

# PSentScore: Evaluating Sentiment Polarity in Dialogue Summarization

Yongxin Zhou, Fabien Ringeval, François Portet  
Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000  
Grenoble, France

LREC-COLING 2024, Torino (Italia), 20-25 May, 2024  
Contact: [yongxin.zhou@univ-grenoble-alpes.fr](mailto:yongxin.zhou@univ-grenoble-alpes.fr)  
Paper: <https://arxiv.org/abs/2307.12371>

# Introduction (1/4)

## Motivation

- **Dialogue Summarization**: distilling the most **crucial information** from human conversations into **concise summaries**
- Mostly focusing on **summarizing factual information**, neglecting the affective content
- **Sentiments** are however crucial for the analysis of **interactive data**, especially in **healthcare** and **customer service**

# Introduction (2/4)

Name	Domain	Language	Guideline Available	Guideline criteria for writing the reference summaries
AMI (McCowan et al. 2005)	Meeting	English	Yes	Abstractive summaries should have the following structure: abstract, decisions, problems/issues, actions. Extractive summaries: identify extracts from the transcript which jointly convey the correct kind of information about the meeting to fit the required purpose. The instructions do not mention emotion/sentiment.
RATP-DECODA (Favre et al. 2015)	Telephone Customer Service	French	No	We contacted the authors and obtained their summary definition, there is no mention of emotion/sentiment.
SAMSum (Gliwa et al. 2019)	Chat	English	Yes	(1) Be rather short, (2) extract important pieces of information, (3) include names of interlocutors, (4) be written in the third person. The instructions do not mention emotion/sentiment.
MEDIASUM (Zhu et al. 2021)	Media Interview	English	No	The reference summaries were downloaded from text descriptions of the input documents (interviews) available on the web.
TWEETSUMM (Feigenblat et al. 2021)	Customer Service	English	Yes	Extractive summary: highlight the most salient sentences in the dialog. Abstractive summaries: one sentence summarizing what the customer conveyed and a second sentence summarizing what the agent responded. The instructions do not mention emotion/sentiment.
QMSum (Zhong et al. 2021)	Multi-domain Meeting	English	Yes	The annotation process consists of three stages: topic segmentation, query generation, and query-based summarization. The instructions do not mention emotion/sentiment.
CSDS (Lin et al. 2021)	Customer Service	Chinese	No	There are three different summaries for each dialogue: an overall summary and two role-oriented summaries (user and agent). Emotion/sentiment is not mentioned.
DIALOGSUM (Chen, Y. Liu, et al. 2021)	Spoken	English	Yes	Convey the most salient information; Be brief; Preserve important named entities within the conversation; Be written from an observer perspective; Be written in formal language. Pay extra attention to the following aspects: Tense Consistency, Discourse Relation, <b>Emotion</b> and Intent Identification.

**Table:** Major datasets for dialogue summarization with their summaries criteria. DialogSum is the only one to include *Emotion* in the guideline.

# Introduction (3/4)

## Example from DialogSum (test\_440)

**Dialogue:** "#Person1#: Man, I'm freaking out! You gotta help me! #Person2#: Whoa, whoa, take it easy, relax. Geez, you're sweating like a pig! What's going on?"

#Person1#: I can't go through with this! I just can't! I'm not ready for marriage! What was I thinking? I'm only thirty five years old! I've got my entire life ahead of me, adventures waiting! I can't settle down yet!

#Person2#: What are you talking about? It wasn't more than a month ago that you were rambling on about how you are tired of living the life of a bachelor and how you envy your friends that have a family!

#Person1#: I know. I know!

#Person2#: Let's think this through. First of all, you cannot leave Amy at the altar. Not only will she be humiliated and kill you, but she is the woman that you love and the woman of your dreams! Why would you want to end that or jeopardize it like this? Second of all, you are just getting cold feet. You know deep down inside that you want to marry her, so cut the crap and do it!

#Person1#: You're right. I'm being crazy! Ok, I'm ready, let's do this!

#Person2#: Great! Phew! That was a close one! You had me going there for a minute I thought I was gonna have to slap some sense into you. "

---

**summary1:** "#Person1# is not ready for marriage. #Person2# reminds him of what he said a month ago and what will happen if he leaves Amy. After listening to #Person2#, #Person1# is ready for marriage." ⇒ **absence of affective information**

---

**summary2:** "#Person1# feels anxious about future marriage and sweats a lot. #Person2# helps #Person1# to think through the current situation and get ready for the marriage." ⇒ **presence of affective information**

---

**summary3:** "#Person1# is freaking out before his marriage with Amy and #Person2# helps him get over it by helping him think through all the details." ⇒ **presence of affective information**

# Introduction (4/4)

## Hypothesis

We hypothesize that sentiments present in a dialogue should be preserved in its summary, both in terms of **proportion** (factual vs. affective) and sentiment **polarity** (global vs. local)

# Introduction (4/4)

## Hypothesis

We hypothesize that sentiments present in a dialogue should be preserved in its summary, both in terms of **proportion** (factual vs. affective) and sentiment **polarity** (global vs. local)

## Contribution

We introduce and assess a set of measures **PSentScore** aimed at quantifying the preservation of affective content in dialogue summaries

# Method (1/4)

## PSent

The score calculates the proportion of affect charged words (positive/negative) in a given text:

$$PSent = (PosN + NegN) / TotalN \quad (1)$$

- $PSentDial$  and  $PSentSumm$ : represent the  $PSent$  in the input dialogue, and reference summaries (or generated summaries)
- We can also compute  $PSent_P$  (resp.  $PSent_N$ ) to denote the proportion of affect charged positive words –  $PosN$  – only (resp. negative words –  $NegN$  – only) in the given texts

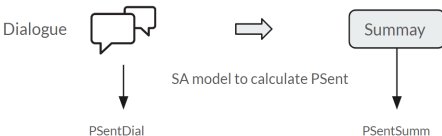
- I am six feet tall. → 0 sentiment
- I **love** being there. → 1/4 positive
- I feel **bad** this morning. → 1/5 negative

# Method (2/4)

## PSentScore

**PSentScore** evaluates how much affective content is preserved in summaries → examines the relationship strength between *PSentDial* and *PSentSumm* using different metrics:

- **Spearman's rank correlation coefficient**: monotonic relationships between two variables (↑)
- **Concordance Correlation Coefficient (CCC)**: strength of the relation between two variables (↑)
- **Mean Absolute Error (MAE)**: average bias between two variables (↓)



$$\Rightarrow \text{PSentScore}(\{ \text{PSentDial}_i, \dots, \text{PSentDial}_N \}, \{ \text{PSentSumm}_i, \dots, \text{PSentSumm}_N \})$$



# Method (3/4)

## Training word-level sentiment analysis models

**Corpus:** Stanford Sentiment Treebank (**SST**), fully labeled parse tree

- 215,154 phrases in the parse tree of 11,855 sentences
- Mapped annotations to a 3-point Likert scale (positive/neutral/negative) – SST3

## Word-level Sentiment Analysis (SA)

- **token-dict.:** dictionary-based classifier – opinion\_lexicon (Hu and B. Liu 2004), with a list of  $\approx 7k$  positive and negative words in English
- **BERT-SST3:** fine-tune BERT model using SST3
- **BERT-DS-SST3:** 1) Run token-dict. to annotate DialogSum training partition (12,460 dialogues,  $\approx 120,744$  turns); 2) Fine-tune BERT on DialogSum; 3) Further fine-tune this model on the training set of SST3 (8544 samples)

# Method (4/4)

## SA models' results

	overall_accuracy	precision	recall	f1
token-dict.	88.82	73.61	60.96	65.64
BERT-SST3	97.87 ( $\pm 0.06$ )	94.43 ( $\pm 0.43$ )	94.07 ( $\pm 0.38$ )	94.24 ( $\pm 0.15$ )
BERT-DS-SST3	97.96 ( $\pm 0.04$ )	94.53 ( $\pm 0.17$ )	94.39 ( $\pm 0.20$ )	94.46 ( $\pm 0.10$ )

**Table:** Performances in terms of accuracy, precision, recall, f1 (%) on the test set of the SST-3 dataset, for different models: `token-dict.`, `BERT-SST3` and `BERT-DS-SST3`. Statistics are given in the following format: mean (standard deviation), based on three runs. Macro results for precision, recall and f1.

- Domain adaptation stabilized its performance with a lower standard deviation than that of `BERT-SST3`
- ⇒ We chose `BERT-DS-SST3` for future use

# Experiments (1/2)

Dataset: DialogSum (Chen, Y. Liu, et al. 2021)

- Annotation guidelines including Emotions
- Used in a challenge

## Filtering Methodology

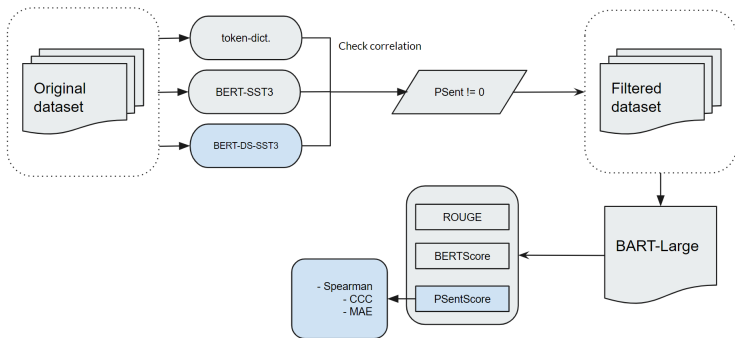
Training target selected by eliminating pairs without affective content in the dialogue or the summary

Part.	Full	Filtered by <b>BERT-DS-SST3</b>	w/ zero value	
			Dial.	Sum.
train	12460	9687 (77.5%)	43	2757
dev	500	391 (78.2%)	2	108
test	500	499 (99.8%)	1	0

**Table:** Statistics for DialogSum dataset. For the Filtered corpus both input dialogue and reference summaries without affective content according to BERT-DS-SST3 were removed. However, for the test partition only the input dialogues were filtered out. Percentage of data kept is shown in parentheses (%).

⇒ Corpus construction: Filtered dataset, Random sampled

# Experiments (2/2)



## Experimental Setup

- Fine-tune BART-Large
- 1) Full set, 2) Random sampled, 3) Filtered dataset

# Results

Model	R1	R2	RL	BERTScore	PSentScore	PSentScore <sub>P</sub>	PSentScore <sub>N</sub>
# samples	500*	500*	500*	500*	499†	491†	419†
Human (Chen, Deng, et al. 2022)	53.35	26.72	50.84	92.63	-	-	-
BART <sub>Large</sub> (Chen, Y. Liu, et al. 2021)	47.28	21.18	44.83	-	-	-	-
GoodBai (Chen, Deng, et al. 2022)	<b>47.61</b>	<b>21.66</b>	45.48	<b>92.72</b>	.357/.289/.027	.341/.307/.024	.397/.358/.014
UoT (Lundberg, Sánchez Viñuela, and Biales 2022)	47.29	21.65	<b>45.92</b>	92.26	.356/.297/.027	.364/.325/.023	.383/.338/.014
IITP-CUNI (Bhattacharjee et al. 2022)	47.26	21.18	45.17	92.70	.348/.289/.031	.311/.280/.027	.397/.295/.018
TCS_WITM (Chauhan et al. 2022)	47.02	21.20	44.90	90.13	.364/.294/.028	<b>.375</b> /.331/.024	.431/.333/.014
baseline-BART <sub>Large</sub>	47.36	21.23	44.88	91.42	.353/.292/.029	.318/.295/.025	.395/.322/.016
baseline_sub-sampled◊	46.94	20.52	44.43	91.29	.351/.294/.028	.351/.319/.024	.410/.352/.016
baseline_Filtered◊	45.78	19.69	43.21	90.83	<b>.435</b> /. <b>348</b> /. <b>027</b>	.370/. <b>352</b> /. <b>023</b>	<b>.449</b> /. <b>373</b> /.015

**Table:** Comparison of results from the DialogSum challenge teams and our BART-Large models fine-tuned on the full (baseline-BART<sub>Large</sub>) and filtered corpus (baseline\_Filtered) of the DialogSum dataset. The baseline\_sub-sampled model has been trained on a corpus of the same size as the Filtered dataset but whose instances have been randomly sampled from the the full DialogSum dataset. ◊ indicates training on partial corpora: training set 9687 (77.5%), dev set 391 (78.2%). *PSentScore* values indicate evaluation results: Spearman (↑) / CCC (↑) / MAE (↓). \* refers to the Full DialogSum test set, and † to the Filtered test set.

- SOTA models: do **not preserve well** affective content within summaries

# Results

Model	R1	R2	RL	BERTScore	PSentScore	PSentScore <sub>P</sub>	PSentScore <sub>N</sub>
# samples	500*	500*	500*	500*	499†	491†	419†
Human (Chen, Deng, et al. 2022)	53.35	26.72	50.84	92.63	-	-	-
BART <sub>Large</sub> (Chen, Y. Liu, et al. 2021)	47.28	21.18	44.83	-	-	-	-
GoodBai (Chen, Deng, et al. 2022)	<b>47.61</b>	<b>21.66</b>	45.48	<b>92.72</b>	.357/.289/.027	.341/.307/.024	.397/.358/.014
UoT (Lundberg, Sánchez Viñuela, and Biales 2022)	47.29	21.65	<b>45.92</b>	92.26	.356/.297/.027	.364/.325/.023	.383/.338/.014
IITP-CUNI (Bhattacharjee et al. 2022)	47.26	21.18	45.17	92.70	.348/.289/.031	.311/.280/.027	.397/.295/.018
TCS_WITM (Chauhan et al. 2022)	47.02	21.20	44.90	90.13	.364/.294/.028	<b>.375</b> /.331/.024	.431/.333/.014
baseline-BART <sub>Large</sub>	47.36	21.23	44.88	91.42	.353/.292/.029	.318/.295/.025	.395/.322/.016
baseline_sub-sampled <sup>◇</sup>	46.94	20.52	44.43	91.29	.351/.294/.028	.351/.319/.024	.410/.352/.016
baseline_Filtered <sup>◇</sup>	45.78	19.69	43.21	90.83	<b>.435</b> /. <b>.348</b> /. <b>.027</b>	.370/. <b>.352</b> /. <b>.023</b>	<b>.449</b> /. <b>.373</b> /.015

**Table:** Comparison of results from the DialogSum challenge teams and our BART-Large models fine-tuned on the full (baseline-BART<sub>Large</sub>) and filtered corpus (baseline\_Filtered) of the DialogSum dataset. The baseline\_sub-sampled model has been trained on a corpus of the same size as the Filtered dataset but whose instances have been randomly sampled from the the full DialogSum dataset. <sup>◇</sup> indicates training on partial corpora: training set 9687 (77.5%), dev set 391 (78.2%). PSentScore values indicate evaluation results: Spearman (↑) / CCC (↑) / MAE (↓). \* refers to the Full DialogSum test set, and † to the Filtered test set.

Careful selection of the training set:

- **Improved preservation** of affective content in the generated summaries
- **Minor reduction** in content-related metrics

# Discussion and Conclusion

## Contributions

- We emphasize the importance of taking affective content into account in dialogue summarization, in addition to factual information
- Proposed PSent and trained SA models → PSentScore, a reference-less metric
- Results: Summarization models often exhibit a **mismatch** between the affective content of the input dialogue and the summary, carefully selecting the training target can decrease this mismatch

# Limitations and future work

## Limitation

- PSent is still gross and focuses on proportion and polarity perspectives, and was evaluated only on one data set
- Our measure relies on a word-level sentiment analysis model which is biased and might not be available for all languages
- While the BERT-DS-SST3 model demonstrated promising performance on the SST corpus, it has not been evaluated on dialogue corpora



# Limitations and future work

## Limitation

- PSent is still gross and focuses on proportion and polarity perspectives, and was evaluated only on one data set
- Our measure relies on a word-level sentiment analysis model which is biased and might not be available for all languages
- While the BERT-DS-SST3 model demonstrated promising performance on the SST corpus, it has not been evaluated on dialogue corpora

## Future work

- Extend our method and conduct large-scale analysis on more dialogue summarization datasets and with more **fine-grained affective categories**
- Extend the use of PSentScore to other tasks, such as summarizing reviews/opinions, generating emotional dialogues, etc

# References I

- ▶ Hu, Minqing and Bing Liu (2004). "Mining and Summarizing Customer Reviews". In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '04. Seattle, WA, USA: Association for Computing Machinery, pp. 168–177. ISBN: 1581138881. DOI: 10.1145/1014052.1014073. URL: <https://doi.org/10.1145/1014052.1014073>.
- ▶ McCowan, I. et al. (Aug. 2005). "The AMI meeting corpus". Undefined. In: *Proceedings of Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research*. Ed. by L.P.J.J. Noldus et al. Noldus Information Technology, pp. 137–140. ISBN: 90-74821-71-5.
- ▶ Favre, Benoit et al. (Sept. 2015). "Call Centre Conversation Summarization: A Pilot Task at Multiling 2015". In: *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Ed. by Alexander Koller et al. Prague, Czech Republic: Association for Computational Linguistics, pp. 232–236. DOI: 10.18653/v1/W15-4633. URL: <https://aclanthology.org/W15-4633>.
- ▶ Gliwa, Bogdan et al. (Nov. 2019). "SAMSum Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization". In: *Proceedings of the 2nd Workshop on New Frontiers in Summarization*. Ed. by Lu Wang et al. Hong Kong, China: Association for Computational Linguistics, pp. 70–79. DOI: 10.18653/v1/D19-5409. URL: <https://aclanthology.org/D19-5409>.
- ▶ Chen, Yulong, Yang Liu, et al. (Aug. 2021). "DialogSum: A Real-Life Scenario Dialogue Summarization Dataset". In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Ed. by Chengqing Zong et al. Online: Association for Computational Linguistics, pp. 5062–5074. DOI: 10.18653/v1/2021.findings-acl.449. URL: <https://aclanthology.org/2021.findings-acl.449>.
- ▶ Feigenblat, Guy et al. (Nov. 2021). "TWEETSUMM - A Dialog Summarization Dataset for Customer Service". In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. Ed. by Marie-Francine Moens et al. Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 245–260. DOI: 10.18653/v1/2021.findings-emnlp.24. URL: <https://aclanthology.org/2021.findings-emnlp.24>.
- ▶ Lin, Haitao et al. (Nov. 2021). "CSDS: A Fine-Grained Chinese Dataset for Customer Service Dialogue Summarization". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Ed. by Marie-Francine Moens et al. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 4436–4451. DOI: 10.18653/v1/2021.emnlp-main.365. URL: <https://aclanthology.org/2021.emnlp-main.365>.
- ▶ Zhong, Ming et al. (June 2021). "QMSum: A New Benchmark for Query-based Multi-domain Meeting Summarization". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Ed. by Kristina Toutanova et al. Online: Association for Computational Linguistics, pp. 5905–5921. DOI: 10.18653/v1/2021.naacl-main.472. URL: <https://aclanthology.org/2021.naacl-main.472>.

# References II

- ▶ Zhu, Chenguang et al. (June 2021). "MediaSum: A Large-scale Media Interview Dataset for Dialogue Summarization". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Ed. by Kristina Toutanova et al. Online: Association for Computational Linguistics, pp. 5927–5934. DOI: 10.18653/v1/2021.naacl-main.474. URL: <https://aclanthology.org/2021.naacl-main.474>.
- ▶ Bhattacharjee, Saprativa et al. (July 2022). "A Multi-Task Learning Approach for Summarization of Dialogues". In: *Proceedings of the 15th International Conference on Natural Language Generation: Generation Challenges*. Ed. by Samira Shaikh, Thiago Ferreira, and Amanda Stent. Waterville, Maine, USA and virtual meeting: Association for Computational Linguistics, pp. 110–120. URL: <https://aclanthology.org/2022.inlg-genchal.16>.
- ▶ Chauhan, Vipul et al. (July 2022). "TCS\_WITM\_2022 @ DialogSum : Topic oriented Summarization using Transformer based Encoder Decoder Model". In: *Proceedings of the 15th International Conference on Natural Language Generation: Generation Challenges*. Ed. by Samira Shaikh, Thiago Ferreira, and Amanda Stent. Waterville, Maine, USA and virtual meeting: Association for Computational Linguistics, pp. 104–109. URL: <https://aclanthology.org/2022.inlg-genchal.15>.
- ▶ Chen, Yulong, Naihao Deng, et al. (July 2022). "DialogSum Challenge: Results of the Dialogue Summarization Shared Task". In: *Proceedings of the 15th International Conference on Natural Language Generation: Generation Challenges*. Ed. by Samira Shaikh, Thiago Ferreira, and Amanda Stent. Waterville, Maine, USA and virtual meeting: Association for Computational Linguistics, pp. 94–103. URL: <https://aclanthology.org/2022.inlg-genchal.14>.
- ▶ Lundberg, Conrad, Leyre Sánchez Viñuela, and Siena Biales (July 2022). "Dialogue Summarization using BART". In: *Proceedings of the 15th International Conference on Natural Language Generation: Generation Challenges*. Ed. by Samira Shaikh, Thiago Ferreira, and Amanda Stent. Waterville, Maine, USA and virtual meeting: Association for Computational Linguistics, pp. 121–125. URL: <https://aclanthology.org/2022.inlg-genchal.17>.

# Example analysis

Reference summary 1: Frank got a new job and is telling Judy not only the heavy schedule but also the benefits of this job.

Reference summary 2: Frank explains to Judy the reason why he took a job at the Post office is that the government offers excellent health insurance benefits for its employees.

Reference summary 3: Frank describes his new job's schedule and insurance at the Post Office to Judy.

**model\_name: baseline-BARTLarge**

Generated summary: Frank tells Judy he got a new job for the Post Office and explains why he applied for such a demanding job.

```
<s> # Person 1 #: Hey, Frank, I heard you got a new job. # Person 2 #: Yeah, Judy, I will be working for the Post Office. It's not a bad job. # Person 1 #: Is it true that you have a heavy work schedule? # Person 2 #: That's right, I am supposed to work at 5 am everyday, and I only get 45 minutes for lunch. # Person 1 #: So, why did you apply for such a demanding job? # Person 2 #: Well, the government offers its employees excellent health insurance benefits. # Person 1 #: Oh, I see. And can your family members utilize the health insurance, too? # Person 2 #: Yeah, that's the best part. All of my children can get free medical care. # Person 1 #: That's a great employment benefit! # Person 2 #: Now you know why I wanted to work for the Post Office! </s>
```

**model\_name: baseline\_Filtered**

Generated summary: Frank tells Judy he got a new job and he thinks it's not a bad job. Frank also explains why he applied for the job.

```
<s> # Person 1 #: Hey, Frank, I heard you got a new job. # Person 2 #: Yeah, Judy, I will be working for the Post Office. It's not a bad job. # Person 1 #: Is it true that you have a heavy work schedule? # Person 2 #: That's right, I am supposed to work at 5 am everyday, and I only get 45 minutes for lunch. # Person 1 #: So, why did you apply for such a demanding job? # Person 2 #: Well, the government offers its employees excellent health insurance benefits. # Person 1 #: Oh, I see. And can your family members utilize the health insurance, too? # Person 2 #: Yeah, that's the best part. All of my children can get free medical care. # Person 1 #: That's a great employment benefit! # Person 2 #: Now you know why I wanted to work for the Post Office! </s>
```

# Example analysis

Reference summary 1: #Person1# is not ready for marriage. #Person2# reminds him of what he said a month ago and what will happen if he leaves Amy. After listening to #Person2#, #Person1# is ready for marriage.

Reference summary 2: #Person1# feels anxious about future marriage and sweats a lot. #Person2# helps #Person1# to think through the current situation and get ready for the marriage.

Reference summary 3: #Person1# is freaking out before his marriage with Amy and #Person2# helps him get over it by helping him think through all the details.

**model\_name: baseline-BARTLarge**

Generated summary: #Person1# wants to break up with Amy. #Person2# tries to persuade #Person1# not to do so and thinks he is just getting cold feet.

```
<s> # Person 1 #: Man, I'm freaking out! You gotta help me! # Person 2 #: Who a, who a, take it easy, relax | Geez, you're sweating like a pig | What's going on? # Person 1 #: I can't go through with this!! I just can't!! I'm not ready for marriage! What was I thinking? I'm only thirty five years old!! I've got my entire life ahead of me, adventures waiting! I can't settle down yet! # Person 2 #: What are you talking about? It wasn't more than a month ago that you were rambling on about how you are tired of living the life of a bachelor and how you envy your friends that have a family! # Person 1 #: I know! I know! # Person 2 #: Let's think this through | First of all, you cannot leave Amy at the altar | Not only will she be humiliated and kill you, but she is the woman that you love and the woman of your dreams! Why would you want to end that or jeopardize it like this? Second of all, you are just getting cold feet | You know deep down inside that you want to marry her, so cut the crap and do it! # Person 1 #: You're right | I'm being crazy | Ok, I'm ready, let's do this! # Person 2 #: Great! Phew! That was a close one! You had me going there for a minute I thought I was gonna have to slap some sense into you | </s>
```

**model\_name: baseline\_Filtered**

Generated summary: #Person1# is not ready for marriage. #Person2# tries to persuade #Person1# not to end that or jeopardize it like this, but #Person1# is still getting cold feet. Finally, he decides to marry Amy.

```
<s> # Person 1 #: Man, I'm freaking out! You gotta help me! # Person 2 #: Who a, who a, take it easy, relax | Geez, you're sweating like a pig | What's going on? # Person 1 #: I can't go through with this!! I just can't!! I'm not ready for marriage! What was I thinking? I'm only thirty five years old!! I've got my entire life ahead of me, adventures waiting! I can't settle down yet! # Person 2 #: What are you talking about? It wasn't more than a month ago that you were rambling on about how you are tired of living the life of a bachelor and how you envy your friends that have a family! # Person 1 #: I know! I know! # Person 2 #: Let's think this through | First of all, you cannot leave Amy at the altar | Not only will she be humiliated and kill you, but she is the woman that you love and the woman of your dreams! Why would you want to end that or jeopardize it like this? Second of all, you are just getting cold feet | You know deep down inside that you want to marry her, so cut the crap and do it! # Person 1 #: You're right | I'm being crazy | Ok, I'm ready, let's do this! # Person 2 #: Great! Phew! That was a close one! You had me going there for a minute I thought I was gonna have to slap some sense into you | </s>
```