Challenges and Contributions in Table-to-text Generation: A Survey

Tathagata Dey and **Pushpak Bhattacharyya**
Department of Computer Science and Engineering,
Indian Institute of Technology Bombay
{tathagata,pb}@cse.iitb.ac.in

Abstract

In the contemporary era of Big Data, an immense volume of data is generated daily, with tables being one of the most prominent and significant forms of structured data. With the advent of generative AI-based systems, text generation has seen considerable advancements. However, the problem of generating coherent and contextually accurate text from tables remains an intriguing and challenging task. Different systems have been developed depending on the type and complexity of the tables involved. Some datasets feature relatively simpler tables to serve as stepping stones, while others present more realistic and intricate tables. Capturing the correct context from these tables, ensuring factual consistency, maintaining coherence, and adhering to fundamental text generation principles are persistent challenges in this field. In this paper, we present a comprehensive survey of table-to-text generation. We outline the importance of this problem and review the significant contributions made thus far. By examining various challenges and datasets, we summarize the approaches taken to address these challenges and the results achieved, providing a thorough overview of the current state and future directions of this research domain.

1 Introduction

Table-to-text conversion has been a longstanding challenge in the field of data-to-text generation. In the era of big data, vast amounts of structured data are generated daily, necessitating efficient and accurate methods to transform this data into comprehensible text. Structured data formats, such as triples and RDFs, are prevalent, but tables remain the most significant and widely used format due to their simplicity and versatility. Tables can encapsulate complex datasets and reveal hidden insights when properly interpreted. Consequently, generating natural language text from tables is crucial for developing modern systems that can automatically interpret and communicate data insights. This capability is essential not only for enhancing data accessibility and usability but also for enabling automated reporting, content creation, and decision support across various domains. The ability to convert structured data into coherent narratives empowers users to derive meaningful conclusions from raw data, thereby bridging the gap between data generation and its practical application in real-world scenarios.

Over the past decade, numerous challenges have emerged in the field of table-to-text generation, each addressing distinct aspects of the problem and presenting unique hurdles to overcome. These challenges have spurred a variety of innovative approaches aimed at improving the accuracy, fluency, and contextual relevance of generated text. This survey comprehensively reviews these challenges and the corresponding methodologies devised to address them. Section 2 delves into the background and motivation of the work, providing a detailed context for the evolution and significance of table-to-text generation. Section 3 presents an extensive overview of the available datasets, highlighting their characteristics, scope, and the specific problems they aim to solve. Section 4 discusses the various approaches adopted to tackle these challenges, including traditional techniques, state-of-the-art neural models, and hybrid methods. Finally, Section 5 showcases the results of these contributions, offering a comparative analysis of different approaches and their effectiveness in addressing the multifaceted challenges of table-to-text generation. Through this structured examination, the survey aims to provide a thorough understanding of the progress made in this field and to identify potential avenues for future research.

2 Motivation and Background of Table-to-text Generation

Text generation from tables holds immense potential in a variety of social applications where merely

describing numerical values falls short of conveying meaningful insights. By transforming structured data into coherent and contextually relevant narratives, these applications can enhance the accessibility and comprehensibility of data for diverse audiences.

For instance, in the realm of news or blog writing, automated text generation can be employed to create detailed articles based on tournament points tables, pricing tables, or voting results tables. Such generated content can provide readers with a clear and concise interpretation of the data, highlighting trends, anomalies, and key takeaways that might not be immediately apparent from the raw numbers alone.

In the context of business and sports reporting, generating summaries from match results, detailed business reports, or sales data can streamline the creation of insightful and engaging narratives. These summaries can assist stakeholders in making informed decisions by presenting data in an easily digestible format, complete with context and analysis that underscore the significance of the figures presented.

Moreover, in fields such as healthcare, meteorology, and law, the ability to generate explanatory text from structured reports can be invaluable. Automated descriptions of healthcare reports can aid in patient understanding and communication, while weather reports can be translated into user-friendly forecasts that emphasize critical information. Similarly, legal documents, which often contain complex data, can be distilled into clear summaries that facilitate comprehension and decision-making.

Overall, the application of table-to-text generation in these areas underscores the technology's versatility and its potential to transform data presentation across multiple domains, thereby enhancing the way information is communicated and understood.

In recent years, table-to-text generation has witnessed significant advancements, driven by the development of more sophisticated models and the availability of extensive datasets. However, the scope of exploration within this domain has largely remained confined to specific areas, predominantly focusing on sports and Wikipedia tables. This narrow focus has led to substantial improvements in generating descriptive and contextually accurate text for these types of data, but it has also highlighted the need for broader applicability. The reliance on sports data and Wikipedia infoboxes has provided a controlled environment to test and refine models, resulting in notable progress in handling structured data and producing coherent narratives. Yet, the limited domain of exploration poses challenges for generalizability and robustness across diverse datasets. Expanding research efforts to encompass a wider range of domains, such as medical records, financial data, and scientific datasets, is essential to further advance the capabilities of table-to-text generation models. By addressing this gap, future research can ensure that the techniques and models developed are versatile and applicable to a broader spectrum of real-world data scenarios, thereby enhancing the overall impact and utility of table-to-text generation technologies.

3 Dataset

Significant contributions have been seen in different datasets of this domain. Table-to-text generation being an integral part of data-to-text generation has been highly correlated to other forms of tasks also. For example, rather simple tabular texts can be perceived as knowledge graphs, which in turn can be interpreted as RDFs. So, in this section we discuss all the possible datasets for this task and also introduce our own contribution regarding data.

3.1 WebNLG

Gardent et al. (2017); Castro Ferreira et al. (2020) introduced the WebNLG benchmark dataset and shared task, aimed at generating natural language descriptions from structured data sources, particularly knowledge graphs. This challenge has been conducted in three editions: 2017, 2020, and 2023. The WebNLG 2017 edition comprises 21,855 data/text pairs and includes 8,372 distinct data inputs across 15 DBPedia categories, providing a comprehensive resource for training and evaluating natural language generation models. The WebNLG Challenge 2020, a follow-up to the 2017 edition, introduced one additional DBPedia category and a new dataset for Russian. This edition includes approximately 8,000 data inputs and 20,800 data-text pairs spanning nine distinct categories, expanding the linguistic and categorical diversity of the dataset. The WebNLG 2023 edition further diversifies the challenge by focusing on four under-resourced languages: Maltese, Irish, Breton, and Welsh. This progression highlights the dataset's evolution in addressing linguistic diversity

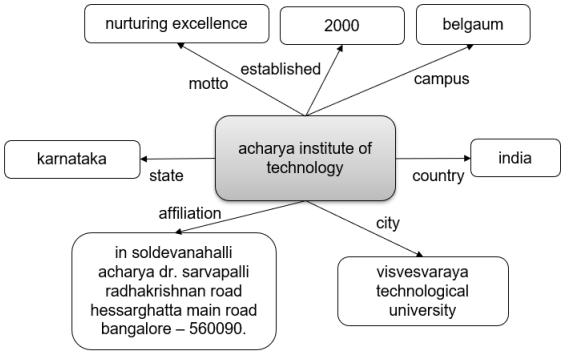


Figure 1: Illustration of triples: Example taken from WebNLG corpus.

and complexity, making it a valuable benchmark for advancing natural language generation from structured data.

Each instance in the WebNLG dataset comprises a set of RDF (Resource Description Framework) triples that can form a star-like knowledge graph, where a central entity is connected to various attributes or related entities. These triples serve as the structured data input for the natural language generation task. For each instance, multiple sentences are provided, effectively describing the given set of triples. This multiplicity of sentences per instance allows for a richer training dataset, capturing various ways to express the same set of facts in natural language. By encompassing diverse linguistic expressions, the dataset facilitates the development of models that can generate fluent, contextually accurate, and varied descriptions from structured data inputs. This structure and detail make the WebNLG dataset an invaluable resource for training and evaluating data-to-text generation models, pushing the boundaries of how effectively machines can translate structured data into human-like descriptions.

3.2 DART

Unlike WebNLG, DART (Nan et al., 2021) is an open-domain Data Record to Text generation dataset that encompasses over 82,000 instances of triples/text pairs, with each triple following the ENTITY-RELATION-ENTITY format. DART is constructed from three diverse sources: (1) Human annotations on Wikipedia tables, which provide rich and varied descriptions; (2) Automatic conversion of questions in WikiSQL (Zhong et al., 2017) into declarative sentences, thereby expanding the dataset's scope and complexity; and (3) Integration

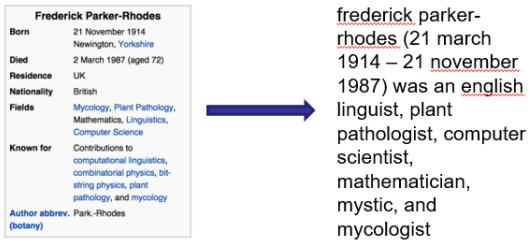


Figure 2: WikiBio Dataset Problem Statement

of existing datasets, including WebNLG 2017 and Cleaned E2E (Dušek et al., 2019), which contribute additional data variety and density. This amalgamation creates a robust and comprehensive dataset suitable for training and evaluating text generation models. Notably, employing DART for data augmentation has led to performance improvements on the WebNLG 2017 dataset across various models. These gains are primarily attributed to the high-quality, human-written sentences included in DART, which enhance the training process by providing more natural and contextually accurate text samples. Consequently, DART stands out as a significant resource for advancing the capabilities of data-to-text generation models in an open-domain context.

3.3 WikiBio

The WikiBio dataset serves as a prominent resource for the table-to-text conversion task (Lebret et al., 2016a). Characterized by tables containing single-column *(Key, Value)* pairs, the dataset effectively frames the task as a triple-to-text conversion challenge. These tables are sourced from Wikipedia Infoboxes, where each instance consists of structured data points paired with their corresponding target text, typically extracted from the first sentence of the respective Wikipedia page. The dataset is partitioned into three distinct sets: the training set comprises $582,660$ instances, while the validation and test sets each contain $72,831$ instances. This segmentation ensures robust evaluation and benchmarking of models across varied datasets, reflecting real-world scenarios where structured data must be transformed into coherent and informative natural language descriptions. Thus, the WikiBio dataset stands as a pivotal resource in advancing research and development in the field of data-to-text generation, facilitating the exploration of effective approaches to handling single-column table data for natural language generation tasks.

The objective of the WikiBio dataset is illus-

trated in Figure 2. Each instance in the dataset features an infobox containing structured information extracted from a biography page. This infobox typically consists of key-value pairs summarizing various aspects of the subject's life or achievements. The corresponding target text aims to provide a natural language description that encapsulates and elaborates upon the information presented in the infobox. This task aligns with the broader goal of table-to-text conversion, where the challenge lies in transforming concise, structured data into coherent and informative textual narratives. By mapping infobox content to descriptive text, the dataset facilitates the development and evaluation of models capable of generating fluent and contextually accurate prose from sparse, tabular representations. Thus, the WikiBio dataset plays a crucial role in advancing research and methodologies for data-to-text generation, offering a standardized benchmark for assessing model performance and innovation in this domain.

3.4 ToTTo

The ToTTo dataset, sourced from Wikipedia, focuses on the task of converting highlighted rows into natural language text (Parikh et al., 2020). Each instance in the dataset involves selecting specific rows from a table, termed as highlighted rows, and generating textual descriptions that independently convey the meaning of these selected cells. This task is designed to ensure that the generated text remains contextually coherent and informative, despite being isolated from the surrounding rows in the table. The dataset comprises a training set with $120,761$ instances, along with development and test sets each containing $7,700$ instances. On average, the highlighted rows contain 3.55 rows and the target text spans approximately 17.4 tokens, reflecting the dataset's focus on generating concise yet informative descriptions from structured data. By providing a standardized benchmark for table-to-text conversion, the ToTTo dataset facilitates the evaluation and development of models capable of accurately and fluently transforming tabular data into natural language text, thereby advancing research in data-driven text generation tasks.

3.5 Wiki Table-to-text

Wiki Table-to-text dataset, akin to the ToTTo dataset (Bao et al., 2018), shares a similar objective of transforming structured data into natural language text. The dataset consists of tables scraped

Table Title: Gabriele Becker
Section Title: International Competitions
Table Description: None

Year	Competition	Venue	Position	Event	Notes
Representing Germany					
1992	World Junior Championships	Seoul, South Korea	10th (semis)	100 m	11.83
1993	European Junior Championships	San Sebastián, Spain	7th	100 m	11.74
			3rd	4x100 m relay	44.60
1994	World Junior Championships	Lisbon, Portugal	12th (semis)	100 m	11.66 (wind: +1.3 m/s)
			2nd	4x100 m relay	44.78
1995	World Championships	Gothenburg, Sweden	7th (q-finals)	100 m	11.54
			3rd	4x100 m relay	43.01

Original Text: After winning the German under-23 100 m title, she was selected to run at the 1995 World Championships in Athletics both individually and in the relay.
Text after Deletion: she at the 1995 World Championships in both individually and in the relay.
Text After Decontextualization: Gabriele Becker competed at the 1995 World Championships in both individually and in the relay.
Final Text: Gabriele Becker competed at the 1995 World Championships both individually and in the relay.

Figure 3: Caption

Figure 4: Wiki Table-to-text Dataset

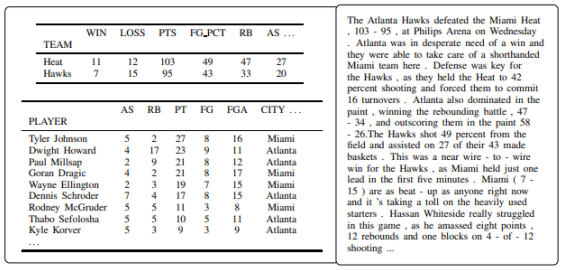
from Wikipedia, where specific rows, termed highlighted rows, are annotated to serve as target text in natural language. These annotations are meticulously crafted to ensure that the generated text is contextually independent from other rows within the table. The dataset encompasses approximately 5,000 randomly selected tables from Wikipedia, each annotated by manual annotators to highlight key information that should be converted into textual descriptions. In total, the dataset contains around 13,318 sentences, providing a diverse and extensive collection of instances for training and evaluating models in the table-to-text generation task. By focusing on context-free annotations and leveraging real-world data from Wikipedia, the Wiki Table-to-text dataset serves as a valuable resource for advancing research and development in natural language generation from structured data sources. It offers a standardized benchmark for evaluating the efficacy and accuracy of models in converting tabular information into coherent and informative textual narratives.

3.6 Rotowire

The Rotowire dataset is specifically designed for the task of generating text from NBA basketball game statistics. It encompasses detailed tables representing various matches and the overall performances of different players. This dataset is distinctive as it pairs human-written NBA basketball game summaries with their corresponding box and line scores, providing a rich source of structured and narrative data. The summaries, sourced from

rotowire.com, form what is referred to as the "rotowire" data. The dataset comprises 4,853 unique summaries covering NBA games played between January 1, 2014, and March 29, 2017, with some games having multiple summaries to enhance diversity and comprehensiveness.

The dataset is split into training, validation, and test sets, containing 3,398, 727, and 728 summaries, respectively. This structured split ensures a robust framework for training and evaluating models. An example of a data instance from the Rotowire dataset is illustrated in Figure 5, which showcases how detailed statistical tables are aligned with narrative summaries. The comprehensive nature of the Rotowire dataset, with its extensive game statistics and human-crafted summaries, makes it an ideal benchmark for evaluating the effectiveness of table-to-text generation models in handling complex and context-rich data.



TEAM	WIN	LOSS	PTS	FG_PCT	RB	AS ...
Heat	11	12	103	49	47	27
Hawks	7	15	95	43	33	20

PLAYER	AS	RB	PT	FG	FGA	CITY ...
Tyler Johnson	5	2	27	8	16	Miami
Dwight Howard	4	17	23	9	11	Atlanta
Paul Millsap	2	9	21	8	12	Atlanta
Goran Dragic	4	2	21	8	17	Miami
Wayne Ellington	2	3	19	7	15	Miami
Dennis Schroder	7	4	17	8	15	Atlanta
Rodney McGruder	5	5	11	3	8	Miami
Thabo Sefolosha	5	5	10	5	11	Atlanta
Kyle Korver	5	3	9	3	9	Atlanta
...						

The Atlanta Hawks defeated the Miami Heat , 103 - 95 , at Philips Arena on Wednesday . Atlanta was in desperate need of a win and they were able to take care of a shorthanded Miami team here . Defense was key for the Hawks , as they held the Heat to 42 percent shooting and forced them to commit 16 turnovers . Atlanta also dominated in the paint , winning the rebounding battle , 47 - 34 , and outscoring them in the paint 58 - 26.The Hawks shot 49 percent from the field and assisted on 27 of their 43 made baskets . This was a near wire - to - wire win for the Hawks , as Miami held just one lead in the first five minutes . Miami (7 - 15) are as beat - up as anyone right now and it 's taking a toll on the heavily used starters . Hassan Whiteside really struggled in this game , as he amassed eight points , 12 rebounds and one blocks on 4 - of - 12 shooting ...

Figure 1: An example data-record and document pair from the ROTOWIRE dataset. We show a subset of the game's records (there are 628 in total), and a selection from the gold document. The document mentions only a select subset of the records, but may express them in a complicated manner. In addition to capturing the writing style, a generation system should select similar record content, express it clearly, and order it appropriately.

Figure 5: Example of Rotowire tables.

4 Architecture and Experiments

There has been some significant amount of work in table-to-text generation previously. In this section we try to summarize the work suitably.

4.1 Wikibio Problem Approaches

Rebuffel et al. (2022) coined a simple language modelling solution to the *Wikibio* problem. The language modelling problem is explained in equation 1.

$$P(s) = \prod_{t=1}^{T} P(w_t | w_1, ..., w_{t-1}) \qquad (1)$$

Here $w_1, ..., w_T$ are T words from the vocabulary.

Table linearization plays a pivotal role in the deployment of sequence-to-sequence models. The task under consideration, utilizing the (Lebret et al., 2016a) dataset, benefits from the simplification of tables inherent to this dataset. In particular, the tables in this dataset are typically reduced to a single column, thereby facilitating their representation as key-value pairs. This simplification is crucial as it reduces the complexity involved in processing and converting tabular data into a linear format suitable for sequence-to-sequence modeling. The one-column structure of these tables ensures that each entry can be directly mapped to a corresponding value, streamlining the linearization process and improving the model's ability to generate coherent and contextually appropriate sequences based on the tabular inputs.

The authors employ both local conditioning and global conditioning techniques to flatten a table, facilitating the transformation of tabular data into a linear sequence suitable for sequence-to-sequence modellings.

Local conditioning refers to the integration of information from the table that is specifically applied to the description of words that have already been generated. This technique leverages the context provided by the previously generated words, ensuring that the model's output remains coherent and contextually relevant. By conditioning on the immediate linguistic context, the model can dynamically adjust its predictions based on the evolving sequence, thus enhancing the fluency and accuracy of the generated text.

In contrast, **global conditioning** encompasses the incorporation of information from all tokens and fields of the table, irrespective of their presence in the previously generated words. This broader approach ensures that the model has access to the entire dataset's content, allowing it to maintain a comprehensive understanding of the tabular information throughout the generation process. By considering the entirety of the table, global conditioning helps the model to produce outputs that are more comprehensive and reflective of the overall data structure, ensuring that no relevant information is overlooked.

In the context of scoring output words, a copy action is employed to ensure the factual consistency of the generated text. This mechanism is crucial for maintaining the integrity of the information conveyed by the model, particularly when transforming structured tabular data into natural language descriptions.

The copy action involves directly incorporating words or phrases from the fact table into the gener-

ated output, thereby preserving the factual accuracy of the information. This is particularly important in domains where precision and reliability of the data are paramount, such as biographical summaries, scientific reports, and other information-rich texts.

On the other hand, Liu et al. (2017) developed a structure-aware model to enhance the task of table-to-text generation. This model builds upon the fundamental principle of field-value or key-value pairs, which serve as the cornerstone for representing the tabular data. However, Liu et al. (2017) advanced this concept by introducing a field gating encoder and a description decoder, thereby improving the model's ability to generate coherent and contextually accurate descriptions.

The objective function of the encoder-decoder architecture, as proposed by Liu et al. (2017), is formally defined by the equation 2. This objective function is crucial for training the model, as it guides the optimization process by quantifying the discrepancy between the generated descriptions and the ground truth data. By minimizing this objective function, the model learns to generate more accurate and contextually appropriate descriptions from the tabular input.

Equation 2 captures the essence of this optimization process, encompassing the various components and interactions within the encoder-decoder framework. It integrates the contributions of both the field gating encoder and the description decoder, ensuring that the model effectively leverages the structure-aware design to produce high-quality textual outputs. Through this advanced approach, Liu et al. (2017) have demonstrated a significant improvement in the capability of sequence-to-sequence models to handle table-to-text generation tasks, highlighting the importance of incorporating structural awareness and selective attention mechanisms in model design.

$$w^*_{1:p} = \underset{w_{1:p}}{\operatorname{argmax}} \prod_{t=1}^{p} P(w_t | w_{0:t-1}, R_{t,n}) \quad (2)$$

The concept of field embedding, as extracted from Lebret et al. (2016b), introduces a sophisticated approach to representing structured data for natural language generation tasks. This technique involves pairing field names and values along with the positional information of tokens, thereby enhancing the model's ability to comprehend and utilize the contextual relationships within the data.

Field embedding extends beyond the simple key-value representation by embedding both the field names and their corresponding values into a continuous vector space. This embedding process captures the semantic relationships between different fields and values, facilitating a more nuanced understanding of the data's structure. By incorporating positional information, the model gains insight into the sequential order and hierarchical relationships present within the table, further enriching its contextual comprehension.

The utility of field embedding lies in its ability to provide a comprehensive understanding of the facts and their context within the table. By embedding field names and values together with their positional information, the model can discern the significance of each field in relation to others, thus improving its ability to generate coherent and factually consistent text. This approach ensures that the generated descriptions are not only fluent and contextually relevant but also aligned with the underlying structure and semantics of the tabular data.

In summary, the concept of field embedding as derived from Lebret et al. (2016b) plays a critical role in enhancing the performance of sequence-to-sequence models for table-to-text generation. By embedding field names, values, and positional information into a unified representation, the model gains a deeper understanding of the data's structure and context, leading to more accurate and contextually appropriate textual outputs. This advanced embedding technique underscores the importance of capturing the detailed relationships within structured data to improve the quality of natural language generation.

The table encoder is designed to effectively encode each word in the table, integrating its corresponding field embedding into a hidden state using a Long Short-Term Memory (LSTM) encoder. This process ensures that both the semantic content of the words and their contextual information derived from field embeddings are captured, providing a rich and comprehensive representation of the tabular data.

The LSTM encoder processes the input sequence of table words and their field embeddings sequentially. Each word, along with its field embedding, is transformed into an embedded vector, which is then fed into the LSTM. The LSTM's hidden states are updated iteratively as it processes each

word in the sequence, capturing both the immediate context and long-term dependencies. This encoding process results in a series of hidden states that encapsulate the detailed structural and semantic information of the table.

To address the structured table during the generation of descriptions, a dual attention mechanism is employed within the LSTM architecture. This mechanism enables the model to perform both local and global addressing, ensuring that the generated descriptions are coherent and contextually accurate.

The combination of these two attention mechanisms within the LSTM architecture allows the description generator to effectively leverage the encoded table information. By dynamically attending to both local and global contexts, the model can produce detailed, accurate, and contextually appropriate descriptions from the structured data.

The architecture, as depicted in Figure 6, illustrates the interplay between the table encoder and the dual-attention LSTM description generator. The figure highlights the flow of information from the table, through the field embeddings and LSTM encoder, to the dual-attention mechanism, and finally to the generated descriptions. This comprehensive framework underscores the importance of integrating advanced encoding and attention mechanisms to enhance the performance of table-to-text generation models.

In summary, the table encoder and dual-attention LSTM architecture work in tandem to encode and generate descriptions from structured tables. The use of field embeddings and attention mechanisms ensures that the model captures the full context and semantics of the tabular data, resulting in high-quality, contextually accurate textual outputs.

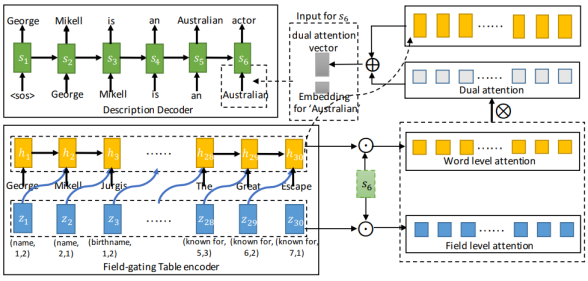


Figure 6: The overall diagram of structure-aware seq2seq architecture for generating description.

On the other hand, the Multi-Branch Decoder (MBD) architecture (Rebuffel et al., 2022) aims to enhance the text generation process by isolating and independently controlling critical co-dependent factors. This architecture incorporates separate decoding modules, or branches, for each control factor—content, hallucination, and fluency. Each branch is responsible for modeling its respective factor, producing an output representation that can be dynamically weighted according to its importance in the final output. The content branch ensures the generated text accurately reflects the factual information from the input data, minimizing errors and inaccuracies. The hallucination branch focuses on controlling and mitigating the generation of information not present in the input, thus enhancing the reliability of the output. Meanwhile, the fluency branch ensures the text is linguistically coherent and natural, maintaining grammatical correctness and readability. By combining these branches' output representations in a weighted manner, the MBD architecture provides precise control over the trade-offs between content accuracy, hallucination reduction, and fluency. This separation and targeted control result in higher quality text generation, with the flexibility to adapt to various scenarios and requirements, ultimately improving the overall reliability and coherence of the generated descriptions.

4.2 ToTTo Challenge Approaches

Along with providing the dataset, Parikh et al. (2020) introduced three approaches to tackle the challenge, with the first approach being the BERT-to-BERT framework. This framework employs a Transformer encoder-decoder model, where both the encoder and decoder are initialized with the BERT (Bidirectional Encoder Representations from Transformers) model. The BERT model, pre-trained on extensive corpora including Wikipedia and the BooksCorpus, is known for its powerful contextual understanding of language. In this approach, the encoder transforms the structured table data into contextualized representations, leveraging BERT's bidirectional context capabilities. Simultaneously, the decoder generates the output sequence, drawing from these rich representations to produce coherent and contextually appropriate text. The BERT-to-BERT framework benefits from the comprehensive pre-training of BERT, which captures intricate linguistic patterns and contextual information, thereby enhancing the accuracy and fluency of the generated descriptions. This approach highlights the effectiveness of initializ-

ing both components of the model with pre-trained BERT, providing a robust solution for generating natural language from structured data.

The second approach consists of a Pointer Generator Network, a sequence-to-sequence (Seq2Seq) model augmented with attention and a copy mechanism. This model, originally designed for the task of text summarization, has been effectively adapted for data-to-text generation. The attention mechanism within the Pointer Generator Network allows the model to focus on different parts of the input sequence as it generates each word in the output, thereby enhancing the contextual relevance of the generated text. The copy mechanism further enhances the model's capability by enabling it to directly copy words from the input sequence to the output, ensuring factual accuracy and preserving key information from the source data. This dual mechanism of attention and copying is particularly advantageous in data-to-text tasks, where maintaining the integrity of the original data while generating coherent and fluent text is critical. The Pointer Generator Network thus combines the strengths of both traditional Seq2Seq models and specialized mechanisms for handling structured data, making it a robust approach for converting tabular data into natural language descriptions.

The third approach is a sequence-to-sequence (Seq2Seq) model that incorporates an explicit content selection and planning mechanism specifically designed for data-to-text generation. This model enhances the traditional Seq2Seq framework by introducing a preliminary step that identifies and selects the most relevant content from the input data before the text generation phase begins. The content selection mechanism ensures that only the most pertinent pieces of information are considered, thereby reducing noise and improving the focus of the generated text. Following this, the planning mechanism organizes the selected content into a coherent structure, outlining the sequence in which the information should be presented. This explicit planning phase guides the generation process, allowing the model to produce text that is not only contextually accurate but also logically and sequentially coherent. By integrating these additional mechanisms, this approach addresses some of the inherent challenges in data-to-text tasks, such as content relevance and structural coherence, thereby enhancing the overall quality and reliability of the generated descriptions.

On the other hand, Kale and Rastogi (2020) developed a transfer learning-based approach utilizing the T5 (Text-to-Text Transfer Transformer) model. This approach leverages the pre-trained T5 checkpoint, which has been extensively trained in a multitask manner. The pre-training of T5 involves an unsupervised "span masking" objective applied to Common Crawl data, where spans of text are masked and the model learns to predict them, thus gaining a deep understanding of language patterns and contextual relationships. Additionally, the T5 model is trained on various supervised tasks including translation, summarization, classification, and question answering. This multitask training regime enables T5 to acquire versatile language generation capabilities and adapt effectively to diverse tasks. By employing this pre-trained T5 model, Kale and Rastogi (2020) harness the benefits of transfer learning, allowing the model to effectively generalize from its extensive training on varied datasets to the specific challenges of data-to-text generation. This approach underscores the efficacy of leveraging pre-trained language models for complex generation tasks, providing a robust foundation for producing high-quality and contextually accurate text from structured data.

T5 has also contributed significantly to addressing the challenges in data-to-text generation, as highlighted by Gehrmann et al. (2021). In their experiments, they employed both T5 and BART (Bidirectional and Auto-Regressive Transformers) models across various data-to-text generation tasks. Although the T5 model's performance on the ToTTo dataset did not surpass the state-of-the-art results, the experiments yielded important insights into the complexities and nuances of the challenge. These insights include understanding the limitations and strengths of pre-trained models like T5 in handling structured data and generating coherent, contextually accurate text. Moreover, the comparative analysis with BART provided further understanding of the models' behaviors and capabilities, contributing valuable knowledge to the field of data-to-text generation. This work underscores the significance of continued experimentation and evaluation with advanced pre-trained models, even when they do not achieve top-tier performance, as they can offer critical learnings and drive further advancements in the domain.

Wang et al. (2022) introduced the transformation-invariant graph masking technology within the

LATTICE framework, designed to enforce the model's structure-awareness and transformation-invariance. This innovative approach ensures that the model can effectively handle variations in the input data's structure, maintaining robust performance across different transformations. The transformation-invariant graph masking technology allows LATTICE to encode the input data in a way that preserves its inherent structure, regardless of transformations applied to it. Additionally, Wang et al. (2022) presented two alternative techniques aimed at enhancing transformation-invariance, providing a comparative analysis with LATTICE. These alternative techniques offer different methodologies for achieving transformation-invariance, enabling a thorough evaluation of their effectiveness relative to the LATTICE framework. By introducing and comparing these approaches, the study contributes valuable insights into the development of structure-aware models that can adapt to varying data transformations, thereby advancing the field of data-to-text generation.

Content-invariant transformations consist of operations that do not alter the content within individual rows or columns but instead modify the table's layout while preserving the semantic equivalence of the (sub-)tables. These transformations enable the presentation of the same information in various table configurations, ensuring that the underlying data remains consistent despite changes in its structural representation. By applying such transformations, models can be trained to recognize and generate text from differently organized tables without losing the integrity of the information.

Pretrained Transformer-based generative models, such as T5 and BART, have demonstrated state-of-the-art (SOTA) performance across various text generation tasks. These models are pretrained on an extensive range of supervised and self-supervised text-to-text tasks, equipping them with robust language understanding and generation capabilities. During pre-training, these models learn to handle a variety of text manipulation tasks, from summarization and translation to classification and question answering. This comprehensive training enables them to generalize effectively to new tasks, including data-to-text generation, where they can leverage their learned representations and generation strategies to produce coherent and contextually accurate text. The ability of these models to handle content-invariant transformations further

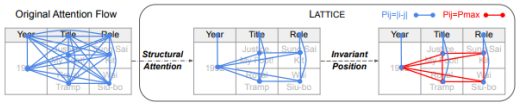


Figure 3: Attention flows of the base model and LATTICE. In this example, we adopt the input format which appends headers to each cell, so headers can be seen as part of the cell content. We omit the attention flows among tokens within a cell, as they are in the same type of the flows between headers and corresponding cells. P_{ij} represents the relative position between tokens at both ends of the attention flow, where i and j are absolute positions of tokens in the linearized table and P_{max} is the max relative position allowed. The base model has a complete attention graph among all cells with relative positions based on linear distance. LATTICE prunes the attention flow based on the table layout and assigns transformation-invariant relative positions between cells.

Figure 7: Attention flows of the base model and LATTICE.

enhances their robustness and versatility, making them highly effective for generating text from structured data presented in different formats.

In the figure 7 the architecture is explained.

4.3 Rotowire Challenge Approaches

In 2017, Wiseman et al. (2017) introduced innovative neural text generation methods to address the RotoWire challenge, which involves generating textual summaries from structured data. This work marked a significant advancement in the field and sparked a surge of research focused on enhancing encoder-decoder models with the ability to copy words directly from the source material. These augmented models leverage a copy mechanism, allowing them to transfer specific information verbatim from the input data to the output text. This capability is crucial for maintaining factual accuracy and ensuring that essential details are accurately reflected in the generated text. The incorporation of the copy mechanism into neural text generation models has proven particularly effective for tasks like the RotoWire challenge, where the preservation of precise information is paramount. By combining the strengths of traditional sequence-to-sequence frameworks with advanced copy mechanisms, these models achieve a higher level of fidelity and coherence in data-to-text generation, addressing some of the key challenges in this domain.

$$p(y_i|y_{1:i-1}, s) = \sum_{z \in \{0,1\}} p(y_i|z, y_{1:i-1}, s) \quad (3)$$

The generation of copied content can be formally defined by equation 3, where the functions 'copy' and 'gen' are parameterized in terms of the decoder RNN's hidden state. These functions assign scores to words, determining whether a word should be generated from the vocabulary or copied directly from the input source. Models incorporating copy-decoders are typically trained to minimize the negative log marginal probability, effectively marginal-

izing out the latent variable associated with the copy mechanism. During training, reconstruction-based techniques can be employed at both the document and sentence levels, enhancing the model's ability to produce accurate and coherent text. Additionally, a fully differentiable approach utilizing the decoder's hidden states has been successfully applied in neural machine translation, demonstrating the efficacy of these methods in generating high-quality translations. This approach leverages the rich contextual information captured by the hidden states, enabling the model to make more informed decisions about when to generate or copy content, thereby improving the overall performance of the text generation process.

Puduppully et al. (2019) developed an innovative content selection and planning-based approach for table-to-text generation. As illustrated in Figure 8, this method employs a content selection gate applied to the table. This gate is responsible for identifying and extracting factual content from the table, ensuring that the most relevant and accurate information is selected for inclusion in the generated text. Following content selection, a planning network is utilized to organize the extracted information in a coherent and logical sequence, facilitating the generation of well-structured and contextually appropriate content. This dual mechanism of content selection and planning enhances the overall quality and factual accuracy of the generated text, addressing one of the critical challenges in table-to-text generation by systematically managing the information flow from structured tables to natural language descriptions.

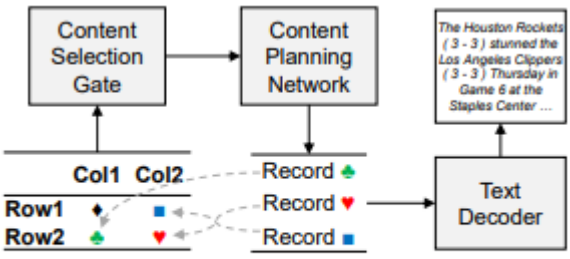


Figure 1: Block diagram of our approach.

Figure 8: Approach of Content selection and planning by Puduppully et al. (2019).

Thus, the ordered plan is fed into a text decoder, which generates the final content. As depicted in Figure 9, the modeling approach taken by the authors begins with a record encoder that processes the unordered table input, extracting relevant facts. These extracted facts then pass through a content selection gate, which evaluates the context of each record to determine its importance relative to other records in the table. By capturing these dependencies among records, the content selection gate mechanism ensures that only the most pertinent information is prioritized for text generation. This structured approach not only enhances the factual accuracy of the generated text but also ensures a coherent narrative flow by appropriately ordering and contextualizing the information extracted from the table.

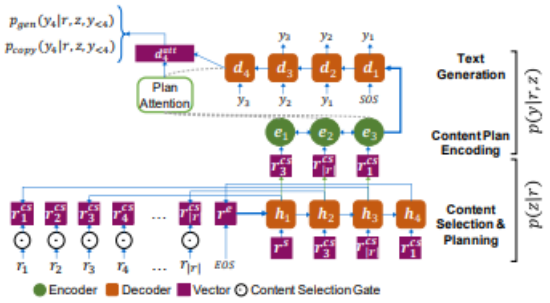


Figure 2: Generation model with content selection and planning; the content selection gate is illustrated in Figure 3.

Figure 9: Modeling approach by Puduppully et al. (2019)

In our generation task, the output text is lengthy but adheres to a canonical structure, ensuring consistency and clarity. The planning process involves mapping the text in the summaries to entities in the input table, including their values and types. Since the output tokens of the content planning stage correspond to positions in the input sequence, a Pointer Network is employed to effectively manage these mappings. The text decoder, which is based on a recurrent neural network with LSTM units, is initialized with the hidden states from the final step of the encoder. This initialization helps maintain contextual information throughout the decoding process, thereby producing coherent and contextually relevant text. This methodology ensures that the generated text not only follows a structured narrative but also accurately reflects the content and context of the input table, thereby enhancing the overall quality and reliability of the output.

Similar contributions are also seen in rotowire challenge by Puduppully and Lapata (2021), Choi et al. (2021), Rebuffel et al. (2019), Li et al. (2021)

and Gong et al. (2019).

5 Results

The results obtained from various dataset challenges exhibit substantial variability, which can be attributed to several factors including the complexity, type, context, and factual consistency of the datasets involved. These challenges often differ in their structural intricacies and the specific requirements they impose on the generation models. For instance, datasets with complex relational structures or those requiring high contextual relevance may present greater challenges for text generation systems. Moreover, the ability to maintain factual consistency is critical, as models must accurately reflect the data in the generated text without introducing errors or hallucinations. Table 1 provides a summary of the performance scores achieved by various approaches across different datasets. This comparison highlights the strengths and limitations of each approach in handling diverse data-to-text generation tasks, offering insights into how different methodologies perform under varying conditions. By analyzing these results, researchers can identify key factors that influence performance and guide the development of more robust and adaptable models for future applications.

In Table 1, we observe that the highest BLEU score achieved on the WikiBio challenge is 44.89. This challenge has predominantly utilized BLEU and ROUGE scores to evaluate the effectiveness of various systems. Among the contributions, the field-gating seq2seq model with dual attention mechanisms has performed exceptionally well, demonstrating superior capabilities in handling the table-to-text generation task by efficiently leveraging field information and attention-based context modeling. However, despite its high performance, the reasons behind the underperformance of the beam search strategy when applied on top of this model remain inadequately explained in the literature. The Multi-Branch Decoder (MBD) architecture also shows commendable results, suggesting its potential in balancing various aspects such as content, hallucination, and fluency during text generation. These findings highlight the importance of incorporating advanced attention mechanisms and robust architectural designs to enhance the quality and accuracy of generated text, while also emphasizing the need for further investigation into optimization techniques like beam search within these frameworks.

On the other hand, for the ToTTo dataset, the T5-3B model has emerged as the top performer, achieving a BLEU score of 49.5. In this context, the PARENT score is also commonly used alongside the BLEU score to provide a more comprehensive evaluation of the models' performance, particularly in capturing content fidelity and fluency. The LATTICE and BERT-to-BERT models also demonstrate competitive results, closely trailing the T5-3B model. The ToTTo challenge has garnered considerable attention and contributions over the years, with researchers continually refining their approaches to tackle the unique complexities of the dataset. This dataset's emphasis on generating context-independent text from highlighted rows of tables has driven the development of sophisticated models that excel in ensuring factual consistency and contextual coherence. The progressive improvements in performance metrics, as evidenced by the high scores of recent models, underscore the ongoing advancements in the field and the potential for further breakthroughs in table-to-text generation methodologies.

Rotowire presents one of the most challenging tasks in table-to-text generation due to the extensive length and complexity of its tables. These intricate tables require models to handle a significant amount of structured data, making the generation task exceptionally demanding. Among the various models evaluated on this dataset, the hierarchical encoder-decoder model with Numeric Reasoning (NR) and Information Retrieval (IR) capabilities has achieved the highest performance, with a BLEU score of 17.96. This model's ability to process and synthesize large amounts of data into coherent narratives underscores the importance of hierarchical structures and specialized reasoning mechanisms in handling such complex datasets. The vast context length inherent to the Rotowire dataset further necessitates the use of encoder-decoder architectures, which can effectively manage long sequences of data and maintain contextual relevance throughout the generated text. This demonstrates that while Rotowire poses significant challenges, it also highlights the critical advancements in model architectures needed to address the intricate demands of complex table-to-text generation tasks.

Challenge	Model	Citation	Bleu	Rouge	Parent	Meteor
Wikibio	Table NLM	(Lebret et al., 2016a)	34.70	25.80		
	Field-gating Seq2seq + dual attention	(Liu et al., 2017)	**44.89**	41.21		
	Field-gating Seq2seq + dual attention + beam search	(Liu et al., 2017)	44.71	41.65		
	MBD	(Rebuffel et al., 2022)	41.56		56.16	
ToTTo	NCP+CC	(Parikh et al., 2020)	19.2		29.2	
	Pointer Generator	(Parikh et al., 2020)	41.6		51.6	
	BERT-to-BERT	(Parikh et al., 2020)	44		52.6	
	T5-3B	(Kale and Rastogi, 2020)	**49.5**		58.4	
	T5	(Gehrmann et al., 2021)				36.3
	LATTICE	(Wang et al., 2022)	48.4		58.4	
Rotowire	Encoder-decoder + conditional copy	(Wiseman et al., 2017)	14.19			
	Neural Content Planning + conditional copy	(Puduppully et al., 2019)	16.50			
	Hierarchical transformer encoder + conditional copy	(Rebuffel et al., 2019)	17.50			
	HierarchicalEncoder + NR + IR	(Li et al., 2021)	**17.96**			
	Force-Copy	(Choi et al., 2021)	17.26			
	Macro	(Gong et al., 2019)	15.46			

Table 1: Comprehensive quantitative evaluation of all the different models and approaches in these challenges.

6 Conclusion

Over the years, the domain of table-to-text generation has witnessed substantial advancements. The aforementioned challenges and corresponding datasets, such as WikiBio, ToTTo, and Rotowire, represent pivotal contributions that have significantly shaped the research landscape. Although there remains considerable scope for enhancement, these works offer a solid foundation for exploring and addressing the intricacies inherent in table-to-text generation tasks. Each dataset and challenge presents unique aspects of complexity, thereby ensuring the developed systems are versatile and applicable across various scenarios. Despite the progress, future research must focus on integrating subjectivity into the generated text and addressing more complex and realistic tables. These areas remain relatively unexplored, presenting exciting opportunities for innovation and development in creating more nuanced and sophisticated table-to-text generation systems.

References

Junwei Bao, Duyu Tang, Nan Duan, Zhao Yan, Yuanhua Lv, Ming Zhou, and Tiejun Zhao. 2018. Table-to-text: Describing table region with natural language.

Thiago Castro Ferreira, Claire Gardent, Nikolai Ilinykh, Chris van der Lee, Simon Mille, Diego Moussallem, and Anastasia Shimorina, editors. 2020. *Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+)*. Association for Computational Linguistics, Dublin, Ireland (Virtual).

Sanghyuk Choi, Jeong in Hwang, Hyungjong Noh, and Yeonsoo Lee. 2021. May the force be with your copy mechanism: Enhanced supervised-copy method for natural language generation.

Ondřej Dušek, David M. Howcroft, and Verena Rieser. 2019. Semantic noise matters for neural natural language generation. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 421–426, Tokyo, Japan. Association for Computational Linguistics.

Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. Creating training corpora for NLG micro-planners. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 179–188, Vancouver, Canada. Association for Computational Linguistics.

Sebastian Gehrmann, Tosin Adewumi, Karmanya Aggarwal, Pawan Sasanka Ammanamanchi, Anuoluwapo Aremu, Antoine Bosselut, Khyathi Raghavi Chandu, Miruna-Adriana Clinciu, Dipanjan Das, Kaustubh Dhole, Wanyu Du, Esin Durmus, Ondřej Dušek, Chris Chinenye Emezue, Varun Gangal, Cristina Garbacea, Tatsunori Hashimoto, Yufang Hou, Yacine Jernite, Harsh Jhamtani, Yangfeng Ji, Shailza Jolly, Mihir Kale, Dhruv Kumar, Faisal Ladhak, Aman Madaan, Mounica Maddela, Khyati Mahajan, Saad Mahamood, Bodhisattwa Prasad Majumder, Pedro Henrique Martins, Angelina McMillan-Major, Simon Mille, Emiel van Miltenburg, Moin Nadeem, Shashi Narayan, Vitaly Nikolaev, Andre Niyongabo Rubungo, Salomey Osei, Ankur Parikh, Laura Perez-Beltrachini, Niranjan Ramesh Rao, Vikas Raunak, Juan Diego Rodriguez, Sashank Santhanam, João Sedoc, Thibault Sellam, Samira Shaikh, Anastasia Shimorina, Marco Antonio Sobrevilla Cabezudo, Hendrik Strobelt, Nishant Subramani, Wei Xu, Diyi Yang, Akhila Yerukola, and Jiawei Zhou. 2021. The GEM benchmark: Natural language generation, its evaluation and metrics. In *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, pages 96–120, Online. Association for Computational Linguistics.

Li Gong, Josep Crego, and Jean Senellart. 2019. Enhanced transformer model for data-to-text generation. In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, pages 148–156, Hong Kong. Association for Computational Linguistics.

Mihir Kale and Abhinav Rastogi. 2020. Text-to-text pre-training for data-to-text tasks. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 97–102, Dublin, Ireland. Association for Computational Linguistics.

Rémi Lebret, David Grangier, and Michael Auli. 2016a. Neural text generation from structured data with application to the biography domain. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1203–1213, Austin, Texas. Association for Computational Linguistics.

Rémi Lebret, David Grangier, and Michael Auli. 2016b. Neural text generation from structured data with application to the biography domain. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1203–1213, Austin, Texas. Association for Computational Linguistics.

Liang Li, Can Ma, Yinliang Yue, and Dayong Hu. 2021. Improving encoder by auxiliary supervision tasks for table-to-text generation. 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 5979–5989, Online. Association for Computational Linguistics.

Tianyu Liu, Kexiang Wang, Lei Sha, Baobao Chang, and Zhifang Sui. 2017. Table-to-text generation by structure-aware seq2seq learning. In *AAAI Conference on Artificial Intelligence*.

Linyong Nan, Dragomir Radev, Rui Zhang, Amrit Rau, Abhinand Sivaprasad, Chiachun Hsieh, Xiangru Tang, Aadit Vyas, Neha Verma, Pranav Krishna, Yangxiaokang Liu, Nadia Irwanto, Jessica Pan, Faiaz Rahman, Ahmad Zaidi, Mutethia Mutuma, Yasin Tarabar, Ankit Gupta, Tao Yu, Yi Chern Tan, Xi Victoria Lin, Caiming Xiong, Richard Socher, and Nazneen Fatema Rajani. 2021. DART: Open-domain structured data record to text generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 432–447, Online. Association for Computational Linguistics.

Ankur Parikh, Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. 2020. ToTTo: A controlled table-to-text generation dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1173–1186, Online. Association for Computational Linguistics.

Ratish Puduppully, Li Dong, and Mirella Lapata. 2019. Data-to-text generation with content selection and planning. AAAI'19/IAAI'19/EAAI'19. AAAI Press.

Ratish Puduppully and Mirella Lapata. 2021. Data-to-text Generation with Macro Planning. *Transactions of the Association for Computational Linguistics*, 9:510–527.

Clement Rebuffel, Marco Roberti, Laure Soulier, Geoffrey Scoutheeten, Rossella Cancelliere, and Patrick Gallinari. 2022. Controlling hallucinations at word level in data-to-text generation. *Data Min. Knowl. Discov.*, 36(1):318–354.

Clément Rebuffel, Laure Soulier, Geoffrey Scoutheeten, and Patrick Gallinari. 2019. A hierarchical model for data-to-text generation.

Fei Wang, Zhewei Xu, Pedro Szekely, and Muhao Chen. 2022. Robust (controlled) table-to-text generation with structure-aware equivariance learning. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational*

Linguistics: Human Language Technologies, pages 5037–5048, Seattle, United States. Association for Computational Linguistics.

Sam Wiseman, Stuart Shieber, and Alexander Rush. 2017. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2253–2263, Copenhagen, Denmark. Association for Computational Linguistics.

Victor Zhong, Caiming Xiong, and Richard Socher. 2017. Seq2sql: Generating structured queries from natural language using reinforcement learning. *CoRR*, abs/1709.00103.