

# SirineZanina\_3LM2

November 9, 2023

## 1 Mini Project Indexation

| nom    | prenom | classe | groupe |
|--------|--------|--------|--------|
| ZANINA | Sirine | 3LM2   | TP2    |

```
[ ]: import os

# Spécifiez le chemin du répertoire où se trouvent vos documents
corpus_directory = 'C:
↳\\Users\\MSI\\Documents\\3eme_anneeCS\\S1\\Indexation\\Project1\\Corpus\\'

corpus_list = [] # Créez une liste vide pour stocker les objets

# Parcourez le répertoire pour collecter les noms de fichiers et leur contenu
for filename in os.listdir(corpus_directory):
    if filename.endswith('.txt'):
        file_path = os.path.join(corpus_directory, filename)
        with open(file_path, 'r', encoding='utf-8') as file:
            document_content = file.read()
            # Créez un objet avec les propriétés "filename" et "content"
            document_object = {"filename": filename, "content":
↳document_content}
            # Ajoutez l'objet à la liste
            corpus_list.append(document_object)

# À présent, corpus_list est une liste d'objets, chacun contenant les
↳propriétés "filename" et "content"

# Imprimez chaque objet sur une nouvelle ligne
for document_object in corpus_list:
    print(document_object)
```

```
{'filename': "L'IA dans l'industrie automobile.txt", 'content': "Les véhicules autonomes sont un exemple de l'application de l'IA dans l'industrie automobile. Ils utilisent des capteurs et des algorithmes pour naviguer sans conducteur humain."}
```

```
{'filename': "L'IA et l'automatisation du travail.txt", 'content':
"L'automatisation alimentée par l'IA peut transformer les emplois et les
industries. Il est important de comprendre son impact sur l'avenir du travail."}
{'filename': "L'éthique de l'intelligence artificielle.txt", 'content': "L'IA
soulève des questions éthiques importantes, notamment en ce qui concerne la vie
privée, la transparence et la prise de décision automatisée. Il est essentiel de
réfléchir à ces questions."}
{'filename': "Les défis de l'IA  biais et équité.txt", 'content': "L'IA peut
être sujette à des biais si les données utilisées pour former les modèles sont
biaisées. Assurer l'équité dans les systèmes d'IA est un défi majeur."}
{'filename': "Les fondements de l'apprentissage automatique.txt", 'content':
"L'apprentissage automatique est l'une des branches les plus importantes de
l'intelligence artificielle. Il repose sur des algorithmes qui permettent aux
ordinateurs d'apprendre à partir de données. Les systèmes d'apprentissage
automatique sont capables d'analyser des ensembles de données, de détecter des
schémas et de prendre des décisions basées sur ces modèles.\n\nL'apprentissage
automatique comprend diverses techniques, telles que la régression, la
classification, le clustering et la reconnaissance de motifs. Ces techniques
sont utilisées dans une variété de domaines, de la prédiction de tendances
financières à la recommandation de produits en ligne. Les algorithmes
d'apprentissage automatique sont alimentés par des données, et plus les données
sont riches et pertinentes, plus les modèles d'apprentissage automatique sont
performants."}
{'filename': "Les futurs développements de l'IA.txt", 'content': "L'IA continue
de progresser, avec des développements passionnants à venir, tels que l'IA
générale artificielle (AGI) qui pourrait rivaliser avec l'intelligence
humaine."}
{'filename': 'Les outils populaires en IA  TensorFlow et PyTorch.txt',
'content': "TensorFlow et PyTorch sont des bibliothèques populaires utilisées
par les chercheurs et les développeurs pour construire des modèles d'IA."}
{'filename': "Les pionniers de l'IA  Alan Turing et John McCarthy.txt",
'content': "Alan Turing et John McCarthy sont des figures clés de l'histoire de
l'IA. Leurs travaux ont jeté les bases de ce que nous connaissons aujourd'hui en
matière d'intelligence artificielle."}
{'filename': "Qu'est-ce que l'intelligence artificielle.txt", 'content':
"L'intelligence artificielle (IA) est un domaine passionnant de l'informatique
qui vise à créer des systèmes capables d'effectuer des tâches qui nécessitent
généralement l'intelligence humaine. Ces systèmes sont conçus pour imiter
certaines capacités humaines, telles que la vision, la compréhension du langage
naturel, l'apprentissage et la résolution de problèmes complexes. L'IA englobe
un large éventail de domaines, de la robotique aux algorithmes d'apprentissage
automatique.\n\nLes chercheurs en IA travaillent sur le développement de modèles
d'IA capables de prendre des décisions éclairées, de résoudre des problèmes
complexes et d'interagir de manière plus naturelle avec les humains. Ils
utilisent des techniques d'apprentissage automatique, de traitement du langage
naturel, de vision par ordinateur et d'apprentissage profond pour créer ces
systèmes intelligents."}
```

```
[ ]: import nltk
nltk.download("punkt")
from nltk.tokenize import word_tokenize
```

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

```
[ ]: # Créez une liste pour stocker le texte prétraité
preprocessed_corpus = []

# Parcourez chaque document dans votre corpus
for document_object in corpus_list:
    # Récupérez le contenu du document
    document_content = document_object["content"]

    # Tokenization (découpage en tokens/mots)
    tokens = word_tokenize(document_content)

    # Ajoutez les tokens au texte prétraité du document
    preprocessed_document = {
        "filename": document_object["filename"], # Nom du fichier
        "tokens": tokens # Liste des tokens du document
    }

    # Ajoutez le document prétraité à la liste du corpus prétraité
    preprocessed_corpus.append(preprocessed_document)

# À ce stade, preprocessed_corpus contient les tokens pour chaque document de
↳ votre corpus
# Le code suivant est en dehors de la boucle, il ne va imprimer que les tokens
↳ du dernier document dans corpus_list
print(tokens)
```

```
["L'intelligence", 'artificielle', '(', 'IA', ')', 'est', 'un', 'domaine',
'passionnant', 'de', 'l'informatique', 'qui', 'vise', 'à', 'créer', 'des',
'systèmes', 'capables', 'd\'effectuer', 'des', 'tâches', 'qui', 'nécessitent',
'généralement', 'l\'intelligence', 'humaine', '.', 'Ces', 'systèmes', 'sont',
'conçus', 'pour', 'imiter', 'certaines', 'capacités', 'humaines', ',', 'telles',
'que', 'la', 'vision', ',', 'la', 'compréhension', 'du', 'langage', 'naturel',
',', 'l\'apprentissage', 'et', 'la', 'résolution', 'de', 'problèmes',
'complexes', '.', 'L\'IA', 'englobe', 'un', 'large', 'éventail', 'de',
'domains', ',', 'de', 'la', 'robotique', 'aux', 'algorithmes',
'd\'apprentissage', 'automatique', '.', 'Les', 'chercheurs', 'en', 'IA',
'travaillent', 'sur', 'le', 'développement', 'de', 'modèles', 'd\'IA',
'capables', 'de', 'prendre', 'des', 'décisions', 'éclairées', ',', 'de',
'résoudre', 'des', 'problèmes', 'complexes', 'et', 'd\'interagir', 'de',
'manière', 'plus', 'naturelle', 'avec', 'les', 'humains', '.', 'Ils',
```

```
'utilisent', 'des', 'techniques', "d'apprentissage", 'automatique', ',', 'de',
'traitement', 'du', 'langage', 'naturel', ',', 'de', 'vision', 'par',
'ordinateur', 'et', "d'apprentissage", 'profond', 'pour', 'créer', 'ces',
'systèmes', 'intelligents', '.']
```

```
[ ]: from nltk.corpus import stopwords
nltk.download('stopwords')
```

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

```
[ ]: True
```

```
[ ]: stop_words = set(stopwords.words('english'))
print(stop_words)
```

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

```
[ ]: from nltk.stem import PorterStemmer
ps = PorterStemmer()
```

```
[ ]: # Créez une liste pour stocker le texte prétraité
preprocessed_corpus = []

# Parcourez chaque document dans votre corpus
for document_object in corpus_list:

    mots_formates = [word.lower() for word in document_object["content"] if
↳ word.isalpha()]
```

```

mots_sans_motsvides = [word for word in mots_formates if word not in
↳stop_words]

mots_normalise = [ps.stem(word) for word in mots_sans_motsvides]

# Créez un nouvel objet avec le nom du document et le texte prétraité
preprocessed_document = {
    "filename": document_object["filename"],
    "content": mots_normalise
}

# Ajoutez l'objet au corpus prétraité
preprocessed_corpus.append(preprocessed_document)

# À présent, preprocessed_corpus contient le texte prétraité pour chaque
↳document de votre corpus
for obj in preprocessed_corpus:
    print(obj)

```

```

{'filename': "L'IA dans l'industrie automobile.txt", 'content': ['l', 'e', 'v',
'é', 'h', 'c', 'u', 'l', 'e', 'u', 'n', 'e', 'n', 'u', 'n', 'e', 'x', 'e', 'p',
'l', 'e', 'e', 'l', 'p', 'p', 'l', 'c', 'n', 'e', 'l', 'n', 'l', 'n', 'u', 'r',
'e', 'u', 'b', 'l', 'e', 'l', 'u', 'l', 'e', 'n', 'e', 'c', 'p', 'e', 'u', 'r',
'e', 'e', 'l', 'g', 'r', 'h', 'e', 'p', 'u', 'r', 'n', 'v', 'g', 'u', 'e', 'r',
'n', 'c', 'n', 'u', 'c', 'e', 'u', 'r', 'h', 'u', 'n']}
{'filename': "L'IA et l'automatisation du travail.txt", 'content': ['l', 'u',
'n', 'l', 'e', 'n', 'é', 'e', 'p', 'r', 'l', 'p', 'e', 'u', 'r', 'n', 'f', 'r',
'e', 'r', 'l', 'e', 'e', 'p', 'l', 'e', 'l', 'e', 'n', 'u', 'r', 'e', 'l', 'e',
'p', 'r', 'n', 'e', 'c', 'p', 'r', 'e', 'n', 'r', 'e', 'n', 'p', 'c', 'u', 'r',
'l', 'v', 'e', 'n', 'r', 'u', 'r', 'v', 'l']}
{'filename': "L'éthique de l'intelligence artificielle.txt", 'content': ['l',
'u', 'l', 'è', 'v', 'e', 'e', 'q', 'u', 'e', 'n', 'é', 'h', 'q', 'u', 'e', 'p',
'r', 'n', 'e', 'n', 'e', 'n', 'e', 'n', 'c', 'e', 'q', 'u', 'c', 'n', 'c', 'e',
'r', 'n', 'e', 'l', 'v', 'e', 'p', 'r', 'v', 'é', 'e', 'l', 'r', 'n', 'p', 'r',
'e', 'n', 'c', 'e', 'e', 'l', 'p', 'r', 'e', 'e', 'é', 'c', 'n', 'u', 'é', 'e',
'l', 'e', 'e', 'e', 'n', 'e', 'l', 'e', 'r', 'é', 'f', 'l', 'é', 'c', 'h', 'r',
'à', 'c', 'e', 'q', 'u', 'e', 'n']}
{'filename': "Les défis de l'IA  biais et équité.txt", 'content': ['l', 'p',
'e', 'u', 'ê', 'r', 'e', 'u', 'j', 'e', 'e', 'à', 'e', 'b', 'l', 'e', 'n', 'n',
'é', 'e', 'u', 'l', 'é', 'e', 'p', 'u', 'r', 'f', 'r', 'e', 'r', 'l', 'e', 'è',
'l', 'e', 'n', 'b', 'é', 'e', 'u', 'r', 'e', 'r', 'l', 'é', 'q', 'u', 'é', 'n',
'l', 'e', 'è', 'e', 'e', 'u', 'n', 'é', 'f', 'j', 'e', 'u', 'r']}
{'filename': "Les fondements de l'apprentissage automatique.txt", 'content':
['l', 'p', 'p', 'r', 'e', 'n', 'g', 'e', 'u', 'q', 'u', 'e', 'e', 'l', 'u', 'n',
'e', 'e', 'b', 'r', 'n', 'c', 'h', 'e', 'l', 'e', 'p', 'l', 'u', 'p', 'r', 'n',
'e', 'e', 'l', 'n', 'e', 'l', 'l', 'g', 'e', 'n', 'c', 'e', 'r', 'f', 'c', 'e',

```

```

'l', 'l', 'e', 'l', 'r', 'e', 'p', 'e', 'u', 'r', 'e', 'l', 'g', 'r', 'h', 'e',
'q', 'u', 'p', 'e', 'r', 'e', 'e', 'n', 'u', 'x', 'r', 'n', 'e', 'u', 'r', 'p',
'p', 'r', 'e', 'n', 'r', 'e', 'à', 'p', 'r', 'r', 'e', 'n', 'n', 'é', 'e', 'l',
'e', 'è', 'e', 'p', 'p', 'r', 'e', 'n', 'g', 'e', 'u', 'q', 'u', 'e', 'n', 'c',
'p', 'b', 'l', 'e', 'n', 'l', 'e', 'r', 'e', 'e', 'n', 'e', 'b', 'l', 'e', 'e',
'n', 'n', 'é', 'e', 'e', 'é', 'e', 'c', 'e', 'r', 'e', 'c', 'h', 'é', 'e', 'e',
'p', 'r', 'e', 'n', 'r', 'e', 'e', 'é', 'c', 'n', 'b', 'é', 'e', 'u', 'r', 'c',
'e', 'è', 'l', 'e', 'l', 'p', 'p', 'r', 'e', 'n', 'g', 'e', 'u', 'q', 'u', 'e',
'c', 'p', 'r', 'e', 'n', 'v', 'e', 'r', 'e', 'e', 'c', 'h', 'n', 'q', 'u', 'e',
'e', 'l', 'l', 'e', 'q', 'u', 'e', 'l', 'r', 'é', 'g', 'r', 'e', 'n', 'l', 'c',
'l', 'f', 'c', 'n', 'l', 'e', 'c', 'l', 'u', 'e', 'r', 'n', 'g', 'e', 'l', 'r',
'e', 'c', 'n', 'n', 'n', 'c', 'e', 'e', 'f', 'c', 'e', 'e', 'c', 'h', 'n', 'q',
'u', 'e', 'n', 'u', 'l', 'é', 'e', 'n', 'u', 'n', 'e', 'v', 'r', 'é', 'é', 'e',
'n', 'e', 'e', 'l', 'p', 'r', 'é', 'c', 'n', 'e', 'e', 'n', 'n', 'c', 'e', 'f',
'n', 'n', 'c', 'è', 'r', 'e', 'à', 'l', 'r', 'e', 'c', 'n', 'n', 'e', 'p', 'r',
'u', 'e', 'n', 'l', 'g', 'n', 'e', 'l', 'e', 'l', 'g', 'r', 'h', 'e', 'p', 'p',
'r', 'e', 'n', 'g', 'e', 'u', 'q', 'u', 'e', 'n', 'l', 'e', 'n', 'é', 'p', 'r',
'e', 'n', 'n', 'é', 'e', 'e', 'p', 'l', 'u', 'l', 'e', 'n', 'n', 'é', 'e', 'n',
'r', 'c', 'h', 'e', 'e', 'p', 'e', 'r', 'n', 'e', 'n', 'e', 'p', 'l', 'u', 'l',
'e', 'è', 'l', 'e', 'p', 'p', 'r', 'e', 'n', 'g', 'e', 'u', 'q', 'u', 'e', 'n',
'p', 'e', 'r', 'f', 'r', 'n']]
{'filename': "Les futurs développements de l'IA.txt", 'content': ['l', 'c', 'n',
'n', 'u', 'e', 'e', 'p', 'r', 'g', 'r', 'e', 'e', 'r', 'v', 'e', 'c', 'e', 'é',
'v', 'e', 'l', 'p', 'p', 'e', 'e', 'n', 'p', 'n', 'n', 'n', 'à', 'v', 'e', 'n',
'r', 'e', 'l', 'q', 'u', 'e', 'l', 'g', 'é', 'n', 'é', 'r', 'l', 'e', 'r', 'f',
'c', 'e', 'l', 'l', 'e', 'g', 'q', 'u', 'p', 'u', 'r', 'r', 'r', 'v', 'l', 'e',
'r', 'v', 'e', 'c', 'l', 'n', 'e', 'l', 'l', 'g', 'e', 'n', 'c', 'e', 'h', 'u',
'n', 'e']}
{'filename': 'Les outils populaires en IA TensorFlow et PyTorch.txt',
'content': ['e', 'n', 'r', 'f', 'l', 'w', 'e', 'p', 'r', 'c', 'h', 'n', 'e',
'b', 'b', 'l', 'h', 'è', 'q', 'u', 'e', 'p', 'p', 'u', 'l', 'r', 'e', 'u', 'l',
'é', 'e', 'p', 'r', 'l', 'e', 'c', 'h', 'e', 'r', 'c', 'h', 'e', 'u', 'r', 'e',
'l', 'e', 'é', 'v', 'e', 'l', 'p', 'p', 'e', 'u', 'r', 'p', 'u', 'r', 'c', 'n',
'r', 'u', 'r', 'e', 'e', 'è', 'l', 'e']}
{'filename': "Les pionniers de l'IA Alan Turing et John McCarthy.txt",
'content': ['l', 'n', 'u', 'r', 'n', 'g', 'e', 'j', 'h', 'n', 'c', 'c', 'r',
'h', 'n', 'e', 'f', 'g', 'u', 'r', 'e', 'c', 'l', 'é', 'e', 'l', 'h', 'r', 'e',
'e', 'l', 'l', 'e', 'u', 'r', 'r', 'v', 'u', 'x', 'n', 'j', 'e', 'é', 'l', 'e',
'b', 'e', 'e', 'c', 'e', 'q', 'u', 'e', 'n', 'u', 'c', 'n', 'n', 'n', 'u', 'j',
'u', 'r', 'h', 'u', 'e', 'n', 'è', 'r', 'e', 'n', 'e', 'l', 'l', 'g', 'e', 'n',
'c', 'e', 'r', 'f', 'c', 'e', 'l', 'l', 'e']}
{'filename': "Qu'est-ce que l'intelligence artificielle.txt", 'content': ['l',
'n', 'e', 'l', 'l', 'g', 'e', 'n', 'c', 'e', 'r', 'f', 'c', 'e', 'l', 'l', 'e',
'e', 'u', 'n', 'n', 'e', 'p', 'n', 'n', 'n', 'e', 'l', 'n', 'f', 'r', 'q', 'u',
'e', 'q', 'u', 'v', 'e', 'à', 'c', 'r', 'é', 'e', 'r', 'e', 'è', 'e', 'c', 'p',
'b', 'l', 'e', 'e', 'f', 'f', 'e', 'c', 'u', 'e', 'r', 'e', 'â', 'c', 'h', 'e',
'q', 'u', 'n', 'é', 'c', 'e', 'e', 'n', 'g', 'é', 'n', 'é', 'r', 'l', 'e', 'e',
'n', 'l', 'n', 'e', 'l', 'l', 'g', 'e', 'n', 'c', 'e', 'h', 'u', 'n', 'e', 'c',

```

```
'e', 'è', 'e', 'n', 'c', 'n', 'ç', 'u', 'p', 'u', 'r', 'e', 'r', 'c', 'e', 'r',
'n', 'e', 'c', 'p', 'c', 'é', 'h', 'u', 'n', 'e', 'e', 'l', 'l', 'e', 'q', 'u',
'e', 'l', 'v', 'n', 'l', 'c', 'p', 'r', 'é', 'h', 'e', 'n', 'n', 'u', 'l', 'n',
'g', 'g', 'e', 'n', 'u', 'r', 'e', 'l', 'l', 'p', 'p', 'r', 'e', 'n', 'g', 'e',
'e', 'l', 'r', 'é', 'l', 'u', 'n', 'e', 'p', 'r', 'b', 'l', 'è', 'e', 'c', 'p',
'l', 'e', 'x', 'e', 'l', 'e', 'n', 'g', 'l', 'b', 'e', 'u', 'n', 'l', 'r', 'g',
'e', 'é', 'v', 'e', 'n', 'l', 'e', 'n', 'e', 'e', 'l', 'r', 'b', 'q', 'u', 'e',
'u', 'x', 'l', 'g', 'r', 'h', 'e', 'p', 'p', 'r', 'e', 'n', 'g', 'e', 'u', 'q',
'u', 'e', 'l', 'e', 'c', 'h', 'e', 'r', 'c', 'h', 'e', 'u', 'r', 'e', 'n', 'r',
'v', 'l', 'l', 'e', 'n', 'u', 'r', 'l', 'e', 'é', 'v', 'e', 'l', 'p', 'p', 'e',
'e', 'n', 'e', 'è', 'l', 'e', 'c', 'p', 'b', 'l', 'e', 'e', 'p', 'r', 'e', 'n',
'r', 'e', 'e', 'é', 'c', 'n', 'é', 'c', 'l', 'r', 'é', 'e', 'e', 'r', 'é', 'u',
'r', 'e', 'e', 'p', 'r', 'b', 'l', 'è', 'e', 'c', 'p', 'l', 'e', 'x', 'e', 'e',
'n', 'e', 'r', 'g', 'r', 'e', 'n', 'è', 'r', 'e', 'p', 'l', 'u', 'n', 'u', 'r',
'e', 'l', 'l', 'e', 'v', 'e', 'c', 'l', 'e', 'h', 'u', 'n', 'l', 'u', 'l', 'e',
'n', 'e', 'e', 'c', 'h', 'n', 'q', 'u', 'e', 'p', 'p', 'r', 'e', 'n', 'g', 'e',
'u', 'q', 'u', 'e', 'e', 'r', 'e', 'e', 'n', 'u', 'l', 'n', 'g', 'g', 'e', 'n',
'u', 'r', 'e', 'l', 'e', 'v', 'n', 'p', 'r', 'r', 'n', 'e', 'u', 'r', 'e', 'p',
'p', 'r', 'e', 'n', 'g', 'e', 'p', 'r', 'f', 'n', 'p', 'u', 'r', 'c', 'r', 'é',
'e', 'r', 'c', 'e', 'è', 'e', 'n', 'e', 'l', 'l', 'g', 'e', 'n']}]}
```

```
[ ]: # Créez un dictionnaire pour stocker les fréquences de termes
term_frequency = {}

# Parcourez chaque document dans le corpus prétraité
for file in preprocessed_corpus:
    # Récupérez le texte prétraité du document
    preprocessed_text = file["content"]

    # Parcourez chaque mot dans le texte prétraité
    for word in preprocessed_text:
        # Vérifiez si le mot est déjà dans le dictionnaire des fréquences de
        ↪ termes
        if word not in term_frequency:
            # Si le mot n'est pas dans le dictionnaire, ajoutez-le avec une
            ↪ fréquence de 1
            term_frequency[word] = 1
        else:
            # Si le mot est déjà dans le dictionnaire, incrémentez sa fréquence
            term_frequency[word] += 1

# À ce stade, term_frequency contient les fréquences de termes pour l'ensemble
↪ du corpus prétraité
print(term_frequency)
```

```
{'l': 149, 'e': 350, 'v': 23, 'é': 49, 'h': 30, 'c': 77, 'u': 110, 'n': 169,
'x': 6, 'p': 81, 'r': 144, 'b': 16, 'g': 36, 'f': 18, 'è': 17, 'q': 26, 'à': 6,
'ê': 1, 'j': 5, 'w': 1, 'â': 1, 'ç': 1}
```

```
[ ]: # Inverted index contenant la fréquence par document
inverted_index = {}

# Parcourez chaque document dans votre corpus
for idx, document_object in enumerate(corpus_list):
    # Récupérez le contenu du document
    document_content = document_object["content"]

    # Parcourez chaque mot dans le contenu du document
    for word in document_content:
        # Vérifiez si le mot est déjà dans l'inverted index
        if word in inverted_index:
            # Si oui, récupérez la liste de postings
            posting_list = inverted_index[word]

            # Vérifiez si le fichier est déjà répertorié dans les postings
            if idx in posting_list:
                # Si oui, augmentez la fréquence du fichier
                posting_list[idx] += 1
            else:
                # Sinon, ajoutez le fichier à la liste des postings avec une
                ↪ fréquence de 1
                posting_list[idx] = 1
        else:
            # Si le mot n'est pas dans l'inverted index, créez une nouvelle
            ↪ entrée
            inverted_index[word] = {idx: 1}

# À présent, inverted_index contient la fréquence par document (index) pour
↪ chaque mot de votre corpus
print(inverted_index)
```

```
{'L': {0: 1, 1: 1, 2: 1, 3: 1, 4: 4, 5: 1, 7: 1, 8: 3}, 'e': {0: 18, 1: 14, 2: 25, 3: 17, 4: 108, 5: 21, 6: 16, 7: 20, 8: 111}, 's': {0: 14, 1: 10, 2: 16, 3: 21, 4: 83, 5: 9, 6: 14, 7: 13, 8: 64}, ' ': {0: 24, 1: 21, 2: 29, 3: 26, 4: 112, 5: 21, 6: 18, 7: 28, 8: 115}, 'v': {0: 2, 1: 2, 2: 3, 4: 2, 5: 5, 6: 1, 7: 1, 8: 7}, 'é': {0: 1, 1: 1, 2: 6, 3: 6, 4: 14, 5: 3, 6: 2, 7: 2, 8: 14}, 'h': {0: 3, 2: 2, 4: 7, 5: 1, 6: 4, 7: 4, 8: 9}, 'i': {0: 11, 1: 10, 2: 13, 3: 9, 4: 47, 5: 13, 6: 6, 7: 12, 8: 52}, 'c': {0: 5, 1: 2, 2: 7, 4: 21, 5: 5, 6: 4, 7: 6, 8: 24}, 'u': {0: 12, 1: 5, 2: 6, 3: 8, 4: 26, 5: 5, 6: 7, 7: 9, 8: 32}, 'l': {0: 10, 1: 8, 2: 7, 3: 6, 4: 33, 5: 10, 6: 8, 7: 9, 8: 45}, 'a': {0: 10, 1: 11, 2: 9, 3: 4, 4: 54, 5: 9, 6: 2, 7: 9, 8: 53}, 't': {0: 11, 1: 13, 2: 13, 3: 9, 4: 55, 5: 7, 6: 6, 7: 10, 8: 50}, 'o': {0: 9, 1: 7, 2: 8, 3: 5, 4: 33, 5: 5, 6: 10, 7: 8, 8: 31}, 'n': {0: 11, 1: 8, 2: 12, 3: 5, 4: 55, 5: 11, 6: 3, 7: 12, 8: 52}, 'm': {0: 5, 1: 7, 2: 4, 3: 4, 4: 21, 5: 2, 6: 1, 7: 1, 8: 23}, 'x': {0: 1, 4: 1, 7: 1, 8: 3}, 'p': {0: 5, 1: 6, 2: 4, 3: 2, 4: 27, 5: 5, 6: 6, 8: 25}, 'd': {0: 7, 1: 4, 2: 4, 3: 6, 4: 40, 5: 3, 6: 5, 7: 6, 8: 32}, '"': {0: 3, 1: 3, 2: 1, 3:
```



```
3, 4: 9, 5: 3, 6: 1, 7: 4, 8: 11}, 'I': {0: 2, 1: 2, 2: 2, 3: 2, 4: 1, 5: 3, 6: 1, 7: 1, 8: 5}, 'A': {0: 1, 1: 1, 2: 1, 3: 3, 5: 3, 6: 1, 7: 2, 8: 4}, 'r': {0: 6, 1: 11, 2: 8, 3: 7, 4: 40, 5: 10, 6: 10, 7: 9, 8: 43}, 'b': {0: 1, 3: 2, 4: 4, 6: 2, 7: 1, 8: 6}, '.': {0: 2, 1: 2, 2: 2, 3: 2, 4: 6, 5: 1, 6: 1, 7: 2, 8: 5}, 'g': {0: 2, 4: 11, 5: 3, 7: 3, 8: 16}, 'f': {1: 1, 2: 1, 3: 2, 4: 5, 5: 1, 7: 2, 8: 5}, 'è': {2: 1, 3: 2, 4: 4, 6: 2, 7: 1, 8: 7}, 'q': {2: 4, 3: 1, 4: 9, 5: 2, 6: 1, 7: 1, 8: 8}, ',': {2: 2, 4: 7, 5: 2, 8: 7}, 'à': {2: 1, 3: 1, 4: 2, 5: 1, 8: 1}, 'è': {3: 1}, 'j': {3: 2, 7: 2}, 'y': {3: 1, 4: 2, 6: 1, 7: 1, 8: 3}, '\n': {4: 2, 8: 2}, 'C': {4: 1, 7: 1, 8: 1}, '(' : {5: 1, 8: 1}, 'G': {5: 1}, ')': {5: 1, 8: 1}, 'T': {6: 2, 7: 1}, 'F': {6: 1}, 'w': {6: 1}, 'P': {6: 1}, 'J': {7: 1}, 'M': {7: 1}, 'â': {8: 1}, 'ç': {8: 1}}
```

```
[ ]: from nltk import ngrams, FreqDist
      from nltk.tokenize import word_tokenize
```

```
[ ]: # Définissez la valeur de n pour les n-grams
n = 3

# Créez une liste pour stocker les n-grams de chaque document
document_ngrams_list = []

# Parcourez chaque document dans votre corpus_list
for document_object in corpus_list:
    # Parcourez chaque document dans votre corpus_list
    document_content = document_object["content"]

    # Tokenize le contenu du document
    document_words = word_tokenize(document_content)

    # Génère des n-grams pour les mots
    word_ngrams = list(ngrams(document_words, n))

    # Génère des n-grams pour les caractères
    char_ngrams = list(ngrams(document_content, n))

    # Calcule les fréquences des mots à l'aide de FreqDist
    word_freq_dist = FreqDist(document_words)

    # Crée un dictionnaire pour stocker les informations sur le document
    document_info = {
        "content": document_content,
        "word_ngrams": word_ngrams,
        "char_ngrams": char_ngrams,
        "word_freq_dist": dict(word_freq_dist)
    }

    # Ajoute document_info à la liste
```









Content: L'IA peut être sujette à des biais si les données utilisées pour former les modèles sont biaisées. Assurer l'équité dans les systèmes d'IA est un défi majeur.

Word n-grams: [("L'IA", 'peut', 'être'), ('peut', 'être', 'sujette'), ('être', 'sujette', 'à'), ('sujette', 'à', 'des'), ('à', 'des', 'biais'), ('des', 'biais', 'si'), ('biais', 'si', 'les'), ('si', 'les', 'données'), ('les', 'données', 'utilisées'), ('données', 'utilisées', 'pour'), ('utilisées', 'pour', 'former'), ('pour', 'former', 'les'), ('former', 'les', 'modèles'), ('les', 'modèles', 'sont'), ('modèles', 'sont', 'biaisées'), ('sont', 'biaisées', '.'), ('biaisées', '.', 'Assurer'), ('.', 'Assurer', "l'équité"), ('Assurer', "l'équité", 'dans'), ("l'équité", 'dans', 'les'), ('dans', 'les', 'systèmes'), ('les', 'systèmes', "d'IA"), ('systèmes', "d'IA", 'est'), ("d'IA", 'est', 'un'), ('est', 'un', 'défi'), ('un', 'défi', 'majeur'), ('défi', 'majeur', '.')] ]

Character n-grams: [(('L', '"', 'I'), ('"', 'I', 'A'), ('I', 'A', ' '), ('A', ' ', 'p'), (' ', 'p', 'e'), ('p', 'e', 'u'), ('e', 'u', 't'), ('u', 't', ' '), ('t', ' ', 'ê'), (' ', 'ê', 't'), ('ê', 't', 'r'), ('t', 'r', 'e'), ('r', 'e', ' '), ('e', ' ', 's'), (' ', 's', 'u'), ('s', 'u', 'j'), ('u', 'j', 'e'), ('j', 'e', 't'), ('e', 't', 't'), ('t', 't', 'e'), ('t', 'e', ' '), ('e', ' ', 'à'), (' ', 'à', ' '), ('à', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'b'), (' ', 'b', 'i'), ('b', 'i', 'a'), ('i', 'a', 'i'), ('a', 'i', 's'), ('i', 's', ' '), ('s', ' ', 's'), (' ', 's', 'i'), ('s', 'i', ' '), ('i', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', 'o'), ('d', 'o', 'n'), ('o', 'n', 'n'), ('n', 'n', 'é'), ('n', 'é', 'e'), ('é', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'u'), (' ', 'u', 't'), ('u', 't', 'i'), ('t', 'i', 'l'), ('i', 'l', 'i'), ('l', 'i', 's'), ('i', 's', 'é'), ('s', 'é', 'e'), ('é', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'p'), (' ', 'p', 'o'), ('p', 'o', 'u'), ('o', 'u', 'r'), ('u', 'r', ' '), ('r', ' ', 'f'), (' ', 'f', 'o'), ('f', 'o', 'r'), ('o', 'r', 'm'), ('r', 'm', 'e'), ('m', 'e', 'r'), ('e', 'r', ' '), ('r', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'm'), (' ', 'm', 'o'), ('m', 'o', 'd'), ('o', 'd', 'è'), ('d', 'è', 'l'), ('è', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 's'), (' ', 's', 'o'), ('s', 'o', 'n'), ('o', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'b'), (' ', 'b', 'i'), ('b', 'i', 'a'), ('i', 'a', 'i'), ('a', 'i', 's'), ('i', 's', 'é'), ('s', 'é', 'e'), ('é', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'A'), (' ', 'A', 's'), ('A', 's', 's'), ('s', 's', 'u'), ('s', 'u', 'r'), ('u', 'r', 'e'), ('r', 'e', 'r'), ('e', 'r', ' '), ('r', ' ', 'l'), (' ', 'l', ' '), ('l', ' ', 'é'), (' ', 'é', 'q'), ('é', 'q', 'u'), ('q', 'u', 'i'), ('u', 'i', 't'), ('i', 't', 'é'), ('t', 'é', ' '), ('é', ' ', 'd'), (' ', 'd', 'a'), ('d', 'a', 'n'), ('a', 'n', 's'), ('n', 's', ' '), ('s', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 's'), (' ', 's', 'y'), ('s', 'y', 's'), ('y', 's', 't'), ('s', 't', 'è'), ('t', 'è', 'm'), ('è', 'm', 'e'), ('m', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'I'), (' ', 'I', 'A'), ('I', 'A', ' '), ('A', ' ', 'e'), (' ', 'e', 's'), ('e', 's', 't'), ('s', 't', ' '), ('t', ' ', 'u'), (' ', 'u', 'n'), ('u', 'n', ' '), ('n', ' ', 'd'), (' ', 'd', 'é'), ('d', 'é', 'f'), ('é', 'f', 'i'), ('f', 'i', ' '), ('i', ' ', 'm'), (' ', 'm', 'a'), ('m', 'a', 'j'), ('a', 'j', 'e'), ('j', 'e', 'u'), ('e', 'u', 'r'), ('u', 'r', ' ')] ]

Word Frequency Distribution: {"L'IA": 1, 'peut': 1, 'être': 1, 'sujette': 1,

'à': 1, 'des': 1, 'biais': 1, 'si': 1, 'les': 3, 'données': 1, 'utilisées': 1, 'pour': 1, 'former': 1, 'modèles': 1, 'sont': 1, 'biaisées': 1, '.': 2, 'Assurer': 1, "l'équité": 1, 'dans': 1, 'systèmes': 1, "d'IA": 1, 'est': 1, 'un': 1, 'défi': 1, 'majeur': 1}

#### Document 5 Information:

Content: L'apprentissage automatique est l'une des branches les plus importantes de l'intelligence artificielle. Il repose sur des algorithmes qui permettent aux ordinateurs d'apprendre à partir de données. Les systèmes d'apprentissage automatique sont capables d'analyser des ensembles de données, de détecter des schémas et de prendre des décisions basées sur ces modèles.

L'apprentissage automatique comprend diverses techniques, telles que la régression, la classification, le clustering et la reconnaissance de motifs. Ces techniques sont utilisées dans une variété de domaines, de la prédiction de tendances financières à la recommandation de produits en ligne. Les algorithmes d'apprentissage automatique sont alimentés par des données, et plus les données sont riches et pertinentes, plus les modèles d'apprentissage automatique sont performants.

Word n-grams: [("L'apprentissage", 'automatique', 'est'), ('automatique', 'est', 'l'une'), ('est', 'l'une', 'des'), ('l'une', 'des', 'branches'), ('des', 'branches', 'les'), ('branches', 'les', 'plus'), ('les', 'plus', 'importantes'), ('plus', 'importantes', 'de'), ('importantes', 'de', 'l'intelligence'), ('de', 'l'intelligence', 'artificielle'), ('l'intelligence', 'artificielle', '.'), ('artificielle', '.', 'Il'), ('.', 'Il', 'repose'), ('Il', 'repose', 'sur'), ('repose', 'sur', 'des'), ('sur', 'des', 'algorithmes'), ('des', 'algorithmes', 'qui'), ('algorithmes', 'qui', 'permettent'), ('qui', 'permettent', 'aux'), ('permettent', 'aux', 'ordinateurs'), ('aux', 'ordinateurs', 'd'apprendre'), ('ordinateurs', 'd'apprendre', 'à'), ('d'apprendre', 'à', 'partir'), ('à', 'partir', 'de'), ('partir', 'de', 'données'), ('de', 'données', '.'), ('données', '.', 'Les'), ('.', 'Les', 'systèmes'), ('Les', 'systèmes', 'd'apprentissage'), ('systèmes', 'd'apprentissage', 'automatique'), ('d'apprentissage', 'automatique', 'sont'), ('automatique', 'sont', 'capables'), ('sont', 'capables', 'd'analyser'), ('capables', 'd'analyser', 'des'), ('d'analyser', 'des', 'ensembles'), ('des', 'ensembles', 'de'), ('ensembles', 'de', 'données'), ('de', 'données', ','), ('données', ',', 'de'), ('', 'de', 'détecter'), ('de', 'détecter', 'des'), ('détecter', 'des', 'schémas'), ('des', 'schémas', 'et'), ('schémas', 'et', 'de'), ('et', 'de', 'prendre'), ('de', 'prendre', 'des'), ('prendre', 'des', 'décisions'), ('des', 'décisions', 'basées'), ('décisions', 'basées', 'sur'), ('basées', 'sur', 'ces'), ('sur', 'ces', 'modèles'), ('ces', 'modèles', '.'), ('modèles', '.', 'L'apprentissage'), ('.', 'L'apprentissage', 'automatique'), ('L'apprentissage', 'automatique', 'comprend'), ('automatique', 'comprend', 'diverses'), ('comprend', 'diverses', 'techniques'), ('diverses', 'techniques', ','), ('techniques', ',', 'telles'), ('', 'telles', 'que'), ('telles', 'que', 'la'), ('que', 'la', 'régression'), ('la', 'régression', ','), ('régression', ',', 'la'), ('', 'la', 'classification'), ('la', 'classification', ','), ('classification', ',', 'le'),

(',', 'le', 'clustering'), ('le', 'clustering', 'et'), ('clustering', 'et', 'la'), ('et', 'la', 'reconnaissance'), ('la', 'reconnaissance', 'de'), ('reconnaissance', 'de', 'motifs'), ('de', 'motifs', '.'), ('motifs', '.', 'Ces'), ('.', 'Ces', 'techniques'), ('Ces', 'techniques', 'sont'), ('techniques', 'sont', 'utilisées'), ('sont', 'utilisées', 'dans'), ('utilisées', 'dans', 'une'), ('dans', 'une', 'variété'), ('une', 'variété', 'de'), ('variété', 'de', 'domaines'), ('de', 'domaines', ','), ('domaines', ', ', 'de'), (' ', 'de', 'la'), ('de', 'la', 'prédiction'), ('la', 'prédiction', 'de'), ('prédiction', 'de', 'tendances'), ('de', 'tendances', 'financières'), ('tendances', 'financières', 'à'), ('financières', 'à', 'la'), ('à', 'la', 'recommandation'), ('la', 'recommandation', 'de'), ('recommandation', 'de', 'produits'), ('de', 'produits', 'en'), ('produits', 'en', 'ligne'), ('en', 'ligne', '.'), ('ligne', '.', 'Les'), ('.', 'Les', 'algorithmes'), ('Les', 'algorithmes', "d'apprentissage"), ('algorithmes', "d'apprentissage", 'automatique'), ("d'apprentissage", 'automatique', 'sont'), ('automatique', 'sont', 'alimentés'), ('sont', 'alimentés', 'par'), ('alimentés', 'par', 'des'), ('par', 'des', 'données'), ('des', 'données', ','), ('données', ', ', 'et'), (' ', 'et', 'plus'), ('et', 'plus', 'les'), ('plus', 'les', 'données'), ('les', 'données', 'sont'), ('données', 'sont', 'riches'), ('sont', 'riches', 'et'), ('riches', 'et', 'pertinentes'), ('et', 'pertinentes', ','), ('pertinentes', ', ', 'plus'), (' ', 'plus', 'les'), ('plus', 'les', 'modèles'), ('les', 'modèles', "d'apprentissage"), ('modèles', "d'apprentissage", 'automatique'), ("d'apprentissage", 'automatique', 'sont'), ('automatique', 'sont', 'performants'), ('sont', 'performants', '.')] ]

Character n-grams: [(('L', '"', 'a'), ('"', 'a', 'p'), ('a', 'p', 'p'), ('p', 'p', 'r'), ('p', 'r', 'e'), ('r', 'e', 'n'), ('e', 'n', 't'), ('n', 't', 'i'), ('t', 'i', 's'), ('i', 's', 's'), ('s', 's', 'a'), ('s', 'a', 'g'), ('a', 'g', 'e'), ('g', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'u'), ('a', 'u', 't'), ('u', 't', 'o'), ('t', 'o', 'm'), ('o', 'm', 'a'), ('m', 'a', 't'), ('a', 't', 'i'), ('t', 'i', 'q'), ('i', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '), ('e', ' ', 'e'), (' ', 'e', 's'), ('e', 's', 't'), ('s', 't', ' '), ('t', ' ', 'l'), (' ', 'l', '"'), ('l', '"', 'u'), ('"', 'u', 'n'), ('u', 'n', 'e'), ('n', 'e', ' '), ('e', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'b'), (' ', 'b', 'r'), ('b', 'r', 'a'), ('r', 'a', 'n'), ('a', 'n', 'c'), ('n', 'c', 'h'), ('c', 'h', 'e'), ('h', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'p'), (' ', 'p', 'l'), ('p', 'l', 'u'), ('l', 'u', 's'), ('u', 's', ' '), ('s', ' ', 'i'), (' ', 'i', 'm'), ('i', 'm', 'p'), ('m', 'p', 'o'), ('p', 'o', 'r'), ('o', 'r', 't'), ('r', 't', 'a'), ('t', 'a', 'n'), ('a', 'n', 't'), ('n', 't', 'e'), ('t', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'l'), (' ', 'l', '"'), ('l', '"', 'i'), ('"', 'i', 'n'), ('i', 'n', 't'), ('n', 't', 'e'), ('t', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'i'), ('l', 'i', 'g'), ('i', 'g', 'e'), ('g', 'e', 'n'), ('e', 'n', 'c'), ('n', 'c', 'e'), ('c', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'r'), ('a', 'r', 't'), ('r', 't', 'i'), ('t', 'i', 'f'), ('i', 'f', 'i'), ('f', 'i', 'c'), ('i', 'c', 'i'), ('c', 'i', 'e'), ('i', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'e'), ('l', 'e', ' '), ('e', ' ', ' '), (' ', ' ', 'I'), (' ', 'I', 'l'), ('I', 'l', ' '), ('l', ' ', 'r'), (' ', 'r', 'e'), ('r', 'e', 'p'), ('e', 'p', 'o'), ('p', 'o', 's'), ('o', 's',



'e'), ('s', 'e', ' '), ('e', ' ', 's'), (' ', 's', 'u'), ('s', 'u', 'r'), ('u',  
 'r', ' '), ('r', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '),  
 ('s', ' ', 'a'), (' ', 'a', 'l'), ('a', 'l', 'g'), ('l', 'g', 'o'), ('g', 'o',  
 'r'), ('o', 'r', 'i'), ('r', 'i', 't'), ('i', 't', 'h'), ('t', 'h', 'm'), ('h',  
 'm', 'e'), ('m', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'q'), (' ', 'q', 'u'),  
 ('q', 'u', 'i'), ('u', 'i', ' '), ('i', ' ', 'p'), (' ', 'p', 'e'), ('p', 'e',  
 'r'), ('e', 'r', 'm'), ('r', 'm', 'e'), ('m', 'e', 't'), ('e', 't', 't'), ('t',  
 't', 'e'), ('t', 'e', 'n'), ('e', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'a'),  
 (' ', 'a', 'u'), ('a', 'u', 'x'), ('u', 'x', ' '), ('x', ' ', 'o'), (' ', 'o',  
 'r'), ('o', 'r', 'd'), ('r', 'd', 'i'), ('d', 'i', 'n'), ('i', 'n', 'a'), ('n',  
 'a', 't'), ('a', 't', 'e'), ('t', 'e', 'u'), ('e', 'u', 'r'), ('u', 'r', 's'),  
 ('r', 's', ' '), ('s', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'a'), (' ', 'a',  
 'p'), ('a', 'p', 'p'), ('p', 'p', 'r'), ('p', 'r', 'e'), ('r', 'e', 'n'), ('e',  
 'n', 'd'), ('n', 'd', 'r'), ('d', 'r', 'e'), ('r', 'e', ' '), ('e', ' ', 'à'),  
 (' ', 'à', ' '), ('à', ' ', 'p'), (' ', 'p', 'a'), ('p', 'a', 'r'), ('a', 'r',  
 't'), ('r', 't', 'i'), ('t', 'i', 'r'), ('i', 'r', ' '), ('r', ' ', 'd'), (' ',  
 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'd'), (' ', 'd', 'o'), ('d', 'o', 'n'),  
 ('o', 'n', 'n'), ('n', 'n', 'é'), ('n', 'é', 'e'), ('é', 'e', 's'), ('e', 's',  
 ' '), ('s', ' ', ' '), (' ', ' ', 'L'), (' ', 'L', 'e'), ('L', 'e', 's'), ('e',  
 's', ' '), ('s', ' ', 's'), (' ', 's', 'y'), ('s', 'y', 's'), ('y', 's', 't'),  
 ('s', 't', 'è'), ('t', 'è', 'm'), ('è', 'm', 'e'), ('m', 'e', 's'), ('e', 's', ' '),  
 ('s', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'a'), (' ', 'a', 'p'), ('a',  
 'p', 'p'), ('p', 'p', 'r'), ('p', 'r', 'e'), ('r', 'e', 'n'), ('e', 'n', 't'),  
 ('n', 't', 'i'), ('t', 'i', 's'), ('i', 's', 's'), ('s', 's', 'a'), ('s', 'a',  
 'g'), ('a', 'g', 'e'), ('g', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'u'), ('a',  
 'u', 't'), ('u', 't', 'o'), ('t', 'o', 'm'), ('o', 'm', 'a'), ('m', 'a', 't'),  
 ('a', 't', 'i'), ('t', 'i', 'q'), ('i', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '),  
 ('e', ' ', 's'), (' ', 's', 'o'), ('s', 'o', 'n'), ('o', 'n', 't'), ('n',  
 't', ' '), ('t', ' ', 'c'), (' ', 'c', 'a'), ('c', 'a', 'p'), ('a', 'p', 'a'),  
 ('p', 'a', 'b'), ('a', 'b', 'l'), ('b', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '),  
 ('s', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'a'), (' ', 'a', 'n'), ('a',  
 'n', 'a'), ('n', 'a', 'l'), ('a', 'l', 'y'), ('l', 'y', 's'), ('y', 's', 'e'),  
 ('s', 'e', 'r'), ('e', 'r', ' '), ('r', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e',  
 's'), ('e', 's', ' '), ('s', ' ', 'e'), (' ', 'e', 'n'), ('e', 'n', 's'), ('n',  
 's', 'e'), ('s', 'e', 'm'), ('e', 'm', 'b'), ('m', 'b', 'l'), ('b', 'l', 'e'),  
 ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '),  
 ('e', ' ', 'd'), (' ', 'd', 'o'), ('d', 'o', 'n'), ('o', 'n', 'n'), ('n',  
 'n', 'é'), ('n', 'é', 'e'), ('é', 'e', 's'), ('e', 's', ' '), ('s', ' ', ' '),  
 (' ', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'd'), (' ', 'd',  
 'é'), ('d', 'é', 't'), ('é', 't', 'e'), ('t', 'e', 'c'), ('e', 'c', 't'), ('c',  
 't', 'e'), ('t', 'e', 'r'), ('e', 'r', ' '), ('r', ' ', 'd'), (' ', 'd', 'e'),  
 ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 's'), (' ', 's', 'c'), ('s', 'c',  
 'h'), ('c', 'h', 'é'), ('h', 'é', 'm'), ('é', 'm', 'a'), ('m', 'a', 's'), ('a',  
 's', ' '), ('s', ' ', 'e'), (' ', 'e', 't'), ('e', 't', ' '), ('t', ' ', 'd'),  
 (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'p'), (' ', 'p', 'r'), ('p', 'r',  
 'e'), ('r', 'e', 'n'), ('e', 'n', 'd'), ('n', 'd', 'r'), ('d', 'r', 'e'), ('r',  
 'e', ' '), ('e', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '),  
 ('s', ' ', 'd'), (' ', 'd', 'é'), ('d', 'é', 'c'), ('é', 'c', 'i'), ('c', 'i',

's'), ('i', 's', 'i'), ('s', 'i', 'o'), ('i', 'o', 'n'), ('o', 'n', 's'), ('n',  
's', ' '), ('s', ' ', 'b'), (' ', 'b', 'a'), ('b', 'a', 's'), ('a', 's', 'é'),  
('s', 'é', 'e'), ('é', 'e', 's'), ('e', 's', ' '), ('s', ' ', 's'), (' ', 's',  
'u'), ('s', 'u', 'r'), ('u', 'r', ' '), ('r', ' ', 'c'), (' ', 'c', 'e'), ('c',  
'e', 's'), ('e', 's', ' '), ('s', ' ', 'm'), (' ', 'm', 'o'), ('m', 'o', 'd'),  
('o', 'd', 'è'), ('d', 'è', 'l'), ('è', 'l', 'e'), ('l', 'e', 's'), ('e', 's',  
'.'), ('s', '. ', '\n'), ('. ', '\n', '\n'), ('\n', '\n', 'L'), ('\n', 'L', '"'),  
('L', '"', 'a'), ('"', 'a', 'p'), ('a', 'p', 'p'), ('p', 'p', 'r'), ('p', 'r',  
'e'), ('r', 'e', 'n'), ('e', 'n', 't'), ('n', 't', 'i'), ('t', 'i', 's'), ('i',  
's', 's'), ('s', 's', 'a'), ('s', 'a', 'g'), ('a', 'g', 'e'), ('g', 'e', ' '),  
('e', ' ', 'a'), (' ', 'a', 'u'), ('a', 'u', 't'), ('u', 't', 'o'), ('t', 'o',  
'm'), ('o', 'm', 'a'), ('m', 'a', 't'), ('a', 't', 'i'), ('t', 'i', 'q'), ('i',  
'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '), ('e', ' ', 'c'), (' ', 'c', 'o'),  
('c', 'o', 'm'), ('o', 'm', 'p'), ('m', 'p', 'r'), ('p', 'r', 'e'), ('r', 'e',  
'n'), ('e', 'n', 'd'), ('n', 'd', ' '), ('d', ' ', 'd'), (' ', 'd', 'i'), ('d',  
'i', 'v'), ('i', 'v', 'e'), ('v', 'e', 'r'), ('e', 'r', 's'), ('r', 's', 'e'),  
('s', 'e', 's'), ('e', 's', ' '), ('s', ' ', 't'), (' ', 't', 'e'), ('t', 'e',  
'c'), ('e', 'c', 'h'), ('c', 'h', 'n'), ('h', 'n', 'i'), ('n', 'i', 'q'), ('i',  
'q', 'u'), ('q', 'u', 'e'), ('u', 'e', 's'), ('e', 's', ' '), ('s', ' ', ' '),  
(' ', ' ', 't'), (' ', 't', 'e'), ('t', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l',  
'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'q'), (' ', 'q', 'u'), ('q',  
'u', 'e'), ('u', 'e', ' '), ('e', ' ', 'l'), (' ', 'l', 'a'), ('l', 'a', ' '),  
('a', ' ', 'r'), (' ', 'r', 'é'), ('r', 'é', 'g'), ('é', 'g', 'r'), ('g', 'r',  
'e'), ('r', 'e', 's'), ('e', 's', 's'), ('s', 's', 'i'), ('s', 'i', 'o'), ('i',  
'o', 'n'), ('o', 'n', ' '), ('n', ' ', ' '), (' ', ' ', 'l'), (' ', 'l', 'a'),  
('l', 'a', ' '), ('a', ' ', 'c'), (' ', 'c', 'l'), ('c', 'l', 'a'), ('l', 'a',  
's'), ('a', 's', 's'), ('s', 's', 'i'), ('s', 'i', 'f'), ('i', 'f', 'i'), ('f',  
'i', 'c'), ('i', 'c', 'a'), ('c', 'a', 't'), ('a', 't', 'i'), ('t', 'i', 'o'),  
('i', 'o', 'n'), ('o', 'n', ' '), ('n', ' ', ' '), (' ', ' ', 'l'), (' ', 'l',  
'e'), ('l', 'e', ' '), ('e', ' ', 'c'), (' ', 'c', 'l'), ('c', 'l', 'u'), ('l',  
'u', 's'), ('u', 's', 't'), ('s', 't', 'e'), ('t', 'e', 'r'), ('e', 'r', 'i'),  
('r', 'i', 'n'), ('i', 'n', 'g'), ('n', 'g', ' '), ('g', ' ', 'e'), (' ', 'e',  
't'), ('e', 't', ' '), ('t', ' ', 'l'), (' ', 'l', 'a'), ('l', 'a', ' '), ('a',  
' ', 'r'), (' ', 'r', 'e'), ('r', 'e', 'c'), ('e', 'c', 'o'), ('c', 'o', 'n'),  
('o', 'n', 'n'), ('n', 'n', 'a'), ('n', 'a', 'i'), ('a', 'i', 's'), ('i', 's',  
's'), ('s', 's', 'a'), ('s', 'a', 'n'), ('a', 'n', 'c'), ('n', 'c', 'e'), ('c',  
'e', ' '), ('e', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'm'),  
(' ', 'm', 'o'), ('m', 'o', 't'), ('o', 't', 'i'), ('t', 'i', 'f'), ('i', 'f',  
's'), ('f', 's', '.'), ('s', '. ', ' '), ('. ', ' ', 'C'), (' ', 'C', 'e'), ('C',  
'e', 's'), ('e', 's', ' '), ('s', ' ', 't'), (' ', 't', 'e'), ('t', 'e', 'c'),  
('e', 'c', 'h'), ('c', 'h', 'n'), ('h', 'n', 'i'), ('n', 'i', 'q'), ('i', 'q',  
'u'), ('q', 'u', 'e'), ('u', 'e', 's'), ('e', 's', ' '), ('s', ' ', 's'), (' ',  
's', 'o'), ('s', 'o', 'n'), ('o', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'u'),  
(' ', 'u', 't'), ('u', 't', 'i'), ('t', 'i', 'l'), ('i', 'l', 'i'), ('l', 'i',  
's'), ('i', 's', 'é'), ('s', 'é', 'e'), ('é', 'e', 's'), ('e', 's', ' '), ('s',  
' ', 'd'), (' ', 'd', 'a'), ('d', 'a', 'n'), ('a', 'n', 's'), ('n', 's', ' '),  
('s', ' ', 'u'), (' ', 'u', 'n'), ('u', 'n', 'e'), ('n', 'e', ' '), ('e', ' ',  
'v'), (' ', 'v', 'a'), ('v', 'a', 'r'), ('a', 'r', 'i'), ('r', 'i', 'é'), ('i',



(' ', 'p'), (' ', 'p', 'l'), ('p', 'l', 'u'), ('l', 'u', 's'), ('u', 's', ' '), ('s', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'm'), (' ', 'm', 'o'), ('m', 'o', 'd'), ('o', 'd', 'è'), ('d', 'è', 'l'), ('è', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'a'), (' ', 'a', 'p'), ('a', 'p', 'p'), ('p', 'p', 'r'), ('p', 'r', 'e'), ('r', 'e', 'n'), ('e', 'n', 't'), ('n', 't', 'i'), ('t', 'i', 's'), ('i', 's', 's'), ('s', 's', 'a'), ('s', 'a', 'g'), ('a', 'g', 'e'), ('g', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'u'), ('a', 'u', 't'), ('u', 't', 'o'), ('t', 'o', 'm'), ('o', 'm', 'a'), ('m', 'a', 't'), ('a', 't', 'i'), ('t', 'i', 'q'), ('i', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '), ('e', ' ', 's'), (' ', 's', 'o'), ('s', 'o', 'n'), ('o', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'p'), (' ', 'p', 'e'), ('p', 'e', 'r'), ('e', 'r', 'f'), ('r', 'f', 'o'), ('f', 'o', 'r'), ('o', 'r', 'm'), ('r', 'm', 'a'), ('m', 'a', 'n'), ('a', 'n', 't'), ('n', 't', 's'), ('t', 's', ' ')]

Word Frequency Distribution: {"L'apprentissage": 2, 'automatique': 5, 'est': 1, 'l'une': 1, 'des': 6, 'branches': 1, 'les': 3, 'plus': 3, 'importantes': 1, 'de': 10, "l'intelligence": 1, 'artificielle': 1, '.': 6, 'Il': 1, 'repose': 1, 'sur': 2, 'algorithmes': 2, 'qui': 1, 'permettent': 1, 'aux': 1, 'ordinateurs': 1, 'd'apprendre': 1, 'à': 2, 'partir': 1, 'données': 4, 'Les': 2, 'systèmes': 1, 'd'apprentissage': 3, 'sont': 5, 'capables': 1, 'd'analyser': 1, 'ensembles': 1, ',': 7, 'détecter': 1, 'schémas': 1, 'et': 4, 'prendre': 1, 'décisions': 1, 'basées': 1, 'ces': 1, 'modèles': 2, 'comprendre': 1, 'diverses': 1, 'techniques': 2, 'telles': 1, 'que': 1, 'la': 5, 'régression': 1, 'classification': 1, 'le': 1, 'clustering': 1, 'reconnaissance': 1, 'motifs': 1, 'Ces': 1, 'utilisées': 1, 'dans': 1, 'une': 1, 'variété': 1, 'domaines': 1, 'prédiction': 1, 'tendances': 1, 'financières': 1, 'recommandation': 1, 'produits': 1, 'en': 1, 'ligne': 1, 'alimentés': 1, 'par': 1, 'riches': 1, 'pertinentes': 1, 'performants': 1}

#### Document 6 Information:

Content: L'IA continue de progresser, avec des développements passionnants à venir, tels que l'IA générale artificielle (AGI) qui pourrait rivaliser avec l'intelligence humaine.

Word n-grams: [("L'IA", 'continue', 'de'), ('continue', 'de', 'progresser'), ('de', 'progresser', ','), ('progresser', ', ', 'avec'), (' ', 'avec', 'des'), ('avec', 'des', 'développements'), ('des', 'développements', 'passionnants'), ('développements', 'passionnants', 'à'), ('passionnants', 'à', 'venir'), ('à', 'venir', ','), ('venir', ', ', 'tels'), (' ', 'tels', 'que'), ('tels', 'que', 'l'IA'), ('que', 'l'IA', 'générale'), ('l'IA', 'générale', 'artificielle'), ('générale', 'artificielle', '('), ('artificielle', '(', 'AGI'), ('(', 'AGI', ')'), ('AGI', ')', 'qui'), (' ', 'qui', 'pourrait'), ('qui', 'pourrait', 'rivaliser'), ('pourrait', 'rivaliser', 'avec'), ('rivaliser', 'avec', 'l'intelligence'), ('avec', 'l'intelligence', 'humaine'), ('l'intelligence', 'humaine', '.')] ]

Character n-grams: [('L', ' ', 'I'), (' ', 'I', 'A'), ('I', 'A', ' '), ('A', ' ', ' '), (' ', 'c'), (' ', 'c', 'o'), ('c', 'o', 'n'), ('o', 'n', 't'), ('n', 't', 'i'), ('t', 'i', 'n'), ('i', 'n', 'u'), ('n', 'u', 'e'), ('u', 'e', ' '), ('e', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'p'), (' ', 'p', 'r'), ('p',

'r', 'o'), ('r', 'o', 'g'), ('o', 'g', 'r'), ('g', 'r', 'e'), ('r', 'e', 's'), ('e', 's', 's'), ('s', 's', 'e'), ('s', 'e', 'r'), ('e', 'r', ' '), ('r', ' ', ' '), (' ', ' ', 'a'), (' ', 'a', 'v'), ('a', 'v', 'e'), ('v', 'e', 'c'), ('e', 'c', ' '), ('c', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', 'é'), ('d', 'é', 'v'), ('é', 'v', 'e'), ('v', 'e', 'l'), ('e', 'l', 'o'), ('l', 'o', 'p'), ('o', 'p', 'p'), ('p', 'p', 'e'), ('p', 'e', 'm'), ('e', 'm', 'e'), ('m', 'e', 'n'), ('e', 'n', 't'), ('n', 't', 's'), ('t', 's', ' '), ('s', ' ', 'p'), (' ', 'p', 'a'), ('p', 'a', 's'), ('a', 's', 's'), ('s', 's', 'i'), ('s', 'i', 'o'), ('i', 'o', 'n'), ('o', 'n', 'n'), ('n', 'n', 'a'), ('n', 'a', 'n'), ('a', 'n', 't'), ('n', 't', 's'), ('t', 's', ' '), ('s', ' ', 'à'), (' ', 'à', ' '), ('à', ' ', 'v'), (' ', 'v', 'e'), ('v', 'e', 'n'), ('e', 'n', 'i'), ('n', 'i', 'r'), ('i', 'r', ' '), ('r', ' ', ' '), (' ', ' ', 't'), (' ', 't', 'e'), ('t', 'e', 'l'), ('e', 'l', 's'), ('l', 's', ' '), ('s', ' ', 'q'), (' ', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '), ('e', ' ', 'l'), (' ', 'l', ' '), ('l', ' ', 'I'), (' ', 'I', 'A'), ('I', 'A', ' '), ('A', ' ', 'g'), (' ', 'g', 'é'), ('g', 'é', 'n'), ('é', 'n', 'é'), ('n', 'é', 'r'), ('é', 'r', 'a'), ('r', 'a', 'l'), ('a', 'l', 'e'), ('l', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'r'), ('a', 'r', 't'), ('r', 't', 'i'), ('t', 'i', 'f'), ('i', 'f', 'i'), ('f', 'i', 'c'), ('i', 'c', 'i'), ('c', 'i', 'e'), ('i', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'e'), ('l', 'e', ' '), ('e', ' ', ' '), (' ', ' '), ('A'), ('(', 'A', 'G'), ('A', 'G', 'I'), ('G', 'I', ' '), ('I', ' '), (' '), (' ', 'q'), (' ', 'q', 'u'), ('q', 'u', 'i'), ('u', 'i', ' '), ('i', ' ', 'p'), (' ', 'p', 'o'), ('p', 'o', 'u'), ('o', 'u', 'r'), ('u', 'r', 'r'), ('r', 'r', 'a'), ('r', 'a', 'i'), ('a', 'i', 't'), ('i', 't', ' '), ('t', ' ', 'r'), (' ', 'r', 'i'), ('r', 'i', 'v'), ('i', 'v', 'a'), ('v', 'a', 'l'), ('a', 'l', 'i'), ('l', 'i', 's'), ('i', 's', 'e'), ('s', 'e', 'r'), ('e', 'r', ' '), ('r', ' ', 'a'), (' ', 'a', 'v'), ('a', 'v', 'e'), ('v', 'e', 'c'), ('e', 'c', ' '), ('c', ' ', 'l'), (' ', 'l', ' '), ('l', ' ', 'i'), (' ', 'i', 'n'), ('i', 'n', 't'), ('n', 't', 'e'), ('t', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'i'), ('l', 'i', 'g'), ('i', 'g', 'e'), ('g', 'e', 'n'), ('e', 'n', 'c'), ('n', 'c', 'e'), ('c', 'e', ' '), ('e', ' ', 'h'), (' ', 'h', 'u'), ('h', 'u', 'm'), ('u', 'm', 'a'), ('m', 'a', 'i'), ('a', 'i', 'n'), ('i', 'n', 'e'), ('n', 'e', ' ')]

Word Frequency Distribution: {"L'IA": 1, 'continue': 1, 'de': 1, 'progresser': 1, ' ': 2, 'avec': 2, 'des': 1, 'développements': 1, 'passionnants': 1, 'à': 1, 'venir': 1, 'tels': 1, 'que': 1, "l'IA": 1, 'générale': 1, 'artificielle': 1, '(': 1, 'AGI': 1, ')': 1, 'qui': 1, 'pourrait': 1, 'rivaliser': 1, "l'intelligence": 1, 'humaine': 1, '.': 1}

#### Document 7 Information:

Content: TensorFlow et PyTorch sont des bibliothèques populaires utilisées par les chercheurs et les développeurs pour construire des modèles d'IA.

Word n-grams: [('TensorFlow', 'et', 'PyTorch'), ('et', 'PyTorch', 'sont'), ('PyTorch', 'sont', 'des'), ('sont', 'des', 'bibliothèques'), ('des', 'bibliothèques', 'populaires'), ('bibliothèques', 'populaires', 'utilisées'), ('populaires', 'utilisées', 'par'), ('utilisées', 'par', 'les'), ('par', 'les', 'chercheurs'), ('les', 'chercheurs', 'et'), ('chercheurs', 'et', 'les'), ('et', 'les', 'développeurs'), ('les', 'développeurs', 'pour'), ('développeurs',

'pour', 'construire'), ('pour', 'construire', 'des'), ('construire', 'des', 'modèles'), ('des', 'modèles', "d'IA"), ('modèles', "d'IA", '.')]
Character n-grams: [('T', 'e', 'n'), ('e', 'n', 's'), ('n', 's', 'o'), ('s', 'o', 'r'), ('o', 'r', 'F'), ('r', 'F', 'l'), ('F', 'l', 'o'), ('l', 'o', 'w'), ('o', 'w', ' '), ('w', ' ', 'e'), (' ', 'e', 't'), ('e', 't', ' '), ('t', ' ', 'P'), (' ', 'P', 'y'), ('P', 'y', 'T'), ('y', 'T', 'o'), ('T', 'o', 'r'), ('o', 'r', 'c'), ('r', 'c', 'h'), ('c', 'h', ' '), ('h', ' ', 's'), (' ', 's', 'o'), ('s', 'o', 'n'), ('o', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'b'), (' ', 'b', 'i'), ('b', 'i', 'b'), ('i', 'b', 'l'), ('b', 'l', 'i'), ('l', 'i', 'o'), ('i', 'o', 't'), ('o', 't', 'h'), ('t', 'h', 'è'), ('h', 'è', 'q'), ('è', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'p'), (' ', 'p', 'o'), ('p', 'o', 'p'), ('o', 'p', 'u'), ('p', 'u', 'l'), ('u', 'l', 'a'), ('l', 'a', 'i'), ('a', 'i', 'r'), ('i', 'r', 'e'), ('r', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'u'), (' ', 'u', 't'), ('u', 't', 'i'), ('t', 'i', 'l'), ('i', 'l', 'i'), ('l', 'i', 's'), ('i', 's', 'é'), ('s', 'é', 'e'), ('é', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'p'), (' ', 'p', 'a'), ('p', 'a', 'r'), ('a', 'r', ' '), ('r', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'c'), (' ', 'c', 'h'), ('c', 'h', 'e'), ('h', 'e', 'r'), ('e', 'r', 'c'), ('r', 'c', 'h'), ('c', 'h', 'e'), ('h', 'e', 'u'), ('e', 'u', 'r'), ('u', 'r', 's'), ('r', 's', ' '), ('s', ' ', 'e'), (' ', 'e', 't'), ('e', 't', ' '), ('t', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', 'é'), ('d', 'é', 'v'), ('é', 'v', 'e'), ('v', 'e', 'l'), ('e', 'l', 'o'), ('l', 'o', 'p'), ('o', 'p', 'p'), ('p', 'p', 'e'), ('p', 'e', 'u'), ('e', 'u', 'r'), ('u', 'r', 's'), ('r', 's', ' '), ('s', ' ', 'p'), (' ', 'p', 'o'), ('p', 'o', 'u'), ('o', 'u', 'r'), ('u', 'r', ' '), ('r', ' ', 'c'), (' ', 'c', 'o'), ('c', 'o', 'n'), ('o', 'n', 's'), ('n', 's', 't'), ('s', 't', 'r'), ('t', 'r', 'u'), ('r', 'u', 'i'), ('u', 'i', 'r'), ('i', 'r', 'e'), ('r', 'e', ' '), ('e', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'm'), (' ', 'm', 'o'), ('m', 'o', 'd'), ('o', 'd', 'è'), ('d', 'è', 'l'), ('è', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'I'), (' ', 'I', 'A'), ('I', 'A', '.')]
Word Frequency Distribution: {'TensorFlow': 1, 'et': 2, 'PyTorch': 1, 'sont': 1, 'des': 2, 'bibliothèques': 1, 'populaires': 1, 'utilisées': 1, 'par': 1, 'les': 2, 'chercheurs': 1, 'développeurs': 1, 'pour': 1, 'construire': 1, 'modèles': 1, "d'IA": 1, '.': 1}

#### Document 8 Information:

Content: Alan Turing et John McCarthy sont des figures clés de l'histoire de l'IA. Leurs travaux ont jeté les bases de ce que nous connaissons aujourd'hui en matière d'intelligence artificielle.

Word n-grams: [('Alan', 'Turing', 'et'), ('Turing', 'et', 'John'), ('et', 'John', 'McCarthy'), ('John', 'McCarthy', 'sont'), ('McCarthy', 'sont', 'des'), ('sont', 'des', 'figures'), ('des', 'figures', 'clés'), ('figures', 'clés', 'de'), ('clés', 'de', "l'histoire"), ('de', "l'histoire", 'de'), ("l'histoire", 'de', "l'IA"), ('de', "l'IA", '.'), ("l'IA", '.', 'Leurs'), ('.', 'Leurs', 'travaux'), ('Leurs', 'travaux', 'ont'), ('travaux', 'ont', 'jeté'), ('ont',

'jeté', 'les'), ('jeté', 'les', 'bases'), ('les', 'bases', 'de'), ('bases', 'de', 'ce'), ('de', 'ce', 'que'), ('ce', 'que', 'nous'), ('que', 'nous', 'connaissons'), ('nous', 'connaissons', 'aujourd'hui'), ('connaissons', 'aujourd'hui', 'en'), ('aujourd'hui', 'en', 'matière'), ('en', 'matière', 'd'intelligence'), ('matière', 'd'intelligence', 'artificielle'), ('d'intelligence', 'artificielle', '.')] ]

Character n-grams: [('A', 'l', 'a'), ('l', 'a', 'n'), ('a', 'n', ' '), ('n', ' ', 'T'), (' ', 'T', 'u'), ('T', 'u', 'r'), ('u', 'r', 'i'), ('r', 'i', 'n'), ('i', 'n', 'g'), ('n', 'g', ' '), ('g', ' ', 'e'), (' ', 'e', 't'), ('e', 't', ' '), ('t', ' ', 'J'), (' ', 'J', 'o'), ('J', 'o', 'h'), ('o', 'h', 'n'), ('h', 'n', ' '), ('n', ' ', 'M'), (' ', 'M', 'c'), ('M', 'c', 'C'), ('c', 'C', 'a'), ('C', 'a', 'r'), ('a', 'r', 't'), ('r', 't', 'h'), ('t', 'h', 'y'), ('h', 'y', ' '), ('y', ' ', 's'), (' ', 's', 'o'), ('s', 'o', 'n'), ('o', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'f'), (' ', 'f', 'i'), ('f', 'i', 'g'), ('i', 'g', 'u'), ('g', 'u', 'r'), ('u', 'r', 'e'), ('r', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'c'), (' ', 'c', 'l'), ('c', 'l', 'é'), ('l', 'é', 's'), ('é', 's', ' '), ('s', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'l'), (' ', 'l', ' '), ('l', ' ', 'h'), (' ', 'h', 'i'), ('h', 'i', 's'), ('i', 's', 't'), ('s', 't', 'o'), ('t', 'o', 'i'), ('o', 'i', 'r'), ('i', 'r', 'e'), ('r', 'e', ' '), ('e', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'l'), (' ', 'l', ' '), ('l', ' ', 'I'), (' ', 'I', 'A'), ('I', 'A', ' '), ('A', ' ', ' '), (' ', ' ', 'L'), (' ', 'L', 'e'), ('L', 'e', 'u'), ('e', 'u', 'r'), ('u', 'r', 's'), ('r', 's', ' '), ('s', ' ', 't'), (' ', 't', 'r'), ('t', 'r', 'a'), ('r', 'a', 'v'), ('a', 'v', 'a'), ('v', 'a', 'u'), ('a', 'u', 'x'), ('u', 'x', ' '), ('x', ' ', 'o'), (' ', 'o', 'n'), ('o', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'j'), (' ', 'j', 'e'), ('j', 'e', 't'), ('e', 't', 'é'), ('t', 'é', ' '), ('é', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'b'), (' ', 'b', 'a'), ('b', 'a', 's'), ('a', 's', 'e'), ('s', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'c'), (' ', 'c', 'e'), ('c', 'e', ' '), ('e', ' ', 'q'), (' ', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '), ('e', ' ', 'n'), (' ', 'n', 'o'), ('n', 'o', 'u'), ('o', 'u', 's'), ('u', 's', ' '), ('s', ' ', 'c'), (' ', 'c', 'o'), ('c', 'o', 'n'), ('o', 'n', 'n'), ('n', 'n', 'a'), ('n', 'a', 'i'), ('a', 'i', 's'), ('i', 's', 's'), ('s', 's', 'o'), ('s', 'o', 'n'), ('o', 'n', 's'), ('n', 's', ' '), ('s', ' ', 'a'), (' ', 'a', 'u'), ('a', 'u', 'j'), ('u', 'j', 'o'), ('j', 'o', 'u'), ('o', 'u', 'r'), ('u', 'r', 'd'), ('r', 'd', ' '), ('d', ' ', 'h'), (' ', 'h', 'u'), ('h', 'u', 'i'), ('u', 'i', ' '), ('i', ' ', 'e'), (' ', 'e', 'n'), ('e', 'n', ' '), ('n', ' ', 'm'), (' ', 'm', 'a'), ('m', 'a', 't'), ('a', 't', 'i'), ('t', 'i', 'è'), ('i', 'è', 'r'), ('è', 'r', 'e'), ('r', 'e', ' '), ('e', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'i'), (' ', 'i', 'n'), ('i', 'n', 't'), ('n', 't', 'e'), ('t', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'i'), ('l', 'i', 'g'), ('i', 'g', 'e'), ('g', 'e', 'n'), ('e', 'n', 'c'), ('n', 'c', 'e'), ('c', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'r'), ('a', 'r', 't'), ('r', 't', 'i'), ('t', 'i', 'f'), ('i', 'f', 'i'), ('f', 'i', 'c'), ('i', 'c', 'i'), ('c', 'i', 'e'), ('i', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'e'), ('l', 'e', ' ')] ]

Word Frequency Distribution: {'Alan': 1, 'Turing': 1, 'et': 1, 'John': 1, 'McCarthy': 1, 'sont': 1, 'des': 1, 'figures': 1, 'clés': 1, 'de': 3,

"l'histoire": 1, "l'IA": 1, '.:': 2, 'Leurs': 1, 'travaux': 1, 'ont': 1, 'jeté': 1, 'les': 1, 'bases': 1, 'ce': 1, 'que': 1, 'nous': 1, 'connaissons': 1, "aujourd'hui": 1, 'en': 1, 'matière': 1, "d'intelligence": 1, 'artificielle': 1}

#### Document 9 Information:

Content: L'intelligence artificielle (IA) est un domaine passionnant de l'informatique qui vise à créer des systèmes capables d'effectuer des tâches qui nécessitent généralement l'intelligence humaine. Ces systèmes sont conçus pour imiter certaines capacités humaines, telles que la vision, la compréhension du langage naturel, l'apprentissage et la résolution de problèmes complexes. L'IA englobe un large éventail de domaines, de la robotique aux algorithmes d'apprentissage automatique.

Les chercheurs en IA travaillent sur le développement de modèles d'IA capables de prendre des décisions éclairées, de résoudre des problèmes complexes et d'interagir de manière plus naturelle avec les humains. Ils utilisent des techniques d'apprentissage automatique, de traitement du langage naturel, de vision par ordinateur et d'apprentissage profond pour créer ces systèmes intelligents.

Word n-grams: [("L'intelligence", 'artificielle', '('), ('artificielle', '(', 'IA)'), ('(', 'IA', ')'), ('IA', ')', 'est'), (')', 'est', 'un'), ('est', 'un', 'domaine'), ('un', 'domaine', 'passionnant'), ('domaine', 'passionnant', 'de'), ('passionnant', 'de', "l'informatique"), ('de', "l'informatique", 'qui'), ("l'informatique", 'qui', 'vise'), ('qui', 'vise', 'à'), ('vise', 'à', 'créer'), ('à', 'créer', 'des'), ('créer', 'des', 'systèmes'), ('des', 'systèmes', 'capables'), ('systèmes', 'capables', 'd'effectuer'), ('capables', 'd'effectuer', 'des'), ("d'effectuer", 'des', 'tâches'), ('des', 'tâches', 'qui'), ('tâches', 'qui', 'nécessitent'), ('qui', 'nécessitent', 'généralement'), ('nécessitent', 'généralement', "l'intelligence"), ('généralement', "l'intelligence", 'humaine'), ("l'intelligence", 'humaine', '.), ('humaine', '.), 'Ces'), ('.', 'Ces', 'systèmes'), ('Ces', 'systèmes', 'sont'), ('systèmes', 'sont', 'conçus'), ('sont', 'conçus', 'pour'), ('conçus', 'pour', 'imiter'), ('pour', 'imiter', 'certaines'), ('imiter', 'certaines', 'capacités'), ('certaines', 'capacités', 'humaines'), ('capacités', 'humaines', ','), ('humaines', ',', 'telles'), ('.', 'telles', 'que'), ('telles', 'que', 'la'), ('que', 'la', 'vision'), ('la', 'vision', ','), ('vision', ',', 'la'), ('.', 'la', 'compréhension'), ('la', 'compréhension', 'du'), ('compréhension', 'du', 'langage'), ('du', 'langage', 'naturel'), ('langage', 'naturel', ','), ('naturel', ',', "l'apprentissage"), ('.', "l'apprentissage", 'et'), ("l'apprentissage", 'et', 'la'), ('et', 'la', 'résolution'), ('la', 'résolution', 'de'), ('résolution', 'de', 'problèmes'), ('de', 'problèmes', 'complexes'), ('problèmes', 'complexes', '.), ('complexes', '.), "L'IA"), ('.', "L'IA", 'englobe'), ("L'IA", 'englobe', 'un'), ('englobe', 'un', 'large'), ('un', 'large', 'éventail'), ('large', 'éventail', 'de'), ('éventail', 'de', 'domaines'), ('de', 'domaines', ','), ('domaines', ',', 'de'), ('.', 'de', 'la'), ('de', 'la', 'robotique'), ('la', 'robotique', 'aux'), ('robotique', 'aux', 'algorithmes'), ('aux', 'algorithmes', "d'apprentissage"),



('algorithmes', 'd'apprentissage', 'automatique'), ('d'apprentissage', 'automatique', '.'), ('automatique', '.', 'Les'), ('.', 'Les', 'chercheurs'), ('Les', 'chercheurs', 'en'), ('chercheurs', 'en', 'IA'), ('en', 'IA', 'travaillent'), ('IA', 'travaillent', 'sur'), ('travaillent', 'sur', 'le'), ('sur', 'le', 'développement'), ('le', 'développement', 'de'), ('développement', 'de', 'modèles'), ('de', 'modèles', 'd'IA'), ('modèles', 'd'IA', 'capables'), ('d'IA', 'capables', 'de'), ('capables', 'de', 'prendre'), ('de', 'prendre', 'des'), ('prendre', 'des', 'décisions'), ('des', 'décisions', 'éclairées'), ('décisions', 'éclairées', ','), ('éclairées', ',', 'de'), ('', 'de', 'résoudre'), ('de', 'résoudre', 'des'), ('résoudre', 'des', 'problèmes'), ('des', 'problèmes', 'complexes'), ('problèmes', 'complexes', 'et'), ('complexes', 'et', 'd'interagir'), ('et', 'd'interagir', 'de'), ('d'interagir', 'de', 'manière'), ('de', 'manière', 'plus'), ('manière', 'plus', 'naturelle'), ('plus', 'naturelle', 'avec'), ('naturelle', 'avec', 'les'), ('avec', 'les', 'humains'), ('les', 'humains', '.'), ('humains', '.', 'Ils'), ('.', 'Ils', 'utilisent'), ('Ils', 'utilisent', 'des'), ('utilisent', 'des', 'techniques'), ('des', 'techniques', 'd'apprentissage'), ('techniques', 'd'apprentissage', 'automatique'), ('d'apprentissage', 'automatique', ','), ('automatique', ',', 'de'), ('', 'de', 'traitement'), ('de', 'traitement', 'du'), ('traitement', 'du', 'langage'), ('du', 'langage', 'naturel'), ('langage', 'naturel', ','), ('naturel', ',', 'de'), ('', 'de', 'vision'), ('de', 'vision', 'par'), ('vision', 'par', 'ordinateur'), ('par', 'ordinateur', 'et'), ('ordinateur', 'et', 'd'apprentissage'), ('et', 'd'apprentissage', 'profond'), ('d'apprentissage', 'profond', 'pour'), ('profond', 'pour', 'créer'), ('pour', 'créer', 'ces'), ('créer', 'ces', 'systèmes'), ('ces', 'systèmes', 'intelligents'), ('systèmes', 'intelligents', '.')] ]

Character n-grams: [('L', '', 'i'), ('', 'i', 'n'), ('i', 'n', 't'), ('n', 't', 'e'), ('t', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'i'), ('l', 'i', 'g'), ('i', 'g', 'e'), ('g', 'e', 'n'), ('e', 'n', 'c'), ('n', 'c', 'e'), ('c', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'r'), ('a', 'r', 't'), ('r', 't', 'i'), ('t', 'i', 'f'), ('i', 'f', 'i'), ('f', 'i', 'c'), ('i', 'c', 'i'), ('c', 'i', 'e'), ('i', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'e'), ('l', 'e', ' '), ('e', ' ', ' '), (' ', ' ', ' '), (' ', 'I'), (' ', 'I', 'A'), ('I', 'A', ' '), ('A', ' '), (' '), (' ', 'e'), (' ', 'e', 's'), ('e', 's', 't'), ('s', 't', ' '), ('t', ' ', 'u'), (' ', 'u', 'n'), ('u', 'n', ' '), ('n', ' ', 'd'), (' ', 'd', 'o'), ('d', 'o', 'm'), ('o', 'm', 'a'), ('m', 'a', 'i'), ('a', 'i', 'n'), ('i', 'n', 'e'), ('n', 'e', ' '), ('e', ' ', 'p'), (' ', 'p', 'a'), ('p', 'a', 's'), ('a', 's', 's'), ('s', 's', 'i'), ('s', 'i', 'o'), ('i', 'o', 'n'), ('o', 'n', 'n'), ('n', 'n', 'a'), ('n', 'a', 'n'), ('a', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'l'), (' ', 'l', 'l'), ('l', 'l', 'i'), ('l', 'i', 'n'), ('i', 'n', 'f'), ('n', 'f', 'o'), ('f', 'o', 'r'), ('o', 'r', 'm'), ('r', 'm', 'a'), ('m', 'a', 't'), ('a', 't', 'i'), ('t', 'i', 'q'), ('i', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '), ('e', ' ', 'q'), (' ', 'q', 'u'), ('q', 'u', 'i'), ('u', 'i', ' '), ('i', ' ', 'v'), (' ', 'v', 'i'), ('v', 'i', 's'), ('i', 's', 'e'), ('s', 'e', ' '), ('e', ' ', 'à'), (' ', 'à', ' '), ('à', ' ', 'c'), (' ', 'c', 'r'), ('c', 'r', 'é'), ('r', 'é', 'e'), ('é', 'e', 'r'), ('e', 'r', ' '), ('r', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 's'), (' ', 's', 'y'), ('s', 'y', 's'), ('y', 's', 't'), ('s',

't', 'è'), ('t', 'è', 'm'), ('è', 'm', 'e'), ('m', 'e', 's'), ('e', 's', ' '),  
 ('s', ' ', 'c'), (' ', 'c', 'a'), ('c', 'a', 'p'), ('a', 'p', 'a'), ('p', 'a',  
 'b'), ('a', 'b', 'l'), ('b', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s',  
 ' ', 'd'), (' ', 'd', '"'), ('d', '"', 'e'), ('"', 'e', 'f'), ('e', 'f', 'f'),  
 ('f', 'f', 'e'), ('f', 'e', 'c'), ('e', 'c', 't'), ('c', 't', 'u'), ('t', 'u',  
 'e'), ('u', 'e', 'r'), ('e', 'r', ' '), ('r', ' ', 'd'), (' ', 'd', 'e'), ('d',  
 'e', 's'), ('e', 's', ' '), ('s', ' ', 't'), (' ', 't', 'â'), ('t', 'â', 'c'),  
 ('â', 'c', 'h'), ('c', 'h', 'e'), ('h', 'e', 's'), ('e', 's', ' '), ('s', ' ',  
 'q'), (' ', 'q', 'u'), ('q', 'u', 'i'), ('u', 'i', ' '), ('i', ' ', 'n'), (' ',  
 'n', 'é'), ('n', 'é', 'c'), ('é', 'c', 'e'), ('c', 'e', 's'), ('e', 's', 's'),  
 ('s', 's', 'i'), ('s', 'i', 't'), ('i', 't', 'e'), ('t', 'e', 'n'), ('e', 'n',  
 't'), ('n', 't', ' '), ('t', ' ', 'g'), (' ', 'g', 'é'), ('g', 'é', 'n'), ('é',  
 'n', 'é'), ('n', 'é', 'r'), ('é', 'r', 'a'), ('r', 'a', 'l'), ('a', 'l', 'e'),  
 ('l', 'e', 'm'), ('e', 'm', 'e'), ('m', 'e', 'n'), ('e', 'n', 't'), ('n', 't', ' '),  
 ('t', ' ', 'l'), (' ', 'l', '"'), ('l', '"', 'i'), ('"', 'i', 'n'), ('i',  
 'n', 't'), ('n', 't', 'e'), ('t', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'i'),  
 ('l', 'i', 'g'), ('i', 'g', 'e'), ('g', 'e', 'n'), ('e', 'n', 'c'), ('n', 'c',  
 'e'), ('c', 'e', ' '), ('e', ' ', 'h'), (' ', 'h', 'u'), ('h', 'u', 'm'), ('u',  
 'm', 'a'), ('m', 'a', 'i'), ('a', 'i', 'n'), ('i', 'n', 'e'), ('n', 'e', ' '),  
 ('e', ' ', ' '), (' ', ' ', 'C'), (' ', 'C', 'e'), ('C', 'e', 's'), ('e', 's', ' '),  
 ('s', ' ', 's'), (' ', 's', 'y'), ('s', 'y', 's'), ('y', 's', 't'), ('s',  
 't', 'è'), ('t', 'è', 'm'), ('è', 'm', 'e'), ('m', 'e', 's'), ('e', 's', ' '),  
 ('s', ' ', 's'), (' ', 's', 'o'), ('s', 'o', 'n'), ('o', 'n', 't'), ('n', 't', ' '),  
 ('t', ' ', 'c'), (' ', 'c', 'o'), ('c', 'o', 'n'), ('o', 'n', 'ç'), ('n',  
 'ç', 'u'), ('ç', 'u', 's'), ('u', 's', ' '), ('s', ' ', 'p'), (' ', 'p', 'o'),  
 ('p', 'o', 'u'), ('o', 'u', 'r'), ('u', 'r', ' '), ('r', ' ', 'i'), (' ', 'i',  
 'm'), ('i', 'm', 'i'), ('m', 'i', 't'), ('i', 't', 'e'), ('t', 'e', 'r'), ('e',  
 'r', ' '), ('r', ' ', 'c'), (' ', 'c', 'e'), ('c', 'e', 'r'), ('e', 'r', 't'),  
 ('r', 't', 'a'), ('t', 'a', 'i'), ('a', 'i', 'n'), ('i', 'n', 'e'), ('n', 'e',  
 's'), ('e', 's', ' '), ('s', ' ', 'c'), (' ', 'c', 'a'), ('c', 'a', 'p'), ('a',  
 'p', 'a'), ('p', 'a', 'c'), ('a', 'c', 'i'), ('c', 'i', 't'), ('i', 't', 'é'),  
 ('t', 'é', 's'), ('é', 's', ' '), ('s', ' ', 'h'), (' ', 'h', 'u'), ('h', 'u',  
 'm'), ('u', 'm', 'a'), ('m', 'a', 'i'), ('a', 'i', 'n'), ('i', 'n', 'e'), ('n',  
 'e', 's'), ('e', 's', ' '), ('s', ' ', ' '), (' ', ' ', 't'), (' ', 't', 'e'),  
 ('t', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '),  
 ('s', ' ', 'q'), (' ', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '), ('e', ' ',  
 'l'), (' ', 'l', 'a'), ('l', 'a', ' '), ('a', ' ', 'v'), (' ', 'v', 'i'),  
 ('v', 'i', 's'), ('i', 's', 'i'), ('s', 'i', 'o'), ('i', 'o', 'n'), ('o', 'n',  
 ' '), ('n', ' ', 'l'), (' ', 'l', 'a'), ('l', 'a', ' '), ('a',  
 ' ', 'c'), (' ', 'c', 'o'), ('c', 'o', 'm'), ('o', 'm', 'p'), ('m', 'p', 'r'),  
 ('p', 'r', 'é'), ('r', 'é', 'h'), ('é', 'h', 'e'), ('h', 'e', 'n'), ('e', 'n',  
 's'), ('n', 's', 'i'), ('s', 'i', 'o'), ('i', 'o', 'n'), ('o', 'n', ' '), ('n',  
 ' ', 'd'), (' ', 'd', 'u'), ('d', 'u', ' '), ('u', ' ', 'l'), (' ', 'l', 'a'),  
 ('l', 'a', 'n'), ('a', 'n', 'g'), ('n', 'g', 'a'), ('g', 'a', 'g'), ('a', 'g',  
 'e'), ('g', 'e', ' '), ('e', ' ', 'n'), (' ', 'n', 'a'), ('n', 'a', 't'), ('a',  
 't', 'u'), ('t', 'u', 'r'), ('u', 'r', 'e'), ('r', 'e', 'l'), ('e', 'l', ' '),  
 ('l', ' ', ' '), (' ', ' ', 'l'), (' ', 'l', '"'), ('l', '"', 'a'), ('"', 'a',  
 'p'), ('a', 'p', 'p'), ('p', 'p', 'r'), ('p', 'r', 'e'), ('r', 'e', 'n'), ('e',

'n', 't'), ('n', 't', 'i'), ('t', 'i', 's'), ('i', 's', 's'), ('s', 's', 'a'),  
('s', 'a', 'g'), ('a', 'g', 'e'), ('g', 'e', ' '), ('e', ' ', 'e'), (' ', 'e',  
't'), ('e', 't', ' '), ('t', ' ', 'l'), (' ', 'l', 'a'), ('l', 'a', ' '), ('a',  
' ', 'r'), (' ', 'r', 'é'), ('r', 'é', 's'), ('é', 's', 'o'), ('s', 'o', 'l'),  
('o', 'l', 'u'), ('l', 'u', 't'), ('u', 't', 'i'), ('t', 'i', 'o'), ('i', 'o',  
'n'), ('o', 'n', ' '), ('n', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e',  
' ', 'p'), (' ', 'p', 'r'), ('p', 'r', 'o'), ('r', 'o', 'b'), ('o', 'b', 'l'),  
('b', 'l', 'è'), ('l', 'è', 'm'), ('è', 'm', 'e'), ('m', 'e', 's'), ('e', 's', ' '),  
'), ('s', ' ', 'c'), (' ', 'c', 'o'), ('c', 'o', 'm'), ('o', 'm', 'p'), ('m',  
'p', 'l'), ('p', 'l', 'e'), ('l', 'e', 'x'), ('e', 'x', 'e'), ('x', 'e', 's'),  
('e', 's', ' '), ('s', ' ', ' '), (' ', ' ', 'L'), (' ', 'L', ' '), ('L', ' '),  
'I'), (' ', 'I', 'A'), ('I', 'A', ' '), ('A', ' ', 'e'), (' ', 'e', 'n'), ('e',  
'n', 'g'), ('n', 'g', 'l'), ('g', 'l', 'o'), ('l', 'o', 'b'), ('o', 'b', 'e'),  
('b', 'e', ' '), ('e', ' ', 'u'), (' ', 'u', 'n'), ('u', 'n', ' '), ('n', ' ',  
'l'), (' ', 'l', 'a'), ('l', 'a', 'r'), ('a', 'r', 'g'), ('r', 'g', 'e'), ('g',  
'e', ' '), ('e', ' ', 'é'), (' ', 'é', 'v'), ('é', 'v', 'e'), ('v', 'e', 'n'),  
('e', 'n', 't'), ('n', 't', 'a'), ('t', 'a', 'i'), ('a', 'i', 'l'), ('i', 'l', ' '),  
'), ('l', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'd'), (' ',  
'd', 'o'), ('d', 'o', 'm'), ('o', 'm', 'a'), ('m', 'a', 'i'), ('a', 'i', 'n'),  
('i', 'n', 'e'), ('n', 'e', 's'), ('e', 's', ' '), ('s', ' ', ' '), (' ', ' ',  
'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'l'), (' ', 'l', 'a'), ('l',  
'a', ' '), ('a', ' ', 'r'), (' ', 'r', 'o'), ('r', 'o', 'b'), ('o', 'b', 'o'),  
('b', 'o', 't'), ('o', 't', 'i'), ('t', 'i', 'q'), ('i', 'q', 'u'), ('q', 'u',  
'e'), ('u', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'u'), ('a', 'u', 'x'), ('u',  
'x', ' '), ('x', ' ', 'a'), (' ', 'a', 'l'), ('a', 'l', 'g'), ('l', 'g', 'o'),  
('g', 'o', 'r'), ('o', 'r', 'i'), ('r', 'i', 't'), ('i', 't', 'h'), ('t', 'h',  
'm'), ('h', 'm', 'e'), ('m', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ',  
'd', ' '), ('d', ' ', 'a'), (' ', 'a', 'p'), ('a', 'p', 'p'), ('p', 'p', 'r'),  
('p', 'r', 'e'), ('r', 'e', 'n'), ('e', 'n', 't'), ('n', 't', 'i'), ('t', 'i',  
's'), ('i', 's', 's'), ('s', 's', 'a'), ('s', 'a', 'g'), ('a', 'g', 'e'), ('g',  
'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'u'), ('a', 'u', 't'), ('u', 't', 'o'),  
('t', 'o', 'm'), ('o', 'm', 'a'), ('m', 'a', 't'), ('a', 't', 'i'), ('t', 'i',  
'q'), ('i', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '), ('e', ' ', '\n'), (' ',  
'\n'), (' \n', '\n', 'L'), (' \n', 'L', 'e'), ('L', 'e', 's'), ('e', 's', ' '),  
'), ('s', ' ', 'c'), (' ', 'c', 'h'), ('c', 'h', 'e'), ('h', 'e', 'r'), ('e',  
'r', 'c'), ('r', 'c', 'h'), ('c', 'h', 'e'), ('h', 'e', 'u'), ('e', 'u', 'r'),  
('u', 'r', 's'), ('r', 's', ' '), ('s', ' ', 'e'), (' ', 'e', 'n'), ('e', 'n', ' '),  
'), ('n', ' ', 'I'), (' ', 'I', 'A'), ('I', 'A', ' '), ('A', ' ', 't'), (' ',  
't', 'r'), ('t', 'r', 'a'), ('r', 'a', 'v'), ('a', 'v', 'a'), ('v', 'a', 'i'),  
('a', 'i', 'l'), ('i', 'l', 'l'), ('l', 'l', 'e'), ('l', 'e', 'n'), ('e', 'n',  
't'), ('n', 't', ' '), ('t', ' ', 's'), (' ', 's', 'u'), ('s', 'u', 'r'), ('u',  
'r', ' '), ('r', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', ' '), ('e', ' ', 'd'),  
(' ', 'd', 'é'), ('d', 'é', 'v'), ('é', 'v', 'e'), ('v', 'e', 'l'), ('e', 'l',  
'o'), ('l', 'o', 'p'), ('o', 'p', 'p'), ('p', 'p', 'e'), ('p', 'e', 'm'), ('e',  
'm', 'e'), ('m', 'e', 'n'), ('e', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'd'),  
(' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'm'), (' ', 'm', 'o'), ('m', 'o',  
'd'), ('o', 'd', 'è'), ('d', 'è', 'l'), ('è', 'l', 'e'), ('l', 'e', 's'), ('e',  
's', ' '), ('s', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'I'), (' ', 'I', 'A'),

('I', 'A', ' '), ('A', ' ', 'c'), (' ', 'c', 'a'), ('c', 'a', 'p'), ('a', 'p', 'a'), ('p', 'a', 'b'), ('a', 'b', 'l'), ('b', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'p'), (' ', 'p', 'r'), ('p', 'r', 'e'), ('r', 'e', 'n'), ('e', 'n', 'd'), ('n', 'd', 'r'), ('d', 'r', 'e'), ('r', 'e', ' '), ('e', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', 'é'), ('d', 'é', 'c'), ('é', 'c', 'i'), ('c', 'i', 's'), ('i', 's', 'i'), ('s', 'i', 'o'), ('i', 'o', 'n'), ('o', 'n', 's'), ('n', 's', ' '), ('s', ' ', 'é'), (' ', 'é', 'c'), ('é', 'c', 'l'), ('c', 'l', 'a'), ('l', 'a', 'i'), ('a', 'i', 'r'), ('i', 'r', 'é'), ('r', 'é', 'e'), ('é', 'e', 's'), ('e', 's', ' '), ('s', ' ', ' '), (' ', ' ', ' '), ('d', ' '), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'r'), (' ', 'r', 'é'), ('r', 'é', 's'), ('é', 's', 'o'), ('s', 'o', 'u'), ('o', 'u', 'd'), ('u', 'd', 'r'), ('d', 'r', 'e'), ('r', 'e', ' '), ('e', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'p'), (' ', 'p', 'r'), ('p', 'r', 'o'), ('r', 'o', 'b'), ('o', 'b', 'l'), ('b', 'l', 'è'), ('l', 'è', 'm'), ('è', 'm', 'e'), ('m', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'c'), (' ', 'c', 'o'), ('c', 'o', 'm'), ('o', 'm', 'p'), ('m', 'p', 'l'), ('p', 'l', 'e'), ('l', 'e', 'x'), ('e', 'x', 'e'), ('x', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'e'), (' ', 'e', 't'), ('e', 't', ' '), ('t', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'i'), (' ', 'i', 'n'), ('i', 'n', 't'), ('n', 't', 'e'), ('t', 'e', 'r'), ('e', 'r', 'a'), ('r', 'a', 'g'), ('a', 'g', 'i'), ('g', 'i', 'r'), ('i', 'r', ' '), ('r', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'm'), (' ', 'm', 'a'), ('m', 'a', 'n'), ('a', 'n', 'i'), ('n', 'i', 'è'), ('i', 'è', 'r'), ('è', 'r', 'e'), ('r', 'e', ' '), ('e', ' ', 'p'), (' ', 'p', 'l'), ('p', 'l', 'u'), ('l', 'u', 's'), ('u', 's', ' '), ('s', ' ', 'n'), (' ', 'n', 'a'), ('n', 'a', 't'), ('a', 't', 'u'), ('t', 'u', 'r'), ('u', 'r', 'e'), ('r', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'e'), ('l', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'v'), ('a', 'v', 'e'), ('v', 'e', 'c'), ('e', 'c', ' '), ('c', ' ', 'l'), (' ', 'l', 'e'), ('l', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'h'), (' ', 'h', 'u'), ('h', 'u', 'm'), ('u', 'm', 'a'), ('m', 'a', 'i'), ('a', 'i', 'n'), ('i', 'n', 's'), ('n', 's', ' '), ('s', ' ', ' '), (' ', ' ', 'I'), (' ', 'I', 'l'), ('I', 'l', 's'), ('l', 's', ' '), ('s', ' ', 'u'), (' ', 'u', 't'), ('u', 't', 'i'), ('t', 'i', 'l'), ('i', 'l', 'i'), ('l', 'i', 's'), ('i', 's', 'e'), ('s', 'e', 'n'), ('e', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', 's'), ('e', 's', ' '), ('s', ' ', 't'), (' ', 't', 'e'), ('t', 'e', 'c'), ('e', 'c', 'h'), ('c', 'h', 'n'), ('h', 'n', 'i'), ('n', 'i', 'q'), ('i', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', ' '), ('a'), (' ', 'a', 'p'), ('a', 'p', 'p'), ('p', 'p', 'r'), ('p', 'r', 'e'), ('r', 'e', 'n'), ('e', 'n', 't'), ('n', 't', 'i'), ('t', 'i', 's'), ('i', 's', 's'), ('s', 's', 'a'), ('s', 'a', 'g'), ('a', 'g', 'e'), ('g', 'e', ' '), ('e', ' ', 'a'), (' ', 'a', 'u'), ('a', 'u', 't'), ('u', 't', 'o'), ('t', 'o', 'm'), ('o', 'm', 'a'), ('m', 'a', 't'), ('a', 't', 'i'), ('t', 'i', 'q'), ('i', 'q', 'u'), ('q', 'u', 'e'), ('u', 'e', ' '), ('e', ' ', ' '), (' ', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 't'), (' ', 't', 'r'), ('t', 'r', 'a'), ('r', 'a', 'i'), ('a', 'i', 't'), ('i', 't', 'e'), ('t', 'e', 'm'), ('e', 'm', 'e'), ('m', 'e', 'n'), ('e', 'n', 't'), ('n', 't', ' '), ('t', ' ', 'd'), (' ', 'd', 'u'), ('d', 'u', 'l'), (' ', 'l', 'a'), ('l', 'a', 'n'), ('a', 'n', 'g'), ('n', 'g', 'a'), ('g', 'a', 'g'), ('a', 'g', 'e'), ('g', 'e', ' '),

( 'e', ' ', 'n'), (' ', 'n', 'a'), ('n', 'a', 't'), ('a', 't', 'u'), ('t', 'u', 'r'), ('u', 'r', 'e'), ('r', 'e', 'l'), ('e', 'l', ' ', ' '), ('l', ' ', ' ', ' '), (' ', ' ', 'd'), (' ', 'd', 'e'), ('d', 'e', ' '), ('e', ' ', 'v'), (' ', 'v', 'i'), ('v', 'i', 's'), ('i', 's', 'i'), ('s', 'i', 'o'), ('i', 'o', 'n'), ('o', 'n', ' '), ('n', ' ', 'p'), (' ', 'p', 'a'), ('p', 'a', 'r'), ('a', 'r', ' '), ('r', ' ', 'o'), (' ', 'o', 'r'), ('o', 'r', 'd'), ('r', 'd', 'i'), ('d', 'i', 'n'), ('i', 'n', 'a'), ('n', 'a', 't'), ('a', 't', 'e'), ('t', 'e', 'u'), ('e', 'u', 'r'), ('u', 'r', ' '), ('r', ' ', 'e'), (' ', 'e', 't'), ('e', 't', ' '), ('t', ' ', 'd'), (' ', 'd', ' '), ('d', ' ', 'a'), (' ', 'a', 'p'), ('a', 'p', 'p'), ('p', 'p', 'r'), ('p', 'r', 'e'), ('r', 'e', 'n'), ('e', 'n', 't'), ('n', 't', 'i'), ('t', 'i', 's'), ('i', 's', 's'), ('s', 's', 'a'), ('s', 'a', 'g'), ('a', 'g', 'e'), ('g', 'e', ' '), ('e', ' ', 'p'), (' ', 'p', 'r'), ('p', 'r', 'o'), ('r', 'o', 'f'), ('o', 'f', 'o'), ('f', 'o', 'n'), ('o', 'n', 'd'), ('n', 'd', ' '), ('d', ' ', 'p'), (' ', 'p', 'o'), ('p', 'o', 'u'), ('o', 'u', 'r'), ('u', 'r', ' '), ('r', ' ', 'c'), (' ', 'c', 'r'), ('c', 'r', 'é'), ('r', 'é', 'e'), ('é', 'e', 'r'), ('e', 'r', ' '), ('r', ' ', 'c'), (' ', 'c', 'e'), ('c', 'e', 's'), ('e', 's', ' '), ('s', ' ', 's'), (' ', 's', 'y'), ('s', 'y', 's'), ('y', 's', 't'), ('s', 't', 'è'), ('t', 'è', 'm'), ('è', 'm', 'e'), ('m', 'e', 's'), ('e', 's', ' '), ('s', ' ', 'i'), (' ', 'i', 'n'), ('i', 'n', 't'), ('n', 't', 'e'), ('t', 'e', 'l'), ('e', 'l', 'l'), ('l', 'l', 'i'), ('l', 'i', 'g'), ('i', 'g', 'e'), ('g', 'e', 'n'), ('e', 'n', 't'), ('n', 't', 's'), ('t', 's', ' ')]

Word Frequency Distribution: {"L'intelligence": 1, 'artificielle': 1, '(' : 1, 'IA': 2, ')': 1, 'est': 1, 'un': 2, 'domaine': 1, 'passionnant': 1, 'de': 10, 'l'informatique': 1, 'qui': 2, 'vise': 1, 'à': 1, 'créer': 2, 'des': 5, 'systèmes': 3, 'capables': 2, "d'effectuer": 1, 'tâches': 1, 'nécessitent': 1, 'généralement': 1, "l'intelligence": 1, 'humaine': 1, '.': 5, 'Ces': 1, 'sont': 1, 'conçus': 1, 'pour': 2, 'imiter': 1, 'certaines': 1, 'capacités': 1, 'humaines': 1, ',': 7, 'telles': 1, 'que': 1, 'la': 4, 'vision': 2, 'compréhension': 1, 'du': 2, 'langage': 2, 'naturel': 2, "l'apprentissage": 1, 'et': 3, 'résolution': 1, 'problèmes': 2, 'complexes': 2, "L'IA": 1, 'englobe': 1, 'large': 1, 'éventail': 1, 'domaines': 1, 'robotique': 1, 'aux': 1, 'algorithmes': 1, "d'apprentissage": 3, 'automatique': 2, 'Les': 1, 'chercheurs': 1, 'en': 1, 'travaillent': 1, 'sur': 1, 'le': 1, 'développement': 1, 'modèles': 1, "d'IA": 1, 'prendre': 1, 'décisions': 1, 'éclairées': 1, 'résoudre': 1, "d'interagir": 1, 'manière': 1, 'plus': 1, 'naturelle': 1, 'avec': 1, 'les': 1, 'humains': 1, 'Ils': 1, 'utilisent': 1, 'techniques': 1, 'traitement': 1, 'par': 1, 'ordinateur': 1, 'profond': 1, 'ces': 1, 'intelligents': 1}

[ ]: