A Novel Approach to Automated Live-Ticker Generation in Football: Using Large Language Models and Audio Data

James Anurathan, Manfred Rössle[a] and Marco Klaiber[b]
Aalen University of Applied Sciences, Beethovenstr. 1, Aalen 73430, Germany

Keywords: Large Language Model, Audio Data, Football, Named Entity Recognition, Natural Language Processing, Live-Ticker.

Abstract: Football (soccer) is one of the most popular sports in the world, with fans enjoying real-time coverage of their favorite team's from anywhere. Explicitly, the progress in the field of Artificial Intelligence (AI) holds great potential to further improve this experience and optimize the delivery of content. In this context, our work investigates the integration of Large Language Models (LLMs) – in our case GPT-4 – with Advanced Speech Recognition (ASR) systems to automate the creation of live football ticker commentary. For this purpose, we present an approach for transcribing live audio commentary from real football matches, utilizing a whisper model to prepare the transcribed text for correct input to the LLM. This approach is leveraged by Named Entity Recognition (NER) and BERT-based models to provide clear, user-friendly, and multilingual texts for live tickers. In addition, we evaluate our approach with an objective and metric-based method to transparently assess the effectiveness of our approach. The study shows the potential of LLMs in automating sports commentary, but also emphasizes the importance of refining entity recognition and addressing content accuracy issues. Future work should focus on improving transcription accuracy, refining NER models, and mitigating LLM hallucinations to develop more reliable and scalable automated live ticker systems.

1 INTRODUCTION

Football (soccer) is considered one of the most popular sports in the world, captivating a large number of people and experiencing strong continuous growth in recent years (Cotta, 2016, Anzer and Bauer, 2022). This popularity developed football into a very lucrative business, generating billions of dollars from various sources, with fans supporting their teams from all over the world (Goes et al., 2019, Ćwiklinski et al., 2021). This global need for the availability of information from football matches has significantly changed the rapid development of digital journalism in recent years with respect to the landscape of sports reporting (Cheng et al., 2024). As one of the most elementary components of sports reporting, live ticker systems have established themselves as an indispensable element (Ojomo and Olomojobi, 2021).

These systems have become an integral part of sports coverage, providing fans with real-time updates and an immersive experience that bridges the gap for those unable to access live broadcasts on television or radio (Ojomo and Olomojobi, 2021). For most football matches, the textual live-tickers focus on highlighting the most important events such as goals, shots, yellow or red cards, and substitutions (Löchtefeld et al., 2015). In addition, impressions of the game and other important information are conveyed as vividly as possible, to simulate the feeling of being almost live in the stadium. However, manual creation of live-tickers remains a time-consuming and demanding task, requiring constant monitoring and quick text generation by reporters (Huang et al., 2020). In addition, the vivid live-tickers are usually only offered for more popular games, with smaller leagues sometimes not having a live-ticker or only being able to report simple events based on structured data, such as event data.

This challenge emphasises the need for automation, which relies heavily on data being available to accurately capture and report key events (Kunert, 2020). Although structured data captures the basic details of a match, unstructured data, such as audio commentary, provides comprehensive information, including context and detailed insights (Behera

[a] https://orcid.org/0000-0002-9038-9317
[b] https://orcid.org/0009-0007-7070-3413

Anurathan, J., Rössle, M. and Klaiber, M.
A Novel Approach to Automated Live-Ticker Generation in Football: Using Large Language Models and Audio Data.
DOI: 10.5220/0013665700003988

In *Proceedings of the 13th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2025)*, pages 134-141
ISBN: 978-989-758-771-9; ISSN: 2184-3201

and Saradhi, 2024, Tuyls et al., 2021). In combination with advances in Artificial Intelligence (AI), in particular Large Language Models (LLMs), a promising solution is emerging to address the automation of live ticker creation (Bonner et al., 2023). This is partly because state-of-the-art LLMs have the ability to integrate multimodal data, including audio, to effectively deliver results (Strand et al., 2024). This is particularly apparent for Automatic Speech Recognition (ASR) systems (Lakomkin et al., 2024). As a result, the combination of the ASR system and LLM offers the possibility to use the mentioned properties of audio commentaries, for high-quality automated live-tickers (Fathullah et al., 2024, Min and Wang, 2024).

Consequently, a research gap exists, as there are currently no studies that effectively integrate LLMs with audio commentary for the automatic generation of live tickers. Our work aims to address this research gap by developing a novel approach that demonstrates how LLMs can be effectively used to generate and optimise automated live tickers based on audio data from football radio broadcasts. To do this, we propose an approach that involves the process of transcribing, processing and structuring audio data to make the data understandable and usable for LLMs, with a focus on accurately capturing and reporting key events. In addition, we are also exploring the potential of translating the transcribed texts into multiple languages, which should improve the accessibility and global reach of the live ticker system. Therefore, this work offers the following contributions:

1. Development of an approach to automatically recognise relevant football events from audio broadcasts.
2. Implementation of a Large Language Model for the conversion of identified events into clear, user-friendly and multilingual texts for live tickers.
3. Objective metrics-based evaluation of the results to transparently assess the effectiveness of our approach.
4. Based on our challenges, recommendations for future research in the field of football analysis and speech processing are given.

2 RELATED WORK

The use of LLMs to generate live tickers from audio sources is a novel area of research in football. Nevertheless, some studies have pursued similar research approaches, which we present below and distinguish from our own approach.

Cook and Karakus (Cook and Karakuş, 2024) introduced the "LLM Commentator", a system designed to automate real-time football commentary using LLMs. Their approach leverages advanced speech processing techniques and raw football data to generate accurate descriptions of match events. A significant contribution of their work is the exploration of fine-tuning strategies for LLMs, specifically tailored to improve the models' performance in capturing and articulating live football commentary. Furthermore, Sarkhoosh et al.(Sarkhoosh et al., 2024) developed the "SoccerSum" framework, which employs LLMs such as GPT-4 to automate the summarisation of football events. Their approach integrates multimodal data, including audio, to generate narrative content. However, in contrast to the focus of our study, "SoccerSum" extends beyond the creation of live tickers by incorporating video analytics and social media content creation. Another approach was presented by Strand et al. (Strand et al., 2024) in the area of football analytics. "SoccerRAG" is a framework that combines Retrieval Augmented Generation (RAG) with LLMs to effectively respond to natural language queries and find relevant information. This framework integrates multiple data modalities, including video, audio, and recorded commentary, to handle complex queries and improve user interaction with sports archives.

In contrast to presented articles, our work focuses on transcribing and analyzing live audio data from football radio broadcasts to capture key events and produce multilingual live-tickers. In addition, we use a NER approach specially fine-tuned for football, as well as BERT to prepare the events in the best possible way, focusing exclusively on the audio data. We also differentiate ourselves by integrating the possibility of real-time translation into other languages to increase global accessibility.

3 METHODOLOGY

For this work, a german four-minute audio file was obtained from *Sportschau.de*[1], a popular sports show in Germany that focuses on commentary on sporting events, especially football. Therefore, we used a match from the German Bundesliga between *VfL Wolfsburg* and *VfB Stuttgart* (Matchday 24, 2024/03/02). The MP3 file contains the highlights of the game and is used as a dataset for our approach. The overall methodology is shown in Fig. 1.

[1] Sportschau.de - Audio file

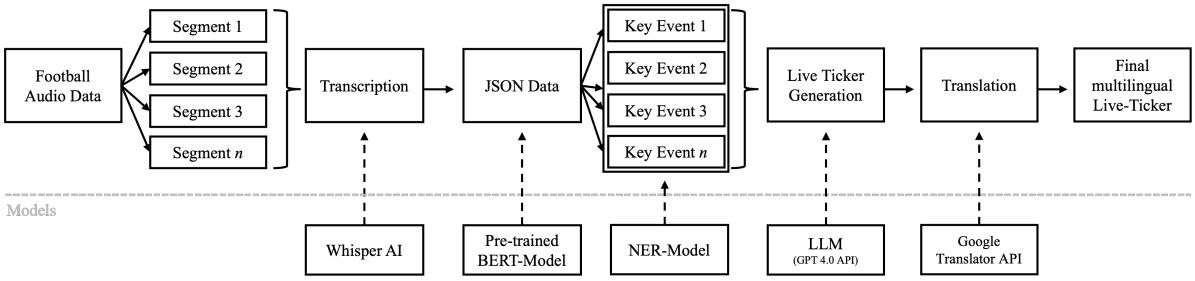

Figure 1: Overview of the individual steps of the proposed approach.

3.1 Transcription of Audio Data

The first step was to provide the audio data in text form; therefore, the data had to be transcribed. The Whisper model (OpenAI, 2025) was used, which is an advanced speech recognition system (ASR) from OpenAI (Radford et al., 2023). Whisper is known for its high speech recognition accuracy and has the ability to process multiple languages, including German, which is why Whisper is considered effective as an ASR system for our approach (Radford et al., 2023,Wills et al., 2023). Of the various versions available, the medium-sized Whisper model was selected based on the trade-off between performance and resource efficiency. In addition, for optimised processing efficiency, the audio file was divided into smaller segments and each segment was then transcribed using Whisper.

3.2 Entity Extraction with BERT-Based NER and Fuzzy Matching

To correctly recognize key events, players, and team names from transcribed audio commentaries, Named Entity Recognition (NER) was used. NER is a core element of Natural Language Processing (NLP), used to classify objects in text (Jehangir et al., 2023). This involves assigning specific tags to each word or token, indicating its role. The integration of NER into the pre-training phase is expected to improve the overall performance of LLMs by providing them with more accurate and structured data inputs (Devlin et al., 2019). An effective model for NER tasks is the Bidirectional Encoder Representations of Transformers (BERT) (Devlin et al., 2019), which was therefore selected for our work. The BERT-based multilingual model is a pre-trained transformer model that captures the context of words in a sentence taking into account the preceding as well as following words, and is also able to accurately recognise and process named objects in multiple languages (Chizhikova et al., 2023). The BERT-based NER model was specifically fine-tuned for the football context in our study.

Since the pronunciation of the player names in the audio file can differ, which can happen especially with non-native names, our dataset was adapted to contain several variants of player names to account for possible transcription differences. Each element in the transcribed text was labeled with specific tags, such as ‘B-PER’ for the beginning of a person’s name, ‘I-PER’ for the continuation of that name, and similar tags for events. The tagged data was then split into training and validation sets to ensure that the NER model could accurately recognise and categorise football-specific entities, which is essential for creating accurate and contextually relevant live ticker updates. In addition to the NER model, fuzzy matching techniques were employed to handle variations and misspellings in player names and other entities. Fuzzy matching enhances the model’s ability to correctly identify entities even when they are not perfectly transcribed, ensuring greater accuracy in the recognition process (Bhasuran et al., 2016).

After training, the NER model was used to process the transcriptions, automatically identifying and labeling the key entities. These labeled entities were then mapped to the correct teams and events, providing the LLM with the structured information needed to generate accurate live-ticker updates. The trained NER model was further applied to extract and classify relevant entities from live audio comments.

3.3 Live-Ticker Generation and Translation

The final step was to create the live ticker commentary using the GPT-4 model. This model was selected because of its advanced capabilities in natural language processing, particularly its ability to understand context, generate coherent text, and integrate transcription data. In addition, a simple prompt was defined that places the model in the role of a sports commentator to capture the context:

> *"You are a sports commentator providing live text commentary. Make sure you correctly report key events, player names, scores, and goal scorers, and assign them to the correct clubs."*

To ensure that the live commentary was accessible to a global audience, the commentary was translated from German to English using the Google Translator API[2]. Google Translator was selected as it provides a fast and reliable balance between translation quality and speed (Baek et al., 2024). In addition, we wanted to outsource the translation task from the LLM to save resources.

3.4 Objective Evaluation Metrics

To ensure an objective evaluation, we integrate three metrics: (1) BERTScore, (2) ROUGE-L, and (3) BLEU. For each metric, we used LLM-generated texts compared to the transcribed original from the audio data, which represents the ground truth. The LLM-generated live-ticker is not intended to reproduce the audio description identically, but to reproduce the most important information on the basis of the audio description, whereby there should be a similarity, which is why the comparison is reasonable. The BERTScore (Zhang et al., 2020) measures the quality of texts based on semantic similarity to the original text. Furthermore, we integrate ROUGE-L, based on the Longest Common Subsequence (LCS) (Lin and Och, 2004), which considers structural similarity at sentence level and automatically identifies the longest n-grams occurring in the sequence (Briman and Yildiz, 2024). In addition, we use BLEU, which analyses the accuracy of n-grams with a focus on the precision of the model output (Tran et al., 2019).

4 RESULTS AND DISCUSSION

Considering the transcription of the audio commentary using the Whisper model, the relevant information was captured and accurately reproduced. This includes, for example, information on the flow and dynamics of the game, as well as specific events such as goals, penalties, and notable moves. Explicitly, the quieter passages were captured almost perfectly. In contrast, the more exciting and louder passages, mostly related to goals, posed difficulties. In addition, the names of the players represent a major challenge for our model, with *Mäle* being transcribed instead of *Maehle* or *Girassi* instead of *Guirassy*, for example. The model struggled to accurately recognize and transcribe player names from the audio commentary, which is a critical aspect of generating precise live-tickers. It can also be assumed that this difficulty may arise with more complex club names, but this was not the case in our work. This highlights the need for more domain-specific language models trained specifically on sports commentary data. Another option would be to provide the model with the entire squad data of the respective teams. For this, a comprehensive dataset of the teams must be collected accordingly, which could be implemented using free-accessible data from Transfermarkt (TM) (Transfermarkt, 2025) or Fbref (FBref, 2025), for example. Alternatively, using a larger and more advanced version of Whisper, such as the "large" model, could potentially improve transcription accuracy. Nevertheless, it is conceivable that this challenge is not only due to the model, but it could also be due to the pronunciation of the players' names by the commentators. For instance, players with names of non-German origin, such as "Guirassy", which has a French pronunciation, might be more challenging to recognize accurately compared to names like "Müller", which have a more straightforward German pronunciation. In addition, different commentators also have different language styles and pronunciations, making exact recognition even more difficult.

In Table 1 excerpts from the LLM-generated live ticker are shown. Our approach was able to report the most relevant match events, for example, goals and scores. Moreover, the disallowed goal was also recognised and recorded in the live ticker. Despite the integration of the NER and fuzzy matching techniques, there are still inconsistencies in the naming of player names. For example, our model incorrectly named the scorer of the 0:1, where *Jührich* (which corresponds to the player *Führich*) was named, although the real scorer was *Guirassy*, which was correctly recorded in the transcription. It should also be noted that our approach generated artificial game events by inserting game minutes that were not present in the original transcribed text, which is referred to as hallucination (Li et al., 2024). Accordingly, the LLM occasionally generated non-existent details for the live commentary, which poses a significant challenge to ensuring accurate and reliable reporting. To mitigate this, the LLM could be fine-tuned specifically for summarizing football commentary, with a strong emphasis on grounding its outputs in actual input data. Integrating a fact-checking mechanism or cross-referencing (Jiang et al., 2024) system within the LLM could also help reduce instances of hallucination. Another

[2]GoogleTranslator API

Table 1: Excerpts of generated live ticker texts and the translation.

Min.	German Live Ticker	English Live Ticker
90' +	Stuttgart sichert sich 3 wichtige Punkte… Tore durch Jürich, Girassi und Wagnumann.	Stuttgart secures 3 important points… Goals by Jürich, Girassi, and Wagnumann.
85'	**TOR!** Matcher erzielt den Anschlusstreffer für Wolfsburg!	**GOAL!** Matcher scores the equalizer for Wolfsburg!
81'	**TOR ABERKANNT!** Wagnumanns Treffer wird aufgrund einer Abseitsposition nicht gegeben.	**GOAL DISALLOWED!** Wagnumann's goal is disallowed due to an offside position.
79'	**TOR!** Ein Flitzer von Wagnumann!	**GOAL!** A streaker from Wagnumann!
72'	Girassi läuft an… Tor! 2:1 VfB Stuttgart.	Girassi runs up… Goal! 2-1 VfB Stuttgart.
70'	Elfmeter für Stuttgart! Mäle bringt Mio zu Fall und sieht Gelb. Girassi legt sich den Ball zurecht.	Penalty for Stuttgart! Mäle brings Mio down and sees yellow. Girassi sets the ball up.
65'	**TOR!** Wolfsburg! Mäle trifft mit einem wunderbaren Schuss aus 18 Metern in das linke obere Eck. 1:1	**GOAL!** Wolfsburg! Mäle scores with a wonderful shot from 18 meters into the top left corner. 1:1
45'	Halbzeit! Stuttgart führt 1:0 in Wolfsburg. Das Tor von Jürich, ein Distanzschuss, der wirklich atemberaubend war.	Halftime! Stuttgart leads 1:0 in Wolfsburg. Jürich's goal, a long-range shot that was truly breathtaking.
15'	**TOR! Stuttgart!** Jürich zieht von links nach rechts und versenkt den Ball aus der zweiten Reihe seelenruhig ins Netz. Super Start für den VfB. Wolfsburg 0, Stuttgart 1.	**GOAL! Stuttgart!** Jürich moves from left to right and calmly sinks the ball into the net from the second row. Great start for VfB. Wolfsburg 0, Stuttgart 1.
1'	**Anpfiff!** Das Spiel ist eröffnet, Wolfsburg hat Anstoß.	**Kick-off!** The game is underway, Wolfsburg has the kick-off.

way to prevent hallucination would be to adjust the prompt, which in our case was kept relatively straightforward, but the addition that the model "should not hallucinate" could improve the results (Tonmoy et al., 2024). However, our approach correctly updated the scores and assigned the goals to the correct teams. Accordingly, the matching at team level is effective, as the model correctly assigns the players to the respective teams. In addition, the translation from German to English has been successfully implemented so that live ticker entries have been translated correctly in terms of content. The incorrect player names have been adopted identically in the translation.

For an objective evaluation of our approach, we used the three presented metrics to compare 10 segments with the reference and LLM-generated text, which is shown in Table 2. It should be noted that the LLM changed the order between the segments 6 – 9, which explains why the values in this area became significantly worse, especially for ROUGE-L and BLEU. To counteract this in the future, it is conceivable to divide the commentary into chunks, for example, whenever a highlight was discovered, whereby all chunks could then be processed in sequential order. The BERTscore, on the other hand, is relatively constant across all segments and has an average value of ≈ 0.70, which shows that our approach is semantically close to the references. On the other hand, an average ROUGE-L of ≈ 0.20 and BLEU of ≈ 5.3 show that there is relatively little exact word choice or n-gram overlap. This pattern is typical when the outputs are heavily paraphrased, content has been rearranged, or differs in length from the ground truth. Overall, the evaluation show that the system is successful in capturing general context and key events, but that there are significant inaccuracies and inconsistencies, particularly in exact word matching and insertion of missing parts. Since the live ticker is not intended to directly reproduce the audio file, the results can be considered a benchmark, which applies in particular to the BERTscore. Improvements in the BERT-based NER model, LLM tuning, and post-processing steps could enhance the quality of live ticker generation, ensuring that outputs are more reliable and closer to actual commentary.

However, we were able to successfully present an initial approach for a multilingual automated live ticker system with Whisper AI, NER, BERT, and an

Table 2: Evaluation of the generated segments.

Segment	BERTScore	ROUGE-L	BLEU
1	0.8150	0.2667	8.91
2	0.6638	0.2014	15.32
3	0.7414	0.2833	22.45
4	0.6814	0.2410	18.75
5	0.6675	0.2117	16.89
6	0.7325	0.1714	4.33
7	0.6541	0.1111	0.9954
8	0.6695	0.0833	1.1174
9	0.7171	0.2143	1.7118
10	0.6898	0.2090	6.8816
Average	**0.70321**	**0.20159**	**5.2703**

LLM. With our research, we are laying the foundations for the automatic generation of live tickers using advanced AI technologies. The live ticker texts generated emphasize the potential of our approach, as the audio files were captured correctly and a meaningful live ticker could be presented.

5 LIMITATIONS AND FUTURE WORK

One challenge was training the NER model to identify key players and game events from the transcribed text. However, the training might not have been sufficient, leading to inconsistent recognition of entities and potential inaccuracies in the live-ticker. To address this, the NER model could benefit from more extensive training on a larger and more diverse dataset. In addition, experimenting with different NER models, e.g. those based on transformer architectures such as RoBERTa (Mehta and Varma, 2023), could lead to better results. Furthermore, it is conceivable to evaluate the effectiveness of individual components, such as the NER model, which can be carried out in the future with ablation studies.

For our approach, we used a game with the highlights audio file, which was sufficient for our study, but it is conceivable that more data could improve the results. The accuracy of both the transcription and entity recognition processes could be limited by the small size of the training data, limiting the model's ability to generalize effectively. Collecting a larger dataset, including a wide range of football matches with varying commentary styles and audio quality, could significantly improve model performance. The use of data augmentation techniques to synthetically increase the size of the dataset could also be beneficial (Moreno-Barea et al., 2020). It is also conceivable that not only highlights are used, but entire broadcasts of matches, which would significantly increase the amount of data but would prioritize the recognition of highlights for the live ticker.

We used three different metrics for the evaluation, namely BERTScore, ROUGE-L, and BLEU, although other metrics or evaluation approaches could be considered in the future. ROUGE-L and BLEU in particular are susceptible to deterioration in results when formulations differ despite semantic correctness, which could potentially obscure the validity of the live ticker passages. Consequently, COMET (Tasnim et al., 2019) or BARTScore (Yuan et al., 2021), which have been used in other domains, could be considered in the future. It should also be considered for future evaluation that, especially for a 90-minute game, a relevant challenge is to find the appropriate highlights that are worth mentioning in a live ticker.

Furthermore, our approach focused on using a previous game to demonstrate feasibility. However, live commentaries are generated on match days, which then have to be converted into a live ticker in real-time. The performance of our approach was sufficient for our use case, but it remains to be seen how the system will handle real-time processing, especially when processing large audio files over 90 minutes. In the future, further optimization of the system for real-time processing could be promoted, which includes streamlining audio segmentation and transcription pipelines, possibly using lighter models that offer a better compromise between accuracy

and speed. Exploring parallel processing techniques (Brakel et al., 2024) and GPU acceleration (Huang et al., 2024) could also improve performance.

6 CONCLUSION

This study demonstrates the potential of integrating LLMs with advanced speech recognition systems to automate live-ticker generation in football. The proposed approach, which includes the implementation of the Whisper model for audio transcription, followed by NER and fuzzy matching techniques, effectively processes live commentary to generate accurate, real-time textual updates. Our approach showcases both the possibilities and challenges involved in creating a robust system for real-time sports commentary automation. The results showed that while the system successfully captured the overall context and key events of a football match, there were notable challenges with player name recognition and the generation of non-existent match details. These inaccuracies underscore the need for further refinement in both the transcription process and the entity recognition models. Specifically, enhancing the training datasets with more diverse and extensive football commentary could improve the system's ability to generalize across different pronunciation variations and commentary styles.

The limitations identified in this study, such as the LLM's tendency to hallucinate and the challenges in real-time processing, point to several avenues for future research. Fine-tuning LLMs to better handle football-specific commentary, improving NER model accuracy, and optimizing real-time processing capabilities are essential steps forward. Furthermore, exploring alternative transformer-based models like RoBERTa (Liu et al., 2019) or developing more domain-specific LLMs (Jeong, 2024) could further enhance system performance. Future research should also consider the scalability of such systems, particularly in multilingual environments and across various sports. Implementing cloud-based distributed computing solutions could address these scalability concerns, allowing simultaneous processing of multiple matches in different languages.

REFERENCES

Anzer, G. and Bauer, P. (2022). Expected passes: Determining the difficulty of a pass in football (soccer) using spatio-temporal data. *Data Mining and Knowledge Discovery*, 36(1):295–317.

Baek, S., Lee, S., and Seok, J. (2024). Strategic Insights in Korean-English Translation: Cost, Latency, and Quality Assessed through Large Language Model. *2024 Fifteenth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 551–553.

Behera, S. R. and Saradhi, V. V. (2024). Visualization of Unstructured Sports Data – An Example of Cricket Short Text Commentary.

Bhasuran, B., Murugesan, G., Abdulkadhar, S., and Natarajan, J. (2016). Stacked ensemble combined with fuzzy matching for biomedical named entity recognition of diseases. *Journal of Biomedical Informatics*, 64:1–9.

Bonner, E., Lege, R., and Frazier, E. (2023). Large Language Model-Based Artificial Intelligence in the Language Classroom: Practical Ideas for Teaching. *Teaching English with Technology*, 23(1):23–41.

Brakel, F., Odyurt, U., and Varbanescu, A.-L. (2024). Model Parallelism on Distributed Infrastructure: A Literature Review from Theory to LLM Case-Studies.

Briman, M. K. H. and Yildiz, B. (2024). Beyond ROUGE: A Comprehensive Evaluation Metric for Abstractive Summarization Leveraging Similarity, Entailment, and Acceptability. *International Journal on Artificial Intelligence Tools*, 33(05):1–23.

Cheng, L., Deng, D., Xie, X., Qiu, R., Xu, M., and Wu, Y. (2024). SNIL: Generating Sports News From Insights With Large Language Models. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–14.

Chizhikova, A., Konovalov, V., Burtsev, M., Kryzhanovsky, B., Dunin-Barkowski, W., Redko, V., and Tiumentsev, Y. (2023). Multilingual Case-Insensitive Named Entity Recognition. In *Advances in Neural Computation, Machine Learning, and Cognitive Research VI*, pages 448–454.

Cook, A. and Karakuş, O. (2024). LLM-Commentator: Novel fine-tuning strategies of large language models for automatic commentary generation using football event data. *Knowledge-Based Systems*, 300:1–22.

Cotta, L. (2016). Using FIFA Soccer video game data for soccer analytics. *Workshop on large scale Sports Analytics*, pages 1–4.

Ćwiklinski, B., Giełczyk, A., and Choraś, M. (2021). Who Will Score? A Machine Learning Approach to Supporting Football Team Building and Transfers. *Entropy*, 23(1):1–12.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Fathullah, Y., Wu, C., Lakomkin, E., Jia, J., Shangguan, Y., Li, K., Guo, J., Xiong, W., Mahadeokar, J., Kalinli, O., Fuegen, C., and Seltzer, M. (2024). Prompting Large Language Models with Speech Recognition Abilities. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 13351–13355.

FBref (2025). Fbref.com. https://fbref.com/en/. Accessed: 08.01.2025.

Goes, F. R., Kempe, M., Meerhoff, L. A., and Lemmink, K. A. (2019). Not Every Pass Can Be an Assist: A

Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches. *Big Data*, 7(1):57–70.

Huang, K. H., Chen, L., and Chang, K. W. (2020). Generating Sports News from Live Commentary: A Chinese Dataset for Sports Game Summarization. In *Proc. 1st Conf. of the Asia-Pacific Chapter of the Ass. for Computational Linguistics and the 10th Int. Joint Conf. on Natural Language Processing*, pages 609–615.

Huang, Y., Wan, L. J., Ye, H., Jha, M., Wang, J., Li, Y., Zhang, X., and Chen, D. (2024). Invited: New Solutions on LLM Acceleration, Optimization, and Application. In *Proceedings of the 61st ACM/IEEE Design Automation Conference*, pages 1–4.

Jehangir, B., Radhakrishnan, S., and Agarwal, R. (2023). A survey on Named Entity Recognition — datasets, tools, and methodologies. *Natural Language Processing Journal*, 3:1–12.

Jeong, C. (2024). Domain-specialized LLM: Financial fine-tuning and utilization method using Mistral 7B. *Journal of Intelligence and Information Systems*, 30(1):93–120.

Jiang, L., Jiang, K., Chu, X., Gulati, S., and Garg, P. (2024). Hallucination Detection in LLM-enriched Product Listings. In Malmasi, S., Fetahu, B., Ueffing, N., Rokhlenko, O., Agichtein, E., and Guy, I., editors, *Proc. Seventh Workshop on e-Commerce and NLP @ LREC-COLING*, pages 29–39.

Kunert, J. (2020). Automation in Sports Reporting: Strategies of Data Providers, Software Providers, and Media Outlets. *Media and Communication*, 8(3):1–11.

Lakomkin, E., Wu, C., Fathullah, Y., Kalinli, O., Seltzer, M. L., and Fuegen, C. (2024). End-to-End Speech Recognition Contextualization with Large Language Models. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 12406–12410.

Li, H., Chi, H., Liu, M., and Yang, W. (2024). Look Within, Why LLMs Hallucinate: A Causal Perspective.

Lin, C.-Y. and Och, F. J. (2004). Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics - ACL '04*, pages 1–8.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.

Löchtefeld, M., Jäckel, C., and Krüger, A. (2015). TwitSoccer: Knowledge-Based Crowd-Sourcing of Live Soccer Events. In *Proceedings of the 14th International Conference on MUM*, pages 1–4.

Mehta, R. and Varma, V. (2023). LLM-RM at SemEval-2023 Task 2: Multilingual Complex NER using XLM-RoBERTa.

Min, Z. and Wang, J. (2024). "Exploring the Integration of Large Language Models into Automatic Speech Recognition Systems: An Empirical Study". In Luo, B., Cheng, L., Wu, Z.-G., Li, H., and Li, C., editors, *Neural Information Processing*, pages 69–84.

Moreno-Barea, F. J., Jerez, J. M., and Franco, L. (2020). Improving classification accuracy using data augmentation on small data sets. *Expert Systems with Applications*, 161:1–14.

Ojomo, O. W. and Olomojobi, O. T. (2021). Viewing the Game Textually: Online Consumption of Live Text Commentary as Alternate Spectatorship Among Nigerian Football Fans. *Communication & Sport*, 9(3):496–521.

OpenAI (2025). Introducing Whisper. https://openai.com/index/whisper/. Accessed: 15.01.2025.

Radford, A., Kim, J. W., Xu, T., Brockman, G., Mcleavey, C., and Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 28492–28518.

Sarkhoosh, M. H., Gautam, S., Midoglu, C., Sabet, S. S., Torjusen, T., and Halvorsen, P. (2024). The SoccerSum Dataset for Automated Detection, Segmentation, and Tracking of Objects on the Soccer Pitch. In *Proceedings of the 15th ACM Multimedia Systems Conference*, MMSys '24, page 353–359.

Strand, A. T., Gautam, S., Midoglu, C., and Halvorsen, P. (2024). SoccerRAG: Multimodal Soccer Information Retrieval via Natural Queries. *arXiv*.

Tasnim, M., Collarana, D., Graux, D., Galkin, M., and Vidal, M.-E. (2019). COMET: A Contextualized Molecule-Based Matching Technique. In *Lecture Notes in Computer Science*, pages 175–185.

Tonmoy, S. M. T. I., Zaman, S. M. M., Jain, V., Rani, A., Rawte, V., Chadha, A., and Das, A. (2024). A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models.

Tran, N., Tran, H., Nguyen, S., Nguyen, H., and Nguyen, T. (2019). Does BLEU Score Work for Code Migration? In *2019 IEEE/ACM 27th International Conference on Program Comprehension (ICPC)*, pages 165–176.

Transfermarkt (2025). transfermarkt.com. https://www.transfermarkt.de/. Accessed: 12.04.2025.

Tuyls, K., Omidshafiei, S., Muller, P., Wang, Z., Connor, J., Hennes, D., Graham, I., and et al. (2021). Game Plan: What AI can do for Football, and What Football can do for AI. *JAIR*, 71:41–88.

Wills, S., Bai, Y., Tejedor-García, C., Cucchiarini, C., and Strik, H. (2023). Automatic Speech Recognition of Non-Native Child Speech for Language Learning Applications (Short Paper). In *arXiv:2306.16710*, pages 1–8.

Yuan, W., Neubig, G., and Liu, P. (2021). BARTScore: Evaluating generated text as text generation. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 27263–27277.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2020). BERTScore: Evaluating Text Generation with BERT.