Improved Fine-Tuned Reinforcement Learning From Human Feedback Using Prompting Methods for News Summarization

Sini Raj Pulari[1], Maramreddy Umadevi[1], Shriram K. Vasudevan[2]*

[1] Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research, Guntur - 522213 (India)
[2] Division of Developer Platform and Evangelism, Software and Advanced Technology Group, Intel India Pvt. Ltd., Bengaluru - 560103 (India)

* Corresponding author: shriram.kris.vasudevan@intel.com

Received 2 August 2024 | Accepted 25 October 2024 | Published 3 February 2025

Abstract

ChatGPT uses a generative pretrained transformer neural network model, which is under the larger umbrella of generative models. One major boom after ChatGPT is the advent of prompt engineering, which is the most critical part of ChatGPT that utilizes Large Language Models (LLM) and helps ChatGPT provide the desired outputs based on the style and tone of interactions carried out with it. Reinforcement learning from human feedback (RLHF) was used as the major aspect for fine-tuning LLM-based models. This work proposes a human selection strategy that is incorporated in the RLHF process to prevent undesirable consequences of the rightful choice of human reviewers for feedback. H-Rouge is a new metric proposed for humanized AI systems. A detailed evaluation of State-of-the-art summarization algorithms and prompt-based methods have been provided as part of the article. The proposed methods have introduced a strategy for human selection of RLHF models which employs multi-objective optimization to balance various goals encountered during the process with H-Rouge. This article will help nuance readers conduct research in the field of text summarization to start with prompt engineering in the summarization field, and future work will help them proceed in the right direction of research.

Keywords

Abstractive Summarization, Extractive Summarization, Natural Language Processing, News Summarization, Prompt Engineering, Reinforcement Learning From Human Feedback (RLHF).

DOI: 10.9781/ijimai.2025.02.001

I. Introduction

SUMMARIZATION Generative models have become an indispensable part of our lives, since the inception of ChatGPT. Chatbot technology has advanced significantly since the launch of ChatGPT. Natural language processing tasks, such as text summarization, question answering, content selection, and query optimization, have all gained a lot of interest and have attracted many researchers. Research and student communities largely depend on ChatGPT to find quick answers to their questions. Text summarization in Natural Language Processing is a vast area of research that has been conducted for a while. Text summarization is the process of generating summaries from enormous amounts of a single document or multi-document or from very large data sources. The main challenge is the generation of an accurate and concise summary as the amount of online data increases exorbitantly. ChatGPT can process and provide summarizations to document text that is fed as input. It is very important to understand the underlying technologies used in ChatGPT and how well these technologies could be integrated in applied research in text summarization [1].

News article summarization is a subset of text summarization problems in Natural Language Processing. People become busy, and many do not have time to read detailed news. A study of our initial part proved that almost 70% of important news is not even noticed by people. They even noticed that they spent less than one minute understanding the crux of the news content.

This makes news article summarization a very important and urgent need in the fast-moving world. In addition to this, the personalization of data happens in such a way that the people are less likely to get the diverse news around the world, which could be important for them [2]. Prompt engineering is the vital part of ChatGPT where the prompts are the instructions fed into the large language models (LLMs) to give desired outputs in the way asked for. Prompts had to be crafted in

Please cite this article as: S. R. Pulari, M. Umadevi, S. K. Vasudevan. Improved Fine-Tuned Reinforcement Learning From Human Feedback Using Prompting Methods for News Summarization, International Journal of Interactive Multimedia and Artificial Intelligence, vol. 9, no. 2, pp. 59-67, 2025, http://dx.doi.org/10.9781/ijimai.2025.02.001

an effective way to get accurate and meaningful results from large language models like Generative pretrained transformer (GPT) models used in ChatGPT. Prompts must be specific and clear. Prompts need to be continuously experimented based on the context and application that is used in. Prompts had to be provided with contextual information for it to generate meaningful and more relevant results. Fine tuning needs to be performed by using the Reinforcement learning from human feedback for continuous improvement in the results obtained till it produces desired results by the LLMs.

The major contributions of this article are how to incorporate prompt engineering concepts in research on news article summarization. There are various transformer-based language models, such as BERT, PEGASUS, BART, T5, and BIGBIRD, that are available and have already been tested. This study expands the learning of large language models in the field of text summarization. Evaluation and comparison of various techniques against prompt engineering-based techniques using Rouge metric will be covered in the results section.

This article will have a positive effect on the academics and research professionals who are working in the field of text summarization and opening a new window with a ray of knowledge to inculcate in their works. The insights from the results will allow more budding researchers with new open problems and in turn to contribute to the society by bringing more AI powered solutions in the field of text summarization and natural language processing.

The article explains related work in the next section, which will provide the knowledge required for enhancing the basics. This is followed by evaluation results and a comparison with contemporary methods using these metrics. Finally, the conclusion, future scope, and limitations of this study are presented.

II. Theoretical Background and Related Work

With the dawn of ChatGPT, prompt engineering started receiving more attention in the field of natural language processing, mainly chatbots. The article titled "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing" by Pengfei Liu et al. has covered all the basic concepts needed to know while prompting [3]. This article is very expansive, and major approaches, techniques, and comparisons claiming the positives for a paradigm shift to prompt-based learning are covered in detail. A common standard implementation framework for prompt engineering was not established in this study. Nevertheless, this article acts as the base article for most of the works in the field of prompt engineering. Several advancements have been made since the publication of this article.

Ding, N., Hu, S. et al. proposed a unified framework called open prompt through the article "OpenPrompt: An open-source framework for prompt-learning" [4]. In this article authors have tried to bring a single unified framework with pre-defined blocks like prompt model, prompt dataset, and prompt trainer. All the modules are not necessary, and they are dependent on the applications where the OpenPrompt is used. The integration of prompt-based techniques has not been completed as part of the project and is in the progressing phase.

Text summarization has been an extensive research area for more than a decade. Abstractive and extractive summarization methods using various language models are on the rise. When extractive summarization only provides summaries extracting content from the textual part, abstractive summarization generates meaningful summaries by considering the contextual meaning of the content and text. Hence, many prefer abstractive summarization, as it includes more in-depth and meaningful contextual summarization. Many hybrid methods that incorporate both extractive and abstractive summarization methods have also gained attention nowadays. These hybrid methods claim that they include major sentences from the textual content along with the context, thereby including the advantages and disadvantages of both methods. In the article "A comprehensive review of automatic text summarization techniques: method, data, evaluation and coding" by Cajueiro et al. has provided a thorough explanation of automatic text summarization methods (ATS) [5]. In this article, prompt engineering for text summarization has not been discussed.

Another major challenge in text summarization is that there is no perfect dataset that includes human summaries as references. In many of the datasets, there are summaries, but evaluation metrics such as Rouge require the generated summary to be compared against human summaries. This is a major gap in the research on text summarization. The closest solution or match is found in literature survey by the article "NEWTS: a corpus for news topic-focused summarization" by Bahrainian, S. A. et al. [6]. This study focuses on topic-based summarization. For an article, they find the relevant topics and the summarizations are given based on the topics identified. Two summaries for each article based on the most relevant topics were given based on experimental prompting and text summarization methods. This study has used only basic prompting methods and could be extended in an elaborate manner for significant applications, which is a future work and limitation.

Prompt engineering could be useful in text summarizations, and there are many methods available to generate prompts and use LLMs to utilize them in the required context. Many articles have discussed prompt templates for many applications [7][8][9]. Conversely, it is not easy to maintain as many templates for specific applications and will be a tedious task. Therefore, the best method is automatic prompt generation based on context. This is highlighted in the article titled "Large language models are human-level prompt engineers" by Zhou, Y., Muresanu et al. thereby proposing a solution of Automatic Prompt Engineer (APE) for automatic instruction generation and selection by formulating it as an optimization problem and highlight the major prompting methods like zero shot learning, few shot learning, chain of thought prompting methods [10].

In this article, we focus on news summarization using prompting methods, which is discussed in the article "News summarization and evaluation in the era of gpt-3" by Goyal, T., Li et al. In this study, the authors compared GPT3 results with other fine-tuned models [11]. The limitation mentioned is the use of reinforcement learning from the human feedback method, which does not actually cover the impact on news summarizations in this article. The contributions made through this article are summarized below.

a) Utilization of various prompting methods compared to other existing language models.
b) The usage of reinforcement learning from human feedback in the news summarization research area.
c) We have proposed a human selection strategy that could improve the reinforcement learning from human feedback method.
d) A Multi Objective pareto front optimization is suggested for the tradeoff between human feed-backs and the reward system.
e) This work has proposed a H-Rouge (RH) metric for considering the human feedback scores in the evaluation of RLHF process.

A detailed comparison of results using the Rouge metric for prompting methods and a detailed understanding and insights from state-of-the-art (SOTA) algorithms in summarization is carefully performed.

III. Problem Formulation and Methodology

This section has three major highlights. The first part talks about the Data collection part where we propose Prompt engineering is a methodology used for fine-tuning and optimizing natural language processing models by introducing the concept of prompts or instructions that are carefully crafted for the desired application, as shown in Fig.1. and Fig. 2.

Fig. 1. Example of how a specific task prompt is added along the input tasks.

Some of the important algorithms in prompt engineering include fine-tuning, masked language modelling, gradient-based optimization, and reinforcement-based learning [12]. Prompting techniques include zero-shot learning, few-shot learning, and a chain of thought learning.

In zero-shot learning, the language model may not have prior training carried out on the specific data that we have provided. The model outputs a response based on a generic understanding of the language. In few-shot learning, a model is provided with more examples to better understand context.

This helps the model present the response in a more appropriate manner. Chain of thought is the most common technique used in chatbots, where there are multilevel conversations as input and response chains. The model maintains the context from previous responses and coherence.

Reinforcement learning from human feedback (RLHF) is a technique used in fine-tuning to improve results. This is done using a reward-based model, where the human review shall be carried out for the model responses to gain appropriate feedback. In turn, the model is fine-tuned to attain better rewards, which will improve the model performance in the desired manner, as shown in Fig. 3.

Fig. 3. Reinforcement learning from Human Feedback (LLM using RLHF).

RLHF helps the language model by introducing human reviews, thereby improving the model summarization accuracy. RLHF mainly uses large dataset and apply supervised learning models for initial training, secondly reinforcement-based reward models are developed based on the comments on the outputs, as the last step the models are fine tuned to get better rewards producing precise and contextually relevant summaries. Human feedback could be given in the form of thumps up or down, scaled results with smileys or incentive-based ones. News summarization could be done in an accurate manner by prompt engineering methods followed by fine-tuning using RLHF techniques [13].

The main limitation or challenge here is human feedback [14]. Even though these methods are highly impactful and promising, human feedback and reward systems are a tradeoff [15]. Human feedback is affected in many ways because it is expensive and time-consuming. As real humans are involved in giving summaries, there is no proper selection process based on the expertise of humans; hence, the results could be biased. The reward model works based on the feedback and ranking of documents by humans. If human feedback is not proper, it affects the entire system of the process, resulting in inaccurate results. Hence, our system proposes two major methodologies to improve the entire human feedback process [16].

A. Proposed Method for Human Selection Strategy (HSS) for Reviews

A graph-based model was proposed for selecting the most appropriate human reviewers. Once the initial supervised model produces an output response, it can be given to the right human reviewers. K-means clustering was applied to the human database. Each human database consists of reviewer data with their interests prioritized.

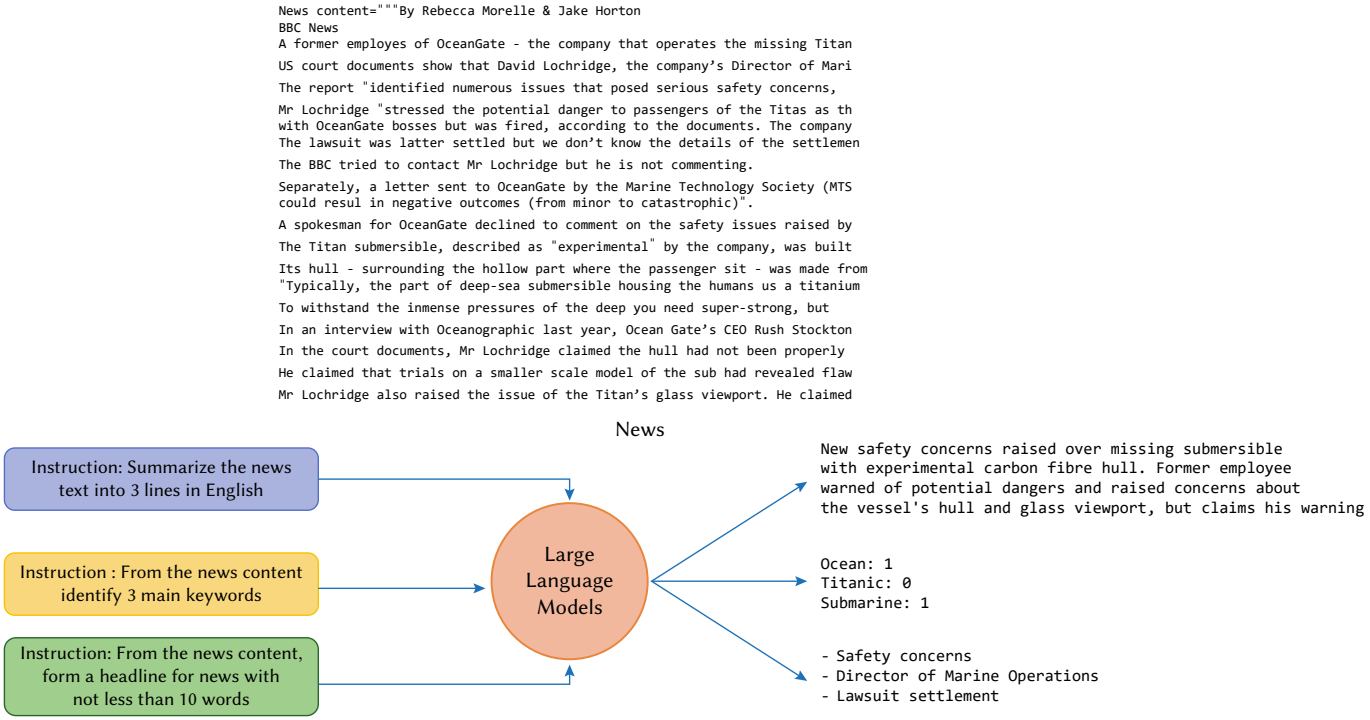


Fig. 2. Significance of the usage of Prompts in a Large Language Model.

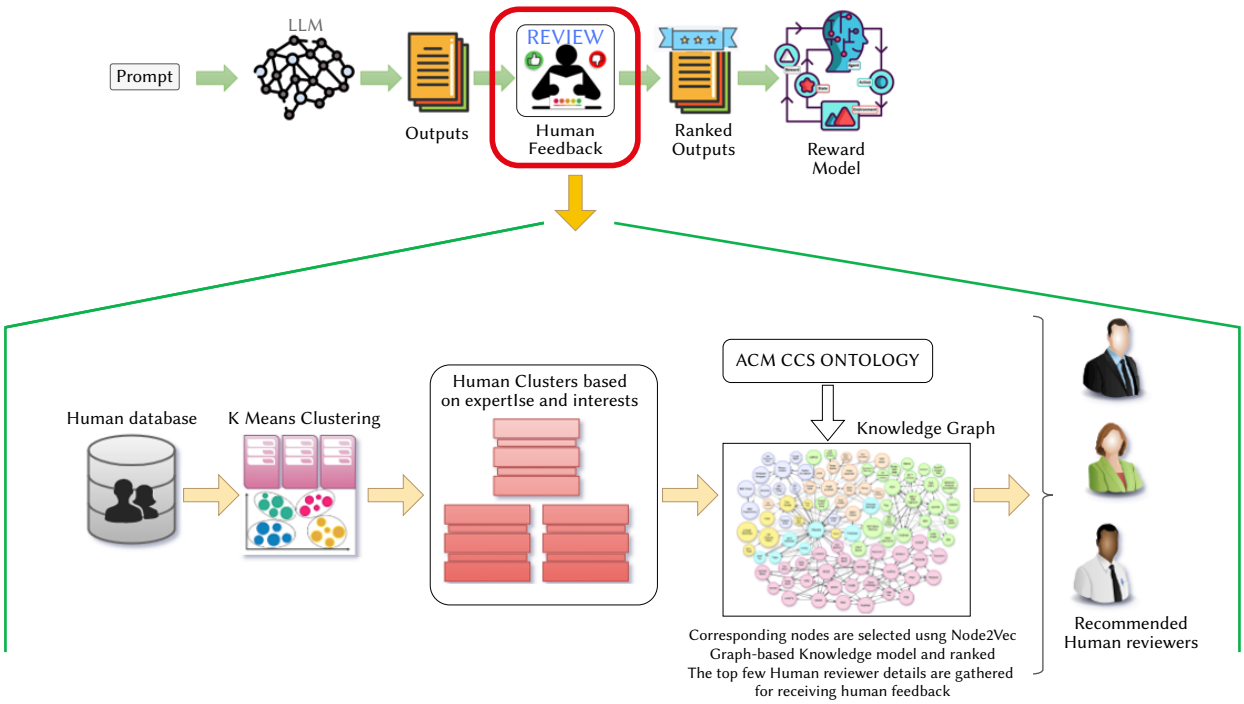


Fig. 4. How the appropriate human reviewers are selected.

K-means topic-based clustering allows human reviewer data to be clustered based on topics or interests by applying a cosine similarity. These clusters were mapped using node2vec based on the ACM CCS ontology, which contains the ontology of computing concepts. Hence, the complete data are made as nodes, and the reviewer information is placed based on the Earth mover's distance. Some nodes have common subsumers which are parent nodes that encompass two or more specific child nodes. Within a hierarchical arrangement, this broader concept would serve as the "parent" element to the more specialized "child" concepts. Each node will have the reviewer ID, interests, cluster-ID information, and the associated weight. The corresponding nodes are selected and ranked according to the top human reviewer information, as shown in Fig. 4.

Semantic similarity is the easiest way to find the similarity between the nodes and could be identified as follows. Consider two concepts C_i and C_j.

$$sim_{graph}(C_i, C_j) = \frac{1}{1+length(C_i,C_j)*K^{IC(C_{LCS})}} \quad (1)$$

Even concept pairs with the same path length can have different least common subsumer (LCS), which contribute to different semantic similarity scores [17]. LCS is the nearest common generalization in a hierarchical structure that is shared by multiple concepts, representing their most precise mutual ancestor. The information content of a concept helps to determine its relevance. This uses a factor K, which uses values [0, 1] that indicate the contribution of IC to the path length, as shown in (1). It captures the relevance of content to the node in the graph. This provides a better similarity for the searched articles.

This method helps comprehend the entire process with simplicity. There is always a trade-off between adding more humans to the system and AI capabilities. The balance must be carefully chosen without any bias, and this method will contribute to the same point. The introduction of this human selection strategy will aid in preventing undesirable consequences and unprecedented scenarios in advance by regulating harmful feedback. In addition, if a wrong reviewer is selected, it will add additional cost to the RLHF system, which could be avoided by using this strategy for the rightful selection of human reviewers based on the keywords, topics, or interests, thus making the whole system personalized as well.

B. Multiobjective Optimization Problem

Human feedback and reward models work hand in hand as the RLHF process progresses. If human feedback is appropriate, the reward model quickly converges and moves to a stopping point. However, if the human feedback is not carefully given or has a series of feedback to be considered, the process also requires a more promising process to find a full stop. Typically, Kullback-Leibler divergence (KL divergence) is used [18]. This challenge could be tackled as an optimization problem which gives Pareto optimal solutions, i.e., set of solutions that define the best tradeoff between competing objectives, and the basic multi objective optimization problem is given by (2)

$$\min_{x \in X} \left(f_1(x), f_2(x), \ldots, f_k(x)\right) \quad (2)$$

where the integer $k \geq 2$ is the number of objectives and the set X is the feasible set of decision vectors as shown in Fig. 5.

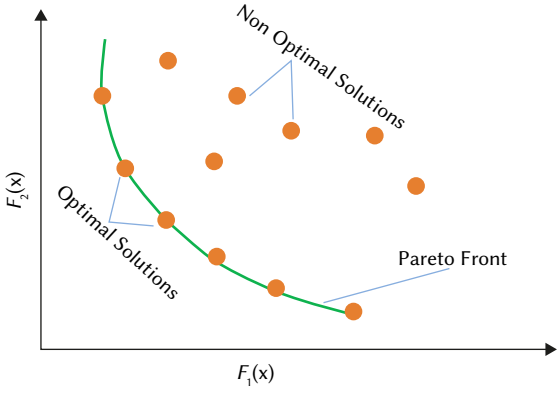


Fig. 5. Multi Objective (Pareto Front) Optimization.

The same problem could be considered as a weighted sum method so that the weight of an objective is chosen in proportion to the relative importance of the objective as given in the formula (3) [19]

$$minimize\ F(x) = \sum_{\emptyset \in X} wf(x) \tag{3}$$

The objective function J has been framed with due consideration being given to the tradeoff between human feedback and reward.

h_j is the component which represents human feedback and represents the system's quality based on the evaluator's assessment as given in (4). This is the average of the scores. A higher value indicates that it has a better performance.

$$h_j = f_h\ (f_1, f_2, \dots, f_n) \tag{4}$$

r_j is the component which represents the reward and represents the system's quality based on the rewards awarded by reinforcement-based techniques and the aim is to maximize reward as in (5).

$$r_j = f_r\ (r_1, r_2, \dots, r_m) \tag{5}$$

As in the optimization problems, we introduce the tradeoff parameter α which determines the relative significance of human feedback and reward component, and it ranges from 0 to 1 as given in (6).

$$J = \alpha * h_j + (1 - \alpha) * r_j \tag{6}$$

is the optimization function of both human feedback and reward components. From the formula when α is 1, human feedback gets maximum priority and when α is 0, optimization will fully depend on reward components and the values in between 0 and 1, determine the tradeoff between both. Optimization helps to find the optimal parameter values, finding the best balance between both the parameters.

Normally, in an RLHF model, there is a Proximal Policy Optimization (PPO) that helps the LLM to learn to generate summaries that are more accurate and score well according to the reward model. This process is iterative and involves several rounds of fine-tuning, which finally provides context–specific summaries. Each iteration makes the model better aligned, so that it reaches an optimal stage in terms of the reward model. By incorporating these in the normal RLHF process, the challenges based on human involvement and preferences can be controlled to a minimal extent [20].

C. Proposed H-Rouge (RH) as a Metric

Rouge is one of the most well-known evaluation metrics used in the field of language models to check the accuracy of output summarizations [21]. Rouges 1, 2, and N are computed based on the precision, recall, and F1-score of the matching 1, 2, and n-grams, where the Rouge uses a reference summary and generated summary. ROUGE-L (RL) is based on the longest common subsequence (LCS) between the generated summary and the reference summary. The longest sequence of words that considered the generated summary and the reference summary was calculated [22]. In RLHF, human feedback is taken; hence, the proposed method adds weightage to human feedback. The proposed H-Rouge (RH) also uses precision, recall, and F1 measures. However, a human that adds weight on a scale of {0 to 1} is given. If more human feedback is given, then the weighted average is added along with the values given by Equation (7).

$$H - Rouge\ (RH) = [Rouge - L + \textstyle\sum_{i=1}^{n} wh_i\ /n]/2 \tag{7}$$

where, i is the specific human feedback and $\sum wh_i$ *is the weighted average of Human added values* and *n* is the number of humans given feedbacks.

This H-Rouge (RH) will give a better understanding and accuracy considering the human feedback explicitly. If humans think the summary is sufficiently close, the value given might be close to 1, and vice versa. The feedback was based on a scale from 0 to 1. If more human feedback is given for the same output of the LLM, then the weighted average of those values will be obtained. The feedback $H = \{f_1, f_2, \dots f_n\}$ values are ranked in the ascending order based on the values obtained. The complete process is shown in Algorithm 1 [22].

Algorithm 1. Improved RLHF Process with people selection method

Input: Training samples

Output: Accurate Summary based on H-Rogue

1. Start the process by selecting the training dataset
2. Use Supervised learning LLMs over the training samples and generate summaries GS = $\{s_1, s_2, \dots, s_n\}$
3. Summaries $\{s_1, s_2, \dots, s_n\}$ are human reviewed and are given the HF = $\{f_1, f_2, \dots, f_n\}$
4. Until the desired reward is met, go to Step 5.
5. Repeat
6. Selecting the humans for giving feedback based on their profile topic match
   - Select the human database
   - Cluster the database based on the topic preferences by K-Means clustering method
   - Forms the human clusters and the number of clusters decided by silhouette or elbow method
   - Comparlng with the ACM CCS ontology, map the topics to the nodes in the knowledge graph
   - Identify the top nodes based on the semantic similarity method
   - Identify the specificity index of the nodes or the information content
   - Rank the nodes and gives recommendations for the selection of human most appropriate for the feedback
7. Once the feedbacks HF = $\{f_1, f_2, \dots, f_n\}$ are obtained, rank the outputs
8. Pass it to the reward models, based on the feedback revert
9. Calculate H-Rogue to see the best and accurate model efficiency
10. Return summary with the highest H-Rogue measure
11. End

IV. Evaluation and Results

In this study, we used CNN/DM dataset slightly modified by adding human-generated summaries to make it more appropriate. This study also uses dynamic news by collecting news from online sources using their respective APIs. The first task focused on creating new prompts for summarization tasks. Some samples are provided below for reference purposes. The prompts given below encompass input, messages and response, as shown in Fig. 6. and Fig. 7. [23][24].

The proposed method was compared with state-of-the-art methods. These were selected because of their demonstrated efficacy in comparable NLP tasks, relevance to our research objectives, capacity to address specific challenges in our problem domain, and potential for comparative analysis against our proposed approach. Furthermore, consideration was given to methods that have exhibited promising results in recent literature, represent diverse algorithmic paradigms, and offer insights into various aspects of language understanding and generation. Currently, there is a need to perform a detailed comparison of various models with prompt based RLHF methods such as BRIO, GPT3, and T0 [25][26]. The SOTA summarization algorithms considered are listed in Table I [27][28][29].

```
prompt = f"""
Your task is to extract revelant information from \
a paper content from the website to give \
feedback to readers.

You may use extractive summarization model BERT to find them

From the content below, delimited by triple quotes \
extract the information relevant to knowledge grapghs and Information cont

Review: ```(Paper_review)```
"""

responde = get_completion(prompt)
print(response)
```

```
This paper proposes a semantic similarity method called wpath that combines structure
and Information Content (IC) of concepts in Knowledge Graphs (KGs). Graph-based IC is
used to compute IC based on the distributions of concepts over instances. The wpath
method outperforms other semantic similarity methods in word similarity datasets and
category classification evaluation.
```

```
prompt = f"""
Determine five topics that are being discussed in the \
following text, wich is delimited by triple backticks.

Make each item one or two words long.

Format your response as a list of items separated by commas.

Text sample: ```{Paper_content}```
"""

responde = get_completion(prompt)
print(response)
```

```
1. Semantic similarity
2. Knoledge Graphs
3. Information Content
4. Corpus-based IC
5. Graph-based IC
```

Fig. 6. Use of a prompt in a summarization scenario.

1.

```
News_content = """By Rebecca Morelle & Jake Horton
BBC News
A former employee of OceanGate - the company that operates the missing T
US court documents show that David Lochridge, the company's Director of
The report "identified numerous issues that posed serious safety concern
Mr Lochridge "stressed the potential danger to passengers of the Titan a
The BBC tried to contact Mr Lochridge but he is not commenting.
Separately, a letter sent to OceanGate by the Marine Technology Society
A spokesman for OceanGate  declined to comment on the safety issues raise
The Titan submersible, described as "experimental" by the company, was b
Its hull - surrounding the hollow part where passengers sit - was made f
```

2.

```
prompt = f"""
Create a headline for the news wich is catchy for the reader
Determine five topics that are being discussed in the \
following text, which is delimited by triple backticks.
Make each item one or two words long.
Please make the heading in Bold
Format your response as a heading followed by a list of items separated by
create another short news from the list of items that are relevant for the
Text sample: ```{News_content}```
"""
responde = get_completion(prompt)
print(response)
```

```
Former employee warned of safety concerns with missing subnersible

- Safety concerns
- Director of Marine Operations
- Lawsuit settlement
- Experimental approach
- Carbon fibre hull

New safety concerns raised over missing submersible with experimental
carbon fibre hull. Former employee warned of potential dangers and raised
concerns about the vessel's hull and Alass viewoort, but claims his warning
```

3.

```
topic_list = [
  "Ocean", "Titanic", "Submarine",
  "employee satisfaction", "Law suit", "government", "Laws", "Safety"
]
```

4.

```
prompt = f"""
Determine whether each item in the following list of \
topics is a topic in the text below, which
is delimited with tri Ie backticks.
Give your answer as list with 0 or 1 for
List of topics: {", ".join(topic_list)}
Text sample: '''{News_content}'''
"""

response = get_completion(prompt)
print(response)
```

```
Ocean: 1
Titanic: 0
Submarine: 1
employee satisfaction: 0
Law suit: 1
government: 0
Laws: 0
Safety: 1
```

1. Shows the BBC News Content extracted from online source
2. Shows how the topics are identified and how short news is created
3. Shows the topic list provided explicitly to check
4. From the item list the prompt checks whether the topic exists in the textual content of news data.

These examples gives you a clear idea of how the prompts can be formulated for news summarization purpose.

Fig. 7. Use of a prompt for news summarization and identification of topics.

TABLE I. SOTA Vs RLHF Methods Comparison

Method Types		Model Name	R1	R2	RL	RH
State of the Art (SOTA)	Abstractive (ABS) and Extractive (EXT) Summarization models	BERT-Base	44.22	20.62	40.38	-
		RoBERTA-Base	44.41	20.86	40.55	-
		BERTSUM-EXT	43.25	20.24	39.63	-
		BERTSUM-ABS	41.72	19.39	38.76	-
		BERTSUM-EXT-ABS	42.13	19.60	39.18	-
		HiBERT	42.31	19.87	38.78	-
	Fine Tuned Models	BART	44.16	21.28	40.90	-
		BART +BERT-Base	45.94	22.32	42.48	-
		PEGASUS -Base	41.79	18.81	38.93	-
	Zero- or few-shot models	BRIO	38.49	17.08	31.44	-
		GPT3-D2	31.86	11.31	24.71	-
		T0	35.06	13.84	28.46	-
RLHF based Prompting Methods	Fine Tuned with RLHF with Human Selection Strategy	BRIO	40.21	19.25	33.45	37.86
		GPT3-D2	33.74	13.41	25.54	28.89
		T0	32.81	13.41	25.96	27.56

From Table I, it is clear that the values from the abstraction methods and select extractive methods, such as BERTSUM-EXT [30], always tend to yield better performance. BART based also provide excellent performance on the CNN/DM dataset [31]. When compared to reinforcement learning from human feedback with and without a human selection strategy, it is noticed that there is a small improvement in the Rouge L value for the BRIO and GPT3-D2 methods with using a human selection strategy. However, when using RLHF with a human selection strategy [32], improvement is better in the Rouge values in the RH (proposed H-Rouge) values, considering the scores according to the feedback given. However, more extensive studies need to be conducted on large-scale data with various datasets, which is progressing.

RH values are not calculated for the SOTA algorithms because they do not provide reinforcement learning from human feedback [33]. This formulated metric can be used for methods that use feedback systems, and scores can be considered. These scores were used as H factors in the RH formula. There may be multiple people reviewing LLM-based summaries and it depends on the complexity and purpose of the summaries to be generated. For the initial trial, the weighted average of the H scores was added along with the Rouge L scores.

The work is progressing to associate and formulate a method to optimize the weight along with the scores depending on the humans to make the formula better and unbiased. The R1, R2, RL and RH values for various prompting methods like BRIO, GPT3-D2 and T0 could be seen from the plot presented in Fig. 8. and summary generated is presented in Fig. 9. BRIO and GPT3-D2 are proven to have a better RH value using the human selection strategy in RLHF fine-tuned methods [34].

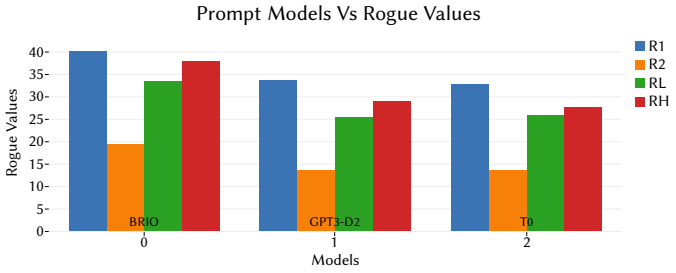


Fig. 8. Rouge {1, 2, L and H} values obtained with BRIO, GPT3-D2 and T0 methods.

Apart from the ChatGPT output kept for visual comparison, both BRIO and GPT3-D2 gave better results than T0. The execution time in seconds taken can be understood by the following graph for the different models used as given in Fig. 10.

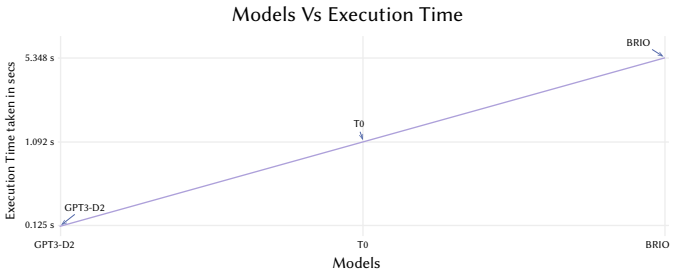


Fig. 10. Execution time Vs Models used.

This study offers significant contributions and showcases the effectiveness of the suggested techniques in enhancing news summarization. Although the initial outcomes are encouraging, additional explanation regarding real-world implementation and a more extensive assessment across varied datasets could reinforce the reason for its practical significance.

V. Conclusion

In this article, we studied various reinforcement learning fine-tuned prompting-based methods for news summarization. All these methods were compared with state-of-the-art summarization algorithms categorized as extractive, abstractive, and fine-tuned models with and without RLHF. In addition, the major contribution of this work is that we have proposed a human selection strategy for the RLHF models used, multi-objective optimization is used for the tradeoff between various objectives introduced in the process, and the third contribution is the proposed evaluation metric H-Rouge (RH). The RH evaluation metric can be used for scenarios in which humans need to provide reviews and feedback. This helps score and obtain an unbiased consideration of the feedback scores through the process. The main advantages of including these human evaluations will lead to systems with an improved user experience, accurate summarizations, and reduced training costs. The proposed human selection strategy helps

Original Text	A former employee of OceanGate - the company that operates the missing Titan submersible - warned of potential safety problems with the vessel as far back as 2018.US court documents show that David Lochridge, the company's Director of Marine Operations, had raised concerns in an inspection report. The report "identified numerous issues that posed serious safety concerns", according to the documents, including the way the hull had been tested. Mr Lochridge "stressed the potential danger to passengers of the Titan as the submersible reached extreme depths". He said his warnings were ignored and called a meeting with OceanGate bosses but was fire, according to the documents.
BRIO	Former employee of OceanGate warned of potential safety problems with the Titan submersible. David Lochridge raised concerns about the way the hull had been tested. He said his warnings were ignored and he was fired from the company in 2018. Court documents show he raised concerns as far back as 2018.
T0	Former OceanGate employee David Lochridge warned of safety problems with the submersible Titan
GPT3-D2	A former employee of OceanGate the company that operates the missing Titan submersible - warned of potential safety problems with the vessel as far back as 2018.US court documents show that David Lochridge, the company's Director of Marine Operations, had raised concerns in an inspection report- The report "identified numerous issues that posed serious safety concerns", according to the documents, including the way the hull had been tested. Mr Lochridge "stressed the potential danger to passengers of the Titan as the submersible reached extreme depths". He said his warnings were ignored and called a meeting with OceanGate bosses but was fired, according to the documents.
GPT3.5 OpenAI Chat Bot	In 2018, David Lochridge, a former employee of OceanGate, the company operating the missing Titan submersible, expressed safety concerns about the vessel. US court documents reveal that Lochridge, who served as the company's Director of Marine Operations, issued a detailed inspection report highlighting multiple safety issues. These issues, as outlined in the documents, encompassed various aspects, including concerns about the testing of the hull. Lochridge emphasized the potential dangers that passengers could face as the submersible delved into extreme depths. Despite his warnings, he alleged that OceanGate ignored his concerns and subsequently terminated his employment after he called for a meeting with the company's leadership.

Fig. 9. Summaries generated for a test news item from BBC News.

users obtain more targeted information based on specific interests, which enhances the overall personalized user experience.

A few of these are listed here to help nascent researchers in the field of news summarization. The future scope of this work includes the prominence of ethical issues when more human evaluation assistance is embedded in AI models. The major challenge is to find a tradeoff between them, probably by formulating an optimization problem by identifying and adjusting the parameters involved in the process. The process can further be extended for its research towards aspect-based summarization, where many common aspects, high-level topics, or popularity-based topics could be considered and how well these algorithms will work for such scenarios [35]. This work is being tested in various diverse applications in the news summarization area of research, helping to develop various humanized AI systems that nudges researchers to dwell deeper into diverse applications with even more diverse user information needs.

References

[1] K. I. Roumeliotis and N. D. Tselikas, "ChatGPT and Open-AI Models: A Preliminary Review," *Future Internet*, vol. 15, no. 6, p. 192, 2023.

[2] N. Wu, M. Gong, L. Shou, S. Liang, and D. Jiang, "Large language models are diverse role-players for summarization evaluation," *ArXiv preprint*, arXiv:2303.15078, 2023.

[3] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1-35, 2023.

[4] N. Ding, S. Hu, W. Zhao, Y. Chen, Z. Liu, H. T. Zheng, and M. Sun, "OpenPrompt: An open-source framework for prompt-learning," *ArXiv preprint*, arXiv:2111.01998, 2021.

[5] V. Deokar and K. Shah, "Automated Text Summarization of News Articles," *International Research Journal of Engineering and Technology*, vol. 8, no. 9, pp. 1-13, 2021.

[6] S. A. Bahrainian, S. Feucht, and C. Eickhoff, "NEWTS: a corpus for news topic-focused summarization," *ArXiv preprint*, arXiv:2205.15661, 2022.

[7] J. Wang, Z. Liu, L. Zhao, Z. Wu, C. Ma, S. Yu, and S. Zhang, "Review of large vision models and visual prompt engineering," *ArXiv preprint*, arXiv:2307.00855, 2023.

[8] J. Wang, E. Shi, S. Yu, Z. Wu, C. Ma, H. Dai, and S. Zhang, "Prompt engineering for healthcare: Methodologies and applications," *ArXiv preprint*, arXiv:2304.14670, 2023.

[9] V. Liu and L. B. Chilton, "Design guidelines for prompt engineering text-to-image generative models," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1-23, Apr. 2022.

[10] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, and J. Ba, "Large language models are human-level prompt engineers," *ArXiv preprint*, arXiv:2211.01910, 2022.

[11] T. Goyal, J. J. Li, and G. Durrett, "News summarization and evaluation in the era of GPT-3," *ArXiv preprint*, arXiv:2209.12356, 2022.

[12] H. Liu, C. Sferrazza, and P. Abbeel, "Languages are rewards: Hindsight finetuning using human feedback," *ArXiv preprint*, arXiv:2302.02676, 2023.

[13] N. Stiennon, L. Ouyang, J. Wu, D. Ziegler, R. Lowe, C. Voss, and P. F. Christiano, "Learning to summarize with human feedback," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3008–3021, 2020.

[14] G. Wu, W. Wu, X. Liu, K. Xu, T. Wan, and W. Wang, "Cheap-fake Detection with LLM using Prompt Engineering," *ArXiv preprint*, arXiv:2306.02776, 2023.

[15] T. K. Gilbert, N. Lambert, S. Dean, T. Zick, A. Snoswell, and S. Mehta, "Reward reports for reinforcement learning," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 84-130, Aug. 2023.

[16] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, and R. Lowe, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27730–27744, 2022.

[17] G. Zhu and C. A. Iglesias, "Computing semantic similarity of concepts in knowledge graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 72–85, 2016.

[18] H. T. Kung, F. Luccio, and F. P. Preparata, "On Finding the Maxima of a Set of Vectors," *Journal of the ACM*, vol. 22, no. 4, pp. 469–476, 1975.

[19] A. Rame, G. Couairon, M. Shukor, C. Dancette, J. B. Gaya, L. Soulier, and M. Cord, "Rewarded soups: towards Pareto-optimal alignment by interpolating weights fine-tuned on diverse rewards," *ArXiv preprint*, arXiv:2306.04488, 2023.

[20] Y. Zhao, R. Joshi, T. Liu, M. Khalman, M. Saleh, and P. J. Liu, "Slic-hf: Sequence likelihood calibration with human feedback," *ArXiv preprint*, arXiv:2305.10425, 2023.

[21] P. J. A. Colombo, C. Clavel, and P. Piantanida, "Infolm: A new metric to evaluate summarization & data2text generation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 10, pp. 10554-10562, Jun. 2022.

[22] T. Oka, P. Patankar, S. Rege, and M. Dixit, "Text summarization of news articles," in *ICT Systems and Sustainability: Proceedings of ICT4SD 2021, Volume 1*, pp. 441-450, Springer Singapore, 2022.

[23] Y. Li, "Iterative improvements from feedback for language models," *ScienceOpen Preprints*, 2023.

[24] A. Ng, "Deep learning specialization," *DeepLearning.AI/Coursera*, 2020.

[25] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, and D. Amodei, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877-1901, 2020.

[26] Y. Liu, P. Liu, D. Radev, and G. Neubig, "BRIO: Bringing order to abstractive summarization," *ArXiv preprint*, arXiv:2203.16804, 2022.

[27] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *ArXiv preprint*, arXiv:1910.13461, 2019.

[28] J. Zhang, Y. Zhao, M. Saleh, and P. Liu, "Pegasus: Pre-training with extracted gap-sentences for abstractive summarization," in *International Conference on Machine Learning*, pp. 11328-11339, Nov. 2020.

[29] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Systems with Applications*, vol. 165, p. 113679, 2021.

[30] T. Wolf, "Huggingface's transformers: State-of-the-art natural language processing," *ArXiv preprint*, arXiv:1910.03771, 2019.

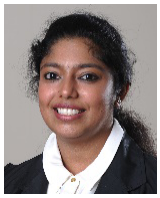
[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[32] Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, and J. Kaplan, "Training a helpful and harmless assistant with reinforcement learning from human feedback," *ArXiv preprint*, arXiv:2204.05862, 2022.

[33] M. M. Afsar, T. Crump, and B. Far, "Reinforcement learning based recommender systems: A survey," *ACM Computing Surveys*, vol. 55, no. 7, pp. 1-38, 2022.

[34] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," *Advances in Neural Information Processing Systems*, vol. 35, pp. 22199-22213, 2022.

[35] O. Ahuja, J. Xu, A. Gupta, K. Horecka, and G. Durrett, "ASPECTNEWS: Aspect-oriented summarization of news documents," *ArXiv preprint*, arXiv:2110.08296, 2021.



Sini Raj Pulari

Sini Raj Pulari received the B.Tech degree in Computer science and engineering in 2007 from University of Calicut, India, the M.E Post graduation in Computer Science and Engineering in 2011 from Anna University, India and master's in business administration from Pondicherry University, India in 2021. She is currently pursuing the Ph.D. in Natural Language Processing Vignan's Foundation for Science, Technology and Research University, India. She is working for academia for past 15 years in the field of artificial intelligence, deep learning, and natural language processing as the major research interests. She has published 20 plus quality research articles and have co-authored two books published under Taylor and Francis publications (CRC Press) in the field of AI and ML. Author's Awards and honors include FHEA (Advance HE), best faculty award, and is an official Quality Matters Peer Reviewer course, Intel certified instructor (Machine Learning), Intel certified Instructor (oneAPI DPC++ essentials).

Umadevi Maramreddy

Umadevi Maramreddy completed Ph. D in Computer Science from University of Hyderabad in 2011 in area of Document forensics. She Worked as JRF and SRF in Government Examiner of Questioned Document, Hyderabad. She has 16 years of experience out of 4 years of research and 12 years of teaching experience. Her Research interests are Printed Document forensics, Image Processing, Soft Computing, Natural Language Processing and Machine Learning. She has published 13 papers in various journal and international conferences.

Shriram K. Vasudevan

Shriram K. Vasudevan has over 17 years of experience in the Industry and Academia together. He holds a Doctorate in embedded systems. He has authored / co-authored 45 books for various publishers including Taylor and Francis, Oxford University Press, and Wiley. He also has been granted 14 patents so far. Shriram is a hackathon enthusiast and has been awarded by Harvard University, AICITE, CII, Google, TDRA Dubai, Govt. Of Saudi Arabia, Govt. Of India and a lot more. He has published more than 150 research articles. He was associated with L&T Technology Services before joining in current role with Intel. Shriram Vasudevan runs a YouTube channel in his name which has more than 49K subscribers and maintains a wide range of playlists on varied topics. Dr. Shriram is a public speaker too and participated in multiple training events. He is oneAPI certified Instructor, ACM Distinguished Speaker and NASSCOM Prime Ambassador. Shriram is a Fellow – IEI, Fellow – IETE and Senior Member – IEEE.