

Live commentary in a football video game generated by an AI

MICHAŁ CZAPLICKI, University of Twente, The Netherlands

This paper is about an innovative approach to creating live football commentary for a video game using Artificial Intelligence (AI) technology. It is a possible solution to the increasing cost and time of computer game production nowadays. I propose using Generative Pre-trained Transformer-3.5 (GPT-3.5) to generate live commentary that dynamically updates with the game state. GPT-3.5 gets as input pre-written football commentary templates filled with game state variables, such as the score, time, player actions, etc. Next, it will be asked to rewrite them as a football commentary. I evaluated the system to assess the accuracy of the generated commentary and enjoyability for the user. My approach could be extended to other sports and video games, making it a valuable asset for the game development industry. The results show that such commentary is overall enjoyable for the players despite some limitations.

Additional Key Words and Phrases: Live commentary, AI, video game, football.

1 INTRODUCTION

Sports video games since the 90s have had the commentary to increase the immersion and engagement of the players. This commentary is normally done by pre-recorded audio. Game developers and writers create a script with a wide range of commentary lines and professional voice actors record these lines in a studio. The pre-recorded audio clips are then integrated into the game. This solution has some limitations, because of the way how it is implemented. The main ones are:

- Limited variation caused by a finite amount of pre-recorded content, which can lead to the reduced enjoyability of the players after spending too much time in the game
- Time constraints caused by the amount of developed and recorded commentary lines. Especially considering the need to cover a wide range of game situations and player actions in multiple languages. This may limit the depth and variety of available commentary in the game.

I suggest an alternative in the form of AI-generated commentary, which can get rid of all the above limitations and add more advantages to it. Firstly, unlike pre-recorded audio, AI-generated commentary can offer virtually limitless variation. The AI model can generate new and unique commentary lines based on the available data and its understanding of the game context. This variety helps prevent repetitive commentary, providing a fresh and dynamic experience for players. Moreover, AI algorithms can analyze real-time game data and generate commentary that adapts to specific in-game situations, player actions, and game outcomes. This dynamic nature of AI-generated commentary could allow for a more immersive and responsive experience, enhancing player engagement. Possibly in

the future, once the AI model is trained, generating new commentary lines could become a more cost-effective process compared to recording and producing pre-recorded audio. This could be particularly beneficial when it comes to localization, as AI-generated commentary can be easily adapted to different languages and regions.

For the purpose of testing my hypothesis, I used GPT natural language model. A natural language model is a type of artificial intelligence model that is designed to understand and generate human language. It uses statistical and probabilistic methods to analyze and generate text, enabling it to understand the structure, context, and meaning of written or spoken language. GPT (Generative Pre-trained Transformer) is a highly advanced language model that has shown remarkable capabilities in generating natural language text. The strength of GPT is its capacity to comprehend and produce writing that is human-like. It can understand sophisticated language patterns, grammatical rules, and context since it has already received pre-training on large amounts of diverse text material.

My goal for the research can be defined in two parts:

- **Goal 1:** Develop a program which is receiving data about the state of a game, populates a template with that information and uses the combination of both as input to the GPT-3.5 language model
- **Goal 2:** Assess the accuracy of the output commentary and enjoyability of the player

To achieve these goals, I will be using the following research questions (RQ) as the basis of my research:

- **RQ1: How can an artificial intelligence program be designed to effectively integrate live game data with the OpenAI GPT-3.5 language model to generate live commentary?**

Working with live game data in research presents many challenges that need to be addressed to ensure the effectiveness and adaptability of my program. The main challenge lies in the dynamic nature of the data, which constantly changes based on the course of the game. It is especially difficult, because of the quick response from the game, and the actions that relate to each other. For example, to correctly comment on some action, the program needs to know the context of the game, so needs to know previous actions. It is crucial to overcome this challenge because existing solutions in the game industry also operate on live data.

- **RQ2: To what extent is the GPT-3.5 generated live commentary accurate?**

I need to check the accuracy of the system to be sure that the generated commentary is describing what is the factual state of the game. For example, if the system output would be: "Player 1 scores with a header" but in the match, a goal was scored from a penalty kick, this would mean that the system is giving incorrect information. If it would happen too often, that would mean my solution is dysfunctional.

TS&IT 39, July 7, 2023, Enschede, The Netherlands

© 2023 University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

- **RQ3: To what extent is the GPT-3.5 generated live commentary enjoyable for the player?**

To assess the enjoyability of AI-generated commentary for players, it is crucial to conduct user studies and gather feedback directly from individuals who are already engaged in playing sports video games. This feedback will help determine whether my proposition is attractive to the target audience and if it surpasses or falls short of the current solutions.

My contribution is twofold. First, I developed a program, that generates live commentary for the football video game. Second, I evaluated the generated text in terms of accuracy and enjoyability for the player.

The structure of the paper is as follows. In Section 3 I will give an overview of the related works in the field of automatic generation of sports commentary; Section 4 will detail the methodology which will be used to answer the aforementioned research questions; Section 5 will show the results of the evaluation of accuracy and enjoyability; section 6 will discuss the results by showing limitations and analyzing future work; eventually, section 7 will conclude the paper. In section 8 I will provide a list of references used in this paper, and in section 9 I will put all of the appendixes.

2 RELATED WORK

Research on AI live commentary began at least as early as 1998. MIKE [1] is an automatic commentary system for soccer. It used analyzing modules to determine the current status of the game (e.g. looking at the average location of all of a team's players taken by whole) and which game event needs to be commented on. Later the Natural Language Generator uses this information to choose, in a set of comment templates, the most fitting one for the current state of the match. Eventually, the NLG system converts it into natural language, by filling it with information from the game. Because of the template-matching mechanism, the system output, during a single game, was very simple and repetitive. Moreover, it was created for Robot Soccer, so its template would not cover some of the actions that can happen during a human football match e.g. scoring goals with a header.

More recently, Lee et al. [2] attempted to create more engaging commentary by focusing on two prototypical roles of sports commentators: the play-by-play commentator, who is describing objective facts that are happening during a game, and the color commentator, who is more subjective and is storytelling about what is happening in the game. The authors created the system SCoReS, which depending on the state of a game, automatically suggests a story for a color commentator. The program is ranking the "stories" in a database and chooses the most fitting to the situation, by using a Machine Learning Ranker, and later a Machine Learning Evaluator checks how "good" the top-ranked story is. However, the system is limited by training data about stories, so after some time they become repetitive, or there is no relevant story to the current state of a game.

Taniguchi et al. [3] generate live soccer-match commentary from play data. The system is receiving multiple data, about a single event that is happening during a game (player name, play category, time, coordinates of the ball, and other details e.g. keeper touched a ball,

whether there was assistance, and so on). Later it uses an encoder-decoder model to generate live commentary. As a decoder, they used a recursive neural network language model. The limitation of this system is that despite getting multiple detailed data about a game, the system was making mistakes while describing events (e.g., failing to correctly distinguish player names or mentioning the same information twice).

3 METHODOLOGIES

This section detail the steps I took to answer each of the research questions. First, I obtained a football simulation environment on which my generated commentary is based and improve it to receive a state of a game from every frame. Next, it fulfilled the input template with game data and provide it to the GPT-3.5 language model, and ask it to rewrite the template as a football commentary. Eventually, I evaluated this implementation in accuracy and enjoyability for the player.

3.1 On answering RQ1

- **Google Football Environment**

For this project, I use a football simulation environment provided by the Google AI team [4]. It was initially intended as a platform for developing Reinforcement Learning, but I used it as a football video game for generating live commentary. I used this environment because it is open-source so I could freely modify and adjust to my needs. In this program, I was able to play a full football match and receive all information about the state of the game at each timestep, by using the command 'observation()'. This command was returning all data about: the ball, home and away teams, and the status of the game (Appendix 1). The most valuable information for me was game-mode, so the event that actually occurs in the game (kickoff, goal kick, free kick/offside, corner, throw-in, penalty).

- **Generation of commentary by GPT**

Next, I incorporated the information in templates, obtaining text that I used to prompt GPT-3.5 [5] with what is happening on the field. GPT is a state-of-the-art natural language processing model developed by OpenAI. It is trained on a dataset of millions of web pages, books, and other text sources to predict the next word in a sentence based on the words that came before it. In the context of GPT, input refers to the textual information provided to the model for processing and generating responses. The input can consist of one or more sentences or paragraphs, and it serves as the context or prompt for the model to generate a relevant and coherent response. For example, if you were interacting with a chatbot powered by GPT, you might input a question like, "What is the weather like today?" The model would then generate a response based on its training and understanding of weather-related information. In order to use it in my system, I decided to create a conversation between my program (later called AGENT) and AI (later called DARIUS). I chose the form of conversation to generate commentary using GPT because it allows for a more interactive and dynamic exchange of ideas.

Unlike static one-way communication, a conversation allows me to engage with the model in a back-and-forth manner, enabling a collaborative approach to generate commentary. Using conversation ensure that GPT knows the context of the game and events that happened before, so it is much more likely that it will not make any contextual mistakes i.e. about the actual score. Since the match in the program, was always between the same teams (Real Madrid and Manchester United), the beginning of the conversation was always the same:

The following is a conversation between AGENT and DARIUS. DARIUS is a football commentator, who is getting information about the match from AGENT. The football match will be between Real Madrid and Manchester United.

AGENT: The match will start in a second. Make a welcome talk, where you tell who is playing, in what stadium (Santiago Bernabeu), and who are the best players in both teams (Ronaldo, and Rashford)

DARIUS:

where GPT-3.5 completion API is giving an output which is a text of 'DARIUS'. An example of such text is:

Welcome everyone to the Santiago Bernabeu, where tonight's match will be between Real and United. On the Real side, we have the world-class talent of Cristiano Ronaldo, who is sure to be a key player tonight. On the United side, we have Marcus Rashford, who is sure to be a standout as well. Let's get ready for an exciting match.

After the start of the game, and generating "welcome" output, the next inputs depend on the course of the match. For example, if there is a goal, the program generates the following input:

AGENT: In [time] minute, there is a goal for [currentBallPos], scored by [lastPlayer], the actual score is [score]

Where the [time] is the actual time in the game in minutes, [currentBallPos] is the team that lastly possessed the ball, [lastPlayer] is the player that last time touched the ball, and [score] is the actual score of the game. The example output of such input could be:

Wow! What a start, Valverde scores for Real just 3 minutes into the match! Can United find an equalizer? We have to wait and see!

It is important to notice, that GPT as a prompt is getting not only newly generated input but also full previously generated conversation. It is designed in this way, to provide the important context of the game, and decrease the likeliness of some contextual mistakes. After 90 minutes of the match, in the end, there is also a final talk, where 'DARIUS' is briefly describing the game, based on the events that happen during it, and says something to farewell the players. An example of the whole conversation during one match can be found in Appendix 2.

- **Audio commentary**

Because in every sports video game the commentary is not only generated as text but also as audio, I implemented a Text-to-Speech program to generate spoken commentary from the output of GPT-3.5. The program that I will use for that is

called OpenTTS [6]. It is an open Text-To-Speech server with multiple voices for numerous languages (almost thirty). Since it is an open-source TTS program, it has some limitations, such as a 'robotic', emotionless voice. Because of that, it was difficult to imitate a sports commentator, whose speech is full of emotions. However, I found a voice among those available that was trying to give some impressions into the spoken text and had low generating time (I had to reduce the sound generation delay to a minimum). In order to implement it in my system, I parsed every generated commentary into the URL, opened it in the browser, and automatically played the generated audio commentary.

3.2 On answering RQ2

Assessing the accuracy of AI-generated live commentary is an important process for ensuring the reliability of the information being presented. To assess it, I will compare the output from the GPT-3.5 language model with what happened in the game, for example; If the output in the 10th minute of a game would be: "Player 1 scores a goal by head", I would check in the game replay if it really happened at this moment of the game. If it is true, I will count it as a correct output, if wrong as incorrect output. Next, I will compare the frequency of correct and incorrect facts over multiple games, which provides me with an estimation of the accuracy of my solution. Eventually, I will try to find patterns and causes of why the program is making mistakes, and thanks to that information I may be able to increase the accuracy by improving the system.

3.3 On answering RQ3

To assess the enjoyability of the player about the commentary, I will conduct an interview with a group of players. I want to reach out to football video game players since they are familiar with video game live commentary. To conduct an interview I will invite a group of players and ask them to play the prototype of a football video game with AI-generated commentary. After that, I will ask them questions regarding the game that they just played. The questions will address factors such as the overall enjoyability of the commentary for the player, how accurate and informative it was, how diverse and varied was the commentary in terms of language and vocabulary, and how responsive to the real-time gameplay commentary was. Based on the interview results, I will be able to measure the overall enjoyability of the players' experience with the game commentary.

4 RESULTS

This section will detail the results, of my work on developing the program and evaluating it in terms of accuracy and enjoyability of the players. In section 5.1, I am going to show the results of the accuracy test and analyze the mistakes that it was making; in section 5.2 I am going to show the results of the conducted interview regarding the enjoyability of the players.

4.1 Results of the accuracy test

In order to assess the accuracy of the developed system, I performed manual checking of the generated commentary within twenty games in a football environment. If the output was entirely correct or had

some minor inaccuracies (e.g. wrong first name of the player, since I provided only surnames to the input), I counted it as correct output. However, if the output had some bigger mistakes or was entirely wrong, I counted it as incorrect output. After manually checking twenty games and 403 outputs, I got a result of 315 correct outputs and 88 incorrect outputs, which is giving 78 percent of accuracy. I analyzed some of the most repeating cases of mistakes and found that they are of two types:

1 Mistakes caused by Google Football Environment (67 % of all mistakes):

The most common mistake, in this category, was that the video game coded free kicks and offsides as the same game state (43 % of environment mistakes). Because of that, I could not differentiate the input for offside from the input for a free kick unless a yellow card was also awarded for the same action. For this specific game state, I always encoded input as offside, since it happens more often than foul without the yellow card, but when it occurs, there was a mistake in output. Another common mistake was that the game was too precise about giving information about the player who last touched the ball (23 % of environment mistakes), and there were situations where it turned out that the goalkeeper scored his own goal, because, during saving, he touched the ball. From that, occurs another problem, namely own goals, which were not perceived as a different state than regular goals (occurred every time when own goal is scored). An example of such incorrectness is: "In 35 minutes, there is a goal for Real Madrid, scored by Karim Benzema, the actual score is 0:1", what is wrong because the team who scored is Manchester United.

In order to solve mistakes from this category and increase the accuracy of the program, the environment should differentiate more game states in the function 'observation()', and not count every touch of the ball in the parameter ['ball-owned-player'], only these that are relevant to the course of the game (when the goalkeeper is touching the ball during a saving, but fails, it is not counted as own goal in rules of football).

2 Mistakes caused by GPT 3.5 (33 % of all mistakes):

The most common mistake, in this category, was that Artificial Intelligence overinterpreted some of the inputs, and because of that, it was adding events that did not happen (71 % of GPT mistakes). It is a known issue of artificial intelligence called 'Artificial Hallucination'. Beutel et al. [8] wrote that "In general, "hallucinations" of ChatGPT or similar large language models (LLMs) are characterized by generated content that is not representative or senseless to the provided source, e.g. due to errors in encoding and decoding between text and representations". In the case of my system, it occurred many times because around 71 % of GPT mistakes were caused by hallucinations. For example, when there was a free kick in the middle of the field, it generated an output: "And there will be a free kick for Real! United has to be careful because Ronaldo can turn it into a goal!", despite that it was almost impossible to score from this position. Another type of incorrectness, was when GPT made some logical mistakes

regarding the rules of the game (19 % of GPT mistakes). For example, in some games it generated output: "And there is a goal for United in 45 minutes scored by Rashford! It is a very important goal in the last minutes of the match!", despite the football game lasting 90 minutes, so it was only half of the match.

In order to solve these mistakes and increase the accuracy of the system, the newest version of the natural language model-GPT 4 [9] could help, especially, since it contains new functionality i.e. generating output from visual inputs. Thanks to that, my program could provide an image of the gameplay to the GPT, and generate commentary from it. Moreover, I could try to develop the input to give AI more information about the game, and action that actually happens. It would likely help with overinterpreting and generating commentary unrelated to what is happening in the game.

4.2 Results of the interview

I did semi-structured interviews with, three players who played the football video game- FIFA 23 for at least 10 to 15 hours this year, so they are experienced with commentary in such games. All of them are students from Poland in their early 20's age. I asked them when they will evaluate different aspects of the commentary, to compare it to FIFA 23 or other football video games that they played. They were not informed that the commentary they are going to listen to in the game is AI generated, but I marked that Text-To-Speech that I used is an open-source program, so it has some limitations regarding the quality of voice. The transcript of the interview is in Appendix 3.

• 1. Overall enjoyability:

All of the players found commentary as a positive aspect of the game. They said that it's adding competition and emotions to the game, and without spoken commentary, it would be weird to play. All of them marked that it is not really far from the commentary of real sports video games (which was one of the main goals of the project). The overall rating they gave was 6.5, so in my opinion, it was an enjoyable experience for them.

• 2. Accuracy of the commentary:

For all of them, in general, the commentary was accurate and did not have bigger discrepancies from the gameplay. However, in every game, there was some incorrectness in the generated commentary: in the first case it mistook a throw-in for the corner, in the second despite the winning score for Manchester United it said that it is a draw, and in the last one, it said that there is a dangerous position for taking a free kick, despite that it was in the middle of the field. The first two cases are GPT's mistakes since it had all information in the input to correctly evaluate the action. The last one is the fault of limited input, which should provide information about the position of the ball in the field. Despite some errors, all of the interviewees evaluate that overall the commentary was accurate.

• 3. Diversification and variation of vocabulary and language:

All of the players noticed that there were phrases and words that were repeated many times. This especially happened to the third interviewee, who heard the exact same phrase two times in one game. Most of the repetition happened when the same action occurred many times within one game (for example, when there were a lot of fouls in one game, AI repeated a lot of times word "penalize"). Moreover, in interview two, the text generator started repeating the phrase "We'll have to wait and see" after every commentary. These repetitions are the limitation that this natural language model has, but despite them, for the first player, there was no sense of monotonous commentary during gameplay.

- **4. Responsiveness and timing:**

All of the players evaluate that they felt a bit of delay between occurring of the action and spoken commentary, especially during offsidess. Moreover, sometimes two or even three commentaries overlap when they happen within a small period of time. These errors result from how text-to-speech was implemented in the game and that it did not interrupt the previous commentary before starting the next one. It is a noticeable limitation, and after some time of playing, it decreases the overall enjoyability of the player. However, since the research was mainly focused on generating text commentary and this limitation is caused by text-to-speech implementation, it is not entirely relevant to my research questions.

5 DISCUSSION

In this section, I will discuss the results of the evaluation, by answering research questions and providing the list of limitations and corresponding future work.

- **RQ1: How can an artificial intelligence program be designed to effectively integrate live game data with the OpenAI GPT-3.5 language model to generate live commentary?**

I designed a program, which incorporate the information from the live match in templates, obtaining text that I used to prompt GPT-3.5 with what is happening on the field. I requested a natural language model to create a conversation, where I input that text, and it in response was generating live commentary. Thanks to the form of conversation, I ensured that GPT knew the context of the game, and generated commentary was also based on the previous actions, that already happened. For the purpose of the interview, I also implemented the Text-To-Speech program, to generate live audio commentary.

- **RQ2: To what extent is the GPT-3.5 generated live commentary accurate?**

The accuracy of the GPT-3.5 generated live commentary was a strong point of the program. From the accuracy test, I performed, the program got 78 percent of accuracy. I categorized 22 percent of mistakes into two categories: those caused by Google Football Environment, and those caused by the natural language model. I found also potential solutions to increase the correctness (see sections 4.1.1 and 4.1.2) but despite these errors, interviewees found this commentary overall accurate.

- **RQ3: To what extent is the GPT-3.5 generated live commentary enjoyable for the player?**

From the interview, I infer that GPT-3.5 generated live commentary is overall enjoyable for the players of sports video games. The program has some limitations, which are reducing the enjoyability, i.e. repeating specific words and phrases, and delay between actions on the screen and generated audio commentary. However, all interviewees said that the generated text is close to the one from real sports video games. The overall rating of enjoyability from the interview was 6.5.

5.1 Limitations and Future Work

1 Commenting on more match events

My program was limited to commenting only on the most important events in the match like goals, cards, penalties, corners, etc. but it omits other important events during a match, for example, shots on goal, or precise long passes, that during a normal game are also commented. This limitation is caused by the input for GPT being chosen depending on a game-mode parameter, which was limited to the most important events. In future work, I would try to extend this program by creating methods that would recognize more types of actions on the field. For example, I could differentiate own goal from the regular goal, by checking which team last touched the ball. Thanks to that, I would be able to develop commentary with new inputs, and so on new outputs. It is very likely, that it would make commentary more varied and would be more enjoyable for the players.

2 Extending input for GPT

Input for GPT was limited only to the most important information about the actual state of the match. For example, the only variables for scoring a goal, were: the time when it was scored, the player that scored, and the team that scored. This limitation is caused because, whenever I was trying to put to GPT more developed input, creating an output was taking too long, so it was not live commentary anymore. In future work, if it would be possible to decrease the waiting time, I would extend the input by giving more information e.g. when scoring a goal, I would provide also who assists, whether a goalkeeper tries to save it, what part of the body does player score a goal with. I believe that it would decrease the repetitiveness of the commentary, and so on making it more enjoyable for the players.

3 Using a newer version of GPT

In my project, I used version 3.5 of GPT with the model 'text-DaVinci-003', which training data is up to June 2021, and it is not even the best model in GPT 3.5 'family' (because 'gpt-3.5-turbo' is newer and more cost-effective). Moreover, there is a new GPT 4 [9], which is much better and more advanced than GPT 3.5. It is estimated to contain trillions of parameters (the previous version had 175 billion) and is able to generate 32.768 tokens (the previous version was able to generate 16.384) [7]. Moreover, GPT-4 can process images as input, which could revolutionize my idea. Instead of text input, I could provide images of the gameplay and ask GPT to

comment on it as a football commentator. However, because I used the free trial that OpenAI is sharing for researchers, I was not able to use the newest version of GPT. In future work, I would try to use GPT 4, and it is very likely, that it would increase the variety of vocabulary and language that was used in the generated commentary and so on decrease the repetitiveness and inaccuracy.

6 CONCLUSION

This research aimed to design a program that would generate live commentary in a football video game, by using the natural language model GPT 3.5, and evaluate this proposition in terms of accuracy and enjoyability of the player. I made a prototype that fills pre-written templates with live data from Google Football Environment, and provide it as input to GPT 3.5. I tested it with players, who are regularly playing video games in order to measure their enjoyability of this solution. I got positive results, since they found it likable, and were feeling that it could replace actual commentary. Of course, I found also numerous limitations like inaccuracies caused by the environment (limited perceiving of football events), and artificial intelligence (artificial hallucinations). In my opinion, to solve these issues, we should explore the hypothesis of using the newest model GPT 4, and its functionality- visual inputs. I believe it could increase the accuracy and variety of the generated commentary.

7 REFERENCES

- 1 Tanaka, K., Nakashima, H., Noda, I., Hasida, K., Frank, I., Matsubara, H. (1998, July). MIKE: An automatic commentary system for soccer. In Proceedings International Conference on Multi Agent Systems (Cat. No. 98EX160) (pp. 285-292). IEEE.
- 2 Lee, G., Bulitko, V., Ludvig, E. A. (2013). Automated story selection for color commentary in sports. IEEE transactions on computational intelligence and ai in games, 6(2), 144-155.
- 3 Taniguchi, Y., Feng, Y., Takamura, H., Okumura, M. (2019, July). Generating live soccer-match commentary from play data. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 7096-7103).
- 4 Karol K. Research Lead and Olivier B. Research Scientist. (2019). Introducing Google Research Football: A Novel Reinforcement Learning Environment. <https://ai.googleblog.com/2019/06/introducing-google-research-football.html>
- 5 OpenAI, (2015-2023), openai.com
- 6 Synesthesiam (2021, October). Open Text to Speech Server, <https://github.com/synesthesiam/opentts/>
- 7 OpenAI, (2015-2023), Models overview, <https://platform.openai.com/docs/models/gpt-4>
- 8 Beutel, G., Geerits, E., Kielstein, J.T. (2023). Artificial hallucination: GPT on LSD?. Crit Care 27, 148 <https://doi.org/10.1186/s13054-023-04425-6>
- 9 OpenAI (2023, March). GPT-4 Technical Report <https://doi.org/10.48550/arXiv.2303.08774>

8 APPENDIXES

8.1 Appendix 1

Ball	Home and Away Team	State of game
Position	Team	Score
Direction	Direction	Time Left
Rotation	Tired factor	Game mode
Owned team	Yellow card	
Owned player	Active	
	Roles	

8.2 Appendix 2

The following is a conversation between AGENT and DARIUS. DARIUS is a football commentator, who is getting information about the match from AGENT. The football match will be between Real and United.

Agent: The match will start in a second. Make a welcome talk, where you tell who is playing, in what stadium (Santiago Bernabeu), and who are the best players in both teams (Ronaldo, and Rashford)

DARIUS: Welcome everyone to the Santiago Bernabeu, where Real and United are about to go head to head in this thrilling match. We have some of the best players in the game today on either side, with Ronaldo for Real and Rashford for United. This should be an exciting match so let's get it started!

AGENT: In 2 minute, the ball was kicked out from the field by Real Madrid player, Camavinga

DARIUS: And there's the kickoff! Camavinga sends the ball out for Real Madrid, and we're off and running here in this match between Real and United.

AGENT: In 7 minute, the ball is kicked on corner by Casemiro, Manchester United will have a chance

DARIUS: Casemiro sends the ball into the box, and Manchester United has a chance to capitalize with a corner kick. Let's see what they can do here!

AGENT: In 12 minute Garnacho from Manchester United is on offside

DARIUS: Looks like Garnacho was just flagged for offside, and Manchester United will have to try again. Back to the drawing board for them!

AGENT: In 14 minute Modric from Real Madrid is on offside

DARIUS: Modric is called for offside, and Real Madrid will have to regroup and look to get back on track. Let's see if they can turn this one around.

AGENT: In 21 minute Rashford from Manchester United is on offside

DARIUS: Rashford is caught offside, and Manchester United will have to look for another way to get the ball into the box. This match is still up for grabs!

AGENT: In 29 minute, Valverde is fouled in penalty box and Real Madrid will have a penalty

DARIUS: Valverde is taken down in the penalty box! Looks like Real Madrid will have a chance to score from the spot here. Let's see how they make use of this opportunity.

AGENT: In 29 minute, there is yellow card for Real

DARIUS: And a yellow card for Real Madrid! That's the second one in this match, and they'll want to make sure they don't give away any more in the coming minutes.

AGENT: In 29 minute, there is goal for Real, scored by Valverde

DARIUS: And Valverde makes no mistake with the penalty! Real Madrid take the lead, and this match is shaping up to be a thrilling one.

AGENT: In 40 minute, there is throwout made by Vinicius from Real Madrid

DARIUS: Vinicius sends the ball out for throw-in, and Real Madrid will have an opportunity here to create a scoring chance. Let's see if they can capitalize!

AGENT: In 50 minute, the ball is kicked on corner by Maguire

DARIUS: Maguire sends the ball into the box, and Manchester United will have a chance to level things up with a corner kick. Let's see what they can do here!

AGENT: In 63 minute Rashford from Manchester United is on offside

DARIUS: Rashford is called for offside, and Manchester United will have to try again. Let's see if they can make something happen here!

AGENT: In 70 minute Maguire from Manchester United is on offside

DARIUS: Maguire is flagged for offside, and Manchester United will have to look for another way to get the ball into the box. Can they find a way to level things up here?

AGENT: In 74 minute, the ball is kicked on corner by Casemiro,

DARIUS: Casemiro sends the ball into the box, and Manchester United has a chance to capitalize with a corner kick. Let's see what they can do here!

AGENT: In 74 minute, the ball was kicked out from the field by Manchester United player, Casemiro

DARIUS: And there's the kick-out, with Casemiro sending the ball out for Manchester United. We're still in a tight match here at the Santiago Bernabeu, so let's see what the next few minutes bring!

AGENT: In 86 minute, the ball was kicked out from the field by Real Madrid player, Benzema

DARIUS: And there's the kick-out from Benzema, and Real Madrid will have an opportunity here to close out this match. Let's see what they can do!

AGENT: It is the end of the match, the score is 1 to 0 for Real Madrid summarize the match

DARIUS: It's all over here at the Santiago Bernabeu, and Real Madrid have come away with a 1-0 win! It was a tight match all the way through, with Ronaldo and Rashford proving to be the stars of the show. Congratulations to Real Madrid on their victory!

8.3 Appendix 3

1 How would you rate the overall enjoyability of the AI-generated commentary in the football video game on a scale of 1 to 10? Can you briefly explain your rating?

Player 1: "I would give the rating 7 out of 10, the commentary did not disturb the game, and sounds like something that the real commentator could say during a game. That is why I think that this commentary gives a lot of credibility to the game. You can feel that there are some competitions and emotions, like in a real football video game."

Player 2: "On a scale from 1 to 10, I would give 7, because the commentary added emotions to the whole gameplay, and gives more sense to the game. Without the commentary, you could feel that it is missing something."

Player 3: "The commentary was a really positive aspect for me. Despite that, the gameplay was far from perfect, because of the visible input lag, the commentary brings it closer to the known-for-me football video games like FIFA. For me, it was 6 out of 10."

2 In terms of accuracy, how well did the AI-generated commentary reflect the events happening in the game? Were there any notable discrepancies or inaccuracies?

Player 1: "In general, the commentary was in line with what was going on the field. It was rather accurate and described what was happening in the game. It could be more precise in one example, where the commentator said that there will be a corner for Real Madrid, but it was a throw-in. Apart from this one incorrectness, the commentary accurately described what happened."

Player 2: "Sometimes, I could hear some discrepancies or inaccuracies in the commentary. For example, once when the enemy team was winning, it said that there is a draw in the game. However, there were not a lot of mistakes, more of the game was properly commented."

Player 3: "In general, there were a lot of actions in the middle of the field, so it does not speak often, but every situation around the goal was described well. In one situation it made a weird mistake because when there was a free kick in the middle of the field, it said that it is a dangerous position for scoring a goal, which is illogical in terms of the football game."

3 How diverse and varied were the vocabulary and language used by the AI-generated commentary? Did it repeat certain phrases or become monotonous after a while?

Player 1: "During gameplay, you could not feel that the commentary was monotonous, but after analyzing the text you can notice that some of the phrases are repeating. However, it does not affect the overall receipt of the commentary."

Player 2: "I heard a lot of repeats of some phrases and words. For example word "equalizer" and the phrase "We'll have to wait and see!" repeated numerous times. Although, there were also some original sentences, especially at the beginning."

Player 3: "In this match, I could notice repeats of some words. Especially "penalize", when there were three fouls in a row. In my opinion, it has a problem when there are

too many same events within a small period of time. It even said the same phrase two times."

4 How responsive was the AI-generated commentary to real-time gameplay? Did it provide timely comments and reactions to significant events in the game?

Player 1: "In general, the commentary was said when some action was happening, but you could feel that there was a small delay or sometimes two commentaries overlap themselves. However, in general, the commentary was said after something happened on a field."

Player 2: "The commentary was mostly on time, but you could hear, especially during offsides, that there was a bit of delay. Although, it was not really disturbing."

Player 3: "There was a situation, where two commentaries overlap themselves because there were two actions very close to each other. I also noticed that there is a little delay between action and the spoken commentary."