```python
import docx
import string
import nltk
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.decomposition import NMF
from sklearn.feature_extraction.text import CountVectorizer

# Replace the following line with the full path to your DOCX file
file_path = "C:\\Users\\JASBIR\\Downloads\\Data.docx"

# Load the DOCX file
doc = docx.Document(file_path)

# Initialize an empty string to store the extracted text
extracted_text = ""

# Iterate through paragraphs and extract text
for paragraph in doc.paragraphs:
    extracted_text += paragraph.text + "\n"

# Print the extracted text
print(extracted_text)
```

Discussions about AI abound. But critics have begun to recognize that these debates have focused mostly on technical issues. That is, there are certain problems that need to be addressed and have technical solutions. Debate rages, accordingly, about how to resolve these issues and move the applications of AI forward. Perhaps Terry Winograd (1996) was correct when he lamented some time ago that interest in philosophy, what he calls high theory, has dissipated.

The application of AI, nonetheless, has pushed the discussion beyond technical devices and their possible uses. For example, critics have begun to recognize that AI can be quite alienating (Dreyfus 1992; Ritzer 1993).   Workers at Amazon have provided an interesting case study. While blaming AI, they claim to be overworked and do not stayed employed for long. They complain regularly about manipulation, stress, job insecurity, and so on. Clearly, AI is not viewed to be their friend (Livingstone 2018; Vicent 2019).

On the other hand, AI can be quite dangerous. Take driverless cars! In this application, through the use of AI cars can learn how to navigate streets, other cars, pedestrians, and occasionally unanticipated obstacles. Any failure can result in a catastrophe, even death. Questions about the ability of AI to master truly complex activities—those with fluid or shifting frames—have come to the forefront (Goodfellow, Bengio, and Courville 2016).

And what about ethical issues? The focus of developing algorithms, for example, is not necessarily on job loss or intrusions into privacy. These issues seems to make context difficult to ignore. Persons tend to become especially nervous when their jobs or privacy are threatened. Although alienation and learning involve context, the ability to survive and the quality of life seem to go to the heart of the matter.

The time appears to be ripe, accordingly, to raise the issue of context in the development of advanced technology. Shoshana Zuboff (2019) strives to initiate this sort of discussion with her recent foray into the impact of late capitalism on AI. In this regard, Kate Crawford (2021) strives to move beyond the "nowhere spaces" where she contends most discussion of AI take place. She believes, for example, that AI is enmeshed in the world's ecology. While these entreaties are interesting and relevant, the context provided by Twentieth Century philosophy is missing. Indeed, the pragmatic framework that is currently the focus of attention of these efforts provides some interesting insights into whether AI can learn or deal with knotty social issues

The emphasis of this manuscript, however, is the anti-Cartesian maneuver that characterizes much of contemporary theory. This change has enormous impact on the potential of AI. After all, Cartesianism is at the core of digitalization and modern data processing. Accordingly, the importance of this change in philosophical orientation for understanding the mind, facts, learning, and communication is a vital consideration. Basic to this reassessment is that the limits of AI become obvious in the absence of Cartesianism.

Throughout the Western intellectual tradition, true knowledge has been viewed to be timeless (Grayling 2019). That is, this information is assumed to be divorced from contexts and other human contingencies. If immersed in these situations, knowledge can never surpass opinion and only supply anecdotal evidence. Therefore, most philosophers sought foundations that are universal to establish firm epistemological and moral principles. Martin Heidegger (1969) refers to this trend as the onto-theological tradition.

In many ways, Cartesianism epitomizes this tendency. In fact, the aim of Cartesians is to advance clear and distinct knowledge, severed from opinion and other sources of human error (Bordo 1987). In this regard, these thinkers are not necessarily unique, although their strategy is novel. Rather than speculate about ethereal metaphysical principles, such as Ideas, gods, or cosmic unity, Cartesians make a straightforward proposal known as dualism.

Pre-Processing:

```python
# Remove punctuation from the text
text = extracted_text.translate(str.maketrans('', '', string.punctuation))

# Tokenize the text
text = text.split()

# Remove stop words
stop_words = set(stopwords.words('english'))
text = ' '.join([word for word in text if word.lower() not in stop_words])

# Print the preprocessed text
print(text)
```

Discussions AI abound critics begun recognize debates focused mostly technical issues certain problems need addressed technical solutions Debate rages accordingly resolve issues move applications AI forward Perhaps Terry Winograd 1996 correct lamented time ago interest philosophy calls high theory dissipated application AI nonetheless pushed discussion beyond technical devices possible uses example critics begun recognize AI quite alienating Dreyfus 1992 Ritzer 1993 Workers Amazon provided interesting case study blaming AI claim overworked stayed employed long complain regularly manipulation stress job insecurity Clearly AI viewed friend Livingstone 2018 Vicent 2019 hand AI quite dangerous Take driverless cars application use AI cars learn navigate streets cars pedestrians occasionally unanticipated obstacles failure result catastrophe even death Questions ability AI master truly complex activities—those fluid shifting frames—have come forefront Goodfellow Bengio Courville 2016 ethical issues focus developing algorithms example necessarily job loss intrusions privacy issues seems make context difficult ignore Persons tend become especially nervous jobs privacy threatened Although alienation learning involve context ability survive quality life seem go heart matter time appears ripe accordingly raise issue context development advanced technology Shoshana Zuboff 2019 strives initiate sort discussion recent foray impact late capitalism AI regard Kate Crawford 2021 strives move beyond "nowhere spaces" contends discussion AI take place believes example AI enmeshed world's ecology entreaties interesting relevant context provided Twentieth Century philosophy missing Indeed pragmatic framework currently focus attention efforts provides interesting insights whether AI learn deal knotty social issues emphasis manuscript however antiCartesian maneuver characterizes much contemporary theory change enormous impact potential AI Cartesianism core digitalization modern data processing Accordingly importance change philosophical orientation understanding mind facts learning communication vital consideration Basic reassessment limits AI become obvious absence Cartesianism Throughout Western intellectual tradition true knowledge viewed timeless Grayling 2019 information assumed divorced contexts human contingencies immersed situations knowledge never surpass opinion supply anecdotal evidence Therefore philosophers sought foundations universal establish firm epistemological moral principles Martin Heidegger 1969 refers trend ontotheological tradition many ways Cartesianism epitomizes tendency fact aim Cartesians advance clear distinct knowledge severed opinion sources human error Bordo 1987 regard thinkers necessarily unique although strategy novel Rather speculate ethereal metaphysical principles Ideas gods cosmic unity Cartesians make straightforward proposal known dualism

Step 1: Text Tokenization and Lowercasing

```
In [3]:   # Tokenize the text into words and convert to lowercase
          tokens = extracted_text.split()
          lowercased_tokens = [token.lower() for token in tokens]

          # Display word count
          print("Word Count After Tokenization and Lowercasing:", len(lowercased_tokens))
          print(lowercased_tokens)
```

```
Word Count After Tokenization and Lowercasing: 625
['discussions', 'about', 'ai', 'abound.', 'but', 'critics', 'have', 'begun', 'to', 'recognize', 'that', 'these', 'debates', 'have', 'focused', 'mostly', 'on', 'technical', 'issues.', 'that', 'is,', 'there', 'are', 'certain', 'problems', 'that', 'need', 'to', 'be', 'addressed', 'and', 'have', 'technical', 'solutions.', 'debate', 'rages,', 'accordingly,', 'about', 'how', 'to', 'resolve', 'these', 'issues', 'and', 'move', 'the', 'applications', 'of', 'ai', 'forward.', 'perhaps', 'terry', 'winograd', '(1996)', 'was', 'correct', 'when', 'he', 'lamented', 'some', 'time', 'ago', 'that', 'interest', 'in', 'philosophy,', 'what', 'he', 'calls', 'high', 'theory,', 'has', 'dissipated.', 'the', 'application', 'of', 'ai,', 'nonetheless,', 'has', 'pushed', 'the', 'discussion', 'beyond', 'technical', 'devices', 'and', 'their', 'possible', 'uses.', 'for', 'example,', 'critics', 'have', 'begun', 'to', 'recognize', 'that', 'ai', 'can', 'be', 'quite', 'alienating', '(dreyfus', '1992;', 'ritzer', '1993).', 'workers', 'at', 'amazon', 'have', 'provided', 'an', 'interesting', 'case', 'study.', 'while', 'blaming', 'ai,', 'they', 'claim', 'to', 'be', 'overworked', 'and', 'do', 'not', 'stayed', 'employed', 'for', 'long.', 'they', 'complain', 'regularly', 'about', 'manipulation,', 'stress,', 'job', 'insecurity,', 'and', 'so', 'on.', 'clearly,', 'ai', 'is', 'not', 'viewed', 'to', 'be', 'their', 'friend', '(livingstone', '2018;', 'vicent', '2019).', 'on', 'the', 'other', 'hand,', 'ai', 'can', 'be', 'quite', 'dangerous.', 'take', 'driverless', 'cars!', 'in', 'this', 'application,', 'through', 'the', 'use', 'of', 'ai', 'cars', 'can', 'learn', 'how', 'to', 'navigate', 'streets,', 'other', 'cars,', 'pedestrians,', 'and', 'occasionally', 'unanticipated', 'obstacles.', 'any', 'failure', 'can', 'result', 'in', 'a', 'catastrophe,', 'even', 'death.', 'questions', 'about', 'the', 'ability', 'of', 'ai', 'to', 'master', 'truly', 'complex', 'activities—those', 'with', 'fluid', 'or', 'shifting', 'frames—have', 'come', 'to', 'the', 'forefront', '(goodfellow,', 'bengio', 'and', 'courville', '2016).', 'and', 'what', 'about', 'ethical', 'issues?', 'the', 'focus', 'of', 'developing', 'algorithms,', 'for', 'example,', 'is', 'not', 'necessarily', 'on', 'job', 'loss', 'or', 'intrusions', 'into', 'privacy.', 'these', 'issues', 'seems', 'to', 'make', 'context', 'difficult', 'to', 'ignore.', 'persons', 'tend', 'to', 'become', 'especially', 'nervous', 'when', 'their', 'jobs', 'or', 'privacy', 'are', 'threatened.', 'although', 'alienation', 'and', 'learning', 'involve', 'context', 'the', 'ability', 'to', 'survive', 'and', 'the', 'quality', 'of', 'life', 'seem', 'to', 'go', 'to', 'the', 'heart', 'of', 'the', 'matter.', 'the', 'time', 'appears', 'to', 'be', 'ripe,', 'accordingly,', 'to', 'raise', 'the', 'issue', 'of', 'context', 'in', 'the', 'development', 'of', 'advanced', 'technology.', 'shoshana', 'zuboff', '(2019)', 'strives', 'to', 'initiate', 'this', 'sort', 'of', 'discussion', 'with', 'her', 'recent', 'foray', 'into', 'the', 'impact', 'of', 'late', 'capitalism', 'on', 'ai.', 'in', 'this', 'regard,', 'kate', 'crawford', '(2021)', 'strives', 'to', 'move', 'beyond', 'the', '"nowhere', 'spaces"', 'where', 'she', 'contends', 'most', 'discussion', 'of', 'ai', 'take', 'place.', 'she', 'believes', 'for', 'example,', 'that', 'ai', 'is', 'enmeshed', 'in', 'the', 'world's', 'ecology.', 'while', 'these', 'entreaties', 'are', 'interesting', 'and', 'relevant,', 'the', 'context', 'provided', 'by', 'twentieth', 'century', 'philosophy', 'is', 'missing.', 'indeed,', 'the', 'pragmatic', 'framework', 'that', 'is', 'currently', 'the', 'focus', 'of', 'attention', 'of', 'these', 'efforts', 'provides', 'some', 'interesting', 'insights', 'into', 'whether', 'ai', 'can', 'learn', 'or', 'deal', 'with', 'knotty', 'social', 'issues', 'the', 'emphasis', 'of', 'this', 'manuscript,', 'however,', 'is', 'the', 'anti-cartesian', 'maneuver', 'that', 'characterizes', 'much', 'of', 'contemporary', 'theory.', 'this', 'change', 'has', 'enormous', 'impact', 'on', 'the', 'potential', 'of', 'ai.', 'after', 'all,', 'cartesianism', 'is', 'at', 'the', 'core', 'of', 'digitalization', 'and', 'modern', 'data', 'processing.', 'accordingly,', 'the', 'importance', 'of', 'this', 'change', 'in', 'philosophical', 'orientation', 'for', 'understanding', 'the', 'mind,', 'facts,', 'learning,', 'and', 'communication', 'is', 'a', 'vital', 'consideration.', 'basic', 'to', 'this', 'reassessment', 'is', 'that', 'the', 'limits', 'of', 'ai', 'become', 'obvious', 'in', 'the', 'absence', 'of', 'cartesianism.', 'throughout', 'the', 'western', 'intellectual', 'tradition,', 'true', 'knowledge', 'has', 'been', 'viewed', 'to', 'be', 'timeless', '(grayling', '2019).', 'that', 'is,', 'this', 'information', 'is', 'assumed', 'to', 'be', 'divorced', 'from', 'contexts', 'and', 'other', 'human', 'contingencies.', 'if', 'immersed', 'in', 'these', 'situations,', 'knowledge', 'can', 'never', 'surpass', 'opinion', 'and', 'only', 'supply', 'anecdotal', 'evidence.', 'therefore,', 'most', 'philosophers', 'sought', 'foundations', 'that', 'are', 'universal', 'to', 'establish', 'firm', 'epistemological', 'and', 'moral', 'principles.', 'martin', 'heidegger', '(1969)', 'refers', 'to', 'this', 'trend', 'as', 'the', 'onto-theological', 'tradition.', 'in', 'many', 'ways,', 'cartesianism', 'epitomizes', 'this', 'tendency.', 'in', 'fact,', 'the', 'aim', 'of', 'cartesians', 'is', 'to', 'advance', 'clear', 'and', 'distinct', 'knowledge,', 'severed', 'from', 'opinion', 'and', 'other', 'sources', 'of', 'human', 'error', '(bordo', '1987).', 'in', 'this', 'regard,', 'these', 'thinkers', 'are', 'not', 'necessarily', 'unique,', 'although', 'their', 'strategy', 'is', 'novel.', 'rather', 'than', 'speculate', 'about', 'ethereal', 'metaphysical', 'principles,', 'such', 'as', 'ideas,', 'gods,', 'or', 'cosmic', 'unity,', 'cartesians', 'make', 'a', 'straightforward', 'proposal', 'known', 'as', 'dualism.']
```

Observation:

- Text tokenization splits text into individual tokens, which are typically words or punctuation marks. Lower casing converts all characters to lowercase. This helps to normalize the text and makes it easier to match words and phrases.

Step 2: Removing Special Characters

```
In [4]:   # Remove punctuation and special characters from the lowercased tokens
          filtered_tokens = [token for token in lowercased_tokens if token not in string.punctuation]

          # Display word count
          print("Word Count After Removing Special Characters:", len(filtered_tokens))
          print(filtered_tokens)
```

```
Word Count After Removing Special Characters: 625
['discussions', 'about', 'ai', 'abound.', 'but', 'critics', 'have', 'begun', 'to', 'recognize', 'that', 'these', 'debates', 'hav
e', 'focused', 'mostly', 'on', 'technical', 'issues.', 'that', 'is,', 'there', 'are', 'certain', 'problems', 'that', 'need', 't
o', 'be', 'addressed', 'and', 'have', 'technical', 'solutions.', 'debate', 'rages,', 'accordingly,', 'about', 'how', 'to', 'reso
lve', 'these', 'issues', 'and', 'move', 'the', 'applications', 'of', 'ai', 'forward.', 'perhaps', 'terry', 'winograd', '(1996)',
'was', 'correct', 'when', 'he', 'lamented', 'some', 'time', 'ago', 'that', 'interest', 'in', 'philosophy,', 'what', 'he', 'call
s', 'high', 'theory,', 'has', 'dissipated.', 'the', 'application', 'of', 'ai,', 'nonetheless,', 'has', 'pushed', 'the', 'discuss
ion', 'beyond', 'technical', 'devices', 'and', 'their', 'possible', 'uses.', 'for', 'example,', 'critics', 'have', 'begun', 't
o', 'recognize', 'that', 'ai', 'can', 'be', 'quite', 'alienating', '(dreyfus', '1992;', 'ritzer', '1993).', 'workers', 'at', 'am
azon', 'have', 'provided', 'an', 'interesting', 'case', 'study.', 'while', 'blaming', 'ai,', 'they', 'claim', 'to', 'be', 'overw
orked', 'and', 'do', 'not', 'stayed', 'employed', 'for', 'long.', 'they', 'complain', 'regularly', 'about', 'manipulation,', 'st
ress,', 'job', 'insecurity,', 'and', 'so', 'on.', 'clearly,', 'ai', 'is', 'not', 'viewed', 'to', 'be', 'their', 'friend', '(livi
ngstone', '2018;', 'vicent', '2019).', 'on', 'the', 'other', 'hand,', 'ai', 'can', 'be', 'quite', 'dangerous.', 'take', 'driverl
ess', 'cars!', 'in', 'this', 'application,', 'through', 'the', 'use', 'of', 'ai', 'cars', 'can', 'learn', 'how', 'to', 'navigat
e', 'streets,', 'other', 'cars,', 'pedestrians,', 'and', 'occasionally', 'unanticipated', 'obstacles.', 'any', 'failure', 'can',
'result', 'in', 'a', 'catastrophe,', 'even', 'death.', 'questions', 'about', 'the', 'ability', 'of', 'ai', 'to', 'master', 'trul
y', 'complex', 'activities—those', 'with', 'fluid', 'or', 'shifting', 'frames—have', 'come', 'to', 'the', 'forefront', '(goodfel
low,', 'bengio,', 'and', 'courville', '2016).', 'and', 'what', 'about', 'ethical', 'issues?', 'the', 'focus', 'of', 'developin
g', 'algorithms,', 'for', 'example,', 'is', 'not', 'necessarily', 'on', 'job', 'loss', 'or', 'intrusions', 'into', 'privacy.',
'these', 'issues', 'seems', 'to', 'make', 'context', 'difficult', 'to', 'ignore.', 'persons', 'tend', 'to', 'become', 'especiall
y', 'nervous', 'when', 'their', 'jobs', 'or', 'privacy', 'are', 'threatened.', 'although', 'alienation', 'and', 'learning', 'inv
olve', 'context,', 'the', 'ability', 'to', 'survive', 'and', 'the', 'quality', 'of', 'life', 'seem', 'to', 'go', 'to', 'the', 'h
eart', 'of', 'the', 'matter.', 'the', 'time', 'appears', 'to', 'be', 'ripe,', 'accordingly,', 'to', 'raise', 'the', 'issue', 'o
f', 'context', 'in', 'the', 'development', 'of', 'advanced', 'technology.', 'shoshana', 'zuboff', '(2019)', 'strives', 'to', 'in
itiate', 'this', 'sort', 'of', 'discussion', 'with', 'her', 'recent', 'foray', 'into', 'the', 'impact', 'of', 'late', 'capitalis
m', 'on', 'ai.', 'in', 'this', 'regard,', 'kate', 'crawford', '(2021)', 'strives', 'to', 'move', 'beyond', 'the', '"nowhere', 's
paces"', 'where', 'she', 'contends', 'most', 'discussion', 'of', 'ai', 'take', 'place.', 'she', 'believes', 'for', 'example,',
'that', 'ai', 'is', 'enmeshed', 'in', 'the', 'world's', 'ecology.', 'while', 'these', 'entreaties', 'are', 'interesting', 'and',
'relevant,', 'the', 'context', 'provided', 'by', 'twentieth', 'century', 'philosophy', 'is', 'missing.', 'indeed,', 'the', 'prag
matic', 'framework', 'that', 'is', 'currently', 'the', 'focus', 'of', 'attention', 'of', 'these', 'efforts', 'provides', 'some',
'interesting', 'insights', 'into', 'whether', 'ai', 'can', 'learn', 'or', 'deal', 'with', 'knotty', 'social', 'issues', 'the',
'emphasis', 'of', 'this', 'manuscript,', 'however,', 'is', 'the', 'anti-cartesian', 'maneuver', 'that', 'characterizes', 'much',
'of', 'contemporary', 'theory.', 'this', 'change', 'has', 'enormous', 'impact', 'on', 'the', 'potential', 'of', 'ai.', 'after',
'all,', 'cartesianism', 'is', 'at', 'the', 'core', 'of', 'digitalization', 'and', 'modern', 'data', 'processing.', 'accordingl
y,', 'the', 'importance', 'of', 'this', 'change', 'in', 'philosophical', 'orientation', 'for', 'understanding', 'the', 'mind,',
'facts,', 'learning,', 'and', 'communication', 'is', 'a', 'vital', 'consideration.', 'basic', 'to', 'this', 'reassessment', 'i
s', 'that', 'the', 'limits', 'of', 'ai', 'become', 'obvious', 'in', 'the', 'absence', 'of', 'cartesianism.', 'throughout', 'th
e', 'western', 'intellectual', 'tradition,', 'true', 'knowledge', 'has', 'been', 'viewed', 'to', 'be', 'timeless', '(grayling',
'2019).', 'that', 'is,', 'this', 'information', 'is', 'assumed', 'to', 'be', 'divorced', 'from', 'contexts', 'and', 'other', 'hu
man', 'contingencies.', 'if', 'immersed', 'in', 'these', 'situations,', 'knowledge', 'can', 'never', 'surpass', 'opinion', 'an
d', 'only', 'supply', 'anecdotal', 'evidence.', 'therefore,', 'most', 'philosophers', 'sought', 'foundations', 'that', 'are', 'u
niversal', 'to', 'establish', 'firm', 'epistemological', 'and', 'moral', 'principles.', 'martin', 'heidegger', '(1969)', 'refer
s', 'to', 'this', 'trend', 'as', 'the', 'onto-theological', 'tradition.', 'in', 'many', 'ways,', 'cartesianism', 'epitomizes',
'this', 'tendency.', 'in', 'fact,', 'the', 'aim', 'of', 'cartesians', 'is', 'to', 'advance', 'clear', 'and', 'distinct', 'knowle
dge,', 'severed', 'from', 'opinion', 'and', 'other', 'sources', 'of', 'human', 'error', '(bordo', '1987).', 'in', 'this', 'regar
d,', 'these', 'thinkers', 'are', 'not', 'necessarily', 'unique,', 'although', 'their', 'strategy', 'is', 'novel.', 'rather', 'th
an', 'speculate', 'about', 'ethereal', 'metaphysical', 'principles,', 'such', 'as', 'ideas,', 'gods,', 'or', 'cosmic', 'unity,',
'cartesians', 'make', 'a', 'straightforward', 'proposal', 'known', 'as', 'dualism.']
```

Observation:

- Special characters, such as punctuation marks, symbols, and emojis, can add noise to the data and make it more difficult for machine learning models to learn. Removing special characters can help to improve the performance of machine learning models.

3.Contraction Expansion

In [5]:
```python
# Contraction expansion dictionary
contraction_mapping = {
    "ain't": "is not",
    "aren't": "are not",
    # Add more contractions and their expansions as needed
}

# Function to expand contractions
def expand_contractions(tokens, mapping):
    expanded_tokens = []
    for token in tokens:
        if token in mapping:
            expanded_tokens.extend(mapping[token].split())
        else:
            expanded_tokens.append(token)
    return expanded_tokens

# Apply contraction expansion to the tokens
expanded_tokens = expand_contractions(filtered_tokens, contraction_mapping)

# Print the text with contractions expanded
print(expanded_tokens)
```

['discussions', 'about', 'ai', 'abound.', 'but', 'critics', 'have', 'begun', 'to', 'recognize', 'that', 'these', 'debates', 'have', 'focused', 'mostly', 'on', 'technical', 'issues.', 'that', 'is,', 'there', 'are', 'certain', 'problems', 'that', 'need', 'to', 'be', 'addressed', 'and', 'have', 'technical', 'solutions.', 'debate', 'rages,', 'accordingly,', 'about', 'how', 'to', 'resolve', 'these', 'issues', 'and', 'move', 'the', 'applications', 'of', 'ai', 'forward.', 'perhaps', 'terry', 'winograd', '(1996)', 'was', 'correct', 'when', 'he', 'lamented', 'some', 'time', 'ago', 'that', 'interest', 'in', 'philosophy,', 'what', 'he', 'calls', 'high', 'theory,', 'has', 'dissipated.', 'the', 'application', 'of', 'ai,', 'nonetheless,', 'has', 'pushed', 'the', 'discussion', 'beyond', 'technical', 'devices', 'and', 'their', 'possible', 'uses.', 'for', 'example,', 'critics', 'have', 'begun', 'to', 'recognize', 'that', 'ai', 'can', 'be', 'quite', 'alienating', '(dreyfus', '1992;', 'ritzer', '1993).', 'workers', 'at', 'amazon', 'have', 'provided', 'an', 'interesting', 'case', 'study.', 'while', 'blaming', 'ai,', 'they', 'claim', 'to', 'be', 'overworked', 'and', 'do', 'not', 'stayed', 'employed', 'for', 'long.', 'they', 'complain', 'regularly', 'about', 'manipulation,', 'stress,', 'job', 'insecurity,', 'and', 'so', 'on.', 'clearly,', 'ai', 'is', 'not', 'viewed', 'to', 'be', 'their', 'friend', '(livingstone', '2018;', 'vicent', '2019).', 'on', 'the', 'other', 'hand,', 'ai', 'can', 'be', 'quite', 'dangerous.', 'take', 'driverless', 'cars!', 'in', 'this', 'application', 'through', 'the', 'use', 'of', 'ai', 'cars', 'can', 'learn', 'how', 'to', 'navigate', 'streets', 'other', 'cars,', 'pedestrians,', 'and', 'occasionally', 'unanticipated', 'obstacles.', 'any', 'failure', 'can', 'result', 'in', 'a', 'catastrophe,', 'even', 'death.', 'questions', 'about', 'the', 'ability', 'of', 'ai', 'to', 'master', 'truly', 'complex', 'activities—those', 'with', 'fluid', 'or', 'shifting', 'frames—have', 'come', 'to', 'the', 'forefront', '(goodfellow,', 'bengio,', 'and', 'courville', '2016).', 'and', 'what', 'about', 'ethical', 'issues?', 'the', 'focus', 'of', 'developing', 'algorithms,', 'for', 'example,', 'is', 'not', 'necessarily', 'on', 'job', 'loss', 'or', 'intrusions', 'into', 'privacy.', 'these', 'issues', 'seems', 'to', 'make', 'context', 'difficult', 'to', 'ignore.', 'persons', 'tend', 'to', 'become', 'especially', 'nervous', 'when', 'their', 'jobs', 'or', 'privacy', 'are', 'threatened.', 'although', 'alienation', 'and', 'learning', 'involve', 'context,', 'the', 'ability', 'to', 'survive', 'and', 'the', 'quality', 'of', 'life', 'seem', 'to', 'go', 'to', 'the', 'heart', 'of', 'the', 'matter.', 'the', 'time', 'appears', 'to', 'be', 'ripe,', 'accordingly,', 'to', 'raise', 'the', 'issue', 'of', 'context', 'in', 'the', 'development', 'of', 'advanced', 'technology.', 'shoshana', 'zuboff', '(2019)', 'strives', 'to', 'initiate', 'this', 'sort', 'of', 'discussion', 'with', 'her', 'recent', 'foray', 'into', 'the', 'impact', 'of', 'late', 'capitalism', 'on', 'ai.', 'in', 'this', 'regard,', 'kate', 'crawford', '(2021)', 'strives', 'to', 'move', 'beyond', 'the', '"nowhere', 's', 'paces"', 'where', 'she', 'contends', 'most', 'discussion', 'of', 'ai', 'take', 'place.', 'she', 'believes,', 'for', 'example,', 'that', 'ai', 'is', 'enmeshed', 'in', 'the', 'world's', 'ecology.', 'while', 'these', 'entreaties', 'are', 'interesting', 'and', 'relevant,', 'the', 'context', 'provided', 'by', 'twentieth', 'century', 'philosophy', 'is', 'missing.', 'indeed,', 'the', 'pragmatic', 'framework', 'that', 'is', 'currently', 'the', 'focus', 'of', 'attention', 'of', 'these', 'efforts', 'provides', 'some', 'interesting', 'insights', 'into', 'whether', 'ai', 'can', 'learn', 'or', 'deal', 'with', 'knotty', 'social', 'issues', 'the', 'emphasis', 'of', 'this', 'manuscript,', 'however,', 'is', 'the', 'anti-cartesian', 'maneuver', 'that', 'characterizes', 'much', 'of', 'contemporary', 'theory.', 'this', 'change', 'has', 'enormous', 'impact', 'on', 'the', 'potential', 'of', 'ai.', 'after', 'all,', 'cartesianism', 'is', 'at', 'the', 'core', 'of', 'digitalization', 'and', 'modern', 'data', 'processing.', 'accordingly,', 'the', 'importance', 'of', 'this', 'change', 'in', 'philosophical', 'orientation', 'for', 'understanding', 'the', 'mind,', 'facts,', 'learning,', 'and', 'communication', 'is', 'a', 'vital', 'consideration.', 'basic', 'to', 'this', 'reassessment', 'is', 'that', 'the', 'limits', 'of', 'ai', 'become', 'obvious', 'in', 'the', 'absence', 'of', 'cartesianism.', 'throughout', 'the', 'western', 'intellectual', 'tradition', 'true', 'knowledge', 'has', 'been', 'viewed', 'to', 'be', 'timeless', '(grayling', '2019).', 'that', 'is,', 'this', 'information', 'is', 'assumed', 'to', 'be', 'divorced', 'from', 'contexts', 'and', 'other', 'human', 'contingencies.', 'if', 'immersed', 'in', 'these', 'situations,', 'knowledge', 'can', 'never', 'surpass', 'opinion', 'and', 'only', 'supply', 'anecdotal', 'evidence.', 'therefore,', 'most', 'philosophers', 'sought', 'foundations', 'that', 'are', 'universal', 'to', 'establish', 'firm', 'epistemological', 'and', 'moral', 'principles.', 'martin', 'heidegger', '(1969)', 'refers', 'to', 'this', 'trend', 'as', 'the', 'onto-theological', 'tradition.', 'in', 'many', 'ways,', 'cartesianism', 'epitomizes', 'this', 'tendency.', 'in', 'fact,', 'the', 'aim', 'of', 'cartesians', 'is', 'to', 'advance', 'clear', 'and', 'distinct', 'knowledge,', 'severed', 'from', 'opinion', 'and', 'other', 'sources', 'of', 'human', 'error', '(bordo', '1987).', 'in', 'this', 'regard,', 'these', 'thinkers', 'are', 'not', 'necessarily', 'unique,', 'although', 'their', 'strategy', 'is', 'novel.', 'rather', 'than', 'speculate', 'about', 'ethereal', 'metaphysical', 'principles,', 'such', 'as', 'ideas,', 'gods,', 'or', 'cosmic', 'unity,', 'cartesians', 'make', 'a', 'straightforward', 'proposal', 'known', 'as', 'dualism.']

Observation:

- Contractions are words that have been shortened by combining two or more words. For example, the contraction "I'm" is a combination of the words "I" and "am". Expanding contractions can help to improve the accuracy of machine learning models by making the text more consistent and easier to understand.

4. Removing Stopwords

In [6]:
```python
# Remove stopwords from the expanded tokens
stop_words = set(stopwords.words('english'))
filtered_tokens = [token for token in expanded_tokens if token not in stop_words]

# Print the text with stopwords removed
print(filtered_tokens)
```

['discussions', 'ai', 'abound.', 'critics', 'begun', 'recognize', 'debates', 'focused', 'mostly', 'technical', 'issues.', 'is,', 'certain', 'problems', 'need', 'addressed', 'technical', 'solutions.', 'debate', 'rages,', 'accordingly,', 'resolve', 'issues', 'move', 'applications', 'ai', 'forward.', 'perhaps', 'terry', 'winograd', '(1996)', 'correct', 'lamented', 'time', 'ago', 'interest', 'philosophy,', 'calls', 'high', 'theory,', 'dissipated.', 'application', 'ai,', 'nonetheless,', 'pushed', 'discussion', 'beyond', 'technical', 'devices', 'possible', 'uses.', 'example,', 'critics', 'begun', 'recognize', 'ai', 'quite', 'alienating', '(dreyfus', '1992;', 'ritzer', '1993).', 'workers', 'amazon', 'provided', 'interesting', 'case', 'study.', 'blaming', 'ai,', 'claim', 'overworked', 'stayed', 'employed', 'long.', 'complain', 'regularly', 'manipulation,', 'stress,', 'job', 'insecurity,', 'on.', 'clearly,', 'ai', 'viewed', 'friend', '(livingstone', '2018;', 'vicent', '2019).', 'hand,', 'ai', 'quite', 'dangerous.', 'take', 'driverless', 'cars!', 'application,', 'use', 'ai', 'cars', 'learn', 'navigate', 'streets,', 'cars,', 'pedestrians,', 'occasionally', 'unanticipated', 'obstacles.', 'failure', 'result', 'catastrophe,', 'even', 'death.', 'questions', 'ability', 'ai', 'master', 'truly', 'complex', 'activities—those', 'fluid', 'shifting', 'frames—have', 'come', 'forefront', '(goodfellow,', 'bengio,', 'courville', '2016).', 'ethical', 'issues?', 'focus', 'developing', 'algorithms,', 'example,', 'necessarily', 'job', 'loss', 'intrusions', 'privacy.', 'issues', 'seems', 'make', 'context', 'difficult', 'ignore.', 'persons', 'tend', 'become', 'especially', 'nervous', 'jobs', 'privacy', 'threatened.', 'although', 'alienation', 'learning', 'involve', 'context,', 'ability', 'survive', 'quality', 'life', 'seem', 'go', 'heart', 'matter.', 'time', 'appears', 'ripe,', 'accordingly,', 'raise', 'issue', 'context', 'development', 'advanced', 'technology.', 'shoshana', 'zuboff', '(2019)', 'strives', 'initiate', 'sort', 'discussion', 'recent', 'foray', 'impact', 'late', 'capitalism', 'ai.', 'regard,', 'kate', 'crawford', '(2021)', 'strives', 'move', 'beyond', '"nowhere', 'spaces"', 'contends', 'discussion', 'ai', 'take', 'place.', 'believes,', 'example,', 'ai', 'enmeshed', 'world's', 'ecology.', 'entreaties', 'interesting', 'relevant,', 'context', 'provided', 'twentieth', 'century', 'philosophy', 'missing.', 'indeed,', 'pragmatic', 'framework', 'currently', 'focus', 'attention', 'efforts', 'provides', 'interesting', 'insights', 'whether', 'ai', 'learn', 'deal', 'knotty', 'social', 'issues', 'emphasis', 'manuscript,', 'however,', 'anti-cartesian', 'maneuver', 'characterizes', 'much', 'contemporary', 'theory.', 'change', 'enormous', 'impact', 'potential', 'ai.', 'all,', 'cartesianism', 'core', 'digitalization', 'modern', 'data', 'processing.', 'accordingly,', 'importance', 'change', 'philosophical', 'orientation', 'understanding', 'mind,', 'facts,', 'learning,', 'communication', 'vital', 'consideration.', 'basic', 'reassessment', 'limits', 'ai', 'become', 'obvious', 'absence', 'cartesianism.', 'throughout', 'western', 'intellectual', 'tradition', 'true', 'knowledge', 'viewed', 'timeless', '(grayling', '2019).', 'is,', 'information', 'assumed', 'divorced', 'contexts', 'human', 'contingencies.', 'immersed', 'situations,', 'knowledge', 'never', 'surpass', 'opinion', 'supply', 'anecdotal', 'evidence.', 'therefore,', 'philosophers', 'sought', 'foundations', 'universal', 'establish', 'firm', 'epistemological', 'moral', 'principles.', 'martin', 'heidegger', '(1969)', 'refers', 'trend', 'onto-theological', 'tradition.', 'many', 'ways,', 'cartesianism', 'epitomizes', 'tendency.', 'fact,', 'aim', 'cartesians', 'advance', 'clear', 'distinct', 'knowledge,', 'severed', 'opinion', 'sources', 'human', 'error', '(bordo', '1987).', 'regard,', 'thinkers', 'necessarily', 'unique,', 'although', 'strategy', 'novel.', 'rather', 'speculate', 'ethereal', 'metaphysical', 'principles,', 'ideas,', 'gods,', 'cosmic', 'unity,', 'cartesians', 'make', 'straightforward', 'proposal', 'known', 'dualism.']

Observation:

- Stopwords are common words that do not add much meaning to the text. For example, the words "the", "is", and "and" are stopwords. Removing stopwords can help to reduce the size of the data and improve the performance of machine learning models.

Step 5: Stemming

```
In [7]:   from nltk.stem import PorterStemmer

          # Initialize the Porter Stemmer
          stemmer = PorterStemmer()

          # Apply stemming to the tokens
          stemmed_tokens = [stemmer.stem(token) for token in filtered_tokens]

          # Display word count after stemming
          print("Word Count After Stemming:", len(stemmed_tokens))
          print(stemmed_tokens)
```

```
Word Count After Stemming: 362
['discuss', 'ai', 'abound.', 'critic', 'begun', 'recogn', 'debat', 'focus', 'mostli', 'technic', 'issues.', 'is,', 'certain', 'p
roblem', 'need', 'address', 'technic', 'solutions.', 'debat', 'rages,', 'accordingly,', 'resolv', 'issu', 'move', 'applic', 'a
i', 'forward.', 'perhap', 'terri', 'winograd', '(1996)', 'correct', 'lament', 'time', 'ago', 'interest', 'philosophy,', 'call',
'high', 'theory,', 'dissipated.', 'applic', 'ai,', 'nonetheless,', 'push', 'discuss', 'beyond', 'technic', 'devic', 'possibl',
'uses.', 'example,', 'critic', 'begun', 'recogn', 'ai', 'quit', 'alien', '(dreyfu', '1992;', 'ritzer', '1993).', 'worker', 'amaz
on', 'provid', 'interest', 'case', 'study.', 'blame', 'ai,', 'claim', 'overwork', 'stay', 'employ', 'long.', 'complain', 'regula
rli', 'manipulation,', 'stress,', 'job', 'insecurity,', 'on.', 'clearly,', 'ai', 'view', 'friend', '(livingston', '2018;', 'vice
nt', '2019).', 'hand,', 'ai', 'quit', 'dangerous.', 'take', 'driverless', 'cars!', 'application,', 'use', 'ai', 'car', 'learn',
'navig', 'streets,', 'cars,', 'pedestrians,', 'occasion', 'unanticip', 'obstacles.', 'failur', 'result', 'catastrophe,', 'even',
'death.', 'question', 'abil', 'ai', 'master', 'truli', 'complex', 'activities—thos', 'fluid', 'shift', 'frames—hav', 'come', 'fo
refront', '(goodfellow,', 'bengio,', 'courvil', '2016).', 'ethic', 'issues?', 'focu', 'develop', 'algorithms,', 'example,', 'nec
essarili', 'job', 'loss', 'intrus', 'privacy.', 'issu', 'seem', 'make', 'context', 'difficult', 'ignore.', 'person', 'tend', 'be
com', 'especi', 'nervou', 'job', 'privaci', 'threatened.', 'although', 'alien', 'learn', 'involv', 'context,', 'abil', 'surviv',
'qualiti', 'life', 'seem', 'go', 'heart', 'matter.', 'time', 'appear', 'ripe,', 'accordingly,', 'rais', 'issu', 'context', 'deve
lop', 'advanc', 'technology.', 'shoshana', 'zuboff', '(2019)', 'strive', 'initi', 'sort', 'discuss', 'recent', 'foray', 'impac
t', 'late', 'capit', 'ai.', 'regard,', 'kate', 'crawford', '(2021)', 'strive', 'move', 'beyond', '"nowher', 'spaces"', 'conten
d', 'discuss', 'ai', 'take', 'place.', 'believes,', 'example,', 'ai', 'enmesh', 'world"', 'ecology.', 'entreati', 'interest', 'r
elevant,', 'context', 'provid', 'twentieth', 'centuri', 'philosophi', 'missing.', 'indeed,', 'pragmat', 'framework', 'current',
'focu', 'attent', 'effort', 'provid', 'interest', 'insight', 'whether', 'ai', 'learn', 'deal', 'knotti', 'social', 'issu', 'emph
asi', 'manuscript,', 'however,', 'anti-cartesian', 'maneuv', 'character', 'much', 'contemporari', 'theory.', 'chang', 'enorm',
'impact', 'potenti', 'ai.', 'all,', 'cartesian', 'core', 'digit', 'modern', 'data', 'processing.', 'accordingly,', 'import', 'ch
ang', 'philosoph', 'orient', 'understand', 'mind,', 'facts,', 'learning,', 'commun', 'vital', 'consideration.', 'basic', 'reasse
ss', 'limit', 'ai', 'becom', 'obviou', 'absenc', 'cartesianism.', 'throughout', 'western', 'intellectu', 'tradition,', 'true',
'knowledg', 'view', 'timeless', '(grayl', '2019).', 'is,', 'inform', 'assum', 'divorc', 'context', 'human', 'contingencies.', 'i
mmers', 'situations,', 'knowledg', 'never', 'surpass', 'opinion', 'suppli', 'anecdot', 'evidence.', 'therefore,', 'philosoph',
'sought', 'foundat', 'univers', 'establish', 'firm', 'epistemolog', 'moral', 'principles.', 'martin', 'heidegg', '(1969)', 'refe
r', 'trend', 'onto-theolog', 'tradition.', 'mani', 'ways,', 'cartesian', 'epitom', 'tendency.', 'fact,', 'aim', 'cartesian', 'ad
vanc', 'clear', 'distinct', 'knowledge,', 'sever', 'opinion', 'sourc', 'human', 'error', '(bordo', '1987).', 'regard,', 'thinke
r', 'necessarili', 'unique,', 'although', 'strategi', 'novel.', 'rather', 'specul', 'ether', 'metaphys', 'principles,', 'idea
s,', 'gods,', 'cosmic', 'unity,', 'cartesian', 'make', 'straightforward', 'propos', 'known', 'dualism.']
```

Observation:

- Stemming can reduce the size of the data and make it easier to match words and phrases. However, it can also lead to loss of information, as it reduces words to their stems.

Step 6: Lemmatization

```
In [8]:   from nltk.stem import WordNetLemmatizer

          # Initialize the WordNet Lemmatizer
          lemmatizer = WordNetLemmatizer()

          # Apply lemmatization to the tokens
          lemmatized_tokens = [lemmatizer.lemmatize(token) for token in filtered_tokens]

          # Display word count after lemmatization
          print("Word Count After Lemmatization:", len(lemmatized_tokens))
          print(lemmatized_tokens)
```

```
Word Count After Lemmatization: 362
['discussion', 'ai', 'abound.', 'critic', 'begun', 'recognize', 'debate', 'focused', 'mostly', 'technical', 'issues.', 'is,', 'c
ertain', 'problem', 'need', 'addressed', 'technical', 'solutions.', 'debate', 'rages,', 'accordingly,', 'resolve', 'issue', 'mov
e', 'application', 'ai', 'forward.', 'perhaps', 'terry', 'winograd', '(1996)', 'correct', 'lamented', 'time', 'ago', 'interest',
'philosophy,', 'call', 'high', 'theory,', 'dissipated.', 'application', 'ai,', 'nonetheless,', 'pushed', 'discussion', 'beyond',
'technical', 'device', 'possible', 'uses.', 'example,', 'critic', 'begun', 'recognize', 'ai', 'quite', 'alienating', '(dreyfus',
'1992;', 'ritzer', '1993).', 'worker', 'amazon', 'provided', 'interesting', 'case', 'study.', 'blaming', 'ai,', 'claim', 'overwo
rked', 'stayed', 'employed', 'long.', 'complain', 'regularly', 'manipulation,', 'stress,', 'job', 'insecurity,', 'on.', 'clearl
y,', 'ai', 'viewed', 'friend', '(livingstone', '2018;', 'vicent', '2019).', 'hand,', 'ai', 'quite', 'dangerous.', 'take', 'drive
rless', 'cars!', 'application,', 'use', 'ai', 'car', 'learn', 'navigate', 'streets,', 'cars,', 'pedestrians,', 'occasionally',
'unanticipated', 'obstacles.', 'failure', 'result', 'catastrophe,', 'even', 'death.', 'question', 'ability', 'ai', 'master', 'tr
uly', 'complex', 'activities—those', 'fluid', 'shifting', 'frames—have', 'come', 'forefront', '(goodfellow,', 'bengio,', 'courvi
lle', '2016).', 'ethical', 'issues?', 'focus', 'developing', 'algorithms,', 'example,', 'necessarily', 'job', 'loss', 'intrusio
n', 'privacy.', 'issue', 'seems', 'make', 'context', 'difficult', 'ignore.', 'person', 'tend', 'become', 'especially', 'nervou
s', 'job', 'privacy', 'threatened.', 'although', 'alienation', 'learning', 'involve', 'context,', 'ability', 'survive', 'qualit
y', 'life', 'seem', 'go', 'heart', 'matter.', 'time', 'appears', 'ripe,', 'accordingly,', 'raise', 'issue', 'context', 'developm
ent', 'advanced', 'technology.', 'shoshana', 'zuboff', '(2019)', 'strives', 'initiate', 'sort', 'discussion', 'recent', 'foray',
'impact', 'late', 'capitalism', 'ai.', 'regard,', 'kate', 'crawford', '(2021)', 'strives', 'move', 'beyond', '"nowhere', 'space
s"', 'contends', 'discussion', 'ai', 'take', 'place.', 'believes,', 'example,', 'ai', 'enmeshed', 'world's', 'ecology.', 'entrea
ty', 'interesting', 'relevant,', 'context', 'provided', 'twentieth', 'century', 'philosophy', 'missing.', 'indeed,', 'pragmati
c', 'framework', 'currently', 'focus', 'attention', 'effort', 'provides', 'interesting', 'insight', 'whether', 'ai', 'learn', 'd
eal', 'knotty', 'social', 'issue', 'emphasis', 'manuscript', 'however,', 'anti-cartesian', 'maneuver', 'characterizes', 'much',
'contemporary', 'theory.', 'change', 'enormous', 'impact', 'potential', 'ai.', 'all,', 'cartesianism', 'core', 'digitalization',
'modern', 'data', 'processing.', 'accordingly,', 'importance', 'change', 'philosophical', 'orientation', 'understanding', 'min
d,', 'facts,', 'learning,', 'communication', 'vital', 'consideration.', 'basic', 'reassessment', 'limit', 'ai', 'become', 'obvio
us', 'absence', 'cartesianism.', 'throughout', 'western', 'intellectual', 'tradition,', 'true', 'knowledge', 'viewed', 'timeles
s', '(grayling', '2019).', 'is,', 'information', 'assumed', 'divorced', 'context', 'human', 'contingencies.', 'immersed', 'situa
tions,', 'knowledge', 'never', 'surpass', 'opinion', 'supply', 'anecdotal', 'evidence.', 'therefore,', 'philosopher', 'sought',
'foundation', 'universal', 'establish', 'firm', 'epistemological', 'moral', 'principles.', 'martin', 'heidegger', '(1969)', 'ref
ers', 'trend', 'onto-theological', 'tradition.', 'many', 'ways,', 'cartesianism', 'epitomizes', 'tendency.', 'fact,', 'aim', 'ca
rtesian', 'advance', 'clear', 'distinct', 'knowledge,', 'severed', 'opinion', 'source', 'human', 'error', '(bordo', 1987).', 'r
egard,', 'thinker', 'necessarily', 'unique,', 'although', 'strategy', 'novel.', 'rather', 'speculate', 'ethereal', 'metaphysica
l', 'principles,', 'ideas,', 'gods,', 'cosmic', 'unity,', 'cartesian', 'make', 'straightforward', 'proposal', 'known', 'dualis
m.']
```

Observations:

- Lemmatization is more sophisticated than stemming and can preserve more of the meaning of the words. However, it can be more computationally expensive.

Step 7: Aggregated Features

In [9]:
```python
# Calculate word count for the final preprocessed text
word_count = len(lemmatized_tokens)
print("Final Word Count:", word_count)
```

```
Final Word Count: 362
```

Observations:

- Aggregated features can be useful for identifying trends and patterns in the data. However, they can also be less informative than individual features.

Step 8: Date-Time Features

In [10]:
```python
import datetime

# Get the current date and time
current_datetime = datetime.datetime.now()

# Display the current date and time
print("Current Date and Time:", current_datetime)
```

```
Current Date and Time: 2023-10-30 22:58:41.321419
```

Observations:

- Date-time features can be useful for predicting future events or identifying seasonal patterns. However, they can also be less informative than text features.

Step 9: TF-IDF Features':

In [11]:
```python
from sklearn.feature_extraction.text import TfidfVectorizer

# Initialize the TF-IDF vectorizer
tfidf_vectorizer = TfidfVectorizer()

# Fit and transform the preprocessed text to calculate TF-IDF features
tfidf_features = tfidf_vectorizer.fit_transform([" ".join(lemmatized_tokens)])

# Display the TF-IDF features
print("TF-IDF Features:")
print(tfidf_features.toarray())
```

```
TF-IDF Features:
[[0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.1098378  0.0366126  0.0732252  0.0366126  0.0366126
  0.1098378  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.54918902 0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0732252  0.0366126  0.0366126  0.0366126  0.0366126  0.1098378
  0.0366126  0.0366126  0.0366126  0.0732252  0.0732252  0.0366126
  0.0366126  0.0732252  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0732252  0.1098378  0.1098378  0.0366126  0.0366126
  0.0366126  0.0366126  0.0732252  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.18306301 0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0732252  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0732252  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.14645041 0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.1098378  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0732252  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0732252  0.0366126  0.0366126
  0.0366126  0.0732252  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.1098378  0.0366126
  0.0366126  0.0732252  0.14645041 0.0732252  0.1098378  0.0366126
  0.0366126  0.1098378  0.0366126  0.0366126  0.0366126  0.0732252
  0.0732252  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0732252  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0732252  0.0366126  0.0366126  0.0732252
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0732252
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0732252  0.0366126  0.0366126  0.0366126  0.0366126
  0.0732252  0.0732252  0.0366126  0.0366126  0.0366126  0.0732252
  0.0366126  0.0366126  0.0366126  0.0366126  0.0732252  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0732252  0.0366126
  0.0732252  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0732252  0.0366126  0.0366126  0.0366126  0.0366126
  0.0732252  0.1098378  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0732252  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0732252  0.0366126  0.0732252  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0732252  0.0366126
  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126  0.0366126
  0.0366126 ]]
```

Observations:

- TF-IDF identifies the importance of words in the text based on their frequency and uniqueness. It helps highlight significant terms, potentially useful for document classification and information retrieval.

Step 11: Character Level TF-IDF:

In [12]:
```python
# Initialize the Character Level TF-IDF vectorizer
char_tfidf_vectorizer = TfidfVectorizer(analyzer='char', ngram_range=(1, 3))

# Fit and transform the preprocessed text to calculate Character Level TF-IDF features
char_tfidf_features = char_tfidf_vectorizer.fit_transform([" ".join(lemmatized_tokens)])

# Display the Character Level TF-IDF features
print("Character Level TF-IDF Features:")
print(char_tfidf_features.toarray())
```

```
Character Level TF-IDF Features:
[[0.46097057 0.01149234 0.00255385 ... 0.00127693 0.00127693 0.00127693]]
```

Observations:

- Word-level TF-IDF extends TF-IDF to individual words.
- It can provide more detailed insights into word importance within the text.
- Character-level TF-IDF considers individual characters as tokens.
- This can be useful for analyzing text where character-level patterns are important, such as in sentiment analysis.

Step 12: Word Embedding Features

In [13]:
```python
import spacy
# Word Embedding Features (using spaCy)
nlp = spacy.load("en_core_web_sm")
doc = nlp(text)
word_embeddings = doc.vector  # This represents the document in vector form

print("9. Word Embedding Features:")
print(word_embeddings)
```

9. Word Embedding Features:
```
[ 2.04537343e-02 -3.28307658e-01 -1.19586006e-01  1.70986623e-01
  1.87846914e-01 -1.44223705e-01  3.73611748e-01  2.82605886e-01
  2.04944536e-01 -2.81335026e-01 -1.12993598e-01  1.70498379e-02
 -4.34800655e-01 -5.29034615e-01 -1.80870846e-01  2.62012273e-01
 -1.50933519e-01 -1.38827518e-01 -5.31255543e-01 -2.12910414e-01
  1.49458079e-02  4.14486110e-01 -3.27145904e-01  1.80062316e-02
  2.89542545e-02 -8.82609114e-02  3.38365257e-01  6.86462343e-01
  3.72882754e-01 -1.86958379e-04  1.42645493e-01 -1.28249958e-01
 -2.56096900e-01 -5.57084102e-03 -6.29138127e-02 -2.44863689e-01
  1.50707960e-01  4.94341403e-02  1.26808599e-01 -1.50142387e-01
 -4.01284575e-01  2.02549174e-01 -2.87601762e-02  3.07297915e-01
  1.40913278e-01  4.57746945e-02  3.00672948e-02 -2.12178808e-02
 -2.35103101e-01 -8.62731114e-02 -5.80132663e-01  6.83139712e-02
  9.13135931e-02 -5.75393558e-01 -3.76059741e-01  2.59555548e-01
  3.14639568e-01  5.30456472e-03 -2.10698345e-03 -4.71865721e-02
 -9.30994749e-02 -1.06954880e-01 -8.92128497e-02 -2.61776805e-01
  2.27242619e-01 -1.58029675e-01  2.35476106e-01  1.45069242e-01
  7.87395164e-02 -1.26975089e-01  5.21830320e-01  2.29009107e-01
  4.61135745e-01 -4.21056598e-01  9.91114974e-02 -8.74110609e-02
  3.74628380e-02 -7.86580622e-01  3.76642756e-02 -6.68925524e-01
 -1.59439594e-01 -1.19335957e-01 -1.67524233e-01  1.92751005e-01
  1.76484779e-01  7.63944834e-02  1.26266807e-01 -1.01006545e-01
 -5.00824034e-01  2.46855497e-01  1.50041968e-01 -1.03717536e-01
  1.00435126e+00  6.79803431e-01 -2.29640827e-01  4.71054018e-02]
```

Step 13: Count Features:

In [14]:
```python
# Count Features
count_vectorizer = CountVectorizer()
count_features = count_vectorizer.fit_transform([text])

print("10. Count Features:")
print(count_features.toarray())
```

```
10. Count Features:
[[ 1  1  1  1  1  1  1  3  1  2  1  1  3  1  1  1  1  1 15  1  1  1  1  2
   1  1  1  1  2  1  1  1  1  2  2  1  1  2  1  1  1  1  3  3  2  1  1  1
   1  2  1  1  1  1  1  1  1  1  1  1  1  4  1  1  1  1  1  1  1  2  1  1
   1  1  1  1  1  1  1  1  1  1  3  1  1  1  1  1  1  1  1  1  1  1  1  1
   1  1  1  1  1  1  1  1  1  3  1  1  1  1  1  1  2  1  1  1  1  1  1  1
   1  1  1  1  1  1  1  1  1  1  1  2  1  1  1  2  1  1  1  1  1  1  1  1
   3  1  1  1  5  2  1  1  1  3  1  1  1  2  2  1  1  1  1  2  1  1  1
   1  1  1  1  1  1  1  1  1  1  2  1  1  2  1  1  1  1  1  1  1  1  1
   2  1  1  1  1  1  1  1  2  1  1  1  1  2  2  1  1  2  1  1  1  1  2
   1  1  1  1  2  1  2  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
   1  1  1  1  1  1  1  1  2  1  1  1  1  2  3  1  1  1  1  2  1  1  1
   1  2  1  2  1  1  1  1  1  1  1  1  1  1  1  1  1  2  1  1  1  1  1  1
   1]]
```

In [15]:
```python
# Define the words you want to count
selected_words = ["AI", "technology", "philosophy", "ethics"]

# Tokenize the text
tokens = text.split()

# Count the occurrences of selected words
word_counts = {word: tokens.count(word) for word in selected_words}

# Print word counts
print("Word Counts:")
for word, count in word_counts.items():
    print(f"{word}: {count}")

# Tokenization (assuming you already have the 'tokens' from the previous steps)
tokens = nltk.word_tokenize(text)

# Count of words
word_count = len(tokens)

# Average word length
avg_word_length = sum(len(word) for word in tokens) / word_count

# Number of unique words
unique_words = set(tokens)
unique_word_count = len(unique_words)

# Print the aggregated features
print("Aggregated Features:")
print("Word Count:", word_count)
print("Average Word Length:", avg_word_length)
print("Unique Word Count:", unique_word_count)
```

```
Word Counts:
AI: 15
technology: 1
philosophy: 2
ethics: 0
Aggregated Features:
Word Count: 362
Average Word Length: 7.044198895027624
Unique Word Count: 294
```

Observation:

- Count features help identify the frequency of specific words, which can be useful for identifying important terms or topics in the text.

Step 14: Bag-of-n-Grams

```python
In [16]: from sklearn.feature_extraction.text import CountVectorizer
         # Bag-of-n-Grams
         n_gram_vectorizer = CountVectorizer(ngram_range=(2, 2))
         n_gram_features = n_gram_vectorizer.fit_transform([text])

         print("11. Bag-of-n-Grams:")
         print(n_gram_features.toarray())
```

```
11. Bag-of-n-Grams:
[[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1
  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1
  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]]
```

Observation:

- Bag-of-n-grams capture sequences of words (n-grams), providing more context and helping identify multi-word phrases in the text. This is especially valuable in text analysis when word order matters.