In [32]:

```
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
```

In [33]:

```
with open('AI.txt', 'r') as file:
    data = file.read().replace('\n', '')

STOP_WORDS = set(stopwords.words('english'))
STOP_WORDS.add('a')
STOP_WORDS.add('they')
STOP_WORDS.add('the')
STOP_WORDS.add('his')
STOP_WORDS.add('.')
STOP_WORDS.add(',')
STOP_WORDS.add('so')
STOP_WORDS.add('and')
STOP_WORDS.add('were')
STOP_WORDS.add('from')
STOP_WORDS.add('that')
STOP_WORDS.add('of')
STOP_WORDS.add('in')
STOP_WORDS.add('only')
STOP_WORDS.add('with')
STOP_WORDS.add('to')

word_tokens = word_tokenize(data)


filtered_sentence1 = [w for w in word_tokens if not w in STOP_WORDS]

print(len(filtered_sentence1))
```

```
511
```

In [35]:

```
with open('ML.txt', 'r') as file:
    data = file.read().replace('\n', '')

STOP_WORDS = set(stopwords.words('english'))

STOP_WORDS.add('a')
STOP_WORDS.add('they')
STOP_WORDS.add('the')
STOP_WORDS.add('his')
STOP_WORDS.add('.')
STOP_WORDS.add(',')
STOP_WORDS.add('so')
STOP_WORDS.add('and')
STOP_WORDS.add('were')
STOP_WORDS.add('from')
STOP_WORDS.add('that')
STOP_WORDS.add('of')
STOP_WORDS.add('in')
STOP_WORDS.add('only')
STOP_WORDS.add('with')
STOP_WORDS.add('to')

word_tokens = word_tokenize(data)

##print(word_tokens)
filtered_sentence2 = [w for w in word_tokens if not w in STOP_WORDS]

print(len(filtered_sentence2))
```

```
707
```

In [40]:

```
print("\n\nFrequency of each word in ai.txt : - \n")
d=dict()
for line in filtered_sentence1:
    line = line.strip()
    line = line.lower()
    words1 = line.split(" ")
    for word in filtered_sentence1:
        if word in d:
            d[word] = d[word] + 1
        else:
            d[word] = 1

for key in list(d.keys()):
    print(key, ":", d[key])
```

```
human : 3577
beings.What : 511
? : 2044
According : 511
father : 511
John : 511
McCarthy : 511
“ : 1022
The : 1533
science : 1022
engineering : 511
making : 1022
especially : 511
programs : 1022
” : 1022
.Artificial : 511
way : 1022
computer-controlled : 511
robot : 511
```

In [41]:

```
print("\n\nFrequency of each word in ml.txt : - \n")
d=dict()
for line in filtered_sentence2:
    line = line.strip()
    line = line.lower()
    words1 = line.split(" ")
    for word in filtered_sentence2:
        if word in d:
            d[word] = d[word] + 1
        else:
            d[word] = 1

for key in list(d.keys()):
    print(key, ":", d[key])
```

```
Frequency of each word in ml.txt : -

Artificial : 1414
Intelligence : 2121
Machine : 707
Learning : 707
: : 6363
Policy : 707
PaperForewordArtificial : 707
intelligence : 9191
technology : 5656
already : 2121
impacting : 707
users : 1414
interact : 2121
affected : 707
Internet : 11312
```

In [45]:

```
#Common words
if len(f1) != 0 | len(f2) != 0:
    for words in f1:
        for wordds in f2:
            if words == wordds:
                print(words)
```

```
perform
perform
various
tasks
tasks
developed
computer
computer
systems
systems
increasing
Artificial
Artificial
Intelligence
Intelligence
Intelligence
creating
intelligent
Artificial
```

In [ ]: