

```
In [ ]: f = open("text1.txt", "r")
content = f.read()
print(content)
path_online_source = "summarized_from_online_source.txt"
path_model_source = "summary_from_model.txt"
text_path = "text1.txt"
```

Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to intelligence of humans and other animals. Example tasks in which this is done include speech recognition, computer vision, translation between (natural) languages, as well as other mappings of inputs.

AI applications include advanced web search engines (e.g., Google Search), recommendation systems (used by YouTube, Amazon, and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Waymo), generative or creative tools (ChatGPT and AI art), automated decision-making, and competing at the highest level in strategic game systems (such as chess and Go).

As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI effect. For instance, optical character recognition is frequently excluded from things considered to be AI, having become a routine technology.

Artificial intelligence was founded as an academic discipline in 1956, and in the years since it has experienced several waves of optimism, followed by disappointment and the loss of funding (known as an "AI winter"), followed by new approaches, success, and renewed funding. AI research has tried and discarded many different approaches, including simulating the brain, modeling human problem solving, formal logic, large databases of knowledge, and imitating animal behavior. In the first decades of the 21st century, highly mathematical and statistical machine learning has dominated the field, and this technique has proved highly successful, helping to solve many challenging problems throughout industry and academia.

The various sub-fields of AI research are centered around particular goals and the use of particular tools. The traditional goals of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception, and the ability to move and manipulate objects. General intelligence (the ability to solve an arbitrary problem) is among the field's long-term goals. To solve these problems, AI researchers have adapted and integrated a wide range of problem-solving techniques, including search and mathematical optimization, formal logic, artificial neural networks, and methods based on statistics, probability, and economics. AI also draws upon computer science, psychology, linguistics, philosophy, and many other fields.

The field was founded on the assumption that human intelligence "can be so precisely described that a machine can be made to simulate it". This raised philosophical arguments about the mind and the ethical consequences of creating artificial beings endowed with human-like intelligence; these issues have previously been explored by myth, fiction, and philosophy since antiquity. Computer scientists and philosophers have since suggested that AI may become an existential risk to humanity if its rational capacities are not steered towards beneficial goals. The term artificial intelligence has also been criticized for overhyping AI's true technological capabilities.

```
In [ ]: import nltk
nltk.download('stopwords')
nltk.download('punkt')
```

```
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize, sent_tokenize
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\Sahsa\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\Sahsa\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

```
In [ ]: # Tokenizing the text

stopWords = set(stopwords.words("english"))
words = word_tokenize(content)
print(stopWords)
print()
print(words)
```

{'can', 'yourselves', 'him', 'wasn', 'ma', 'below', 'not', "didn't", 'them', 'mightn', 'more', 'only', 'any', 'few', "you'll", 'being', 'ourselves', 'all', 'weren', 'from', 'there', "won't", "needn't", 're', 'haven', 'between', 'myself', 'out', 'with', "hasn't", 'while', 'up', "isn't", "doesn't", "mustn't", 'an', 'aren', 'very', 'before', 'hadn', 'down', 'why', 'mustn', 'had', 'on', 'theirs', 'because', 'against', 'once', 'shouldn', 'herself', 'where', 'who', 'off', 'if', 'after', "should've", 'what', "hadn't", "you'd", 'other', 'm', 'over', 'each', 'and', 'how', 'so', "it's", 'just', 'of', 'for', 'their', 'as', 'such', 'same', 'he', 'itself', 'than', 'during', "she's", 'didn', 'won', 'under', 'needn', "shouldn't", 'ain', 'those', 'they', "shan't", 'whom', 'some', "don't", 'now', 'my', 'it', 'will', 's', 'did', 'does', 'll', "you've", 'too', 'is', 'her', 'then', 'shan', 'until', 'ours', 'both', 'into', 'above', 'your', "that'll", 'o', 'are', "wouldn't", "couldn't", 'y', 'the', 'in', "aren't", 'doing', 'yours', 'do', 'its', 'themselves', 'been', 'd', 'was', 'himself', 'having', 'don', 'should', 'hasn', "haven't", 'isn', 'which', 'our', 'were', "mightn't", 'here', 'that', 'own', "wasn't", 'his', 'further', 'nor', 'couldn', 'about', 'but', 'doesn', 'hers', 'she', 'be', 'to', 'when', 'or', 'these', 'wouldn', 'this', "weren't", 'have', 'has', 'we', 'at', 'i', 'you', 'through', 'again', 'most', 'yourself', 've', 't', "you're", 'me', 'no', 'by', 'am', 'a'}

['Artificial', 'intelligence', '(', 'AI', ')', 'is', 'intelligence', 'demonstrated', 'by', 'machines', ',', 'as', 'opposed', 'to', 'intelligence', 'of', 'humans', 'and', 'other', 'animals', '.', 'Example', 'tasks', 'in', 'which', 'this', 'is', 'done', 'include', 'speech', 'recognition', ',', 'computer', 'vision', ',', 'translation', 'between', '(', 'natural', ')', 'languages', ',', 'as', 'well', 'as', 'other', 'mappings', 'of', 'inputs', '.', 'AI', 'applications', 'include', 'advanced', 'web', 'search', 'engines', '(', 'e.g.', ',', 'Google', 'Search', ')', ',', 'recommendation', 'systems', '(', 'used', 'by', 'YouTube', ',', 'Amazon', ',', 'and', 'Netflix', ')', ',', 'understanding', 'human', 'speech', '(', 'such', 'as', 'Siri', 'and', 'Alexa', ')', ',', 'self-driving', 'cars', '(', 'e.g.', ',', 'Waymo', ')', ',', 'generative', 'or', 'creative', 'tools', '(', 'ChatGPT', 'and', 'AI', 'art', ')', ',', 'automated', 'decision-making', ',', 'and', 'competing', 'at', 'the', 'highest', 'level', 'in', 'strategic', 'game', 'systems', '(', 'such', 'as', 'chess', 'and', 'Go', ')', '.', 'As', 'machines', 'become', 'increasingly', 'capable', ',', 'tasks', 'considered', 'to', 'require', 'intelligence', 'are', 'often', 'removed', 'from', 'the', 'definition', 'of', 'AI', ',', 'a', 'phenomenon', 'known', 'as', 'the', 'AI', 'effect', '.', 'For', 'instance', ',', 'optical', 'character', 'recognition', 'is', 'frequently', 'excluded', 'from', 'things', 'considered', 'to', 'be', 'AI', ',', 'having', 'become', 'a', 'routine', 'technology', '.', 'Artificial', 'intelligence', 'was', 'founded', 'as', 'an', 'academic', 'discipline', 'in', '1956', ',', 'and', 'in', 'the', 'years', 'since', 'it', 'has', 'experienced', 'several', 'waves', 'of', 'optimism', ',', 'followed', 'by', 'disappointment', 'and', 'the', 'loss', 'of', 'funding', '(', 'known', 'as', 'an', 'AI', 'winter', ')', ',', 'followed', 'by', 'new', 'approaches', ',', 'success', ',', 'and', 'renewed', 'funding', '.', 'AI', 'research', 'has', 'tried', 'and', 'discarded', 'many', 'different', 'approaches', ',', 'including', 'simulating', 'the', 'brain', ',', 'modeling', 'human', 'problem', 'solving', ',', 'formal', 'logic', ',', 'large', 'databases', 'of', 'knowledge', ',', 'and', 'imitating', 'animal', 'behavior', '.', 'In', 'the', 'first', 'decades', 'of', 'the', '21st', 'century', ',', 'highly', 'mathematical', 'and', 'statistical', 'machine', 'learning', 'has', 'dominated', 'the', 'field', ',', 'and', 'this', 'technique', 'has', 'proved', 'highly', 'successful', ',', 'helping', 'to', 'solve', 'many', 'challenging', 'problems', 'throughout', 'industry', 'and', 'academia', '.', 'The', 'various', 'sub-fields', 'of', 'AI', 'research', 'are', 'centered', 'around', 'particular', 'goals', 'and', 'the', 'use', 'of', 'particular', 'tools', '.', 'The', 'traditional', 'goals', 'of', 'AI', 'research', 'include', 'reasoning', ',', 'knowledge

e', 'representation', ',', 'planning', ',', 'learning', ',', 'natural', 'language', 'processing', ',', 'perception', ',', 'and', 'the', 'ability', 'to', 'move', 'and', 'manipulate', 'objects', '.', 'General', 'intelligence', '(', 'the', 'ability', 'to', 'solve', 'an', 'arbitrary', 'problem', ')', 'is', 'among', 'the', 'field', '"', 's', 'long-term', 'goals', '.', 'To', 'solve', 'these', 'problems', ',', 'AI', 'researchers', 'have', 'adapted', 'and', 'integrated', 'a', 'wide', 'range', 'of', 'problem-solving', 'techniques', ',', 'including', 'search', 'and', 'mathematical', 'optimization', ',', 'formal', 'logic', ',', 'artificial', 'neural', 'networks', ',', 'and', 'methods', 'based', 'on', 'statistics', ',', 'probability', ',', 'and', 'economics', '.', 'AI', 'also', 'draws', 'upon', 'computer', 'science', ',', 'psychology', ',', 'linguistics', ',', 'philosophy', ',', 'and', 'many', 'other', 'fields', '.', 'The', 'field', 'was', 'founded', 'on', 'the', 'assumption', 'that', 'human', 'intelligence', '``', 'can', 'be', 'so', 'precisely', 'described', 'that', 'a', 'machine', 'can', 'be', 'made', 'to', 'simulate', 'it', '"', '., 'This', 'raised', 'philosophical', 'arguments', 'about', 'the', 'mind', 'and', 'the', 'ethical', 'consequences', 'of', 'creating', 'artificial', 'beings', 'endowed', 'with', 'human-like', 'intelligence', ';', 'these', 'issues', 'have', 'previously', 'been', 'explored', 'by', 'myth', ',', 'fiction', ',', 'and', 'philosophy', 'since', 'antiquity', '.', 'Computer', 'scientists', 'and', 'philosophers', 'have', 'since', 'suggested', 'that', 'AI', 'may', 'become', 'an', 'existential', 'risk', 'to', 'humanity', 'if', 'its', 'rational', 'capacities', 'are', 'not', 'steered', 'towards', 'beneficial', 'goals', '.', 'The', 'term', 'artificial', 'intelligence', 'has', 'also', 'been', 'criticized', 'for', 'overhyping', 'AI', '"', 's', 'true', 'technological', 'capabilities', '']

```
In [ ]: # Creating a frequency table to keep the
        # score of each word
```

```
freqTable = dict()
for word in words:
    word = word.lower()
    if word in stopWords:
        continue
    if word in freqTable:
        freqTable[word] += 1
    else:
        freqTable[word] = 1
```

```
In [ ]: print(freqTable)
```

```
{'artificial': 5, 'intelligence': 9, '(': 10, 'ai': 14, ')': 10, 'demonstrated': 1, 'machines': 2, ',': 50, 'opposed': 1, 'humans': 1, 'animals': 1, '.': 17, 'example': 1, 'tasks': 2, 'done': 1, 'include': 3, 'speech': 2, 'recognition': 2, 'computer': 3, 'vision': 1, 'translation': 1, 'natural': 2, 'languages': 1, 'well': 1, 'mappings': 1, 'inputs': 1, 'applications': 1, 'advanced': 1, 'web': 1, 'search': 3, 'engines': 1, 'e.g.': 2, 'google': 1, 'recommendation': 1, 'systems': 2, 'used': 1, 'youtube': 1, 'amazon': 1, 'netflix': 1, 'understanding': 1, 'human': 3, 'siri': 1, 'alexa': 1, 'self-driving': 1, 'cars': 1, 'waymo': 1, 'generative': 1, 'creative': 1, 'tools': 2, 'chatgpt': 1, 'art': 1, 'automated': 1, 'decision-making': 1, 'competing': 1, 'highest': 1, 'level': 1, 'strategic': 1, 'game': 1, 'chess': 1, 'go': 1, 'become': 3, 'increasingly': 1, 'capable': 1, 'considered': 2, 'require': 1, '``': 3, '""': 3, 'often': 1, 'removed': 1, 'definition': 1, 'phenomenon': 1, 'known': 2, 'effect': 1, 'instance': 1, 'optical': 1, 'character': 1, 'frequently': 1, 'excluded': 1, 'things': 1, 'routine': 1, 'technology': 1, 'founded': 2, 'academic': 1, 'discipline': 1, '1956': 1, 'years': 1, 'since': 3, 'experienced': 1, 'several': 1, 'waves': 1, 'optimism': 1, 'followed': 2, 'disappointment': 1, 'loss': 1, 'funding': 2, 'winter': 1, 'new': 1, 'approaches': 2, 'success': 1, 'renewed': 1, 'research': 3, 'tried': 1, 'discarded': 1, 'many': 3, 'different': 1, 'including': 2, 'simulating': 1, 'brain': 1, 'modeling': 1, 'problem': 2, 'solving': 1, 'formal': 2, 'logic': 2, 'large': 1, 'databases': 1, 'knowledge': 2, 'imitating': 1, 'animal': 1, 'behavior': 1, 'first': 1, 'decades': 1, '21st': 1, 'century': 1, 'highly': 2, 'mathematical': 2, 'statistical': 1, 'machine': 2, 'learning': 2, 'dominated': 1, 'field': 3, 'technique': 1, 'proved': 1, 'successful': 1, 'helping': 1, 'solve': 3, 'challenging': 1, 'problems': 2, 'throughout': 1, 'industry': 1, 'academia': 1, 'various': 1, 'sub-fields': 1, 'centered': 1, 'around': 1, 'particular': 2, 'goals': 4, 'use': 1, 'traditional': 1, 'reasoning': 1, 'representation': 1, 'planning': 1, 'language': 1, 'processing': 1, 'perception': 1, 'ability': 2, 'move': 1, 'manipulate': 1, 'objects': 1, 'general': 1, 'arbitrary': 1, 'among': 1, "'s": 2, 'long-term': 1, 'researchers': 1, 'adapted': 1, 'integrated': 1, 'wide': 1, 'range': 1, 'problem-solving': 1, 'techniques': 1, 'optimization': 1, 'neural': 1, 'networks': 1, 'methods': 1, 'based': 1, 'statistics': 1, 'probability': 1, 'economics': 1, 'also': 2, 'draws': 1, 'upon': 1, 'science': 1, 'psychology': 1, 'linguistics': 1, 'philosophy': 2, 'fields': 1, 'assumption': 1, 'precisely': 1, 'described': 1, 'made': 1, 'simulate': 1, 'raised': 1, 'philosophical': 1, 'arguments': 1, 'mind': 1, 'ethical': 1, 'consequences': 1, 'creating': 1, 'being': 1, 'endowed': 1, 'human-like': 1, ';': 1, 'issues': 1, 'previously': 1, 'explored': 1, 'myth': 1, 'fiction': 1, 'antiquity': 1, 'scientists': 1, 'philosophers': 1, 'suggested': 1, 'may': 1, 'existential': 1, 'risk': 1, 'humanity': 1, 'rational': 1, 'capacities': 1, 'steered': 1, 'towards': 1, 'beneficial': 1, 'term': 1, 'criticized': 1, 'overhyping': 1, 'true': 1, 'technological': 1, 'capabilities': 1}
```

```
In [ ]: # Creating a dictionary to keep the score of each sentence
sentences = sent_tokenize(content)
sentenceValue = dict()
```

```
for sentence in sentences:
    for word, freq in freqTable.items():
        if word in sentence.lower():
            if sentence in sentenceValue:
                sentenceValue[sentence] += freq
            else:
                sentenceValue[sentence] = freq
```

```
In [ ]: print(sentenceValue)
```

{'Artificial intelligence (AI) is intelligence demonstrated by machines, as oppose d to intelligence of humans and other animals.': 128, 'Example tasks in which this is done include speech recognition, computer vision, translation between (natural) languages, as well as other mappings of inputs.': 110, 'AI applications include ad vanced web search engines (e.g., Google Search), recommendation systems (used by Y ouTube, Amazon, and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Waymo), generative or creative tools (ChatGPT and AI ar t), automated decision-making, and competing at the highest level in strategic gam e systems (such as chess and Go).': 148, 'As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition o f AI, a phenomenon known as the AI effect.': 112, 'For instance, optical character recognition is frequently excluded from things considered to be AI, having become a routine technology.': 96, 'Artificial intelligence was founded as an academic di scipline in 1956, and in the years since it has experienced several waves of optim ism, followed by disappointment and the loss of funding (known as an "AI winter"), followed by new approaches, success, and renewed funding.': 143, 'AI research has tried and discarded many different approaches, including simulating the brain, mod eling human problem solving, formal logic, large databases of knowledge, and imita ting animal behavior.': 117, 'In the first decades of the 21st century, highly mat hematical and statistical machine learning has dominated the field, and this techn ique has proved highly successful, helping to solve many challenging problems thro ughout industry and academia.': 103, 'The various sub-fields of AI research are ce ntered around particular goals and the use of particular tools.': 56, 'The traditi onal goals of AI research include reasoning, knowledge representation, planning, l earning, natural language processing, perception, and the ability to move and mani pulate objects.': 113, "General intelligence (the ability to solve an arbitrary pr oblem) is among the field's long-term goals.": 68, 'To solve these problems, AI re searchers have adapted and integrated a wide range of problem-solving techniques, including search and mathematical optimization, formal logic, artificial neural ne tworks, and methods based on statistics, probability, and economics.': 127, 'AI al so draws upon computer science, psychology, linguistics, philosophy, and many othe r fields.': 100, 'The field was founded on the assumption that human intelligence "can be so precisely described that a machine can be made to simulate it".': 41, 'This raised philosophical arguments about the mind and the ethical consequences o f creating artificial beings endowed with human-like intelligence; these issues ha ve previously been explored by myth, fiction, and philosophy since antiquity.': 12 1, 'Computer scientists and philosophers have since suggested that AI may become a n existential risk to humanity if its rational capacities are not steered towards beneficial goals.': 60, "The term artificial intelligence has also been criticized for overhyping AI's true technological capabilities.": 58}

```
In [ ]: sumValues = 0
        for sentence in sentenceValue:
            sumValues += sentenceValue[sentence]
```

```
In [ ]: # Average value of a sentence from the original text

average = int(sumValues / len(sentenceValue))

# Storing sentences into our summary.
summary = ''
for sentence in sentences:
    if (sentence in sentenceValue) and (sentenceValue[sentence] > (1.2 * average)):
```

```
summary += " " + sentence
print(summary)
```

Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to intelligence of humans and other animals. AI applications include advanced web search engines (e.g., Google Search), recommendation systems (used by YouTube, Amazon, and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Waymo), generative or creative tools (ChatGPT and AI art), automated decision-making, and competing at the highest level in strategic game systems (such as chess and Go). Artificial intelligence was founded as an academic discipline in 1956, and in the years since it has experienced several waves of optimism, followed by disappointment and the loss of funding (known as an "AI winter"), followed by new approaches, success, and renewed funding. To solve these problems, AI researchers have adapted and integrated a wide range of problem-solving techniques, including search and mathematical optimization, formal logic, artificial neural networks, and methods based on statistics, probability, and economics. This raised philosophical arguments about the mind and the ethical consequences of creating artificial beings endowed with human-like intelligence; these issues have previously been explored by myth, fiction, and philosophy since antiquity.

```
In [ ]: len(summary)
```

```
Out[ ]: 1294
```

```
In [ ]: import spacy
```

```
In [ ]: nlp = spacy.load("en_core_web_sm")
```

```
In [ ]: doc = nlp(summary)
```

```
In [ ]: print(doc)
```

Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to intelligence of humans and other animals. AI applications include advanced web search engines (e.g., Google Search), recommendation systems (used by YouTube, Amazon, and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Waymo), generative or creative tools (ChatGPT and AI art), automated decision-making, and competing at the highest level in strategic game systems (such as chess and Go). Artificial intelligence was founded as an academic discipline in 1956, and in the years since it has experienced several waves of optimism, followed by disappointment and the loss of funding (known as an "AI winter"), followed by new approaches, success, and renewed funding. To solve these problems, AI researchers have adapted and integrated a wide range of problem-solving techniques, including search and mathematical optimization, formal logic, artificial neural networks, and methods based on statistics, probability, and economics. This raised philosophical arguments about the mind and the ethical consequences of creating artificial beings endowed with human-like intelligence; these issues have previously been explored by myth, fiction, and philosophy since antiquity.

```
In [ ]: summary_length = len(doc)
print(summary_length)
```

```
239
```

```
In [ ]: #reading summary obtained from online sources
with open (path_online_source, "r") as f:
```

```
data = f.read()
```

```
In [ ]: # summarized = nlp(data)
```

```
In [ ]: print(data)
```

AI is the intelligence demonstrated by machines, as opposed to humans and other animals. Examples of AI tasks include speech recognition, computer vision, translation between languages, and other mappings of inputs. AI applications include advanced web search engines, recommendation systems, understanding human speech, self-driving cars, generative or creative tools, automated decision-making, and competing at the highest level in strategic game systems. AI was founded in 1956 and has experienced several waves of optimism, followed by disappointment and the loss of funding, followed by new approaches, success, and renewed funding. AI research has tried and discarded many different approaches, including simulating the brain, modeling human problem solving, formal logic, large databases of knowledge, and imitating animal behavior.

AI research has been dominated by highly mathematical and statistical machine learning in the first decades of the 21st century. It has been successful in solving many challenging problems throughout industry and academia. AI researchers have adapted and integrated a wide range of problem-solving techniques, including search and mathematical optimization, formal logic, artificial neural networks, and methods based on statistics, probability, and economics. AI also draws upon computer science, psychology, linguistics, philosophy, and many other fields. The field was founded on the assumption that human intelligence can be so precisely described that a machine can be made to simulate it, raising philosophical arguments about the mind and the ethical consequences of creating artificial beings endowed with human-like intelligence.

Computer scientists and philosophers have suggested that AI may become an existential risk to humanity if its rational capacities are not steered towards beneficial goals. AI has also been criticized for overhyping its true technological capabilities.

```
In [ ]: f.close()
```

```
In [ ]: summarized = nlp(data)
```

```
In [ ]: print(summarized)
```


AI is the intelligence demonstrated by machines, as opposed to humans and other animals. Examples of AI tasks include speech recognition, computer vision, translation between languages, and other mappings of inputs. AI applications include advanced web search engines, recommendation systems, understanding human speech, self-driving cars, generative or creative tools, automated decision-making, and competing at the highest level in strategic game systems. AI was founded in 1956 and has experienced several waves of optimism, followed by disappointment and the loss of funding, followed by new approaches, success, and renewed funding. AI research has tried and discarded many different approaches, including simulating the brain, modeling human problem solving, formal logic, large databases of knowledge, and imitating animal behavior.

AI research has been dominated by highly mathematical and statistical machine learning in the first decades of the 21st century. It has been successful in solving many challenging problems throughout industry and academia. AI researchers have adapted and integrated a wide range of problem-solving techniques, including search and mathematical optimization, formal logic, artificial neural networks, and methods based on statistics, probability, and economics. AI also draws upon computer science, psychology, linguistics, philosophy, and many other fields. The field was founded on the assumption that human intelligence can be so precisely described that a machine can be made to simulate it, raising philosophical arguments about the mind and the ethical consequences of creating artificial beings endowed with human-like intelligence.

Computer scientists and philosophers have suggested that AI may become an existential risk to humanity if its rational capacities are not steered towards beneficial goals. AI has also been criticized for overhyping its true technological capabilities.

```
In [ ]: online_length = len(summarized)
        print(online_length)
```

322

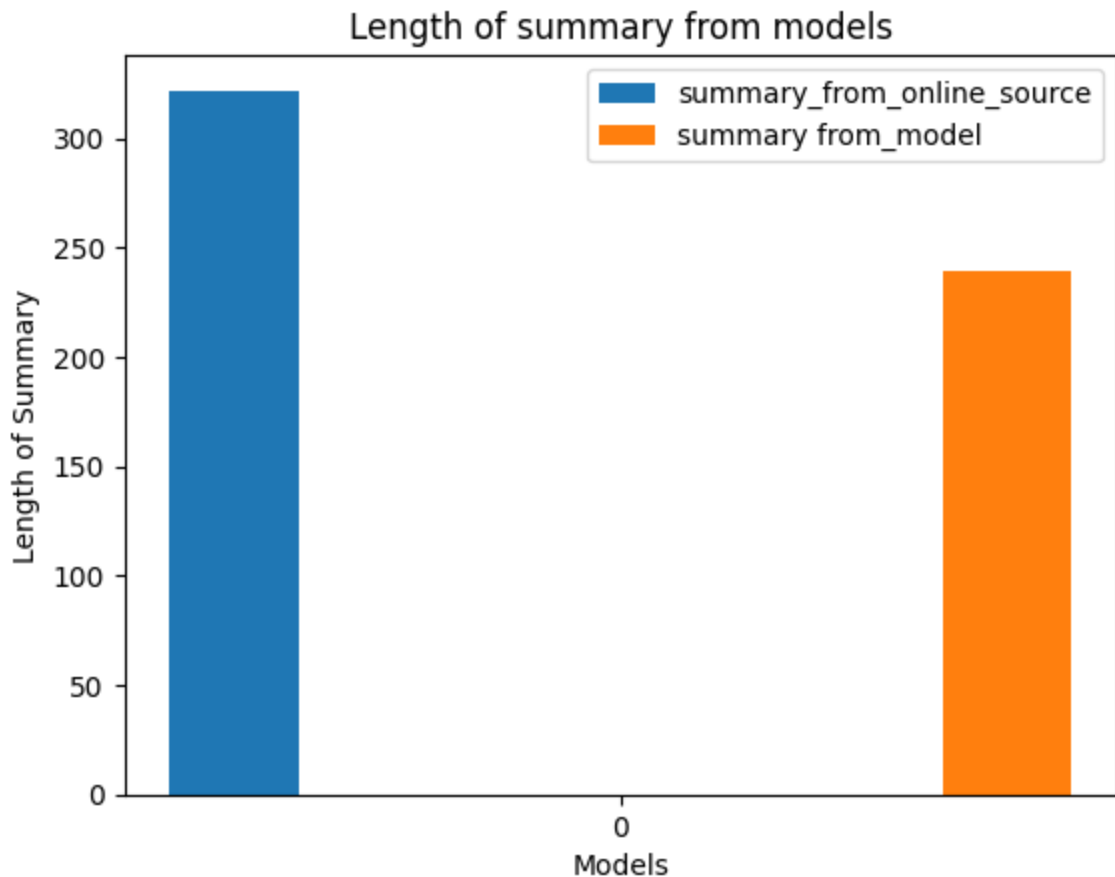
```
In [ ]: import numpy as np
        import matplotlib.pyplot as plt
```

```
In [ ]: x =[online_length]
```

```
In [ ]: y = [summary_length]
```

```
In [ ]: X_axis = np.arange(1)
```

```
In [ ]: plt.bar(X_axis - 0.3, x, 0.1, label = 'summary_from_online_source')
        plt.bar(X_axis + 0.3, y, 0.1, label = 'summary from_model')
        plt.xticks(X_axis)
        plt.xlabel("Models")
        plt.ylabel("Length of Summary")
        plt.title("Length of summary from models")
        plt.legend()
        plt.show()
```



```
In [ ]: import math
import string
import sys
```

```
In [ ]: # reading the text file
# This function will return a List of the lines of text in the file.
def read_file(filename):

    try:
        with open(filename, 'r') as f:
            data = f.read()
        return data

    except IOError:
        print("Error opening or reading input file: ", filename)
        sys.exit()
```

```
In [ ]: # splitting the text lines into words
# translation table is a global variable mapping upper case to lower case and punct
translation_table = str.maketrans(string.punctuation+string.ascii_uppercase,
                                   " "*len(string.punctuation)+string.ascii_lower
```

```
In [ ]: # returns a List of the words in the file
def get_words_from_line_list(text):

    text = text.translate(translation_table)
```

```
word_list = text.split()
```

```
return word_list
```

```
In [ ]: # counts frequency of each word returns a dictionary which maps the words to their  
def count_frequency(word_list):
```

```
    D = {}
```

```
    for new_word in word_list:
```

```
        if new_word in D:
```

```
            D[new_word] = D[new_word] + 1
```

```
        else:
```

```
            D[new_word] = 1
```

```
    return D
```

```
In [ ]: # returns dictionary of (word, frequency) pairs from the previous dictionary.  
def word_frequencies_for_file(filename):
```

```
    line_list = read_file(filename)
```

```
    word_list = get_words_from_line_list(line_list)
```

```
    freq_mapping = count_frequency(word_list)
```

```
    print("File", filename, ":", )
```

```
    print(len(line_list), "lines, ", )
```

```
    print(len(word_list), "words, ", )
```

```
    print(len(freq_mapping), "distinct words")
```

```
    return freq_mapping
```

```
In [ ]: # returns the dot product of two documents
```

```
def dotProduct(D1, D2):
```

```
    Sum = 0.0
```

```
    for key in D1:
```

```
        if key in D2:
```

```
            Sum += (D1[key] * D2[key])
```

```
    return Sum
```

```
In [ ]: # returns the angle in radians between document vectors
```

```
def vector_angle(D1, D2):
```

```
    numerator = dotProduct(D1, D2)
```

```
    denominator = math.sqrt(dotProduct(D1, D1)*dotProduct(D2, D2))
```

```
    return math.acos(numerator / denominator)
```

```
In [ ]: def documentSimilarity(filename_1, filename_2):
```

```
    sorted_word_list_1 = word_frequencies_for_file(filename_1)
```

```
sorted_word_list_2 = word_frequencies_for_file(filename_2)
distance = vector_angle(sorted_word_list_1, sorted_word_list_2)

print("The distance between the documents is: % 0.6f (radians)"% distance)

return [len(sorted_word_list_1),len(sorted_word_list_2)]
```

```
In [ ]: with open (path_model_source, "a") as f:
        f.write(summary)
```

```
In [ ]: f.close()
```

```
In [ ]: word = documentSimilarity(path_model_source, path_online_source)
```

```
File summary_from_model.txt :
1294 lines,
187 words,
130 distinct words
File summarized_from_online_source.txt :
1931 lines,
274 words,
179 distinct words
The distance between the documents is: 0.605326 (radians)
```

```
In [ ]: plt.bar(X_axis - 0.4, word[0], 0.2, label = 'summary_from_model')
plt.bar(X_axis + 0.4, word[1], 0.2, label = 'summary from_online_source')
plt.xticks(X_axis)
plt.xlabel("Models")
plt.ylabel("Unique Words")
plt.title("Distinct words in summary")
plt.legend()
plt.show()
```

