APPLIED MACHINE LEARNING

Computational Data Science Program

Natei Ermias Benti (PhD)

natei.ermias@aau.edu.et

+251917089988

March, 2024

OUTLINE

- **Machine learning (ML) Basic Concepts**
  - Supervised ML
    - Classification
    - Regression
  - Unsupervised Learning
    - Clustering
    - Dimensionality Reduction
- **Data Preparation and Feature Engineering**
  - Data Acquisition
  - Data Preprocessing
  - Methods to impute missing values
  - Outlier/Anomaly Detection
  - Feature Engineering
  - Feature Selection
  - Overfitting/under-fitting
  - bias/variance trade-off
  - Learning Curve
  - Mean Removal
  - MinMax Scaling
  - Binarization
  - Label Encoder
- **Regression**
  - Introduction
  - Cost function and Gradient Descent
  - Basic idea: Regression and its applications
  - Linear regression
  - Types of Errors and Better regression models
  - Polynomial Linear Regression
  - Regularization (Rigde and LASSO regression)
- **Classification**
  - Introduction
  - Types and applications of classification algorithms
  - Logistic regression
  - Random Forest
  - k-Nearest Neighbors
  - Naive Bayes
  - SVM
  - Decision trees
  - Gini Index and Information Gain
  - Ensembe techniques (XGboost, AdaBoost and RankBoost)

Unsupervised Machine learning
- Why unsupervised learning and its Applications
- Clustering algorithms
- K-means clustering
- Dimensionality Reduction
- Principal Component Analysis (PCA)

Neural Networks

TOOLKITS LAB

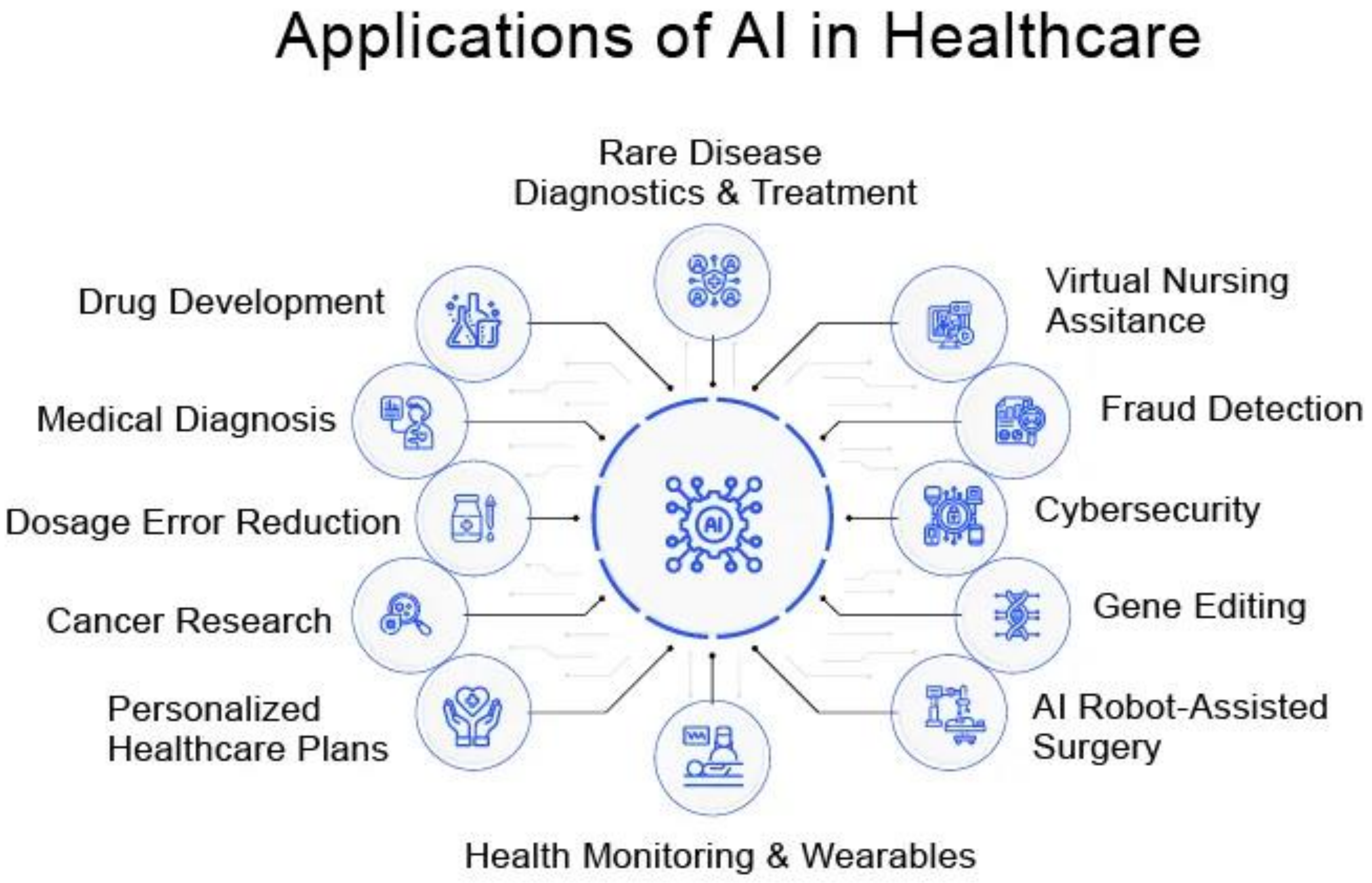
- Toolkit Lab
  - Anaconda / miniconda
  - Jupiter lab / Jupiter notebook
  - Markdown
  - ipython
  - scikit-learn
  - Pytroch
  - Tensorflow
  - google colab
  - binder

INTRODUCTION . . .

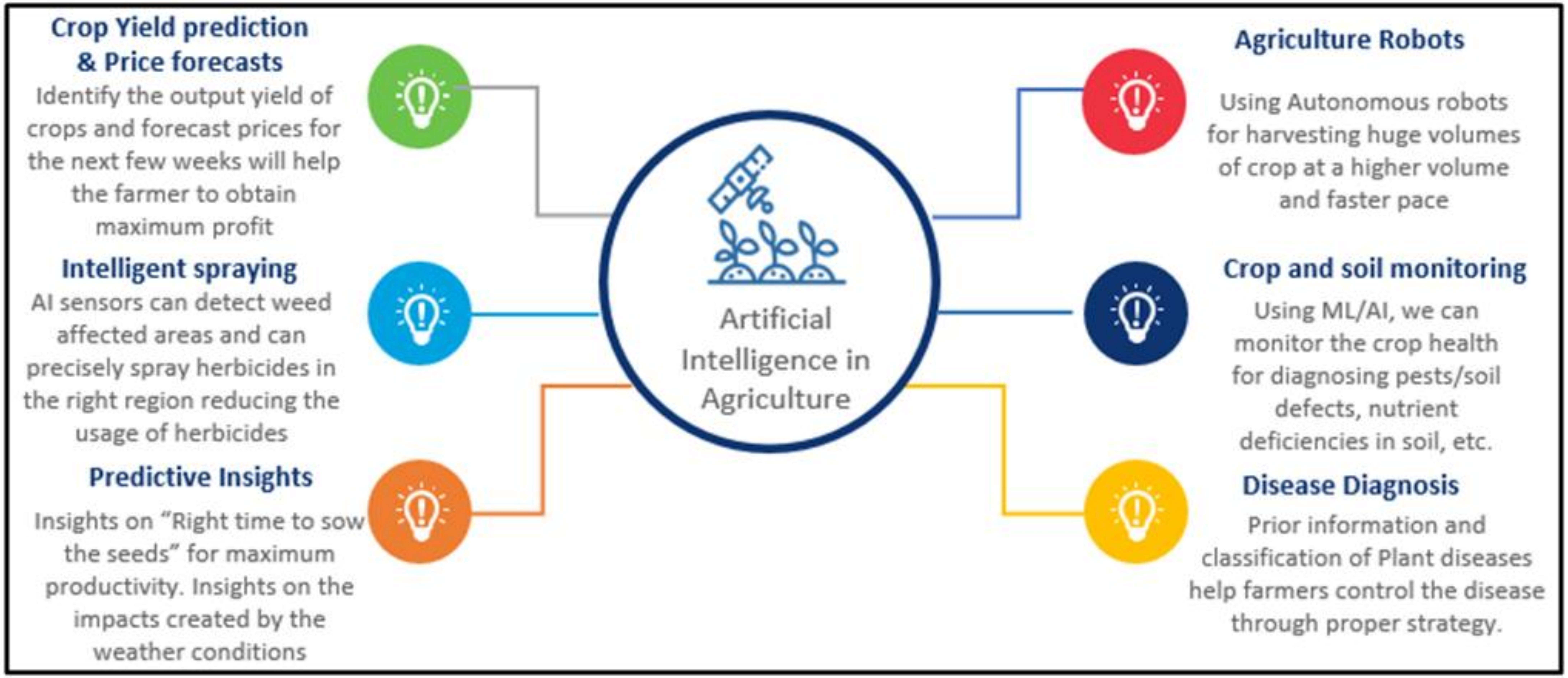
- AI was first coined in 1956 by John McCarthy at the Dartmouth Conference.
- It is defined as the science and engineering of making intelligent machines in a sense AI is a technique of getting machines to work and behave like human.

INTRODUCTION . . .

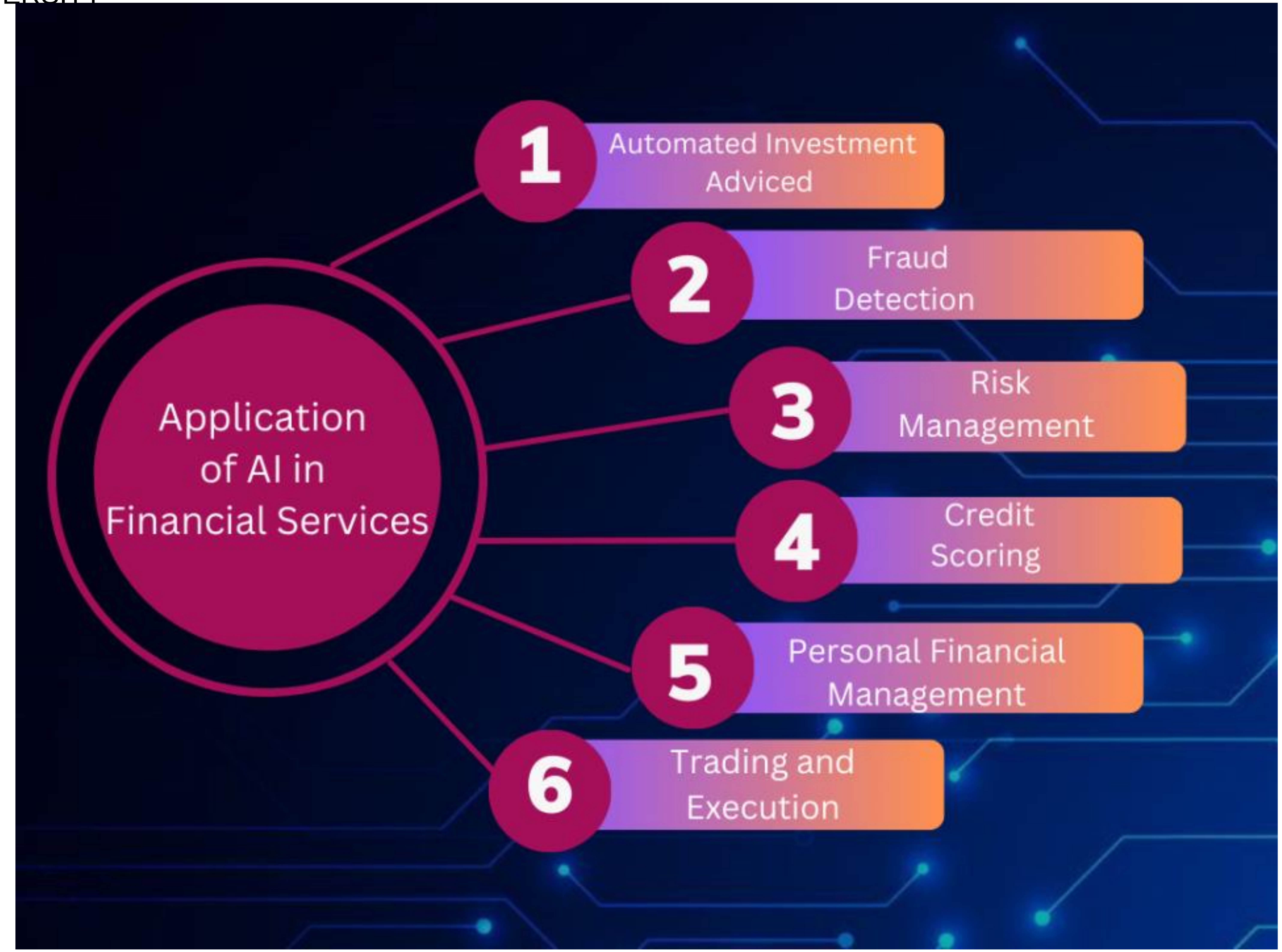
INTRODUCTION . . .



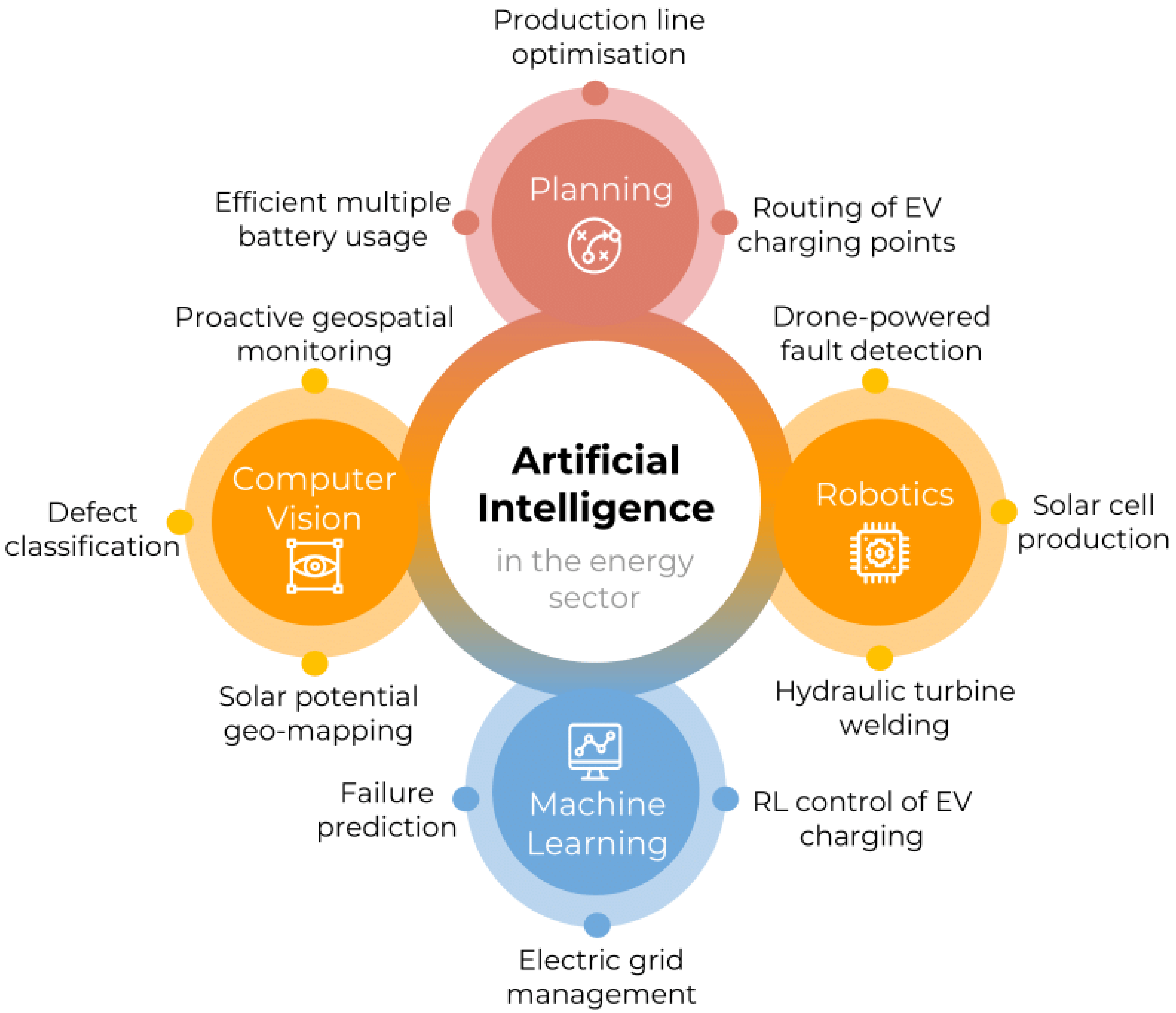
INTRODUCTION . . .

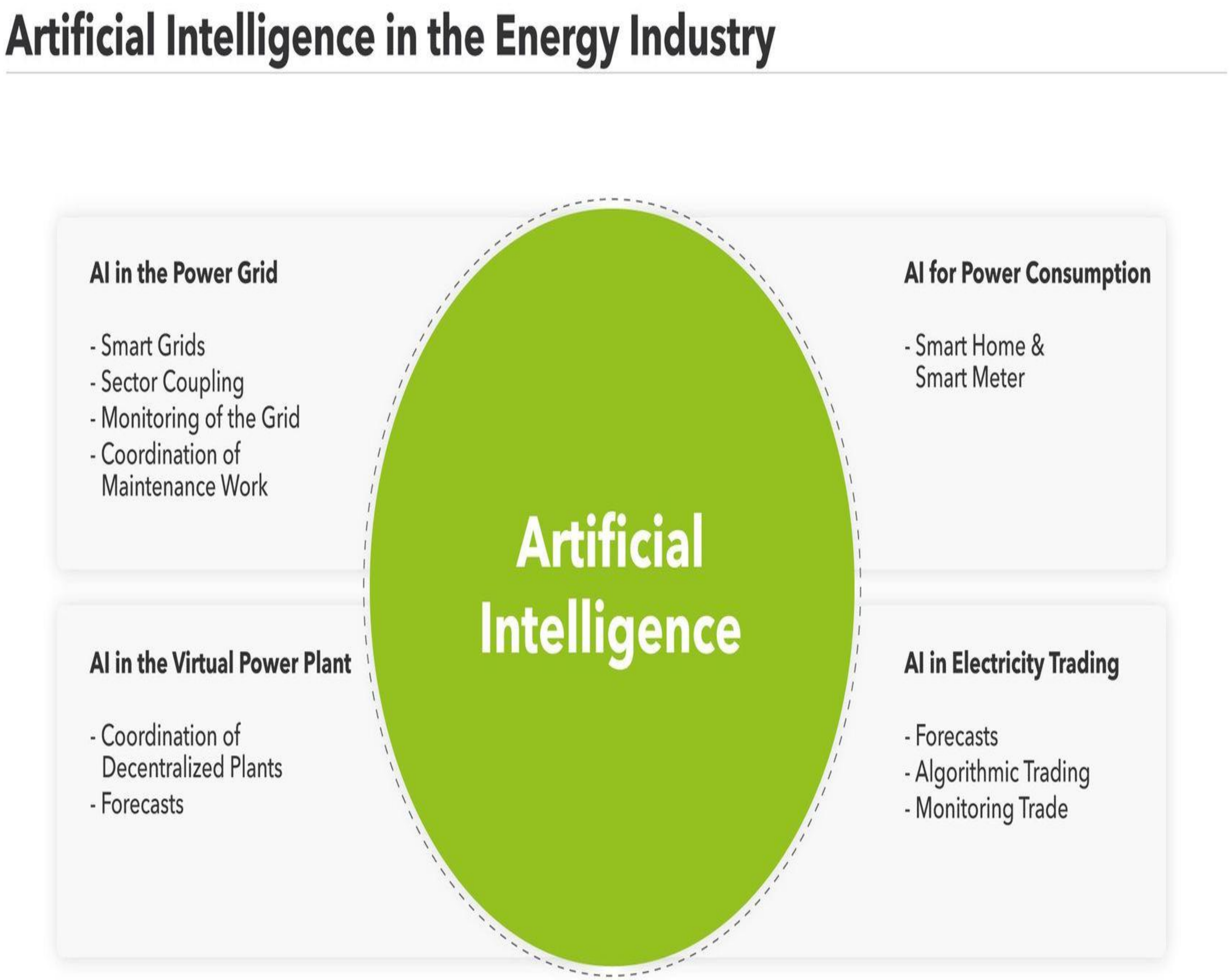


INTRODUCTION . . .

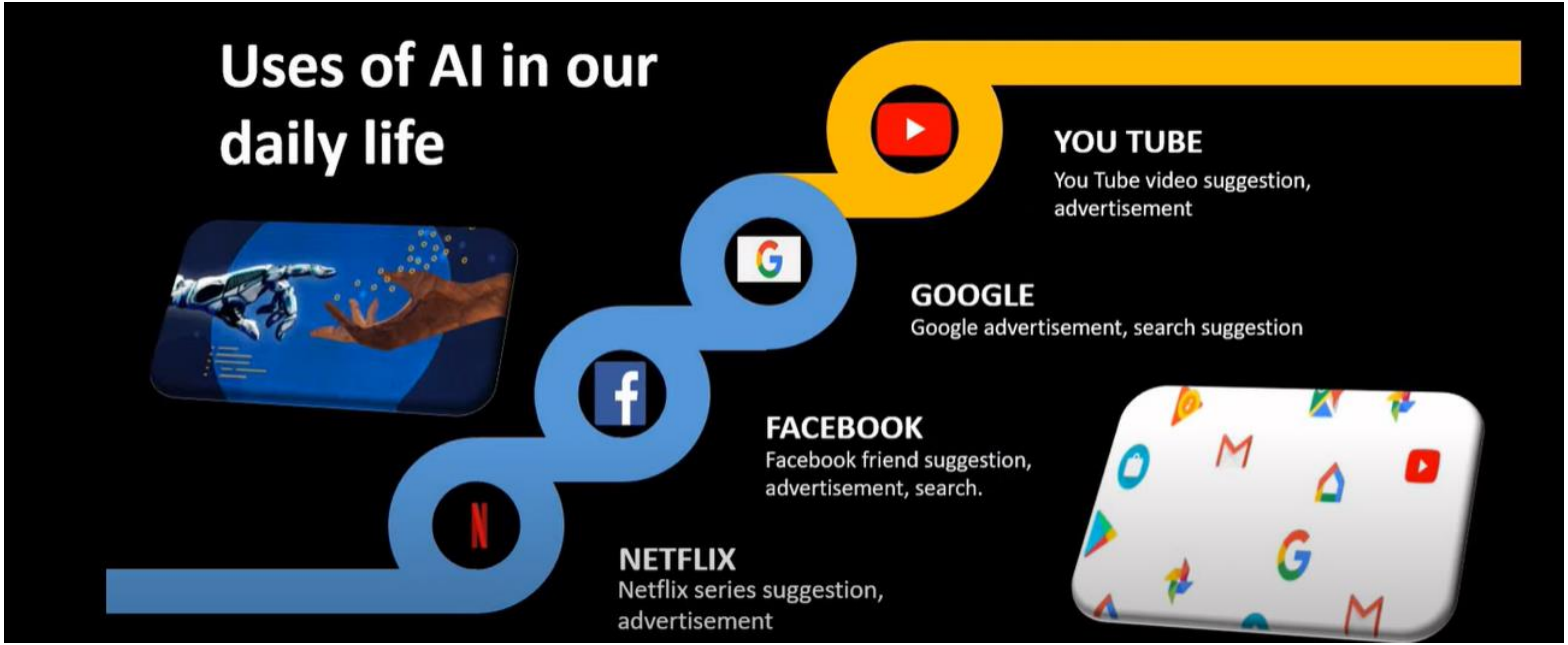


INTRODUCTION . . .





INTRODUCTION . . .



INTRODUCTION . . .

Artificial Intelligence

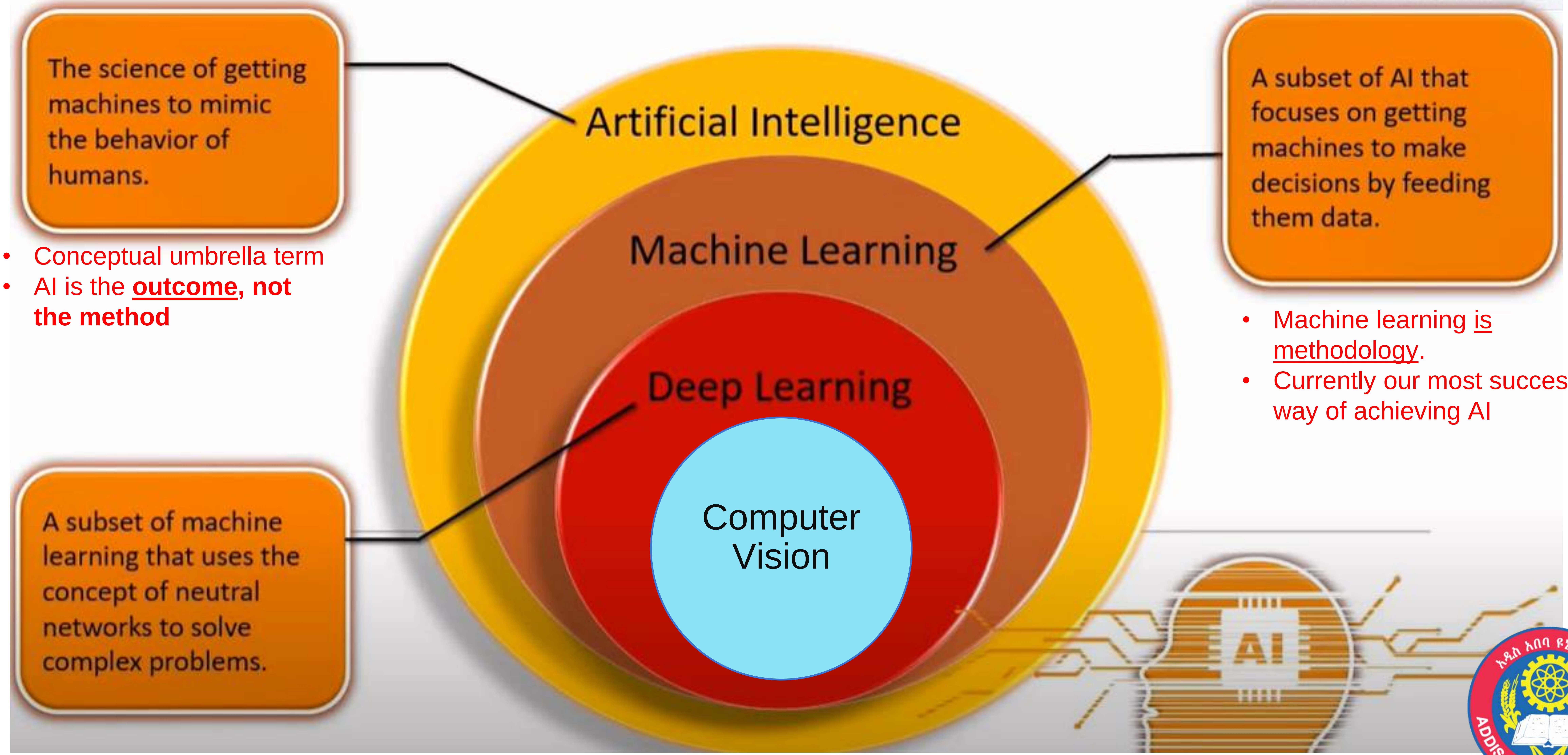
Common Misconception

Machine learning

Deep learning

People often tend to think that AI, ML, DL are same thing since they have common application

INTRODUCTION . . .



- Conceptual umbrella term
- AI is the **<u>outcome</u>, not the method**

- Machine learning <u>is methodology</u>.
- Currently our most successful way of achieving AI

INTRODUCTION . . .

- AI can be broadly classified into different subfields, and two major categories are natural language processing (NLP) and computer vision.

NLP:

- NLP focuses on the interaction between computers and human language.
- It involves the development of algorithms and models that enable machines to understand, interpret, and generate human language.
- NLP applications include machine translation, sentiment analysis, chatbots, speech recognition, and text summarization.

Computer Vision:

- Computer vision is the field of AI that enables machines to interpret and make decisions based on visual data.
- It involves the development of algorithms and models to enable machines to understand images and videos.
- Computer vision applications include image recognition, object detection, facial recognition, autonomous vehicles, and medical image analysis.

Artificial Intelligence (AI)

AI

- ❑ AI focuses on creating intelligent machines that can perform human-like tasks.
- ❑ AI involves developing algorithms and software that simulate human cognitive functions.
- ❑ AI can perform tasks such as learning, problem-solving, decision-making, and perception.

❑ AI systems can operate in different ways.

1. Rule-based systems rely on predetermined rules to make decisions.
2. Machine learning systems can learn and adapt from data.

Different Types of AI

Types of AI based on characteristics and applications

Narrow or weak AI

- is designed to perform a specific task or set of tasks.
- Examples of narrow AI include speech recognition, image recognition, and chatbots.
- Narrow AI is the most common form of AI today.

General or strong AI

- is designed to perform any intellectual task that a human can do.
- This type of AI would have human-like cognitive abilities, such as problem-solving and reasoning.
- General AI does not yet exist, but it is a goal for many researchers in the field.

Artificial super intelligence

- is an AI that surpasses human intelligence in all areas.
- This type of AI would be capable of solving complex problems and making decisions that humans cannot.
- Artificial super-intelligence is a theoretical concept, but its development raises ethical and societal concerns.

AI …

Types of AI based on memory

Reactive machines

- **Reactive machines** react to stimuli from the environment but do not have memory or past experiences.

Limited memory machines

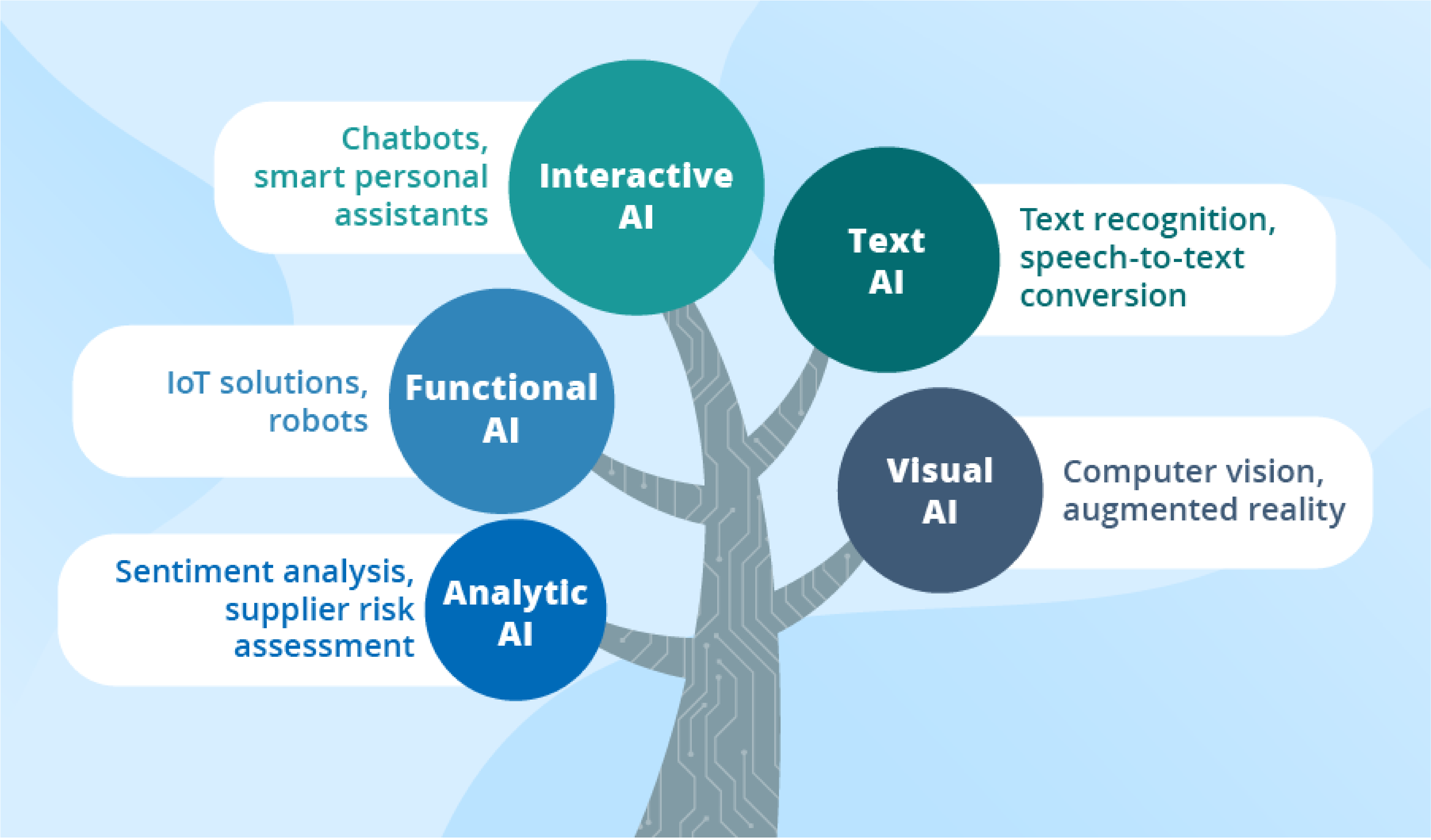
- **Limited memory machines** can make decisions based on past experiences and current data.

Self-aware AI

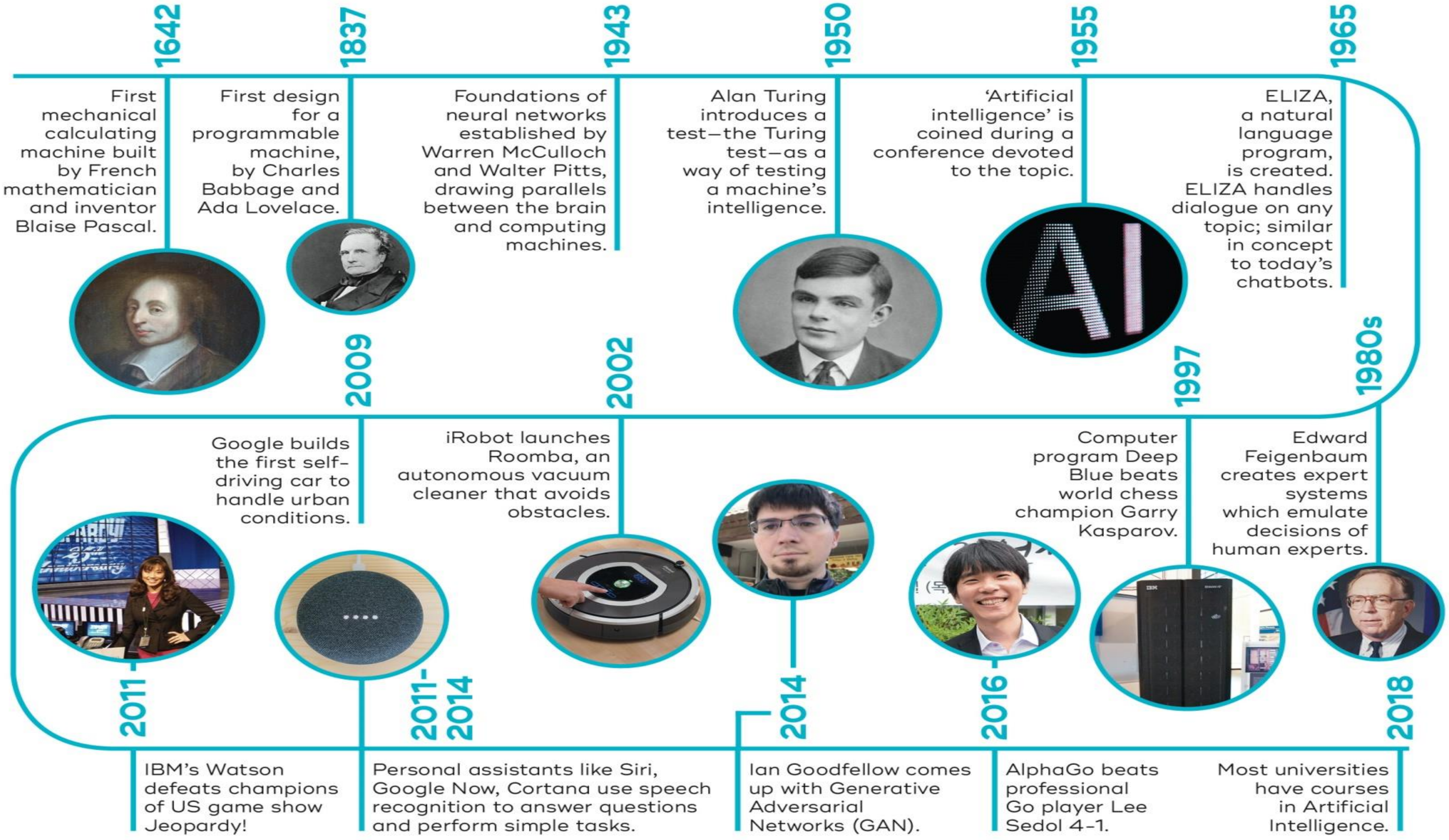
- **Self-aware AI** is a hypothetical concept that involves machines with consciousness and emotions.

- Understanding the different types of AI is important in developing appropriate applications and addressing ethical and societal concerns.
- As AI continues to evolve, it is important to consider the potential implications of each type of AI and how it can be developed in a responsible and ethical manner.

AI …


Chatbots,
smart personal
assistants
Interactive
AI
Text
AI
Text recognition,
speech-to-text
conversion
IoT solutions,
robots
Functional
AI
Visual
AI
Computer vision,
augmented reality
Sentiment analysis,
supplier risk
assessment
Analytic
AI

History of AI



History of AI …

- In the 21st century, advancements in AI have led to the development of deep learning and natural language processing technologies.
- These technologies have enabled AI systems to achieve human-like performance in tasks such as image recognition and language translation.

- Along with progress and success, the field of AI has also faced challenges and ethical concerns.
- These include concerns about bias in AI algorithms, the impact of AI on employment and the workforce, and the potential for misuse of AI in areas such as surveillance and warfare.

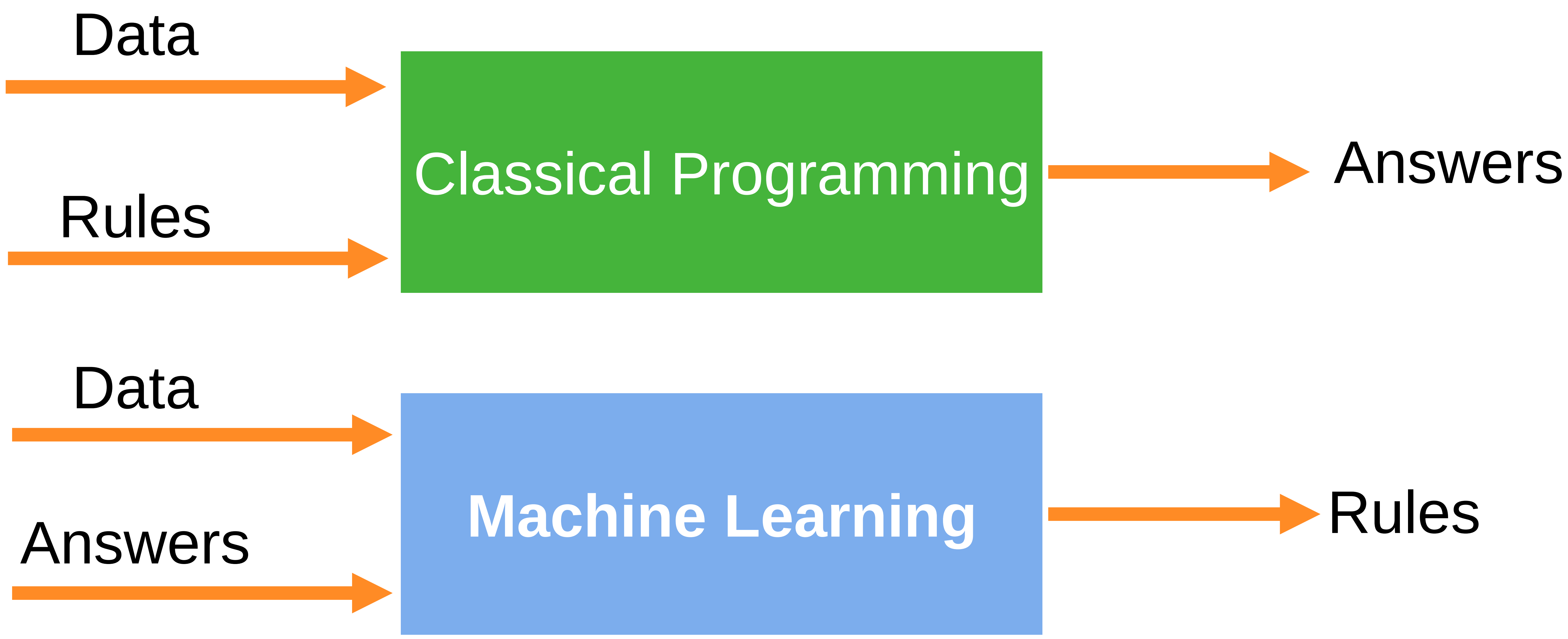
- Despite the challenges and concerns, AI remains a rapidly growing field with the potential to revolutionize the way we live and work.
- As the history of AI has shown, continued research and development are necessary to unlock its full potential and address its ethical implications.

Machine Learning

- Arthor Samuel(1959): Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed
- Tom Mitchell(1998): A computer program is said to learn from Experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improve with experience E.

ML

Machine Learning …



- At a high level, ML systems look a tons of data and come up with rules to predict outcomes for unseen data.
- Fundamentally, machine learning involves building mathematical models to help understand data.

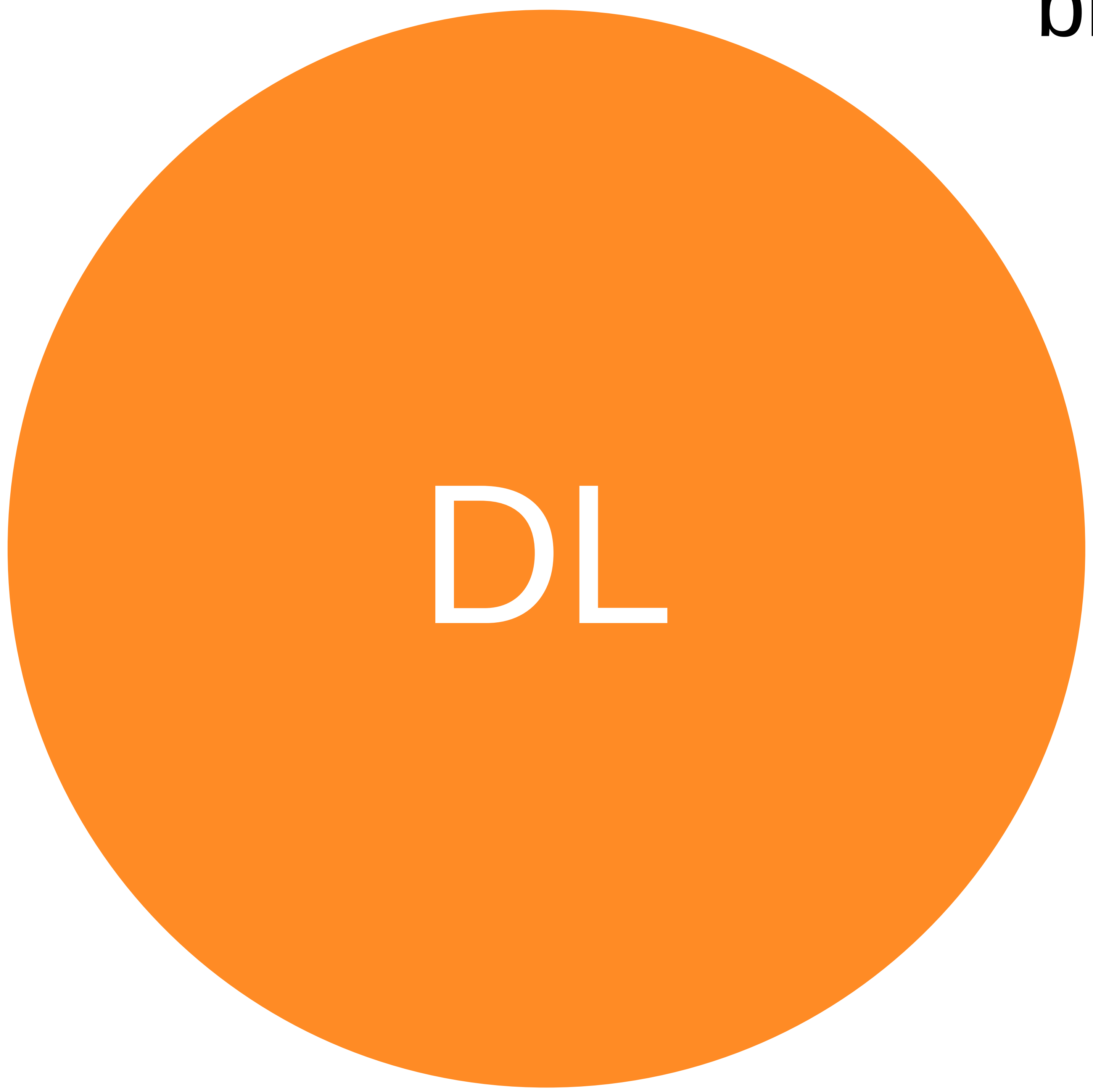
MACHINE LEARNING …

- ML systems learn how to combine input to produce useful predictions on never-before-seen data.
- A label is the thing we're predicting—the y variable in simple linear regression. The label could be the future price of wheat, the kind of animal shown in a picture, the meaning of an audio clip, or just about anything.
- A feature is an input variable—the x variable in simple linear regression. A simple machine learning project might use a single feature, while a more sophisticated machine learning project could use millions of features, specified as:

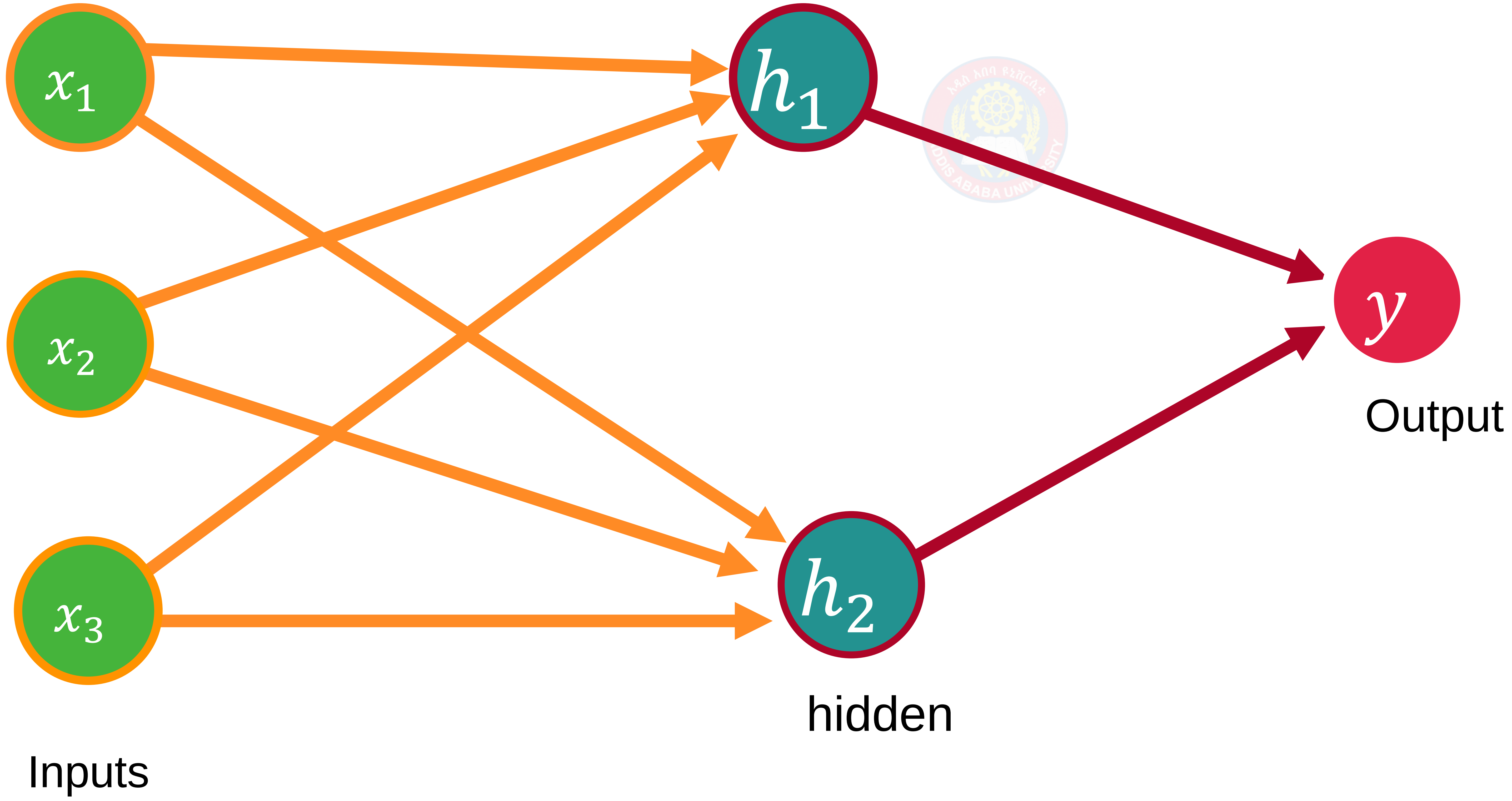
$$x_1, x_2, \cdots, x_N$$

ML

Deep LEARNING



- Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.



DEEP LEARNING …

- A linear layer applied by a linear transformation as:

$$y = wx + b \qquad (1.1)$$

where w and b are weights and bias respectively which are learnable parameters in the neural network

DL

MACHINE LEARNING MODEL

- A model is a mathematical representation of the relationship between features and labels in a dataset, which can be used to make predictions on new, unseen data.

- The process of creating or learning the model is referred to as training.
- During training, the model is presented with labeled examples of data, and the algorithm adjusts the model's parameters to gradually learn the relationship between the features and labels.
- In other words, the model is optimized to make accurate predictions based on the training data.

- Once the model has been trained, it can be used for inference, which means applying the model to new, unlabeled examples to make predictions.
- In this phase, the model takes the features of the new example as input and produces an output, which is a predicted label. For example, if the model is trained to detect spam emails, during inference, it can predict whether an incoming email is spam or not by analyzing its features.

MACHINE LEARNING MODEL

- The ML modeling process insists a modeling phase of four stages:

1 **Feature engineering and model selection**

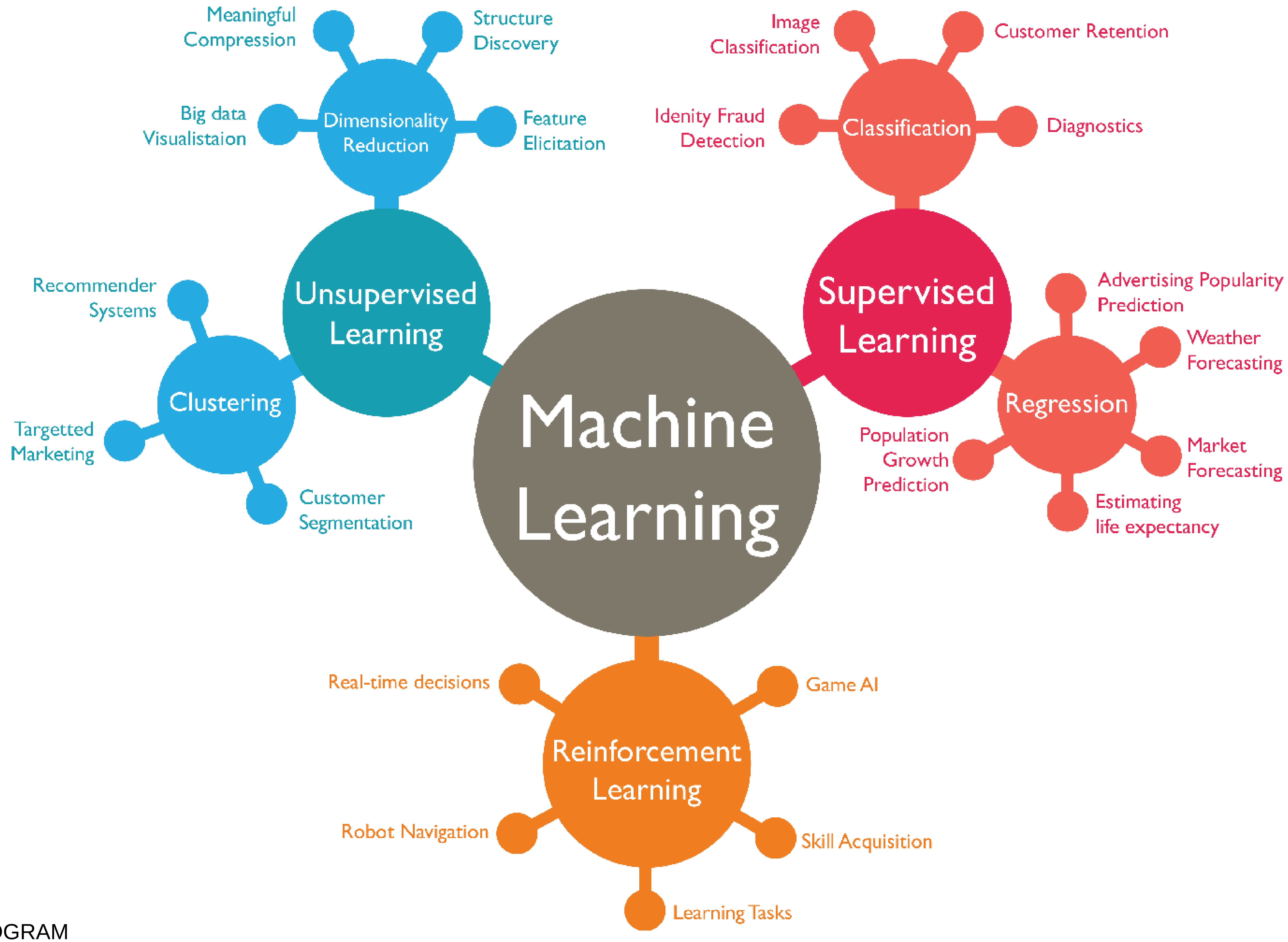
2 **Training the model**

3 **Model validation and selection**

4 **Applying the trained model to unseen data**

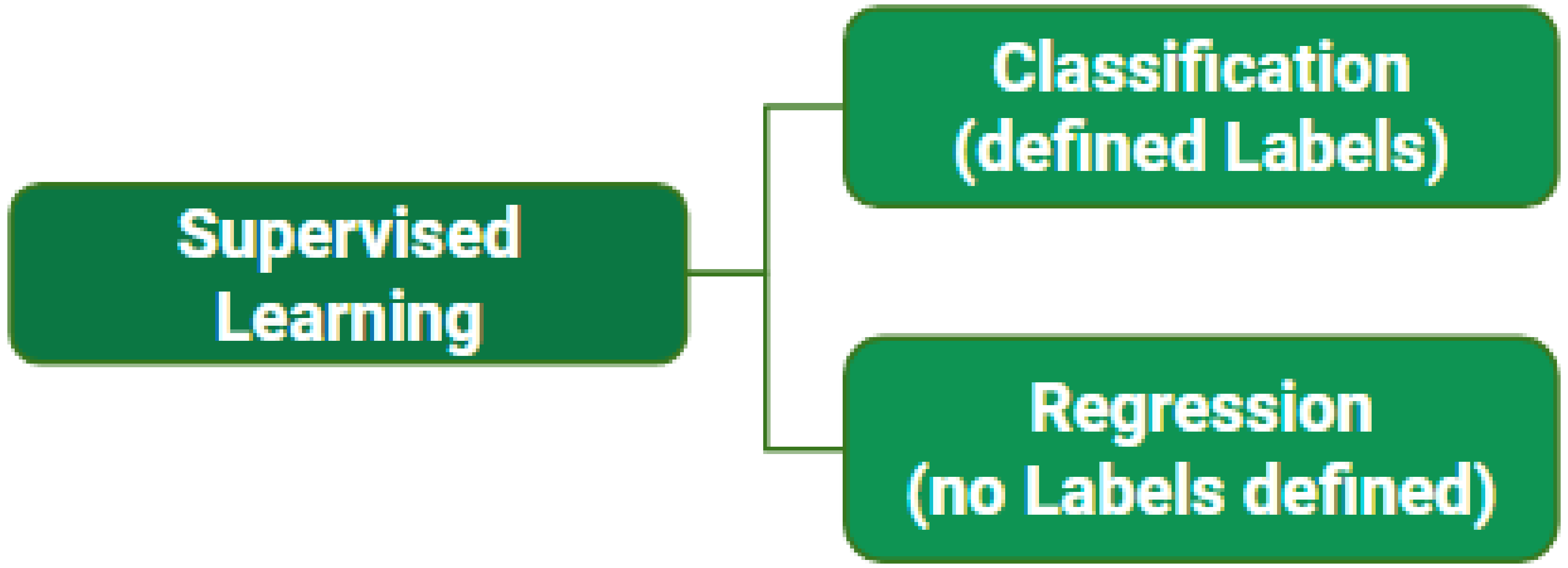
TYPE OF MACHINE LEARNING

The three domains of ML



SUPERVISED LEARNING

- Most of the successful use cases of machine learning and deep learning space fall under supervised learning.

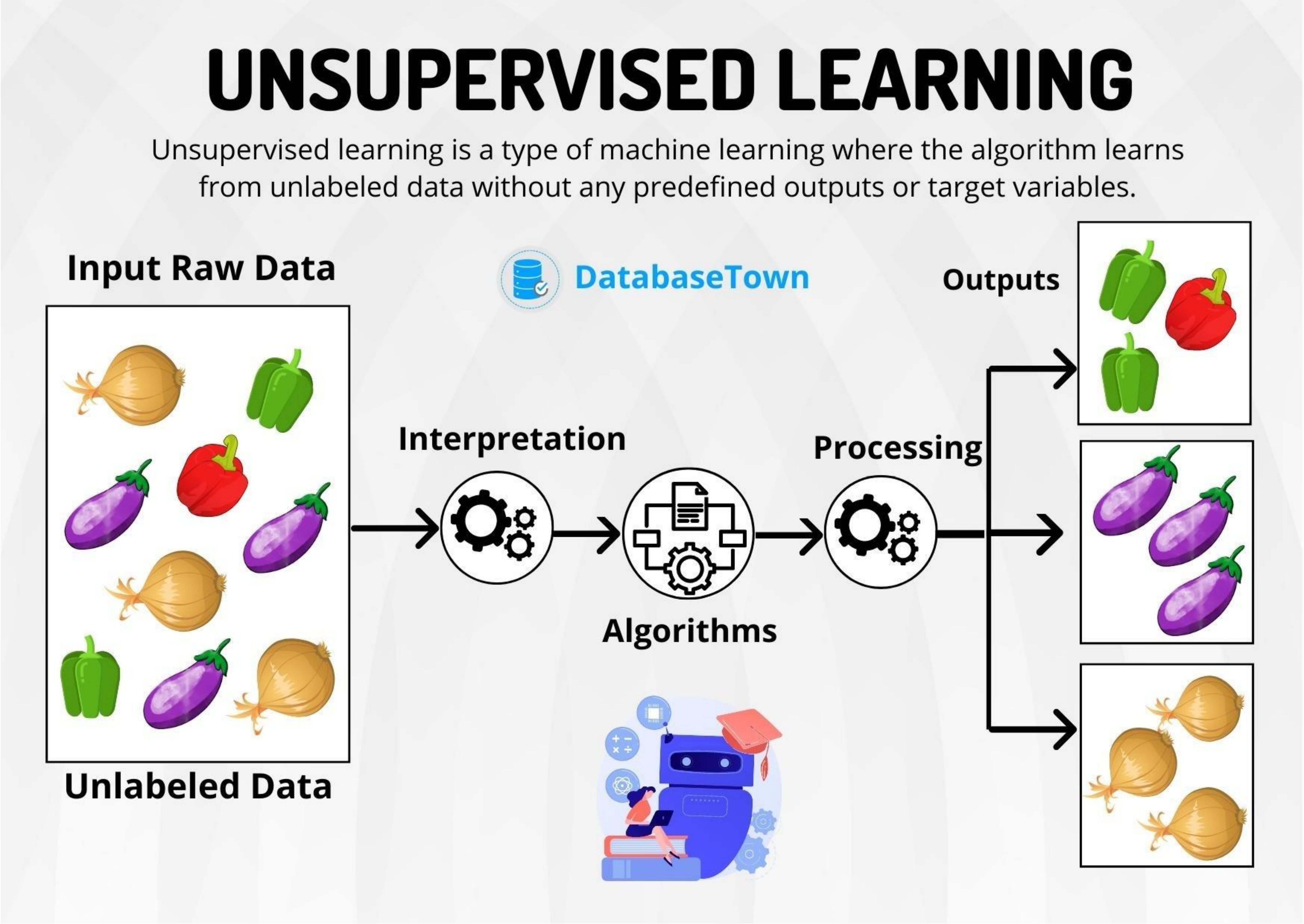


SUPERVISED LEARNING

- A regression model predicts continuous values. For example, regression models make predictions of that answer questions like the following:
  - What is the value of a house in Addis Ababa?
  - What is the probability of having rain tomorrow in Addis Ababa?
- A classification model predicts a discrete(categorical) values. For example, classification models make predictions of that answer the questions like the following:
  - is this image of a cat or dog
  - is a given whether cloudy, rainy, or sunny

UNSUPERVISED LEARNING

- When there is no label data, unsupervised learning techniques help in understanding the data by visualizing and compressing.



- The two most commonly-used techniques in unsupervised learning are:
  - **Clustering** : helps in grouping all similar data-points together
  - **Dimensionality Reduction**: helps in reducing the number of dimensions(features) so that we can visualize the high dimensional data to find hidden patterns.

Reinforcement Learning

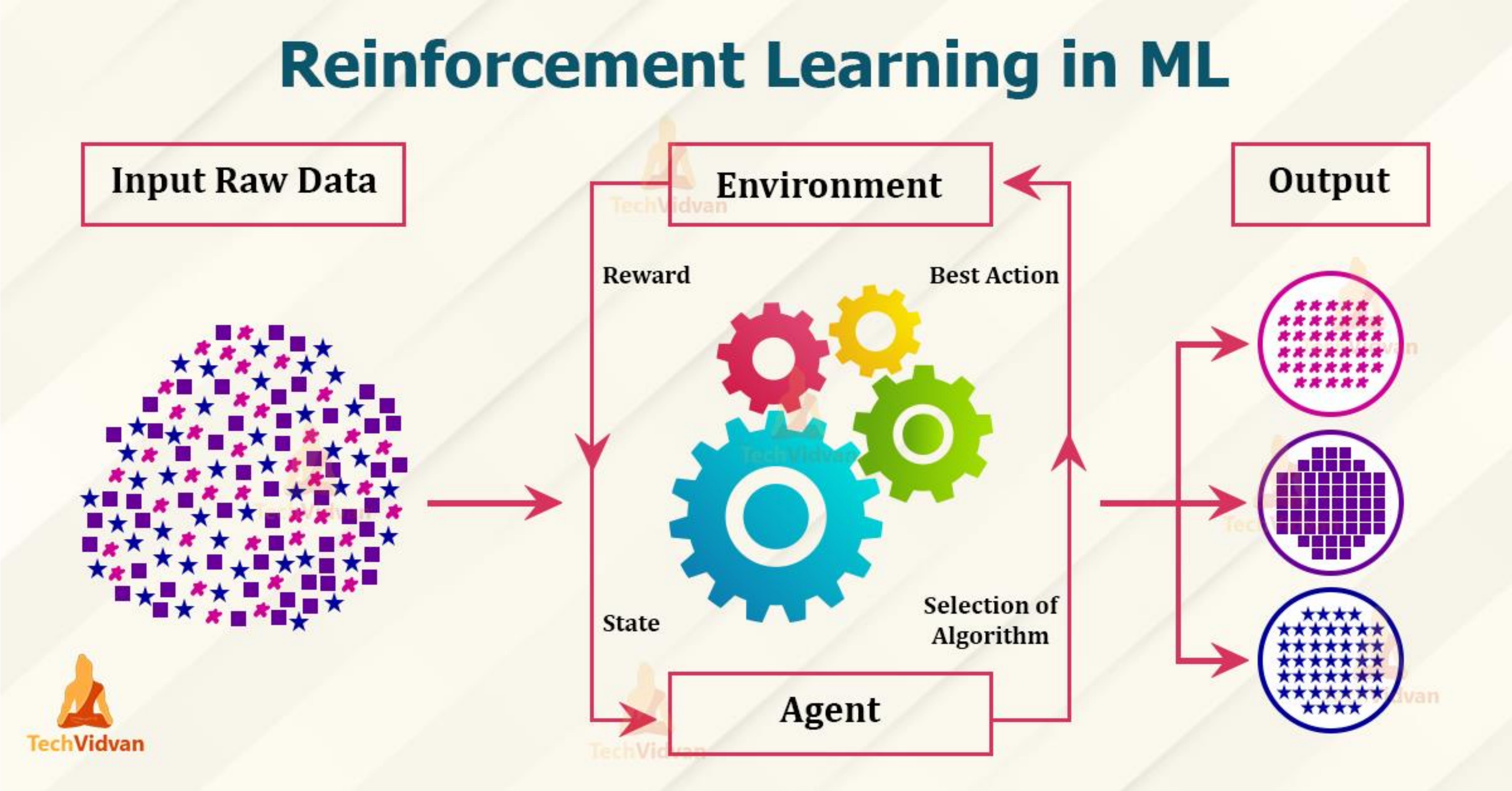
- Reinforcement learning involves training a machine learning model through trial and error, where the model receives feedback in the form of rewards or penalties.

- It is inspired by the way humans and animals learn from feedback in their environment.

- In reinforcement learning, the algorithm is tasked with learning an optimal policy that maximizes a reward signal over time.
- The agent interacts with the environment and receives feedback in the form of rewards or penalties.

- The goal of reinforcement learning is to maximize the cumulative reward over time.
- This requires the agent to balance exploration (trying new actions to discover rewards) and exploitation (using known actions to maximize rewards).

Reinforcement Learning



- Reinforcement learning involves an agent interacting with an environment to learn a policy that maximizes a cumulative reward.
- The agent takes actions in the environment, receives feedback in the form of rewards, and learns to make decisions that lead to the best outcomes over time.

EVALUATING ML MODELS

- To avoid the problem of an information leak and improve generalization, it is often a common practice to split the datasets into three different parts: training dataset, validation dataset and test dataset
  1. Trying the algorithm on the training dataset
  2. Perform hyper-parameter tuning based on the validation dataset
  3. Perform the first two steps interactively until the expected performance is achieved
  4. After freezing the algorithm and the hyper-parameters, evaluate it on the test dataset

EVALUATION METRICS(NEXT)

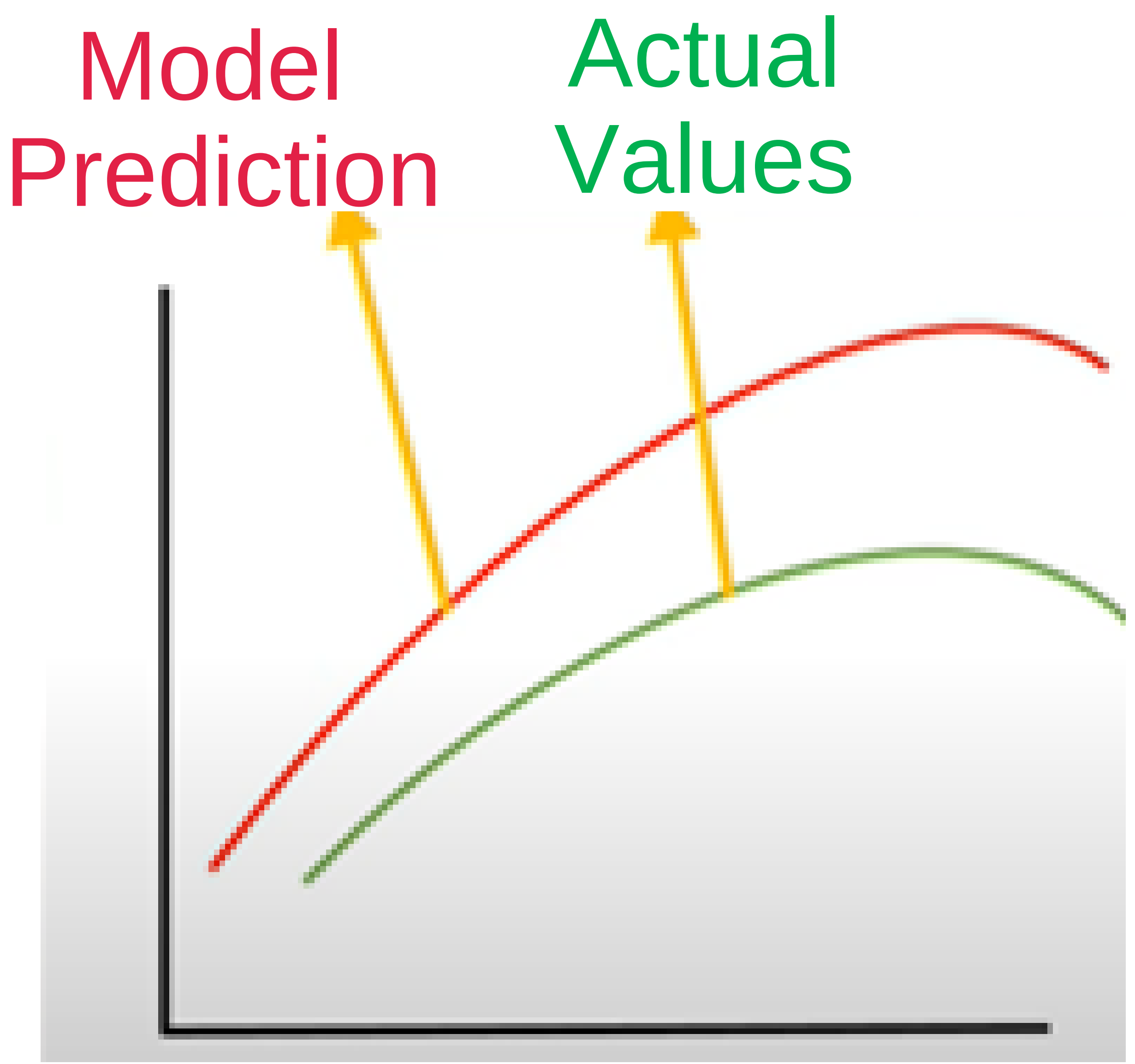
- Performance metrics are a part of every machine learning pipeline.
- It is important to keep measure the performance of trained model how well it generalizes on the unseen data.
- The machine learning fall within different categories and we have different metrics for these categories.
- Metrics are different from loss functions:
  - Loss functions show a measure of model performance using some king of optimization like gradient descent
  - Metrics are used to monitor and measure the performance of a model which don't need to be differentiable
- These metrics can be broken down to either Classification or Regression

Importance of Error Calculation

- An error can be described as an action which is inaccurate or wrong

- In ML, error is used to see how accurately our model can predict on data it uses to learn; as well as new unseen data.

  - Based on error, we choose the ML model which performs best for a particular dataset.



Errors in ML

There are two main types of errors present in ML model.

Irreducible Error

- Are errors which will always be present in a ML model, due to unknown variables, and who's values cannot be reduced

Reducible Error

- Are those errors whose values can be further reduced to improve a model.
- They are caused because our model's output function does not match the desired output function and can be optimized

Errors in ML …

- Reducible errors can be further divided into its two main constituent errors

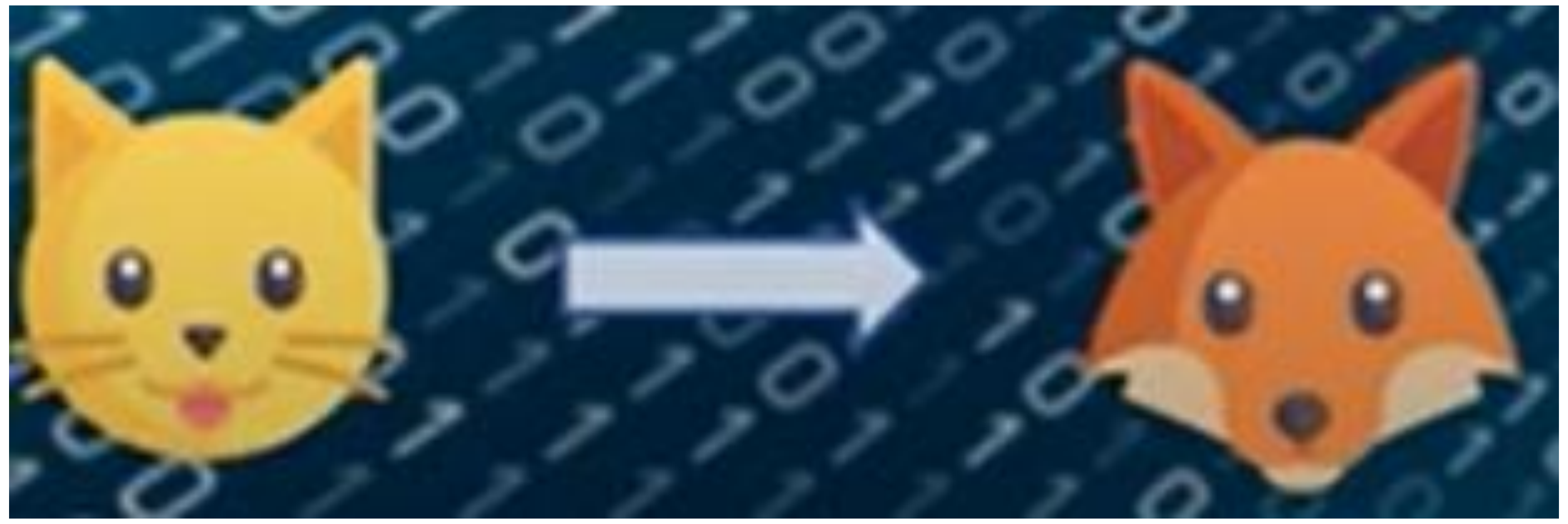
Reducible errors

- **Bias**
- **Variance**

Bias and its effects

- Bias is the difference between our actual and predicted values.
- Bias are the simple assumptions that our model makes about our data to be able to predict on new data

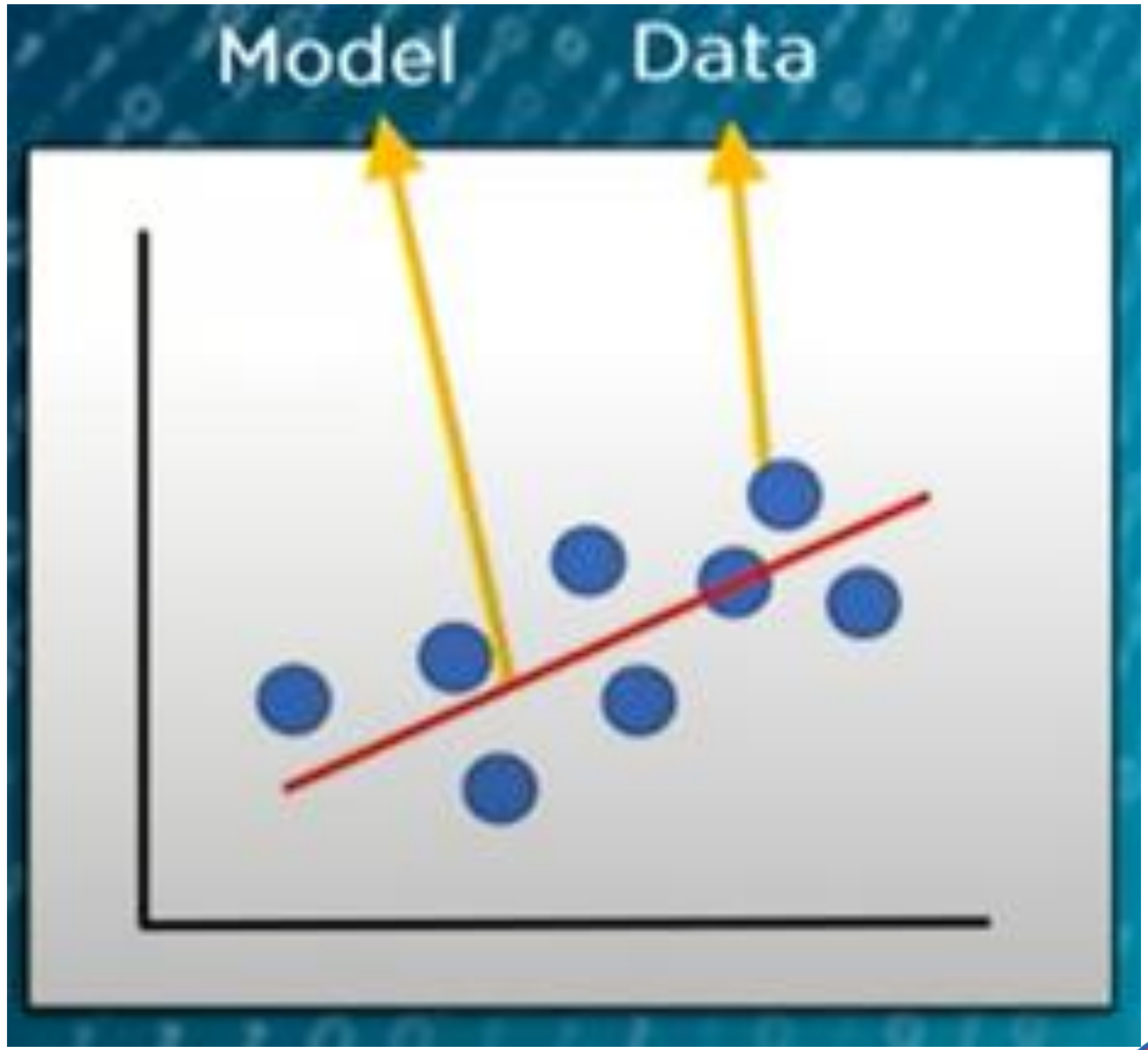
Below example shows our model making wrong assumptions and mistaking a cat for a Fox



Actual Predicted

Bias and its effects . . .

- When the Bias is high, assumptions made by our model are too basic, the model can't capture the important features of our data

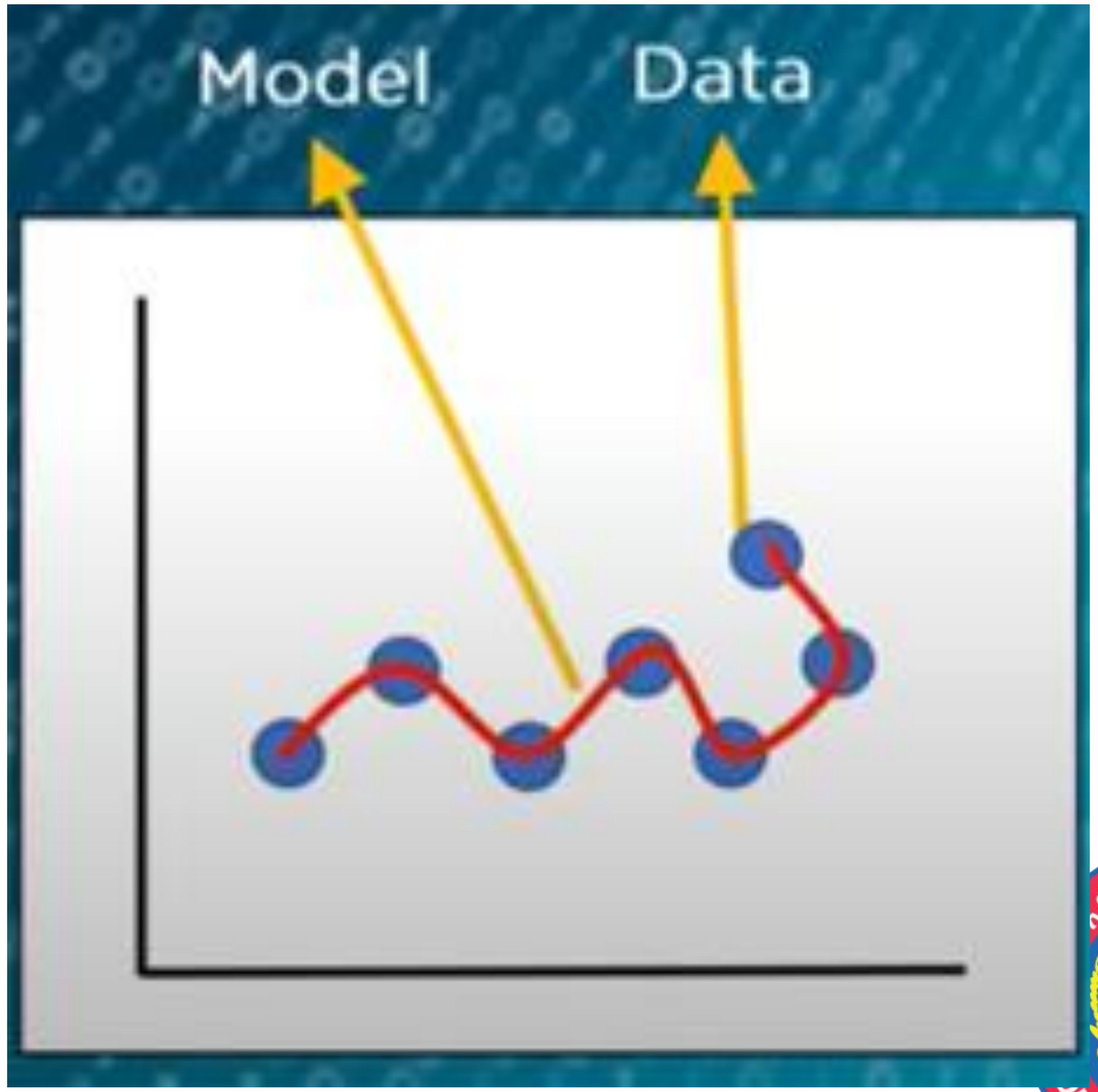


- The model can't predict on the data provided to it, let alone on new data.

- This is called **Underfitting**.

Variance and its effects

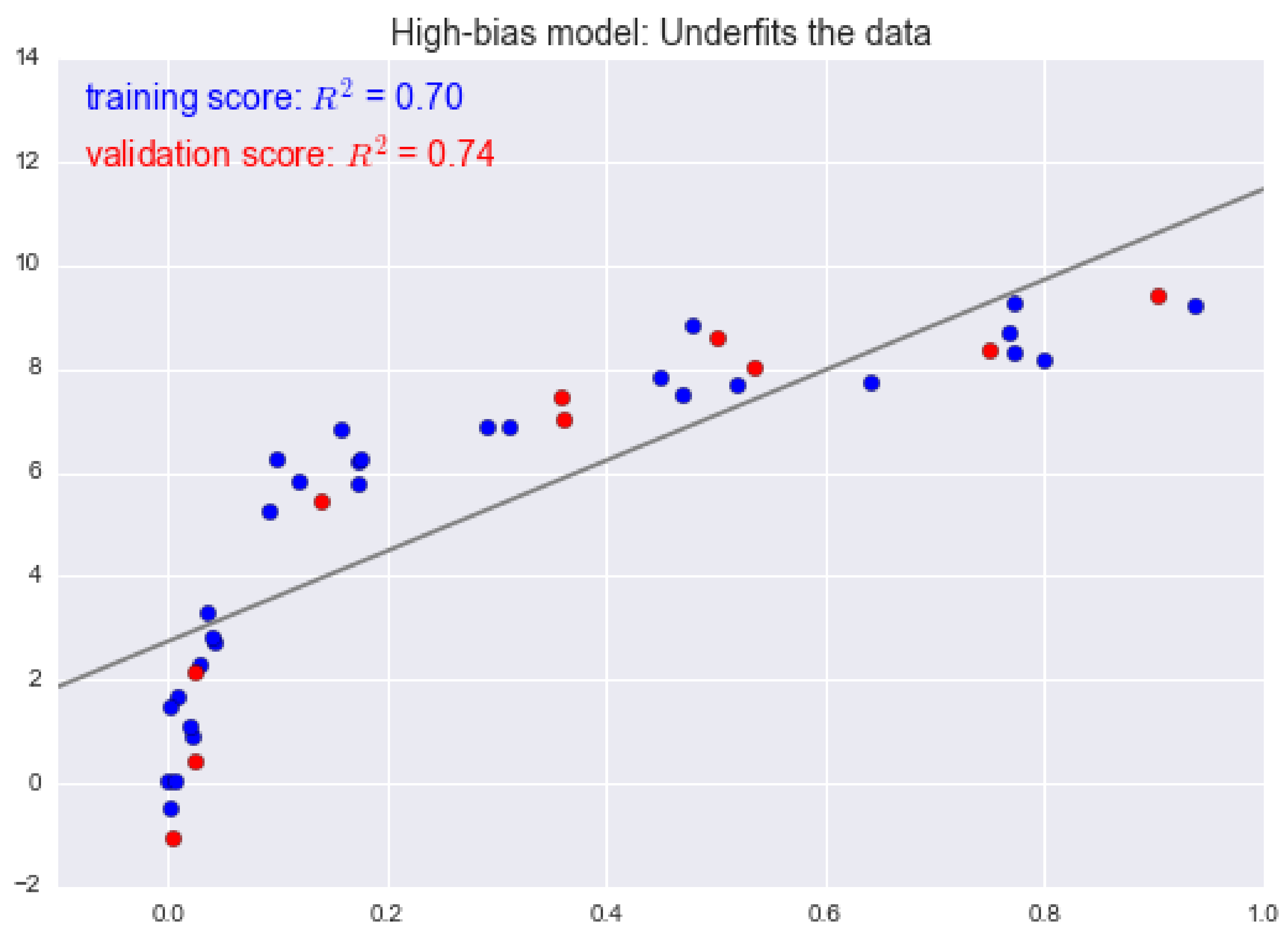
- ❑ Variance can be defined as the model's sensitivity to fluctuates in the data.
- ❑ Our model may learn from noise.
- ❑ This will cause our model to consider trivial features as important

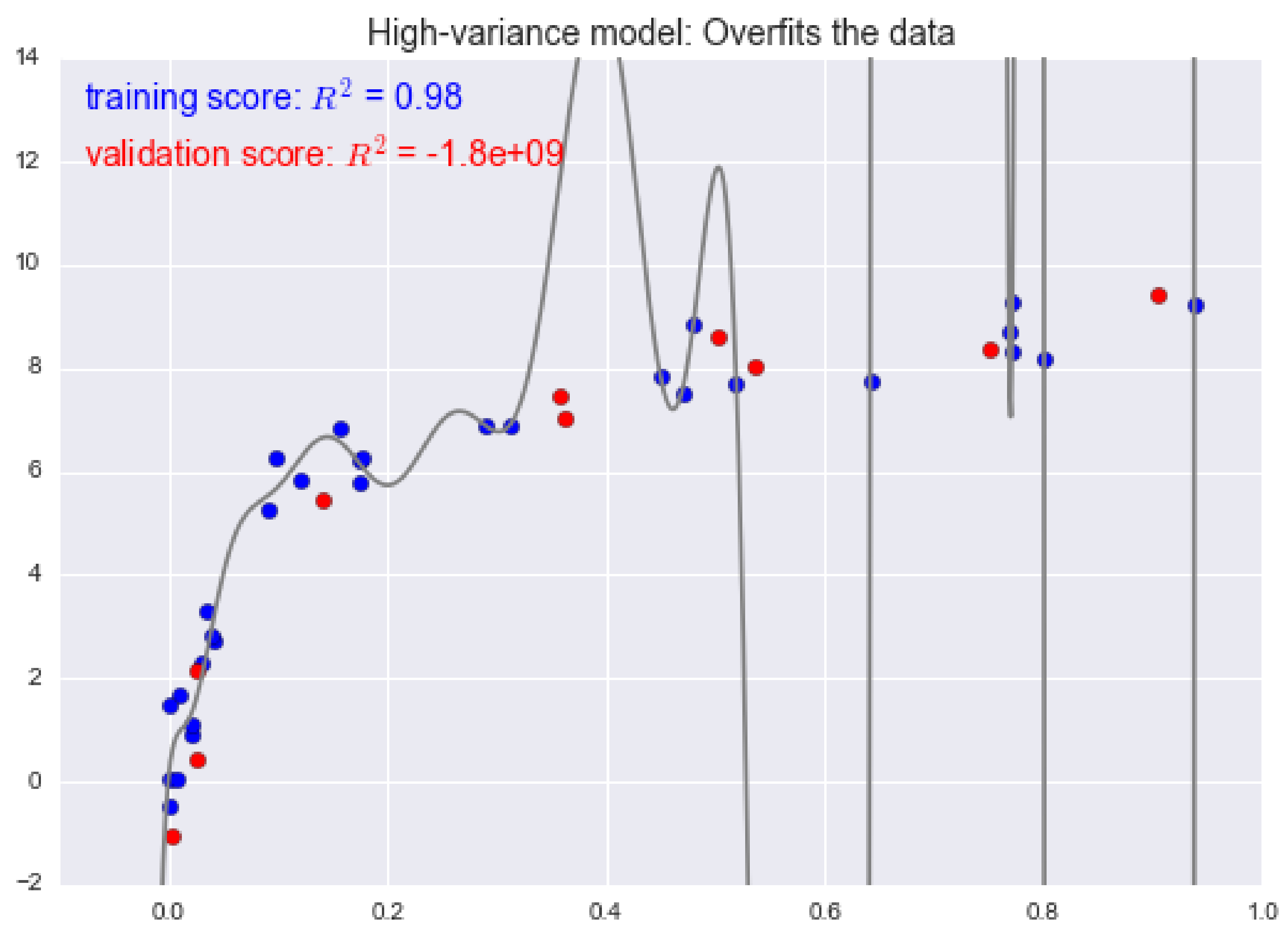


- ❑ When the Variance is high, our model will capture all the features of the data given to it, will tune itself to the data, and predict on it very well

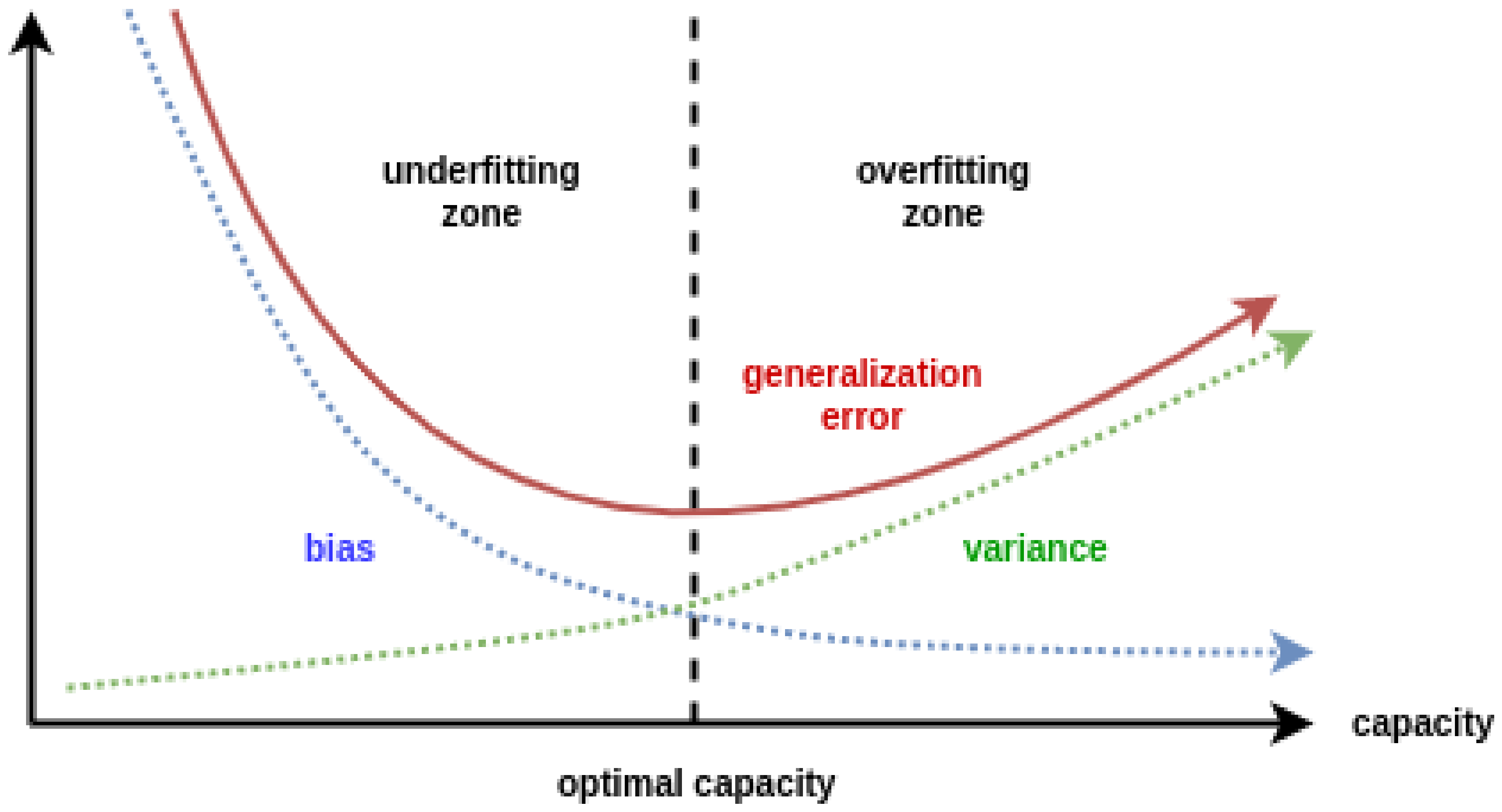
THE BIAS -VARIANCE TRADE-OFF

- Fundamentally, the question of “ the best model” is about finding the sweet spot in the trade-off between bias and variance.
- For high bias models, the performance of the model on the validation set is similar to the performance in the training set.
- For high variance models, the performance of the model on the validation set is far worse than the performance of the training set.



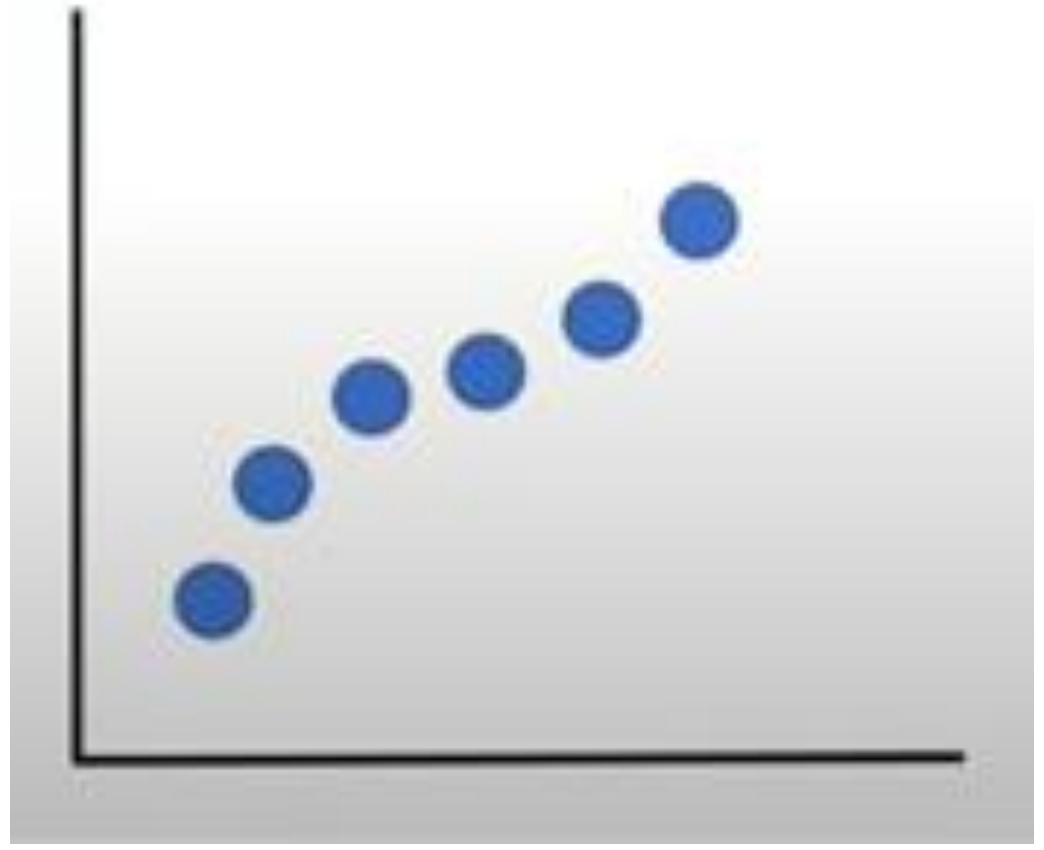


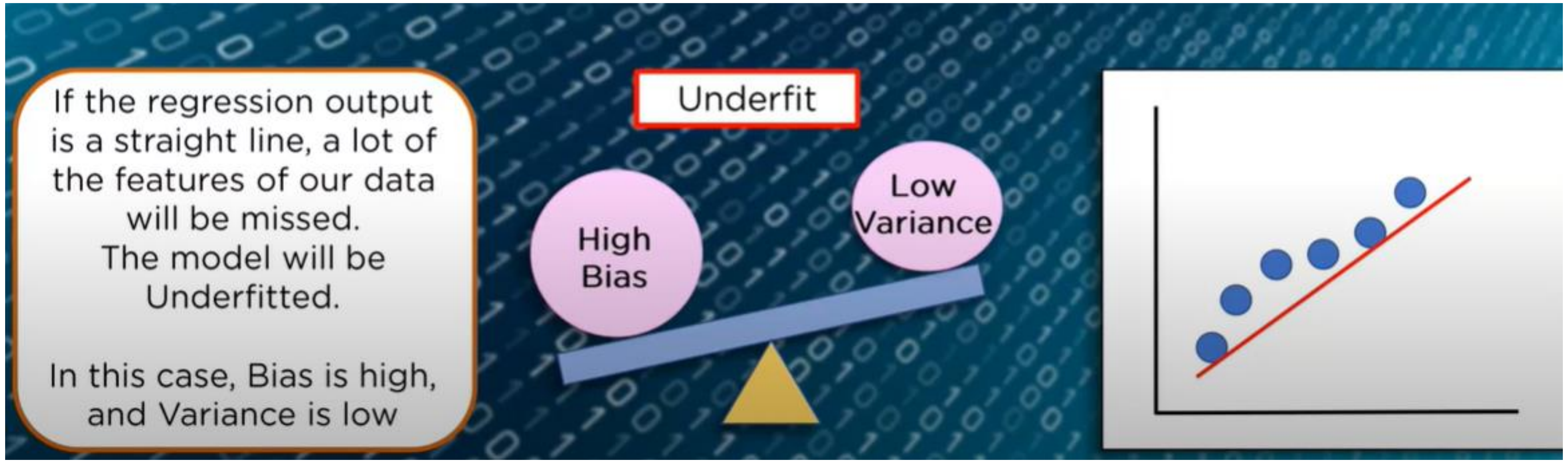
- To optimize the error in our model, we need to find the right balance between bias and variance.
- This is called Bias-Variance Trade-off.

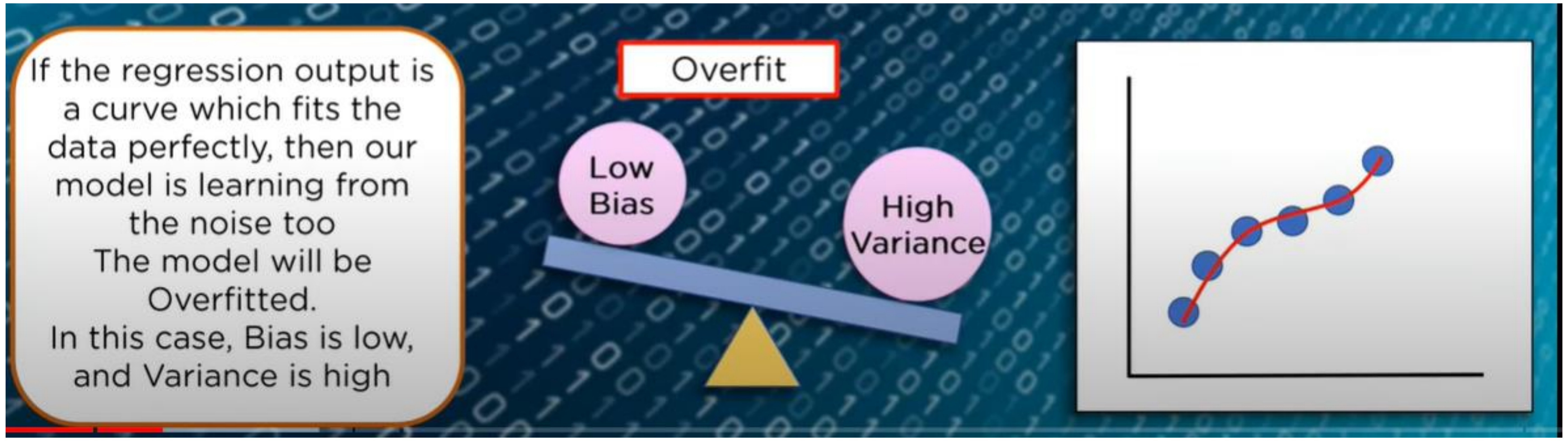


BIAS -VARIANCE TRADE-OFF . . .

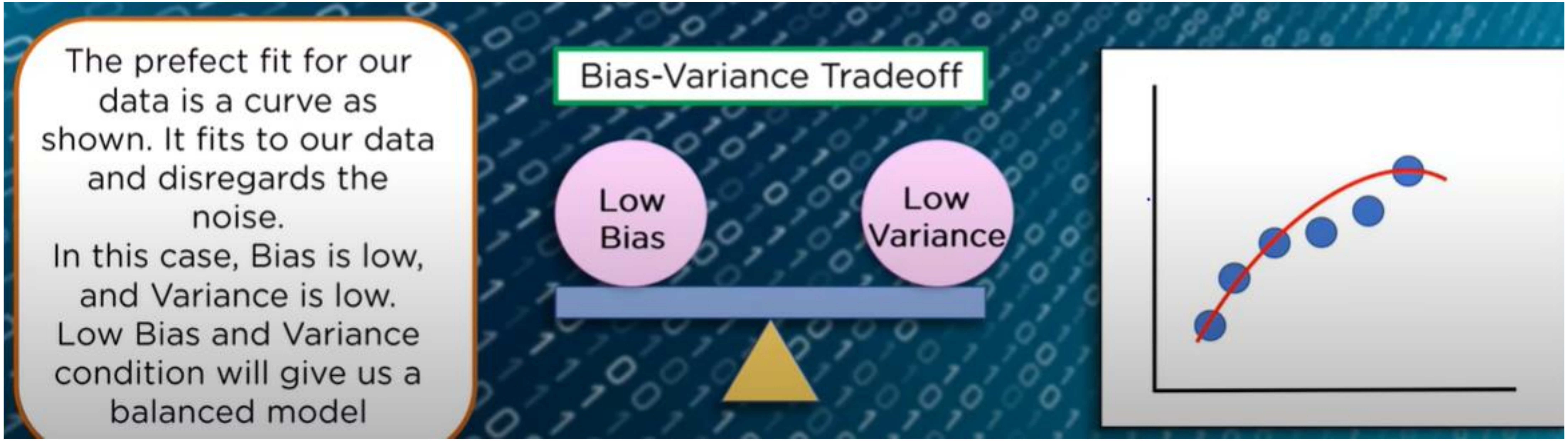
- Regression is a model which finds a relationship between output and input variables by deciding which of the inputs are important and giving more weighted to them.







BIAS -VARIANCE TRADE-OFF . . .

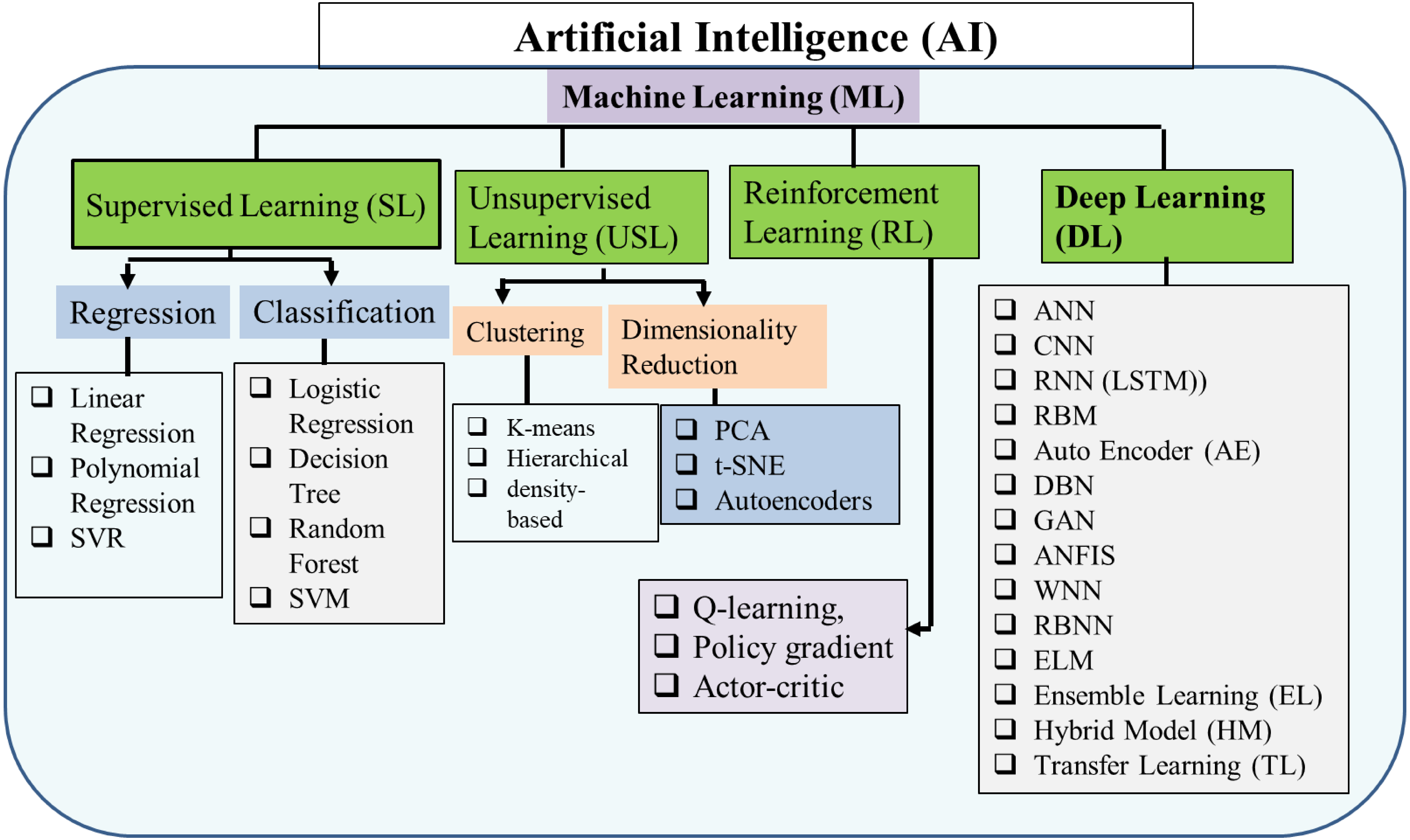


OVER-FITTING AND UNDER-FITTING

- Overfitting, or not generalizing, is a common problem in machine learning and deep learning.
- A particular algorithm overfits when it performs well on the training dataset but fails to perform on unseen or validation and test datasets.
- Underfitting refers to a model that can neither model the training data nor generalize to new data.
- There are different techniques that can be used to avoid the algorithm overfitting. Some of the techniques are:
  - Getting more data
  - Applying regularizer

LEARNING CURVE

- One important aspect of model complexity is that the optimal model will generally depend on the size of the training data.
- A plot of the training/validation score with respect to the size of the training set is known as a learning curve
- A model will never except by chance give a better score to the validation set that the training set.


Artificial Intelligence (AI)
Machine Learning (ML)
Supervised Learning (SL)
Unsupervised Learning (USL)
Reinforcement Learning (RL)
Deep Learning (DL)
Regression
Classification
Clustering
Dimensionality Reduction
Linear Regression
Polynomial Regression
SVR
Logistic Regression
Decision Tree
Random Forest
SVM
K-means
Hierarchical
density-based
PCA
t-SNE
Autoencoders
Q-learning,
Policy gradient
Actor-critic
ANN
CNN
RNN (LSTM))
RBM
Auto Encoder (AE)
DBN
GAN
ANFIS
WNN
RBNN
ELM
Ensemble Learning (EL)
Hybrid Model (HM)
Transfer Learning (TL)

Hands-on

DATA PREPROCESSING

- **Preparing data for machine learning and deep learning involves several steps including data collection, cleaning and preprocessing, feature engineering, data splitting, cross-validation, and data augmentation.**

- **These steps ensure that the data is in the correct format for building accurate and effective machine learning and deep learning models.**

- **Data preprocessing and cleaning are essential steps in preparing data for machine learning and deep learning models. These steps involve identifying and correcting errors, removing outliers, normalizing and scaling the data, encoding categorical data, and sampling and balancing imbalanced datasets.**

DATA PREPROCESSING

- Data preprocessing is a process in which we make the data more suitable for the ML algorithms to train on. The following are some of the commonly-used data preprocessing steps:
  - Binarization: convert our numerical values into boolean values
  - Mean removal:
  - MinMax Scaling
  - Normalization

BINARIZATION

- This is a process when we want to convert our numerical values into boolean values. for instance:

```
import numpy as np
from sklearn.preprocessing import Binarizer

data = np.array([[5.2, -3, 3.5],
                [-2.0,7.0,-6.2],
                [-7.4,-9.9,-5.4]])

#Binarize data

binarized = Binarizer(threshold=2.0,).transform(data)
print("\n Binarized data: \n", binarized)
```

BINARIZATION

- If you run the code, you will get the following output

```
Binarized data:
 [[1., 0.0, 1.],
  [0.,1.0,0.0],
  [0.,0.,0.0]])
```

#Binarize data

- All the values above 2.0 becomes 1 and the remains values become 0.0

MEAN REMOVAL

- Removing the mean is a common preprocessing technique used in machine learning.
- It helps to center each feature mean on zero in order to remove bias from the featured in feature vectors

```
import numpy as np
from sklearn.preprocessing import scale

data = np.array([[5.2, -3, 3.5],
               [-2.0,7.0,-6.2],
               [-7.4,-9.9,-5.4]])
#print mean and standard deviation
print("\n Before:")
print("mean = ", data.mean(axis=0))
print("Standard deviation=", data.std(axis=0))
#Remove mean
scaled = scale(data)
print("\n After:")
print("Mean =", scaled.mean(axis=0))
print("Standard deviation =", scaled.std(axis=0))
```

MEAN REMOVAL

- If you run the code, we will get the following printed on our terminal where the mean is very close to 0 and the standard deviation is 1.

Before:
mean = [-1.4 -1.96666667 -2.7]
Standard deviation= [5.16139516 6.93797921 4.39621049]

After:
Mean = [7.40148683e-17 0.00000000e+00 1.11022302e-16]
Standard deviation = [1. 1. 1.]

MINMAX SCALING

- When the value of each feature varies between many random values, it becomes important to scale those features so that it is a level playing field for the ML algorithm to train on.

```
import numpy as np
from sklearn.preprocessing import MinMaxScaler
data = np.array([[5.2, -3, 3.5],
                 [-2.0,7.0,-6.2],
                 [-7.4,-9.9,-5.4]])
#Min max Scaling
minmax_scaler = MinMaxScaler(feature_range=(0,1))
minmax_scaled = minmax_scaler.fit_transform(data)
print("\n Min max scaled data= \n", minmax_scaled
```

MINMAX SCALING

- The minimal Scaler in our previous code will generate the following print on our terminal.

```
Min max scaled data=
[[1.         0.40828402 1.        ]
 [0.42857143 1.         0.        ]
 [0.         0.         0.08247423]]
```

NORMALIZATION

- Normalization modify the values in the feature vectors so that we can measure them on a common scale.
- The most common forms of normalization aim to modify the values so that they sum up to one(1).
- L1 normalization which refers to Least Absolute Deviations works by making sure that the sum of absolute values is 1 in each row.
- L2 normalization which refers to least squares works by making sure that the sum of squares is 1.
- In general L1 normalization technique is considered more robust than L2 normalization.

NORMALIZATION

```
import numpy as np
from sklearn.preprocessing import normalize

data = np.array([[5.2, -3, 3.5],
                [-2.0,7.0,-6.2],
                [-7.4,-9.9,-5.4]])

#normalize data

l1_norm = normalize(data, norm='l1')
l2_norm = normalize(data, norm='l2')

print("\n L1 normalized data: \n", l1_norm)
print("\n L2 normalized data: \n", l2_norm)
```

NORMALIZATION

- The output of the previous normalization code print the following result on our terminal:

L1 normalized **data**:

```
[[ 0.44444444 -0.25641026  0.2991453 ]
[-0.13157895  0.46052632 -0.40789474]
[-0.32599119 -0.43612335 -0.23788546]]
```

L2 normalized **data**:

```
[[ 0.74829827 -0.43171054  0.5036623 ]
[-0.20915194  0.73203177 -0.648371  ]
[-0.54863001 -0.73397799 -0.40035163]]
```

LABEL ENCODEING

- When we perform classification, we usually deal with a lot of categorical labels.
- These labels can be in the form of words, numbers or something else.
- However, the ML algorithms expect them to be numbers

LABEL ENCODEING

Label mapping:
cloudy --> 0
rainy --> 1
sunny --> 2

Encoded values = [2, 0, 2]
Decoded labels = ['sunny', 'cloudy', 'sunny']