The Name of the Title Is Hope

YUPENG HE

Table-to-text (Table2Text) generation is a pivotal task in natural language generation, aiming to convert structured tabular data into fluent, contextually appropriate, and informative textual descriptions. Over time, this field has evolved from early sequence-to-sequence neural models to more sophisticated techniques that incorporate content selection, planning, and template-based strategies to improve factual accuracy and controllability. The introduction of large-scale datasets and robust evaluation protocols has further propelled advancements in model benchmarking and comparison. Recent research has explored generation order, insertion-based decoding, and logical reasoning, expanding the adaptability of Table2Text systems to open-domain and complex settings. Despite significant progress, challenges remain, particularly in ensuring factual consistency, preventing hallucination, and maintaining semantic faithfulness. This survey provides a comprehensive overview of the development, methodologies, benchmark datasets, and evaluation methods in Table2Text generation. It concludes by highlighting current limitations and suggesting directions for future research toward more faithful, controllable, and semantically grounded Table2Text systems.

ACM Reference Format:
Yupeng HE. 2025. The Name of the Title Is Hope. 1, 1 (April 2025), 13 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

The increasing prevalence of structured data in the form of tables—spanning domains such as finance, sports, healthcare, and encyclopedic resources—has fueled significant interest in the automatic generation of natural language text from tabular inputs. Table-to-text (Table2Text) generation aims to transform such structured data into fluent, contextually appropriate, and informative textual descriptions, facilitating data accessibility, report automation, and knowledge base augmentation.

Early Table2Text systems relied on rule-based and template-based approaches (e.g., Angeli et al. [2], Reiter and Dale [20]), which incorporated handcrafted rules

Author's Contact Information: Yupeng HE, trovato@corporation.com.

ACM XXXX-XXXX/2025/4-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

and predefined templates for mapping table entries to text. While effective for narrow, domain-specific tasks, these methods struggled to scale to diverse schemas and open-domain scenarios due to their rigidity and high engineering cost.

The advent of neural sequence-to-sequence (seq2seq) models marked a paradigm shift in Table2Text generation. Pioneering work by Lebret et al. [11] demonstrated that neural models could learn to generate biographies from Wikipedia infobox tables, leveraging attention mechanisms to align table fields with generated text. Subsequent research, such as Wiseman et al. [23], highlighted the challenges of producing coherent multi-sentence outputs and maintaining factual consistency.

To address these challenges, methods integrating explicit content selection and text planning were proposed (Puduppully et al. [17]), enabling models to first choose salient table entries and then plan their organization in the output. Further advances explored hybrid approaches that combined neural generation with template-based constraints (Wiseman et al. [24]), balancing flexibility with faithfulness and controllability.

The introduction of large-scale datasets—such as WikiBio (Lebret et al. [11]), WebNLG (Gardent et al. [7]), and ToTTo (Parikh et al. [16])—provided robust benchmarks for model development and comparison. These datasets vary in complexity, domain, and the degree of reasoning required, pushing models to generalize across diverse table schemas and generate factually accurate text.

Recent studies have advanced the field by exploring generation order (Sha et al. [21]), non-monotonic and insertion-based decoding (Gu et al. [8]), and logical reasoning over open-domain tables (Chen et al. [5]). These innovations have broadened the applicability of Table2Text systems and improved their ability to produce coherent, diverse, and semantically faithful outputs. Evaluation practices have also evolved, with an increasing emphasis on both automatic metrics (such as BLEU, ROUGE) and human assessments of factual accuracy and semantic alignment.

Despite these advances, important challenges remain. Models can still hallucinate facts, struggle with novel or complex table schemas, or fail to maintain logical consistency across sentences. Ensuring controllable, explainable, and faithful generation is an active area of research.

This survey provides a comprehensive overview of Table2Text generation, tracing its evolution from early rule-based systems to modern neural architectures. We summarize key methodologies, benchmark datasets, and evaluation protocols, and conclude by discussing open challenges and promising future research directions.

2 Related Work

Research in table-to-text (Table2Text) generation has traversed a rich landscape—from early rule-based systems to the latest neural and retrieval-augmented models. This section reviews influential work, emphasizing each contribution's core advances and remaining limitations.

2.1 Foundations and Early Neural Models

The field of table-to-text (Table2Text) generation began with early rule-based and template-driven approaches (Angeli et al. [2], Reiter and Dale [20]), which enabled interpretable and precise text generation from structured data by encoding expert knowledge and domain-specific rules. These systems were effective for fixed or narrow schemas but struggled to generalize to new domains and could not handle the diversity or ambiguity present in open-domain tables.

The advent of neural sequence-to-sequence (seq2seq) models (Bahdanau et al. [3], Sutskever et al. [22]) marked a transformative step forward. Lebret et al. [11], Mei et al. [15] were among the first to show that neural models could learn mappings from tables to text using attention mechanisms, reducing manual feature engineering and increasing scalability. However, these early models often suffered from hallucination—generating unsupported facts—and faced challenges in maintaining factual consistency and logical ordering, particularly in longer outputs (Wiseman et al. [23]).

2.2 Content Selection, Pretrained Models, and Benchmarks

To address these challenges, subsequent research introduced explicit content selection and planning modules (Liu et al. [14], Puduppully et al. [17]), which improved alignment between the source table and generated text by controlling which entries to mention and their ordering. The emergence of large-scale datasets such as WebNLG (Gardent et al. [7]), WikiBio (Lebret et al. [11]), ToTTo (Parikh et al. [16]) and LogicNLG (Chen et al. [5]) provided robust benchmarks and drove progress on both factuality and reasoning.

The introduction of pretrained language models (PLMs) further elevated Table2Text generation. Models like BART (Lewis et al. [12]), T5 (Raffel et al. [19]), and GPT-3.5/4 (Achiam et al. [1], Brown et al. [4]) have been adapted for structured data tasks, achieving strong results in fluency and generalization. Retrieval-augmented methods, such as RAG (Lewis et al. [12]) and the recently proposed RARR (Zhang et al., 2024), further enhance faithfulness by grounding generation in external knowledge.

2.3 Recent Advances: Faithfulness, Reasoning, and Evaluation

Recent studies have focused on addressing persistent issues of faithfulness, compositional reasoning, and robust evaluation. Several new papers published in 2024 demonstrate the ongoing advances in this field. For example, Zhang and Chen [25] proposed HD-RAG, a retrieval-augmented generation method for hybrid documents containing text and hierarchical tables, showing improvements in factual accuracy and reasoning over complex table structures. Another 2024 work by Fang et al. [6] investigates bias in biography generation from structured data, using counterfactual approaches to analyze model behavior. The Dual-Transformer Table-to-Text Generation Model with Timeline Integration (Qi et al. [18]) introduces a temporal-aware architecture to capture timeline relationships in tabular data. Meanwhile, research into multilingual Table2Text NLG has also emerged, focusing on the challenge of achieving attributability and reasoning in low-resource languages (Haussmann [9]).

Despite significant advancements in Table2Text generation—marked by the transition from rigid, domain-specific systems to flexible, data-driven models with improved factuality and reasoning capabilities—critical challenges persist. State-of-the-art models, including those augmented with reasoning or retrieval modules, remain prone to hallucination, particularly when processing ambiguous, large-scale, or compositionally complex tables. Furthermore, automatic evaluation metrics frequently exhibit weak correlation with human judgments, especially in assessing factual accuracy and logical coherence. As the field progresses, addressing these limitations through scalable verification mechanisms, enhanced reasoning frameworks, and human-centered evaluation methodologies remains a pressing research priority. Ultimately, while contemporary approaches have expanded the scope and adaptability of Table2Text systems, achieving controllable, faithful, and semantically rich generation—particularly for open-domain and structurally intricate tables—continues to pose a fundamental research challenge.

3 Methodology

Table-to-text (Table2Text) generation has progressed through a series of methodological innovations, each designed to address the unique challenges of transforming structured tabular data into coherent, faithful, and contextually appropriate natural language. In this section, we delve into several representative methodologies that have shaped the field, analyzing their technical underpinnings, strengths, and limitations.

3.1 Sequence-to-Sequence Neural Models with Attention

The introduction of neural sequence-to-sequence (seq2seq) architectures marked a turning point in Table2Text generation. Lebret et al. [11] pioneered this direction by representing table records as sequences of field-value pairs and using an encoder-decoder framework with attention to generate biographical sentences from Wikipedia infoboxes. Mei et al. [15] extended this idea with selective generation, introducing a coarse-to-fine alignment mechanism that first selects relevant records and then generates text conditioned on those selections.

These models encode the table as a sequence (or set) of embeddings and generate text token by token, with the attention mechanism dynamically focusing on salient table entries. While seq2seq models with attention have demonstrated strong performance in learning the mapping from table schema to text, they tend to

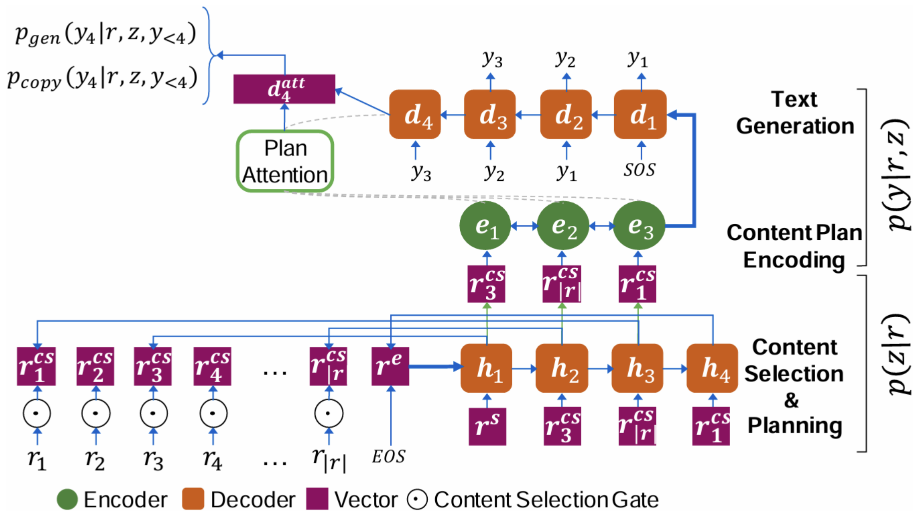

Fig. 1. Generation model with content selection and planning.

suffer from factual hallucinations, especially when dealing with large or complex tables. Furthermore, they typically lack explicit mechanisms for content selection or logical reasoning, which can lead to outputs that omit important information or include irrelevant details.

3.2 Content Selection and Planning-based Approaches

A significant breakthrough in Table2Text generation was the explicit modeling of content selection and planning. This line of work addresses the limitations of vanilla seq2seq models, which often entangle what content to say (content selection) and how to say it (surface realization), leading to omissions, redundancy, or hallucination. Below, we select two papers as examples to illustrate.

Puduppully et al. [17] introduced a two-stage neural architecture (see Figure 1). (1) Content Selector: Given a table, a hierarchical encoder (using GRUs) first encodes each table record and then aggregates across records. An attention-based content selector predicts which records should be verbalized. (2) Content Planner: The selected records are then ordered using a pointer network, which outputs a content plan—a sequence of records representing the intended narrative flow. Finally, a seq2seq decoder with attention generates the text, guided by the content plan. This separation allows the model to reason explicitly about coverage and logical order, reducing redundancy and increasing factual accuracy.

Liu et al. [14]: proposed a structure-aware seq2seq model that integrates content selection into the encoder (see Figure 2). The model uses a graph neural network to capture relationships among table records, enhancing the encoder's ability to

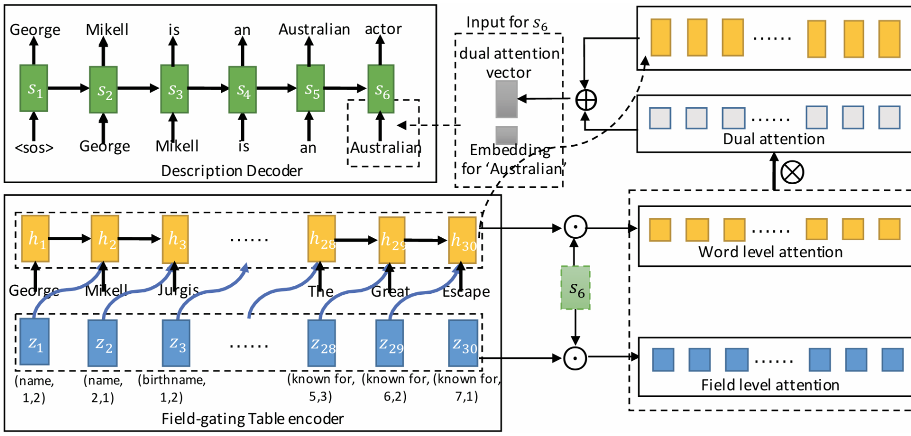

Fig. 2. The overall diagram of structure aware seq2seq architecture for generating description.

distinguish salient from non-salient content. The decoder attends over this enriched representation, promoting better alignment between input and output.

The innovation points of these papers can be summarized as: (1) Modular decomposition of selection, planning, and realization. (2) Use of pointer networks for content ordering. (3) Graph neural encoders to capture table structure and dependencies. (4) Improved factual faithfulness through explicit modeling of content coverage.

3.3 Pretrained Language Models and Table-specific Pretraining

The advent of large pretrained language models (PLMs) and table-specific pretraining has dramatically improved Table2Text performance, especially in terms of fluency, generality, and reasoning.

TAPAS (Herzig et al. [10]) adapts BERT to table input by flattening tables row-wise and using special embeddings to encode row, column, and segment positions (see Figure 3). It is pre-trained on large-scale table question answering and cell selection tasks, enabling the model to learn table structure and cross-cell reasoning. For generation, TAPAS is fine-tuned with an encoder-decoder head or used as an encoder feeding a generative model.

TAPEX (Liu et al. [13]) extends PLM-based table encoding by pretraining a neural SQL executor (see Figure 4). The model learns to execute SQL-like queries on tables, developing a strong internal representation of tabular logic. During Table2Text generation, TAPEX can reason about complex operations (e.g., aggregation, filtering), which boosts performance on tasks requiring compositionality.

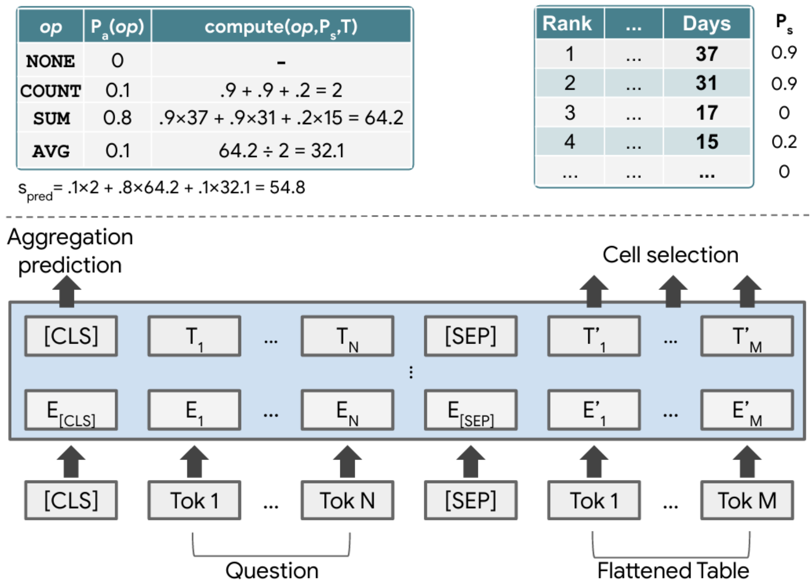

Fig. 3. TAPAS model (bottom) with example model outputs for the question: "Total number of days for the top two". Cell prediction (top right) is given for the selected column's table cells in bold (zero for others) along with aggregation prediction (top left).

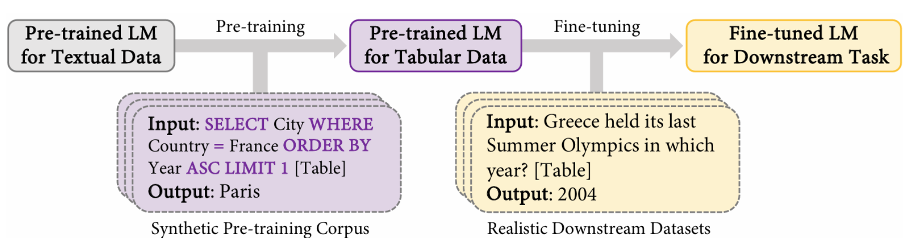

Fig. 4. The schematic overview of the paper's method. For the sake of brevity, the table content in the input is simplified with the symbol [Table].

3.4 Retrieval-Augmented and Reasoning-Enhanced Generation

Recent approaches augment generation models with retrieval and explicit reasoning capabilities, aiming to further improve factuality and handle knowledge-intensive or compositional tasks.

HD-RAG (Zhang and Chen [25]) introduces a hybrid retrieval-augmented framework where the model retrieves relevant textual passages and hierarchical table segments before and during generation (see Figure 5). It employs a dual-encoder architecture: one for textual context and another for tabular context. The generator fuses these sources using cross-attention, grounding the output in both retrieved

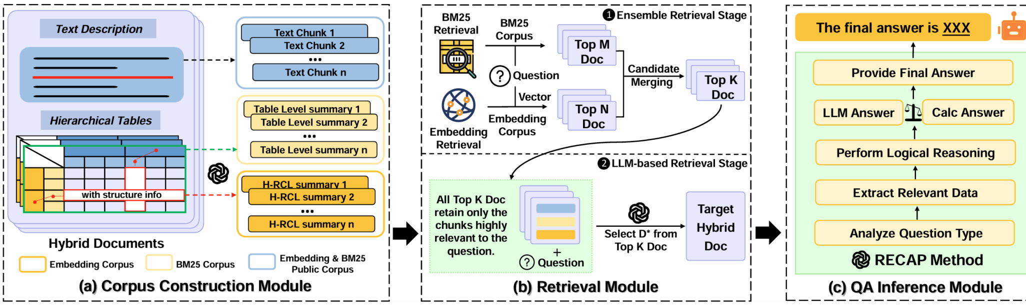

Fig. 5. The Overview of HD-RAG Framework.

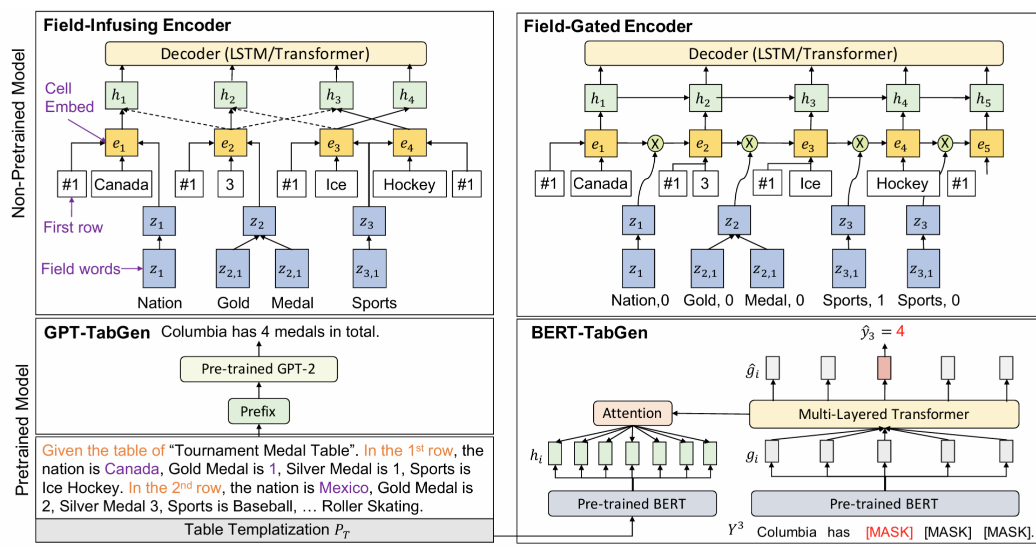

Fig. 6. The Non-pretrained and Pre-trained generation models.

evidence and the input table, which is especially effective for hybrid documents that interleave text and tables.

LogicNLG (Chen et al. [5]) focuses on generating text that involves logical operations (e.g., comparison, filtering, aggregation) over tables. The model extends standard seq2seq with tagging and reasoning modules that predict which logical operations to apply to which table entries (see Figure 6). This approach enables the generation of statements that require multi-step inference, pushing the boundary of compositional text generation.

The innovation points of these papers can be summarized as: (1) Dual-encoder architectures for multi-source retrieval (text + table). (2) Cross-attention fusion for evidence integration. (3) Explicit logic tagging and operation prediction for

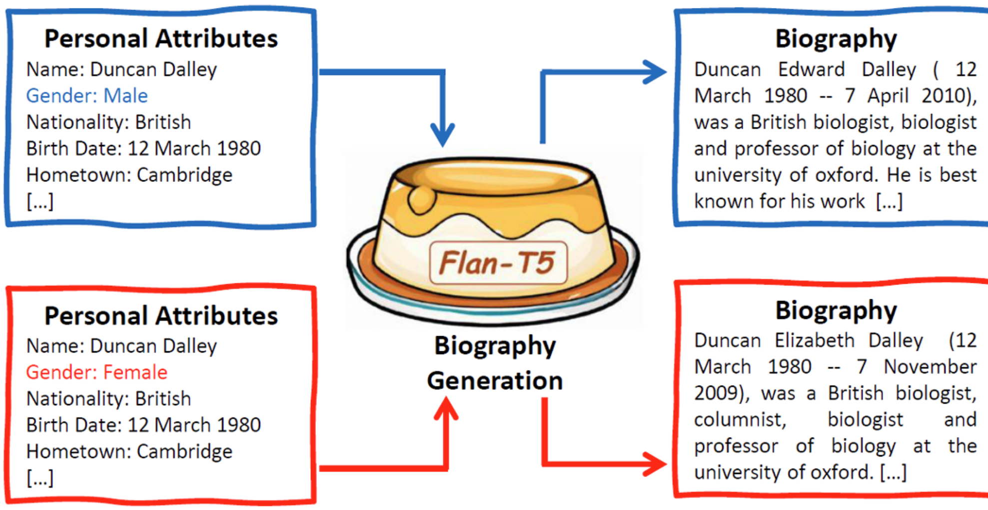

Fig. 7. An example from the Synthbio dataset. We measure semantic matching and sentiment in the true and generated biography (top-right) based on the personal attributes (top-left). Counterfactuals (bottom-right) replace the personal attribute (male, top-left) with a different one (female, bottom-left).

reasoning tasks. (3) Handling of hybrid documents and complex, multi-hop table reasoning.

3.5 Multilingual and Bias-aware Table-to-Text Generation

As the field expands to multilingual and fairness-sensitive applications, new methods target language diversity and bias mitigation.

Haussmann [9] explores the challenge of generating faithful, attributable text from tables in low-resource languages. The approach involves cross-lingual pretraining and alignment methods, leveraging parallel datasets and multilingual PLMs to enable robust generation even for languages with limited labeled data. The evaluation focuses on both fluency and factual alignment across languages.

Fang et al. [6] conduct a counterfactual study on bias in biography generation (see Figure 7). Their methodology involves generating paired biographies with counterfactual attribute swaps (e.g., gender, ethnicity) and measuring output differences. The study reveals persistent biases in Table2Text models and motivates the integration of debiasing modules or data balancing techniques.

The innovation points of these papers can be summarized as: (1) Cross-lingual transfer and multilingual pretraining for low-resource settings. (2) Counterfactual evaluation and debiasing techniques for fairness. (3) Focus on factual alignment and representational equity across languages and groups.

Collectively, these methodologies illustrate the field's ongoing efforts to balance fluency, factuality, and reasoning. While neural and pretrained models have improved naturalness and coverage, explicit content planning, retrieval augmentation, and logical reasoning remain crucial for generating faithful and informative text from complex tables. The integration of multilingual, bias-aware, and evaluation-sensitive approaches represents the next frontier for robust and equitable Table2Text generation.

4 Conclusion

This survey has presented a comprehensive overview of the Table2Text field, which focuses on generating coherent and contextually relevant natural language descriptions from structured table data. We have systematically reviewed the progression of Table2Text techniques, beginning with early rule-based and template methods, through to neural approaches leveraging sequence-to-sequence architectures, attention mechanisms, and pre-trained language models. The survey also highlighted the evolution of datasets and evaluation metrics that have been crucial in benchmarking progress in this area.

A major thrust of recent research centers on neural models, particularly those employing encoder-decoder frameworks with various enhancements to better capture table structure, handle rare entities, and ensure factual consistency. The emergence of large-scale pre-trained models and the integration of external knowledge have further advanced the field, yielding significant improvements in fluency and informativeness.

5 Limitations and Future Work

6 Limitations

Despite the rapid progress, several key limitations remain in the Table2Text field, as underscored by the papers reviewed in this survey:

Structural Representation: Many models struggle to fully capture the complex and hierarchical structures present in real-world tables, especially those with multiple headers, nested cells, or irregular formats. Simple linearization may lose critical relational information.

Factual Accuracy and Faithfulness: Neural models, particularly those based on large language models, are prone to hallucinating information or generating text that does not faithfully reflect the input table. Ensuring strict factual alignment remains a major challenge.

Generalization and Robustness: Current models often overfit to training datasets and may not generalize well to out-of-domain tables, unseen schemas, or tables with varying linguistic content.

Interpretability and Controllability: The black-box nature of deep neural models makes it difficult to interpret or control generated outputs, limiting their applicability in high-stakes or user-facing scenarios.

Evaluation Metrics: Existing automatic metrics (e.g., BLEU, ROUGE, METEOR) are limited in their ability to assess factual correctness, coverage, and coherence. Human evaluation, while more informative, is labor-intensive and costly.

Data Scarcity and Diversity: While several datasets exist, there remains a lack of large, diverse, and high-quality datasets that cover a wide range of domains, languages, and table types. This limits the ability of models to generalize and adapt.

6.1 Future Directions

Looking ahead, several promising directions emerge for advancing the Table2Text field:

Improved Structural Encoding: Developing models that can more effectively capture and leverage complex table structures, including hierarchical, multi-dimensional, and nested tables, remains a key area for future work.

Faithful and Controllable Generation: Research into models that guarantee faithfulness, possibly via constrained decoding, explicit content selection, or verifiability mechanisms, will be crucial. Incorporating user controls or prompts to guide generation could further enhance usability.

Cross-Domain and Multilingual Table2Text: Building models that can generalize across domains and support multilingual generation can broaden the applicability of Table2Text systems.

Unified Pre-training and Transfer Learning: Leveraging large-scale, table-augmented pre-training or multi-task learning may help models transfer knowledge across tasks and domains, improving both data efficiency and robustness.

Better Evaluation Paradigms: Developing new automatic metrics that better capture factual accuracy, coverage, and logical consistency is essential. Leveraging advances in natural language inference and fact verification may provide more reliable evaluation.

Human-in-the-Loop and Interactive Systems: Integrating human feedback during training or deployment can improve model performance, interpretability, and trustworthiness, especially in critical applications.

In summary, while significant progress has been achieved in Table2Text generation, addressing the above limitations and exploring these future directions will be vital for developing robust, accurate, and widely applicable Table2Text systems.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] Gabor Angeli, Percy Liang, and Dan Klein. 2010. A simple domain-independent probabilistic approach to generation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. 502–512.

[3] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[5] Wenhu Chen, Jianshu Chen, Yu Su, Zhiyu Chen, and William Yang Wang. 2020. Logical natural language generation from open-domain tables. *arXiv preprint arXiv:2004.10404* (2020).

[6] Biaoyan Fang, Ritvik Dinesh, Xiang Dai, and Sarvnaz Karimi. 2024. Born Differently Makes a Difference: Counterfactual Study of Bias in Biography Generation from a Data-to-Text Perspective. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 409–424.

[7] Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. The WebNLG challenge: Generating text from RDF data. In *10th International Conference on Natural Language Generation*. ACL Anthology, 124–133.

[8] Jiatao Gu, Qi Liu, and Kyunghyun Cho. 2019. Insertion-based decoding with automatically inferred generation order. *Transactions of the Association for Computational Linguistics* 7 (2019), 661–676.

[9] Aden Haussmann. 2025. The Challenge of Achieving Attributability in Multilingual Table-to-Text Generation with Question-Answer Blueprints. *arXiv preprint arXiv:2503.23204* (2025).

[10] Jonathan Herzig, Paweł Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Martin Eisenschlos. 2020. TaPas: Weakly supervised table parsing via pre-training. *arXiv preprint arXiv:2004.02349* (2020).

[11] Rémi Lebret, David Grangier, and Michael Auli. 2016. Neural text generation from structured data with application to the biography domain. *arXiv preprint arXiv:1603.07771* (2016).

[12] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461* (2019).

[13] Qian Liu, Bei Chen, Jiaqi Guo, Morteza Ziyadi, Zeqi Lin, Weizhu Chen, and Jian-Guang Lou. 2021. TAPEX: Table pre-training via learning a neural SQL executor. *arXiv preprint arXiv:2107.07653* (2021).

[14] Tianyu Liu, Kexiang Wang, Lei Sha, Baobao Chang, and Zhifang Sui. 2018. Table-to-text generation by structure-aware seq2seq learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[15] Hongyuan Mei, Mohit Bansal, and Matthew R Walter. 2015. What to talk about and how? selective generation using lstms with coarse-to-fine alignment. *arXiv preprint arXiv:1509.00838* (2015).

[16] Ankur P Parikh, Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. 2020. ToTTo: A controlled table-to-text generation dataset. *arXiv*

preprint arXiv:2004.14373 (2020).

[17] Ratish Puduppully, Li Dong, and Mirella Lapata. 2019. Data-to-text generation with content selection and planning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 6908–6915.

[18] Qi Qi, Yajun Du, Jia Liu, Xianyong Li, Xiaoliang Chen, and Yan-Li Lee. [n. d.]. Dual-Transformer Table-to-Text Generation Model with Timeline Integration. *Available at SSRN 5216329* ([n. d.]).

[19] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research* 21, 140 (2020), 1–67.

[20] Ehud Reiter and Robert Dale. 1997. Building applied natural language generation systems. *Natural Language Engineering* 3, 1 (1997), 57–87.

[21] Lei Sha, Lili Mou, Tianyu Liu, Pascal Poupart, Sujian Li, Baobao Chang, and Zhifang Sui. 2018. Order-planning neural text generation from structured data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[22] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems* 27 (2014).

[23] Sam Wiseman, Stuart M Shieber, and Alexander M Rush. 2017. Challenges in data-to-document generation. *arXiv preprint arXiv:1707.08052* (2017).

[24] Sam Wiseman, Stuart M Shieber, and Alexander M Rush. 2018. Learning neural templates for text generation. *arXiv preprint arXiv:1808.10122* (2018).

[25] Chi Zhang and Qiyang Chen. 2025. HD-RAG: Retrieval-Augmented Generation for Hybrid Documents Containing Text and Hierarchical Tables. *arXiv preprint arXiv:2504.09554* (2025).