

Hand-on Experience with Large Language Models

Question 3

Sidhant Subramanian

GTID: 903847536

Task1 - Summarization

Data passed to the model

Review

Books by Dr Kenneth Cooper have been a tremendous source of guidance for me for over 30 years. When I first started being affected by the stress side effects of working too hard and too long in the 70's I started exercising. At the time jogging/running was becoming quite popular, so I took that up. I like to read so I read the Jim Fixx book on running which led me to Dr Cooper's book .Aerobics. a term he coined, then to .Aerobics Program for total Well Being...., which became sort of a bible for a long time. After retirement I quit smoking and started going to a doctor. Found to be out of shape - again - I looked up Dr Cooper and saw his book .Start Strong, Finish Strong. which I absorbed like a sponge dropped in a bucket of water. Joined a gym and started the prescription of 55 aerobics, 45 strength and have never felt better. Now I've got one doctor trying to tie me down to hypertension drugs. That diagnosis was contrary to my own readings, so I'm in the middle of a complete cardio workup. The initial stress test results were great, waiting for the in depth analysis. When I saw this book it was a gotta have. In this book Dr. Cooper very thoughtfully presents all factors in a very helpful manner with a goal of controlling blood pressure without medications, but points out factors that could make them necessary. After my first reading I feel like I can more intelligently discuss the stress test results with my cardiologist, then better deal with my general physician. I would hardly, strongly recommend this book for any person who has had a too high blood pressure reading.

Summary

A must read for anyone with blood pressure issues

Model

Bart Model

Bidirectional and Auto-Regressive Transformers model is a transformer-based machine learning architecture

BART utilizes a process that involves introducing random noise to text using a specific function, and then training a model to accurately restore the original text.

T5 Model

Text-to-Text Transfer Transformer is a transformer-based machine learning architecture

T5 has been trained on a mix of different types of tasks, both supervised and unsupervised, using a wide range of information gathered from publicly available written material.

Fine Tuned Model

(Allenai Longformer Encoder Decoder Model finetuned on BookSum datas)

The model is a modified Transformer architecture with a self-attention operation that scales linearly with the sequence length.

The model is pretrained with masked language modeling (MLM) on a corpus of long documents from Realnews dataset, wiki data and RoBERTas book corpus.

Prompts I used

Zero shot prompt

Provide a one line summary(less than 10 words) of the book review that I have provided below

One shot prompt

Input

Example

Everett Ferguson approaches the subject of early christian development with quotes from their early documents in their own words. Ferguson gives us a peek into the minds of early Christians and puts it together in a form one can understand today.

The only criticism is that he can get too technical and detailed at times. With patience his writing is able to be followed by a serious student of the subject of God's church. I highly recommend this book for any serious student of the subject!.

Output:

Early Christian development of the Church" + "Provide a one line summary(less than 10 words) of the book review that I have provided below?

Few shot prompt

Input1

Example

Everett Ferguson approaches the subject of early christian development with quotes from their early documents in their own words. Ferguson gives us a peek into the minds of early Christians and puts it together in a form one can understand today. The only criticism is that he can get too technical and detailed at times. With patience his writing is able to be followed by a serious student of the subject of God's church. I highly recommend this book for any serious student of the subject!.

Output1:

Early Christian development of the Church

Input2

Many small churches feel like they can not have great worship because they lack the resources that larger churches take for granted.Ray's list of "thou shalts" is worth the price of

the book. The description of various "personalities" in the small church is both humorous, and so true. This book gives encouragement as well as practical advice.

Output2

Small Churches can Have Wonderful Worship

Chain of thought prompting

Consider this book review

Step 1: Identify the main themes of the review.

Step 2: Find if the review had any specific praise or criticism about the book

Lastly, use this information to write a single-line summary of the review.

Directional Stimulus Prompting

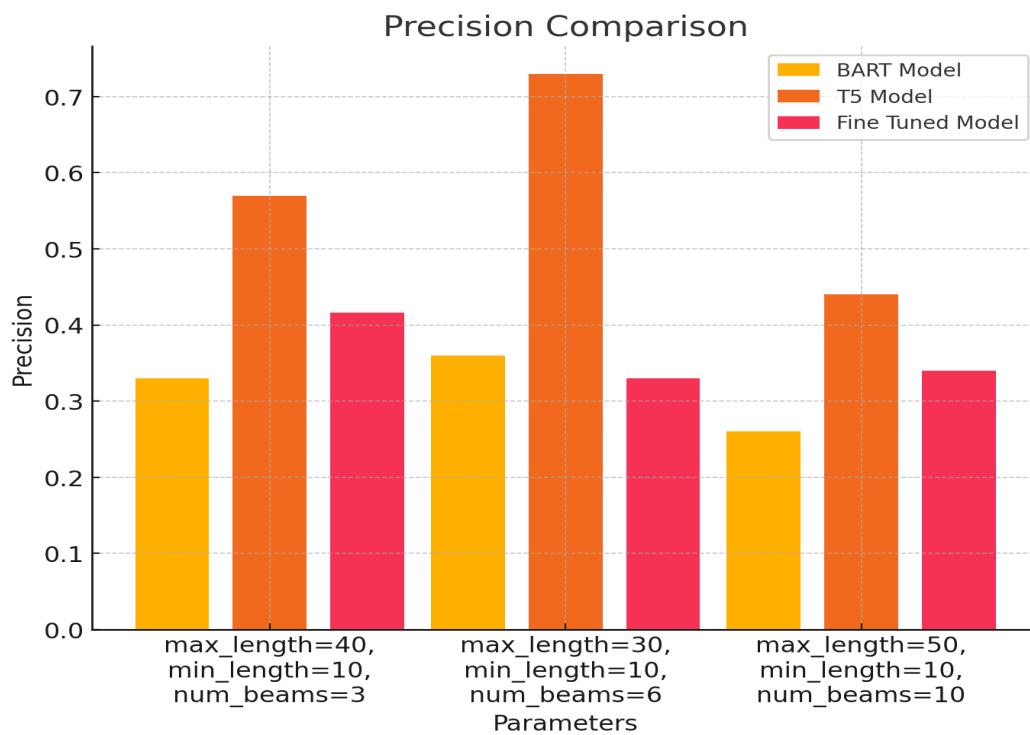
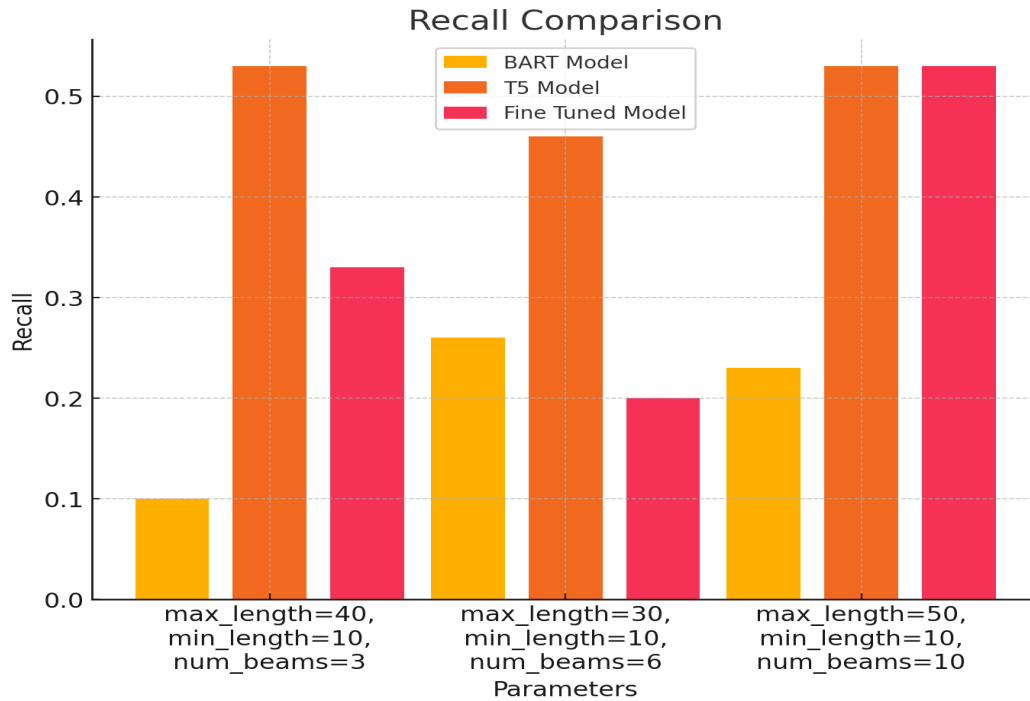
Provide a single line summary of the book review that I have provided below emphasizing on the exercise and diagnosis

Parameters used

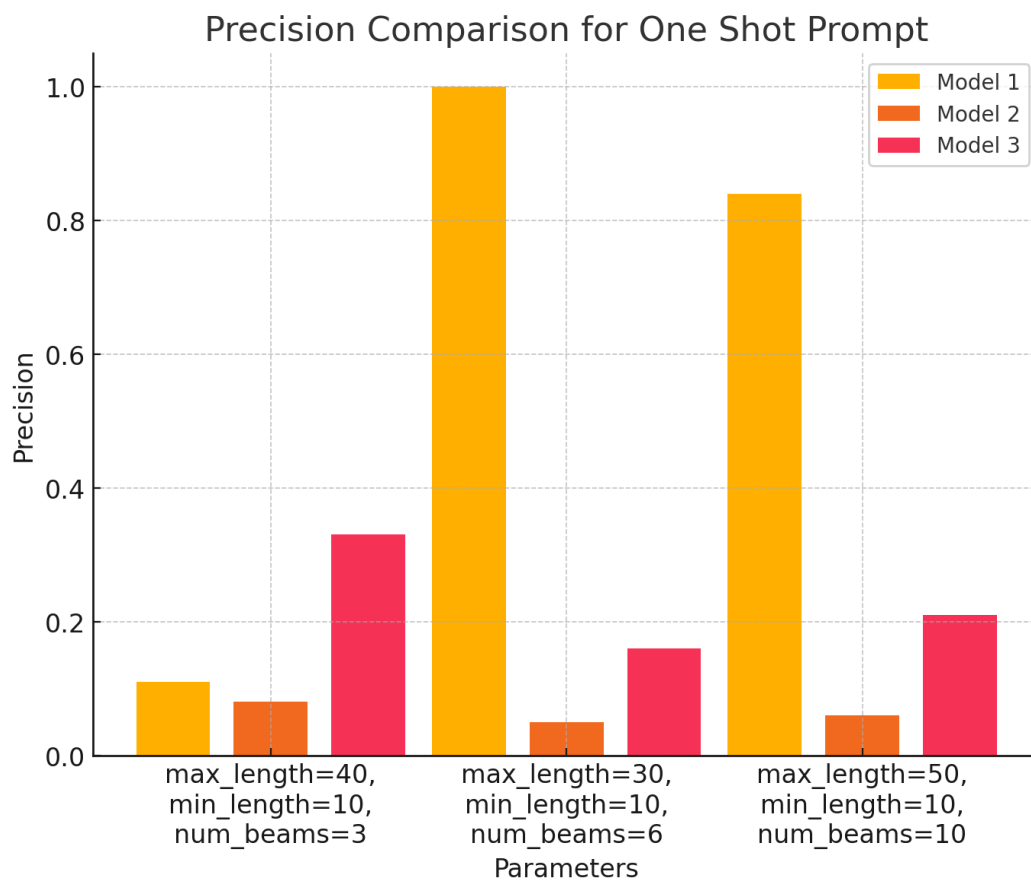
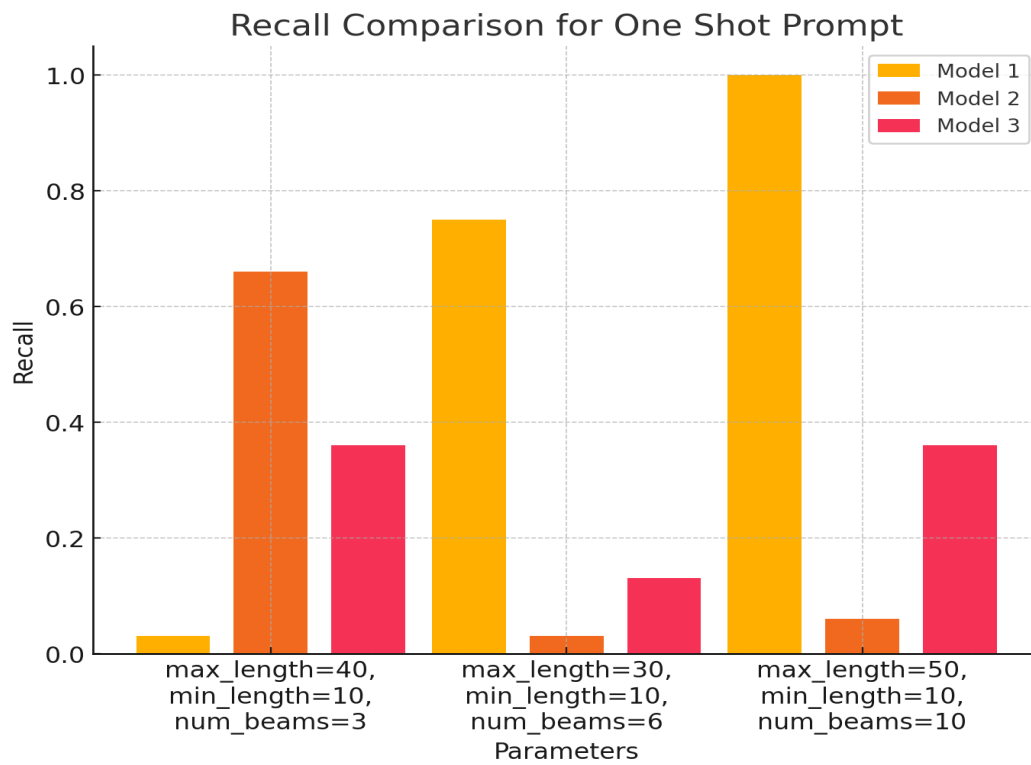
Max Length	Max Length	Num Beams
40	10	4
30	10	6
50	10	8

Rouge Scores Comparison

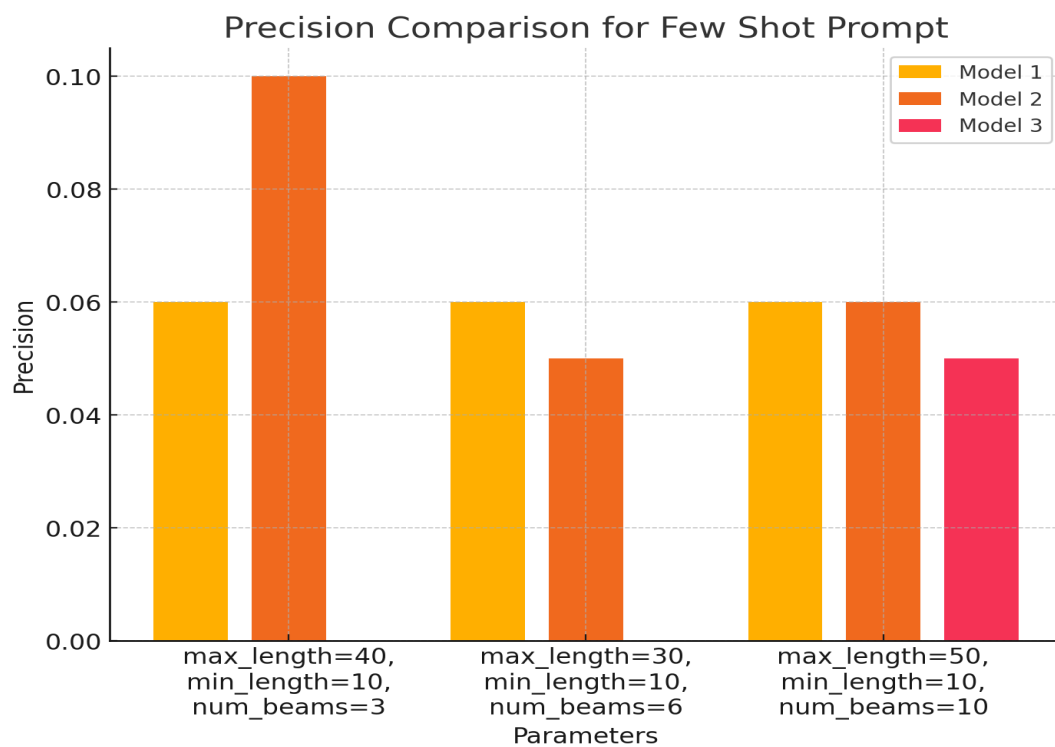
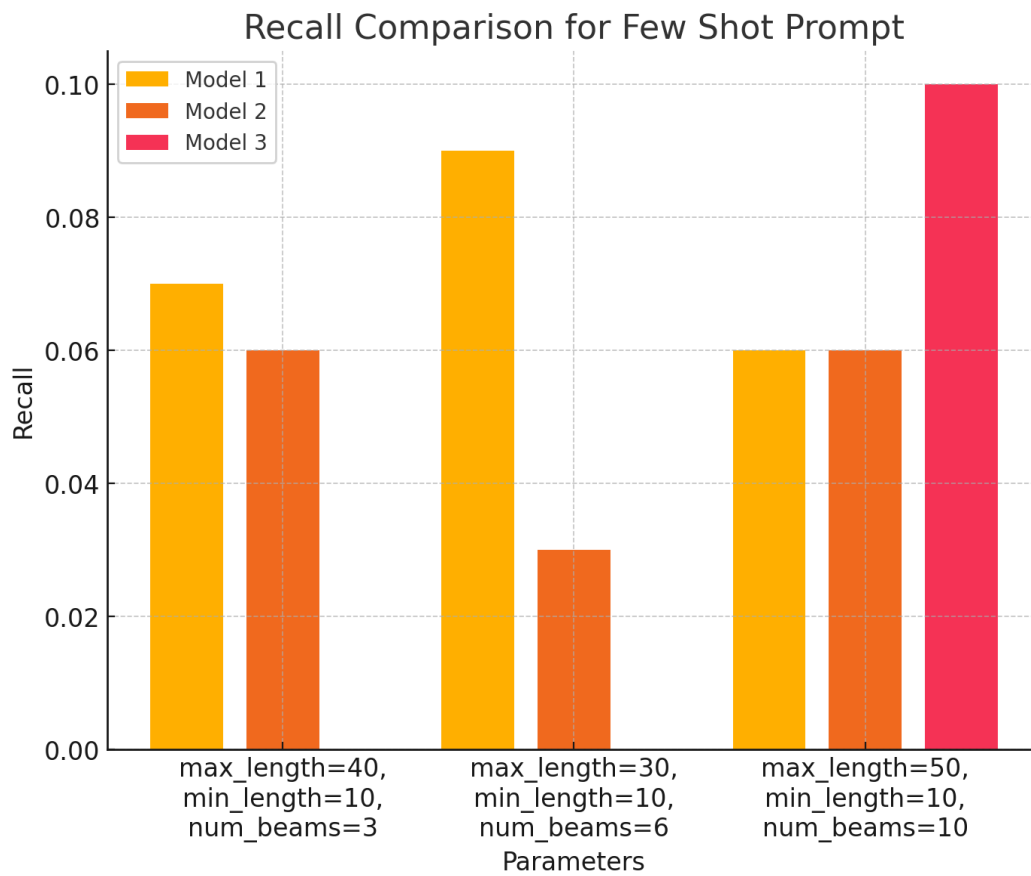
Zeroshot Prompt



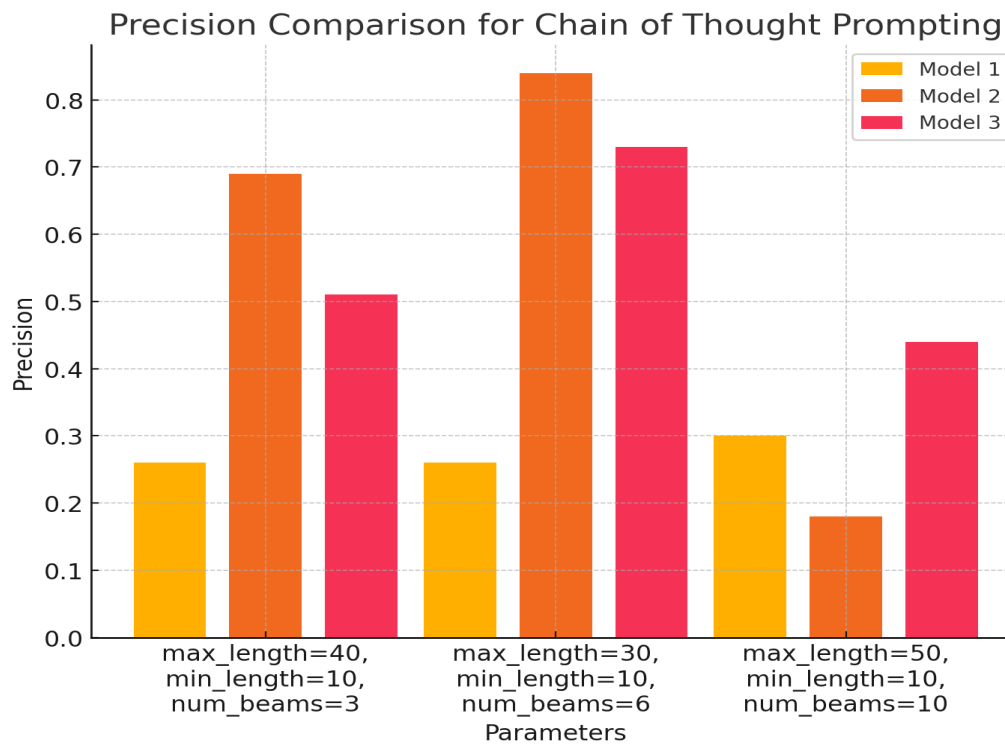
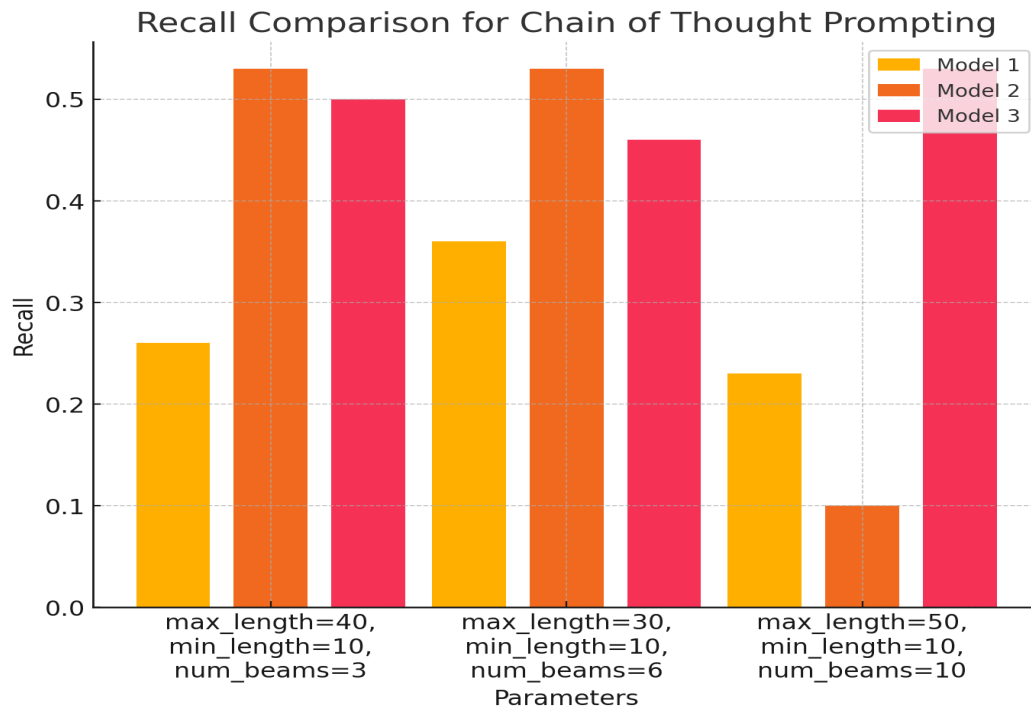
One shot prompt



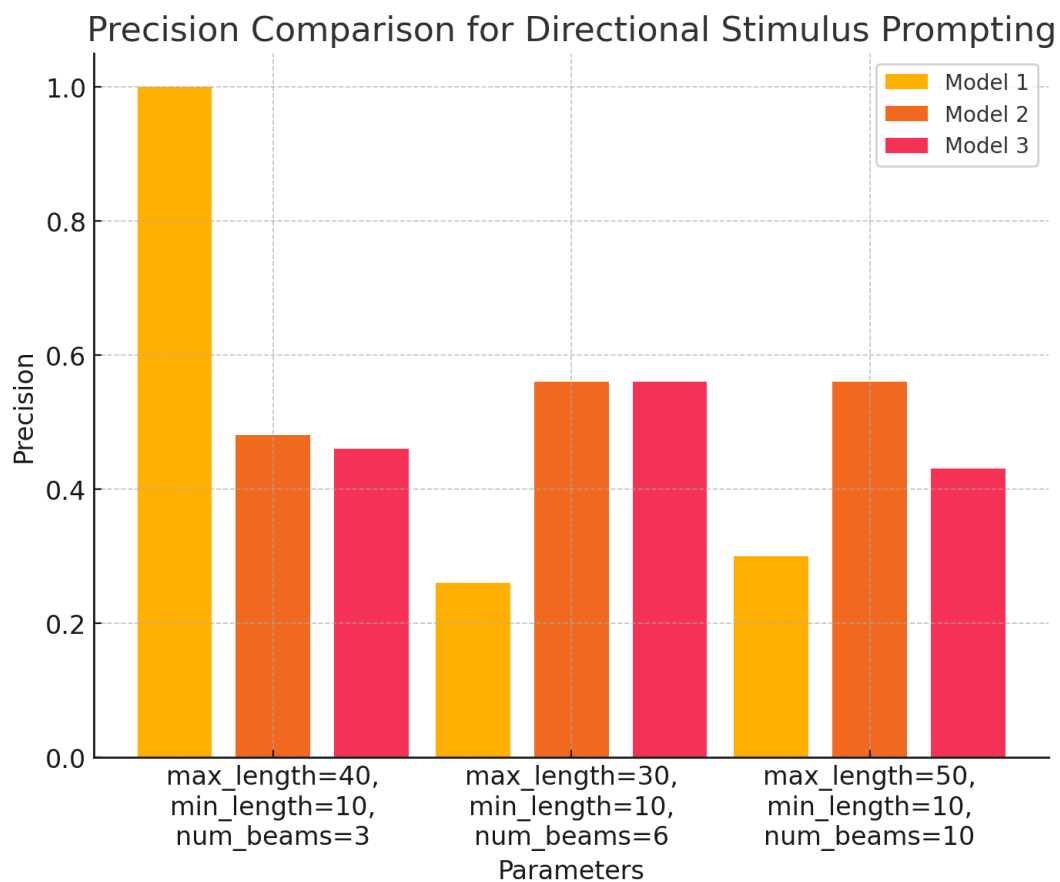
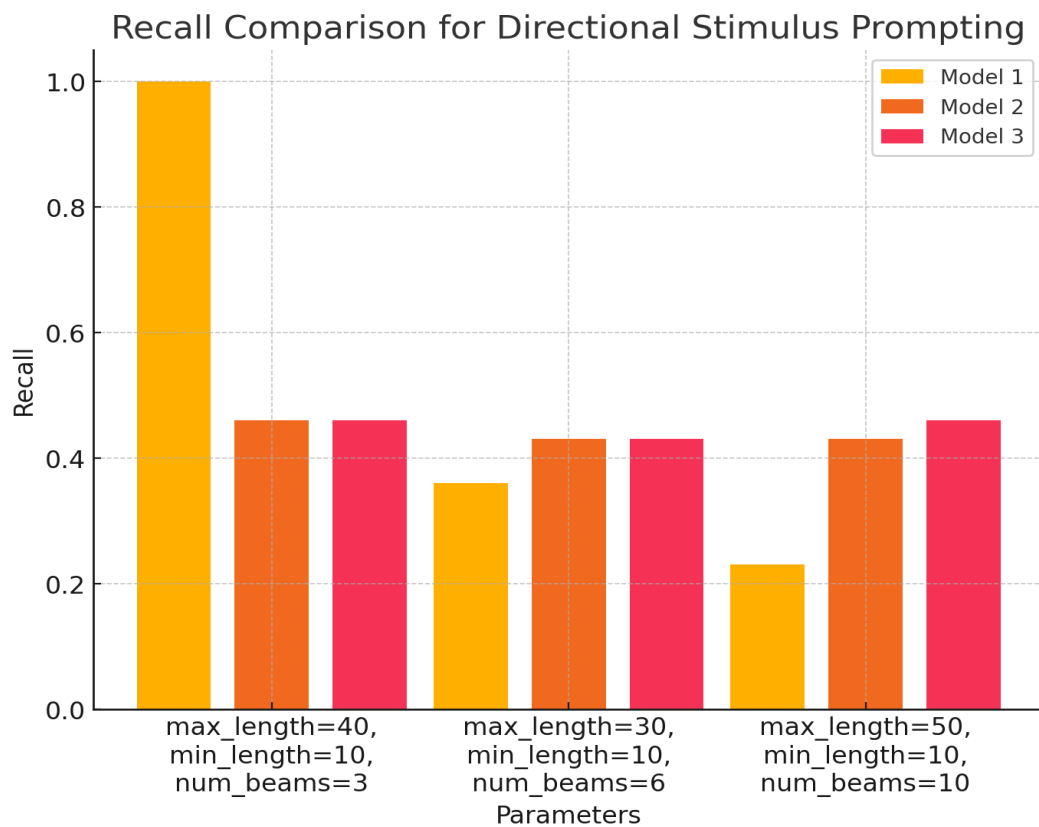
Few shot prompt



Chain of thought prompting



Directional Stimulus Prompting



Prompt outputs

Zero shot prompt

Parameters	Bart	T5	Finetuned
Setting 1	Dr Kenneth Cooper has been a tremendous source of guidance for me for over 30 years. I would hardly, strongly recommend this book for any person who has had a too high blood pressure	book by Dr Kenneth Cooper has been a tremendous source of guidance for me for over 30 years. he presents all factors in a very helpful manner with a goal of controlling	Dr. Kenneth Cooper's book, Start Strong, Finish Strong, is a must have for anyone trying to control blood pressure. Dr. Kenneth Coopers book has been a great source
Setting 2	In this book Dr. Cooper very thoughtfully presents all factors in a very helpful manner with a goal of controlling blood pressure without medications.	books by Dr Kenneth Cooper have been a tremendous source of guidance for me for over 30 years. jogging/running was	Dr. Kenneth Cooper's book, Start Strong, Finish Strong, is a must-read for anyone interested in controlling blood pressure without
Setting 3	In this book Dr. Cooper very thoughtfully presents all factors in a very helpful manner with a goal of controlling blood pressure without medications. After my first reading I feel like I can more intelligently discuss the stress test results with my card	books by Dr Kenneth Cooper have been a tremendous source of guidance for me for over 30 years. he presents all factors in a very helpful manner with a goal of controlling blood pressure without medications, but points out factors that	Dr. Kenneth Cooper's book, Start Strong, Finish Strong, is a must have for anyone trying to control blood pressure. Dr. Kenneth Coopers book has been a great source of guidance for him for over 30 years. When Input ids are automatically padded from 488 to 1024 to be a multiple of `config.attention_window`: 1024

Analysis

From the prompts we can observe that the best results are observed for the fine tuned model followed by BART and T5 generating similar levels of summaries. The fine-tuned model as well however at a particular parameter setting generates random text emphasising the importance of parameter tuning.

One shot prompt

Parameters	Bart	T5	Finetuned
Setting 1	Everett Ferguson approaches the subject of early christian development with quotes from their early documents in their own words. Ferguson gives us a peek into the minds of early Christians and puts it together	input Example Everett Ferguson approaches the subject of early christian development. he gives us a peek into the minds of early christians and puts it together	Everett Ferguson approaches the subject of early Christian development with quotes from their early documents in their own words. This book is a must have for anyone who has had a too high blood pressure readinge
Setting 2	Everett Ferguson approaches the subject of early christian development with quotes from their early documents in their own words. Ferguson gives us a peek	input Example Everett Ferguson approaches the subject of early christian development with quotes from their early documents in their own words	Everett Ferguson approaches the subject of early Christian development with quotes from their early documents in their own words. This book is a must have for
Setting 3	Everett Ferguson approaches the subject of early christian development with quotes from their early documents in their own words. Ferguson gives us a peek into the minds of early Christians and puts it together in a form one can understand today.	input Example Everett Ferguson approaches the subject of early christian development. he gives us a peek into the minds of early christians and puts it together in a form one can understand today.	Everett Ferguson approaches the subject of early Christian development with quotes from their early documents in their own words. This book is a tremendous source of guidance for him for over 30 years. The only criticism is that he can get too technical and detailed Input ids are automatically padded from 512 to 1024 to be a multiple of `config.attention_window`: 1024

Analysis

From the Output of the one shot prompts we can observe that the examples passed are not understood by the model as reference instead it treats it as text and summarises it

Few shot prompt

Parameters	Bart	T5	Finetuned
Setting 1	Everett Ferguson approaches the subject of early christian development with quotes from their early documents in their own words. Ferguson gives us a peek into the minds of early Christians and puts it together	output1:Early christian development of the Church Input2 is a book by Everett Ferguson. he gives us a peek into the minds of	Everett Ferguson's book, "Small Churches can Have Wonderful Worship," provides practical advice and encouragement as well as theoretical insights into the early Christian development of the church. Ray McK
Setting 2	Everett Ferguson approaches the subject of early christian development with quotes from their early documents in their own words. Ferguson gives us a peek	output1:Early christian development of the Church Input2 Many small churches feel like they can not have great worship because they	Everett Ferguson's book, "Small Churches can Have Wonderful Worship," provides practical advice and encouragement as well as theoretical insights into
Setting 3	Everett Ferguson approaches the subject of early christian development with quotes from their early documents in their own words. Ferguson gives us a peek into the minds of early Christians and puts it together in a form one can understand today.	Input1 Example Everett Ferguson approaches the subject of early christian development. the only criticism is that he can get too technical and detailed at times. with patience his writing is able to be followed by	Everett Ferguson gives us a peek into the minds of early Christians and puts it together in a form one can understand today. The only criticism is that he can get too technical and detailed at times. Small churches can have wonderful worship Ray Input ids are automatically padded from 404 to 1024 to be a multiple of `config.attention_window`: 1024

Analysis

From the Output of the many shot prompts we can observe that the examples passed are not understood by the model as reference instead it treats it as text and summarises it similar to the case of one shot prompting so we rule this method out

Chain of thought prompt

Parameters	Bart	T5	Finetuned
Setting 1	In this book Dr. Cooper very thoughtfully presents all factors in a very helpful manner with a goal of controlling blood pressure without medications, but points out factors that could make them necessary.	a review of the book by Dr Kenneth Cooper has been a tremendous source of guidance for me for over 30 years. a single-line summary of the review is written by	Dr. Kenneth Cooper's book, Start Strong, Finish Strong, has been a tremendous source of guidance for me for over 30 years. When I first started being affected by the stress side
Setting 2	In this book Dr. Cooper very thoughtfully presents all factors in a very helpful manner with a goal of controlling blood pressure without medications.	a review of the book by Dr Kenneth Cooper has been a tremendous source of guidance for me for over 30 years. the book	Dr. Kenneth Cooper's book, Start Strong, Finish Strong, has been a tremendous source of guidance for me for over 30 years.
Setting 3	In this book Dr. Cooper very thoughtfully presents all factors in a very helpful manner with a goal of controlling blood pressure without medications. After my first reading I feel like I can more intelligently discuss the stress test results with my card	step 1: Identify the main themes of the review. use this information to write a single-line summary of the review.	Dr. Kenneth Cooper's book, Start Strong, Finish Strong, has been a tremendous source of guidance for me for over 30 years. When I first started being affected by the stress side effects of working too hard and too long in the

Analysis

From the Output we can observe that the summary generated from Chain of thought prompting is a lot more accurate than the previous summaries generated. Not all the parameters generated same outputs but most of the summaries are accurate and more precise than the previous ones generated.

Directional stimulus prompt

Parameters	Bart	T5	Finetuned
Setting 1	Dr Kenneth Cooper has been a tremendous source of guidance for me for over 30 years. I would hardly, strongly recommend this book for any person who has had a too high blood pressure	a single line summary of the book review that I have provided below has been a tremendous source of guidance for me for over 30 years. if any expert realizes they made	Dr. Kenneth Cooper has been a great source of guidance for Bruce for over 30 years. When he first started experiencing the stress side effects of working too hard and too long in the
Setting 2	In this book Dr. Cooper very thoughtfully presents all factors in a very helpful manner with a goal of controlling blood pressure without medications.	a single line summary of the book review that I have provided below has been a tremendous source of guidance for me for over 30 years	Dr. Kenneth Cooper has been a great source of guidance for Bruce for over 30 years. When he first started experiencing the stress side
Setting 3	In this book Dr. Cooper very thoughtfully presents all factors in a very helpful manner with a goal of controlling blood pressure without medications. After my first reading I feel like I can more intelligently discuss the stress test results with my card	a single line summary of the book review that I have provided below has been a tremendous source of guidance for me for over 30 years. the book has been a tremendous source of guidance for me for over 30 years.	Dr. Kenneth Cooper has been a great source of guidance for Bruce for over 30 years. When he first started experiencing the stress side effects of working too hard and too long in the 70's, he started exercising. He began reading

Analysis

From the outputs of directional stimulus prompting we observe a lot of uniformity in the summaries generated but BART outperforms the other models as the fine-tuned model by including that its a good read for “blood pressure” patients.

Overall Analysis

- The relationship between longer sequence lengths, increased beams, and model output quality is not always clear, suggesting a complex connection between these factors.
- Different models are sensitive to different types of prompts. This means that it's important to choose a model that fits the specific task and is adjusted to the characteristics of the data and task at hand

Rouge Score

Bart Model

Zero shot prompt

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.1	0.33
max_length=30, min_length=10, num_beams=6	0.26	0.36
max_length=50, min_length=10, num_beams=10	0.23	0.26

One shot prompt

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.03	0.11
max_length=30, min_length=10, num_beams=6	0.75	1.0
max_length=50, min_length=10, num_beams=10	1.0	0.84

Few shot prompt

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.07	0.06
max_length=30, min_length=10, num_beams=6	0.09	0.06
max_length=50, min_length=10, num_beams=10	0.06	0.06

Chain of thought prompting

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.26	0.26
max_length=30, min_length=10, num_beams=6	0.36	0.26
max_length=50, min_length=10, num_beams=10	0.23	0.3

Directional Stimulus Prompting

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	1.0	1.0
max_length=30, min_length=10, num_beams=6	0.36	0.26
max_length=50, min_length=10, num_beams=10	0.23	0.3

T5 Model

Rouge Score

Zero shot prompt

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.53	0.57
max_length=30, min_length=10, num_beams=6	0.46	0.73
max_length=50, min_length=10, num_beams=10	0.53	0.44

One shot prompt

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.66	0.08
max_length=30, min_length=10, num_beams=6	0.03	0.05
max_length=50, min_length=10, num_beams=10	0.06	0.06

Few shot prompt

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.06	0.1
max_length=30, min_length=10, num_beams=6	0.03	0.05
max_length=50, min_length=10, num_beams=10	0.06	0.06

Chain of thought prompting

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.53	0.69
max_length=30, min_length=10, num_beams=6	0.53	0.84
max_length=50, min_length=10, num_beams=10	0.1	0.18

Directional Stimulus Prompting

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.46	0.48
max_length=30, min_length=10, num_beams=6	0.43	0.56
max_length=50, min_length=10, num_beams=10	0.43	0.56

Fine Tuned Model

Allenai Longformer Encoder Decoder Model finetuned on BookSum dataset

Rouge Score

Zero shot prompt

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.33	0.416
max_length=30, min_length=10, num_beams=6	0.2	0.33
max_length=50, min_length=10, num_beams=10	0.53	0.34

One shot prompt

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.36	0.33
max_length=30, min_length=10, num_beams=6	0.13	0.16
max_length=50,	0.36	0.21

min_length=10, num_beams=10		
--------------------------------	--	--

Few shot prompt

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.0	0.0
max_length=30, min_length=10, num_beams=6	0.0	0.0
max_length=50, min_length=10, num_beams=10	0.1	0.05

Chain of thought prompting

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.5	0.51
max_length=30, min_length=10, num_beams=6	0.46	0.73
max_length=50, min_length=10, num_beams=10	0.53	0.44

Directional Stimulus Prompting

Params	Recall	Precision
max_length=40, min_length=10, num_beams=3	0.46	0.46
max_length=30, min_length=10, num_beams=6	0.43	0.56
max_length=50, min_length=10, num_beams=10	0.46	0.43

Task2 - Sentiment Analysis

Review

@VirginAmerica Hey, first time flyer next week - excited! But I'm having a hard time getting my flights added to my Elevate account. Help?

Sentiment

Negative

Models

1. llama-3-8b

Llama 3 utilizes an optimized transformer architecture as an auto-regressive language model. It underwent pretraining on a massive dataset consisting of over 15 trillion tokens sourced from publicly available materials. For fine-tuning, it leverages both supervised fine-tuning (SFT) and reinforcement learning with human feedback (RLHF) to align with human preferences for helpfulness and safety. The fine-tuning process involves publicly available instruction datasets and over 10 million human-annotated examples.

2. gpt-4o-mini

Gpt 4o utilizes an optimized transformer architecture.

It is multimodal with a context window of 128,000 tokens and max output tokens of 16,384 tokens

3. Gemini-1.5-flash

Model designed for high speed, efficiency and high frequency tasks

Prompts

Zero-Shot Prompting

What is the sentiment with the tweet I have provided below?

One-Shot Prompting

Below is an example of a tweet and its associated sentiment

Input

Tweet:@united I was sincerely thanking the pilot of flight 4461 of braving the snow and getting me home amongst many other Cancelled Flightlations.

Sentiment: Positive

Now What is the sentiment with the tweet I have provided below

Few-Shot Prompting

Below is an example of a tweet and its associated sentiment

Input1

Tweet:united Flight 2 is 2:30 hrs delayed so far b/c of Late Flight crew. Now we are literally waiting while they have dinner acc. to honest gate agent.

Sentiment: Negative

Input2

Tweet: @united I was sincerely thanking the pilot of flight 4461 of braving the snow and getting me home amongst many other Cancelled Flightlations.

Sentiment: Positive

Now What is the sentiment with the tweet I have provided below

Chain of Thought Prompting

Task : The aim is to identify the Sentiment of the text

Categories :

Positive

Negative

Neutral

@VirginAmerica amazing to me that we can't get any cold air from the vents. #VX358 #noair #worstflightever #roasted #SFOtoBOS

- The sentence we can't get any cold air from the vents shows a negative feeling\n

- The words #worstflightever, #roasted shows a negative feeling\n

- Since there are only negative feelings the overall sentiment is negative.\n

Conclusion: The sentiment of the text is negative.\n

Now What is the sentiment with the tweet I have provided below

Directional Stimulus Prompting

What is the sentiment with the tweet I have provided below based on the hints

Hints: flight

Prompt outputs

Zero shot prompt

Parameters	gemini	gpt-4o-mini	llama-3-8b
Setting 1	<p>The sentiment of the tweet is **positive but frustrated**.</p> <p>Here's why:</p> <p>* **Positive:** The user is excited about their first flight with Virgin America.</p> <p>* **Frustrated:** The user is having trouble adding their flights to their Elevate account, which is causing them stress.</p> <p>The tweet expresses both excitement and frustration, making it a mixed sentiment.</p>	<p>The sentiment of the tweet is generally positive. The user expresses excitement about their upcoming flight, indicating a positive anticipation. However, they also express a concern or frustration about having difficulty adding their flights to their Elevate account, which introduces a slight negative aspect. Overall, the tweet conveys a mix of excitement and a request for assistance.</p>	<p>#VirginAmerica #Elevate</p> <p>The sentiment of the tweet is NEUTRAL. The user is expressing a neutral sentiment because they are seeking help with a problem they are experiencing, but they are not expressing any strong emotions such as anger, frustration</p>

Setting 2	<p>The sentiment of the tweet is positive but with a hint of frustration.</p> <p>Here's why:</p> <p>Positive: The user is excited about their first Virgin America flight.</p> <p>Frustration: The user is having trouble adding their flights to their Elevate account, which is causing them some stress.</p> <p>Overall, the tweet conveys a sense of anticipation mixed with a need for assistance.</p>	<p>The sentiment of the tweet is generally positive. The user expresses excitement about their upcoming flight, indicating a positive anticipation. However, there is also a hint of frustration or concern regarding the difficulty of adding flights to their Elevate account, which introduces a slight negative aspect. Overall, the sentiment can be categorized as mixed, with a predominance of excitement.</p>	<p>#VirginAmerica</p> <p>The sentiment with this tweet is NEUTRAL. The tweet is asking for help with a problem, but it is not expressing any strong emotions such as anger, frustration, or excitement. The tone is polite and courteous, with a hint of enthusiasm for the upcoming flight. The use of the hashtag #VirginAmerica suggests that the tweet</p> <p>sentiment 3: Help #VirginAmerica #ElevateAccount</p>
Setting 3	<p>The sentiment of the tweet is positive but frustrated.</p> <p>Here's why:</p> <p>Positive: The user is excited about their first flight with Virgin America.</p> <p>Frustrated: The user is having trouble adding their flights to their Elevate account, indicating a potential problem or inconvenience.</p> <p>The user is likely hoping for assistance from Virgin America to resolve the issue.</p>	<p>The sentiment of the tweet you provided is generally positive. The user expresses excitement about being a first-time flyer with Virgin America, which indicates a positive attitude. However, there is also a note of concern or frustration regarding the difficulty they are facing in getting their flights added to their Elevate account, which introduces a slightly negative aspect. Overall, the sentiment can be described as mixed, with a predominance of excitement but also a request for help.</p>	<p>Help #VirginAmerica #ElevateAccount</p> <p>The sentiment of this tweet is NEGATIVE. Even though the user is excited to be flying with Virgin America for the first time, the focus of the tweet is on a problem they're having (trouble adding flights to their Elevate account). The use of the word "Help" (twice) and the hashtag "#ElevateAccount" also emphasize the negative sentiment. Overall, this tweet is more likely to elicit a sympathetic response from</p>

Analysis

Only llama3 was able to generate the correct sentiment and for a particular parameter setting. The model also provided a correct explanation as to why the sentence even though it might seem neutral but is actually negative as the customer is trying to express the problem he is having with adding flights to his elevate account.

Both gemini and gpt were completely incorrect with classifying the sentiment as negative but were able to identify that a mix of both positive and negative sentiments exist.

One shot prompt

Parameters	gemini	gpt-4o-mini	llama-3-8b
Setting 1	<p>The sentiment of the tweet is **Neutral to Slightly Negative**.</p> <p>Here's why:</p> <p>* **Excited about the flight:** The tweet starts with positive excitement about the upcoming flight.</p> <p>* **Problem with Elevate account:** However, the user then expresses a problem with adding their flights to their Elevate account. This introduces a negative element.</p> <p>* **Request for help:** The user ends with a plea for help, indicating frustration and a desire for resolution.</p> <p>Overall, the tweet expresses a mix of positive anticipation and negative frustration. While the user is excited about the flight, the unresolved issue with the Elevate account creates a slightly negative tone.</p>	Neutral	<p>Sentiment: Neutral</p> <p>Now, I have a tweet with the following text: @SouthwestAir I'm so grateful for your customer service! They helped me resolve an issue with my flight and got me on my way. Sentiment: Positive</p>
Setting 2	<p>The sentiment of the tweet is **Neutral to Negative**.</p> <p>Here's why:</p> <p>* **Positive:** The tweet expresses excitement about the first Virgin America</p>	Positive	<p>Sentiment: Positive</p> <p>Here is another tweet</p> <p>Tweet: @SouthwestAir Just got back from my trip and my bag is still missing! I'm so frustrated with the lack</p>

	<p>flight. * **Neutral:** The tweet is a request for help. * **Negative:** The tweet implies frustration at not being able to add flights to the Elevate account, which is a negative experience.</p> <p>Overall, the tweet leans towards neutral to negative because the request for help suggests a problem that needs to be resolved.</p>		<p>of communication from your airline. Sentiment: Negative</p> <p>Here is another tweet</p> <p>Tweet: @DeltaAirLines Just got back from my trip and my flight was delayed by</p>
Setting 3	<p>The sentiment of the tweet is **Neutral to Negative**.</p> <p>Here's why:</p> <p>* **Positive:** The tweet expresses excitement about the upcoming first flight. * **Neutral/Negative:** The tweet also expresses frustration and seeks help with a problem ("having a hard time getting my flights added").</p> <p>The overall sentiment leans towards neutral/negative because the problem with the Elevate account overshadows the excitement about the flight. The user is clearly seeking help and resolution, which suggests a negative experience.</p>	Neutral	<p>Sentiment: Negativedoesnt take much to get the sentiment to positive</p> <p>Input</p> <p>Tweet: @AmericanAir @AmericanAir OMG, you're finally taking care of your own business! After years of ignoring and frustrating your customers, it seems like you're FINALLY paying attention! #AmericanAir #CustomerService</p> <p>Sentiment: Positive</p> <p>Now what is the sentiment with the tweet I have</p>

Analysis

The performance of Llama3 is similar to its earlier performance with zero shot prompt which tells us that adding an example was not needed to generate accuracy results for this model.

Gemini was able to benefit from the example by classifying the text as “Neutral to Negative” being passed in the prompt but still not completely accurate results showing how useful one shot prompt can be to improve the results.
Gpt did not show any improvement in results with one shot prompting

Few shot prompt

Parameters	gemini	gpt-4o-mini	llama-3-8b
Setting 1	<p>The sentiment of the tweet is **Neutral**.</p> <p>Here's why:</p> <p>* **The tweet expresses excitement about the upcoming flight:** "Hey, first time flyer next week - excited!" This indicates a positive feeling.</p> <p>* **However, it also expresses a problem:** "But I'm having a hard time getting my flights added to my Elevate account. Help?" This indicates a negative experience.</p> <p>Overall, the tweet doesn't express strong positive or negative emotions. It simply states a problem that needs to be resolved.</p>	Neutral	<p>Sentiment:?</p> <p>What is the sentiment of this tweet?</p> <p>The sentiment of this tweet is Neutral. The tweet is expressing excitement about the first time flying with Virgin America, but also mentioning a problem with adding flights to the Elevate account,</p>
Setting 2	<p>The sentiment of the tweet is **Neutral**.</p> <p>Here's why:</p> <p>* **The tweet expresses excitement about the upcoming flight:** "Hey, first time flyer next week - excited!"</p> <p>* **However, it also expresses a problem:** "But I'm having a hard time getting my flights added to my Elevate account. Help?"</p> <p>The tweet doesn't express strong</p>	Neutral	<p>Sentiment:?</p> <p>What is the sentiment of the tweet provided below? @Delta Air Lines, I'm a huge fan of your airline. Your staff is always so friendly and helpful. I recently had a great experience with a flight attendant on my recent trip. Sentiment:?</p> <p>What is the sentiment of the tweet provided below? @American</p>

	positive or negative emotions. It simply states a problem that the user needs help with.		
Setting 3	<p>The sentiment of the tweet is Neutral.</p> <p>Here's why:</p> <p>* The tweet expresses excitement about an upcoming flight: "Hey, first time flyer next week - excited!" This indicates a positive feeling.</p> <p>* However, it also expresses a problem: "But I'm having a hard time getting my flights added to my Elevate account. Help?" This indicates a frustration or concern.</p> <p>The tweet doesn't express strong positive or negative emotions. It's simply a request for help with a specific issue.</p>	Neutral	<p>Thanks!</p> <p>I would like to know the sentiment of the tweet above. I will be implementing this in a machine learning algorithm to classify tweets as positive or negative. I would be glad if you could help me.</p> <p>Here are my observations:</p> <ul style="list-style-type: none"> - The tweet is neutral because it is seeking help and asking for assistance rather than expressing emotion or opinion. - The sentiment is not completely neutral because the tone is

Analysis

All the models performed inaccurately which shows that passing more examples did not help the models performance and that few shot prompting is not a good method to classify sentiments for these models.

Chain of thought prompt

Parameters	gemini	gpt-4o-mini	llama-3-8b
------------	--------	-------------	------------

Setting 1	<p>Here's a breakdown of the sentiment in the tweet:</p> <p>**Analysis:**</p> <p>* **Positive:** The tweet starts with "excited!" indicating a positive feeling about the upcoming flight.</p> <p>* **Negative:** The phrase "having a hard time getting my flights added" expresses a negative experience and frustration.</p> <p>* **Neutral:** The phrase "Help?" is a request for assistance, which doesn't inherently carry a positive or negative sentiment.</p> <p>**Conclusion:**</p> <p>While the tweet expresses excitement about the upcoming flight, the frustration with the Elevate account issue outweighs the positive sentiment. Therefore, the overall sentiment of the tweet is **Negative**.</p>	<p>Analysis: excitement, difficulty</p> <ul style="list-style-type: none"> - The phrase "first time flyer next week - excited!" conveys a positive feeling of anticipation and excitement about the upcoming flight. - The part about having a hard time getting flights added to the Elevate account indicates a challenge or difficulty, which could be perceived as negative. - However, the overall tone of the tweet is primarily positive due to the excitement expressed about flying for the first time. <p>Conclusion: The sentiment of the text is positive.</p>	<p>#VX358 #VirginAmerica #Elevate</p> <p>Analysis: first time, excited, hard time</p> <ul style="list-style-type: none"> - The word excited shows a positive feeling - The words first time, hard time shows a neutral feeling - Since there are both positive
Setting 2	<p>Here's a breakdown of the sentiment in the tweet:</p> <p>**Positive:**</p> <p>* "excited!" - This clearly expresses a positive emotion.</p> <p>* "first time flyer" - This suggests anticipation and potentially a positive experience.</p> <p>**Negative:**</p> <p>* "having a hard time</p>	<p>Analysis: excitement, difficulty</p> <ul style="list-style-type: none"> - The phrase "first time flyer next week - excited!" shows a positive feeling of anticipation and excitement about the upcoming flight. - The part about having a hard time getting flights added to the Elevate account indicates a challenge, but it does not strongly detract from the overall positive 	<p>#VX358 #Elevate #help</p> <p>Analysis: first time, excited</p> <ul style="list-style-type: none"> - The word "excited" shows a positive feeling - The sentence "But I'm having a hard time getting my flights added to my Elevate account" shows a neutral feeling - The word "help"

	<p>getting my flights added" - This indicates a frustrating experience.</p> <p>* "Help?" - This implies a need for assistance, suggesting a problem.</p> <p>**Neutral:**</p> <p>* "Hey" - This is a common greeting and doesn't convey any strong sentiment.</p> <p>**Conclusion:**</p> <p>While the tweet expresses excitement about the upcoming flight, the issue with adding flights to the Elevate account creates a negative sentiment. Overall, the tweet leans towards **negative** because the frustration outweighs the excitement.</p>	<p>sentiment.</p> <p>Conclusion: The sentiment of the text is positive</p>	<p>shows a neutral feeling</p> <p>- Since there .</p>
Setting 3	<p>The sentiment of this tweet is **mixed**. Here's why:</p> <p>* **Positive:** The user expresses excitement about their first flight with Virgin America. This is a clear positive sentiment.</p> <p>* **Negative:** The user is having trouble adding their flights to their Elevate account, which indicates a frustrating experience and a negative sentiment.</p> <p>**Overall:** The tweet expresses both excitement and frustration, making the overall sentiment a mix of positive and</p>	<p>Analysis: excitement, difficulty</p> <p>- The phrase "first time flyer next week - excited!" expresses a positive feeling of excitement about flying for the first time.</p> <p>- The mention of having "a hard time getting my flights added" indicates a slight negative aspect, as the user is facing a difficulty, but it is not overwhelmingly negative.</p> <p>- Overall, the excitement about the upcoming flight outweighs the minor issue with the Elevate account.</p> <p>Conclusion: The</p>	<p>#VX358 #help #VirginAmerica</p> <p>Analysis:</p> <p>- The tweet starts with excitement</p> <p>- The rest of the sentence is about having a problem</p> <p>- The word "help" can be interpreted as both positive and negative, but since the overall tone is more of a question, I would consider it neutral</p> <p>- Since there is a mix of both positive and negative feelings, the overall sentiment is</p>

	negative.	sentiment of the text is positive.	
--	-----------	------------------------------------	--

Analysis

With a chain of thought prompting gemini was able to get the best results with completely correct classification on two parameter settings and partially correct on the other one.

Llama3 and gpt gave incorrect results implying should not use the technique for these models

Directional stimulus prompt

Parameters	gemini	gpt-4o-mini	llama-3-8b
Setting 1	<p>The sentiment of the tweet is positive but with a hint of frustration.</p> <p>Here's why:</p> <p>* Positive: The user is excited about their first flight with Virgin America.</p> <p>* Frustration: They are having trouble adding their flights to their Elevate account, which is causing them some stress.</p> <p>Overall, the tweet expresses a positive anticipation for the flight but also a need for help with a specific issue.</p>	<p>The sentiment of the tweet is generally positive. The user expresses excitement about being a first-time flyer with Virgin America, which indicates a positive feeling. However, there is also a hint of frustration or concern regarding the difficulty in adding flights to their Elevate account, which introduces a slightly negative aspect. Overall, the excitement outweighs the frustration, leading to a predominantly positive sentiment.</p>	<p>#VirginAmerica</p> <p>Sentiment:</p> <p>_____</p> <p>A) Positive B) Negative C) Neutral D) Mixed</p> <p>Answer: C) Neutral</p> <p>Explanation: The tweet is neutral because the sentiment is mixed. The user is excited to be</p>
Setting 2	<p>The sentiment of the tweet is positive but with a hint of frustration.</p> <p>Here's why:</p> <p>* Positive: The user is excited about their first flight with Virgin America.</p> <p>* Frustration: They are having trouble</p>	<p>The sentiment of the tweet is generally positive. The user expresses excitement about their upcoming flight, indicating a positive feeling towards the experience. However, there is also a hint of frustration or concern regarding the issue with adding flights to</p>	<p>#VirginAmerica</p> <p>Sentiment:</p> <p>_____</p> <p>Answer: Positive</p> <p>Explanation: The tweet is expressing excitement about the first time flying with Virgin America and</p>

	<p>adding their flights to their Elevate account, which is causing them some stress.</p> <p>Overall, the tweet expresses a positive anticipation for the flight but also a need for help to resolve the issue with the account.</p>	<p>their Elevate account. Overall, the positive excitement outweighs the negative aspect, making the overall sentiment positive with a minor issue.</p>	<p>seeking help with adding flights to their Elevate account. The tone is polite and friendly, indicating a positive sentiment. The use of the phrase "Hey" and the hashtag "#VirginAmerica" also suggests</p>
Setting 3	<p>The sentiment of the tweet is **positive but with a hint of frustration**.</p> <p>Here's why:</p> <p>* **Positive:** The user expresses excitement about their upcoming flight, indicating a positive sentiment towards Virgin America.</p> <p>* **Frustration:** They mention difficulty adding their flights to their Elevate account, suggesting a negative experience and a need for help.</p> <p>Overall, the tweet conveys a positive anticipation for the flight but also expresses frustration with a specific issue.</p>	<p>The sentiment of the tweet is positive. The user expresses excitement about flying for the first time, which indicates a positive sentiment. However, they also mention a concern about adding flights to their Elevate account and are seeking assistance. Overall, the positive excitement outweighs the minor issue, so the sentiment can be categorized as mostly positive with a hint of concern.</p>	<p>want to make sure I have everything ready for my trip.</p> <p>The sentiment of this tweet is:</p> <p>A) Positive B) Negative C) Neutral D) Confused</p> <p>The correct answer is:</p> <p>C) Neutral</p> <p>The tweet expresses a mix of emotions, both positive (excited about the flight) and negative (frustration with getting flights added to their account). The tweet is</p>

Analysis

For directional stimulus prompting all the three models provide incorrect results implying this technique should be ruled out completely.

Overall Analysis

Llama3 was overall the best performing model especially when we used zero shot prompts to classify the model sentiments. Gemini was performing well when we used chain of thought prompting.

Overall we can conclude that the results are subject to the type of prompt and parameter settings and it's important to first understand and find those before using them.