DeepCon: An End-to-End Multilingual Toolkit for Automatic Minuting of Multi-Party Dialogues

Aakash Bhatnagar[$], **Nidhir Bhavsar**[&], **Muskaan Singh**[#], **Petr Motlicek**[#]
[$]Boston University, Boston, Massachusetts
[&]Navrachana University Vadodara, India
[#] Speech and Audio Processing Group, IDIAP Research Institute, Martigny, Switzerland
aakash07@bu.edu, nidbhavsar989@gmail.com,(msingh,petr.motlicek)@idiap.ch

Abstract

In this paper, we present our minuting tool *DeepCon*, an end-to-end toolkit for minuting the multiparty dialogues of meetings. It provides technological support for (multilingual) communication and collaboration, with a specific focus on Natural Language Processing (NLP) technologies: Automatic Speech Recognition (ASR), Machine Translation (MT), Automatic Minuting (AM), Topic Modelling (TM) and Named Entity Recognition (NER). To the best of our knowledge, there is no such tool available. Further, this tool follows a microservice architecture, and we release the tool as open-source, deployed on Amazon Web Services (AWS). We release our tool open-source here http://www.deepcon.in.

1 Introduction and Related Work

Due to the COVID-19 pandemic, a substantial part of the working population has seen a significant increase in virtual meetings, especially people working in Information Technology (IT) industry and academia. By all means, meetings are the most vital component to ensure collaborative work and efficient to-and-fro communications. Natural Language Processing (NLP) technologies provide users with a holistic experience in these online interactions. (i) remote conferences or meetings discussions held over an online platform are extremely important in today's globalized world and need interpretation. (ii) Coherent translations of larger documents and dialogues and efficient systems with many sources or target languages. (iii) Summarizing meetings in the form of structured minutes from speech can potentially save up to 80%[1] of time. With all these in mind, we designed an easy-to-use and clean interface that provides (i) Automatic Minuting with dynamic length controlled outputs. (ii) Isometric Translation for five different languages: French, German, Russian, Italian, and Hindi (iii) Topic Extraction

To the best of our knowledge, there is no such tool available. However, some applications, such as Deeptalk[2], provide the fastest way to transform text from chats, emails, surveys, reviews, and social networks into a real business. It is a tool for end users that provides an interactive user interface for topic detection, sentiment analysis, auto-tagging, analytical interpretation, and summarization. However, this tool lacks in providing users with audio and video support. Wordcab[3] intelligence adds customizable call summaries to applications so that users can revisit conversations in a fraction of the time. It is developer tool that allows developers to integrate their API and easily generate different types of summaries. This tool focuses on generating complex and customized summaries from the given transcript. Happyscribe[4] provides automatic and human transcription services convert audio to text with 85-99% accuracy in 120+ languages and 45+ formats. It provides strong API integration that enables users to handle multi-lingual input. This tool is primarily built for generating transcripts from audio files or automatic subtitles. Happyscribe does not provide any summarization and topic modeling capabilities that limit this platform's scope. Hendrix AI[5] is your intelligent, award-winning AI assistant for meetings that automatically transcribes meeting notes, captures action items and other data points, and uses machine learning to identify productivity insights. Hendrix AI is a highly analytical tool for Zoom meetings. It gives users various outputs such as concise summaries, actionable outcomes, most commonly discussed topics, and meeting effectiveness. Hendrix AI is limited to the Zoom platform's

[1] https://elitr.github.io/automatic-minuting/

[2] https://www.deep-talk.ai

[3] https://wordcab.com

[4] https://www.happyscribe.com

[5] https://hendrix.ai/features

scope and cannot be used with other platforms and offline recordings.

(A) Meeting transcript segment:

(PERSON0) Mhm. I will try to get my presentation. All right. Yeah. Mm hmm. Mhm, mm hmm. Hi guys. Sorry. Okay. Excuse me. Mhm, mm hmm. Okay. And then the person. Mm hmm. It is, you know? Yes. Okay. Mhm. But I'm prepared. I'm going in there. Right.
(PERSON2) Mm hmm, mm hmm. Oh, wow,
(PERSON0) amazing.
(PERSON2) It's working. Okay. Thank you. Mhm, mm hmm. Yeah, mm hmm. Okay.
(PERSON1) Yeah, you can flip it. Mhm. Mhm. So, good morning everyone.
(PERSON2) Um
(PERSON2) hmm. Mhm. Mhm. So, the general outline of the, the project will be the first to go through a functional design phase. You all get task in this in this space and ah then we will meet again and discuss this functional design And the same holds for the two faces. After this, the conceptual design and detailed design In with the final project should get 76. Alright. But first we will do from training. Okay. In all, in front of you,
(PERSON4) you are designed we must keep in mind that The selling price of the product will be about €25

(B) Meeting minutes by DeepCon:

- PERSON0 will try to get his presentation
- PERSON3, PERSON2, Person1 and Person0 are attending the kick off meeting of their latest project
- PERSON1, PERSON2 and PERSON3 discuss the general outline of the project
- The project will go through a functional design phase
- After this, conceptual design and detailed design in with the final project should get 76
- PERSON2, Person3, Person4 and Person0 are designing a project for their company
- The selling price of the product will be about €25

(C) Isometric Translated French Minutes:

- PERSON0 va essayer de faire sa présentation
- PERSON3, Person2, Person1 et Person0 assistent à la réunion de lancement de leur dernier projet
- PERSON1, Person2 et Person3 discutent des grandes lignes du projet
- Le projet passera par une phase de conception fonctionnelle
- Après cela, le design conceptuel et le design détaillé avec le projet final devraient avoir 76
- PERSON2, Person3, Person4 et Person0 conçoivent un projet pour leur entreprise
- Le prix de vente du produit sera d'environ 25 €

(D) Generated Topics:

- tools to communicate
- actual project plan
- functional design phase
- million to make
- thinking television targeting

(E) Translated French Topics:

- outils pour communiquer
- plan de projet réel
- phase de conception fonctionnelle
- millions à faire
- cibler la télévision

Figure 1: An example from our DeepCon tool showing (A) a segment of a meeting transcript (B) along with corresponding generated minutes (C) isometric translated French minutes (d) generated named entity (e) translated French named entity. We have utilized the AMI meeting corpus and anonymized, "PERSON*number*" and "PROJECT*number*" denote persons' and projects' placeholders, respectively.

2 DeepCon

In previous sections, we discuss the various tools that provide similar features as DeepCon. Most of the tools are made as an API service for developers. A tool like Deeptalk, accessible directly to users, does not provide audio and video support. DeepCon, however, provides an easy-to-use interface for end users to utilize advanced transformer models without coding. DeepCon also provides users with multiple features and support of audio and video files on one platform, which are not provided by various other tools.

As seen in Figure 2, DeepCon have various different components. In the following subsections we elaborate on each of the major components in our proposed system.

2.1 Automatic Speech Recognition (ASR)

For generating meeting transcripts, we use Amazon Transcribe [6] which is a Speech-to-text service offered by Amazon AWS. For English, the speech recognition model achieves a WER of 6.2%. The ASR-generated transcripts follow time-sequence order, with both speaker and utterances stated separately. In our DeepCon, we define a post-processing function that aligns speaker roles with corresponding utterances, as shown in figure 1. The user can set a range of $\{2, 10\}$ speakers.

2.2 Automatic Minuting

The meeting summarization module generates minutes, given a transcript. Minuting is primarily concerned with capturing and providing a third-person perspective of essential points raised throughout the meeting. Manual minuting also has drawbacks where the minutes' format and language vary through different annotators. Our tool, DeepCon provides an end-to-end solution to generate consistent and robust meeting minutes.

We use a finetuned BART-large model (Lewis et al., 2019) [7]. We test various summarization models such as T5 (Raffel et al., 2019), Pegasus (Zhang et al., 2019a), RoBERTa2RoBERTa (Liu et al., 2019). However, the BART-based pipeline outperformed the others. This could be because BART utilizes GPT-2 architecture. Further, we fine-tuned

[6] https://aws.amazon.com/transcribe/

[7] https://huggingface.co/lidiya/bart-large-xsum-SAMSum

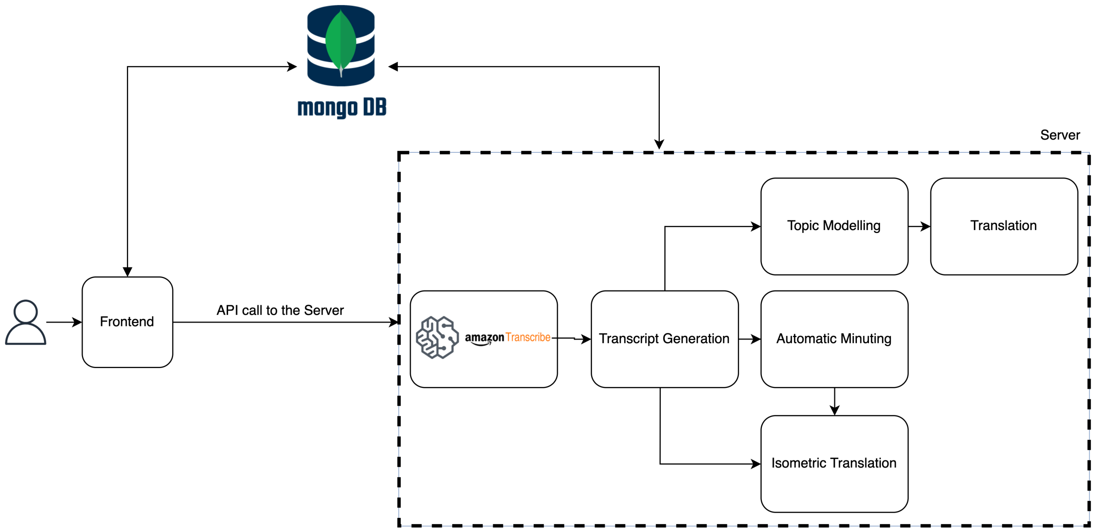
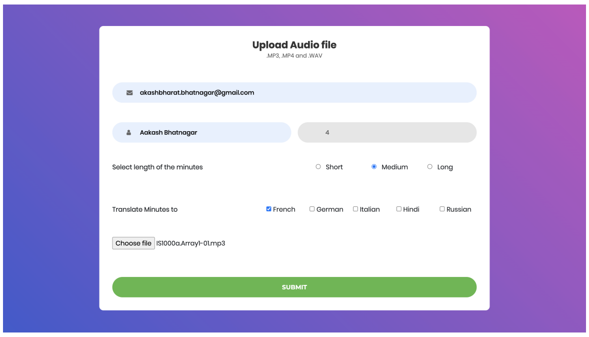


Figure 2: This system architecture diagram represents the pipeline of DeepCon. The first step is for the user to submit the audio file with the desired attributes. This audio file is uploaded on AWS S3, and the link is updated on our MongoDB database. Then, our back-end processes the audio file and generates transcripts using Amazon Transcribe. Further, we use our fine-tuned model for automatic minuting, isometric translation, and topic extraction.



Upload Audio file
.MP3, .MP4 and .WAV

akashbharat.bhatnagar@gmail.com

Aakash Bhatnagar

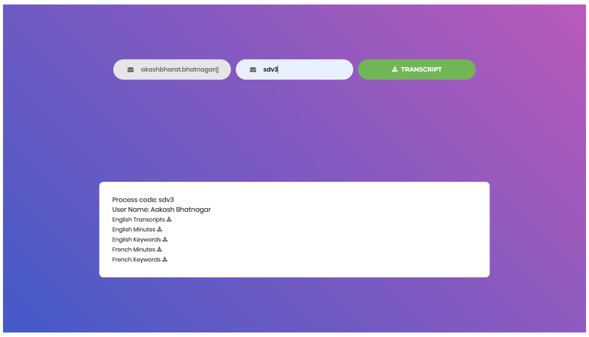
4

Select length of the minutes ○ Short ● Medium ○ Long

Translate Minutes to ☑ French □ German □ Italian □ Hindi □ Russian

Choose file IS1000a.Array1-01.mp3

Figure 3: Interface for users to select parameters and upload their audio files. On this page, users enter a name and email ID where they want to receive the process code notification. Users can also select from 3 types of summaries, i.e., short, medium, and long. There is also an option to choose 5 languages per user's requirement.



akashbharat.bhatnagar@

sdv3

Process code: sdv3
User Name: Aakash Bhatnagar
English Transcripts
English Minutes
English Keywords
French Minutes
French Keywords

Figure 4: This interface is accessed by the user when the processing of their file is completed. Users can enter the process code here and get the files. All the files here are in text format and can be downloaded easily.

this model on both XSum (Narayan et al., 2018) and SAMSum (Gliwa et al., 2019) datasets. XSum dataset includes short summaries of articles and discussions, whereas SAMSum is a standard dialogue summarization dataset. Training over these datasets provides the BART model the robustness to generate short precise summaries of conversations.

As depicted in Figure 5 Automatic Minuting functionality is divided into 3 major segments. i) We analyze and apply various preprocessing techniques to the generated ASR transcripts, including segmenting the input text into much smaller chunks. (ii) We apply the summarization using a finetuned BART model. (iii) Finally, we use an unsupervised redundancy elimination method to obtain ideal minutes.

The current summarization algorithms are not trained to remove redundancy from a long dialogue discourse and are also restricted to a specific input length for improved text production. Thus to eliminate such redundancies, we specify a few custom rules. We try to eliminate repetitions, pauses, and vocal sounds. We also remove stopwords defined using the publicly available AMI corpus.

We evaluate the generated minutes on the SAMSum Corpus. Table 1 provides scores obtained on the validation & test set measured across various automatic evaluation metrics including R-1, R-2, R-L and R-Lsum (Vasilyev et al., 2020). It is evident that our models achieve higher evaluation scores

	rouge 1	rouge 2	rouge L	rouge Lsum
Validation	54.39	29.81	45.15	49.94
Test	53.31	28.35	44.09	48.92

Table 1: Evaluation results of the summarization model on SAMSum validation and test dataset. We use the ROUGE score for evaluating automatic minutes generated from the text.

Team	rouge 1	rouge 2	rouge L
Auto Minuters	0.25±0.06	0.06±0.03	0.14±0.04
Hitachi	0.26±0.09	0.08±0.03	0.15±0.04
Ours	0.33±0.08	0.08±0.04	0.19±0.06

Table 2: Automated evaluation scores for the best performing system at the AutoMin 2021 shared task.

across all metrics.

Table 2 compares our results on the test set with the two of the best performing system submissions in the AutoMin Shared Task (Ghosal et al., 2021). As depicted our system outperforms the Yamaguchi et al. (2021) & Mahajan et al. (2021).

2.3 Isometric Machine Translation

The Machine translation module allows users to generate transcripts, minutes, and topics in five languages. For all these languages, we provide users with isometric translation output. Isometric MT is the concept of generating translation output that falls within the range of ±10% of the Length Ratio (ratio of the target text and source text). This feature helps to generate synchronous outputs upon text-to-speech conversion. For implementing isometric translation, we develop a multitask learning model similar to Bhatnagar et al. (2022). We use fine-tuned OPUS-MT (Tiedemann and Thottingal, 2020) model for translation and fine-tuned mBART (Liu et al., 2020) for paraphrasing. We use WMT (Bojar et al., 2018) and MuST-C (Di Gangi et al., 2019) dataset for fine-tuning MT models, and PAWS-X (Yang et al., 2019), Opusparcus (Creutz, 2018) and Tapaco (Scherrer, 2020) dataset for paraphrasing. We utilize the IIT-B Hindi-English (Kunchukuttan et al., 2017) dataset for En-Hi translation.

As shown in Figure 6 our system architecture for Isometric Machine Translation utilizes prompt engineering technique during the machine translation & paraphrasing of input text. (i) For translating the input text, we try to maintain a target-to-source length ratio close to 1. This is worked out using the verbosity control feature, where in the finetuned OPUS-MT model tries to localize the source text based on the pre-calculated LR ratio using the predefined short, normal, long prompts. (ii) We also utilize the paraphrasing module to enhance the vocabulary and modify the length of already translated text. For generating output we apply "normal" prompt during the translation phase and reverse-prompts during the paraprhasing phase. Reverse-prompts are applied to alter the length of the translated sentence. We use similar method as done by Bhatnagar et al. (2022).

Language	BLEU Score	BERT Score	Length Ratio	Length Range (%)
de	29.9	0.83	1.05	51.95
it	34	0.84	1.04	57.03
fr	41.2	0.85	1.04	61.81
ru	21.7	0.83	0.97	62.47
hi	11.9	0.84	0.94	42.52

Table 3: Results obtained by the Isometric Machine Translation module on MuST-C test dataset, evaluated using the BLEU metric, BERT score, Length Ratio & Length Range.

We evaluate our MT system for the language de, fr, it & ru on the MuST-C test dataset and use the IIT-B Hindi-English test dataset for hi. As shown in Table 3, we are able to get high BERT score (Zhang et al., 2019b), length ratio, and length range for the language ru, fr & it. As shown in the table 4 we also compare our system using the BLEU metric evaluated on the same MuST-C dataset. We compare our results with that proposed by the Wang et al. (2020). As shown, our proposed system outperforms by a subsequent margin.

	fr	de	it	ru
fairseq S2T T-Sm	32.9	22.7	22.7	15.3
fairseq S2T Multi. T-Md	34.9	24.5	24.6	16.0
Ours	41.2	29.0	34	21.7

Table 4: BLEU Score evaluated on MuST-C test dataset and provides a systematic comparison between the fairseq S2T (Wang et al., 2020) and our proposed system. Here T-Sm and Multi. T-Md are both transformer-based models, later being trained jointly on 8 Languages.

2.4 Topics Modeling

DeepCon also provides a feature for automatic topic extraction based on Named Entity Recognition that extracts the top-k repeating n-grams from the transcripts. We use Yake[8] library for extracting named entities. Our system also supports multilinguality as a feature for generated keywords. Upon

[8] https://github.com/LIAAD/yake

generation, the system can apply translation across all the languages mentioned earlier.

3 Pilot Study

To assess DeepCon, we experiment it with IS1000a recording of the AMI Meeting Corpus (Carletta et al., 2005) that contains 4 speakers. The first step to use DeepCon is for users to login using their credentials. If users are not registered they can use the sign-up link[9] on the landing page. Next as shown in the Figure 3 users can upload the *.mp3* or *.mp4* file and select the desired attributes. Users can also add name, email, and number of speakers. We have provided some advanced options that can control the length and translations of the audio file. Users can select the length of meeting-minutes from the three options: *short*, *medium*, and *long*. Users can also select amongst five languages for translation. Once users choose the appropriate options, they can upload the audio file via the upload button at the bottom and click on submit. After clicking the submit button, our front-end micro-service uploads the recording on the AWS S3 bucket and sends a *post request* to our back-end micro-service that contains the user's details along with the unique process code. Once this process is done, users get a confirmation email about the submission of the job. Once the back-end processing is completed, users again receive an email notification from our back-end microservice. This notification informs users that their processing is completed. Users can also download the outputs from our results page as seen in 4. The link of this page is also mailed to the user.

We gave the users a feedback form and question based on transcript quality, adequacy, grammaticality correctness, and fluency for English minutes. We received an average score of 4.57, 4.71, 4.57, and 4.85 out of 5, respectively. For MT quality estimation, we provided the users with the questions for quality in French, German, Russian, Hindi, and Italian Translation and received an average score of 5, 4.83, 4.8, 5, 4.6 out of 5, respectively. The average quality of generated topics is 4.85 out of 5, and the overall user interface received a score of 4.85 with an additional comment of "Improvements can be made in its ability to differentiate between the voice of respective speakers. Also, there was a slight deviation from the actual words in the transcript, that can be improved as well".

4 Design Choices

As mentioned in previous sections, we make use of microservice architecture. We utilize this methodology for three main reasons: As we utilize large pre-train models, it is not an efficient choice to do real-time processing for users. Because every meeting recording can vary in length and users can select many optimization options, real-time processing can take considerable time and slows down the website for other users. With recent advancements in DevOps technology like Kubernetes, microservice architecture has proven to be the most efficient, fault-tolerant, and robust deployment architecture. We use AWS EKS[10] as a container orchestration tool to deploy a highly scalable application. Microservice architecture also enables an easy development process of the application. As we have independent services, changing one will not affect others, and it can lead to a faster and more efficient software development process. We also plan to make this app open source, and we believe microservice architecture will enable developers to work in the module of their interest.

5 Conclusion and Future Work

In this paper, we present a tool for better management of meeting recordings by providing users the ability to generate meeting minutes, topics, and entities in six different languages. Our development architecture is designed in a way that it can be scaled and optimized if traffic increases on the application. We can also extend this application to accommodate more languages like Spanish, Romanian, Telugu, etc. The microservice architecture can further be abstracted by introducing a microservice for each functionality. This abstraction can result in more robustness and efficiency during high workloads.

This application can also be extended as an API service for developers to integrate in their systems. One major application can be of building native apps for online meeting platforms like Zoom, Microsoft Teams, and Google Meet

6 Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 833635 (project ROX-

[9] https://forms.gle/Wsbe6ASdggQNxjsHA

[10] https://aws.amazon.com/eks/

ANNE: Real-time network, text, and speaker analytics for combating organized crime, 2019-2022).

References

Aakash Bhatnagar, Nidhir Bhavsar, Muskaan Singh, and Petr Motlicek. 2022. Hierarchical multi-task learning framework for isometric-speech language translation. In *ACL*.

Ondřej Bojar, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Philipp Koehn, and Christof Monz. 2018. Findings of the 2018 conference on machine translation (WMT18). In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, pages 272–303, Belgium, Brussels. Association for Computational Linguistics.

Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, and Mccowan Wilfried Post Dennis Reidsma. 2005. The ami meeting corpus: A pre-announcement. In *In Proc. MLMI*, pages 28–39.

Mathias Creutz. 2018. Open subtitles paraphrase corpus for six languages. *CoRR*, abs/1809.06142.

Mattia A. Di Gangi, Roldano Cattoni, Luisa Bentivogli, Matteo Negri, and Marco Turchi. 2019. MuST-C: a Multilingual Speech Translation Corpus. 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 2012–2017, Minneapolis, Minnesota. Association for Computational Linguistics.

Tirthankar Ghosal, Ondřej Bojar, Muskaan Singh, and Anja Nedoluzhko. 2021. Overview of the First Shared Task on Automatic Minuting (AutoMin) at Interspeech 2021. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 1–25.

Bogdan Gliwa, Iwona Mochol, Maciej Biesek, and Aleksander Wawer. 2019. Samsum corpus: A human-annotated dialogue dataset for abstractive summarization. *CoRR*, abs/1911.12237.

Anoop Kunchukuttan, Pratik Mehta, and Pushpak Bhattacharyya. 2017. The IIT bombay english-hindi parallel corpus. *CoRR*, abs/1710.02855.

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.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *CoRR*, abs/2001.08210.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Parth Mahajan, Muskaan Singh, and Harpreet Singh. 2021. Team AutoMinuters @ AutoMin 2021: Leveraging state-of-the-art Text Summarization model to Generate Minutes using Transfer Learning. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 34–40.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. *CoRR*, abs/1808.08745.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *CoRR*, abs/1910.10683.

Yves Scherrer. 2020. TaPaCo: A corpus of sentential paraphrases for 73 languages. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6868–6873, Marseille, France. European Language Resources Association.

Jörg Tiedemann and Santhosh Thottingal. 2020. OPUS-MT – building open translation services for the world. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480, Lisboa, Portugal. European Association for Machine Translation.

Oleg V. Vasilyev, Vedant Dharnidharka, and John Bohannon. 2020. Fill in the BLANC: human-free quality estimation of document summaries. *CoRR*, abs/2002.09836.

Changhan Wang, Yun Tang, Xutai Ma, Anne Wu, Dmytro Okhonko, and Juan Miguel Pino. 2020. fairseq S2T: fast speech-to-text modeling with fairseq. *CoRR*, abs/2010.05171.

Atsuki Yamaguchi, Gaku Morio, Hiroaki Ozaki, Ken ichi Yokote, and Kenji Nagamatsu. 2021. Team Hitachi @ AutoMin 2021: Reference-free Automatic Minuting Pipeline with Argument Structure Construction over Topic-based Summarization. In *Proc. First Shared Task on Automatic Minuting at Interspeech 2021*, pages 41–48.

Yinfei Yang, Yuan Zhang, Chris Tar, and Jason Baldridge. 2019. PAWS-X: A cross-lingual adversarial dataset for paraphrase identification. *CoRR*, abs/1908.11828.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2019a. PEGASUS: pre-training with extracted gap-sentences for abstractive summarization. *CoRR*, abs/1912.08777.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2019b. Bertscore: Evaluating text generation with BERT. *CoRR*, abs/1904.09675.

A Appendix

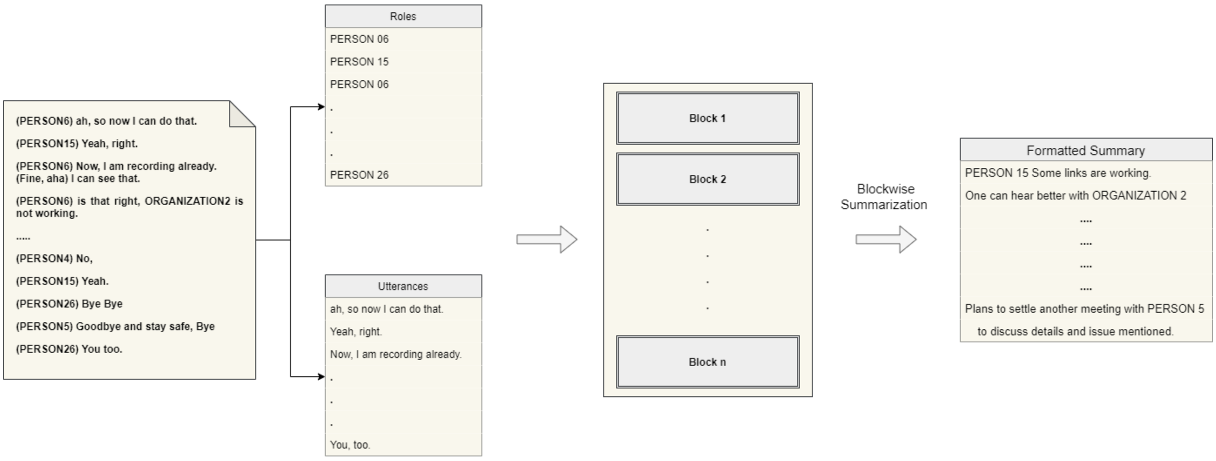

(PERSON6) ah, so now I can do that.
(PERSON15) Yeah, right.
(PERSON6) Now, I am recording already. (Fine, aha) I can see that.
(PERSON6) is that right, ORGANIZATION2 is not working.
.....
(PERSON4) No,
(PERSON15) Yeah.
(PERSON26) Bye Bye
(PERSON5) Goodbye and stay safe, Bye
(PERSON26) You too.
Roles
PERSON 06
PERSON 15
PERSON 06
PERSON 26
Utterances
ah, so now I can do that.
Yeah, right.
Now, I am recording already.
You, too.
Block 1
Block 2
Block n
Blockwise Summarization
Formatted Summary
PERSON 15 Some links are working.
One can hear better with ORGANIZATION 2
Plans to settle another meeting with PERSON 5 to discuss details and issue mentioned.

Figure 5: This figure is an architectural representation of the automatic minuting functionality. Here the ASR output is segmented into multiple chunks of text according to the sequence length accepted by the BART model. Next, the finetuned BART model process these chunks. Finally, we stack these chunks and perform redundancy elimination to generate the meeting minutes.

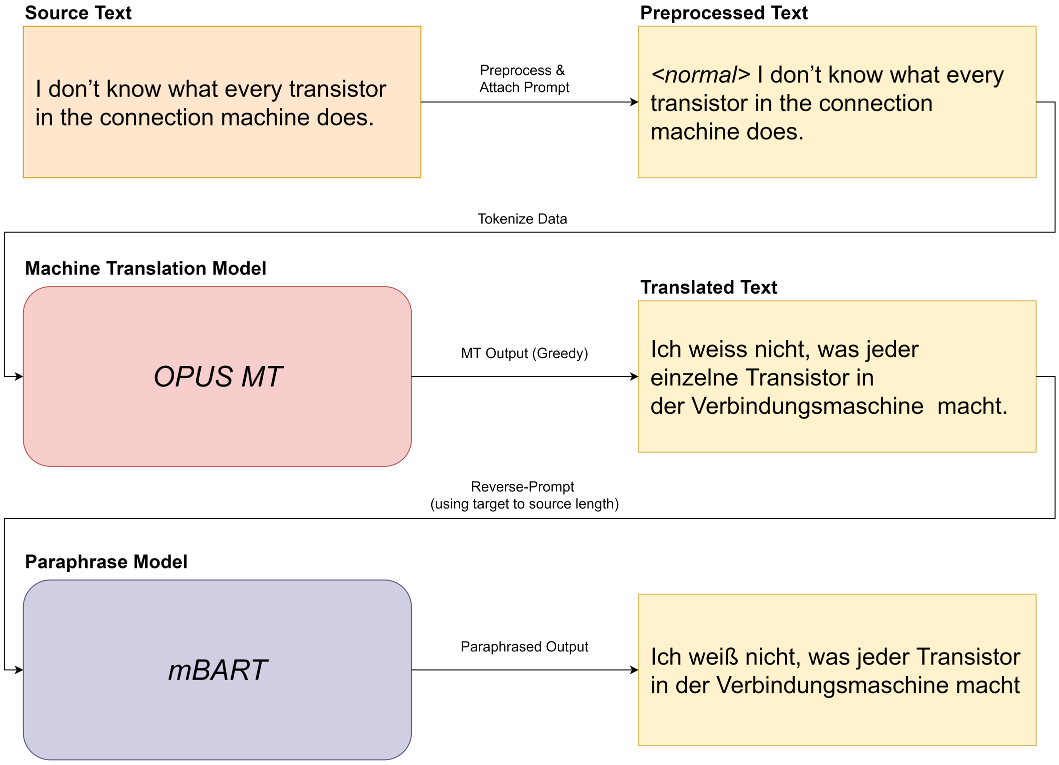

Source Text
I don't know what every transistor in the connection machine does.
Preprocess & Attach Prompt
Preprocessed Text
<normal> I don't know what every transistor in the connection machine does.
Tokenize Data
Machine Translation Model
OPUS MT
MT Output (Greedy)
Translated Text
Ich weiss nicht, was jeder einzelne Transistor in der Verbindungsmaschine macht.
Reverse-Prompt (using target to source length)
Paraphrase Model
mBART
Paraphrased Output
Ich weiß nicht, was jeder Transistor in der Verbindungsmaschine macht

Figure 6: This figure shows the pipeline we use to generate isometric MT outputs from finetuned BART models. The first step, as seen, is attaching a 'normal' prompt to the input source sentence. This will help the translation model generate output with a length ratio close to 1. The next step is to attach a reverse prompt and send input to the paraphrasing module. Based on these reverse prompts, the paraphrasing module tries to shorten the long sentences and lengthen the short ones.