Amardeep Singh

E23CSEU2189

```
# ================= IMPORTS =================
from collections import defaultdict
import random
import string

# ================= LOAD DATASET FROM FILE =================
with open("sample text.txt", "r", encoding="utf-8") as f:
    text = f.read()

# ================= PREPROCESSING =================
# Convert to lowercase and remove punctuation
text = text.lower().translate(
    str.maketrans("", "", string.punctuation)
)

# Split into words
words = text.split()

# ================= BUILD BIGRAM MODEL =================
bigram_model = defaultdict(list)

for i in range(len(words) - 1):
    bigram_model[words[i]].append(words[i + 1])

# ================= TEXT GENERATION FUNCTION =================
def generate_bigram(seed, length=30):
    result = [seed]

    for _ in range(length):
        next_words = bigram_model.get(result[-1])
        if not next_words:
            break
        result.append(random.choice(next_words))

    return " ".join(result)

# ================= OUTPUT =================
# Choose a valid random seed from dataset
seed = random.choice(words)

print("Generated Text (Bigram):\n")
print(generate_bigram(seed))
```

```
Generated Text (Bigram):

intelligence also raises ethical and algorithm design will continue to play an even greater role in large datasets neural networks to the education sector artificial intelligence is a meaningful way applications
```