Knowledge Representation

NLP

"Knowledge"

Unstructured
Ambiguous
Lots and lots of it!

Structured
Precise, Actionable
Specific to the task

Humans can read them, but
… very slowly
… can't remember all
… can't answer questions

Computers can use
… quickly answer questions
… memory is not a problem
… don't get tired

ALGONQUIN COLLEGE

What Is Natural Language Processing?
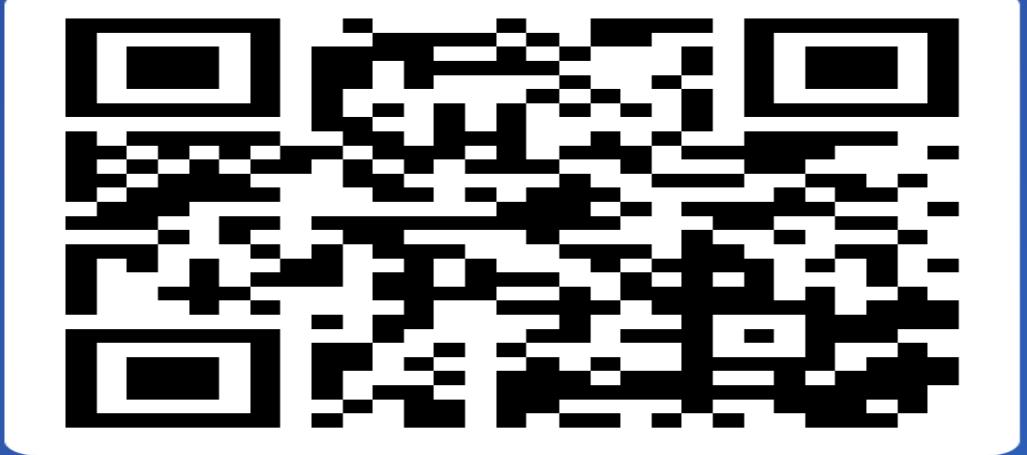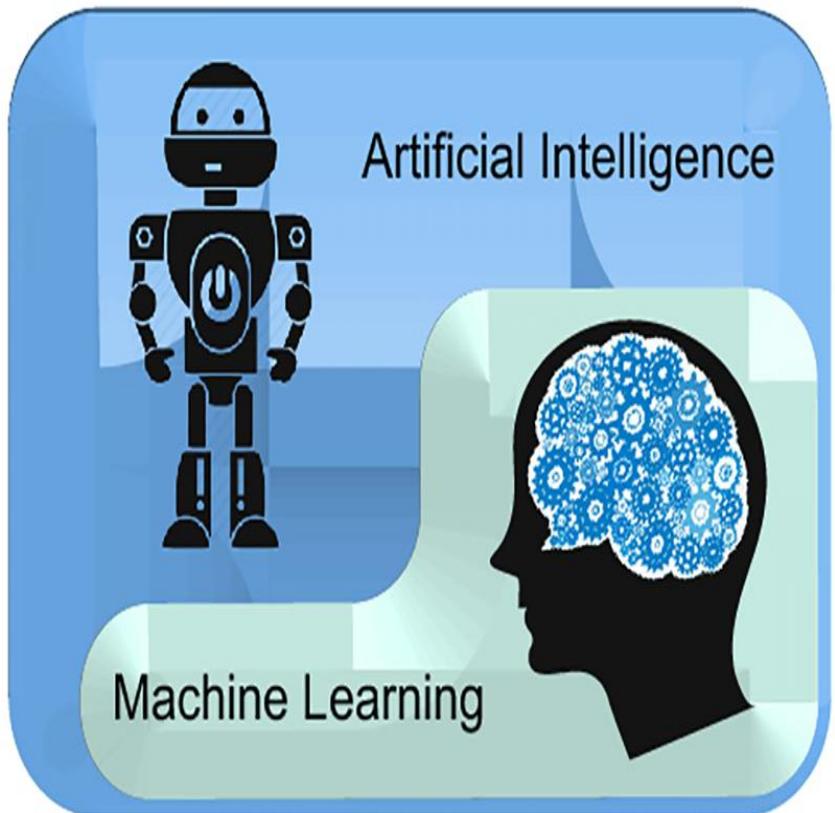
- ❑ Subfield of linguistics, computer science, and artificial intelligence concerned with the **interactions between computers and human language.**

- ❑ How to program computers to process and analyze large amounts of **natural language data.**

Motivations

- Enables computers to understand and process human language

- Helping businesses to analyze customer feedback and improve customer service

- Making it easier for individuals with **disabilities** to interact with computers and access information

- Extract insights and make sense from data

Discussion Activity

Join at
slido.com
#1046 936

23

Interconnections Between AI, ML, DL, and NLP
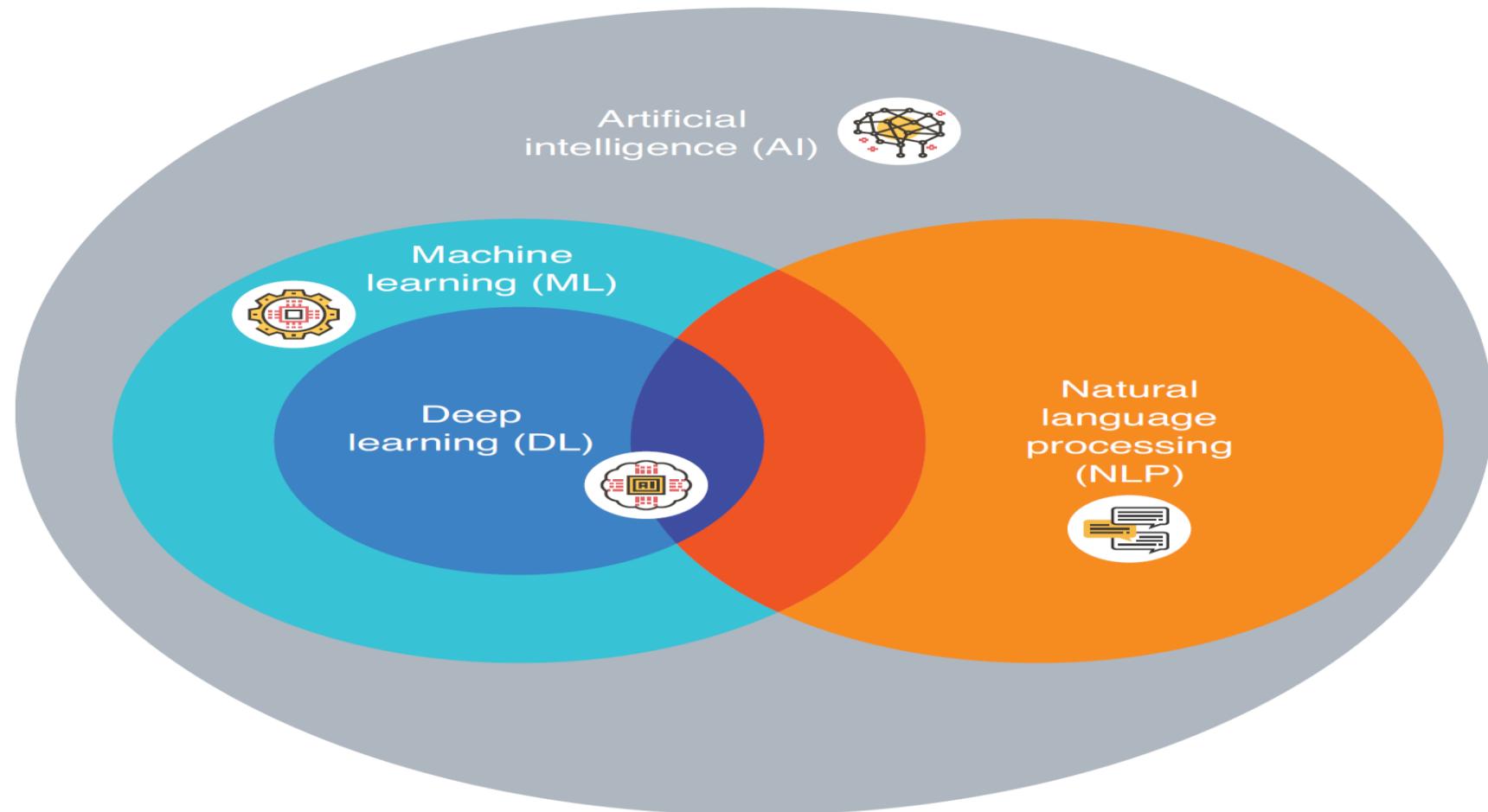
History of NLP

Genesis and Evolution of
Natural Language Processing

1950s
Basic language translation tasks were executed

1970s & 1980s
Systems were programmed with language rules and lexicons

2000s
Marked the advent of large-scale data-driven NLP

2017
Introduction of Transformer model architecture

| 1950s | 1960s | 1970s 1980s | 1990s | 2000s | 2010s | 2017 | 2023 |

1960s
Development of the first NLP applications which could perform simple language processing tasks

1990s
Shift towards statistical NLP

2010s
Advancements in deep learning and neural networks

2023
Mass adoption of Generative LLM (Large Language Models) based application

Image source: https://commtelnetworks.com/exploring-the-impact-of-natural-languageprocessing-on-cni-operations

ALGONQUIN COLLEGE

25

Turing's test for Artificial Intelligence



Ability to understand and generate language ~ intelligence
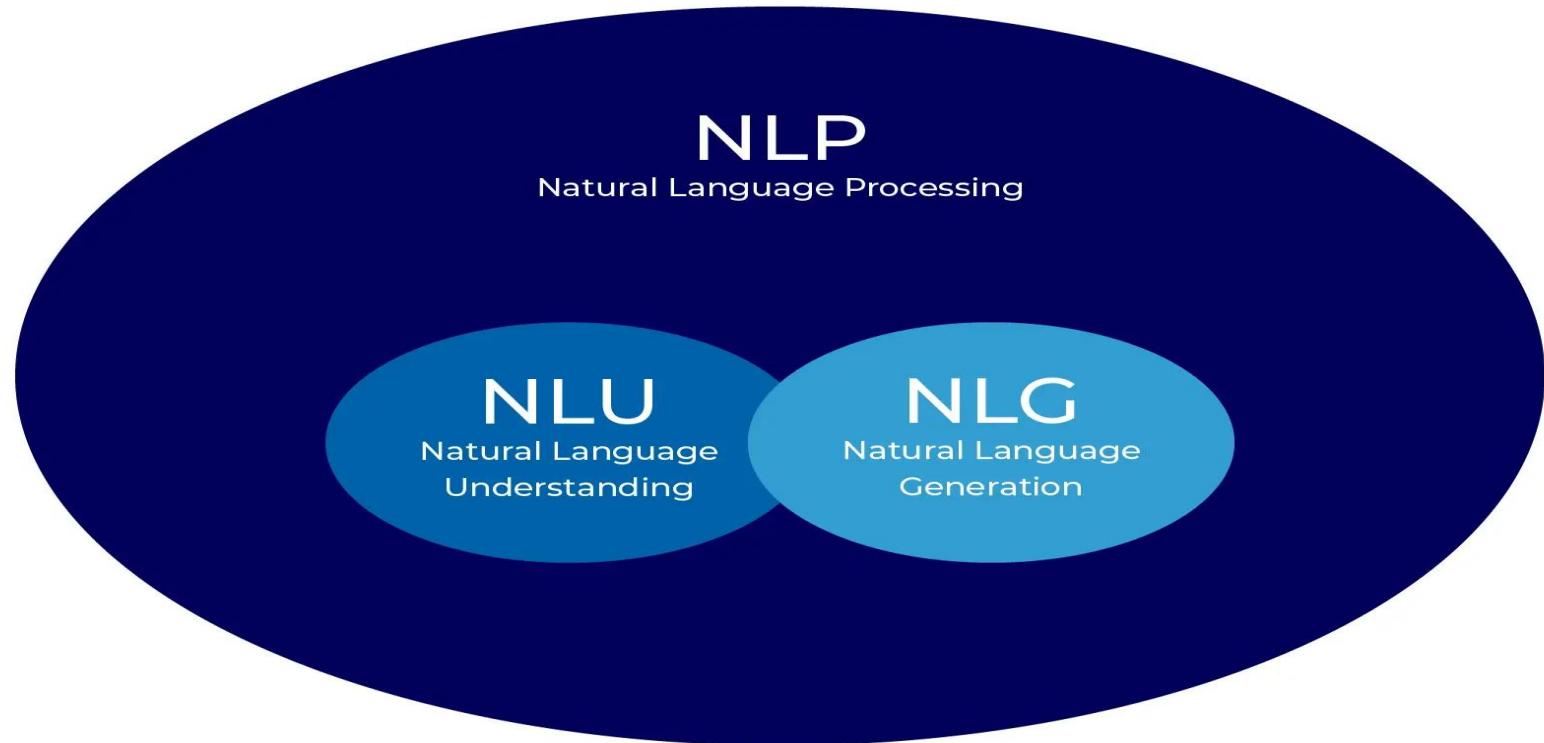
Lesson Agenda

❑ Course CSI

❑ What is Natural Language Processing (NLP)?

❑ Applications of Natural Language Processing

❑ Challenges in Natural Language Processing

❑ Approaches to NLP

❑ NLP Libraries in Python

NLP Applications

**NLP=NLU+NLG**

NLP Applications: Speech Recognition
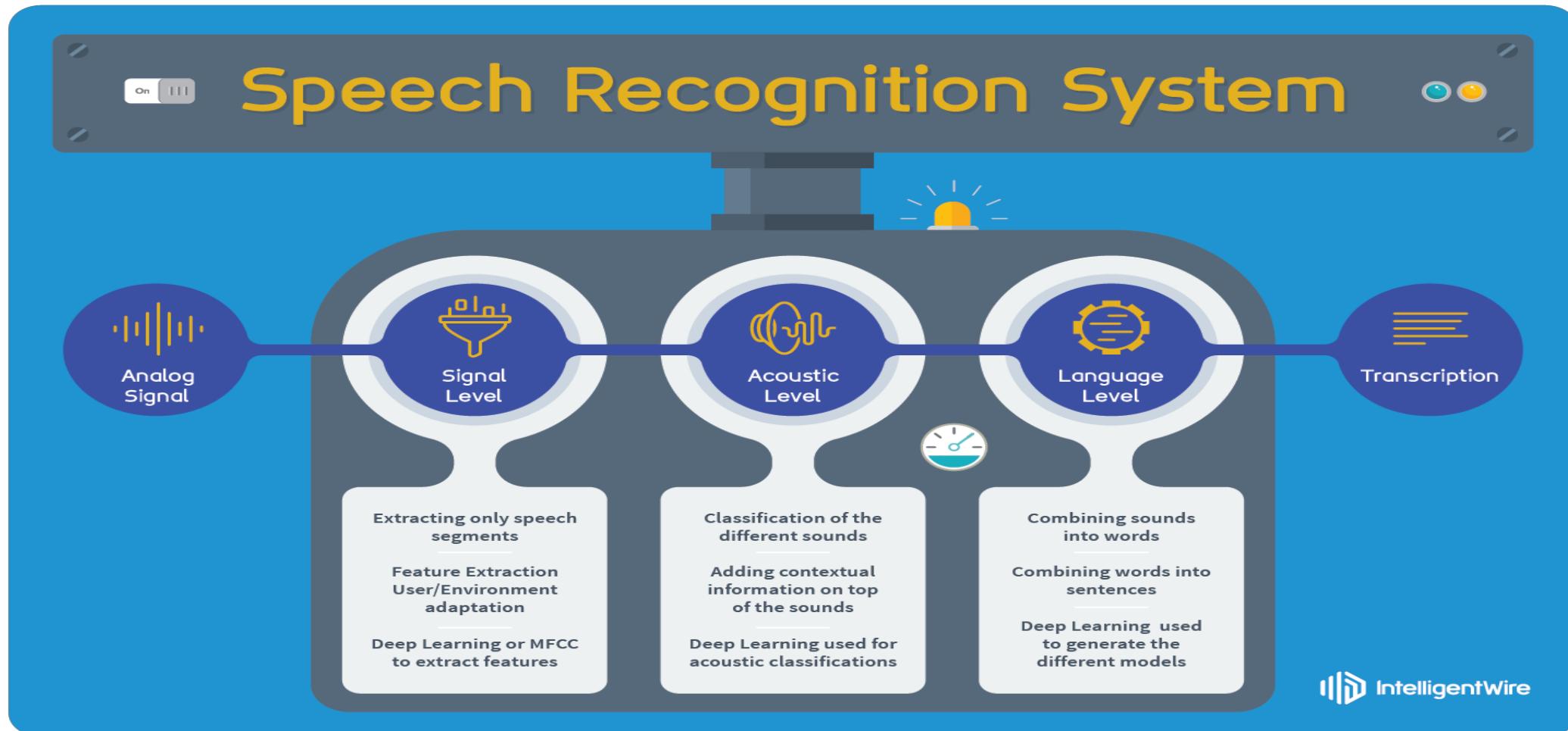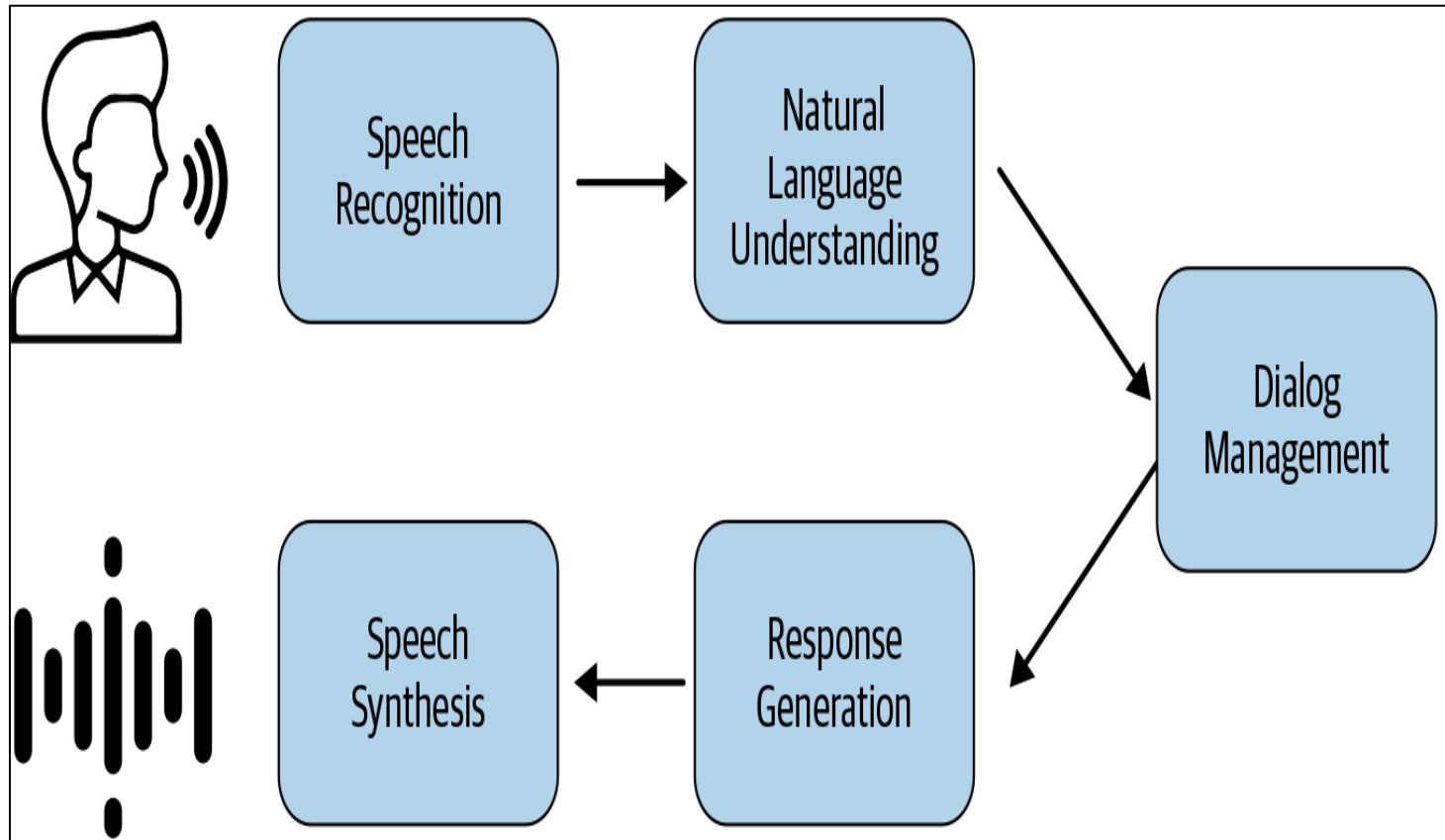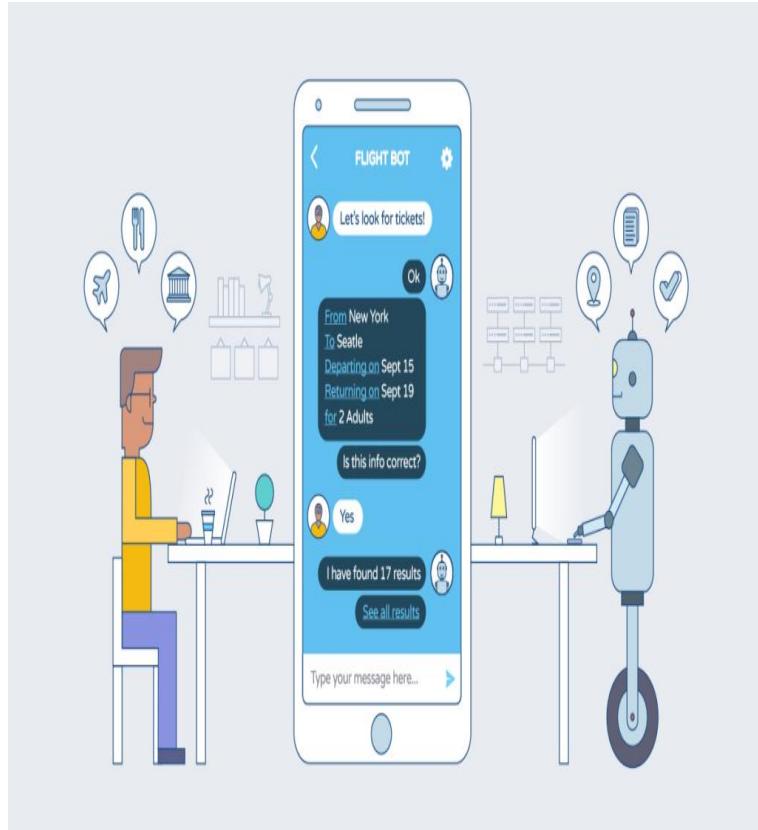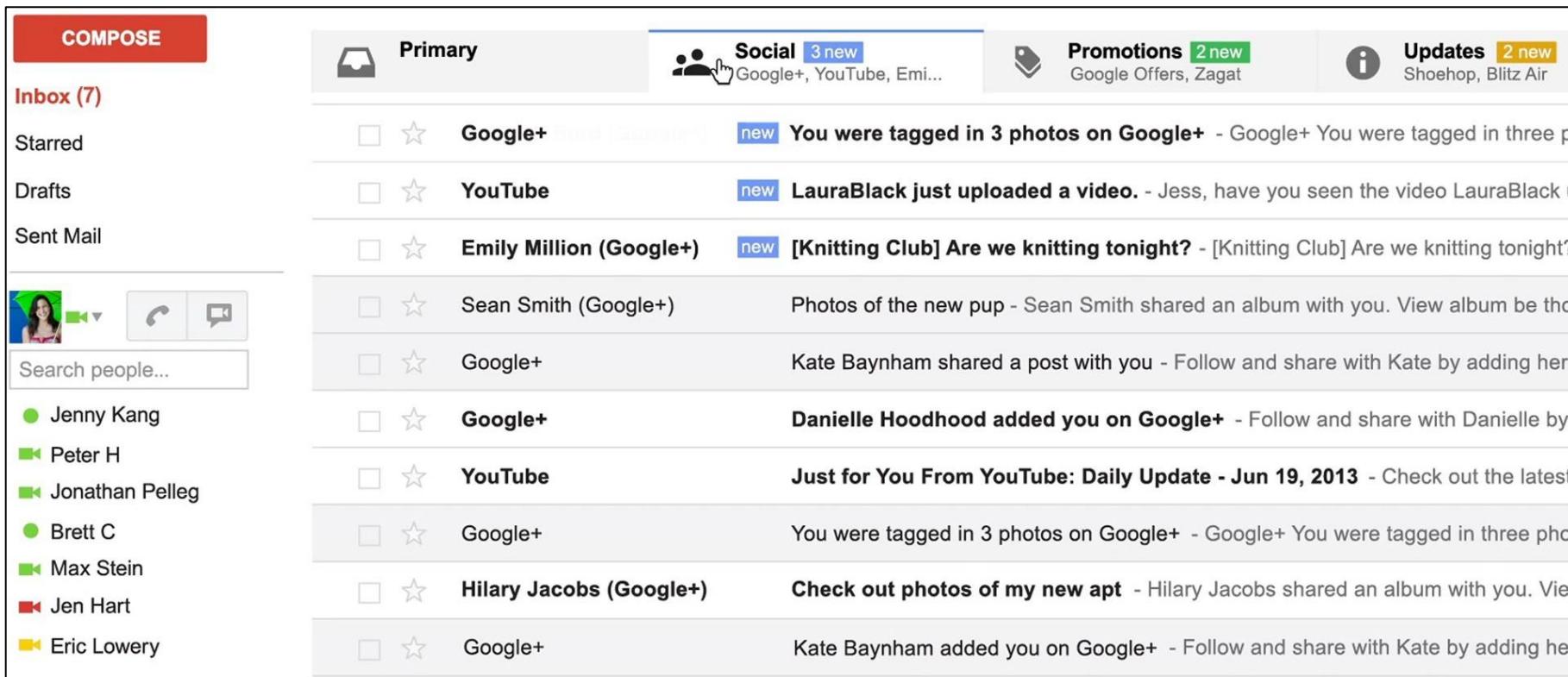
NLP Applications: Conversation Agents and Chatbots

NLP Applications :Text Classification

- spam / not spam
- priority level
- category (primary / social / promotions / updates)

# NLP Applications :Sentiment Analysis
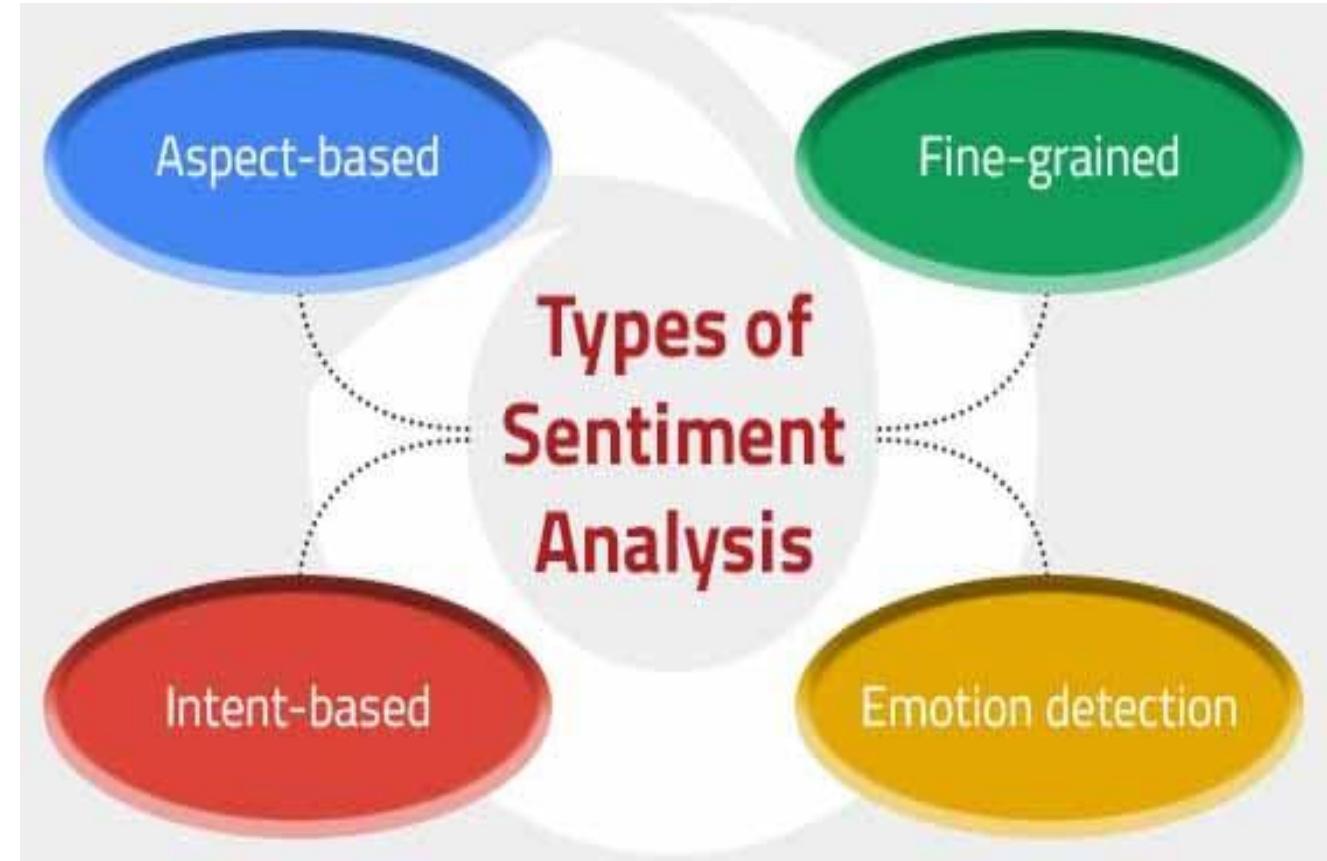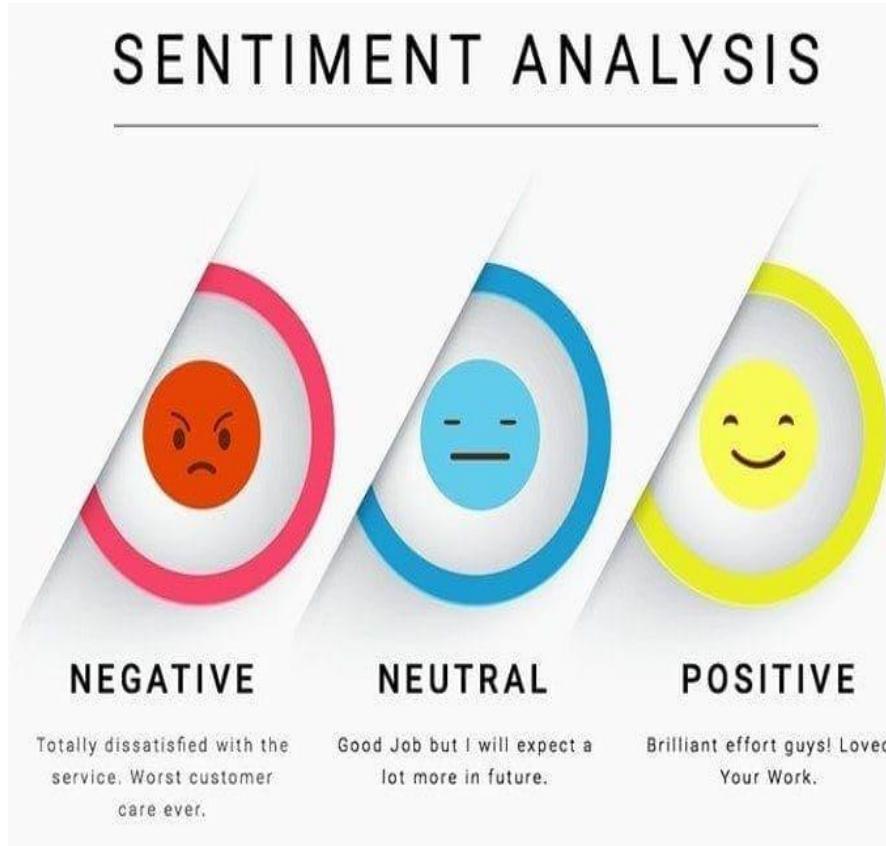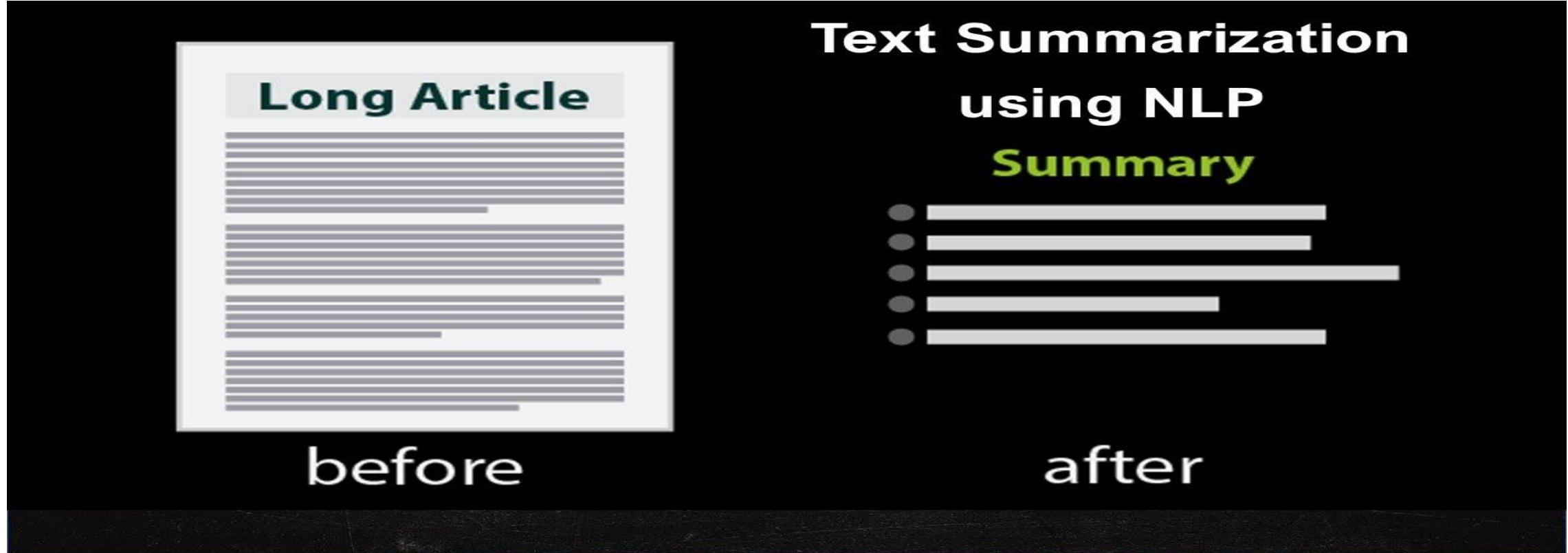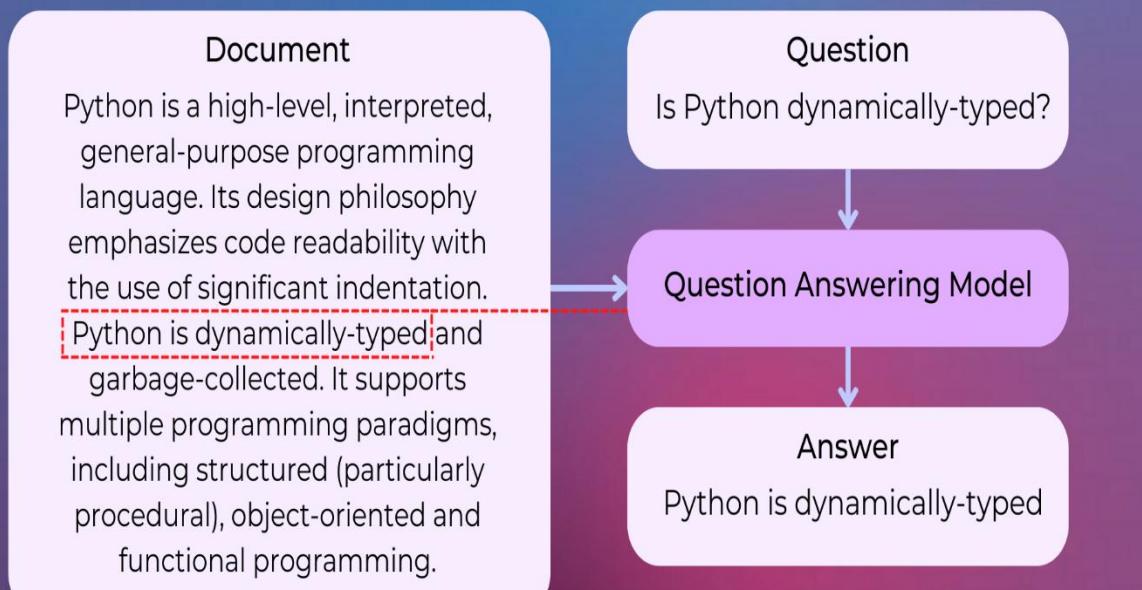
NLP Applications :Text Summarization

NLP Applications :Question Answering

Image source https://www.nlplanet.org/course-practical-nlp/02-practical-nlp-first-tasks/17-question-answering /
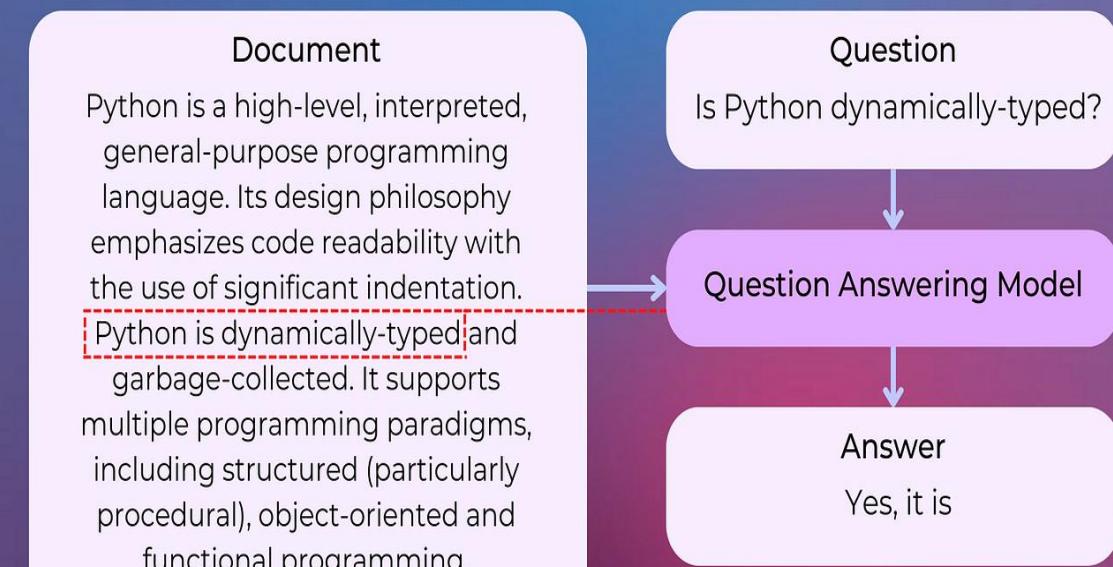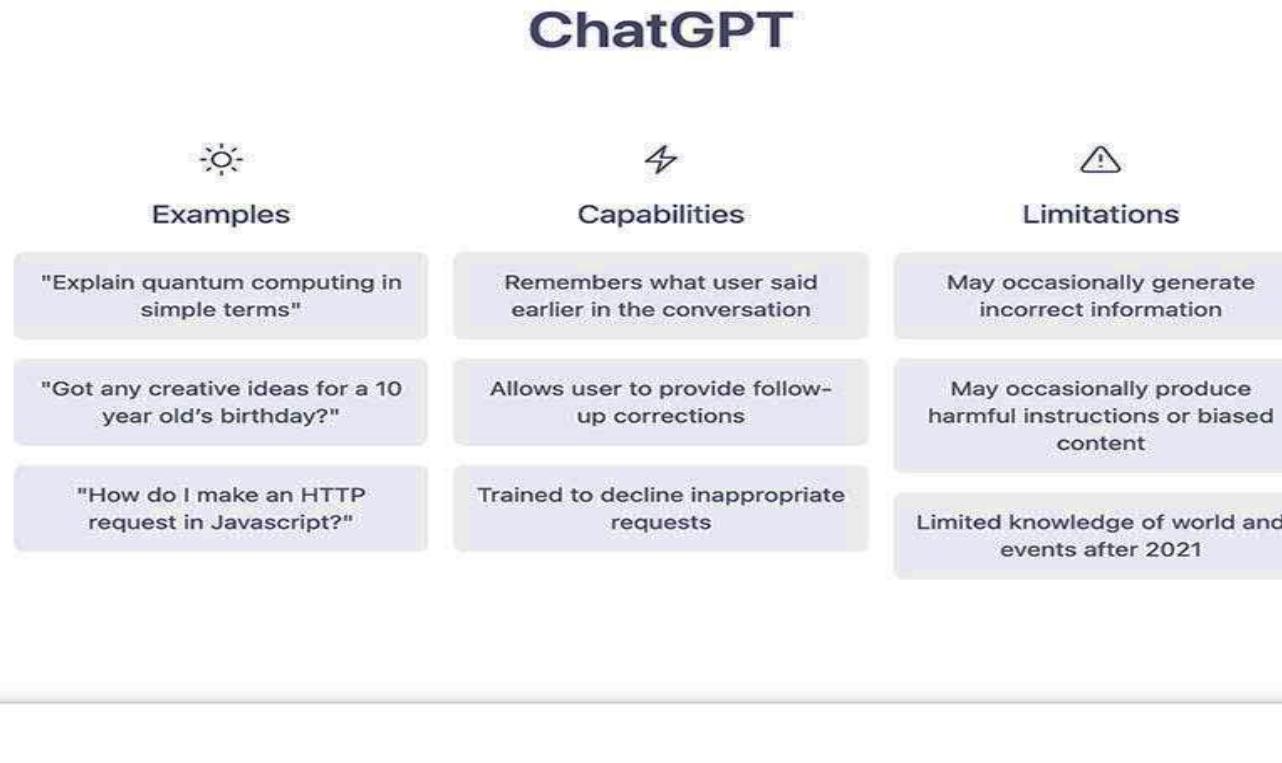
Generative AI: OpenAI's ChatGPT

ChatGPT

☀️ Examples	⚡ Capabilities	⚠️ Limitations
"Explain quantum computing in simple terms"	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?"	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?"	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them safer.

Lesson Agenda

❑ Course CSI

❑ What is Natural Language Processing (NLP)?

❑ Applications of Natural Language Processing

❑ Challenges in Natural Language Processing

❑ Approaches to NLP

❑ NLP Libraries in Python

Challenges in NLP

Ambiguity

Sparsity

Variation

Common knowledge

Language is Ambiguous :Words Have Many Meanings

- word "bass" can refer to a type of fish or a low-frequency sound.

- The word "bank" can refer to a financial institution or the edge of a river.
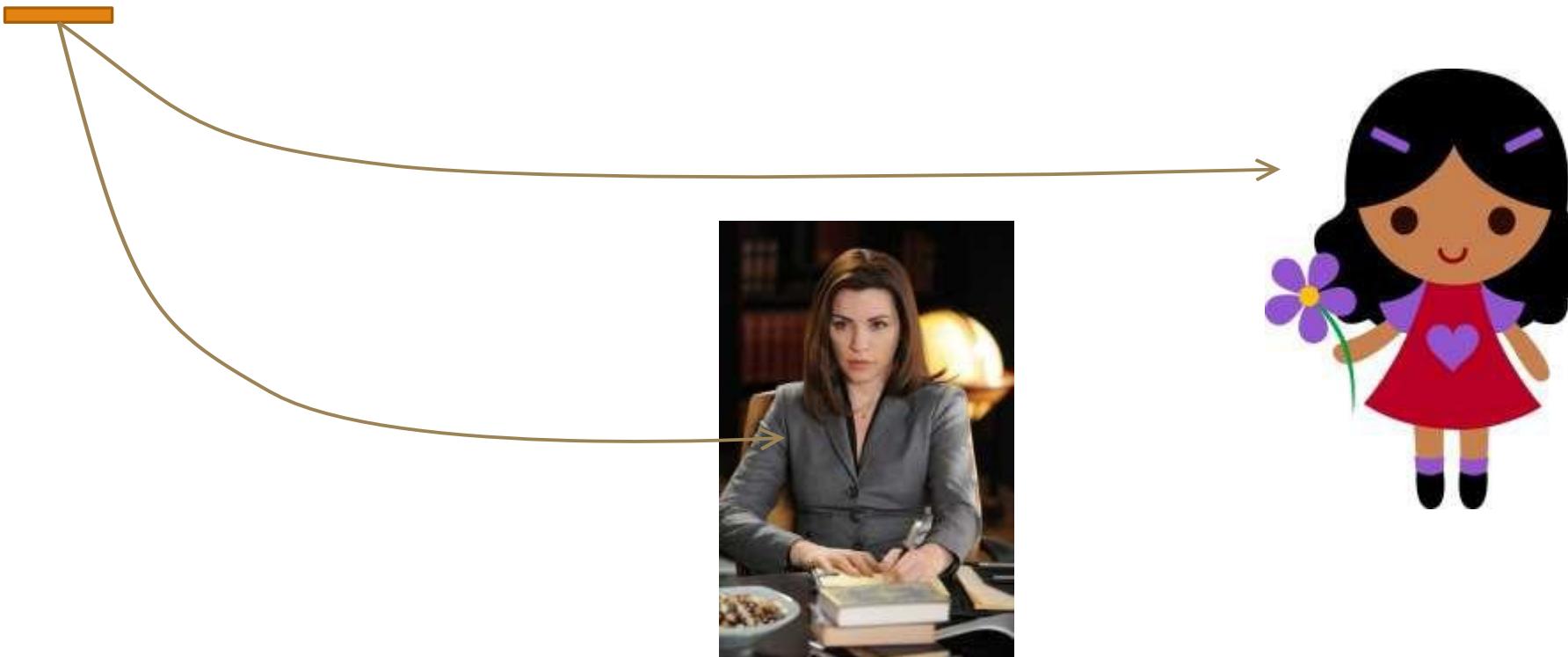
Language is Ambiguous :Attachment Ambiguities

She saw the man with the telescope.

Language is Ambiguous :Coreference Ambiguities

My girlfriend and I met my lawyer for a drink but **she** became ill and had to leave.

Challenges in NLP

Ambiguity
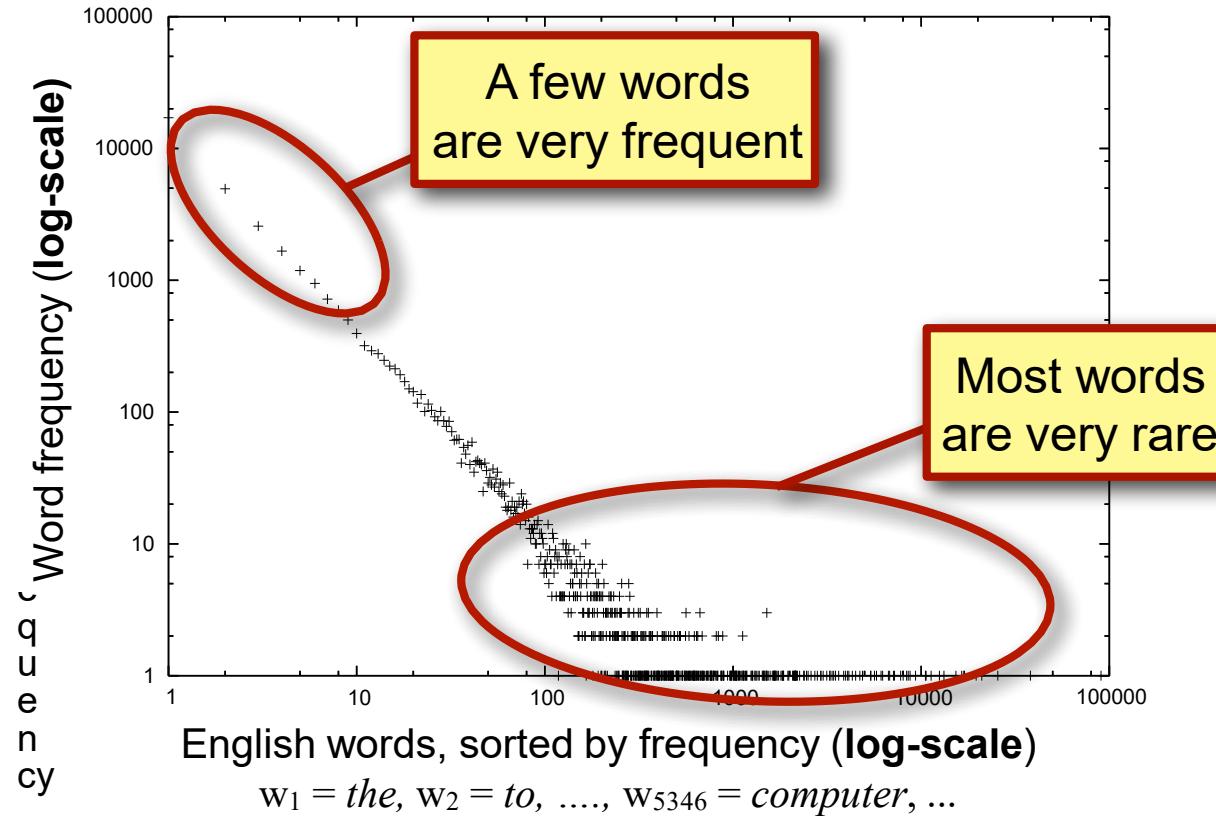
Sparsity

Variation
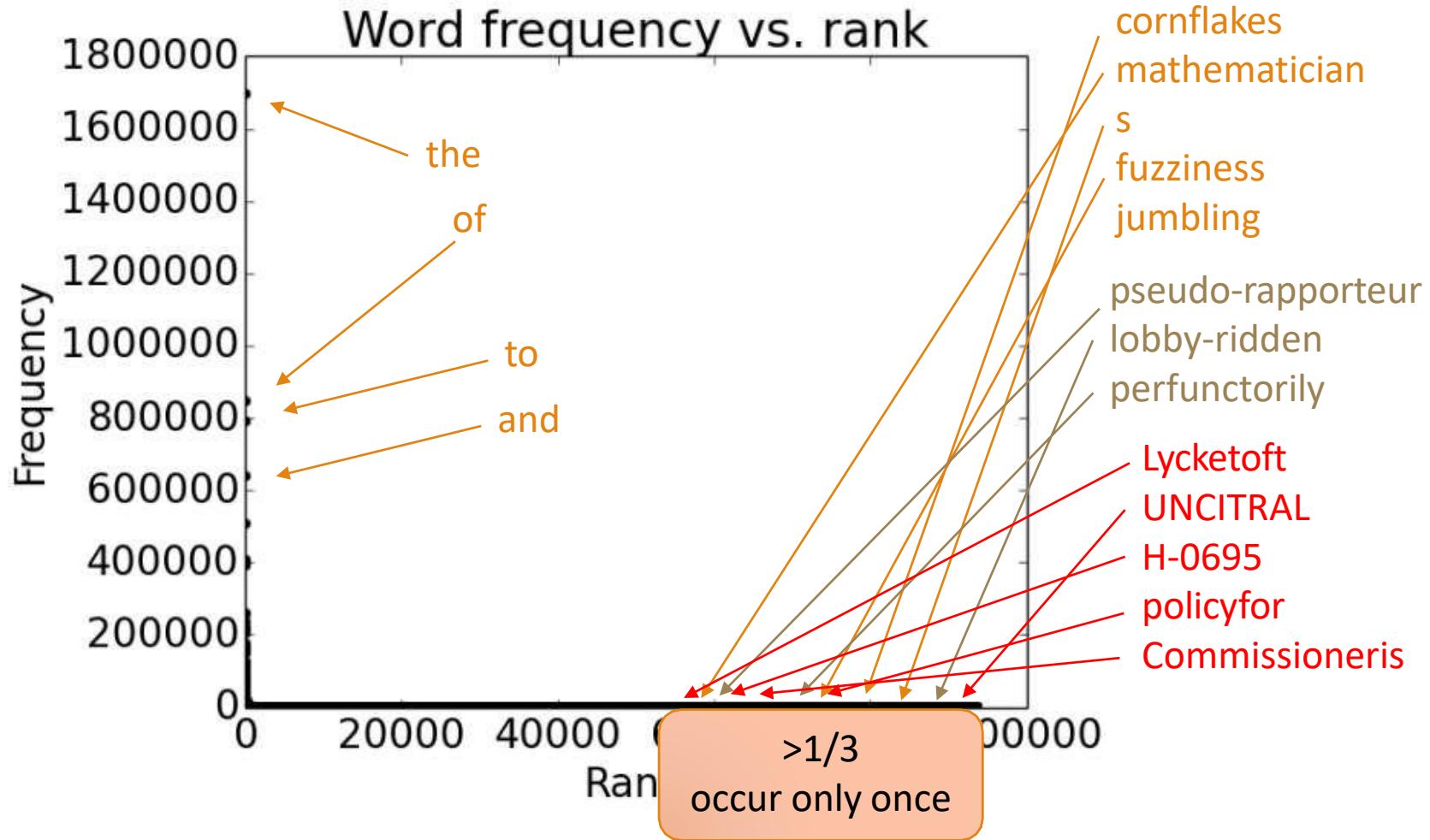
Common knowledge

Zipf's law: the long tail

How many words occur once, twice, 100 times, 1000 times?

A few words are very frequent

Most words are very rare

Word frequency (**log-scale**)

English words, sorted by frequency (**log-scale**)

$w_1 = the$, $w_2 = to$, ..., $w_{5346} = computer$, ...

Sparsity

Challenges in NLP

Ambiguity

Sparsity

Variation

Common knowledge

Variation within the same language

- *Variation within the same language*

  - Lexical variation

    - She gave the book to Tom **vs**. She gave Tom the book

  - Regional or Geographic Variation

  - Social Variation

  - Stylistic Variation

  - Generational Variation

- *Variation between different languages*

Challenges in NLP

Ambiguity

Sparsity

Variation
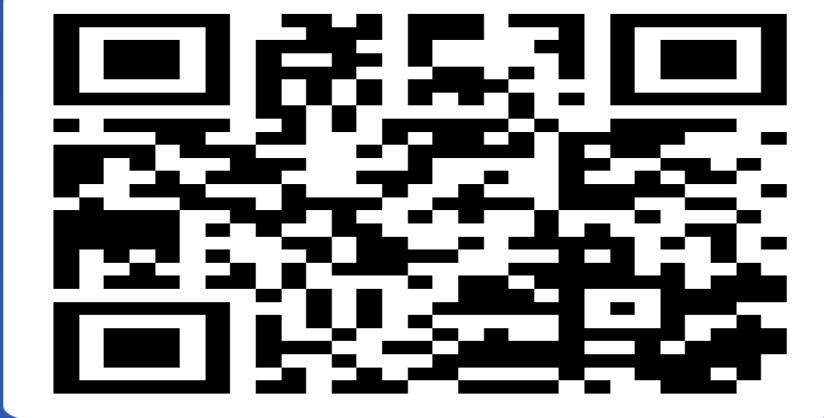
Common knowledge

Common knowledge

It is the set of all facts that most humans are aware of

- ➢ man bit dog

- ➢ dog bit man

Other Challenges ?

Join at
slido.com
#3209 585

Lesson Agenda

❑ Course CSI

❑ What is Natural Language Processing (NLP)?

❑ Applications of Natural Language Processing

❑ Challenges in Natural Language Processing

❑ Approaches to NLP

❑ NLP Libraries in Python

Approaches to NLP

- **Heuristics-Based NLP**
  - Regular Expression

- **Machine Learning for NLP**
  - Supervised
  - Unsupervised

- **Deep Learning for NLP**

  - Recurrent neural networks

  - Long short-term memory

- **Transformers**

Lesson Agenda

❑ Course CSI

❑ What is Natural Language Processing (NLP)?

❑ Applications of Natural Language Processing

❑ Challenges in Natural Language Processing

❑ Approaches to NLP

❑ NLP Libraries in Python

NLP Libraries in Python

52

Summary

❑ Language to Knowledge

❑ Clearing the Confusion: AI vs. Machine Learning vs. Deep Learning Differences

❑ Lots of applications…

❑ It's quite difficult

❑ Varied, sparse, and lots of ambiguities

What's Next?

Text preprocessing and exploratory analysis

- Regular expression
- Tokenization
- Stemming
- Removing stop words
- Part Of Speech tagging (POS)

Q&A