

Question Generation using Natural Language Processing Techniques



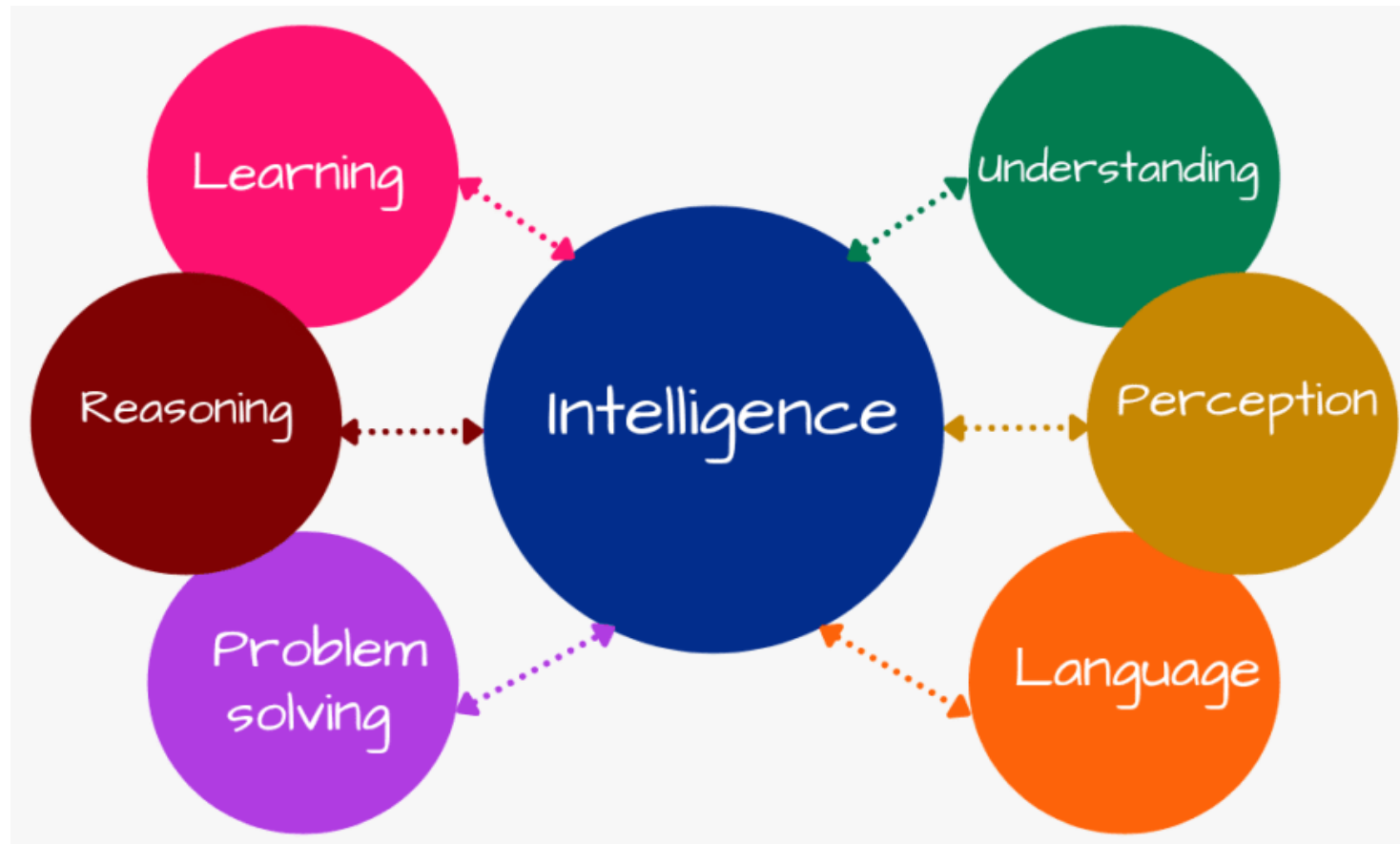
ACISN'23

– PRESENTED BY DR AID AICHA

<https://sites.google.com/view/acisn23>

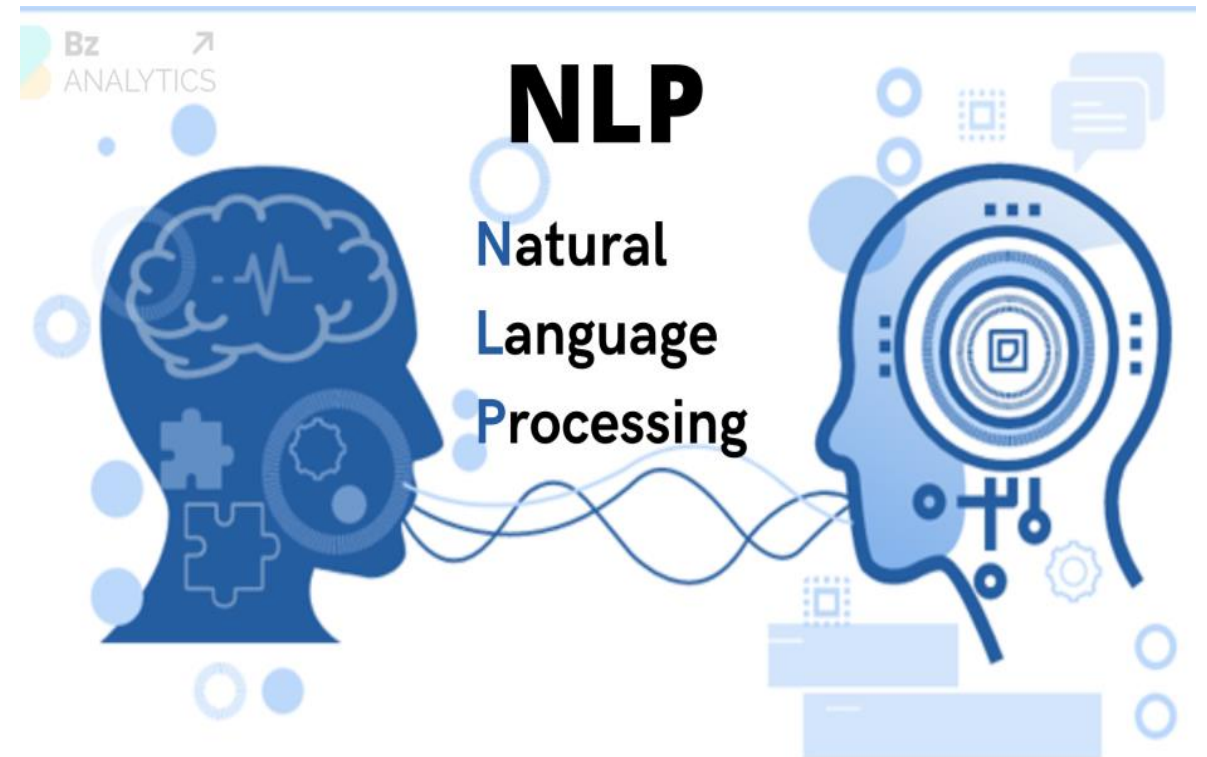
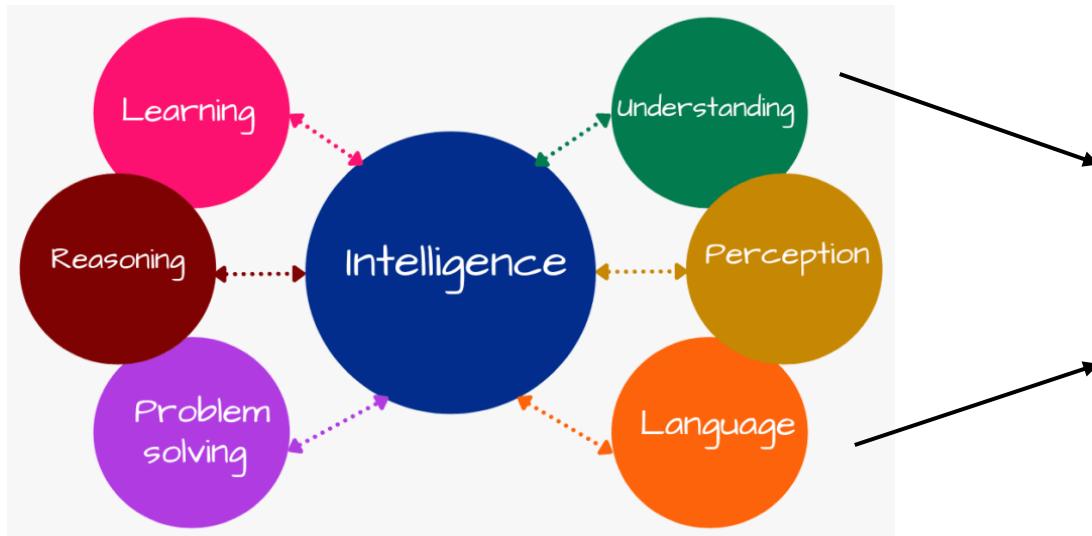
Artificial Intelligence & NLP

- Imitation and simulation of human reasoning and **human intelligence** in machines and computer systems.



Artificial Intelligence & NLP

- NLP is a field of artificial intelligence that focuses on the interaction between computers and human language, enabling machines to **understand**, interpret, and **generate** human-like text.



NLP Tasks and Applications

- NLP is a field of artificial intelligence that focuses on the interaction between computers and human language, enabling machines to **understand**, interpret, and **generate** human-like text.
- Some key NLP **tasks** include:



Machine
Translation



Sentiment
Analysis



Information
extraction



Question
Answering



Automatic
Summarization



Text
Classification



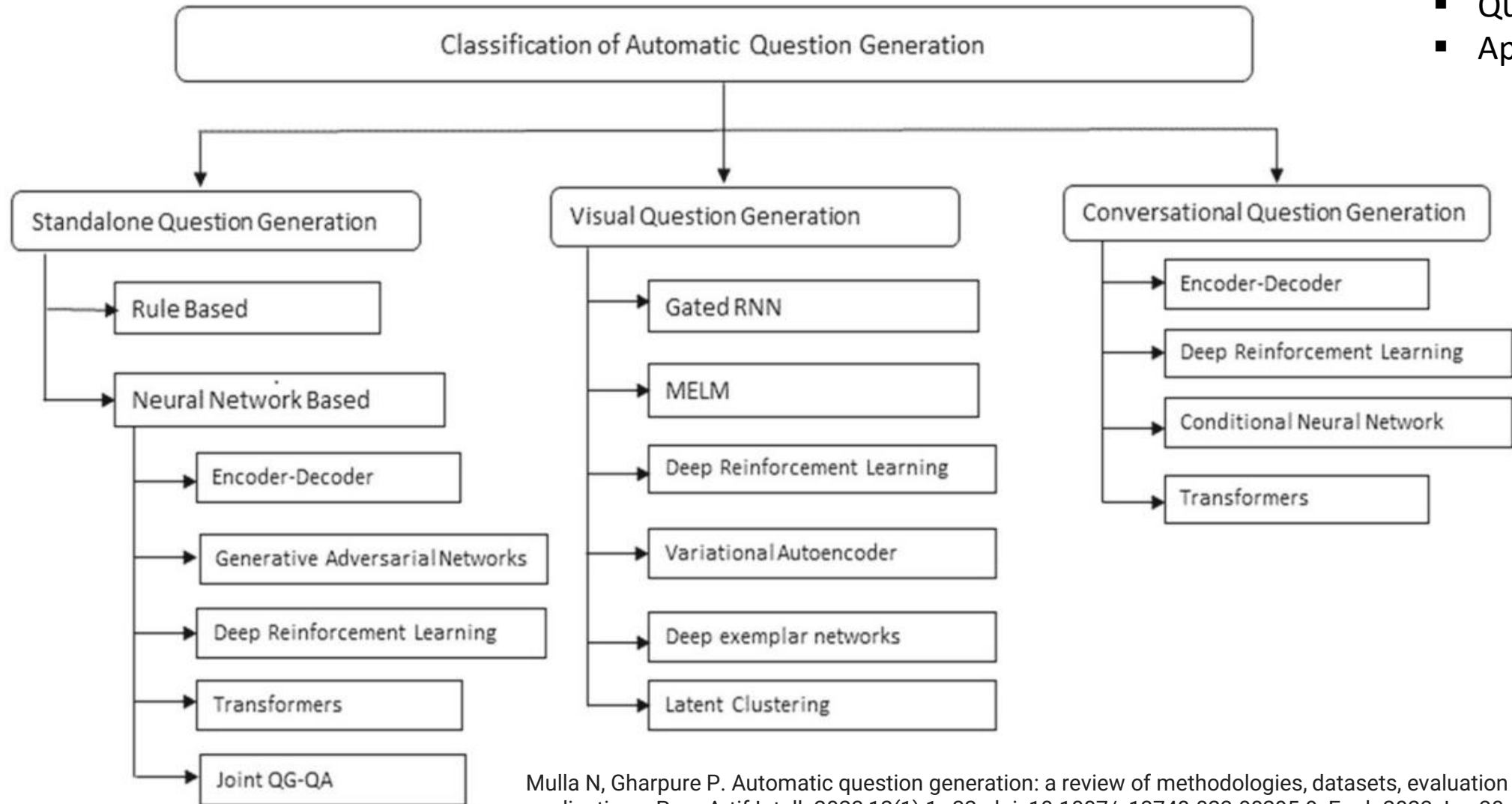
Question
Generation

Question Generation in NLP

- Question Generation refers to the task of automatically **creating questions from a given text or images**.
- Question Generation, Visual Question Generation, and Conversational Question Generation.
- The goal is to generate **relevant** and **contextually accurate** (**correct** syntactically, semantically as well as meaningful) questions that a human might ask about provided information and context.
- Question Generation can be **closed(specific) domain** or **open domain**.
- **Used for** educational purposes, content creation, tests and quizzes, self-assessment, enhancing user engagement in chatbot interactions, conversational systems, etc.
- Various **questions types**: multiple-choice, yes/no, matching-the-following, fill-in-the-blanks, factoid and causal, wh-questions, Extractive/Abstractive, etc.

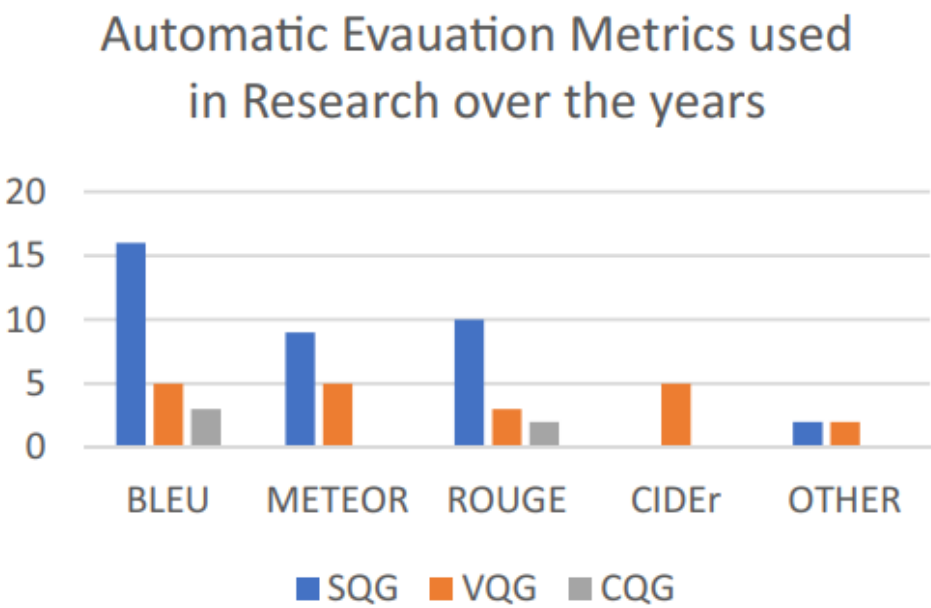
Question Generation in NLP

- Input
- Domain
- Context
- Question Type
- Application



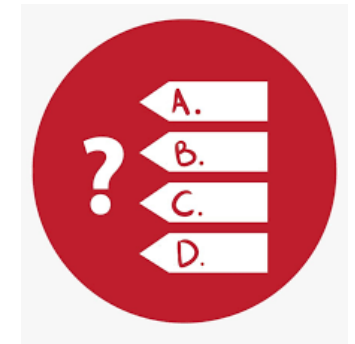
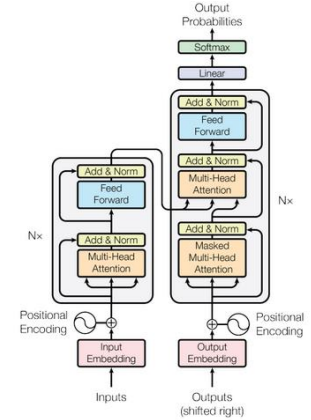
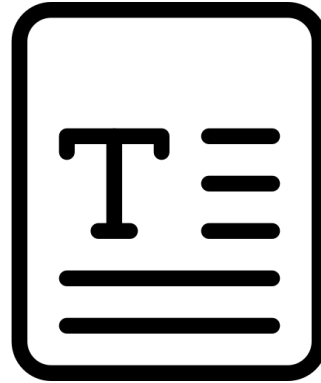
Question Generation in NLP

Dataset name	Source
SQuAD (Stanford Question Answering Dataset)[46]	Wikipedia
CMU Q/A Dataset[108]	Wikipedia
News QA [47]	CNN news articles
DeepMind Q&A Dataset [109]	CNN + Daily Mail
WIKIQA [110]	Bing Query logs
MSMARCO [111]	Bing Query logs
RACE [112]	English exams
LearningQ[106]	TED-Ed and Khan Academy
NarrativeQA[107]	Stories and human-generated summaries
Natural Questions [113]	Wikipedia and aggregated queries through Google Search engine

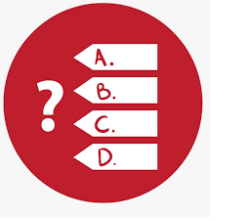


Question Generation in NLP – Case Study

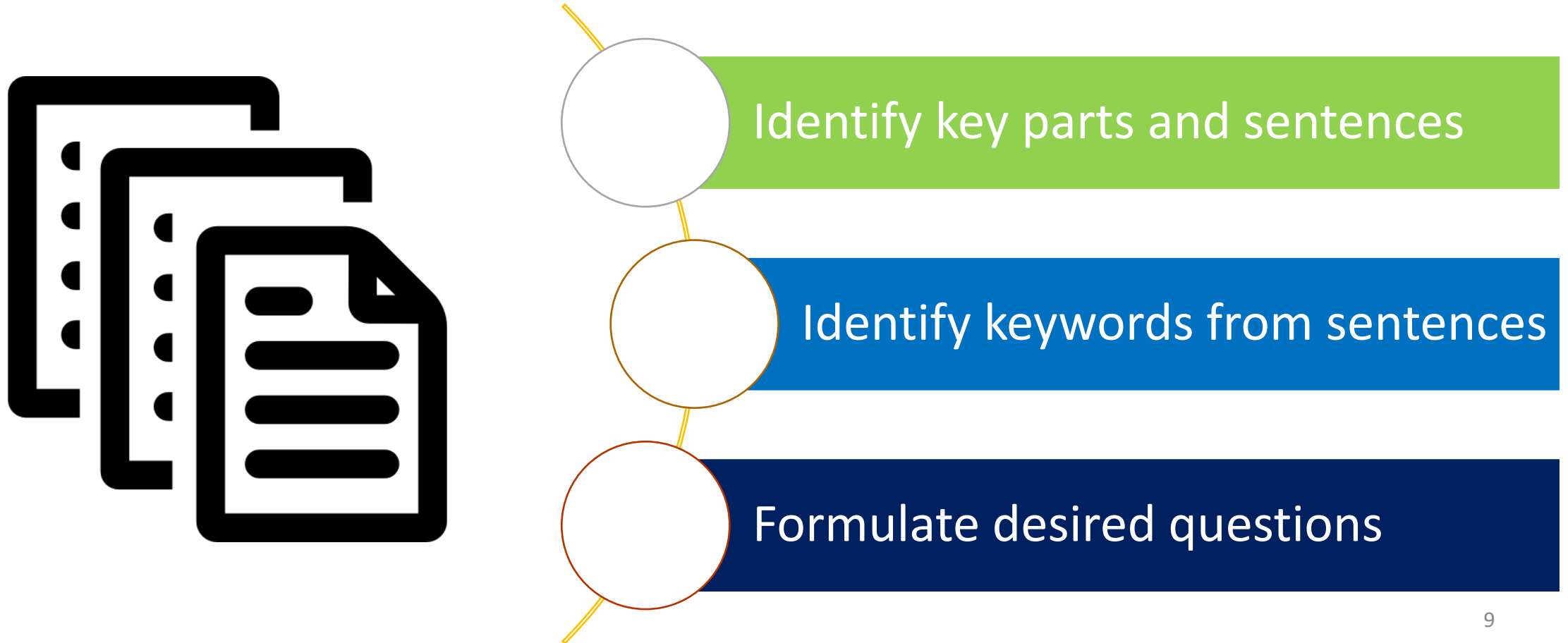
- **Input** : Text
- **Dataset** : SQuAD
- **Purpose** : Education
- **Technique** : DL – Transformers
- **Model** : Pretrained T5
- **Output** - Question Type : MCQ, Fill-in-the-blank, Yes/No



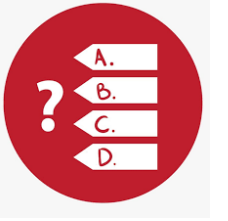
Question Generation in NLP – Case Study



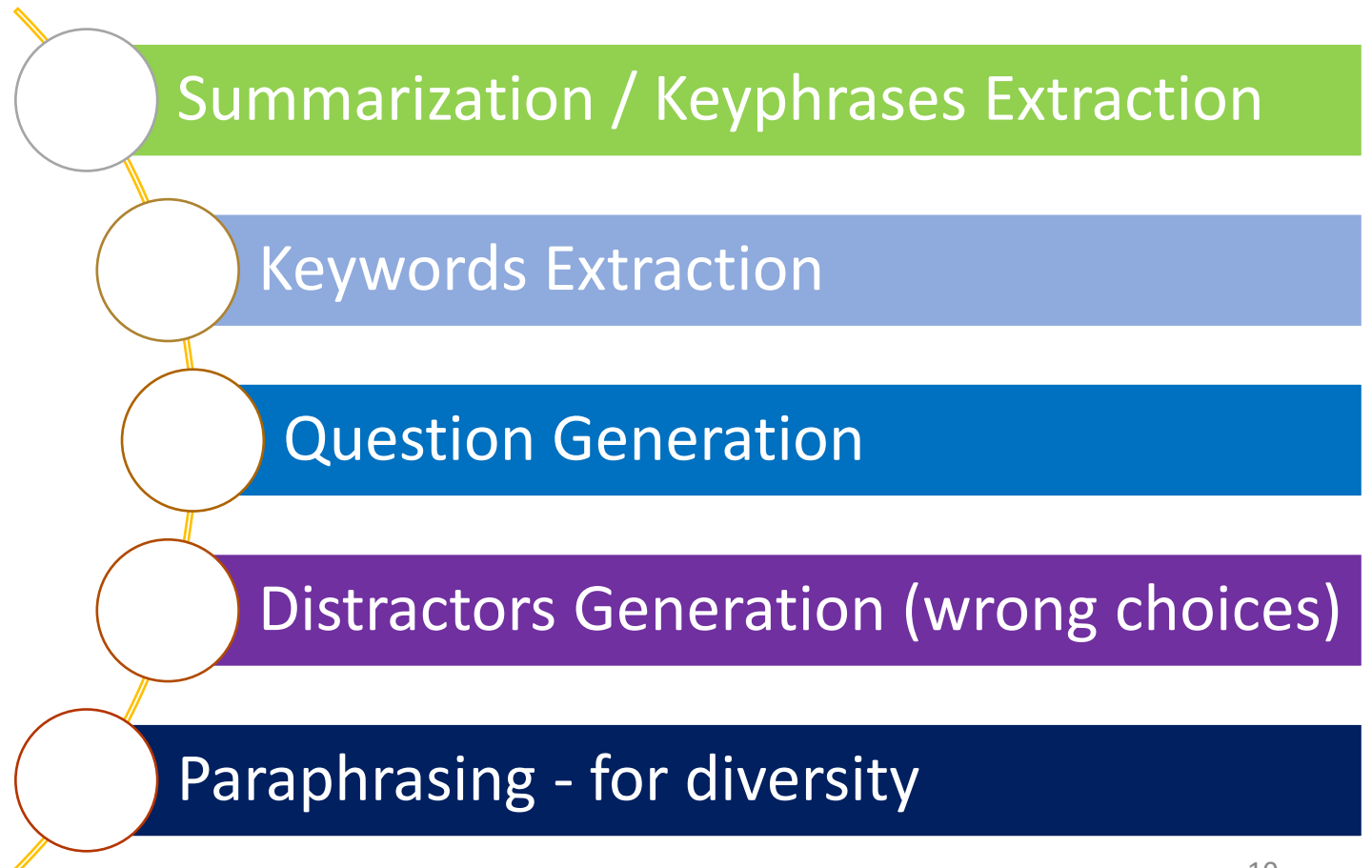
- Pipeline Steps – from a human perspective:



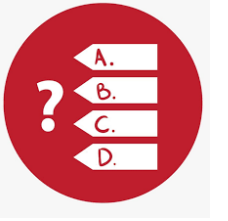
Question Generation in NLP – Case Study



■ Pipeline Steps – from NLP perspective:



Question Generation in NLP – Case Study



- Pipeline Steps – from NLP perspective:

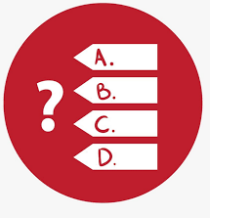
Chapter 6 Social Studies

The Amazon River

The Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

The total length of the river as measured from the headwaters of the Ucayali-Apurímac river system in southern Peru is at least 4,000 miles (6,400 km), which makes it slightly shorter than the Nile River but still the equivalent of the distance from New York City to Rome. Its westernmost source is high in the Andes Mountains, within 100 miles (160 km) of the Pacific Ocean, and its mouth is in the Atlantic Ocean, on the northeastern coast of Brazil. However, both the length of the Amazon and its ultimate source have been subjects of debate since the mid-20th century, and some claim that the Amazon is longer than the Nile.

Question Generation in NLP – Case Study



- Pipeline Steps – from NLP perspective:

Chapter 6 Social Studies The Amazon River

The Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

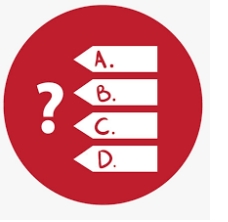
The total length of the river as measured from the headwaters of the Ucayali-Apurímac river system in southern Peru is at least 4,000 miles (6,400 km), which makes it slightly shorter than the Nile River but still the equivalent of the distance from New York City to Rome. Its westernmost source is high in the Andes Mountains, within 100 miles (160 km) of the Pacific Ocean, and its mouth is in the Atlantic Ocean, on the northeastern coast of Brazil. However, both the length of the Amazon and its ultimate source have been subjects of debate since the mid-20th century, and some claim that the Amazon is longer than the Nile.



Keyphrases
Keywords
Extraction

The **Amazon** in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

Question Generation in NLP – Case Study



■ Pipeline Steps – from NLP perspective:

Chapter 6 Social Studies The Amazon River

The Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

The total length of the river as measured from the headwaters of the Ucayali-Apurímac river system in southern Peru is at least 4,000 miles (6,400 km), which makes it slightly shorter than the Nile River but still the equivalent of the distance from New York City to Rome. Its westernmost source is high in the Andes Mountains, within 100 miles (160 km) of the Pacific Ocean, and its mouth is in the Atlantic Ocean, on the northeastern coast of Brazil. However, both the length of the Amazon and its ultimate source have been subjects of debate since the mid-20th century, and some claim that the Amazon is longer than the Nile.



Keyphrases
Keywords
Extraction

The **Amazon** in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

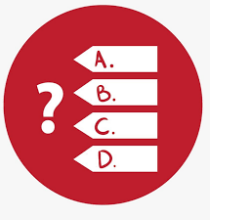


Question
Generation

Which is the world's second longest river ?



Question Generation in NLP – Case Study



■ Pipeline Steps – from NLP perspective:

Chapter 6 Social Studies The Amazon River

The Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

The total length of the river as measured from the headwaters of the Ucayali-Apurímac river system in southern Peru is at least 4,000 miles (6,400 km), which makes it slightly shorter than the Nile River but still the equivalent of the distance from New York City to Rome. Its westernmost source is high in the Andes Mountains, within 100 miles (160 km) of the Pacific Ocean, and its mouth is in the Atlantic Ocean, on the northeastern coast of Brazil. However, both the length of the Amazon and its ultimate source have been subjects of debate since the mid-20th century, and some claim that the Amazon is longer than the Nile.



Keyphrases
Keywords
Extraction

The **Amazon** in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.



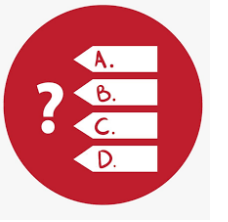
Question
Generation

Which is the world's second longest river ?

Distractors
Generation

- | | |
|---|-------------|
| a | Mississippi |
| b | Amazon |
| c | Nile |
| d | Yangtze |

Question Generation in NLP – Case Study



- Pipeline Steps – from NLP perspective:

Chapter 6 Social Studies The Amazon River

The Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

The total length of the river as measured from the headwaters of the Ucayali-Apurímac river system in southern Peru is at least 4,000 miles (6,400 km), which makes it slightly shorter than the Nile River but still the equivalent of the distance from New York City to Rome. Its westernmost source is high in the Andes Mountains, within 100 miles (160 km) of the Pacific Ocean, and its mouth is in the Atlantic Ocean, on the northeastern coast of Brazil. However, both the length of the Amazon and its ultimate source have been subjects of debate since the mid-20th century, and some claim that the Amazon is longer than the Nile.

Which is the world's second longest river ?

- a** Mississippi
- b** Amazon
- c** Nile
- d** Yangtze

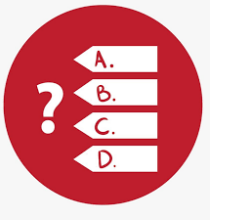


Paraphrasing

- What river holds the title of being the second-longest in the world ?
- What river ranks as the world's second-longest ?

- a** Mississippi
- b** Amazon
- c** Nile
- d** Yangtze

Question Generation in NLP – Case Study



- Pipeline Steps – from NLP perspective:

Chapter 6 Social Studies The Amazon River

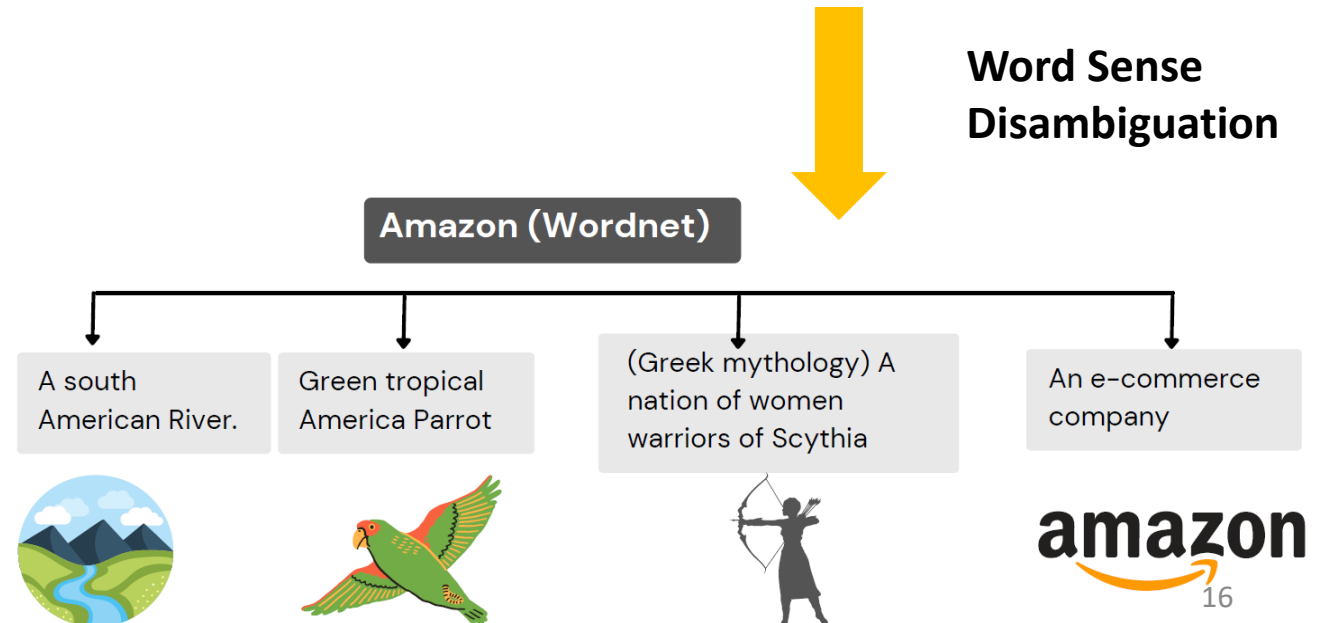
The Amazon in South America is the largest river by discharge volume of water in the world, and the second-longest river in the world.

The total length of the river as measured from the headwaters of the Ucayali-Apurímac river system in southern Peru is at least 4,000 miles (6,400 km), which makes it slightly shorter than the Nile River but still the equivalent of the distance from New York City to Rome. Its westernmost source is high in the Andes Mountains, within 100 miles (160 km) of the Pacific Ocean, and its mouth is in the Atlantic Ocean, on the northeastern coast of Brazil. However, both the length of the Amazon and its ultimate source have been subjects of debate since the mid-20th century, and some claim that the Amazon is longer than the Nile.

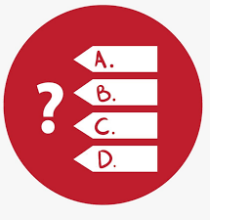
Which is the world's second longest river ?

- a** Mississippi
- b** Amazon
- c** Nile
- d** Yangtze

Word Sense
Disambiguation

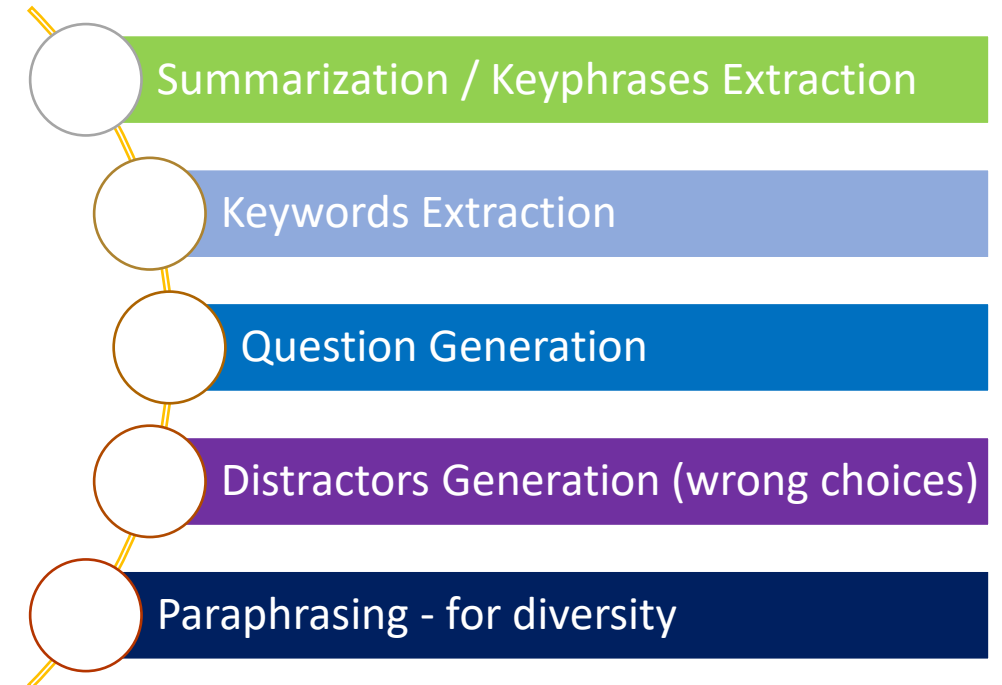


Question Generation in NLP – Case Study

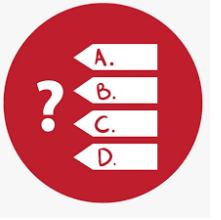


■ Pipeline Steps – Techniques & Tools:

1. Summarization: transformer models like BERT & T5.
2. Keywords and Keyphrases Extraction and Sentence Mapping: PKE, Yake, Rake, KeyBert, FlashText, etc.
3. Question Generation: transformer models like T5.
4. Word Sense Disambiguation: Bert-WSD, PYWSD (MaxSim, Lesk algos), etc.
5. Distractors generation: WordNet, Sense2Vec, ConceptNet, etc.
6. Paraphrasing: transformer models like GPT.



Question Generation in NLP – Case Study



■ Pipeline Steps – Coding Demo:

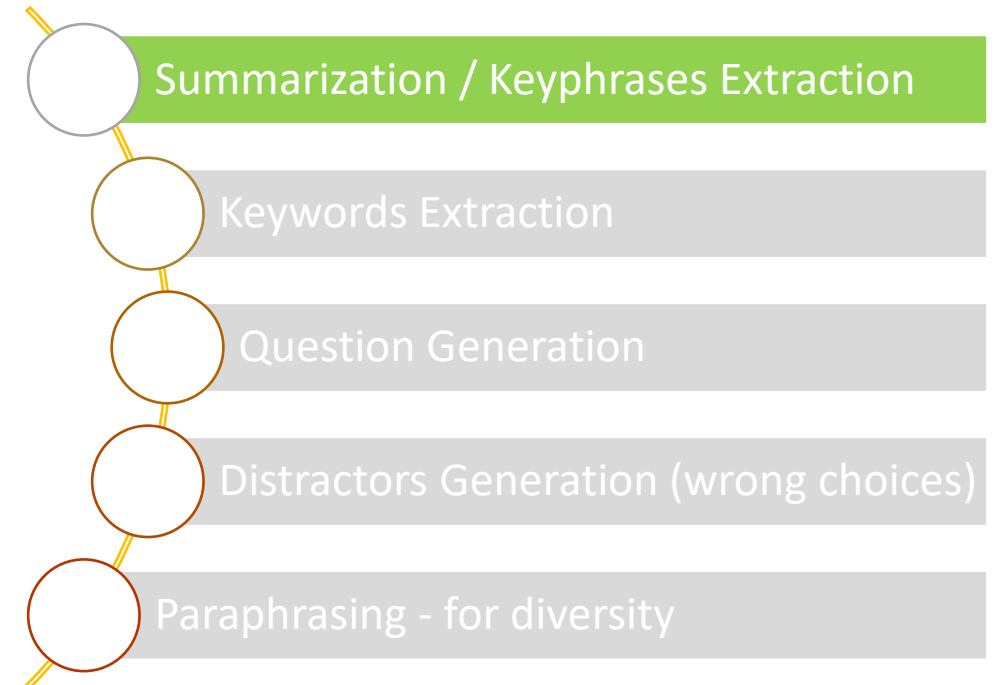
```
✓ [34] summary_model = T5ForConditionalGeneration.from_pretrained('t5-base')
11s summary_tokenizer = T5Tokenizer.from_pretrained('t5-base')

encoding = tokenizer.encode_plus(text, max_length=max_len, pad_to_max_length=False, tr

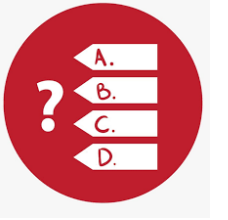
input_ids, attention_mask = encoding["input_ids"], encoding["attention_mask"]

outs = model.generate(input_ids=input_ids,
                      attention_mask=attention_mask,
                      early_stopping=True,
                      num_beams=3,
                      num_return_sequences=1,
                      no_repeat_ngram_size=2,
                      min_length = 75,
                      max_length=300)

dec = [tokenizer.decode(ids, skip_special_tokens=True) for ids in outs]
summary = dec[0]
```



Question Generation in NLP – Case Study



■ Pipeline Steps – Coding Demo:

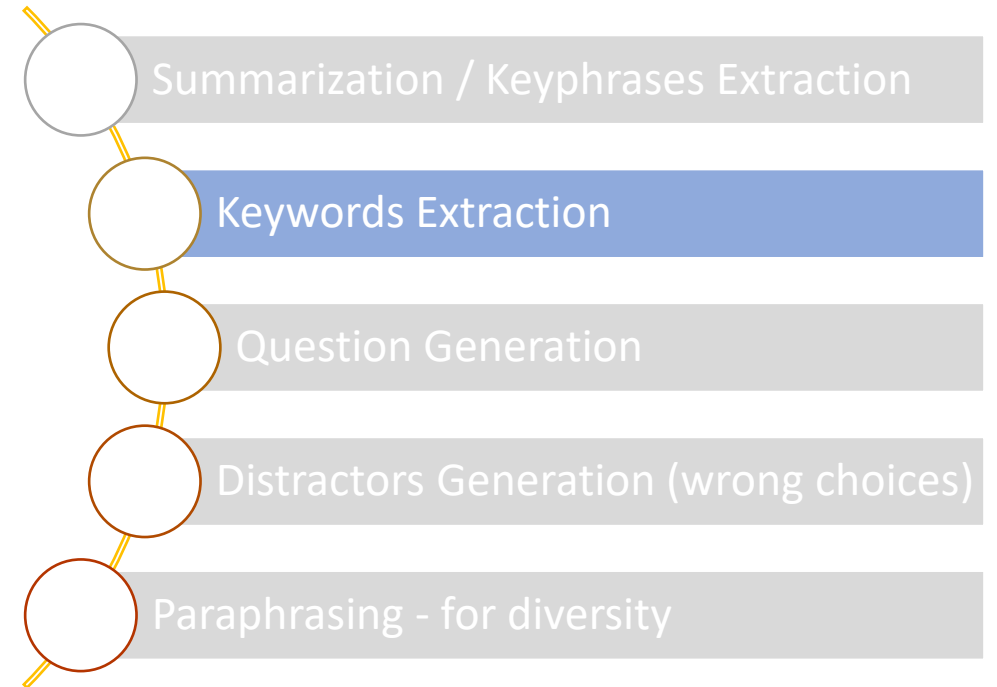
```
extractor = pke.unsupervised.MultipartiteRank()
extractor.load_document(input=content, language='en')

stoplist = list(string.punctuation)
stoplist += ['-lrb-', '-rrb-', '-lcb-', '-rcb-', '-lsb-', '-rsb-']
stoplist += stopwords.words('english')

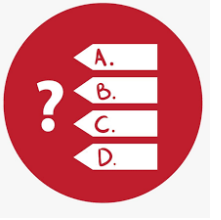
pos = {'PROPN', 'NOUN'}
extractor.candidate_selection(pos=pos)

extractor.candidate_weighting(alpha=1.1, threshold=0.75, method='average')
keywords = extractor.get_n_best(n=15)

for val in keywords:
    out.append(val[0])
```



Question Generation in NLP – Case Study



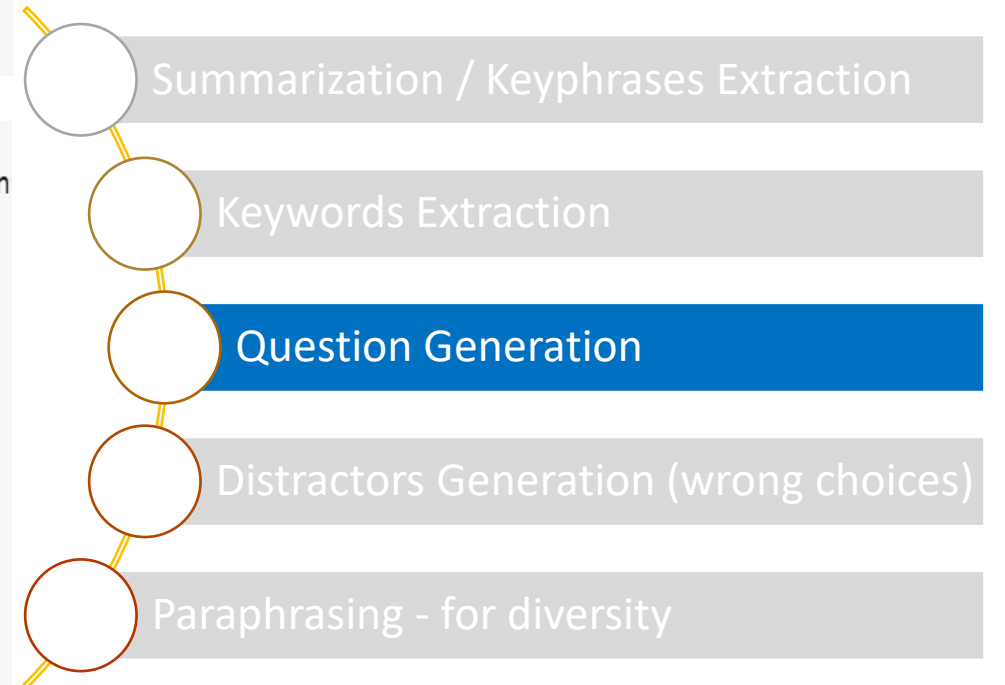
■ Pipeline Steps – Coding Demo:

```
question_model = T5ForConditionalGeneration.from_pretrained('ramsrigouthamg/t5_squad_v1')
question_tokenizer = T5Tokenizer.from_pretrained('ramsrigouthamg/t5_squad_v1')
```

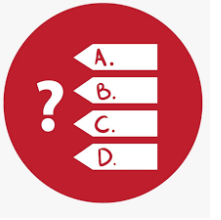
```
text = "context: {} answer: {}".format(context, answer)
encoding = tokenizer.encode_plus(text, max_length=384, pad_to_max_length=False, truncation
input_ids, attention_mask = encoding["input_ids"], encoding["attention_mask"]
```

```
outs = model.generate(input_ids=input_ids,
                      attention_mask=attention_mask,
                      early_stopping=True,
                      num_beams=5,
                      num_return_sequences=1,
                      no_repeat_ngram_size=2,
                      max_length=72)
```

```
dec = [tokenizer.decode(ids, skip_special_tokens=True) for ids in outs]
question = dec[0].replace("question:", "")
```



Question Generation in NLP – Case Study

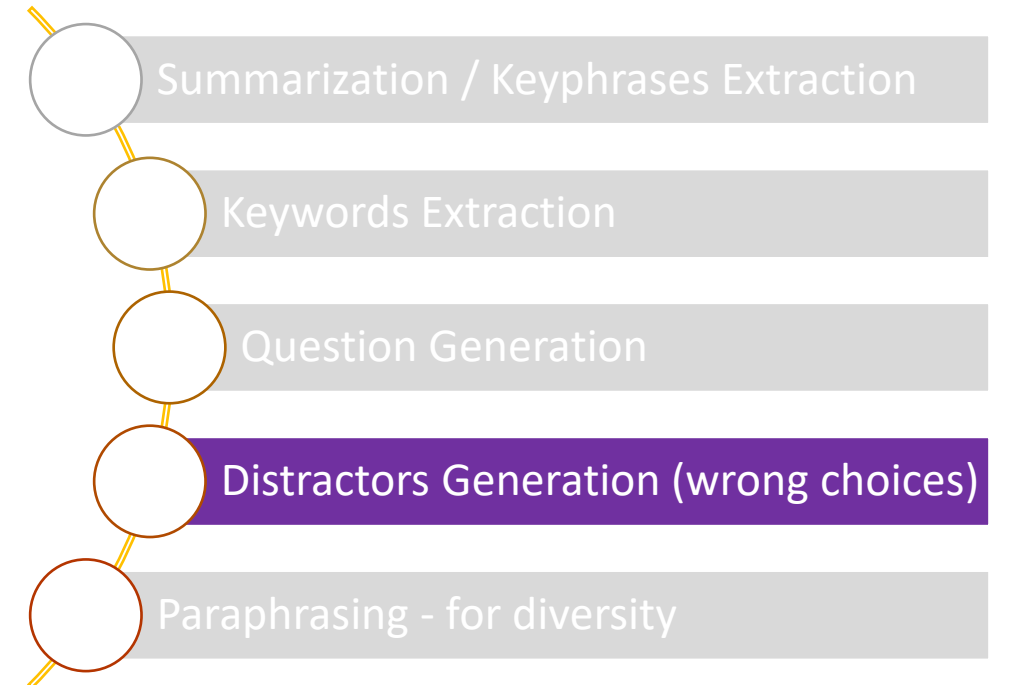


■ Pipeline Steps – Coding Demo:

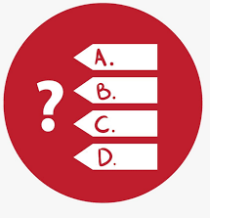
```
distractors=[]
try:
    syn = wn.synsets(word,'n')[s]
    word= word.lower()
    orig_word = word
    if len(word.split()) > 0:
        word = word.replace(" ", "_")

    hypernym = syn.hypernyms()
    if len(hypernym) == 0:
        return distractors

    for item in hypernym[0].hyponyms():
        name = item.lemmas()[0].name()
        if name == orig_word:
            continue
        name = name.replace("_", " ")
        name = " ".join(w.capitalize() for w in name.split())
        if name is not None and name not in distractors:
            distractors.append(name)
except:
    print ("Wordnet distractors not found")
```



Question Generation in NLP – Case Study

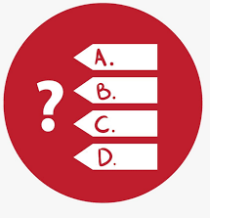


■ Pipeline Steps – Coding Demo:

Original Text >>

A Lion lay asleep in the forest, his great head resting on his paws. A timid little Mouse came upon him unexpectedly, and in her fright and haste to get away, ran across the Lion's nose. Roused from his nap, the Lion laid his huge paw angrily on the tiny creature to kill her. "Spare me!" begged the poor Mouse. "Please let me go and some day I will surely repay you." The Lion was much amused to think that a Mouse could ever help him. But he was generous and finally let the Mouse go. Some days later, while stalking his prey in the forest, the Lion was caught in the toils of a hunter's net. Unable to free himself, he filled the forest with his angry roaring. The Mouse knew the voice and quickly found the Lion struggling in the net. Running to one of the great ropes that bound him, she gnawed it until it parted, and soon the Lion was free. "You laughed when I said I would repay you," said the Mouse. "Now you see that even a Mouse can help a Lion."

Question Generation in NLP – Case Study



■ Pipeline Steps – Coding Demo:

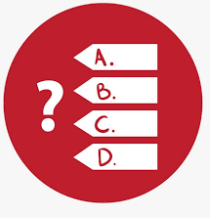
Original Text >>

A Lion lay asleep in the forest, his great head resting on his paws. A timid little Mouse came upon him unexpectedly, and in her fright and haste to get away, ran across the Lion's nose. Roused from his nap, the Lion laid his huge paw angrily on the tiny creature to kill her. "Spare me!" begged the poor Mouse. "Please let me go and some day I will surely repay you." The Lion was much amused to think that a Mouse could ever help him. But he was generous and finally let the Mouse go. Some days later, while stalking his prey in the forest, the Lion was caught in the toils of a hunter's net. Unable to free himself, he filled the forest with his angry roaring. The Mouse knew the voice and quickly found the Lion struggling in the net. Running to one of the great ropes that bound him, she gnawed it until it parted, and soon the Lion was free. "You laughed when I said I would repay you," said the Mouse. "Now you see that even a Mouse can help a Lion."

Summarized Text >>

A lion lay asleep in the forest, his great head resting on his paws. The timid little mouse ran across the lion's nose and begged him to let him go. "please let me go and some day i will surely repay you," said the mouse. He was generous and finally let the mouse go; some days later, while stalking his prey, the lion was caught in an angry net.

Question Generation in NLP – Case Study



■ Pipeline Steps – Coding Demo:

Original Text >>

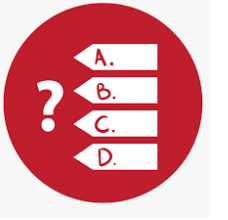
A Lion lay asleep in the forest, his great head resting on his paws. A timid little Mouse came upon him unexpectedly, and in her fright and haste to get away, ran across the Lion's nose. Roused from his nap, the Lion laid his huge paw angrily on the tiny creature to kill her. "Spare me!" begged the poor Mouse. "Please let me go and some day I will surely repay you." The Lion was much amused to think that a Mouse could ever help him. But he was generous and finally let the Mouse go. Some days later, while stalking his prey in the forest, the Lion was caught in the toils of a hunter's net. Unable to free himself, he filled the forest with his angry roaring. The Mouse knew the voice and quickly found the Lion struggling in the net. Running to one of the great ropes that bound him, she gnawed it until it parted, and soon the Lion was free. "You laughed when I said I would repay you," said the Mouse. "Now you see that even a Mouse can help a Lion."

Summarized Text >>

A lion lay asleep in the forest, his great head resting on his paws. The timid little mouse ran across the lion's nose and begged him to let him go. "please let me go and some day i will surely repay you," said the mouse. He was generous and finally let the mouse go; some days later, while stalking his prey, the lion was caught in an angry net.

Keywords: ['lion', 'mouse', 'forest', 'net', 'paws', 'day', 'hunter', 'toils', 'roaring', 'voice', 'prey', 'nose', 'nap', 'head', 'creature']

Question Generation in NLP – Case Study



■ Pipeline Steps – Coding Demo:

```
What animal lay asleep in the forest?  
Lion
```

```
What animal ran across the Lion's nose?  
Mouse
```

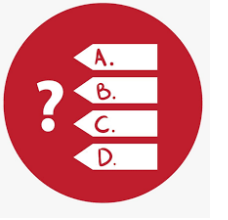
```
Where did a Lion lay asleep?  
Forest
```

```
What was the Lion caught in while stalking his prey?  
Net
```

```
Where was the Lion's great head resting?  
Paws
```

```
Keywords: ['lion', 'mouse', 'forest', 'net', 'paws', 'day', 'hunter', 'toils', 'roaring', 'voice', 'prey', 'nose', 'nap', 'head', 'creature']
```

Question Generation in NLP – Case Study



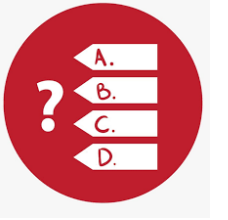
■ Pipeline Steps – Coding Demo:

```
What animal lay asleep in the forest?  
Lion
```

```
Original word:  Lion  
['Cheetah', 'Jaguar', 'Leopard', 'Liger', 'Saber-toothed Tiger', 'Snow Leopard', 'Tiger', 'Tiglon']
```

```
Keywords: ['lion', 'mouse', 'forest', 'net', 'paws', 'day', 'hunter', 'toils', 'roaring', 'voice', 'prey', 'nose', 'nap', 'head', 'creature']
```

Question Generation in NLP – Case Study



■ Pipeline Steps – Coding Demo:

```
Synset('cricket.n.01') : leaping insect; male makes chirping noises by rubbing the forewings together
```

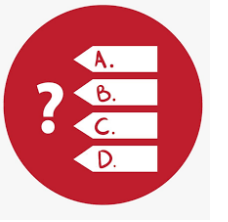
```
Synset('cricket.n.02') : a game played with a ball and bat by two teams of 11 players; teams take turns trying to score runs
```

```
original word: Cricket  
['Grasshopper']
```

```
original word: Cricket  
['Ball Game', 'Field Hockey', 'Football', 'Hurling', 'Lacrosse', 'Polo', 'Pushball', 'Ultimate Frisbee']
```

```
Keywords: ['lion', 'mouse', 'forest', 'net', 'paws', 'day', 'hunter', 'toils', 'roaring', 'voice', 'prey', 'nose', 'nap', 'head', 'creature']
```

Question Generation in NLP – Case Study



■ Pipeline Steps – Coding Demo:

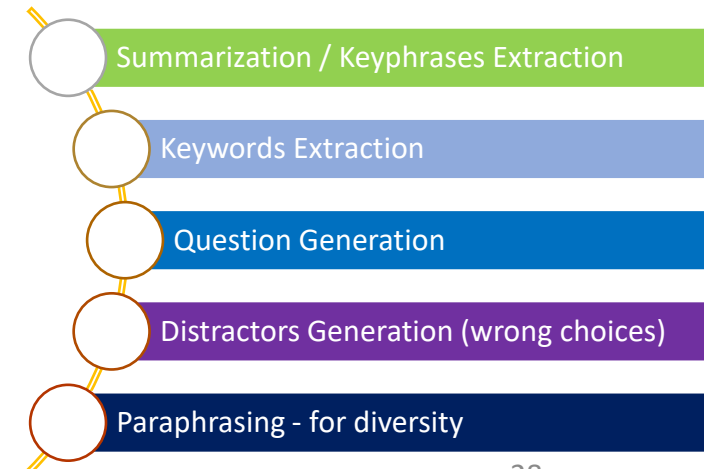
```
✓ [36] # [Question, Answer, Distractors List]
6 s

gen_questions = []

for answer in imp_keywords:
    ques = get_question(text, answer, question_model, question_tokenizer)
    distractors_calculated = get_distractors_wordnet(answer, 0)
    gen_questions.append([ques, answer.capitalize(), distractors_calculated])
```

```
✓ 0 s ▶ gen_questions[0]

➡ [' What animal lay asleep in the forest?',
   'Lion',
   ['Cheetah',
    'Jaguar',
    'Leopard',
    'Liger',
    'Saber-toothed Tiger',
    'Snow Leopard',
    'Tiger',
    'Tiglon']]
```



Thank You



DO DIGITAL. STAY HUMAN.

github.com/GitTeaching/Language_Detection_ML_API/blob/main/MCQ_QuestGen.ipynb



SCAN ME



References

- Mulla N, Gharpure P. Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications. Prog Artif Intell. 2023;12(1):1–32. doi: 10.1007/s13748-023-00295-9. Epub 2023 Jan 30. PMID: PMC9886210.
- Ramsri Goutham Golla, Question Generation using Natural Language Processing.
- Transformers, T5 models. https://huggingface.co/docs/transformers/model_doc/t5