# DSE8-Prediction

DSE - Data Science Course by Data Scientists

© Botnoi Group

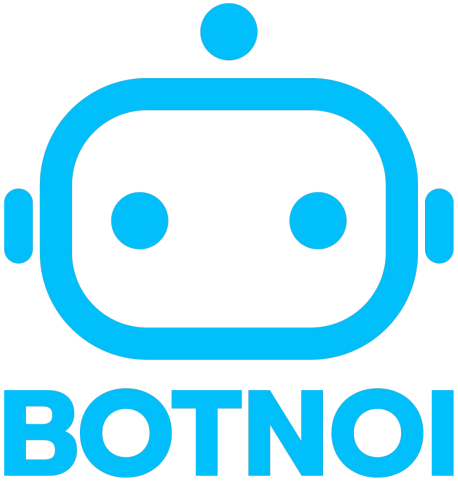
# About myself


## Head of Digital Human

**Witthawin Sripheanpol**

- B.Eng. Institute of Field Robotics, King Mongkut's University of Technology Thonburi
- Project Manager & Lead Lecturer
  iGenius Robot Education, Thailand
- Senior Data Scientist & Machine Learning Engineer
  Botnoi Group, Thailand
- Head of Digital Human
  Botnoi Group, Thailand
- Co-Founder
  FB: HowKnow IKnow

© Botnoi Group




We localise AI

# Agenda

1. What is Predictive Model?
2. How it work?
3. How to do Predictive Model on Production

[Workshop] Streamlit & Production

© Botnoi Group

1

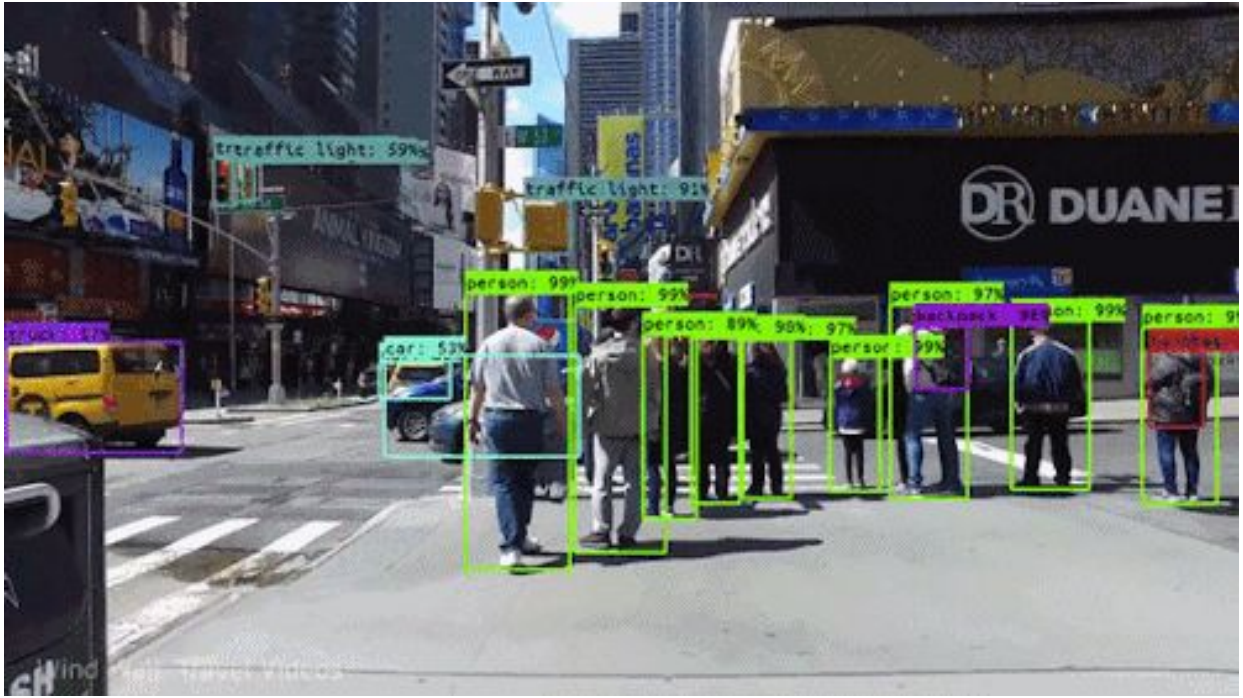
# What Is

# Predictive Model?


© Botnoi Group

# Computer Vision

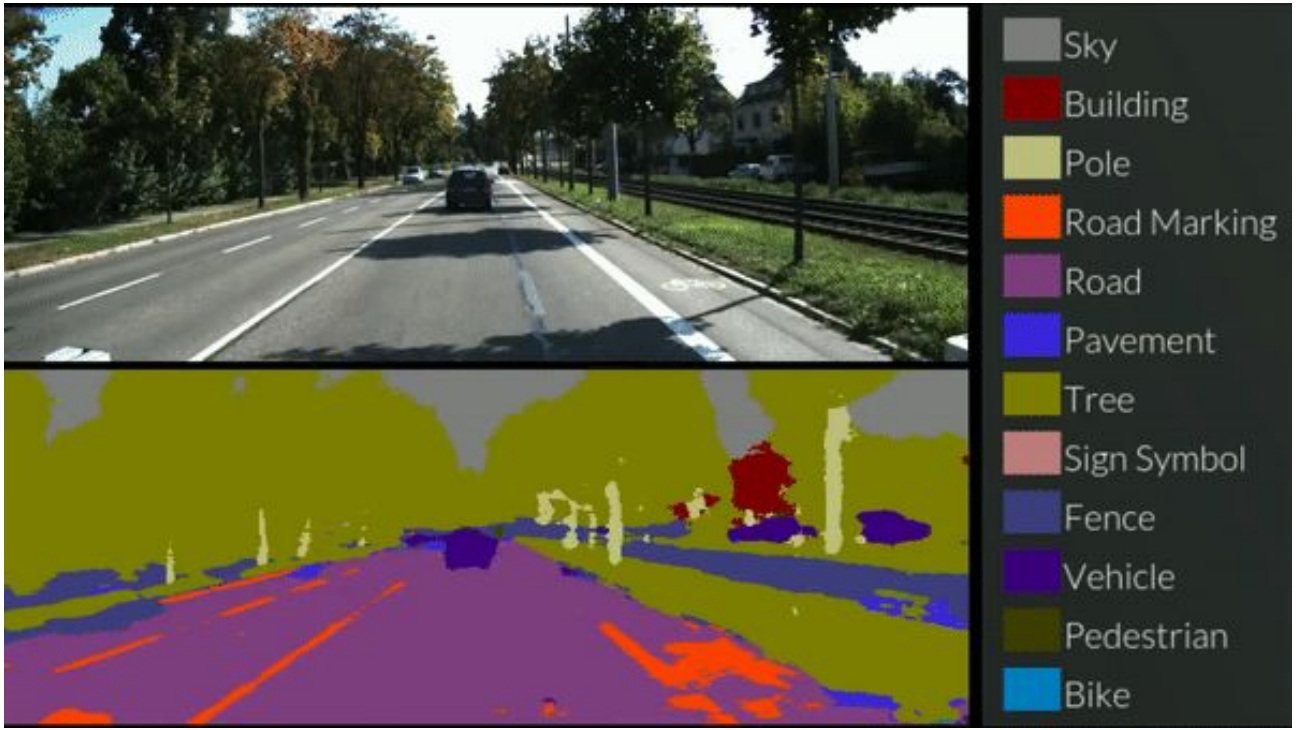
(Object Detection)




© Botnoi Group

# Computer Vision

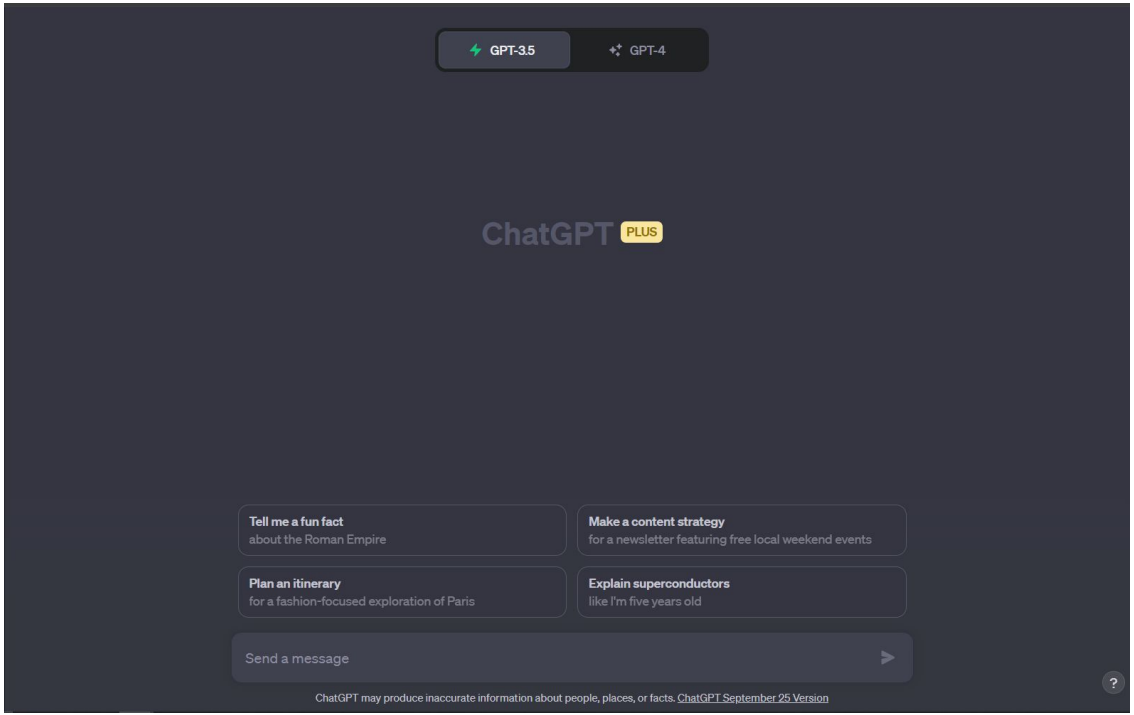
(Image Segmentation)




© Botnoi Group

# Natural Language Processing

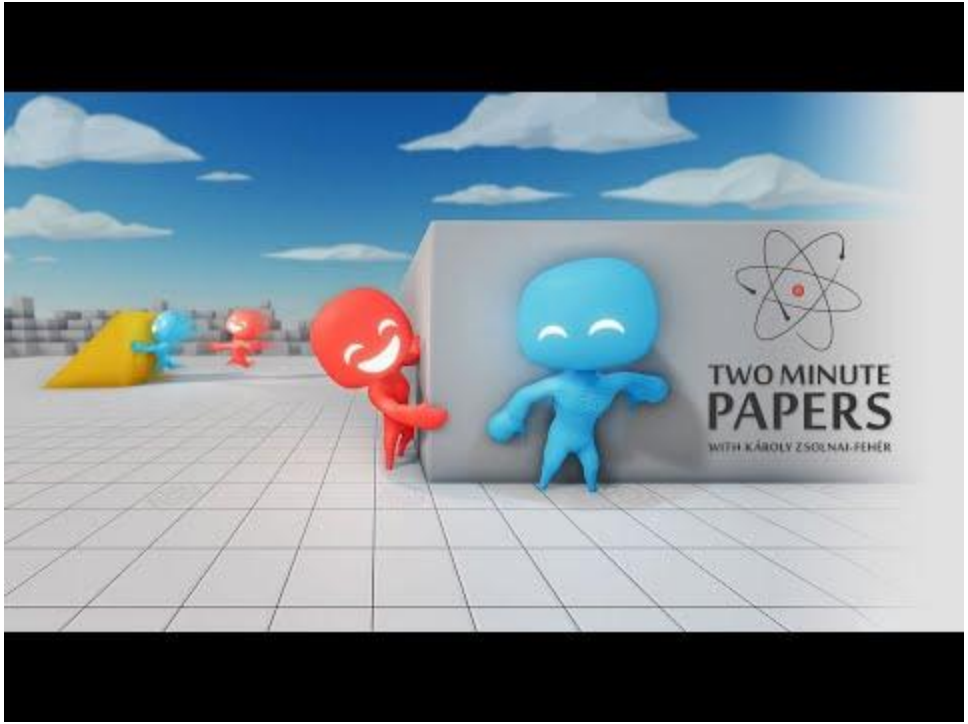
(AI understand text, message)




© Botnoi Group

# Game & Simulation

(AI understand Mechanics and Motion)




© Botnoi Group

# Robotics & Automation

(AI understand Mechanics and Motion)

© Botnoi Group

# Recommendation

(AI understand behavior)

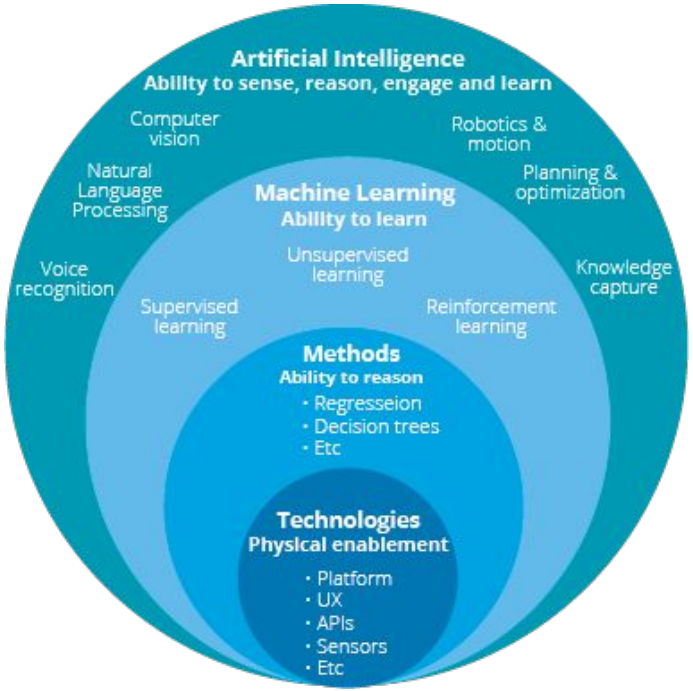


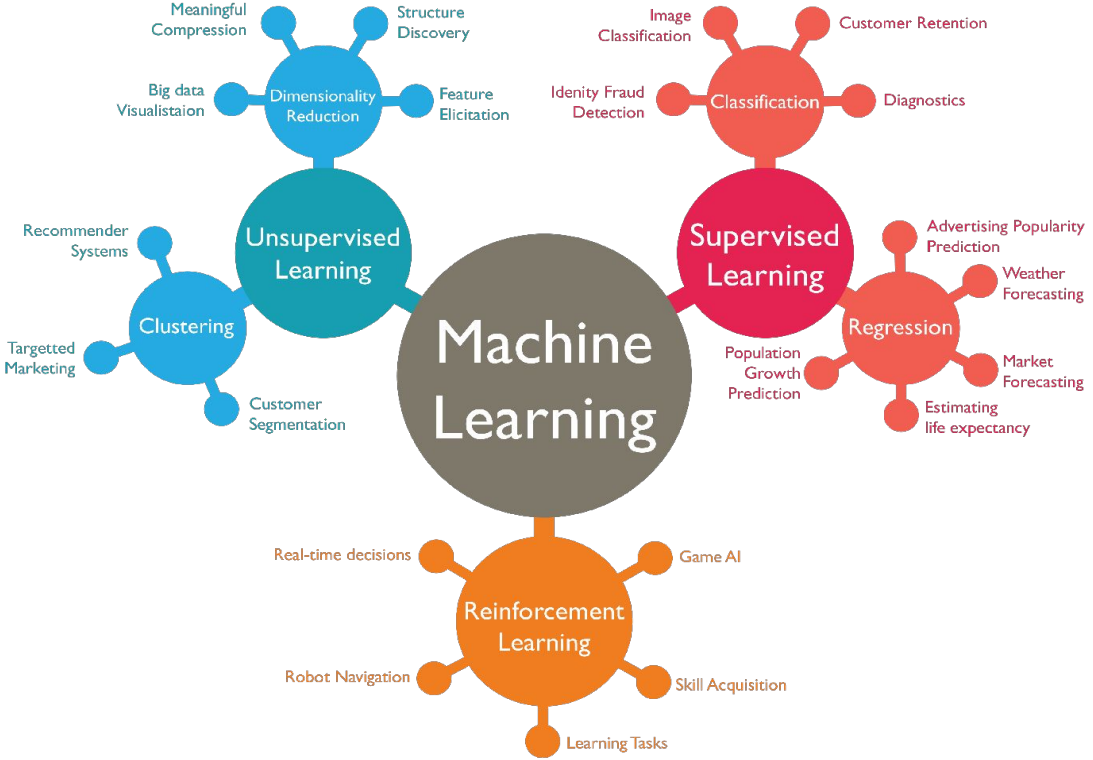
© Botnoi Group

2

# How it work?

# Recap : Machine Learning

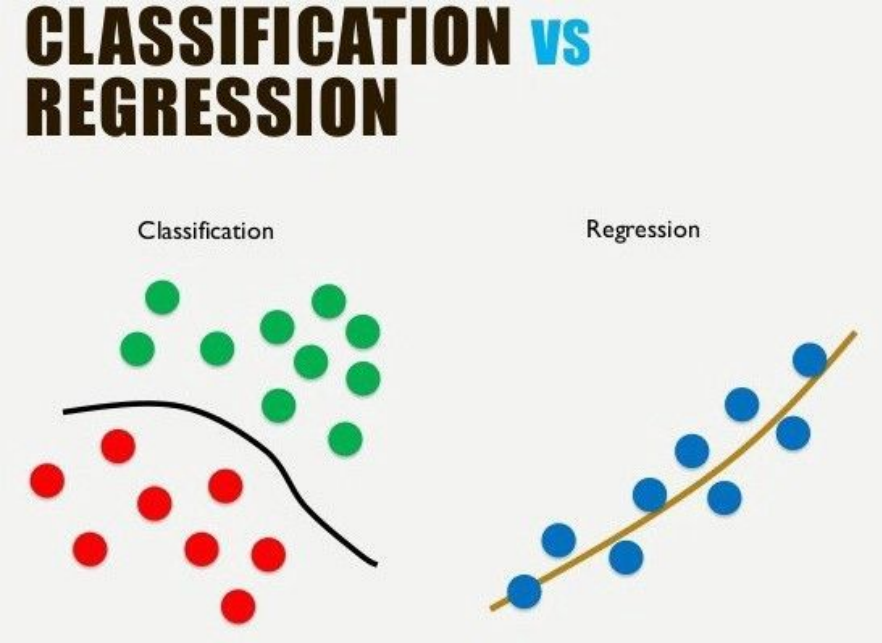



© Botnoi Group

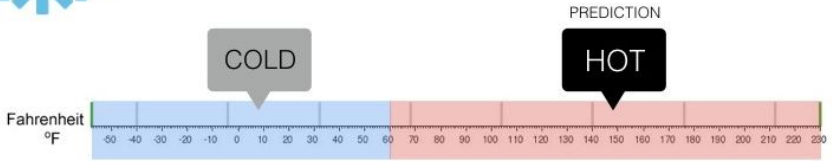
# Supervised Learning

## Classification vs Regression



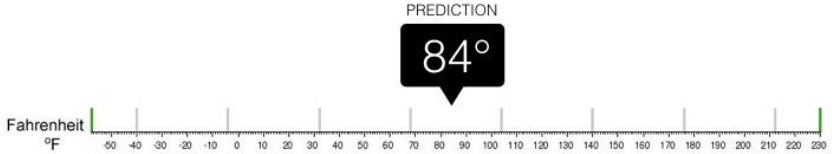

**Classification**

Will it be Cold or Hot tomorrow?




**Regression**

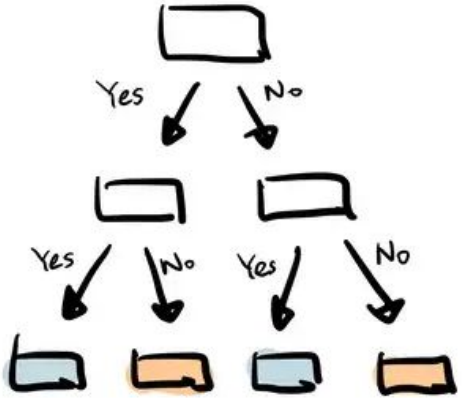
What is the temperature going to be tomorrow?

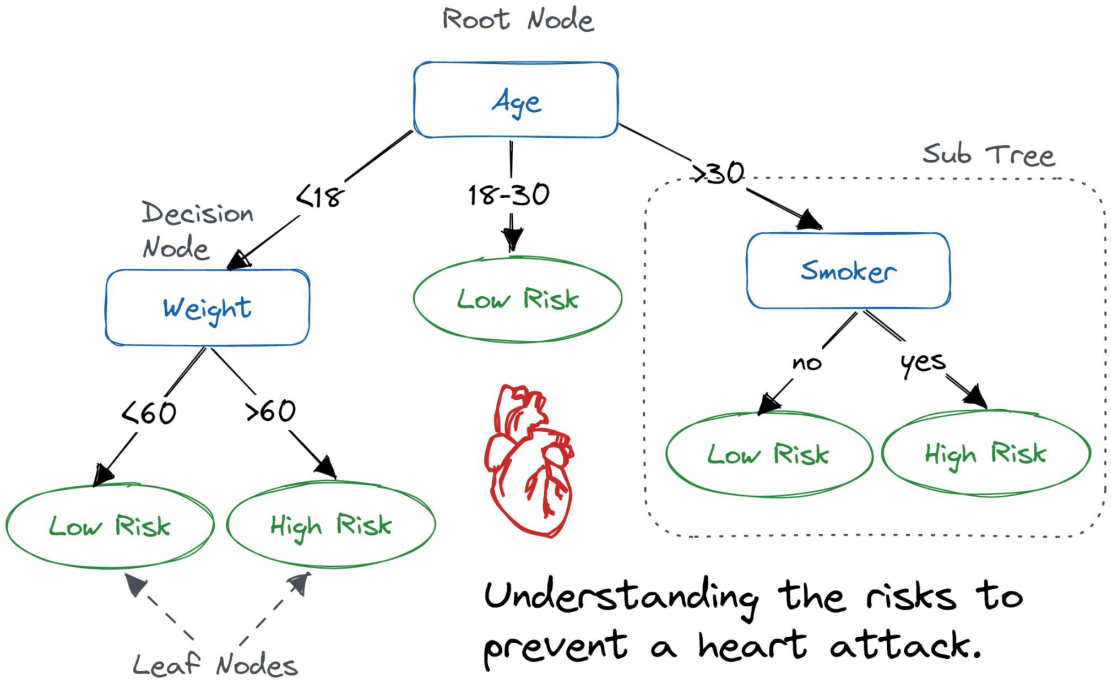



© Botnoi Group

# Classification Algorithm

(Basic Algorithm : Decision Tree)



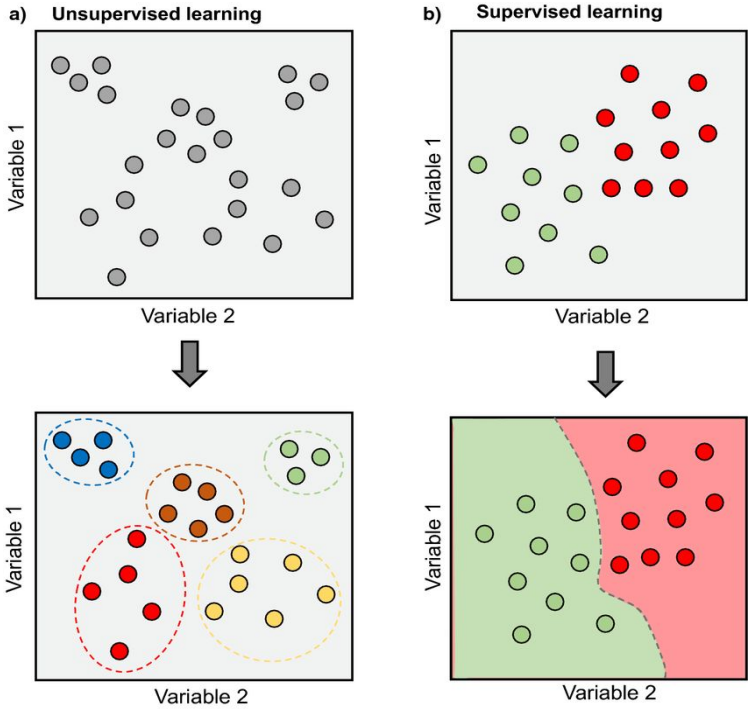



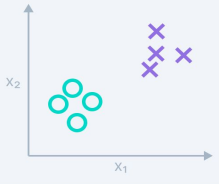
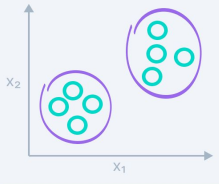

https://scikit-learn.org/stable/supervised_learning.html#supervised-learning

© Botnoi Group

# Unsupervised Learning

(unknown label)



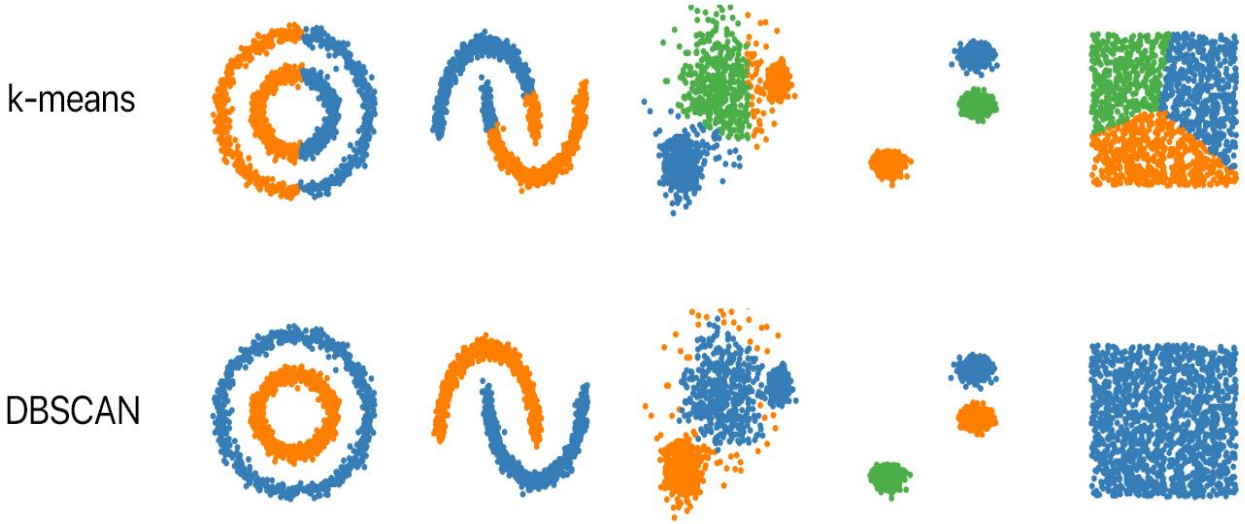

| Supervised learning | Unsupervised learning |
|---|---|
| Input data is labeled | Input data is unlabeled |
| Has a feedback mechanism | Has no feedback mechanism |
| Data is classified based on the training dataset | Assigns properties of given data to classify it |
| Divided into Regression & Classification | Divided into Clustering & Association |
| Used for prediction | Used for analysis |
| Algorithms include: decision trees, logistic regressions, support vector machine | Algorithms include: k-means clustering, hierarchical clustering, apriori algorithm |
| A known number of classes | A unknown number of classes |
|  |  |

V7 Labs

© Botnoi Group

# Clustering Algorithm

(Basic Algorithm for grouping data)

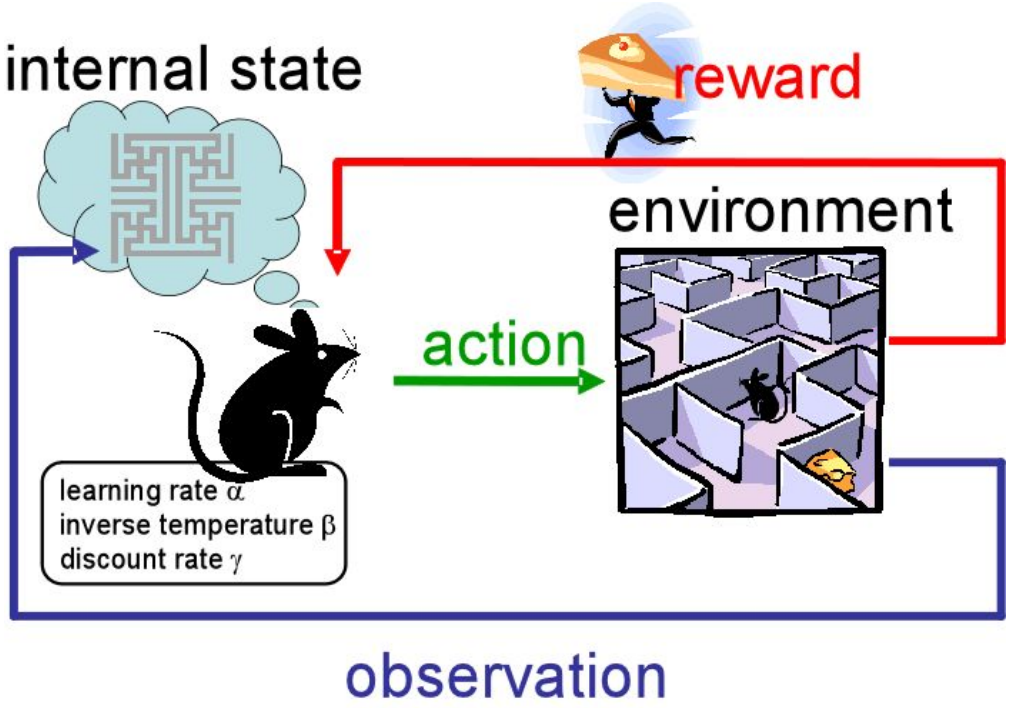



https://scikit-learn.org/stable/modules/clustering.html#clustering

© Botnoi Group

# Reinforcement Learning

(I don't know but I'll maximize reward)




© Botnoi Group

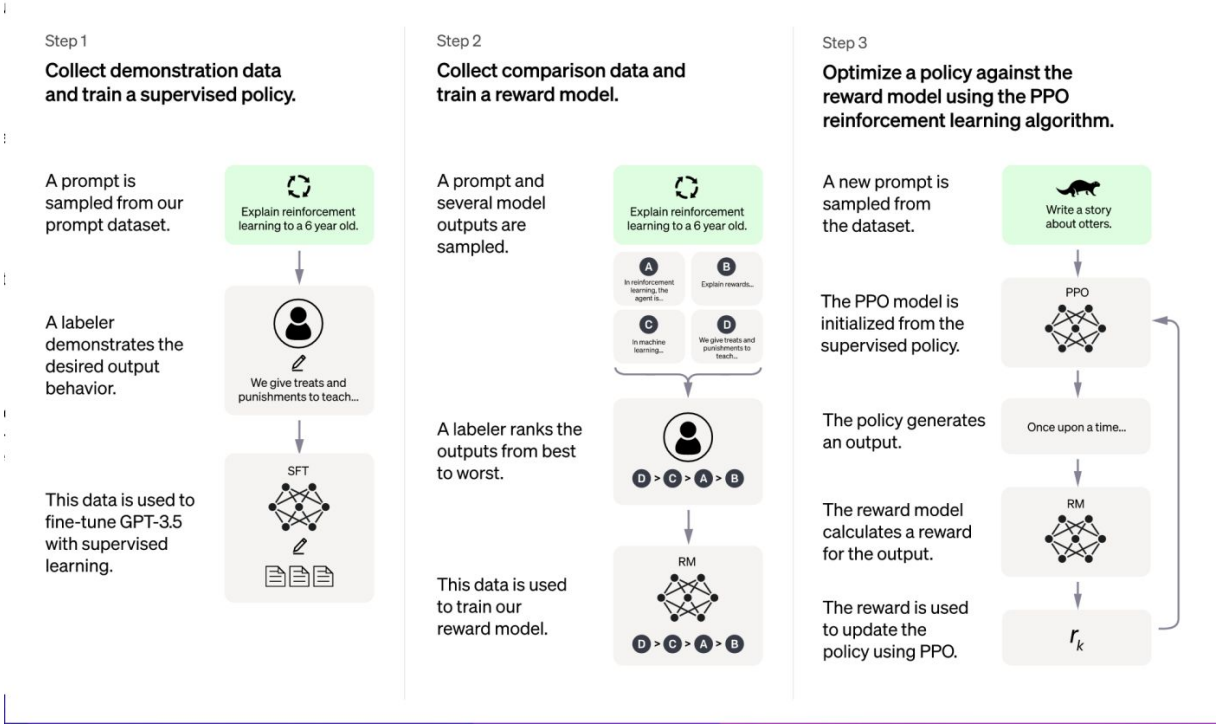
# Reinforcement Learning

(Sample Computing)

© Botnoi Group

# Reinforcement Learning

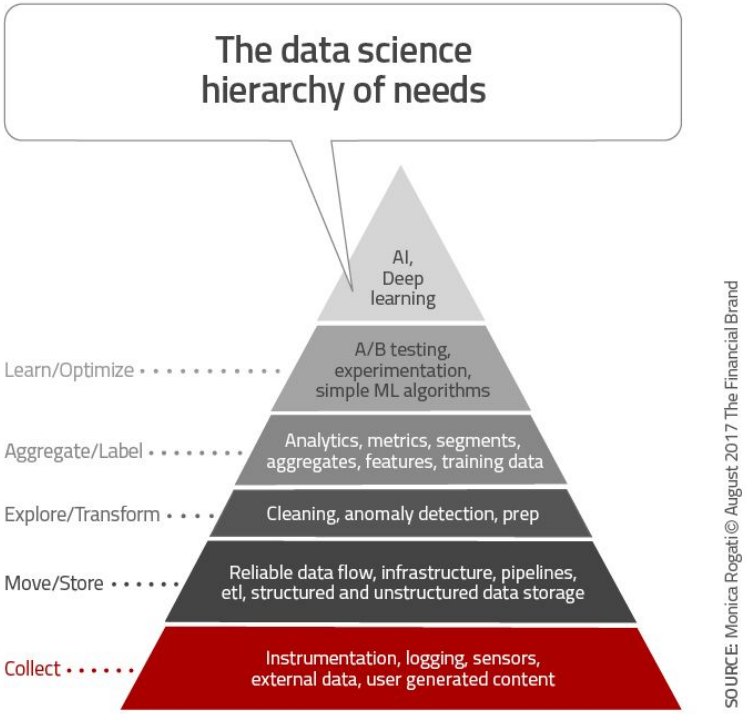
(Deep RL in GPT & Other Large Model)




© Botnoi Group

# Data to AI Application

(AI driven by Data)




© Botnoi Group

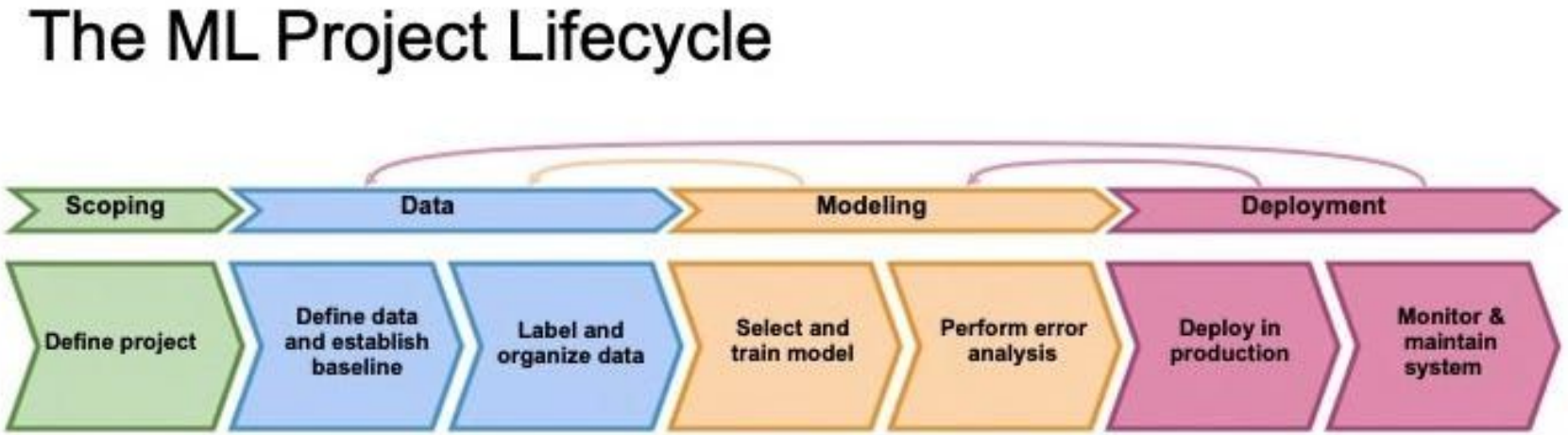
3

# How to do

# Prediction

# Machine Learning Pipeline

(ML & Prediction Lifecycle)




© Botnoi Group

# Problem

(Define Project for solve problem)

© Botnoi Group

# Data Collection

(collect many data)

**Structured data**

**Characteristics**

Predefined data models
Easy to search
Text-based
Shows what's happening

**Resides in**

Relational databases
Data warehouses

**Stored in**

Rows and columns

**Examples**

Dates, phone numbers, social security numbers, customer names, transaction info

**Unstructured data**

**Characteristics**

No predefined data models
Difficult to search
Text, pdf, images, video
Shows the why

**Resides in**

Applications
Data warehouses and lakes

**Stored in**

Various forms

**Examples**

Documents, emails and messages, conversation transcripts, image files, open-ended survey answers

**Semi-structured data**

**Characteristics**

Loosely organized
Meta-level structure that can contain unstructured data
HTML, XML, JSON

**Resides in**

Relational databases
Tagged-text format

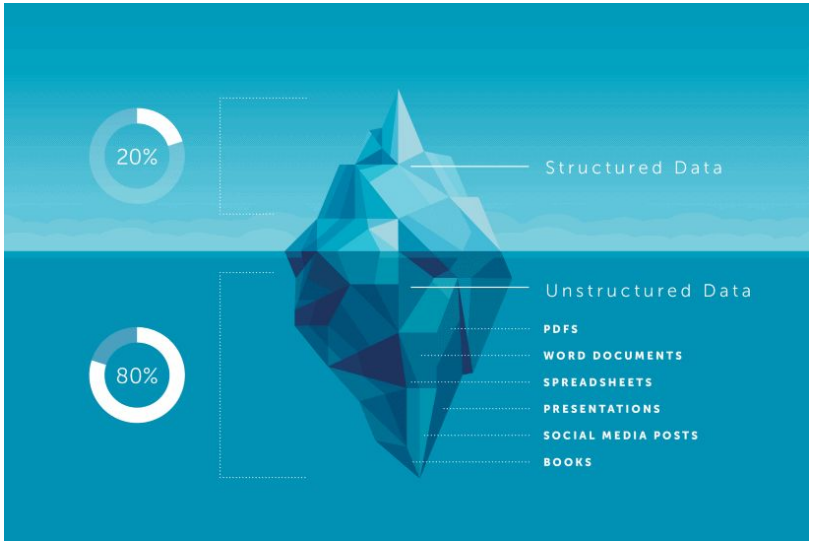
**Stored in**

Abstracts & figures

**Examples**

Server logs, tweets organized by hashtags, emails sorting by folders (inbox; sent; draft)

LEVITY




© Botnoi Group

# Data Collection

(Public Dataset)

**Hugging Face**

Google

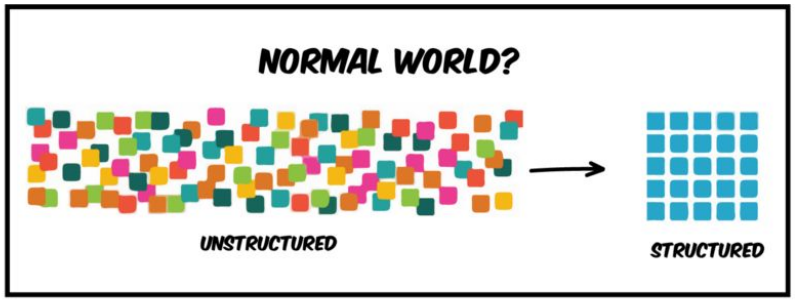
kaggle

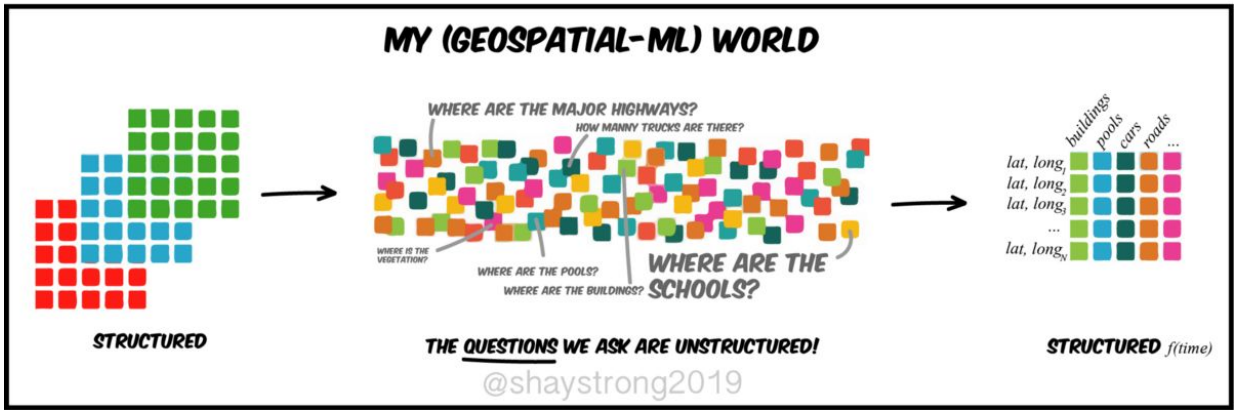
https://pub.towardsai.net/best-datasets-for-machine-learning-data-science-computer-vision-nlp-ai-c9541058cf4f

© Botnoi Group

# Data Preparation (ETL)

(Extract + Transform + Load)

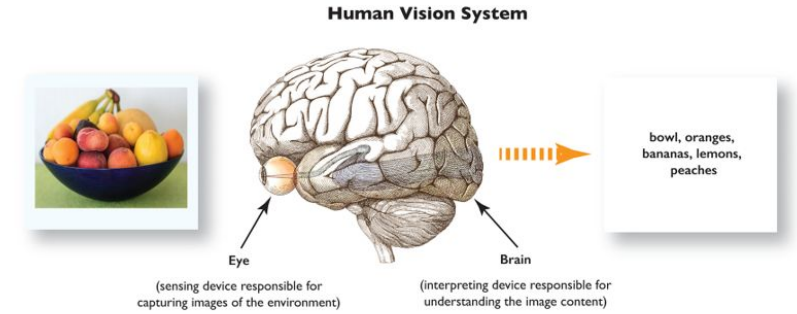


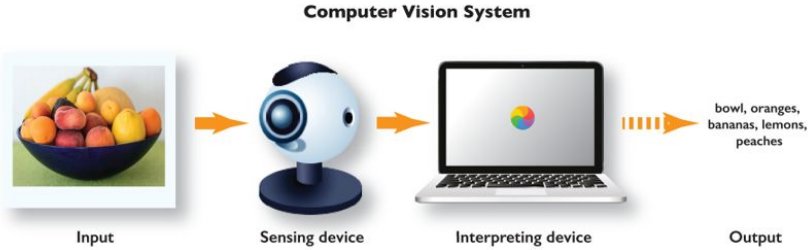




© Botnoi Group

# Computer Vision

(AI understand Media)



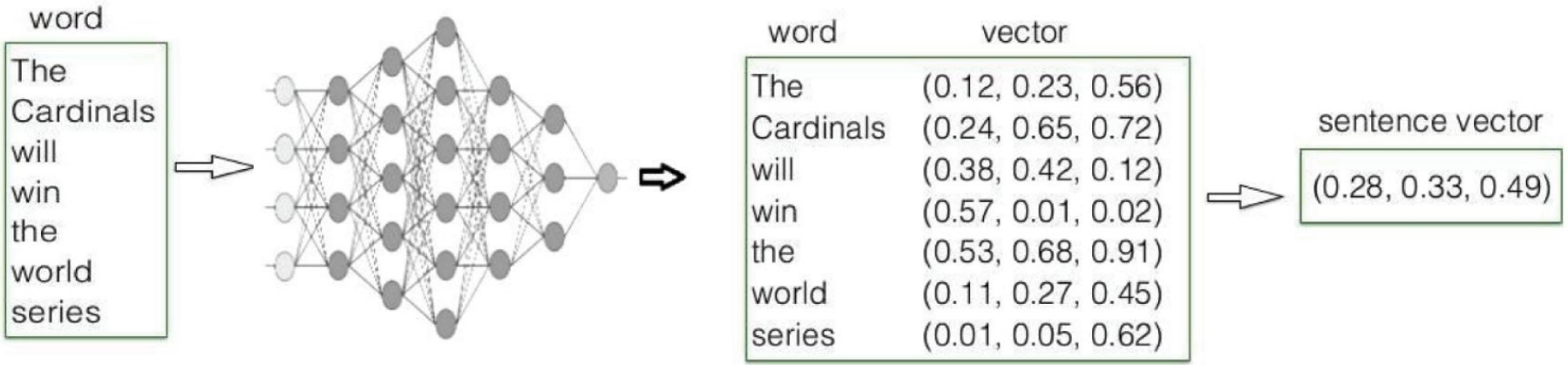




| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0 | 3 | 2 | 5 | 4 | 7 | 6 | 9 | 8 |
| 3 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 1 | 0 | 3 | 2 | 5 | 4 | 7 | 6 |
| 4 | 5 | 3 | 0 | 1 | 2 | 3 | 4 | 5 |
| 7 | 4 | 5 | 2 | 3 | 0 | 1 | 2 | 3 |
| 6 | 5 | 4 | 3 | 2 | 1 | 0 | 3 | 2 |
| 9 | 6 | 7 | 4 | 5 | 2 | 3 | 0 | 1 |
| 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |

© Botnoi Group

# Natural Language Processing

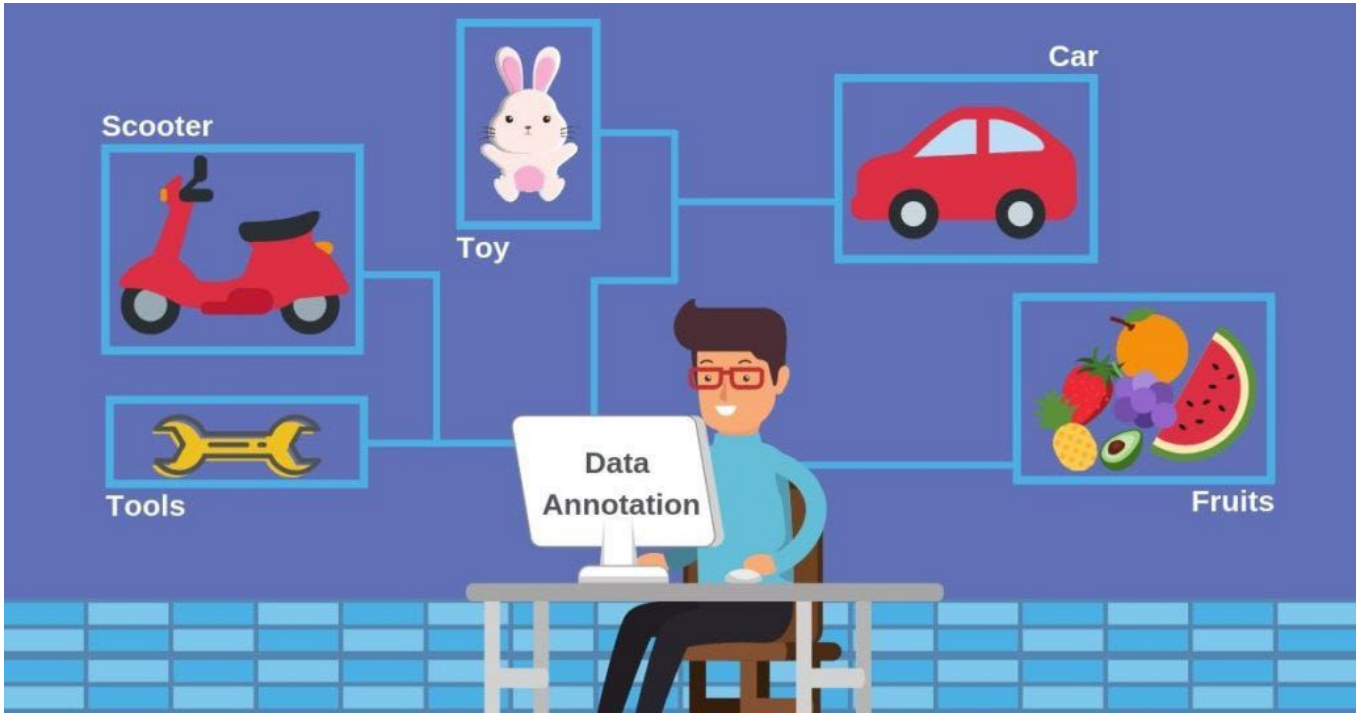
(AI understand text, message)




© Botnoi Group

# Data Labeling

(make choice or answer)

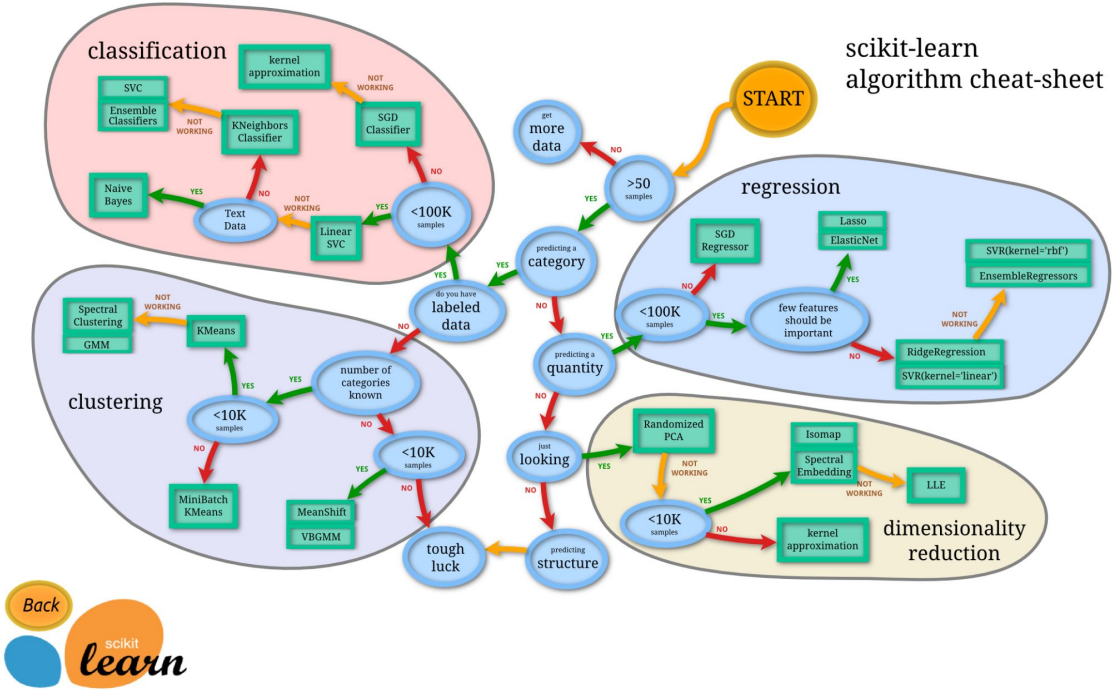



https://www.superannotate.com/blog/best-data-labeling-tools

© Botnoi Group

# Machine Learning Algorithm

(Select Algorithm)

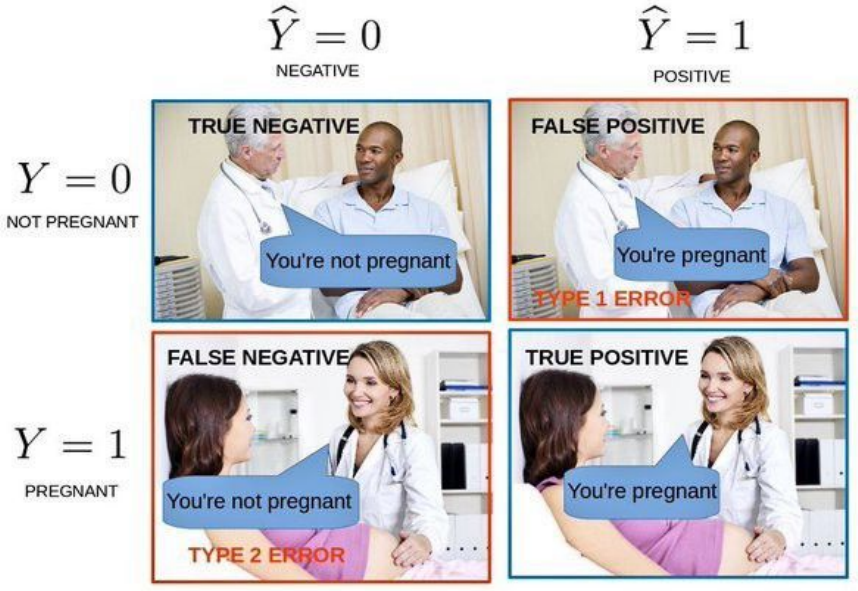



https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

© Botnoi Group

# Performance Classification

(Confusion Matrix)




- True Positive (TP) : Actual=0 & Predict=0
- True Negative (TN) : Actual=0 & Predict=1
- False Positive (FP) : Actual=1 & Predict=0
- False Negative (FN) : Actual=1 & Predict=1

**accuracy (ACC)**

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

**precision or positive predictive value (PPV)**

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} = 1 - \mathrm{FDR}$$

**sensitivity, recall, hit rate, or true positive rate (TPR)**

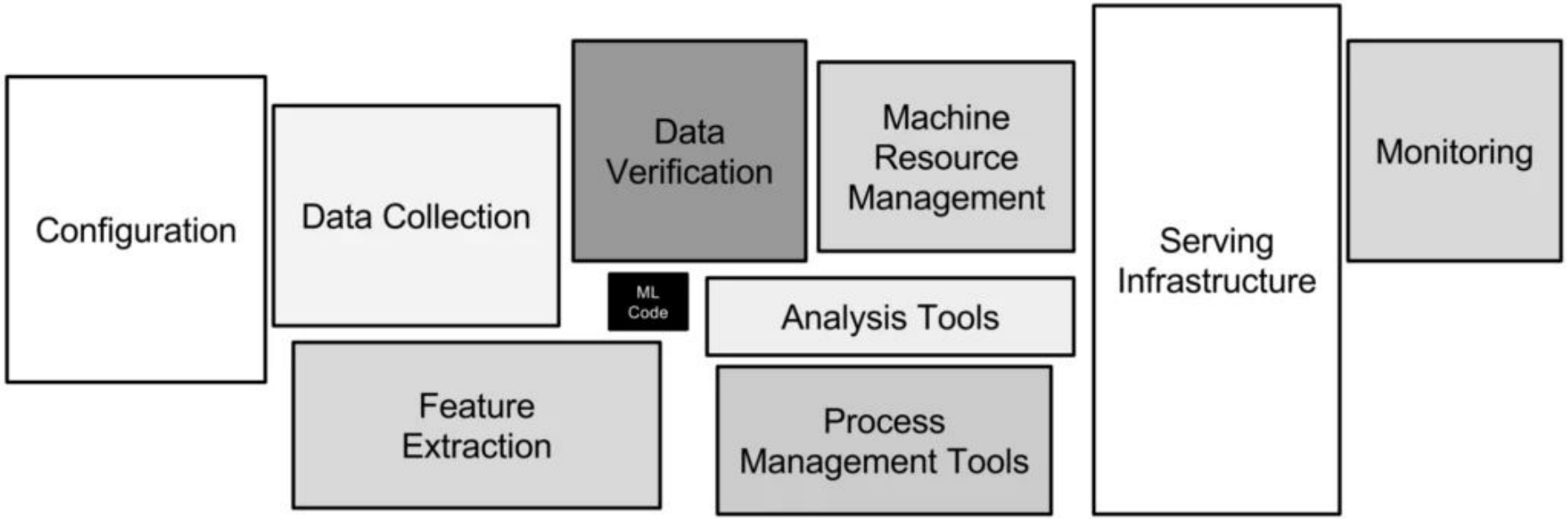
$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = 1 - \mathrm{FNR}$$

https://en.wikipedia.org/wiki/Confusion_matrix

© Botnoi Group

# Machine Learning Pipeline

(ML is 10% of Project)




© Botnoi Group

# Predictive Production

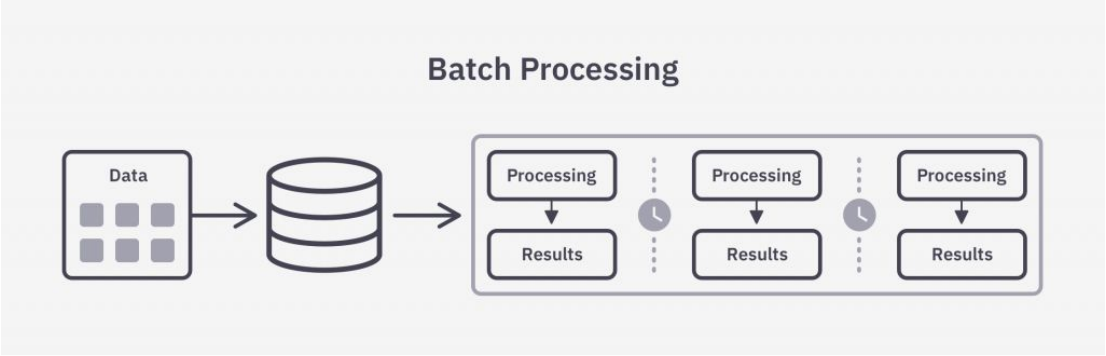
**“The study needs to match application”**

© Botnoi Group

# Bonus

# Batch vs Stream Processing

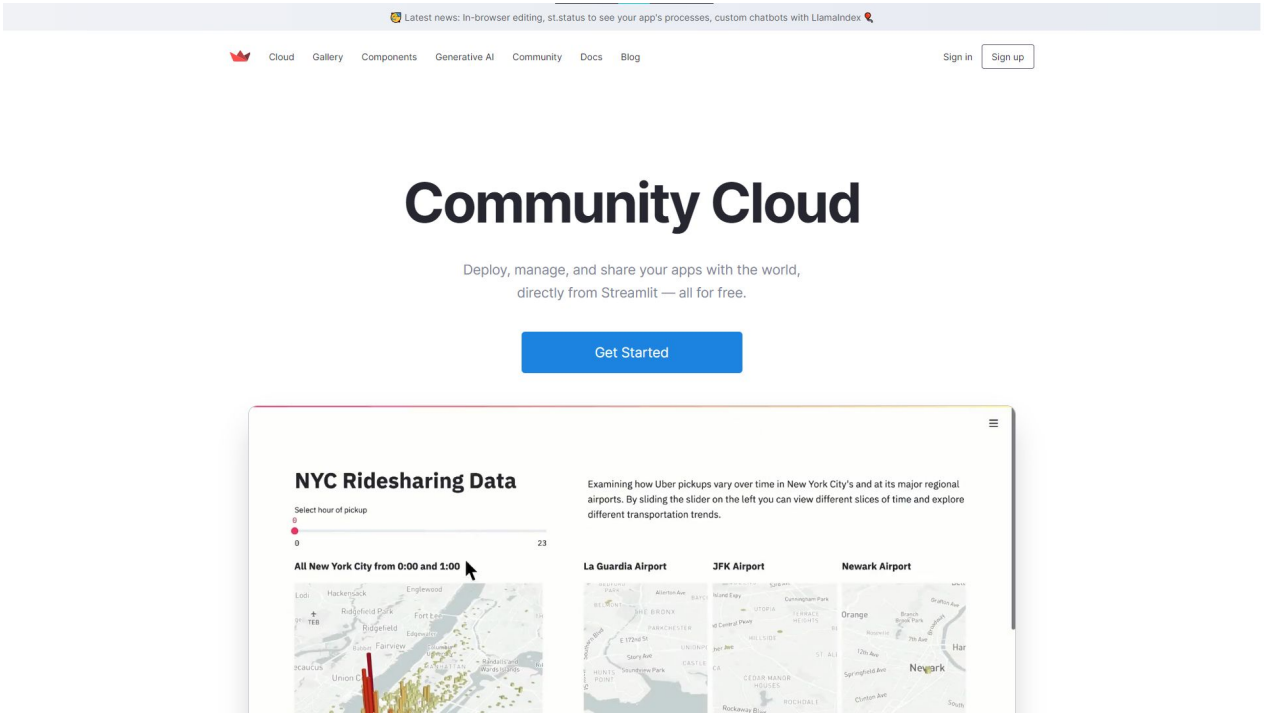
(life cycle production)




© Botnoi Group

# Test Production

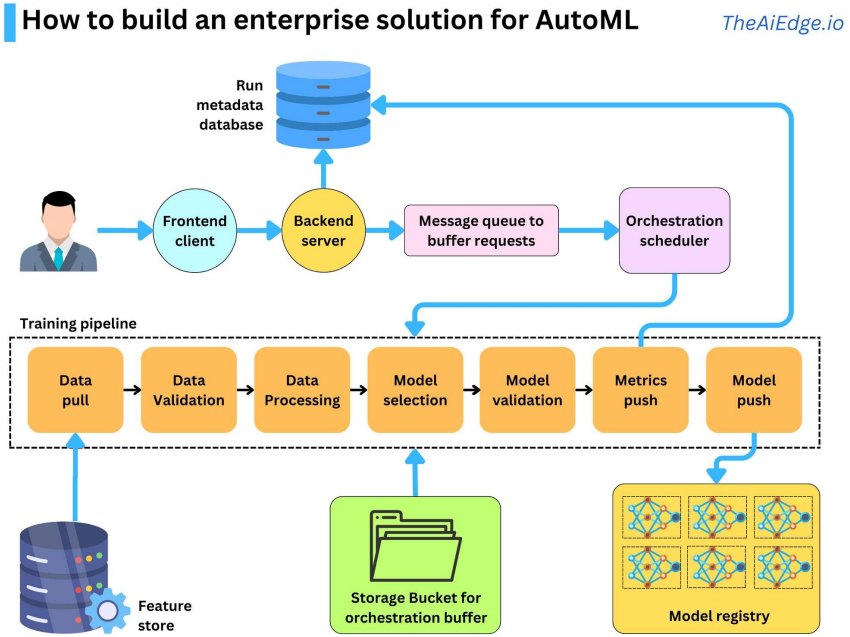
(POC on Streamlit Cloud)




© Botnoi Group

# AutoML

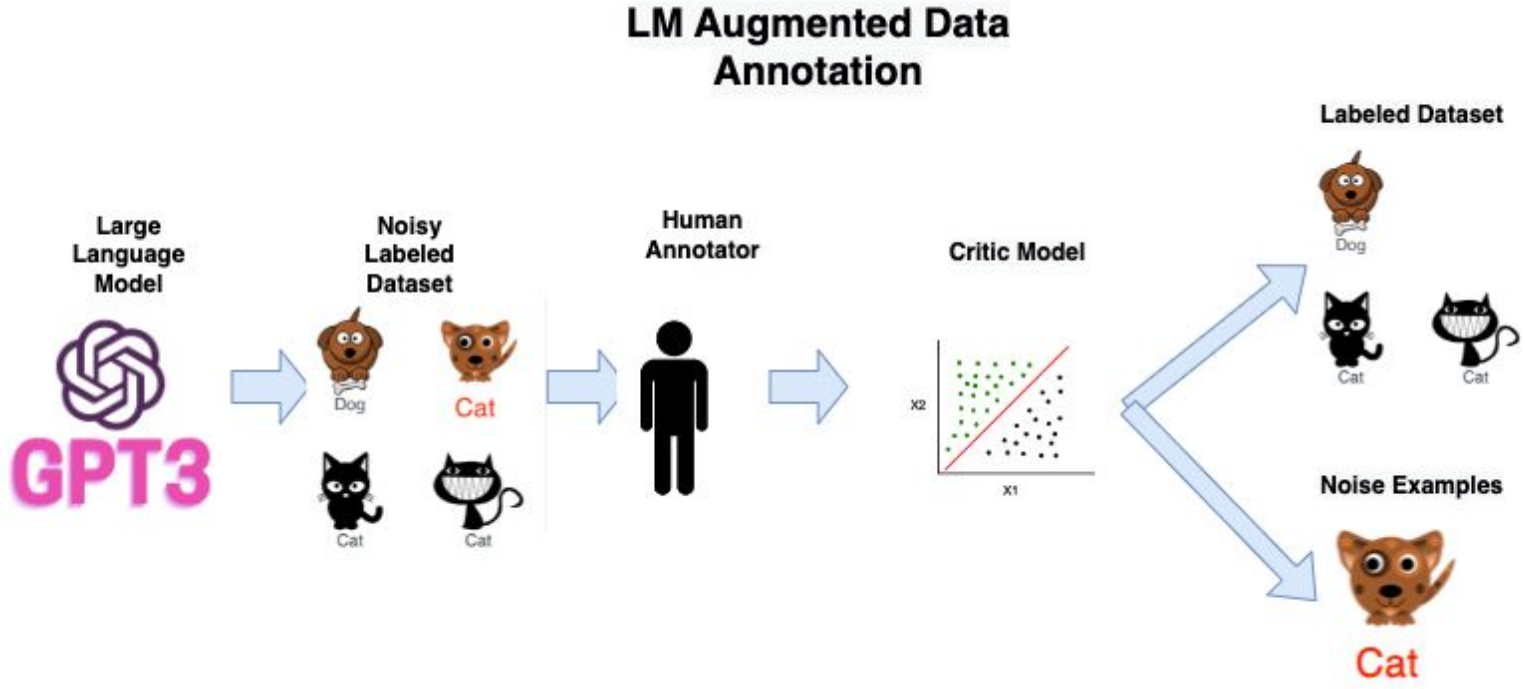
(Advance ML Pipeline)




© Botnoi Group

# How LLM Labeling Data

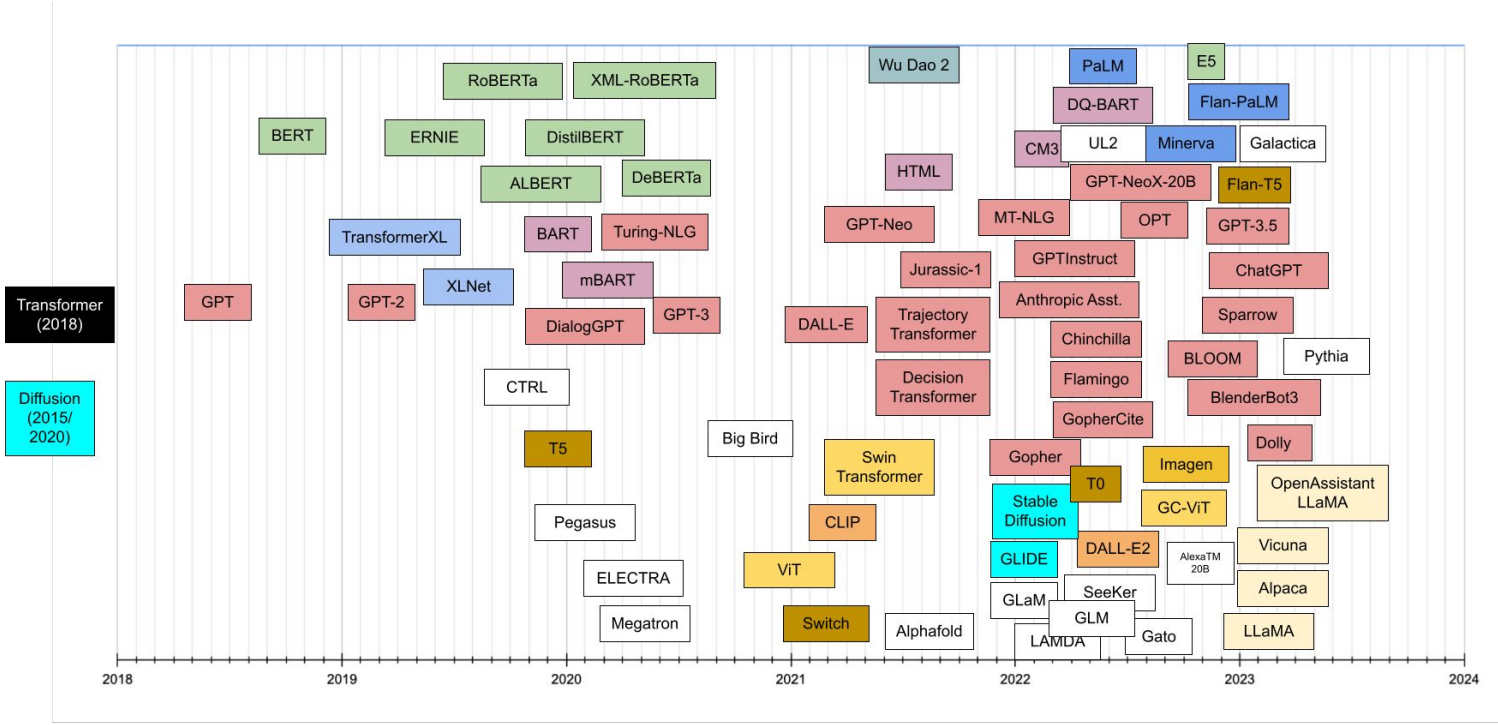
(self-supervised learning)




© Botnoi Group

# Trend of AI Generative

(AI Trend 2023)




© Botnoi Group