SAMSUNG

Samsung Innovation Campus

Artificial Intelligence Course

Together for Tomorrow!
Enabling People
Education for Future Generations

Chapter 1.

Introduction to Artificial Intelligence

Artificial Intelligence Course

Chapter Description

Chapter objectives

- ✓ Learn about the concepts of Artificial Intelligence and Machine Learning as well as types, procedures, and limitations of Machine Learning-Based data analyses.
- ✓ Be able to analyze the internal and external environment of Artificial Intelligence-based corporations and be able to analyze the ways of utility as well as the commercialize business models in order to expand the scope of Artificial Intelligence service application in Artificial Intelligence-based industries.
- ✓ Be able to analyze trends in Artificial Intelligence-related technologies and changes in market demand.
- ✓ Learn about the entire course roadmap of Artificial Intelligence.

Chapter contents

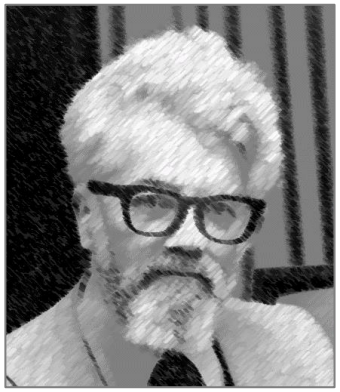
- ✓ Unit 1. The Concept of Artificial Intelligence
- ✓ Unit 2. Applications of Artificial Intelligence
- ✓ Unit 3. Techniques in Artificial Intelligence
- ✓ Unit 4. Artificial Intelligence: Trends and Markets
- ✓ Unit 5. Course Roadmap

Unit 1.

The Concept of Artificial Intelligence

What is Artificial Intelligence

Definition



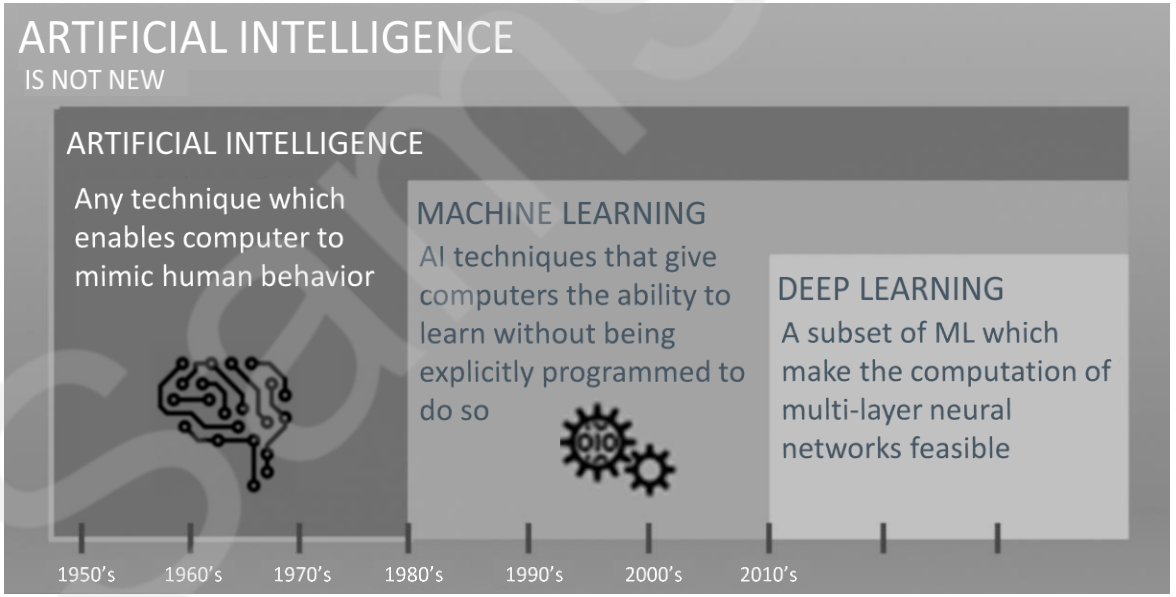
John McCarthy(1927–2011)

- Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart". They work on their own without being encoded with commands. John McCarthy came up with the name "artificial intelligence" in 1955.
- These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions) and self-correction.

https://www.wired.com/2011/10/john-mccarthy-father-of-ai-and-lisp-dies-at-84/

Building an Intelligent System that Transforms Data into Knowledge

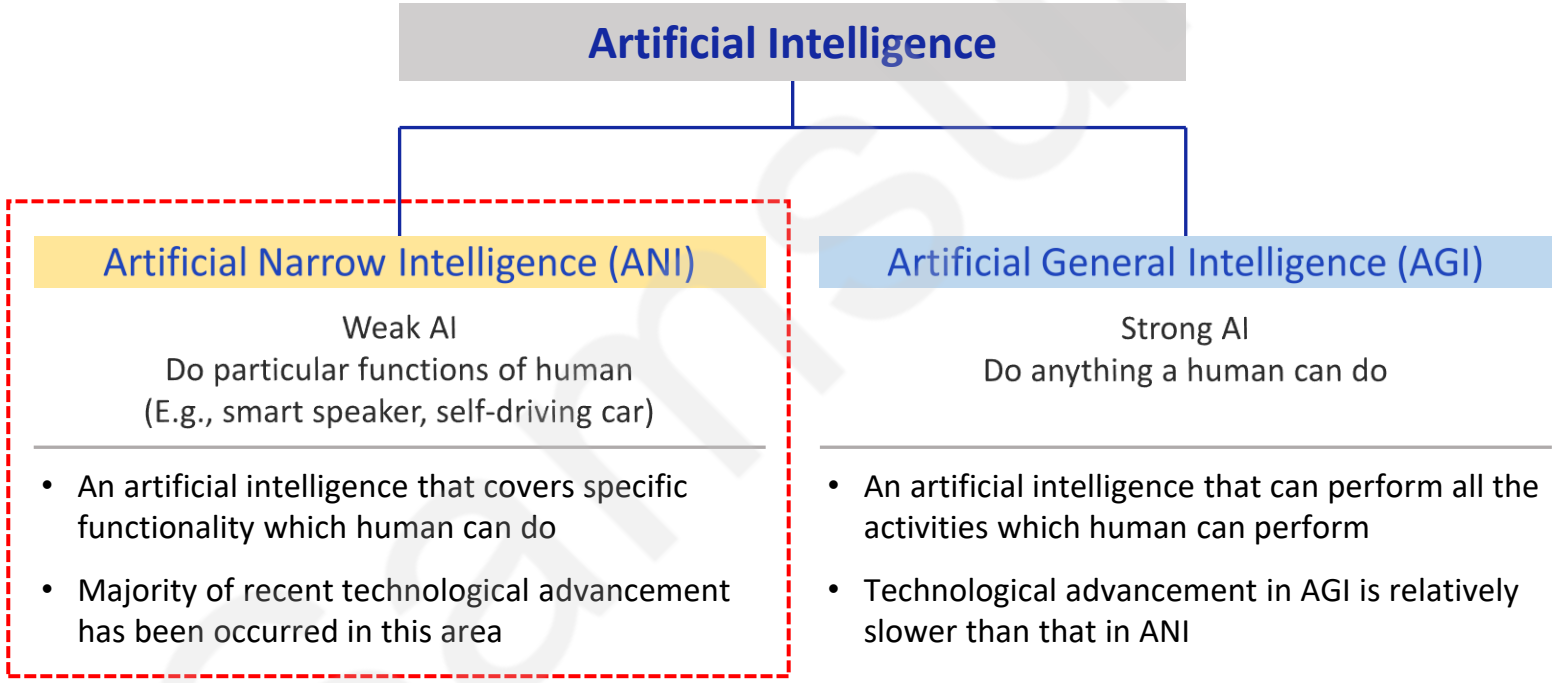
- There is an abundance of structured and unstructured data in the modern tech era.
- Machine learning emerges in the late 20th century as a sub-field of Artificial Intelligence (AI) related to self-learning algorithms that extract and predict knowledge from data.
- Humans manually analyze large amounts of data to induce rules and make models.
- Machine learning can gradually improve predictive models and data-based decision-making performance by extracting knowledge more efficiently from data.



Unit 1.

The Concept of Artificial Intelligence

Type of Artificial Intelligence



Main focus on this lecture and the entire course

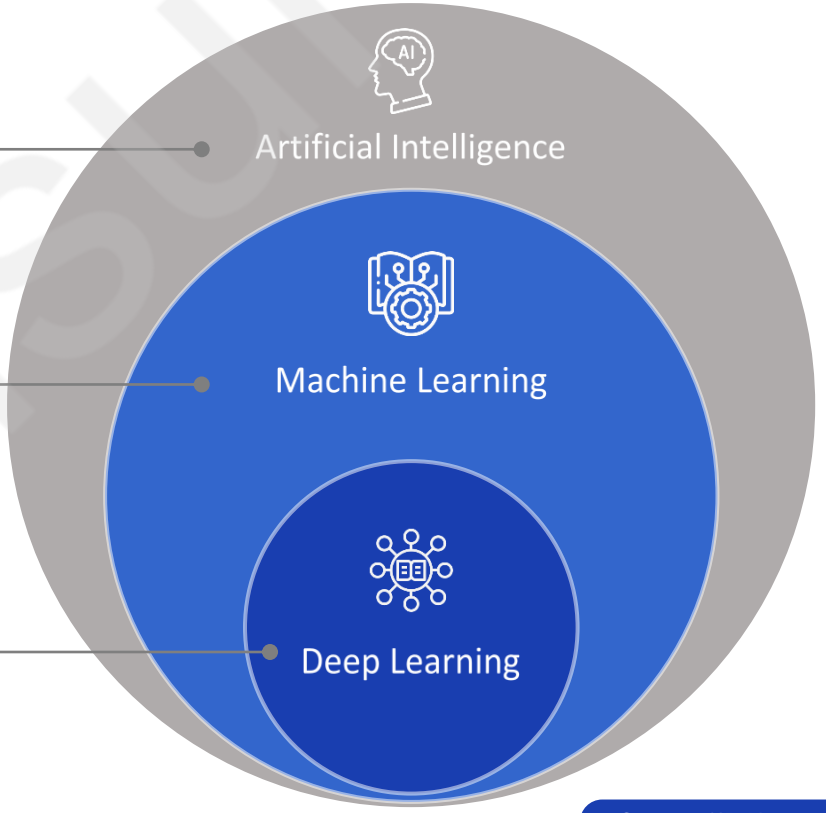
What are subset of AI?

Subset of Artificial Intelligence

Artificial Intelligence
Any technique which enables computers to mimic human behavior.

Machine Learning
Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

Deep Learning
Subset of ML which make the computation of multi-layer neural networks feasible.



Source: KD Nuggets

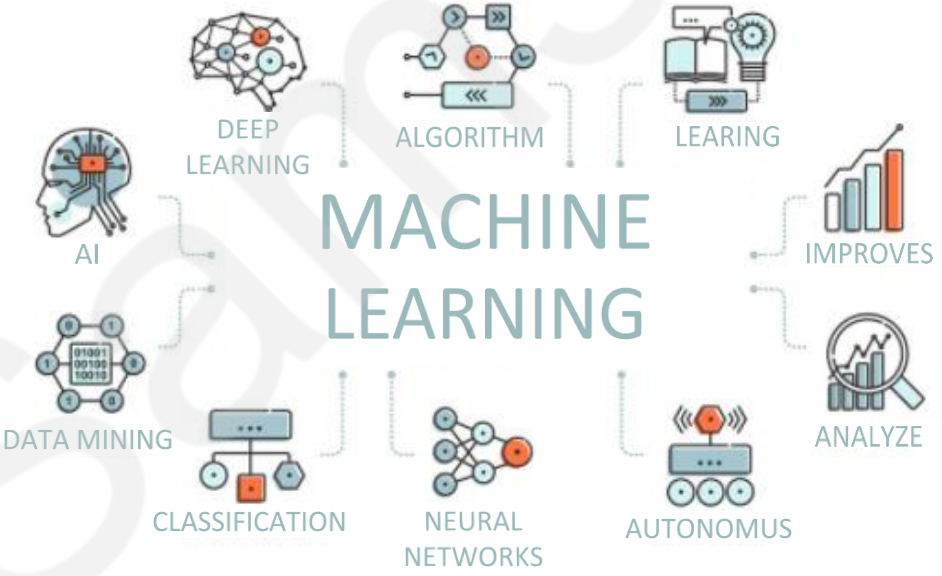
Unit 1.

The Concept of Artificial Intelligence

Definition of Machine Learning

Machine Learning is regarded as a field of artificial intelligence and is the study of computer algorithms that automatically improve through examples and experiences.

- "A field of research that develops algorithms that allow machines to learn from data and execute actions that are not explicitly specified by code" - Arthur Samuel, 1959
- "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improved with experience E." - Tom Mitchell, 1977



- “A field of research that develops algorithms that allow machines to learn from data and execute actions that are not explicitly specified by code“ - Arthur Samuel, 1959
- “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improved with experience E.“ - Tom Mitchell, 1977

- The task T is to classify dogs and cats, and the performance P represents a measure of classifying dogs and cats. E can be said to be “learning” if the performance of classifying dogs and cats gradually improves through experience, or data (10,000 photos).

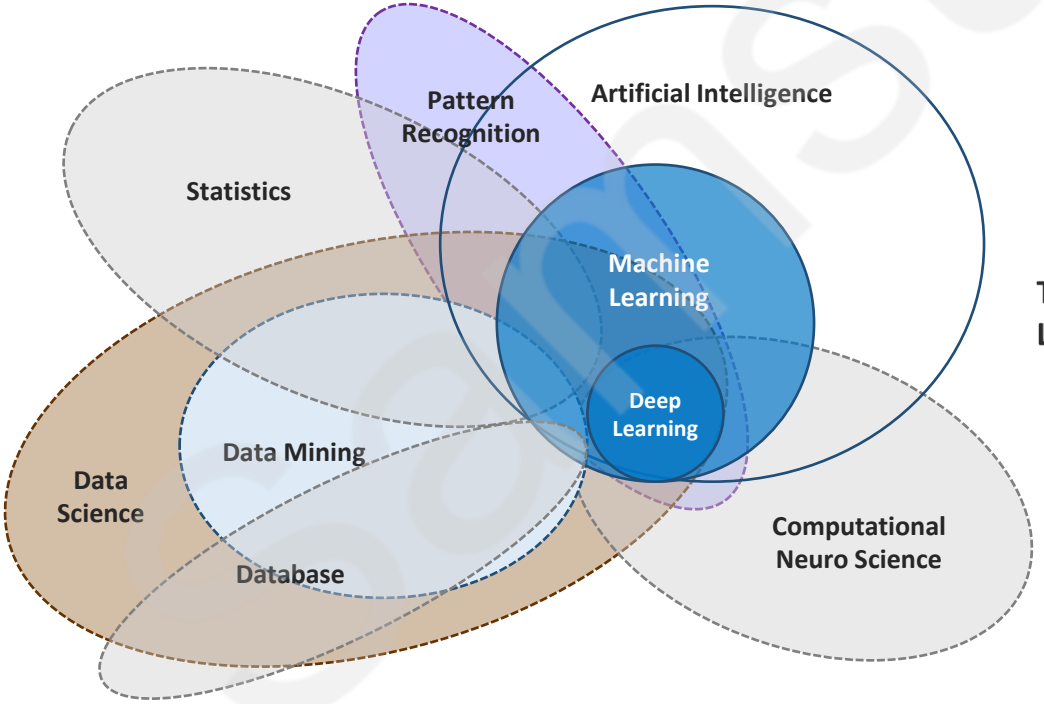
Unit 1.

The Concept of Artificial Intelligence

Disciplines Related to Machine Learning

Studies Related to Machine Learning

- Machine learning is an interdisciplinary field that combines academic backgrounds and achievements in various fields such as probability and statistics, computer science, database theory, cognitive science, neuroscience, and patter recognition, rather than limited to technology of methodology in any one field.



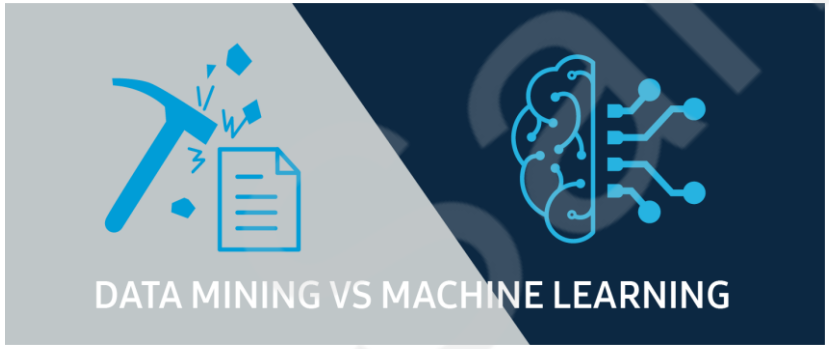
The Relationship Between Machine Learning and Various Disciplines

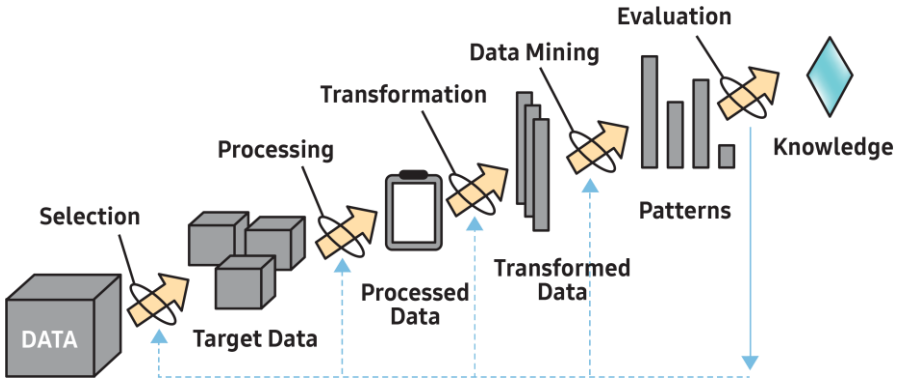
Machine Learning vs. Statistics

- **Statistics** are traditionally regarded as the theoretical foundation that provides a scientific and systematic method for converting data into information. The areas particularly emphasized in statistics are inference and verification, and various methodologies and theories have been established to **explain how well the given data conform to the researcher's hypothesis and theory** (or how accurately the values derived from observations estimate the actual population parameters).
- Meanwhile, **machine learning** is mainly used to **solve tasks that are difficult to design or difficult to program explicit algorithms**. Most machine learning algorithms are first used to **quantify complex relationships by identifying the feature of potential mechanisms generated by data**, and then to **make predictions on new data using this identified pattern**.
- At first glance, the approaches of statistics and machine learning may look opposite to each other, but besides the emphases on some differences in aspects or perspectives, the methodologies that form the basis of each discipline is very similar. In fact, it can be said that **many methodologies of machine learning are based on statistical learning based on statistics**.

Machine Learning vs. Data Mining

- What is Data Mining?:
  - It refers to a mining operation of digging useful information [gold nuggets] from a large data warehouse [stone pile].
  - It is a series of processes that help companies make decisions to secure competitiveness by finding and modeling relationships, patterns, and rules between data existing in large amounts of data.
- Machine learning is also deeply related to data mining in that it extracts useful rules, knowledge expression, or judgement criteria from data.
- If **data mining is a process of systematically and automatically discovering meaningful rules or patterns** within large-scale stored data and intellectualizing them, **machine learning** differs in that it is a process in which **computer programs learn, make predictions, and research and build algorithms for said process**.





Statistical Analysis vs. Data Mining

- Traditional Statistical Analysis
  - There is a target group, and various assumptions such as the distribution or model of the population are premised; and analysis is conducted under this prerequisite.

    → The process of inferring the entire parameter of a population through observation of a sample.
- Data Mining
  - No prerequisites are required for the distribution or model inevitably involved in the sample survey/experiment.

    → The process of extracting necessary information/knowledge using the entire data of the population.
  - There is a prerequisite that it needs to be large amounts of data.

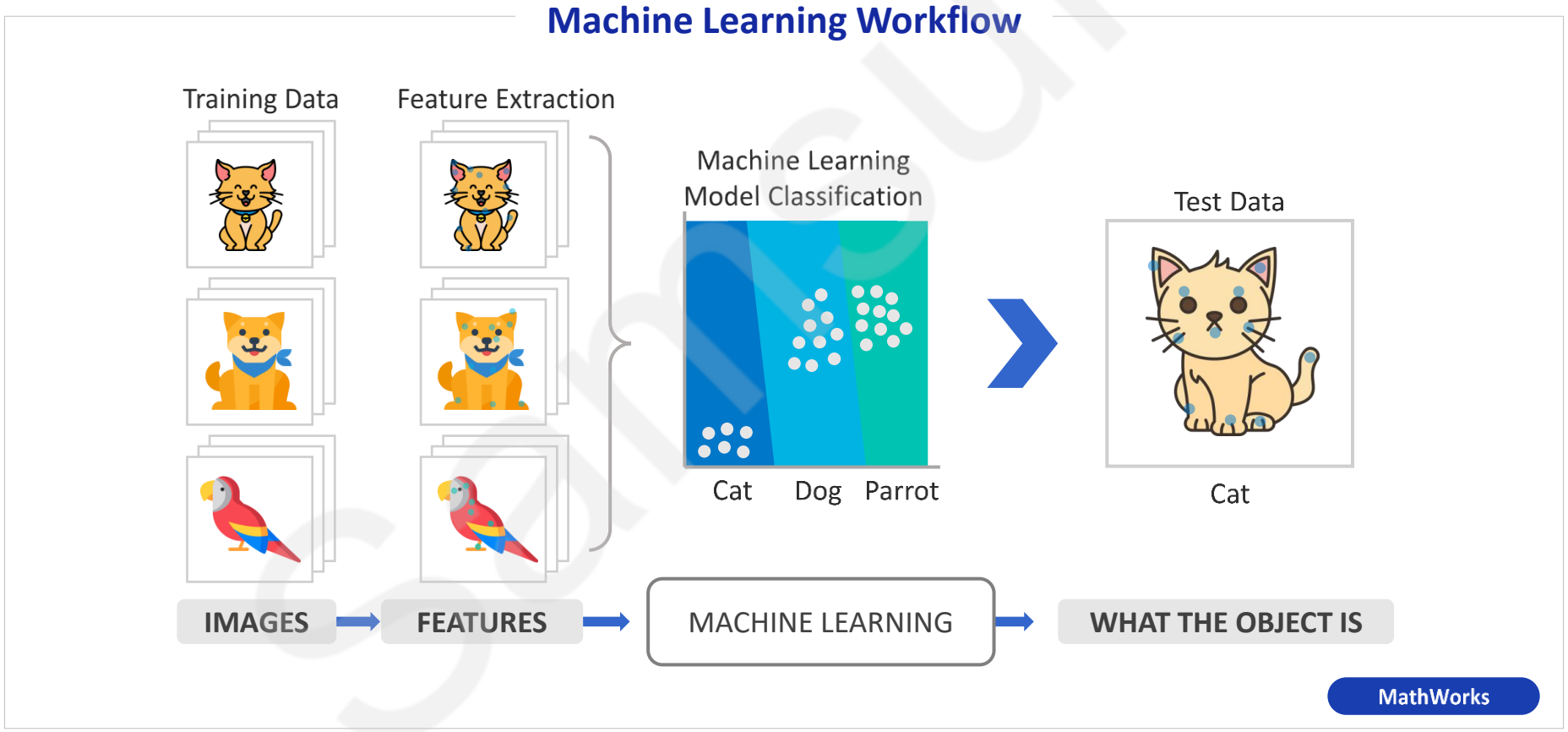
Unit 1.

The Concept of Artificial Intelligence

- 1.1. Definition of Artificial Intelligence
- 1.2. Types and Subsets of Artificial Intelligence
- 1.3. Definition of Machine Learning
- 1.4. Disciplines Related to Machine Learning
- 1.5. Types and Choices of Machine Learning-Based Data Analysis
- 1.6. Procedures for Machine Learning-Based Data Analysis
- 1.7. Reasons For Machine Learning
- 1.8. Limitations of Machine Learning

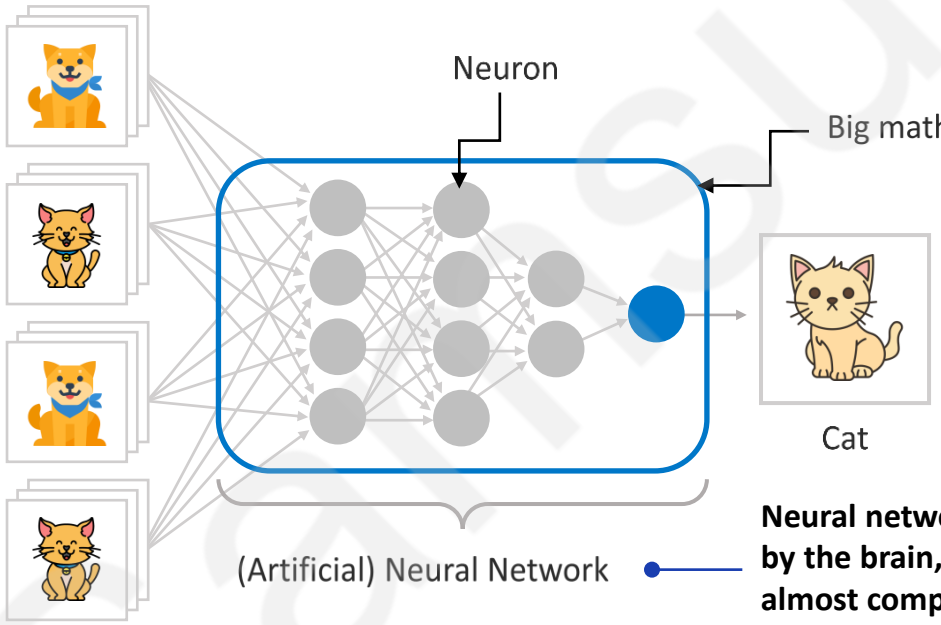
Types and Choices of Machine Learning-Based Data Analysis

Machine Learning vs. Deep Learning



Machine Learning vs. Deep Learning

Deep Learning Workflow



Neural networks were originally inspired by the brain, but the details of how they work are almost completely unrelated to how biological brains work.

MathWorks, Coursera

Types of Machine Learning-Based Data Analysis

- When analyzing large amounts of data by applying machine learning technology, patterns that were not visible can be found– this is called data mining. Data mining deals with the following four problems.

Prediction

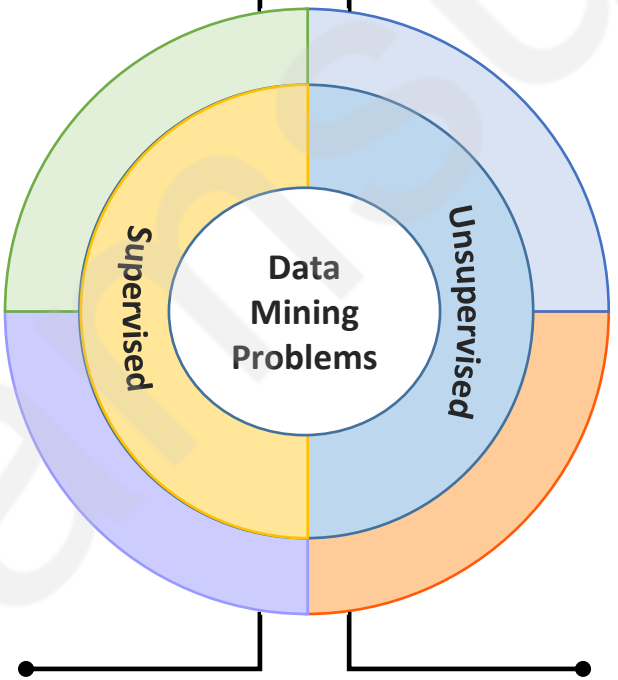
Making a model based on the given data, then using the model for new cases in order to make predictions

Ex Predicting the quality specifications as a result of ingredients and environment (pressure, temperature, humidity etc.) on the fishing industry.

Classification

Determining where a particular case belongs in a given series of classified categories

Ex Quality Rating from good/normal/bad. Determining the quality of new products.



Clustering

Comparing the properties of the data and forming clusters based on similar characteristics

Ex Clusterization of processes with similar characteristics out of various processes.

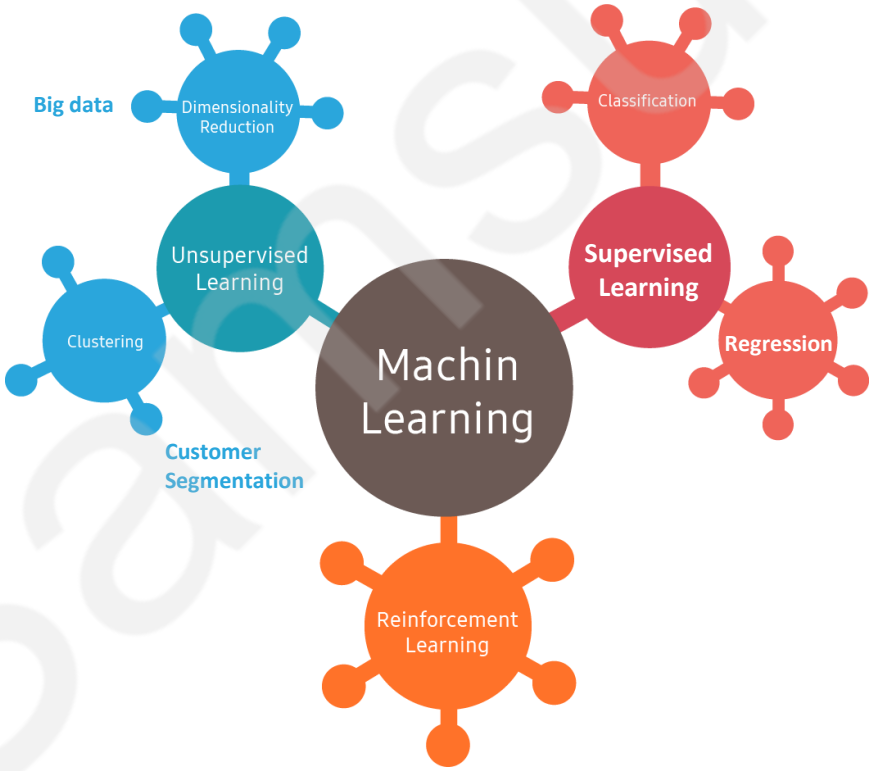
Association Rule

Identifying the attributes or relationships between items that one pattern of appearance implies the appearance of another pattern

Ex Predicting what will happen to the entire process when there is an abnormal pattern in one process.

Types of Machine Learning-Based Data Analysis

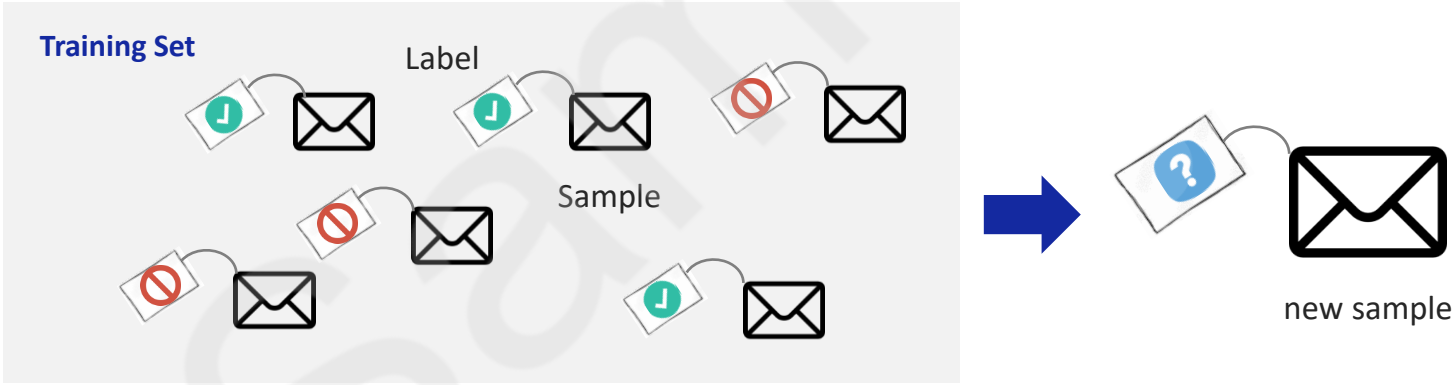
- There may be various criteria for classifying machine learning-based data analysis techniques from various perspectives, but in general, they are classified into supervised learning and self-learning(or unsupervised learning), depending on the existence of objective variables (or response variables, output target values, etc.)



Supervised Learning

- Supervised learning focuses on expressing the relationship between explanatory variables (expressed as independent variables, features, etc.) and objective variables (expressed as response variables, dependent variables, target variables, output values) as well as predicting future observations. It is mainly suitable for solving problems such as recognition, classification, diagnosis, and prediction.
- Main techniques in supervised learning can be reclassified into classification and numerical prediction methods depending on the objective variable (or response variable, dependent variable) and whether its explanation or prediction is a numerical (quantitative variable) or categorical (qualitative variable).

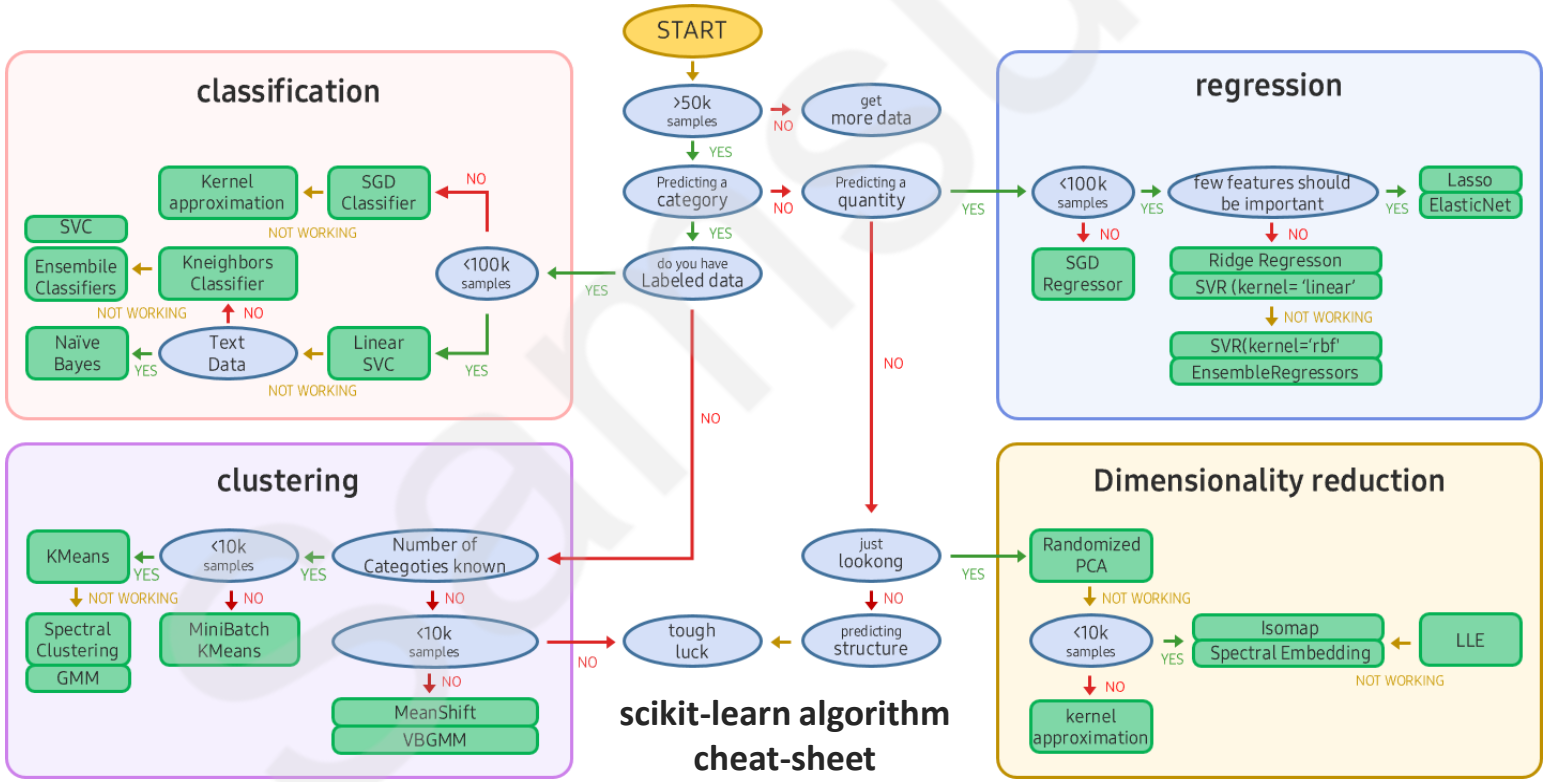
Labeled Training Set for Spam Classification (Example of Supervised Learning)



- The training data injected into the algorithm includes a desired answer called the label.

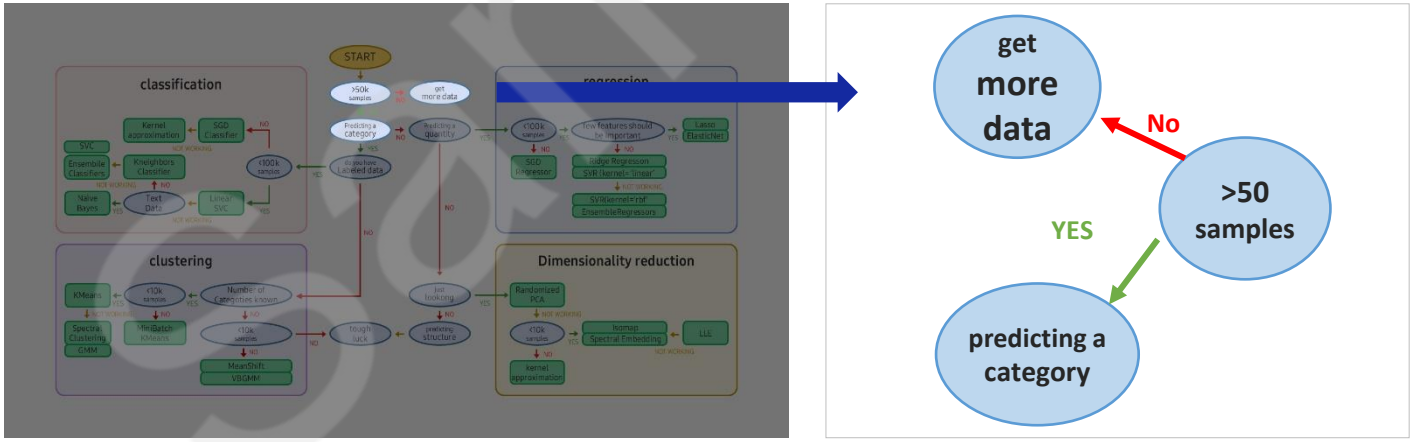
Supervised Learning

- Data mining is a combination of statistics and machine learning (or artificial intelligence). The problem to the solved by machine learning is in line with the problem of data mining.
- The figure below shows a guide to solving problems with Scikit-learn, a Python machine learning library.



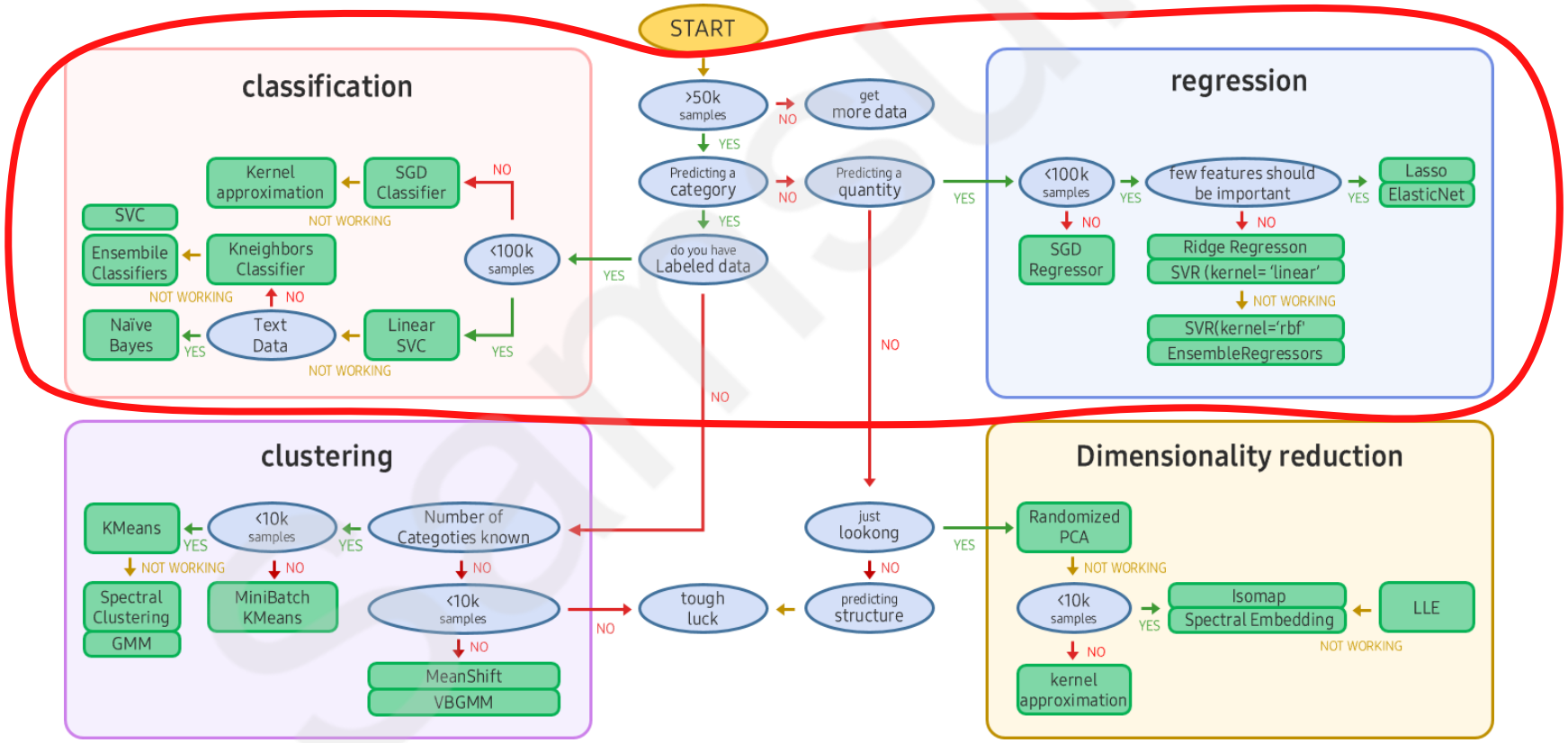
In the Case of Insufficient Data

- Looking at the figure below, statistics-based analysis is inevitable when there is little data.
- Statistics show that approximately 30 samples are understood to be minimal to estimate the characteristics (parameter) of the population.
- As the number of samples increases, they become closer to the parameters. Some assumptions can be made based on approximately 30 data, which is the minimum level.
- The assumption here is that if there are more than 30 samples through the law of algebra, normal distribution is assumed and followed.
- Large amounts of data are essential in data mining or machine learning, and the criteria are considered to be approximately 100,000 or more.



Supervised Learning

- In the figure below, it can be seen that if there is an answer among the predicted questions, and the answer is a numerical type, it is a regression. If the answer is a categorical type, it is a classification.

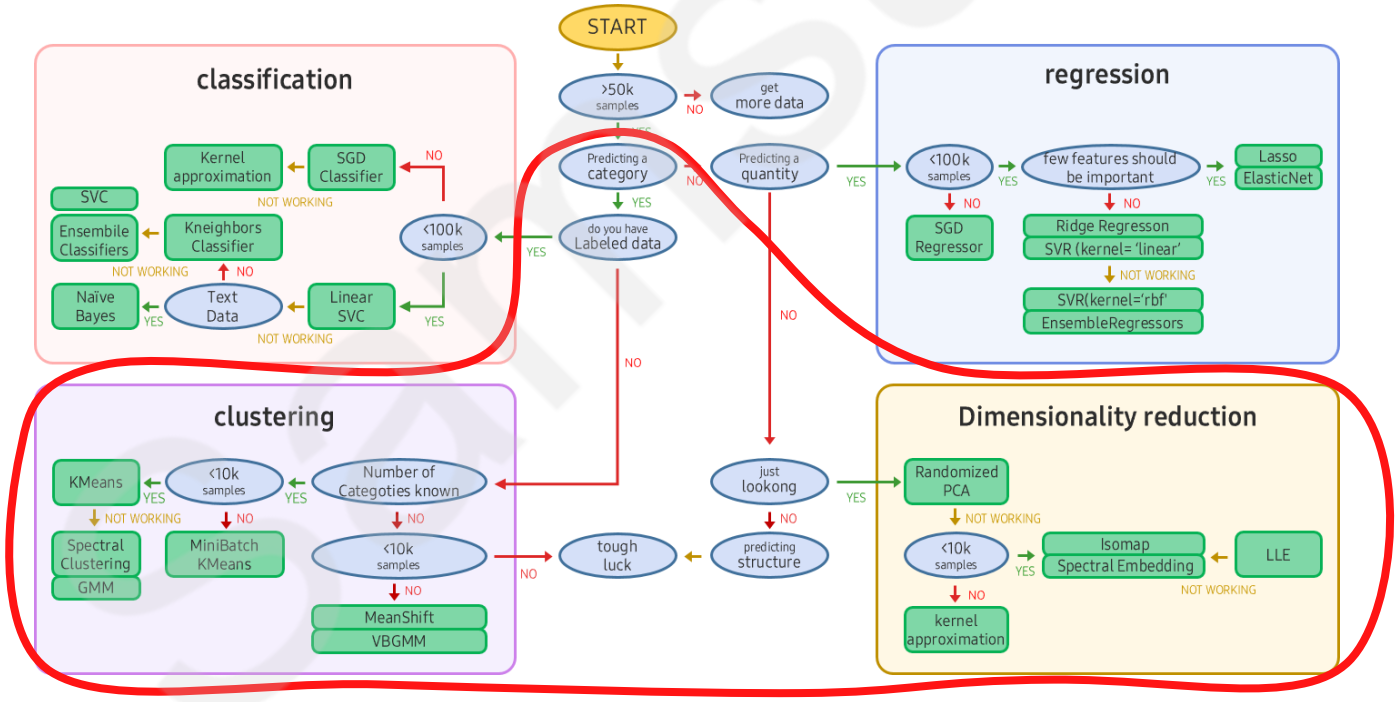


Main Methods and Algorithms of Supervised Learning

Type	Regression (Numerical Prediction)
K-Nearest Neighbors	Linear Regression
Logistic Regression	Extended Regression Analysis (ex: Polynomial Regression, Nonlinear Regression, Penalized Regression, etc.)
Artificial Neural Network	Artificial Neural Network
Decision Tree	Decision Tree
Support Vector Machine	Support Vector Machine (Regression)
Naïve Bayes	PLS (Partial Least Squares)
Ensemble Method (Random Forest, etc.)	Ensemble Method (Random Forest, etc.)

Self-Learning (or Unsupervised Learning)

- Self-learning or unsupervised learning refers to a form in which learning is performed without information on objective variables (or response variables, dependent variables, target variables, and output values), and is mainly used for problems such as description, characteristic derivation, and pattern derivation.
- In general, compared to supervised learning, which has clear and distinct predictive purposes, self-learning techniques have a stronger nature of data mining to search for useful information or patterns without prior information.



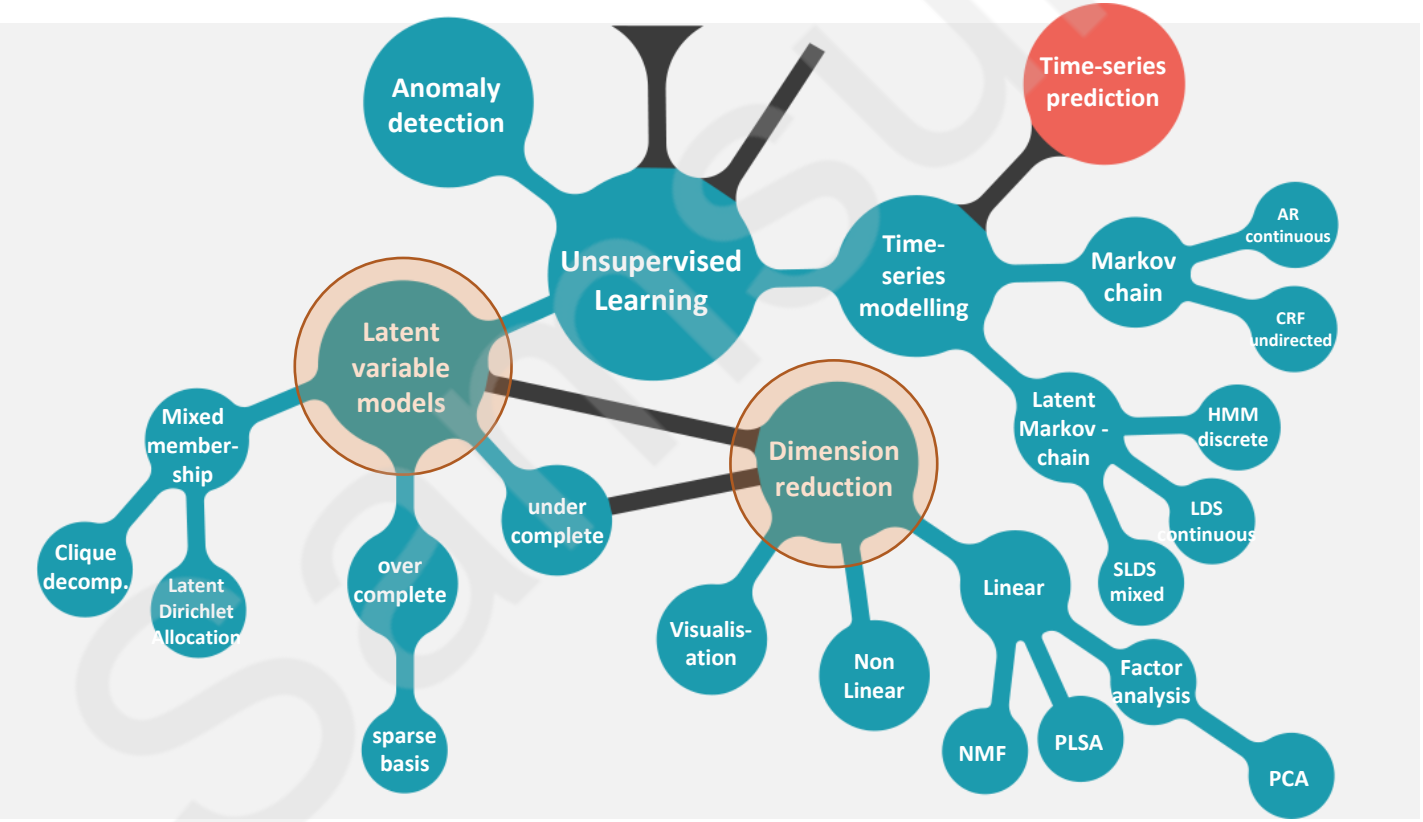
Unsupervised Learning

- Unsupervised Learning: No label in training data, so the system must learn without any help.
- Important Unsupervised Learning Algorithms

Cluster	Visualization and Dimension Reduction	Association Rule Learning
• K-Means • DBSCAN • Hierarchical Clustering Analysis (HCA) • Anomaly Detection and Outlier Detection • One-Class SVM • Isolation Forest	• Principal Component Analysis(PCA) • Kernel PCA • Local Linear Embedding • t-SNE	• Apriori • Eclat

Self-Learning (or Unsupervised Learning)

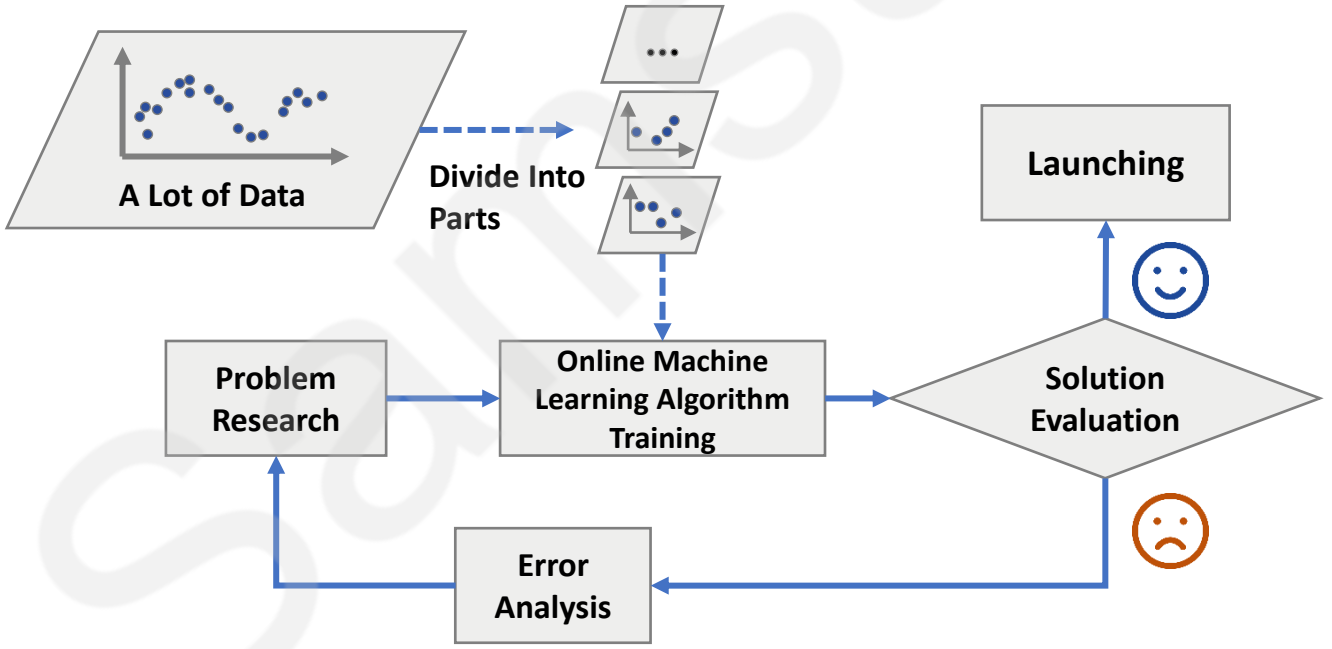
- Main techniques of self-learning, or unsupervised learning, include clustering, dimensional reduction, correlation analysis, and self-learning artificial neural network (SOM etc.), and their utilization is very diverse.



Batch Learning and Online Learning

- **Batch Learning:** The system cannot learn gradually.
- **Online Learning:** The system is trained by sequentially injecting data one by one or in small batches called a mini-batch.

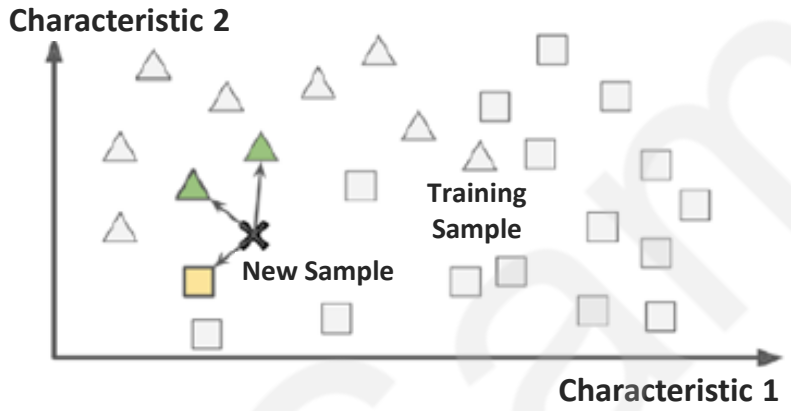
Processing Large Amounts of Data Using Online Learning



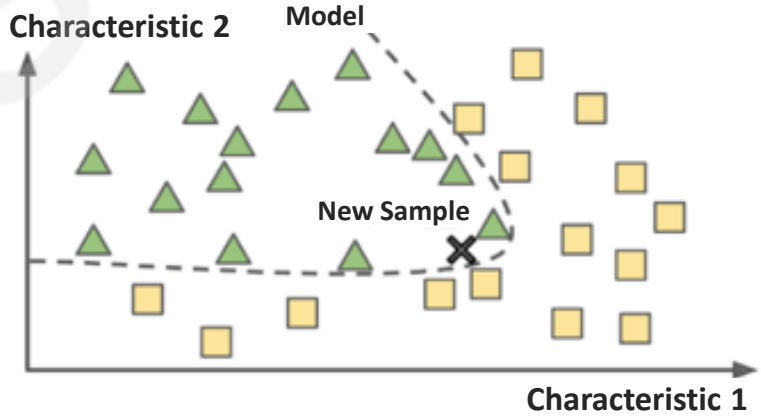
Instance Based Learning Vs. Model Based Learning

- **Instance Based Learning:** The systems learns by remembering training samples. It is generalized by comparing new data and learned samples using similarity measurements.
- **Model Based Learning:** The system makes a model from a sample and uses it for prediction.

Instance Based Learning



Model Based Learning



Unit 1.

The Concept of Artificial Intelligence

Machine Learning-Based Data Analysis: Plans and Procedures

When performing machine learning-based data analysis, the following procedure is generally followed.

01. Understanding the Business and Defining the Problem

02. Collecting Data

03. Data Pre-processing and Searching

04. Data for Model Training

05. Model Performance Evaluation

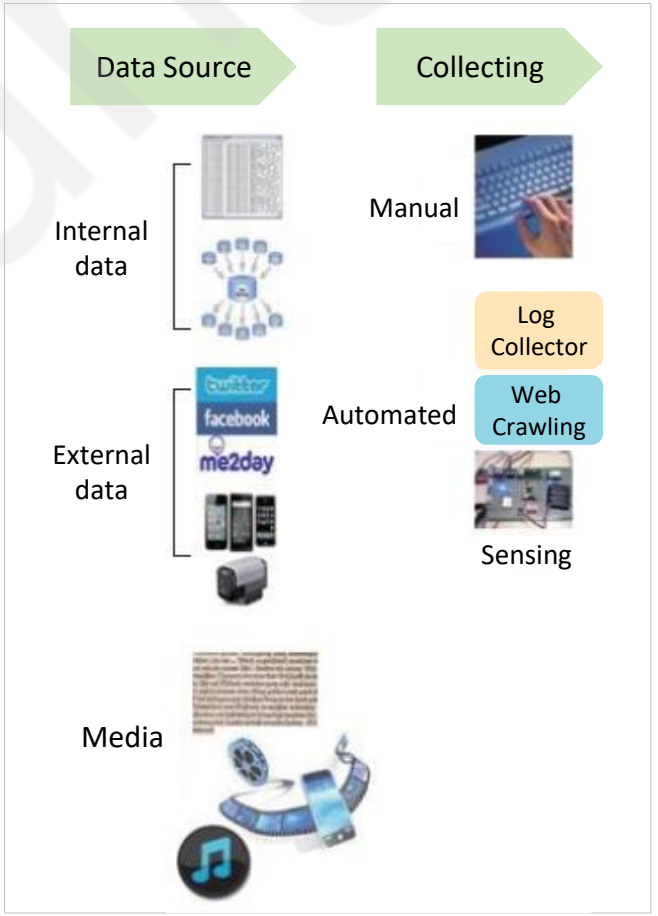
06. Improving Model Performance and Market Application

01. Understanding the Business and Defining the Problem

- The top priority in machine learning-based data analysis is to **define the problem to be solved**.
- Identify the problem by repeating the process of understanding the problem you are trying to solve, understanding the business domain, and grasping the concepts of the problem. Through repeatedly redefining the problem and finding solutions, your understanding of the specific purpose and necessary data will become clear.
- Also, in the process of defining the problem and brainstorming the required data types, **it naturally leads to a tentative decision-making process** on which machine learning techniques will be applied.

02. Collecting Data

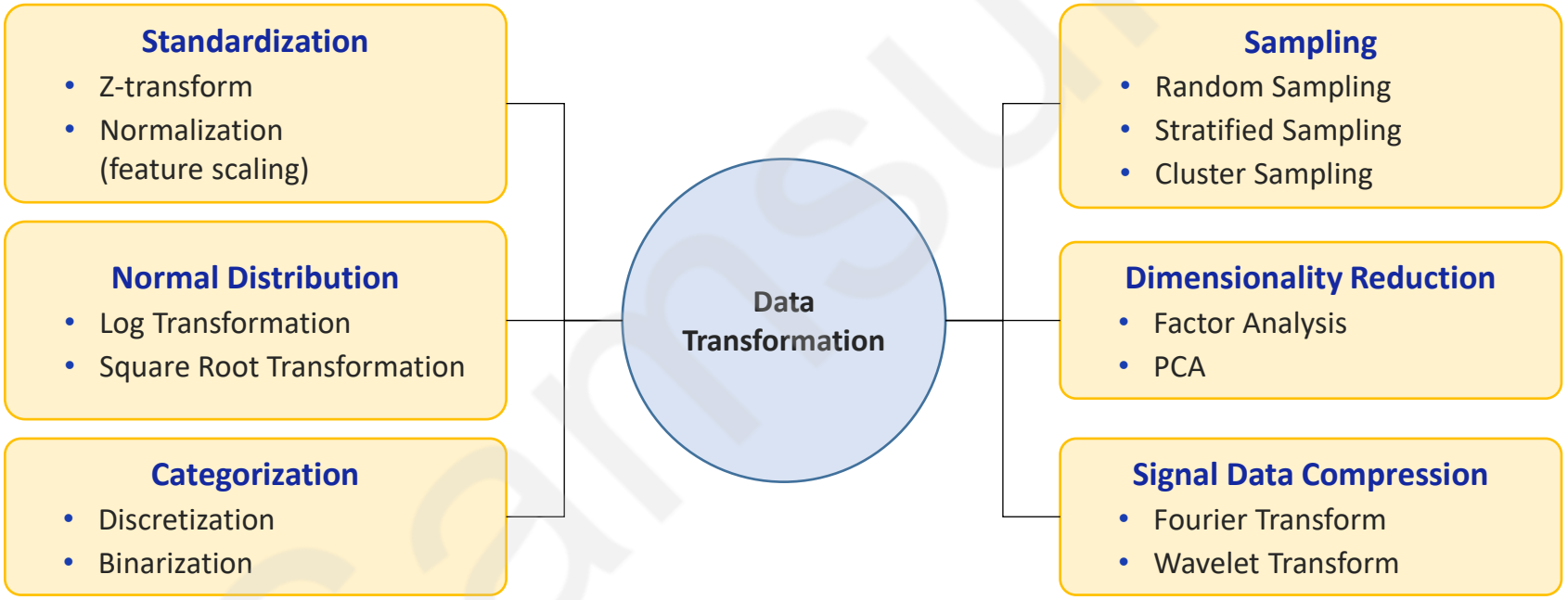
- If the issue of analysis to is clear, **necessary data for the analysis must be collected**.
- If the data is already given, you may go straight to thinking about the appropriate analysis, but even in those cases, additional data may be generated, so collecting necessary data is always a factor to be considered for proper analysis.
- It is common to extract data from an internal data store (data warehouse or data mart) through SQL or from a Hadoop- based big data platform, but in some cases, external data may be required, so it may be necessary to collect data through web scraping or API.



03. Data Pre-Processing and Searching

- If the necessary data is collected, the data is preprocessed and converted into a form suitable for applying machine learning. The quality of machine learning-based data analysis results depends on the required techniques or algorithms, but **data preprocessing, conversion, and search steps are very important because they affect the quality of the data**.
- In fact, like most analyses, this stage can be said to be the stage that takes the most time and effort among machine-learning-based data analysis processes.

03. Data Pre-Processing and Searching

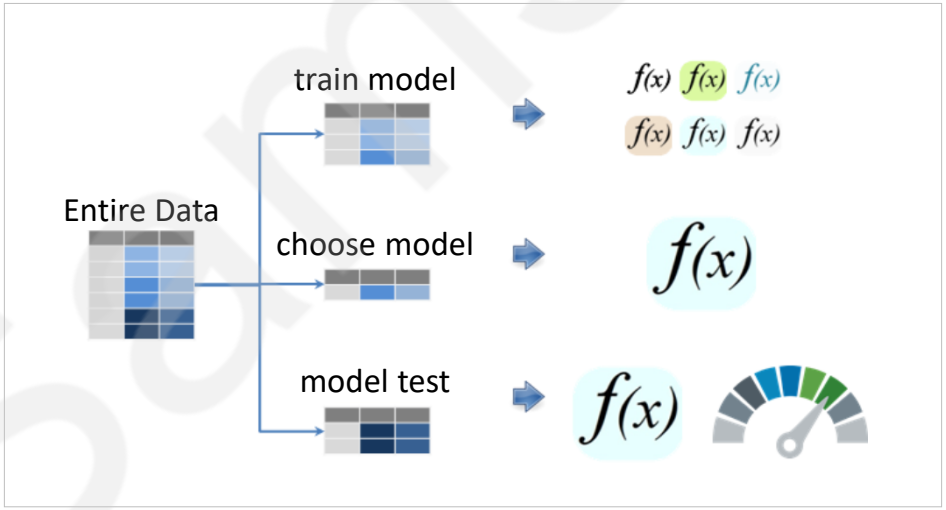


03. Data Pre-Processing and Searching

- **Normal Distribution**
  - **Log Transform:** If the input data represents an inverse function distribution, convert it to a normal distribution using log.
  - **Square Root Transform:** transforming non-normal distribution data into a normal distribution using square roots.
- **Categorization**
  - **Discretization:** Categorizing continuous variables into multiple sections.
  - **Binarization:** Transforming into dummy variables with two values of 0 and 1.
- **Sampling**
  - Simple Randomization
  - Systematic Sampling
  - Stratified Randomization
  - Cluster Sampling
  - Multistage Sampling
- **Dimensionality Reduction**
  - Factor Analysis: Finding common factors by finding potential variables
  - Principle Component Analysis

04. Data for Model Training

- This is **a step of learning data by applying the intended machine learning technique** after data preprocessing and searching.
- In the case of supervised learning, data can be divided into learning data and verification/evaluation data for model training, or model training can be conducted after designing for cross-validation.
- In the case of self-learning (or unsupervised learning), since it does not have a purpose value, it is a process of deriving patterns through analysis rather than model training.



05. Model Performance Evaluation

- In general, machine learning-based learning models tend to produce biased results in training data used for model training, so machine learning algorithms **use evaluation datasets** to evaluate the accuracy of the model.
- In the case of self-learning, it is common not to have an evaluation data set, so performance is evaluated by focusing on the interpretability of statistics or rules derived from the analysis process rather than cross-validation.

		True/Actual		
		Cat	Fish	Hen
Predicted	Cat	4	6	3
	Fish	1	2	0
	Hen	1	2	6

https://towardsdatascience.com/multi-class-metrics-made-simple-prat-i-precision-and-recall-9250280bddc2

06. Improving Model Performance and Market Application

- An issue that needs solving by a single machine learning analysis process is rarely solved at once, and **the model's performance is improved by continuously changing model parameters and estimation methods**.
  - To compare performance with the original algorithms, other algorithms are sometimes applied.
  - **Judging model performance**: The judgment on how satisfactory the model performance is can vary depending on the issue or business domain that needs analysis. In any case, there is no absolute criterion for judging that the model performance has improved sufficiently.
  - **Improving model performance**: Analysts need to compare and apply various algorithms, change parameters and estimation methods even within the same algorithm, and seek to improve model performance until they are judged to be satisfactory.
  - **Market application of the satisfied model**: After this process, if the results are derived enough to satisfy the model, it can be applied to the originally intended business issue. In some cases, additional development work for automation or system linkage may be required.

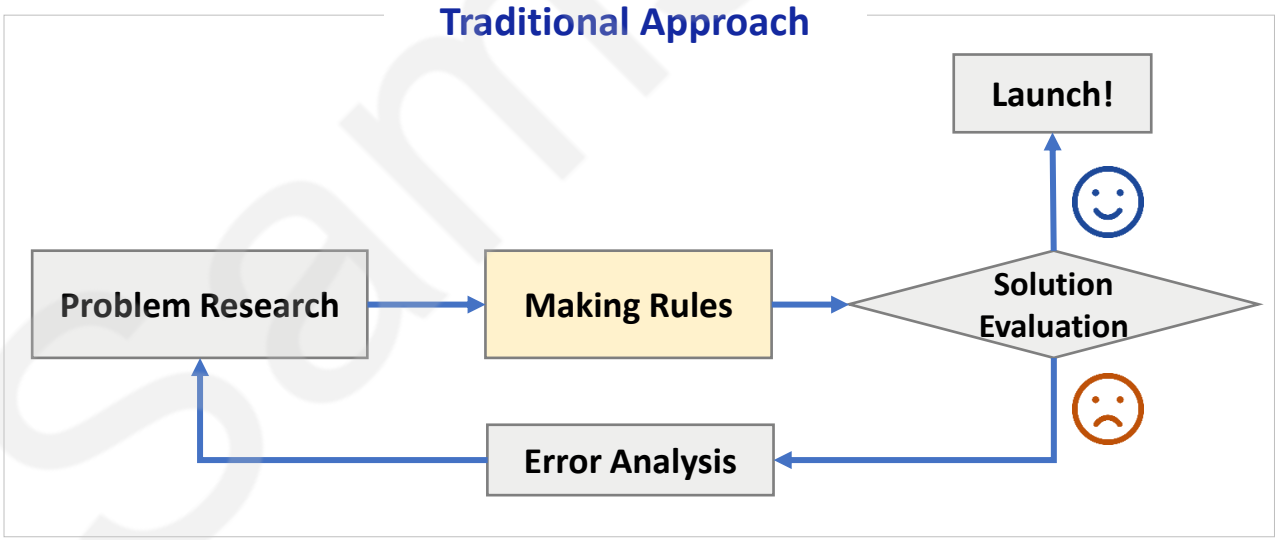
Unit 1.

The Concept of Artificial Intelligence

Reasons For Machine Learning

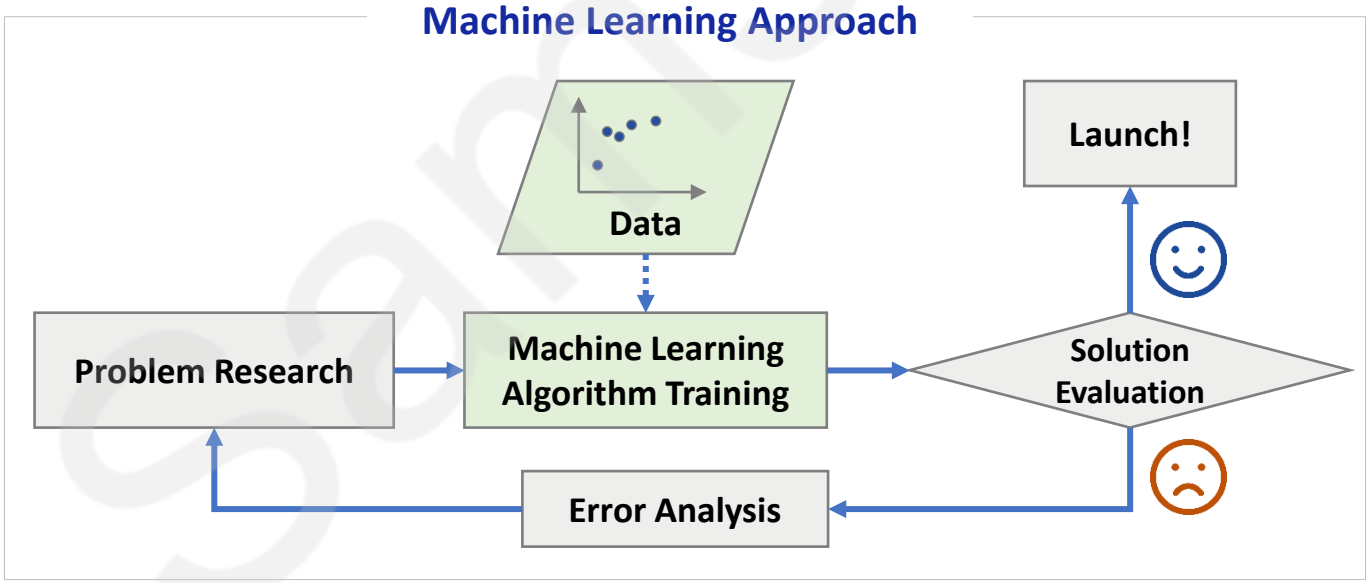
Why Use Machine Learning?

- With traditional programming techniques, the rules become longer and more complex, making it very difficult to maintain. However, spam filters based on machine learning techniques detect patterns that appear more frequently in spam than regular mails, automatically learning which words and phrases are good criteria for judging spam mails.
- Machine learning is suitable for fields that are too complex or have no known algorithms in traditional ways (e.g., speech recognition).



Strengths of Machine Learning

- Problem: Existing solutions require a lot of manual adjustment and rules. ML: One machine learning model can simplify the code and increase performance compared to traditional methods.
- Problem: Complex problems with no solutions in traditional ways. ML: Solutions can be found with the best machine learning techniques.
- Fluid Environment: Machine learning systems can adapt to new data.
- Gain insight from complex problems and large amounts of data.



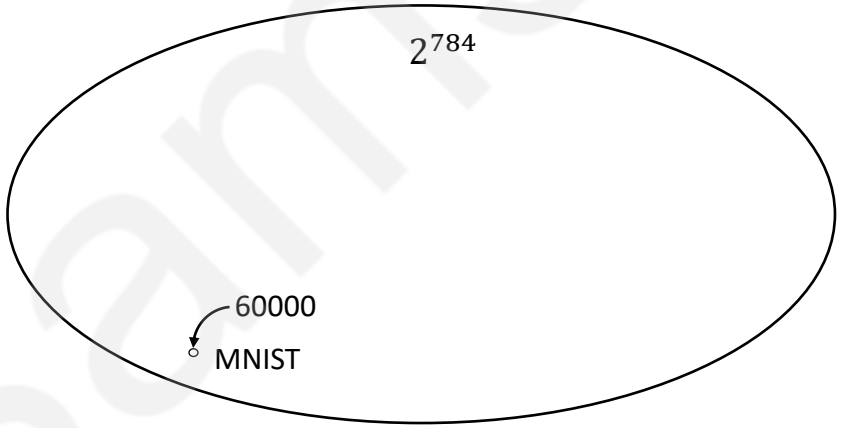
Unit 1.

The Concept of Artificial Intelligence

Limitations of Machine Learning

Insufficient Training Data

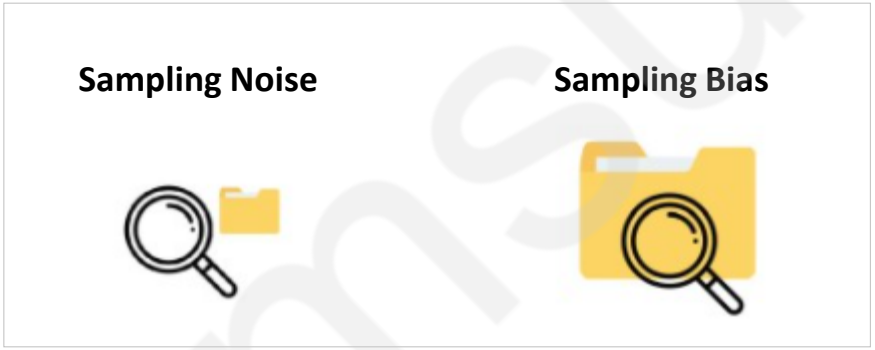
- Most machine learning algorithms require a lot of data to work well.
- Even in very simple problems, thousands of data are needed and millions of complex problems such as image or voice recognition may be needed (In the case you can't reuse a model that's already been made).
- Since collecting additional training data is not always easy or cheap, the algorithm cannot yet be ignored.



Rare Database in a Vast Space

Training Data Without **Representation**

- **Sampling Noise:** data without representation by chance
- **Sampling Bias:** very large samples are not representative if sampling method is wrong



Poor Quality Data

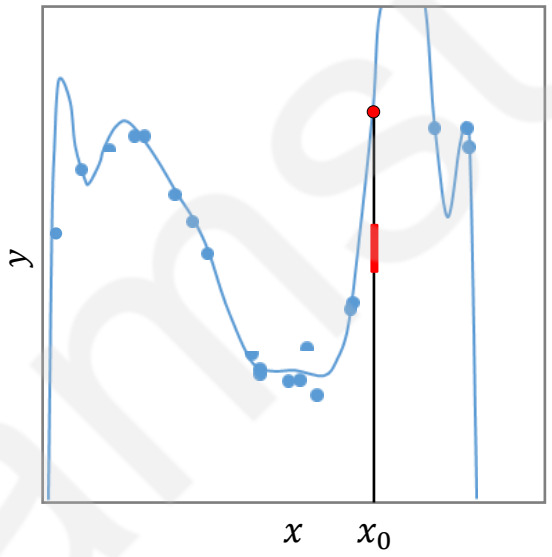
- If the training data is full of errors, outliers, and noise, the machine learning system will not work well because it is difficult to find the inherent patterns.
- **If cleaning is needed:** If it is clear that some samples are outliers, it is better to ignore or fix them

Irrelevant Features

- The system can only learn when the features and features related to the training data are sufficient.
- A key element of a successful machine learning project is to find good feature to use for training (feature engineering).
- Feature selection: Choose the most useful feature among the characteristics you are losing.
- Feature extraction: Combining features creates more useful features (reducing the dimension).

Overfitting Training Data

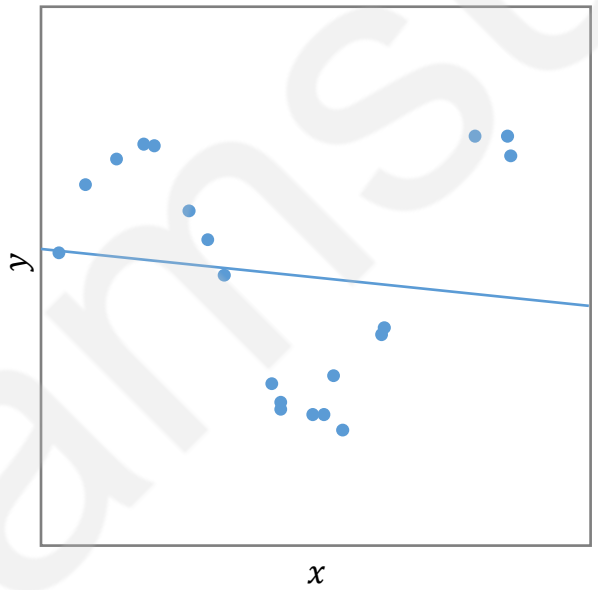
- Regularization : Putting constraints on the model to simplify the model and reduce the risk of overfitting.



Graph of an Inaccurate Prediction Due to Overfitting

Underfitting Training Data

- Select a more powerful model with more model parameters.
- Provide better features in learning algorithms (feature engineering).
- Reduce model constraints (reduce the regulation hyperparameters).



Unit 2.

Applications of Artificial Intelligence

- 2.1. Applications of Artificial Intelligence
- 2.2. Image Recognition
- 2.3. Computer Vision & Machine Vision
- 2.4. Speech Intelligence

Applications of Artificial Intelligence

- Thanks to machine learning, email spam filters, convenient text and voice recognition software, reliable web search engines, and safe and efficient autonomous vehicles will be available.
- There has also been great progress in medical applications.

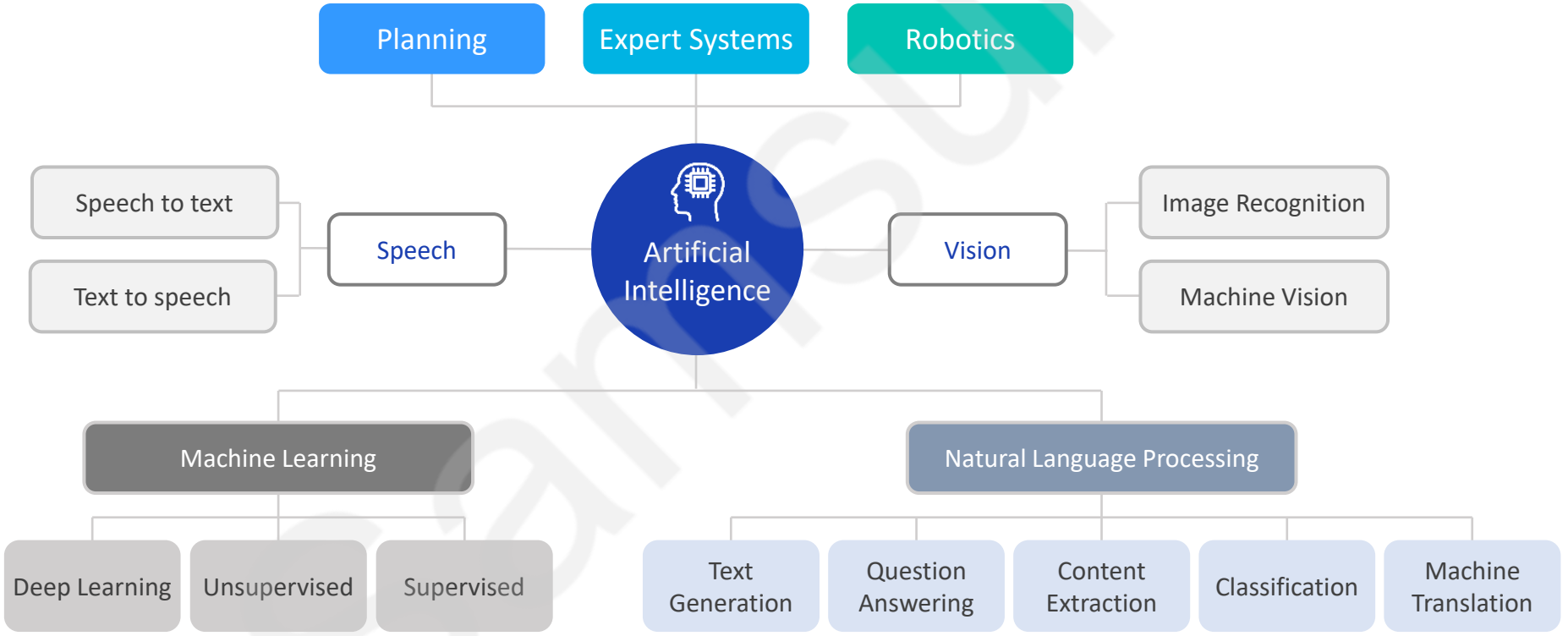
  Ex A deep learning model can diagnose skin cancer with a near-human accuracy.

  Andre Esteva, et al., "**Dermatologist-level classification of skin cancer with deep neural networks**", *Nature,* volume 542(2017), pages 115–118

  https://www.nature.com/articles/nature21056

- Deep learning predicted the 3D protein structure, surpassing the performance of physics-based methods for the first time.

Machine learning plays a very large role in our daily lives and will become increasingly important. The following are the various applications of Artificial Intelligence.

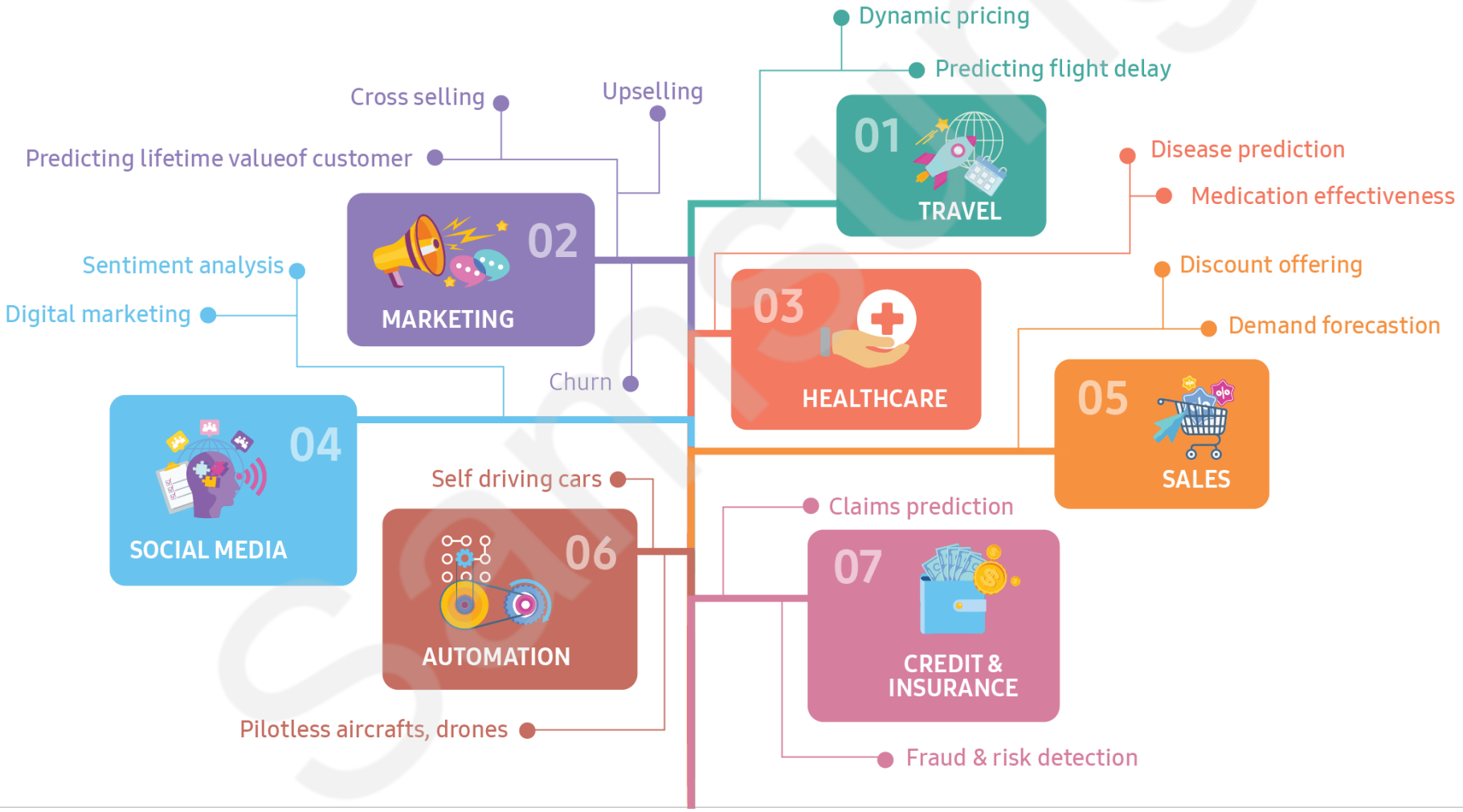


Hackernoon

Applications of Artificial Intelligence

- Image Classification: Automatically classifies product images by analyzing them on the production line
- Semantic Segmentation: Scanning the brain to diagnose tumors
- Text Classification (Natural Language Processing): Automatically classifying news articles
- Text Classification: Automatically distinguish negative comments in the discussion forum
- Text Summary: Automatically Summarize Long Documents
- Understanding Natural Language: Making a chatbot or a personal secretary
- Regression Analysis: Predicting the company's revenue for next year
- Voice Recognition: Apps that respond to voice commands
- Outlier Detection: Detecting fraudulent credit card transactions.
- Cluster Work: Divide customers based on purchase history and plan different marketing strategies for each set
- Data Visualization: Expressing a complex dataset in a clear and meaningful graph
- Recommendation System: Recommend products that customers may be interested based on past purchase history.
- Reinforcement Learning: Making intelligent game bots.

Applications of Artificial Intelligence



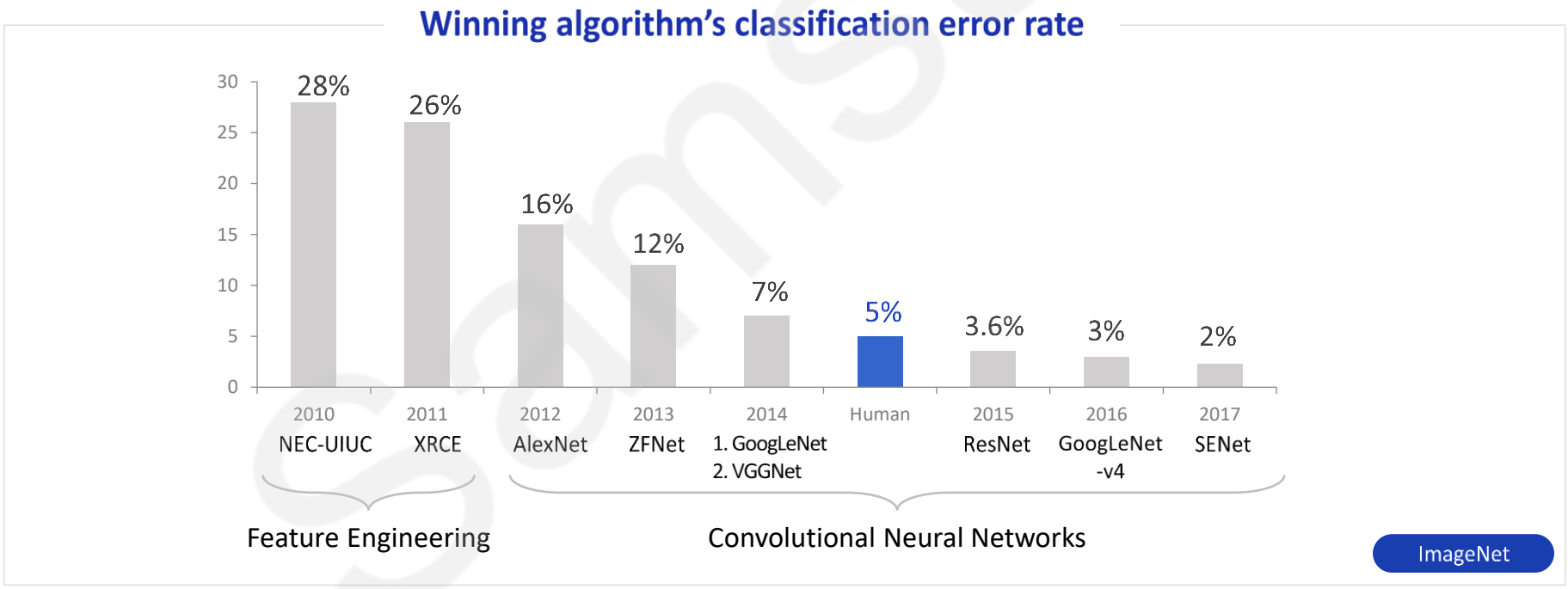
Unit 2.

Applications of Artificial Intelligence

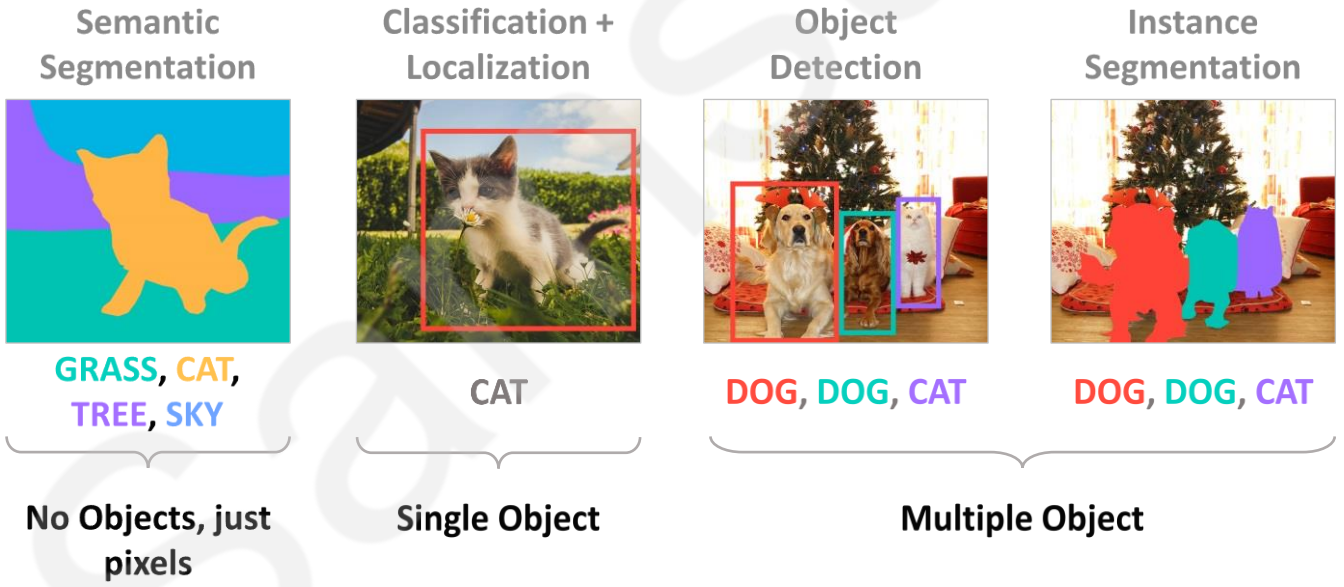
- 2.1. Applications of Artificial Intelligence
- 2.2. Image Recognition
- 2.3. Computer Vision & Machine Vision
- 2.4. Speech Intelligence

Image Recognition

- The benchmark ImageNet classification error rate fell from 26 percent in 2011 to 3.1 percent in 2016, surpassing far beyond the human capability of 5% of error rate.
- It was not until the adoption of Convolutional Neural Networks that accurate of image recognition was dramatically improved.



- Image recognition refers to technologies that identify places, logos, people, objects, buildings, and several other variables in images.
- Apart from image recognition, computer vision also includes event detection, object recognition, learning, image reconstruction and video tracking.



https://mlwhiz.com/blog/2018/09/22/object_detection/

Unit 2.

Applications of Artificial Intelligence

Computer Vision & Machine Vision

- Computer vision is an interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos.
- Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information.
- The applications of computer vision are various. They include agriculture, geoscience, biometrics, augmented reality, medical image analysis, robotics, industrial quality inspection, security and surveillance.

Scientific Research

Natural Resources Management

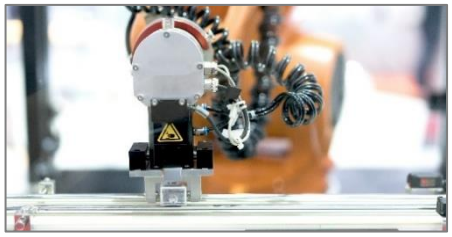
Protection and Security

Marketing

Wikipedia, Vidolab

Computer Vision & Machine Vision

- Machine vision as a systems engineering discipline can be considered distinct from computer vision, a form of computer science. It attempts to integrate existing technologies in new ways and apply them to solve real world problems
- Machine vision refers to many technologies, software and hardware products, integrated systems, actions, methods and expertise
- Machine vision (MV) is the technology and methods used to provide imaging-based automatic inspection and analysis for such applications as automatic inspection, process control, and robot guidance, usually in industry



Guidance

Gauging

Defect Detection

Packaging Inspection

Wikipedia, Vidolab, DevisionX

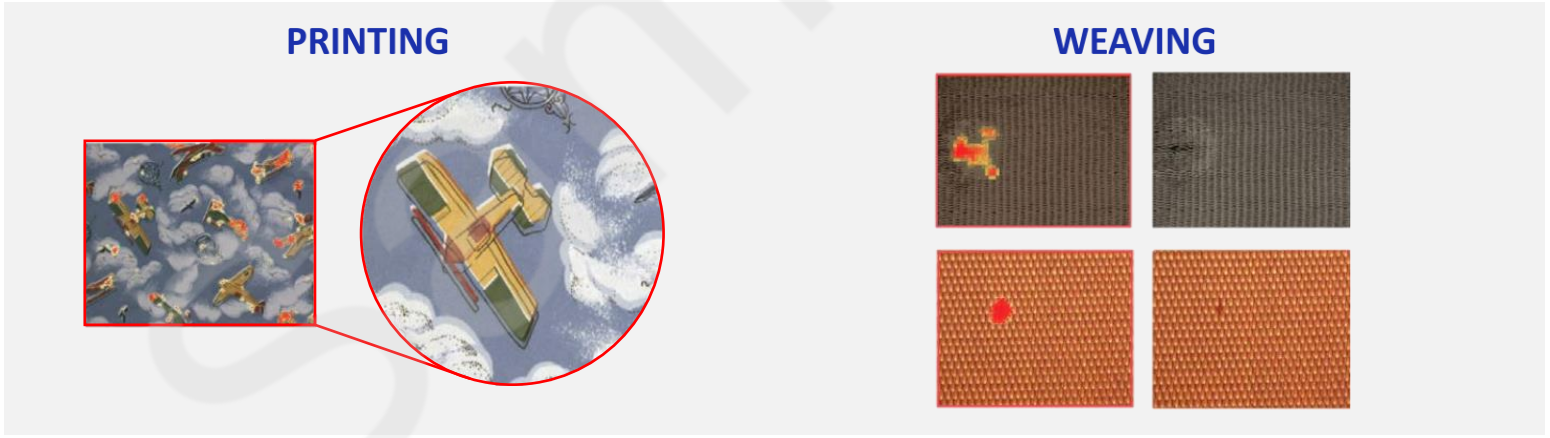
Acquire Automation – Machine Vision

- Acquire Automation has a wide range of automated vision solutions that can ideally pack and verify a company's product quality at the level required by the government or regulatory body.
- For example, the manufactured bottle is inspected at 360 degrees to ensure that the product is properly packaged and that all components are properly included.
- In detail, functions such as whether the cap is closed, seal, position and inclination, label identification, color, and barcode checks are performed.
- It also collects and provides data through machine vision technology to provide real-time production statistics and facilitate process control.

https://accquireautomation.com/

Cognex – VisionPro ViDi

- Cognex is a machine vision systems, software, and sensor manufacturer, and has launched Vision Pro ViDi, a deep learning-based image analysis software optimized for factory automation.
- VisionPro ViDi is capable of solving defect detection, texture and material classification, assembly inspection, and text decipher, which are difficult to program with rule-based algorithms.
- Since it is used for a special purpose of analyzing industrial images, only a few hundred images are required, and the training and verification time is short. The computing costs are also low.
- It was targeted for non-expert use in the vision field and developed according to actual factory conditions.



Cognex's brochure

Focal

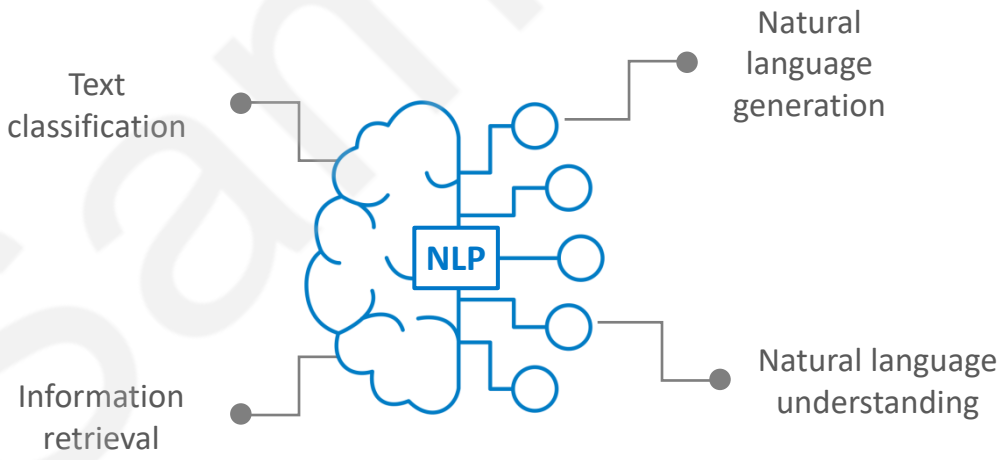
- Focal is an AI and computer vision-based stock detection systems service.
- Small and inexpensive cameras are installed in each store passage and photos are taken once every 30 minutes.
- They detect and recognize out-of-stock products in the image and provides a chart to the employees.
- It claims that it takes about four hours a day for human employees to recognize out-of-stock products.
- As labor costs increase, unmanned retail stores are increasing, but the reality is that the work of checking and replenishing inventory has become slower as manpower decreases.
- Furthermore, when checking out, items can be scanned and calculated with a local camera on the conveyor belt instead of a barcode scan. If employees focus on packaging, transaction time can be reduced by 60%.
- By providing a comparative analysis of inventory status of sold-out products and competitive products by the hour in the form of a chart, we can obtain insights on store operation.

Unit 2.

Applications of Artificial Intelligence

Natural Language Processing

- **Natural language processing (NLP)** is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.
- Challenges in NLP frequently involve speech recognition, natural language understanding, and natural language generation.
- Machine Translation, Information Retrieval, Question Answering, Information Extraction, Summarization are main application of natural language processing techniques.



Wikipedia, Expert System

Speech & Voice Recognition

- **Speech Recognition**: recognize words, sentences, and contents spoken by anyone.
  - General dictation, transcribing, using a computer hands-free, medical transcription, automated customer service, etc.

- **Voice Recognition**: recognize the accent, pitch, or intonation of a person regardless of language he or she speaks.
  - Speaker verification and speaker identification.

Barbie – Hello Barbie

- Barbie, the world-renowned doll manufacturer, has developed an AI toy that can talk.
- Hello Barbie analyzes and responds to the children's speech using natural language processing and AI.
- When you press the conversation button on the doll, the built-in microphone records what the child says, sends it to the server, and the AI selects an appropriate response and sends it back to the doll's speaker.
- Hello Barbie can remember what the children said and respond again later.

  Ex If you say that Taylor Swift is your favorite singer, the AI will repeat the same answer a few weeks later.

- The AI can predict conversations of children aged 3 to 9 and exchange conversations with children up to 200 times.

Personetics - Assist

- Assist is a domain-specific advisory chatbot service that combines knowledge of consumer finance and financial industries with interactive AI functions.
- It is gaining popularity from global banks as it can conveniently provide a 24-hour customer service without labor costs.
- Assist integrates the latest customer transaction data into the conversation to provide responses based on individual financial activities and banking history.
- It also uses predictive analysis to predict customer questions and problems and provides timely insights and advice one step ahead of the customers.
- Assist is built based on natural language processing, and can replace tasks such as remittance, reservation, and password change.
- It can also distribute customer interaction channels to message platforms such as Facebook Messenger and AI assistant platforms such as Amazon's Alexa.

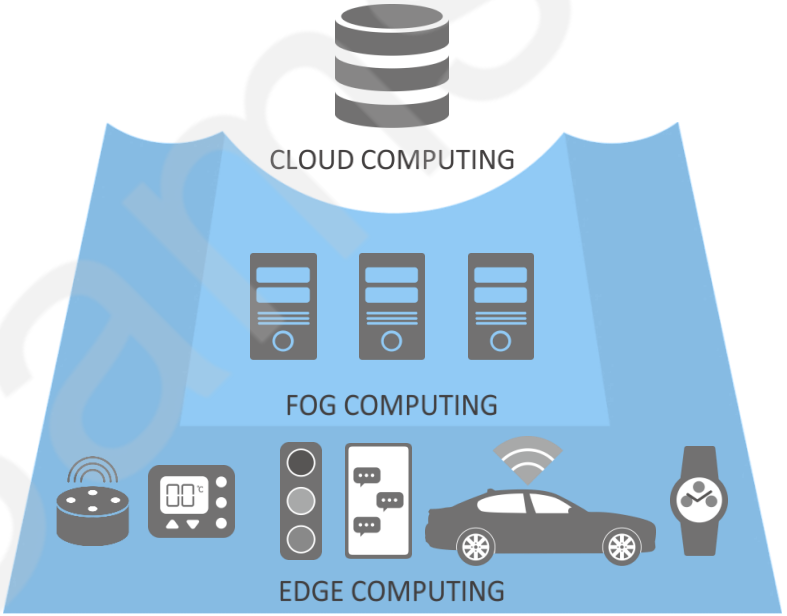
Unit 3.

Techniques in Artificial Intelligence

- 3.1. Edge AI
- 3.2. Medical Imaging & Diagnostics
- 3.3. Autonomous Vehicle
- 3.4. Reinforcement Learning
- 3.5. Conversational AI
- 3.6. GAN, XAI, Synthetic Training Data

Definition and Concept

- Running AI algorithms on edge devices, such as a smartphone or a car or even a wearable device, instead of communicating with a central cloud or server provides devices the ability to process information locally and respond more quickly to situations.
- For example, an autonomous vehicle has to respond in real-time to what's happening on the road, and function in areas with no internet connectivity. Decisions are time-sensitive and latency could prove fatal.



CB Insights

Initiatives

				
Edge AI use case	In-home smart cameras can recognize that a person(s) has entered an area	On-device facial recognition and object recognition, where user data doesn't leave the device	Instantaneous driving decisions	Vision for baby monitors, drones, robots, and other devices that can respond to situations without internet connection
	nest aws IQ cameras, Deep Lens	HUAWEI AI processor	TESLA custom AI chip	intel Myriad X

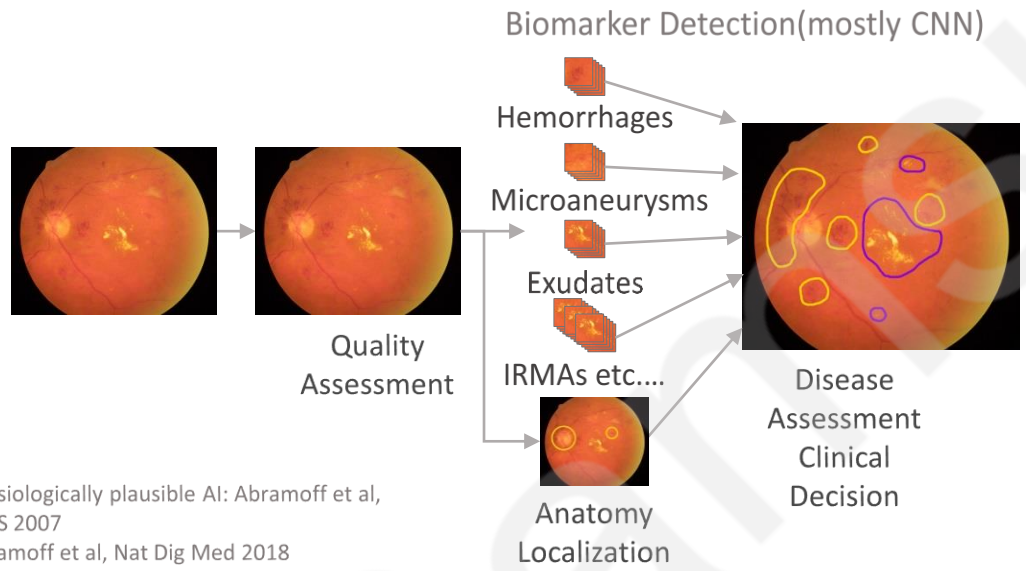
CB Insights

Unit 3.

Techniques in Artificial Intelligence

AI-as-a-medical-device

Autonomous AI algorithm based on biomarkers



Physiologically plausible AI: Abramoff et al, IOVS 2007
Abramoff et al, Nat Dig Med 2018

CB Insights, IDx

- In April 2018, the FDA approved AI based software that screens patients for diabetic retinopathy without the need for a second opinion from a medical specialist.
- The software, IDx-DR, correctly identified patients with “more than mild diabetic retinopathy” 87.4% of the time, and identified those who did not have it 89.5% of the time.
- IDx is one of the many AI based software products approved by the FDA for clinical commercial applications in recent months.

AI-as-a-medical-device

Diabetic Retinopathy

- FDA clearance
- AI based software that screens patients for diabetic retinopathy without confirmation of medical specialist

Liver and Lung AI Lesion

- FDA clearance
- GE Ventures-backed startup
- Oncology AI suite initially focused on spotting lung and liver lesions.

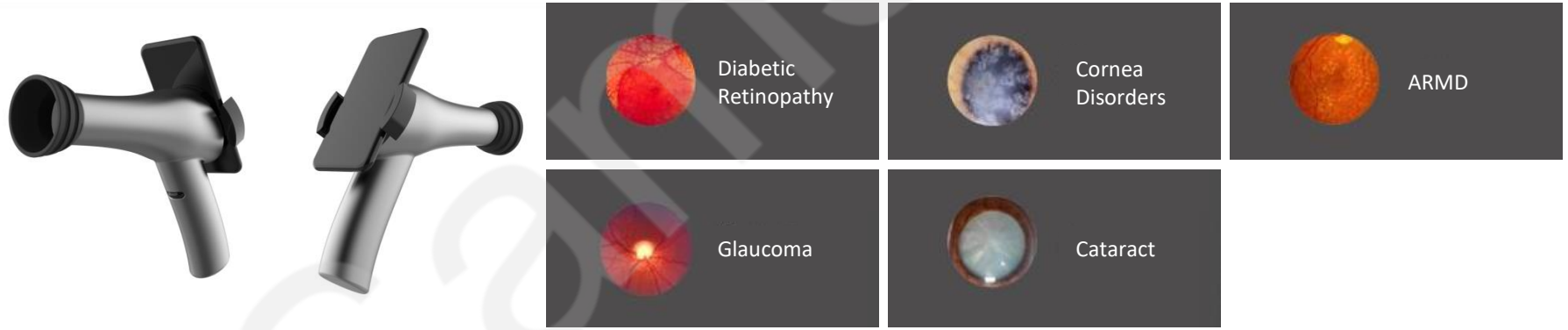
CT-Scan Analysis for Strokes

- FDA clearance
- Analyze CT scans and notify healthcare providers of potential strokes in patients
- $21M Series A round from Google Ventures and Kleiner Perkins Caufield & Byers

CB Insights

Portable Ophthalmoscope Based on AI (Project Powered by Samsung)

- Developed a portable diagnostic device and diagnosis support algorithm based on image processing for underserved people in Vietnam.
- 40,000 images are being accumulated per month.
- Detect anomaly from eye images based on Machine learning and classify the image for triage.



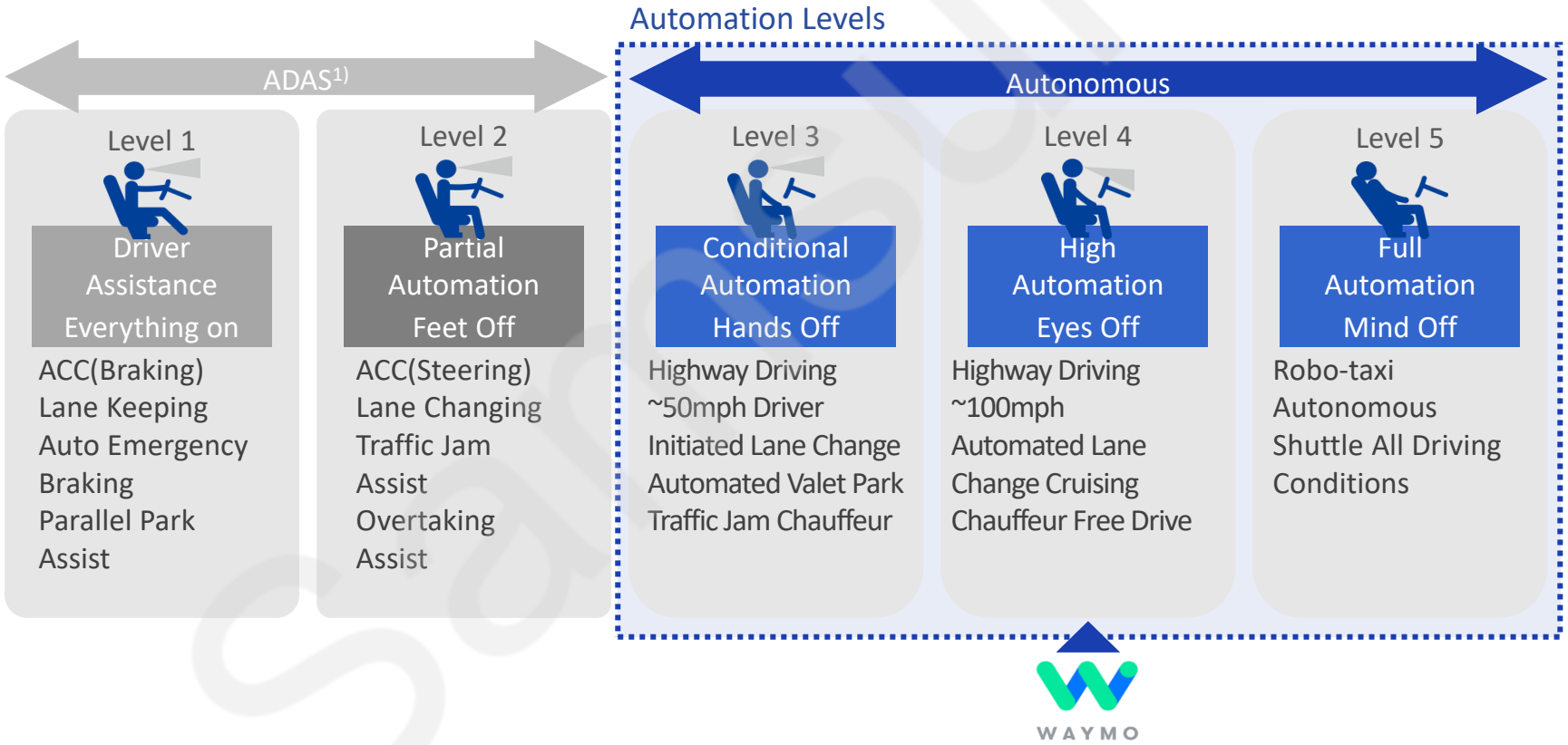
Samsung, Project BOM

Unit 3.

Techniques in Artificial Intelligence

Autonomous Vehicle

Level of Automation

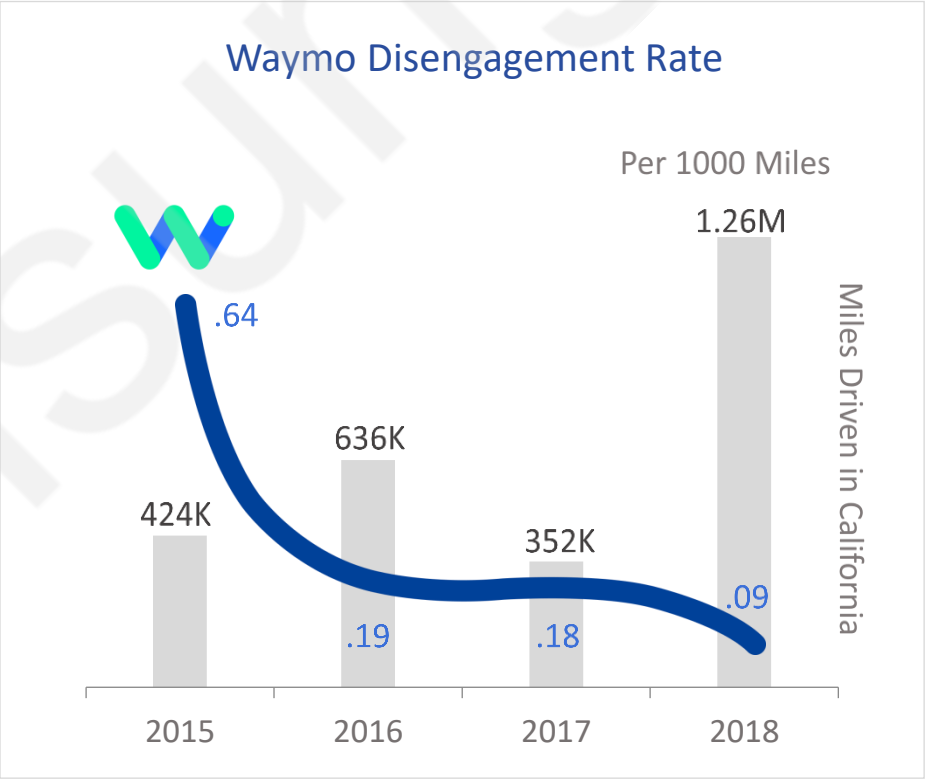


Note: 1) ADAS : Advanced Driver Assistance Systems

IEEE Spectrum, SAE International, Evercore ISI

Miles per Disengagement

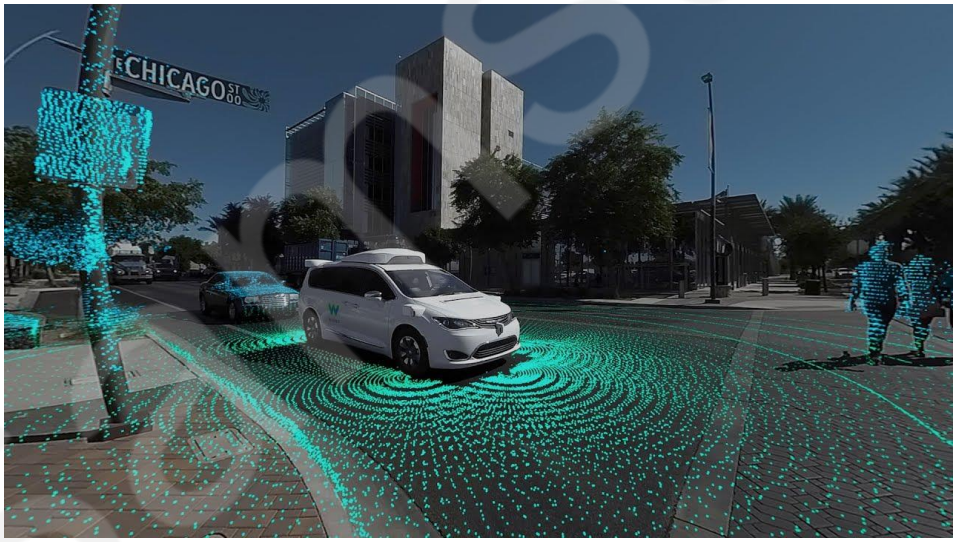
- All companies actively testing self-driving cars on public roads in California are required to disclose the number of miles driven and the frequency in which human drivers were forced to take control of their driverless vehicles - also known as a “disengagement”.
- Waymo said its autonomous system only disengaged at a rate of once every 11,017 miles. GM Cruise reported the second-lowest disengagement rate, with a safety driver taking over every 5,205 miles
- Other startups making progress include Zoox (0.50 disengagements per 1,000 miles), Nuro (0.97 disengagements per 1,000 miles), and Pony.ai (0.98 disengagements per 1,000 miles).



The Verge, Forbes

Miles per Disengagement

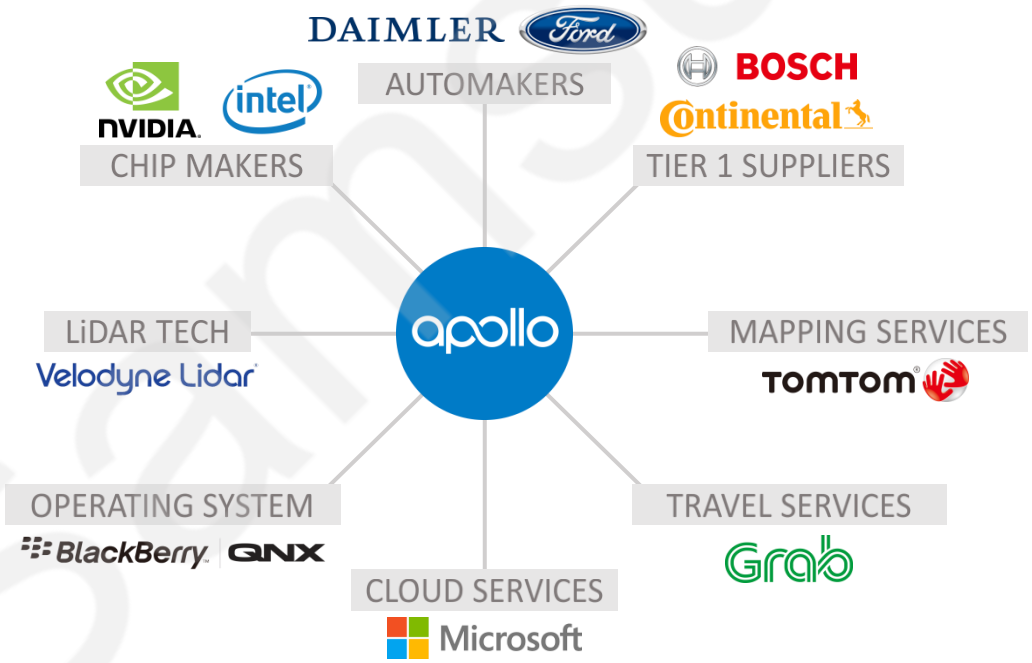
- Google has made a name for itself in the auto space. Its self-driving project Waymo is the first autonomous vehicle developer to deploy a commercial fleet of AVs.
- Investors remain confident in companies developing the full autonomous driving stack, pouring hundreds of millions of dollars into GM's Cruise Automation ($750M from Honda in October 2018 and $900M from SoftBank in May prior) and Zoox ($500M in July 2018). Other startups here include Drive.ai, Pony.ai, and Nuro.



https://www.forbes.com/sites/alanohnsman/2018/03/02/waymo-is-millions-of-miles-ahead-in-robot-car-tests-does-it-need-a-billion-more/#c854b381ef4c

China Market

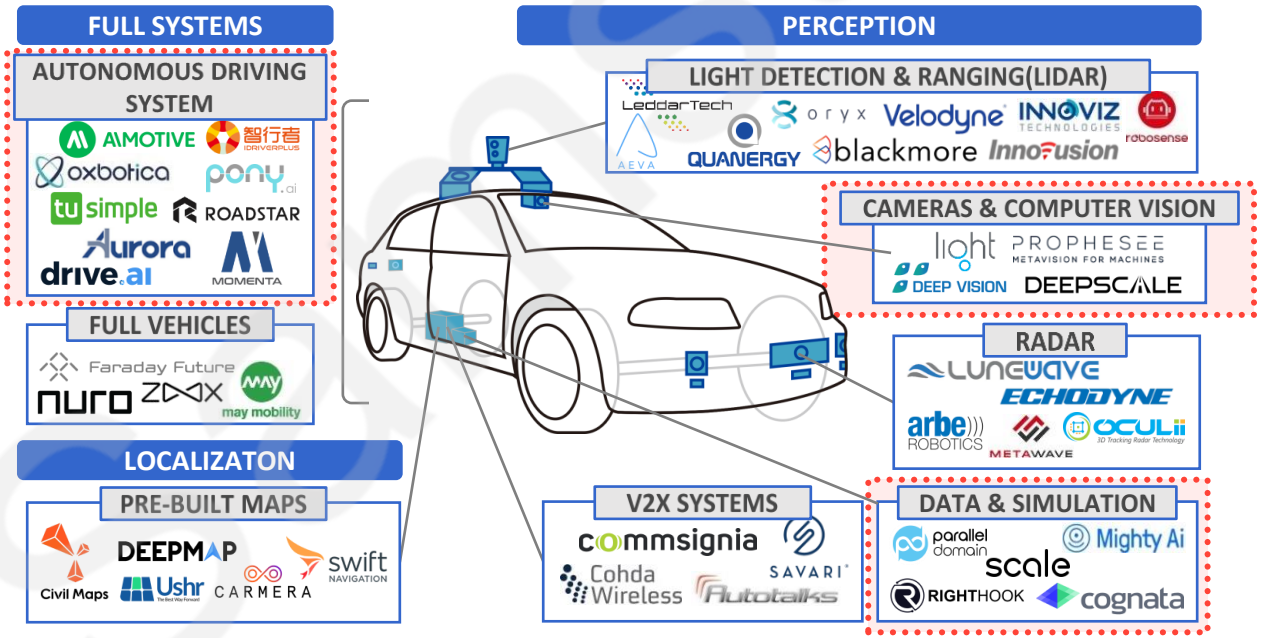
- The Chinese science ministry announced last year that the nation's first wave of open AI platforms will rely heavily on Baidu for autonomous driving
- In April 2017, Baidu announced a one-of-a-kind open platform, called Apollo, for autonomous driving solutions, roping in partners from across the globe



CB Insights

Unbundling the Autonomous Vehicle

AUTONOMOUS DRIVING SYSTEMS	COMPUTER VISION	DATA & SIMULATION
• Drive.ai • Momenta • Pony.ai	• DeepScale • Prophesee	• Cognata • NVIDIA



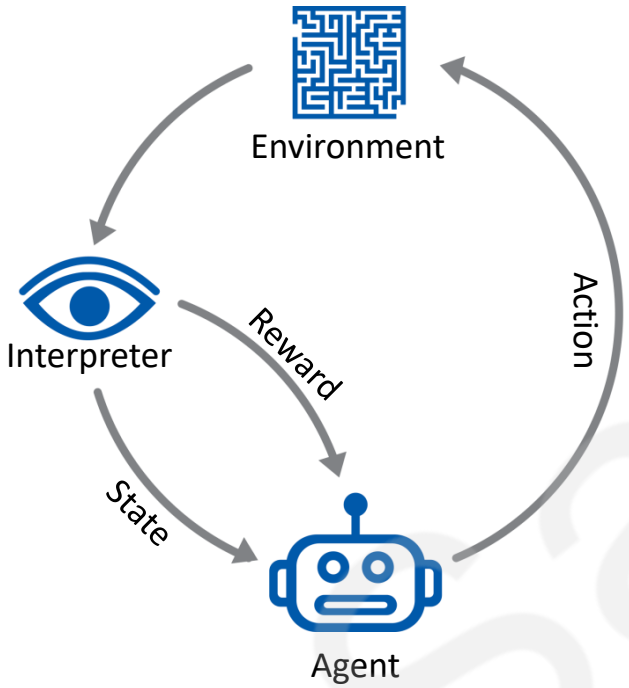
Insights

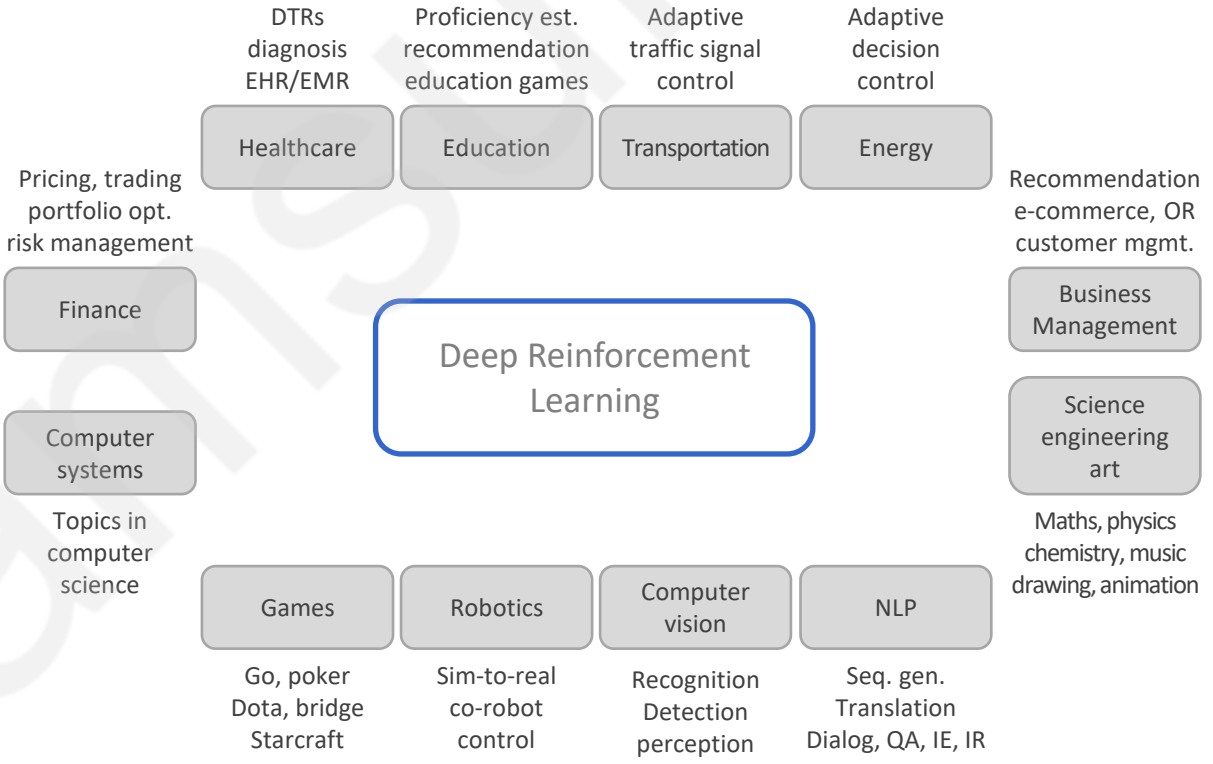
Unit 3.

Techniques in Artificial Intelligence

Reinforcement Learning

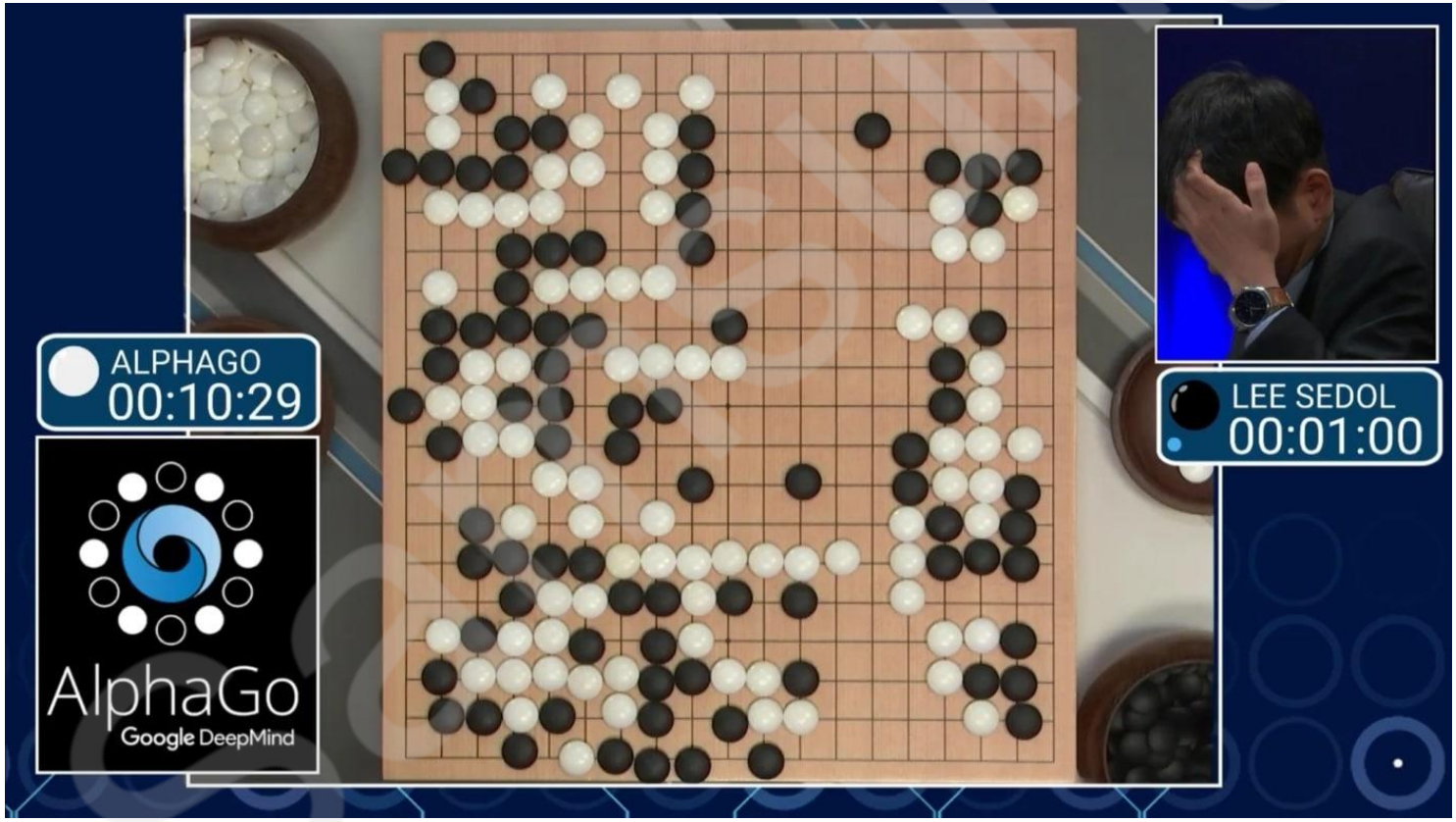
Definition and Application





Yuxi Li, Deep Reinforcement Learning, arXiv, 2018

Definition and Application



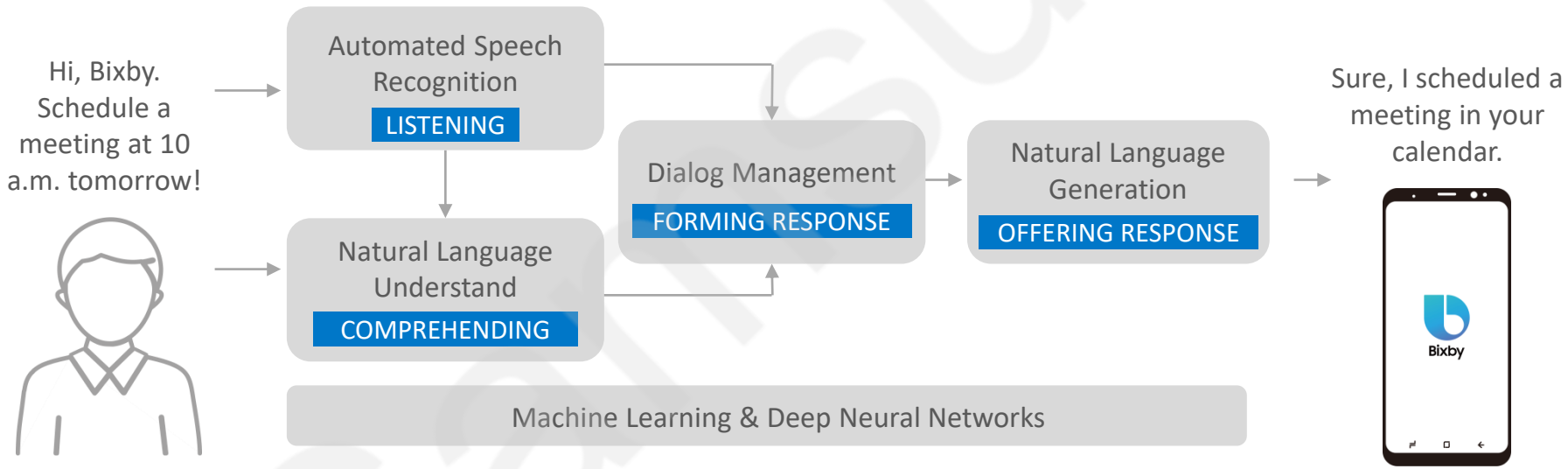
https://dimensionless.in/reinforcement-learning-super-mario-alphago/

Unit 3.

Techniques in Artificial Intelligence

Conversational AI

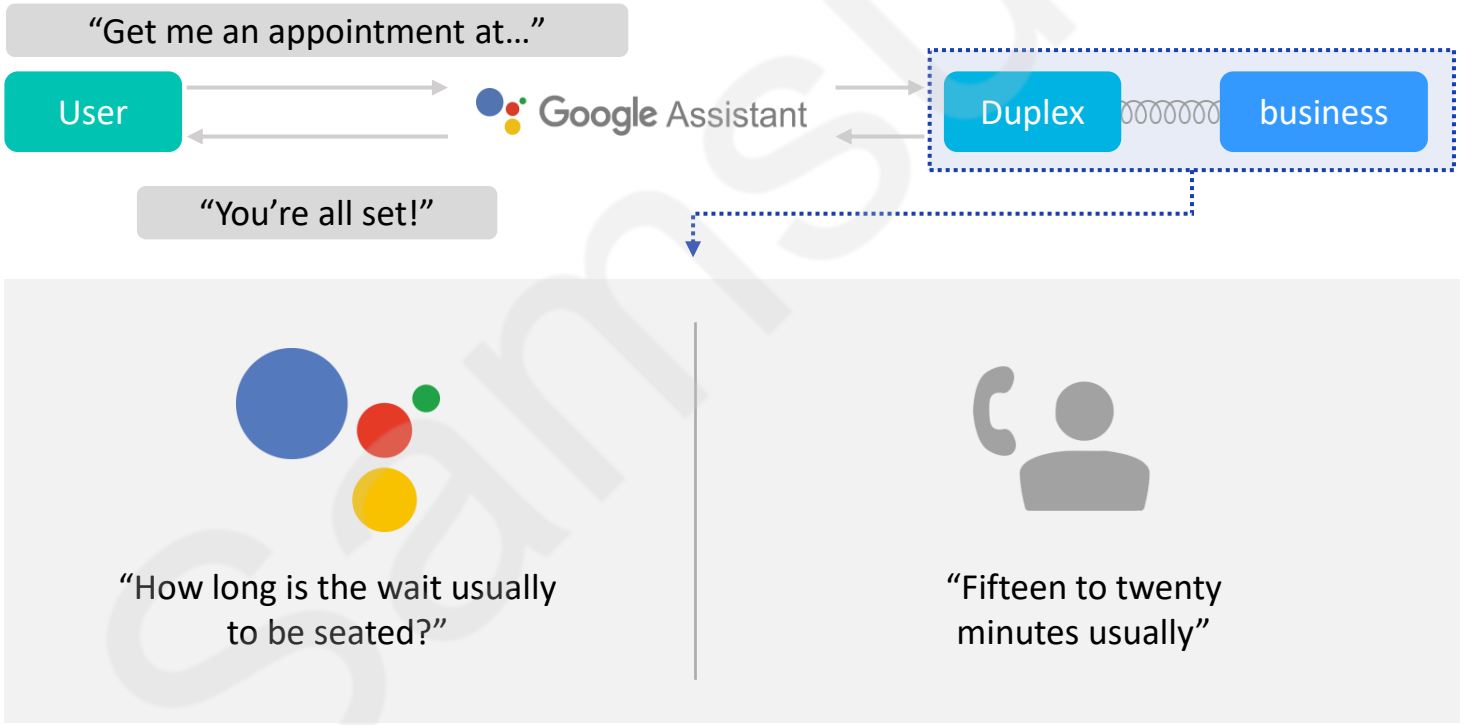
Voice Assistant



CB Insights, Interactions

Duplex

- Duplex can make phone calls and reservations on behalf of the user, communicating like a real human. Google added Duplex to its new phone, Pixel 3. It has turned the Pixel 3 into an AI powerhouse, including a “screen call” option that allows the Google Assistant to screen for spam callers.



CB Insights, Google

Unit 3.

Techniques in Artificial Intelligence

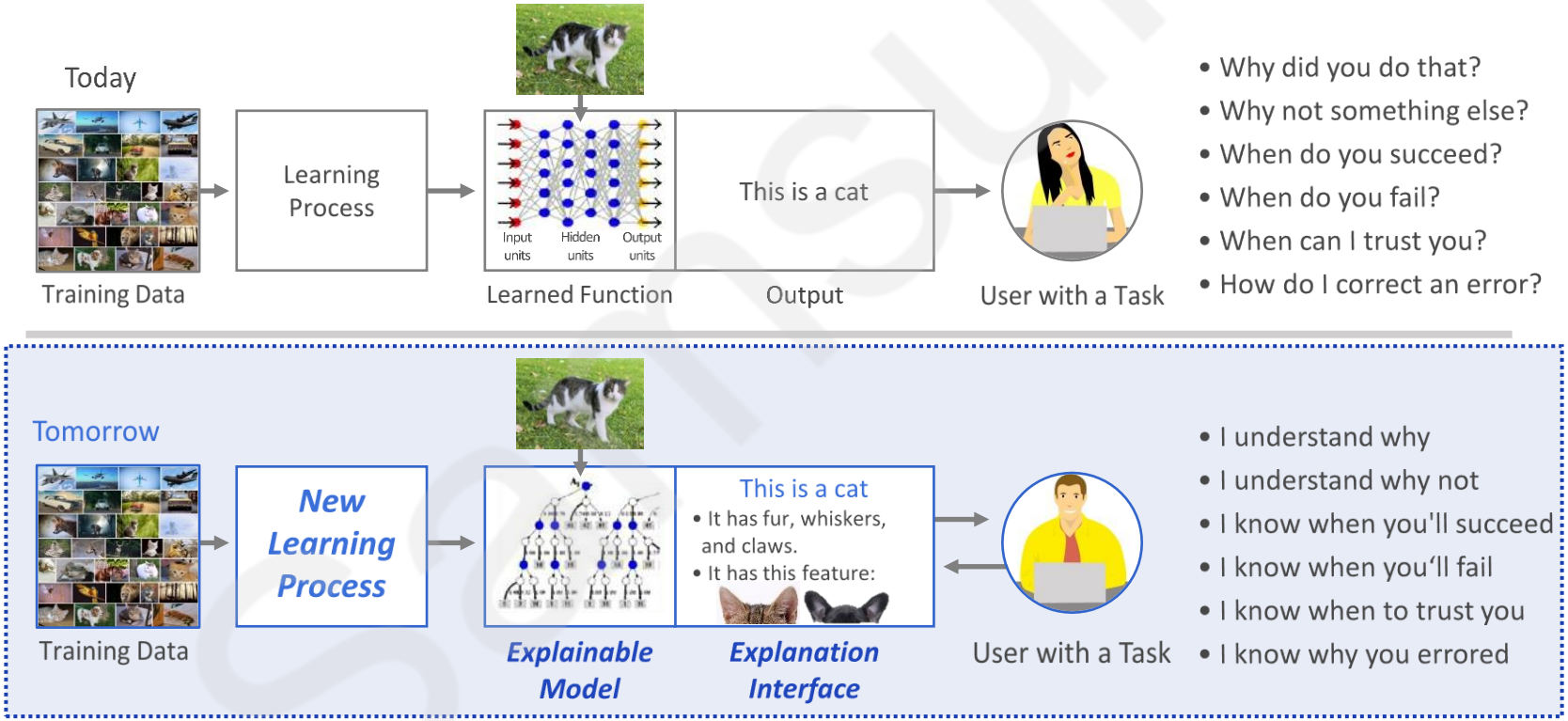
Generative Adversarial Network (GAN)

Deep fake

- Pictures on this page are all generated by artificial intelligence.
- Generative Adversarial Network, also known as GAN, brought a groundbreaking innovations in artificial intelligence field, especially in generating fake images and videos.
- In 2018, researchers at Carnegie Mellon University created a video in which two people showed exactly the same facial expression simultaneously –
  it is called deep fake.
- With recycled-GAN algorithm, the quality of
  fake video was significantly enhanced.

XAI (eXplainable AI)

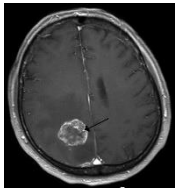
What are we trying to do?



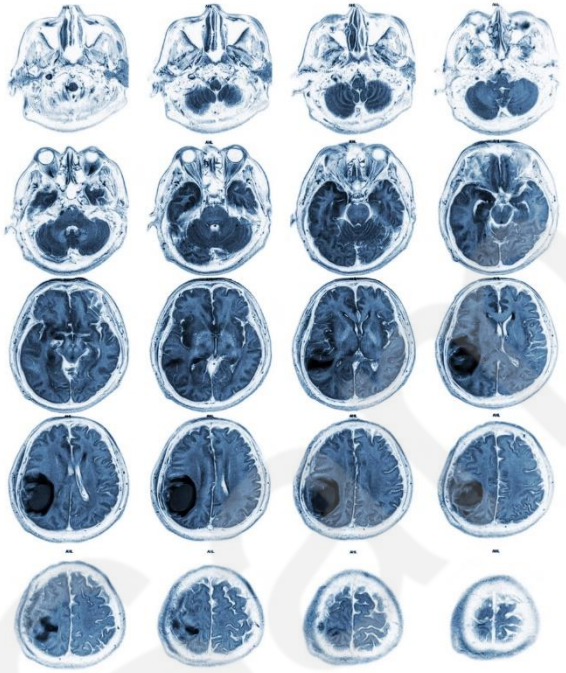
DARPA

Synthetic Training Data

Synthetic Brain Tumor MRI images



Original Brain Tumor Image



Synthetic Brain Tumor Image
for training tumor segmentation algorithm

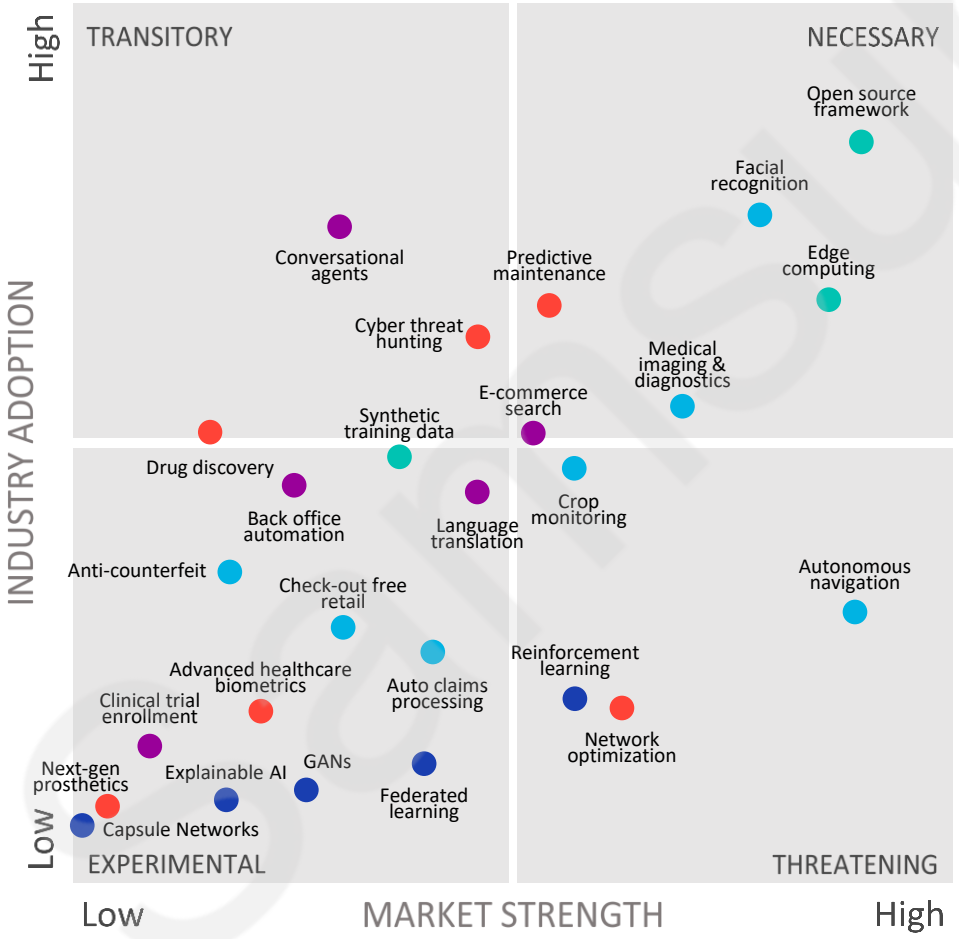
- Researchers generated synthetic medical image of brain tumor by training a generative adversarial network algorithm with publicly available MRI data
- With synthetically generated data,
  - they were able to improve the accuracy of tumor segmentation.
  - and they used the generative model as an anonymization tool which mitigates the risk of using patients' medial data without permission.

DARPA

Unit 4.

Artificial Intelligence: Trends and Markets

Trend Matrix



- Application: Computer vision
- Application: Natural language Processing / synthesis
- Application: Predictive intelligence
- Architecture
- Infrastructure

Artificial Intelligence Trends 2019, CB Insights

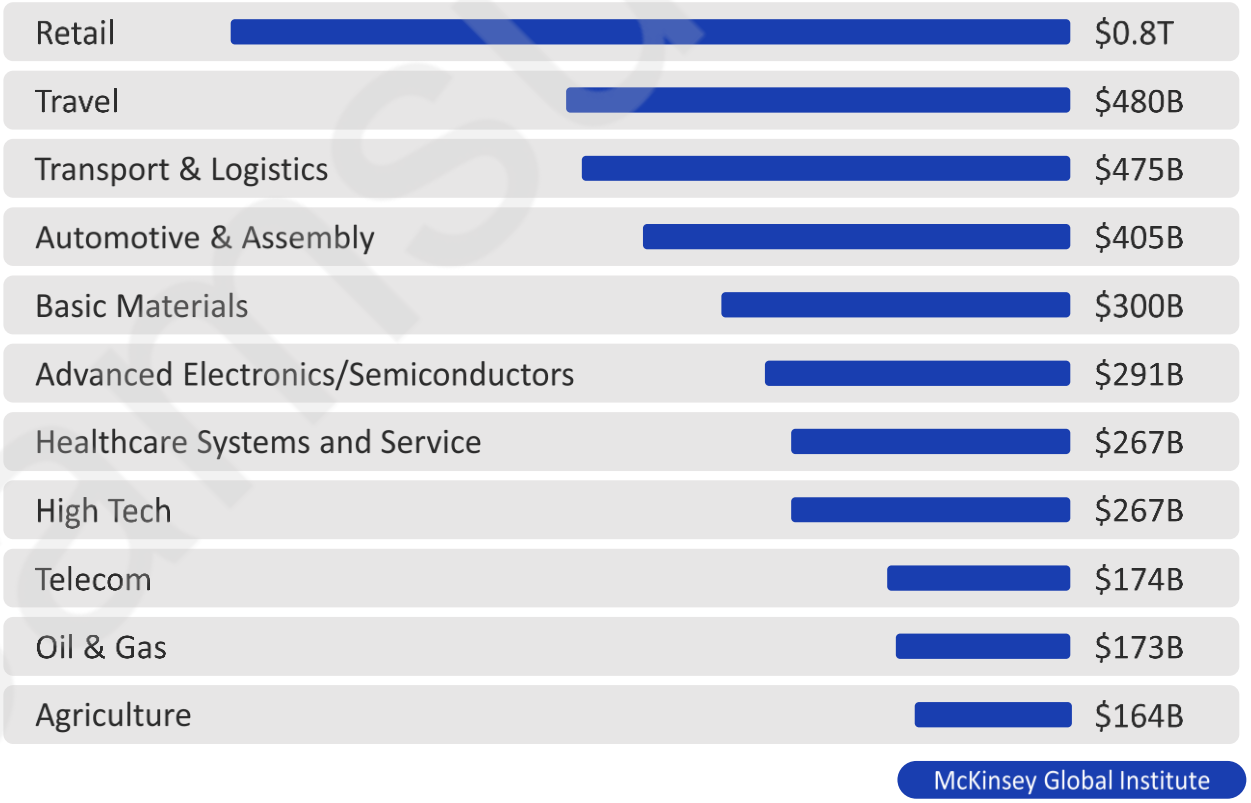
Unit 4.

Artificial Intelligence: Trends and Markets

AI Value Creation by 2030

Innovation created by artificial intelligence (AI) started from software industry and will soon revolutionize all the other industries. Value created by artificial intelligence will reach 13 trillion dollars by 2030.

$13 Trillion



Unit 4.

Artificial Intelligence: Trends and Markets

AI in Sustainable Energy

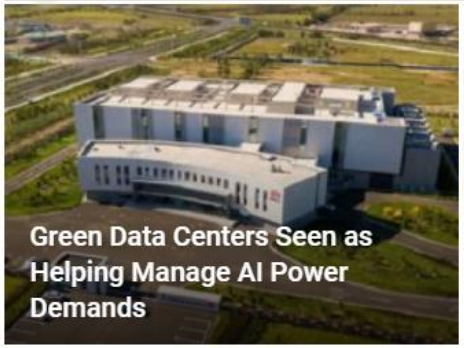
Green AI establishes a connection between computing power and carbon emissions, allowing one to manage the carbon cost of large AI computing demands to some extent.

- Today, greenhouse emissions from the information and communication technology (ICT) industry account for 2% of total emissions (as much as all air traffic), warning that failure to confirm this will increase up to 14% of global emissions. AI itself can be used for eco-friendly projects.
- Examples

- Space Intelligence, a British startup, applies machine learning and AI to satellite data to solve environmental problems such as re-seasoning, allowing the industry to take corrective action.

- Google has promised not to use carbon by 2030. At the recent Google I/O 2021 event, the company explained these efforts in detail by announcing plans to use geothermal energy to power data centers in Nevada.
- Kate Brandt, Google's head of sustainability, said, "We are the first company to sign a contract for next-generation geothermal power."

AI in Sustainable Energy



https://www.aitrends.com/category/energy/

Unit 4.

Artificial Intelligence: Trends and Markets

AI in Financial Services

As cyber attacks and identity fraud losses increased sharply in 2020, the pandemic set the stage of combining AI and biometric recognition to make telecommuting a standard and achieve a higher level of protection.

- Banks around the world saw a 238% surge in cyber attacks between February and April of 2020.
- According to a study by Javelin Strategy & Research, ID fraud losses rose to $56 billion last year as fraudsters created synthetic IDs using stolen personal information, as accounted by Pymnts.com.
- In addition, the number of automated bot attacks increased by 100 million between July and December for companies in various industries.
- Companies that strive for better protection risk making their customer's lives more difficult. It' been reported about 40% of financial institutions often mistake legitimate customers' online behavior for fraudulent behavior.

Security Benefits of AI and Biometric Recognition

- AI can be used to prevent digital identity fraud, such as providing higher accuracy and speed when verifying an individual's identity or integrating biometric data so that cybercriminals cannot access information simply by providing credentials.

Pandemic-changed Consumer Financial Behavior, Causes Identity Fraud

- The global pandemic has had a dramatic impact on consumer financial behavior. In 2020, consumers spent more time at home, there was less trading than the previous year, and they relied heavily on streaming services, digital commerce, and digital payments.
- They also communicated more via email and text about both work and personal life.

https://www.aitrends.com/category/financial-services/

Unit 4.

Artificial Intelligence: Trends and Markets

AI in Government

Wisconsin Department of Workforce Development(DWD) summarized the 2020 backlogs for thousands of unemployment insurance claims with the help of AI cloud computing tools that surged during the pandemic.

- DWD handles an average of 157,000 claims per week and has paid a total of $2 billion in unemployment benefits since the outbreak of the pandemic. According to GCN's recent account, the agency can now release funds within three business days.
- The process took weeks or months, according to Brent Mitchell, the head of state and local governments in Google's public sector. Recovery was helpful by DWD hiring, contracting , or rearranging 1,300 people to help with backlogs.
- **DWD used Google Cloud's DocAI service along with AI and machine learning products.** DocAI is used to add value by suggesting ways to automate the extraction of sensitive data from documents and simplify workflows. **The Human-in-the-Loop AI function increases accuracy by adding human review.**

"With design thinking, close partnerships with state officials, and combination of modern technologies, DWD's solutions are customized to maximize benefits to components," said Mitchell.

AI in Government

https://www.aitrends.com/category/ai-in-government/

Unit 4.

Artificial Intelligence: Trends and Markets

AI in Healthcare

In February, the U.S. Food and Drug Administration (FDA) announced a plan regarding the direction in which manufacturers, seeking FDA approval for innovative medical devices using AI, should pursue. As a part of this medical device execution plan, the FDA announced a AI and machine learning software.

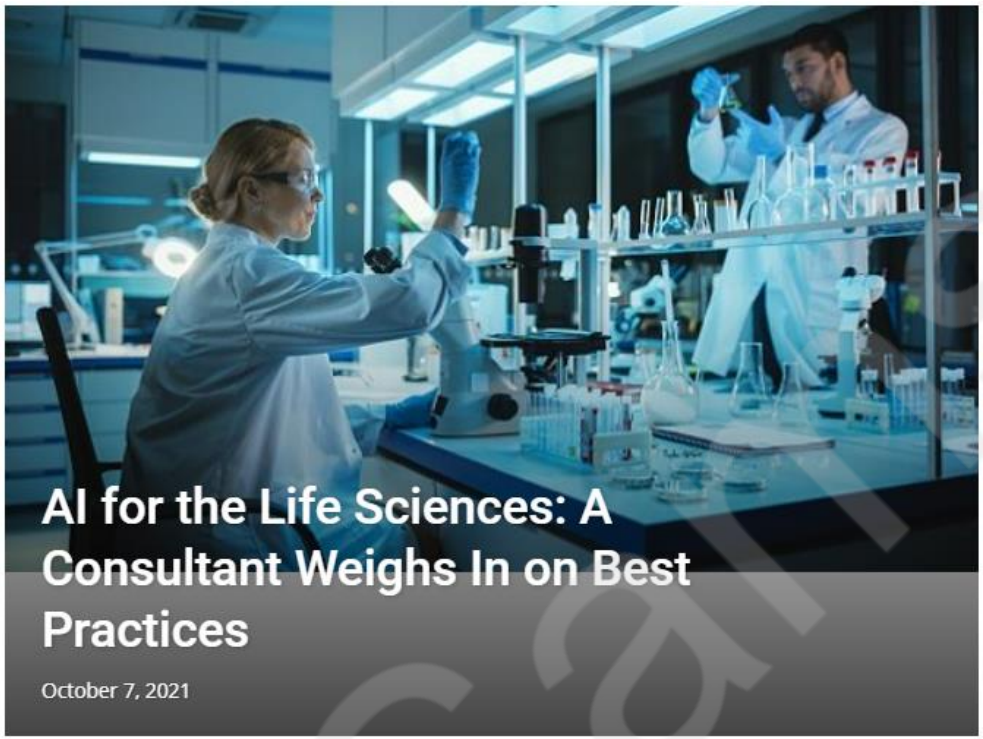
- "Artificial Intelligence (AI) and machine learning (ML) technologies have the potential to transform healthcare by drawing new and important insights from vast amounts of data generated during daily healthcare service," said the author of the report.
- This report is a follow-up to the FDA's 2019 efforts to begin discussions on how to proceed with medical devices using AI in response to a request to update the approval process. "The FDA's traditional paradigm for regulating medical devices is not designed for adaptive artificial intelligence and machine learning technologies," says the author of the FDA report.
- The FDA's action plans look for manufacturers' commitments to maintain device performance with processes for periodic updates as these devices develop. In this way, the FDA and manufacturers can evaluate software products from pre-marketing development to post-marketing performance.
- This allows the FDA to accommodate the power of AI and ML-based software on medical devices while ensuring the safety of the patient.

AI in Healthcare

The author identified promising target uses of AI SaMD in the field of medical devices, including the following.

- **Heart Disease Diagnosis**: A machine learning algorithm that integrates age and gender paired with high-sensitivity heart troponin I concentration (myocardial ischemic injury index) was used to train AI platforms using data from 3,013 patients.
  - The platform was, then, tested on 7,998 patients suspected of myocardia infarction. It was found to surpass doctors with a 82.5% sensitivity and 92.2% specificity.
- **Detecting Retinopathy:** Diabetic retinopathy (DR) is one of the main causes of preventable blindness worldwide. The U.S. Academy of Research Ophthalmology tested an artificial intelligence engine that distinguishes healthy fundi in people with DRs using 75,137 public fundus images in diabetic patients. As a result, the sensitivity and specificity were at an impressive 94% and 98%, respectively.
- **Biosensors for Vital Sign Monitoring**: Biosensor-based devices generate vast data sets. AI can be used to predict trends and disease probabilities.
  - A good example is the integration of AI into a heart monitoring-based biosensor for field treatment (POC) diagnosis. Machine learning algorithms are used with microchip-based heart biosensors for real-time health monitoring and provide timely and accurate clinical decisions.

AI in Healthcare



https://www.aitrends.com/category/healthcare/

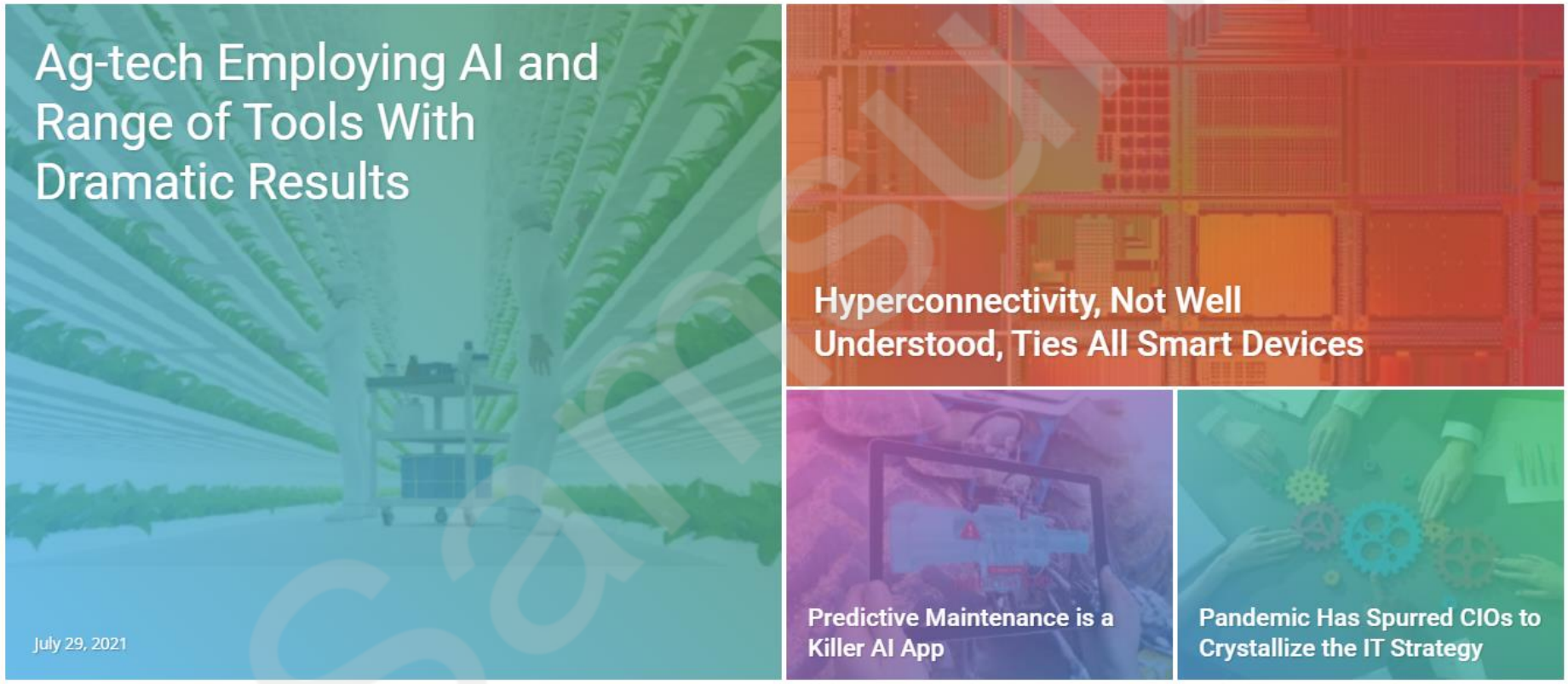
Unit 4.

Artificial Intelligence: Trends and Markets

AI in Healthcare

- Agricultural Technology (ag-tech) startups in Plenty, San Francisco, use AI and robots to plant crops vertically indoors throughout the year.
  - **Plenty's vertical farm approach can produce the same amount of fruit and vegetables from only 2 acres compared to a flat farm of 720 acres.**
  - Plenty is one of hundreds of **ag-tech startups** that receive billions of dollars in investment in the capital market using new technology approaches, including AI, drones, robots, and IoT sensors.
  - Plenty's climate-controlled farm is lined with vertically growing plants hanging from the ceiling. LED lights imitating the sun illuminate plants. Robots move around them. AI manages all variables of water, temperature, and light. AI continues to learn and optimize how to grow better crops.
  - In addition, vertical farms are located in urban areas and can produce local food, removing the need for transport miles. Benefits of locally produced crops include reduced CO2 emissions from transport vehicles and potentially reduced consumer prices.
  - "The collapse of the supply chain caused by COVID-19 and natural turmoil, such as the fires in California this year, show that the need for predictable and durable product supply can only come from vertical agriculture," said Storey.

AI in Healthcare



https://www.aitrends.com/category/iot/

Unit 5.

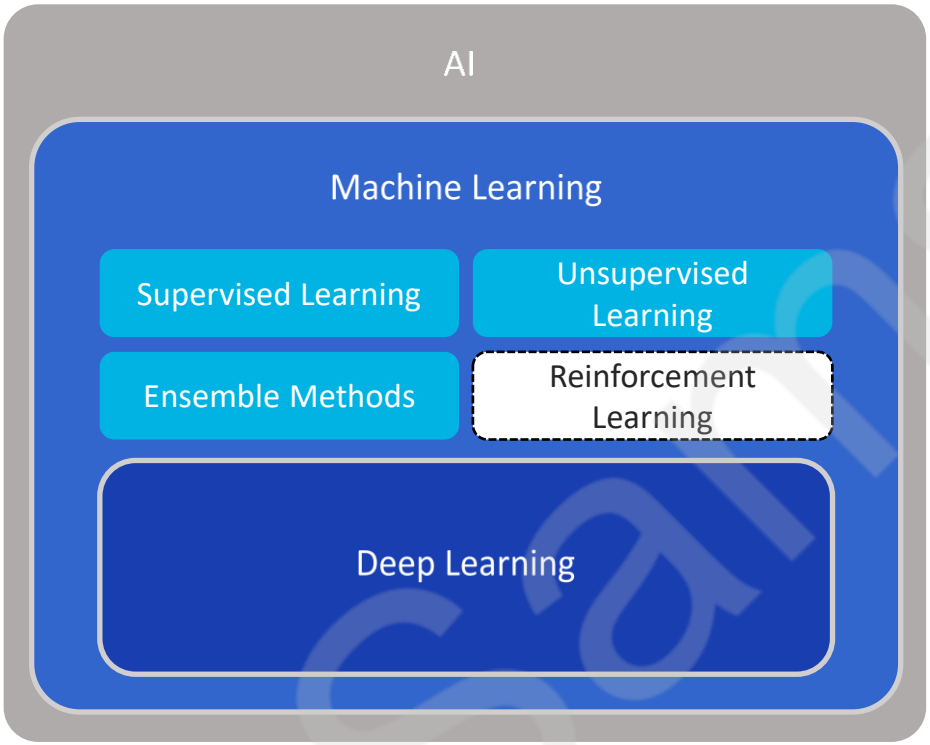
Course Roadmap

5.1. Artificial Intelligence Course Roadmap

5.2. Category of Machine Learning Techniques (Full)

Artificial Intelligence Course Roadmap

Course Coverage



- Artificial Intelligence course in Samsung Innovation Campus will deal with the majority of key concepts and techniques of AI, except reinforcement learning.
- Since reinforcement learning is an advanced AI technique, it is not easy to learn in a short-term period. For those who have interest in the topic, it is highly recommended to start learning by yourself during the course or after completing the course.
- The course will be focused mainly on practical knowledge on how to code by yourself.

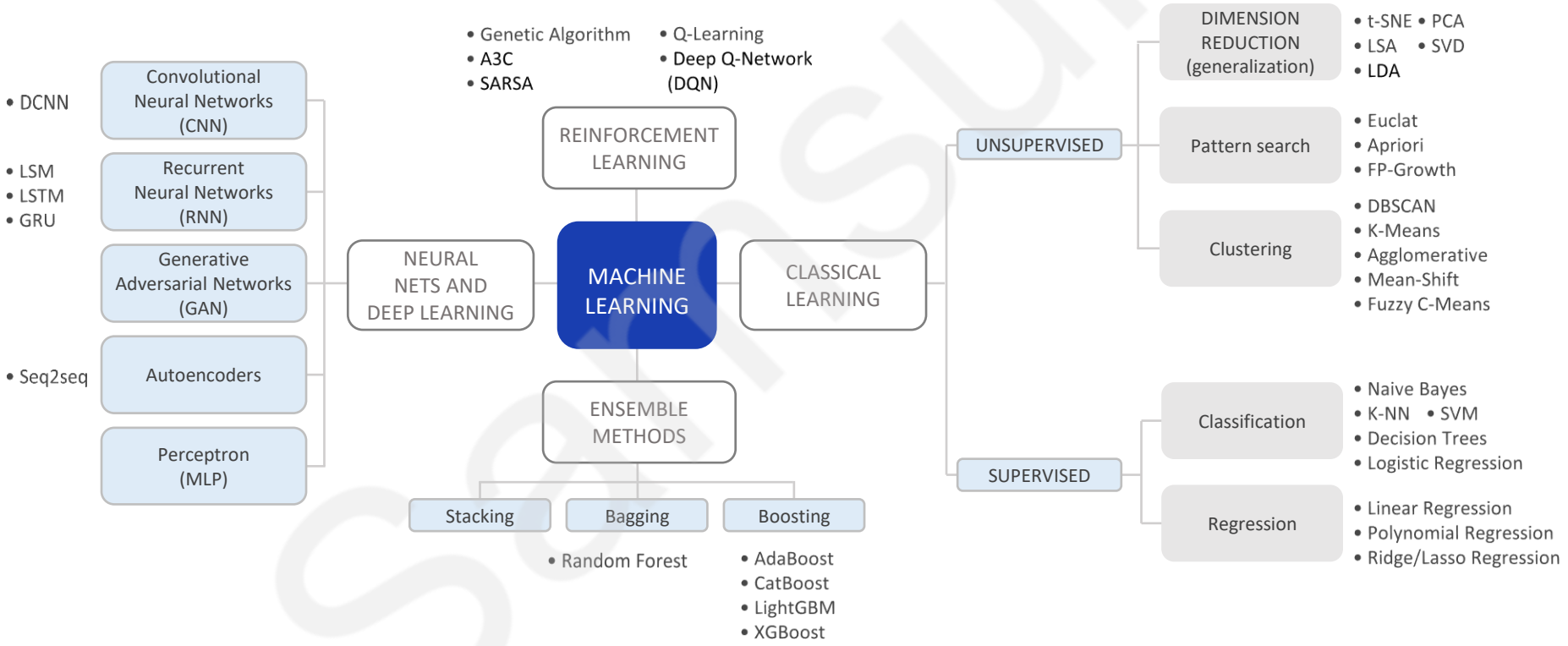
Unit 5.

Course Roadmap

5.1. Artificial Intelligence Course Roadmap

5.2. Category of Machine Learning Techniques (Full)

Category of Machine Learning Techniques (Full)



SAMSUNG

Together for Tomorrow!

Enabling People

Education for Future Generations