

LLMs_Hypothetical_Answers_ReRank_QA_Task_NEWS_API

January 21, 2024

1 How to Use LLMs to Generate Hypothetical Answers and Improve Result Ranking (Re-ranking) in a QA Task with the NEWS API.?

2 Install Lib

```
[ ]: !pip install openai
from google.colab import userdata
openai_api_key = userdata.get('OPENAI_API_KEY')
news_api_key = userdata.get('NEWS_API_KEY')

from openai import OpenAI
client = OpenAI(api_key=openai_api_key)
```

```
[7]: import time
from datetime import date, timedelta
import json
import pandas as pd
pd.set_option('display.max_colwidth', None)
from IPython import display

import os
import requests

from scipy import spatial
```

3 Build prompt of queries using GPT-3.5:

```
[64]: def get_completion_gpt(input, gpt_model = "gpt-3.5-turbo-1106"):
    completion = client.chat.completions.create(
        model=gpt_model,
        messages=[
            {"role": "system", "content": "Output only valid JSON"},
            {"role": "user", "content": input},
        ],
        response_format={ "type": "json_object" }
```

```

    )

    text = completion.choices[0].message.content
    parsed = json.loads(text)
    return parsed

```

```

[31]: user_query = "What are the events that impacted the bitcoin price lately?."

input = f"""
You have access to a search NEWS API that returns recent news articles.
Make a list of search questions that match this topic.
Try using different keywords that are related to the topic to make your
    ↪ questions more general.
Create many queries, some with specific terms and others without them.
Be imaginative and make as many queries as you can. More queries will help you
    ↪ find better results.
Pick 10 of these queries.
For example, you can include queries like ['keyword_1 keyword_2', 'keyword_1',
    ↪ 'keyword_2'].

# User question: {user_query}

# Format: {"queries": ["query_1", "query_2", "query_3"]}}
"""

gpt_model = "gpt-3.5-turbo-1106"
# To be able to use json response_format, only 2 models support this :
# gpt-4-1106-preview and gpt-3.5-turbo-1106

parsed = get_completion_gpt(input, gpt_model = gpt_model)
parsed

```

```

[31]: {'queries': ['Bitcoin price impact events',
    'Recent factors affecting Bitcoin price',
    'Bitcoin price fluctuations causes',
    'Bitcoin value influencers',
    'Cryptocurrency market impact on Bitcoin',
    'Market forces affecting Bitcoin',
    'Key events influencing Bitcoin price',
    'Bitcoin price movement triggers',
    'Recent developments and Bitcoin price',
    'Bitcoin value shift catalysts']}

```

```

[55]: queries = parsed['queries']
queries.append(user_query)
queries

```

```
[55]: ['Bitcoin price impact events',
       'Recent factors affecting Bitcoin price',
       'Bitcoin price fluctuations causes',
       'Bitcoin value influencers',
       'Cryptocurrency market impact on Bitcoin',
       'Market forces affecting Bitcoin',
       'Key events influencing Bitcoin price',
       'Bitcoin price movement triggers',
       'Recent developments and Bitcoin price',
       'Bitcoin value shift catalysts',
       'What are the events that impacted the bitcoin price lately?..']
```

4 Fetch news articles from NEWS API for each query:

```
[53]: def search_news(query, news_api_key= news_api_key,num_articles=5, from_datetime=
      ↪ "2024-01-01",to_datetime = "2024-01-20"):
      response = requests.get(
          "https://newsapi.org/v2/everything",
          params={
              "q": query,
              "apiKey": news_api_key,
              "pageSize": num_articles,
              "sortBy": "relevancy",
              "from": from_datetime,
              "to": to_datetime,
          },
      )

      return response.json()
```

```
[79]: articles = []
      for query in queries:
          result = search_news(query)
          if result['status'] == 'ok':
              articles = articles + result['articles']
          else:
              raise Exception(result["message"])

      len(articles)
```

```
[79]: 43
```

```
[83]: # remove duplicates
      articles = list({article["url"]: article for article in articles}.values())
      len(articles)
```

[83]: 41

```
[84]: #Display some articles:
print("Total number of articles:", len(articles))

for article in articles[0:5]:
    print("Title:", article["title"])
    print("Url:", article["url"])
    print("Description:", article["description"])
    print("Content:", article["content"][0:300] + "...")
    print()
```

Total number of articles: 41

Title: Bitcoin ETFs finally approved after a chaotic, 'embarrassing' 24 hours for SEC

Url: <https://www.marketwatch.com/story/bitcoin-etfs-finally-approved-after-a-chaotic-embarrassing-24-hours-for-sec-e42233d5>

Description: The U.S. Securities and Exchange on Wednesday greenlighted several exchange traded funds investing directly in bitcoin for the first time. But the 24 hours leading up to the approval were chaotic.

Content: On Wednesday, the U.S. Securities and Exchange Commission for the first time greenlighted several exchange-traded funds investing directly in bitcoin. But the 24 hours leading up to that approval wer... [+2536 chars]...

Title: How Will A Bitcoin Spot ETF Affect Bitcoin's Price?

Url: <https://www.forbes.com/sites/digital-assets/2024/01/08/how-will-a-bitcoin-spot-etf-affect-bitcoins-price/>

Description: The approval of a bitcoin spot ETF is expected to increase the demand for bitcoin, leading to a supply squeeze.

Content: Photo Illustration by Omar Marques

SOPA Images/LightRocket via Getty Images

What Is A Bitcoin Spot ETF?

A bitcoin spot ETF, traded on stock exchanges, holds bitcoin as its primary asset, unlike fu... [+3410 chars]...

Title: How Will A Bitcoin Spot ETF Impact Bitcoin's Price?

Url: <https://www.forbes.com/sites/digital-assets/2024/01/08/how-will-a-bitcoin-spot-etf-impact-bitcoins-price/>

Description: The approval of a bitcoin spot ETF is expected to increase the demand for bitcoin, leading to a supply squeeze.

Content: Photo Illustration by Omar Marques

SOPA Images/LightRocket via Getty Images

What Is A Bitcoin Spot ETF?

A bitcoin spot ETF, traded on stock exchanges, holds bitcoin as its primary asset, unlike fu... [+3405 chars]...

Title: Unpacking The Impact: Bitcoin ETF Sparks New Market Activity And Price

Volatility Levels

Url: <https://www.ibtimes.com/unpacking-impact-bitcoin-etf-sparks-new-market-activity-price-volatility-levels-3722132>

Description: According to a research analyst, "The price action of BTC, coupled with volume data and the performance of certain altcoins, demonstrates the adherence to the typical 'buy the rumor, sell the news' pattern associated with major market events."

Content: In the wake of the recent debut of the inaugural cluster of spot Bitcoin exchange-traded funds (ETFs) within the U.S., the cryptocurrency market has undergone notable shifts, characterized by heighte... [+5022 chars]...

Title: Why is 2032 pivotal for Bitcoin

Url: <https://erickimphotography.com/blog/2024/01/18/why-is-2032-pivotal-for-bitcoin/>

Description: Controlled supply A fixed money supply, or a supply altered only in accord with objective and calculable criteria, is a necessary condition to a meaningful just price of money.[1] -Fr. Bernard W. Dempsey, S.J. (1903-1960)

Every 4 years... halving effect? What c...

Content: A fixed money supply, or a supply altered only in accord with objective and calculable criteria, is a necessary condition to a meaningful just price of money.[1]

Fr. Bernard W. Dempsey, S.J. (1903-1... [+4800 chars]...

5 Re-rank

Recommend to the LLM that it provides an optimal hypothetical answer for us to use in reevaluating our search results.

This will prioritize answers that closely resemble the ideal answer rather than those resembling our original question at the top.

```
[62]: hypoth_answer = f"""
"Make up an answer to the user's question. We'll use this made-up answer to
↳sort the search results.

Imagine you have all the details to answer, but don't use real facts.
Instead, use things like 'EVENT affected something' or 'NAME mentioned
↳something on DATE' as placeholders.

User question: {user_query}

Format: {"hypotheticalAnswer": "hypothetical answer text"}
"""
```

```
[65]: parsed_hypotheset_answer = get_completion_gpt(hypoth_answer)
parsed_hypotheset_answer
```

```
[65]: {'hypotheticalAnswer': 'Several major market movements and regulatory
developments have significantly influenced the recent fluctuations in bitcoin
price. A prominent figure in the industry made a statement about the future of
cryptocurrency, which sparked considerable market activity. Additionally, an
unexpected EVENT led to a sharp increase in trading volume and subsequent price
volatility. These occurrences have contributed to the current state of bitcoin
price.'}
```

```
[67]: hypoth_answer_llm = parsed_hypothes_answer['hypotheticalAnswer']
      hypoth_answer_llm
```

```
[67]: 'Several major market movements and regulatory developments have significantly
influenced the recent fluctuations in bitcoin price. A prominent figure in the
industry made a statement about the future of cryptocurrency, which sparked
considerable market activity. Additionally, an unexpected EVENT led to a sharp
increase in trading volume and subsequent price volatility. These occurrences
have contributed to the current state of bitcoin price.'
```

6 Embedding

```
[87]: def get_embeddings(input):
      response = client.embeddings.create(model="text-embedding-ada-002",
      ↪input=input)
      return [data.embedding for data in response.data]
```

6.1 The hypotheticalAnswer

```
[86]: embedding_hypoth = get_embeddings(hypoth_answer_llm)[0]
      embedding_hypoth[:5]
```

```
[86]: [-0.01261220220476389,
      -0.030588287860155106,
      0.01759662665426731,
      -0.011190837249159813,
      -0.022085824981331825]
```

6.2 The content articles

```
[90]: articles_prepare_embedd = [
      f"{article['title']} {article['description']} {article['content'][:
      ↪100]}"
      for article in articles
      ]

print(f"Length of articles to embed: {len(articles_prepare_embedd)}")
```

```

article_embeddings = get_embeddings(articles_prepare_embedd)
print(f"Length of embedding vector {len(article_embeddings)}")
article_embeddings[0][:5]

```

Length of articles to embed: 41
Length of embedding vector 41

```

[90]: [-0.00592326233163476,
      -0.022001618519425392,
      -0.012379975989460945,
      -0.041531145572662354,
      -0.029092619195580482]

```

7 Against Hypothetical Answer ==> Re-Rank

7.1 Calculate cosine similarity

```

[92]: similarity_score_func=lambda x, y: 1 - spatial.distance.cosine(x, y)

# Calculate cosine similarity
cosine_similarities = []
for article_embedding in article_embeddings:
    cosine_similarities.append(similarity_score_func(embedding_hypoth,
↪article_embedding))

cosine_similarities[0:10]

```

```

[92]: [0.800271713019792,
      0.8329731514528427,
      0.8351398278778469,
      0.8958925735044341,
      0.8035452144391221,
      0.8128721153398575,
      0.6859686244322137,
      0.8217590233146307,
      0.8175247397202389,
      0.8174839989896512]

```

7.2 Get the most relevant articles

```

[96]: scored_articles = zip(articles, cosine_similarities)
sorted_articles = sorted(scored_articles, key=lambda x: x[1], reverse=True)

print("Top 5 articles:", "\n")

```

```

for article, score in sorted_articles[0:5]:
    print("Title:", article["title"])
    print("Url:", article["url"])
    print("Date of publication:", article["publishedAt"])
    print("Description:", article["description"])
    print("Content:", article["content"][0:100] + "...")
    print("Score:", score)
    print()

```

Top 5 articles:

Title: Unpacking The Impact: Bitcoin ETF Sparks New Market Activity And Price Volatility Levels

Url: <https://www.ibtimes.com/unpacking-impact-bitcoin-etf-sparks-new-market-activity-price-volatility-levels-3722132>

Date of publication: 2024-01-15T18:17:34Z

Description: According to a research analyst, "The price action of BTC, coupled with volume data and the performance of certain altcoins, demonstrates the adherence to the typical 'buy the rumor, sell the news' pattern associated with major market events."

Content: In the wake of the recent debut of the inaugural cluster of spot Bitcoin exchange-traded funds (ETFs...

Score: 0.8958925735044341

Title: Bitcoin's Rollercoaster: Surges Then Faces Sudden Dip As Options Expiry Looms

Url: <http://www.newsbtc.com/bitcoin-news/bitcoins-rollercoaster-surges-then-faces-sudden-dip-as-options-expiry-looms/>

Date of publication: 2024-01-12T22:00:01Z

Description: According to Crypto trader Mags, Bitcoin, the flagship cryptocurrency, has recently demonstrated a notable +55% surge, breaking out from a smaller broadening wedge pattern. Mags shared insights on this significant price movement, observing Bitcoin's breakout ...

Content: According to Crypto trader Mags, Bitcoin, the flagship cryptocurrency, has recently demonstrated a n...

Score: 0.8499605958081285

Title: Bitcoin's Road To Record Heights: Analysts Forecast Post-Halving Surge

Url: <http://www.newsbtc.com/bitcoin-news/bitcoin-road-to-record-heights-analysts-forecast-post-halving-surge/>

Date of publication: 2024-01-02T17:30:06Z

Description: Analysts have recently focused on Bitcoin's potential trajectory after its anticipated halving event. Adrian Zduńczyk, a prominent crypto trader, has shared a bullish outlook, suggesting that Bitcoin could soar to new all-time highs post-halving. Zduńczyk's p...

Content: Analysts have recently focused on Bitcoin's potential trajectory after its anticipated halving event...

Score: 0.8442352622295201

Title: Bitcoin ETF Approval Triggers \$1.2 Billion Trading Volume And New Highs For BTC Price

Url: <http://www.newsbtc.com/bitcoin-news/bitcoin-etf-approval-triggers-1-2-billion-trading-volume-and-new-highs-for-btc-price/>

Date of publication: 2024-01-11T15:41:15Z

Description: In a highly anticipated move, the United States Securities and Exchange Commission (SEC) approved all 11 Bitcoin ETF applications, and the market response has been nothing short of remarkable. The approval has led to significant trading volume and propelled B...

Content: In a highly anticipated move, the United States Securities and Exchange Commission (SEC) approved al...

Score: 0.8440754845414833

Title: The hullabaloo over crypto has quietened despite the good year it had

Url: <https://www.livemint.com/opinion/online-views/the-hullabaloo-over-crypto-has-quietened-despite-the-good-year-it-had-11704205040952.html>

Date of publication: 2024-01-03T02:30:07Z

Description: Bitcoin's price shot up in 2023 but few seemed to care. The recent cryptocurrency bust led wave-riding investors to other bets like small-cap stocks and derivatives, even as tighter monetary policies meant less asset inflation globally.

Content: In 2023, Bitcoin, the most popular crypto, rallied 155% in US dollar terms. The Bloomberg Galaxy Cry...

Score: 0.8360295084231848

8 Against Original User Query

```
[99]: embedding_original_query = get_embeddings(user_query)[0]
      embedding_original_query[:5]
```

```
[99]: [-0.03228180855512619,
      -0.015388915315270424,
      0.024735063314437866,
      -0.006432190537452698,
      -0.033463504165410995]
```

```
[100]: similarity_score_func=lambda x, y: 1 - spatial.distance.cosine(x, y)

      # Calculate cosine similarity
      cosine_similarities_original = []
      for article_embedding in article_embeddings:
          cosine_similarities_original.
      ↪append(similarity_score_func(embedding_original_query, article_embedding))
```

```
cosine_similarities_original[0:10]
```

```
[100]: [0.7799895223331812,  
        0.7867212899062556,  
        0.7877645981167399,  
        0.8337490000010337,  
        0.7804363357046923,  
        0.7995206266400686,  
        0.7132686302759849,  
        0.7933271463912787,  
        0.7924005032912812,  
        0.7758823627797351]
```

The similarity scores are much lower than the one against the hypothetical answer:

```
[101]: scored_articles_original = zip(articles, cosine_similarities_original)
scored_articles_original = sorted(scored_articles_original, key=lambda x: x[1],
    ↪reverse=True)

print("Top 5 articles:", "\n")
for article, score in scored_articles_original[0:5]:
    print("Title:", article["title"])
    print("Url:", article["url"])
    print("Date of publication:", article["publishedAt"])
    print("Description:", article["description"])
    print("Content:", article["content"][0:100] + "...")
    print("Score:", score)
    print()
```

Top 5 articles:

Title: Unpacking The Impact: Bitcoin ETF Sparks New Market Activity And Price Volatility Levels

Url: <https://www.ibtimes.com/unpacking-impact-bitcoin-etf-sparks-new-market-activity-price-volatility-levels-3722132>

Date of publication: 2024-01-15T18:17:34Z

Description: According to a research analyst, "The price action of BTC, coupled with volume data and the performance of certain altcoins, demonstrates the adherence to the typical 'buy the rumor, sell the news' pattern associated with major market events."

Content: In the wake of the recent debut of the inaugural cluster of spot Bitcoin exchange-traded funds (ETFs...

Score: 0.8337490000010337

Title: Bitcoin's Rollercoaster: Surges Then Faces Sudden Dip As Options Expiry Looms

Url: <http://www.newsbtc.com/bitcoin-news/bitcoins-rollercoaster-surges-then->

faces-sudden-dip-as-options-expiry-looms/

Date of publication: 2024-01-12T22:00:01Z

Description: According to Crypto trader Mags, Bitcoin, the flagship cryptocurrency, has recently demonstrated a notable +55% surge, breaking out from a smaller broadening wedge pattern. Mags shared insights on this significant price movement, observing Bitcoin's breakout ...

Content: According to Crypto trader Mags, Bitcoin, the flagship cryptocurrency, has recently demonstrated a n...

Score: 0.8237980543552765

Title: The hullabaloo over crypto has quietened despite the good year it had

Url: <https://www.livemint.com/opinion/online-views/the-hullabaloo-over-crypto-has-quietened-despite-the-good-year-it-had-11704205040952.html>

Date of publication: 2024-01-03T02:30:07Z

Description: Bitcoin's price shot up in 2023 but few seemed to care. The recent cryptocurrency bust led wave-riding investors to other bets like small-cap stocks and derivatives, even as tighter monetary policies meant less asset inflation globally.

Content: In 2023, Bitcoin, the most popular crypto, rallied 155% in US dollar terms. The Bloomberg Galaxy Cry...

Score: 0.8227674587809878

Title: Bitcoin's Road To Record Heights: Analysts Forecast Post-Halving Surge

Url: <http://www.newsbtc.com/bitcoin-news/bitcoin-road-to-record-heights-analysts-forecast-post-halving-surge/>

Date of publication: 2024-01-02T17:30:06Z

Description: Analysts have recently focused on Bitcoin's potential trajectory after its anticipated halving event. Adrian Zduńczyk, a prominent crypto trader, has shared a bullish outlook, suggesting that Bitcoin could soar to new all-time highs post-halving. Zduńczyk's p...

Content: Analysts have recently focused on Bitcoin's potential trajectory after its anticipated halving event...

Score: 0.8149606810462926

Title: How Low Will Bitcoin Price Drop? 10xResearch Reveals Potential Low Levels

Url: <https://biztoc.com/x/7ef40dd4ef24c30c>

Date of publication: 2024-01-15T08:20:33Z

Description: The post How Low Will Bitcoin Price Drop? 10xResearch Reveals Potential Low Levels appeared first on Coinpedia Fintech News Bitcoin has always stood out as a benchmark for market trends and investor sentiment. Recent developments, particularly concerning Bitc...

Content: The post How Low Will Bitcoin Price Drop? 10xResearch Reveals Potential Low Levels appeared first on...

Score: 0.8082506049412768

The list of the articles with hypothetical answer seems to give more relevant answers to the question

9 Final Answer

9.1 Final answer with hypothetical re-rank:

```
[121]: from IPython import display

[124]: formatted_top_results = [
    {
        "title": article["title"],
        "description": article["description"],
        "url": article["url"],
    }
    for article, _score in sorted_articles[0:5]
]

final_input = f"""
Generate an answer to the user's question based on the given search results.
TOP_RESULTS: {formatted_top_results}
USER_QUESTION: {user_query}

Include as much information as possible in the answer. Reference the relevant_
↪search result urls as markdown links.
"""

gpt_model = "gpt-3.5-turbo-1106"

completion = client.chat.completions.create(
    model=gpt_model,
    messages=[
        {"role": "user", "content": final_input},
    ],
)

display.display(display.Markdown(completion.choices[0].message.content))
```

Recent events that have impacted the price of Bitcoin include the approval of Bitcoin ETF applications by the United States Securities and Exchange Commission, which has led to a significant trading volume and propelled the price of Bitcoin to new highs ([source](#)). Additionally, analysts have been focusing on Bitcoin's potential trajectory after its anticipated halving event, with some predicting a surge to new all-time highs post-halving ([source](#)). The price action of Bitcoin, coupled with volume data and the performance of certain altcoins, has also demonstrated adherence to the typical 'buy the rumor, sell the news' pattern associated with major market events ([source](#)). These events have contributed to significant price volatility and market activity in the Bitcoin space.

9.2 Final Answer with original question similarities

```
[125]: formatted_top_results = [
    {
        "title": article["title"],
        "description": article["description"],
        "url": article["url"],
    }
    for article, _score in scored_articles_original[0:5]
]

final_input = f"""
Generate an answer to the user's question based on the given search results.
TOP_RESULTS: {formatted_top_results}
USER_QUESTION: {user_query}

Include as much information as possible in the answer. Reference the relevant
↳search result urls as markdown links.
"""

gpt_model = "gpt-3.5-turbo-1106"

completion = client.chat.completions.create(
    model=gpt_model,
    messages=[
        {"role": "user", "content": final_input},
    ],
)

display.display(display.Markdown(completion.choices[0].message.content))
```

Based on recent research and analysis, several events have impacted the price of Bitcoin. These include the anticipation of a Bitcoin ETF, which sparked new market activity and volatility levels. Additionally, Bitcoin experienced a significant surge and subsequent dip, possibly due to options expiry looming. There was also a noticeable increase in Bitcoin's price in 2023, despite the cryptocurrency bust, leading investors to alternative assets. Furthermore, analysts have forecasted a post-halving surge for Bitcoin. For more details on these events, you can refer to the following URLs:

- Unpacking The Impact: Bitcoin ETF Sparks New Market Activity And Price Volatility Levels: ibtimes.com
- Bitcoin's Rollercoaster: Surges Then Faces Sudden Dip As Options Expiry Looms: newsbtc.com
- The hullabaloo over crypto has quietened despite the good year it had: livemint.com
- Bitcoin's Road To Record Heights: Analysts Forecast Post-Halving Surge: newsbtc.com
- How Low Will Bitcoin Price Drop? 10xResearch Reveals Potential Low Levels: biztoc.com

10 Final Thoughts

In this notebook, you have discovered :

- How to use LLMs to generate various queries in regards to the user query ==> **Prompt builder with JSON results**
- Then, how to use these queries to fetch news from [NEW API](#). (btw, you need an API KEY)
- Also, we will use again LLMs to help us build a **Hypothetical Answer** against which, we w'll rank the news article. We'll use embeddding and cosine similarity to get the most relevant answers. Comparison will be made with the original query.
- Then, we will ask the LLM the original user query and see what happens:
“What are the events that impacted the bicoin price lately?.”