LECTURER: Nghia Duong-Trung

ARTIFICIAL INTELLIGENCE

TOPIC OUTLINE

UNIT 4.3

MODERN ARTIFICIAL INTELLIGENCE SYSTEMS

STUDY GOALS

On completion of this unit, you will have…

- … understanding of basic components within Natural Language Processing (NLP) components.
- … knowledge of main steps within Computer Vision (CV) applications.
- … a basic idea of common features extracted from images.

EXPLAIN SIMPLY

1. Describe the concept of stemming within NLP using your own words with the example of change.
2. Explain the relation between AI, ML and Computer Vision.
3. Can you list some of the typical features used within Computer Vision?

NATURAL LANGUAGE PROCESSING (NLP)

Speech recognition

identification of words in spoken language

Language understanding

extraction of meaning from words and sentences, and reading comprehension

Language generation

ability to express information

NATURAL LANGUAGE PROCESSING (NLP)

Speech recognition

Virtual assistants

Language understanding

Spam classification

Sentiment analysis

Language generation

Summaries

Chatbots

Semantics

Pen = writing tool

Semantics

Word stemming

{Change, changing, changer, changed} = **chang**

NLP COMPONENTS

Semantics

Word stemming

Part-of-speech-recognition

I	**like**	**reading**	**books**
Pronoun	*Verb*	*Verb*	*Noun*

NLP COMPONENTS

Semantics	Word stemming	Part-of-speech-recognition	Named entities recognition (NER)

Jane Doe is studying at IU since 2020.

Person *Org* *Date*

NLP COMPONENTS

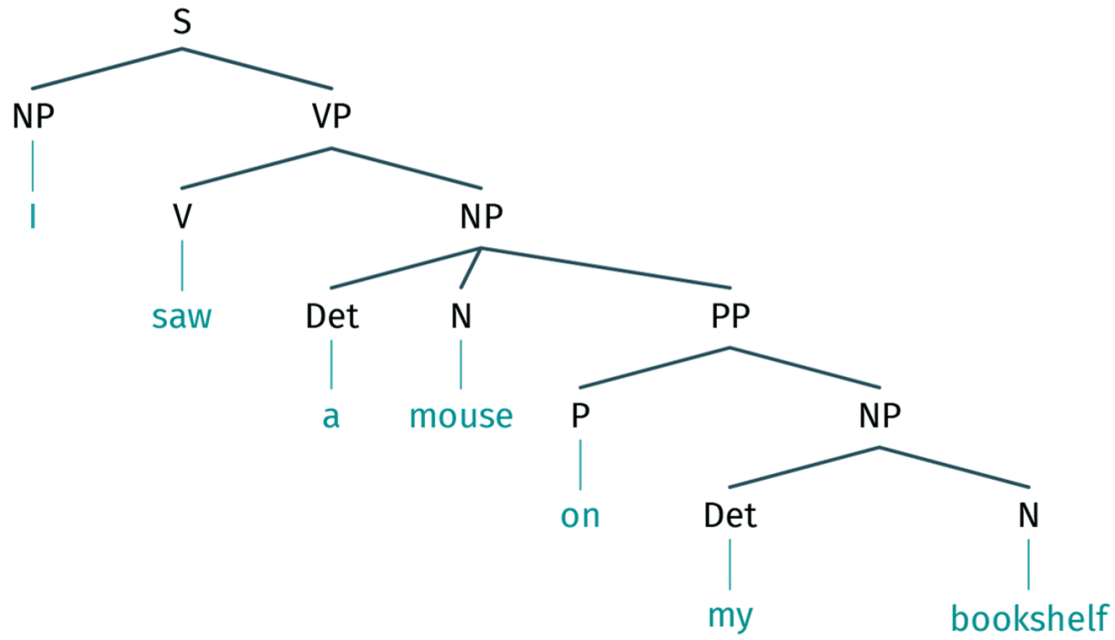
Semantics

Word stemming

Part-of-speech-recognition

Named entities recognition (NER)

Parsing



Source of the graphic: IU International University, Course Book DLMAIAI01.

NLP COMPONENTS

Semantics

Word stemming

Part-of-speech-recognition

Named entities recognition (NER)

Parsing

Sentiment analysis

The product is good.

The product can be improved.

The product is useless.

NLP COMPONENTS

Semantics	Word stemming	Part-of-speech-recognition	Named entities recognition (NER)
Parsing	Sentiment analysis	Question and answering	

Question: Who was the first president of the United States?

Answer: George Washington

SUBDOMAINS OF NLP

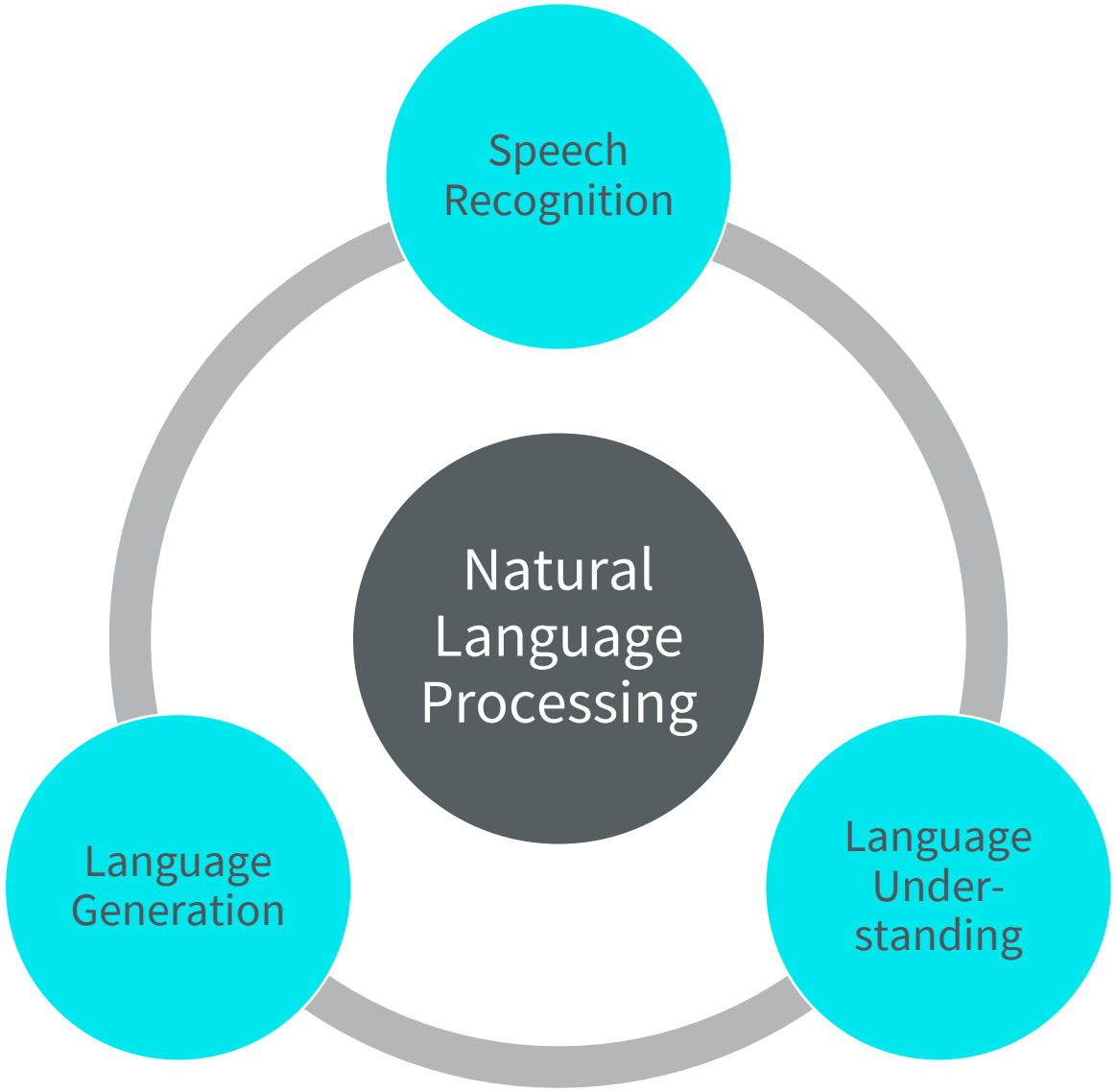


Image Source: Custom Depiction

SPEECH RECOGNITION

- Speech Recognition is a subfield of computational linguistics dealing with the recognition and translation of spoken language into text by computers, a process known as “speech to text” in some cases.
- The phrase “speech recognition” refers to the process of converting spoken words into text in general; however, subfields such as voice recognition and speaker identification specialize in identifying both the spoken content and the speaker’s identity.

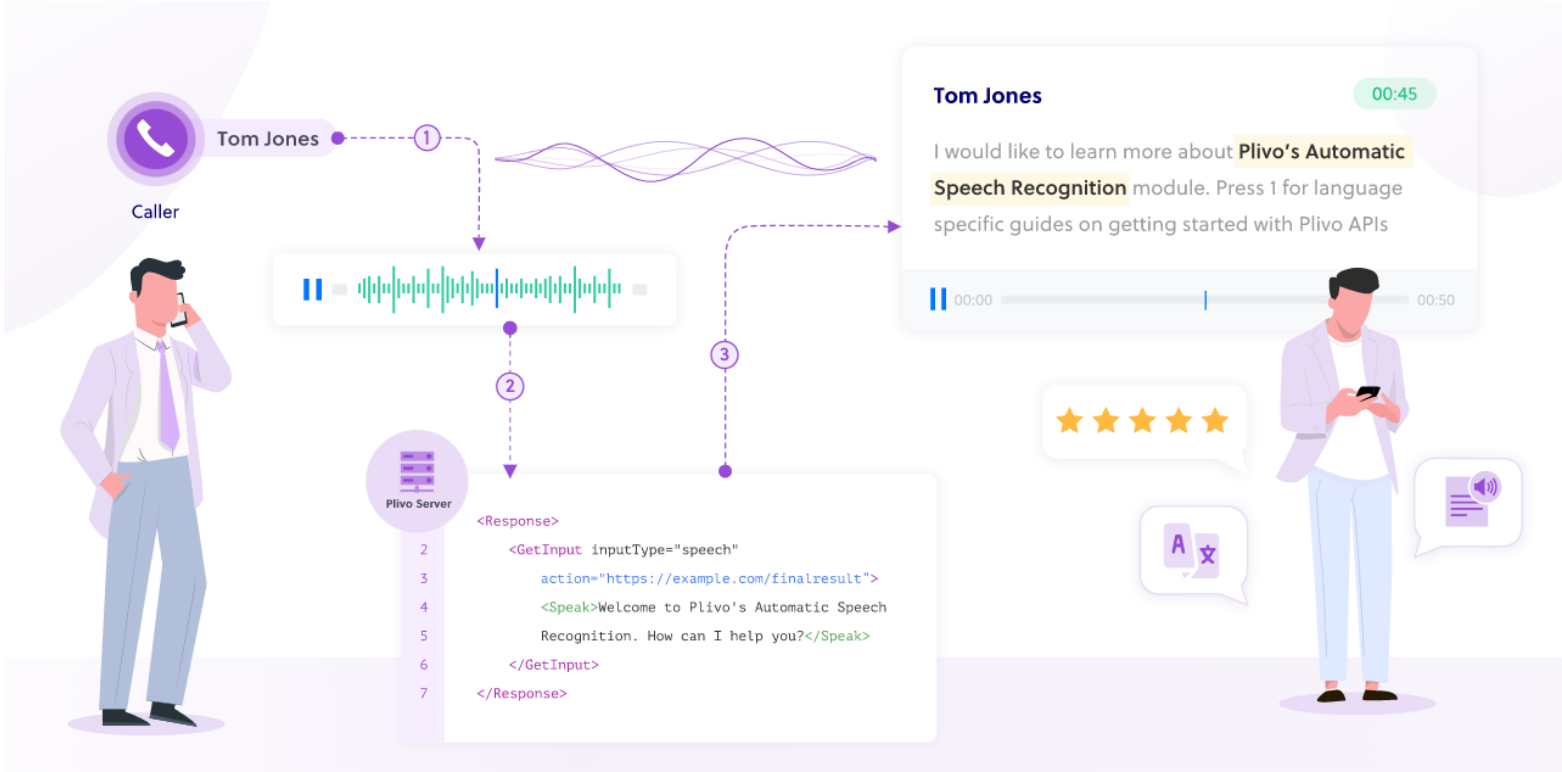


Image Source: Medium.com, Plivo.com

There are two types of speech recognition systems, speaker-dependent, and speaker-independent

- Speaker-dependent systems are designed in such a way that training is required. This works by having a speaker read text into the system or a succession of discrete vocabulary. The algorithm will then analyze the vocal recordings and link them to the text collection.
- Speaker independent systems are speech recognition systems that do not rely on vocal training.

There are two types of models used in speech recognition systems:

- Acoustic Model: A file containing statistical representations of each of the various sounds that make up a word is known as an acoustic model. A phoneme is a label given to each of these statistical representations. There are approximately 40 distinct sounds in the English language that is suitable for speech recognition, resulting in 40 separate phonemes.

- Language Model: To discriminate between words that sound similar, sounds are matched with word sequences. We presume our audio sample is grammatically and semantically sound, even if it is not grammatically perfect or has skipped words. As a result, incorporating a language model into decoding can enhance ASR accuracy.

Steps involved in the process of speech recognition:

- **Analog-to-Digital Conversion**: In most cases, speech is recorded and available in analog format. To convert analog voice to digital utilizing sampling and quantization techniques, standard sampling techniques or devices are available. A one-dimensional vector of voice samples, each of which is an integer, is typically used to represent digital speech.
- **Speech Pre-processing**: Background noise and long periods of quiet are common in a recorded conversation. Identification and removal of silent frames, as well as signal processing techniques to reduce/eliminate noise, are all part of speech pre-processing. Following pre-processing, the speech is divided into 20-second frames for subsequent feature extraction stages.
- **Feature Extraction**: It is the conversion of speech frames into a feature vector that specifies which phoneme or syllable is being spoken.
- **Word Selection**: The sequence of phonemes/features is translated into the spoken word using a language model/probability model.

SPEECH RECOGNITION & NATURAL LANGUAGE PROCESSING (NLP)

- The combination of linguistics and machine learning (ML) is known as Natural Language Processing (NLP). To produce actionable results, NLP seeks to understand human-human and human-computer interactions in the form of language (voice or text). NLP is an ML application in which machines "learn" to understand the natural language from millions of example datasets.
- Speech recognition is concerned with turning speech input into text, whereas NLP is concerned with "understanding" language in order to feed subsequent activities
- Applications
  - Automatic subtitle with speech recognition
  - Mobile telephony, including mobile email
  - People with disabilities
  - Home automation
  - Virtual assistant
  - Dictation

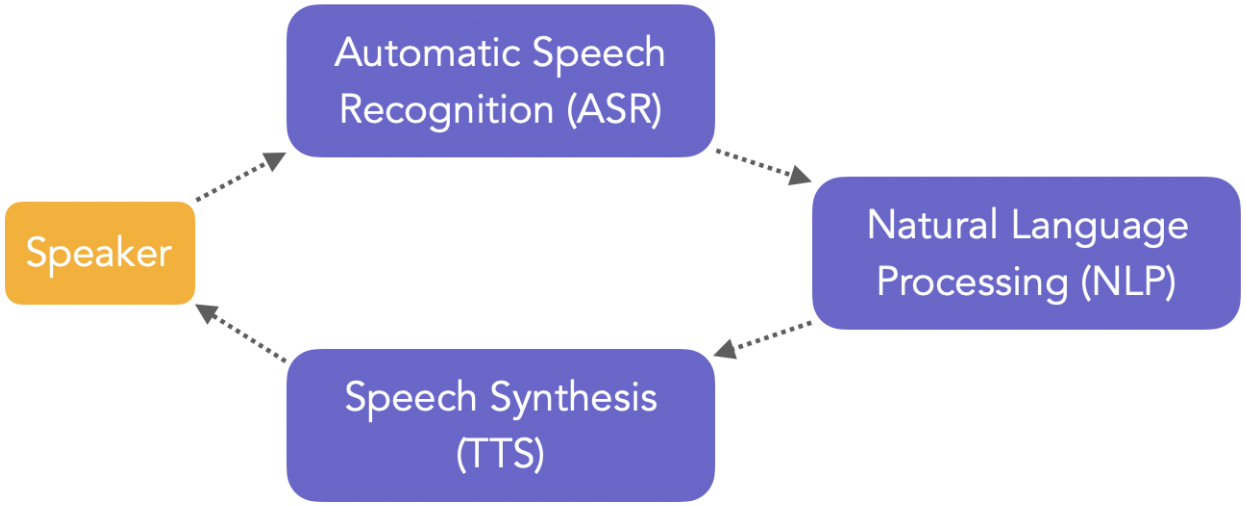


Image source: Medium.com

(NATURAL) LANGUAGE GENERATION

- Natural language generation (NLG) is a software process that produces natural language output.
- NLG is characterized as "the subfield of artificial intelligence and computational linguistics that is concerned with the construction of computer systems than can produce understandable texts in English or other human languages from some underlying non-linguistic representation of information".
  - Common applications of NLG methods include the production of various reports, for example weather and patient reports; image captions; and chatbots.

(NATURAL) LANGUAGE GENERATION

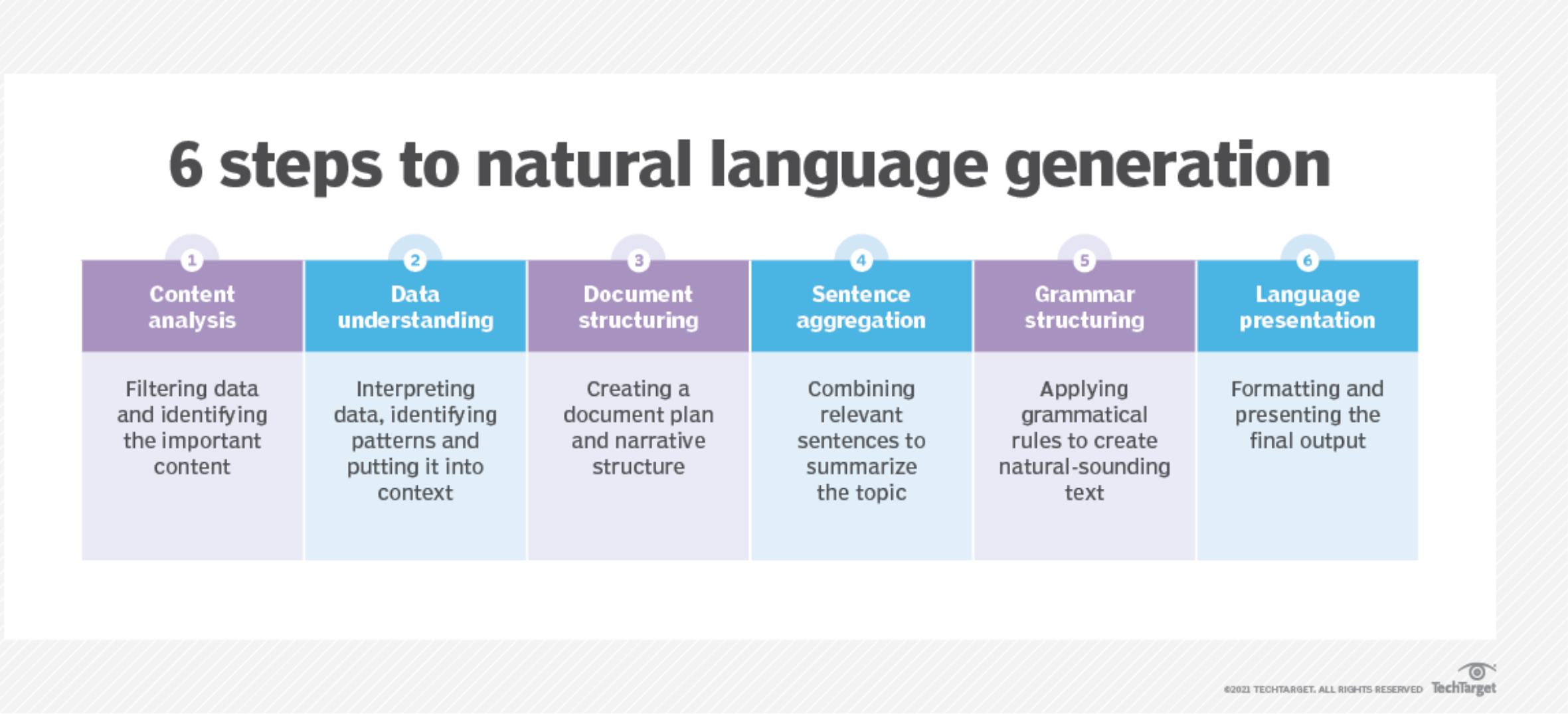


Image source: techtarget.com

(NATURAL) LANGUAGE GENERATION

Natural language generation is being used in an array of ways. Some of the many uses include the following:

- generating the responses of chatbots and voice assistants such as Google's Alexa and Apple's Siri;
- converting financial reports and other types of business data into easily understood content for employees and customers;
- automating lead nurturing email, messaging and chat responses;
- personalizing responses to customer emails and messages;
- generating and personalizing scripts used by customer service representatives;
- aggregating and summarizing news reports;
- reporting on the status of internet of things devices; and
- creating product descriptions for e-commerce webpages and customer messaging.

(NATURAL) LANGUAGE GENERATION

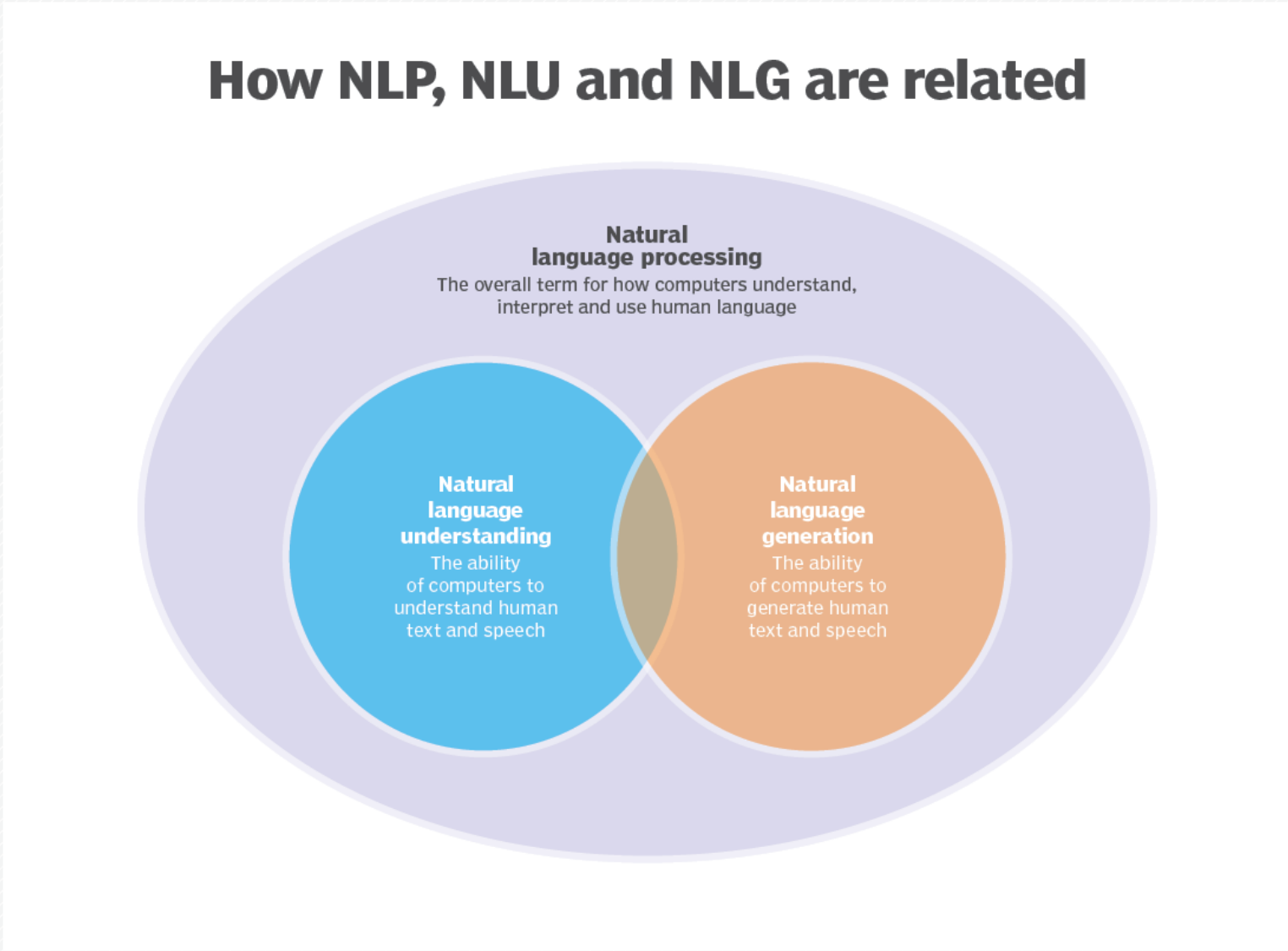


Image source: techtarget.com

(NATURAL) LANGUAGE GENERATION

NLG relies on machine learning algorithms and other approaches to create machine-generated text in response to user inputs. Some of the methodologies used include the following:

- Recurrent neural network
- Long short-term memory
- Transformer
  - Generative pre-trained transformer (GPT)
  - Bidirectional encoder representations from transformers (BERT)
  - XLNet

Image source: techtarget.com

(NATURAL) LANGUAGE UNDERSTANDING OR NATURAL-LANGUAGE INTERPRETATION (NLI)

- Natural-language understanding (NLU) or natural-language interpretation (NLI) is a subtopic of natural-language processing in artificial intelligence that deals with machine reading comprehension. Natural-language understanding is considered an AI-hard problem.
  - what human language means, rather than simply what individual words say.
  - comprehend and respond accurately to the sentiments expressed in natural language text.
  - Natural Language Understanding deconstructs human speech using trained algorithms until it forms a structured ontology, or a set of concepts and categories that have established relationships with one another.
  - computational linguistics

(NATURAL) LANGUAGE UNDERSTANDING OR NATURAL-LANGUAGE INTERPRETATION (NLI)

Extract information from the data source

- “ICE ticket Berlin to München” 15.11, 9am”
  - ICE: type of train in Germany
  - Ticket: intent to buy a train ticket
  - Berlin, München: locations
  - 15.11, 9am: date and time

NLU deconstructs human speech/text using trained algorithms until it forms a structured ontology, or a set of concepts and categories that have established relationships with one another.

- Intent recognition, entity recognition

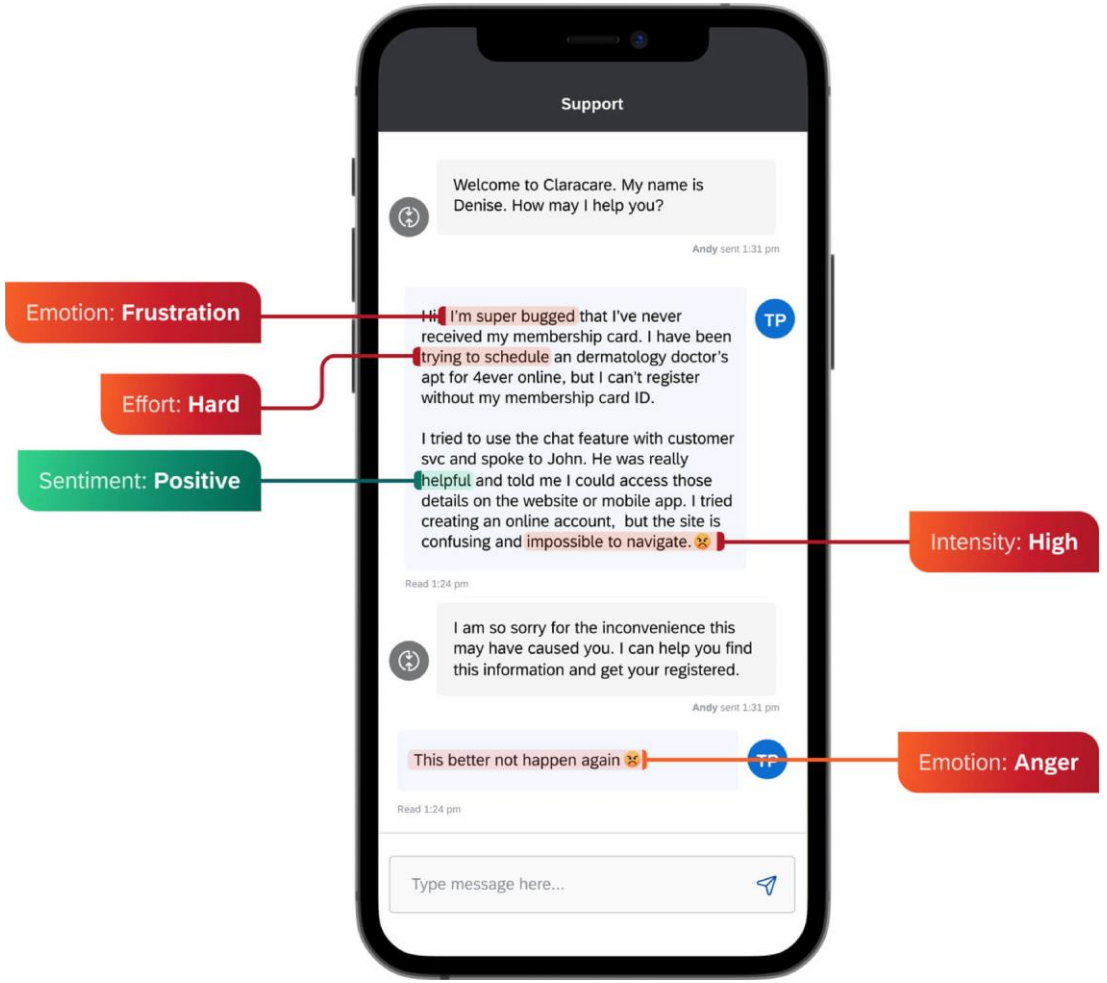
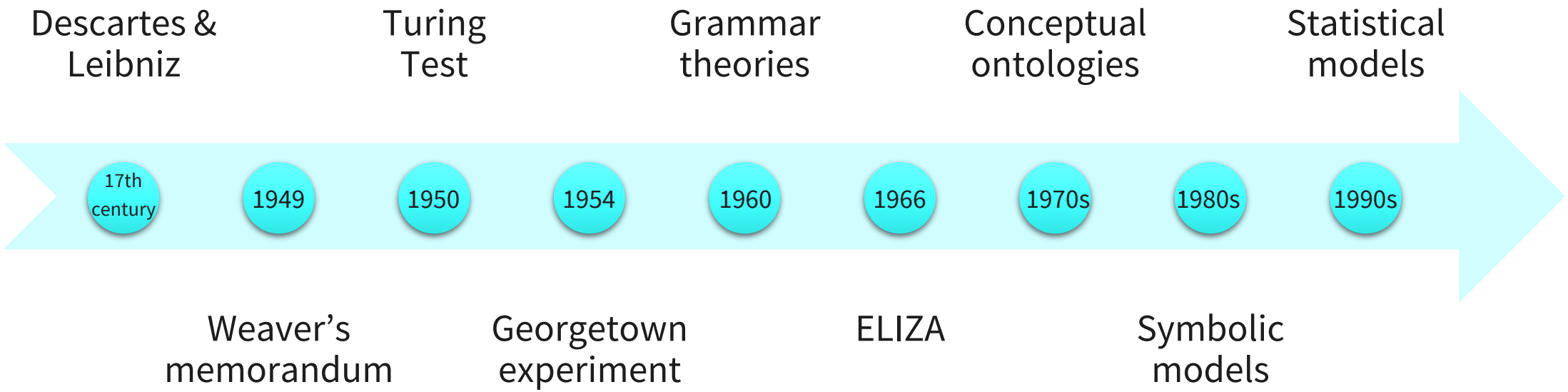


Image source: Qualtrics.com

HISTORICAL BACKGROUND



https://aylien.com/blog/a-review-of-the-recent-history-of-natural-language-processing

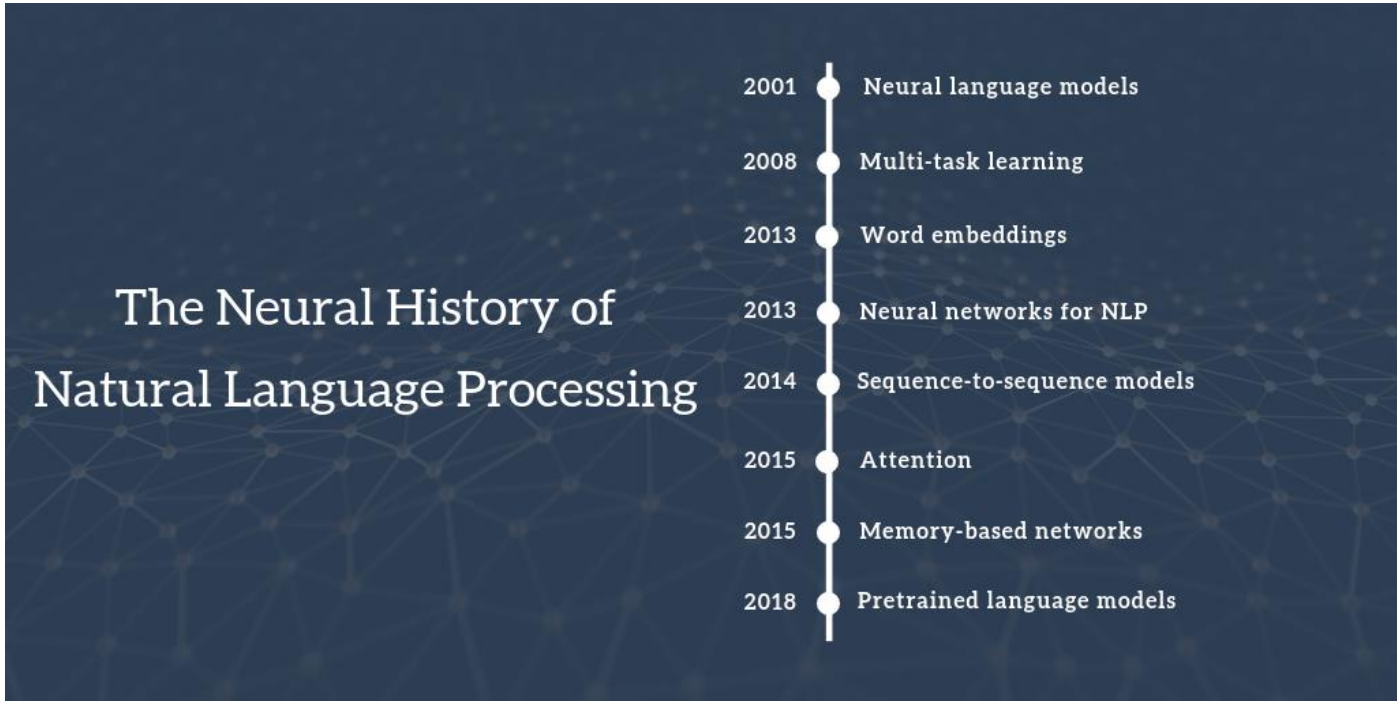
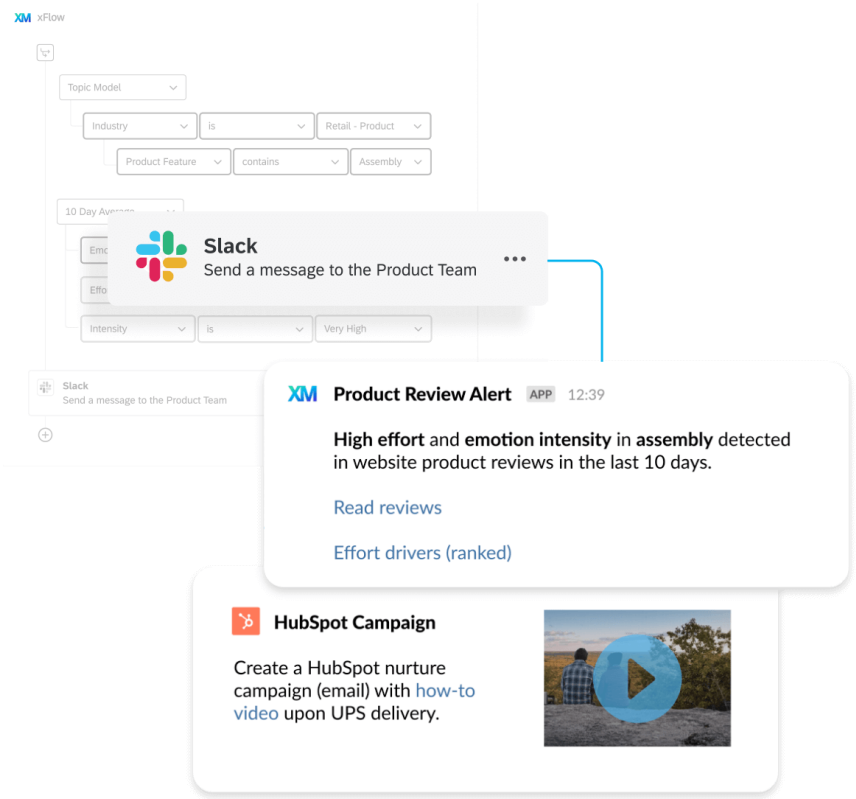


Image Source: Custom Depiction

TYPICAL AREAS OF APPLICATION FOR NLP



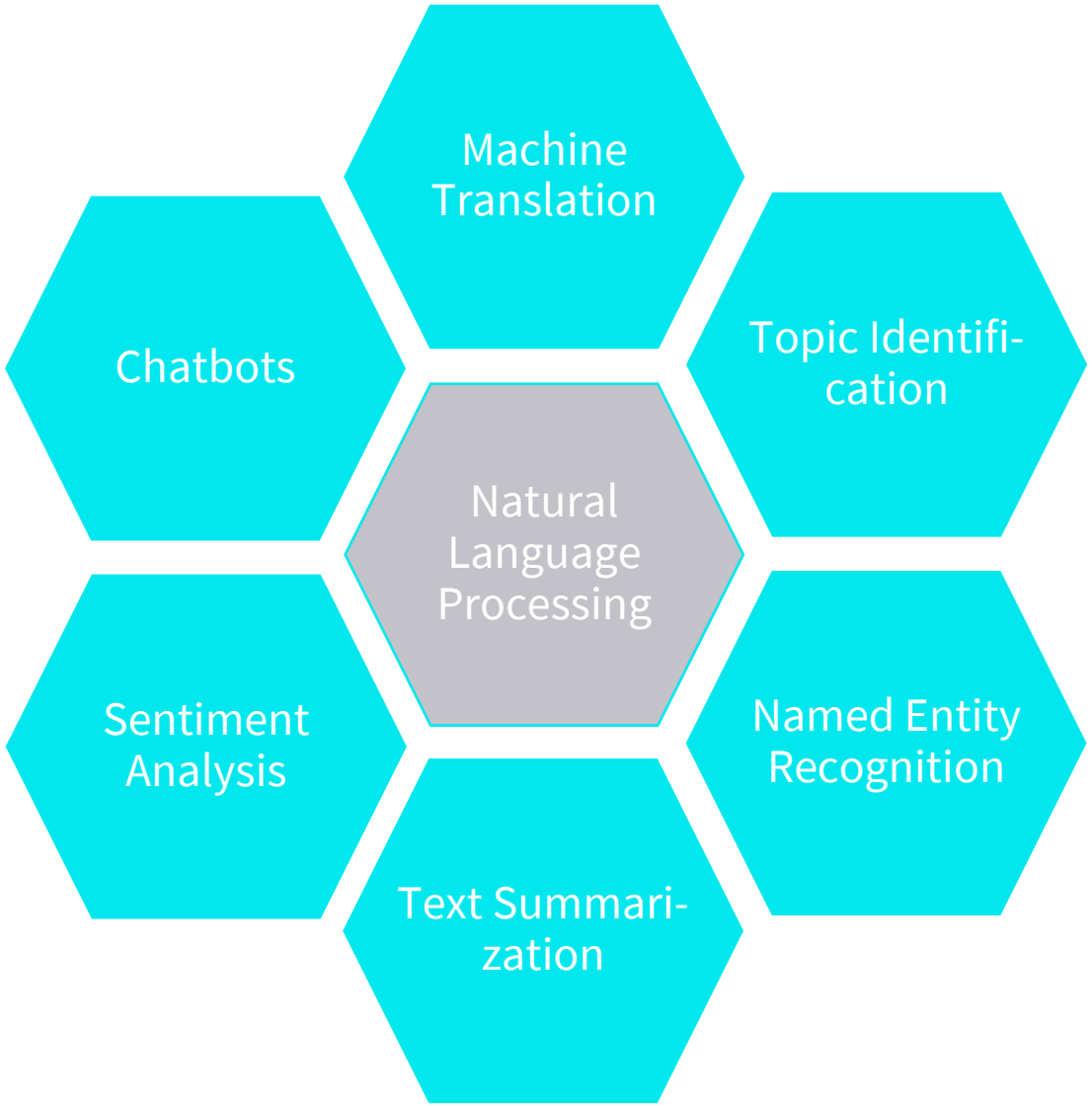
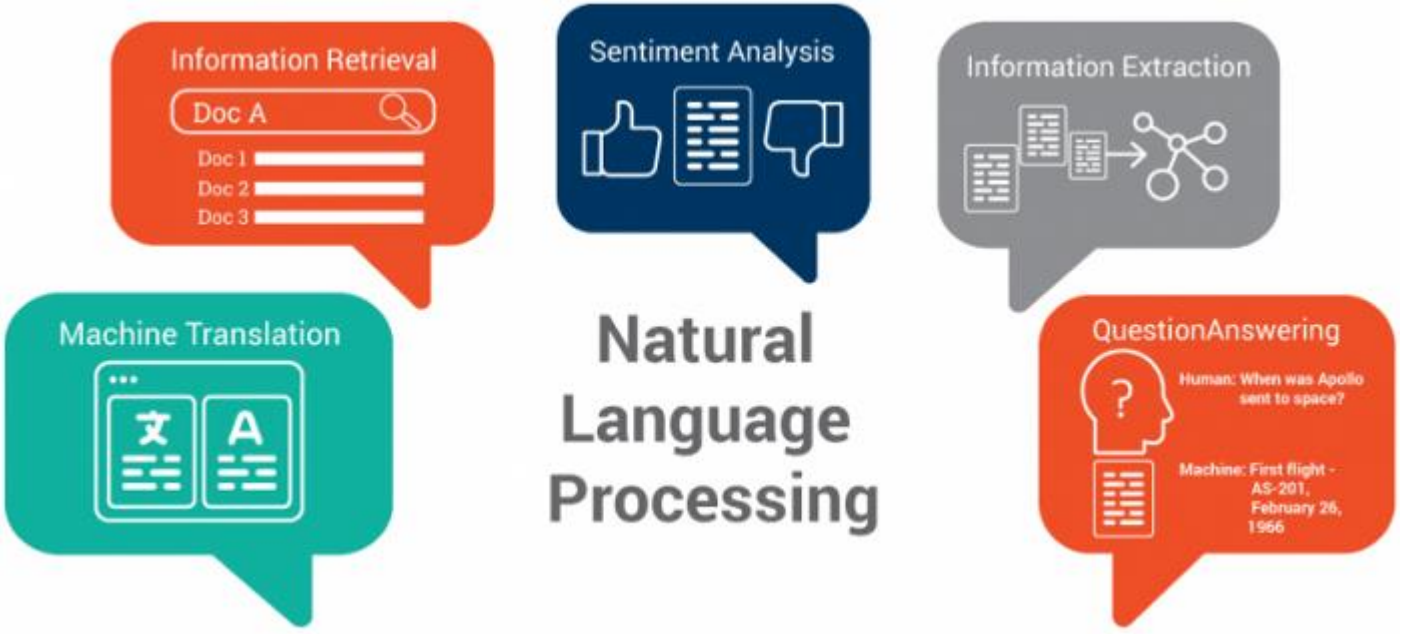


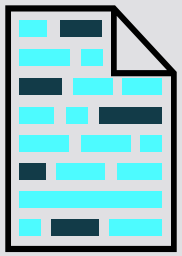
Image Source: Custom Depiction

STATISTICS

Worldwide revenue from the natural language processing (NLP) market is forecast to increase rapidly in the next few years. The NLP market is predicted be almost 14 times larger in 2025 than it was in 2017, increasing from around three billion U.S. dollars in 2017 to **over 43 billion in 2025**. Natural language processing (NLP) is a branch of artificial intelligence (AI) that helps computers understand, interpret and manipulate human language. (https://www.statista.com/statistics/607891/worldwide-natural-language-processing-market-revenues/#:~:text=The%20NLP%20market%20is%20predicted,interpret%20and%20manipulate%20human%20language.)



TOPIC IDENTIFICATION



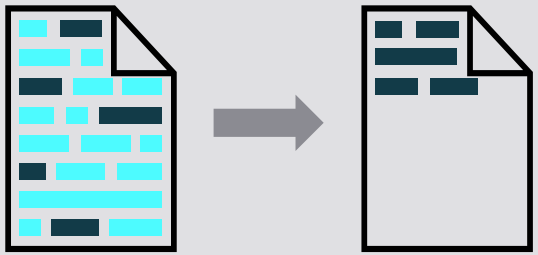
→ <topic>

"One quick way to summarize the milestones in AI history is to list the Turing Award winners: Marvin Minsky (1969) and John McCarthy (1971) for defining the foundations of the field based on representation and reasoning; Allen Newell and Herbert Simon (1975) for symbolic models of problem solving and human cognition" (Russell & Norvig, 2021, p. 35).

→ history

Text Source: Russell, S., & Norvig, P. (2021), p. 35.
Image Source: Custom Depiction.

TEXT SUMMARIZATION

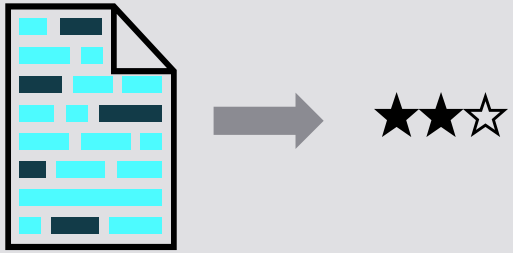


"Mr and Mrs Dursley, of number four, Privet Drive, were proud to say that they were perfectly normal, thank you very much. They were the last people you'd expect to be involved in anything strange or mysterious, because they just didn't hold with such nonsense" (Rowling, 1998, p. 1).

Mr and Mrs Dursley were perfectly normal. They were the last people to be involved in anything mysterious.

Text Source: Rowling (1998), p. 1.
Image Source: Custom Depiction.

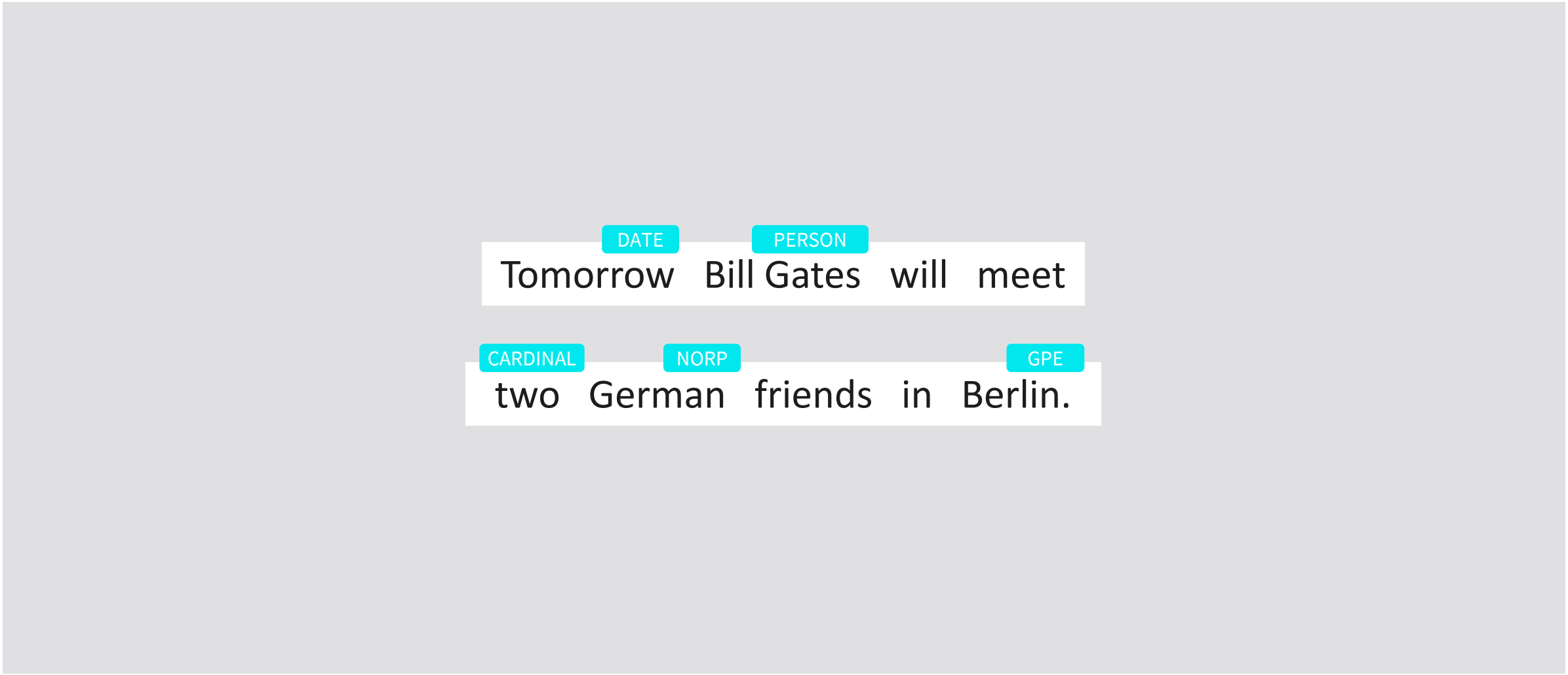
SENTIMENT ANALYSIS



“Exactly what I thought it would be. Perfect early Father's Day gift for my husband – we had so much fun playing ping pong at a resort recently I just had to get it for him.”

Text Source: Jay, L. (2016).
Image Source: Custom Depiction.

NAMED ENTITY RECOGNITION



CHATBOTS

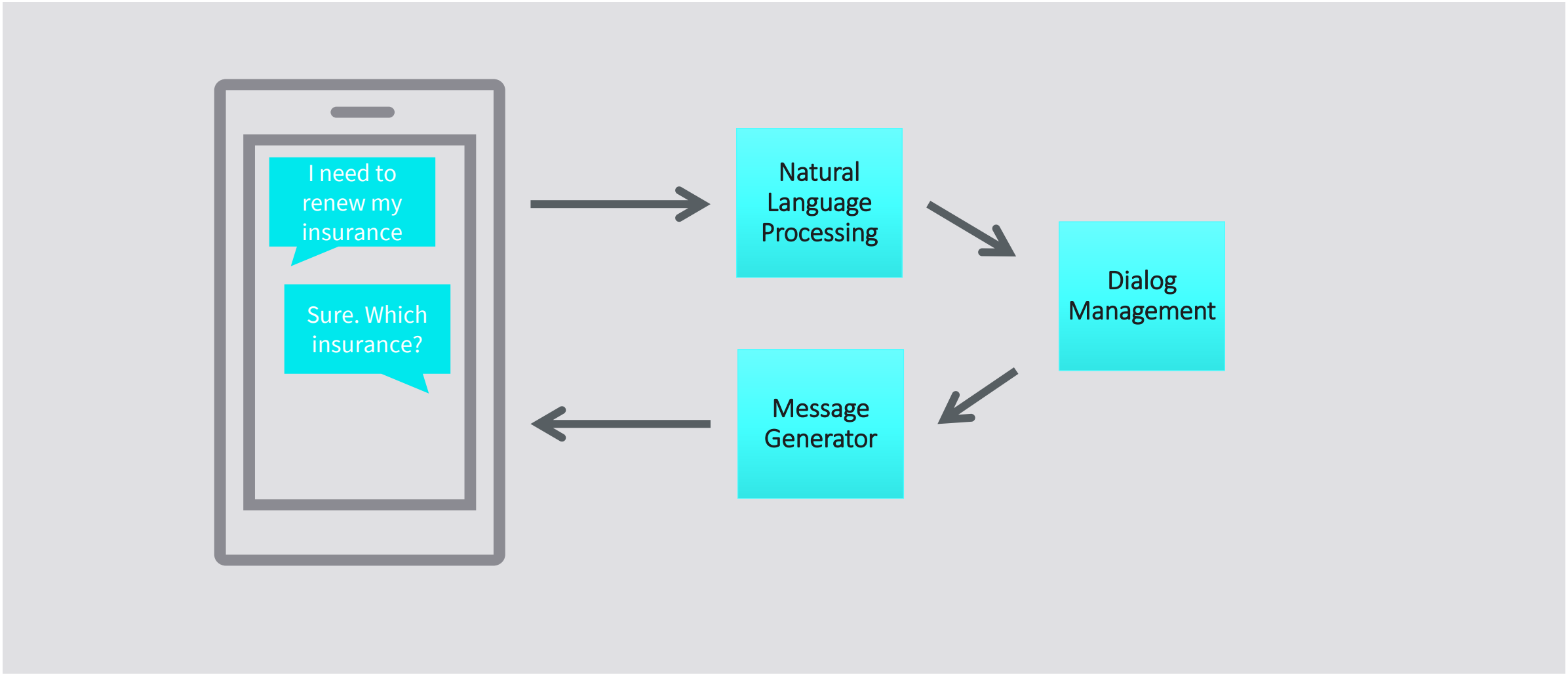


Image Source: Custom Depiction.

MACHINE TRANSLATION

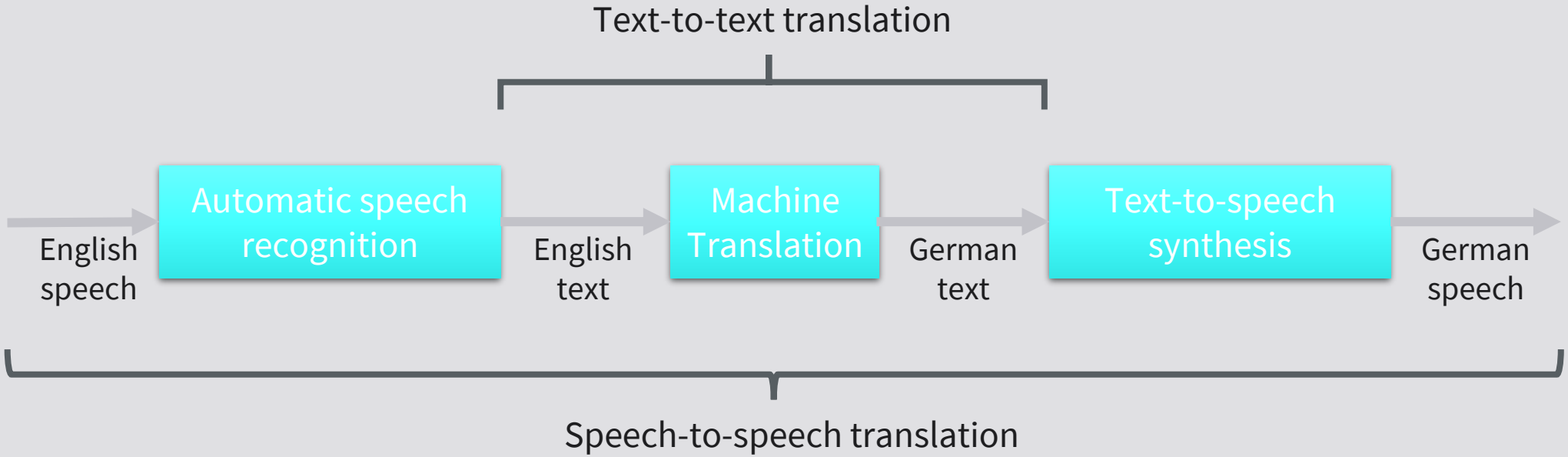
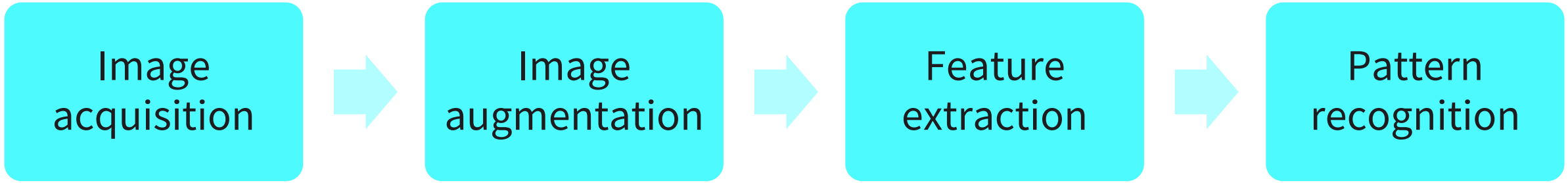


Image Source: Custom Depiction.

Computer Vision = understand the content of images and videos



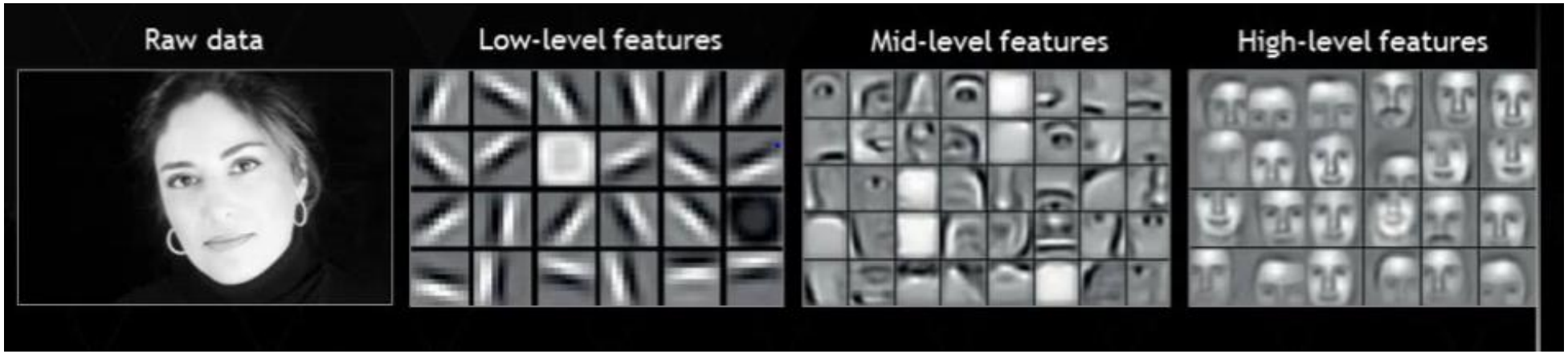
- Camera vision
- Pinhole cameras
- Film cameras
- Digital cameras

- Image sharpening
- Contrast enhancement

- Extract higher-level image features

- Infer meaningful image content

Extract salient features from images such as edges, colors, blobs and ridges:



Source of the image: Tomodan & Caleanu, 2018.

COMPUTER VISION DEMO

Source of the video: YouTube, 2018, https://www.youtube.com/watch?v=G2VaJvNNp4k.

HISTORICAL DEVELOPMENTS

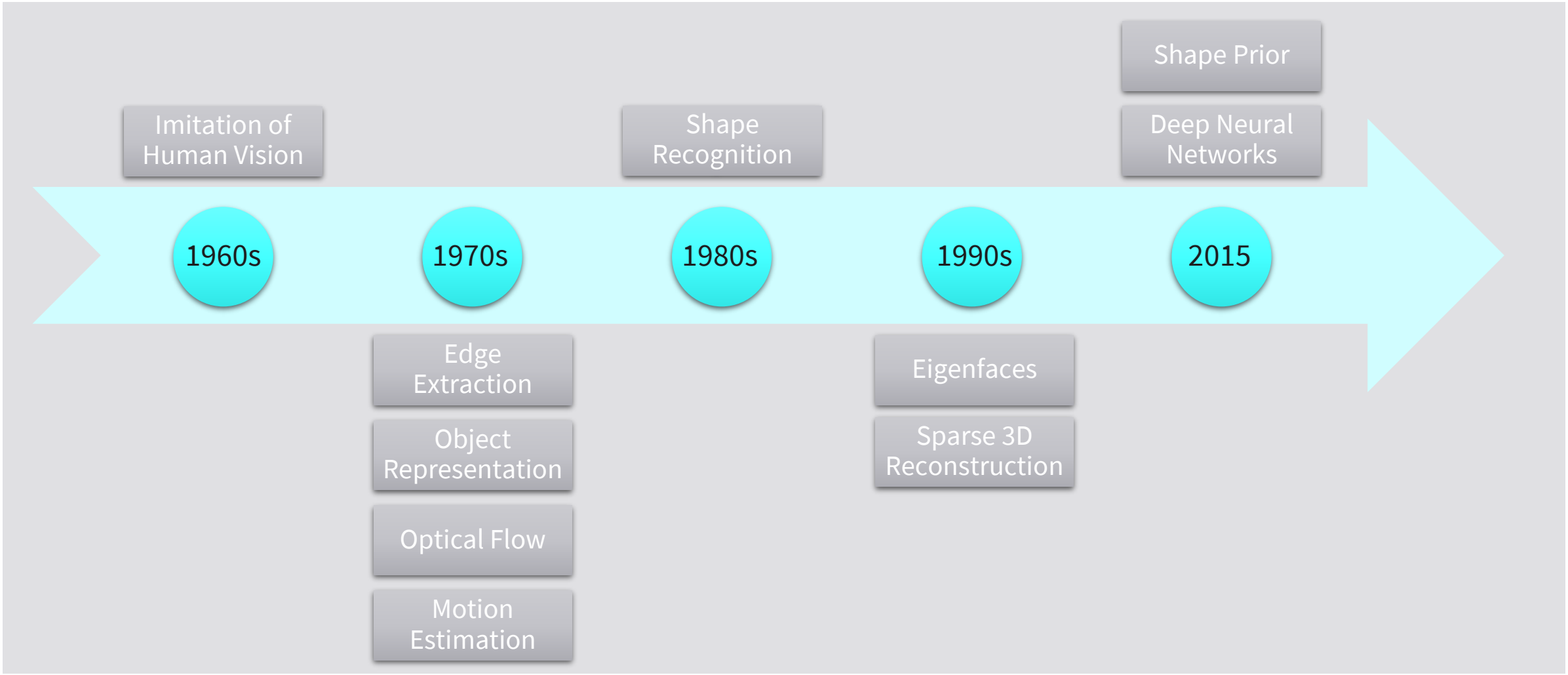


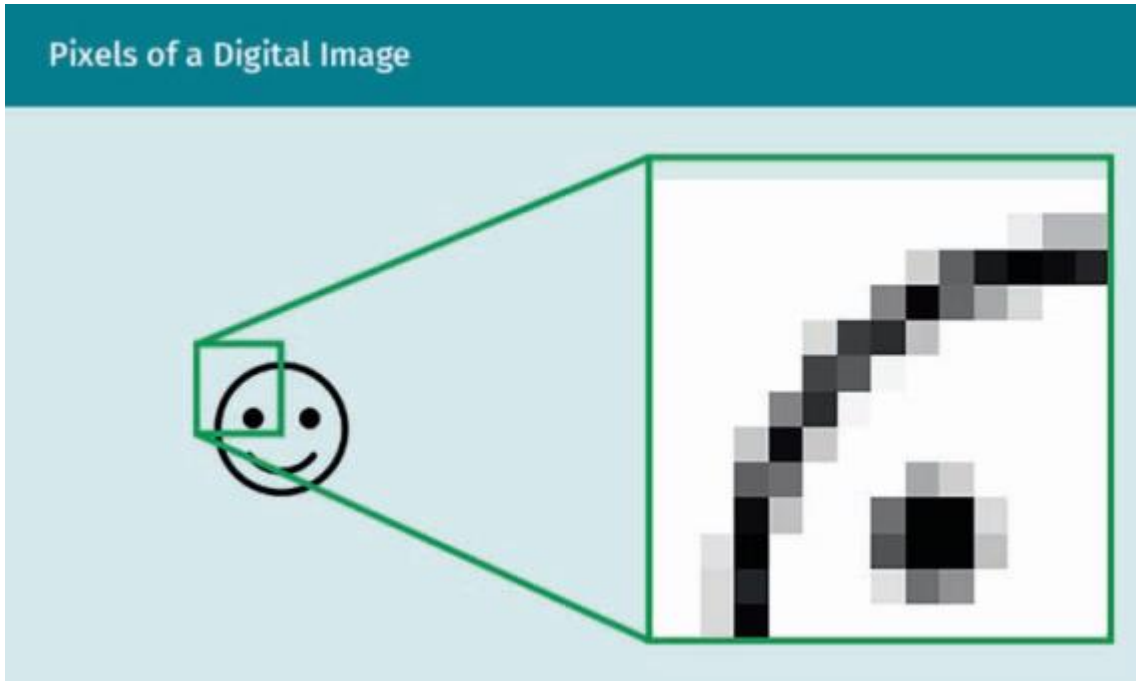
Image Source: Custom Depiction

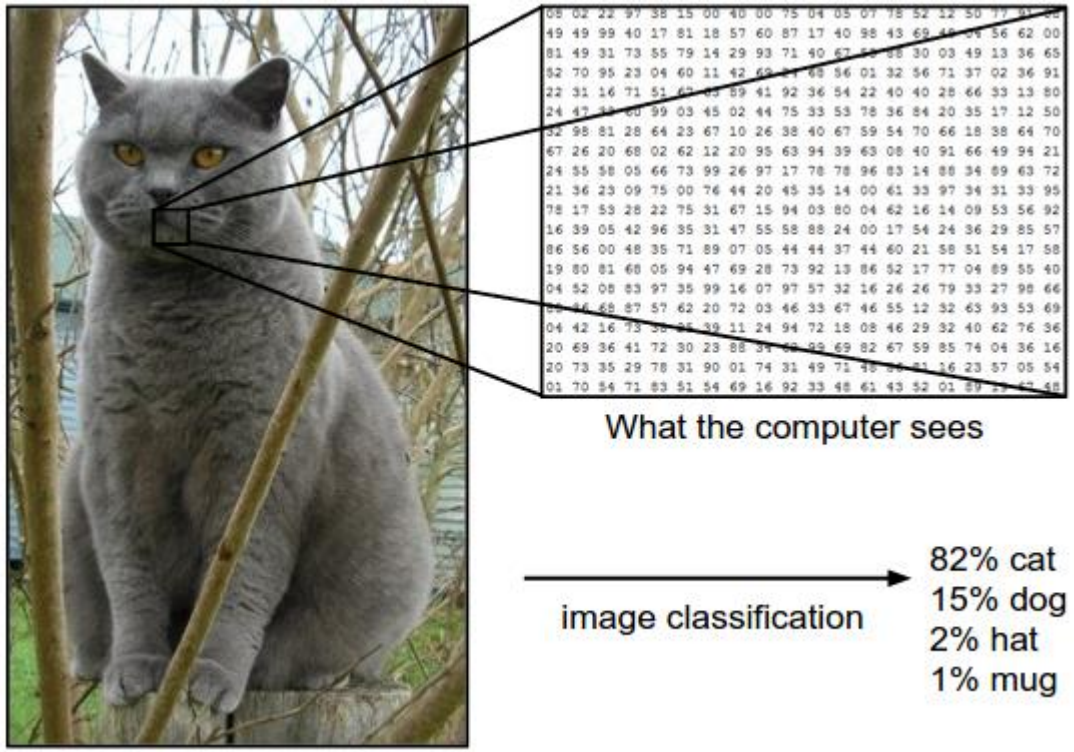
Term: pixel

color representations

monochrome

real, high, true, deep color





TYPICAL COMPUTER VISION TASKS

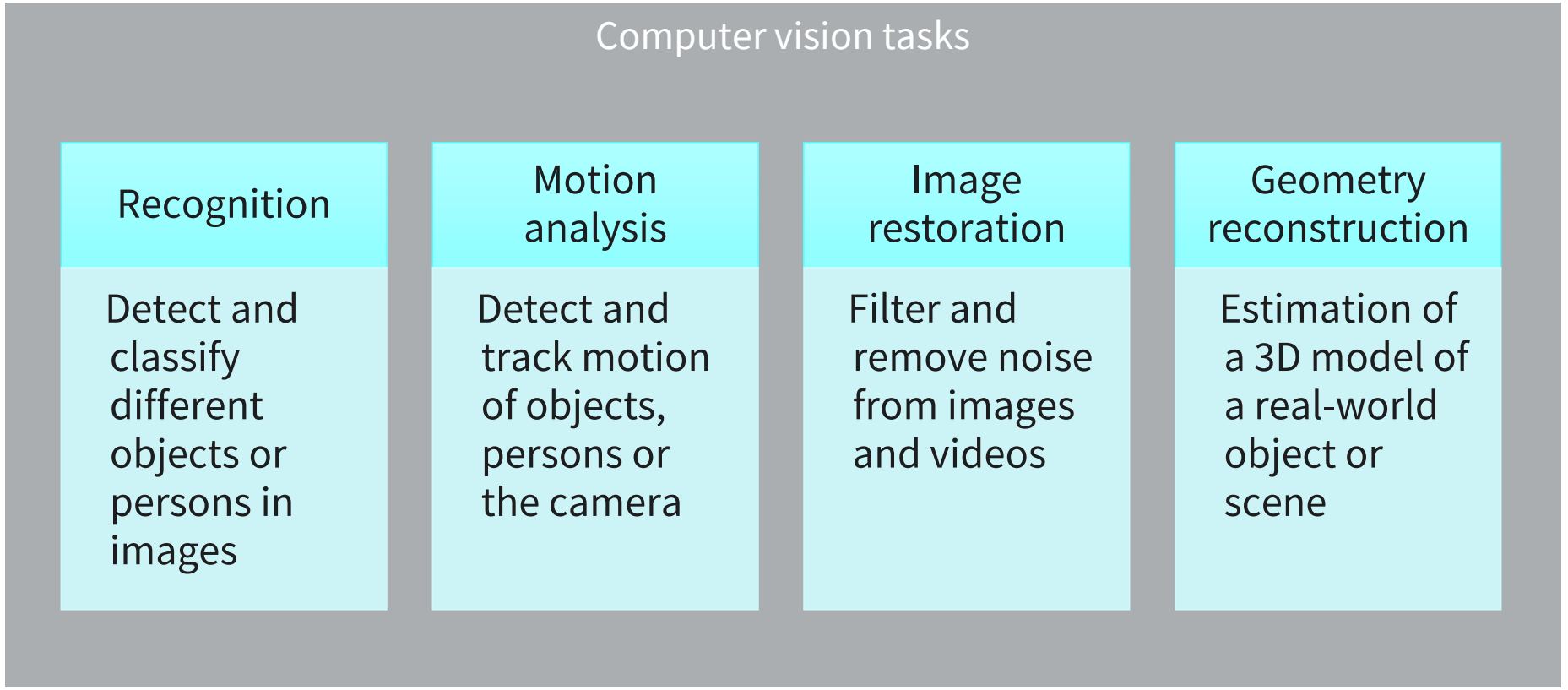


Image Source:

10 INTERESTING COMPUTER VISION TASKS: OBJECT DETECTION

- identify and locate an object of particular interest from still image or video data.

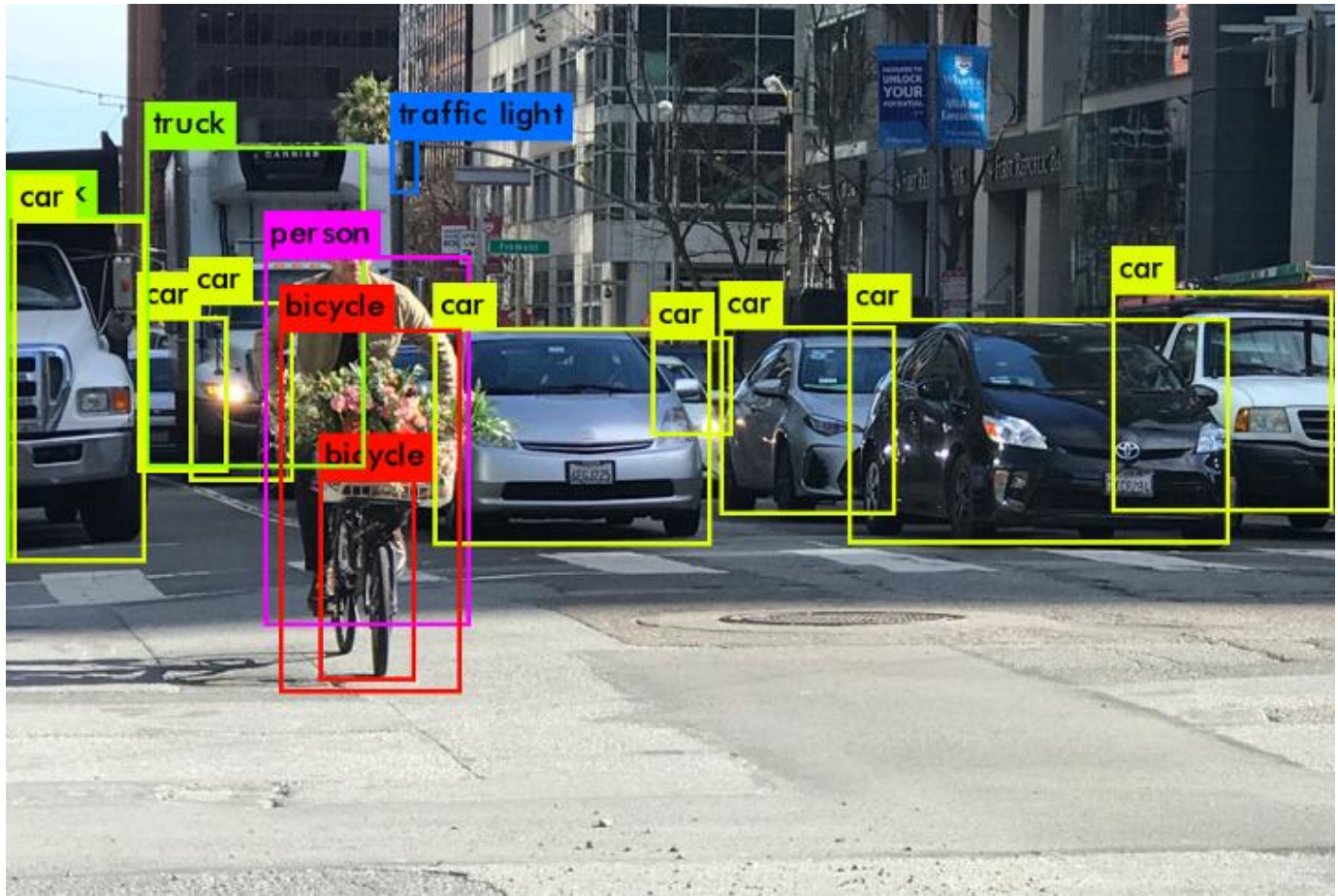


Image source: towardsdev.com

10 INTERESTING COMPUTER VISION TASKS: IMAGE SEGMENTATION

- The process of partitioning an image into multiple segments. the aim is to transform the image in a representation that is easier to analyze, there are many flavors of image segmentation

Image source: towardsdev.com

10 INTERESTING COMPUTER VISION TASKS: POSE ESTIMATION

Estimating the position and the orientation of the joints of a human body. this has a wide range of applications in AR/VR (augmented and virtual reality)

Image source: towardsdev.com

10 INTERESTING COMPUTER VISION TASKS: IMAGE GENERATION

Generating new images from an existing data set, this task has been made popular by the use Generative adversarial networks (GANs) to generate seemingly real images or videos which were actually fake.

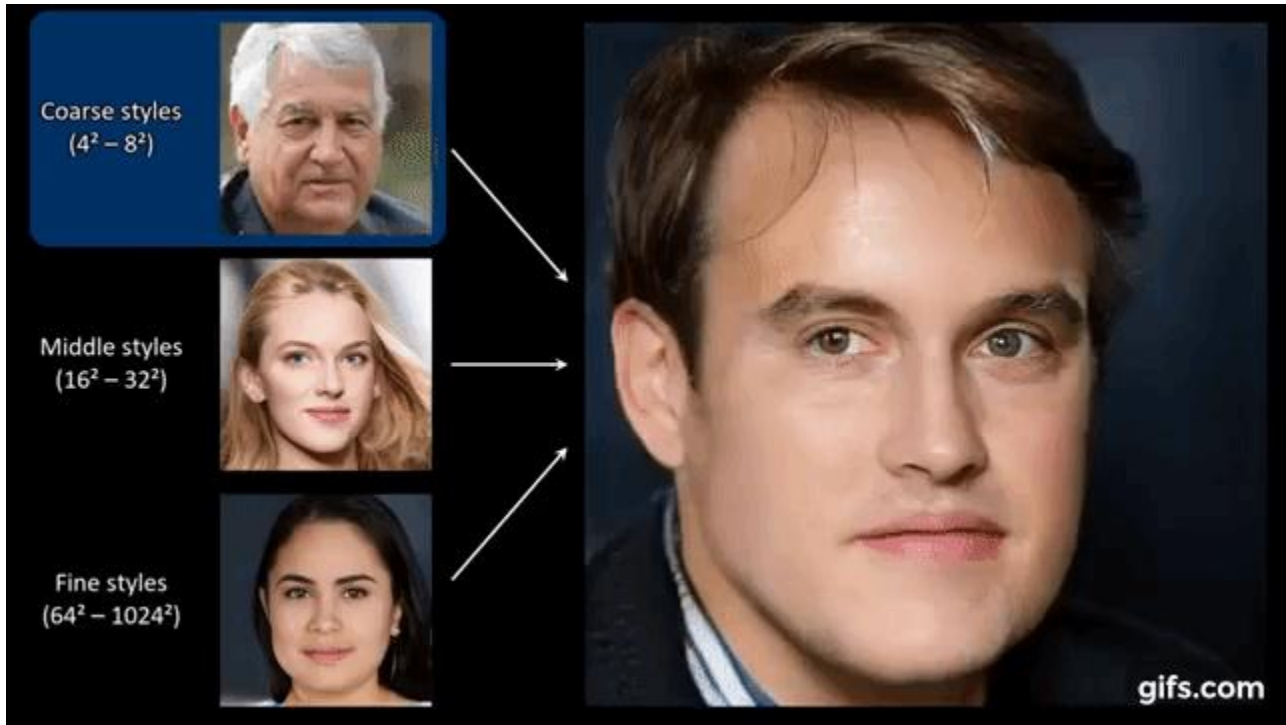


Image source: towardsdev.com

Identifying classes of objects within the image

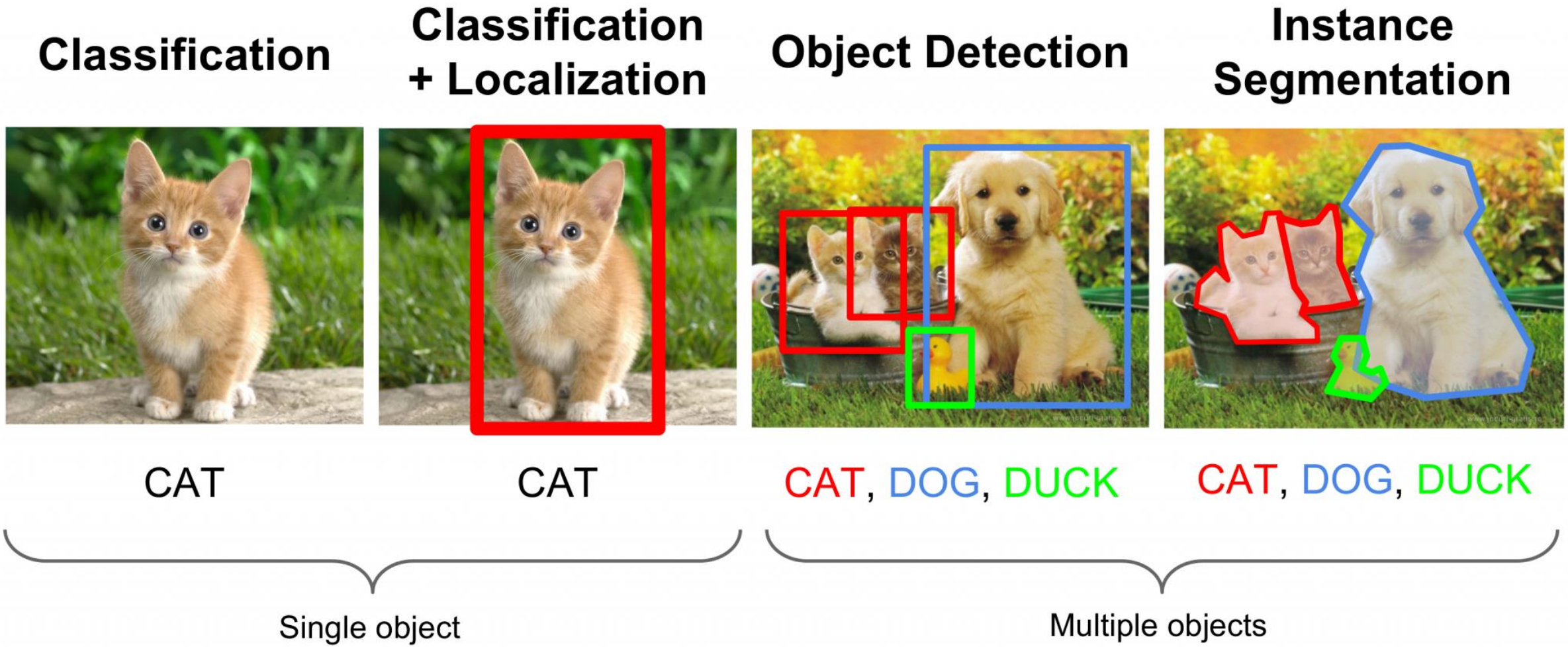


Image source: analyticsindiamag.com

maintaining a predictive model performance even if the data distribution the model was trained on has changed. domain adaptation uses labeled data in one or more data distribution and tries to solve a new task with data belonging to a different distribution

Image source: towardsdev.com

10 INTERESTING COMPUTER VISION TASKS: DE-NOISING

predicting the original image by removing noise from a noise-contaminated image

Image source: towardsdev.com

10 INTERESTING COMPUTER VISION TASKS: SUPER-RESOLUTION

taking an input from a low resolution image and up-scaling it to a higher resolution

Image source: towardsdatascience.com

10 INTERESTING COMPUTER VISION TASKS: 3D FACE ANIMATION

reconstructing an animated 3D human face from a still image

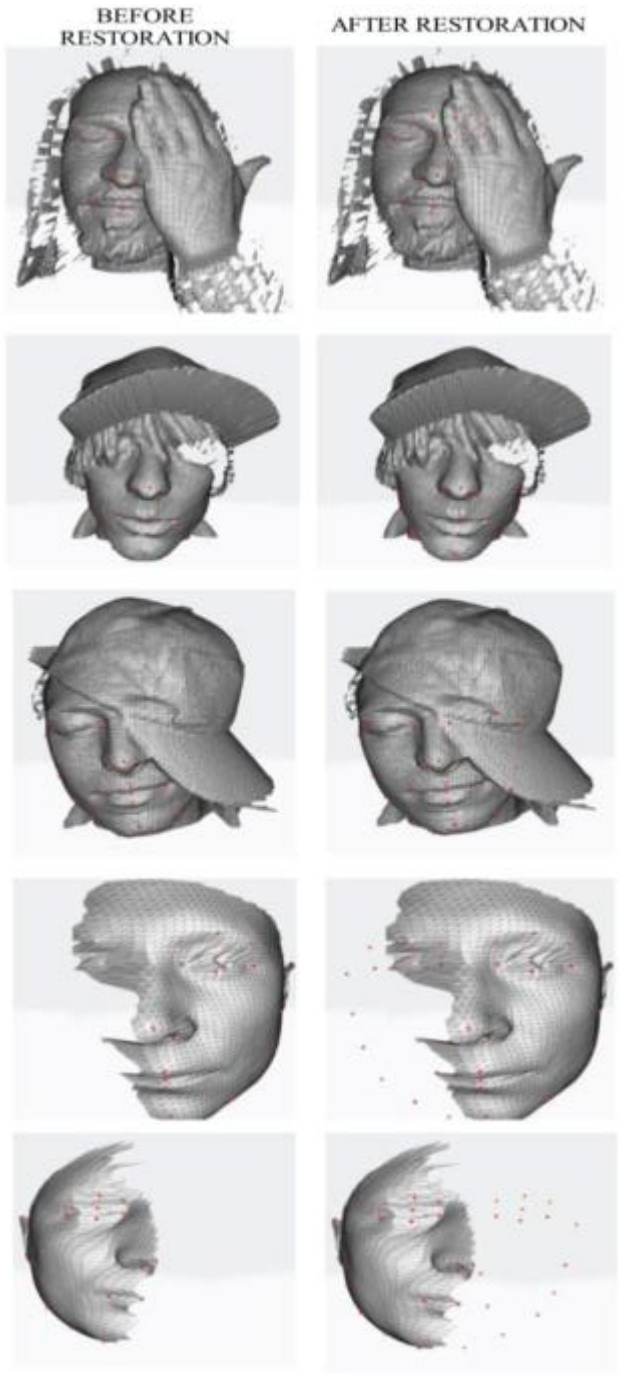


Image source: towardsdev.com, https://link.springer.com/article/10.1007/s11831-021-09705-4

predicting the movement of an object from observations made in a video stream

Image source: towardsdev.com

IMAGES AS FUNCTIONS

A function that can map a two-dimensional coordinate (x,y) to a specific color value.

On the x-axis we begin on the left with a value of 0 and continue to the right until the maximum width of an image is reached.

On the y-axis, we begin with 0 at the top and reach the height of an image at the bottom.

f(42,100)

Filter:

Function which receives an image as an input, applies modifications, and returns the filtered image as an output.

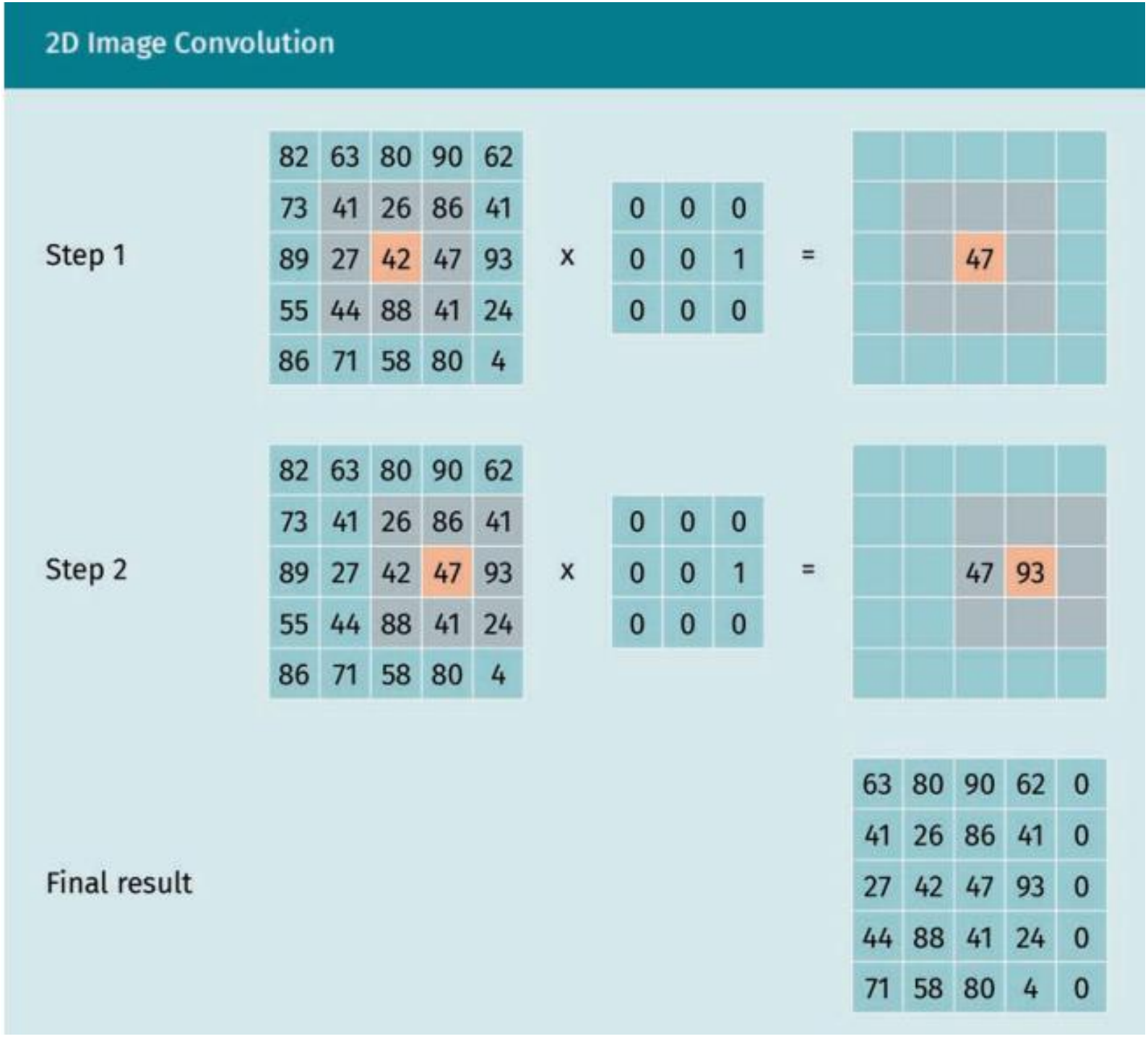
Example: 2D image convolution

Convolution of an image I with a kernel k with a size of n and a center coordinate a:

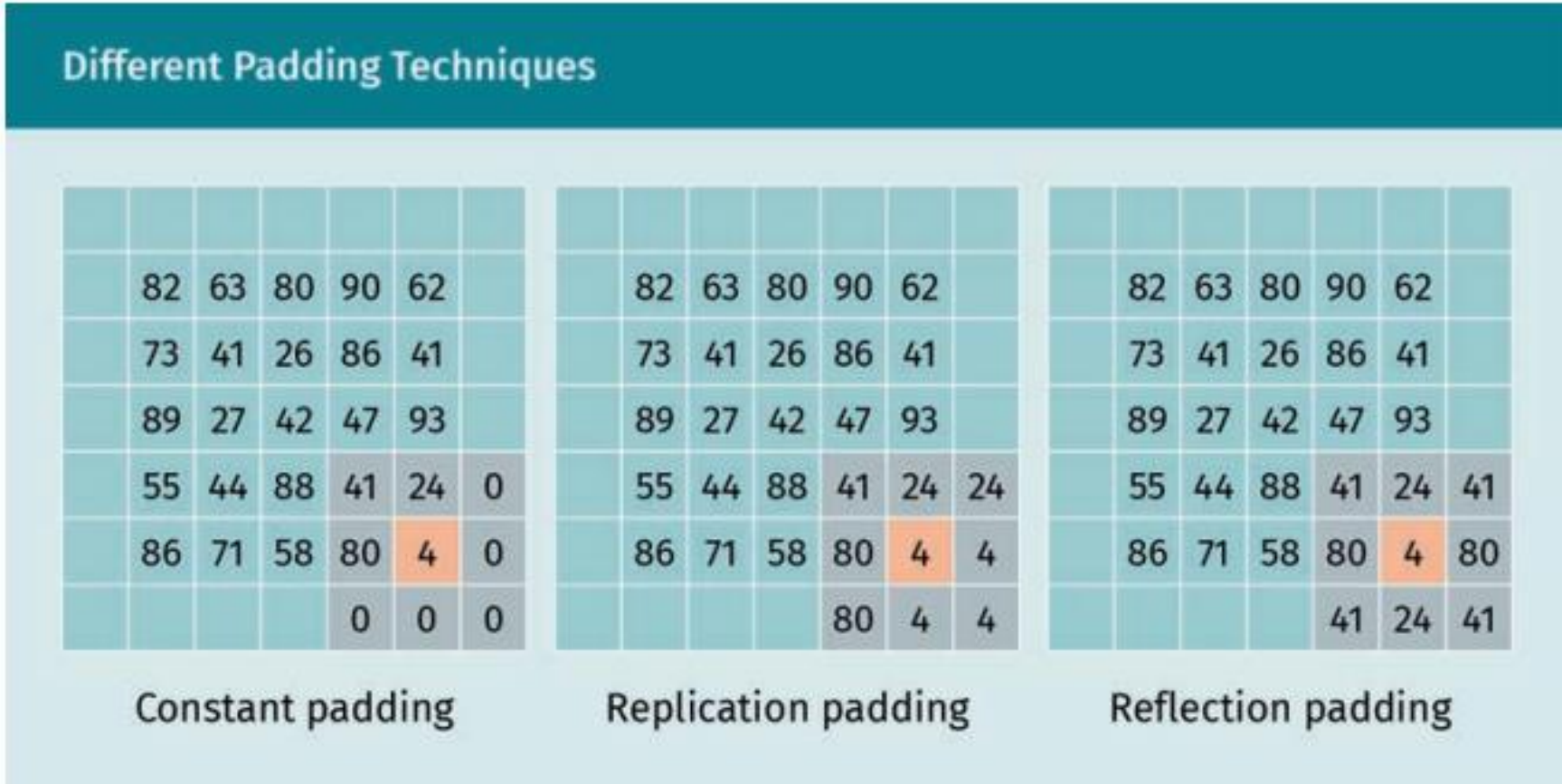
$$\mathrm{I} \cdot (x, y) = \sum_{i=1}^{n} \sum_{i=j}^{n} \mathrm{I}(x - 1 + a, y - j + a) k(i, j)$$

$\mathrm{I} \cdot (x, y)$: value of the resulting image $\mathrm{I} \cdot$ at position (x, y), I is the original image.

FILTERS


2D Image Convolution
Step 1
82 63 80 90 62
73 41 26 86 41
89 27 42 47 93
55 44 88 41 24
86 71 58 80 4
x
0 0 0
0 0 1
0 0 0
=
47
Step 2
82 63 80 90 62
73 41 26 86 41
89 27 42 47 93
55 44 88 41 24
86 71 58 80 4
x
0 0 0
0 0 1
0 0 0
=
47 93
Final result
63 80 90 62 0
41 26 86 41 0
27 42 47 93 0
44 88 41 24 0
71 58 80 4 0

PADDING TECHNIQUES



The calibration makes it possible to extract distortion from the images

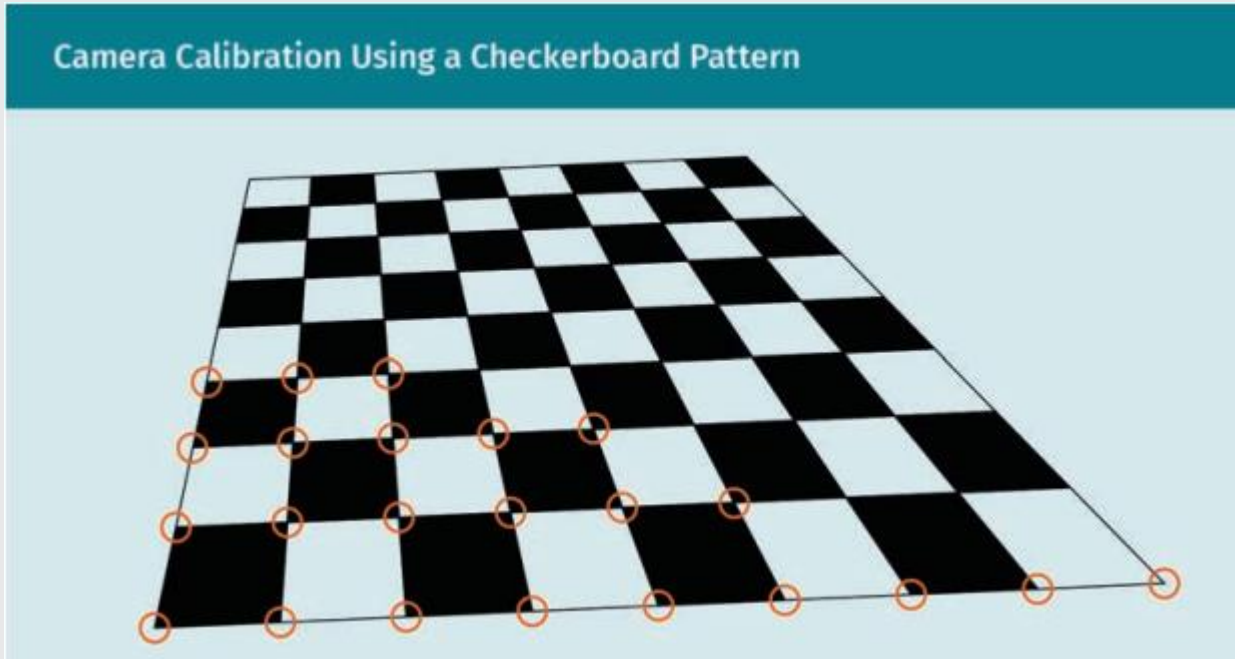


Image Source: Custom Depiction

Feature Detection

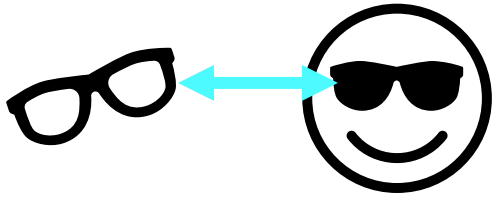
- Blobs (blue)
- Edges (red)
- Corners (yellow)

Feature Description

Semantic information

- Common algorithms:
  - BRIEF
  - ORB
  - SIFT
  - SURF

Feature Matching



- Identification of similar features in different images

https://medium.com/data-breach/introduction-to-surf-speeded-up-robust-features-c7396d6e7c4e
https://docs.opencv.org/3.4/df/dd2/tutorial_py_surf_intro.html

Image Source: Custom Depiction

IMPORTANT CHARACTERISTICS OF FEATURE DETECTION ALGORITHMS

SEMANTIC SEGMENTATION

Original image

Segmentation map

Segmentation overlay

Background Chair Coffee table

You now have…

- … understanding of basic components within Natural Language Processing (NLP) components.
- … knowledge of main steps within Computer Vision (CV) applications.
- … a basic idea of common features extracted from images.

SESSION 5

TRANSFER TASK

TRANSFER TASK

Brainstorm about factors of images that make Computer Vision difficult. While doing so, reflect on Human Vision throughout daily life and what information the human brain needs to translate.

TRANSFER TASK
PRESENTATION OF THE RESULTS

Please present your results.

The results will be discussed in plenary.

1. Speech-to-Text
   a) has no relation to natural language processing.
   b) is one of many sub-fields of natural language processing.
   c) is what natural language processing is all about.
   d) has no practical application.

2. Artificial General Intelligence
   a) is a natural extension of weak artificial intelligence.
   b) is provably impossible.
   c) will require a considerably more variegated set of abilities than what weak artificial intelligence approaches can currently provide.
   d) is guaranteed to come about in the near future.

3. Computer vision refers to
   a) image acquisition by computers.
   b) computer graphics.
   c) the semantic understanding of visual scenes.
   d) filter operations such as smoothing or de-noising.

LIST OF SOURCES

Tomodan, E.R. & Caleanu, C. (2018). *Bag of Features vs Deep Neural Networks for Face Recognition*. 2018 International Symposium on Electronics and Telecommunications (ISETC). DOI: 10.1109/ISETC.2018.8583846.