Applications of Large Language Models in Finance

Case Studies in Financial Generative AI

Ivan Bakrac, July 2023

Welcome!

Ivan Bakrac

LinkedIn

Professional Experience

- Director of Digital Assets Market Strategy | Head of Treasury at ConsenSys
- Head of Multi-Sector Fixed Income Portfolio Management at Invesco
- VP | Fixed Income Portfolio Manager at Franklin Templeton
- Co-Head of Quantitative Research at Franklin Templeton

Academic Experience

- Adjunct Faculty - Fintech: AI & Data Science, Berkeley College School of Business
- Adjunct Faculty - Fintech: Blockchain & DeFi, Stevens Institute of Technology
- Affiliated Faculty - CRAFT: Center for Advancing Financial Technologies at Stevens

Specializations

FinTech, disruptive & emerging technologies including financial machine learning, AI, predictive analytics and business data science, digital assets, cryptocurrencies and blockchain applications

Education

MS in Financial Data Science from Stevens Institute of Technology

MS in Financial Engineering from NYU Tandon School of Engineering

BBA in Quantitative Finance, Baruch Zicklin School of Business

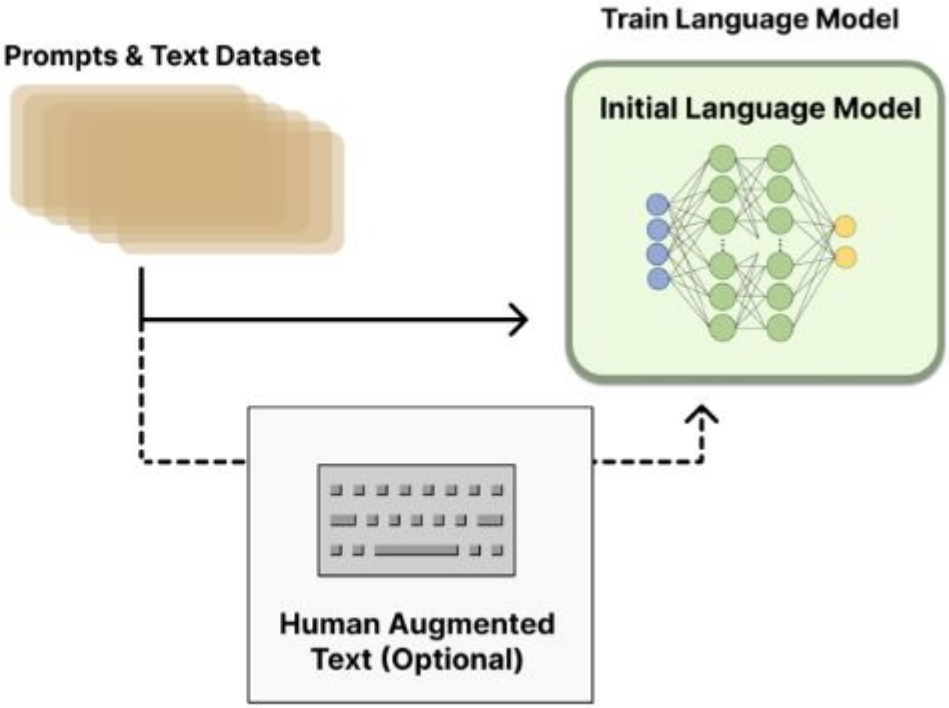
Agenda

- Quick Chat GPT Architecture Overview
- Case Study 1: Summarizing Earnings Call Features
- Case Study 2: Extending Earnings Features to Sentiment Features
- Case Study 3: Earnings Topics Generation with Sentiment & Rank Labels
- Case Study 4: Using Large Language Model Embeddings as Stock Features
- Case Study 5: Fine Tuning Model - Building Earnings Call KPI Features App
- Case Study 6: Stock Domain Queries w/LangChain & Using Chains for Features
- Case Study 7: Sentiment & News Stock Relevance in Long Short Investing
- Reference Material

Quick Chat GPT Architecture Overview

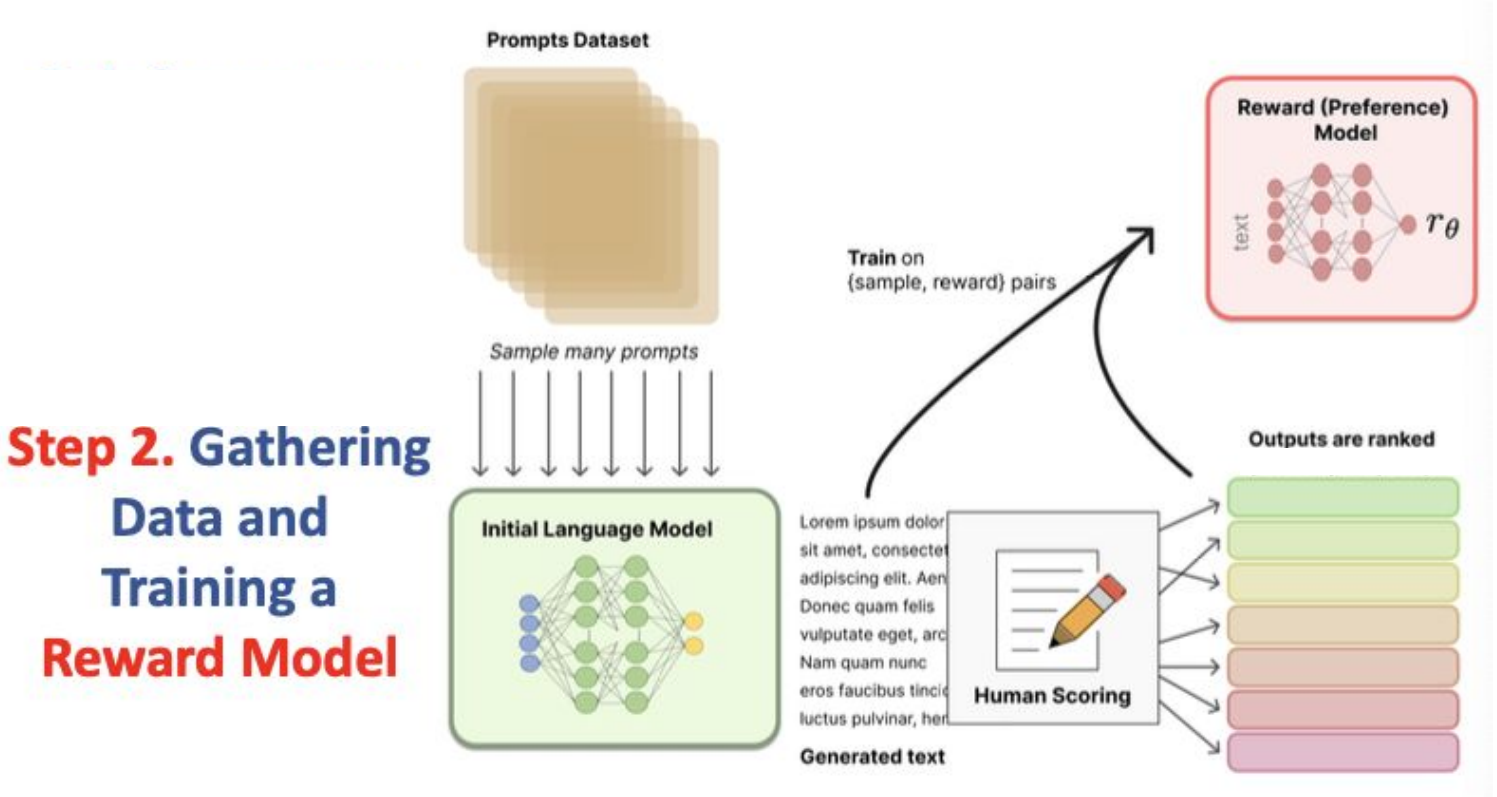
Collect Data and Train a Supervised Policy

Step 1. Pretraining a Language Model (LM)



Source: https://huggingface.co/blog/rlhf

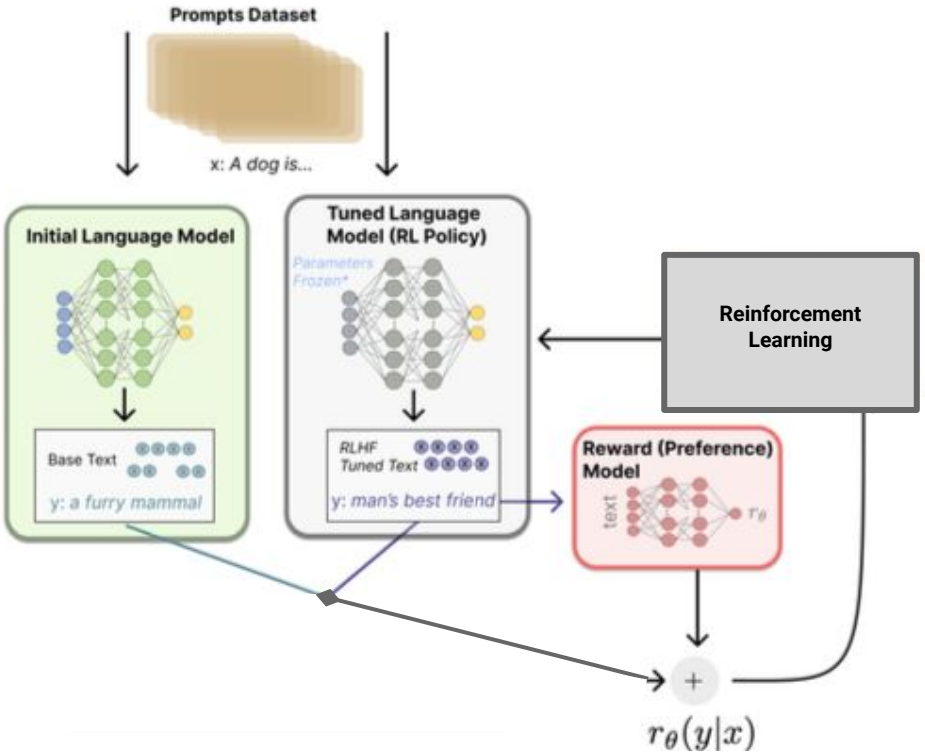
Collect Data and Train a Rewards Model



Source: https://huggingface.co/blog/rlhf

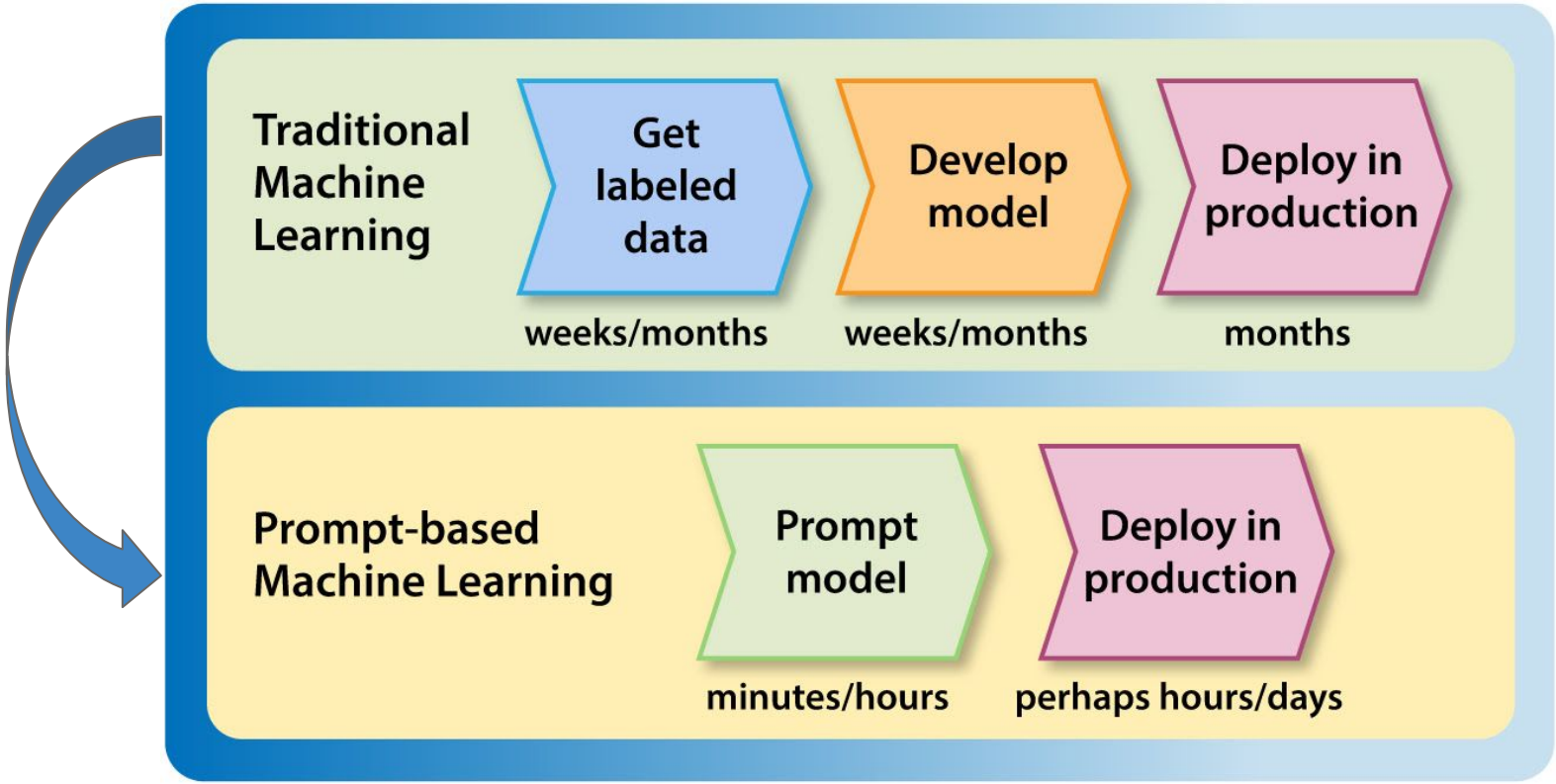
Optimize Policy against the Reward Model

Step 3. Fine-tuning the LM with Reinforcement Learning



Source: https://huggingface.co/blog/rlhf

Potential to Shorten Development Process

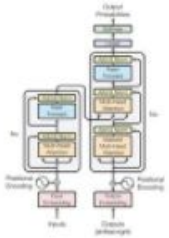


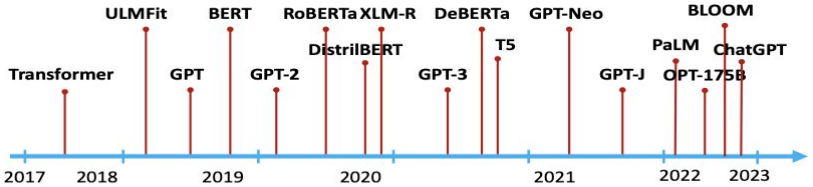
Source: https://learn.deeplearning.ai

The Transformers Revolution

Transformer Architecture Showed Self-Attention is Sufficient and no Recurrence is Required

Attention Is All You Need showed that RNNs are not required!





"Self-attention mechanism enables the model to weigh the importance of different elements in an input sequence and dynamically adjust their influence on the output"

Source: Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.

Case Study 1: Summarizing Earnings Call Features

Earnings Transcript

Data Preparation

- Step 1: Initial text file size: 48,657 words
- Step 2: Divide text file into chunks to meet token limit
- Step 3: Write the summary prompt for each section

NVIDIA Transcript as of 2/23

"Operator\n\nGood afternoon. My name is Emma, and I will be your conference operator today. At this time, I would like to welcome everyone to the NVIDIA's third quarter earnings call. [Operator instructions] Simona Jankowski, you may begin your conference.\n\nSimona Jankowski -- Vice President, Investor Relations\n\nThank you. Good afternoon, everyone, and welcome to NVIDIA's conference call for the third quarter of fiscal 2023. With me today from NVIDIA are Jen-Hsun Huang, president and chief executive officer; and Colette Kress, executive vice president and chief financial officer. I'd like to remind you that our call is being webcast live on NVIDIA's investor relations website.\n\n\nThe webcast will be available for replay until the conference call to discuss our financial results for the fourth quarter and fiscal 2023. The content of today's call is NVIDIA's property. It can't be reproduced or transcribed without our prior written consent. During this call, we may make forward-looking statements based on current expectations.\n\nThese are subject to a number of significant risks and uncertainties, and our actual results may differ materially. For a discussion of factors that could affect our future financial results and business, please refer to the disclosure in today's earnings release our most recent Forms 10-K and 10-Q and the reports that we may file on Form 8-K with the Securities and Exchange Commission. All our statements are made as of today, November 16, 2022, and based on information currently available to us. Except as required by law, we assume no obligation to update any such statements.\n\nDuring this call, we will discuss non-GAAP financial measures. You can find a reconciliation of these non-GAAP financial measures to GAAP financial measures in our CFO commentary, which is posted on our website. With that, let me turn the call over to Colette.\n\nColette Kress -- Executive Vice President and Chief Financial Officer\n\n\nThanks, Simona. Q3 revenue was $5.93 billion, down 12% sequentially and down 17% year on year. We delivered record data center and automotive revenue. while our gaming and pro visualization platforms declined as we work through channel inventory corrections and challenging external conditions.\n\nStarting with data center. Revenue of $3.83 billion was up 1% sequentially and 31% year-on-year. This reflects very solid performance in the face of macroeconomic challenges new export controls and lingering supply chain disruptions. Year-on-year growth was driven primarily by leading U.S.\n\ncloud providers and a broadening set of consumer Internet companies for workloads such as large language models, recommendation systems and generative AI. As the number and scale of public cloud computing and Internet service companies deploying NVIDIA AI grows our traditional hyperscale definition will need to be expanded to convey the different end market use cases. We will align our data center customer commentary going forward accordingly. Other vertical industries, such as automotive and energy, also contributed to growth with key workloads relating to autonomous driving, high-performance computing, simulations and analytics.\n\nDuring the qu

Word count: 48,657

Prompt

Prompt

```
for i in range(len(sentences)):
    prompt = f"""
    Your task is to generate a short summary

    Summarize the review below, delimited by triple \
    backticks in at most 20 words.

    Review: ```{sentences[i]}```
    """

    response = get_completion(prompt)
    print(i, response, "\n")
```

Response

```
Markdown(summary1)
```

0 NVIDIA's Q3 earnings call discussed a decline in gaming and pro visualization platforms, while data center and automotive revenue hit record highs. The company also announced collaborations with Microsoft and Oracle to offer AI training and inference at scale to enterprises. Gaming revenue was down due to lower sell-in to partners to align channel inventory levels with demand expectations. The company's new Ada Lovelace GPU architecture had an exceptional launch, with the first ADA GPU selling out quickly in many locations.

1 NVIDIA's RTX games and applications now exceed 350, with strong fundamentals expected to continue in the long term. The company added over 85 games to its GeForce NOW cloud gaming service in Q3, bringing the total to over 1,400. Despite near-term challenges, NVIDIA's long-term opportunity remains intact, fuelled by AI simulation, computationally intensive design and engineering workloads. The company expects modest sequential growth driven by automotive, gaming and data centre, with revenue expected to be $6bn, plus or minus 2% in Q4.

2 NVIDIA CEO discusses surging demand in AI, including deep recommender systems, large language models, and generative AI. The company is making progress in NVIDIA AI enterprise and moving its software stack to the cloud. The company also took a $702 million inventory reserve for data center products due to changes in expected demand in China.

3 NVIDIA's Q3 earnings call revealed continued growth in data center business, driven by AI and accelerated computing. The company's networking business is heavily indexed to high-performance computing and AI adoption, with a focus on high-end solutions for data centers that move a lot of data. NVIDIA's new data processing unit, Grace, is designed for large data processing at high speeds and is expected to have production samples available in Q1 2022.

4 NVIDIA CEO Jen-Hsun Huang discussed the importance of networking in data centers and the efficiency gains achieved through NVIDIA's InfiniBand and Magnum IO software stack. He also highlighted the significant demand for the company's new Hopper GPU, which is expected to ramp up quickly due to its cost benefits and improved performance for AI models. Additionally, CFO Colette Kress discussed plans to report hyperscaler revenue in a new way to better reflect the complexity of the data center business.

5 NVIDIA CEO Jen-Hsun Huang discussed the company's growth in the adoption of its products by internet service companies and the cloud computing industry. He also addressed the impact of the crypto market on the company's GPU sales, stating that NVIDIA does not expect blockchain to be an important part of its business going forward. The company's new platforms, including the MRTX and Hopper, are off to a strong start, and NVIDIA networking is enjoying record results. Huang emphasized the importance of accelerated computing in meeting the demands of AI, as general-purpose computing has slowed to a crawl

Word count: 604

Case Study 2: Extending Earnings Features to Sentiment Features

Prompt

Prompt

```
for i in range(len(sentences)):
    prompt = f"""
    What is the sentiment of the following review,

    Give your answer as a single word, either "positive" \
    or "negative".

    Review: ```{sentences[i]}```
    """

    response = get_completion(prompt)
    print(i, response, "\n")
```

Summary Features	Sentiment
0 NVIDIA's Q3 earnings call discussed a decline in gaming and pro visualization platforms, while data center and automotive revenue hit record highs. The company also announced collaborations with Microsoft and Oracle to offer AI training and inference at scale to enterprises. Gaming revenue was down due to lower sell-in to partners to align channel inventory levels with demand expectations. The company's new Ada Lovelace GPU architecture had an exceptional launch, with the first ADA GPU selling out quickly in many locations.	Neutral
1 NVIDIA's RTX games and applications now exceed 350, with strong fundamentals expected to continue in the long term. The company added over 85 games to its GeForce NOW cloud gaming service in Q3, bringing the total to over 1,400. Despite near-term challenges, NVIDIA's long-term opportunity remains intact, fuelled by AI simulation, computationally intensive design and engineering workloads. The company expects modest sequential growth driven by automotive, gaming and data centre, with revenue expected to be $6bn, plus or minus 2% in Q4.	Positive
2 NVIDIA CEO discusses surging demand in AI, including deep recommender systems, large language models, and generative AI. The company is making progress in NVIDIA AI enterprise and moving its software stack to the cloud. The company also took a $702 million inventory reserve for data center products due to changes in expected demand in China.	Neutral
3 NVIDIA's Q3 earnings call revealed continued growth in data center business, driven by AI and accelerated computing. The company's networking business is heavily indexed to high-performance computing and AI adoption, with a focus on high-end solutions for data centers that move a lot of data. NVIDIA's new data processing unit, Grace, is designed for large data processing at high speeds and is expected to have production samples available in Q1 2022.	Neutral
4 NVIDIA CEO Jen-Hsun Huang discussed the importance of networking in data centers and the efficiency gains achieved through NVIDIA's InfiniBand and Magnum IO software stack. He also highlighted the significant demand for the company's new Hopper GPU, which is expected to ramp up quickly due to its cost benefits and improved performance for AI models. Additionally, CFO Colette Kress discussed plans to report hyperscaler revenue in a new way to better reflect the complexity of the data center business.	Neutral
5 NVIDIA CEO Jen-Hsun Huang discussed the company's growth in the adoption of its products by internet service companies and the cloud computing industry. He also addressed the impact of the crypto market on the company's GPU sales, stating that NVIDIA does not expect blockchain to be an important part of its business going forward. The company's new platforms, including the MRTX and Hopper, are off to a strong start, and NVIDIA networking is enjoying record results. Huang emphasized the importance of accelerated computing in meeting the demands of AI, as general-purpose computing has slowed to a crawl.	Positive

Case Study 3: Earnings Topics Generation with Sentiment & Rank Labels

Prompt (1 of 2)

Prompt

```
for i in range(len(sentences)):
    prompt = f"""

    Identify the following items from the review text:

    - Determine three topics that are being discussed in the \
      following text, which is delimited by triple backticks. \
      Make each topic one or two words long.

    - Deterime Sentiment (positive or negative) for each topic
    - Rank Sentiment with score between 1 to 5 with 1\

      as very negative and 5 as very positive

    The review is delimited with triple backticks. \
    Format your response as a JSON object with \
    "Sentiment", "Topic" , "Rank" as the keys.



    text: ```{sentences[i]}```
    """

    response = get_completion(prompt)
    print(i, response, "\n")
```

Response

```
1 {
    "Sentiment": {
        "Topic 1": "Positive",
        "Topic 2": "Positive",
        "Topic 3": "Neutral"
    },
    "Topic": {
        "Topic 1": "Gaming",
        "Topic 2": "GeForce NOW cloud gaming service",
        "Topic 3": "Automotive"
    },
    "Rank": {
        "Topic 1": 5,
        "Topic 2": 4,
        "Topic 3": 3
    }
}
```

Topic 1: Gaming - Positive sentiment with a rank of 5
Topic 2: GeForce NOW cloud gaming service - Positive sentiment with a rank of 4
Topic 3: Automotive - Neutral sentiment with a rank of 3

Prompt (2 of 2)

Prompt

```
prompt = f"""
Translate the following python dictionary from JSON to an HTML \
table with column headers and title: {data_json_all}
"""
response = get_completion(prompt)
print(response)
```

Response

```
<table>
  <caption>Dictionary</caption>
  <thead>
    <tr>
      <th>Topic</th>
      <th>Sentiment</th>
      <th>Rank</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <td>Introduction and Legal Information</td>
      <td>Neutral</td>
      <td>3</td>
    </tr>
    <tr>
      <td>Financial Results</td>
      <td>Negative</td>
      <td>2</td>
    </tr>
    <tr>
      <td>Product Updates and Announcements</td>
      <td>Positive</td>
      <td>4</td>
    </tr>
    <tr>
```

Topic Generation with Sentiment & Rank

Summary Output

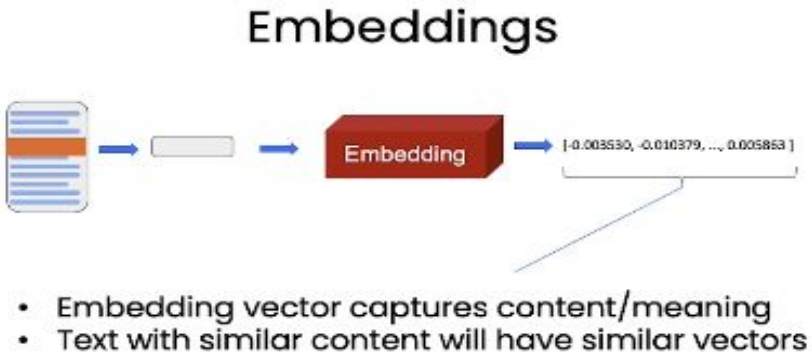
Dictionary

Topic	Sentiment	Rank
Introduction and Legal Information	Neutral	3
Financial Results	Negative	2
Product Updates and Announcements	Positive	4
Gaming	Positive	5
Cloud Computing	Positive	4
Automotive	Neutral	3
General Purpose Computing	Negative	2
AI - Deep Recommender Systems and Large Language Models	Positive	4
Generative AI	Positive	5
Inventory Charge	Neutral	3
Data Center Business	Positive	4
Networking Business	Positive	5
Networking	Positive	4
Compute	Neutral	3
Hyperscale	Positive	4
NVIDIA's adoption in internet service companies	Positive	4
Cloud computing	Positive	5
Blockchain and GPU liquidation	Negative	2

Case Study 4: Using Large Language Model Embeddings as Stock Features

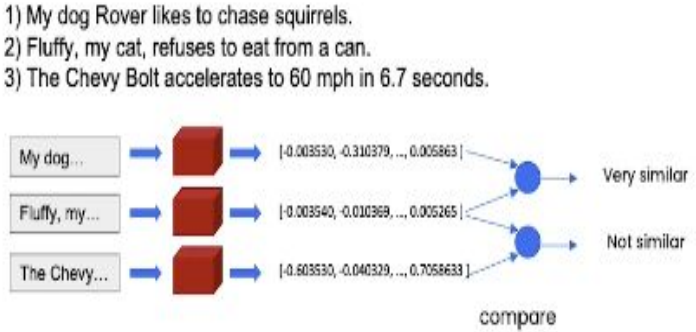
Embeddings as Representation

Embeddings as Features



- Embedding vector captures content/meaning
- Text with similar content will have similar vectors

Example



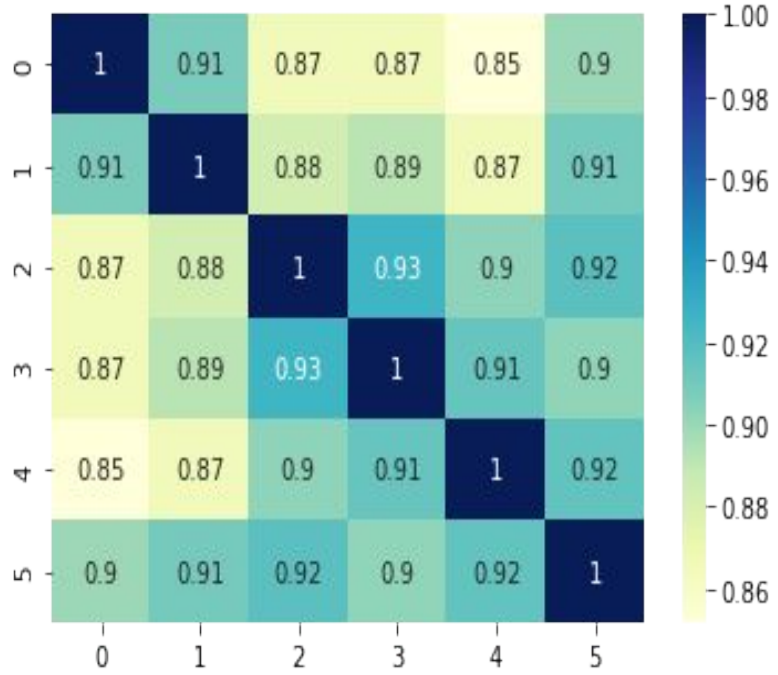
Source: https://learn.deeplearning.ai

Embeddings as Stock Features

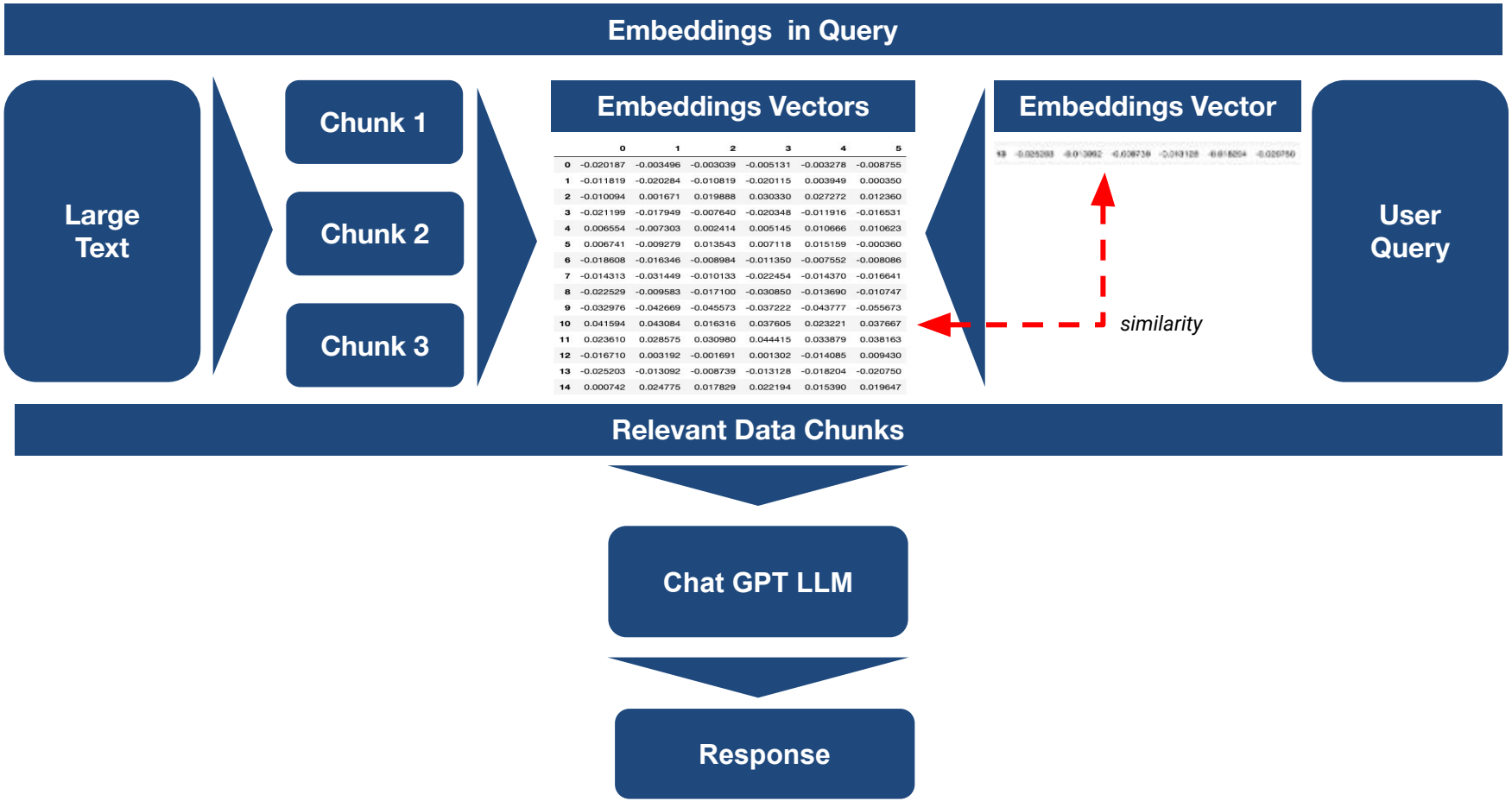
NVIDIA Embeddings Vectors

	0	1	2	3	4	5
0	-0.020187	-0.003496	-0.003039	-0.005131	-0.003278	-0.008755
1	-0.011819	-0.020284	-0.010819	-0.020115	0.003949	0.000350
2	-0.010094	0.001671	0.019888	0.030330	0.027272	0.012360
3	-0.021199	-0.017949	-0.007640	-0.020348	-0.011916	-0.016531
4	0.006554	-0.007303	0.002414	0.005145	0.010666	0.010623
5	0.006741	-0.009279	0.013543	0.007118	0.015159	-0.000360
6	-0.018608	-0.016346	-0.008984	-0.011350	-0.007552	-0.008086
7	-0.014313	-0.031449	-0.010133	-0.022454	-0.014370	-0.016641
8	-0.022529	-0.009583	-0.017100	-0.030850	-0.013690	-0.010747
9	-0.032976	-0.042669	-0.045573	-0.037222	-0.043777	-0.055673
10	0.041594	0.043084	0.016316	0.037605	0.023221	0.037667
11	0.023610	0.028575	0.030980	0.044415	0.033879	0.038163
12	-0.016710	0.003192	-0.001691	0.001302	-0.014085	0.009430
13	-0.025203	-0.013092	-0.008739	-0.013128	-0.018204	-0.020750
14	0.000742	0.024775	0.017829	0.022194	0.015390	0.019647

Translated to Similarity Matrix



Building Query with Embeddings LLM



	0	1	2	3	4	5
0	-0.020187	-0.003496	-0.003039	-0.005131	-0.003278	-0.008755
1	-0.011819	-0.020284	-0.010819	-0.020115	0.003949	0.000350
2	-0.010094	0.001671	0.019888	0.030330	0.027272	0.012360
3	-0.021199	-0.017949	-0.007640	-0.020348	-0.011916	-0.016531
4	0.006554	-0.007303	0.002414	0.005145	0.010666	0.010623
5	0.006741	-0.009279	0.013543	0.007118	0.015159	-0.000360
6	-0.018608	-0.016346	-0.008984	-0.011350	-0.007552	-0.008086
7	-0.014313	-0.031449	-0.010133	-0.022454	-0.014370	-0.016641
8	-0.022529	-0.009583	-0.017100	-0.030850	-0.013690	-0.010747
9	-0.032976	-0.042669	-0.045573	-0.037222	-0.043777	-0.055673
10	0.041594	0.043084	0.016316	0.037605	0.023221	0.037667
11	0.023610	0.028575	0.030980	0.044415	0.033879	0.038163
12	-0.016710	0.003192	-0.001691	0.001302	-0.014085	0.009430
13	-0.025203	-0.013092	-0.008739	-0.013128	-0.018204	-0.020750
14	0.000742	0.024775	0.017829	0.022194	0.015390	0.019647

Embeddings in Domain Specific Query

Data Preparation

- Initial text file size: 48,657
- Divide text file into chunks to meet token limit
- Specify heading and title
- Leverage **Embeddings**

Text

	title	heading	content
0	NVDAQ422	0	Operator\n\nGood afternoon. My name is Emma, a...
1	NVDAQ422	1	RTX games and applications now exceeds 350.\n\...
2	NVDAQ422	2	general purpose computing slows.\n\nThe second...
3	NVDAQ422	3	can still get it.\n\nAnd I guess the second pa...
4	NVDAQ422	4	I want to expand on the networking question a ...
5	NVDAQ422	5	and Chief Executive Officer\n\nYes. Let me dou...

Prompt Response

```
answer_query_with_context("What are some of the current and looming threats to the business", df_text2, df)
```

```
'The current threats to the business include supply constraints and difficulty in setting up data centers. Looming
threats include the availability of secondhand and used graphics cards, and the potential for blockchain to become
an important part of the business.'
```

```
answer_query_with_context("Provide key information about revenue for the quarter", df_text2, df)
```

```
'Revenue for the quarter was $5.93 billion, down 12% sequentially and down 17% year on year. Data center revenue wa
s $3.83 billion, up 1% sequentially and 31% year-on-year. Gaming revenue was $1.57 billion, down 23% sequentially a
nd down 51% from a year ago.'
```

Case Study 5: Fine Tuning Model - Building Earnings KPI App Assistant

Stock KPI – Key Performance Indicators

	Key Performance Indicators	Prompt
1	Revenue	Provide key information about revenue for the quarter
2	Profitability	Provide key information about profits and losses (P&L) for the quarter
3	Industry Trends	Provide key information about industry trends for the quarter
4	Gaming	Provide key information about Gaming discussed on the call
5	GPU	Provide key information about GPUs discussed on the call
6	AI	Provide key information about AI discussed on the call
7	Data Centers	Provide key information about data centers discussed on the call
8	M&A	Provide any information about mergers and acquisitions discussed on the call
9	Guidance	Provide key information about guidance discussed on the call

Stock KPI Prompt Response (1 of 2)

KPI	Prompt Response
1	Revenue for the quarter was $5.93 billion, down 12% sequentially and down 17% year on year. Data center revenue was $3.83 billion, up 1% sequentially and 31% year-on-year. Gaming revenue was $1.57 billion, down 23% sequentially and down 51% from a year ago.
2	Provide key information about profits and losses (P&L) for the quarter : Revenue for the quarter was $5.93 billion, down 12% sequentially and down 17% year on year. We delivered record data center and automotive revenue, while our gaming and pro visualization platforms declined as we work through channel inventory corrections and challenging external conditions.
3	Provide key information about industry trends for the quarter : Data center revenue was up 1% sequentially and 31% year-on-year, driven primarily by leading U.S. cloud providers and a broadening set of consumer Internet companies for workloads such as large language models, recommendation systems and generative AI. Networking posted strong growth driven by hyperscale customers and easing supply constraints. Gaming revenue was down 23% sequentially and down 51% from a year ago, reflecting lower sell-in to partners to help align channel inventory levels with current demand expectations.
4	Provide key information about Gaming discussed on the call : Gaming revenue was $1.57 billion, down 23% sequentially and down 51% year on year. This reflects lower sell-in to partners to help align channel inventory levels with current demand expectations. The first ADA GPU, the GeForce RTX 4090 became available in mid-October and received positive feedback from the gaming community. RTX 40 Series GPUs features DLSS 3, the neuro rendering technology that uses AI to generate entire frames for faster game play. Portal, now reimagined with full ray tracing and DLSS 3 has made it on Steam's top 100 most wish-listed gains.

Stock KPI Prompt Response (1 of 2)

KPI	Prompt Response
5	Provide key information about GPUs discussed on the call : The call discussed the A100, H-100, RTX 4090, RTX 4080, and the L4 GPUs, all of which are based on the new hopper architecture. The A100 and H-100 are data center GPUs, the RTX 4090 and 4080 are gaming GPUs, and the L4 GPU is based on the ADA Lovelace architecture.
6	During the call, AI was discussed in relation to data center, automotive, energy, and other vertical industries. It was mentioned that the U.S. government announced new restrictions impacting exports of A100 and H-100 based products to China, and that demand in China remains soft. It was also mentioned that NVIDIA has started shipping their flagship 100 data center GPU based on the new hopper architecture, and that A100-based systems are available from leading server makers. Additionally, it was mentioned that NVIDIA and Microsoft are collaborating to build an advanced cloud-based AI supercomputer, and that NVIDIA and Oracle are working together to offer AI training and inference at scale to thousands of enterprises.
7	Provide key information about data centers discussed on the call : Jen-Hsun Huang discussed how NVIDIA's InfiniBand and Magnum IO software stack can improve the efficiency of a data center by 10%. He also discussed the A800 hardware, which ensures that it always meets U.S. government's clear test for export control. Colette Kress mentioned that the compute for the quarter was about flat with the previous quarter, and that NVIDIA is working to help cloud service providers stand up supercomputers with hopper
8	Provide any information about mergers and acquisitions (M&A) discussed on the call : I don't know. 
9	Provide key information about guidance discussed on the call : We expect that demand in China will remain soft in the current quarter. We believe Channel inventories are on track to approach normal levels as we exit Q4. We sold out quickly in many locations and are working hard to keep up with demand.

Extend KPI to Add More Features

Prompt

```
JSONresponse=[]
for i in range(len(KPI)):
    prompt = f"""

    Identify the following items from the text:

    - Determine Sentiment (positive, negative or neutral)
    - Rank Sentiment with score between 1 to 5 with 1\
      as very negative and 5 as very positive

    The text is delimited with triple backticks. \
    Format your response as a JSON object with \
    "{KPI[i]}","{responseKPI[i]}" "Sentiment", "Rank" as the keys




    text: ```{responseKPI[i]}```
    """
    response = get_completion(prompt)
    JSONresponse.append(response)

    print(response, "\n")
```

Json Response

{
"Provide key information about revenue for the quarter": "Revenue for the quarter was $5.93 billion, down 12% sequentially and down 17% year on year. Data center revenue was $3.83 billion, up 1% sequentially and 31% year-on-year. Gaming revenue was $1.57 billion, down 23% sequentially and down 51% from a year ago.",
"Sentiment": "Negative",
"Rank": 2
}

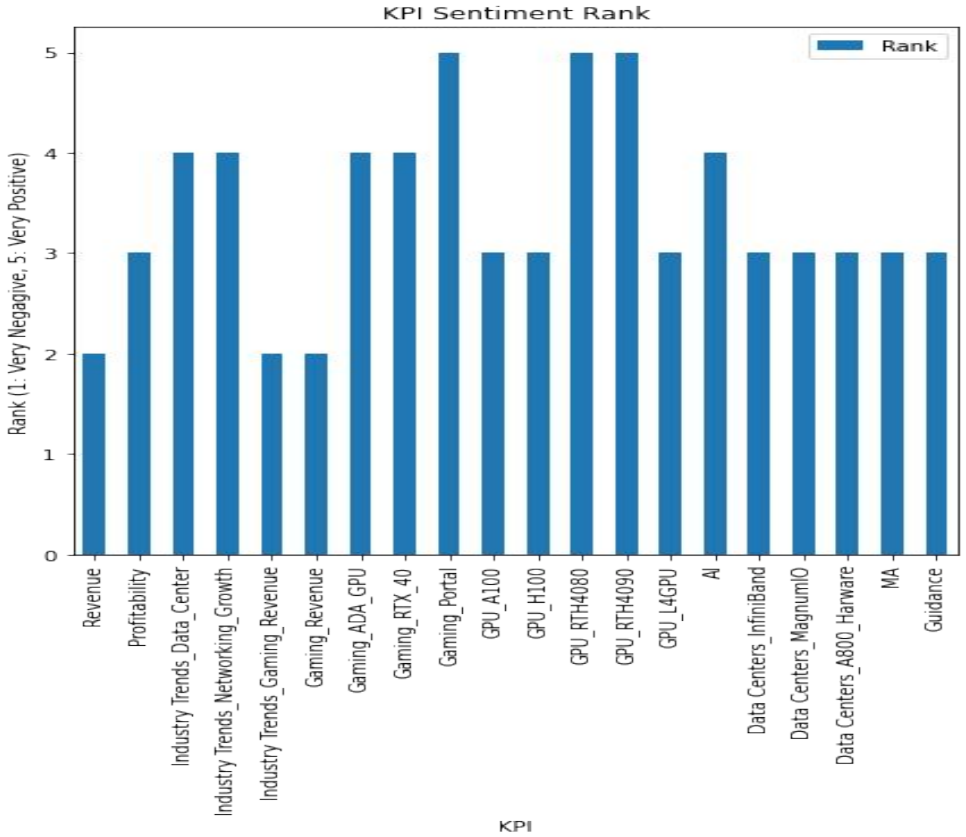
{
"Provide key information about profits and losses (P&L) for the quarter": "Revenue for the quarter was $5.93 billion, down 12% sequentially and down 17% year on year. We delivered record data center and automotive revenue, while our gaming and pro visualization platforms declined as we work through channel inventory corrections and challenging external conditions.",
"Sentiment": "Neutral",
"Rank": 3
}

Stock KPI Prompt Final Product Refined

KPI Sentiment

	Index	KPI	Sentiment	Rank
0	1.0	Revenue	Negative	2
1	2.0	Profitability	Neutral	3
2	3.1	Industry Trends_Data_Center	Positive	4
3	3.2	Industry Trends_Networking_Growth	Positive	4
4	3.3	Industry Trends_Gaming_Revenue	Negative	2
5	4.1	Gaming_Revenue	Negative	2
6	4.2	Gaming_ADA_GPU	Positive	4
7	4.3	Gaming_RTX_40	Positive	4
8	4.4	Gaming_Portal	Positive	5
9	5.1	GPU_A100	Neutral	3
10	5.2	GPU_H100	Neutral	3
11	5.3	GPU_RTH4080	Positive	5
12	5.4	GPU_RTH4090	Positive	5
13	5.5	GPU_L4GPU	Neutral	3
14	6.0	AI	Positive	4
15	7.1	Data Centers_InfiniBand	Neutral	3
16	8.1	Data Centers_MagnumIO	Neutral	3
17	9.1	Data Centers_A800_Harware	Neutral	3
18	8.0	MA	Neutral	3
19	9.0	Guidance	Neutral	3

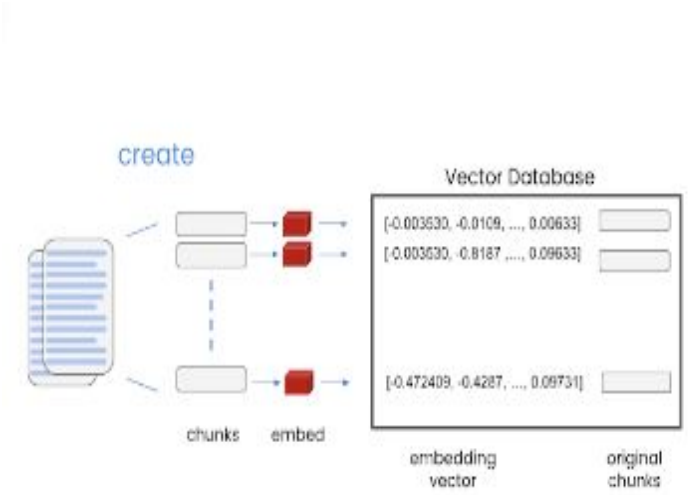
KPI Sentiment Rank



Case Study 6: Stock Domain Queries w/LangChain & Using Chains for Features

Using LangChain for Chains and QA

LangChain Vector Database



Source: https://learn.deeplearning.ai

Prompt (1 of 2)

```
from langchain.chat_models import ChatOpenAI
from langchain.prompts import ChatPromptTemplate
from langchain.chains import LLMChain
```

```
llm = ChatOpenAI(temperature=0.9,openai_api_key=openai.api_key)
```

```
qa_stuff = RetrievalQA.from_chain_type(
    llm=llm,
    chain_type="stuff",
    retriever=retriever,
    verbose=True)
```

```
query1 = """
Provide key information about revenue for the quarter,\n
"""
```

```
response1 = qa_stuff.run(query1)
```

```
> Entering new RetrievalQA chain...
> Finished chain.
```

Query Output

```
Markdown(response1)
```

For the first quarter of fiscal 2024, Nvidia's revenue was $7.19 billion, which was up 19% sequentially and down 13% year

Using LangChain for Chains Build Up and QA

Prompt (2 of 2)

```
Markdown(response1)
```

For the first quarter of fiscal 2024, Nvidia's revenue was $7.19 billion, which was up 19% sequentially and down 13% year

```
prompt = ChatPromptTemplate.from_template(
    " Describe Sentiment (positive, negative or neutrall in \
    {response_}"
)
# chain 2
chain = LLMChain(llm=llm, prompt=prompt)
```

```
response_=response1
```

```
chain.run(response_)
```

```
'The sentiment in this statement is mixed. The use of specific numbers and percentages indicates a factual, neutral
tone. However, the decrease in revenue from the previous year may be seen as negative, while the increase from the
previous quarter may be seen as positive. Overall, the sentiment could be interpreted as slightly negative due to t
he decline in year-over-year revenue.'
```

Prompt Response

```
'The sentiment in this statement is mixed. The use of specific numbers and percentages indicates a factual, neutral
tone. However, the decrease in revenue from the previous year may be seen as negative, while the increase from the
previous quarter may be seen as positive. Overall, the sentiment could be interpreted as slightly negative due to t
he decline in year-over-year revenue.'
```

Case Study 7: Sentiment & News Stock Relevance in Long Short Investing

NVIDIA & Intel Earnings Calls Sentiment

News Sentiment & Stock Specific Relevance

NVIDIA News Data 5.31.23

	Text	Sentiment	Relevance	Sentiment_Label	Signal
0	Generative AI meets gaming in Nvidia's demo, p...	positive	1.0	2	2.0
1	Foxconn is ecstatic you're all going gaga for ...	positive	0.2	2	0.4
2	AI Boom Gets Fresh Gut Check as Nvidia Drops M...	negative	1.0	0	0.0
3	AI Boom Gets Fresh Gut Check as Nvidia Drops M...	negative	1.0	0	0.0
4	AI Boom Gets Fresh Gut Check as Nvidia Drops M...	negative	1.0	0	0.0
5	AI Boom Gets Fresh Gut Check as Nvidia Drops M...	negative	1.0	0	0.0
6	AI Boom Gets Fresh Gut Check as Nvidia Drops M...	negative	1.0	0	0.0
7	AI Boom Gets Fresh Gut Check as Nvidia Drops M...	negative	1.0	0	0.0
8	Nvidia's CEO planning trip to meet China tech ...	neutral	0.8	1	0.8
9	Stock market today: Wall Street slips as stock...	negative	0.2	0	0.0
10	Stock market today: Asian shares mostly rise a...	positive	0.2	2	0.4
11	Future Transportation, Vol. 3, Pages 708-725: ...	neutral	0.5	1	0.5
12	Nvidia joins the exclusive $1 trillion club	positive	1.0	2	2.0
13	Stock market today: Asian shares mostly rise a...	positive	0.2	2	0.4
14	Stock market today: Asian shares mostly rise a...	positive	0.2	2	0.4
15	Super MicroCompute 2023, Server & Storage Arra...	neutral	0.5	1	0.5
16	Stock market today: Asian shares mostly rise a...	positive	0.2	2	0.4
17	Stock market today: Asian shares mostly rise a...	positive	0.2	2	0.4
18	Stocks Rise on China Data Surprise, Fed Pause ...	positive	0.2	2	0.4
19	US Stocks: Wall St falls as labor data spurs r...	negative	0.2	0	0.0
20	Stock market today: Asian shares mostly rise a...	positive	0.2	2	0.4

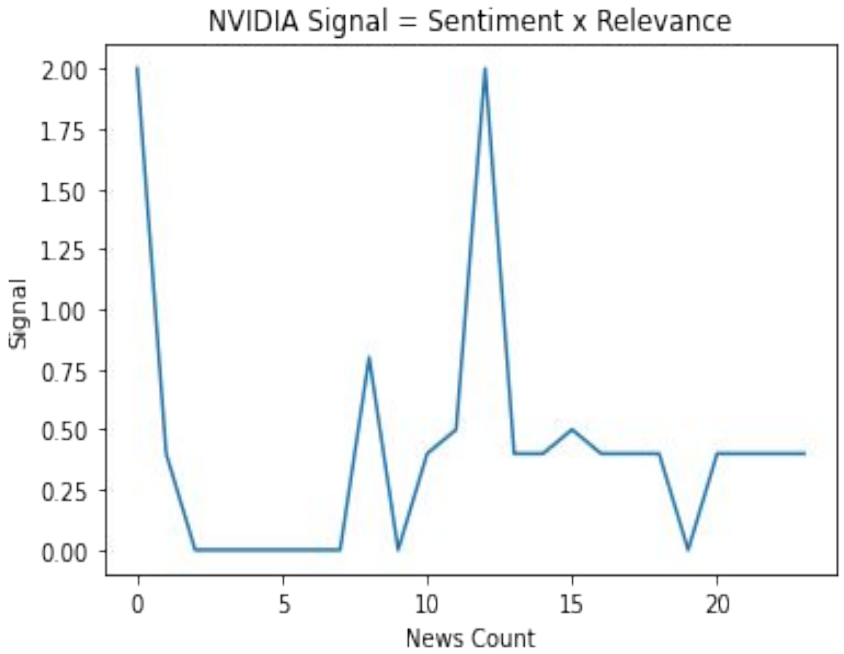
Intel News Data 5.31.23

	Text	Sentiment	Relevance	Sentiment_Label	Signal
0	DUNCAN McLEOD: Why Nvidia is the hottest game ...	positive	0.2	2	0.4
1	Intel shares jump 5% as chipmaker sees Q2 reve...	positive	1.0	2	2.0
2	Tower-Intel deal stalls ISMC's India chip plan	negative	0.5	0	0.0
3	Vending Machine Market size to grow by USD 12,...	positive	0.2	2	0.4
4	Vending Machine Market size to grow by USD 12,...	positive	0.0	2	0.0
5	Vending Machine Market size to grow by USD 12,...	positive	0.2	2	0.4
6	Vending Machine Market size to grow by USD 12,...	positive	0.2	2	0.4
7	Vending Machine Market size to grow by USD 12,...	positive	0.2	2	0.4
8	Vending Machine Market size to	Relevance x Sentiment			0.4
9	Vending Machine Market size to grow by USD 12,...	positive	0.2	2	0.4
10	Vending Machine Market size to grow by USD 12,...	positive	0.2	2	0.4
11	Vending Machine Market size to grow by USD 12,...	positive	0.2	2	0.4
12	Vending Machine Market size to grow by USD 12,...	positive	0.2	2	0.4
13	Nvidia to join $1 Trillion club	positive	0.5	2	1.0
14	India's semiconductor dreams dashed: $3 Billio...	negative	0.5	0	0.0
15	WWDC 2023: Apple Tests New High-End Macs With ...	neutral	0.2	1	0.2
16	AWS is on a mission to port open RAN to Arm an...	neutral	0.5	1	0.5
17	[7/12 (Wed) Fukuoka Chamber of Commerce and In...	neutral	0.2	1	0.2
18	Asus ROG Strix SCAR 18 (2023): brand's newest ...	neutral	0.0	1	0.0
19	Vending Machine Market size to grow by USD 12,...	positive	0.0	2	0.0
20	Three firms pay one-third of all corporation tax	neutral	0.5	1	0.5

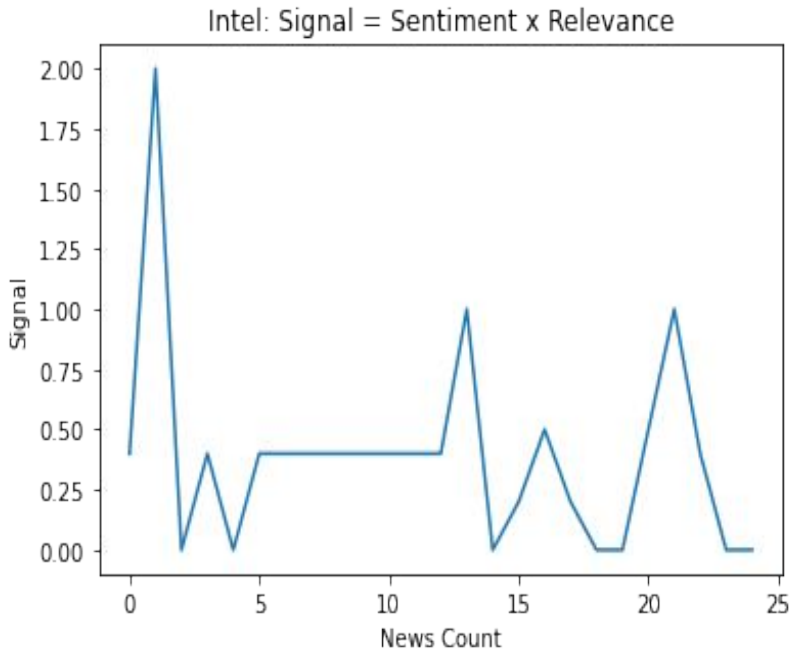
Source of data feed : raven pack

News Sentiment & Stock Specific Relevance

Nvidia Extracted Signal 5.31.23



Intel Extracted Signal 5.31.23



Prompt

Using Prompt for Zero Shot Learning

Zero-Shot Learning

no demonstrations are allowed, and the model is only given a natural language instruction describing the task

```
stock = "NVIDIA"

nvidia_text=open('nvidia_text','w')

for index, row in nvidia.tail(25).iterrows():

    """ 1. Derive Sentiment """

    prompt_sentiment = f"""
    What is the sentiment of the following text,
    Give your answer as a single word, either "positive" \
    "negative" or "neutral"

    text: ```{row['TITLE']}```

    """



    """ 2. Derive Stock Specific Relevance """


    prompt_relevance= f"""

    Determine if the text is relevant for {stock} by ranking it between 0 and 1. \
    If very relevant, label it as 1. If least relevant, label it as 0

    text: ```{row['TITLE']}```

    """


    response_sentiment = get_completion(prompt_sentiment)
    response_relevance = get_completion(prompt_relevance)

    response_string = row['TITLE'] + ";" + response_sentiment + ";" + response_relevance
```

Reference Material

References

ChatGPT Cheat Sheet

OpenAI Cookbook

Prompt Engineering

Prompt Engineering with LangChain

System Design with ChatGPT

Disclaimer

The information provided in this presentation does not, and is not intended to constitute investment advice; instead, all information, content, and materials available in this presentation site are for general informational purposes only.

The views expressed at, or through, this presentation are those of the individual authors writing in their individual capacities only – not those of their respective employers, or committee/task force as a whole. All liability with respect to actions taken or not taken based on the contents of this site are hereby expressly disclaimed.