```
'''Fun with embeddings'''

# Imports
from openai import OpenAI
import numpy as np
import pandas as pd


# 1) Compute embeddings

def get_embeddings(chunks):
    response = OpenAI().embeddings.create(input=chunks, model="text-embedding-3-large")
    return [np.array(record.embedding) for record in response.data]


words = ['employment', 'inflation', 'stocks', 'bubbles']

embeddings = get_embeddings(words)

print(embeddings[0])
print(len(embeddings[0]))


# 2) Compute distances

def cosine_similarity(a, b):
    return np.dot(a, b) / (np.linalg.norm(a) * np.linalg.norm(b))

def distance(a, b):
    return 1 - cosine_similarity(a, b)

for word, e in zip(words, embeddings):
    d = distance(e, embeddings[-1])
    print(f'{word:<20}: {d:.2}')


# 3) Warning
# Of course ideally we would like to compute distances across certain dimensions...
# - optimism-vs-pessimism
# - certainty-vs-doubt
# - formality, language, etc. etc.

# Which one is closer to "bubbles"?
print()
words = ['burbujas', 'bubble', 'tulip mania', 'ubbles', 'bubbles']
embeddings = get_embeddings(words)
for word, e in zip(words, embeddings):
    d = distance(e, embeddings[-1])
    print(f'{word:<20}: {d:.2}')


# 4) Classification
```

```
ecb_speeches = [
'raises the probability that we are also underestimating inflation today',
'inflation will return to our two per cent target by the third quarter of 2025',
'continued high inflation persistence currently remains the largest risk to price stability in the euro area',
'price-setting dynamics could make high inflation stickier',
'a decline in profit margins translating into lower inflationary pressures'
]

embeddings = get_embeddings(ecb_speeches)

statements = ['inflation will be high', 'inflation will be low']
e_high, e_low = get_embeddings(statements)

print()
for text, e in zip(ecb_speeches, embeddings):
    category = 'HIGH' if distance(e, e_high) < distance(e, e_low) else 'LOW '
    print(category, text)


exit()


'''
Classify the following statements into HIGH and LOW types, depending on whether they assess inflation might be high or low in the future:

raises the probability that we are also underestimating inflation today
inflation will return to our two per cent target by the third quarter of 2025
continued high inflation persistence currently remains the largest risk to price stability in the euro area
price-setting dynamics could make high inflation stickier
a decline in profit margins translating into lower inflationary pressures
'''
```