1. Import the necessary libraries

In [1]:

```
import nltk
nltk.download("punkt")
nltk.download("stopwords")
```

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

Out[1]:

```
True
```

In [2]:

```
from nltk import word_tokenize
from nltk import sent_tokenize
#!pip install autocorrect
#!pip install spellchecker
from spellchecker import SpellChecker
import re
from string import punctuation
#from autocorrect import Speller
from nltk.corpus import stopwords
from nltk.stem.snowball import SnowballStemmer
```

2. Load the text in data.text to a variable text. Apply lower casing.

In [3]:

```
data_path = "C:\spark\MCA\Semester1\E3_NLP\input\lab_e2
\Artificial_Intelligence.txt"
text = ""
with open (data_path, "rb") as f:
    text = f.read()
text = text.lower()
```

In [4]:

```
text = str(text)
```

3. Create a vocabulary for the text.

In [5]:

```
vocab = text.split(" ")
```

4. Apply the word tokenization process to the text and store in wordTokens

In [6]:

```
wordTokens = word_tokenize(text)
```

a. How is str.split() different from word tokenizer?

split() split a string into a list of strings after breaking the given string by the specified separator.

split() on given text gave 2140 sub-strings which included only words

word_tokenize() divide strings into lists of substrings. For example, tokenizers can be used to find the words and punctuation in a string

word_tokenize() on given strings gave 2462 sub-strings which included both words and punctuations

```
In [7]: print("Number of splits using split():", len(vocab))
        print("Number of splits using word_tokenize():", len(wordTokens))
```

```
Number of splits using split(): 2140
Number of splits using word_tokenize(): 2430
```

b. How many tokens are there now?

```
In [8]: print("Number of splits using word_tokenize():", len(wordTokens))
```

```
Number of splits using word_tokenize(): 2430
```

c. Print the tokens

```
In [9]: print(wordTokens)
```

```
["b'ai", 'is', 'undoubtedly', 'one', 'of', 'the', 'biggest', 'tech', 'trends', 'at', 't
he', 'moment', ',', 'and', 'during', '2021', 'it', 'will', 'become', 'an', 'even', 'mor
e', 'valuable', 'tool', 'for', 'helping', 'us', 'to', 'interpret', 'and', 'understand',
'the', 'world', 'around', 'us', '.', 'the', 'volume', 'of', 'data', 'we', 'are', 'colle
cting', 'on', 'health-care', ',', 'infection', 'rates', ',', 'and', 'the', 'success', '
of', 'measures', 'we', 'take', 'to', 'prevent', 'the', 'spread', 'of', 'infection', 'wi
ll', 'continue', 'to', 'increase', '.', 'this', 'means', 'that', 'machine', 'learning',
'algorithms', 'will', 'become', 'better', 'informed', 'and', 'increasingly', 'sophistic
ated', 'in', 'the', 'solutions', 'they', 'uncover', 'for', 'us', '.', '\\r\\nfrom', 'co
mputer', 'vision', 'systems', 'monitoring', 'the', 'capacity', 'of', 'public', 'areas',
'to', 'analyzing', 'the', 'interactions', 'uncovered', 'through', 'contact', 'tracing',
'initiatives', ',', 'self-learning', 'algorithms', 'will', 'spot', 'connections', 'and
', 'insights', 'that', 'would', 'go', 'unnoticed', 'by', 'manual', 'human', 'analysis',
'.', 'they', 'will', 'help', 'us', 'predict', 'demand', 'for', 'services', 'from', 'hos
pitals', 'and', 'other', 'healthcare', 'providers', ',', 'and', 'allow', 'administrator
s', 'to', 'make', 'better', 'decisions', 'about', 'when', 'and', 'where', 'to', 'deploy
', 'resources', '.', '\\r\\nfor', 'business', ',', 'the', 'challenge', 'will', 'be', 't
o', 'understand', 'the', 'changing', 'patterns', 'of', 'customer', 'behavior', '.', 'mo
re', 'human', 'activity', 'will', 'take', 'place', 'online', '\\xe2\\x80\\x93', 'from',
'shopping', 'and', 'socializing', 'to', 'virtual', 'working', 'environments', ',', 'mee
tings', ',', 'and', 'recruitment', '.', 'during', '2021', 'we', 'can', 'expect', 'the',
'tools', 'we', 'use', 'to', 'analyze', 'these', 'behavioral', 'shifts', 'to', 'become',
'more', 'sophisticated', 'and', 'increasingly', 'fit', 'the', 'budget', 'and', 'infrast
ructure', 'requirements', 'of', 'more', 'and', 'more', 'organizations', '.', '\\r\\nsom
e', 'of', 'the', 'best', 'circuits', 'to', 'drive', 'ai', 'in', 'the', 'future', 'may',
'be', 'analog', ',', 'not', 'digital', ',', 'and', 'research', 'teams', 'around', 'the
', 'world', 'are', 'increasingly', 'developing', 'new', 'devices', 'to', 'support', 'su
ch', 'analog', 'ai.\\r\\n\\r\\nthe', 'most', 'basic', 'computation', 'in', 'the', 'deep
', 'neural', 'networks', 'driving', 'the', 'current', 'explosion', 'in', 'ai', 'is', 't
he', 'multiply-accumulate', '(', 'mac', ')', 'operation', '.', 'deep', 'neural', 'netwo
rks', 'are', 'composed', 'of', 'layers', 'of', 'artificial', 'neurons', ',', 'and', 'in
', 'mac', 'operations', ',', 'the', 'output', 'of', 'each', 'one', 'of', 'these', 'laye
rs', 'is', 'multiplied', 'by', 'the', 'values', 'of', 'the', 'strengths', 'or', '``', '
weights', "''", 'of', 'their', 'connections', 'to', 'the', 'next', 'layer', ',', 'which
', 'then', 'sums', 'up', 'these', 'contributions.\\r\\n\\r\\nmodern', 'computers', 'hav
e', 'digital', 'components', 'devoted', 'to', 'mac', 'operations', ',', 'but', 'analog
', 'circuits', 'theoretically', 'can', 'perform', 'these', 'computations', 'for', 'orde
rs', 'of', 'magnitude', 'less', 'energy', '.', 'this', 'strategy\\xe2\\x80\\x94known',
'as', 'analog', 'ai', ',', 'compute-in-memory', 'or', 'processing-in-memory\\xe2\\x80\\
x94often', 'performs', 'these', 'multiply-accumulate', 'operations', 'using', 'non-vola
tile', 'memory', 'devices', 'such', 'as', 'flash', ',', 'magnetoresistive', 'ram', '(',
'mram', ')', ',', 'resistive', 'ram', '(', 'rram', ')', ',', 'phase-change', 'memory',
'(', 'pcm', ')', 'and', 'even', 'more', 'esoteric', 'technologies.\\r\\n\\r\\none', 'te
am', 'in', 'korea', ',', 'however', ',', 'is', 'exploring', 'neural', 'networks', 'base
d', 'on', 'praseodymium', 'calcium', 'manganese', 'oxide', 'electrochemical', 'ram',
'(', 'ecram', ')', 'devices', ',', 'which', 'act', 'like', 'miniature', 'batteries',
',', 'storing', 'data', 'in', 'the', 'form', 'of', 'changes', 'in', 'their', 'conductan
ce', '.', 'study', 'lead', 'author', 'chuljun', 'lee', 'at', 'the', 'pohang', 'universi
ty', 'of', 'science', 'and', 'technology', 'in', 'korea', 'notes', 'that', 'neural', 'n
etwork', 'hardware', 'often', 'has', 'different', 'demands', 'during', 'training', 'ver
sus', 'during', 'applications', '.', 'for', 'instance', ',', 'low', 'energy', 'barriers
', 'help', 'neural', 'networks', 'learn', 'quickly', ',', 'but', 'high', 'energy', 'bar
riers', 'help', 'them', 'retain', 'what', 'they', 'learned', 'for', 'use', 'during', 'a
pplications.\\r\\n\\r\\n', "''", 'heating', 'up', 'their', 'devices', 'almost', '100',
'degrees', 'c', 'warmer', 'during', 'training', 'brought', 'out', 'the', 'characteristi
cs', 'that', 'are', 'good', 'for', 'training', ',', "''", 'says', 'electrical', 'engine
er', 'john', 'paul', 'strachan', ',', 'head', 'of', 'the', 'peter', 'gr\\xc3\\xbcnberg
', 'institute', 'for', 'neuromorphic', 'compute', 'nodes', 'at', 'the', 'j\\xc3\\xbclic
h', 'research', 'center', 'in', 'germany', ',', 'who', 'did', 'not', 'participate', 'in
', 'this', 'study', '.', '``', 'when', 'it', 'cooled', 'down', ',', 'they', 'got', 'the
', 'advantages', 'of', 'longer', 'retention', 'and', 'lower', 'current', 'operation',
'.', 'by', 'just', 'adjusting', 'one', 'knob', ',', 'heat', ',', 'they', 'could', 'see
', 'improvements', 'on', 'multiple', 'dimensions', 'of', 'computing', '.', "''", 'the',
'researchers', 'detailed', 'their', 'findings', 'at', 'the', 'annual', 'ieee', 'interna
tional', 'electron', 'devices', 'meeting', '(', 'iedm', ')', 'in', 'san', 'francisco',
'on', 'dec.', '14.\\r\\n\\r\\none', 'key', 'question', 'this', 'work', 'faces', 'is', '
what', 'kind', 'of', 'deterioration', 'this', 'ecram', 'may', 'face', 'after', 'multipl
e', 'cycles', 'of', 'heating', 'and', 'cooling', ',', 'strachan', 'notes', '.', 'still
', ',', '``', 'it', 'was', 'a', 'very', 'creative', 'idea', ',', 'and', 'their', 'work
', 'is', 'a', 'proof', 'of', 'concept', 'that', 'there', 'could', 'be', 'some', 'potent
ial', 'with', 'this', 'approach', '.', '``', '\\r\\n\\r\\nanother', 'group', 'investiga
ted', 'ferroelectric', 'field-effect', 'transistors', '(', 'fefets', ')', '.', 'study',
'lead', 'author', 'khandker', 'akif', 'aabrar', 'at', 'the', 'university', 'of', 'notre
', 'dame', 'explained', 'that', 'fefets', 'store', 'data', 'in', 'the', 'form', 'of', '
electric', 'polarization', 'within', 'each', 'transistor.\\r\\n\\r\\na', 'challenge', '
fefets', 'face', 'is', 'whether', 'they', 'can', 'still', 'display', 'the', 'analog', '
behavior', 'valuable', 'to', 'ai', 'applications', 'as', 'they', 'scale', 'down', ',',
'or', 'whether', 'they', 'will', 'abruptly', 'switch', 'to', 'a', 'binary', 'mode', 'wh
ere', 'they', 'only', 'store', 'one', 'bit', 'of', 'information', ',', 'with', 'the', '
polarization', 'either', 'one', 'state', 'or', 'the', 'other.\\r\\n\\r\\n', "''", 'the
', 'strength', 'of', 'this', 'team\\', "'s", 'work', 'is', 'in', 'their', 'insight', 'i
nto', 'the', 'materials', 'involved', ',', "''", 'says', 'strachan', ',', 'who', 'did',
'not', 'take', 'part', 'in', 'this', 'research', '.', '``', 'a', 'ferroelectric', 'mate
rial', 'can', 'be', 'thought', 'of', 'as', 'a', 'block', 'made', 'of', 'many', 'little
', 'domains', ',', 'just', 'as', 'a', 'ferromagnet', 'can', 'be', 'thought', 'up', 'as
', 'up', 'and', 'down', 'domains', '.', 'for', 'the', 'analog', 'behavior', 'they', 'de
sire', ',', 'they', 'want', 'all', 'these', 'domains', 'to', 'slowly', 'align', 'either
', 'up', 'or', 'down', 'in', 'response', 'to', 'an', 'applied', 'electric', 'field',
',', 'and', 'not', 'get', 'a', 'runaway', 'process', 'where', 'they', 'all', 'go', 'up
', 'or', 'down', 'at', 'once', '.', 'so', 'they', 'physically', 'broke', 'up', 'their',
'ferroelectric', 'superlattice', 'structure', 'with', 'multiple', 'dielectric', 'layers
', 'to', 'reduce', 'this', 'runaway', 'process', '.', '``', '\\r\\n\\r\\nthe', 'system
', 'achieved', 'a', '94.1', '%', 'online', 'learning', 'accuracy', ',', 'which', 'compa
red', 'very', 'well', 'against', 'other', 'fefet', 'and', 'rram', 'technologies', ',',
'findings', 'that', 'scientists', 'detailed', 'on', 'dec.', '14', 'at', 'the', 'iedm',
'conference', '.', 'strachan', 'notes', 'future', 'research', 'can', 'seek', 'to', 'opt
imize', 'properties', 'such', 'as', 'current', 'levels.\\r\\n\\r\\na', 'novel', 'microc
hip', 'from', 'scientists', 'in', 'japan', 'and', 'taiwan', 'made', 'using', 'c-axis-al
igned', 'crystalline', 'indium', 'gallium', 'zinc', 'oxide', '.', 'study', 'co-author',
'satoru', 'ohshita', 'at', 'semiconductor', 'energy', 'laboratory', 'co.', 'in', 'japan
', 'notes', 'their', 'oxide', 'semiconductor', 'field-effect', 'transistors', '(', 'osf
ets', ')', 'displayed', 'ultra-low-current', 'operations', 'below', '1', 'nano-ampere',
'per', 'cell', 'and', 'operation', 'efficiencies', 'of', '143.9', 'trillion', 'operatio
ns', 'per', 'second', 'per', 'watt', ',', 'the', 'best', 'reported', 'to', 'date', 'in
', 'analog', 'ai', 'chips', ',', 'findings', 'detailed', 'on', 'dec.', '14', 'at', 'the
', 'iedm', 'conference', '.', '``', 'these', 'are', 'extremely', 'low-current', 'device
s', ',', "''", 'strachan', 'says', '.', '``', 'since', 'the', 'currents', 'needed', 'ar
e', 'so', 'low', ',', 'you', 'can', 'make', 'circuit', 'blocks', 'larger\\xe2\\x80\\x94
they', 'get', 'arrays', 'of', '512', 'by', '512', 'memory', 'cells', ',', 'whereas', 't
he', 'typical', 'numbers', 'for', 'rram', 'are', 'more', 'like', '100', 'by', '100.', '
that\\', "'s", 'a', 'big', 'win', ',', 'since', 'larger', 'blocks', 'get', 'a', 'quadra
tic', 'advantage', 'in', 'the', 'weights', 'they', 'store', '.', '``', 'when', 'the', '
osfets', 'are', 'combined', 'with', 'capacitors', ',', 'they', 'can', 'retain', 'inform
ation', 'with', 'more', 'than', '90', '%', 'accuracy', 'for', '30', 'hours', '.', '``',
'that', 'could', 'be', 'a', 'long', 'enough', 'time', 'to', 'move', 'that', 'informatio
```

```
n', 'to', 'some', 'less', 'volatile', 'technology\\xe2\\x80\\x94tens', 'of', 'hours', '
of', 'retention', 'is', 'not', 'a', 'dealbreaker', ',', "''", 'strachan', 'says', '.',
'all', 'in', 'all', ',', '``', 'these', 'new', 'technologies', 'that', 'researchers', '
are', 'exploring', 'are', 'all', 'proof', 'of', 'concept', 'cases', 'that', 'raise', 'n
ew', 'questions', 'about', 'challenges', 'they', 'may', 'face', 'in', 'their', 'future
', ',', "''", 'strachan', 'says', '.', '``', 'they', 'also', 'show', 'a', 'path', 'to',
'the', 'foundry', ',', 'which', 'they', 'need', 'for', 'high-volume', ',', 'low-cost',
'commercial', 'products.\\xe2\\x80\\x9d\\r\\n\\r\\n2021', 'was', 'the', 'year', 'in', '
which', 'the', 'wonders', 'of', 'artificial', 'intelligence', 'stopped', 'being', 'a',
'story', '.', 'which', 'is', 'not', 'to', 'say', 'that', 'ieee', 'spectrum', "didn\\'
t", 'cover', 'ai\\xe2\\x80\\x94we', 'covered', 'the', 'heck', 'out', 'of', 'it', '.', '
but', 'we', 'all', 'know', 'that', 'deep', 'learning', 'can', 'do', 'wondrous', 'things
', 'and', 'that', 'it\\', "'s", 'being', 'rapidly', 'incorporated', 'into', 'many', 'in
dustries', ';', 'that\\', "'s", 'yesterday\\', "'s", 'news', '.', 'many', 'of', 'this',
'year\\', "'s", 'top', 'articles', 'grappled', 'with', 'the', 'limits', 'of', 'deep', '
learning', '(', 'today\\', "'s", 'dominant', 'strand', 'of', 'ai', ')', 'and', 'spotlig
hted', 'researchers', 'seeking', 'new', 'paths.\\r\\n\\r\\nhere', 'are', 'the', '10', '
most', 'popular', 'ai', 'articles', 'that', 'spectrum', 'published', 'in', '2021', ',',
'ranked', 'by', 'the', 'amount', 'of', 'time', 'people', 'spent', 'reading', 'them',
'.', 'several', 'came', 'from', 'spectrum\\', "'s", 'october', '2021', 'special', 'issu
e', 'on', 'ai', ',', 'the', 'great', 'ai', 'reckoning.\\r\\n\\r\\n1', '.', 'deep', 'lea
rning\\', "'s", 'diminishing', 'returns', ':', 'mit\\', "'s", 'neil', 'thompson', 'and
', 'several', 'of', 'his', 'collaborators', 'captured', 'the', 'top', 'spot', 'with', '
a', 'thoughtful', 'feature', 'article', 'about', 'the', 'computational', 'and', 'energy
', 'costs', 'of', 'training', 'deep', 'learning', 'systems', '.', 'they', 'analyzed', '
the', 'improvements', 'of', 'image', 'classifiers', 'and', 'found', 'that', '``', 'to',
'halve', 'the', 'error', 'rate', ',', 'you', 'can', 'expect', 'to', 'need', 'more', 'th
an', '500', 'times', 'the', 'computational', 'resources', '.', "''", 'they', 'wrote',
':', '``', 'faced', 'with', 'skyrocketing', 'costs', ',', 'researchers', 'will', 'eithe
r', 'have', 'to', 'come', 'up', 'with', 'more', 'efficient', 'ways', 'to', 'solve', 'th
ese', 'problems', ',', 'or', 'they', 'will', 'abandon', 'working', 'on', 'these', 'prob
lems', 'and', 'progress', 'will', 'languish', '.', "''", 'their', 'article', "isn\\'t",
'a', 'total', 'downer', ',', 'though', '.', 'they', 'ended', 'with', 'some', 'promising
', 'ideas', 'for', 'the', 'way', 'forward.\\r\\n\\r\\n2', '.', '15', 'graphs', 'you', '
need', 'to', 'see', 'to', 'understand', 'ai', 'in', '2021', ':', 'every', 'year', ',',
'the', 'ai', 'index', 'drops', 'a', 'massive', 'load', 'of', 'data', 'into', 'the', 'co
nversation', 'about', 'ai', '.', 'in', '2021', ',', 'the', 'index\\', "'s", 'diligent',
'curators', 'presented', 'a', 'global', 'perspective', 'on', 'academia', 'and', 'indust
ry', ',', 'taking', 'care', 'to', 'highlight', 'issues', 'with', 'diversity', 'in', 'th
e', 'ai', 'workforce', 'and', 'ethical', 'challenges', 'of', 'ai', 'applications', '.',
'i', ',', 'your', 'humble', 'ai', 'editor', ',', 'then', 'curated', 'that', 'massive',
'amount', 'of', 'curated', 'data', ',', 'boiling', '222', 'pages', 'of', 'report', 'dow
n', 'into', '15', 'graphs', 'covering', 'jobs', ',', 'investments', ',', 'and', 'more',
'.', 'you\\', "'re", 'welcome.\\r\\n\\r\\n3', '.', 'how', 'deepmind', 'is', 'reinventin
g', 'the', 'robot', ':', 'deepmind', ',', 'the', 'london-based', 'alphabet', 'subsidiar
y', ',', 'has', 'been', 'behind', 'some', 'of', 'the', 'most', 'impressive', 'feats', '
of', 'ai', 'in', 'recent', 'years', ',', 'including', 'breakthrough', 'work', 'on', 'pr
otein', 'folding', 'and', 'the', 'alphago', 'system', 'that', 'beat', 'a', 'grandmaster
', 'at', 'the', 'ancient', 'game', 'of', 'go', '.', 'so', 'when', 'deepmind\\', "'s", '
head', 'of', 'robotics', 'raia', 'hadsell', 'says', 'she\\', "'s", 'tackling', 'the', '
long-standing', 'ai', 'problem', 'of', 'catastrophic', 'forgetting', 'in', 'an', 'attem
pt', 'to', 'build', 'multi-talented', 'and', 'adaptable', 'robots', ',', 'people', 'pay
', 'attention.\\r\\n\\r\\n4', '.', 'the', 'turbulent', 'past', 'and', 'uncertain', 'fut
ure', 'of', 'artificial', 'intelligence', ':', 'this', 'feature', 'article', 'served',
'as', 'the', 'introduction', 'to', 'spectrum\\', "'s", 'special', 'report', 'on', 'ai',
',', 'telling', 'the', 'story', 'of', 'the', 'field', 'from', '1956', 'to', 'present',
'day', 'while', 'also', 'cueing', 'up', 'the', 'other', 'articles', 'in', 'the', 'speci
al', 'issue', '.', 'if', 'you', 'want', 'to', 'understand', 'how', 'we', 'got', 'here',
',', 'this', 'is', 'the', 'article', 'for', 'you', '.', 'it', 'pays', 'special', 'atten
tion', 'to', 'past', 'feuds', 'between', 'the', 'symbolists', 'who', 'bet', 'on', 'expe
rt', 'systems', 'and', 'the', 'connectionists', 'who', 'invented', 'neural', 'networks
', ',', 'and', 'looks', 'forward', 'to', 'the', 'possibilities', 'of', 'hybrid', 'neuro
-symbolic', 'systems.\\r\\n\\r\\n5', '.', 'andrew', 'ng', 'x-rays', 'the', 'ai', 'hype
', ':', 'this', 'short', 'article', 'relayed', 'an', 'anecdote', 'from', 'a', 'zoom', '
q', '&', 'a', 'session', 'with', 'ai', 'pioneer', 'andrew', 'ng', ',', 'who', 'was', 'd
eeply', 'involved', 'in', 'early', 'ai', 'efforts', 'at', 'google', 'brain', 'and', 'ba
idu', 'and', 'now', 'leads', 'a', 'company', 'called', 'landing', 'ai', '.', 'ng', 'spo
ke', 'about', 'an', 'ai', 'system', 'developed', 'at', 'stanford', 'university', 'that
', 'could', 'spot', 'pneumonia', 'in', 'chest', 'x-rays', ',', 'even', 'outperforming',
'radiologists', '.', 'but', 'there', 'was', 'a', 'twist', 'to', 'the', 'story.\\r\\n\\
r\\n6', '.', 'openai\\', "'s", 'gpt-3', 'speaks', '!', '(', 'kindly', 'disregard', 'tox
ic', 'language', ')', ':', 'when', 'the', 'san', 'francisco-based', 'ai', 'lab', 'opena
i', 'unveiled', 'the', 'language-generating', 'system', 'gpt-3', 'in', '2020', ',', 'th
e', 'first', 'reaction', 'of', 'the', 'ai', 'community', 'was', 'awe', '.', 'gpt-3', 'c
ould', 'generate', 'fluid', 'and', 'coherent', 'text', 'on', 'any', 'topic', 'and', 'in
', 'any', 'style', 'when', 'given', 'the', 'smallest', 'of', 'prompts', '.', 'but', 'it
', 'has', 'a', 'dark', 'side', '.', 'trained', 'on', 'text', 'from', 'the', 'internet',
',', 'it', 'learned', 'the', 'human', 'biases', 'that', 'are', 'all', 'too', 'prevalent
', 'in', 'certain', 'portions', 'of', 'the', 'online', 'world', ',', 'and', 'can', 'the
refore', 'has', 'an', 'awful', 'habit', 'of', 'unexpectedly', 'spewing', 'out', 'toxic
', 'language', '.', 'your', 'humble', 'ai', 'editor', '(', 'again', ',', 'that\\', "'
s", 'me', ')', 'got', 'very', 'interested', 'in', 'the', 'companies', 'that', 'are', 'r
ushing', 'to', 'integrate', 'gpt-3', 'into', 'their', 'products', ',', 'hoping', 'to',
'use', 'it', 'for', 'such', 'applications', 'as', 'customer', 'support', ',', 'online',
'tutoring', ',', 'mental', 'health', 'counseling', ',', 'and', 'more', '.', 'i', 'wante
d', 'to', 'know', ':', 'if', 'you\\', "'re", 'going', 'to', 'employ', 'an', 'ai', 'trol
l', ',', 'how', 'do', 'you', 'prevent', 'it', 'from', 'insulting', 'and', 'alienating',
'your', 'customers', '?', '\\r\\n\\r\\n7', '.', 'fast', ',', 'efficient', 'neural', 'ne
tworks', 'copy', 'dragonfly', 'brains', ':', 'what', 'do', 'dragonfly', 'brains', 'have
', 'to', 'do', 'with', 'missile', 'defense', '?', 'ask', 'frances', 'chance', 'of', 'sa
ndia', 'national', 'laboratories', ',', 'who', 'studies', 'how', 'dragonflies', 'effici
ently', 'use', 'their', 'roughly', '1', 'million', 'neurons', 'to', 'hunt', 'and', 'cap
ture', 'aerial', 'prey', 'with', 'extraordinary', 'precision', '.', 'her', 'work', 'is
', 'an', 'interesting', 'contrast', 'to', 'research', 'labs', 'building', 'neural', 'ne
tworks', 'of', 'ever-increasing', 'size', 'and', 'complexity', '(', 'recall', '#', '1',
'on', 'this', 'list', ')', '.', 'she', 'writes', ':', '``', 'by', 'harnessing', 'the',
'speed', ',', 'simplicity', ',', 'and', 'efficiency', 'of', 'the', 'dragonfly', 'nervou
s', 'system', ',', 'we', 'aim', 'to', 'design', 'computers', 'that', 'perform', 'these
', 'functions', 'faster', 'and', 'at', 'a', 'fraction', 'of', 'the', 'power', 'that', '
conventional', 'systems', 'consume.', "''", '\\r\\n\\r\\n8', '.', 'deep', 'learning', "
isn\\'t", 'deep', 'enough', 'unless', 'it', 'copies', 'from', 'the', 'brain', ':', 'in
', 'a', 'former', 'life', ',', 'jeff', 'hawkins', 'invented', 'the', 'palmpilot', 'and
', 'ushered', 'in', 'the', 'smartphone', 'era', '.', 'these', 'days', ',', 'at', 'the',
'machine', 'intelligence', 'company', 'numenta', ',', 'he\\', "'s", 'investigating', 't
he', 'basis', 'of', 'intelligence', 'in', 'the', 'human', 'brain', 'and', 'hoping', 'to
', 'usher', 'in', 'a', 'new', 'era', 'of', 'artificial', 'general', 'intelligence',
'.', 'this', 'q', '&', 'a', 'with', 'hawkins', 'covers', 'some', 'of', 'his', 'most', '
controversial', 'ideas', ',', 'including', 'his', 'conviction', 'that', 'superintellige
nt', 'ai', "doesn\\'t", 'pose', 'an', 'existential', 'threat', 'to', 'humanity', 'and',
'his', 'contention', 'that', 'consciousness', "isn\\'t", 'really', 'such', 'a', 'hard',
'problem.\\r\\n\\r\\n9', '.', 'the', 'algorithms', 'that', 'make', 'instacart', 'roll',
':', 'it\\', "'s", 'always', 'fun', 'for', 'spectrum', 'readers', 'to', 'get', 'an', 'i
nsider\\', "'s", 'look', 'at', 'the', 'tech', 'companies', 'that', 'enable', 'our', 'li
ves', '.', 'engineers', 'sharath', 'rao', 'and', 'lily', 'zhang', 'of', 'instacart',
',', 'the', 'grocery', 'shopping', 'and', 'delivery', 'company', ',', 'explain', 'that
', 'the', 'company\\', "'s", 'ai', 'infrastructure', 'has', 'to', 'predict', 'the', 'av
ailability', 'of', '``', 'the', 'products', 'in', 'nearly', '40,000', 'grocery', 'store
s\\xe2\\x80\\x94billions', 'of', 'different', 'data', 'points', ',', "''", 'while', 'al
so', 'suggesting', 'replacements', ',', 'predicting', 'how', 'many', 'shoppers', 'will
', 'be', 'available', 'to', 'work', ',', 'and', 'efficiently', 'grouping', 'orders', 'a
nd', 'delivery', 'routes.\\r\\n\\r\\n10', '.', '7', 'revealing', 'ways', 'ais', 'fail',
':', 'everyone', 'loves', 'a', 'list', ',', 'right', '?', 'after', 'all', ',', 'here',
'we', 'are', 'together', 'at', 'item', '#', '10', 'on', 'this', 'list', '.', 'spectrum
', 'contributor', 'charles', 'choi', 'pulled', 'together', 'this', 'entertaining', 'lis
t', 'of', 'failures', 'and', 'explained', 'what', 'they', 'reveal', 'about', 'the', 'we
aknesses', 'of', 'today\\', "'s", 'ai', '.', 'the', 'cartoons', 'of', 'robots', 'gettin
g', 'themselves', 'into', 'trouble', 'are', 'a', 'nice', 'bonus.\\r\\n\\r\\nso', 'there
', 'you', 'have', 'it', '.', 'keep', 'reading', 'ieee', 'spectrum', 'to', 'see', 'what
', 'happens', 'next', '.', 'will', '2022', 'be', 'the', 'year', 'in', 'which', 'researc
hers', 'figure', 'out', 'solutions', 'to', 'some', 'of', 'the', 'knotty', 'problems', '
we', 'covered', 'in', 'the', 'year', 'that\\', "'s", 'now', 'ending', '?', 'will', 'the
y', 'solve', 'algorithmic', 'bias', ',', 'put', 'an', 'end', 'to', 'catastrophic', 'for
getting', ',', 'and', 'find', 'ways', 'to', 'improve', 'performance', 'without', 'busti
```

5- Apply the sentence tokenization process to the text and store in sentTokens.

In [10]:

```
sentTokens = sent_tokenize(text)
```

In [11]:

```
readLines = []
with open (data_path, "rb") as f:
    readLines.append(str(f.readlines()))
```

a. How is readlines() different from sentence tokenizer.

The readlines() method returns a list containing each line in the file as a list item. It splits input using "\n" new line character

The sent_tokenize() marks the end and beginning of sentence at what characters and punctuation and returns each sentences from the given input.

b. How many tokens are there now?

In [12]:

```
print("Number of splits using sent_tokenize():", len(sentTokens))
print("Number of splits using readlines():", len(readLines[0]))
```

```
Number of splits using sent_tokenize(): 76
Number of splits using readlines(): 14183
```

c. Print the tokens and compare them with the readlines.

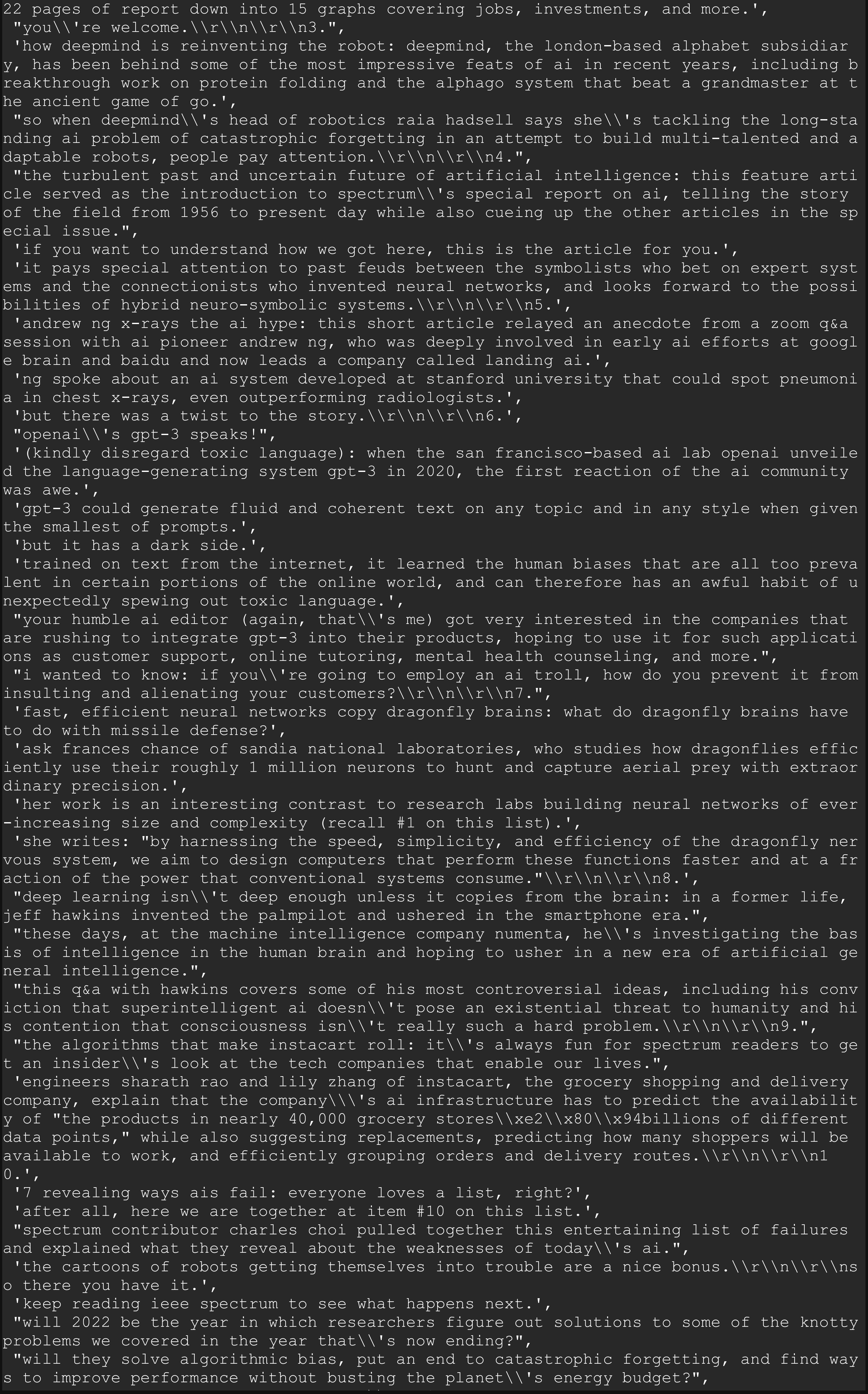
In [13]:

```
sentTokens
```

Out[13]:

```
["b'ai is undoubtedly one of the biggest tech trends at the moment, and during 2021 it
will become an even more valuable tool for helping us to interpret and understand the w
orld around us.",
 'the volume of data we are collecting on health-care, infection rates, and the success
of measures we take to prevent the spread of infection will continue to increase.',
 'this means that machine learning algorithms will become better informed and increasin
gly sophisticated in the solutions they uncover for us.',
 '\\r\\nfrom computer vision systems monitoring the capacity of public areas to analyzi
ng the interactions uncovered through contact tracing initiatives, self-learning algori
thms will spot connections and insights that would go unnoticed by manual human analysi
s.',
 'they will help us predict demand for services from hospitals and other healthcare pro
viders, and allow administrators to make better decisions about when and where to deplo
y resources.',
 '\\r\\nfor business, the challenge will be to understand the changing patterns of cust
omer behavior.',
 'more human activity will take place online \\xe2\\x80\\x93 from shopping and socializ
ing to virtual working environments, meetings, and recruitment.',
 'during 2021 we can expect the tools we use to analyze these behavioral shifts to beco
me more sophisticated and increasingly fit the budget and infrastructure requirements o
f more and more organizations.',
 '\\r\\nsome of the best circuits to drive ai in the future may be analog, not digital,
and research teams around the world are increasingly developing new devices to support
such analog ai.\\r\\n\\r\\nthe most basic computation in the deep neural networks drivi
ng the current explosion in ai is the multiply-accumulate (mac) operation.',
 'deep neural networks are composed of layers of artificial neurons, and in mac operati
ons, the output of each one of these layers is multiplied by the values of the strength
s or "weights" of their connections to the next layer, which then sums up these contrib
utions.\\r\\n\\r\\nmodern computers have digital components devoted to mac operations,
but analog circuits theoretically can perform these computations for orders of magnitud
e less energy.',
 'this strategy\\xe2\\x80\\x94known as analog ai, compute-in-memory or processing-in-me
mory\\xe2\\x80\\x94often performs these multiply-accumulate operations using non-volati
le memory devices such as flash, magnetoresistive ram (mram), resistive ram (rram), pha
se-change memory (pcm) and even more esoteric technologies.\\r\\n\\r\\none team in kore
a, however, is exploring neural networks based on praseodymium calcium manganese oxide
electrochemical ram (ecram) devices, which act like miniature batteries, storing data i
n the form of changes in their conductance.',
 'study lead author chuljun lee at the pohang university of science and technology in k
orea notes that neural network hardware often has different demands during training ver
sus during applications.',
 'for instance, low energy barriers help neural networks learn quickly, but high energy
barriers help them retain what they learned for use during applications.\\r\\n\\r\\n"he
ating up their devices almost 100 degrees c warmer during training brought out the char
acteristics that are good for training," says electrical engineer john paul strachan, h
ead of the peter gr\\xc3\\xbcnberg institute for neuromorphic compute nodes at the j\\x
c3\\xbclich research center in germany, who did not participate in this study.',
 '"when it cooled down, they got the advantages of longer retention and lower current o
peration.',
 'by just adjusting one knob, heat, they could see improvements on multiple dimensions
```

```
of computing."',
 'the researchers detailed their findings at the annual ieee international electron dev
ices meeting (iedm) in san francisco on dec. 14.\\r\\n\\r\\none key question this work
faces is what kind of deterioration this ecram may face after multiple cycles of heatin
g and cooling, strachan notes.',
 'still, "it was a very creative idea, and their work is a proof of concept that there
could be some potential with this approach.',
 '"\\r\\n\\r\\nanother group investigated ferroelectric field-effect transistors (fefet
s).',
 'study lead author khandker akif aabrar at the university of notre dame explained that
fefets store data in the form of electric polarization within each transistor.\\r\\n\\
r\\na challenge fefets face is whether they can still display the analog behavior valua
ble to ai applications as they scale down, or whether they will abruptly switch to a bi
nary mode where they only store one bit of information, with the polarization either on
e state or the other.\\r\\n\\r\\n"the strength of this team\\\'s work is in their insig
ht into the materials involved," says strachan, who did not take part in this researc
h.',
 '"a ferroelectric material can be thought of as a block made of many little domains, j
ust as a ferromagnet can be thought up as up and down domains.',
 'for the analog behavior they desire, they want all these domains to slowly align eith
er up or down in response to an applied electric field, and not get a runaway process w
here they all go up or down at once.',
 'so they physically broke up their ferroelectric superlattice structure with multiple
dielectric layers to reduce this runaway process.',
 '"\\r\\n\\r\\nthe system achieved a 94.1% online learning accuracy, which compared ver
y well against other fefet and rram technologies, findings that scientists detailed on
dec. 14 at the iedm conference.',
 'strachan notes future research can seek to optimize properties such as current level
s.\\r\\n\\r\\na novel microchip from scientists in japan and taiwan made using c-axis-a
ligned crystalline indium gallium zinc oxide.',
 'study co-author satoru ohshita at semiconductor energy laboratory co. in japan notes
their oxide semiconductor field-effect transistors (osfets) displayed ultra-low-current
operations below 1 nano-ampere per cell and operation efficiencies of 143.9 trillion op
erations per second per watt, the best reported to date in analog ai chips, findings de
tailed on dec. 14 at the iedm conference.',
 '"these are extremely low-current devices," strachan says.',
 '"since the currents needed are so low, you can make circuit blocks larger\\xe2\\x80\\
x94they get arrays of 512 by 512 memory cells, whereas the typical numbers for rram are
more like 100 by 100. that\\\'s a big win, since larger blocks get a quadratic advantag
e in the weights they store.',
 '"when the osfets are combined with capacitors, they can retain information with more
than 90% accuracy for 30 hours.',
 '"that could be a long enough time to move that information to some less volatile tech
nology\\xe2\\x80\\x94tens of hours of retention is not a dealbreaker," strachan says.',
 'all in all, "these new technologies that researchers are exploring are all proof of c
oncept cases that raise new questions about challenges they may face in their future,"
strachan says.',
 '"they also show a path to the foundry, which they need for high-volume, low-cost comm
ercial products.\\xe2\\x80\\x9d\\r\\n\\r\\n2021 was the year in which the wonders of ar
tificial intelligence stopped being a story.',
 "which is not to say that ieee spectrum didn\\'t cover ai\\xe2\\x80\\x94we covered the
heck out of it.",
 "but we all know that deep learning can do wondrous things and that it\\'s being rapid
ly incorporated into many industries; that\\'s yesterday\\'s news.",
 "many of this year\\'s top articles grappled with the limits of deep learning (toda
y\\'s dominant strand of ai) and spotlighted researchers seeking new paths.\\r\\n\\r\\n
here are the 10 most popular ai articles that spectrum published in 2021, ranked by the
amount of time people spent reading them.",
 "several came from spectrum\\'s october 2021 special issue on ai, the great ai reckoni
ng.\\r\\n\\r\\n1.",
 "deep learning\\'s diminishing returns: mit\\'s neil thompson and several of his colla
borators captured the top spot with a thoughtful feature article about the computationa
l and energy costs of training deep learning systems.",
 'they analyzed the improvements of image classifiers and found that "to halve the erro
r rate, you can expect to need more than 500 times the computational resources."',
 'they wrote: "faced with skyrocketing costs, researchers will either have to come up w
ith more efficient ways to solve these problems, or they will abandon working on these
problems and progress will languish."',
 "their article isn\\'t a total downer, though.",
 'they ended with some promising ideas for the way forward.\\r\\n\\r\\n2.',
 '15 graphs you need to see to understand ai in 2021: every year, the ai index drops a
massive load of data into the conversation about ai.',
 "in 2021, the index\\'s diligent curators presented a global perspective on academia a
nd industry, taking care to highlight issues with diversity in the ai workforce and eth
ical challenges of ai applications.",
 'i, your humble ai editor, then curated that massive amount of curated data, boiling 2
```



```
22 pages of report down into 15 graphs covering jobs, investments, and more.',
 "you\\'re welcome.\\r\\n\\r\\n3.",
 'how deepmind is reinventing the robot: deepmind, the london-based alphabet subsidiar
y, has been behind some of the most impressive feats of ai in recent years, including b
reakthrough work on protein folding and the alphago system that beat a grandmaster at t
he ancient game of go.',
 "so when deepmind\\'s head of robotics raia hadsell says she\\'s tackling the long-sta
nding ai problem of catastrophic forgetting in an attempt to build multi-talented and a
daptable robots, people pay attention.\\r\\n\\r\\n4.",
 "the turbulent past and uncertain future of artificial intelligence: this feature arti
cle served as the introduction to spectrum\\'s special report on ai, telling the story
of the field from 1956 to present day while also cueing up the other articles in the sp
ecial issue.",
 'if you want to understand how we got here, this is the article for you.',
 'it pays special attention to past feuds between the symbolists who bet on expert syst
ems and the connectionists who invented neural networks, and looks forward to the possi
bilities of hybrid neuro-symbolic systems.\\r\\n\\r\\n5.',
 'andrew ng x-rays the ai hype: this short article relayed an anecdote from a zoom q&a
session with ai pioneer andrew ng, who was deeply involved in early ai efforts at googl
e brain and baidu and now leads a company called landing ai.',
 'ng spoke about an ai system developed at stanford university that could spot pneumoni
a in chest x-rays, even outperforming radiologists.',
 'but there was a twist to the story.\\r\\n\\r\\n6.',
 "openai\\'s gpt-3 speaks!",
 '(kindly disregard toxic language): when the san francisco-based ai lab openai unveile
d the language-generating system gpt-3 in 2020, the first reaction of the ai community
was awe.',
 'gpt-3 could generate fluid and coherent text on any topic and in any style when given
the smallest of prompts.',
 'but it has a dark side.',
 'trained on text from the internet, it learned the human biases that are all too preva
lent in certain portions of the online world, and can therefore has an awful habit of u
nexpectedly spewing out toxic language.',
 "your humble ai editor (again, that\\'s me) got very interested in the companies that
are rushing to integrate gpt-3 into their products, hoping to use it for such applicati
ons as customer support, online tutoring, mental health counseling, and more.",
 "i wanted to know: if you\\'re going to employ an ai troll, how do you prevent it from
insulting and alienating your customers?\\r\\n\\r\\n7.",
 'fast, efficient neural networks copy dragonfly brains: what do dragonfly brains have
to do with missile defense?',
 'ask frances chance of sandia national laboratories, who studies how dragonflies effic
iently use their roughly 1 million neurons to hunt and capture aerial prey with extraor
dinary precision.',
 'her work is an interesting contrast to research labs building neural networks of ever
-increasing size and complexity (recall #1 on this list).',
 'she writes: "by harnessing the speed, simplicity, and efficiency of the dragonfly ner
vous system, we aim to design computers that perform these functions faster and at a fr
action of the power that conventional systems consume."\\r\\n\\r\\n8.',
 "deep learning isn\\'t deep enough unless it copies from the brain: in a former life,
jeff hawkins invented the palmpilot and ushered in the smartphone era.",
 "these days, at the machine intelligence company numenta, he\\'s investigating the bas
is of intelligence in the human brain and hoping to usher in a new era of artificial ge
neral intelligence.",
 "this q&a with hawkins covers some of his most controversial ideas, including his conv
iction that superintelligent ai doesn\\'t pose an existential threat to humanity and hi
s contention that consciousness isn\\'t really such a hard problem.\\r\\n\\r\\n9.",
 "the algorithms that make instacart roll: it\\'s always fun for spectrum readers to ge
t an insider\\'s look at the tech companies that enable our lives.",
 'engineers sharath rao and lily zhang of instacart, the grocery shopping and delivery
company, explain that the company\\\'s ai infrastructure has to predict the availabilit
y of "the products in nearly 40,000 grocery stores\\xe2\\x80\\x94billions of different
data points," while also suggesting replacements, predicting how many shoppers will be
available to work, and efficiently grouping orders and delivery routes.\\r\\n\\r\\n1
0.',
 '7 revealing ways ais fail: everyone loves a list, right?',
 'after all, here we are together at item #10 on this list.',
 "spectrum contributor charles choi pulled together this entertaining list of failures
and explained what they reveal about the weaknesses of today\\'s ai.",
 'the cartoons of robots getting themselves into trouble are a nice bonus.\\r\\n\\r\\ns
o there you have it.',
 'keep reading ieee spectrum to see what happens next.',
 "will 2022 be the year in which researchers figure out solutions to some of the knotty
problems we covered in the year that\\'s now ending?",
 "will they solve algorithmic bias, put an end to catastrophic forgetting, and find way
s to improve performance without busting the planet\\'s energy budget?",
```

```
In [14]:  readLines[0]
```

```
Out[14]: '[b\'AI is undoubtedly one of the biggest tech trends at the moment, and during 2021 it
will become an even more valuable tool for helping us to interpret and understand the w
orld around us. The volume of data we are collecting on health-care, infection rates, a
nd the success of measures we take to prevent the spread of infection will continue to
increase. This means that machine learning algorithms will become better informed and i
ncreasingly sophisticated in the solutions they uncover for us. \\r\\n\', b\'From compu
ter vision systems monitoring the capacity of public areas to analyzing the interaction
s uncovered through contact tracing initiatives, self-learning algorithms will spot con
nections and insights that would go unnoticed by manual human analysis. They will help
us predict demand for services from hospitals and other healthcare providers, and allow
administrators to make better decisions about when and where to deploy resources. \\r\\
n\', b\'For business, the challenge will be to understand the changing patterns of cust
omer behavior. More human activity will take place online \\xe2\\x80\\x93 from shopping
and socializing to virtual working environments, meetings, and recruitment. During 2021
we can expect the tools we use to analyze these behavioral shifts to become more sophis
ticated and increasingly fit the budget and infrastructure requirements of more and mor
e organizations. \\r\\n\', b\'Some of the best circuits to drive AI in the future may b
e analog, not digital, and research teams around the world are increasingly developing
new devices to support such analog AI.\\r\\n\', b\'\\r\\n\', b\'The most basic computat
ion in the deep neural networks driving the current explosion in AI is the multiply-acc
umulate (MAC) operation. Deep neural networks are composed of layers of artificial neur
ons, and in MAC operations, the output of each one of these layers is multiplied by the
values of the strengths or "weights" of their connections to the next layer, which then
sums up these contributions.\\r\\n\', b\'\\r\\n\', b\'Modern computers have digital com
ponents devoted to MAC operations, but analog circuits theoretically can perform these
computations for orders of magnitude less energy. This strategy\\xe2\\x80\\x94known as
analog AI, compute-in-memory or processing-in-memory\\xe2\\x80\\x94often performs these
multiply-accumulate operations using non-volatile memory devices such as flash, magneto
resistive RAM (MRAM), resistive RAM (RRAM), phase-change memory (PCM) and even more eso
teric technologies.\\r\\n\', b\'\\r\\n\', b\'One team in Korea, however, is exploring n
eural networks based on praseodymium calcium manganese oxide electrochemical RAM (ECRA
M) devices, which act like miniature batteries, storing data in the form of changes in
their conductance. Study lead author Chuljun Lee at the Pohang University of Science an
d Technology in Korea notes that neural network hardware often has different demands du
ring training versus during applications. For instance, low energy barriers help neural
networks learn quickly, but high energy barriers help them retain what they learned for
use during applications.\\r\\n\', b\'\\r\\n\', b\'"Heating up their devices almost 100
degrees C warmer during training brought out the characteristics that are good for trai
ning," says electrical engineer John Paul Strachan, head of the Peter Gr\\xc3\\xbcnberg
Institute for Neuromorphic Compute Nodes at the J\\xc3\\xbclich Research Center in Germ
any, who did not participate in this study. "When it cooled down, they got the advantag
es of longer retention and lower current operation. By just adjusting one knob, heat, t
hey could see improvements on multiple dimensions of computing." The researchers detail
ed their findings at the annual IEEE International Electron Devices Meeting (IEDM) in S
an Francisco on Dec. 14.\\r\\n\', b\'\\r\\n\', b\'One key question this work faces is w
hat kind of deterioration this ECRAM may face after multiple cycles of heating and cool
ing, Strachan notes. Still, "it was a very creative idea, and their work is a proof of
concept that there could be some potential with this approach."\\r\\n\', b\'\\r\\n\',
b\'Another group investigated ferroelectric field-effect transistors (FEFETs). Study le
ad author Khandker Akif Aabrar at the University of Notre Dame explained that FEFETs st
ore data in the form of electric polarization within each transistor.\\r\\n\', b\'\\r\\
n\', b\'A challenge FEFETs face is whether they can still display the analog behavior v
aluable to AI applications as they scale down, or whether they will abruptly switch to
a binary mode where they only store one bit of information, with the polarization eithe
r one state or the other.\\r\\n\', b\'\\r\\n\', b\'"The strength of this team\\\'s work
is in their insight into the materials involved," says Strachan, who did not take part
in this research. "A ferroelectric material can be thought of as a block made of many l
ittle domains, just as a ferromagnet can be thought up as up and down domains. For the
analog behavior they desire, they want all these domains to slowly align either up or d
own in response to an applied electric field, and not get a runaway process where they
all go up or down at once. So they physically broke up their ferroelectric superlattice
structure with multiple dielectric layers to reduce this runaway process."\\r\\n\',
b\'\\r\\n\', b\'The system achieved a 94.1% online learning accuracy, which compared ve
ry well against other FEFET and RRAM technologies, findings that scientists detailed on
Dec. 14 at the IEDM conference. Strachan notes future research can seek to optimize pro
perties such as current levels.\\r\\n\', b\'\\r\\n\', b\'A novel microchip from scienti
sts in Japan and Taiwan made using c-axis-aligned crystalline indium gallium zinc oxid
e. Study co-author Satoru Ohshita at Semiconductor Energy Laboratory Co. in Japan notes
their oxide semiconductor field-effect transistors (OSFETs) displayed ultra-low-current
operations below 1 nano-ampere per cell and operation efficiencies of 143.9 trillion op
erations per second per watt, the best reported to date in analog AI chips, findings de
tailed on Dec. 14 at the IEDM conference. "These are extremely low-current devices," St
```

```
rachan says. "Since the currents needed are so low, you can make circuit blocks large
r\\xe2\\x80\\x94they get arrays of 512 by 512 memory cells, whereas the typical numbers
for RRAM are more like 100 by 100. That\\\'s a big win, since larger blocks get a quadr
atic advantage in the weights they store. "When the OSFETs are combined with capacitor
s, they can retain information with more than 90% accuracy for 30 hours. "That could be
a long enough time to move that information to some less volatile technology\\xe2\\x8
0\\x94tens of hours of retention is not a dealbreaker," Strachan says. All in all, "the
se new technologies that researchers are exploring are all proof of concept cases that
raise new questions about challenges they may face in their future," Strachan says. "Th
ey also show a path to the foundry, which they need for high-volume, low-cost commercia
l products.\\xe2\\x80\\x9d\\r\\n\', b\'\\r\\n\', b"2021 was the year in which the wonde
rs of artificial intelligence stopped being a story. Which is not to say that IEEE Spec
trum didn\'t cover AI\\xe2\\x80\\x94we covered the heck out of it. But we all know that
deep learning can do wondrous things and that it\'s being rapidly incorporated into man
y industries; that\'s yesterday\'s news. Many of this year\'s top articles grappled wit
h the limits of deep learning (today\'s dominant strand of AI) and spotlighted research
ers seeking new paths.\\r\\n", b\'\\r\\n\', b"Here are the 10 most popular AI articles
that Spectrum published in 2021, ranked by the amount of time people spent reading the
m. Several came from Spectrum\'s October 2021 special issue on AI, The Great AI Reckoni
ng.\\r\\n", b\'\\r\\n\', b\'1. Deep Learning\\\'s Diminishing Returns: MIT\\\'s Neil Th
ompson and several of his collaborators captured the top spot with a thoughtful feature
article about the computational and energy costs of training deep learning systems. The
y analyzed the improvements of image classifiers and found that "to halve the error rat
e, you can expect to need more than 500 times the computational resources." They wrote:
"Faced with skyrocketing costs, researchers will either have to come up with more effic
ient ways to solve these problems, or they will abandon working on these problems and p
rogress will languish." Their article isn\\\'t a total downer, though. They ended with
some promising ideas for the way forward.\\r\\n\', b\'\\r\\n\', b"2. 15 Graphs You Need
to See to Understand AI in 2021: Every year, The AI Index drops a massive load of data
into the conversation about AI. In 2021, the Index\'s diligent curators presented a glo
bal perspective on academia and industry, taking care to highlight issues with diversit
y in the AI workforce and ethical challenges of AI applications. I, your humble AI edit
or, then curated that massive amount of curated data, boiling 222 pages of report down
into 15 graphs covering jobs, investments, and more. You\'re welcome.\\r\\n", b\'\\r\\
n\', b"3. How DeepMind Is Reinventing the Robot: DeepMind, the London-based Alphabet su
bsidiary, has been behind some of the most impressive feats of AI in recent years, incl
uding breakthrough work on protein folding and the AlphaGo system that beat a grandmast
er at the ancient game of Go. So when DeepMind\'s head of robotics Raia Hadsell says sh
e\'s tackling the long-standing AI problem of catastrophic forgetting in an attempt to
build multi-talented and adaptable robots, people pay attention.\\r\\n", b\'\\r\\n\',
b"4. The Turbulent Past and Uncertain Future of Artificial Intelligence: This feature a
rticle served as the introduction to Spectrum\'s special report on AI, telling the stor
y of the field from 1956 to present day while also cueing up the other articles in the
special issue. If you want to understand how we got here, this is the article for you.
It pays special attention to past feuds between the symbolists who bet on expert system
s and the connectionists who invented neural networks, and looks forward to the possibi
lities of hybrid neuro-symbolic systems.\\r\\n", b\'\\r\\n\', b\'5. Andrew Ng X-Rays th
e AI Hype: This short article relayed an anecdote from a Zoom Q&A session with AI pione
er Andrew Ng, who was deeply involved in early AI efforts at Google Brain and Baidu and
now leads a company called Landing AI. Ng spoke about an AI system developed at Stanfor
d University that could spot pneumonia in chest x-rays, even outperforming radiologist
s. But there was a twist to the story.\\r\\n\', b\'\\r\\n\', b"6. OpenAI\'s GPT-3 Speak
s! (Kindly Disregard Toxic Language): When the San Francisco-based AI lab OpenAI unveil
ed the language-generating system GPT-3 in 2020, the first reaction of the AI community
was awe. GPT-3 could generate fluid and coherent text on any topic and in any style whe
n given the smallest of prompts. But it has a dark side. Trained on text from the inter
net, it learned the human biases that are all too prevalent in certain portions of the
online world, and can therefore has an awful habit of unexpectedly spewing out toxic la
nguage. Your humble AI editor (again, that\'s me) got very interested in the companies
that are rushing to integrate GPT-3 into their products, hoping to use it for such appl
ications as customer support, online tutoring, mental health counseling, and more. I wa
nted to know: If you\'re going to employ an AI troll, how do you prevent it from insult
ing and alienating your customers?\\r\\n", b\'\\r\\n\', b\'7. Fast, Efficient Neural Ne
tworks Copy Dragonfly Brains: What do dragonfly brains have to do with missile defense?
Ask Frances Chance of Sandia National Laboratories, who studies how dragonflies efficie
ntly use their roughly 1 million neurons to hunt and capture aerial prey with extraordi
nary precision. Her work is an interesting contrast to research labs building neural ne
tworks of ever-increasing size and complexity (recall #1 on this list). She writes: "By
harnessing the speed, simplicity, and efficiency of the dragonfly nervous system, we ai
m to design computers that perform these functions faster and at a fraction of the powe
r that conventional systems consume."\\r\\n\', b\'\\r\\n\', b"8. Deep Learning Isn\'t D
eep Enough Unless It Copies From the Brain: In a former life, Jeff Hawkins invented the
PalmPilot and ushered in the smartphone era. These days, at the machine intelligence co
mpany Numenta, he\'s investigating the basis of intelligence in the human brain and hop
ing to usher in a new era of artificial general intelligence. This Q&A with Hawkins cov
ers some of his most controversial ideas, including his conviction that superintelligen
```

```
t AI doesn\'t pose an existential threat to humanity and his contention that consciousn
ess isn\'t really such a hard problem.\\r\\n", b\'\\r\\n\', b\'9. The Algorithms That M
ake Instacart Roll: It\\\'s always fun for Spectrum readers to get an insider\\\'s look
at the tech companies that enable our lives. Engineers Sharath Rao and Lily Zhang of In
stacart, the grocery shopping and delivery company, explain that the company\\\'s AI in
frastructure has to predict the availability of "the products in nearly 40,000 grocery
stores\\xe2\\x80\\x94billions of different data points," while also suggesting replacem
ents, predicting how many shoppers will be available to work, and efficiently grouping
orders and delivery routes.\\r\\n\', b\'\\r\\n\', b"10. 7 Revealing Ways AIs Fail: Ever
yone loves a list, right? After all, here we are together at item #10 on this list. Spe
ctrum contributor Charles Choi pulled together this entertaining list of failures and e
xplained what they reveal about the weaknesses of today\'s AI. The cartoons of robots g
etting themselves into trouble are a nice bonus.\\r\\n", b\'\\r\\n\', b"So there you ha
ve it. Keep reading IEEE Spectrum to see what happens next. Will 2022 be the year in wh
ich researchers figure out solutions to some of the knotty problems we covered in the y
ear that\'s now ending? Will they solve algorithmic bias, put an end to catastrophic fo
rgetting, and find ways to improve performance without busting the planet\'s energy bud
```

Number of splits in given text using sent_tokenize(): 87

Number of splits in given text using readlines(): 50

sent_tokenize() uses punctuation marks as seperator

readlines() uses '\n' as seperator

6. Apply spelling correction on each word tokens and print the initial 15 misspelled tokens as well as the corrected tokens. Keep a count of corrected tokens.

In [15]:

```python
spell = SpellChecker()
misspelled_tkn = {}
corrected_words_tkn = []
temp = wordTokens
mwords = spell.unknown(wordTokens)
for w_tkn in mwords:
    corrected_tkn = spell.correction(w_tkn)
    if(w_tkn != corrected_tkn):
        corrected_words_tkn = [re.sub(r'\b'+w_tkn+'\b', corrected_tkn, word) for word in temp]
        temp = corrected_words_tkn
        misspelled_tkn[w_tkn] = corrected_tkn
top_15_misspell = list(misspelled_tkn.items())[:15]
```

In [16]:

```python
len(corrected_words_tkn)
```

Out[16]:

```
2430
```

In [17]:

```python
corrected_words_tkn[0:20]
```

Out[17]:

```
["b'ai",
 'is',
 'undoubtedly',
 'one',
 'of',
 'the',
 'biggest',
 'tech',
 'trends',
 'at',
 'the',
 'moment',
 ',',
 'and',
 'during',
 '2021',
 'it',
 'will',
 'become',
 'an']
```

In [18]:

```
print(*top_15_misspell, sep='\n')
```

```
('akif', 'akin')
('raia', 'rain')
('connectionists', 'connectionist')
('multi-talented', 'multitalented')
('alphago', 'alpha')
('numenta', 'nugent')
('co.', 'cop')
('co-author', 'coauthor')
('strachan', 'astrakhan')
('x-rays', 'rays')
('it\\', 'it')
("b'ai", 'bhai')
('index\\', 'index')
('year\\', 'years')
('superlattice', 'superlative')
```

a. Finally print all the corrected tokens

In [19]:

```
print(*list(misspelled_tkn.items()))
```

```
('akif', 'akin') ('raia', 'rain') ('connectionists', 'connectionist') ('multi-talented
', 'multitalented') ('alphago', 'alpha') ('numenta', 'nugent') ('co.', 'cop') ('co-auth
or', 'coauthor') ('strachan', 'astrakhan') ('x-rays', 'rays') ('it\\', 'it') ("b'ai", '
bhai') ('index\\', 'index') ('year\\', 'years') ('superlattice', 'superlative') ('hadse
ll', 'hansel') ('yesterday\\', 'yesterday') ('mit\\', 'mit') ("'s", 'is') ('khandker',
'handler') ('iedm', 'ied') ('symbolists', 'symbolises') ("'re", 'are') ("didn\\'t", "di
dn't") ('ferroelectric', 'neuroelectric') ('insider\\', 'insider') ('fefets', 'feets')
('chuljun', 'chulbul') ('openai', 'open') ('aabrar', 'hablar') ('c', 'i') ('he\\', 'he
') ('company\\', 'company') ('health-care', 'healthcare') ('ferromagnet', 'ferromagneti
c') ('classifiers', 'classifies') ('satoru', 'story') ("''", 'i') ('team\\', 'team') ('
ohshita', 'ashita') ('osfets', 'sets') ('learning\\', 'learning') ('mram', 'ram') ('pla
net\\', 'planet') ('long-standing', 'longstanding') ("doesn\\'t", "doesn't") ('baidu',
'haidu') ('ecram', 'cram') ('spotlighted', 'spotlight') ('you\\', 'you') ('consume.', '
consumed') ('today\\', 'today') ('rram', 'roam') ('q', 'i') ('spectrum\\', 'spectrum')
('that\\', 'that') ('ieee', 'ieve') ('dec.', 'deck') ('``', 'i') ('pcm', 'pum') ('fefet
', 'feet') ("isn\\'t", "isn't") ('sandia', 'sandra') ('deepmind', 'deeming') ('she\\',
'she') ('let\\', 'let') ('ng', 'no')
```

7. Remove stop words and punctuation characters from the corrected token list.

In [20]:

```
list_stop_words = stopwords.words("english")
filtered_tokens = [w for w in corrected_words_tkn if w not in
list_stop_words and w not in punctuation]
print(len(filtered_tokens))
```

```
1343
```

8. Stem each tokens

In [21]:

```
stemmer = SnowballStemmer("english")
stem_tkn = []
for w_tkn in filtered_tokens:
    corrected_tkn = stemmer.stem(w_tkn)
    stem_tkn.append(corrected_tkn)
```

9. Now create a vocabulary for the pre-processed text.

In [22]:

```
vocab_pre_processed = stem_tkn
print(vocab_pre_processed[:20])
```

```
["b'ai", 'undoubt', 'one', 'biggest', 'tech', 'trend', 'moment', '2021', 'becom', 'even
', 'valuabl', 'tool', 'help', 'us', 'interpret', 'understand', 'world', 'around', 'us',
'volum']
```

10. Compare and observe vocabularies generated by Q3 and Q9.

In [23]:

```
print("Length of original text = ", len(wordTokens))
print("Length of pre processed text = ", len(vocab_pre_processed))
```

```
Length of original text =  2430
Length of pre processed text =  1343
```

In [24]:

```
print(wordTokens[:50])
```

```
["b'ai", 'is', 'undoubtedly', 'one', 'of', 'the', 'biggest', 'tech', 'trends', 'at', 't
he', 'moment', ',', 'and', 'during', '2021', 'it', 'will', 'become', 'an', 'even', 'mor
e', 'valuable', 'tool', 'for', 'helping', 'us', 'to', 'interpret', 'and', 'understand',
'the', 'world', 'around', 'us', '.', 'the', 'volume', 'of', 'data', 'we', 'are', 'colle
cting', 'on', 'health-care', ',', 'infection', 'rates', ',', 'and']
```

In [25]:

```
print(vocab_pre_processed[:50])
```

```
["b'ai", 'undoubt', 'one', 'biggest', 'tech', 'trend', 'moment', '2021', 'becom', 'even
', 'valuabl', 'tool', 'help', 'us', 'interpret', 'understand', 'world', 'around', 'us',
'volum', 'data', 'collect', 'health-car', 'infect', 'rate', 'success', 'measur', 'take
', 'prevent', 'spread', 'infect', 'continu', 'increas', 'mean', 'machin', 'learn', 'alg
orithm', 'becom', 'better', 'inform', 'increas', 'sophist', 'solut', 'uncov', 'us', '\\
r\\nfrom', 'comput', 'vision', 'system', 'monitor']
```

In vocab_pre_processed, we can observe that the words have been converted to its root words, punctuations have been removed, misspelled words are corrected and stop-words have been removed. Length of original text = 2462 and Length of pre processed text = 1341