

Diversity and Inclusion Summer 2023

Inducing Positive Perspectives with Text Reframing

Caleb Ziems, Minzhi Li

Akshat Gupta

“Gratitude is not only the gratest of virtues, but the parent of all others”

-Marcus Tullius Cicero

What is Positive Reframing?

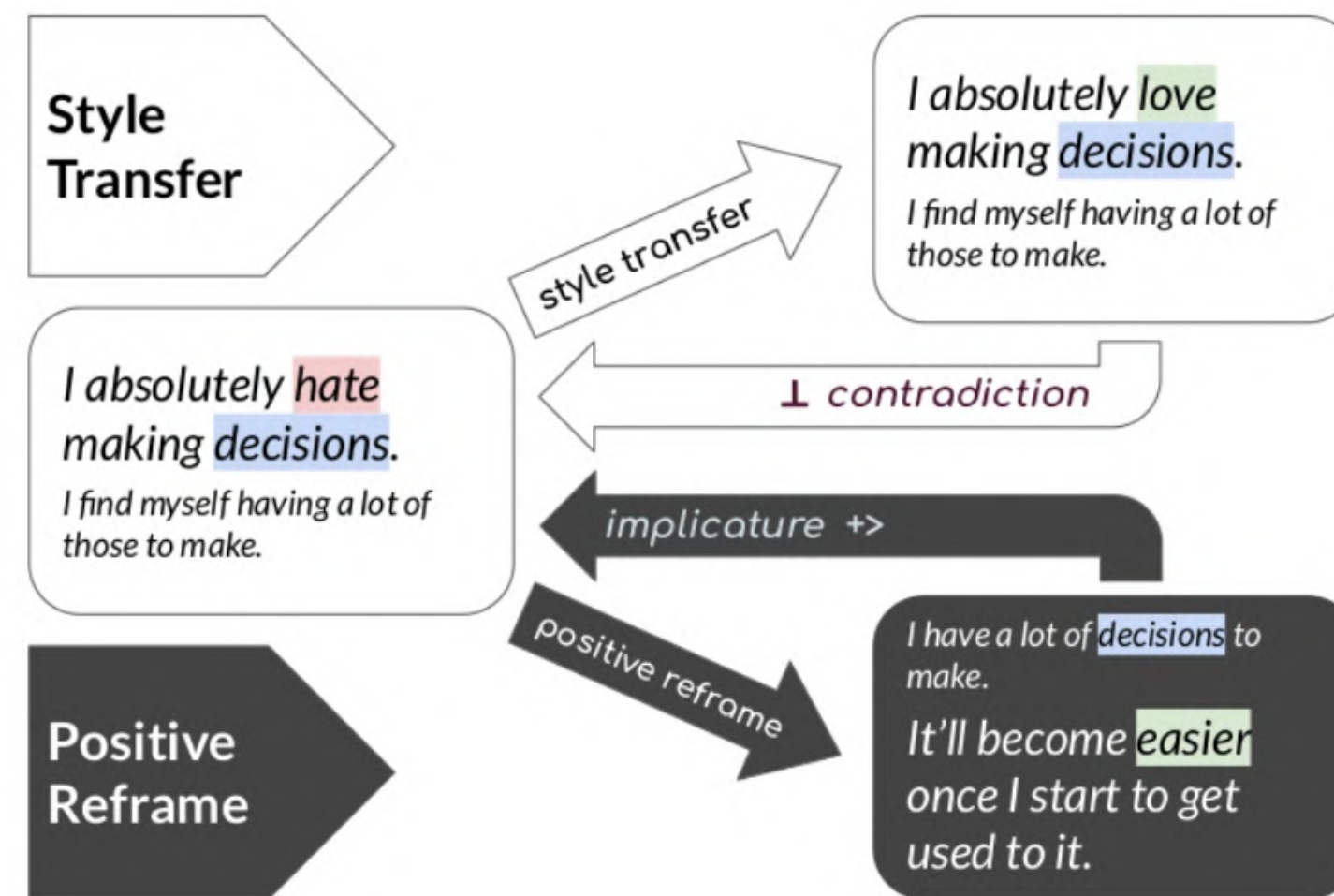
Reframe a negative text by inducing a complementary positive viewpoint.

Example

Original: This was a bland dish.

Reframed: I have made dishes that are much tastier than this one.

Positive Reframing vs Sentiment Style Transfer



Why Positive Reframing?

Studies show positive effects on:

- Emotional and psychological wellbeing
- Improved Cognitive performance
- Reduce group conflicts
- Beneficially shape customer relations
- Expressing optimism and gratitude reduces depressive symptoms

Positive reframing framework

- Not only merely swapping antonyms for negative words
- Inserting unfounded positive language
- Reconstruct cognitive distortions based on positive psychology

Growth Mindset

Example

Original: " I'm such a lazy procrastinator "

Reframed: " I'm determined to learn better time management "

Impermanence

Example

Original: " Since I failed this test, I must be too stupid for school."

Reframed: " This wasn't the test score I hoped for, but everyone slips now and then."

Neutralizing

Example

Original: "Wendy's customer service is terrible. "

Reframed: " Wendy's customer service could use some improvement. "

Self-affirmation

Example

Original: "Sometimes I get these impulses to just throw a tantrum. Like throw/break things, cry and scream"

Reframed: "It's normal for to feel overwhelmed sometimes but I know I am strong to handle and go through it"

Optimism

Example

Original: " I've completely worked myself to the bone this week, burning the candle at both ends. "

Reframed: " It's been a long week, but now I can kick back, relax, and enjoy my favourite shows because it's the weekend. "

Thankfulness

Example

Original: " Is it sad that I don't wanna be at home and wish that work could call me in early? "

Reframed: " I am thankful that I have a job that makes me want to get out of bed every day. "

Dataset

- Sourced from Twitter API using hashtag #stressed
- #stressed is chosen to signal anxiety, negative affect, and hopelessness
- Samples were of 280 characters in length
- Data comes from 1 million tweets written between 2012 and 2021
- Replies and retweets and tweets of less than 30 characters are excluded
- Overly positive and hate speech tweets are also excluded

Annotation

- 204 crowd workers reframed 8687 texts with two workers assigned to each task.
- Annotators decide strategy for annotation independently or in combination.
- The rating was given on a scale of 1-5 about the positiveness of the original tweet.

Data Quality

- To ensure good quality, Inter annotator agreement scores were used.
- Intra-class correlation for one-way random effects between the three raters and find moderate inter-rater agreement across these categories(min 0.32, max 0.68)
- The average genuineness score is 4 out of 5

Dataset Overview

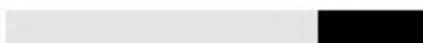

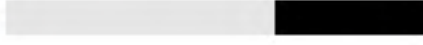


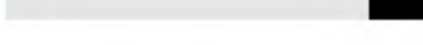
Label Distribution	Count	Label	Description	ICC	Gen
25.4% 	2,120	Growth Mindset	Viewing a challenging event as an opportunity for the author specifically to grow or improve themselves.	0.59	3.77
19.5% 	1,625	Impermanence	Saying bad things don't last forever, will get better soon, and/or that others have experienced similar struggles.	0.60	4.03
36.1% 	3,015	Neutralizing	Replacing a negative word with a neutral word. For example, "This was a terrible day" becomes "This was a long day."	0.32	3.53
48.7% 	4,069	Optimism	Focusing on things about the situation itself, in that moment, that are good (not just forecasting a better future).	0.44	3.89
10.1% 	841	Self-affirmation	Talking about what strengths the author already has, or the values they admire, like love, courage, perseverance, etc.	0.42	3.75
13.0% 	1,085	Thankfulness	Expressing thankfulness or gratitude with key words like appreciate, glad that, thankful for, good thing, etc.	0.68	3.95

Table 1: Summary statistics for POSITIVE PSYCHOLOGY FRAMES. (Left) Distribution of the non-exclusive labels across all 8,349 annotations shows a preference for *optimism* and *neutralizing* strategies. (Right) The quality of annotations is shown by moderate Intra-class Correlation (ICC), with reasonable *genuineness* (Gen) metrics for 100 randomly sampled datapoints.

Task formulation

- In positive reframing, the problem is conditional generation
- Maximize the standard language modelling objective over the string

$$\frac{1}{N} \sum_{i=0}^N \log p(g_i | g_{0:i-1})$$

Task formulation

- g_i is a string of length N
- $\langle \text{BOS} \rangle$ is the start token, to tokenized source $S_{1:n}$
- $\langle \text{STRG} \rangle$ is a strategy (grow, impermanence, etc)
- $\langle \text{REFR} \rangle$ is tokenized reframed target ends with $\langle \text{EOS} \rangle$

$$\begin{aligned} \mathbf{g} &= \{s, \psi_t, t\} \\ &= \{\langle \text{BOS} \rangle, s_1, s_2, \dots, s_n, \\ &\quad \langle \text{STRG} \rangle, \psi_{\text{grow}}, \psi_{\text{imp}}, \dots, \psi_{\text{thank}}, \\ &\quad \langle \text{REFR} \rangle, t_1, t_2, \dots, t_m, \langle \text{EOS} \rangle\} \end{aligned}$$

Task formulation

- In decoding, three settings were used:
 - Unconstrained Generation
 - Controlled Generation
 - Strategy Prediction(a form of Generation)
- Controlled generation is conditioned on desired strategies
- In prediction, the model will also predict the strategies it used to generate its frame
- In an unconstrained setting, reframing is done without any strategy

Experimental Setup

- The dataset is partitioned in the ratio of 8:1:1
- Train: 6679, Dev: 835, Test:835
- Fine-tuning is done on GPT, GPT-2, LSTM, CopyNMT, BART, and T5.
- For all models, greedy decoding is used.

Experimental Setup

- Retrieval
 - Random Retrieval
 - SBERT Retrieval
- Few-shot Learning
 - GPT3
 - GPT-Neo

Evaluation

- Semantic Similarity using BLEU, ROUGE, and BERTScore
- The average of a maximum of two scores is taken
- Average change in sentiment according to text blob is reported
- In Human Evaluation, 50 items were distributed to 3 evaluators who score the reframed sentences for 3 criteria on a scale of 1-5
- Criteria are Meaning preservation, Positivity and Fluency of generated text

Results

		Automatic Evaluation							Human Evaluation			
	Model	R-1	R-2	R-L	BLEU	BScore	Δ TB	Avg. Len	Meaning	Positivity	Fluency	
Retrieval	Random	9.6	3.6	8.4	0.17	84.8	0.36	20.0	2.79	3.03	3.60	
	SBERT	15.2	1.9	12.8	1.47	87.6	0.36	17.7	3.45	3.97	4.16	
Few-shot	GPT-3	18.3	3.4	15.5	2.9	88.2	0.44	17.3	3.73	4.17	4.27	
	GPT-Neo	18.7	3.4	16.0	3.0	88.2	0.40	17.6	3.69	4.16	4.21	
Unconstrained	$p(t s)$	GPT	13.3	1.8	11.3	1.1	86.4	0.37	21.1	3.55	3.91	4.08
		GPT-2 No-pretrain	13.2	1.3	11.4	0.66	89.6	0.37	16.9	3.11	3.66	3.96
		GPT-2	20.9	4.6	17.7	4.2	88.5	0.35	20.0	3.58	4.01	4.18
		Seq2Seq-LSTM	15.7	1.4	12.4	0.73	85.6	0.49	25.8	3.33	4.15	4.10
		CopyNMT	20.8	5.0	18.0	4.0	85.7	0.32	16.1	3.57	3.69	3.91
		T5	27.4	9.8	23.8	8.7	88.7	0.38	35.3	4.09	3.79	4.06
		BART	27.7	10.8	24.3	10.3	89.3	0.23	24.4	4.13	3.81	4.15
Predict	$p(t, \psi_t s)$	T5	27.5	10.5	24.0	11.0	89.0	0.23	25.1	4.10	3.64	4.11
		BART	27.3	10.2	24.1	9.85	89.4	0.32	23.4	4.09	3.95	4.11
Control	$p(t s, \psi_t)$	T5	27.7	10.0	23.9	8.8	88.8	0.36	35.0	4.11	3.89	4.07
		BART	28.8	10.9	25.1	10.1	89.6	0.27	24.7	4.23	4.07	4.27
Human		100	100	100	100	100	0.35	17.4	3.80	3.82	4.18	

Table 2: **Positive reframing results** measured by **Meaning** including **ROUGE-1 (R-1)**, **ROUGE-1 (R-2)**, **ROUGE-L (R-L)**, **BLEU**, **BERTScore (BScore)**, **Positivity** via **Δ TextBlob (Δ TB)** and **Fluency**. State-of-the-art models can generate meaning-preserving reframes in the unconstrained setting $p(t|s)$ and strategy-predictive setting $p(t, \psi_t|s)$ as well as when we condition the generation to use the reframing strategy from the ground truth $p(t|s, \psi_t)$. The best in-category performance is **bolded**; best overall performance is **highlighted**.

Results: Automatic Evaluation

- The unconstrained BART model gives the highest quality of positive reframes
- GPT and no-pre-train GPT2 provided the worst quality
- **Reason:** Due to pretrained on different contextual text
- In a controlled generation and strategy prediction, the two best-performing models are BART and T5
- Controlled generation improves performance
- All supervised models outperformed the simple retrieval baseline
- Few shot GPT-3, GPT-Neo could not match supervised models

Results: Human Evaluation

- For BART and T5 models, average scores are very high
- BART and T5 model surpassing the human gold standard in all unconstrained, predictive, and controlled settings
- The controlled BART model scored 4.07 in positivity and 4.27 in Fluency with meaning preservation
- Few shot systems fail to preserve the meaning of the original sentence despite having 4.17 in positivity and 4.27 in fluency

Results: Qualitative Investigation

Original Text	Strategy	Positive Reframe
(a) Could someone just give me like \$1000? It would change my life, stressing about rent, bills and food money is just the worst.	growth	I need to learn how to manage my money better so that I don't have to stress about rent, bills and food money. <i>Time to focus on making a budget, so I don't have to stress about rent, bills, and food money.</i>
(b) I just went back to school today And I'm already stressed cause we have MIDTERMS NEXT WEEK !!! AND THIS WEEKEND – JAM PACKED :-((impermanence	I just went back to school today and I'm already stressed because we have midterms next week and this weekend – JAM PACKED <i>Well, just think by the end of next week we'll be all done with the midterms!</i>
(c) talacre beach really need to sort there checking in routeen caravans are never ready 30mins later and still sat in the car waiting ..	neutralizing	The caravans at Talacre beach need to sort out their check-in process, so the kids don't get stuck in the car. <i>Talacre beach's check in process could do with some improvement to reduce the wait time for customers.</i>
(d) So glad that tomorrow is Friday. This has seriously been the longest week of my life! tired	optimism	I'm glad that tomorrow is Friday. This week has been long, but I'm looking forward to the weekend. <i>I'm glad the weekend is coming up, so I can rest.</i>
(e) Sometimes I get these impulses to just throw a tantrum. Like throw/break things, cry and scream.	self-affirmation	Sometimes I get these impulses to just throw a tantrum. Like throw/break things, cry and scream. But I'm strong, and I know I can handle it. <i>It's normal for to feel overwhelmed sometimes but I know I am strong to handle and go through it.</i>
(f) The fact that my mom's been calling me every night to calm me down from school says so much. I've needed it these past few days	thankfulness	I'm really lucky to have such a caring mum who is willing to call me every night when I'm having a hard time. <i>My mom has been calling me every night to calm me down from school. I've needed it these past few days. I'm thankful for her.</i>

Table 3: Example reframes. We examine one original text for each of the strategies in our dataset along with its BART-generated positive reframe (first row of each block) and its ground truth reframe (*italics* in second row of each block). Even when the generations differ from the ground truth, the model's reframes are largely successful at shifting the perspective while still maintaining the underlying meaning of the original text.

Results: Qualitative Investigation

- For the best controlled BART model Qualitative investigation is done
- The model does positive reframing without losing the underlying meaning
- Transformer-based models appear to be capable of solving this with success

Error Analysis

- For 100 randomly sampled data points from best controlled BART model
- Errors are divided into four classes
- **Insubstantial changes(26%):** The model only swap a few negative words
 - Eg. I hate it→ I don't like it.
- **Contradiction to the premise(9%):** Failure of meaning preservation
 - Eg. Feel like crying, this math class is impossible to pass → This math class is hard, but I know I can pass it.

Error Analysis

- **Self-contradictions(6%):** Contradicting the meaning
 - Eg. I don't like opening up to people, but I'm glad I have the courage to do it.
- **Hallucinations(2%):** with unmotivated perspectives
 - Eg. mentioning a *good night's sleep when the original text is about something else*.

Frame Strategy Classification

- It is multilabel multiclass classification problem
- Experimentation with BERT, RoBERTA, and XLNet
- All models can learn classification with $F1 > 0.60$ except XLNet
- The impermanence and self-affirmation strategies are more challenging

Strategy	BERT	RoBERTA	XLNet	Support
Thankfulness	0.71	0.69	0.71	109
Neutralizing	0.59	0.60	0.49	302
Optimism	0.72	0.71	0.72	400
Impermanence	0.55	0.55	0.54	157
Growth	0.61	0.63	0.67	221
Self Affirmation	0.43	0.44	0.39	76

Table 4: Strategy classification F1 scores

Conclusion

- Supervised learning with transformer models works best among all
- But the best models are still prone to errors
- Qualitative analysis is done on 100 samples which are very small to conclude anything concrete in terms of errors
- Previous approaches lack meaning preservation which is achieved with reasonable success in this study

Thank you
Questions?

Model Comparison(Appendix)

Original Text		Always stressing and thinking about loads of things at once need I take it one at a time overload stressed need to rant.
Ground Truth		Loads of things on my mind, I need to make a list, prioritise and work through it all calmly and I will feel much better.
Model Generation	Random	Annoyed because the USPS lost 1 of the 8 flat rate boxes of books I shipped from California to Kentucky. Hope it shows up :-/
	SBERT	Really stressed at the moment but I know that this will pass and I'll get everything sorted out.
	Seq2seq-LSTM	i have a lot of things to do it is a good night's sleep and i can get through it.
	CopyNMT	Always stressing and thinking easy of things of things at things at all, but I know that I need it.
	GPT	i ' m so relieved that i ' m going to get to organize my thoughts in such a short time ! i hope i am done with the list of things i have to do today , so that i can go back to bed .
	GPT-2 No-pretrain	I'm going to try to seek a new job to get some newspect me through today.
	GPT-2	I should be more energetic in my thoughts, and not worry too much.
	T5	I am stressed and thinking about loads of things at once, I need to take it one at a time stressed need to rant. I'm sure I'll be able to focus on it.
	BART	I am stressed and thinking about loads of things at once, but I am going to take it one at a time.

Table 5: **A model comparison** for reframing the same text. Here, we observe significant diversity in the generations, suggesting that these different architectures could serve as complementary systems in an effort to restruture and transform negative text on the web.

Model Comparison(Appendix)

Original Text		I'm doing really well in some class and bad in other classes I hate this.
GPT-2 Generations	Growth Mindset	I'm doing really well in some class and bad in other classes I hate this. But I'm doing better than I was at the age of 11.
	Impermanence	I'm having a hard time in some classes and I hope to get through the rest of the semester.
	Neutralizing	I'm doing really well in some class and bad in other classes I hate this. But I'm going to go clear it out.
	Optimism	I need to take a good hard look at myself and realize that I am quite capable of doing better in other classes.
	Self-affirmation	I'm really happy about my class, I'm doing really well in some class, bad or not, I can do better in another.
	Thankfulness	I'm so lucky that I'm able to get my classes to a good start and get off the couch.
T5 Generations	Growth Mindset	I'm doing really well in some class and bad in other classes I hate this, but I'll be able to do it again.
	Impermanence	I'm doing really well in some class and bad in other classes I hate this. But I'll be able to do my best in the future.
	Neutralizing	I'm doing really well in some class and bad in other classes I hate this. But I can't wait to see if I can do it.
	Optimism	I'm doing really well in some class and bad in other classes I hate this. I hope I'll be able to do better in the future.
	Self-affirmation	I'm doing really well in some class and bad in other classes I hate this. It's a good thing to do, but I'll be able to do it.
	Thankfulness	I'm doing really well in some class and bad in other classes I hate this. But I can't wait to see if I can do it.
BART Generations	Growth Mindset	I'm doing really well in some class and bad in other classes. I'm going to try to improve my grades.
	Impermanence	I'm doing really well in some class and bad in other classes, but I'm sure it will all work out in the end.
	Neutralizing	I'm doing really well in some class and bad in other classes I don't like this.
	Optimism	I'm doing really well in some class and bad in other classes, but I'm sure it will all work out.
	Self-affirmation	I'm doing really well in some class and bad in other classes, but I know I can do better.
	Thankfulness	I'm doing really well in some class and bad in other classes, but I'm thankful that I have the opportunity to study.

Table 6: **A model comparison** for reframing the same text using different controlling strategy. Here, we observe models can learn some information from the input strategy label and make distinctive generations, especially for the ‘thankfulness’, ‘growth’ and ‘self-affirmation’ strategies.