

Improving Long Dialogue Summarization with Semantic Graph Representation

Yilun Hua* Zhaoyuan Deng* Kathleen McKeown

Columbia University

{yh3228, zd2286}@columbia.edu

kathy@cs.columbia.edu

Abstract

Although Large Language Models (LLMs) are successful in abstractive summarization of short dialogues, summarization of long dialogues remains challenging. To address this challenge, we propose a novel algorithm that processes complete dialogues comprising thousands of tokens into topic-segment-level Abstract Meaning Representation (AMR) graphs, which explicitly capture the dialogue structure, highlight salient semantics, and preserve high-level information. We also develop a new text-graph attention to leverage both graph semantics and a pretrained LLM that exploits the text. Finally, we propose an AMR node selection loss used jointly with conventional cross-entropy loss, to create additional training signals that facilitate graph feature encoding and content selection. Experiments show that our system outperforms the state-of-the-art models on multiple long dialogue summarization datasets, especially in low-resource settings, and generalizes well to out-of-domain data.

1 Introduction

Summarization of long dialogues is receiving more attention as virtual meetings become prevalent and advanced speech recognition technologies become accessible to accurately transcribe dialogues. Although the state-of-the-art (SOTA) models for abstractive summarization achieved remarkable performance on non-conversational documents via pre-training on datasets involving Wikipedia, books, stories, and news, generating high-quality summaries for long dialogue transcripts remains challenging. In this paper, we present the use of a semantic representation, Abstract Meaning Representation (AMR), to address some of the challenges in long dialogue summarization and we demonstrate its use on two different genres of long dialogues: meetings and screenplays of TV episodes.

One major challenge with dialogues is their complex structure in combination with their informal nature. [Chen and Yang \(2021a\)](#) discussed the impact of repetitions, false-starts, and hesitations on dialogue structure and suggested that key information can be spread across different portions of the dialogue. Multiple speakers and diverse spoken language styles also complicate the interactions involving speakers and coreferences. For example, it may be challenging to associate speakers with their opinions and actions as well as with their reactions to other speaker utterances. Admittedly, recent encoder-decoder LLMs like BART ([Lewis et al., 2020](#)) achieve good performance on short dialogues given adequate finetuning. Long dialogues, however, have more content as well as speakers and this brings greater complexity to dialogue structure, leading to performance degradation for encoder-decoder LLMs.

A second challenge lies in the transformer architecture of LLMs, which often draws spurious correlations ([Tu et al., 2020](#); [Kaushik et al., 2019](#)) and easily overfits small or homogeneous training sets. This limitation is particularly pertinent to long dialogue summarization, as it often involves low-resource domains. Meeting summarization datasets are particularly small. In fact, major meeting summary datasets, AMI ([Carletta et al., 2006](#)) and ICSI ([Janin et al., 2003](#)), have sizes of 137 and 59 transcripts respectively. In these very low-resource scenarios, a greater risk of overfitting is present.

To address these challenges, we propose a novel Abstract Meaning Representation (AMR) for long dialogues, to capture diverse entity interactions and complex structures. AMR, as a semantic graph representation, captures the most salient semantic knowledge using concept nodes and preserves inter-concept relations with labeled edges. It is believed to convey information largely orthogonal to what conventional models exploit from the text input ([Song et al., 2019](#)). AMRs thus provide reliable

*These authors contributed equally to this work.

semantic and structural cues to alleviate overfitting, which we show is particularly helpful for the low-resource setting. Our approach generalizes the sentence AMR introduced by [Banarescu et al. \(2013\)](#) and specifically designs an algorithm for building topic-segment AMRs for long spans of conversations to preserve global information.

Our approach to incorporating AMRs for summarization also differs from previous research. Existing methods of AMR-based summarization rely on graph-to-graph operations for content selection and additional text generation modules ([Liu et al., 2015](#); [Lee et al., 2021](#)), an approach that does not surpass state-of-the-art encoder-decoder LLMs. Instead, our work leverages both graph semantics and a pretrained encoder-decoder model. We develop a novel text-graph attention that exploits structural and global information in AMR to improve text encoding. We also propose a node selection loss used jointly with the standard cross-entropy loss for additional training signals.

Our work is orthogonal to the existing research on dialogue-specific pretraining ([Zhong et al., 2021](#)), which improves models’ familiarity with dialogue styles and has achieved the state-of-the-art for long-dialogue summarization with its two model variants, DialogLM and DialogLED. With our novel architecture, our AMR-based system can make full use of an LLM’s pretrained weights while providing better global information aggregation and fine-grained structural cues, from topic segments to speaker-action associations and inter-utterance coreferences. Our code is available at <https://github.com/Bobby-Hua/summarization-via-semantic-graph>.

In sum, the contributions of our work are:

- a new algorithm to build AMRs for long dialogues and their application to summarization
- a novel AMR node selection loss for better graph encoding and content selection
- new SOTA results on 3 datasets, up by +1.24 in Rouge-1, +1.53 in Rouge-2, and +1.67 in Rouge-L (pooling best results from 3 datasets)

2 Related Work

Dialogue Summarization Previous approaches to dialogue summarization also seek to improve models’ awareness of dialogue structures and interactions. [Chen et al. \(2021a\)](#) combines fact regularization via subject-verb-object (SVO) fact triplets

with modeling of the relationship between summary sentences and the positions of their supporting utterances in the source text. Other features, such as topic segmentation ([Li et al., 2019](#); [Chen and Yang, 2020](#)), dialogue acts ([Goo and Chen, 2018](#)), and conversation stages ([Chen and Yang, 2020](#)) have also been used in different models. The state-of-the-art in short dialogue summarization further suggests modeling discourse dependency and speaker-action relations ([Chen and Yang, 2021b](#)).

Our dialogue AMR is a more comprehensive representation of dialogue structures, which captures not only many useful features from the previous research but also fine-grained semantic structure and inter-utterance coreferences. The increased granularity of our representation allows it to effectively guide the encoding of text tokens via our novel cross-attention mechanism.

Summarization with AMR Current research that applies AMRs to summarization has primarily relied on graph-to-graph operations to produce summary AMR graphs. [Liu et al. \(2015\)](#) first introduces summarization via AMRs by merging sentence AMRs into document AMRs and conducting subgraph selection with integer linear programming. Other research adopts similar graph-to-graph methods, introducing new heuristics and algorithms for summary AMR generation and applying more recent generative models to generate fluent summaries ([Dohare et al., 2018](#); [Hardy and Vlachos, 2018](#); [Lee et al., 2021](#)). The graph-to-graph algorithms in these models are incompatible with the encoder-decoder LLMs behind current SOTA summarization models. Our work instead leverages both graph semantics and a pretrained encoder-decoder model, via our novel text-graph attention and additional node selection loss.

Dialogue AMR [Bai et al. \(2021\)](#) is the only work on dialogue AMR to the best of our knowledge. Its algorithm connects sentence AMRs with a dummy node and models coreference and identical concepts with additional edges. They only apply their dialogue AMRs to short conversations and for dialogue relation extraction and response generation. Our work is different in that we apply topic segmentation to process long dialogues and use node merging to allow a single AMR graph to capture a longer text span. Node merging, along with our speaker and utterance nodes, also creates a hierarchy of nodes with different centrality to help the graph encoder aggregate information at different levels.

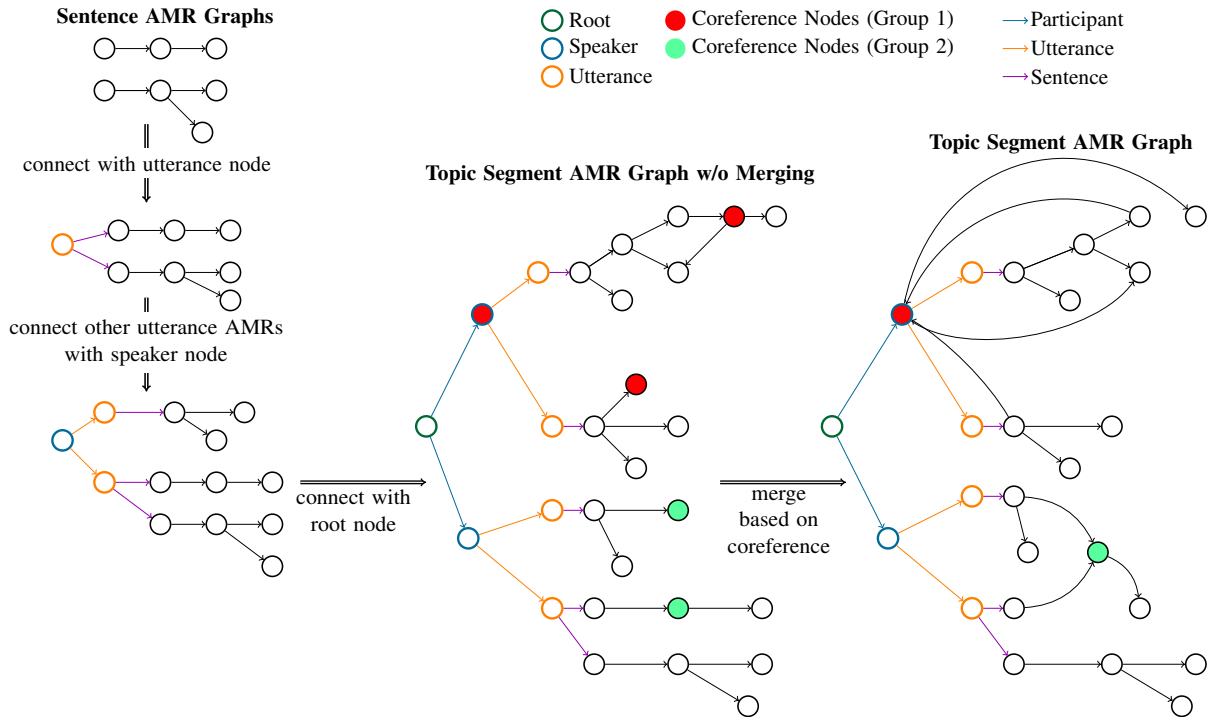


Figure 1: Process of Topic Segment AMR Generation from Sentence AMRs

Finally, our model differs from Bai et al. (2021)’s feature fusion and dual-attention model variants in terms of architecture, suggesting a promising way to combine AMRs with the LLM encoder and decoder.

3 Constructing AMR Representation For Long Dialogue

3.1 Topic Segmentation

A topic segment in a long dialogue is a consecutive sequence of topically-coherent lines. Topic segmentation can provide valuable high-level understanding of complex long dialogue structures, thus contributing important insights for summarization. Within a topic segment, speakers may respond to each other, request and perform actions, and refer to entities mentioned by others.

To make sure an AMR graph captures coherent semantics, we first perform topic segmentation on long dialogue. We adopt an existing strategy (Chen and Yang, 2020) that combines the classic topic segment algorithm C99 (Choi, 2000) with SentenceBERT (Reimers and Gurevych, 2019), to segment long dialogues into topic segments with reasonable lengths.

3.2 Topic Segment AMRs

To construct topic segment AMRs after segmentation, we adapt the steps in Bai et al. (2021) to build dialogue AMRs and we additionally introduce speaker nodes, utterance nodes, and a new procedure of node merging leveraging coreference relations. Given a topic segment consisting of multiple utterances, we use the AMR parser by Cai and Lam (2020a) to obtain an AMR graph for each utterance. And then as illustrated by Figure 1, we construct the topic segment AMR by connecting utterance AMRs with utterance nodes, speaker nodes, and a topic segment node (the root node) with appropriate edges, and perform node merging based on coreferences. For topic segment s_i , $AMR_algo(s_i) = \{V_i, E_i\}$.

Compared with Bai et al. (2021), we want our proposed speaker nodes and utterance nodes to encode the fine-grained hierarchical information from different levels of the topic segment AMR (synthesizing multiple utterances by one speaker, multiple sentences within one utterance, etc.). This is made possible by our graph encoder, which exploits speaker and utterance nodes’ abundant and unique interactions with other AMR concepts (see Section 4.1).

We perform node merging based on coreference relations to obtain the final topic segment AMR

graphs. We use a coreference resolution model by Dobrovolskii (2021) to obtain coreference relations between words, and JAMR (Flanigan et al., 2014) to obtain alignment between concepts and words. For a set of coreferencing nodes that do not contain speaker nodes, we merge them and select the most frequently appearing non-pronoun concept in the set as the concept for the merged node. Otherwise, we merge them and use the speaker node concept.

Through merging, a topic segment AMR graph can contain a single node for each distinct instance of a concept mentioned in the sentence AMR graphs. Previous work on node merging in AMR graphs (Liu et al., 2015) aims at reducing graph size and directly producing summary AMR graphs, so they merge identical concepts regardless of whether they are the same entity. Lee et al. (2021) shows that merging identical concepts results in undesirable information loss. In many cases, nodes representing different instances of an entity will be merged across sentences regardless of whether those nodes actually refer to the same entity. Lee et al. (2021) also shows that merging based on the combination of concept and coreference works best at producing summary AMR graphs. In our work, we only merge coreferencing nodes, as we want to maximize the preservation of information while allowing frequently occurring entities to have greater graph centrality in the topic segment AMR, so that relevant information can be explicitly aggregated and later encoded.

4 Model Architecture

Our model consists of an AMR graph encoder, a text encoder with text-graph cross-attention, and a decoder. The full architecture is illustrated in Figure 2.

4.1 AMR Encoder

To exploit AMR’s rich structural information and entity interactions, we apply Cai and Lam (2020b)’s graph transformer to encode a list of topic segment AMRs, $\mathcal{G} = \{\{V_1, E_1\}, \{V_2, E_2\}, \dots\}$. First, the relationship r_{ij} between two AMR concepts (nodes) v_i, v_j is encoded by the shortest path between them using a GRU. Then, we encode the segment AMR in the graph encoder, where every concept node attends to every other concept node with a modified attention mechanism informed by the encoded relationship between them. Overall, the operations of the AMR encoder F_G on the i -th

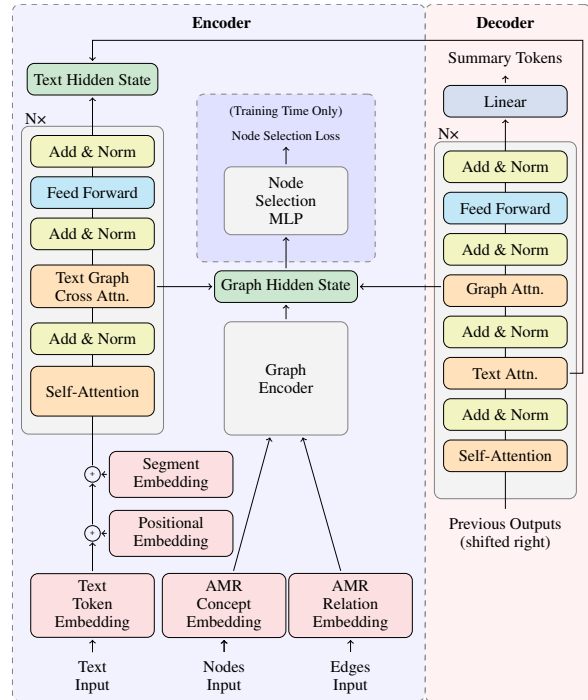


Figure 2: Model Architecture

segment (of m_i nodes) can be denoted as

$$H_i = F_G(\{V_i, E_i\}) = \{h_0^G, h_1^G, \dots, h_{m_i}^G\} \quad (1)$$

The graph output (hidden states) for the entire dialogue is a list of hidden states $F_G(\mathcal{G}) = \{H_1, H_2, \dots\}$, which will be passed on to the text encoder and the decoder.

4.2 Text Encoder

Our text encoder takes as input the text token embeddings $\mathcal{X} = \{x_0, \dots, x_n\}$ and the graph hidden states $F_G(\mathcal{G})$. For the text tokens, we adopt the simple sliding-window self-attention (Beltagy et al., 2020), which allows for a linear computation complexity instead of the $O(n^2)$ complexity in conventional transformers, and thus is more suitable for long sequence modeling.

To incorporate the graph information, we propose a new text-graph cross-attention to help the encoding of local tokens with 1) dialogue structural features explicitly exposed by AMRs and 2) the global semantic information from all the topic segments. We describe our text-graph attention and its advantages in the sections below.

4.2.1 Text-graph Cross-attention

Since our dialogues consist of thousands of tokens, we do not want all of them to have cross-attention

to the AMRs (to avoid high computation complexity). Instead, we add special [BOU] (beginning-of-utterance) tokens to every utterance and only allow the cross-attention from the [BOU]s to all the graph hidden states. Specifically, in every text encoder layer, after a [BOU] token has full attention to the tokens in its window (sliding-window self-attention), it cross-attends to all the topic segment AMRs of the entire dialogue. This step enriches the [BOU] embedding with the structural and global information from AMRs. The enriched [BOU] embedding, when sent to the next encoder layer, will guide the encoding of the surrounding text tokens as they attend the [BOU] via the sliding-window self-attention. This way, even if most of the text tokens do not have cross-attention on AMR graphs, they still benefit from the structural and global information indirectly, via the [BOU] hidden states.

Topic-segment Embedding To improve the text-graph cross-attention’s information selection, we want an utterance’s cross-attention to distinguish if an AMR graph represents the utterance’s own topic segment (i.e. the segment that contains this utterance) vs. AMRs of other topic segments. Intuitively, good cross-attention should treat the topic segments differently depending on their relevance to an utterance, by assigning greater attention weights to the AMR of an utterance’s own topic segment, for example. Therefore, we apply learnable segment embeddings (Devlin et al., 2019). For the i -th segment, the embedder produces a segment embedding E_s , which is added to the graph output of every concept in the segment. The same embedding is also added to all the [BOU] embeddings, $E_{[BOU]}$, corresponding to this segment. All segment embeddings are added to the $E_{[BOU]}$ and AMR hidden states when the text encoder first receives these inputs. Figure 3 illustrates the topic-segment embeddings.

Advantages of our text-graph cross-attention

This method captures the relational information from AMRs, allowing for a better grasp of dialogue structures. Also, the root node, utterance nodes, and important entity nodes have aggregated information to different extents depending on their centrality, which helps text-graph attention aggregate structural features across different levels of granularity.

This cross-attention also allows the global semantic information from the complete dialogue AMR to guide the encoding of local tokens, since

each [BOU] embedding attends to the concepts from all the topic segments of the conversation. The segment embeddings also help our text-graph attention extract relevant patterns more easily by looking at topic segments of different lengths while distinguishing its own segment from others.

Finally, cross-attention to AMR graphs is more efficient than directly attending to all the tokens in the text. AMR abstracts away the unimportant tokens and only keeps the salient information. It thus compresses the input sequence in terms of token numbers. For a typical dialogue, the number of AMR nodes ranges from half to two-thirds of the token number, leading to significantly lower computational complexity for the cross-attention on AMRs than a self-attention to all the text tokens, which is a common way to introduce global information found in LLMs like Longformer.

4.3 Decoder

We use a transformer decoder to generate the summary sequence. At each decoding stage t , self-attention is applied to hidden states of the previous $t - 1$ generated tokens. Then, the model synthesizes information through two cross-attentions, to the text hidden states and graph hidden states respectively. In this way, the AMR information not only benefits text encoding but also directly contributes to the generation of summaries.

During training, the decoder produces a standard cross-entropy loss L_t based on the teacher-forcing training strategy (Bengio et al., 2015).

4.4 Node Selection Loss

Additionally, we propose an auxiliary task of summary node selection at training time to create additional training signals for our AMR encoder and learn features useful for content selection. For each dialogue, we derive a summary AMR from the gold summary. We then apply a multilayer perceptron classifier to the graph hidden states of dialogue AMR concepts and make binary predictions on whether a node should appear in the gold summary AMR. We use a binary cross-entropy as the node selection loss L_s and train this auxiliary task jointly with our summarization task. Thus, our final loss is $L = L_t + L_s$.

The node selection loss is only for feature learning, unlike the previous graph-to-graph AMR-based summarization system (Lee et al., 2021), which selects nodes to construct a summary AMR and generates the summary therefrom. Our node

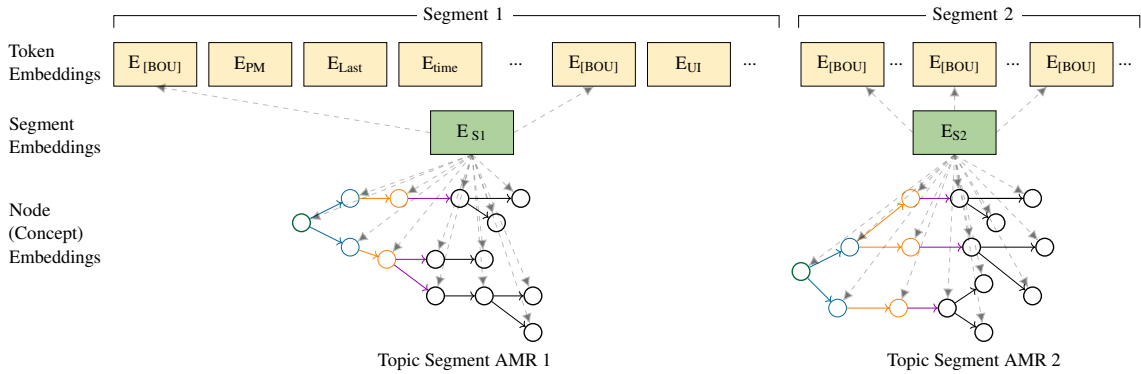


Figure 3: Topic Segment Embedding. Dotted arrows indicate adding the E_{s_1} to the [BOU] or concept embedding

selection loss here directly forces the AMR encoder to extract semantic information from AMRs. This change is particularly useful since our text encoder and decoder are initialized with extensively pre-trained weights but the graph encoder is initialized with random weights. We hope the node selection loss can help mitigate this gap and help the graph encoder receive meaningful gradient updates from the very beginning of training. Finally, the node selection loss also intuitively helps extract graph features useful for content selection, which are utilized by the downstream text encoder-decoder via cross-attentions.

5 Experiments

5.1 Long Dialogue Summarization Datasets

We trained and evaluated our models on the AMI (Carletta et al., 2006), ICSI (Janin et al., 2003), and ForeverDreaming (Chen et al., 2021b) datasets. AMI includes 137 transcripts from product design meetings and ICSI includes 57 transcripts from academic group meetings, both have gold summaries written by professionals. ForeverDreaming has transcripts from 4348 episodes of 66 TV shows and community-contributed gold summaries.

5.2 Implementation

During our development stage, we found our system had no significant performance difference whether our text encoder and decoder used the configuration of DialogLED-large or that of DialogLED-base. Thus, our full system only uses the base configuration and has 217 million trainable parameters in total, half the size of DialogLED-large (460 million). We report the baseline results from both DialogLED-base and the state-of-the-art DialogLED-large for the sake of completeness. We use random weights for the AMR encoder and the

pre-trained weights from DialogLED-base to initialize our text encoder and decoder. We extend DialogLED’s vocabulary and resize its embedding matrix to include additional AMR-specific tokens.

We adopt a learning rate of $2e-5$, Adam optimizer, and a batchsize of 32. Other implementation details are described in Appendix A.

5.3 Baselines

In addition to DialogLED models, other baselines we use are BART-Large (Lewis et al., 2019), UNILM (Dong et al., 2019), and DialogLM-sparse (Zhong et al., 2021), all of which have shown strong results in long dialogue summarization.

6 Results

We report ROUGE scores (Lin, 2004) as automatic evaluation metrics and also report human evaluation results. To test the statistical significance of the improvement brought by our model, we use the Almost Stochastic Order test (ASO) (Del Barrio et al., 2018; Dror et al., 2019) as implemented by Ulmer et al. (2022). ASO tests the statistical significance by evaluating stochastic dominance. ASO is more suitable for deep learning models because it does not require t-test’s normal distribution assumption, which likely does not hold for neural networks (Dror et al., 2019). We also report t-test results for the sake of completeness.

6.1 Performance on in-domain Test Sets

As shown in Table 1 and Table 2, our system achieves new SOTA results on all datasets. Compared with the previous SOTA DialogLED-large, our model brings a statistically significant improvement in most of the Rouge metrics. For the meeting datasets, the most remarkable improvements include a 1.24 increase in Rouge-1 for AMI, a 1.53

Models	AMI			ICSI		
	R-1	R-2	R-L	R-1	R-2	R-L
BART-large [◇]	51.77	18.83	49.67	46.23	10.17	44.83
UNILM-base [◇]	51.92	18.42	49.89	46.75	11.39	45.13
DialogLM-S [◇]	53.72	19.61	51.83	49.56	12.53	47.08
LED-large [◇]	53.52	19.28	51.08	48.98	12.38	46.30
DialogLED-B	53.26	19.99	51.11	50.25	12.70	47.14
DialogLED-L [◇]	54.80	20.37	52.26	50.11	13.23	47.25
DialogLED-L	54.15(.34)	20.00(.45)	51.83(.39)	50.14(.62)	12.71(.30)	46.67(.68)
Ours	55.39(.25) ^{*†}	20.33(.45) ^{*†}	52.93(.32) ^{*†}	50.98(.17)	14.24(.51) ^{*†}	48.34(.47) [†]

Table 1: [◇] denotes results reported by [Zhong et al. \(2021\)](#). -S, -B, and -L stands for -sparse, -base, and -large, respectively. For DialogLED-large, we report both [Zhong et al. \(2021\)](#)'s results and the results we reproduced (average of 3 random seeds). For significance tests, we use a confidence level of 0.95 and consider the improvement significant if ASO's violation ratio is < 0.2 , as recommended by [Ulmer et al. \(2022\)](#). * denotes t-test $p < 0.05$; † denotes violation ratio < 0.2 .

Models	ForeverDreaming		
	R-1	R-2	R-L
BART-large [◇]	33.82	7.48	29.07
UNILM-base [◇]	32.16	5.93	27.27
DialogLM-S [◇]	35.75	8.27	30.76
LED-large [◇]	35.47	8.13	30.28
DialogLED-B	35.92	8.38	30.30
DialogLED-L [◇]	36.70	8.68	31.38
DialogLED-L	36.71 (0.30)	8.63 (0.12)	31.42 (0.25)
Ours	37.10 (0.16)	8.81 [†] (0.09)	31.85 [†] (0.21)

Table 2: ForeverDreaming Results

increase in Rouge-2 for ICSI, and a 1.67 increase in Rouge-L for ICSI. The improvement is greater when we compare our model with DialogLED-base, which has the same pre-trained weights for the text encoder-decoder as ours. This suggests that our AMR graphs have enhanced the small DialogLED-base backbone, outperforming its "large" counterpart. For the ForeverDreaming dataset, our model also outperforms DialogLED-large, though the margin is smaller.

The smaller performance increase on ForeverDreaming than on AMI and ICSI supports our intuition that AMRs are more helpful under low-resource settings, where the risk of overfitting is higher. AMRs expose relevant semantic information and abstract away the syntactic/stylistic patterns, which helps prevent models from drawing

spurious correlations. As discussed in 5.1, ForeverDreaming is a larger dataset with diverse training instances, so it may already contain enough counterexamples for the potential spurious correlations, which intuitively contributes to the model's robustness ([Tu et al., 2020](#)) and thus reduces the performance gap between the baseline and our model.

6.1.1 Ablation Results

We perform ablation experiments on AMI and ICSI, which required less training time and are most representative of the low-resource dialogue summarization setting we are interested in. The results are presented in Table 3. Overall, the node selection loss, encoder text-graph attention, decoder graph attention, and the dialogue AMR all contributed to our model's performance. Among the architecture changes we proposed, removing the encoder text-graph attention resulted in the greatest performance degradation for both datasets. The benefits of the node selection loss were also substantial and comparable to the improvement brought by the text-graph attention. The conventional decoder graph attention also contributed to the metrics, though removing it did bring a slight increase in Rouge-2 for AMI. Since the magnitude of this increase is small, we still consider the decoder graph attention an important component in our system. Finally, removing AMRs made the model equivalent to DialogLED-base and thus led to the lowest performance in the table.

Models	AMI			ICSI		
	R-1	R-2	R-L	R-1	R-2	R-L
Ours (full)	55.49	20.95	53.20	51.22	15.00	49.03
- Node Selection Loss	54.98	20.31	52.52	51.03	13.49	48.60
- Enc. Text-Graph Attn	54.61	20.43	52.31	50.53	14.14	47.99
- Dec. Graph Attn	54.88	21.11	52.44	50.35	14.50	47.30
- AMR	53.92	19.22	51.40	49.90	12.26	46.40

Table 3: Ablation Results

Datasets	Succ.	Flu.	Spec.	Faith.
AMI	.64	.54	.72	.58
ForeverDreaming	.52	.50	.58	.55

Table 4: Human evaluation on **Succinctness**, **Fluency**, **Specificity**, and **Faithfulness**. Results in the percentage of times raters prefer instances generated by our model over DialogLED-L

6.1.2 Human Evaluation

For human evaluation, we ask 8 university students to make pairwise comparisons of the summaries generated by our system and by DialogLED-large. For each source dialogue, we present its two summaries in a random order (summary 1 from our model and summary 2 from the baseline, or the opposite). This way, the students do not know which system produced a specific summary. We use the full test set of 20 meetings from AMI and a random subset of 30 episodes from ForeverDreaming.

Student raters compare summaries according to four metrics: succinctness (e.g., does not contain redundant information), fluency (e.g., does not contain grammatical error), specificity (e.g., does not contain too general or uninformative statements) (Louis and Nenkova, 2011), and faithfulness (e.g., does not contain false information to the source text) (Chen and Yang, 2021b; Zhong et al., 2021). We use specificity to replace the commonly used metric, informativeness (covers the most important content (Chen and Yang, 2021b)), since for long dialogues with multiple topics, it is difficult to decide if a piece of information in the summary is important. Specific instructions and definitions of these terms can be found in Appendix B. As shown in Table 4, our system that utilized semantic information from AMR graphs generated better summaries with respect to all four metrics in both datasets. We believe our AMR graphs helped the model utilize the salient information and thus generate more

Datasets	Models	R-1	R-2	R-L
AMI	DialogLED	33.43	7.02	31.60
	Ours	37.10	7.02	35.32
ICSI	DialogLED	30.48	4.58	28.86
	Ours	35.43	6.22	33.70

Table 5: Out-of-domain evaluation results of the model trained on ForeverDreaming

specific and faithful summaries. We observe less improvement in fluency, likely due to the fact that our model and the baseline have the same decoder structure. The improvement is also smaller for the ForeverDreaming dataset, which is consistent with our ROUGE scores.

6.2 Out-of-domain Evaluation

To examine how our model generalizes, we used our model trained on ForeverDreaming and directly evaluated it on the test sets of AMI and ICSI. We noticed that the gold summary styles significantly differ for the meetings and the TV shows. Thus, we combined the TV-finetuned text and graph encoders with a decoder that only has pretrained DialogLED weights. This helped the decoder produce a more neutral style. This setup still constitutes a meaningful scenario where we have some training data for the TV domain but no data for the meeting domain.

As shown in Table 5, in the out-of-domain setting, our system outperforms the baseline by an even greater margin than in the in-domain setting. These results further support our claim that our model is good at learning relevant features that are generalizable. The improved generalization brought by AMRs also suggests a promising future direction in using semantic graph representations to improve models’ zero-shot performance and apply to a broad set of tasks under low-resource settings.

7 Conclusion

In this work, we propose a novel AMR algorithm to capture long dialogue structures. We develop a text-graph cross-attention and node selection loss to effectively extract structural features and integrate them into an encoder-decoder LLM for summarization. Empirically, we show that our system outperforms the SOTA models and achieves particularly promising results in low-resource settings and out-of-domain evaluation.

Limitations

Constructing semantic graphs, in general, requires multiple additional tools, which inevitably introduce errors. In this paper, we worked to reduce potential errors. We used the SOTA AMR parser and coreference resolution model and adopted mechanisms to reduce error propagation to our final graphs. However, we have not measured errors involving topic segmentation and AMR parsing due to the expensive human annotations required. It will be helpful to investigate how these errors can impact system performance when they are combined with encoder-decoder LLMs and if the AMR encoder is robust to small errors in AMR graphs. We leave these investigations as future work.

Finetuning LLMs for long dialogues requires many GPU hours and energy. Therefore, our hyperparameter search was limited to 3 different values for learning rates, 3 for warmup steps, and 2 for batchsize. A more extensive search may bring additional improvement to our model.

Ethics Statement

It is possible that unintended use of our system could amplify the impact of offensive language and bias in online discussions, as the salient opinions can be extracted and become more visible to the public. We propose using a toxicity classifier on our output to identify and suppress biased and offensive summaries. Despite the improved faithfulness of our system, automatic summaries in general may contain factual errors or inconsistencies with the opinions of the speaker. Thus, anyone who uses an automatic summarization system must proceed with caution and refer to the source document when necessary.

Acknowledgement

The authors thank the anonymous reviewers and Daniel Bauer for their constructive feedback and comments on this work. The authors are also grateful to Zhijie Xu, Samantha Mayers, Chengyao Gu, and Duaa Tariq for their assistance and to all the volunteers who participated in the human evaluation. This work is supported in part by Columbia University’s Class of 1939 Columbia College Summer Research Fellowship.

References

- Xuefeng Bai, Yulong Chen, Linfeng Song, and Yue Zhang. 2021. [Semantic representation for dialogue modeling](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4430–4445, Online. Association for Computational Linguistics.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. [Abstract Meaning Representation for sembanking](#). In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.
- Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. [Longformer: The long-document transformer](#). *CoRR*, abs/2004.05150.
- Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. 2015. Scheduled sampling for sequence prediction with recurrent neural networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS’15, page 1171–1179, Cambridge, MA, USA. MIT Press.
- Deng Cai and Wai Lam. 2020a. [AMR parsing via graph-sequence iterative inference](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1290–1301, Online. Association for Computational Linguistics.
- Deng Cai and Wai Lam. 2020b. [Graph transformer for graph-to-sequence learning](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):7464–7471.
- Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried Post, Dennis Reidsma, and Pierre Wellner. 2006. The AMI Meeting Corpus: A Pre-announcement. In *Machine*

- Learning for Multimodal Interaction*, pages 28–39, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Jiaao Chen and Diyi Yang. 2020. [Multi-view sequence-to-sequence models with conversational structure for abstractive dialogue summarization](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4106–4118, Online. Association for Computational Linguistics.
- Jiaao Chen and Diyi Yang. 2021a. [Structure-Aware Abstractive Conversation Summarization via Discourse and Action Graphs](#). *arXiv:2104.08400 [cs]*. ArXiv: 2104.08400.
- Jiaao Chen and Diyi Yang. 2021b. [Structure-aware abstractive conversation summarization via discourse and action graphs](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1380–1391, Online. Association for Computational Linguistics.
- Wang Chen, Piji Li, Hou Pong Chan, and Irwin King. 2021a. [Dialogue summarization with supporting utterance flow modeling and fact regularization](#). *CoRR*, abs/2108.01268.
- Yulong Chen, Yang Liu, Liang Chen, and Yue Zhang. 2021b. [DialogSum: A real-life scenario dialogue summarization dataset](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 5062–5074, Online. Association for Computational Linguistics.
- Freddy Y. Y. Choi. 2000. [Advances in domain independent linear text segmentation](#). In *1st Meeting of the North American Chapter of the Association for Computational Linguistics*.
- Eustasio Del Barrio, Juan A Cuesta-Albertos, and Carlos Matrán. 2018. An optimal transportation approach for assessing almost stochastic order. In *The Mathematics of the Uncertain*, pages 33–44. Springer.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Vladimir Dobrovolskii. 2021. [Word-level coreference resolution](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7670–7675, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Shibhansh Dohare, Vivek Gupta, and Harish Karnick. 2018. [Unsupervised semantic abstractive summarization](#). In *Proceedings of ACL 2018, Student Research Workshop*, pages 74–83, Melbourne, Australia. Association for Computational Linguistics.
- Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. *Unified Language Model Pre-Training for Natural Language Understanding and Generation*. Curran Associates Inc., Red Hook, NY, USA.
- Rotem Dror, Segev Shlomov, and Roi Reichart. 2019. [Deep dominance - how to properly compare deep neural models](#). In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 2773–2785. Association for Computational Linguistics.
- Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. 2014. [A discriminative graph-based parser for the Abstract Meaning Representation](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1436, Baltimore, Maryland. Association for Computational Linguistics.
- Chih-Wen Goo and Yun-Nung Chen. 2018. [Abstractive dialogue summarization with sentence-gated modeling optimized by dialogue acts](#). In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 735–742.
- Hardy Hardy and Andreas Vlachos. 2018. [Guided neural language generation for abstractive summarization using Abstract Meaning Representation](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 768–773, Brussels, Belgium. Association for Computational Linguistics.
- A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters. 2003. [The icsi meeting corpus](#). In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, volume 1, pages I–I.
- Divyansh Kaushik, Eduard H. Hovy, and Zachary C. Lipton. 2019. [Learning the difference that makes a difference with counterfactually-augmented data](#). *CoRR*, abs/1909.12434.
- J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.
- Fei-Tzin Lee, Chris Kedzie, Nakul Verma, and Kathleen R. McKeown. 2021. [An analysis of document graph construction methods for AMR summarization](#). *CoRR*, abs/2111.13993.

- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2019. [BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). *CoRR*, abs/1910.13461.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. [BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.
- Manling Li, Lingyu Zhang, Heng Ji, and Richard J. Radke. 2019. [Keep meeting summaries on topic: Abstractive multi-modal meeting summarization](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2190–2196, Florence, Italy. Association for Computational Linguistics.
- Chin-Yew Lin. 2004. [ROUGE: A package for automatic evaluation of summaries](#). In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A. Smith. 2015. [Toward abstractive summarization using semantic representations](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1077–1086, Denver, Colorado. Association for Computational Linguistics.
- Annie Louis and Ani Nenkova. 2011. [Automatic identification of general and specific sentences by leveraging discourse annotations](#). In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 605–613, Chiang Mai, Thailand. Asian Federation of Natural Language Processing.
- Ani Nenkova and Rebecca Passonneau. 2004. [Evaluating content selection in summarization: The pyramid method](#). In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, pages 145–152, Boston, Massachusetts, USA. Association for Computational Linguistics.
- 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 and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.
- Linfeng Song, Daniel Gildea, Yue Zhang, Zhiguo Wang, and Jinsong Su. 2019. [Semantic neural machine translation using amr](#). *Transactions of the Association for Computational Linguistics*, 7:19–31.
- Lifu Tu, Garima Lalwani, Spandana Gella, and He He. 2020. [An empirical study on robustness to spurious correlations using pre-trained language models](#). *Transactions of the Association for Computational Linguistics*, 8:621–633.
- Dennis Ulmer, Christian Hardmeier, and Jes Frellsen. 2022. [deep-significance-easy and meaningful statistical significance testing in the age of neural networks](#). *arXiv preprint arXiv:2204.06815*.
- Ming Zhong, Yang Liu, Yichong Xu, Chenguang Zhu, and Michael Zeng. 2021. [Dialoglm: Pre-trained model for long dialogue understanding and summarization](#). *CoRR*, abs/2109.02492.

A Implementation Details

We conducted hyperparameter search over learning rates [1e-5, 2e-5, 1e-4], batchsizes [16, 32], and warmup steps [50, 100, 1000]. For the graph encoder, we use 4 layers, 8 graph attention heads, and 8 relation attention heads. For other hyperparameters, we adopt the same values as DialogLED.

We train our model for 20000 updates (1000 warmup steps) for ForeverDreaming and 400 updates for the meeting datasets (100 warmup steps). Training takes 5 hours for the meeting datasets and a week for the ForeverDreaming on one Tesla A100 GPU.

B Human Evaluation Details

Detailed definitions for our four metrics:

Succinctness how much redundant information is there in the summary? For example, the same piece of information should not appear multiple times in slightly paraphrased forms. Please note that the simple repetition of words within a sentence should be considered as lacking fluency instead of succinctness. (see the definition of fluency)

Fluency how natural do the sentences seem to humans and how many grammatical mistakes do they have? Sentences with erroneously repeated words or phrases are also considered not fluent. (e.g., “I have have not ..” or “they they will not ...”)

Specificity how specific is the information in the summaries? Sentences could be general or specific: general sentences are broad statements made about a topic, while specific sentences contain details and can be used to support or explain the general sentences further. (Louis and Nenkova, 2011)

Faithfulness does the summary contain information not supported by the source text? Does the summary associate actions and opinions with the correct individuals? (Chen and Yang, 2021b; Zhong et al., 2021)

We instruct the raters to perform the following steps:

- Read the full source text first to get a general impression. For each instance, work on one criterion at a time and record which summary is better in terms of each criterion.
- For Succinctness and Fluency, you may make your judgment without consulting the source text again to save time.
- For Specificity and Faithfulness, separate the summaries into Summarization Content Units (SCUs) first (Nenkova and Passonneau, 2004). The goal is to split sentences into small phrases, each conveying a stand-alone piece of information.
- For example, the following summary is split into 4 SCUs: [the project manager recapped the decisions made in the previous meeting]¹. [the industrial designer presented the working design]² and [discussed the interior workings of a remote]³ and [how to incorporate the corporate image into the design]⁴.
- Then, for each SCU, check if they are specific and/or unfaithful according to the definition above. Summary with a higher number of specific SCUs is considered better in specificity. Summary with a lower number of unfaithful SCUs is considered better in faithfulness.
- Finally, record your evaluation in the provided spreadsheet.

Our raters consist of 8 university students, primarily native speakers, with some international students who are fluent in English.

As shown in Table 6, we measured the inter-annotator agreement in our human evaluation using the Fleiss’ kappa score. According to Landis and Koch (1977)’s interpretation of kappa scores, our scores suggest that there was mostly a fair (0.21-0.40) to moderate (0.41-0.6) agreement among our human annotators, depending on the specific evaluation category and dataset.

Datasets	Succ.	Flu.	Spec.	Faith.
AMI	.42	.21	.46	.21
ForeverDreaming	.34	.08	.31	.26

Table 6: Fleiss’ kappa score for inter-annotator agreement in our human evaluation on **Succinctness**, **Fluency**, **Specificity**, and **Faithfulness**.

C Examples

We provide some example summaries produced by the baseline system (DialogLED-L) and our system, alongside the corresponding source texts. For better readability, we present the summary as a contiguous text instead of a list of Summarization Content Units (SCUs). We highlight the text spans we found specific (as defined by our specificity criterion) in yellow and mark the spans we found unfaithful with the red text color. When applicable, we use superscripts to mark the corresponding conflicting text span in the source text to explain unfaithfulness. Hallucinated information is also marked as unfaithful but does not have a superscript to mark related text spans in the source text. Additionally, we underline the redundant text spans in the summary. We didn’t mark problems with fluency as our model has similar performance in fluency compared with the baseline.

Example 1:

In this example, our system’s output was consistently considered to have higher specificity and faithfulness than the baseline’s by our human raters.

Output from Baseline System (DialogLED-L):

the project manager opened the meeting and stated the agenda. the marketing expert presented research on users of remote controls, showing that remotes are being considered ugly. **the group also indicated that users aged sixteen to forty five years often get lost in the room.** the research showed that younger users like to have a fancy-looking remote. the user interface designer discussed the basic layout of the device and suggested that the remote should look very user-friendly. the industrial designer presented requirements for the project , **including the number of buttons , power , and teletext**². the project manager briefed the group on some new requirements and led them in a discussion in which the group discussed their target group and what functions they would like to see in the remote they will produce. the target

group will be people aged 16 to 40¹. the remote will be used only for televisions. the corporate image will be displayed in the product. the remotes will have buttons for the most frequently used functions. there will be a docking station for the lcd screen. remotes only have the general functions of a remote. teletext will not be included in the design³. remote will use a scroll button⁴. the elderly mobile phone will not have internet access. the maximum production cost will be 12.50 euros. the group decided to include teletext for the first time in their remote design. they decided to eliminate the docking station⁶ and keep the buttons for channel selection and volume selection. they also decided to ignore the audio signal⁵ and to focus on the main functions of the remote. it will be too complex and the production costs will go sky-high. the users of 16 to forty-five years old were generally quite pleased with the design and the interface.

Output from Our System:

the project manager opened the meeting and introduced the new requirements for the project. the marketing expert presented research on users of remote controls and indicated that users aged sixteen to forty-five typically do not like remotes. the user interface designer discussed the interior workings of a remote and stressed the importance of a simple design. he suggested that the remote should look very simple and be easy to learn. the industrial designer gave a presentation about the functions that will be included in the design. the group discussed the option to include teletext, but decided that it was not a useful feature³. the project manager briefed the group on some new requirements and led them in a discussion about the remote's basic layout and its basic layout, and suggested that they should focus on the target group and not on the teletext function. they also discussed the docking station which will allow the remote to be used only for televisions and which would not increase the production cost. they discussed the possibility of using a locator function and decided that the buttons for the most frequently used functions and the lcd screen should be included. the remote will have a docking station for the tv functions. there will be no teletext feature on the remote³ because it is too complex and the time to market will be too big, if we are to have a fancy looking remote. the group also complained that remotes often get lost in the

room. they complained that there was not enough information about how a remote works and that elderly people need to adapt to new devices. the designers of the project had problems with the group's tendency to mess things up and to

Source Text:

product_manager: everybody found his place again?

marketing_expert: yes.

product_manager: that's. so this is our second meeting. and still failing? now we're going into the functional design. important thing of this phase is that we're going to try to get an agreement about the user requirements, technical function design, and the working design. so that we can move onto the second phase. but first this phase. first an announcement. there's a little adaptation in the air conditioning system. so there's our ghost mouse again. that that means that you can have a little trouble with, little trouble with the air conditioning, that's because of this it's in wing c _ and e _ . it should be over in a while, couple of days. but it's going to be cold anyway, so i don't think you're gonna need it.

marketing_expert: no.

product_manager: then our agenda. now first the opening. this time i will take the minutes. you're going to have a presentation. all of you. and we've got forty minutes for the whole prese for the whole presentations. so i suggest we take about seven minutes per presentation, and then we can have a little discussion about the new project requirements which have been sent to me. and then the decision on the control functions which we wanna include and those which we don't wanna include. we've got forty minutes for all of it. i suggest let's start with the first presentation. who wants to be first?

marketing_expert: think i'll go first.

product_manager: just maybe it's easier if you will tell your presentation as. just which function you have and what you're gonna talk about.

marketing_expert: my name is freek van ponnen. i 'm the market expert. but you already knew that. i've done some research. we have we have been doing research in a usability lab where we observed users operating remote controls. we let them fill out a questionnaire. we had one hundred of these test subjects. in addition we did some market research. see what the market consists of. what ages are involved. these are

three quite astonishing results,. remotes are being considered ugly. f seventy five percent of the people questioned indicated that they thought their remote were was ugly. and an additional eighty percent indicated that they would spend more money on a fancy - looking remote control. so in addition remotes were not very functional. fifty percent of the people indicated they only loo used about ten percent of the buttons on a remote control. and fifty percent of the people indicated that their remote tended to get lost in their room. some things. then we did some research to the most relevant functions. channel selection and volume selection both got a ten on a scale of one to ten for relevancy. the power button got a nine. and teletext got a six and a half. so these are the most important functions on a remote control. then there are some one - time use function. that's what i like to call them. that audio settings, video settings, and channel settings buttons. which are not really used very frequently, but are still considered to be of some importance. channel selection was also indicated to be used very frequently. one hundred and sixty eight times per hour. then these are the this is the market. sixty percent of the market consists of users between the ages sixteen and forty f six. main characteristic of this group is that they're very critical on the remote control. they like to use new f new functions. but they also are very critical. they won't spend their money very easily. the users of forty six to sixty five years cons the make up forty percent of the market. they are not really very interested in features. but they do tend to spend their money a lot easier. what this indicates for our design. we should make a remote for the future. and this means we would have to focus on the age ages sixteen to **forty five**¹. this also makes up most the biggest part of the market, so that will also be where our main profit would be gettable. this would mean we would have to make a fancy design. the results also indicated that about one quarter of the people questioned thought that the remote control caused r_s _ r_s_i _ . this is certainly something to take into account. and thirty four percent thought that it was hard to learn a n how to operate a new control, remote control. so these are two factors that should be included in the design. besides that the remote must look very. and the functionality as a lot of people indicated, they only use about ten percent of the buttons, we should make very few buttons.

this will also be beneficial to the design of the remote. the most frequently used buttons should be emphasised. especially the channel selection and audio selection buttons. they're used most and so they should be robust. they shouldn't break down easily. then as mo as a lot of people indicated that their remote got lost in the room, it might be and i say might be because it would certainly boost the production costs a lot. but it might be a good idea to make a docking station. and this would, could get a button in it which would send a signal to the remote which would then beep. so you'd know where it is in the room. and in addition to this it could recharge the batteries in the remote if you put it in. then a surprisingly great deal of people w indicated that an l_c_d _ screen in the remote control would be preferred. this was mostly people in the age of sixteen to twenty five. but up till forty five it remains feasible. this would also greatly increase the production costs but these are just some small factors we could consider. that would be all.

product_manager: anybody have any questions until now?

marketing_expert: any questions?

product_manager: about functional requirements?

user_interface_designer: mm - hmm.

industrial_designer: no.

product_manager: that's clear. now to the second.

user_interface_designer: i've been looking at the user interface of it. f for the techno f functions of it.

product_manager: you can take your time. we've got plenty of time,

user_interface_designer:?

marketing_expert: you should go to the top thingy. slide show.

product_manager: there it is.

user_interface_designer: we must use the general functions of the remote control. i've do i've done a little research on the internet and not much information about it, about interface but i've been thinking about a simple manner to put a lot of functions in one remote control. so you've got a lot of devi devices like d_v_d _ television, stereo. so but it must be user - friendly. you c you can't put a lot of functions in one. in one remote control. product_manager: one remote.

user_interface_designer: but got many functions in one remote control, you can see, this is quite simple remote control. few buttons but this re

remote control got a lot of buttons. people don't like it, so what i was thinking about was keep the general functions like they are. like the on - off button. keep it like a red button. everybody knows it so you don't have to change that. my personal preferences. use a display for specific functions of the different device. what i was thinking about was you've got this the remote control and you got here the general functions, like the on - off button sound and here you've got a small display. it's a touchscreen. you got a general function of the device for a device player or so the play for playing reverse. and you got here real buttons for selecting a device. this button is for a device or for every device you've got a small part display of a part buttons. you never got all the buttons on one device. that's my idea about it. and let's see. so a touchscreen. and then the buttons the real buttons we have to use. we better use quite large buttons for everybody have to use it so even old people young people. we must keep buttons quite simple and quite large.. that was my part of it.

product_manager: anybody has questions about the technical functions?

industrial_designer: if we are gonna use a touchscreen we're gonna go way above the twelve and a half euros.

user_interface_designer: no i don't. you got quite a cheap touchscreen. so it's not in colour.

product_manager: touchscreen.

user_interface_designer: it's just one colour i seen something on the internet not today but a few weeks ago. you got quite an a touchscreen and it's for twenty euros or less. so it's possible.

product_manager: that's.

marketing_expert: it would certainly make a fancy design.

industrial_designer: but the it wouldn't be very robust. it's very fragile and you can get scratches on it.

marketing_expert: that is true.

user_interface_designer: that's true.

product_manager: maybe we can first listen to your presentation?

marketing_expert: we would have to look into that.

product_manager: and then we have a little discussion about the requirements and design.

industrial_designer: that's.

product_manager: it's going to it's not too much.

industrial_designer: i've got a presentation about

the working design. first about how it works. it's really simple. everybody knows how a remote works. the user presses a button. the remote determines what button it is, uses the infrared to send a signal to the tv. the tv switches to the frequency, or what function it is. so we've got the plate. it gets conductive disks for every button. when the user presses a button, a signal got sent, goes to the led and transmits transmits its to the tv. it's a very simple device, technically speaking. this is a schematic overview. you've got the buttons. the power source. and when a button gets pressed, its goes to the chip. the chip controls the infrared bulb and perhaps a normal bulb. when you press a button you can actually see your pressed button. we should use default materials, simple plastics. keep the inner workings simple, so it's robust. we should focus on aesthetics, the design and the user interface, because if you're going to use high - tech materials the price is going to go sky - high. and you only have to design a remote once, and if you use high - tech materials it come back in every product. it's, in my idea, it's gonna be smart to invest in design and not in the product itself. that's it.

product_manager: now i hope everybody has a little bit more insight in the functions we all have and what we are doing right now. i'm the project manager so i'm here to mess things up and tell you some new requirements. that's, we've got to design a remote which is only suitable for tv. that's because it will be too complex and the time to market will be too big, if we wanna have it for more functions. so it has to be simple. **another point is we have to skip the teletext, because in the world of upcoming internet we think teletext is going to be a thing of the past. and it's a function we don't need in our remote control².** internet is also mentioned in a function we can use. maybe also on televisions it will be available as. another one is the customer is forty plus. that's the market we have to target, because we are going to develop a new product which is specially designed for the younger customers. this is a bit pity for the marketing expert. because he was aiming on the younger persons. so we have to find a market which is above forty plus but which will suit our remote control, and the other way round. and we have to be very attent in putting the corporate image in our product. so it has to be visible in our design, in the way our device works. and we have

to be very clear on this point as. i suggest let's have a discussion on the control functions.

marketing_expert: is there any discussion possible about the new product requirement?

product_manager: we can see if we can find a way between the functions we wanna use and the market we wanna reach with our product.

marketing_expert: you're saying that teletext is gonna be an old feature and it's not gonna be used anymore anyway pretty soon. and new t_v_s will have internet access on them. but if you're targeting people of forty plus, the chance that they will have a t_v _ with internet access within the next like twenty years is very slim. in addition people indicated that teletext simply was an important feature for the remote control. so it's pretty dumb to put no teletext feature on it. i 'm against it.

product_manager: against the no teletext?

marketing_expert: besides that, the market for forty plus is like pretty small. but if i s if i see this, it's we're just gonna go for another

product_manager: it's it is

user_interface_designer: forty

product_manager: standard remote.

marketing_expert: pretty

product_manager: no we can

marketing_expert: and not innovative

product_manager: we can do a lot with the design and the simple buttons

marketing_expert: remote control.

product_manager: which were also mentioned. if we put a lot of effort in those, we can make a remote control with just two or three buttons. or just a remote which is suitable for the market we wanna reach because it is forty percent of the market. if you look in holland at the whole generation of forty plus, fifty plus, it's the biggest share of the whole population now.

marketing_expert: yes but it's not the biggest part of the market.

product_manager: no.

marketing_expert: and besides that, they're not very critical so they don't really care what the remote control is like. they'll just take the first thing they see and which looks acceptable.

product_manager: but don't you think that if we make a remote which is typically made for this market, that people think the people think that's the device i 've looked for although i didn't realise it. let's try it.

marketing_expert: that would be the case in the sixteen to forty five age category. because they are critical and they want to have a fancy remote control. people of forty plus, they want it to work, but as soo as soon as it works it's with them.

industrial_designer: that if we're if we put our marketing right we can sell this just like i if you 've heard about it in the news, the elderly mobile phone?

product_manager: it's a big success.

industrial_designer: if we make a remote control just l with that idea in mind, we could make tons of money,.

product_manager: very big success.

marketing_expert: i haven't heard of it.

product_manager: so as.

industrial_designer: we don't have to focus on the design then but on functionality. we just change our focus on the project, and we can sell this.

product_manager: i simply think that the new products we are gonna make, spef specifically design, are designed for younger people, so maybe we can focus ourself on the elderly people. and we have to see what requirements we need for those remote controls. what you told is the channel selection is important. volume selection, power and teletext.

marketing_expert: but the board tends to disagree.

product_manager: we haven't voted yet, teletext can be a function as. but only if it won't higher the cost, because i if it will be a lot more money to implement teletext as, but i don't think it will be a problem. or is teletext a

user_interface_designer: but deaf people need teletext for subtitles. so it's

marketing_expert: also.

product_manager: i suggest

marketing_expert: it'd definitely be a bad idea not to include teletext.

user_interface_designer: it's

product_manager: is anybody really against teletext? no? just that, that **we just keep the teletext³**. that's a good idea as, especially for the subtitles. maybe we can make that another point of advantage in our remote control, if we make a k a button ex for big subtitles, which is instantly on the remote control. for elderly people they can think, i wanna have subtitles, and they push the button and they get the big subtitles.

industrial_designer: that's a good idea.

product_manager: teletext can v can be very useful

in our advantage. functionality should be few buttons, you said. that's very important we have a few buttons.

marketing_expert: mm - hmm.

product_manager: so to keep it simple.

marketing_expert: but i don't think that's really an issue any more might be.

user_interface_designer: if it's only for televi

marketing_expert: but it, if it's only for t_v _ you're not gonna need a lot of buttons anyway. you need a one to zero button, next channel, previous channel, volume up, volume down, and some teletext buttons but

product_manager: but do you need

user_interface_designer: so we can s we can skip the display,

marketing_expert: if you only l

user_interface_designer: so we don't need it.

product_manager: but do you need the buttons for one to zero.

marketing_expert: nah.

product_manager: maybe c we can

marketing_expert: think if you're gonna include teletext you do. many people like to use that.

product_manager: maybe we can use

marketing_expert: if you should, if you want to switch from channel one to like thirty five, you don't wanna push the next channel button thirty five times.

product_manager: no, maybe we can implement the scroll button? or a joystick like? there are other ways too. just look if you look at telephones. the sony telephone has a scroll button which is very useful in searching names or

marketing_expert: that's true but i don't think there are many t_v_s that can switch channels that fast. and so you would need like the t_v _ would need an a function where you can actually view all channels and scroll through it. and if many channels would do have that. if many t_v_s have that.

industrial_designer: and besides that it's if we're gonna focus on elderly people they'll have to adapt. **they're not used to using scroll buttons. so perhaps we should s stick to the basic layout⁴.**

product_manager: the numbers. they can see how much buttons there are going to be on the display, and if it's too much we can reconsider it. but there won't be very much buttons. or there don't have to be a lot.

marketing_expert: but i don't think if you're gonna

make a remote control only to operate a t_v _, you there's not much you can gain on having as few buttons as possible. there are pretty many remote controls that can only operate a t_v _, which already only have the minimum number of buttons. i don't think there's much to be gained in that area. product_manager: the number of buttons? it's very important in the design. you can make a very fancy design with putting the buttons on the right places. and if you have less buttons you can do a lot more with

marketing_expert: that is true but there's simply not much to gain on the competition when you're making a remote control only for to operate only the t_v _.

product_manager: to operate only the t_v _.

marketing_expert: if you have a remote control only to operate a t_v _ there's simply not a lot of buttons required. there's not a lot of functions required so most existing remote controls simply don't have a lot of buttons either.

user_interface_designer: no.

product_manager: so.

marketing_expert: it would be very hard to actually gain on the competition here. that would that would cost a big marketing expedition

product_manager:

marketing_expert: which was one of the arguments to make it only for the t_v _ because we didn't have the time to market a lot.

product_manager: you suggest we could better focus on the docking station. like other functions. instead of f of less buttons.

marketing_expert: maybe., mean we need a good way to position all the buttons and but i don't think we should spend very much time in that.

product_manager: no. do you think the docking station will is allowed in the budget we have?

industrial_designer: it should be possible yes. if it's not too fancy.

product_manager: it can be

industrial_designer: and if the remote stays rather small, it should be possible

product_manager: because that's that's a good advantage point as. if we have a fancy - looking docking station

industrial_designer: yes.

product_manager: or very that's a requirement. docking station.

industrial_designer: we're just gonna focus on the extras?

product_manager: so.
marketing_expert: maybe we should do some research into what elderly people like to have in a like to have extra in a new remote control.
product_manager: that's a good point. you said they easily get lost as.
marketing_expert: fifty percent of the people indicated that remote control tended to get lost.
product_manager: so maybe we should implement the audio sign.,
marketing_expert: that was what i suggested.
industrial_designer: like with your key - chain, if you whistle it goes it makes a sound.
marketing_expert: you have it on
user_interface_designer: hm.
marketing_expert: you have it's on some phones too, which have a docking station. and you just press a button and the phone goes ringing.
product_manager: so
marketing_expert: so where it is.
product_manager: **audio signal should be possible as. it's not too expensive.**⁵ another point is the l_c_d _ screen. i if that will rise the cost too much, because
industrial_designer: y i **we'll have to choose between the docking station or the screen,**⁶
product_manager: it will be too much as.
industrial_designer: it's
marketing_expert: since a lot of people indicated that a new remote control is hard to learn, and we're focusing on elderly people here which tend to have a hard time understanding new devices, it might be a good idea to have just a little screen on it, which would explain a button if you press it, which would tell you what it does. and it wouldn't have to be touchscreen or a very expensive screen,
product_manager: based. just the l_c_d _ . just the normal screen.
marketing_expert: just a small screen
product_manager: that's a good idea.
marketing_expert: with two
product_manager: some extra info. feedback. that's a good idea as.
marketing_expert: but if that would
product_manager: as the small screen.
marketing_expert: that would fit into the costs.
product_manager: extra button info. that should be possible as. let's see what did we say. more. should be fancy to, fancy design, easy to learn. few buttons, we talked about that. **docking station**⁶, l_c_d _ . general functions and default materials.

that's a good idea as, because elderly people don't mind if it's a titanium cover or just a plastic one. so that doesn't really matter. we need
marketing_expert: probably elderly people would be a little bit more careful with their remote controls than youngsters.
product_manager: let's specify the target group. because are we talking about elderly people or people from forty to eighty. because what we're going to design now is for people above sixty, maybe. what do we want? if we want a with the telephone for elderly people, we can target the real elderly people.
marketing_expert: that would be a if we should do something like that it would be a, it would be really good for the image of the company. and, there would be a good market for it.
product_manager: that's the
marketing_expert: if we're able to really bring an innovative product.
product_manager: now you're talking about sixty to eighty.
marketing_expert: the really
product_manager: sixty.
marketing_expert: but i'd have to look into that a little more.
product_manager: and different cultures. are we
industrial_designer: i don't think they have different television sets in every country.
product_manager: we've got five minutes left just now. small warning.
marketing_expert: and with the little screen in it, which explains the buttons. you could we n it would be a lot easier to
product_manager: should
marketing_expert: adapt it to different cultures.
product_manager: in different languages,. or you have to put a language button in it,
user_interface_designer: right.
product_manager: but that will be a bit unnecessary.
user_interface_designer: no.
product_manager: it's better to put it on different markets with it all. that's the target. then a few small things. i will put the minutes in the project document folder. what we're going to do for the next meeting is the industrial designer will do the components concept, user interface designer the user interface concept, and the trend - watching. just keep in mind the things we've said about the target group, requirements, and the trends which

are going on. and specific instructions will be sent to you by the personal coach. i for this meeting. and we have a lunch - break now.
marketing_expert: that's good.
product_manager: that's a good thing.

Example 2:

In this example, our human raters gave mixed opinions on the succinctness, specificity, and faithfulness of our system's outputs, compared with the baseline's.

Output from Baseline System (DialogLED-L): the project manager presented the prototype to the team and discussed the features the prototype contained. the team then discussed the production costs of the remote and the possibility of adding a feature which would make the remote double-curved or single-curved. the project manager discussed the interior workings of a remote with the team members. they decided to make the buttons on the top, the menu on the bottom, and the parental control. the marketing expert gave an evaluation of the project process and found that the prototype was innovative and innovative and that the team had room to be creative. the user interface designer and the industrial designer presented their prototype, which was evaluated on a scale of 1 (excellent) to 8¹ (very bad). the team felt that the remote was over budget and that there was not enough room for creativity. they also complained about the energy source and the cost of the batteries. they complained about having to fill out a questionnaire in order to use the smartboard. after the evaluation, the team discussed their initial ideas about the features they would like to include in the remote. they eventually decided to include a feature called the touch-screen, which would allow the remote to be used with only one hand. the remote will be made of rubber instead of rubber. the corporate logo will be incorporated into the design. there will be no feature for the remote except for a feature to make it easier to find when lost in the house. the production cost will be under 12.50 euro. whether to add a feature on the remote or not is dependent on how much the cost is.

Output from Our System:

the project manager opened the meeting by stating the agenda and then giving the prototype presentation. the user interface designer and the industrial designer presented their prototype design for the remote, which was designed to look like

a children's remote. the marketing expert gave an evaluation of the prototype and found that the remote was technologically innovative but was not user-friendly². the group decided to make the remote single-curved or single-shaped. the remote has an lcd screen and is made of a special rubber casing. the designers discussed the features they would like to incorporate into the design, including a feature to make it easier to find when lost. the project manager then led the group in an evaluation in which the group evaluated the project process, finding that they were happy with their design but were unsatisfied with the first meeting. the team then discussed the production costs and what features they should retain or lose in order to stay under the 12.50 euro mark. after the evaluation, the group discussed whether the remote should be single- curved or double- curved. the remote will not feature a teletext button. the corporate logo will be displayed on the remote. the remote will feature a special color. the menu will be covered in a fruity, wood-like material. the buttons will be made in a special colour. the group had problems with the digital pen. the prototype was evaluated on a seven-point scale and received a 7.5. the original design was over budget but was able to meet the target user group. the new design was more suitable for younger users than the existing design was. the change in the color of the remote to use a more fanciful, more technologically innovative design did not satisfy the id c,

Source Text:

product_manager: wouldn't wanna be project manager., what we going to do., once again i'm gonna take minutes. so, no presentation for me., first we have a prototype presentation by g _ and g _ . afterwards some
user_interface_designer: yo.
marketing_expert: j _ and j _ .
product_manager: eval evalu
industrial_designer: evaluation.
product_manager: evaluation
user_interface_designer: evaluation criteria.
marketing_expert: evaluation.
product_manager: s. evaluation crit criteria., in combination with the finance i received a an excel file which we have to fill in later on., you see., and then we must see if we stay under the twelve and a half euro.
marketing_expert: interesting.,
product_manager: so, that's a big

industrial_designer: mm - hmm. that's gonna be t problem.

product_manager: l so let's it we have must,

user_interface_designer: some creative

product_manager: we must have some time for that because it will be, quite a lot of mathematics.

user_interface_designer:

product_manager: and after that, an evaluation of the process how we have done it here with the smartboard, with the with our laptops, with the all this. and afterwards, we closing. once again, forty minutes, so let's start. i would g give the word to g _ and g _ for the prototype presentation.

user_interface_designer: shall i give a short introduction and then

industrial_designer: .

product_manager: j _ and j _.

user_interface_designer: j _ and j _.

marketing_expert: jane and jane. guys, take it away.

user_interface_designer: take it away.

industrial_designer: hi.

user_interface_designer: this was our first concept. we decided to use a single touch - screen. so, we've worked out this concepts, how to hold it, where to put the buttons and. and, we began with a form of shape, that is easy to hold w in one hand, left or right handed. so, we made i it a little bit less thick and it has some ar artistic meaning. no? this isn't nothing. idea maybe is better., during the meeting i showed you the concept of placing the buttons on top, usable with your thumb, and the menu structure, if necessary, with your other hand, so it's just gonna hold it easily. and it has to be acce accessible with your other hand too,. so we began working out a concept.

industrial_designer: and as you saw, we would just have the basic remote with the panel l_c_d _ screen., these would be the main buttons, h you could change them later on in your own profile if you want to. but, it's standard they will be delivered with this set - up. we have the more advanced menu setting right here. we have the sub - menus and stu. we made a top, or a front view. just so like you wanna back view. as you can see, this there, there are two weird bumps in it. this is for the added effect of y youth and dynamic. and this is for the artistic effect., what we figured is we'll show you a picture later on you have more b a better idea after that. but, idea is for to stay in balance with these two. and so when you put

it on the table, it will just lay down. it won't roll around or. but it will lie more in your hand like an old telephone maybe, or like these old phones. y you may get the idea. so thi this is about how we figured it should be. the s panel we g you would hide with some more rubber layers, like we discussed early on., you would s you wouldn't see the straight panel, but more fluidly and round.

user_interface_designer: the panel just goes like this. but the overlaying layer is a little bit curved and.

product_manager: no,.

industrial_designer: and, in these bumps you could actually put some electronics that would you can make a more thinner design, and that would actually look very,. and, about the colour, what have

user_interface_designer: we added that this can be held with your hands for this maximum is om, one and a half centimetres. so, you have room here for your battery and maybe even other electronic chips. s and you can just be the layer of the touchscreen and some have some wires underneath it to make it as thin as possible in the middle for good grip.

industrial_designer: f, as colours, do you have the picture in now, this is the idea about the bumps., you can see there's a v a very youthful dynamic exterior. it you just want to hold it you are young and dynamic like us.

marketing_expert: 's l it's like an easter egg.

industrial_designer: it's like an e but this is for children. we we want a more adult version. but, this is like a remote control for children.

product_manager: it's called a weemote

industrial_designer: a weemote.

marketing_expert: weemote.

product_manager: weemote.

industrial_designer: hey, that's actually a brilliant marketing stand., but

marketing_expert: what i w got in mind.

industrial_designer: so this actually basic the idea. we we just want to build a more adult vers adult version of this.

product_manager: imagine that.

marketing_expert: mm - hmm.

industrial_designer: and and for colours, we figured starting with basic colours like white or metallic grey. those are the technological colours actually,

user_interface_designer: it would be best to appeal to a broad public and make the covers

exchangeable,
industrial_designer: so it d
user_interface_designer: so the young people will buy an orange and a red and blue and a purple,
industrial_designer: or blue or whatever.
user_interface_designer: but when the o older people go in the shop and they see an orange remote control, it would be less appealing than a white one. and young people, we think, are a little bit more flexible, they think, i'll buy for a couple of euros some noi hip
marketing_expert: maybe it's an idea to sell it without a cover, so that you can pick a cover in the shop.
user_interface_designer: a cover is necessary, als otherwise you'll just have the l_c_d_ screen.
marketing_expert: .
user_interface_designer: so, there must be some cheap standard cover, maybe white,
marketing_expert:
user_interface_designer: that's could comes with it and you can buy, so we can make extra money.
product_manager: but you d you mustn't forget that our target aim is younger people.
marketing_expert: oui.
product_manager: we had decided to put some flashy fruity colours in it, and in the survey from milan and paris it came out that the d the older people are more willing to spend money on extra features. so it will be a better idea to have some flashy fruity colours as a standard,
user_interface_designer: the other way around, you mean.
product_manager: and for the people who really want a more sophisticated, more traditional look, they're willing to pay that. they want they want more luxury, but they have the money to do it and they want to b to buy that.
user_interface_designer: mm - hmm.
product_manager: so, maybe it's an idea to put that as an extra and not as a standard.
industrial_designer: , maybe, perhaps you're right., i would actually agree with this sounds logical.
user_interface_designer: .
marketing_expert: an another idea., maybe we could develop a cover with wood style. they'll the elder users as.
user_interface_designer: a colour of a wood style, a white c and a couple of h hip fruity colours. and lea l delivered standard with a fruity colour, but not too much.

industrial_designer: nah.
marketing_expert: yes. not not too.
user_interface_designer: this is banana and mango, not purple or p orange and yellow.
marketing_expert: exactly.
product_manager: but, the mai th the standard must be some attractive flashy colours.
marketing_expert: or blue or
product_manager: not too, but w a little,
user_interface_designer: mm - hmm.
product_manager: because that's our aim.
industrial_designer: li like this. this isn't this isn't too much, is it?
user_interface_designer: . no.
industrial_designer: i f
product_manager: the buttons don't have to be all of
industrial_designer: the buttons,
marketing_expert: so.
industrial_designer: i
marketing_expert: except for the buttons it's it could be a standard model.
product_manager: it
industrial_designer: something like this would be., that's it from us.
marketing_expert: it's my time now.
user_interface_designer: it's my turn.
product_manager: the marketing expert.
industrial_designer: uh - oh.
marketing_expert: during the design life - cycle we made lot of requirements and trend analysis and., now is the time to evaluate our prototype concept to the past requirements. so we are going to evaluate the design according to the past user requirements and trends analysis., we're going to do that with a **seven point scale**¹. opening a word document now. one, i have to expla explain something. we have to be consensive about things. so, it has to be a group decision.
product_manager:
marketing_expert:?
product_manager: so we gon we gonna evaluate the
marketing_expert: we're going to vote. we
product_manager: the thing we saw.
marketing_expert: the prototype.
product_manager: just saw.
marketing_expert: one. the remote control is designed for people with age below forty.
product_manager: seven?
marketing_expert: seven is false.

product_manager: true.
marketing_expert: b one or
product_manager: one.
industrial_designer: why?
marketing_expert: most true?
industrial_designer: it's not just designed for people under the age of forty. it's also designed for people above forty.
marketing_expert: so
industrial_designer: so
marketing_expert: so a o one is appropriate?
user_interface_designer: no no, a little more in the middle.
marketing_expert: or, more like a four.
user_interface_designer: no, three or.
industrial_designer: i have i've
marketing_expert: three.
industrial_designer: two or three, because it's not just the qu question is aimed at is it designed for people with age below forty. but it's also designed for people of age above forty. so,
marketing_expert: exactly.
industrial_designer: i'll say it's about three.
user_interface_designer: it will be primary appealing to m minus forty, but also appealing to
marketing_expert: but also for,, second. the remote control is beautiful.
user_interface_designer: it's
marketing_expert: acco according to us, it's one? or
user_interface_designer: it's the marketing angle on television. we have a wonderful
marketing_expert: p s of c you have to be very positive and enthusiastic about your own product.
user_interface_designer: it's also fancy then.
marketing_expert: the remote control looks fancy.
industrial_designer: yes.
marketing_expert: one?
user_interface_designer: we have a perfect remote.
marketing_expert: good. four. the remote control has big, clear channel switching buttons.
user_interface_designer: yes. they have to agree but
product_manager: yes.
industrial_designer: leads to user face.,
user_interface_designer: i'm the user interface expert.
marketing_expert: daniel., teletext buttons and volume buttons?
user_interface_designer: no.
product_manager: no teletext buttons. teletext is in

the menu.
user_interface_designer: you you've different menu.
industrial_designer: false.
marketing_expert: false?
user_interface_designer: and volume is impo
marketing_expert: and volume?
product_manager: volume is true.
marketing_expert: true. big and clear?
industrial_designer: the they are big and clear.
user_interface_designer: big and clear.
product_manager: big and clear.
user_interface_designer: but you could make a teletext button six.
marketing_expert: hey.
user_interface_designer: otherwise, the people who read this are gonna think we have no teletext button.
marketing_expert: hide.
industrial_designer: but the teletext button., you can ch that's in a menu.
marketing_expert: it's it's not
industrial_designer: so, it's w, it
marketing_expert: it
industrial_designer: it isn't entirely unclear,
marketing_expert: j
industrial_designer: but so, i wouldn't give it a seven.
user_interface_designer: no.
industrial_designer: i would give it a more a five or a six.
marketing_expert: five?
industrial_designer: i don i. what do you think, mister project manager?
marketing_expert: it's.
product_manager: ., i agree. i was thinking very black and white.
user_interface_designer: black and red.
product_manager: j _.
user_interface_designer: don't forget to save it.
marketing_expert: red. volume. the remote control is easy to be found.
user_interface_designer: when we put in fancy colours,
product_manager: fruity.
user_interface_designer: and
industrial_designer: it has these all these fruity colours and it has a strange shape. so, if you so if you have trouble finding it
user_interface_designer: but, it's not making any sound, have we deciding?

marketing_expert:
user_interface_designer: so
marketing_expert: but if you put your normal remote control under your bed, or you throw this remote control under your bed, is it better findable?
user_interface_designer: it'll make a difference. we have the better re i., so. my remote control's black.
marketing_expert: a li little bit maybe?
user_interface_designer: a little bit, but.
product_manager: we p we can do it glow in the dark.
marketing_expert: four? fi
product_manager: so, if it's in the dark place, you still see it glowing.
user_interface_designer: k.
marketing_expert: i
user_interface_designer: fo fo five is.
marketing_expert: five. it's it's it doesn't really make a lot of
industrial_designer: then i'll go for four. because four is between three and also between true and false.
user_interface_designer: you're right.
marketing_expert: yes, but five is between four and six.
industrial_designer: so i'll go for four.
product_manager: you must see it as, w according to the other remote controls, there may be there in your t_v _ room, this one will stand out.,
industrial_designer: wha
marketing_expert: b _.
industrial_designer: that's a better question actually.
product_manager: exa that's what it's about.
marketing_expert: it's
user_interface_designer: if your fifteen remotes in a drawer, you find it,?
product_manager: if it if this lying on your couch, you're you think what's that for kinda orange thing. so
marketing_expert: but but the survey under users was that they really lost it.
user_interface_designer: that's stupid.
marketing_expert: like, no not seeing it, but lost it in the house.
user_interface_designer: but when you lost it you're just not
marketing_expert: but.,
industrial_designer: if i if you see a strange shape lying somewhere, then you'd recognise it as, whoa,

that is strange.
product_manager: that's our remote control.
user_interface_designer: mostly when you lose your remote control, it's under your
marketing_expert: ., i agree.
industrial_designer: what is that.,
user_interface_designer: most of times when you lose it you're sitting on it.
industrial_designer: so it's
marketing_expert: eight, the remote control has fresh, fruity colours.
product_manager: true.
user_interface_designer: i would call choose two, we decided not to make two f fresh colours, as it would not.
marketing_expert: not too flashy. the remote control is made of soft material.
industrial_designer: rubber, is soft.
product_manager: but not too soft we have decided.
user_interface_designer: kinda soft, but not this.
marketing_expert: three?
product_manager: three.,
user_interface_designer: easy to use,
product_manager: easy to use. **one²**.
user_interface_designer: **very afford²**.
marketing_expert: easy to use?
product_manager: can it be zero?
industrial_designer: i don, it is
marketing_expert: **top easy to use?²** it's it's not the most easy to use
user_interface_designer: you can do two, because
industrial_designer: no.
marketing_expert: it can be easier.
user_interface_designer: it can be easier. but then you're l
industrial_designer: it could.
marketing_expert: jus just with ten buttons, that's the easiest.
user_interface_designer: but then you'll lose function f, functionality and our fancy look, so.
industrial_designer: functional ability.
marketing_expert: but the most easy to use is just with one button
user_interface_designer: but it is r it is rather easy to use, because you have the primary buttons always visible.
marketing_expert: on t, but easy n not the most easy to use.,
industrial_designer: no, it's it i'll go for two. my vote's on two.

marketing_expert: two?
product_manager: two.
user_interface_designer: m mine too.
marketing_expert: we also have to compare it to the remote controls on the market nowadays. so
user_interface_designer: but waits just a minutes. inspiration.
product_manager: what's the time? we also have to do the evaluation, the production costs and.
user_interface_designer: these are the m regular remotes.
marketing_expert: i'm hurrying., eleven. the remote control is innovative.
user_interface_designer: yes, true, one.
product_manager: you're agree, tim?
marketing_expert: a very
product_manager: you haven't seen a more innovative thing in paris?
marketing_expert: the remote control has m remova removable from multilux.
user_interface_designer: yes, one. very multifunctional.
marketing_expert: the remote control, i it has speech recognition.
product_manager: false.
industrial_designer: yes,
marketing_expert: false.
industrial_designer: it
user_interface_designer: this is used with speech recognition, this.
marketing_expert: the remote control has built - in games?
user_interface_designer: but, maybe make it two, because the games are in a sub - menu and not it's not an entire game.
product_manager: but they are built in, so it's one.
industrial_designer: they are built in. not down.
marketing_expert: and the last, paren parental advisory function.
industrial_designer: you really like the parental advisory.
product_manager: freak.
marketing_expert: yes, i do.
user_interface_designer: th did you make this or the
industrial_designer: bu
marketing_expert: save as.
industrial_designer: he made it.
user_interface_designer: it changes it maybe.
marketing_expert: i will
user_interface_designer: yay.

marketing_expert: do the math.
user_interface_designer: dear.
marketing_expert: now it's your turn.
product_manager: we'll see.
user_interface_designer:?
product_manager: we have now to c to calculate the production cost. if it's under twelve and half euro, then it's. but i is it if it is b? no, this isn't right. so,. redesign.
user_interface_designer: if they're under twelve fifty.
product_manager: if they under. no., it's
user_interface_designer: ? cau so it's.
product_manager: if the costs are under twelve and a half euro, then we can ra move on to the project evaluation, as we have experienced it. otherwise, we have do have to do a little redesign thingy. so, we have to fill in the numbers of the component components. we have to fil, want to do it in and see if we stay under the twelve and a half euro. so, do we have a hand dynamo? that's zero.
industrial_designer: me, too.
user_interface_designer: battery, yes. one.
product_manager: battery, one?
user_interface_designer: one.,
product_manager: kinetic, one?
user_interface_designer: kinetic, one.
industrial_designer:
product_manager: solar cells, zero., simple chip on print?
industrial_designer: n no.
marketing_expert: no.
product_manager: no? no, advanced chip.
marketing_expert: advanced chip.
product_manager: sample sensor sample speaker?
user_interface_designer: no, the advanced chip is
marketing_expert: advanced chip is three.
product_manager: three?
marketing_expert: three euros.,
product_manager: but it's one thing, it's three euro.
industrial_designer: we have one. we have one
marketing_expert: one piece.,
product_manager: what's the sample sensor?
industrial_designer: no, sev zero., that's
marketing_expert: speech recognition.,
user_interface_designer: you give it a sample, one.
industrial_designer: speech recognition and s
product_manager: zero., uncurved flat.
industrial_designer: a zero.
product_manager: but is it s it's not made from a

single uncurved thingy and then
user_interface_designer: you no.
industrial_designer: no.
marketing_expert: thingy.
product_manager: no?. so it's only once double -
curved.
marketing_expert: yes,
user_interface_designer: the layer around it fits
around the bubbles on the o on the back of the
marketing_expert: three. eight.
product_manager: we're now in a problem, we
have reached eleven euro yet.
user_interface_designer: we don't have anything
else. so
marketing_expert: go on. just go on.
product_manager: but we have
industrial_designer: just go on. then we'll see wha
how much we are over budget.
user_interface_designer: could you step a little to
the right ma
marketing_expert: two.
product_manager: rubber.
user_interface_designer: or
product_manager: you.
industrial_designer: zero.
product_manager: titanium, no?
user_interface_designer: and zero. special is the
special colour?
product_manager: special colour?
industrial_designer: i don't. no, this is a standard
colour.
product_manager: but we want to make the wood
colours, that
marketing_expert: s
industrial_designer: yes, this is a special colour.
user_interface_designer: if you're honest, we'll
type one, special colour.
marketing_expert: but d but daniel, tha that's
another brand.
industrial_designer: that's an add - on.
marketing_expert: that's another article to sell.
product_manager: but we going to, that's true. but,
it's it
user_interface_designer: j maybe we'll finish the
list first and then look back, aye?
marketing_expert: that doesn't account for this.
producing this.
product_manager: the push - button, no.
user_interface_designer: scroll wheel, no.
product_manager: scroll - wheel, no. yes, one.,
button, no. no, the

user_interface_designer: is it
product_manager: we don't have a s
user_interface_designer: no.
industrial_designer: these three.
product_manager: no.
industrial_designer: we're only four euro over
budget.
product_manager: . so, what's the thing we can
change?
industrial_designer: but, other case, we can make
it single - curved or uncurved.
product_manager: can i say something?
user_interface_designer: single - curves.
product_manager: no, can i say something as
project manager? the kinetic thing, can we just
skip it,
user_interface_designer: just cut off the kine
product_manager: because you have to shake it,
but that's not really innovative.
industrial_designer: .,
user_interface_designer: we just put a good battery
it. mobile phones nowadays.
marketing_expert: daniel. daniel,
product_manager: yo.
marketing_expert: what do you think about here.
product_manager: yes.
marketing_expert: what do you think about putting
a battery in it, but also selling like the covers, a
docking station just apart from the thing, so that
you can put
product_manager: mm - hmm.
marketing_expert: rechargeable batteries in it and
just
user_interface_designer: but you can use recharge-
able batteries anyway, just you s you have to
recharge them manual.
product_manager: and not really.
marketing_expert: .
product_manager: but we if you forget about the
kinetic,
marketing_expert: just an idea.
user_interface_designer: that's a cost reduc
product_manager: if we do that, we shall.
industrial_designer: you can go from double -
curved to single - curved. and that would solve the
budget problem.
product_manager: b but i but the single - curved is
just
user_interface_designer: so we have to bake the ba
back flat, and then no, it's just one curve and not a
back curved. or

industrial_designer: it's just, the single - curve that exactly.,
product_manager: so that's wh tha that's one option.
user_interface_designer: or are these two curves?
product_manager: and then w, and then we could have it, but it's its' it's r it is the main point of the look.
user_interface_designer: but what else do we have to cut out? no advanced chip, that's a little bit of problem.
industrial_designer: we going to cut
product_manager: no, tho that can be done. so,
user_interface_designer: although, can we make it with a regular chip?
product_manager: a little less conversation.
user_interface_designer: curvy.
marketing_expert: hey, those ar arcs, why are there for? the blue
product_manager: fill in just a
user_interface_designer: explanation.
product_manager: explanation.
user_interface_designer: twelve fifty.
product_manager: delete it for you if you want.
marketing_expert: no, no.
product_manager: so, if we do this, we're on twelve and a half euro. and we're done.
user_interface_designer: but does it fit with our design?
product_manager: the only thing that don't
user_interface_designer: do we have to u adapt it? it's single - curves.
product_manager: single - curved, but there's a curve in it. so
user_interface_designer: w could we just make the bubbles cut off the back, and then we're has
industrial_designer: we just make it flat. but, you do l
marketing_expert: but, wha, look, what is the if you make it double - curved, it costs one euro more.
user_interface_designer: more. you make it optional.
marketing_expert: but no, but does it have a lot of extra
industrial_designer: function.
user_interface_designer: functional.
marketing_expert: fun function more like
industrial_designer: worth, does it have added worth?
user_interface_designer: there's an a athe aesthetic

value, but not functionality.
industrial_designer: no,
product_manager: it's really a static value.
marketing_expert: aesthetic., you make like eleven and a half euros profit instead of twelve and a half. but i if twelve and a half is a fixed price.
product_manager: it is.
industrial_designer: let's assume it is.
marketing_expert: no, we can't go above that.
industrial_designer: we we should assume it i that it is. but i figured that the kinetic would be a marketing promotion.
marketing_expert: . then it's.
industrial_designer: r if you promote a kinetic i kinetic remote control, that would b sell better than an a normal remote control.
product_manager: do you think?, now you can shake your remote control.
marketing_expert: so.
industrial_designer: no, y you can go into your neighbour and tell him, ha, my k remote control is kinetic.
product_manager: kinetic.
industrial_designer: you have standard old battery control remote con
marketing_expert: what a what about all the m the environment freaks?
user_interface_designer: but it doesn't fit in our cost profile. so
marketing_expert: not freaks, the envi
industrial_designer: true.
marketing_expert: it look like this one.
user_interface_designer: you ma can make an specialised extra gold version.
product_manager: who because if you want to go to kinetic, you're on thirteen and a half and you must go to flat, and now it's more of a compromise thing. and if you make the single curve ha just a big curve, then it's
user_interface_designer: just one big curve.
product_manager: one big good curve. i was going to say nasty words, but i don't. this is strange. wood is m is cheaper than rubber. we thought that wood would be more expensive.
user_interface_designer: this american figures. you just cut down some trees.
product_manager: maybe. but that this is it?., this is it.
marketing_expert: this is it.
user_interface_designer: whoever makes a remote control out of titanium.

product_manager: i'm gonna save it.
industrial_designer: it is possible, but you can't use double curves for titanium. that's one of the functionality
product_manager: considering we have
marketing_expert: no, we have to do all those hours again. go back. one back? costs on no redesign.
product_manager: we were above, so we did a little redesign
user_interface_designer: we sue. we, we'll start her all o all over again.
product_manager: now it's about time to talk about this project., some things. were there room for was there room for creativity in our meetings or in your individual meetings?
marketing_expert: i didn't. that there was a lot of room for it. but, that's mainly because of the information that was delivered to us. it was just fixed information and get your from there, and i couldn't go on i on the internet and search my own. bu
product_manager: that's true. i agree with that.
industrial_designer: i th you two, especially you and daniel, you d you both had the less creative roles in the project.
product_manager: that's true.
user_interface_designer: for us, there was a lot of creativity.
industrial_designer: m
user_interface_designer: we could just sign up an remote if we liked.
industrial_designer: jeroen and i, we had a more design we could have more we had more room for creativity than you two.
marketing_expert: , that's true.
product_manager: how about the leadership?
industrial_designer: ha.
marketing_expert: leadership was crappy.
user_interface_designer: crappy. cra
industrial_designer: nah, the leadership wasn't crappy, it was the leader that was crappy.
product_manager: very much.
industrial_designer: nah. no, the leadership was.
product_manager: now we're done.
user_interface_designer: example of crappy leadershi
marketing_expert: no, leadership was
industrial_designer: it was good.
marketing_expert: the first meeting was a little bit of unstructured meeting., you could have but, it

was your first, no disrespect, but you could have structure it a l little bit more. so like, i was talking most of the time the first meeting, and
user_interface_designer: you could of said, shut up you fool.
product_manager: i notice it too. i was a i was also very unhappy, very unsatisfied about the
marketing_expert: about me.
product_manager: about the first meeting. so, i hope the other meetings
industrial_designer: try to learn from your mistake. and we will never do it again.
marketing_expert: you made up.
product_manager: get better and the last two meetings also we reached some good decisions about talk
marketing_expert: no, it you did better.
user_interface_designer: more consensus. ev everybody w was agreeing every
marketing_expert: much more constructive.
product_manager: so that's., teamwork?, maybe that's only, it's for us, because
user_interface_designer: we work together on a project, but everybody has his own task.
product_manager: and it's wo more like presentation and some points were discussed.
user_interface_designer: so, it is a little bit
product_manager: but, really teamwork were you two
user_interface_designer: no, it went.
marketing_expert: two guys.
product_manager: the
industrial_designer: that went w it went. it's it's just
user_interface_designer: stupid stupid pen, but
product_manager: no hard feelings.
industrial_designer: nah., we had some trouble with the pen, but
product_manager: now you must push a little while.
user_interface_designer: but draw something difficult.
industrial_designer: but us
user_interface_designer: d just write your name right now. try to write your name, in writing letters,?, normally, this the w block letter sign it,?
product_manager: o just
user_interface_designer: just just write your name in one line. if it's a little bit too small bit quicker now.
product_manager: you can be you can go quicker,

then it won't notice it.
user_interface_designer: it didn't he knows how it works.,
product_manager: i follow the master class for the smartboard, so that's the main issue., so about this one you were you're dealing with,
marketing_expert: means.
product_manager: the digital pen.
user_interface_designer: y., th the i the idea is great, but it doesn't work properly.
product_manager:
marketing_expert: digital pen, th the first time i did individual work, i used it. but, a and the first two meetings i brought it with me, but i didn't use it after the first meeting.
industrial_designer: no, i have it working. but,
marketing_expert: it's it's not real use for me.
industrial_designer: no, it doesn't have that much added value to the
user_interface_designer: nee. as as i said a m a c few moments ago, it i would like, myself, to write with a normal pen, because must, it's almost the same concept, but you can just sim more simply put it on our scanner. i it's the same concept as the pen, where you f have to download the software or s very slow.
product_manager: m. and it is still your own handwriting popping up in word.
user_interface_designer: yes. no, and it doesn't give any added value.
product_manager: no, that's true. no.
marketing_expert: not really, no.
product_manager: and the smartboard is useful, but the pen is i not user - friendly.,
user_interface_designer: not user - friendly.
product_manager: be it takes a lot of time to draw things and to write things,
user_interface_designer: and it's not very precise.
product_manager: and that's the
user_interface_designer: we're trying to m to
marketing_expert: like when you do this.
user_interface_designer: it may, and tr try to wri write your name in a normal size.,. smaller.
marketing_expert: smaller?
user_interface_designer: smaller. just like when you're writing on a letter.
product_manager: but that's not th the you when you at a foreign audience, you b don't gonna wr write small.
user_interface_designer: a as you saw on this drawing, just open this one or that one. it's th it,

we had more problems even here when we trying to draw these buttons, it's almost impossible to get clear when you're
product_manager: but maybe there's some function with no, it isn't. with
user_interface_designer: and the eraser was another problem. it w t is this large. and when you try to erase this line, y
product_manager: i'm gonna erase my name.
user_interface_designer: i'm gonna erase my name there.
product_manager: it's a big eraser.
marketing_expert: new ideas?
user_interface_designer: m abo what new ideas?
product_manager: the idea of the touch - screen is
marketing_expert: do you? go on.
product_manager: i'm just
user_interface_designer: new ideas about the working of this software, about the project, about the remote controls or
marketing_expert: , so.
product_manager: i what.
user_interface_designer: hello.
marketing_expert: did you heard what he said? i don't i what., i have some figure. here. the eva the evaluation, the mean number is one point eight six.
product_manager: that's interesting.
marketing_expert: so that's fairly good.,
product_manager: because what does it mean?
user_interface_designer: so true.
marketing_expert: that
product_manager: all the mo, are between one and two.
marketing_expert: all the requirements are true or very true, right.
product_manager: ,, expert. but the new ideas found for wi with working with this software?
user_interface_designer: not really, just they have to improve it.
product_manager: not really,?
user_interface_designer: the concept is, but it has to be quicker., it is still opening my programme, n almo almost my entire computer is locked up during the process and it, just takes too many time. people will still feel the need to write it quickly on a page and not download it and save it, and
product_manager: you had expected it to be more
user_interface_designer: more user - friendly. when you use a pen, you can just draw like you d draw normally, and you do
marketing_expert: may maybe the idea you

proposed is a screen here. and draw it, and it's placed over there.

user_interface_designer: that I would be easier. or at least when you don't have to adapt to the technology, just you can write in the way you normally write. and now you have to keep constantly in mind that you're drawing on this screen.

marketing_expert: ? no.

user_interface_designer: and that's a very bad concept.

marketing_expert: very bad.

user_interface_designer: nah, i it's my opinion that i

marketing_expert: this is better than regular flip - overs,

user_interface_designer: it's can be saved easier. but if you're in normal flip - over you a lot of people write text. there's no text option. and writing text, you've got ta really do your best to write some

product_manager: and maybe some functions for circle or a square. you have to draw it yourself now.

user_interface_designer: or maybe even insert picture. if you have some presentation, and you have some f

industrial_designer: or text function. just t type text, and that would be excellent. mean

marketing_expert: but insert image isn't available?

user_interface_designer: then you could

marketing_expert: here. picture from scanner, clip - art.

product_manager: that can be done already. but not the predefined squares

user_interface_designer: so you can

marketing_expert: hyperlink? hey, what if you do like hyperlink?

user_interface_designer: with w_w _ dot google dot com.

marketing_expert: type type it?

user_interface_designer: maybe.

marketing_expert: re real reaction dot n_l _ . yes, is now is.

user_interface_designer: you'll just make a link in product_manager: that's.

user_interface_designer: there's one way to

product_manager: is it

marketing_expert: double - click it.

user_interface_designer: maybe if you're not using the eraser

product_manager: here.

marketing_expert: you're erasing.

user_interface_designer: something else th, arrow.

product_manager: here, that.

marketing_expert: double - click it.

product_manager: so you have

user_interface_designer: it's

product_manager: as you saw, you have a little, you can you can go there. so there the functionality is there, but it's not ideal, and it's very it costs a lot of time to

user_interface_designer: to use,.

product_manager: to use. and that's a pity, if you have thirty, forty minutes for this things, and we are now with four people,

user_interface_designer: and that's m

product_manager: but it, imagine you are here you're with the ten people and everyone

user_interface_designer: that's mostly the case, from the over here with the managements you get two minutes to make your case, and if you have to do all this kind

product_manager: two minutes of drawing,.

user_interface_designer: you'll rather use power-point and work it out in advance. and th the one or two things you have to draw when you're there, just use a flip - board.

marketing_expert: what i really miss also is a d is a turtle is a decision system like with the evaluation, you have to polls like, what do you want, a one, a two, maybe a l a little application like give your own number and click one two three four five six seven.

product_manager: just like he said with the a screen which you can write, also a voting mechanism.

marketing_expert: j ju, v voting application. just a little group decision application.

industrial_designer: but, problem is, you can't discuss anything you ca you can, but you will discuss a lot less than I like we did now. we w one person s maybe said three. but, i we said, no i w th think two, because this and this, and then you can react on it. but if you put a three on it, just figure, everybody knows what i'm knowing, so they'll all just put a two on.

marketing_expert: . but, you can still discuss about it,

industrial_designer: but it would.

marketing_expert: but click it in an application, that's a lot easier to process.

industrial_designer: the for processing part.

marketing_expert: the digit.

industrial_designer: but then, the idea of one person entering it and the rest discussing it, that isn't that bad idea, actually. not that your opinion isn't valued, but still.

product_manager: so

user_interface_designer: yo, manager.

marketing_expert: that's it?

product_manager: just about,

user_interface_designer: when are w

product_manager: because

user_interface_designer: when are we going to produce it?

product_manager: tomorrow?, the costs are within the budget.

marketing_expert: celebration.

product_manager: the project is evaluated. but, before we going to celebrate, i have a little question which you can't answer, because there must be some end report. i am busy with the end report right now. you might thinking what the hell was he doing

marketing_expert: what is an end report?

product_manager: about all the meetings, what we have decided, a r a report of this day., that must be made, but i, here is standing whoa, we can celebrate now, but the end report is

marketing_expert: you ha you have ten minutes left, i read. you have now ten minutes left to finish up the end report.

product_manager: that can be done. maybe we can do it together. you can see what i've so i s i will put it on a story - board. you can see it. because it will it must be

marketing_expert: you you already made a beta version, or

product_manager: it's a three with seventy five

marketing_expert: pages.

product_manager: just about.

industrial_designer: y yikes. seventy five pages.

product_manager: just a moment. end report.

marketing_expert: daniel. do you want a chair maybe?

user_interface_designer: a chairman.

product_manager: no no,

marketing_expert: hey?

product_manager: you can s you can read it and here it is. end report.

industrial_designer: so you finished it actually, and so we just have to read it and say yes or no?

product_manager: this not nit it read - only. but it's not fully finished yet.

marketing_expert: five minutes for finishing.

product_manager: this is about the functional design, the things.

marketing_expert: management expert, you have to change that.

product_manager: i'm when i said it, i remember i had it here.

marketing_expert: marketing.

user_interface_designer: it's a read - only version.

marketing_expert: but you can save it u the under another name.

product_manager: marketing expert, about the three functions where which are most used and which must immediately be visible on our remote control., it must be simple to use, very clear what to do, and at the younger people. so, this is really about wh what kind things must be in it and

marketing_expert: maybe the

product_manager: i understand you, talk a little bit dutch.

marketing_expert: the you. no, you have to put, switch channels at the top, because that's the most used function and teletext at the second

product_manager: oka, i really didn't knew that.

marketing_expert: nay, a volume changing, second.

product_manager: so, this one's first.

marketing_expert: s switch, yes.

product_manager: you go there and you go there. so.,

marketing_expert: go on.

product_manager: maybe then do it one two three.

marketing_expert: yes, very good.

user_interface_designer: one two three.

product_manager: if the order is in is important, that's the word for.

marketing_expert: the order.

product_manager: then the conceptual design., all the things we have discussed, the energy, which turn out to be batteries, so that's

user_interface_designer: maybe you can add it later that we decided in the end because of the cost.

product_manager: because., he here it is still double - curve, the rubber, the flashy, the fruity, the removable.

marketing_expert: single - curves.

product_manager: the buttons

industrial_designer: it's not double.

user_interface_designer: a single - curved.

industrial_designer: it's not double anymore,?
product_manager: not double anymore.
marketing_expert: no,.
product_manager: nay but that this is what
marketing_expert: was initial, the plan. the initial
plan.
product_manager: and the added functions like
tetris snake, it's under the parental control, the
touch - screen. so, it's just a summary of what we
had discussed
industrial_designer: on thing one small thing, the
added functions., was it included in the cost? i
don't,?
user_interface_designer: it's very cheap.
product_manager: it's very cheap,
industrial_designer: it's very cheap.
product_manager: you maybe you come at
marketing_expert: no, it's not very cheap, but that
product_manager: it's very necessary.
marketing_expert: no, but it's a development
inside corporation. like, w we don't have to buy
parental control. our own people can make that,.
product_manager: but it still has some
marketing_expert: it has some cost, but
product_manager: , but so we can discuss that
user_interface_designer: he'll do it in his free time.
so
product_manager: we can discuss that things with
the board of
user_interface_designer: of directur or directors.
product_manager: what's it what's the company
called? keep forgetting it.
marketing_expert: real reaction.
product_manager: real reaction. real reaction,.
user_interface_designer: you can ask your personal
product_manager: so anybody misses something
here about
user_interface_designer: the end conclusion.
product_manager: that's what i'm gonna write b
between now.
marketing_expert: still the end conclusion. that's
all,.
product_manager: but, i in here nothing
user_interface_designer: the decision to make the
buttons on the top, and the menu on the
product_manager: mm - hmm. yes.
user_interface_designer: on the bottom. and
clearly, touch - screen you've mentioned.
product_manager: touch - screen i've mentioned.
marketing_expert: but the decisions are put in the
conclusion, . why we decided to use a flat l_c_d _.

user_interface_designer: this n.
product_manager: i gonna redesign something
now.
user_interface_designer: party party.
product_manager: because it will, five minutes
from to finish meeting.
user_interface_designer: before you change
anything maybe you save it first.
marketing_expert: save it.
user_interface_designer: you can't you can select
file and export and then those j _ peg files.
marketing_expert: but then you had the same
thing.
user_interface_designer: eight. and we have,
another blank one.
marketing_expert: example of children remote.
user_interface_designer: did we change anything?
it's a new commercial logo.
marketing_expert: that's a pity.
user_interface_designer: don't save it, aye?
product_manager: that's, tim. finish meeting now.,
i'm
user_interface_designer: why are only the first five
smartboard files
product_manager: i'm going to finish my end
report.
user_interface_designer: saved?
industrial_designer: hereby is the meeting is
finished.
marketing_expert: you declare.
product_manager: i am the one who can say that.?
hereby the meeting is finished.

D Used Artifacts

We include the list of artifacts we use. They are all open-source and we followed their intended use.

- Pyrouge (1.5.5):
<https://github.com/bheinzerling/pyrouge>
MIT license
- PyTorch (1.8):
<https://pytorch.org>
BSD-3 license
- Transformers (4.25):
<https://github.com/huggingface/transformers>
Apache License 2.0
- NetworkX (2.7.1):
<https://github.com/networkx/networkx>
BSD-3 license

ACL 2023 Responsible NLP Checklist

A For every submission:

- A1. Did you describe the limitations of your work?
Limitations Section
- A2. Did you discuss any potential risks of your work?
Ethics Section
- A3. Do the abstract and introduction summarize the paper’s main claims?
Abstract; 1. Introduction
- A4. Have you used AI writing assistants when working on this paper?
Left blank.

B Did you use or create scientific artifacts?

2, 3, 4, 5

- B1. Did you cite the creators of artifacts you used?
References
- B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
C
- B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
C
- B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
Not applicable. Left blank.
- B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Not applicable. Left blank.
- B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
Section 4 and 5

C Did you run computational experiments?

Left blank.

- C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
4, A

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?

4, A

- C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?

5

- C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?

C

D Did you use human annotators (e.g., crowdworkers) or research with human participants?

Left blank.

- D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?

B

- D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?

Not applicable. Left blank.

- D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?

5, B

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

Left blank.

- D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?

B