

Text Summarization using Transformer-based Language Models

Ishrat Ahmed

University of Pittsburgh

Vail Systems

Text Summarization

Process of producing a shorter version of long texts.

Application: medical, news, law, research, etc.

Different approaches e.g., Text Rank, training deep learning models, using pre-trained large language models, etc.

Difficult because need to generate texts unlike other NLP tasks where the output is a single label or value (classification/regression).



Types of Text Summarization



Extractive – Words and phrases are directly extracted from the text.



Abstractive - Words and phrases are generated semantically consistent, ensuring the key information of the original text is maintained.

The sun was setting on a beautiful autumn day in October. Ricky was lying lazily in the garden when a quick brown fox jumped over him. Ricky is our beloved family dog...

EXTRACTIVE

A quick brown fox jumped over him.

ABSTRACTIVE

The quick brown fox jumped over the lazy dog.

Fig: Extractive summary vs. abstractive summary [1]

Summarization of Call Transcripts

Extracted summary and
intent from call transcripts.

Used the latest GPT-3 model.

Call ID	Intent	Keywords	Entities	Sentiment
e6ab570d-28e0-467e-b3af-2df390adadab	Help the customer with their purchase.	customer, purchase		Neutral
e9a3cafe-b799-4431-b4ec-5d26b15c68ec	Place a take-out order.	take, out, order		Neutral
dff52267-e2a3-4b5e-ad9b-f9d8cefb0b59	The caller is asking for the customer's phone number so they can follow up with them about their account.	caller, customer, phone, number, account		Neutral
e8e14a3f-1f34-4769-91a7-d0f2a2ed74bb	The call seems to be about a possible business opportunity, though it is difficult to say for certain.	call, business, opportunity		Neutral
e62d4cf6-8f40-4fed-a45c-5075041070e7	Ask a question about a purchase on Amazon.	question, purchase		Positive
e6deafb4-29ef-48b8-855f-815531296609	Make a payment on an [REDACTED] account.	payment, Furniture, account		Neutral
e4a24eb1-c953-4b67-b6ad-a591241c92b6	Take a fast food order.	food, order		Neutral
e4cf37a1-77cf-4471-80f9-ce196ae34f71	The call is to collect credit card information for a large order that is being delivered today.	call, credit, card, information, order, today		Neutral

Call ID	e4a24eb1-c953-4b67-b6ad-a591241c92b6
Intent	Take a fast food order.
Summary	The caller is ordering two kids meals, one with chicken and one with steak, and requests no vegetables. The total cost will be \$11.52 and the food will be ready for pickup in
Intent Keywords (GPT-3)	<ul style="list-style-type: none">• fast food• take a order• intent
Summary Keywords (GPT-3)	<ul style="list-style-type: none">• 15 minutes fast food• chicken• steak
Transcript	<p>For your convenience if you would prefer to prepay please have your credit card ready at the end of the call . Maybe you can't play at [REDACTED] I have the last four digits of your phone number to be [REDACTED]. OK . Welcome back [REDACTED] Can I . Um what can I . Can I get a zip code . [REDACTED] what can I get for you today . Can I get a kids low mane with chicken . No vegetables . And then a kids low Maine with steak no vegetables . Okay . Into Santa tie Donna or fudge brownie today . OK so I have one order of the kids old man with chicken no vegetables and one order the kids bowl made with steak . No vegetables that bring your total to eleven fifty two and that'll be ready for pickup and about twelve minutes . Okay . Thank you . Thank you for going away and have a good day .</p>

Summary produced by GPT-3

Challenges with GPT-3



Cost: (\$0.02/1k tokens*;
~\$0.01/query)



Privacy and stability:
external API call



Domain adaptation:
difficult and expensive

* <https://help.openai.com/en/articles/6485334-openai-api-pricing-update-faq>

Growth of Language Models

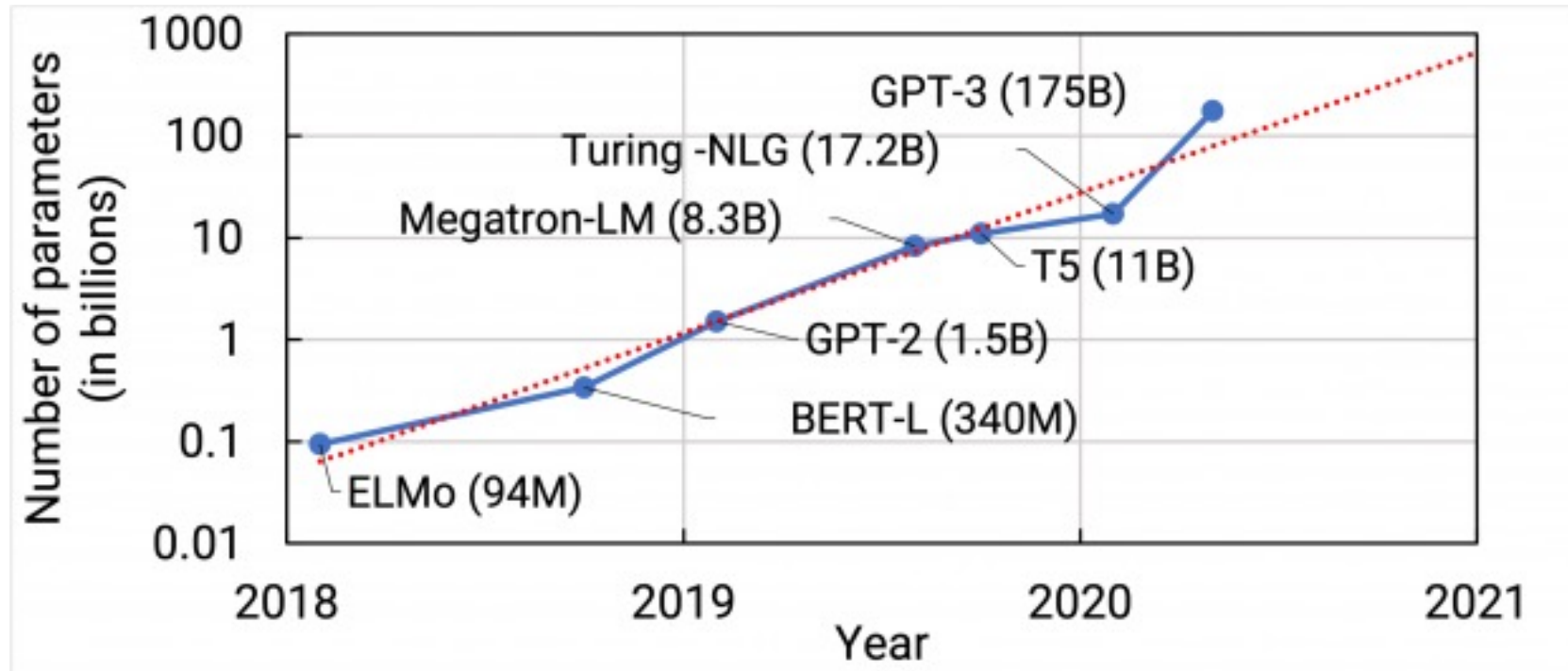


Fig: Trend of state-of-the-art NLP model sizes with time [2]

[2] <https://developer.nvidia.com/blog/scaling-language-model-training-to-a-trillion-parameters-using-megatron/>

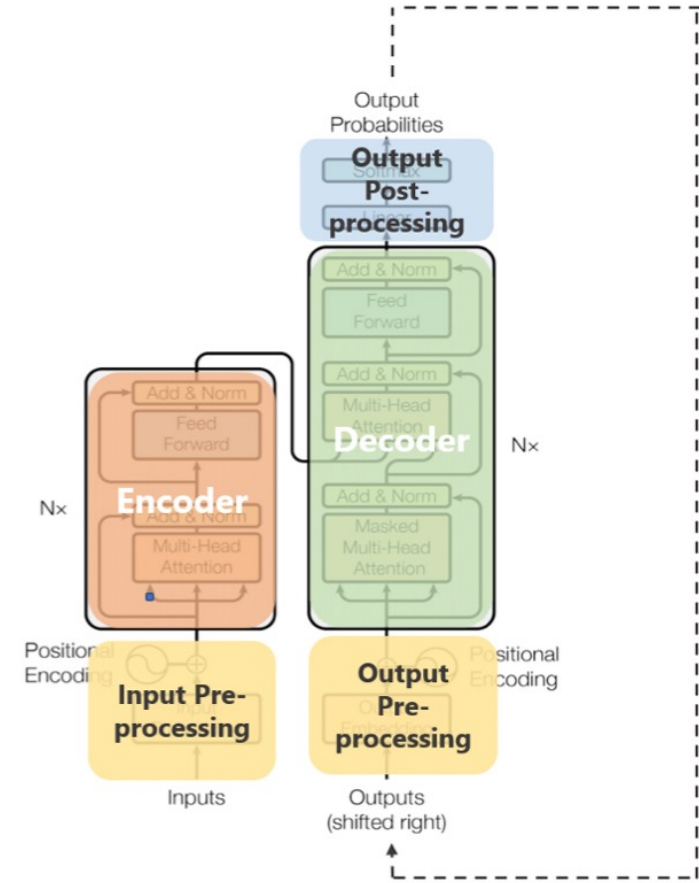
Transformer-based Language Models

A type of neural network with an **encoder-decoder** architecture.

Language models can use the transformer architecture in part (e.g., BERT encoder only, GPT decoder only) or as a whole (e.g., BART, T5)

Common pre-trained models for summarization BART, T5, Pegasus

- BART-large-Xsum, 400M parameters
- T5-small, 60M parameters
- Pegasus-Xsum, 568M parameters



An abstracted version of the Transformer [3]

Dataset

Training Dataset:

- 5389 call transcripts between customers and agents
- source: financial service and food service

Evaluation Dataset:

- 63 call transcripts between customers and agents

Ground Truth:

- Output from GPT-3

No PII data, all transcripts are numerically redacted at the source

Call ID	e62d4cf6-8f40-4fed-a45c-5075041070e7
Intent	Ask a question about a purchase on Amazon.
Summary	A customer is asking a question about a purchase on Amazon from November 26, 2019.
Transcript	Thank you for calling [REDACTED]. You're speaking to [REDACTED]. How are you doing today. Thank you. How are you. I'm doing great ma'am. Thank you so much for asking me. I know your first and last name please. [REDACTED]. Perfect. Thank you for verifying your name ma'am. How may I help you today. I just have a question on a payment or a purchase on OK November. 26 2019 for sixteen dollars and 41 cents. It says Let [REDACTED] me. did. I just wanted to know exactly what it is. The rise of magic says what it says. OK so on [REDACTED] it was for a book. Oh you said the book meant that I meant. I've just unmasking what the purchase was for on [REDACTED]. OK. That's a book man. The Rise of [REDACTED] up to one OK thank you very much. That was it. man. All right. Thank you so much for calling [REDACTED] and got your Amazon dot com stalker. Please take it and have a great day ma'am. Happy New Year. Happy New Year.

Call ID	e4a24eb1-c953-4b67-b6ad-a591241c92b6
Intent	Take a fast food order.
Summary	The caller is ordering two kids meals, one with chicken and one with steak, and requests no vegetables. The total cost will be \$11.52 and the food will be ready for pickup in
Transcript	For your convenience if you would prefer to prepay please have your credit card ready at the end of the call. Maybe you can't play at [REDACTED]. I have the last four digits of your phone number to be [REDACTED]. OK. Welcome back [REDACTED]. Can I. Um what can I. Can I get a zip code. [REDACTED]. what can I get for you today. Can I get a kids low mane with chicken. No vegetables. And then a kids low Maine with steak no vegetables. Okay. Into Santa tie Donna or fudge brownie today. OK so I have one order of the kids old man with chicken no vegetables and one order the kids bowl made with steak. No vegetables that bring your total to eleven fifty two and that'll be ready for pickup and about twelve minutes. Okay. Thank you. Thank you for going away and have a good day.

Pre-Trained Model Output

call ID	e8b7858f-a3e4-4c32-a771-5d5d8e153343.txt
Transcript (truncated)	Hi . Can I order some food for pickup please ... welcome back . What can I get for you I need one penny Rosa with no tomatoes . better cheese or Parmesan 5 7 1 parmesan and then add tofu and then to regular mac and cheese . anything else ma'am . So, take about nine minutes or pick up here in ... Street . Your total is 20 dollars 32 cents . Okay . Thank you . But my .
BART (Xsum) output	Carrie is back on the BBC and this week she's talking about mac and cheese and how it's been made <i>in the kitchen of her new flat in London</i> .
Pegasus (Xsum) output	This is a transcript of a phone call from a woman who wants to order food from a local restaurant.
T5-small output	I need one penny Rosa with no tomatoes or Parmesan 5 7 1 parmesan and then add tofu and dann to regular mac and cheese - anything else ma'am!
GPT-3 output	A woman orders food for pickup and is told her total.

Fine-Tuning Model

Pre-trained models

- allow us to use state-of-the-art models without having to train one from scratch
- reduces computation costs

However, pre-trained models does not generalize well → fine-tune models

This requires multiple iterations of training to identify the best architecture and hyperparameters

Time to fine-tune models on A100:

- Fine-tuned T5 model took 24.56 mins
- Fine-tuned BART took 10.22 min

Fine-Tuned Model Output

call ID	e4a24eb1-c953-4b67-b6ad-a591241c92b6
Transcript	For your convenience if you would prefer to prepay please have your credit card ready at the end of the call . Maybe you can't play at ... I have the last four digits of your phone number to be OK . Welcome back Can I . Um what can I . Can I get a zip code . and what can I get for you today . Can I get a kids low mane with chicken . No vegetables . And then a kids low Maine with steak no vegetables . Okay . Into Santa tie Donna or fudge brownie today . OK so I have one order of the kids old man with chicken no vegetables and one order the kids bowl made with steak . No vegetables that bring your total to eleven fifty two and that'll be ready for pickup and about twelve minutes . Okay . Thank you . Thank you for going away and have a good day .
Pre-trained BART (xsum) output	For the first time in more than a decade we have an online ordering option <i>for you to play the game Pokemon Go on your mobile phone or tablet.</i>
Fine-tune BART (xsum) output	The message is asking if the person would like to prepay for their order, and if so, to have their credit card ready. The total cost of the order is \$11.52.
Pre-trained T5 output	. I have one order of the kids old man with chicken no vegetables and then a kids low Maine with steak no veggies, and another order the children bowl made from steak - that'll be
Fine-tune T5 output	If you would like to prepay for prepay, please have your credit card ready at the end of the call.
GPT-3 output	The caller is ordering two kids meals, one with chicken and one with steak, and requests no vegetables. The total cost will be \$11.52 and the food will be ready for pickup in

Some not-so-great examples...

The person is not making much sense. (GPT-3)

And I think every woman is a monster . I wasn't I mean just remember . I started to it either Maybe I'm perfect I don't want to put on my . (fine-tuned BART)

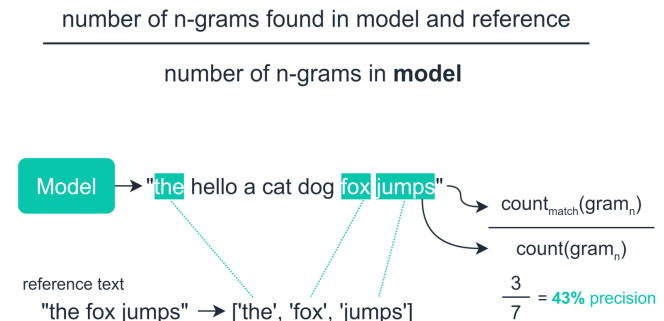
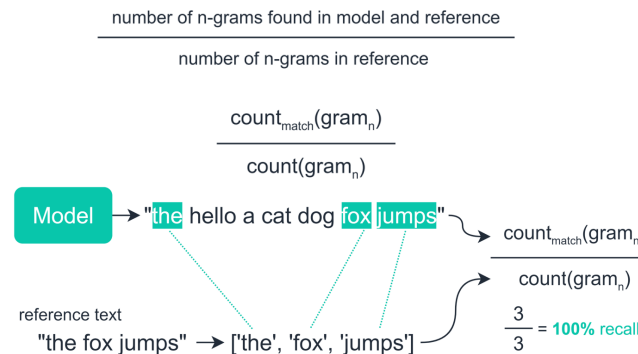
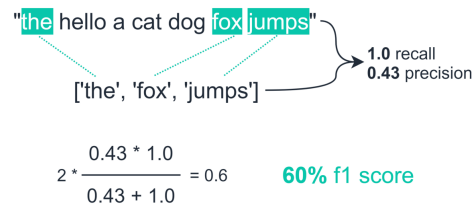
This customer is playing on their phone before Tl;dr of a call. (fine-tuned T5)

Evaluation (1)

Quantitative Evaluation:

ROUGE score (Recall-Oriented Understudy for Gisting Evaluation), e.g., ROUGE-N (n=n-grams), ROUGE-L (L=Longest common subsequence)

$$2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$



A ROUGE score close to zero indicates poor similarity between candidate and references .

A ROUGE score close to one indicates strong similarity between candidate and references .

Evaluation (2)

model	Rouge-1	Rogue-2	Rouge-L
Pre-trained BART	24.5	5.7	16.6
Fine-tuned BART	37.8	16.8	31.0
Pre-trained T5-small	21.9	5.1	15.2
Fine-tuned T5-small	30.6	10.3	24.1

- 37.8% unigrams match between the reference and candidate summary.
- 16.8% bigrams match between the reference and candidate summary.
- 31% LCS match between the reference and candidate summary, i.e., longer shared sequence indicates more similarity between the two sequences.

Evaluation (3)

Qualitative Evaluation

- 7 evaluators scored the model-generated summaries in terms of main content of the transcript.
- Each of them scored 20 model-generated summaries.
- The scores ranged from 1-10 with 10 being highest in correctness.

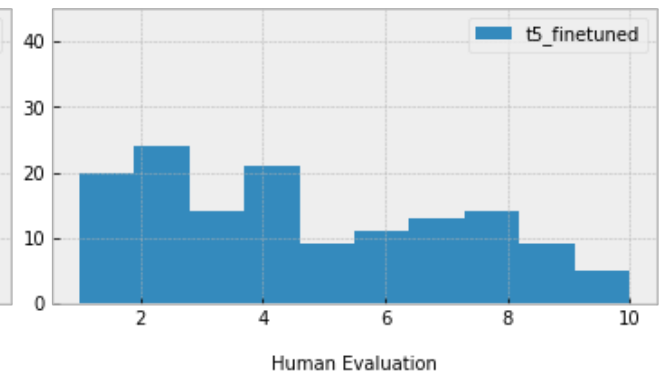
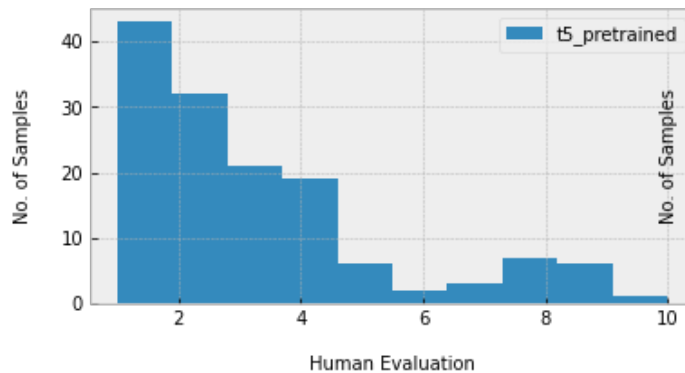
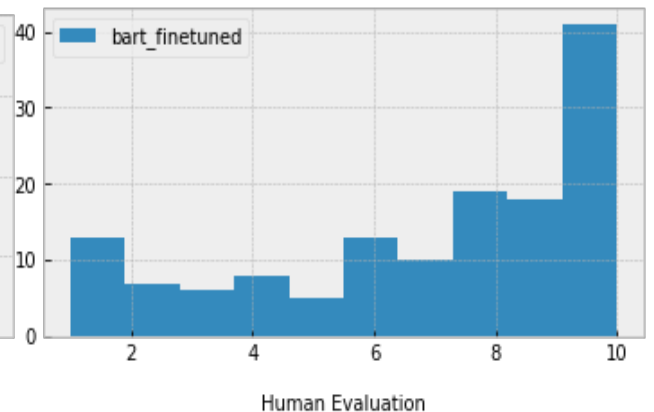
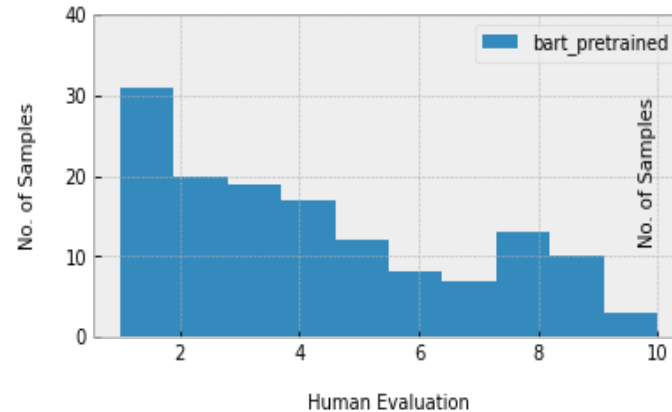
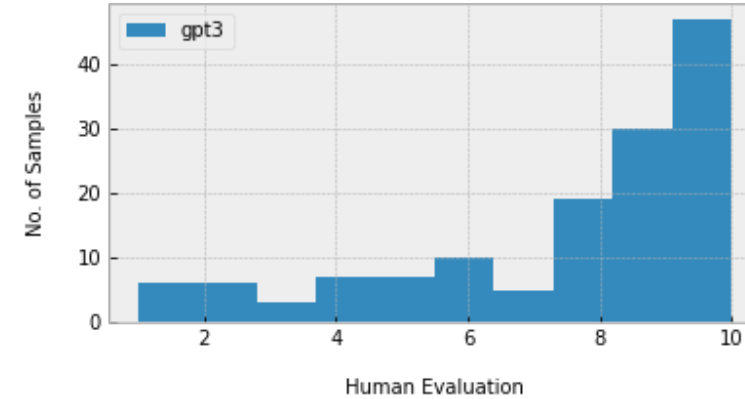
CallID	e62d4cf6-8f40-4fed-a45c-5075041070e7
Summary A	A customer is asking a question about a purchase on Amazon from November 26, 2019.
Summary B	A woman has called ██████ to ask an Amazon customer service representative to help her with a purchase she made on the site last month, only to discover she had been stalked by a stalker.
Summary C	This person is asking for a payment or a purchase on OK November 26 to make it through the Amazon dot com website. It is a customer service representative and customer service representative.
Summary D	for calling ██████ and got your Amazon dot com stalker. Please take it and have a great day ma'am - Happy New Year!
Summary E	The caller is asking about a purchase they made on November 26, and the customer service representative says it was for a book.
Transcript	Thank you for calling ██████. You're speaking to ██████. How are you doing today. Thank you. How are you. I'm doing great ma'am. Thank you so much for asking me. I know your first and last name please. ██████. Perfect. Thank you for verifying your name ma'am. How may I help you today. I just have a question on a payment or a purchase on OK November. 26 2019 for sixteen dollars and 41 cents. It says Let Amazon me. did. I just wanted to know exactly what it is. The rise of magic says what it says. OK so on November 26 it was for a book. Oh you said the book meant that I I meant. I've just unmasking what the purchase was for on November 26. OK. That's a book man. The Rise of Magic's Chronicle chronicles up to one OK thank you very much. That was it. man. All right. Thank you so much for calling ██████ and got your Amazon dot com stalker. Please take it and have a great day ma'am. Happy New Year. Happy New Year.

Qualitative Evaluation Results

GPT-3 has over majority samples with a score 8-10.

Bart fine-tuned behave similarly as GPT-3.

Both pre-trained model has more samples with lower score.

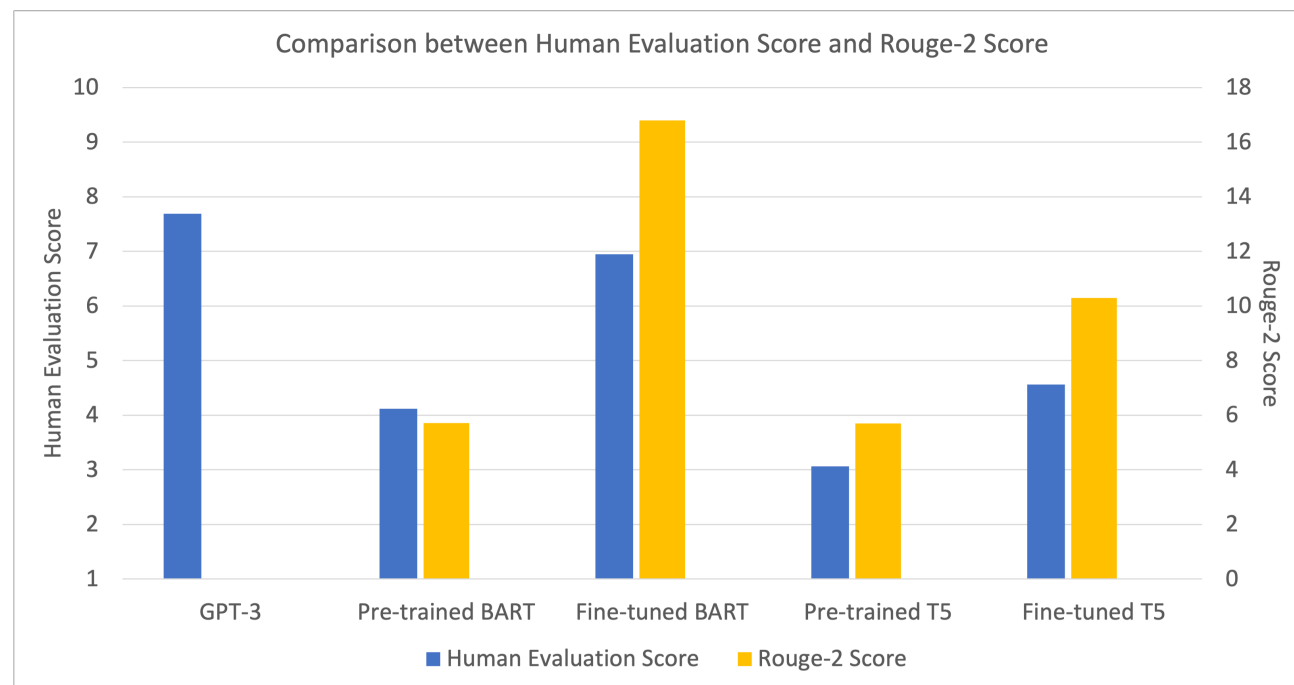


Rouge Score vs. Qualitative Evaluation

Calculated the mean score of each model generated summaries.

Similar trend in both rouge scores and mean human scores.

Fine-tune BART generated summaries are closer to GPT-3 generated summaries.



Models	GPT-3	Pre-trained BART	Fine-tuned BART	Pre-trained T5-small	Fine-tuned T5-small
Human Evaluation Scores	7.69	4.12	6.95	3.06	4.56

Conclusion

Generated summaries using open-source language models and compared with the summaries generated using GPT-3.

Fine-tune BART seemed to generate summaries closer to GPT-3.

Conducted human evaluation to validate our experimental results.

Since open-source models generated decent summaries, we can use to process voicemail call transcripts.