

Commentary Generation from Data Records of Multiplayer Strategy Esports Game

Zihan Wang

The University of Tokyo
zwang@tkl.iis.u-tokyo.ac.jp

Naoki Yoshinaga

Institute of Industrial Science,
The University of Tokyo
ynaga@iis.u-tokyo.ac.jp

Abstract

Esports, a sports competition on video games, has become one of the most important sporting events. Although esports play logs have been accumulated, only a small portion of them accompany text commentaries for the audience to retrieve and understand the plays. In this study, we therefore introduce the task of generating game commentaries from esports' data records. We first build large-scale esports data-to-text datasets that pair structured data and commentaries from a popular esports game, League of Legends. We then evaluate Transformer-based models to generate game commentaries from structured data records, while examining the impact of the pre-trained language models. Evaluation results on our dataset revealed the challenges of this novel task. We will release our dataset to boost potential research in the data-to-text generation community.¹

1 Introduction

Esports (Hamari and Sjöblom, 2017; Reitman et al., 2020), a sports competition using video games, has become popular and gained a larger audience than ever. However, the individual gameplays accompany a few metadata such as player names, which prevents their audience from finding games with strategies of interest and understanding the intention of skillful actions. Although textual game commentaries will help the audience retrieve games by a natural language query and better understand the player's actions (Figure 1) (Lavelle, 2010), it is costly for human experts to provide individual games with play-by-play commentaries. As a result, only a small fraction of esports games with play logs accompany textual commentaries.

To enhance the audience's experience in watching such esports games, the technology of data-to-text generation can be used to generate game

Screenshot of "WARD_PLACED" event (for explanation):



Input (one-minute structured data; an excerpt):

```
{
  ...
  { "type": "WARD_PLACED",
    "timestamp": 905433,
    "wardType": "YELLOWTRINKET",
    "creatorId": 6
  },
  ...
}
```

Output (play-by-play commentary; an excerpt):

... even in a map state g2 can get exclusive vision on an area then suddenly the Nautilus veigar will have a lot of zone control but so behind in map control ...

Figure 1: Game commentary based on one-minute data records (a series of events in JSON format). The screenshot is to help interpret the input (not a part of the input).

commentaries from structured data records. In the literature, data-to-text generation has been applied in summary generation from box- and line-scores of basketball games (Wiseman et al., 2017) and play-by-play commentary generation for board games (Modgil et al., 2013; Kameko et al., 2015). Compared to the basketball data, esports data contains detailed game records and play-by-play commentaries. Compared to board game data, esports data is usually not turn-based. In conclusion, there is a lack of research considering the characteristics of esports data in existing studies.

In this study, we introduce the task of generating game commentaries from structured data records

¹<https://github.com/ArnoZWang/esports-data-to-text>

(Figure 1) for one of the most popular esports games, League of Legends (LoL). Broadly, the overall workflow for addressing this new task includes three main processes: we first build a large-scale data-to-text generation dataset consisting of commentaries obtained from subtitles of YouTube contest videos on esports games and corresponding structured data records obtained using LoL official APIs; we then use a Transformer encoder-decoder model (Vaswani et al., 2017) and its pre-trained variants to tackle the task; we also set evaluation metrics of esports data-to-commentary generation.

We evaluate the performance of esports data-to-commentary generation on our proposed datasets using the metrics regarding the characteristics of esports data and discuss the main challenges of this task to be addressed in the future.

The contributions of this paper are as follows:

- We set up a task of data-to-commentary generation for one of the most popular esports, League of Legends; we have built large-scale datasets to facilitate the research on this task.
- We designed evaluation criteria for esports data-to-commentary generation, which reflects the purposes of the game commentaries.
- We evaluated several strong baselines including Llama2 (Touvron et al., 2023) and discussed the challenges of the task through examples of generated commentaries.

2 Related Work

In this section, we review data-to-text generation tasks on sports games, board games, and video games to highlight the characteristics of our task.

Summary generation for sports game records

Many studies have focused on generating textual summaries of physical sports games from their score records (Wiseman et al., 2017; van der Lee et al., 2017; Dou et al., 2018; van der Lee et al., 2018; Taniguchi et al., 2019; Puduppully et al., 2019a; Rebuffel et al., 2020), mainly including basketball and soccer games. The essential difference between these tasks and our task is that their data only records certain important values (score, player number, etc.), while esports data provides details of the games. As a result, the average length of commentaries per game is much longer for our LoL dataset than those for the sports game records, which challenges a generation model to understand individual actions in the games.

Play-by-play commentary generation for board games

Some studies tackled the task of generating play-by-play commentaries with grounded move expressions from chess and shogi (Japanese chess) (Modgil et al., 2013; Kameko et al., 2015; Jhamtani et al., 2018). For both esports and chess, we can reproduce the whole game from the data records. Nevertheless, in board games, two players alternately perform one move in turn, whereas in esports games, multiple players can simultaneously perform actions in real-time, which challenges a model to interpret simultaneous actions.

Commentary generation from esports game videos

Several studies aimed to produce textual summaries and commentaries from a game video (Khan and Pawar, 2015; Pasunuru and Bansal, 2018; Tanaka and Simo-Serra, 2021; Zhang et al., 2024). In particular, regarding racing games, Ishigaki et al. (2021) generated live commentary from game video data with structured telemetry data, mostly on numerical values of the player’s car and game progress. Although this study utilizes game data records, they provide only partial information on the gameplays and are meant to supplement visual data. Meanwhile, our task focuses on League of Legends (Tanaka and Simo-Serra, 2021), one of the most popular multiplayer strategic esports games (Zhang et al., 2022), and generates commentaries from comprehensive structured data records with detailed descriptions of games with more strategic content.

3 Esports Data-to-text Datasets

We have constructed and will release large-scale data-to-text datasets for one of the most popular esports games, League of Legends (LoL), which is also a demonstration sports event in the 2018 and 2022 Asian Games for its popularity (Hallmann and Giel, 2018; Jenny et al., 2017). The large-scale dataset is the core building block to assess the feasibility of data-to-text technology on the task. Therefore, in this study, we build two datasets, LoL19 and LoL19-21, from all games in the highest-level tournaments of the 2019 and 2019 to 2021 Season World Championship of League of Legends, respectively.

In this section, we first introduce the basics of the target game, LoL, and then explain methods of collecting data records and textual commentaries of LoL. We also explain several data preprocessing methods to improve the collected datasets.

Event type	Proportion		# keys	Explanation
	LoL19 (core)	LoL19-21 (extended)		
ITEM_PURCHASED	25.1%	24.9%	3	The player purchases an item
ITEM_SOLD	1.8%	1.8%	3	The player sells an item
ITEM_UNDO	0.6%	0.6%	4	The player cancels the purchase of an item
ITEM_DESTROYED	20.2%	20.4%	3	The player destroys an item
BUILDING_KILL	1.3%	1.3%	8	The team destroys an enemy building
CHAMPION_KILL	2.4%	2.5%	5	The player defeats an enemy champion
SKILL_LEVEL_UP	13.8%	13.4%	4	The player upgrades a skill
WARD_PLACED	24.5%	24.7%	4	The player places a ward
WARD_KILL	9.8%	9.9%	3	The player destroys an enemy ward
ELITE_MONSTER_KILL	0.6%	0.6%	4 or 5	The team defeats an elite monster

Table 1: Statistics of all ten types of game events in our LoL datasets.

3.1 Basics of League of Legends

In this study, we choose LoL as our research target because of its popularity and representative role as a multiplayer strategy game for esports. In LoL games, each player controls one game character called “champion” with unique abilities that will improve during the game progress and contribute to the team’s overall strategy (Cannizzo and Ramírez, 2015). Two teams compete in one game map, each of which team consists of five players. The goal is to destroy the opponent’s base while protecting their own. As the game progresses, the champions can beat the enemy champions, defeat non-player units called “monsters,” and destroy buildings to earn resources. They then use the resources to improve their abilities by purchasing items and upgrading skills.

Compared to the physical sports games and board games, LoL is more complicated on real-time game actions, complexity of rules, and data record size per game. These factors are the main obstacles to data-to-text generation.

3.2 Data Extraction and Preprocessing

To build the core dataset named LoL19, we target the games of the highest-level tournament in the 2019 Season World Championship of League of Legends. We collect the structured data records of the gameplays from the LoL official API site² as input and extract subtitles of YouTube videos on the gameplays as output commentaries. This data is strictly paired with the game IDs provided by the game’s Match History site.³ Later, we will introduce the method to process the collected data and build large-scale data-to-text dataset.

²<https://developer.riotgames.com/>

³https://lol.fandom.com/wiki/2019_Season_World_Championship/Match_History

Retrieving Data Records on Gameplays

In LoL games, every move made by each player is recorded, and the records are available at the LoL official API site.² From the LoL data, we can strictly restore the entire game from the structured data records, which is impossible in sports games like basketball and soccer. However, the complete data has redundant information, and the large data volume is a heavy burden for storage and subsequent processing.

Therefore, we choose another data type provided by the official API, “event-based data frame.” In this data, individual gameplays are recorded based on associated events. Each event is defined as an update of certain game status and includes the key named “type,” which denotes a type of event. Different types of events have various sets of keys; Table 1 lists all event types with their proportions in our LoL dataset. For example, the “WARD_PLACED” events involve information about “wardType,” while the “ITEM_PURCHASED” events do not have this key. The event-based data frames are stored in JSON format, as shown in Figure 1.

Retrieving Textual Commentaries

We collect YouTube subtitles of the LoL contest videos, which are linked from every contest game in the Match History site, as the output of this task. The subtitles are split into sentences (precisely, utterances) using line breaks given by YouTube’s automatic speech recognition (ASR) as clues.

Then, we randomly selected 200 examples obtained with the following data formatting and manually confirmed their qualities. The resulting word error rate (WER) was 6.8, which is comparable to human performance on common ASR datasets,⁴ confirming the data is clean to use for evaluation.

⁴https://github.com/syhw/wer_are_we

	esports data2text		esports video2text	basketball	chess
	LoL19 (core)	LoL19-21 (extended)	LoL-V2T	RotoWire	GameKnot
Number of games (matches)	220	650	157	4,853	11,578
Number of examples	3,490	10,590	9,723	4,853	298,008
Average number of events in input	49.13	48.58	-	-	-
Average number of tokens of input	540.47	541.10	-	628.00	25.73
Average number of tokens of output	374.68	373.89	15.4	337.10	20.55

Table 2: Statistics of our esports data-to-text datasets and common datasets for similar tasks.

Data Formatting

The average length of the commentary of one LoL game is over 10K words, which is much longer than the outputs of the existing data-to-text datasets. As a result, we cannot exploit the common Transformer architecture (Vaswani et al., 2017) for this task. Therefore, we decompose the obtained pairs of structured data and commentaries into pieces of shorter lengths. We first split the sequence of events by the unit duration of one minute of game-play. We then split the sequence of commentaries by matching their timings with the duration of each subsequence of events.

Next, we address the format difference between the input data (nested list in JSON format) and natural language text to make common encoder-decoder models (Sutskever et al., 2014; Vinyals et al., 2016) applicable. Specifically, we linearize the structured input. For each key-value pair in the top-level list of each event in the JSON format, we recursively concatenate the value and the key with a delimiter “|” while inserting a space between individual key-value pairs. Following this procedure, “WARD_PLACED” event in the JSON format (Figure 1) is linearized into the following sequence:

```
WARD_PLACED|type 905433|timestamp
YELLOWTRINKET|wardType
6|creatorId
```

Table 2 lists the statistics of the resulting dataset, LoL19 and other data-to-text datasets (§ 2) such as LoL-V2T (Tanaka and Simo-Serra, 2021), RotoWire (Wiseman et al., 2017), and chess (Jhamtani et al., 2018) for comparison. We also collected the data from all games in the 2020 and 2021 Season World Championship of League of Legends to extend the LoL19 dataset (LoL19-21). Our datasets have a comparable number of examples to LoL-V2T and RotoWire and have a comparable number of input and output tokens to RotoWire. To obtain our LoL datasets, we will release the scripts for collecting and processing the data.

4 Esports Data-to-text Generation

In this section, we first perform experiments on esports data-to-text generation using the LoL19 dataset and several Transformer (Vaswani et al., 2017)-based models. Then, we analyze the system outputs to reveal the challenges of this task.

4.1 Settings

Datasets We use the core dataset for evaluation. We first split the games into train, validation, and test sets with a ratio of 8:1:1, according to the chronological order of the individual games, resulting in 2790:350:350 examples. Since the core dataset exclusively comprises data from the 2019 Season World Championship, the terminology used within the games, such as player names, is guaranteed to be consistent across the datasets.

Models We compared models based on Transformer (Vaswani et al., 2017), T5 (Raffel et al., 2020), and variations of Llama2 in the experiments. **Transformer** is an encoder-decoder model, implemented by OpenNMT⁵ library. **T5**⁶ is a pre-trained generative model on text-to-text tasks. **Llama2** is a pre-trained large language model ranging in scale from 7B to 70B parameter (Touvron et al., 2023).

Training For **Transformer** and **T5**, we set decoder dropout of 0.5, training steps of 10,000, and learning rate of 0.001; the other hyperparameters follow their default settings. For **Llama2-7B** and **-13B**, we finetune them using QLoRA (Dettmers et al., 2024) and 4-bit precision. We set LoRA dropout of 0.1, training steps and learning rate the same with Transformer, and the other hyperparameters follow Huggingface Llama2⁷ document. For **Llama2-70B**, we apply in-context learning (ICL) (Floridi and Chiriatti, 2020) without updating model weights. The ICL prompt is as follows:

⁵<https://github.com/OpenNMT/OpenNMT-py>

⁶<https://huggingface.co/t5-base>

⁷https://huggingface.co/docs/transformers/model_doc/llama2

Strategic depth	score
Based on the criteria for obtaining a score of 4, the strategic considerations are inspiring, providing insights to help learn from the skillful players and teams	5
Based on the criteria for obtaining a score of 3, the strategic considerations are sufficient and closely related to the game moment described by the structured data	4
Based on explaining the facts, the commentary also reflects several strategic considerations, such as the player’s intention and the team’s arrangement	3
The commentary only reflects the core event of the game moment described by the structured data, without providing any strategic consideration	2
The commentary reflects no facts or only a few facts of the game moment described by the structured data	1

Table 3: Scoring criteria of the strategic depth evaluation.

Models	sacreBLEU ↑	Text distance ↓	ROUGE-L ↑	BERTScore ↑	BARTScore ↑	Strategic depth ↑
Gold	100	0	100	100	0	3.164
Transformer	1.4	70.22	13.62	79.06	-5.27	2.312
T5 (Raffel et al., 2020)	3.5	71.46	13.55	81.67	-5.36	2.790
Llama2-7B (Touvron et al., 2023)	5.1	69.01	14.98	83.16	-5.02	2.994
<i>w/o finetune</i>	0.1	98.84	0.65	63.64	-5.90	2.076
Llama2-13B	11.0	63.49	16.94	86.10	-4.61	3.064
<i>w/o finetune</i>	0.2	90.03	1.00	66.56	-5.86	2.170
Llama2-70B (ICL)	6.2	68.82	11.93	83.54	-4.77	2.916

Table 4: Experimental results on the LoL19 esports data-to-text generation dataset.

You are an expert of League of Legends esports games. Please read the input data records and describe them in natural language commentary as output. Input: [insert input data here] Output:

4.2 Evaluation Metrics

Following the existing data-to-text tasks on sports game summary (Puduppully et al., 2019b; Rebuffel et al., 2020; Tang et al., 2023), board game commentary (Jhamtani et al., 2018), and racing game commentary (Ishigaki et al., 2021), we adopt **sacreBLEU** (Papineni et al., 2002; Post, 2018), **text distance** (normalized Damerau-Levenshtein⁸) (Brill and Moore, 2000), and **ROUGE-L** (Lin, 2004),⁹ along with **BERTScore** (Zhang et al., 2020)¹⁰ and **BARTScore** (Yuan et al., 2021)¹¹ for evaluation. These automatic metrics reflect the quality of generated results over correctness and fluency.

Considering the characteristics of multiplayer strategy esports games, assessing the strategic depth of game commentaries is important. The **strategic depth** is thus designed to measure the extent to which the system output provides useful

information on the players’ actions. Although the automatic metrics above can arguably measure the general qualities of the system output, we also want the output to contain strategically relevant commentaries, such as reflecting the players’ intentions and the team’s arrangement regarding the combat.

Because it is difficult to estimate the strategic depth, we gather human scores using criteria tailored for esports commentaries, as detailed in Table 3. These scores are collected from five graduate students as human annotators, who understand game rules, the content of contest games, and game commentaries of LoL. We calculate the average number of their scores as the results.

4.3 Results

Table 4 lists the results on our LoL19 dataset. Compared to the standard Transformer, T5 improves sacreBLEU and strategic depth scores. Llama2-13B exhibits the best overall performance, which confirms the enhancement resulting from pre-trained models with larger sizes. We also compare the performance of Llama2-7B and -13B without finetune as an ablation study. For Llama2-70B, although in-context learning does not result in high text similarity, the generated results exhibit the potential of producing inspiring content.

⁸<https://github.com/life4/textdistance>

⁹<https://github.com/pltrdy/rouge>

¹⁰<https://pypi.org/project/bert-score>

¹¹<https://github.com/neulab/BARTScore>

Screenshot of “ITEM_PURCHASED” event (for explanation):



Input: ... ITEMPURCHASED|type 783445|timestamp 2|participantId 1033|itemId...

Reference: ... swing Tristana in a very good position I was talking about it yesterday like the build is very forgiving ...

Transformer: ... I think the rift we are and I think it was slightly changing a little bit more ...

T5: ... that’s why you have a look at the fact that 10 minutes into the game ...

Llama2-7B: ... I think that is something that would have been a little bit more active on the map ...

Llama2-13B: ... they have a lot of pressure in the mid lane which means that they can walk towards the bottom lane of the map if they want to contest ...

Llama2-70B (ICL): ... there’s some interesting activity ... with items being destroyed and purchased ...

Figure 2: Excerpts of system outputs of an ITEM_PURCHASED event. The screenshot is to help interpret the input (not a part of the input).

Figures 2 and 3 show examples of system outputs, which confirm the difficulty of associating past events with ongoing events. In the first example, the reference output revisits what happened in the past of this game because the ongoing event is not informative enough, while the system outputs mainly focus on only the current moment. In the second example, although the rest of the players are gathered in another area of the map, the lone player in the top lane was taking an enemy building down a little while ago, which directly affected the direction of the whole game. The commentary is expected to be more helpful by reflecting on this moment from the past rather than solely concentrating on the current game moment. In these cases, using finer segments of inputs may enable a more accurate generation. However, it also leads to the loss of context; the system output thus fails to generate content related to the game’s history. It is also challenging to maintain the balance between the size of inputs and the amount of context.

Meanwhile, the current focus of this task relies on the modalities of structured data records and tex-

Screenshot of “BUILDING_KILL” event (for explanation):



Input: ... BUILDINGKILL|type 1315932|timestamp x981y10441|position 8|killerId...

Reference: ... the attacks bit like they can’t go anywhere it’s always just just like right click the head ...

Transformer: ... well you can walk away from the bottom Lane you’re welcome back ...

T5: ... if you look at the mini-map ... you have to look at the mini-map ...

Llama2-7B: ... you can see the Kabat does want to be able to get to the side lane ...

Llama2-13B: ... at this point in the game you can see that top lane is going to be collected from the side ...

Llama2-70B (ICL): ... we do know that the building was a top-lane turret ... and was destroyed by a player ...

Figure 3: Excerpts of system outputs of an BUILDING_KILL event. The screenshot is to help interpret the input (not a part of the input).

tual game commentaries, while visual inputs like screenshots and video clips can provide a contribution to generation. There is a potential to integrate visual inputs for cross-modal generation.

5 Conclusions

This study set up the task of generating game commentaries from structured data of the multiplayer strategy game, League of Legends. We built and will release the first large-scale data-to-text generation dataset on strategic esports games. Next, we also discussed evaluation metrics for our task to measure the quality and strategic depth of the system outputs. Then, we explored the performance of Transformer and its pre-trained variants for this task, revealing the challenges of the task such as associating past events with ongoing events.

We will address the remaining issues in the future. The unique challenges include i) linking game history like past events related to current events, and ii) integrating visual inputs, including screenshots and video clips, to this task to perform cross-modal understanding and generation.

Limitations

This work mainly focuses on applying data-to-text generation technology in the esports area. Although this paper introduces a novel dataset collected from a representative esports game League of Legends, it lacks the consideration of other esports contests and game genres in the current stage. We plan to continue testing the feasibility of our proposed methods on other esports game data.

Ethics Statement

In the data collection process, we have strictly followed the policies of RiotGames API and YouTube. The former is the publisher of LoL game records. The later provides subtitles of LoL contest videos, which we used as game commentaries in our work. Further ethical concerns related to the game content (e.g., video game content rating) can refer to the ESRB Rating (<https://www.esrb.org/ratings/32211/league-of-legends/>); the LoL is rated as “Teen,” which confirms the game content is suitable ages 13 and up.

Acknowledgements

This work was partially supported by JST, CREST Grant Number JPMJCR19A, Japan and JSPS KAKENHI Grant Number JP21H03494.

References

- Eric Brill and Robert C. Moore. 2000. [An improved error model for noisy channel spelling correction](#). In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 286–293, Hong Kong. Association for Computational Linguistics.
- Alejandro Cannizzo and Esmitt Ramírez. 2015. [Towards procedural map and character generation for the MOBA game genre](#). *Ingeniería y Ciencia*, 11(22):95–119.
- Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2024. [QLoRA: Efficient finetuning of quantized LLMs](#). *Advances in Neural Information Processing Systems*, 36.
- Longxu Dou, Guanghui Qin, Jinpeng Wang, Jin-Ge Yao, and Chin-Yew Lin. 2018. [Data2Text studio: Automated text generation from structured data](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 13–18, Brussels, Belgium. Association for Computational Linguistics.
- Luciano Floridi and Massimo Chiriatti. 2020. [GPT-3: Its nature, scope, limits, and consequences](#). *Minds and Machines*, 30:681–694.
- Kirstin Hallmann and Thomas Giel. 2018. [eSports – Competitive sports or recreational activity?](#) *Sport management review*, 21(1):14–20.
- Juho Hamari and Max Sjöblom. 2017. [What is esports and why do people watch it?](#) *Internet research*, 27(2).
- Tatsuya Ishigaki, Goran Topic, Yumi Hamazono, Hiroshi Noji, Ichiro Kobayashi, Yusuke Miyao, and Hiroya Takamura. 2021. [Generating racing game commentary from vision, language, and structured data](#). In *Proceedings of the 14th International Conference on Natural Language Generation*, pages 103–113, Aberdeen, Scotland, UK. Association for Computational Linguistics.
- Seth E Jenny, R Douglas Manning, Margaret C Keiper, and Tracy W Olrich. 2017. [Virtual\(ly\) athletes: where esports fit within the definition of “sport”](#). *Quest*, 69(1):1–18.
- Harsh Jhamtani, Varun Gangal, Eduard Hovy, Graham Neubig, and Taylor Berg-Kirkpatrick. 2018. [Learning to generate move-by-move commentary for chess games from large-scale social forum data](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1661–1671, Melbourne, Australia. Association for Computational Linguistics.
- Hirotaaka Kameko, Shinsuke Mori, and Yoshimasa Tsuruoka. 2015. [Learning a game commentary generator with grounded move expressions](#). In *2015 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 177–184. IEEE.
- Yasmin S Khan and Soudamini Pawar. 2015. [Video summarization: survey on event detection and summarization in soccer videos](#). *International Journal of Advanced Computer Science and Applications*, 6(11).
- Katherine L Lavelle. 2010. [A critical discourse analysis of black masculinity in NBA game commentary](#). *The Howard Journal of Communications*, 21(3):294–314.
- Chin-Yew Lin. 2004. [ROUGE: A package for automatic evaluation of summaries](#). In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Sanjay Modgil, Francesca Toni, Floris Bex, Ivan Bratko, Carlos I Chesnevar, Wolfgang Dvořák, Marcelo A Falappa, Xiuyi Fan, Sarah Alice Gaggl, Alejandro J García, et al. 2013. [The added value of argumentation](#). *Agreement technologies*, pages 357–403.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia,

- Pennsylvania, USA. Association for Computational Linguistics.
- Ramakanth Pasunuru and Mohit Bansal. 2018. [Game-based video-context dialogue](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 125–136, Brussels, Belgium. Association for Computational Linguistics.
- Matt Post. 2018. [A call for clarity in reporting BLEU scores](#). In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.
- Ratish Puduppully, Li Dong, and Mirella Lapata. 2019a. [Data-to-text generation with content selection and planning](#). In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6908–6915.
- Ratish Puduppully, Li Dong, and Mirella Lapata. 2019b. [Data-to-text generation with entity modeling](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2023–2035, Florence, Italy. Association for Computational Linguistics.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. [Exploring the limits of transfer learning with a unified text-to-text transformer](#). *J. Mach. Learn. Res.*, 21(140):1–67.
- Clément Rebuffel, Laure Soulier, Geoffrey Scuttheeten, and Patrick Gallinari. 2020. [A hierarchical model for data-to-text generation](#). In *European Conference on Information Retrieval*, pages 65–80. Springer.
- Jason G Reitman, Maria J Anderson-Coto, Minerva Wu, Je Seok Lee, and Constance Steinkuehler. 2020. [Esports research: A literature review](#). *Games and Culture*, 15(1):32–50.
- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. [Sequence to sequence learning with neural networks](#). In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.
- Tsunehiko Tanaka and Edgar Simo-Serra. 2021. [LoL-V2T: Large-scale esports video description dataset](#). In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 4557–4566.
- Xiangru Tang, Yiming Zong, Yilun Zhao, Arman Cohan, and Mark Gerstein. 2023. [Struc-Bench: Are large language models really good at generating complex structured data?](#) *arXiv preprint arXiv:2309.08963*.
- Yasufumi Taniguchi, Yukun Feng, Hiroya Takamura, and Manabu Okumura. 2019. [Generating live soccer-match commentary from play data](#). In *Proceedings of the thirty-third AAAI Conference on Artificial Intelligence*, pages 7096–7103.
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. [Llama 2: Open foundation and fine-tuned chat models](#). *arXiv preprint arXiv:2307.09288*.
- Chris van der Lee, Emiel Krahmer, and Sander Wubben. 2017. [PASS: A Dutch data-to-text system for soccer, targeted towards specific audiences](#). In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 95–104, Santiago de Compostela, Spain. Association for Computational Linguistics.
- Chris van der Lee, Bart Verduijn, Emiel Krahmer, and Sander Wubben. 2018. [Evaluating the text quality, human likeness and tailoring component of PASS: A Dutch data-to-text system for soccer](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 962–972, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. 2016. [Order matters: Sequence to sequence for sets](#). In *The fourth International Conference on Learning Representations*, San Juan, Puerto Rico.
- 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.
- Weizhe Yuan, Graham Neubig, and Pengfei Liu. 2021. [Bartscore: Evaluating generated text as text generation](#). In *Advances in Neural Information Processing Systems*, volume 34, pages 27263–27277. Curran Associates, Inc.
- Dawei Zhang, Sixing Wu, Yao Guo, and Xiangqun Chen. 2022. [MOBA-E2C: Generating MOBA game commentaries via capturing highlight events from the meta-data](#). In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 4545–4556, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. [BERTScore: Evaluating Text Generation with BERT](#). In *The eighth International Conference on Learning Representations*.

Zhihao Zhang, Feiqi Cao, Yingbin Mo, Yiran Zhang, Josiah Poon, and Caren Han. 2024. [Game-MUG: Multimodal oriented game situation understanding and commentary generation dataset](#). *arXiv preprint arXiv:2404.19175*.