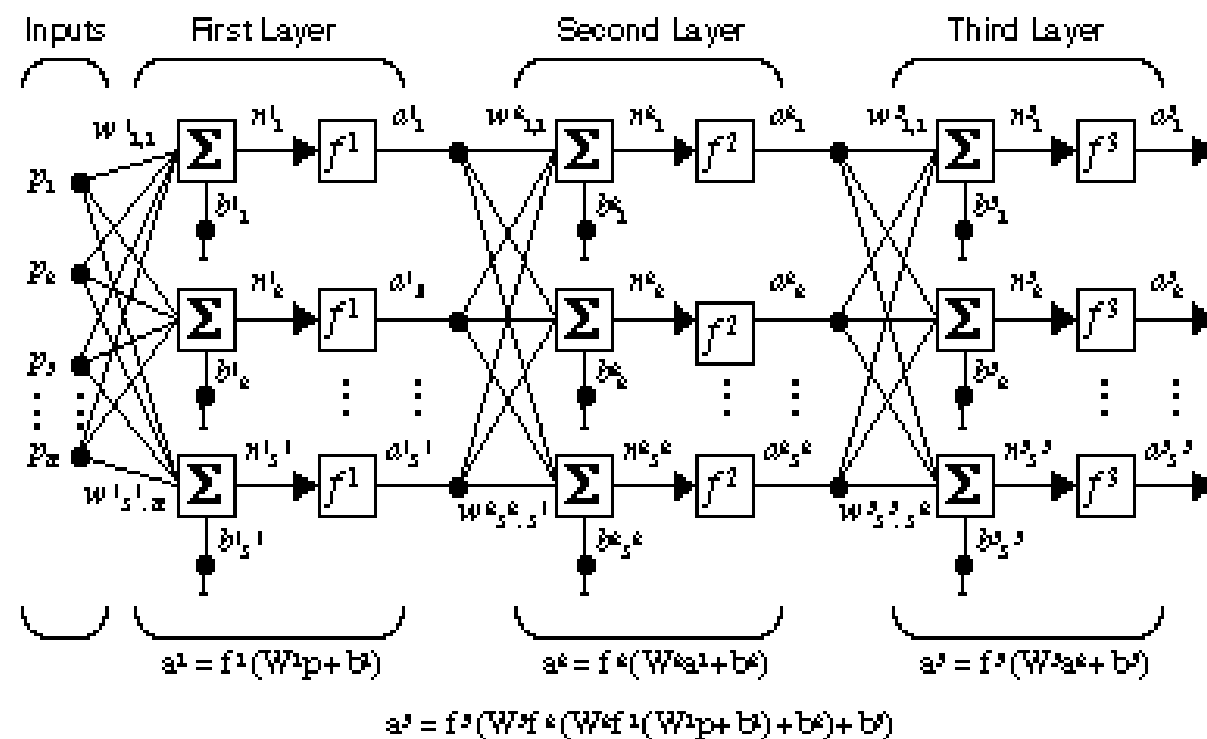
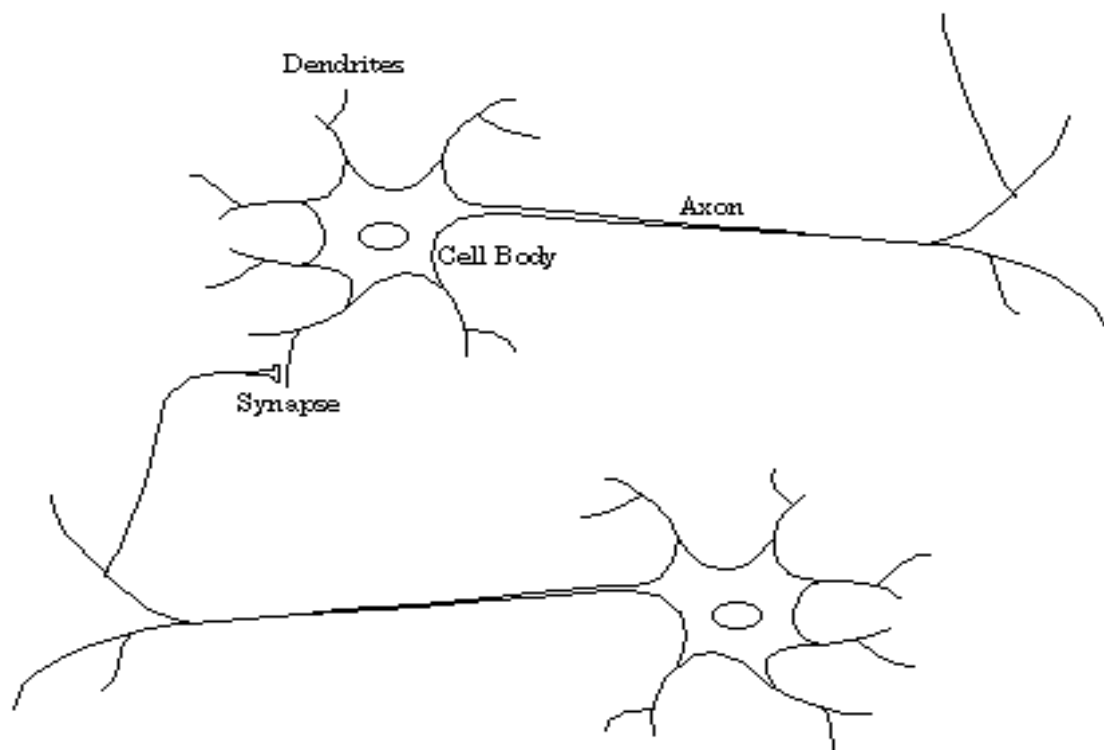


The Physical Symbol System Hypothesis. A physical symbol system has the necessary and sufficient means for general intelligent action.

— Allen Newell —

Cách tiếp cận để xây dựng AI

- Bottom-up: Neural and Brain Science → Connectionism (Sub-symbolic)
 - Artificial Neural Networks → Deep learning, Uncertainty modeling (probabilistic models), ...



Thành Tựu AI trong quá khứ

- MYCIN (1984, Stanford).
- Proved a mathematical conjecture (Robbins conjecture) unsolved for decades.
- During the 1991 Gulf War, US forces deployed an AI logistics planning and scheduling program that involved up to 50,000 vehicles, cargo, and people
- Deep Blue defeated the reigning world chess champion Garry Kasparov in 1997.
- Gulf War 2 (2003), Artificial War.
- NASA's on-board autonomous planning program controlled the scheduling of operations for a spacecraft.
- Human identification through eyes detection and analysis at Heathrow airport using evolutionary computation technique.
- Washing machine generation using NeuroFuzzy Technology, Robotic Vacuum-Cleaner, ...
-

AI ngày nay

Weak AI được ứng dụng để thay thế và hỗ trợ con người trong:

- Thu nhận thông tin từ dữ liệu (lớn):

- Xử lý, phân tích, tổng hợp, hiểu thông tin từ ảnh, video (Computer vision).
- Xử lý, phân tích, tổng hợp, hiểu âm thanh (Speech Recognition).
- Xử lý, phân tích, tổng hợp, hiểu ngôn ngữ (Natural Language Processing).

- Hỗ trợ ra quyết định:

- Dự đoán, dự báo.
- Tối ưu.
- Gợi ý.

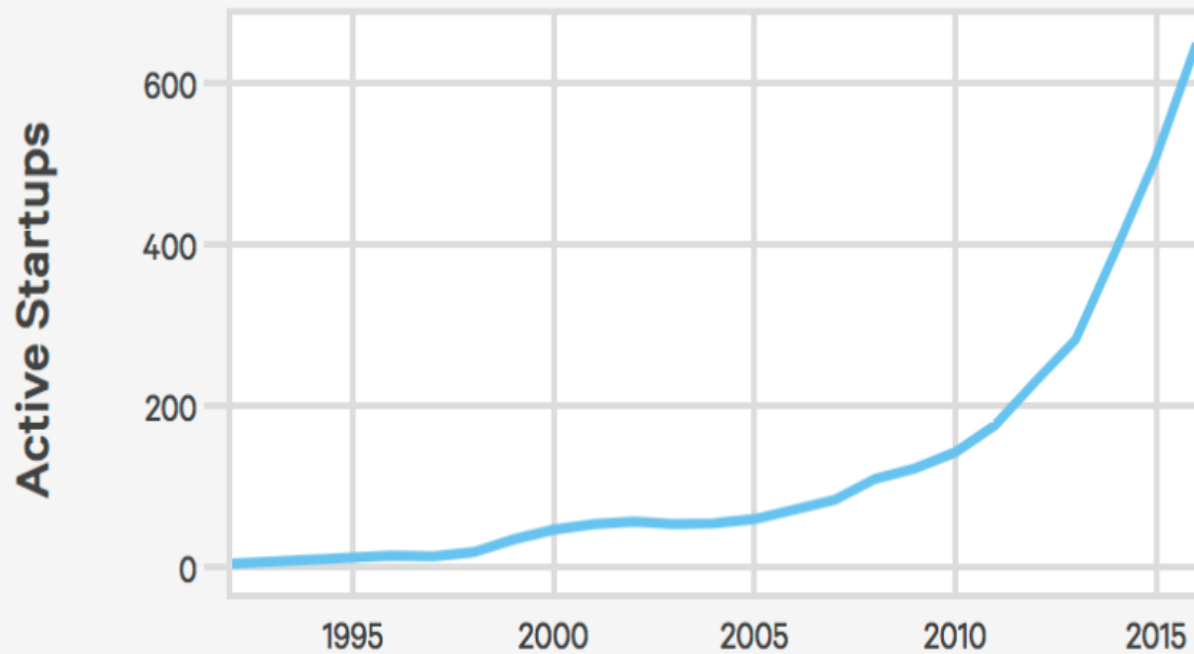
AI ngày nay

- Nhiều hệ thống AI dựa trên Deep Learning đã bằng hoặc vượt khả năng của con người trong một số lĩnh vực hẹp :
 - Computer Vision, Pattern Recognition, Speech Recognition, Machine Translation, Game Playing,
- AI đã được ứng dụng trong mọi ngóc ngách của cuộc
 - AI grant projects (Google Brain, OpenAI, ...)
- AI đã thay đổi nhiều ngành nghề truyền thống và chuyển đổi nhiều công ty lớn (Google Facebook, Amazon, ...) và trở thành 01 ngành công nghiệp mới!

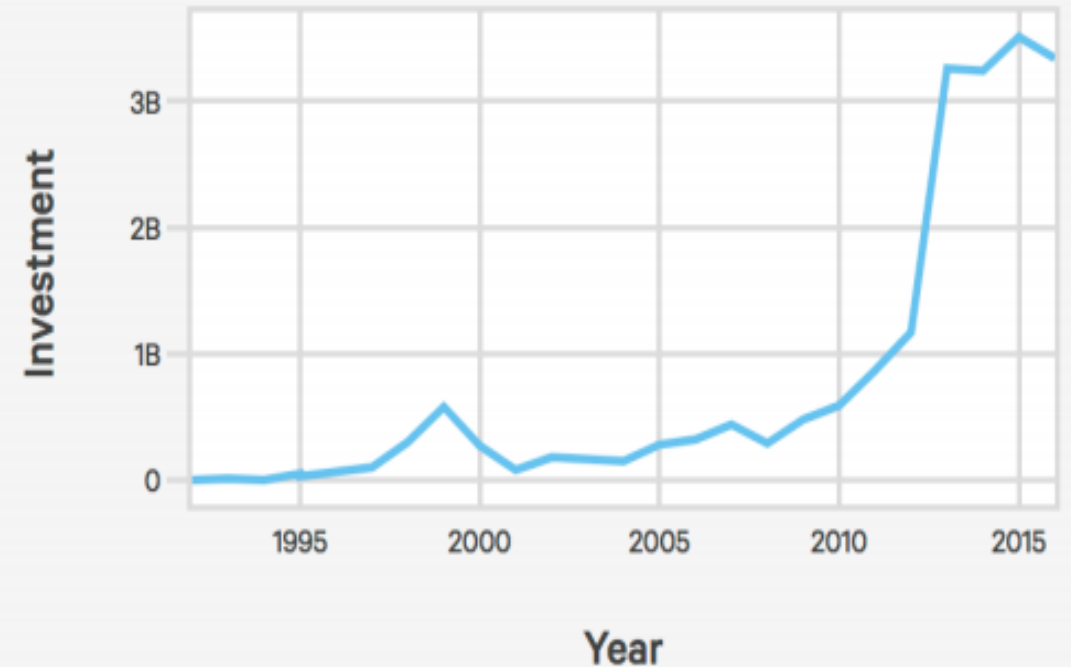


AI ngày nay (US AI Index 2017)

Startups Developing AI Systems

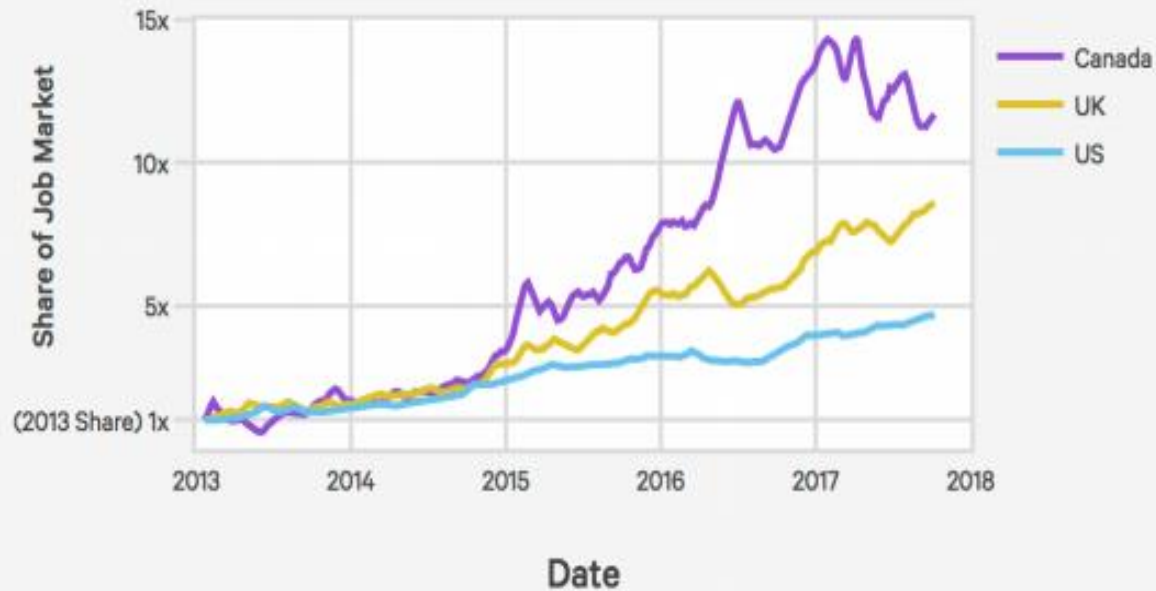


Annual VC Investment in AI Startups

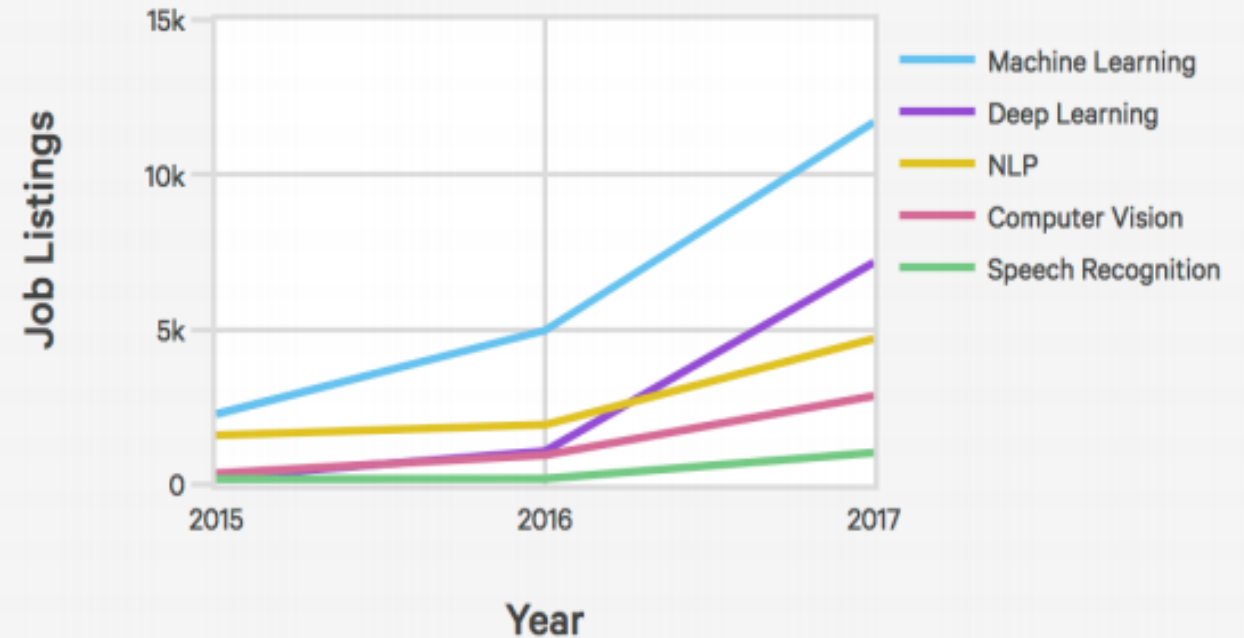


AI ngày nay (US AI Index 2017)

Share of Jobs Requiring AI Skills (Indeed.com)

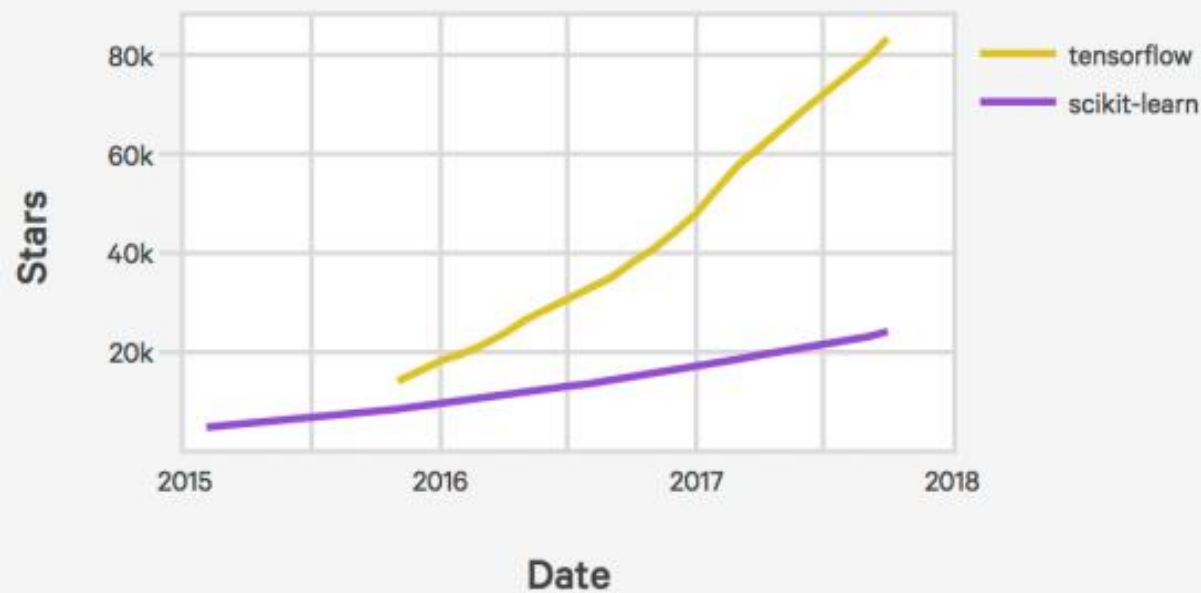


Job Openings, Skills Breakdown (Monster.com)

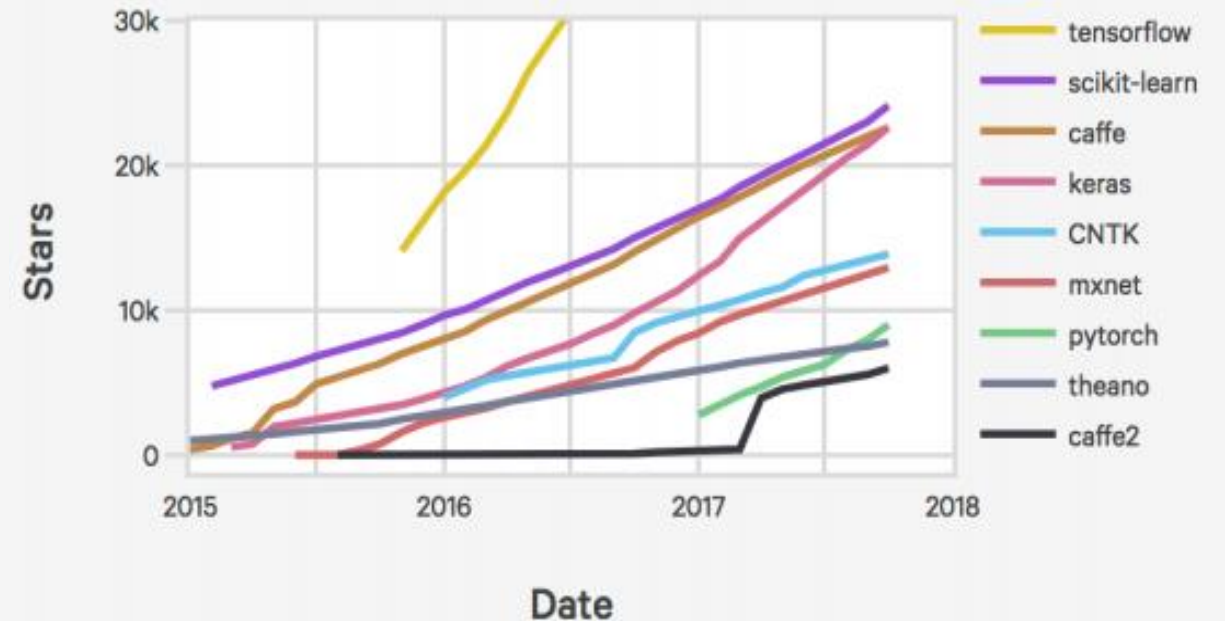


AI ngày nay (US AI Index 2017)

GitHub Stars for TensorFlow & Scikit-Learn



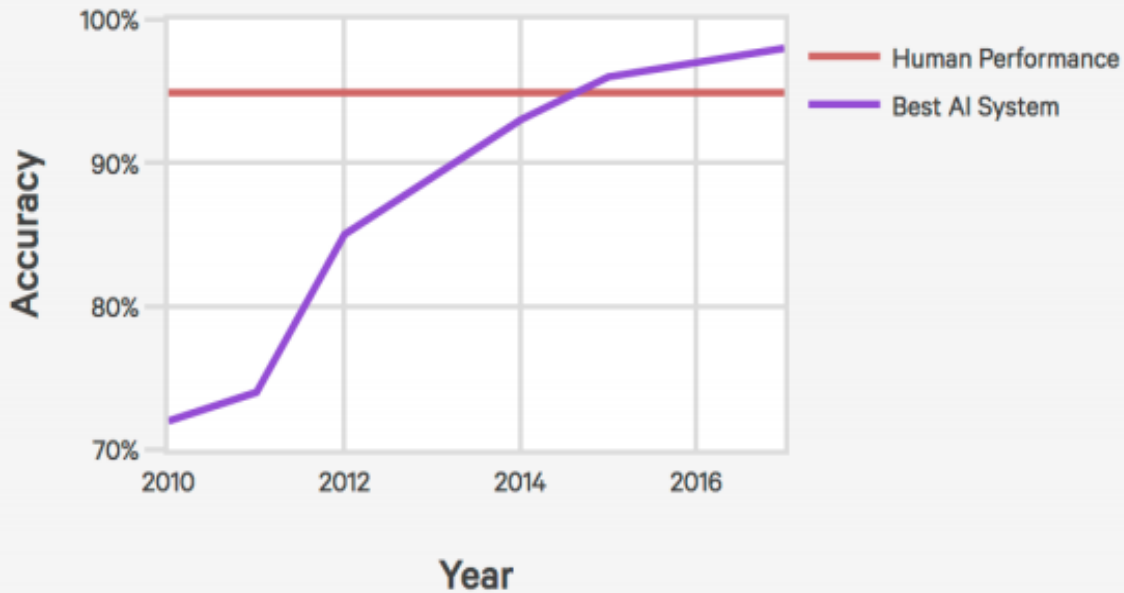
GitHub Stars of AI Software Libraries



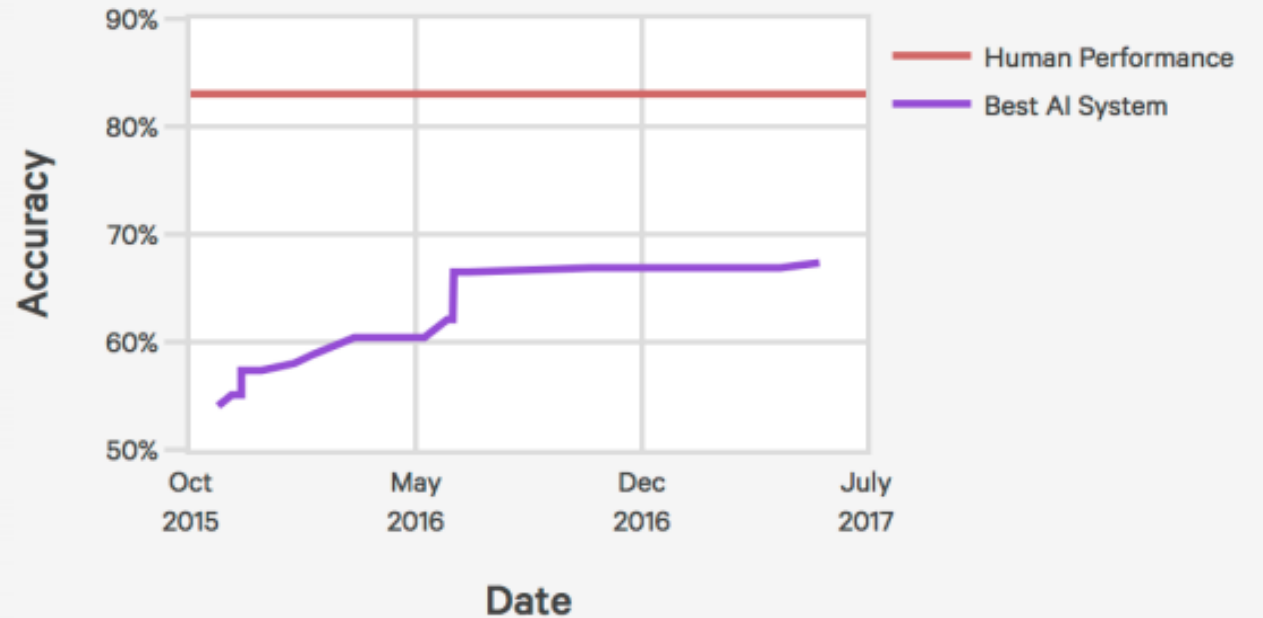
AI ngày nay (US AI Index 2017)

Human vs. Weak AI - Vision:

Object Detection, LSVRC Competition



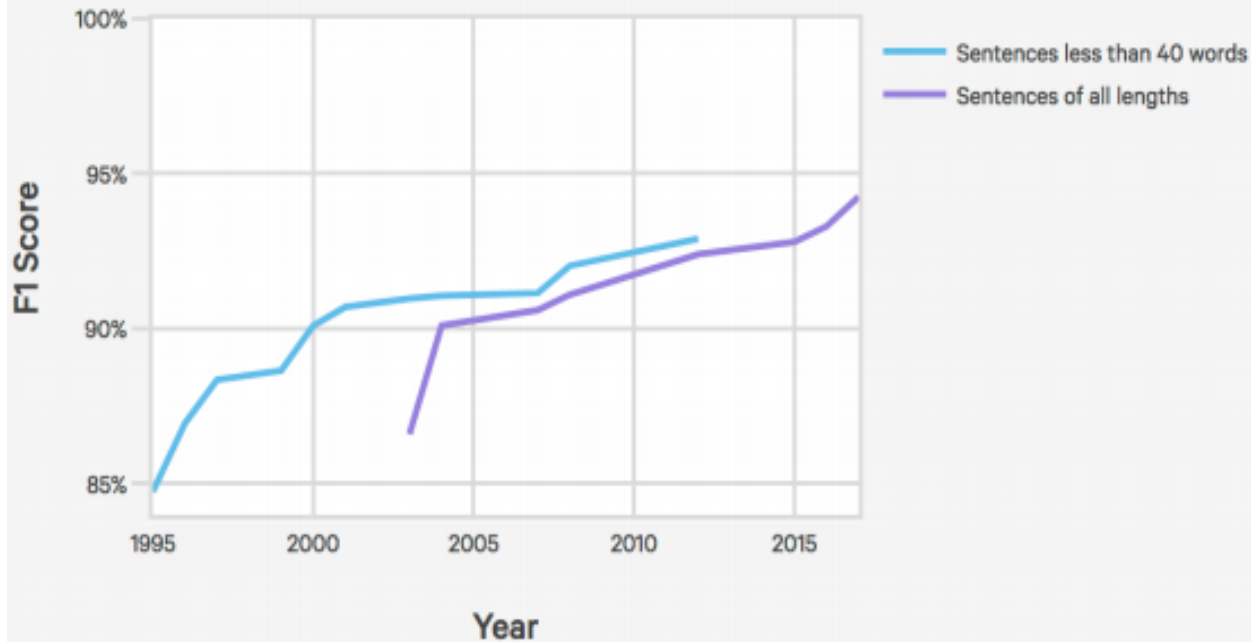
Visual Question Answering, VQA 1.0



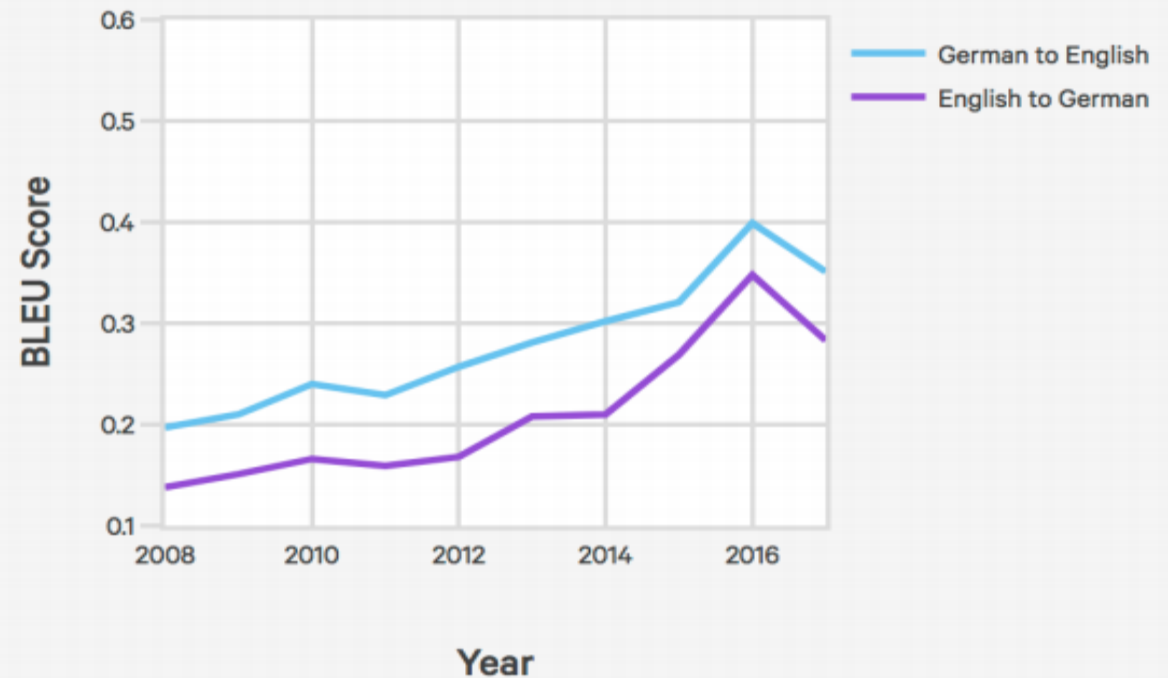
AI ngày nay (US AI Index 2017)

Human vs. Weak AI - NLP:

Constituency Parsing, Penn Treebank WSJ



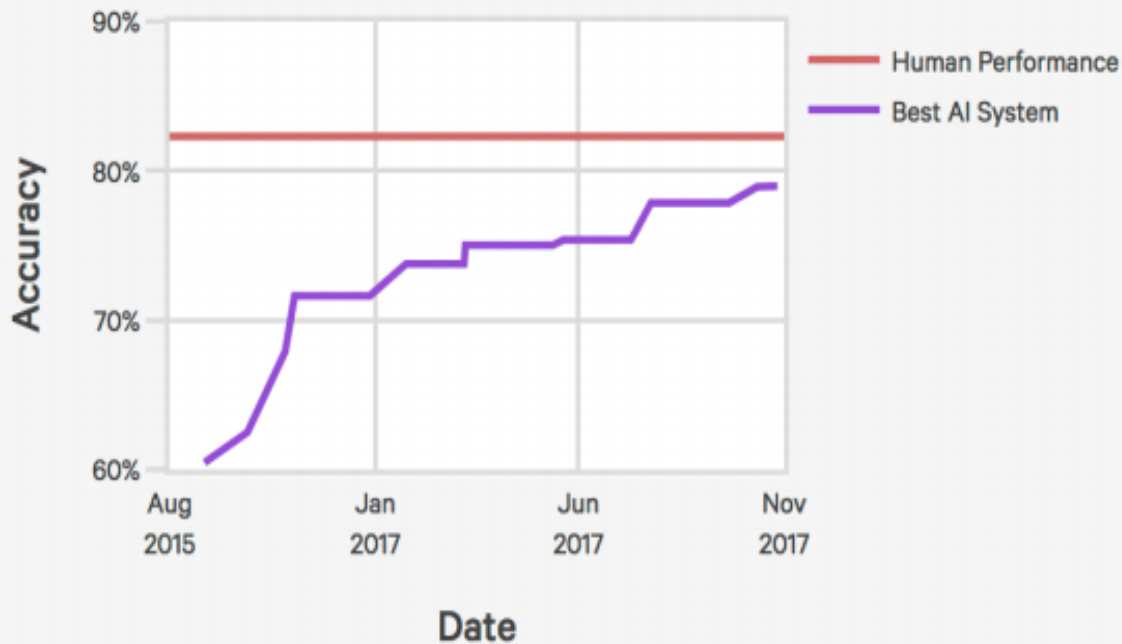
News Translation, WMT Competition



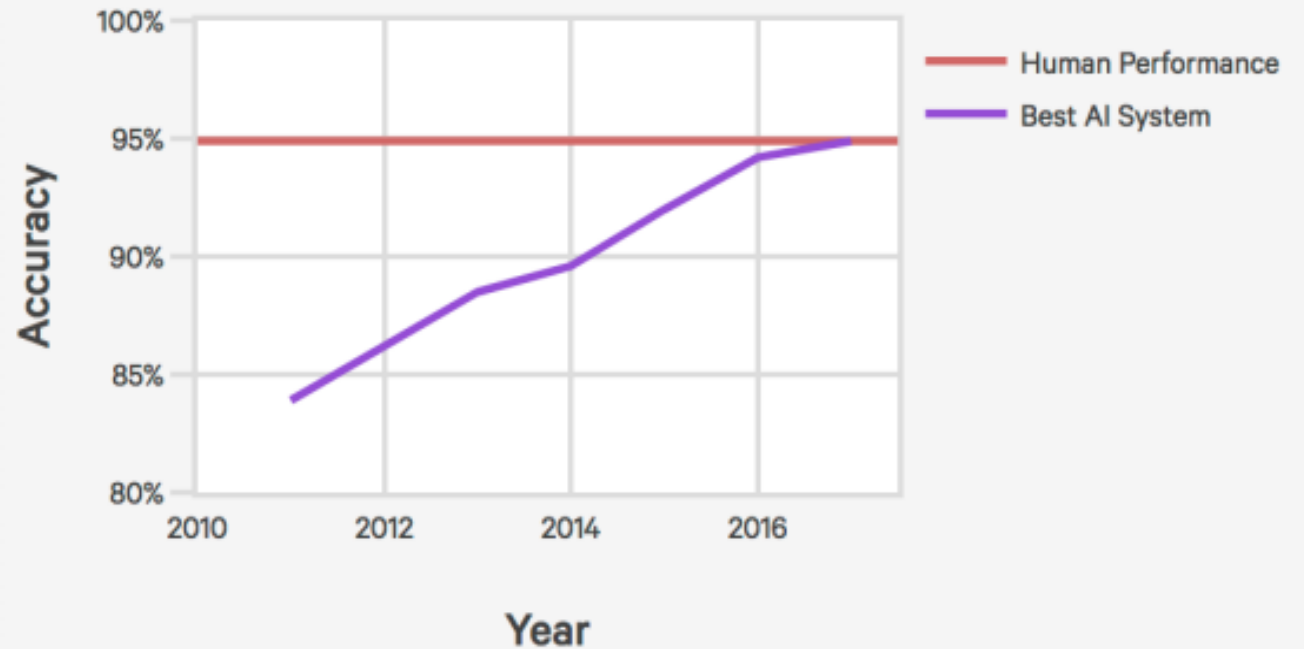
AI ngày nay (US AI Index 2017)

Human vs. Weak AI – NLP, Speech Recognition:

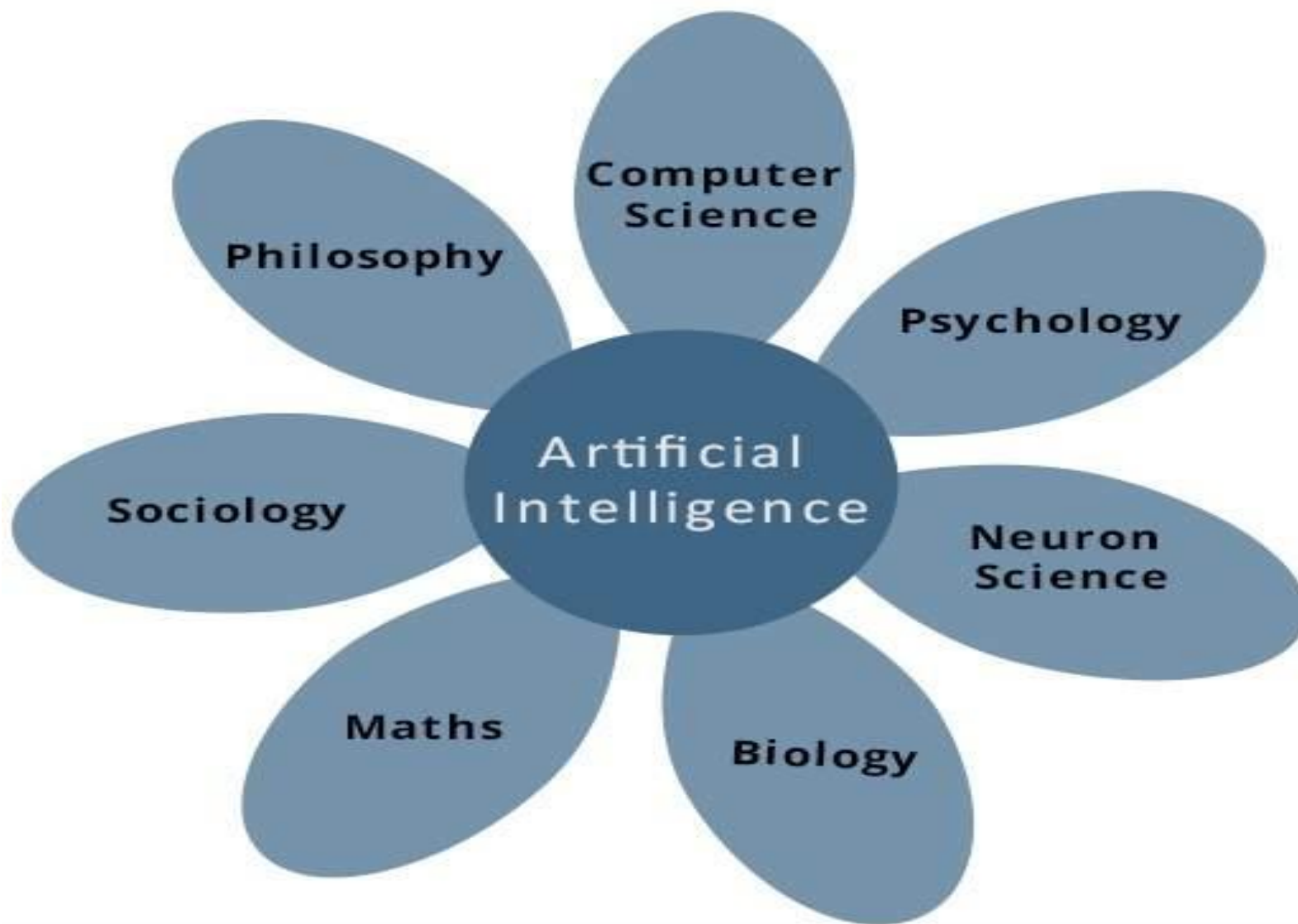
Question Answering, SQuAD v1.1



Speech Recognition, Switchboard HUB5'00



Các lĩnh vực liên quan đến AI



Các lĩnh vực thuộc AI

- Machine Learning.
- Data Science, Mining and Knowledge Discovery.
- Computer Vision.
- Natural Language Processing.
- Speech Recognition.
- Evolutionary and Natural Computation.
- Fuzzy Computation and Technologies.
- Artificial Life.
- Knowledge-Based Systems.
- Automated Reasoning.
- Logic and Constraint Programming.
- Intelligent Planning.
-

Lược Sử AI

Early Explorations (50s-60s)

- **Nhiều ý tưởng và mô hình quan trọng cho AI.**
Evolutionary Computation, Neural Networks, Logic based AI, Heuristics, Semantic Representation, Statistical Methods, ...
- **Xác lập các bài toán trong AI**
Pattern Recognition, Natural Language Processing, Planning, Automated Proving, General Problem Solver, Game Playing,
- **Sự kiện quan trọng:**
Dartmouth Summer Meeting (1956)



Lược Sử AI

Symbolism (70s-mid 80s)

■ Hệ thống và kỹ thuật AI:

- Knowledge Representation and Reasoning.
- Mobile Robot and Automated Planning Systems.
- Expert Systems (Knowledge based Systems).
- MYCIN (Stanford University)
- Logic Programming.
- ...

■ Các bài toán AI:

- Computer Vision.
- Speech Recognition.
- Natural Language Processing.
- Game Playing.

■ ...

Lược Sử AI


Symbolism (70s-mid 80s)

■ Sự kiện đáng chú ý:

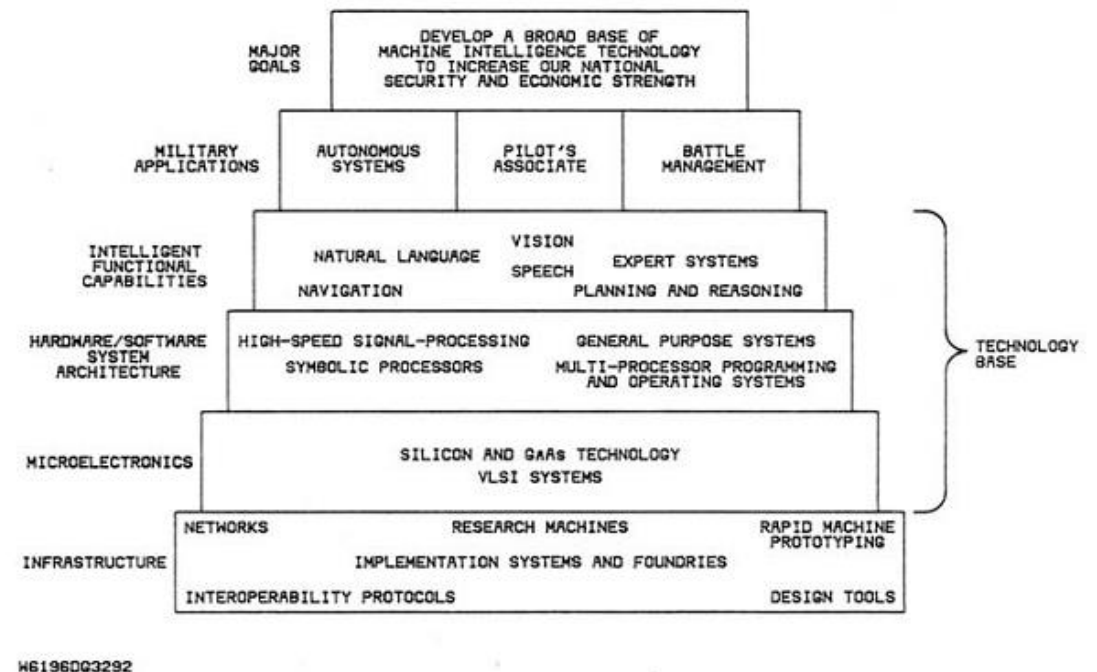
- Japanese Project of Fifth Generation Computers.
- DARPA's Strategic Computing Program (USA).

FIFTH GENERATION COMPUTERS

- Capable of parallel processing
- Support voice recognition
- Understand natural language



Buzzie.com



Lược Sử AI

Connectionism/Sub-symbolic (mid 80s-2006)

Kỹ thuật AI:

■ **Neural Networks**

- Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J., Learning representations by back-propagating errors, Nature, 323 (6088): 533–536, 1986.

■ **Hybrid Intelligence Computational Intelligence, Soft-computing.**

- Neural Networks, Evolutionary Computation, Fuzzy Technologies.

■ **Simulation-based Intelligence.**

- Intelligent Multi-agent Systems, Natural Computation, Swarm Intelligence,

■ **Probabilistic and Statistical Methods**

- Probabilistic Graphical Models, Statistical Learning and Inference, Support Vector Machines,

Lược Sử AI

Connectionism/Sub-symbolic (mid 80s-2006)

AI Problems (dominated by probabilistic and statistical methods):

- Machine Learning.
- Computer Vision.
- Natural Language Processing.
- Speech Recognition.

....

Resurgence of Connectionism/Neural Networks (2006-)

- **Deep Neural Networks, Deep Learning**
 - Hinton, G. E.; Osindero, S.; Teh, Y. W., A Fast Learning Algorithm for Deep Belief Nets, Neural Computation. 18 (7): 1527–1554, 2006.
- **AI industry**
- **AI platforms, AI Tools, AI Companies!**

Những Thách Thức Chính của AI

- **AI PROBLEMS IN THE PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE (31/08/1955).**
 1. AUTOMATIC COMPUTERS.
 2. HOW CAN A COMPUTER BE PROGRAMMED TO USE A LANGUAGE.
 3. NEURON NETS.
 4. THEORY OF THE SIZE OF A CALCULATION.
 5. SELF-IMPROVEMENT.
 6. ABSTRACTIONS.
 7. RANDOMNESS AND CREATIVITY.

Những Thách Thức Chính của AI

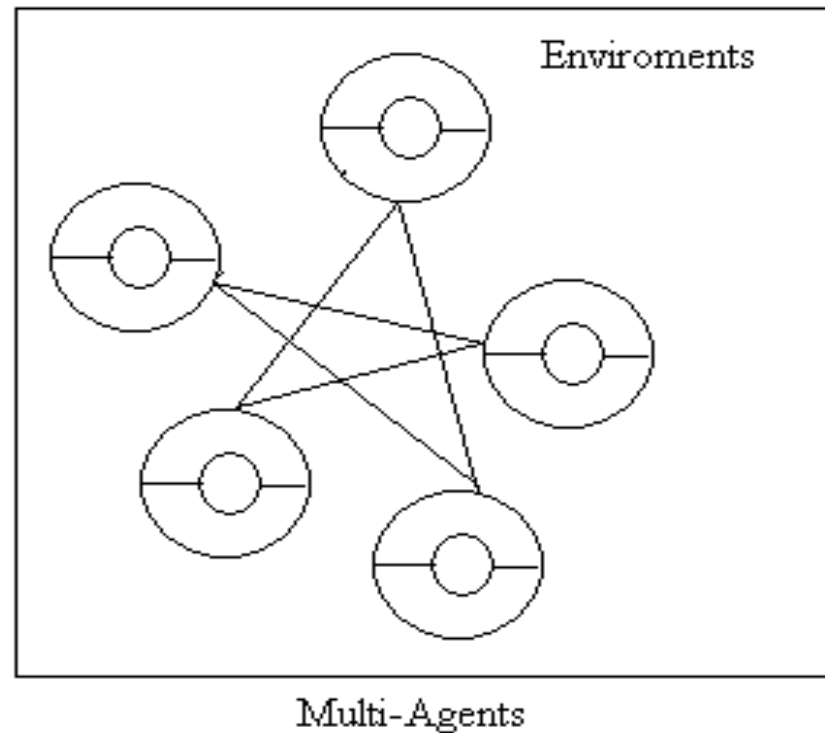
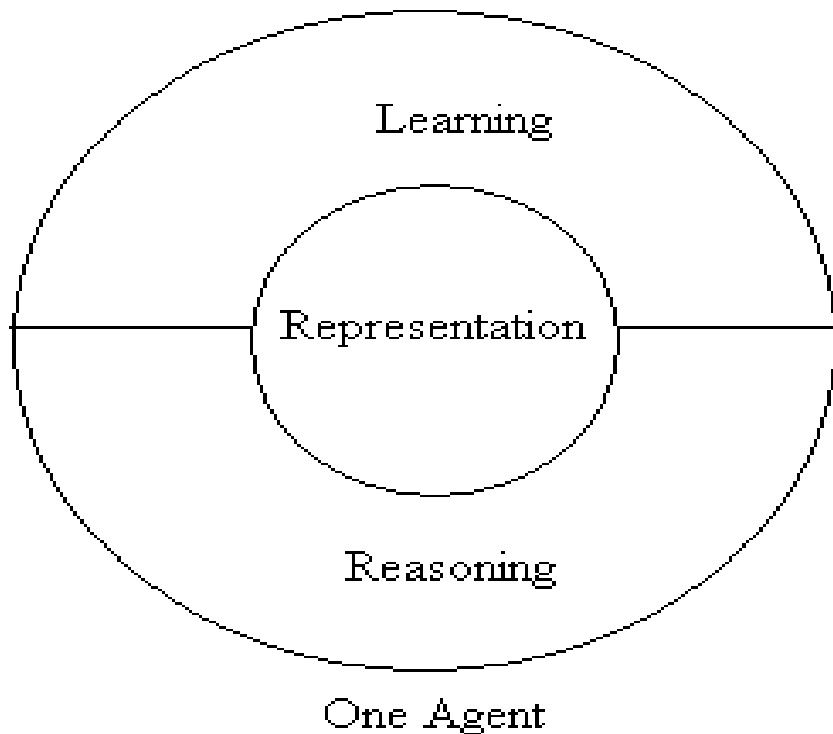
ELON MUSK'S OPENAI CHALLENGES (2016):

1. FINDING COVERT AI SYSTEMS.
2. BUILDING AN AI THAT CAN PROGRAM.
3. USING AI FOR CYBER DEFENSE.
4. MAKING REALLY COMPLEX SIMULATIONS.



Những Thành Phần Cốt Lõi của AI

- **Biểu diễn (Representation).**
- **Lập Luận (Reasoning).**
- **Học (Learning).**
- **Tương Tác (Interaction).**



HỌC MÁY CƠ BẢN

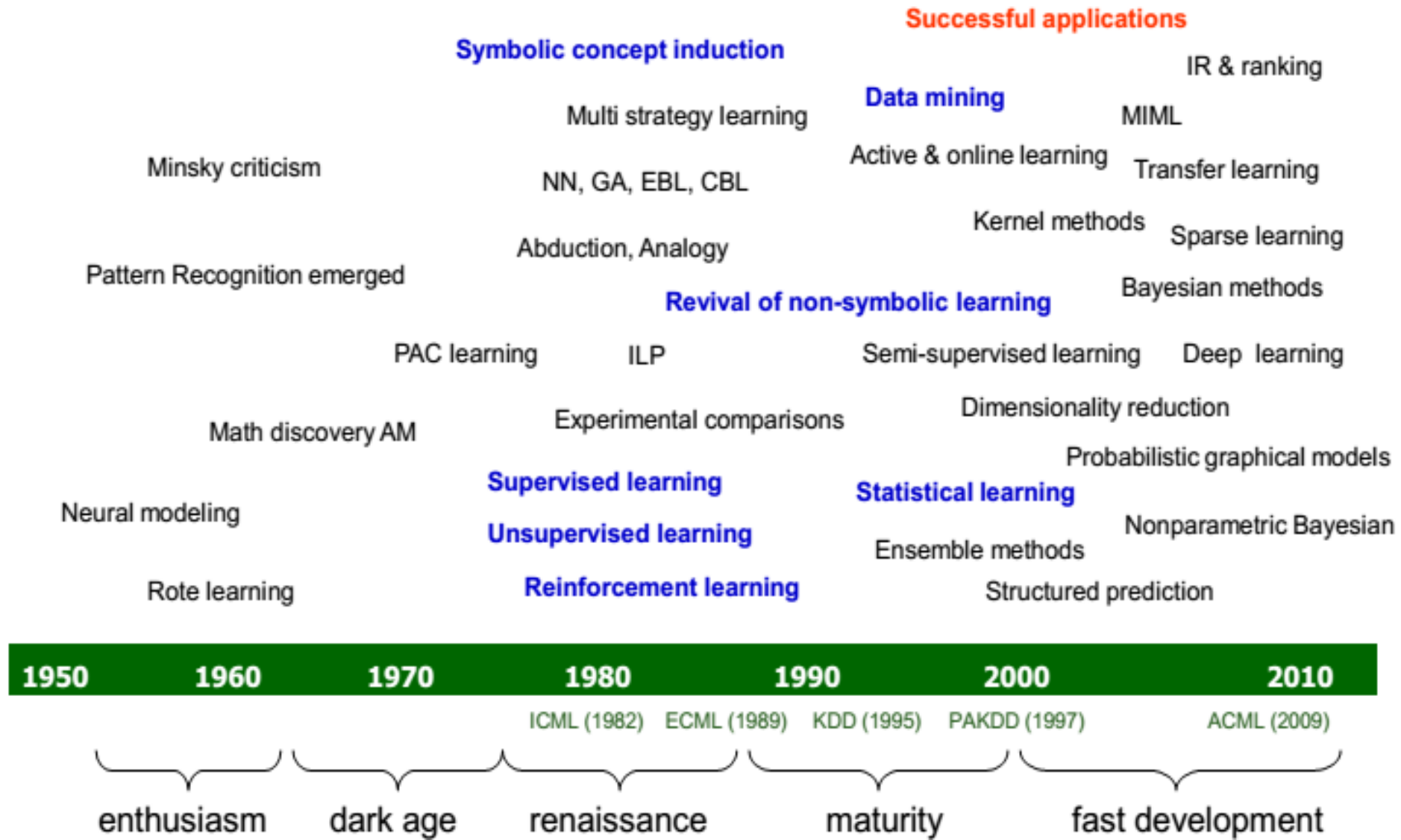
Tại sao lại “Học”?

- Học máy: cho phép chương trình có khả năng tự động, tối ưu hóa dựa trên dữ liệu hoặc kinh nghiệm thay vì các thuật toán.
- Học khi nào:
 - Tri thức con người không hoặc chưa có (tìm đường trên sao hỏa),
 - Con người không thể giải thích tri thức của mình (nhận dạng giọng nói)
 - Giải pháp cần thay đổi liên tục (định tuyến trong mạng máy tính)
 - Giải pháp cho các trường hợp đặc biệt (đối với từng người sử dụng)

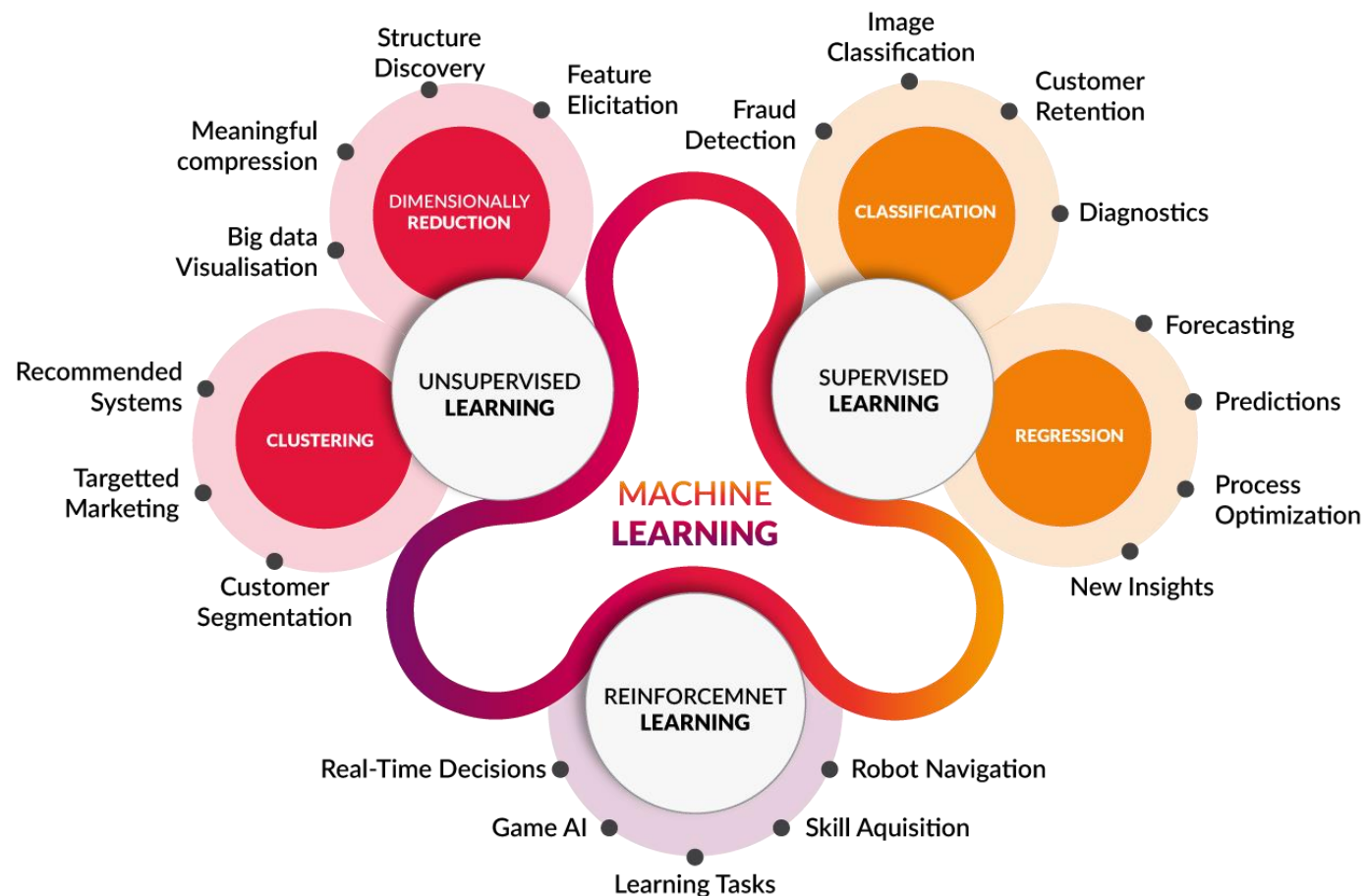
Học như thế nào

- Tạo ra các mô hình từ dữ liệu thực tế thay vì lập trình theo thuật toán cứng.
- Dữ liệu rất nhiều và có thể rẻ (data warehouse, data marts)
- Tri thức là hữu hạn và đắt đỏ (đặc biệt tri thức chuyên gia)
- Ví dụ: Bác sĩ chẩn đoán bệnh, bác sĩ đọc MRI
- Nhiệm vụ: Xây dựng mô hình xấp xỉ dữ liệu

Quá trình Phát Triển (Pre-Deep Learning)

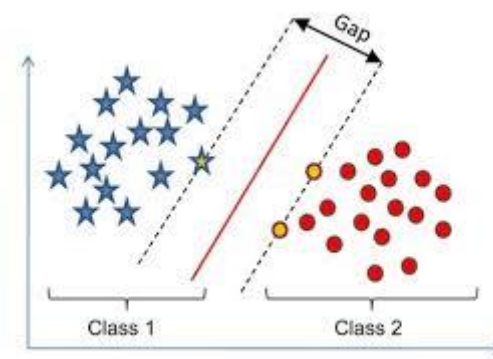
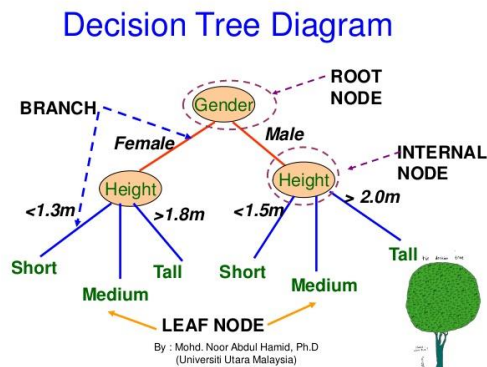
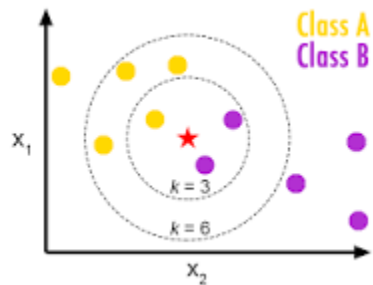


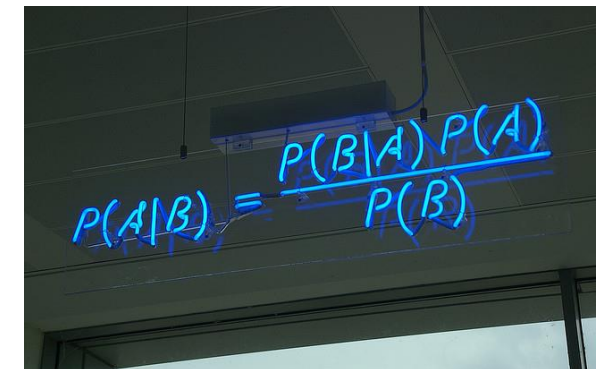
Các phương pháp học máy



Học có giám sát

- Học có giám sát (Supervised learning)
 - Bài toán phân lớp đơn giản kNN và Hồi quy tuyến tính
 - Cây phân lớp (Decision Tree)
 - Support Vector Machines (SVM)
 - Naïve Bayes

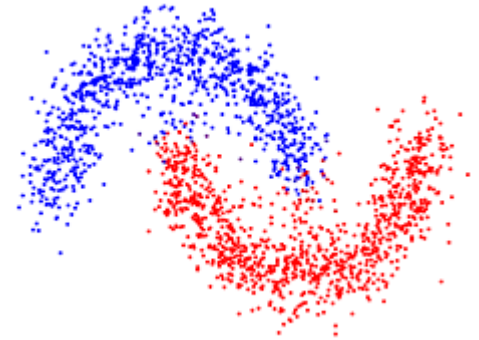
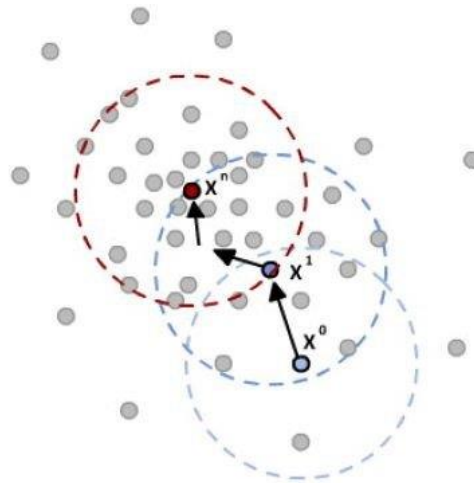
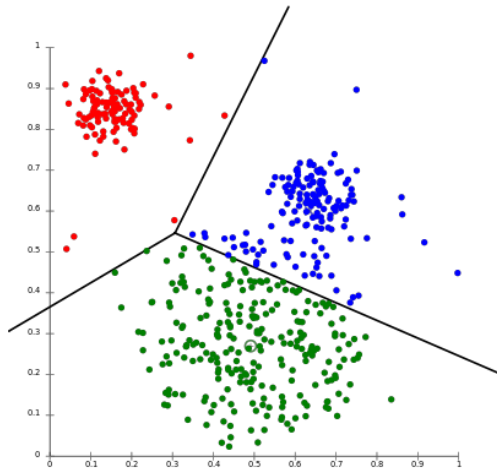




$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

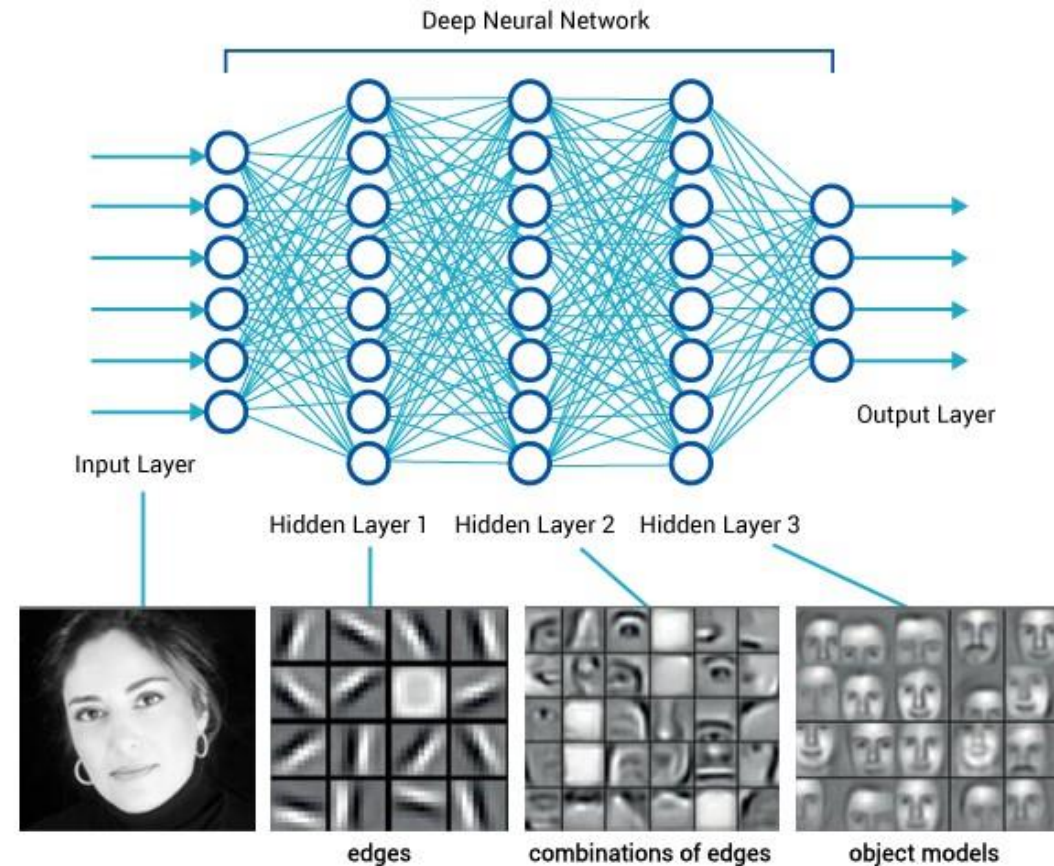
Học không giám sát

- K-means
- Mean-shift
- Phân cụm phổ (Spectral Clustering)



Giới thiệu về Deep Learning

- Mạng nhân tạo đa lớp (Multiple Layer Perceptron)
- Auto encoder
- Deep learning



CÁC DỰ ÁN

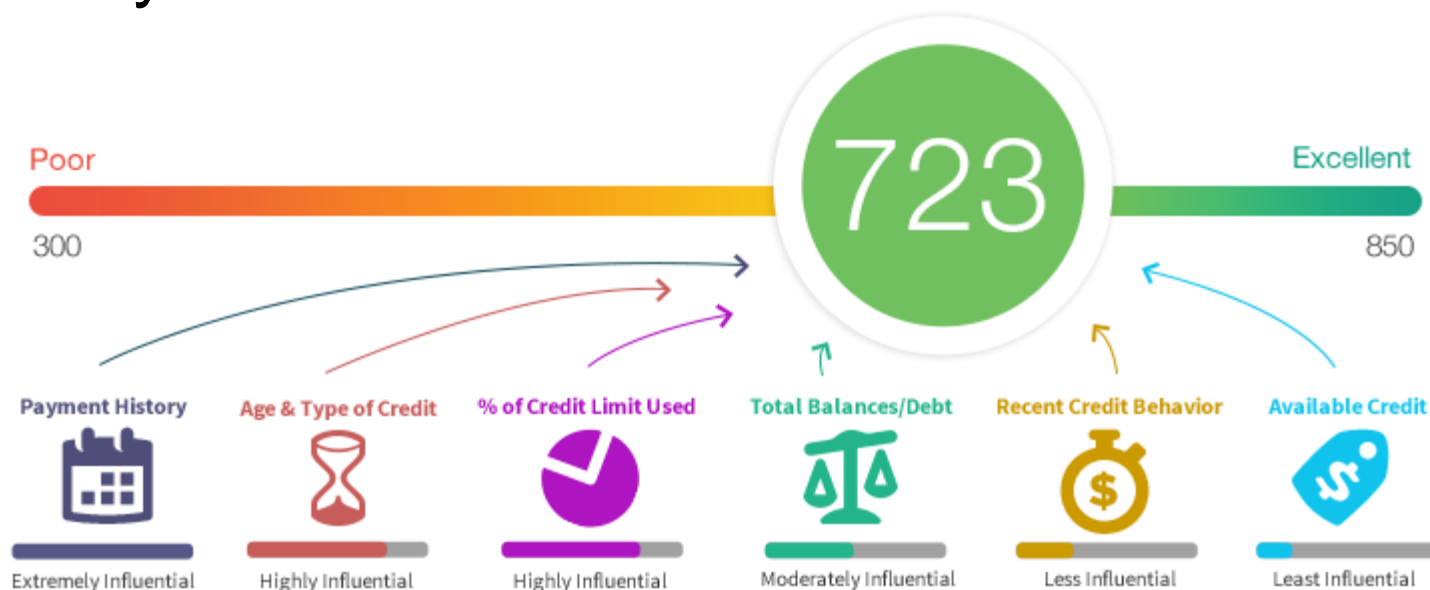
Bài toán định giá bất động sản

- Các thông tin ảnh hưởng tới giá trị của bất động sản
- Thu thập và Phân tích dữ liệu
- Định nghĩa độ đo tương tự
- Xây dựng mô hình học máy
 - **kNN và Hồi quy**
- Đánh giá kết quả



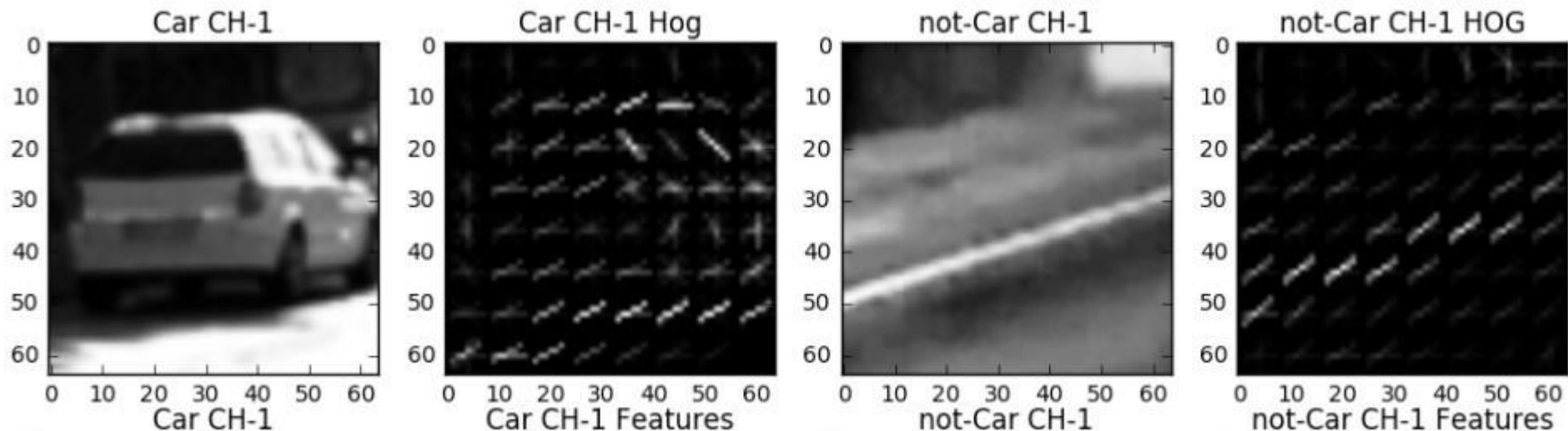
Đánh giá tín dụng cá nhân

- Giới thiệu bài toán Đánh giá tín dụng cá nhân
- Thu thập và Phân tích dữ liệu
- Xây dựng mô hình học máy
 - **Cây Quyết định**
- Đánh giá kết quả



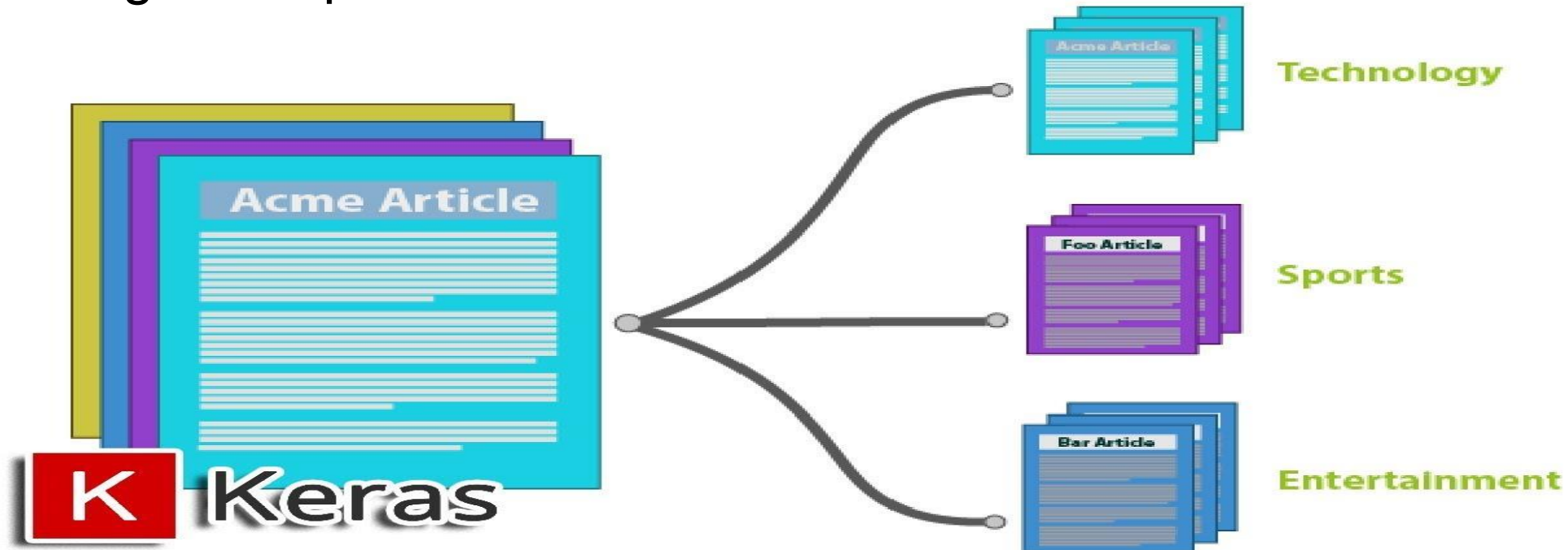
Phân loại phương tiện giao thông

- Ứng dụng phân loại phương tiện giao thông trong hệ thống camera thông minh
- Thu thập và Phân tích dữ liệu
- Xây dựng mô hình học máy
 - **Support Vector Machine**
- Đánh giá kết quả



Phân loại tin tức

- Ứng dụng phân loại tin tức trong các hệ thống xử lý tin tức thông minh
- Thu thập và Phân tích dữ liệu
- Xây dựng mô hình học máy
 - **Naïve Bayesian.**
- Đánh giá kết quả



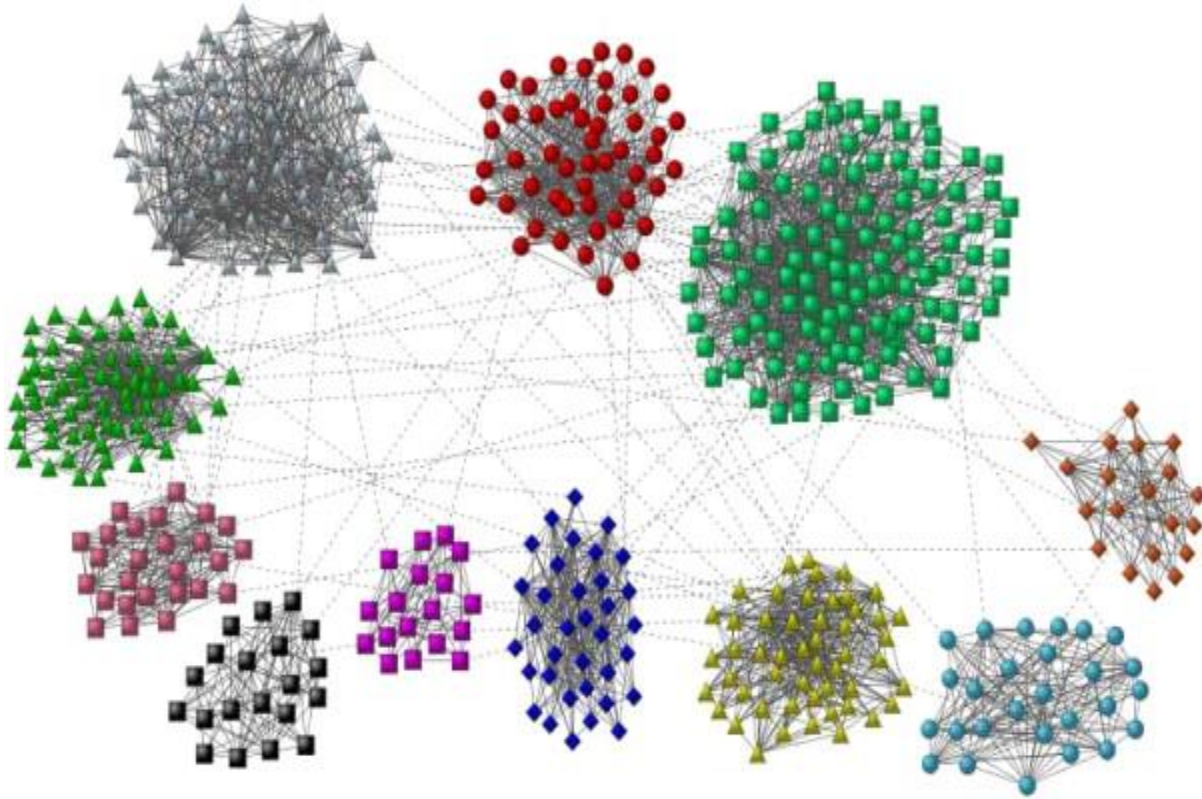
Bài toán phân vùng ảnh

- Ứng dụng trong phân vùng (segmentation) các ảnh để nhận dạng.
- Thu thập và Phân tích dữ liệu
- Xây dựng mô hình học máy.
 - **K-means, Mean-shift.**
- Đánh giá kết quả.



Phân tích và khai phá mạng xã hội

- Ứng dụng trong phân tích dữ liệu mạng xã hội (phát hiện cộng đồng).
- Thu thập và Phân tích dữ liệu
- Xây dựng mô hình học máy.
 - **Spectral Clustering**
- Đánh giá kết quả.

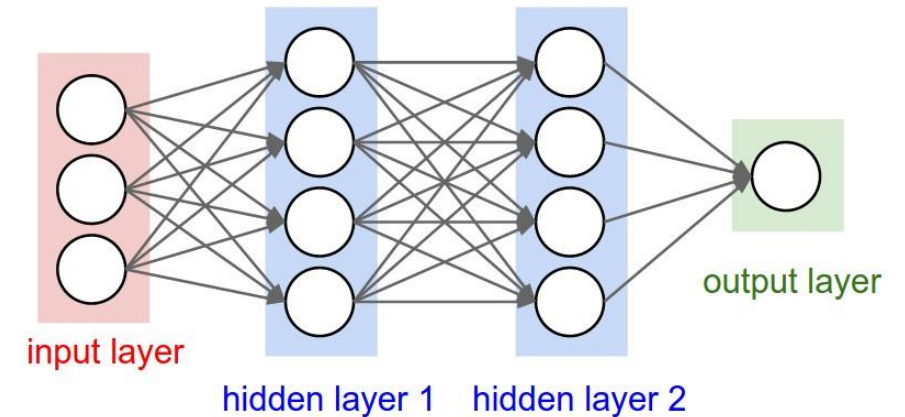


Community Detection in
Social Media



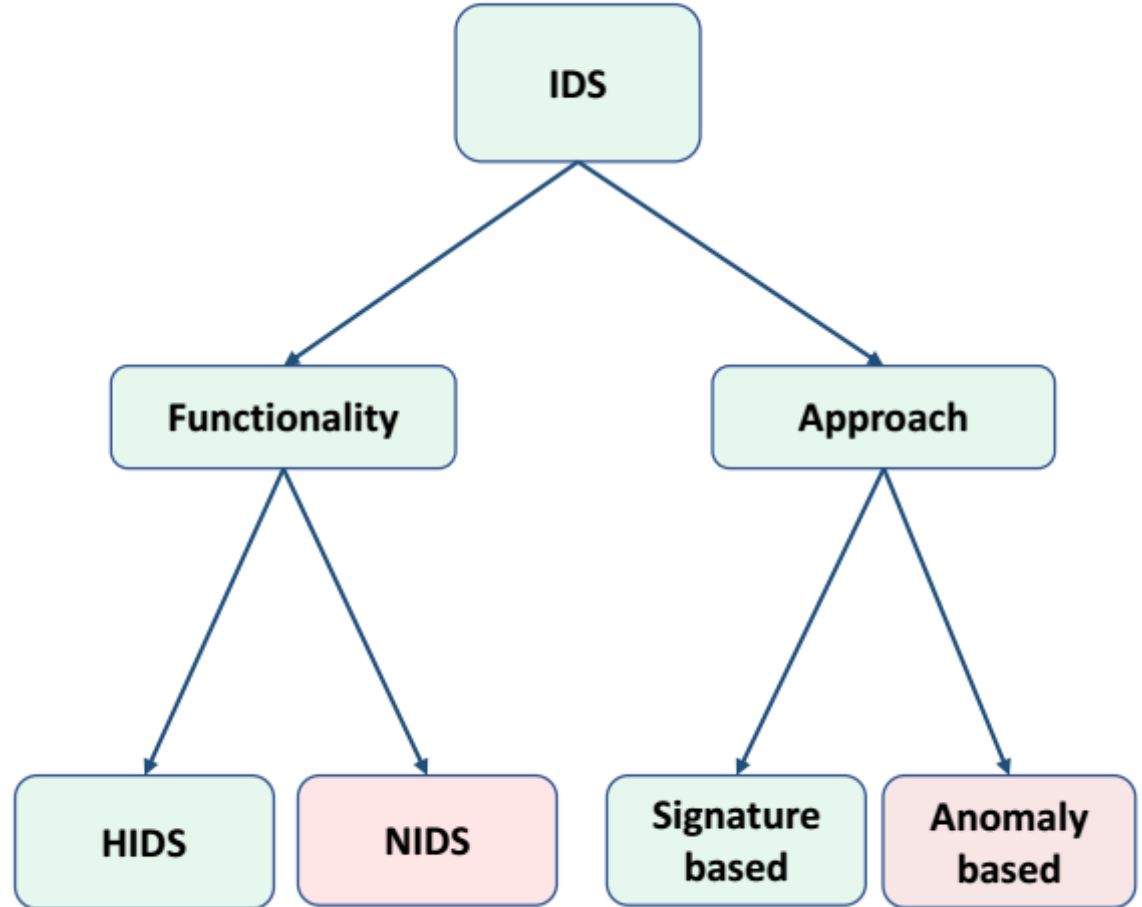
Nhận dạng chữ viết tay

- MNIST gồm 60,000 ảnh cho huấn luyện và 10,000 ảnh cho kiểm thử mô hình.
- Mọi ảnh chữ số viết tay trong tập dữ liệu được chuẩn hóa về kích thước, cụ thể là (28x28)
- Xây dựng mạng Neural fully connected với 2 lớp ẩn sử dụng Tensorflow



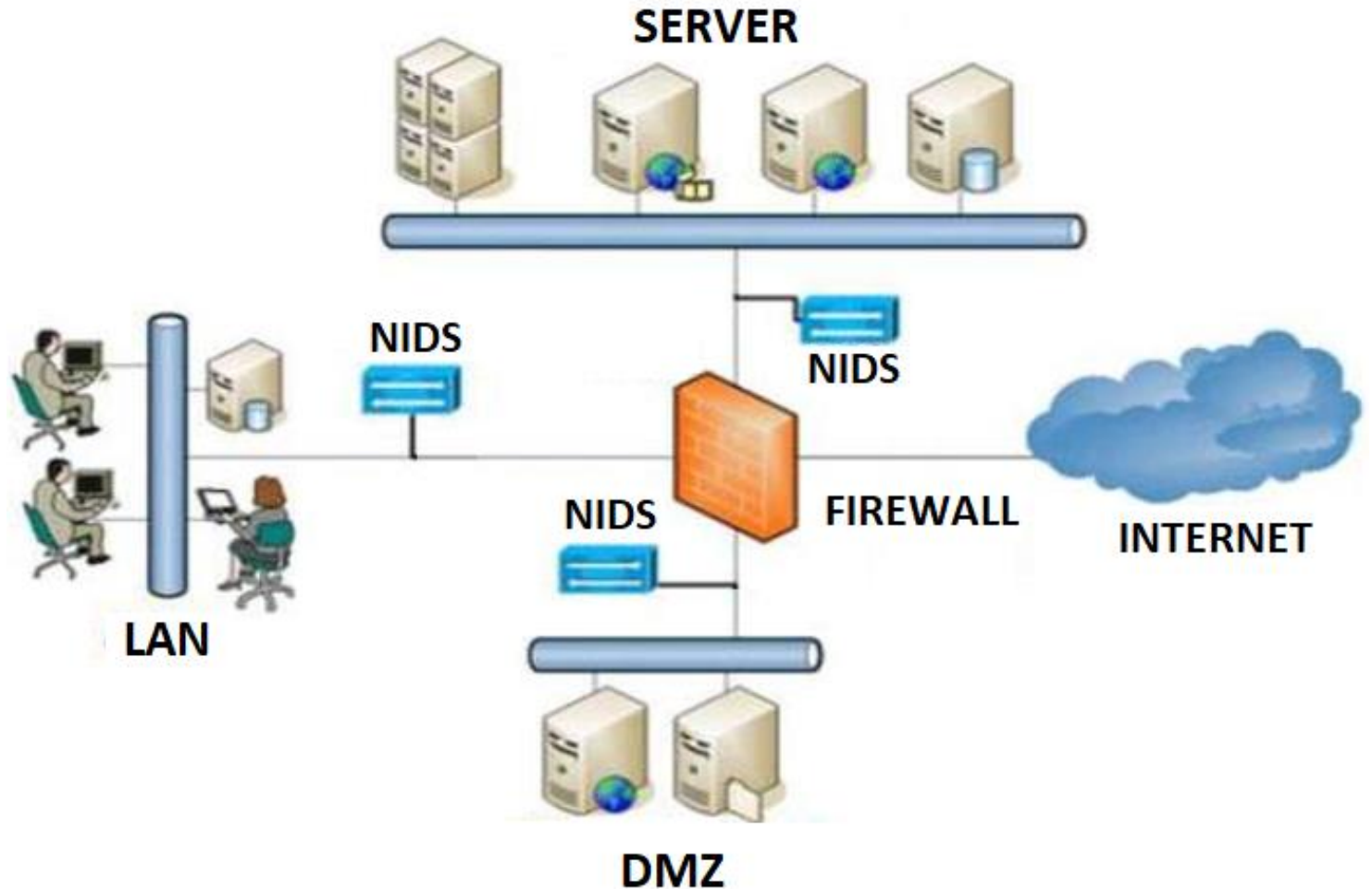
Phát hiện xâm nhập mạng

- Phát hiện xâm nhập (Intrusion Detection System - IDS): hệ thống tự động phát hiện xâm nhập mạng, hệ thống máy tính dựa trên phân tích dữ liệu.
 - **Net IDS (NIDS):** phát hiện xâm nhập mạng: tấn công mạng, truy cập trái phép... từ luồng dữ liệu mạng.
 - **Host IDS (HIDS):** phát hiện xâm nhập hệ thống: viruses, mã độc, những truy cập trái phép,..., từ log systems hoặc các lời gọi API.
- Phương pháp tiếp cận:
 - Signature based approach
 - Anomaly based approach
 - Sử dụng Auto-Encoders



Phát hiện xâm nhập mạng

- **Ví dụ:** hình bên mô tả 3 hệ thống phát hiện xâm nhập giám sát ba vùng mạng: DMZ, LAN và SERVER.
- Các **NIDS** hoạt động như một firewall mềm, nằm sau hoặc trước firewall.
- Autoencoder được dùng như một classifier
 - Mô hình phát hiện xâm nhập dùng AE:
 - Phase 1: huấn luyện mô hình
 - Phase 2: phát hiện xâm nhập

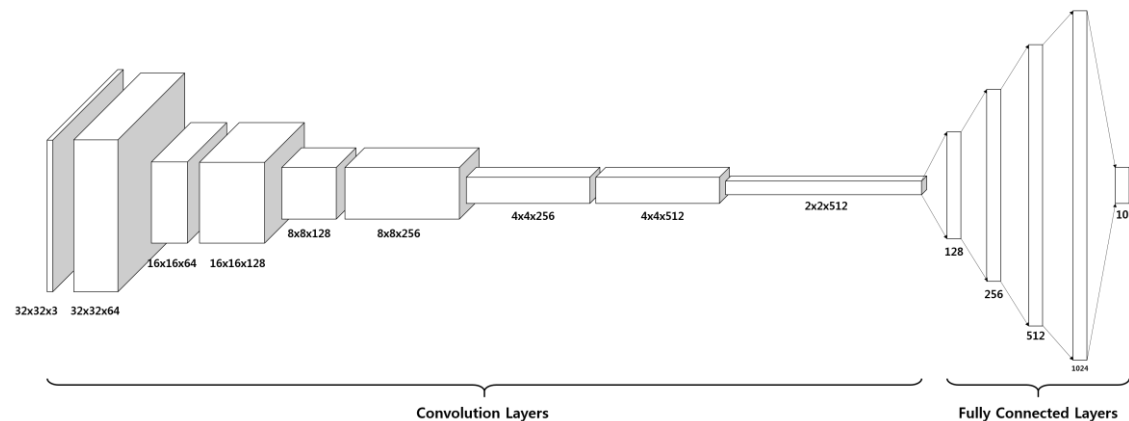
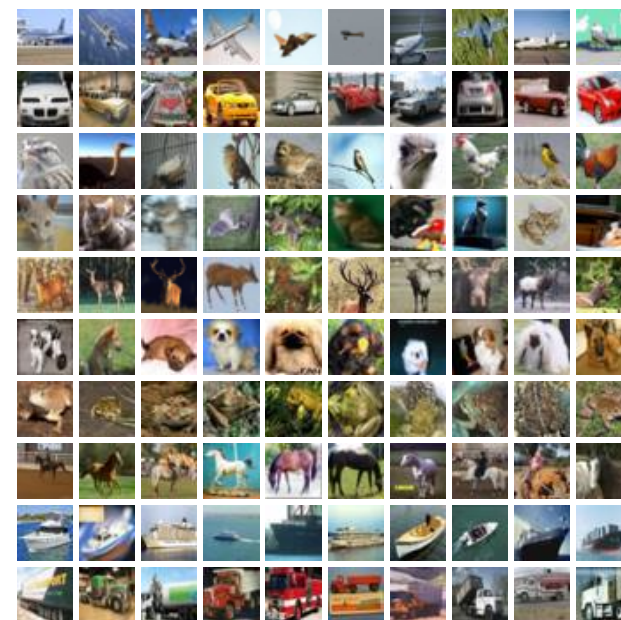




Phân loại đối tượng trong ảnh

- Dữ liệu trong chương trình là bộ dữ liệu chuẩn CIFAR-10
 - 60000 ảnh kích thước 32x32
 - 10 lớp đối tượng
 - Được lựa chọn từ 80 triệu ảnh
- Xây dựng mô hình deep learning (Convolutional Neural Networks) cho phân loại đối tượng.

airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck



HỌC MÁY CƠ BẢN

AI Academy Vietnam

CẢM ƠN!