Task 3 : Combat Online Plagiarism with AI

```
import re
from collections import Counter
from math import sqrt



def preprocess_text(text):
    # Lowercase the text and remove non-alphanumeric characters
    text = re.sub(r'\W', ' ', text.lower())
    # Tokenize the text into words
    words = re.findall(r'\w+', text)
    return Counter(words)




def cosine_similarity(vec1, vec2):
    intersection = set(vec1) & set(vec2)
    numerator = sum([vec1[x] * vec2[x] for x in intersection])

    sum1 = sum([vec1[x] ** 2 for x in vec1])
    sum2 = sum([vec2[x] ** 2 for x in vec2])

    denominator = sqrt(sum1) * sqrt(sum2)

    if not denominator:
        return 0.0
    else:
        return float(numerator) / denominator




def detect_plagiarism(original_text, suspicious_text, similarity_threshold=0.8):
    original_vector = preprocess_text(original_text)
    suspicious_vector = preprocess_text(suspicious_text)

    similarity = cosine_similarity(original_vector, suspicious_vector)

    if similarity >= similarity_threshold:
        return True, similarity
    else:
        return False, similarity




# Example usage
original_text = """The internet is flooded with content, making it challenging to spot plagiarism.
Our AI-powered tool can help authors and news organizations quickly detect copied content,
safeguarding original work. Don't let plagiarism go unnoticed; empower yourself with our
plagiarism detection software."""

suspicious_text = """The online world has an abundance of information, which poses a challenge
for identifying plagiarism. Our advanced AI tool aids authors and news outlets in swiftly
identifying duplicated content, ensuring the protection of original work. Avoid overlooking
plagiarism; enhance your capabilities with our plagiarism detection software."""

is_plagiarized, similarity_score = detect_plagiarism(original_text, suspicious_text)

if is_plagiarized:
    print(f"Plagiarism detected! Similarity score: {similarity_score:.2f}")
else:
    print("No plagiarism detected.")
```

```
No plagiarism detected.
```