An Empirical Study of Prompt Engineering with Large Language Models for Hope Detection in English and Spanish

Nguyen Thi Thuy[1,2,*], Dang Van Thin[1,2]

[1]University of Information Technology-VNUHCM, Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

[2]Vietnam National University, Ho Chi Minh City, Vietnam

Abstract

Hope is one of the exceptional human capacities that allows for flexible anticipation of future events and possible expected outcomes. Hope speech detection is a crucial task in understanding and fostering positive discourse within online communities. Shared task HOPE at IberLEF 2024 encompasses two distinct but complementary subtasks, each addressing different facets of hope speech. Task 1 is called as "Hope for Equality, Diversity, and Inclusion" and aims to detect hope speech related to Equality, Diversity, and Inclusion. Task 2, "Hope as Expectations" focuses on expectations and desirable and undesirable facts. In this paper, we present our unsupervised approach to solve both tasks. Our main idea is to leverage the power of a large language model, ChatGPT 3.5, and the prompting technique to solve two tasks. We used ChatGPT 3.5 with three prompting techniques: zero-shot, few-shot (one-shot, three-shot), and chain of thought, combined with six different information strategies. Our best prompts achieved top 1 in task 1 with a Macro F1-score of 0.7161. However, the performance did not meet expectations in the sub-tasks of task 2.

Keywords

Hope speech detection, ChatGPT, Prompt Engineering, Spanish Language, English Language, Few-shot prompt, Large language models

1. Introduction

Hope is one of the exceptional human capacities that allows for flexible anticipation of future events and possible expected outcomes. These visions significantly influence emotions, behaviour, and mood (Bruininks and Malle [1]). Nowadays, people often post articles to share their feelings, thoughts, and opinions on social media. But in addition to posts with positive messages, there are also massive posts of offensive messages posted against people because of their race, colour, ethnicity, gender, sexual orientation, nationality, or religion. Research by Chakravarthi [2] has pointed out that posts on social media have a significant effect on vulnerable groups. It plays an essential role in shaping the individual's personality and view of

IberLEF 2024, September 2024, Valladolid, Spain

*Corresponding author.
21521514@gm.uit.edu.vn (N. T. Thuy); thindv@uit.edu.vn (D. V. Thin)
https://nlp.uit.edu.vn/ (D. V. Thin)
0000-0001-8340-1405 (D. V. Thin)

society (Burnap et al. [3], Kitzie [4], Milne et al. [5]) during their use of social media. Therefore hope speech detection becomes a meaningful task in the Natural language processing (NLP) community. It helps to prevent toxic posts and spread hope speech, which is the type of speech that can relax a hostile environment (Palakodety et al. [6]) and that helps, gives suggestions, and inspires good feelings in several people when they are in times of illness, stress, loneliness, or depression (Chakravarthi [2]).

Inspired by applying Large language models (LLMs) or ChatGPT to solve Sentiment analysis tasks (Fatouros et al. [7], Belal et al. [8], Sudirjo et al. [9], Zhang et al. [10]) and the efficiency of applying Zero-shot ChatGPT for hope speech setection in the HOPE at IberLEF 2023 (Ngo and Tran [11]), we only use ChatGPT 3.5 to solve both tasks. However, we not only use Zero-shot as [9] but also carry out experiments on many prompting techniques and various strategies to provide information to find the most effective prompts to solve this task.

Prompt Engineering helps to effectively design and improve prompts to get better results on different tasks with LLMs. There are many techniques have been proven to be effective when working with LLMs, in this study we only focus on three techniques Zero-shot, Few-shot, and chain of thought promptings. In the prompting context, the term "shot" refers to the demonstration or example of what exactly the user wants LLMs to do. Zero-shot means no examples are provided for LLMs, and instruction tuning has been shown to improve zero-shot learning (Wei et al. [12]). Although LLMs can achieve remarkable results with the Zero-shot technique, however, Few-shot - prompting with some illustrations is better for complex situations. (Min et al. [13]) present some tips exemplars when doing few-shots such as the label space and the distribution of the input text specified by the demonstrations are both important, the format prompting also plays a key role in performance, selecting random labels from a true distribution of labels also helps. Chain of thought (CoT) prompting enables final results to be derived through intermediate inference steps (Wei et al. [12]).

2. Task description

The shared task on hope speech detection was held many times in the past at the second workshop on Language Technology for Equality, Diversity and Inclusion (LT-EDI-2022), as a part of ACL 2022 (Chakravarthi et al. [14]), at LT-EDI-2023, within RANLP 2023 and the shared task HOPE in IberLEF 2023 (Kumaresan et al. [15]). This new edition - HOPE in IberLEF 2024 (García-Baena et al. [16], Chiruzzo et al. [17]) is novel when researching Hope through two perspectives corresponding with two tasks in this shared task

- **Task 1: Hope for Equality, Diversity and Inclusion**
  This task is related to the inclusion of vulnerable groups and focuses on the study of the detection of hope speech in pursuit of Equality, Diversity, and Inclusion. Given a tweet written in Spanish, participants must identify whether it contains hope speech or not. Specifically, this task is divided into two subtasks, but participants will participate in both of them at the same time.
  With given training data on LGTBI tweets from García-Baena et al. [18] (tweets related to vulnerable groups are the Lesbian, Gay, Bisexual, and Transgender community), participants must classify each tweet in the test set in one of the following categories:

  - hs: if the tweet contains hope speech.
  - nhs: if the tweet does not contain hope speech.

  where the test is set in subtask 1.a only contains LGTBI-related tweets while subtask 1.b contains tweets on unknown domains.

- **Task 2: Hope as expectations.**
  Hope is characterized as "openness of spirit toward the future, a desire, expectation, and wish for something to happen or to be true" that remarkably affects a human's state of mind, emotions, behaviours, and decisions (Bruininks and Malle [1], Balouchzahi et al. [19]). This task focuses on expectations and desirable and undesirable facts. Specifically, it is divided into two subtasks.
  - Subtask 2a: Binary hope speech detection from English and Spanish Texts
    Given training data, the participant will classify the text into two categories. In this problem, each text will be assigned one of the following labels :
    * Hope: tweets that convey a mention of hope.
    * Not Hope: tweets that do not convey hope, expectation, or desire.
  - Subtask 2b: Multiclass hope speech detection from English and Spanish Texts (Balouchzahi et al. [20])
    This subtask is similar to Subtask 2a. However, it becomes more challenging when each tweet is considered Hope. The participant must specify the specific type of Hope to which it belongs, whether Generalized Hope, Realistic Hope, or Unrealistic Hope, where:
    * Generalized Hope: is expressed as a general hopefulness and optimism that is not directed toward any specific event or outcome (Wiles et al. [21], Webb [22], Ezzy [23], Lohne and Severinsson [24], Lohne and Severinsson [24], Smith and Sparkes [25]).
    * Realistic Hope: is about expecting something reasonable, meaningful, and possible thing to happen (Wiles et al. [21], Eaves et al. [26], Garrard and Wrigley [27]).
    * Unrealistic Hope: is usually in the form of wishing for something to become true, even though the possibility of happening is remote or significantly less or even zero (Webb [22], Eliott and Olver [28], Links and Kramer [29]).
    * Not Hope: tweets that do not convey hope.

3. Approach

We designed prompts to solve this shared task and experimented to research two aspects of the prompt design technique are Providing information strategies and Prompting techniques with ChatGPT 3.5.

3.1. Providing information strategies

This aspect aims to research which information is essential and efficient for Large language models to solve this shared task. We propose four kinds of information and combine them to create six providing information strategies.

Strategy 1: <information 1>

Strategy 2: <information 2> <information 1>

Strategy 3: <information 3>\n<information 1>

Strategy 4: <information 4> <information 1>

Strategy 5: <information 2>\n<information 3>\n<information 1>

Strategy 6: <information 4> <information 2>\n<information 3>\n<information 1>

Figure 1: Syntax of the combination of 4 information to create 6 Providing information strategies

- **Information 1 - Request:** This information simply points out what we want Large language models to do.
- **Information 2 - Concept of problem:** This information provides the concept or definition of the key problem LMMs must solve.
- **Information 3 - Meaningful of classes:** In this study, we apply ChatGPT for classification tasks, so the meaning of classes might be the essential information.
- **Information 4 - Role defining:** In most situations, it is better to clearly understand that Someone must do something with a specific role. In this study, we defined ChatGPT as an NLP engineer expert and a language specialist, and it is labelled to create a dataset for this shared task. We use this role throughout subtasks in this shared task.

Fig 1 illustrates the way we use four kinds of information above to create six information strategies. In which:

- **Strategy 1 - Only request:** This strategy simply provides directly what we want Large language models to do without any additional information.
- **Strategy 2 - Concept of problem and Request:** Except for the request information, this strategy also provides the concept of key problem LMMs must do.
- **Strategy 3 - Meaningful of classes and Request:** Except for the request information, this strategy also provides the meaning of each class LLMs must classify.
- **Strategy 4 - Role defining and Request:** Before giving the request, this strategy provides information about the role of LLMs with the hope that LLMs will understand more about their context and tasks.
- **Strategy 5 - Combination 1:** Simply, this strategy is the combination of Information 2, 3 and 1.
- **Strategy 6 - Combination 2:** Simply, this strategy is the combination of Information 2, 3, 4, and 1.

3.2. Prompting techniques

This aspect focuses on how to give the request and instructions for Large language models to generate the answer. In this study, we experimented with three prompting techniques: zero-shot, few-shot, and chain of thought (CoT). Additionally, we experimented with One-shot and Three-shot for the few-shot prompt setting. Specifically, each technique is as follows:

Normal answer format:
Tweet: “<given tweet>”
Label:

CoT answer format:
Tweet: “<given tweet>”
Explanation:
Label:

Zero-shot prompting:
<Strategy>\n<normal answer format>

One-shot prompting:
<Strategy>\n<One-shot demontration>\n<normal answer format>

Three-shot prompting:
<Strategy>\n<Three-shot demontration>\n<normal answer format>

CoT prompting:
<CoT Strategy>\n<CoT answer format>\n<CoT answer format>

Figure 2: Syntax of the combination of Strategy, Prompting technique, and answering format to create complete Prompting.

- **Zero-shot:** With the Zero-shot technique, the final prompts are the strategies of the Providing information strategies aspect without providing additional information.
- **Few-shot:** Except for the instructions information of the Providing information strategies, this technique will provide some examples to demonstrate what exactly the LMMs must do and what our expected output is. In this study, with the One-shot technique, we will provide one demonstration corresponding to each class. Similar to the Three-shot technique, we will provide three demonstrations corresponding to each class.
- **Chain of thought:** Applying the chain of thought technique for general classification tasks is a form that requires LLMs to provide explanations for their classification results.

Because of the differences in the content of the expected answers in the Chain of thought and other techniques. We need to define two separate answering formats: the normal answer format and the CoT answer format. The harmony between Strategy, Prompting techniques, and answering format to create the final promptings is presented in Figure 2.

4. Experimental Setup

4.1. Dataset and Evaluation Metrics

We solve this shared task by using an unsupervised method, and during the process of carrying out the experiment, we nearly don't need the training data except to take some samples of training to serve for the Few-shot technique. We join all subtasks in this shared task and evaluate each subtask according to the metrics of the organizer.

4.2. Prompting design

Because we join all subtasks in this shared task, each subtask has a different goal, concept, or label, so to be convenient for us to present full prompting for each subtask, we will pre-define some variables to store some part of the text. In this report, We only present design prompting results for task 1; pre-define and complete prompting for all subtasks can be found in this

- **request: request for the Zero-shot and Few-shot technique for task 1**

Please classify whether a given text conveys hope or not. Only return the label from [Hope Speech, Non Hope Speech] without any other text.

- **request_CoT: request for the Chain of thought technique for task 1**

Please classify whether a given tweet conveys hope or not. Only return one label from [Hope Speech, Non Hope Speech] and explain your classification results.

- **concept: the concept of Hope in task 1**

Hope speech is the type of speech that is able to relax a hostile environment and that helps, gives suggestions and inspires for good to a number of people when they are in times of illness, stress, loneliness or depression.

- **class: Meaningful of classes in task 1**

Describe the sentiment of the given text using one of these two attributes: "Hope Speech", "Non Hope Speech". Knowing that a text is considered as "Hope Speech" if the text: (1) explicitly supports the social integration of minorities or otherwise distinct groups in society; (2) promotes tolerance, equality, positivity, and healing for all. And knowing that a text is considered "Non Hope Speech" if it: expresses the negative sentiment, discrimination, violence; or uses insults toward people.

- **one-shot: one demonstration for task 1**

***Tweet**: "Hay que tratar de aislar democrÃ¡ticamente los argumentos polÃticos que justifican a machistas, xenÃ³fobos, racistas y las expresiones de odio a los colectivos LGTBI. El odio genera violencia y la violencia no tiene medida en sus consecuencias"*
***Label**: Hope Speech*
***Tweet**: "Los fundadores del socialismo,-Karl Marx: "Los homosexuales son peores que pederastas".- Friedrich Engels: â€œLa homosexualidad es una aberraciÃ³n de la burguesÃa degeneradaâ€.- IÃ³sif Stalin: "La homosexualidad es un vicio burguÃs""*
***Label**: Non Hope Speech*

- **three-shot: three demonstrations for task 1.** It is similar to one-shot_1, however, the number of demonstrations in each class is three

Figure 3: Pre-defined variables on creating all prompting throughout all techniques and strategies for task 1

repository[1]. Detailed pre-variable to create prompting for task 1 as Fig 3. In Table 1, we show our detail syntax for final prompting in 6 strategies with three techniques: Zero-shot, One-shot, and Chain of thought for task 1. Because of paper length limitations, other subtasks are not presented in this table. But it's easy to deduce based on the pre-defined variables and syntax we've provided at this repository[1]. Additionally, in Figure 4 we present full prompting according to One-shot prompting technique and providing information Strategy 5 - prompting helps us achieve top 1 on the test set of task1.

5. Main result

Table 2 shows the Macro Recision, Macro Recall, and Macro F1-score of all prompting we designed on the validation set for task 1. Through this table, we can see that all prompts achieved positive performance. Although there are small differences between classifications, on average, strategy 3 and strategy 6 generally perform better than the others, especially as they often significantly improve over strategy 1. Strategy 1 is better than Strategy 2 in almost all techniques, although Strategy 2 includes the concept of Hope. This unfortunate situation may be because the concept we have defined is somewhat non-standard compared to the data set, causing ChatGPT to be confused in the process of providing the final classification results.

[1]https://github.com/thuynguyen2003/Shared-task-HOPE-at-IberLEF-2024

Concept of problem	Hope speech is the type of speech that is able to relax a hostile environment and that helps, gives suggestions and inspires for good to a number of people when they are in times of illness, stress, loneliness or depression.
Meaningful of classes	Describe the sentiment of the given text using one of these two attributes: "Hope Speech", "Non Hope Speech". Knowing that a text is considered as "Hope Speech" if the text: (1) explicitly supports the social integration of minorities or otherwise distinct groups in society; (2) promotes tolerance, equality, positivity, and healing for all. And knowing that a text is considered "Non Hope Speech" if it: expresses the negative sentiment, discrimination, violence; or uses insults toward people.
Request	Please classify whether a given text conveys hope or not. Only return the label from [Hope Speech, Non Hope Speech] without any other text.
One-shot	Tweet: "Hay que tratar de aislar democrÃ¡ticamente los argumentos polÃticos que justifican a machistas, xenÃ³fobos, racistas y las expresiones de odio a los colectivos LGTBI. El odio genera violencia y la violencia no tiene medida en sus consecuencias" Label: Hope Speech Tweet: "Los fundadores del socialismo,-Karl Marx: "Los homosexuales son peores que pederastas".-Friedrich Engels: â€œLa homosexualidad es una aberraciÃ³n de la burguesÃa degeneradaâ€.- IÃ³sif Stalin: "La homosexualidad es un vicio burguÃs"" Label: Non Hope Speech
Normal answer format	**Tweet: <given tweet>** **Label:**

Figure 4: Prompting according to One-shot prompting technique and Strategy 5 to solve task1.

The average on strategies of each prompting technique, the performances are similar. However, there is a special in the Chain of thought technique, Strategy 3, 4, 5, and 6 have significant improvements compared to Strategy 1, and 2. Besides, Strategies 4, 5, and 6 in this technique also give stable performance compared to the remaining techniques. This may be because the Chain of thought technique requires more information to make inferences. In this report result, the performance of the Few-shot (One-shot and Three-shot) technique does not improve performance compared to Zero-shot but this is completely within our expectations. Because the task in task 1 is described quite clearly, along with the meaning of the classes fed by strategy 3, it is enough for LLMs to understand its task. Providing additional demonstration with a few shots increases the size of the input text, which can affect the accuracy of LLMs during processing.

Table 3 reports the Macro F1-score on the validation set for subtask 2a and subtask 2b for both English and Spanish. In task 2, Performance seems more stable across techniques and strategies. Three-shot is usually better in all cases, and among strategies, Strategy 3 and Strategy 5 often give the most stable and highest results. If in binary classification tasks (task 1 and subtask 2. a), the Few-shot technique and strategy 3 have not made a difference from Zero-shot and strategy 1. Then, in multi-classification tasks 2. b, Few-shot provides some demonstrations, and strategy 3 has additional information on the meaning of classes compared to strategy 1, which has shown significantly improved results. This demonstrates the effectiveness of providing the meaning of classes and illustrative examples for LLMs in complex tasks.

According to the performance report of both task 1 and task 2, if there is only concept and request in the prompt (strategy 2), then concept information is not really effective in this task. However, when combining the other information, such as the meaning of classes

Table 1

Syntax of prompts created from various Prompting techniques and Providing information strategies of task 1

Strategy	Technique	Prompting syntax	Strategy	Technique	Prompting syntax
Strategy 1	**Zero-shot**	<request> <normal answer format>	**Strategy 2**	**Zero-shot**	<concept><request> <normal answer format>
	One-shot	<request> <one-shot> <normal answer format>		**One-shot**	<concept><request> <one-shot> <normal answer format>
	CoT	<request_CoT> <CoT answer format>		**CoT**	<concept><request_CoT> <CoT answer format>
Strategy 3	**Zero-shot**	<class> <request> <normal answer format>	**Strategy 4**	**Zero-shot**	<role><request_1> <normal answer format>
	One-shot	<class> <request> <one-shot> <normal answer format>		**One-shot**	<role><request> <one-shot> <normal answer format>
	CoT	<class> <request_CoT> <CoT answer format>		**CoT**	<role><request> <CoT answer format>
Strategy 5	**Zero-shot**	<concept> <class> <request> <normal answer format>	**Strategy 6**	**Zero-shot**	<role><concept> <class> <request> <normal answer format>
	One-shot	**<concept>** **<class>** **<request>** **<one-shot>** **<normal answer format>**		**One-shot**	<role><concept> <class> <request> <one-shot> <normal answer format>
	CoT	<concept> <class> <request_CoT> <CoT answer format>		**CoT**	<role><concept> <class> <request_CoT> <CoT answer format>

or the role of LLMs (strategy 5, strategy 6), it is somewhat improved and more stable. With role definition (strategy 4) for LLMs, in English tasks, it works quite poorly, actually reducing accuracy compared to strategy 1 (just giving requests). However, we have an intuition about the effectiveness of defining roles for LLMs in Spanish tasks. It cannot be denied that when role information stands alone in strategy 4, it still produces quite bad results, but when combined with other information in strategy 6 when working with Spanish, it can be effective use. We are not saying that defining the role helps LLMs produce good results with Spanish. We would like to suggest that defining the role as "an NLP engineer expert as a language specialist" when working with Spanish helps Strategy 4 be more competitive when working with English.

Because of the limitation of submission, we can't conduct all the prompting that we designed on the testing set. Table 4 presents the best result and our rank on the test set on all of the subtasks in promptings we submitted on CodaLab. For all subtasks, one-shot techniques are the best, and providing an information strategy is like what we observed on the validation set. We achieved top 1 in task 1, while the results achieved in task 2 were not satisfactory. It is possible that achieving top 1 in task 1 is due to the nature of the test set containing out-of-domain data compared to training data. While applying LLMs can be effective in classification tasks such

Table 2
Performance of all prompting on the validation set for task 1 on Macro Recision, Macro Recall, and Macro F1-score

Technique	Strategy	Macro			Technique	Strategy	Macro		
		Precision	Recall	F1-score			Precision	Recall	F1-score
Zero shot	Strategy 1	78.99	78.50	78.41	Three-shot	Strategy 1	82.31	79.00	78.45
	Strategy 2	77.28	72.00	70.58		Strategy 2	82.31	79.00	78.45
	Strategy 3	**83.50**	**83.50**	**83.50**		Strategy 3	82.55	80.00	79.60
	Strategy 4	80.39	76.00	75.10		Strategy 4	**82.89**	**81.00**	**80.72**
	Strategy 5	82.21	82.00	81.97		Strategy 5	80.20	76.50	75.76
	Strategy 6	82.66	82.50	82.48		Strategy 6	82.55	80.00	79.60
One-shot	Strategy 1	80.08	77.00	76.40	CoT	Strategy 1	77.60	71.50	69.83
	Strategy 2	77.28	72.00	70.58		Strategy 2	75.37	70.00	68.32
	Strategy 3	83.07	82.00	81.85		Strategy 3	83.77	83.50	83.47
	Strategy 4	82.18	80.00	79.66		Strategy 4	82.90	82.50	82.45
	Strategy 5	81.10	79.00	78.64		Strategy 5	**84.50**	**84.50**	**84.50**
	Strategy 6	**84.89**	**84.00**	**83.90**		Strategy 6	82.21	82.00	81.97

Table 3
Performance of all prompting on the validation set for task 2 on Macro F1-score

Technique	Strategy	Subtask 2.a F1-score		Subtask 2.b F1-score		Technique	Strategy	Subtask 2.a F1-score		Subtask 2.b F1-score	
		English	Spanish	English	Spanish			English	Spanish	English	Spanish
Zero shot	Strategy 1	72.41	66.26	41.83	38.52	Three-shot	Strategy 1	75.68	69.16	45.35	39.20
	Strategy 2	73.29	66.90	40.45	37.44		Strategy 2	73.86	67.66	46.40	38.17
	Strategy 3	74.97	70.16	48.30	44.74		Strategy 3	**76.82**	72.34	48.37	42.19
	Strategy 4	68.95	63.38	35.63	35.40		Strategy 4	76.85	71.24	48.27	40.25
	Strategy 5	**75.39**	**71.23**	46.08	44.97		Strategy 5	75.74	68.97	46.86	41.14
	Strategy 6	72.42	68.68	46.20	41.30		Strategy 6	75.75	**72.53**	**48.90**	**42.50**
One-shot	Strategy 1	74.31	68.57	39.93	38.76	CoT	Strategy 1	70.33	6657	41.15	38.06
	Strategy 2	73.80	67.16	38.63	40.25		Strategy 2	69.92	64.67	39.71	35.63
	Strategy 3	**75.00**	68.76	46.22	42.78		Strategy 3	73.26	**70.28**	43.45	40.10
	Strategy 4	74.25	69.90	42.52	41.49		Strategy 4	71.31	64.77	35.22	33.01
	Strategy 5	74.60	68.79	**46.23**	**46.03**		Strategy 5	**73.74**	69.65	44.52	**42.46**
	Strategy 6	72.88	**71.45**	45.44	44.25		Strategy 6	71.21	68.94	**44.74**	**42.46**

Table 4
Final result and rank on the test set on all of the subtasks public on CodaLab

Subtask	Technique	Strategy	Rank	Macro		
				Precision	Recall	F1-score
Task 1	One-shot	Strategy 5	**1**	**0.64**	**0.82**	**0.72**
Subtask 2a - English	One-shot	Strategy 5	9	0.75	0.74	0.73
Subtask 2a - Spanish	One-shot	Strategy 6	11	0.73	0.73	0.73
Subtask 2b - English	One-shot	Strategy 5	9	0.46	0.43	0.42
Subtask 2b - Spanish	One-shot	Strategy 6	7	0.47	0.51	0.48

as saving time, training costs, and the need to worry about training data. However, It is still inferior to supervised learning methods when data are in a specific domain. In particular, LLMs work based on the information we provide in the prompts, so a close description of the concept and goal of the test set is a must. Describe information of test set non-standard will directly affect the results of the method.

6. Conclusion and Future Work

In this study, we designed and carried out experiments on various promptings. With each technique and strategy, we evaluate and explain the results. We achieved top 1 by using prompting which combined of Zero-shot prompting technique and providing information strategy 5, however, the results achieved in task 2 are quite modest. Applying LLMs like ChatGPT 3.5 to solve hope speech detection can bring many benefits, but we need to focus on describing the data. The concepts and problem objectives must be compatible with the data. At the same time, it is necessary to evaluate the complexity of the problem to choose the appropriate technique and strategy to take full advantage of the power of LLMs in tasks like this.

References

[1] P. Bruininks, B. F. Malle, Distinguishing hope from optimism and related affective states, Motivation and emotion 29 (2005) 324–352.

[2] B. R. Chakravarthi, Hopeedi: A multilingual hope speech detection dataset for equality, diversity, and inclusion, in: Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media, 2020, pp. 41–53.

[3] P. Burnap, G. Colombo, R. Amery, A. Hodorog, J. Scourfield, Multi-class machine classification of suicide-related communication on twitter. online social networks and media, 2, 32-44, 2017.

[4] V. Kitzie, " i pretended to be a boy on the internet": Navigating affordances and constraints of social networking sites and search engines for lgbtq+ identity work, First Monday (2018).

[5] D. N. Milne, G. Pink, B. Hachey, R. A. Calvo, Clpsych 2016 shared task: Triaging content in online peer-support forums, in: Proceedings of the third workshop on computational linguistics and clinical psychology, 2016, pp. 118–127.

[6] S. Palakodety, A. R. KhudaBukhsh, J. G. Carbonell, Hope speech detection: A computational analysis of the voice of peace, arXiv preprint arXiv:1909.12940 (2019).

[7] G. Fatouros, J. Soldatos, K. Kouroumali, G. Makridis, D. Kyriazis, Transforming sentiment analysis in the financial domain with chatgpt, Machine Learning with Applications 14 (2023) 100508.

[8] M. Belal, J. She, S. Wong, Leveraging chatgpt as text annotation tool for sentiment analysis, arXiv preprint arXiv:2306.17177 (2023).

[9] F. Sudirjo, K. Diantoro, J. A. Al-Gasawneh, H. K. Azzaakiyyah, A. M. A. Ausat, Application of chatgpt in improving customer sentiment analysis for businesses, Jurnal Teknologi Dan Sistem Informasi Bisnis 5 (2023) 283–288.

[10] W. Zhang, Y. Deng, B. Liu, S. J. Pan, L. Bing, Sentiment analysis in the era of large language models: A reality check, arXiv preprint arXiv:2305.15005 (2023).

[11] A. Ngo, H. T. H. Tran, Zootopi at hope2023iberlef: Is zero-shot chat-gpt the future of hope speech detection, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF

2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEURWS. org, 2023.
[12] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, Q. V. Le, Finetuned language models are zero-shot learners, arXiv preprint arXiv:2109.01652 (2021).
[13] S. Min, L. Xinxi, H. Ari, A. Mikel, L. Mike, H. Hannaneh, Z. Luke, Rethinking the role of demonstrations: What makes in-context learning work?, arXiv preprint arxiv:2202.12837 (2022).
[14] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, S. Cn, J. P. McCrae, M. Á. García, S. M. Jiménez-Zafra, R. Valencia-García, P. Kumaresan, R. Ponnusamy, et al., Overview of the shared task on hope speech detection for equality, diversity, and inclusion, in: Proceedings of the second workshop on language technology for equality, diversity and inclusion, 2022, pp. 378–388.
[15] P. K. Kumaresan, B. R. Chakravarthi, S. Cn, M. Á. García-Cumbreras, S. M. Jiménez-Zafra, J. A. García-Díaz, R. Valencia-García, M. Hardalov, I. Koychev, P. Nakov, et al., Overview of the shared task on hope speech detection for equality, diversity, and inclusion, in: Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion, 2023, pp. 47–53.
[16] D. García-Baena, F. Balouchzahi, S. Butt, M. Á. García-Cumbreras, A. Lambebo Tonja, J. A. García-Díaz, S. Bozkurt, B. R. Chakravarthi, H. G. Ceballos, V.-G. Rafael, G. Sidorov, L. A. Ureña-López, A. Gelbukh, S. M. Jiménez-Zafra, Overview of hope at iberlef 2024: Approaching hope speech detection in social media from two perspectives, for equality, diversity and inclusion and as expectations, Procesamiento del Lenguaje Natural 73 (2024).
[17] L. Chiruzzo, S. M. Jiménez-Zafra, F. Rangel, Overview of IberLEF 2024: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS.org, 2024.
[18] D. García-Baena, M. Á. García-Cumbreras, S. M. Jiménez-Zafra, J. A. García-Díaz, R. Valencia-García, Hope speech detection in spanish: The lgbt case, Language Resources and Evaluation (2023) 1–28.
[19] F. Balouchzahi, G. Sidorov, A. Gelbukh, Polyhope: Two-level hope speech detection from tweets, Expert Systems with Applications 225 (2023) 120078.
[20] F. Balouchzahi, G. Sidorov, A. Gelbukh, PolyHope: Two-level hope speech detection from tweets, Expert Systems with Applications 225 (2023) 120078. doi:10.1016/j.eswa.2023.120078.
[21] R. Wiles, C. Cott, B. E. Gibson, Hope, expectations and recovery from illness: a narrative synthesis of qualitative research, Journal of advanced nursing 64 (2008) 564–573.
[22] D. Webb, Modes of hoping, History of the human sciences 20 (2007) 65–83.
[23] D. Ezzy, Illness narratives: time, hope and hiv, Social science & medicine 50 (2000) 605–617.
[24] V. Lohne, E. Severinsson, Hope during the first months after acute spinal cord injury, Journal of Advanced Nursing 47 (2004) 279–286.
[25] B. Smith, A. C. Sparkes, Men, sport, spinal cord injury, and narratives of hope, Social science & medicine 61 (2005) 1095–1105.
[26] E. R. Eaves, M. Nichter, C. Ritenbaugh, Ways of hoping: navigating the paradox of hope

and despair in chronic pain, Culture, Medicine, and Psychiatry 40 (2016) 35–58.
[27] E. Garrard, A. Wrigley, Hope and terminal illness: False hope versus absolute hope, Clinical Ethics 4 (2009) 38–43.
[28] J. Eliott, I. Olver, The discursive properties of “hope”: A qualitative analysis of cancer patients’ speech, Qualitative health research 12 (2002) 173–193.
[29] M. Links, J. Kramer, Breaking bad news: realistic versus unrealistic hopes, Supportive Care in Cancer 2 (1994) 91–93.