Carrot and Stick: Inducing Self-Motivation with Positive & Negative Feedback

Anonymous ACL submission

Abstract

Positive thinking is thought to be an important component of self-motivation in various practical fields such as education and the workplace. Previous work, including sentiment transfer and positive reframing, has focused on the positive side of language. However, self-motivation that drives people to reach their goals has not yet been studied from a computational perspective. Moreover, negative feedback has not yet been explored, even though positive and negative feedback are both necessary to grow self-motivation. To facilitate self-motivation, we propose **CA**rrot and **STIC**k (***CASTIC***) dataset, consisting of $12,590$ sentences with 5 different strategies for enhancing self-motivation. Our data and code are publicly available at here.

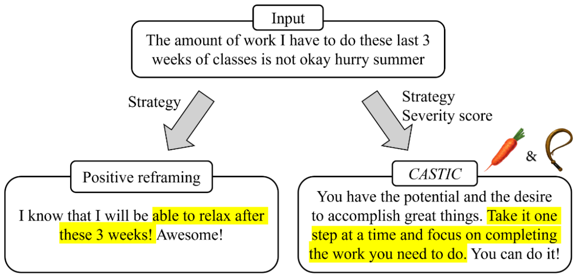

Figure 1: An example of positive reframing (Ziems et al., 2022) and feedback generated using our *CASTIC* framework.

1 Introduction

Interest in positive psychological aspects of language has growing in the field of NLP. Ziems et al. (2022), Sharma et al. (2023), and Maddela et al. (2023) introduce the task of *Positive Reframing*, aiming to shift the negative perspective of a statement into a positive one without altering the original content. Njoo et al. (2023) proposed a new benchmark analyzing how empowerment is conveyed in language.

Previous research has only focused on reframing negative thoughts into positive ones, ignoring the value of non-positive language. In this work, we propose a new approach to appropriately utilize negative (and positive) language as feedback, so as to induce *self-motivation* via stimulation.

Self-motivation is an internal drive that leads a person to take action towards a goal, which is significant in various real-world domains such as education and business. One popular theoretical approach to motivation is Maslow (1958), proposing that motivation is derived from five basic needs: physiological, safety, belongingness & love, esteem, and self-actualization, which are hierarchically organized.

Researchers have attempted to enhance the motivation of people by giving feedback relevant to their situations. In Kim and Lee (2019), it was found that when students received negative feedback, they achieved more accurate self-assessment of skills compared to positive feedback, while positive feedback enhanced students' self-efficacy and boosted confidence in their ability to achieve goals. Hence, the findings from Kim and Lee (2019) suggest that a balanced use of positive and negative feedback is necessary to optimize self-motivation. Wisniewski et al. (2020) also concluded that positive feedback was effective in enhancing confidence and motivation while negative feedback helped to clearly identify areas of deficiency and motivate improvement. Although negative feedback might seem demotivating, it helped in identifying areas that need improvement, guiding future efforts, and avoiding past mistakes. The analysis of the results for each type of feedback in Wisniewski et al. (2020) also aligns with Kim and Lee (2019), indicating that for optimal self-motivation, both positive and negative feedback should be used in a balanced manner.

In this work, we introduce a new benchmark named **CA**rrot and **STIC**k (***CASTIC***) meant measure induction of self-motivation. We address the

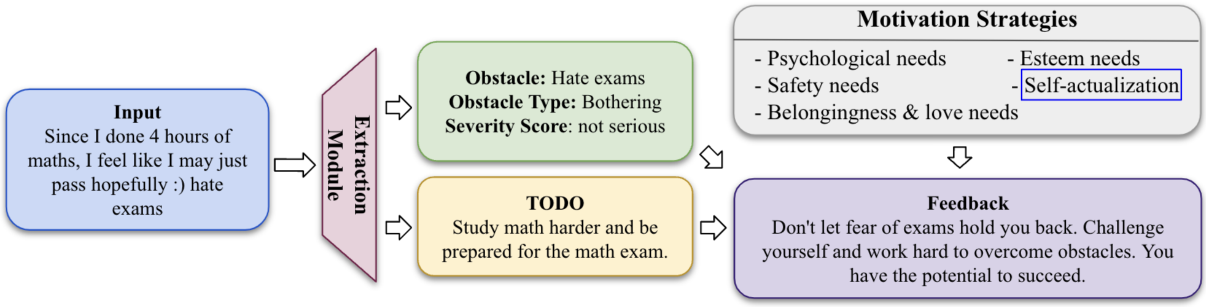

Figure 2: Overall procedure of generating *CASTIC* dataset

task by providing both positive and negative feedback. As far as we can determine, dealing with negative aspects of language in the context of motivation is methodologically novel within NLP. The proposed dataset is generated with a three step procedure and evaluated with both quantitative and qualitative approaches.

2 *CASTIC* Dataset

Large Language Models tend to generate sentences with a positive bias (Chen et al., 2023). However, from the perspective of motivation, unconditional positive support is not always what is needed. Stimulating feedback relevant to the person's circumstance is more effective in achieving goals. Therefore, instead of always giving positive feedback, the model should provide either negative or positive feedback depending on how seriously the obstacle interferes with the task to be done. To give language models this ability, we propose ***CASTIC*** dataset that provides appropriate feedback for given sentences. In this section, we present our overall procedure for data generation and provide a taxonomy of the types of obstacles and strategies for giving feedback.

2.1 Data Collection

The overall procedure of generating *CASTIC* is provided in Fig. 2. The prompt for extracting *TODO*, *Obstacle*, and generating the final feedback sentence are provided in Appendix A.3.

Input sentence We use input sentences from Positive Psychology Frames (POSREF, (Ziems et al., 2022) collected from Twitter with simple keyword #stressed. We use only the `original text` column from the dataset.

Obstacle and TODO Extraction Module We use Orion-14B-Chat (Chen et al., 2024) with an Apache-2.0 License to generate datasets as it is an open-source large language model (LLM) with outstanding performance in comprehensive evaluations and supporting even extremely long texts (Chen et al., 2024). The model first extracts *TODO* and *Obstacle* from the input sentence. *TODO* is the goal that people aim to achieve and *Obstacle* is the challenge or obstacle that hinder people from achieving *TODO*. The model gets to respond "None" when there is no specific *TODO* in a given sentence, and the feedback is generated considering only *TODO*.

Obstacle Type and Severity Score We annotate which of the seven categories the extracted *Obstacle* belongs to. It is worth noting that a sentence can have multiple obstacles and therefore can have multiple types. The *Severity Score* is assigned corresponding to the obstacle type in Table 1. *Severity Score* means how seriously the *Obstacle* blocks the person from *TODO*. If the sentence has multiple obstacles, the sentence is considered "not serious" only when all the obstacle types are considered "not serious". Categories and corresponding severity scores were determined according to the criteria by which we manually checked and classified all input data.

Feedback Generation To generate feedback inducing self-motivation, we use five motivation strategies referenced from the widely known Motivation Theory's "five needs" (Maslow, 1958). A detailed explanation of each need is provided in Appendix A.2. Feedback was created with LLM (Orion-14B-Chat) using the *TODO*, *Obstacle*, and *Severity Score* from the previous step and each of the five needs. The severity score determines whether the feedback is positive or negative, and each of five needs determines which aspects to emphasize to motivate the person. We reviewed each feedback generated by the model.

Obstacle Types	Definition & Example	Severity score
Relationships	Conflict situations with nearby people *fight with friends, mother's nagging*	not serious
Health	Physical or mental illness *migraine, stomachache, burn out*	serious
Fear, overwhelmed	Anxiety about what will happen in the future or previous failings *Anxiety about past failures*	serious
Lack of resource	A lack of supplies needed for work *lack of internet connection, lost laptop*	serious
Annoyance	Irritation by trivial, annoying situations *noisy circumstance*	not serious
Rest, Entertainment	Lack of motivation due to desire for simple entertainment *game, movie, dating*	not serious
etc	Any situation other than the above *Internet/banking system error, bad weather*	serious

Table 1: The seven types of obstacles blocking someone from reaching their goal and the corresponding severity score. The definition is indicated at the top and a corresponding example is given in italics.

Obstacle Types	Train #	Validation #
Relationships	110	80
Health	1,000	305
Fear	5,165	1,205
Lack of resource	235	80
Annoyance	2,015	785
Rest	20	50
etc	1,455	135
serious	7,855	1,725
not serious	2,145	915
Total	10,000	2,640

Table 2: Summary statistics of each obstacle type in the CASTIC dataset.

2.2 Data Distribution

We evaluate the statistics of seven obstacle types in our CASTIC dataset in Table 2. As one sample can have multiple obstacle types, the total number does not indicate the number of distinct samples. The statistics of frequently appearing words in the dataset is provided in Appendix A.6.

3 Self-Motivation Framework

3.1 Task Formulation

To generate feedback, we extract the **Obstacle** O_i and **TODO** T_i from the input sentence. Then, the **Severity Score** SS_i is assigned based on the **Obstacle Type** OT_i of the O_i. Then, **Feedback Type** FT_i is labeled as either positive or negative according to the severity score.

$$FT_i = \begin{cases} \text{Positive if } SS_i = \text{serious}, \\ \text{Negative if } SS_i = \text{not serious} \end{cases}$$

Based on **Motivation Strategy** M_i, the final **Feedback** F_i to induce self-motivation is generated.

$$F_i = \{FT_i, M_i, O_i, T_i\}$$

3.2 Evaluation

3.2.1 Experimental Setup

Dataset The CASTIC dataset contains 9,990 samples in the train split and 2,600 samples in the validation split. All the data are in English.

Model We use BART-L (Lewis et al., 2019), GPT-2 (Radford et al., 2019), M2M-100 (Fan et al., 2021), T5 (Raffel et al., 2020) model to test the dataset. The number of parameters per model is described in Table 3.

3.2.2 Evaluation Metric

Quantitative metric In various studies, the BLEU (Papineni et al., 2002) and ROUGE (Lin, 2004) metrics are utilized to evaluate the semantic similarity with the ground truth. In this paper, we use BLEU, ROUGE-1, ROUGE-2, ROUGE-L, and BERTScore (Zhang et al., 2019) for qualitative results following previous work (Ziems et al., 2022). Cosine similarity between generated output and ground truth is measured using the sentence transformer all-MiniLM-L6-v2 (Reimers and Gurevych,

Fine-tune	Model	Param.	R-1	R-2	R-L	BLEU	BScore	Sim	PPL
w/o Fine-tune	GPT	116M	11.79	0.47	8.35	0.08	82.20	0.121	-
	M2M-100	483M	2.10	0.16	1.89	3.20	75.66	0.088	176.36
	T5	60M	0.49	0.00	0.49	1.66	84.52	0.452	295.87
	Falcon	7B	10.59	0.76	7.49	0.19	82.07	0.19	106.02
	Mistral	7B	12.47	1.04	8.76	0.29	82.77	0.16	202.30
	BART-L	406M	18.10	3.76	13.00	1.81	84.94	0.458	188.65
w/ Fine-tune	GPT	116M	27.68	6.62	20.11	3.77	88.11	0.429	69.52
	M2M-100	483M	30.12	8.76	22.44	5.99	88.74	0.437	32.84
	T5	60M	30.28	10.04	**23.74**	5.5	88.76	0.480	25.40
	Falcon	7B	29.63	8.59	20.84	4.59	88.30	**0.522**	33.49
	Mistral	7B	27.96	**13.19**	23.66	2.64	88.63	0.496	27.61
	BART-L	406M	**33.93**	10.04	23.59	**7.00**	**88.98**	0.498	**24.11**

Table 3: **Self-Motivation results** Performance of models with and without fine-tuning on ***CASTIC*** dataset on ROUGE-1 (R-1), ROUGE-1 (R-2), ROUGE-L (R-L), BLEU, and BERTScore (BScore). Param., Sim, PPL indicates the number of parameters of each model, cosine similarity and perplexity, respectively.

Strategy	GPT	GPT-2	M2M-100	T5	BART-L
Physiological Needs	62.49	76.17	74.9	77.51	74.45
Safety Needs	72.23	75.26	73.31	74.64	71.94
Love and Belonging	72.83	76.97	76.2	76.67	65.95
Self-actualization	67	75.32	75.87	74.95	72.32
Esteem Needs	78.21	80.41	79.01	80.8	68.88
AVG.	70.55	76.83	75.86	**76.91**	70.71
STD.	6.01	2.12	**2.09**	2.48	3.32

Table 4: F1 score (%) of motivation strategies classification.

Model	w/ Fine-tune		w/o Fine-tune	
	Motivation	Fluency	Motivation	Fluency
GPT	2.3	3.57	1.17	2.03
M2M-100	2.32	3.85	1.00	1.00
BART-L	2.54	4.34	1.82	3.06
T5	3.03	4.18	2.12	2.47
Mistral	3.05	3.82	1.82	2.67
Falcon	2.99	4.35	2.03	2.10

Table 5: **LLM Evaluation results** The average rate of generated feedbacks on with and without fine-tuned models in the terms of how motivating and fluent the feedback is.

2019). Perplexity (Bengio et al., 2000) is measured using GPT-2 (Radford et al., 2019) from Hugging Face. It is worth noting that we discard the empty generation samples when measuring the scores.

Qualitative metric Following Chiang and Lee (2023), we use GPT-3.5 (Brown et al., 2020) to evaluate the effect of our dataset on inducing self-motivation. The prompt is illustrated in Appendix A.4. From the model's generated feedback, we randomly sample 100 sentences and ask GPT-3.5 how motivating (Motivation) and how fluent (Fluency) the feedback is. The rating scale is from 1-5 with 1 being the lowest.

3.3 Experimental Result

Overall Result In Table 3, we evaluate the result of the experiment. The models can learn each motivation strategy and generate feedback well. We illustrate the example of generated feedback in Table 7 of Appendix A.7. Overall, BART-L shows the best performance both in zero-shot and fine-tuning experiments.

LLM Evaluation Result We compare the score for models that are fine-tuned on the ***CASTIC*** dataset and pre-trained models without fine-tuning. In Table 5, we illustrate the self-motivation feedback resulting from GPT-3.5. The models fine-tuned with our dataset show better performance compared to the others. Specifically, the average motivation score for the fine-tuned models is 2.7, whereas models without fine-tuning achieve an average motivation score of 1.66. Additionally, in terms of fluency, the fine-tuned models attain an average score of 4, outperforming the zero-shot result. The result indicates that the model fine-tuned on our dataset generates fluent outputs.

Motivation Strategy Classification In Table 4, we evaluate F1 score per 5 motivation strategies. We add one additional linear layer that outputs 5 classes corresponding to each motivation type. The result shows that each model can learn and distinguish the characteristics between motivation strategies with over 70% F1 score for every strategy.

Limitation

We acknowledge that the severity score, which is determined based on the severity of the obstacle, can be subjective. However, just as people previously judged the negative and positive levels of words and annotated them to obtain negative scores in the field of Sentimental analysis, we present the standards by creating our own dataset and annotating it.

More significantly, we did not test the output of the trained models on human participants to determine, empirically, whether they induced greater levels of self-motivation.

Ethics Statement

In this work, we used POSREF (Ziems et al., 2022) which is a publicly available dataset. The creators of POSREF have already considered ethical issues when creating the dataset, but we additionally have manually checked every input sentence and filtered out inappropriate ones. We didn't find any obvious ethical concerns, such as violent or offensive content. We used the dataset consistent with the intended use. We used LLM in the process of creating and validating the dataset and we performed the verification of the output ourselves, meaning there were no issues with human annotators.

References

Yoshua Bengio, Réjean Ducharme, and Pascal Vincent. 2000. A neural probabilistic language model. *Advances in neural information processing systems*, 13.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Du Chen, Yi Huang, Xiaopu Li, Yongqiang Li, Yongqiang Liu, Haihui Pan, Leichao Xu, Dacheng Zhang, Zhipeng Zhang, and Kun Han. 2024. Orion-14b: Open-source multilingual large language models. *arXiv preprint arXiv:2401.12246*.

Jiangjie Chen, Wei Shi, Ziquan Fu, Sijie Cheng, Lei Li, and Yanghua Xiao. 2023. Say what you mean! large language models speak too positively about negative commonsense knowledge. *arXiv preprint arXiv:2305.05976*.

Cheng-Han Chiang and Hung-yi Lee. 2023. Can large language models be an alternative to human evaluations? *arXiv preprint arXiv:2305.01937*.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, et al. 2021. Beyond english-centric multilingual machine translation. *Journal of Machine Learning Research*, 22(107):1–48.

Eun Jung Kim and Kyeong Ryong Lee. 2019. Effects of an examiner's positive and negative feedback on self-assessment of skill performance, emotional response, and self-efficacy in korea: a quasi-experimental study. *BMC medical education*, 19:1–7.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Mounica Maddela, Megan Ung, Jing Xu, Andrea Madotto, Heather Foran, and Y-Lan Boureau. 2023. Training models to generate, recognize, and reframe unhelpful thoughts. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13641–13660, Toronto, Canada. Association for Computational Linguistics.

Abraham Harold Maslow. 1958. A dynamic theory of human motivation.

Lucille Njoo, Chan Park, Octavia Stappart, Marvin Thielk, Yi Chu, and Yulia Tsvetkov. 2023. TalkUp: Paving the way for understanding empowering language. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 9334–9354, Singapore. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks.

In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Ashish Sharma, Kevin Rushton, Inna Lin, David Wadden, Khendra Lucas, Adam Miner, Theresa Nguyen, and Tim Althoff. 2023. Cognitive reframing of negative thoughts through human-language model interaction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9977–10000, Toronto, Canada. Association for Computational Linguistics.

Benedikt Wisniewski, Klaus Zierer, and John Hattie. 2020. The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in psychology*, 10:487662.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Caleb Ziems, Minzhi Li, Anthony Zhang, and Diyi Yang. 2022. Inducing positive perspectives with text reframing. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3682–3700, Dublin, Ireland. Association for Computational Linguistics.

A Appendix

A.1 Implementation Detail

Motivation Result We train BART-L and M2M-100 with learning rate 1e-4, and batch size 32 for 5 epochs with a maximum sequence length of 128. For GPT and T5, we use learning rate 3e-5, and a weight decay of 0.01 for 5 epochs. For training and inference, we use a NVIDIA H100 80GB GPU. We run the experiment once for all the models.

Motivation Strategy Classification We train each model with learning rate 1e-4 and batch size 32 for 2 epochs with output class number as 5. For training and inference, we use a NVIDIA H100 80GB GPU. We run the experiment once for all the models.

A.2 Maslow's Motivation Theory

In this section, we provide detailed explanation of Maslow's Motivation Theory (Maslow, 1958)'s five needs.

Physiological Needs are desires to maintain a constant, normal state such as with respect to homeostasis, hormones, and nutrients. We applied this strategy for desiring rest, sleep, food, water, air and homeostasis.

Safety Needs are the desires for a safe environment such as financial and health security, stability in employment, protection from accidents, and a safe environment. The need derives from human's nature or preferring a safe, orderly, predictable, organized world and disfavoring unexpected, unmanageable, or other dangerous things.

Love & Belonging Needs are desires for love and affection from friends, lovers or spouses. Humans strive to have affectionate relations with people.

Esteem Needs are desires for stable, firmly based, high evaluation of themselves. Humans want to feel real respect from themselves and others for their capacities and achievements.

Self-actualization refers to the desire for self-fulfillment in actualizing a person's full potential. Humans want to become everything that one is capable of becoming.

A.3 Prompt for Dataset Generation

A.3.1 Obstacle and TODO Extraction

We extract *Obstacle* and *TODO* of the negative input sentence with Orion-14B-Chat using the prompt in Figure 3.

A.3.2 Generating Feedback

We generate feedback inducing self-motivation with Orion-14B-Chat using prompt in Figure 3. The type of feedback, positive/negative determined by severity score was applied in the feedback.

A.4 Prompt for LLM Evaluation

Motivation We evaluate how motivating the feedback generated by model trained in our dataset is with Chat-GPT using prompt in Figure 5 following (Chiang and Lee, 2023).

Fluency We evaluated how fluent the generated feedback by model trained in our dataset is using prompt in Figure 6.

A.5 Samples of CASTIC dataset

In Table 6, we illustrate the example of dataset.

A.6 Word frequency analysis

The word frequency figure illustrates the distribution of terms within our CASTIC dataset, providing key insights into the linguistic patterns and content characteristics of the corpus. As depicted in Figure 7,8,9,10, the x-axis represents individual words ordered by their frequency, while the y-axis shows the occurrence count without prompt words. The figure reveals several prominent trends: notably, common words such as 'goal', 'love', and 'focus' dominate the corpus, reflecting the dataset's

```
template = f"Question: \
    Given a statement: 'I just want a night at home! Is that too hard to
                              ask for?! Stop stressing me out! homework' \
    Tell what the person has as TODO and what is the obstacle that
    disturbing him from doing it in summary. \
    If there is no clear obstacle, answer 'None' for obstacle. \
    Answer TODO and Obstacle in a different line. \
    Answer: "
```

Figure 3: Prompt for Obstacle and TODO extraction

```
motivation_def = {
'physiological needs': "'Physiological needs' motivates a person when a person desire
for rest, sleep, homeostasis, maintaining a constant, normal state.",
'safety needs': "'Safety needs' motivates a person when he desires for financial and
health security, stability in employment, protection from accidents, and a safe
environment and prefers a safe, orderly, predictable, organized world.",
'love and belonging': "'Love and belonging needs' motivates a person when a person
feel keenly the absence of friends, lover or sprouse.",
'self-actualization': "'Self-actualization needs' motivates a person when he desires for
self-fulfillment, namely, to the tendency for him to become actualized in what he is
potentially.",
'esteem needs': "'Esteem needs' motivates a person when a person desire for feeling
real capacity, achievement, respect from others"}

for motiv_name, motiv in motivation_def.items():
        template = f"Question: \
        Obstacle: 'Hate exams', \
        TODO: 'Study math harder and be prepared for the math exam.', \
        {motiv} \
        Generate positive feedback to encourage him to do TODO from given obstacle
that blocking him following the meaning of '{motiv_name}'. \
        Please answer in 50 characters or less. Never include the word '{motiv_name}' in
the answer. "
```

Figure 4: Prompt for Generating Feedback

focus on emotional motivation. These findings underscore the dataset's thematic concentration and provide a foundation for understanding its linguistic structure and contextual relevance within the field of natural language processing. Moreover, comparative analyses with benchmark datasets reveal distinctive word usage patterns unique to our dataset, highlighting its potential contributions to advancing research in automated language understanding and generation.

A.7 Qualitative result

In table, we examine one original text for each of the motivation strategies in our dataset along with the BART model.

Please rate the Feedback based on Input.

The goal of this task is to rate Feedback based on Input.

NOTE: Please take the time to **fully read** and **understand** the story fragment. We will reject submissions from workers that are clearly spamming the task.

Input

I just want a night at home! Is that too hard to ask for?! Stop stressing me out! homework

Feedback

Take a night at home to recharge and focus on completing your homework. Your body and mind will thank you for the rest and relaxation.

Given Input from user, Please rate from 1 to 5 **how motivating** the following feedback is. (on a scale of 1-5, with 1 being the lowest?) .

Figure 5: Prompt for evaluating motivation with LLM

Please rate the Feedback based on Input.

The goal of this task is to rate Feedback.

NOTE: Please take the time to **fully read** and **understand** the story fragment. We will reject submissions from workers that are clearly spamming the task.

Feedback

Take a night at home to recharge and focus on completing your homework. Your body and mind will thank you for the rest and relaxation.

Please rate from 1 to 5 **how fluent** the text of the Feedback is. (on a scale of 1-5, with 1 being the lowest?)

Figure 6: Prompt for evaluating fluency with LLM

Input	Obstacle	TODO	Obstacle Type	Severity score	Motivation strategy	Feedback
Still home from school sick gonna pay big time when i get back should be doing homework all weekend on the plane :(	Sick	Take some rest and do homework all weekend on the plane.	Health	serious	physiological need	Take care of yourself by resting and doing your homework on the plane. Your body will thank you for it!
					safety needs	Take care of yourself, prioritize your health and well-being. Rest and study on the plane will help you feel recharged and ready to tackle any obstacles that come your way.
					love and belonging	Take care of yourself and focus on your studies. Your well-being is important for your future relationships.
					self-actualization	Take care of yourself and prioritize your well-being. Rest and study will help you achieve your goals and grow as a person.
					esteem needs	Rest and study will boost your health and increase your confidence, giving you the chance to show your true potential and achieve success.

Table 6: Examples of feedback in the dataset.

Input	Motivation strategy	Feedback
I already hate using computers so having to apply for a job online is not for me.	physiological need	You have the capability and determination to succeed in applying for a job online. Take it one step at a time and focus on what you need to accomplish. *Think about how much easier it will be to apply for a job onlinewithout having to leave your house! You can take care of business and still maintain a constant, normal state.*
	safety needs	You have the capability and determination to succeed in applying for a job online. Take it one step at a time and focus on what you can control. You can do it! *Don't let fear of computers hold you back. Apply for that job online and take control of your future.*
	love and belonging	Feeling lonely? Apply for a job online and make new connections. You'll find love and belonging, which are important for your overall well-being *Feeling lonely? Finding a job online could connect you with new friends and a sense of community.*
	self-actualization	You have the potential and the desire to succeed. Let's focus on applying for a job online and fulfilling your potential. *Don't let fear of computers hold you back. Take control and apply for that job online. You have the potential to succeed and this is your chance to show it.*
	esteem needs	You have the capability and determination to succeed. Let's focus on applying for the job online and show your skills. *Feeling capable and respected is important, but hating computers won't get you far. It's time to overcome your fear and apply for that job online.*

Table 7: Examples of generated feedback of BART in controlled setting. Prediction feedback corresponds to each motivation strategy is at the first row and the ground truth follows in italic.

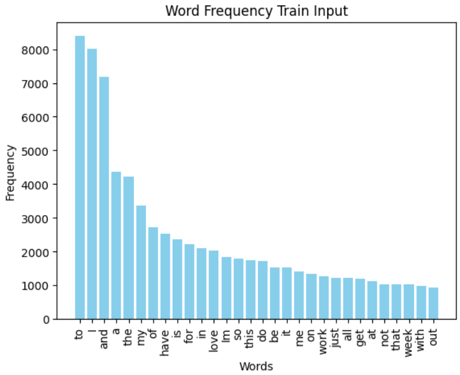

Figure 7: Word Frequeny Analysis For Train Dataset: Input

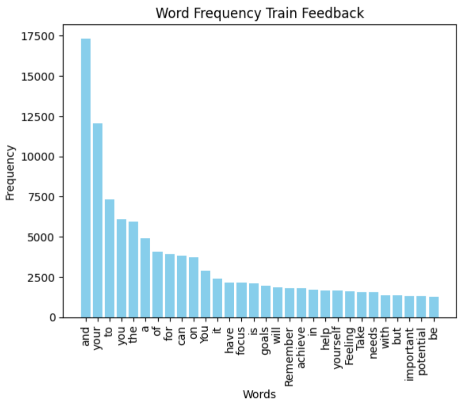

Figure 8: Word Frequeny Analysis For Train Dataset: Feedback

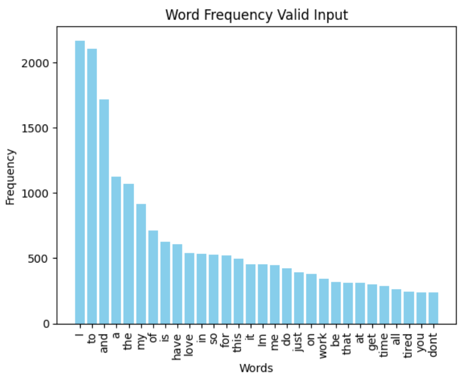

Figure 9: Word Frequeny Analysis For Valid Dataset: Input

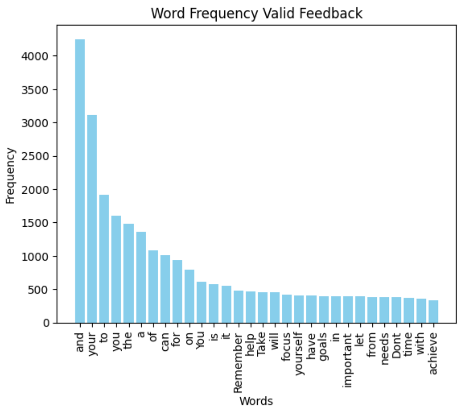

Figure 10: Word Frequeny Analysis For Valid Dataset: Feedback