Hewlett Packard
Enterprise

MODULE 3:
AI - BRINGING INTELLIGENCE TO PREDICTIVE MODELS

Month 1, 2022

AGENDA

1. INTRODUCTION TO ARTIFICIAL INTELLIGENCE
2. CAPABILITIES AND EVALUATION OF AI
3. COMPUTER VISION
4. NATURAL LANGUAGE PROCESSING
5. SPEECH

LESSONS:

Lesson 1: Overview of AI Capabilities

- Introduction to capabilities of AI
- AI Capability tests

Lesson 2: Computer Vision

- Representation of image.
- Types of Images
- Text extraction from documents.
- Image Classification
- Object Recognition

Lesson 3: Natural Language Processing

- Components of NLP
- Phases of NLP
  - Morphological and Lexical Analysis
  - Syntactic Analysis
  - Semantic Analysis
  - Pragmatic Analysis
- Language understanding intelligence
- Virtual customer service assistance

Lesson 4: Speech Processing

Speech Recognition
Speech Synthesis
ANN for speech processing

LESSON 1:

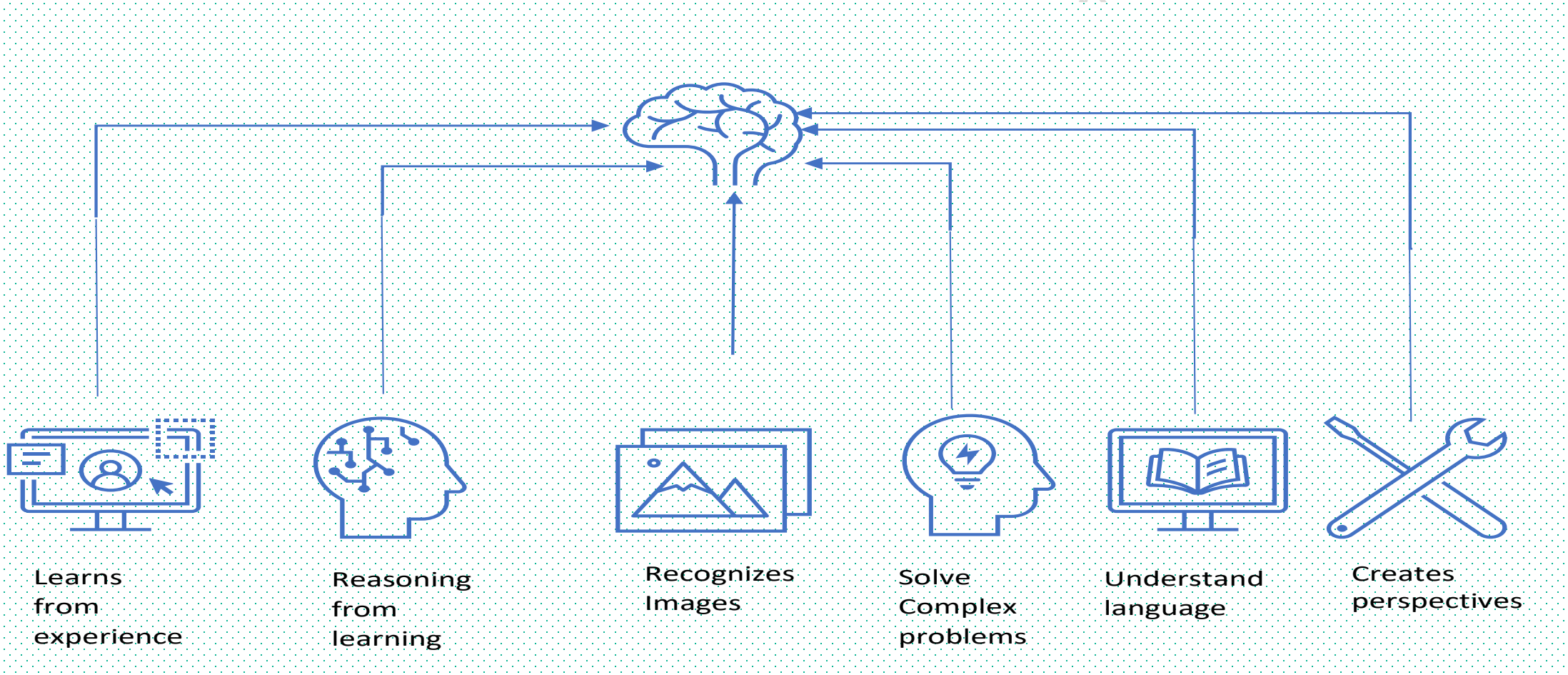
Overview of AI Capabilities

WHAT IS ARTIFICIAL INTELLIGENCE?

Computer Programs that imitates human cognitive capabilities

- Making decisions based on previous data and experience
- Identification of abnormal events
- Understanding visual input
- Written and spoken language
- Engaging in intelligent conversations

CAPABILITIES OF ARTIFICIAL INTELLIGENCE

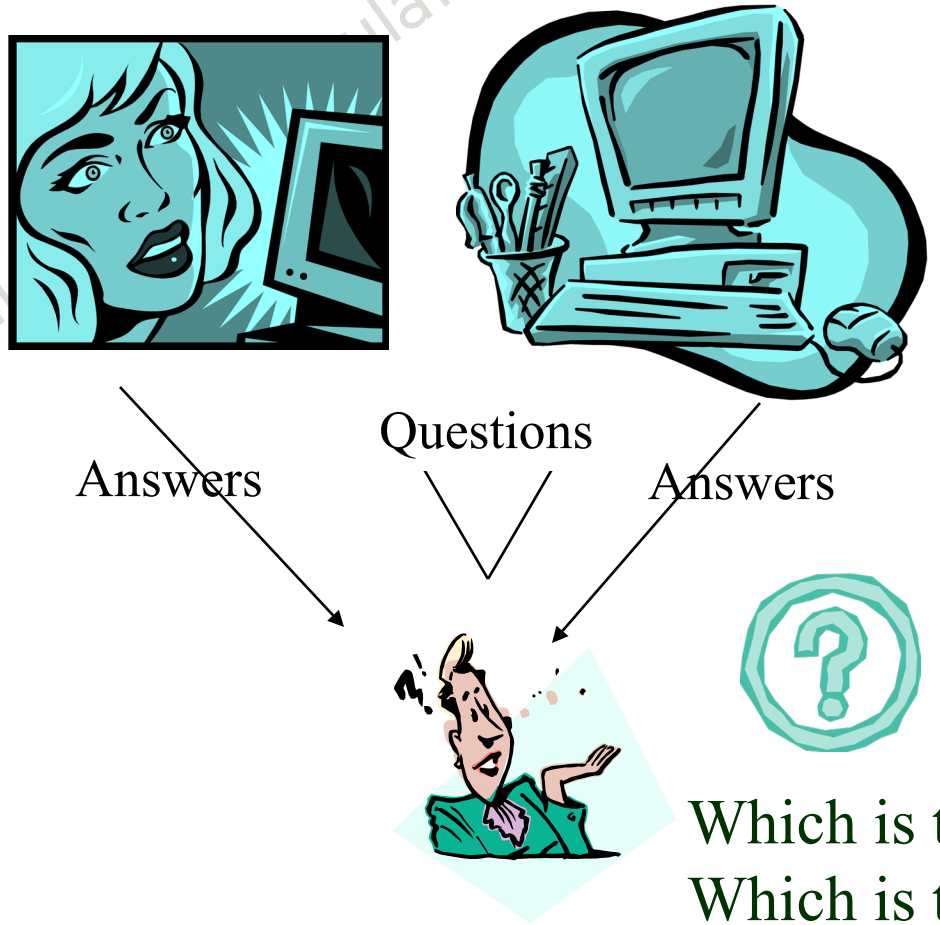


AI CAPABILITIES – BROAD CATEGORIES

- Personalization and profiling : personalized suggestions in Apps.
- Predictions: Marketing, Healthcare etc. to assess future trends.
- Pattern recognition and anomaly detection: Suggests probable future moves and detect abnormalities.
- Natural language Processing: Analyze and communicate in human language.
- Object identification: Detect people, abnormal objects etc.
- Goal achievement: Experiential learning , e.g., Gaming

AI TEST - THE TURING TEST

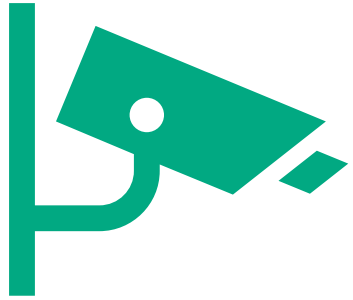
- Alan Turing devised a test for intelligence called the Imitation Game
  - Ask questions of two entities, receive answers from both
  - If subject can't tell which of the entities is human and which is a computer program, consider the computer to be intelligent



LESSON 2:

Computer Vision

COMPUTER VISION



- Computers can see the world through digital images
- Photographs, live camera feeds, or recorded video files
- Digital image is an array of numeric values indicating the color intensity of each pixel in the image.
- Each number is a value between 0 and 255, representing the intensity of each pixel
- For color images, there are three layers of pixel values for red, green, and blue intensities

REPRESENTATION OF DIGITAL IMAGES

- An image can be considered as a function of two real variables, b (x, y) where b is the brightness at the coordinate position (x, y).
- An analog image b (x, y) in a 2D space is converted into a digital image b[m, n] through digitization.

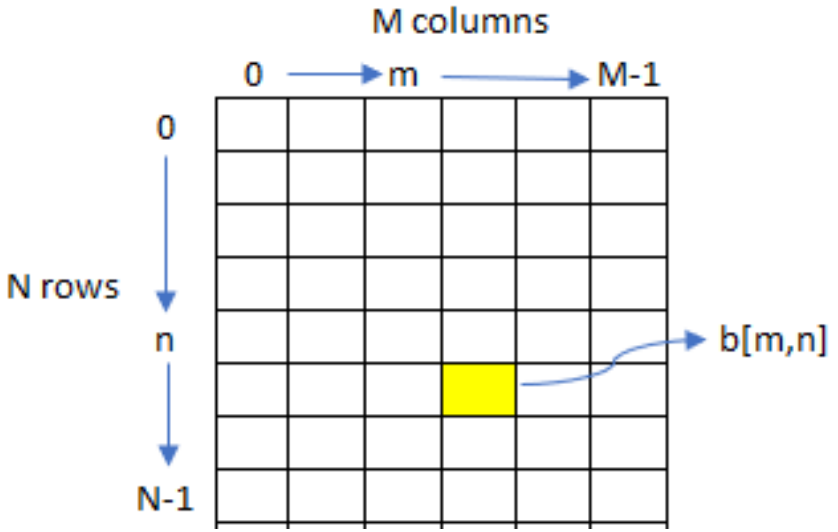


IMAGE ANALYSIS
Image Representation

Parameter	Typical values	Symbol
Columns	256,512,768,1024,1920	M
Rows	256,512,525,625,1024,108(	N
Gray Levels	2,64,256,1024,4096,16384	L

Table 1: Common digital image parameters value

- Pixels are the basic element of any image , also known as picture elements.
- Pixel values are binary words of length k and can represent up to 2^k different values.
- The value k is known as bit depth or depth of the image.

IMAGE ANALYSIS
Types of Images

Binary Images

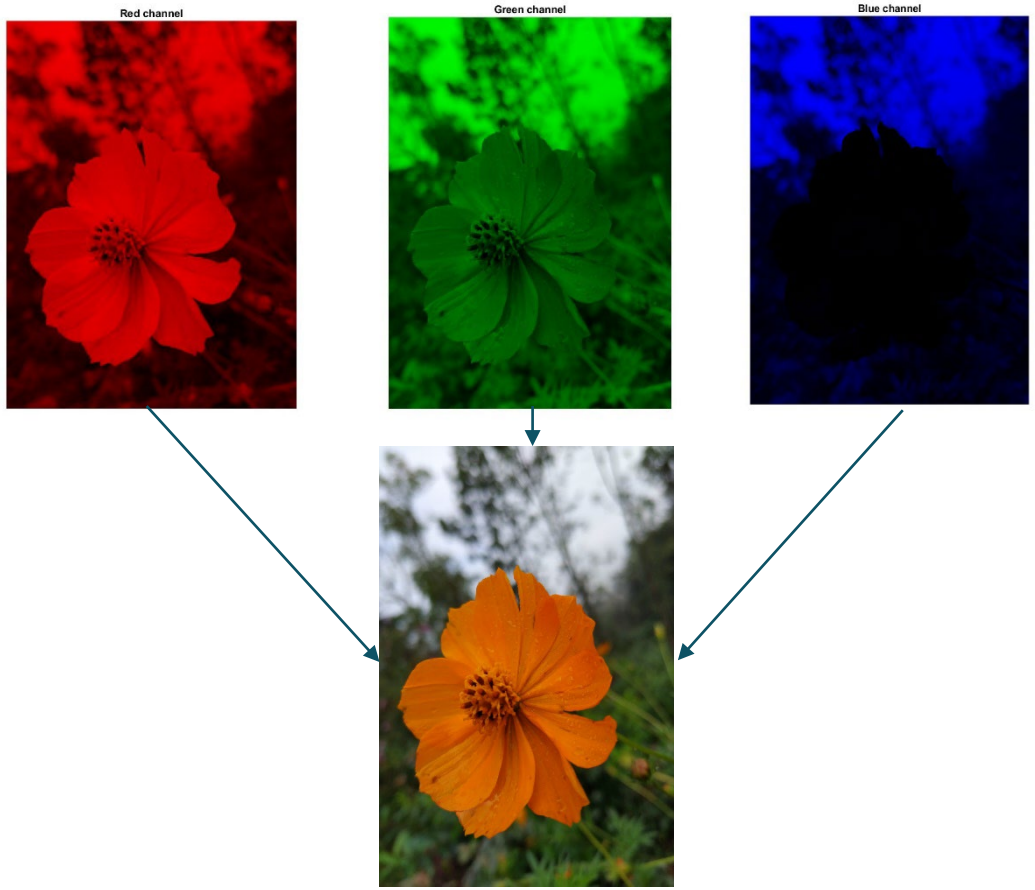
- The various types of images are binary, grayscale, and color images.
- Binary images pixel values can take only two values black or white as shown in Figure 2.
- These are encoded using a single bit (0/1) per pixel.

Grayscale Images

- The brightness, intensity or depth of the image is represented in a grayscale image.
- The image data are positive whole numbers in the range $0 \ldots 2^k - 1$
- It cannot be a negative value as it represents intensity or density of light energy.

IMAGE ANALYSIS

Types of Images –Color Images



- Color images are based on primary colors red, green, and blue (RGB).
- The range is around [0,255] and makes use of 8 bits for each color component.
- By superimposing three colors, red, green, and blue human eyes can be “fooled” into seeing most of the colors which can be recognized perceptually

IMAGE ANALYSIS
Types of Images

(1,0,0) = red	(0,1,0) = green	(1,1,0) = yellow	(0,0,1) = blue
(0,0,0) = black	(0,1,1) = cyan	(1,1,1) = white	(0.5,0.5,0.5)= grey

- The primary colors can be combined to form different colors as shown in Table 2.
- If the image is of dimension 240x400 and each pixel is of 8 bits,
- then the size of the image will be 240 x 400 x 8 = 768000 bits.
- As 1 bit = 8 bytes, this value can be converted to bytes by 768000 / 8 = 96000 bytes.

HOW DO HUMANS RECOGNIZE..?

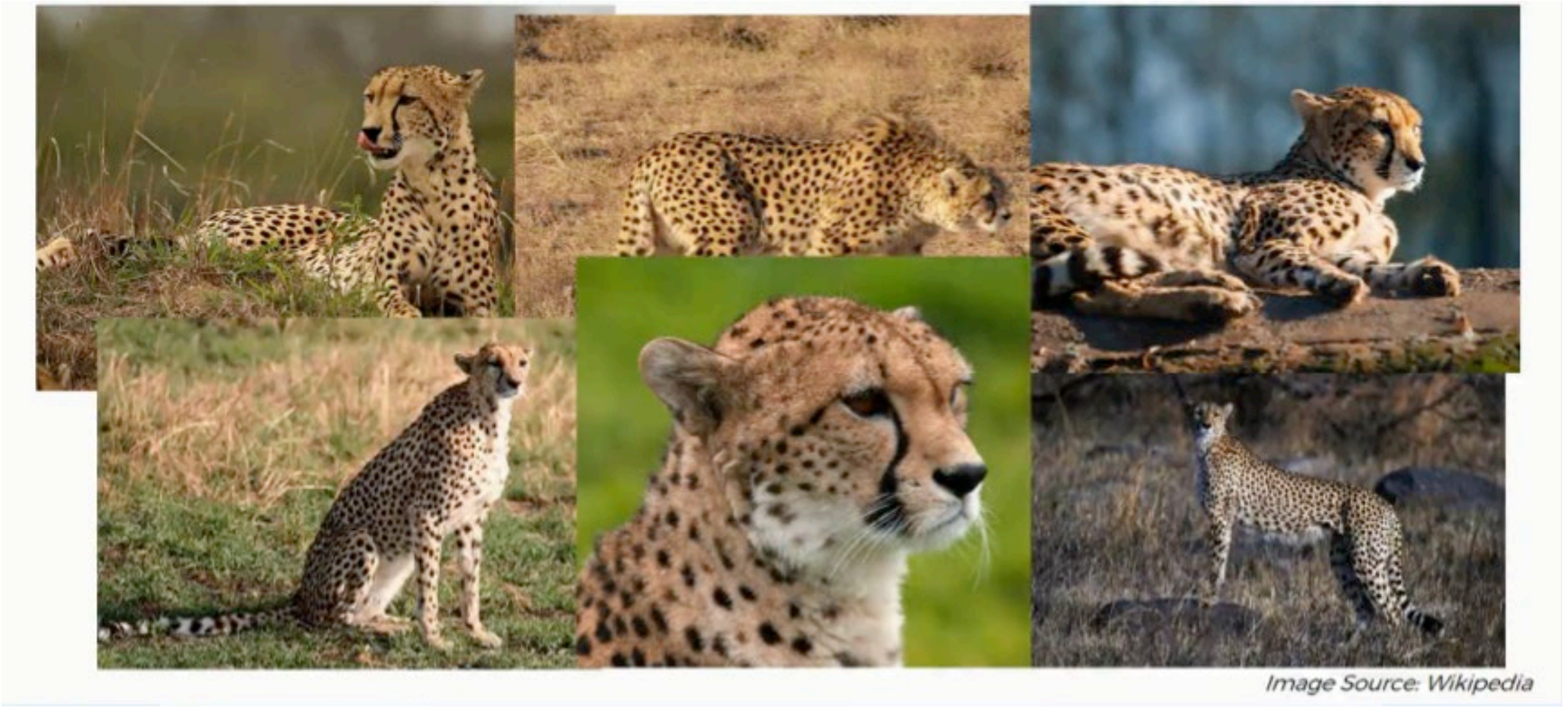
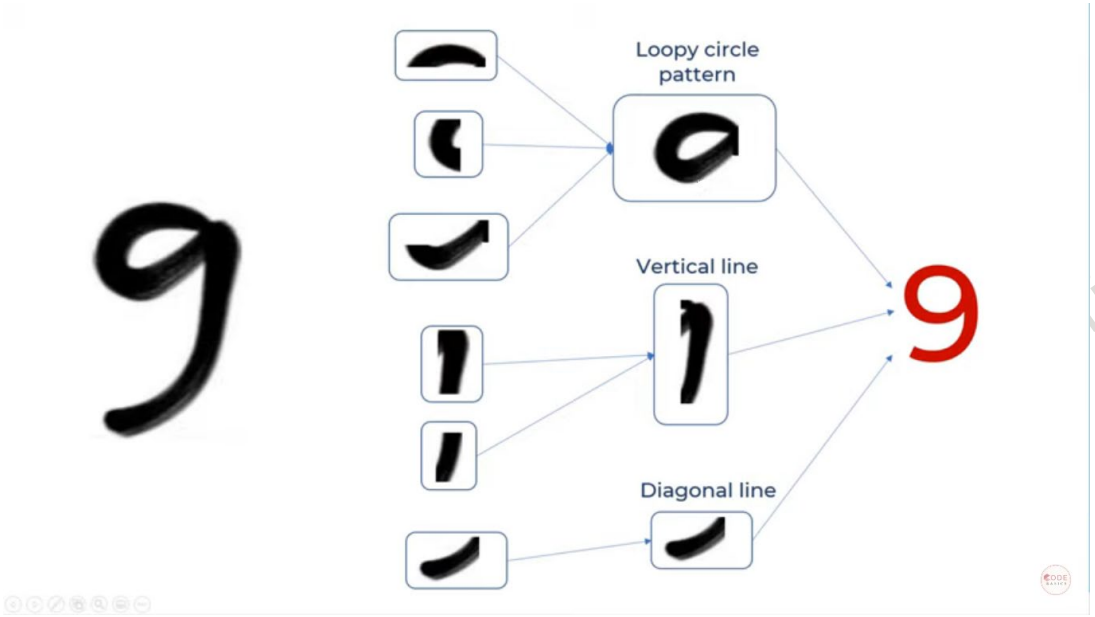


Image Source: Wikipedia

HOW DO HUMANS RECOGNIZE..?

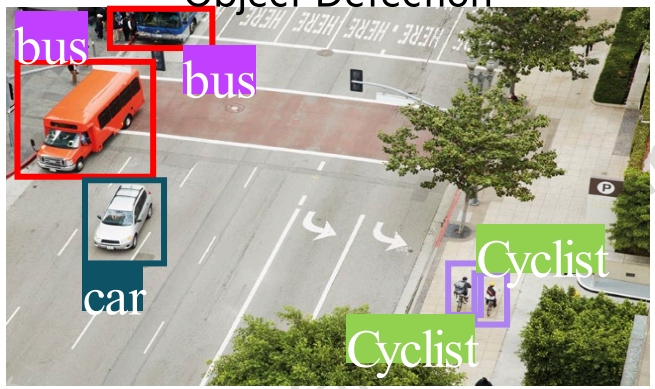


Humans recognize variety..!!

COMPUTER VISION

Image Analysis

Object Detection



Semantic Segmentation

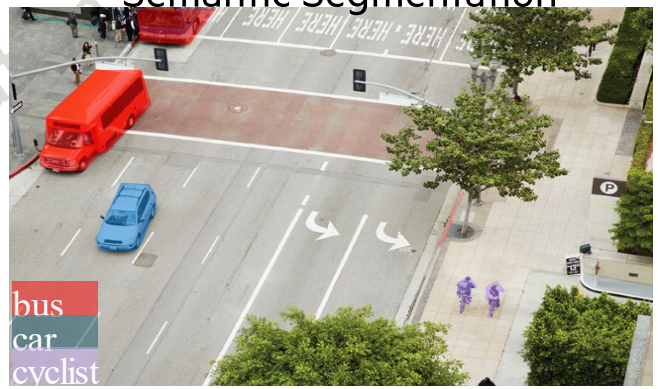
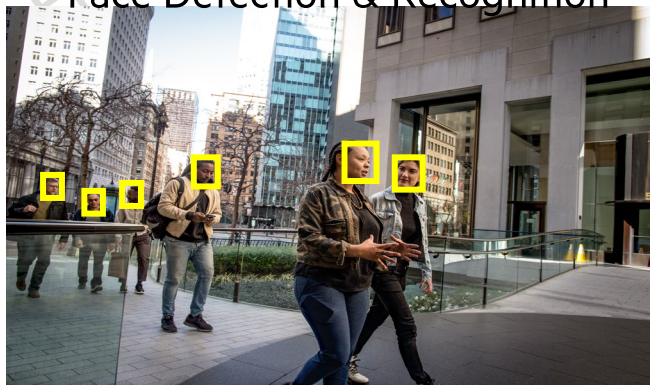


Image Classification

Face Detection & Recognition



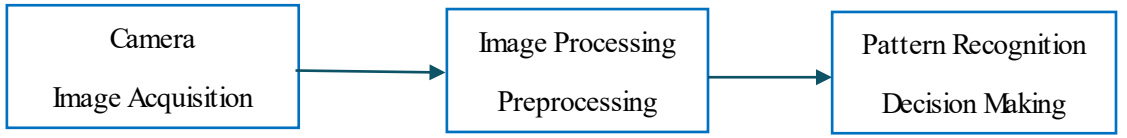
Optical Character Recognition

Source: Microsoft

IMAGE ANALYSIS
Objectives

- To interpret real images
- To identify and reconstruct
  - Texture characteristics
  - Scene illumination
  - Color characteristics
  - Shape information.
- A caption is generated about the details of the image.
- Technique used: Computer vision

Human vision

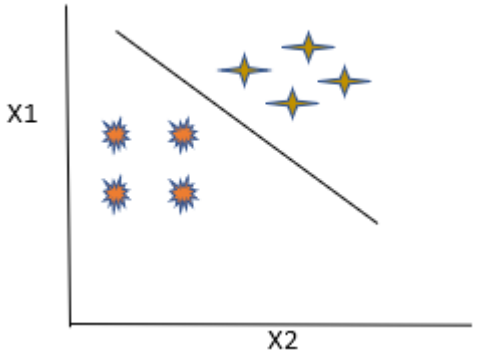


Computer Vision

IMAGE CLASSIFICATION
Types Based on number of classes

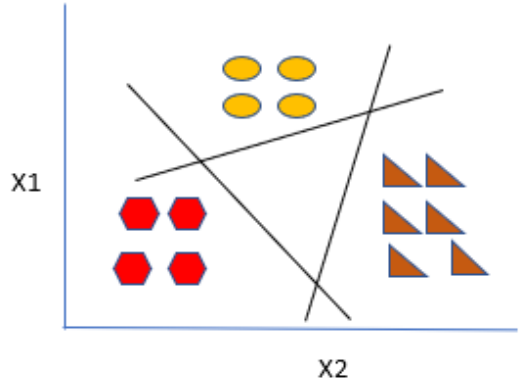
Binary Classification

- 2 classes



Multi class classification

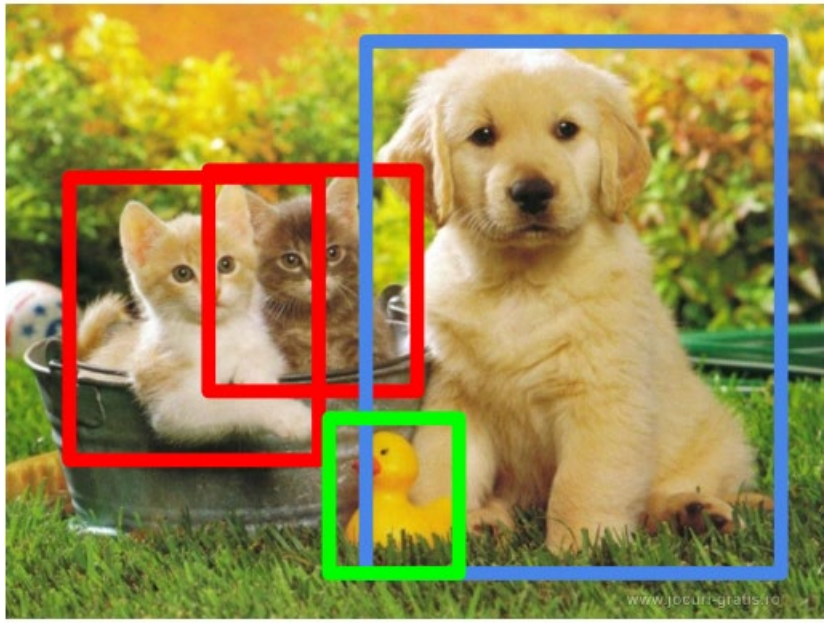
- More than 2 classes



- Boundary is created to differentiate classes
- Types: Based on the presence of labeled data
  - Supervised Classification: labeled: eg: KNN, DT,SVM
  - Unsupervised Classification: Unlabeled : eg: K-means, ISODATA

OBJECT DETECTION

- The task of assigning a label and a bounding box to all objects in the image



CAT, DOG, DUCK

OBJECT DETECTION

- Earlier classification-based approaches

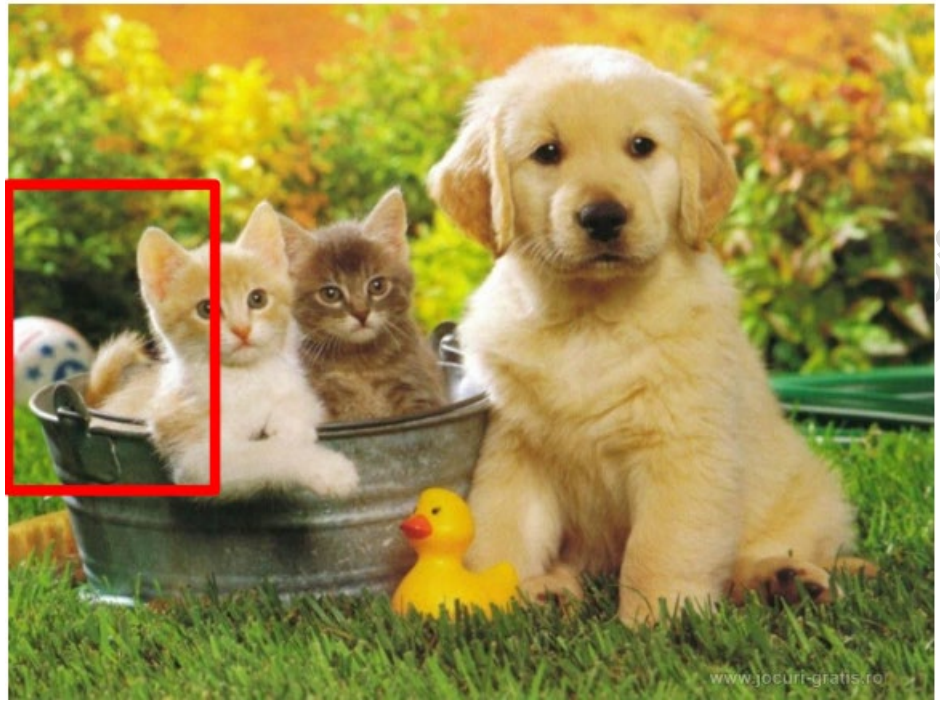
Classes = [cat, dog, duck]

Cat ? NO

Dog ? NO

Duck? NO

OBJECT DETECTION



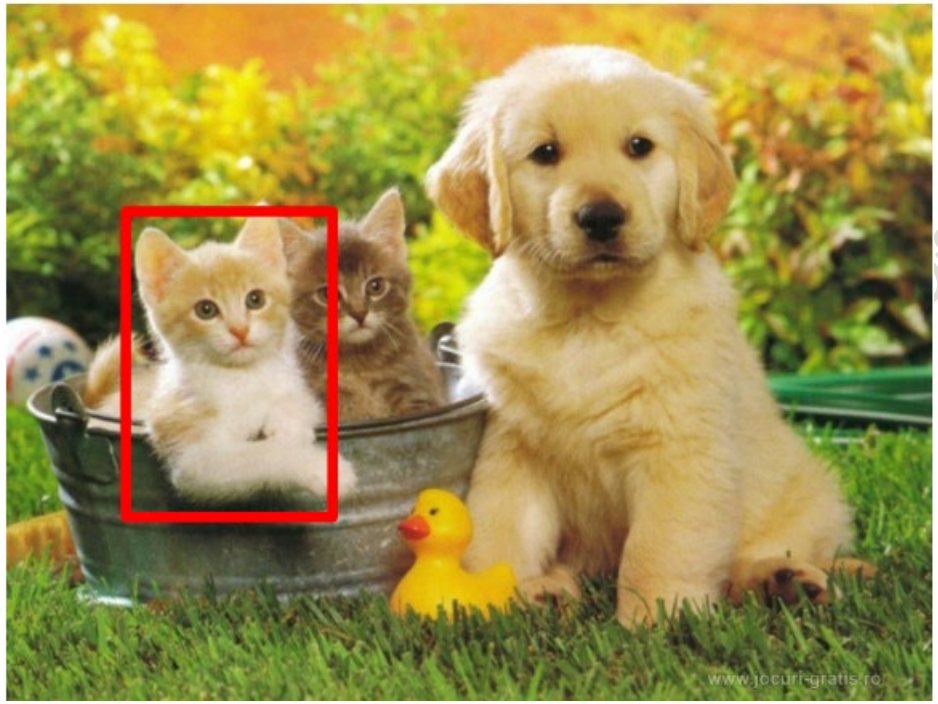
Classes = [cat, dog, duck]

Cat ? NO

Dog ? NO

Duck? NO

OBJECT DETECTION



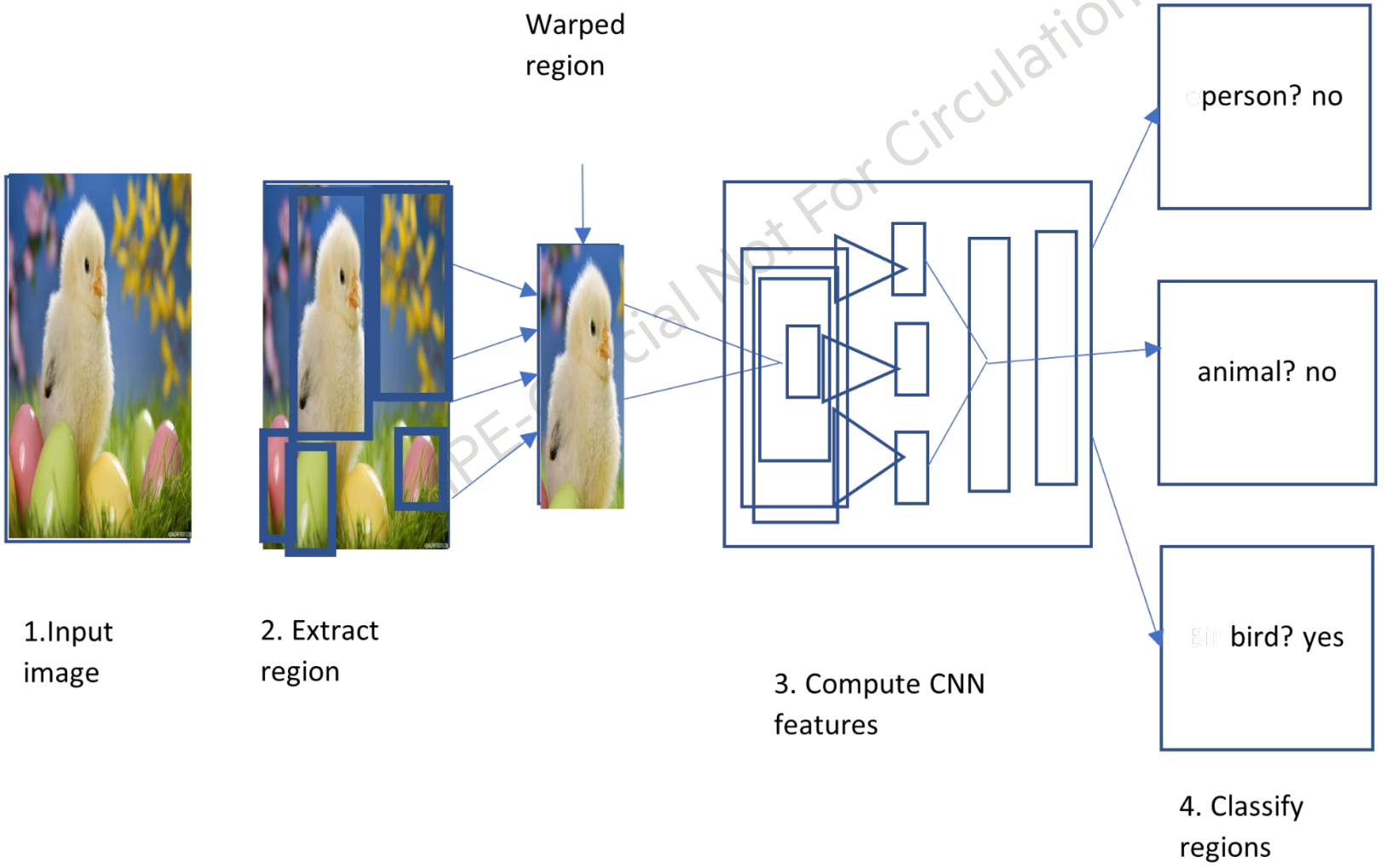
Classes = [cat, dog, duck]

Cat ? YES

Dog ? NO

Duck? NO

OBJECT DETECTION - REGION-BASED CONVOLUTIONAL NEURAL NETWORKS



TEXT EXTRACTION FROM IMAGES AND DOCUMENTS

- Text extraction from images and documents
- Plays vital role in information retrieval

Resources available:

- Journals
- Records
- Scanned documents
- Web pages
- Tv programs
- Currencies
- Business cards

Use Cases

- Document analysis
- Identifying number plates of vehicles
- Video analysis
- Detecting objects
- Content retrieval

TEXT EXTRACTION FROM DOCUMENT

- Detect the location of text:
- The text available in any scanned document:
    - various fonts, alignment, shapes, colours etc.
- Images in documents : Document image, Scene image and caption image
- Caption Texts: Imposed upon videos and images
- Scene texts: Texts present in the scene itself when the video or image is being recorded/created.
- Optical character Recognition (OCR):
    - Attributes considered by OCR: Text density, Text structure, Font and Artifacts.
    - Combination of AI with OCR: Better accuracy and analysis

TEXT EXTRACTION FROM DOCUMENT

1. Region-based method :
    - Sliding window is used to detect text
        - Cannot be generalised
2. Connected component based (CC- based)
    - Small sections of images are grouped to form large components.
3. Edge Based Method: Edges of letter and digits are identified.

TEXT EXTRACTION FROM DOCUMENT

4. Texture-based Method: Textures and properties of texts.

   • Parameters considered :Correlation, contrast, entropy etc.

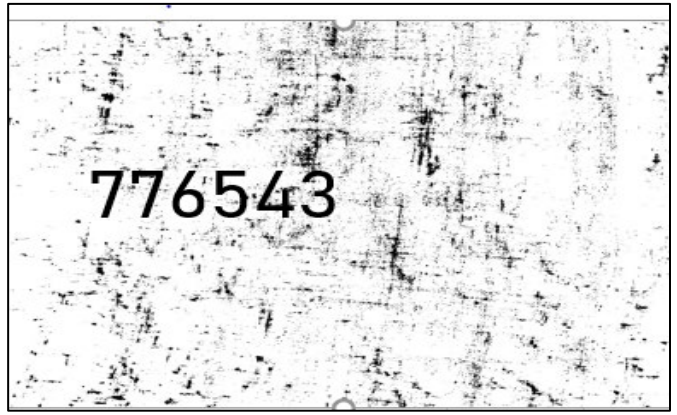
   • Generic, but high computational cost

5. Hybrid Techniques:

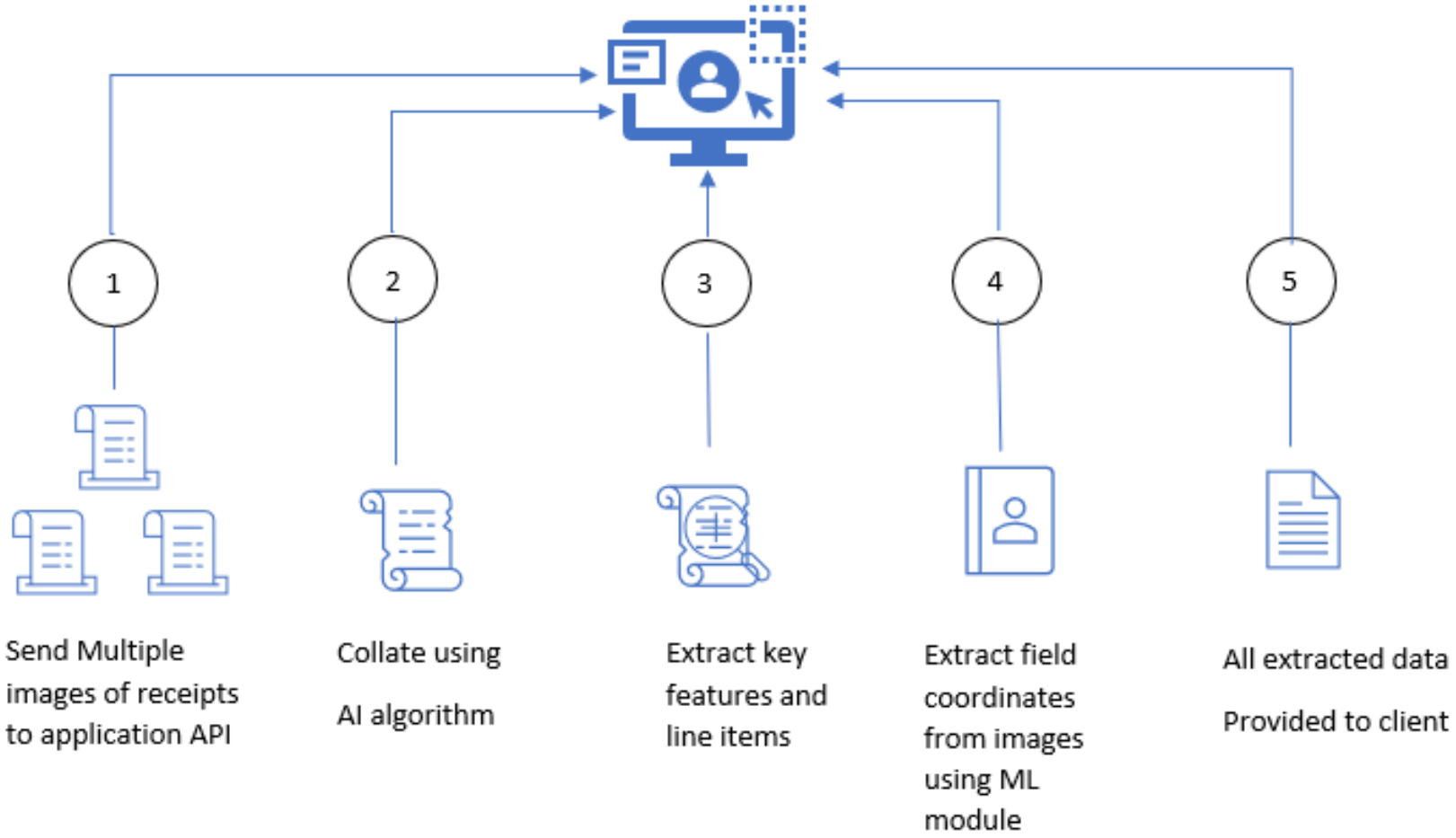
   • Combination of Region based, and Texture based methods.

6. Morphological Based Method: Text related features.

   • The statistical aspects of various words and usages of a particular languages.

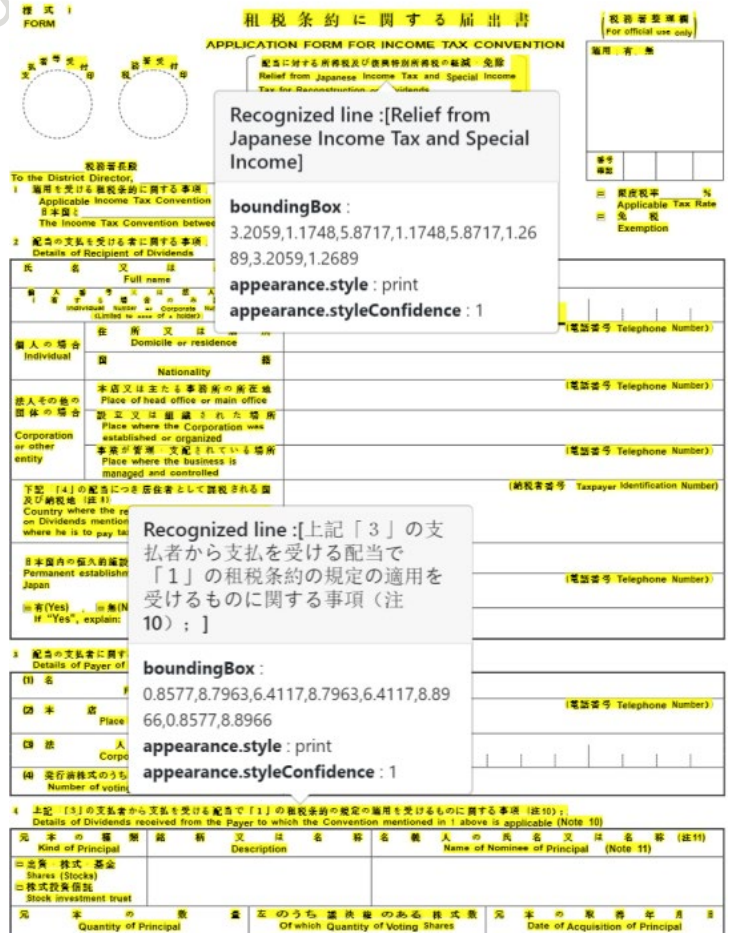


FORM RECOGNITION



FORM RECOGNITION

- Form Recognition helps to provide a well structure database that can be searched easily for information.
- In form recognizer key-values are extracted from form.



CONVOLUTIONAL NEURAL NETWORKS FOR COMPUTER VISION

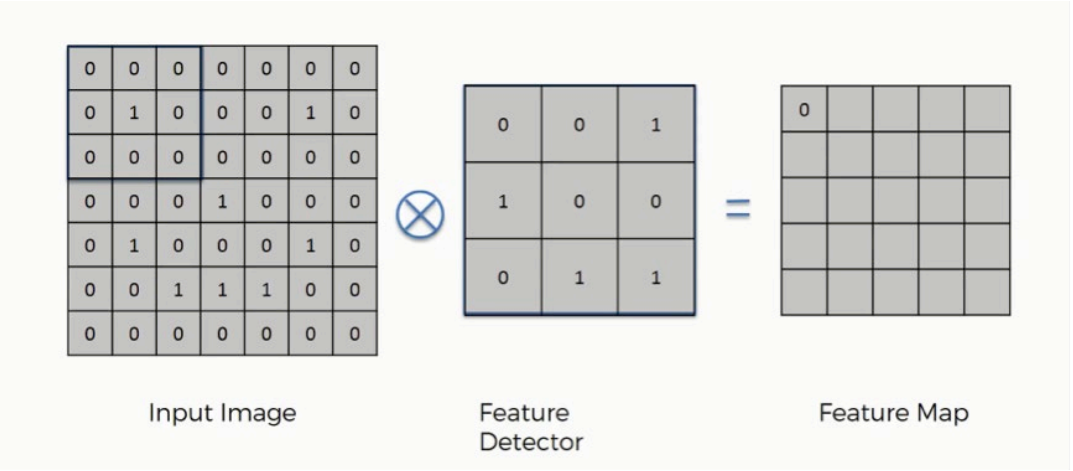
•How do self-driving cars recognize other cars as well as pedestrians and street objects?

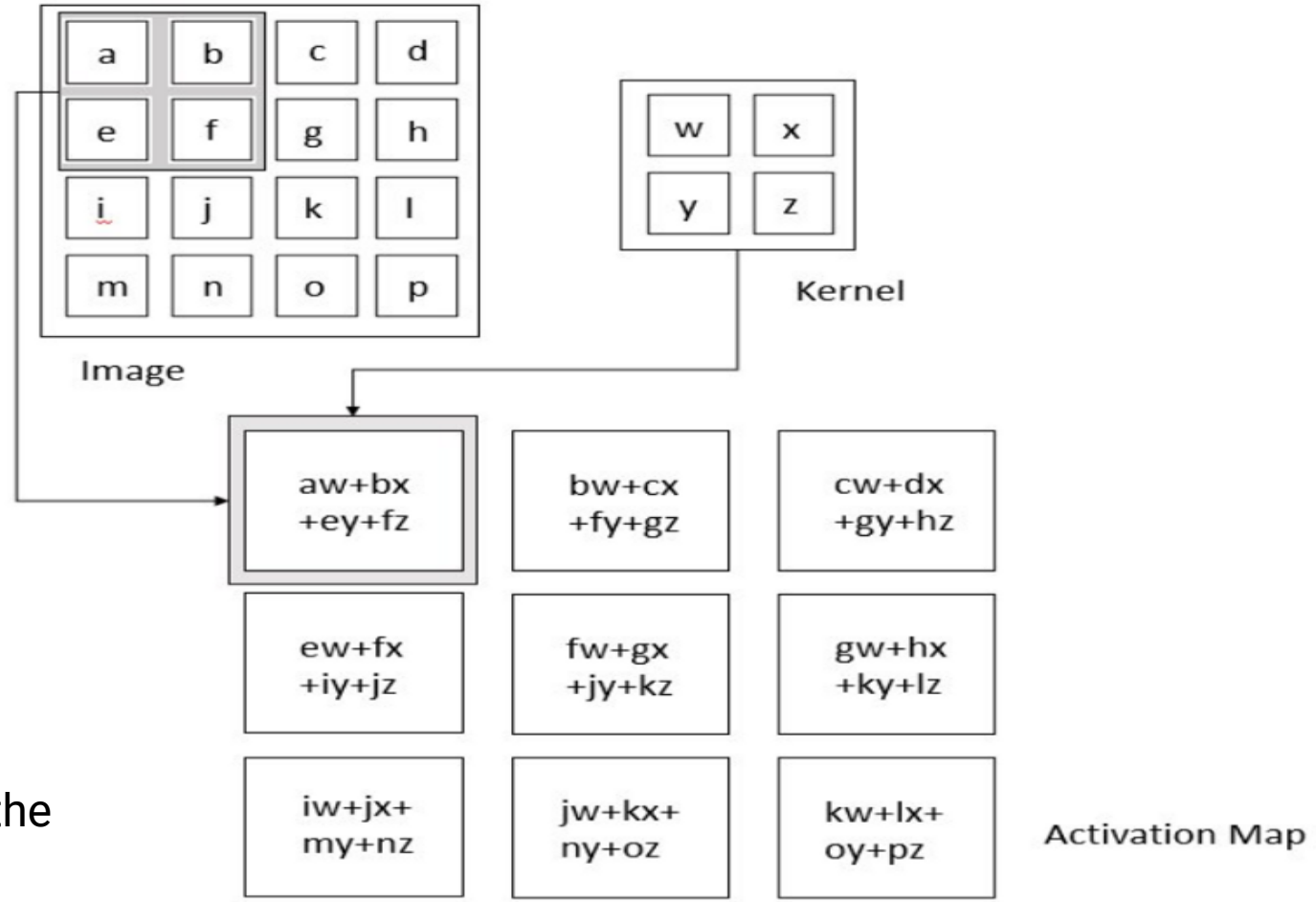
•How did Facebook go from making you tag people in images yourself, to being able to identify your friends and automatically tag them as it does now?

And the answer to both questions would be:

Convolutional Neural Networks.

CONVOLUTION LAYER OPERATION

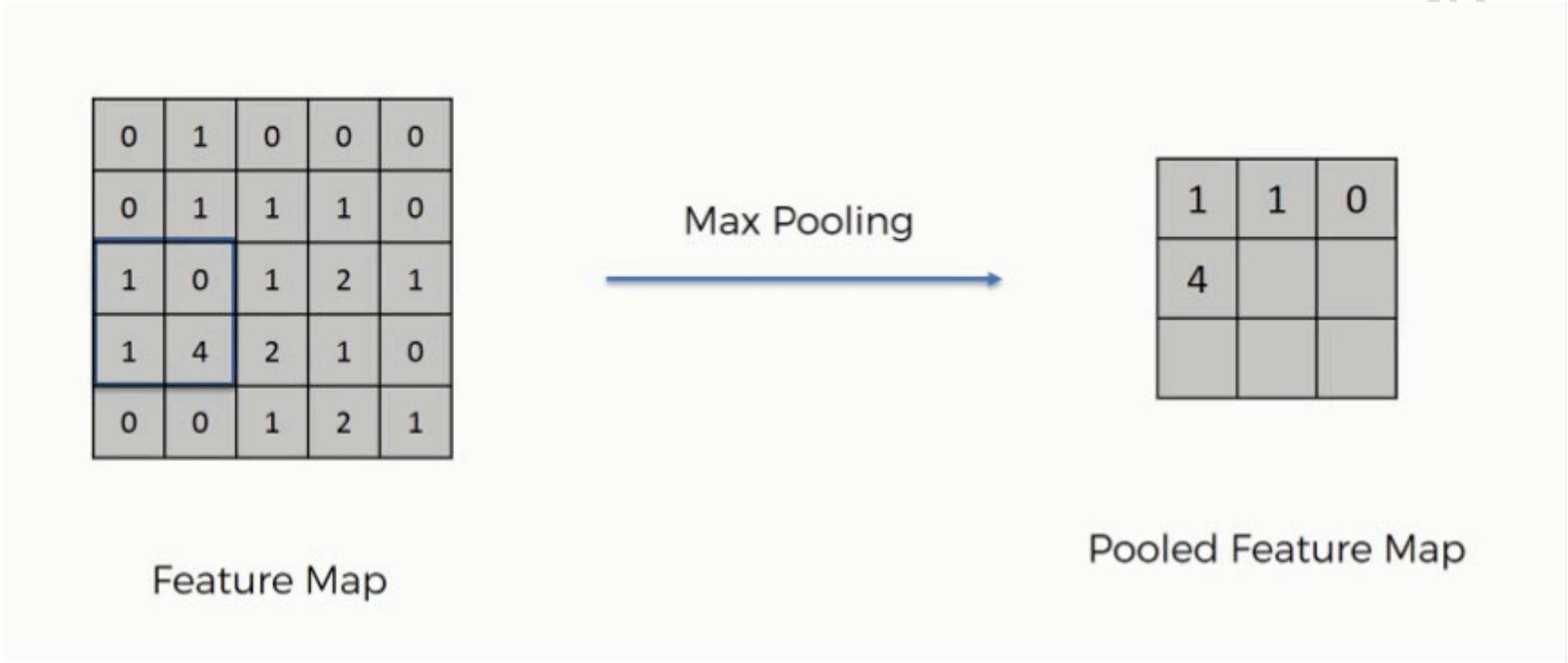




There are several uses that we gain from deriving a feature map:
Reducing the size of the input image,

Larger your strides (the movements across pixels), the smaller your feature map.

POOLING LAYER



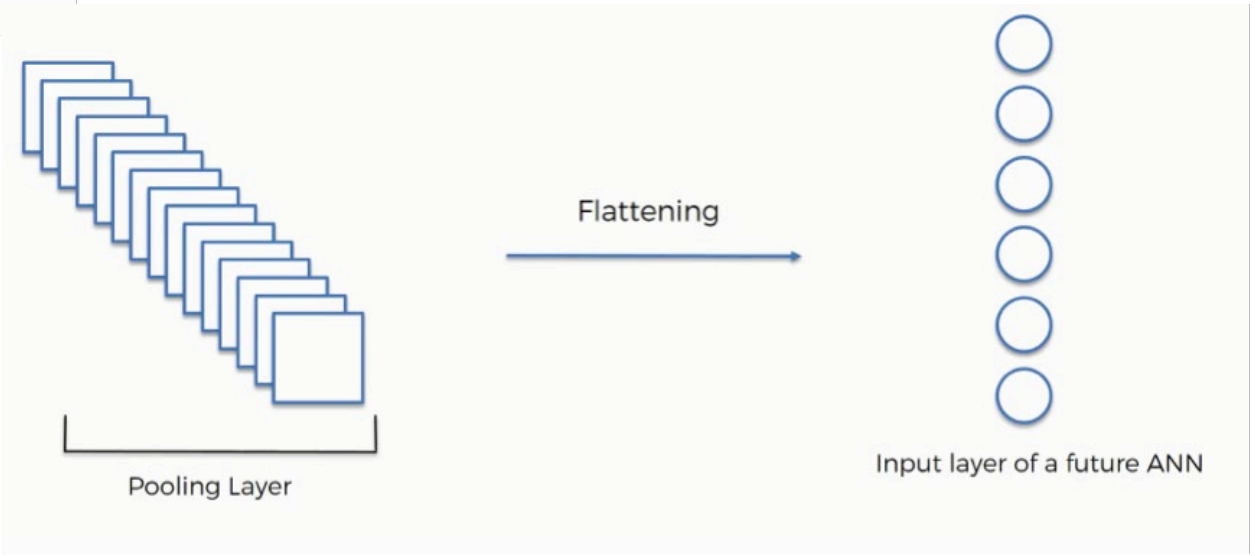
Types of Pooling

Mean pooling

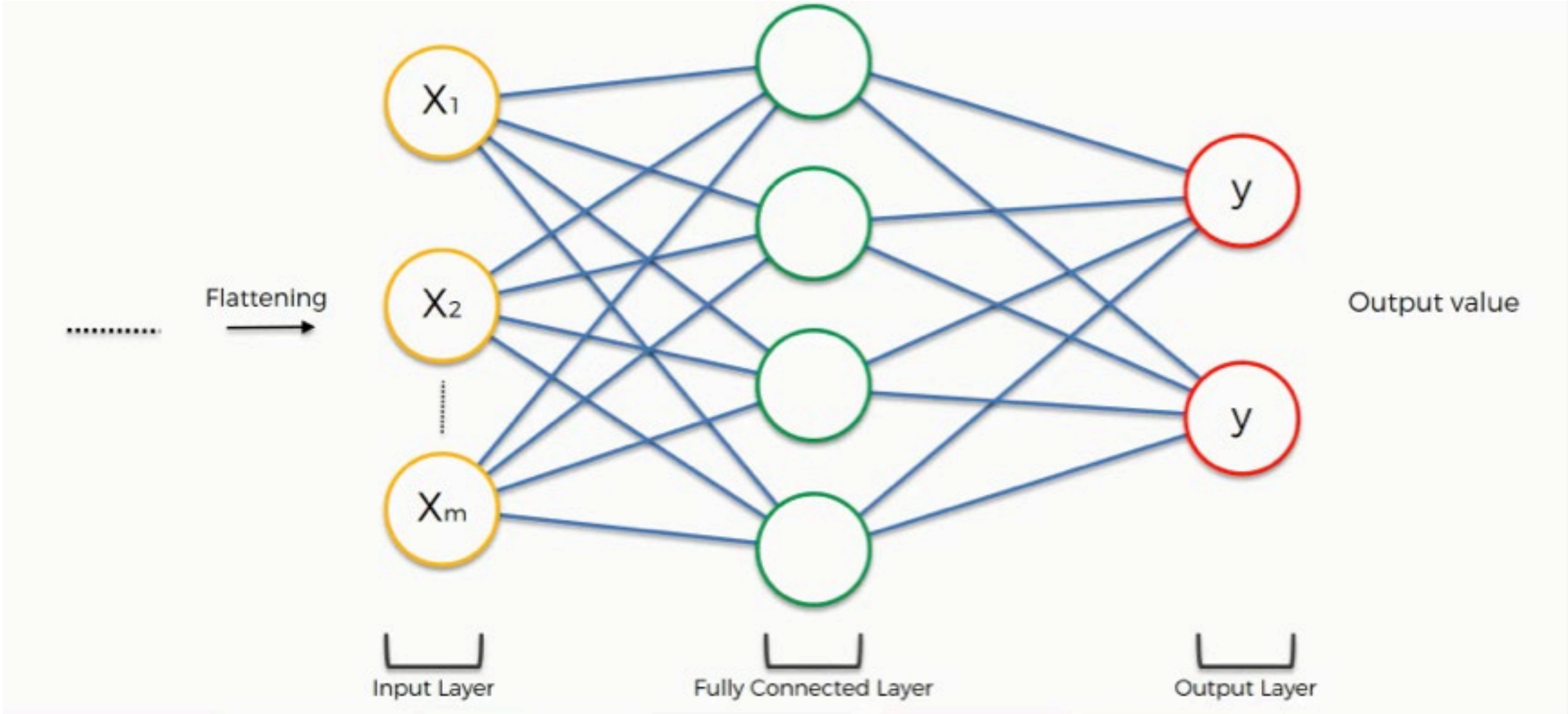
Max pooling

Sum pooling

FLATTENING LAYER



FULLY CONNECTED LAYER



CNN APPLICATIONS IN COMPUTER VISION

- Image Classification
- Image Segmentation
- Object Detection
- Image Caption Generation
- Text Recognition

DEMO

- GitHub link

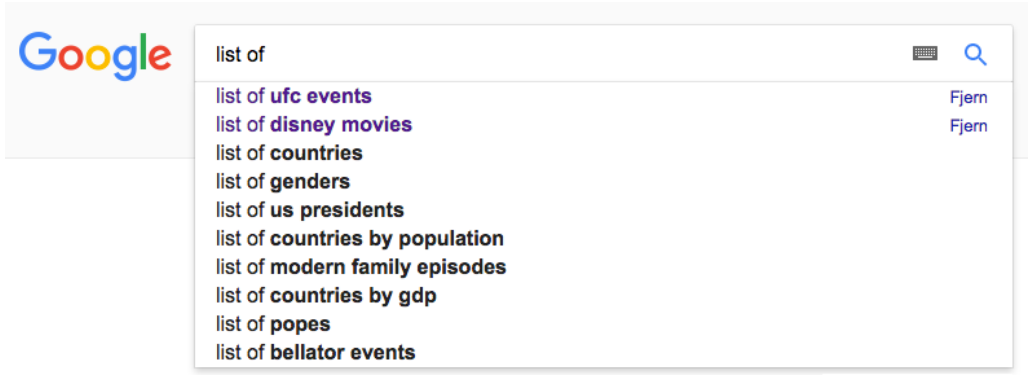
LESSON-3

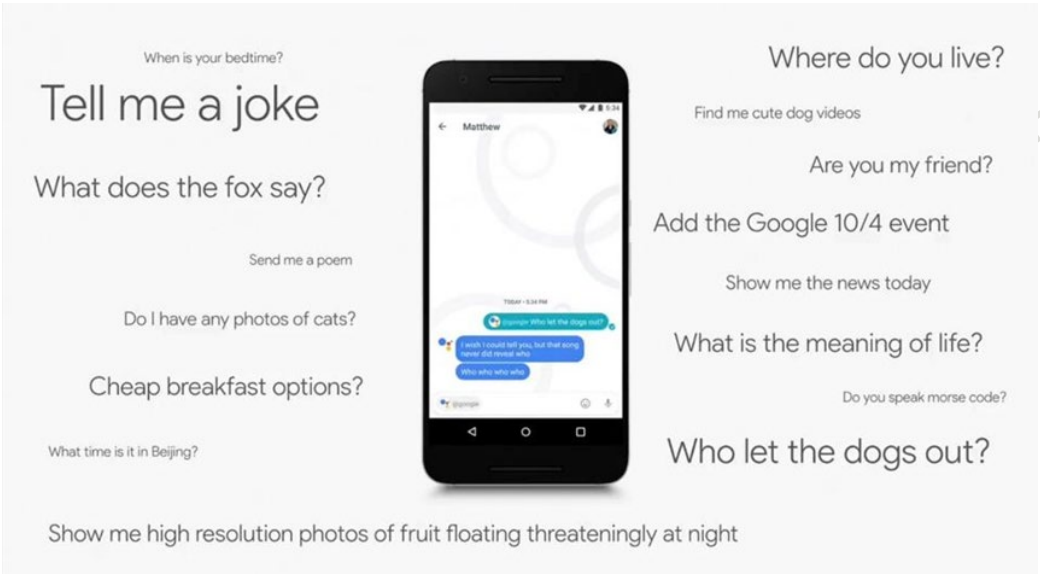
NATURAL LANGUAGE PROCESSING (NLP)

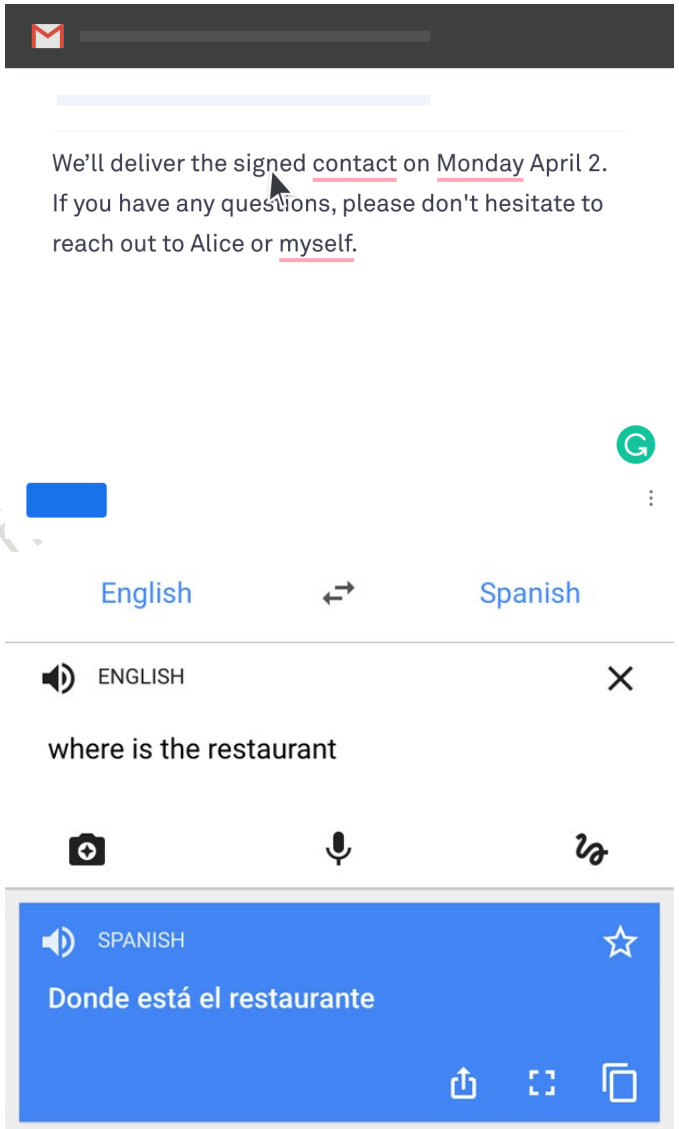
NATURAL LANGUAGE PROCESSING – AN INSIGHT

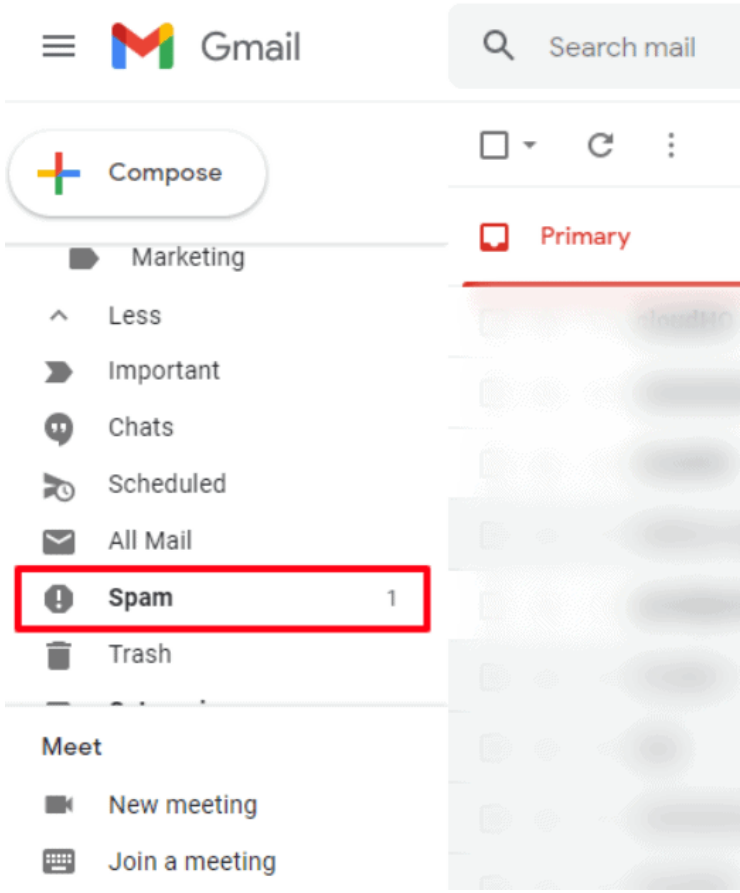
- Teaching machines to understand human language.
- Deals with making sense of written and spoken language.
- To make computers learn our language rather than we learn theirs.
- A culmination of Artificial Intelligence, Computer science & Linguistics
- Enables the extraction of structured data from natural language
- Deals with how computers and humans interact in their natural language.

Have you ever used NLP products?

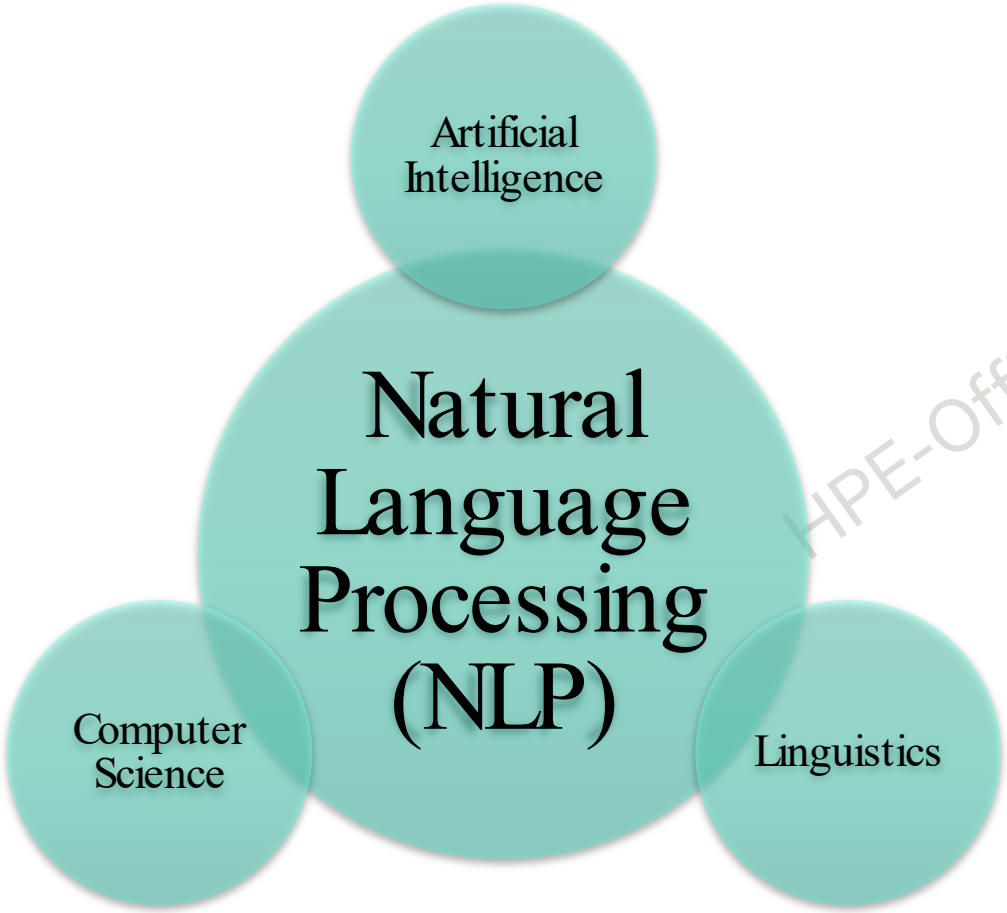








Applications of NLP



- Translators (Eg: Google translate)
- Conversion personal assistants (Eg: Alexa)
- Interactive Voice Response (IVR)
- Verification of grammatical correctness in a sentence (Eg: Grammarly)

Major areas of NLP

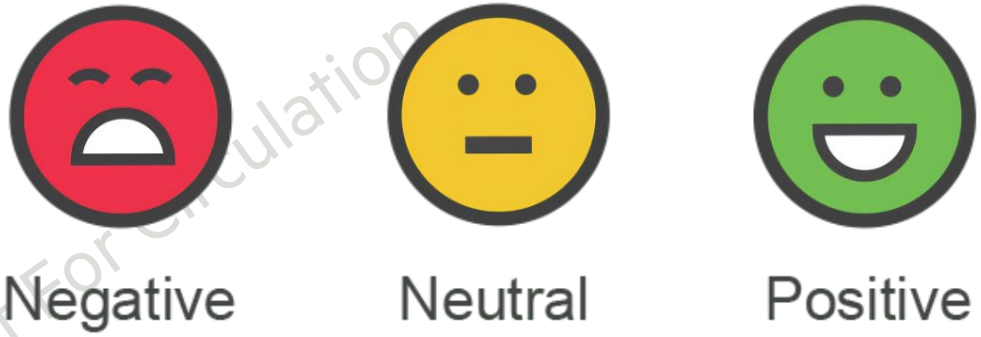
Text analysis and entity recognition

Sentiment analysis

Speech recognition and synthesis

Machine translation

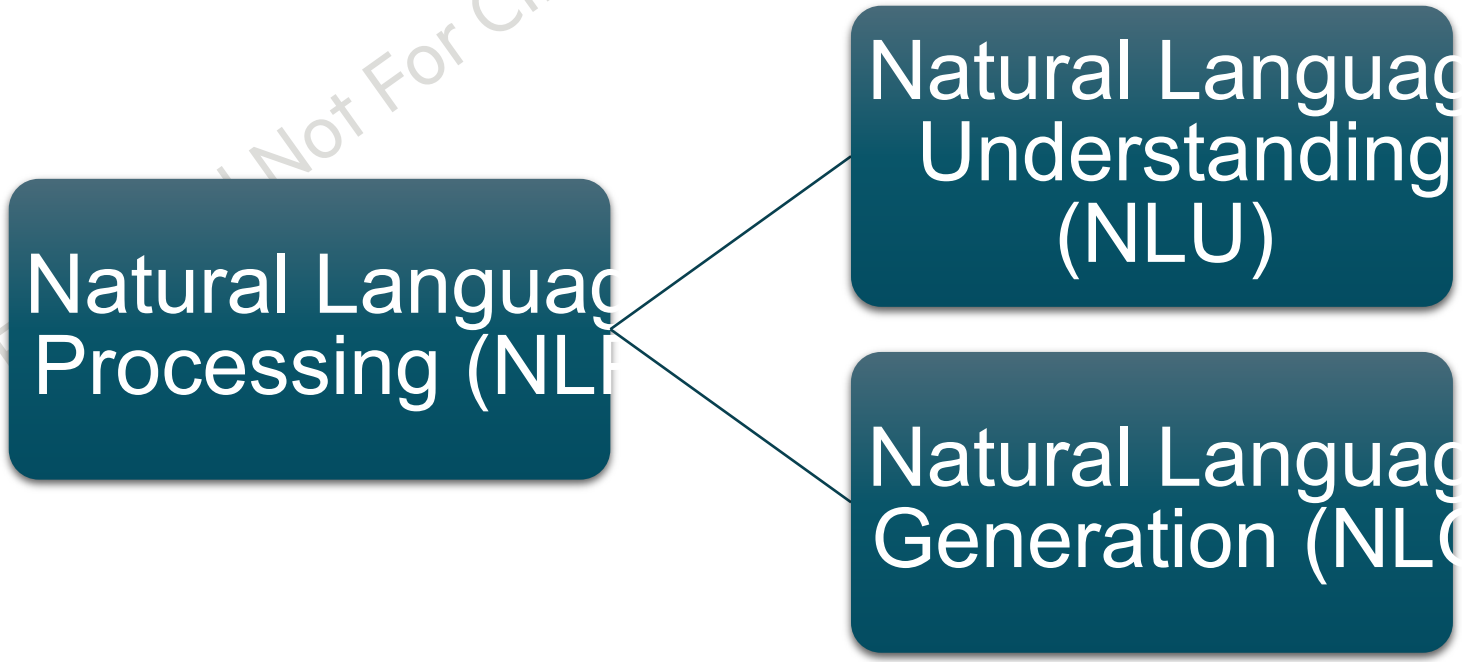
Semantic language modeling



COMPONENTS OF NLP

Natural Language Understanding (NLU) is the process of reading and interpreting language.

Natural Language Generation (NLG) is the process of writing or generating language.

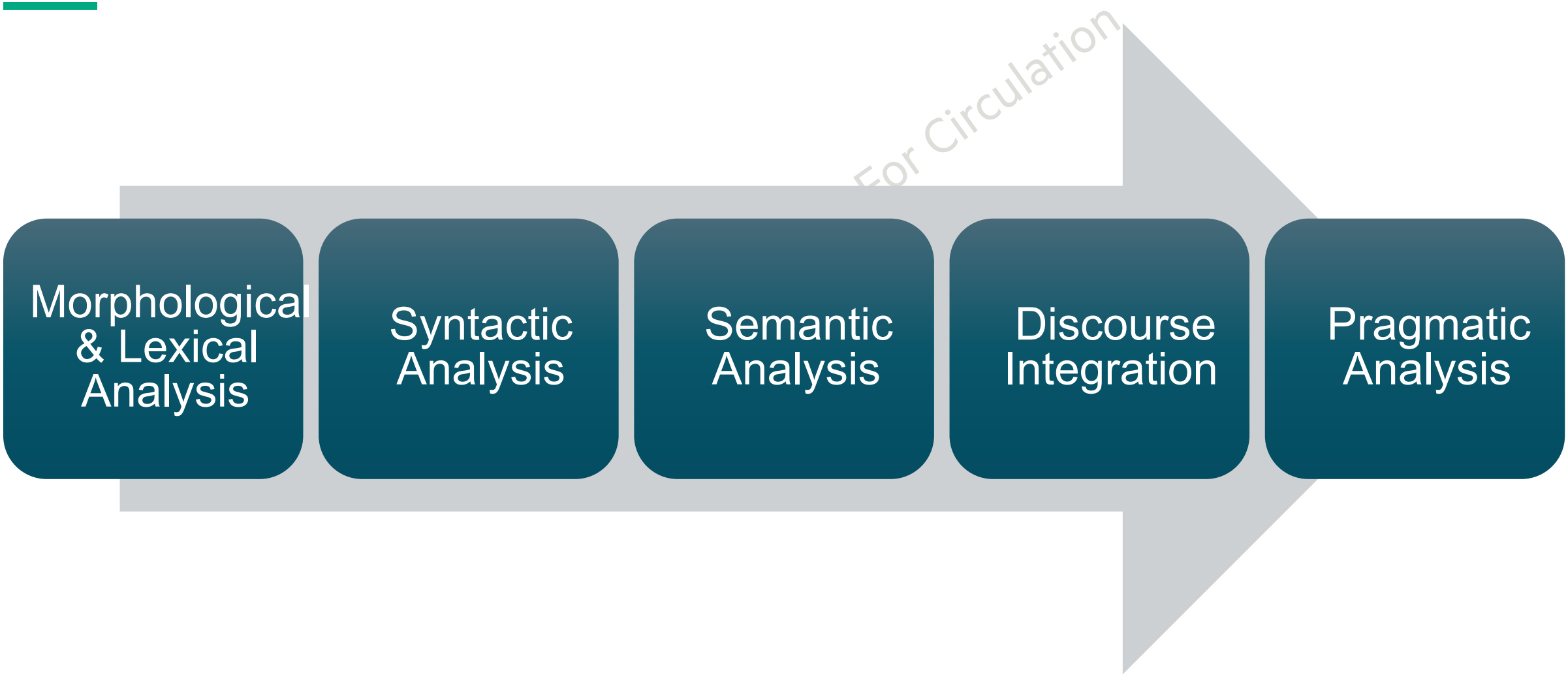


STEPS REQUIRED TO BUILD AN NLP MODEL

In general, there are three steps in creating an NLP-based model:

- Text Preprocessing
- Feature Extraction
- Modeling

TEXT PREPROCESSING



TEXT PREPROCESSING

Text Preprocessing can be divided into four major steps:

- Cleaning
- Tokenization
- Normalization
- Stop Words Removal

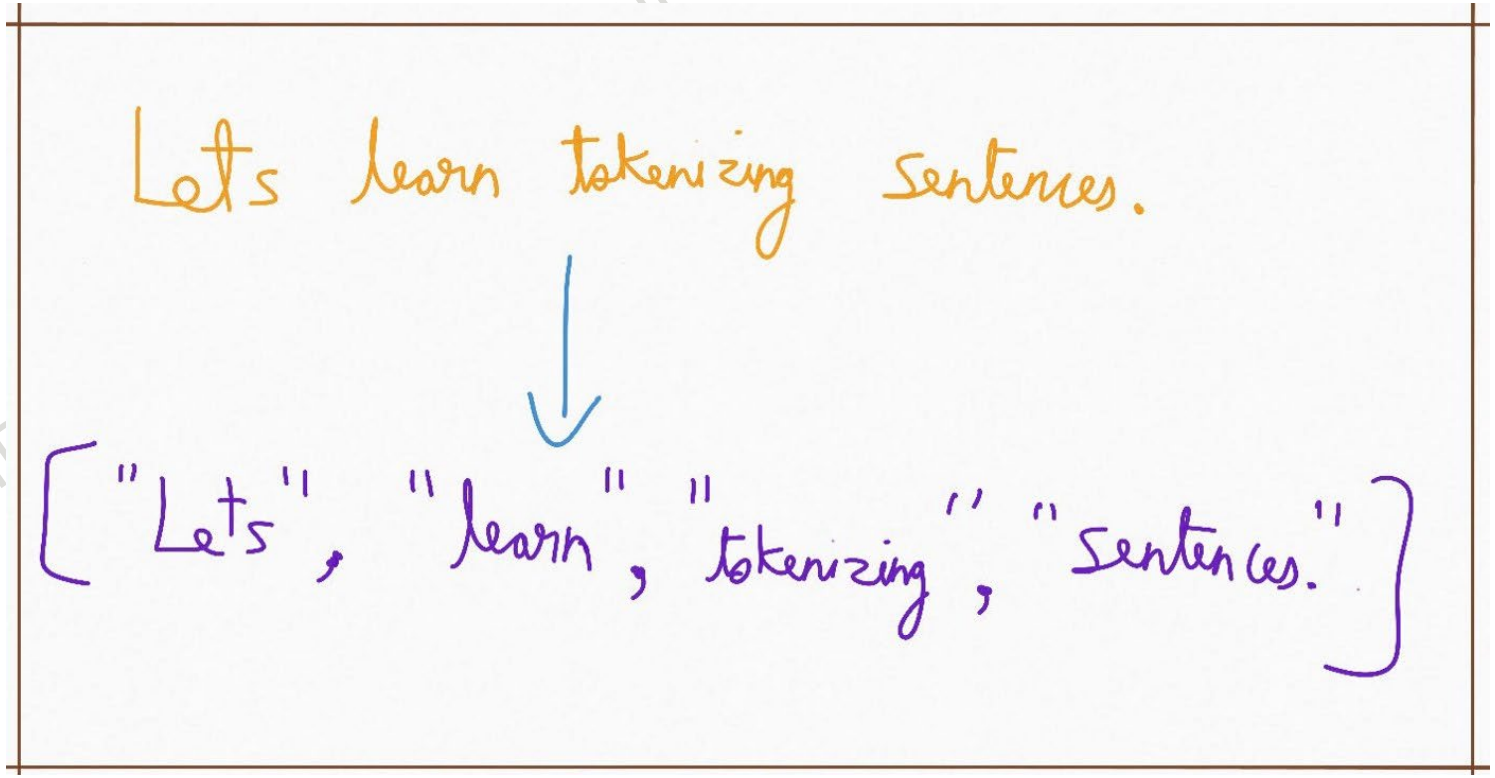
CLEANING

- Most prevalent types of noise found in a text data
- When working with documents from internet sources the text included within the HTML tags
- Numbers often do not contribute much meaning to a given piece of text and are therefore filtered out.
- Links can be of many forms and most of them consist of strange symbols or short-codes that can be present in your document.

TOKENIZATION

- Splitting a text into parts called tokens and dropping certain characters, like punctuation marks

- Tokenization varies with languages



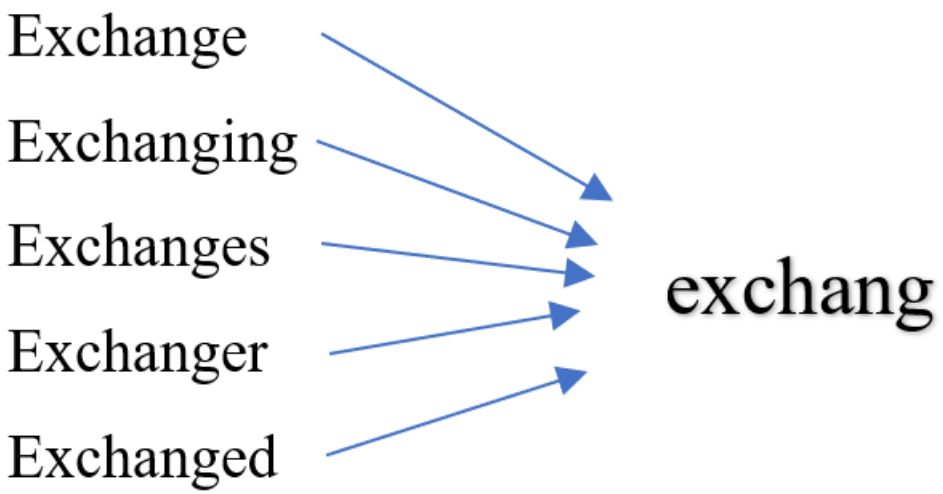
NORMALIZATION

- An important type of textual noise is about the multiple representations exhibited by a single word.

- Lemmatization and stemming are text normalization techniques used in Natural Language Processing (NLP).

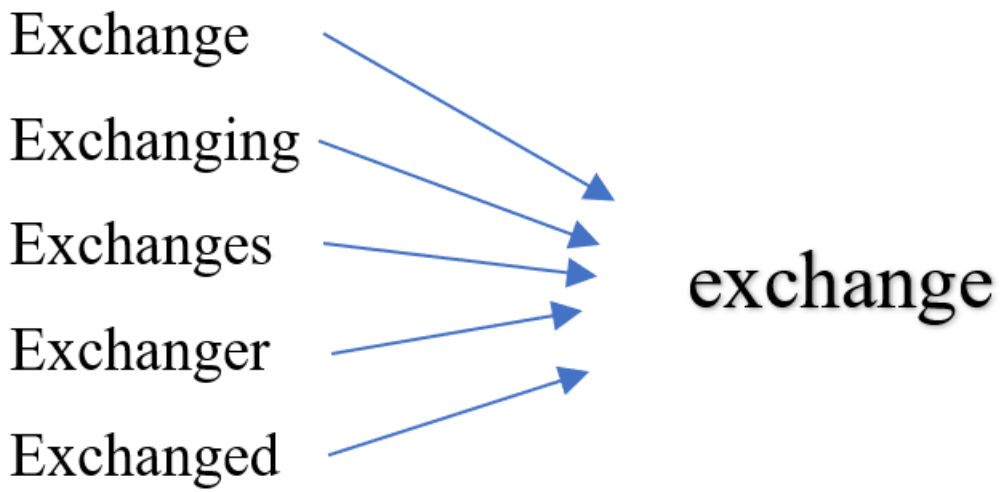
STEMMING

- Reducing the word to its base form
- In English, prefixes must always be derivational
- For eg: a prefix will create a new word, such as "eco" in "ecosystem"



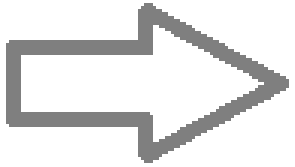
LEMMATIZATION

- Lemma is the base dictionary form of a word.
- Lemmatization is removing intonational endings to produce an accurate word from a dictionary.
- Words with the same meaning as their root are standardized by changing their past tense into the present (e.g., "went" is changed to "go"), and suitable synonyms (e.g., "best" is changed to "good").



STOP WORDS REMOVAL

The	~~The~~
quick	quick
brown	brown
fox	fox
jumped	jumped
over	over
the	~~the~~
lazy	lazy
dog	dog



- Stop words are those words that d not contribute to the deeper meaning of the phrase.
- They are the most common words such as: the, a, and is.
- For some applications like documentation classification, it ma make sense to remove stop words

SYNTACTIC ANALYSIS (PARSING)

- The arrangement of sentences properly to make a grammatical sense is known as syntax.
- The alignment of natural language with grammatical rules is assessed using syntactic analysis.
- This involves the following steps
  - Word Segmentation: It creates distinct units from large pieces of continuous text.
  - Part-of-speech tagging: For every word, parts-of-speech is identified.
  - Parsing: Analysis of Grammatical correctness for the provided sentence.
  - Sentence breaking: If it is a large piece of text, place sentence boundaries.

SEMANTIC ANALYSIS

- Checks for the meaningfulness of the given text.
- Verifies the semantically correct sentence and finds the meaning of the sentence
- Meaningful insights are retrieved from the text.

DISCOURSE INTEGRATION

- This stage checks for the discontinuity between the preceding and succeeding sentences

PRAGMATIC ANALYSIS

- Pragmatic analysis checks the references obtained in the previous phase (semantic) and correlates to the actual object/events in the given context.
- For example, "Place the ball on the floor in the basket" can have two semantic analyses, and one of them is chosen by a pragmatic analyzer depending upon the context.

MODELING

Building a machine learning or deep learning model is the next step after you have a numerical representation for your text documents.

Here are some models that you can work with:

- **Classical ML Algorithms**

  Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest (RF) are a few examples of classical machine learning models that perform well with text data for a variety of tasks, including sentiment classification and document classification, etc.

RECURRENT NEURAL NETWORKS (RNNS)

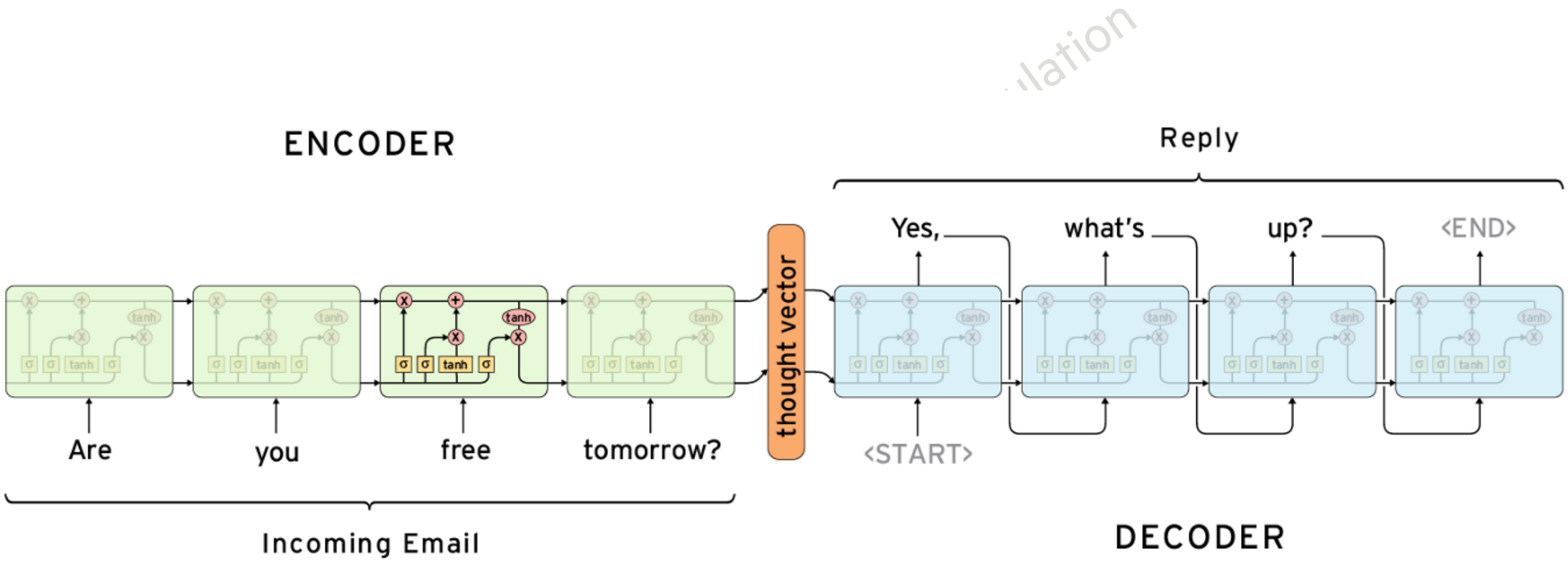
- RNNs are one of many useful deep learning algorithms that have been demonstrated to perform well with text data in addition to machine learning-based methods. Here are two of the most practical RNN types:

  a) Long Short Term Memory (LSTM)

  b) Gated Recurrent Unit (GRU)

- Some of the most well-known problems in NLP, such as neural machine translation, have been addressed using these architectures.

LSTM


ENCODER
Reply
Yes,
what's
up?
<END>
thought vector
Are
you
free
tomorrow?
<START>
Incoming Email
DECODER

DEMO

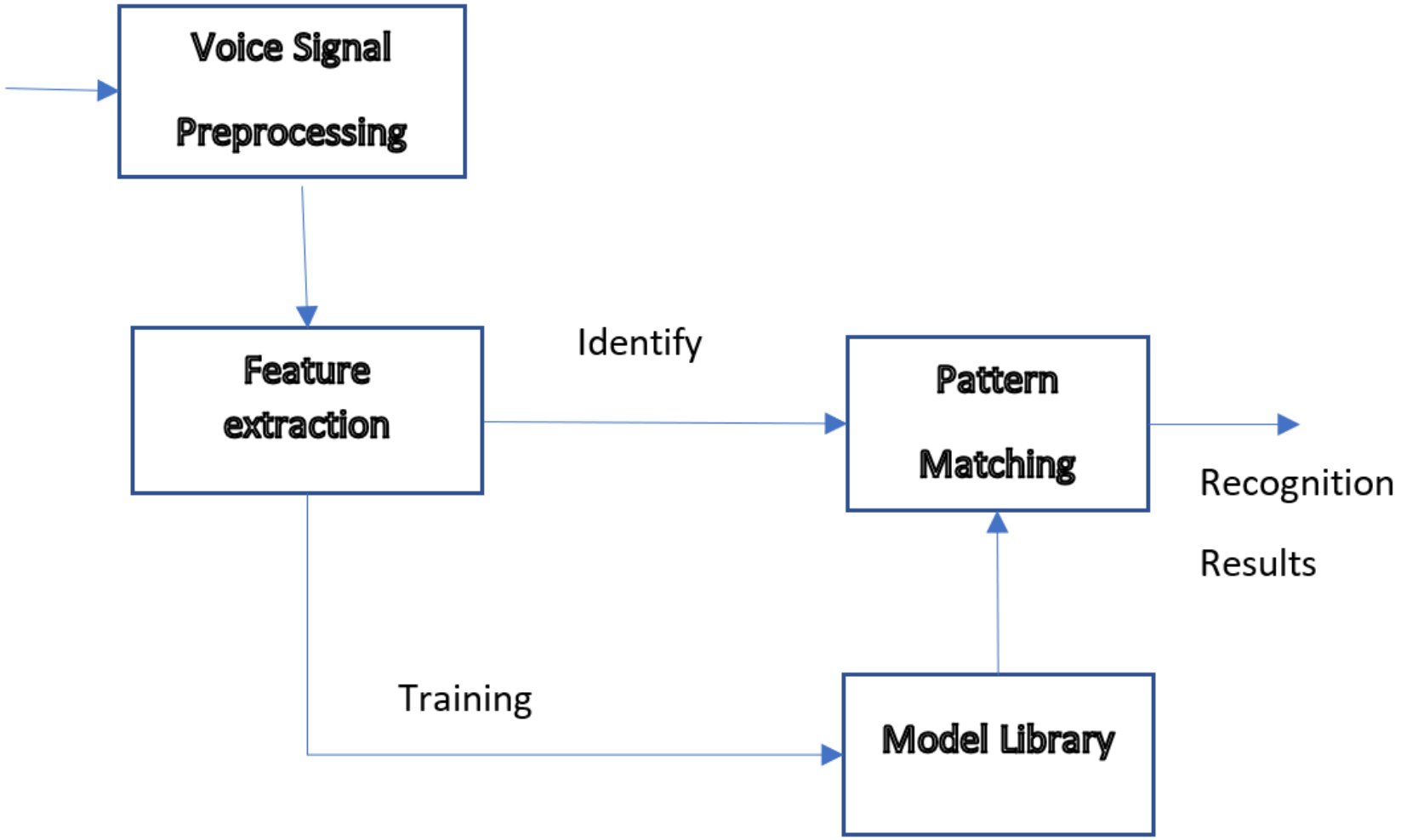
LESSON-4

SPEECH PROCESSING

SPEECH RECOGNITION

- Converting human sound signals to words or instructions is known as speech recognition.
- The speech recognition algorithms consist of feature extraction, acoustic, language, and unique algorithms.
- The speech signals that are collected are sent to feature extraction.
- The speech features that are obtained are sent to the model library module.

SPEECH RECOGNITION



PHONEME

- An instance of a phoneme makes a word's pronunciation and meaning distinctive from another word.
- For instance, the /s/ in 'soar' separates it from the /r/ in 'roar,' as it differs in both pronunciation and meaning from 'soar.'
- The smallest unit of sound that distinguishes one word's pronunciation and meaning from another is called a phoneme.

PROSODY

- In addition to providing meaning beyond words, prosody also provides semantic information.
- For example, when describing an upward motion, speakers naturally raise the pitch of their voice.

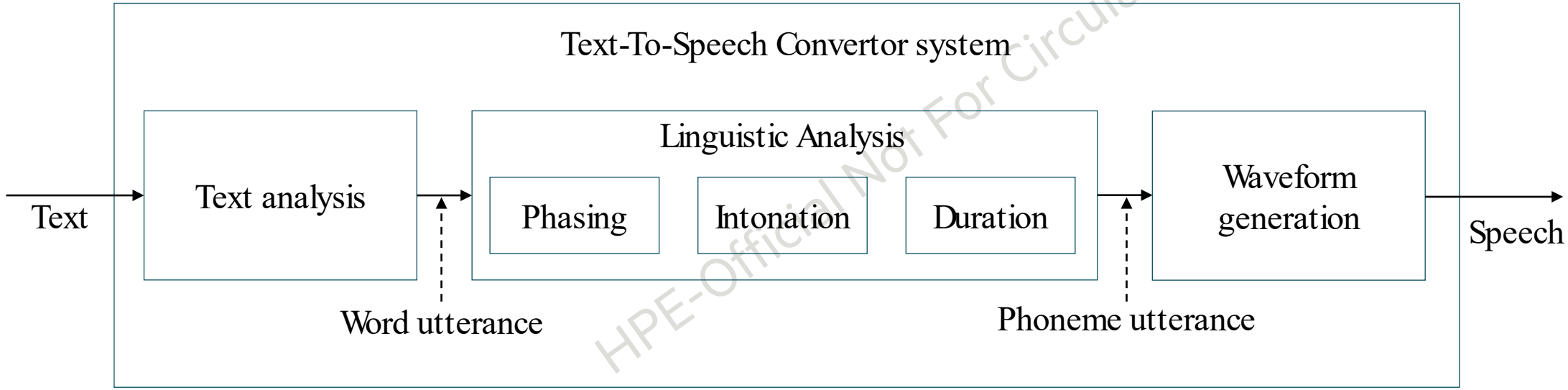
MEL-SPECTROGRAM

- The dimensionality of audio's short-time Fourier transform (STFT) is reduced by applying a non-linear transformation to the frequency axis.
- This method emphasizes low-frequency details that are crucial for distinguishing speech from noise de-emphasizes high-frequency details.

SPEECH SYNTHESIS

- Speech synthesis is the production of human speech artificially.
- It is a text-to-speech converter, whereas speech recognition is a speech-to-text converter.
- It is a technology used for intelligent speech interaction systems.
- Technologies available
  - Concatenation synthesis
  - Parametric synthesis

TEXT TO SPEECH CONVERTOR



TEXT TO SPEECH CONVERTOR

- A text-to-speech convertor is made up of two components namely, the front-end and the back-end modules.

- The front end is responsible for text normalization/preprocessing/tokenization after which each word is assigned a phonetic transcription thereby dividing and marking the text into prosodic units such as those of phrases, clauses and sentences.

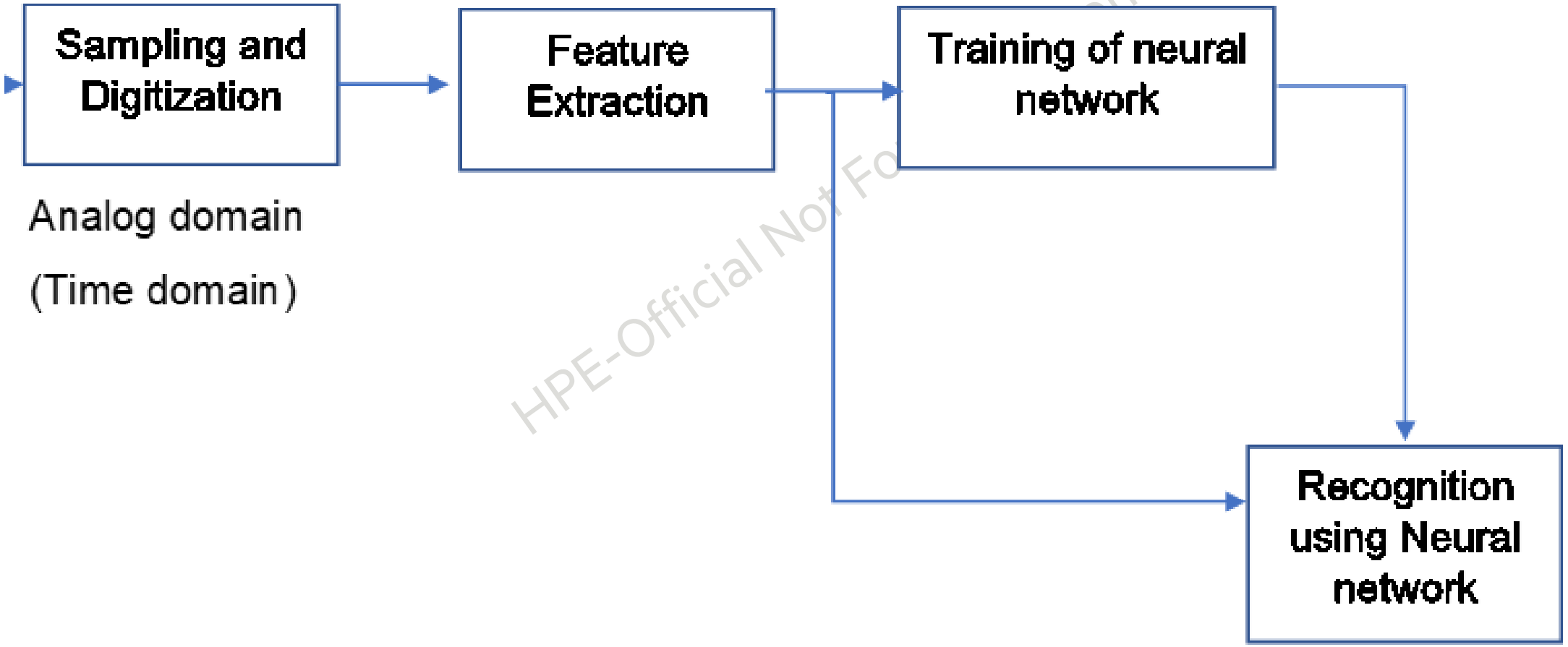
- Both the phonetic transcription and the prosodic information together result in a symbolic linguistic representation which is the output of the front-end module.

- The backend module, also called as the synthesizer converts the symbolic linguistic representation into audio output by computing the target prosody (pitch, contour and phoneme durations).

PARAMETRIC SYNTHESIS

- This approach uses recorded human voices which are modified using a set of functions and parameters.
- The two parts of statistical parametric synthesis are training and synthesis.
- The audio sample characteristics such as fundamental frequency (voice source), frequency spectrum (vocal tract), and duration(prosody)of speech are extracted during training.
- A statistical model like Hidden Markov Model (HMM) is later used to estimate those parameters.
- A set of parameters are generated using HMM's from the target text sequence.
- The final speech is then synthesized from these parameters.
- Deep learning-based methods can improve the quality of speech generated

ANN FOR SPEECH PROCESSING



DEMO

- GitHub link

THANK YOU