
*Subject (**Chatbot for economic
education**)*

*2022-1st Semester
Undergraduate Research*

Student ID.	20182014
Name	Mai Xuan Man
Department	CSE
Project Advisor	Hyoung hun Kim

I. Introduction

1) Topic and Purpose of Research

I. Introduction

I chose to work on a topic about chatbot. Specifically a chatbot will take the task of answering questions related to the economic information such as financial advice education. The motivation for this work is that economic knowledge is very useful but frequently, it does not receive enough attention from people in daily lives, except for experts and scholars. I hope that this work can contribute to raising people's awareness about the role of economic knowledge and skills in personal life.

II. Related works

Currently, there's many popular chatbots in the market, especially ChatGPT and Good Bard. These tools are obviously too powerful, and they can cover almost every range of information needed. They can answer queries in various domains such as coding, health advice, abstract ideas, etc.

Beside these two main tools, there are several tools that's already popular in the market such as Kore.ai and Kasisto. Kore.ai offers a conversational AI platform that can be customized for various industries, including finance and banking. Their platform allows you to build chatbots that specialize in providing financial services, wealth management, investment advice, and more. And Kasisto provides an AI platform specifically designed for the financial industry. Their chatbot, KAI, is capable of understanding and responding to finance-related queries, providing personalized financial advice, account management, and assisting with financial transactions.

II. Main Subject

1) Research Planning

I. Data Collection Method

I use the OpenAI crawler to collect all the main and subpages and pdf files from the economicprinciples.org site. All the files collected are automatically converted into the text files. This creates a problem in the file format of the pdf files such as below:

<https://economicprinciples.org/downloads/cwo-a-deeper-look-at-capital-wars.pdf.txt>

These files are further changed to pdf format and extracted to txt format using a third party application (pdftotext.com)

```

PS C:\Users\Admin\Desktop\thesis proj> python .\count.py
economicprinciples.org/bw-populism-the-phenomenon.txt
Number of words present in given file: 28863
economicprinciples.org/bwam102317.txt
Number of words present in given file: 43673
economicprinciples.org/cwo-a-deeper-look-at-capital-wars.txt
Number of words present in given file: 8285
economicprinciples.org/cwo-large-drivers-of-life-expectancy-through-time.txt
Number of words present in given file: 3311
economicprinciples.org/economicprinciples.org.txt
Number of words present in given file: 2283
economicprinciples.org/economicprinciples.org...txt
Number of words present in given file: 2283
economicprinciples.org/economicprinciples.org_.txt
Number of words present in given file: 2283
economicprinciples.org/economicprinciples.org_subscribe.txt
Number of words present in given file: 542
economicprinciples.org/economicprinciples.org_Why-and-How-Capitalism-Needs-To-Be-Reformed.txt
Number of words present in given file: 8561
economicprinciples.org/MMT_MP3_MK.txt
Number of words present in given file: 556
economicprinciples.org/Paradigm-Shifts.txt
Number of words present in given file: 17898
economicprinciples.org/Primer-on-Universal-Basic-Income.txt
Number of words present in given file: 9418
Total number of words present in all given files: 127956

```

Fig1.The number of words crawled from the site and pdf file is 127956 words.

II. Processing data

Since the csv file format is used to deal with data, the text files need to be converted to CSV files. This step requires looping through all the text files in the generated directory. After opening each file, empty spacings in the file are removed to reduce cluttering. the modified text is appended to a list. The text then is loaded into an empty Pandas data frame and written to a CSV file.

III. Method

Here I use OpenAI API to build the chatbot. To use the API, it requires 3 steps: generate embeddings, create context and finally answer questions.

1. Generate Embeddings

From the CSV file, text is then split down into tokens by breaking down into sentences and words before being used to retrieve the embedding from the API.

The API can handle input up to 8191 tokens for embedding. To stay below the limit, the text in the CSV file needs to be broken down into multiple rows depending on the length of the row.

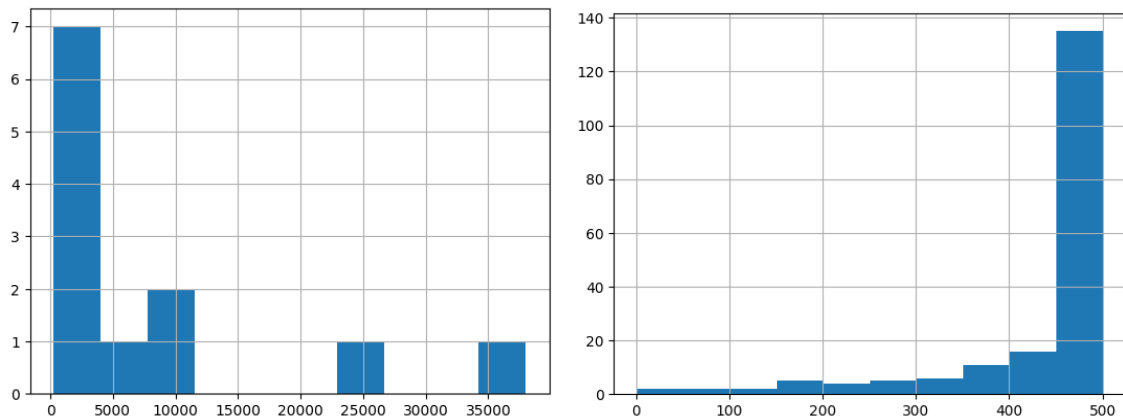


Fig2.The token before(left) and after (right) being chunked

Then, the tokens are ready to be fed into the API (ada-002 model) and generate the embeddings.

	text	n_tokens	embeddings
0	nciples.or	497	[-0.0006809101323597133, -0.012081833556294441, 0.007251798175275326, -0.02323273755609989, -0.00029681780142709613, -0.0008837, 0.018888, -0.02609, -0.00093, 0.028198, -0.00812, 0.002592, -0.00038, 0.002879, 0.004516, 0.021789, 0.036778, 0.025904, -0.00034
1	Watching	483	[-0.0202490221709013, -0.030071210116147995, 0.015640253201127052, -0.023755988106131554, 0.0047229803167283535, -0.0048.23E+08, -0.02214, -0.00044, 0.008358, -0.0369, 0.003544, 0.029776, -0.01027, -0.00337, 0.023474, 0.005509, 0.007679, 0.026725, 0.029265, 0.017602
2	They cove	485	[-0.005826363805681467, -0.028763506561517715, 0.02187429554760456, -0.013125980272889137, -0.007317155133932829, 0.0021.75E+16, -0.00767, 0.005749, -0.02321, -0.0101, 0.043973, 0.000388, -0.00415, 0.010453, 0.010236, 0.003238, 0.02802, 0.026561, 0.00563, 0.001762
3	nciples.or	495	[0.005741580855101347, -0.009123165160417557, 0.01033872552216053, -0.04174284636974335, -0.010965902358293533, 0.0032649, -8.06E-05, 0.01738, -0.0214, -0.00424, 0.031941, -0.01862, 0.001162, 0.004122, -0.00526, 0.008819, 0.016591, -0.00372, -0.00494, -0.01288
4	I believe t	479	[-0.003090199548751116, -0.005741477943956852, -0.0036501470021903515, -0.04257536679506302, -0.024553773924708366, 0.004.94E+11, -0.00753, -0.00726, 0.006965, -0.01949, -0.01068, 0.037954, 0.003131, -0.00791, 0.008094, -0.00365, 0.011121, 0.015685, 0.012774, -0.00628
5	Because I	486	[-0.016110556200146675, -0.026599613949656487, 0.015779880806803703, -0.040130894631147385, -0.012393753044307232, 0.004.65E+14, 0.013915, -0.02369, -0.00907, 0.016984, 0.008188, -0.00393, -0.00627, -0.01138, 0.008822, 0.010661, 0.005307, -0.0241, 0.009259, -0.01034
6	The best r	497	[-0.010917908512055874, -0.01825162209570408, 0.01319936104118824, -0.028805820271372795, -0.0054787928238511086, -0.018.89E+14, -0.00201, -0.00489, -0.0031, -0.03502, -0.01145, 0.032774, 0.002157, 0.000716, 0.001052, -0.01154, 0.023555, 0.018463, 0.019376, -0.01172
7	ThatÃcÃ	495	[-4.514694228419103e-05, -0.023140890523791313, 0.010236380621790886, -0.04058635234832764, -0.001957803964614868, -0.00-0.00747, -0.01986, 0.002219, -0.01593, -0.01356, -0.01742, 0.026579, 0.005583, 0.017266, -0.01476, -0.0141, 0.030196, 0.015072, 0.006026, -0.00365
8	So letÃcÃ	500	[0.0006296807550825179, 0.0020809012930840254, 0.007517880294471979, -0.03076404519379139, 0.003282829886302352, -0.013.73E+13, -0.01903, 0.014996, -0.00817, 0.010488, -0.01484, 0.021788, -0.00457, 0.018285, 0.006805, -0.00649, 0.006176, -0.00108, 0.005274, -0.0281
9	Here are a	464	[-0.007886589504778385, 0.006047045812010765, 0.009848329238593578, -0.04187686741352081, -0.015693917870521545, -0.006

Fig3.A portion of generated embeddings

2. Create Context

With the embedding available, the next step is to create a context for a question. The question is fed into API ada-002 model to generate the question embedding. This embedding is compared with existing embedding to retrieve the most relevant text from scraped websites.

3. Answer questions

From the relevant context, a match answer will be returned. Here API text-davinci-003 model are used to extract the most concise and exact answer by polishing irrelevant information from the context. If the question asks about information beyond the domain of available embedding, the answer will be return "I don't know"

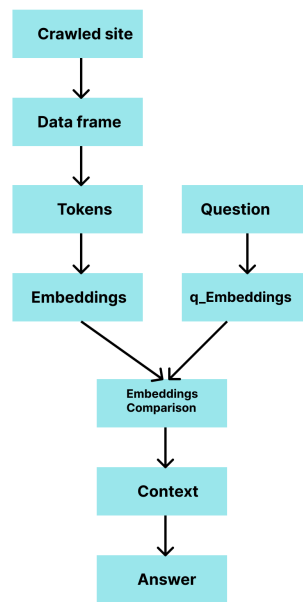


Fig4.The processing flow of the Q&A chatbot

III. Conclusion and Discussion

1) Research Results

The chatbot is able to answer questions in detail within the knowledge provided by the economicprinciples.org site. There are plenty of questions it is able to answer and there are questions it is not. The following are some outcomes of the chatbot, which include successful cases(chatbot able to answer) and unsuccessful cases (chatbot unable to give an answer):

Successful case:

```

[21] answer_question(df, question="why we need economy?")

The current context is:
Because I loved playing the markets I chose to be a global macro investor, which is what Iâve been for about 50 years. That required me to gain a practi
###
When the Federal Reserve (and most other central banks) buys financial assets to put money in the economy in order to stimulate the economy, the sellers c
###
nciples.org Why and How Capitalism Needs To Be Reformed.      Economic Principles - Why and How Capitalism Needs To Be Reformed      Economic Principl
The answer is:
'Ve need an economy to provide basic care, good education, and employment opportunities to people, and to create a system that rewards people for their p
roductive activities with money that can be used to get all that money can buy. Capitalism is an effective motivator of people and allocator of resources
that raises people's living standards.'
```

```
11s ✓ answer_question(df, question="how economy knowledge is applied in personal finance?")

The current context is:
Because I loved playing the markets I chose to be a global macro investor, which is what Iâve been for about 50 years. That required me to gain a practi

###

When the Federal Reserve (and most other central banks) buys financial assets to put money in the economy in order to stimulate the economy, the sellers c

###

Such decisions would of course be up to the people on the bipartisan commission and the leadership to decide and are way too complicated an engineering ex

The answer is:

'Economy knowledge can be applied in personal finance by understanding how different economic systems work and how they can affect personal finances. Thi
s includes understanding how the Federal Reserve and other central banks buy financial assets to put money in the economy in order to stimulate the econo
my, how taxation works, and how to coordinate monetary and fiscal policies. Additionally, understanding the cause and effect relationships and historical
comparisons of different economic systems can help inform personal finance decisions.'
```

```
10s ✓ answer_question(df, question="what have america done in economic crisis?")

The current context is:
As inflation accelerated and the economy weakened in 1969-70, the Fed could not afford to maintain a tight monetary policy, so the balance of payments wor

###

As a result, while in the 1970s, countries with dollar surpluses recycled their dollars via American financial market intermediaries which lent recklessly

###

Bolstered by a decade of prosperity, in the '60s Americans lost their cautiousness and kept their confidence. Since the world was our oyster, conquering c

###

Those companies and people who are richer have greater buying power, which motivates those who seek profit to shift their resources to produce what the ha

The answer is:

'In 2008, the debt and economic crises led the Fed (and other central banks) to ease until interest rates hit 0%, at which point the Fed "printed" more m
oney faster, and spent it buying a broader range of financial assets, than it ever had before. As a result, at the end of the decade, the stock market, g
old, and commodity prices immediately recovered strongly, and the dollar plunged. Also in 2008, as in 1932, the old, right-of-center administration was v
oted out of office and a new left-of-center administration entered office, large government spending programs were initiated, and large budget deficits r
esulted.'
```

```
4s ✓ answer_question(df, question="what activities bring the most finalcial profit?")

The current context is:
How they turn out will have significant effects on how the conflicts raised in this report will be dealt with, which will influence how money will flow be

###

Those companies and people who are richer have greater buying power, which motivates those who seek profit to shift their resources to produce what the ha

###

Such decisions would of course be up to the people on the bipartisan commission and the leadership to decide and are way too complicated an engineering ex

###

That is because they don't provide adequate income. In fact, most of them won't provide any income, so they are worthless for that purpose. They just prov

The answer is:

'Equity and equity-like investments, such as leveraged private equity, leveraged real estate, and venture capital, are likely to bring the most financial
profit.'
```

```
11s ✓ answer_question(df, question="how education make effect on economy of a country?")

The current context is:
Because I loved playing the markets I chose to be a global macro investor, which is what Iâve been for about 50 years. That required me to gain a practi

###

xliv The health consequences and economic costs of low education and poverty are terrible: For example, for those in the bottom 60% premature deaths are c

###

Those companies and people who are richer have greater buying power, which motivates those who seek profit to shift their resources to produce what the ha

The answer is:

'Poor education, a poor culture (one that impedes people from operating effectively together), poor infrastructure, and too much debt cause bad economic
results. The health consequences and economic costs of low education and poverty are terrible, and can lead to an increase in premature deaths, an increa
se in people putting off medical treatment due to cost, and worse health for those who are unemployed or making less than $35,000 per year. These condi
tions pose an existential risk for the US, as they weaken the US economically, threaten to bring about painful and counterproductive domestic conflict, and
undermine the United States' strength relative to that of its global competitors.'
```

Unsuccessful case:

```
[26] answer_question(df, question="what are the richest country nowadays?")

The current context is:
Those companies and people who are richer have greater buying power, which motivates those who seek profit to shift their resources to produce what the ha

###

Thatâs for the population as a whole. For most of those in the lower 60%, the prospects are worse. As shown below, the income gap is about as high as ev

###

USA Interest Rates Monetary Base (%GDP) 20% 24% 15% 19% 10% 14% 5% 9% 4% 0% 60 80 00 8 20 EUR Interest Rates JPN Monetary Base (%GDP) Interest Rates Monet

The answer is:

'I don't know.'
```

```
answer_question(df, question="what currency is strongest?")

The current context is:
By looking at who has what assets and liabilities, asking yourself who the central bank needs to help most, and figuring out what they are 7 most likely t

###

That is because they don't provide adequate income. In fact, most of them won't provide any income, so they are worthless for that purpose. They just prov

###

As a result, the world is leveraged long, holding assets that have low real and nominal expected returns that are also providing historically low returns

The answer is:

'I don't know.'
```

```
answer_question(df, question="which country can have more benefit in enonomic crisis?")

The current context is:
The most relevant causes that are leading to the effects we are seeing are: The high debt levels that led to the 2008 debt crisis (and have since increas

###

Because I loved playing the markets I chose to be a global macro investor, which is what Iâve been for about 50 years. That required me to gain a practi

###

Such decisions would of course be up to the people on the bipartisan commission and the leadership to decide and are way too complicated an engineering ex

The answer is:

'I don't know.'
```

There are at most 5 contexts returned. This increases the chance of finding the most concise answer. Depending on the question, there are cases where the answer can be extracted from the context. And for the question beyond the reach of site information the returned context is totally irrelevant and thus the answer is "I don't know".

2) Discussion

Compared to the previous version of embedding(v1) presented in the slide, where some text files were omitted, the current embedding (v2) uses all the text files generated from the economicprinciple site. And for the answerable question, v1 and v2 give slightly different answers. For the question that v1 gives "I don't know", then v2 also returns "I don't know".

The possible reason for such a pattern is that all the information covered in the site is consistently deep into certain economic areas(global economic pattern). But when the question looks for information irrelevant to that topic, the answer is still "I don't know".

The solution for this is to provide more information by scraping more sites in the economic area that cover a wider range of economic fields. This will increase the chance to return an answer with high correctness.

✂ References

1/ ChatGPT

2/ OpenAI API: [Web Q&A - OpenAI API](#)

2/ Source code: [chatbot=1st_trials_\)_edited.ipynb - Colaboratory \(google.com\)](#)

3/ Data folder:

<https://drive.google.com/drive/folders/1XR5pOUut2flEcR2FmLCYt2NndW11J2DD?usp=sharing>