A REVIEW ON ABSTRACTIVE MEETING SUMMARIZATION: TOWARDS SMARTER MEETING ASSISTANTS

Kratika Sudarshan Halankar, Kiran Kumar S, Khushi R, Kiran Kumar
Student, Department of Information Science and Engineering,
Alvas Institute of Engineering and Technology, Moodbidri, Karnataka, India.

Guide: Mr. Pradeep Nayak
Assistant Professor, Department of Information Science and Engineering,
Alvas Institute of Engineering and Technology, Moodbidri, Karnataka, India.

ABSTRACT:

A system capable of accurately identifying and summarizing the key points of a conversation would be highly beneficial across various real-world scenarios, including business meetings, medical consultations, and customer service interactions. Recent advancements in deep learning, particularly the emergence of encoder-decoder architectures, have greatly enhanced language generation systems, paving the way for improved abstractive summarization, a technique especially effective for summarizing multi-party discussions. This paper provides an overview of the key challenges associated with abstractive meeting summarization and reviews the datasets, models, and evaluation methods that have been developed to address these challenges.

Keywords: Abstractive summarization, Smart meeting summarizer, Deep learning, Encoder-decoder architecture, Natural language processing, Meeting transcription, multi-party conversation, Automatic summarization, Artificial intelligence, Speech-to-text systems.

I. INTRODUCTION:

Meetings are an essential part of workplace communication but often demand considerable time and organisational effort (Mroz et al., 2018). Research indicates that employees and managers in the United States spend between 6 and 23 hours weekly in meetings (Rogelberg et al., 2007). With the widespread adoption of virtual meetings during the COVID-19 pandemic, sessions have become longer and more frequent, resulting in fatigue and reduced retention of information (Fauville et al., 2021). These challenges have created a growing need for intelligent tools capable of automatically generating concise, actionable summaries of meeting discussions (Edmunds and Morris, 2000; Elciyar, 2021).The concept of automatic text summarisation, first introduced in the 1950s (Luhn, 1958), has evolved into two main approaches: **extractive** and **abstractive** summarisation (Gambhir and Gupta, 2017). Extractive methods identify and select the most relevant sentences from the source material, whereas abstractive methods powered by neural architectures like encoder-decoder models (Sutskever et al., 2014; Vaswani et al., 2017) generate new, fluent sentences that convey the same meaning (See et al., 2017; Lewis et al., 2020). While initial research focused primarily on structured texts such as articles or reports, recent attention has turned toward meeting summarisation, an area that stands to benefit significantly from advancements in deep learning and natural language processing.

However, meeting summarisation presents unique difficulties compared to traditional text summarisation. Conversations in meetings are often unstructured, involve multiple speakers, contain overlapping dialogue, and require understanding of context, speaker intent, and roles (Kryscinski et al., 2019). These complexities make it necessary to develop specialised methods tailored to spontaneous, multi-party discussions rather than merely fine-tuning existing text-based summarisation models (Zechner, 2002).

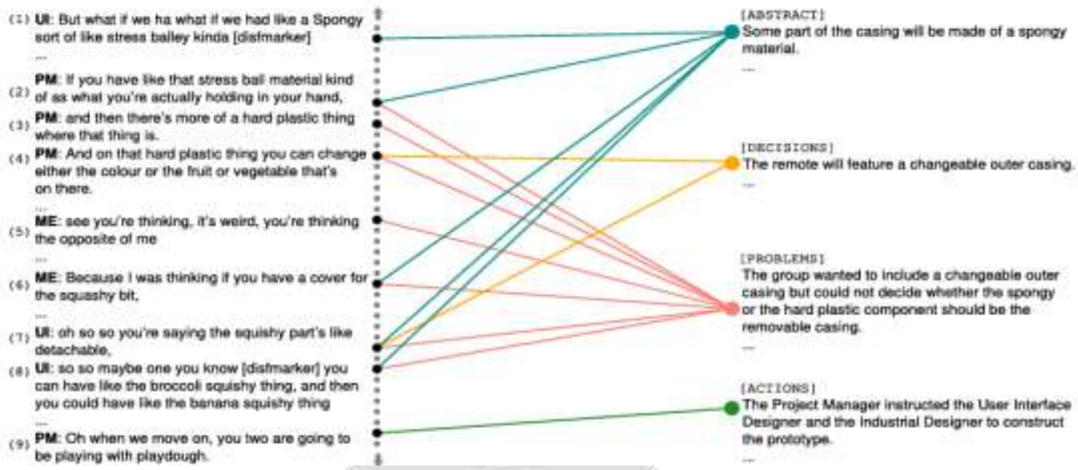

Figure 1: An example (Shang et al., 2020a) of ground truth human annotations from the ES2011c AMI meeting. The grey nodes on the left represent transcripted utterances, while the black nodes represent annotators' selection of relevant utterances for summary. One sentence from each part of the abstractive summary appears on the right. The coloured arcs connect a sentence from the abstract summary to a specific abstractive community, which is the set of clauses from the extractive summary summarised by the sentence in the abstract summary.

In **Section 2**, the paper outlines the main challenges involved in analysing and summarising meeting-style conversations. **Sections 3 to 5** then delve into the existing tools and techniques designed to overcome these challenges, focusing on **datasets**, **evaluation methods**, and **summarisation systems** (Shang, 2021). Each area brings its own complexities, and the discussion highlights the advantages and shortcomings of various existing approaches. The paper concludes by identifying promising **future research directions** that could enhance the development of more effective meeting summarisation methods.

II. The Difficulties of Meeting-style Presentations and Summaries

Figure 1 illustrates the key characteristics of meeting-style interactions and recurring themes often found in meeting summaries. The examples are drawn from the **AMI corpus** (Mccowan et al., 2005), a widely used dataset containing simulated workplace discussions among four participants: a project manager (PM), marketing expert (ME), user interface designer (UI), and industrial designer (ID). These conversations document the collaborative design of a television remote control, following the team from concept generation to final product development.

The figure compares **extractive** and **abstractive** human-generated summaries for one meeting (ES2011c). Coloured links show how sentences in the abstractive summary correspond to related extractive statements. This highlights how **abstractive summarisation** merges information from multiple utterances into concise and coherent representations of the discussion’s main ideas.

1. The Nature of Meeting-style Speech.

Meeting dialogues are **spontaneous and unstructured**, occurring in real time as participants collaborate. As a result, transcripts often contain pauses, repetitions, incomplete thoughts, and informal expressions. These natural speech patterns reduce clarity and make the text noisier. Even after editing, the transcripts retain some of these imperfections, reflecting the informal nature of human conversation.

Meeting transcripts also tend to be **longer and more detailed** than typical written materials. For example, an AMI meeting averages about **4,757 words**, while its summary contains roughly **322 words**. In contrast, a CNN/Daily Mail news article averages **781 words** with a **56-word** summary (Hermann et al., 2015). This difference illustrates the scale and complexity of summarising spoken interactions.

Another difficulty arises from **multi-speaker participation**. Meetings involve several individuals with varying communication styles and overlapping dialogue, making speaker identification and context tracking challenging. Furthermore, personal references like "I" or "you" must be transformed into **third-person expressions** so that summaries are clear and understandable to readers who were not present.

2. **The Preference for Abstractive Summarization.**
   Earlier systems relied mainly on **extractive approaches**, which identified key sentences directly from transcripts (Tur et al., 2009; Riedhammer et al., 2008). While effective for structured documents, these methods often produced fragmented summaries that lacked flow. In contrast, **abstractive summarisation** reformulates the content into smoother, narrative-like summaries that better capture meaning and intent (Murray et al., 2010). However, this method is more demanding, as it requires contextual understanding, synthesis of ideas, and fluent language generation rather than simple sentence extraction.
3. **Heterogeneous Meeting Formats**.
   Meetings differ in **purpose and structure**, ranging from decision-making and brainstorming to information sharing. Because of this diversity, a single summarisation technique cannot effectively serve all scenarios (Nedoluzhko and Bojar, 2019). For example, decision-focused meetings may require detailed accounts of alternatives and outcomes, while status updates may need only brief topic summaries. Thus, effective meeting summarisation systems must be **adaptive** and tailored to each meeting type and objective.
4. **Subjectivity.**
   Subjectivity plays a major role in summarising meeting content. Different annotators may interpret the importance of information differently or express similar ideas in distinct ways. This variability makes evaluation challenging, particularly for abstractive summaries, which allow multiple valid expressions of the same content. Moreover, even the same annotator may produce slightly different summaries at different times (Rath et al., 1961).

III. Meeting Data Sets

Supervised learning is still the most popular method for meeting summaries, although it has limitations. It relies significantly on big, high-quality training datasets and frequently develops models tailored to specific languages or domains. Because meetings have distinct characteristics and summary styles, implementing efficient summarisation algorithms necessitates datasets tailored specifically to meeting circumstances. This section first discusses what constitutes a strong meeting dataset before reviewing the major meeting corpora currently available.

1. Data Set Conception and Creation

Creating a high-quality meeting summarisation dataset is challenging because it must capture **real, spontaneous conversations** including interruptions, informal speech, and natural pauses while remaining structured enough for models to process. Real meetings often involve multiple speakers and shifting topics, making it difficult to maintain balance between realism and organisation.

Researchers must also decide which meeting types to include while **avoiding sensitive or private discussions**. Once recordings are made, **speech recognition errors** and transcription issues can arise, often requiring manual correction. Additional preprocessing such as cleaning disfluencies and adding detailed annotations like dialogue acts demands significant time and expertise.

2. Extant Data Sets

Building large, high-quality meeting summarisation datasets is challenging due to the time, cost, and privacy concerns involved. As a result, only a few major English datasets exist **AMI**, **ICSI**, and

ELITR totalling about 280 hours of meeting recordings. Among these, **ELITR** also provides around 50 hours of Czech data. All three offer carefully corrected transcripts, reducing speech recognition errors and improving reliability.

The **AMI Corpus** features 137 scenario-based meetings where participants role-play in a fictional company designing a remote control. Though structured, these sessions capture natural interactions and include multimodal elements like slides, adding summarisation complexity.

The **ICSI Corpus** contains 75 real academic meetings, showcasing authentic discussions among researchers and students. However, their technical nature makes summarisation difficult without domain-specific understanding.

The **ELITR Corpus** offers realistic, multilingual project meetings but with some redacted segments for privacy. Unlike AMI and ICSI, ELITR summaries are unstructured, making it useful for studying how annotators interpret the same meeting differently. All three corpora are publicly available with detailed annotations, supporting ongoing research in meeting summarisation.

IV. Evaluation Methods

Evaluating meeting summaries is challenging, especially for **abstractive summarisation**, where systems generate new sentences instead of directly extracting them. A good summary should accurately capture the meeting's key points and align semantically with human-written references. However, since abstractive summaries often use different wording, evaluation requires understanding meaning rather than just matching words. **ROUGE** (Lin, 2004) is the most common metric, but it mainly checks word overlap and often fails to recognise correct rephrasing. For instance, a sentence that uses synonyms or paraphrasing might score lower despite being accurate.

To address this, newer metrics like **BERTScore**, **MoverScore**, and **BARTScore** use contextual embeddings to assess meaning similarity. Even more refined models, such as **DATScore** and **FrugalScore**, enhance evaluation by learning from large datasets. Still, no metric is fully reliable studies show that they sometimes disagree on which summary is better and often overlook **factual accuracy**, which is vital for meeting records that influence real decisions.

V. Systems for Meeting Summarization

Ones (1999) describes a three-stage framework for the summarisation process, which is still a valuable model for understanding how meeting summarisation systems operate. Stages are:

Interpretation (I): In this stage, the input whether text or speech is analysed and mapped into a structured representation that captures additional information required to interpret its meaning.
Transformation (T): The system transforms and refines the representation, finding the most relevant information and organising it into a condensed summary form.
Generation (G): Finally, the system translates this structured summary into natural language writing that is easy to read and comprehend.
During our literature research, we discovered that most existing meeting summarisation studies focus on enhancing one of these three steps to solve the previously described issues (such as disfluency, speaker overlap, and subjectivity). As a result, we employ this three-step pipeline: Interpretation, Transformation, and Generation as the foundation for categorising various meeting summarisation methods in the following sections.

1. Interpretation

(1) a. PM: Which colour should the removable cover be?
b. ID: I believe we should give a few options, such as raspberry, lime, and blueberry, as well as black for those who prefer a neutral colour.

c. UI: This seems nice to me.
c. ME: I agree.
e. PM: Okay, we'll go with it.

This exchange clearly indicates that a decision has been made, but how can we know? The final sentence (1-e) confirms the choice but does not clarify what it is. To really grasp it, we must consider the entire discourse, noting that (1-b) answers the query in (1-a) and (1-e) acknowledges agreement from (1-c) and (1-d). If (1-c) and (1-d) had disagreed and offered something different, (1-e) would most likely corroborate their alternate proposal.

This example demonstrates that to properly grasp a meeting, we must look at statements in context and understand how each comment contributes to the flow of the conversation. Some meeting summarisation systems (described in Section 5.1.1) solve this by including annotations that explicitly state these associations. Others, such as HMNet (Zhu et al., 2020) and DDAMS (Feng et al., 2020), concentrate on who said what, which aids in the conversion of first- and second-person phrases (such as "I" or "you") into third-person language appropriate for summaries.
Furthermore, as explained in Section 5.1.2, some advanced systems employ nonverbal data such as participants' eye gaze or gestures to better grasp the context and flow of interaction during meetings.

- **Discursive Information**
  Most abstractive meeting summarisation algorithms view discussions as basic text sequences, ignoring how utterances connect to one another. However, knowing semantic relations such as enquiries, answers, and acknowledgments is critical for correct descriptions.
  Recent models employ discourse structure theories like RST and SDRT, which depict conversations as graphs that connect related statements. Feng et al. (2020) generated summaries by combining such graphs with a Graph Convolutional Network (GCN) and a Pointer-Generator Network (PGN).
  These methods are also useful for decision and action-item summarisation, as demonstrated in research such as Fernández et al. (2008) and Purver et al. (2007), which identify and summarise crucial decision-related talks.
- **Multimodality**

  Meetings often take place in visual settings where gestures, expressions, and eye movements convey key information. Including these nonverbal cues can make summaries more accurate than relying only on text. Research supports this for example, **Li et al. (2019)** found that tracking participants' visual focus improves abstractive summaries, while **Erol et al. (2003)** showed that combining visual and audio cues with text enhances extractive results. However, adding such multimodal data can be time-consuming and error-prone, especially when done manually.

  To overcome this, recent work focuses on **generation-based models** that create summaries directly from data. Still, **interpretive models**, such as those using discourse graphs, remain valuable for capturing structure and meaning in longer, complex meetings.

2. **Transformation**
   Like traditional multi-step text summarisation methods, this approach converts meeting transcripts into *intermediate representations* that facilitate summarisation. First, utterances *are grouped* according to shared subjects, related fields, or common enquiries. This segmentation helps focus on distinct areas of a conference, making summaries more organised and meaningful.
   An *extractive step* is frequently used to remove irrelevant items before or after grouping. Because different meetings require distinct summary formats for example, *doctor-patient consultations* may require SOAP-style notes this segmentation makes it easy to create *domain-specific summaries*.

Furthermore, breaking down long meetings into smaller, *topic-based segments* helps overcome processing constraints and enhances summarisation efficiency.

- **Topic Segments**

  Topic segmentation breaks lengthy discussions into smaller, focused sections, each covering a specific theme. For example, **Banerjee et al. (2015)** used the **LCSeg** tool to divide meeting transcripts into topic-based segments and then selected the most meaningful sentences to form concise summaries. This ensures that important ideas from every part of the discussion are retained. The same approach has also proven effective for summarizing **medical** and **casual conversations**, highlighting its adaptability across different types of dialogues.

- **Abstractive Communities**

  Abstractive community detection, introduced by **Murray et al. (2012)**, is used in meeting summarization to group related sentences that together express one main idea. A logistic regression model identifies links between utterances, forming a graph where connected groups called **abstractive communities** represent shared meanings. Later work, such as **Shang et al. (2020a)**, improved this using advanced clustering on neural embeddings. **Mehdad et al. (2013)** built on this by filtering out less useful sentences and generating concise summaries using a **multi-sentence compression graph**, producing clear and coherent statements. Later studies (**Shang et al., 2018; Park & Lee, 2022**) further enhanced the technique, showing it's especially effective for summarizing **decision-focused discussions** in meetings.

- **Template-related Clusters**

  Sometimes meeting summaries need to follow a specific format or structure. To achieve this, **Oya et al. (2014)** introduced a **template-based abstractive summarization** method that uses clustering and compression techniques to form patterns like *"[speaker] discussed [action] and [content]."* The system breaks the meeting into topics, picks suitable templates, and fills in details to produce clear, structured summaries. Later, **Krishna et al. (2021)** expanded on this by designing summaries based on the **AMI format** Abstract, Decision, Problems, and Action. Their model identifies key sentences and organizes them into the right sections. This structured approach has also proven effective for generating **clinical summaries** of doctor-patient interactions.

- **Query-related Clusters**

  To handle different meeting styles and personal preferences, **Mehdad et al. (2014)** introduced a **query-based meeting summarization** system that lets users generate summaries tailored to specific interests. The system identifies statements most relevant to a user's query and condenses them into clear sentences using compression techniques. Later, the **QMSum dataset** (**Zhong et al., 2021**) expanded this idea by adding multiple query-summary pairs for meetings, along with a model that locates query-relevant content and generates summaries automatically. While flexible, these systems often struggle with the spontaneous and scattered nature of real meetings. Improving topic segmentation and context understanding could lead to more accurate and useful summaries in the future.

3. **Generation**

Large *pretrained transformer-based language models* like as *BART*, *T5*, and others have recently made significant advances in numerous *natural language generation (NLG)* problems. However, when applied directly to *meeting summarisation*, these models frequently *fall short*. This is because *meeting conversations* deviate significantly from the well-structured text that these models were trained on. Furthermore, such models can only handle *limited input lengths*, which are significantly shorter than the *average time and length of meeting transcripts*.

- **Long Input Processing**

A common approach to handle long meeting transcripts is to split them into smaller parts before summarizing. **Koay et al. (2021)** used a sliding window with BART to summarize each chunk and then merge the results, while **Zhang et al. (2022)** proposed **SUMMN**, a multi-stage "split-then-summarize" model that refines summaries step by step. Though effective, these methods often lose context between segments.

To overcome this, researchers explored **Transformer-based models** designed for long sequences. **Longformer** (Beltagy et al., 2020) adapts self-attention to handle lengthy inputs efficiently, and studies show it performs well for meeting data. **Hierarchical Transformers** like **HMNet** (Zhu et al., 2020) and **HAT** (Rohde et al., 2021) further improve performance by mirroring meeting structure processing words into utterances and utterances into full transcripts to produce more coherent and context-aware summaries.

- **Domain Adaptation**

Domain adaptation focuses on enhancing language models by tailoring them to specific contexts instead of relying only on general pretraining. **Gururangan et al. (2020)** showed that additional pretraining on domain-specific data significantly boosts performance compared to using general-purpose models. For instance, **Zhong et al. (2022)** introduced **DialogLM**, trained on dialogue-specific data with simulated interruptions and speaker changes, helping the model better capture real conversational flow. Similarly, **Zou et al. (2021)** proposed a **multi-source pretraining method** for low-resource dialogue summarisation, combining in-domain and external data to improve generalisation.

Overall, **generation-based models** using domain-adapted pretraining perform well in **abstractive summarisation**, even with limited data. However, **interpretation-based systems** still offer stronger contextual understanding. Bridging these two approaches remains a key direction for future research.

VI. Future Directions

- **Multi-task Learning.**
  Future meeting summarisation research can benefit from multi-task learning, which trains models to perform numerous related tasks at once, such as dialogue act classification, discourse parsing, community detection, and multimodal understanding. This method has been demonstrated to improve both efficiency and accuracy, as mastering one task can benefit another by exchanging important information (Caruana, 1997). Studies such as those by Li et al. (2019) and Feng et al. (2021b) demonstrate the promise of this approach, which future systems may expand to make summarisation more context-aware and data-efficient.
- **Factual Consistency.**
  Maintaining factual accuracy is a significant difficulty when summarising meetings. According to Cao et al. (2018), over 30% of neural model descriptions contain false or incorrect facts. These problems are frequently caused by the model's failure to track dialogue flow, handle informal exchanges, or properly resolve coreferences (Tang et al., 2022).
  Future work should focus on decreasing hallucinated content and ensuring summaries accurately reflect the meeting's actual substance. Furthermore, there is a pressing need for new evaluation metrics that can quantify factual consistency specifically for dialogues, as current techniques built

for short texts fail to reflect the complexity of long, multi-speaker transcripts (Zechner and Waibel, 2000; Huang et al., 2021).

- **Weak to No Supervision.**
  Most contemporary meeting summarisation methods rely primarily on supervised learning, which requires big, annotated datasets. However, this dependence limits their versatility by making them language and particular dataset. A possible future path is to investigate weakly supervised, semi-supervised, or even unsupervised methods that can eliminate the need for expensive labelled data. Autoencoders and utterance-level attention are two techniques that could help identify key parts of a meeting without explicit supervision. So yet, these approaches have received little attention, but they hold enormous promise for future research.
- **Spoken Language Models.**
  While huge pretrained models like BERT, T5, and GPT have transformed text-based NLP, they frequently struggle with spoken dialogue data since it is informal and unstructured. Future developments could focus on constructing spoken-language-specific models, such as DialoGPT (Zhang et al., 2020b) and DialogLM (Zhong et al., 2022), to better interpret multi-speaker discussions. Such models have the potential to considerably improve the performance and accuracy of abstractive meeting summarisation systems by capturing the intricacies of real-world speech patterns.
- **Commonsense Incorporation.**
  Humans naturally employ commonsense knowledge to understand and summarise information for example, understanding that a piano is played by pushing keys. However, machines lack this intuitive intuition. Integrating commonsense knowledge graphs into language models helps improve reasoning and understanding, closing the gap between interpretation and generation-based systems (Ilievski et al., 2021).

  Incorporating such prior knowledge is especially useful in professional meetings, where participants frequently rely on shared understanding and domain-specific terminology. By incorporating commonsense and contextual information into summarisation models, systems can capture implicit meanings and provide more coherent, human-like summaries (Zhou et al., 2018; Xiachong et al., 2021).

- **Prompting Paradigm.**

The emergence of large language models like **GPT-3, Gopher, and PaLM** has introduced a new paradigm called **prompt-based learning**. Instead of fine-tuning models for every task, users can now provide simple natural-language instructions such as *“Summarise the above article in three sentences”* and obtain high-quality results with little or no training data.

This approach, enhanced through **instruction tuning** (e.g., FLAN, T0, BLOOMZ) and **human feedback training** (e.g., ChatGPT), has achieved performance comparable to traditional fine-tuning methods (Stiennon et al., 2020; Goyal et al., 2022).

Although still rarely applied to **meeting summarisation**, prompt-based methods show strong potential for producing **customised summaries** that highlight specific aspects like decisions, actions, or discussions. However, evaluating such summaries remains a challenge, as traditional metrics like **ROUGE** and **BERTScore** struggle to capture the quality of more **creative and abstract outputs**. Developing new evaluation strategies and more effective prompts could significantly advance this field.

VII. Conclusion

The paper provides a detailed overview of recent advancements in abstractive meeting summarization. It effectively highlights key research challenges and introduces a taxonomy based on the Interpretation Transformation Generation framework proposed by Jones (1999). The authors also compare existing approaches and outline potential directions for future research.

VIII. REFERENCES

Rennard, V., Shang, G., Hunter, J., & Vazirgiannis, M. (2023). *Abstractive meeting summarization: A survey*. *Transactions of the Association for Computational Linguistics, 11*, 861–884. https://doi.org/10.1162/tacl_a_00578

James Allen and Mark Core. 1997. *Draft of DAMSL: Dialog act markup in several layers.* https://www.cs.rochester.edu/research/cisd/resources/damsl

Nicholas Asher. 1993. *Reference to Abstract Objects in Discourse.* Dordrecht, Boston, and London: Kluwer. https://doi.org/10.1007/978-94-011-1715-9

Nicholas Asher, Julie Hunter, Mathieu Morey, Farah Benamara, and Stergos Afantenos. 2016. *Discourse structure and dialogue acts in multiparty dialogue: The STAC corpus.* In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2721–2727, Portorož, Slovenia. European Language Resources Association (ELRA).

Siddhartha Banerjee, Prasenjit Mitra, and Kazunari Sugiyama. 2015. *Generating abstractive summaries from meeting transcripts.* In *Proceedings of the 2015 ACM Symposium on Document Engineering*, pages 51–60. https://doi.org/10.1145/2682571.2797061

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. *Longformer: The long-document transformer.* arXiv preprint arXiv:2004.05150.

James C. Bezdek, Robert Ehrlich, and William Full. 1984. *FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences,* 10(2–3):191–203. https://doi.org/10.1016/0098-3004(84)90020-7

Manik Bhandari, Pranav Narayan Gour, Atabak Ashfaq, and Pengfei Liu. 2020a. *Metrics also disagree in the low scoring range: Revisiting summarization evaluation metrics.* In *Proceedings of the 28th International Conference on Computational Linguistics (COLING)*, pages 5702–5711, Barcelona, Spain (Online). https://doi.org/10.18653/v1/2020.coling-main.501

Manik Bhandari, Pranav Narayan Gour, Atabak Ashfaq, Pengfei Liu, and Graham Neubig. 2020b. *Re-evaluating evaluation in text summarization.* In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9347–9359, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.751

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei.

2020. *Language models are few-shot learners.*
In *Advances in Neural Information Processing Systems (NeurIPS 33)*, pages 1877–1901. Curran Associates, Inc.

Trung Bui, Matthew Frampton, John Dowding, and Stanley Peters. 2009. *Extracting decisions from multi-party dialogue using directed graphical models and semantic similarity.*
In *Proceedings of the SIGDIAL 2009 Conference*, pages 235–243, London, UK. Association for Computational Linguistics.
https://doi.org/10.3115/1708376.1708410

Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. 2010. *Towards an ISO standard for dialogue act annotation.*
In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David Traum. 2012. *ISO 24617-2: A semantically-based standard for dialogue annotation.*
In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 430–437, Istanbul, Turkey. European Language Resources Association (ELRA).

Ziqiang Cao, Furu Wei, Wenjie Li, and Sujian Li. 2018. *Faithful to the original: Fact aware neural abstractive summarization.*
In *Thirty-Second AAAI Conference on Artificial Intelligence.*

Rich Caruana. 1997. *Multitask learning. Machine Learning,* 28(1):41–75.
https://doi.org/10.1023/A:1007379606734

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.
https://doi.org/10.18653/v1/2020.emnlp-main.336

Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. *Learning a similarity metric discriminatively, with application to face verification.*
In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 539–546. IEEE.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam M. Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Benton C. Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier García, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Díaz, Orhan Firat, Michele

Catasta, Jason Wei, Kathleen S. Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. 2022. *PaLM: Scaling language modeling with pathways. arXiv preprint* arXiv:2204.02311.

Mark G. Core and James Allen. 1997. *Coding dialogs with the DAMSL annotation scheme.* In AAAI Fall Symposium on Communicative Action in Humans and Machines, volume 56, pages 28–35. Boston, MA.

Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc Le, and Ruslan Salakhutdinov. 2019. *Transformer-XL: Attentive language models beyond a fixed-length context.*
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2978–2988, Florence, Italy. Association for Computational Linguistics.

Shibhansh Dohare, Harish Karnick, and Vivek Gupta. 2017. *Text summarization using abstract meaning representation.*
arXiv preprint arXiv:1706.01678.

Moussa Kamal Eddine, Guokan Shang, and Michalis Vazirgiannis. 2022. *Datscore: Evaluating translation with data augmented translations.*
Findings of EACL 2023.

Angela Edmunds and Anne Morris. 2000. *The problem of information overload in business organisations: A review of the literature.*
International Journal of Information Management, 20(1):17–28.
https://doi.org/10.1016/S0268-4012(99)00051-1

Kemal Elciyar. 2021. *Overloading in lockdown: Effects of social, information and communication overloads in COVID-19 days.*
İnönü Üniversitesi İletişim Fakültesi Elektronik Dergisi (İNİF E-Dergi), 6(1):329–342.
https://doi.org/10.47107/inifedergi.872896

Berna Erol, Dar-Shyang Lee, and Jonathan Hull. 2003. *Multimodal summarization of meeting recordings.*
In *2003 International Conference on Multimedia and Expo (ICME'03)*, volume 3, pages III–25. IEEE.
https://doi.org/10.1109/ICME.2003.1221239

Alexander Fabbri, Faiaz Rahman, Imad Rizvi, Borui Wang, Haoran Li, Yashar Mehdad, and Dragomir Radev. 2021. *ConvoSumm: Conversation summarization benchmark and improved abstractive summarization with argument mining.*
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 6866–6880, Online. Association for Computational Linguistics.
https://doi.org/10.18653/v1/2021.acl-long.535

Geraldine Fauville, Mufan Luo, Anna C. M. Queiroz, Jeremy N. Bailenson, and Jeff Hancock. 2021. *Zoom exhaustion & fatigue scale.*
Computers in Human Behavior Reports, 4:100119.
https://doi.org/10.1016/j.chbr.2021.100119

Xiachong Feng, Xiaocheng Feng, and Bing Qin. 2021a. *A survey on dialogue summarization: Recent advances and new frontiers.*
In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI 2022).*
https://doi.org/10.24963/ijcai.2022/764

Xiachong Feng, Xiaocheng Feng, Bing Qin, and Xinwei Geng. 2020. *Dialogue discourse-aware graph model and data augmentation for meeting summarization.*
In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI 2021).*
https://doi.org/10.24963/ijcai.2021/524

Xiachong Feng, Xiaocheng Feng, Libo Qin, Bing Qin, and Ting Liu. 2021b. *Language model as an annotator: Exploring DialoGPT for dialogue summarization.*
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 1479–1491, Online. Association for Computational Linguistics.
https://doi.org/10.18653/v1/2021.acl-long.117

Raquel Fernández, Matthew Frampton, John Dowding, Anish Adukuzhiyil, Patrick Ehlen, and Stanley Peters. 2008. *Identifying relevant phrases to summarize decisions in spoken meetings.*
In *Ninth Annual Conference of the International Speech Communication Association (Interspeech 2008).*
https://doi.org/10.21437/Interspeech.2008-17

Katja Filippova. 2010. *Multi-sentence compression: Finding shortest paths in word graphs.*
In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 322–330, Beijing, China. Coling 2010 Organizing Committee.

Michel Galley, Kathleen R. McKeown, Eric Fosler-Lussier, and Hongyan Jing. 2003. *Discourse segmentation of multi-party conversation.*
In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 562–569, Sapporo, Japan. Association for Computational Linguistics.
https://doi.org/10.3115/1075096.1075167

Mahak Gambhir and Vishal Gupta. 2017. *Recent automatic text summarization techniques: A survey.*
Artificial Intelligence Review, 47(1):1–66.
https://doi.org/10.1007/s10462-016-9475-9

Prakhar Ganesh and Saket Dingliwal. 2019. *Restructuring conversations using discourse relations for zero-shot abstractive dialogue summarization.*
arXiv preprint arXiv:1902.01615.

Nikhil Garg, Benoit Favre, Korbinian Reidhammer, and Dilek Hakkani-Tür. 2009. *ClusterRank: A graph-based method for meeting summarization.*
Technical Report, Idiap.
https://doi.org/10.21437/Interspeech.2009-456

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. IEEE.

Tanya Goyal, Junyi Jessy Li, and Greg Durrett. 2022. *News summarization and evaluation in the era of GPT-3.*
arXiv preprint arXiv:2209.12356.

Jonathan L. Gross and Jay Yellen. 2003. *Handbook of Graph Theory.* CRC Press.
https://doi.org/10.1201/9780203490204

Yanzhu Guo, Chloé Clavel, Moussa Kamal Eddine, and Michalis Vazirgiannis. 2022. *Questioning the validity of summarization datasets and improving their factual consistency.*
In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5716–5727, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. *Don't stop pretraining: Adapt language models to domains and tasks.*
In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics.
https://doi.org/10.18653/v1/2020.acl-main.740

Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. *Teaching machines to read and comprehend.*
Advances in Neural Information Processing Systems (NeurIPS 28).

Elad Hoffer and Nir Ailon. 2015. *Deep metric learning using triplet network.*
In *International Workshop on Similarity-Based Pattern Recognition*, pages 84–92. Springer.
https://doi.org/10.1007/978-3-319-24261-3_7

Yichong Huang, Xiachong Feng, Xiaocheng Feng, and Bing Qin. 2021. *The factual inconsistency problem in abstractive text summarization: A survey.*
arXiv preprint arXiv:2104.14839.

Filip Ilievski, Pedro Szekely, and Bin Zhang. 2021. *CSKG: The commonsense knowledge graph.*
In *European Semantic Web Conference*, pages 680–696. Springer.
https://doi.org/10.1007/978-3-030-77385-4_41

Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, et al. 2003. *The ICSI meeting corpus.*
In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, volume 1, pages I–I. IEEE.

Karen Sparck Jones. 1999. *Automatic summarizing: Factors and directions.*
In *Advances in Automatic Text Summarization*, page 1.

Dan Jurafsky, Liz Shriberg, and Debra Biasca. 1997. *Switchboard SWBD-DAMSL shallow discourse-function annotation coders manual.*
Institute of Cognitive Science Technical Report.

Moussa Kamal Eddine, Guokan Shang, Antoine Tixier, and Michalis Vazirgiannis. 2022. *FrugalScore: Learning cheaper, lighter and faster evaluation metrics for automatic text generation.*
In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1305–1318, Dublin, Ireland. Association for Computational Linguistics.
https://doi.org/10.18653/v1/2022.acl-long.93

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q., & Salakhutdinov, R. (2019). **Transformer-XL: Attentive language models beyond a fixed-length context.** *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2978–2988. Florence, Italy. https://doi.org/10.18653/v1/P19-1285

Dohare, S., Karnick, H., & Gupta, V. (2017). **Text summarization using abstract meaning representation.** *arXiv preprint arXiv:1706.01678*.

Edmunds, A., & Morris, A. (2000). **The problem of information overload in business organisations: A review of the literature.** *International Journal of Information Management*, 20(1), 17–28. https://doi.org/10.1016/S0268-4012(99)00051-1

Eddine, M. K., Shang, G., & Vazirgiannis, M. (2022). **Datscore: Evaluating translation with data augmented translations.** *Findings of EACL 2023*.

Eddine, M. K., Tixier, A., & Vazirgiannis, M. (2021). **BARThez: A skilled pretrained French sequence-to-sequence model.** *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 9369–9390. https://doi.org/10.18653/v1/2021.emnlp-main.740

Eddine, M. K., Shang, G., Tixier, A., & Vazirgiannis, M. (2022). **FrugalScore: Learning cheaper, lighter and faster evaluation metrics for automatic text generation.** *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1305–1318. https://doi.org/10.18653/v1/2022.acl-long.93

Elciyar, K. (2021). **Overloading in lockdown: Effects of social, information and communication overloads in COVID-19 days.** *İnönü Üniversitesi İletişim Fakültesi Elektronik Dergisi*, 6(1), 329–342. https://doi.org/10.47107/inifedergi.872896

Erol, B., Lee, D.-S., & Hull, J. (2003). **Multimodal summarization of meeting recordings.** *Proceedings of the International Conference on Multimedia and Expo (ICME)*, Vol. 3, III–25. IEEE. https://doi.org/10.1109/ICME.2003.1221239

Fabbri, A., Rahman, F., Rizvi, I., Wang, B., Li, H., Mehdad, Y., & Radev, D. (2021). **ConvoSumm: Conversation summarization benchmark and improved abstractive summarization with argument mining.** *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, 6866–6880. https://doi.org/10.18653/v1/2021.acl-long.535

Fauville, G., Luo, M., Queiroz, A. C. M., Bailenson, J. N., & Hancock, J. (2021). **Zoom exhaustion & fatigue scale.** *Computers in Human Behavior Reports*, 4, 100119. https://doi.org/10.1016/j.chbr.2021.100119

Feng, X., Feng, X., & Qin, B. (2021a). **A survey on dialogue summarization: Recent advances and new frontiers.** *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI 2022)*. https://doi.org/10.24963/ijcai.2022/764

Feng, X., Feng, X., Qin, B., & Geng, X. (2020). **Dialogue discourse-aware graph model and data augmentation for meeting summarization.** *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI 2021)*. https://doi.org/10.24963/ijcai.2021/524

Feng, X., Feng, X., Qin, L., Qin, B., & Liu, T. (2021b). **Language model as an annotator: Exploring DialoGPT for dialogue summarization.** *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, 1479–1491. https://doi.org/10.18653/v1/2021.acl-long.117

Fernández, R., Frampton, M., Dowding, J., Adukuzhiyil, A., Ehlen, P., & Peters, S. (2008). **Identifying relevant phrases to summarize decisions in spoken meetings.** *Interspeech 2008*. https://doi.org/10.21437/Interspeech.2008-17

Filippova, K. (2010). **Multi-sentence compression: Finding shortest paths in word graphs.** *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, 322–330.

Galley, M., McKeown, K. R., Fosler-Lussier, E., & Jing, H. (2003). **Discourse segmentation of multi-party conversation.** *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 562–569. https://doi.org/10.3115/1075096.1075167

Gambhir, M., & Gupta, V. (2017). **Recent automatic text summarization techniques: A survey.** *Artificial Intelligence Review*, 47(1), 1–66. https://doi.org/10.1007/s10462-016-9475-9

Ganesh, P., & Dingliwal, S. (2019). **Restructuring conversations using discourse relations for zero-shot abstractive dialogue summarization.** *arXiv preprint arXiv:1902.01615*.

Garg, N., Favre, B., Reidhammer, K., & Hakkani-Tür, D. (2009). **ClusterRank: A graph-based method for meeting summarization.** *Interspeech 2009*. https://doi.org/10.21437/Interspeech.2009-456

Goo, C.-W., & Chen, Y.-N. (2018). **Abstractive dialogue summarization with sentence-gated modeling optimized by dialogue acts.** *2018 IEEE Spoken Language Technology Workshop (SLT)*, 735–742. IEEE.

Goyal, T., Li, J. J., & Durrett, G. (2022). **News summarization and evaluation in the era of GPT-3.** *arXiv preprint arXiv:2209.12356*.

Gross, J. L., & Yellen, J. (2003). **Handbook of Graph Theory.** CRC Press. https://doi.org/10.1201/9780203490204

Guo, Y., Clavel, C., Eddine, M. K., & Vazirgiannis, M. (2022). **Questioning the validity of summarization datasets and improving their factual consistency.** *Proceedings of EMNLP 2022*, 5716–5727.

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). **Don't stop pretraining: Adapt language models to domains and tasks.** *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8342–8360. https://doi.org/10.18653/v1/2020.acl-main.740

Hermann, K. M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., & Blunsom, P. (2015). **Teaching machines to read and comprehend.** *Advances in Neural Information Processing Systems (NeurIPS 28)*.

Hoffer, E., & Ailon, N. (2015). **Deep metric learning using triplet network.** *Similarity-Based Pattern Recognition Workshop*, 84–92. Springer. https://doi.org/10.1007/978-3-319-24261-3_7

Huang, Y., Feng, X., Feng, X., & Qin, B. (2021). **The factual inconsistency problem in abstractive text summarization: A survey.** *arXiv preprint arXiv:2104.14839*.

Ilievski, F., Szekely, P., & Zhang, B. (2021). **CSKG: The commonsense knowledge graph.** *European Semantic Web Conference*, 680–696. Springer. https://doi.org/10.1007/978-3-030-77385-4_41

Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., et al. (2003). **The ICSI meeting corpus.** *ICASSP 2003*, Vol. 1, I–I. IEEE.

Jurafsky, D., Shriberg, L., & Biasca, D. (1997). **Switchboard SWBD-DAMSL shallow discourse-function annotation coders manual.** *Institute of Cognitive Science Technical Report.*

Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C., & Socher, R. (2019). **CTRL: A conditional transformer language model for controllable generation.** *arXiv preprint arXiv:1909.05858*.

Koay, J. J., Roustai, A., Dai, X., Burns, D., Kerrigan, A., & Liu, F. (2020). **How domain terminology affects meeting summarization performance.** *Proceedings of COLING 2020*, 5689–5695. https://doi.org/10.18653/v1/2020.coling-main.499

Koay, J. J., Roustai, A., Dai, X., & Liu, F. (2021). **A sliding-window approach to automatic creation of meeting minutes.** *NAACL Student Research Workshop 2021*, 68–75. https://doi.org/10.18653/v1/2021.naacl-srw.10

Kost, D. (2020). **You're right! You are working longer and attending more meetings.** *Harvard Business School Working Knowledge.*

Krishna, K., Khosla, S., Bigham, J., & Lipton, Z. C. (2021). **Generating SOAP notes from doctor-patient conversations using modular summarization techniques.** *Proceedings of ACL-IJCNLP 2021*, 4958–4972. https://doi.org/10.18653/v1/2021.acl-long.384

Kryscinski, W., Keskar, N. S., McCann, B., Xiong, C., & Socher, R. (2019). **Neural text summarization: A critical evaluation.** *EMNLP-IJCNLP 2019*, 540–551. https://doi.org/10.18653/v1/D19-1051

Kryscinski, W., McCann, B., Xiong, C., & Socher, R. (2020). **Evaluating the factual consistency of abstractive text summarization.** *EMNLP 2020*, 9332–9346. https://doi.org/10.18653/v1/2020.emnlp-main.750

Lascarides, A., & Asher, N. (2008). **Segmented discourse representation theory: Dynamic semantics with discourse structure.** In *Computing Meaning* (pp. 87–124). Springer. https://doi.org/10.1007/978-1-4020-5958-2_5

Abigail See, Peter J. Liu, and Christopher D. Manning. (2017). *Get to the point: Summarization with pointer-generator networks.* In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1073–1083. Vancouver, Canada: Association for Computational Linguistics.

Ananya B. Sai, Akash Kumar Mohankumar, and Mitesh M. Khapra. (2022). *A survey of evaluation metrics used for NLG systems. ACM Computing Surveys (CSUR)*, 55(2), 1–39. https://doi.org/10.1145/3485766

Antoine Tixier, Polykarpos Meladianos, and Michalis Vazirgiannis. (2017). *Combining graph degeneracy and submodularity for unsupervised extractive summarization.* In *Proceedings of the Workshop on New Frontiers in Summarization*, pp. 48–58. Copenhagen, Denmark: Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-4507

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. (2017). *Attention is all you need. Advances in Neural Information Processing Systems*, 30.

Chenguang Zhu, Ruochen Xu, Michael Zeng, and Xuedong Huang. (2020). *A hierarchical network for abstractive meeting summarization with cross-domain pretraining.* In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 194–203. https://doi.org/10.18653/v1/2020.findings-emnlp.19

Feng Xiachong, Feng Xiaocheng, and Qin Bing. (2021). *Incorporating commonsense knowledge into abstractive dialogue summarization via heterogeneous graph networks.* In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pp. 964–975. https://doi.org/10.1007/978-3-030-84186-7_9

Gokhan Tur, Andreas Stolcke, Lynn Voss, John Dowding, Benoit Favre, Raquel Fernández, Matthew Frampton, Michael Frandsen, Clint Frederickson, Martin Graciarena, Dilek Hakkani-Tur, Donald Kintzing, Kyle Leveque, Shane Mason, John Niekrasz, Stanley Peters, Matthew Purver, Korbinian Riedhammer, Elizabeth Shriberg, and Fan Yang. (2009). *The CALO meeting speech recognition and understanding system*, pp. 69–72. https://doi.org/10.1109/SLT.2008.4777842

Gokhan Tur, Andreas Stolcke, Lynn Voss, Stanley Peters, Dilek Hakkani-Tur, John Dowding, Benoit Favre, Raquel Fernández, Matthew Frampton, Mike Frandsen, Clint Frederickson, Martin Graciarena, Donald Kintzing, Kyle Leveque, Shane Mason, John Niekrasz, Matthew Purver, Korbinian Riedhammer, Elizabeth Shriberg, Jing Tien, Dimitra Vergyri, and Fan Yang. (2010). *The CALO meeting assistant system. IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1601–1611. https://doi.org/10.1109/TASL.2009.2038810

Gerard Salton, Amit Singhal, Mandar Mitra, and Chris Buckley. (1997). *Automatic text structuring and summarization. Information Processing & Management*, 33(2):193–207. https://doi.org/10.1016/S0306-4573(96)00062-3

Guokan Shang. (2021). *Spoken Language Understanding for Abstractive Meeting Summarization.* Ph.D. thesis, Institut Polytechnique de Paris.

Guokan Shang, Wensi Ding, Zekun Zhang, Antoine Tixier, Polykarpos Meladianos, Michalis Vazirgiannis, and Jean-Pierre Lorré. (2018). *Unsupervised abstractive meeting summarization with multi-sentence compression and budgeted submodular maximization.* In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 664–674. Melbourne, Australia. https://doi.org/10.18653/v1/P18-1062

Guokan Shang, Antoine Tixier, Michalis Vazirgiannis, and Jean-Pierre Lorré. (2020a). *Energy-based self-attentive learning of abstractive communities for spoken language understanding.* In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pp. 313–327.

Guokan Shang, Antoine Tixier, Michalis Vazirgiannis, and Jean-Pierre Lorré. (2020b). *Speaker-change aware CRF for dialogue act classification.* In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 450–464. https://doi.org/10.18653/v1/2020.coling-main.40

Klaus Zechner. (2002). *Automatic summarization of open-domain multiparty dialogues in diverse genres. Computational Linguistics*, 28(4):447–485. https://doi.org/10.1162/089120102762671945

Klaus Zechner and Alex Waibel. (2000). *Minimizing word error rate in textual summaries of spoken language.* In *1st Meeting of the North American Chapter of the Association for Computational Linguistics.*

Lu Wang and Claire Cardie. (2011). *Summarizing decisions in spoken meetings.* In *Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages*, pp. 16–24.

Lu Wang and Claire Cardie. (2012). *Focused meeting summarization via unsupervised relation extraction.* In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 304–313.

Lu Wang and Claire Cardie. (2013). *Domain-independent abstract generation for focused meeting summarization.* In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1395–1405.

Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. (2018). *Modeling relational data with graph convolutional networks.* In *European Semantic Web Conference*, pp. 593–607. Springer. https://doi.org/10.1007/978-3-319-93417-4_38

Ming Zhong, Da Yin, Tao Yu, Ahmad Zaidi, Mutethia Mutuma, Rahul Jha, Ahmed Hassan Awadallah, Asli Celikyilmaz, Yang Liu, Xipeng Qiu, and Dragomir Radev. (2021). *QMSum: A new benchmark for query-based multi-domain meeting summarization.* In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5905–5921. https://doi.org/10.18653/v1/2021.naacl-main.472

Ming Zhong, Yang Liu, Yichong Xu, Chenguang Zhu, and Michael Zeng. (2022). *DialogLM: Pre-trained model for long dialogue understanding and summarization.* In *Proceedings of the AAAI Conference on Artificial Intelligence*, 36, 11765–11773. https://doi.org/10.1609/aaai.v36i10.21432

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F. Christiano. (2020). *Learning to summarize with human feedback. Advances in Neural Information Processing Systems*, 33:3008–3021.

Oleg Vasilyev, Vedant Dharnidharka, and John Bohannon. (2020). *Fill in the BLANC: Human-free quality estimation of document summaries.* In *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, pp. 11–20. https://doi.org/10.18653/v1/2020.eval4nlp-1.2

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. (2015). *Pointer networks. Advances in Neural Information Processing Systems*, 28.

Steven G. Rogelberg, Cliff Scott, and John Kello. (2007). *The science and fiction of meetings. MIT Sloan Management Review*, 48(2):18–21.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. (2020a). *BERTScore: Evaluating text generation with BERT.* In *8th International Conference on Learning Representations (ICLR 2020)*.

Victor Sanh et al. (2021). *Multitask prompted training enables zero-shot task generalization. arXiv preprint* arXiv:2110.08207.

Weizhe Yuan, Graham Neubig, and Pengfei Liu. (2021). *BartScore: Evaluating generated text as text generation. Advances in Neural Information Processing Systems*, 34:27263–27277.

Xiàngrù Tang et al. (2022). *CONFIT: Toward faithful dialogue summarization with linguistically-informed contrastive fine-tuning.* In *Proceedings of NAACL 2022*, pp. 5657–5668. https://doi.org/10.18653/v1/2022.naacl-main.415

Yicheng Zou, Bolin Zhu, Xingwu Hu, Tao Gui, and Qi Zhang. (2021). *Low-resource dialogue summarization with domain-agnostic multi-source pretraining.* In *Proceedings of EMNLP 2021*, pp. 80–91. https://doi.org/10.18653/v1/2021.emnlp-main.7

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. (2020b). *DIALOGPT: Large-scale generative pre-training for conversational response generation.* In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 270–278. https://doi.org/10.18653/v1/2020.acl-demos.30

Zhou Zhao et al. (2019b). *Abstractive meeting summarization via hierarchical adaptive segmental network learning.* In *The World Wide Web Conference*, pp. 3455–3461. https://doi.org/10.1145/3308558.3313619