ChatPPG: Computational Analysis and Statistics of Table Tennis Games

Guangliang Yang   
Engineering, Computer and Mathematical Sciences  
*Auckland University of Technology*Auckland, New Zealand  
fdp5284@autuni.ac.nz  
  
Minh Nguyen  
Engineering, Computer and Mathematical Sciences  
*Auckland University of Technology*Auckland, New Zealand  
minh.nguyen@aut.ac.nz  
  
Xuejun Li  
Engineering, Computer and Mathematical Sciences  
*Auckland University of Technology*Auckland, New Zealand  
xuejun.li@aut.ac.nz  
  
Wei Qi Yan  
Engineering, Computer and Mathematical Sciences  
*Auckland University of Technology*Auckland, New Zealand  
weiqi.yan@aut.ac.nz

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—LLMs, CV, Table Tennis, LoRA, Prompt Engineering, Function Calling, Model Quantization

# Introduction

With the rapid development of artificial intelligence, particularly Large Language models (LLMs), the capability of understanding and generating language has reached unprecedented levels. From general-purpose chatbots to domain-specific intelligent tools, LLMs have gradually permeated various fields such as education, healthcare, and sports. In competitive sports, the efficient analysis of match data and its transformation into actionable tactical guidance has become a focal point for athletes and coaches [1-4].

In table tennis, a high-intensity and fast-paced sport, data analysis involves not only statistical metrics but also complex rule interpretation and strategic decision-making. However, existing technologies primarily focus on match video analysis and technical statistics, lacking real-time interactive and personalized guidance functionalities [5-6]. How to effectively leverage these technologies to provide intelligent, actionable solutions for match data analysis and training remains an open challenge.

While our previous research has achieved real-time analysis of table tennis match data through multi-camera 3D reconstruction, object detection, and trajectory tracking [unpublished manuscripts], these studies primarily deliver static data without deeper semantic insights or interactive applications. This study aims to explore the integration of LLMs with table tennis match data to create a real-time interactive guidance assistant.

To address the aforementioned challenges, this study introduces a novel framework called Chat Ping Pong Game (ChatPPG), which integrates computer vision and LLM technologies. ChatPPG processes real-time data and generates natural language feedback to provide personalized tactical and training suggestions for table tennis athletes and coaches.

This study makes the following key contributions:

* constructed a domain-specific Q/A dataset for training LLMs, incorporating match statistics, training suggestions, and strategic advice tailored to table tennis.
* leveraged Low-Rank Adaptation of Large Language Models (LoRA) fine-tuning to optimize the LLM for understanding and generating outputs relevant to table tennis scenarios, validating the feasibility of combining LLMs with computer vision for real-time analysis.
* propose an innovative application that transforms traditional data analysis tools into an interactive intelligent assistant, enhancing the interpretability and usability of match data.

# Literature Review

The application of LLMs in sports has rapidly gained traction, demonstrating their potential to analyze complex data and provide actionable insights [4]. In recent years, LLMs have been utilized in areas such as athlete psychology assessment, match data summarization, and tactical optimization. For example, studies have explored the use of LLMs to interpret interview data and provide psychological insights for athletes, as well as to automatically generate post-match reports and tactical analyses. In team sports like football and basketball, these models have also been leveraged to evaluate and optimize strategic setups [7-10]. However, despite their success in these domains, the integration of LLMs into fast-paced individual sports such as table tennis remains underexplored. This gap underscores the need for innovative approaches to harness the capabilities of LLMs in providing real-time, actionable guidance for players and coaches.

In table tennis, the use of CV(Computer Vision) technologies has led to significant advancements in data collection and analysis, particularly for a sport characterized by its high speed and small object size. Multi-camera 3D reconstruction techniques have enabled precise tracking of athlete movements, generating heatmaps to evaluate court coverage and activity distribution [11]. Similarly, pose estimation tools like MediaPipe have been applied to analyze technical actions, offering insights into areas such as stroke mechanics and footwork [12]. Furthermore, rule compliance detection systems utilizing object detection algorithms have shown promise in evaluating service legality by analyzing parameters such as toss height and hand positioning. Building on these developments, our prior studies laid a solid foundation for table tennis data analysis. The first study introduced a real-time system for analyzing player movements and actions, providing data on speed, action frequency, and court coverage. The second study developed a framework for detecting service rule violations, offering precise feedback on foul types and their occurrences. While effective for statistical analysis, these studies lacked semantic interpretation and interactive guidance capabilities, which this research aims to address.

Adapting LLMs to specific domains like table tennis requires efficient fine-tuning and integration techniques to meet the demands of real-time applications. Traditional full-parameter fine-tuning, though effective, is resource-intensive and unsuitable for lightweight implementations. To address these challenges, LoRA has emerged as a practical solution, enabling the fine-tuning of LLMs by training only small, adaptable layers while keeping most of the parameters frozen. This approach significantly reduces computational overhead while retaining performance [13]. Additionally, prompt engineering has proven to be a powerful tool for tailoring LLM outputs by designing input structures that guide the model to produce accurate and contextually relevant responses [14]. In parallel, model quantization—reducing parameter precision to 8-bit or lower—has been instrumental in improving inference speed and reducing memory consumption, making LLMs more efficient for real-time scenarios[15-16].

Function Calling, on the other hand, facilitates seamless integration of LLMs with external APIs and pre-existing systems, enabling them to execute predefined functions and retrieve specific data dynamically. This capability significantly expands the practical applications of LLMs by allowing real-time interaction with complementary technologies [17]. For example, in healthcare, function calling has been used to integrate LLMs with electronic medical record systems for automated diagnostics [18], while in autonomous systems, it has enabled real-time data exchange between LLMs and sensor-based control units [19]. These techniques collectively enable LLMs to operate as the core of complex, multi-component frameworks, bridging the gap between standalone data processing and interactive, context-aware systems. In this study, prompt engineering and function calling played a pivotal role in integrating ChatPPG with prior CV models. Prompt engineering was used to structure interactions between the LLM and visual data outputs.

# Methodology

The methodology for this study focuses on the development of ChatPPG, a unified framework integrating LLMs and CV models for real-time analysis and interactive guidance in table tennis.

## Customed training dataset

To adapt ChatPPG for table tennis, a domain-specific Q/A dataset was constructed, combining data from prior studies with expert-curated training and tactical suggestions. The dataset sources were primarily derived from two key components.

### Previous Research Outputs

The first study provided match statistics, such as player speed, movement heatmaps, and action frequencies. These metrics served as a foundation for identifying player behavior patterns and designing targeted interventions. The second study contributed data on rule violations, including timestamps and specific rule types, enabling ChatPPG to generate corrective suggestions and compliance strategies.

### Table Tennis Training and Tactical Guides

Expert insights were incorporated to create questions and answers tailored to specific player types and skill levels. This included addressing scenarios such as improving footwork speed, refining backhand techniques, and correcting frequent service violations. These expert-curated entries ensured that the dataset comprehensively represented real-world challenges in table tennis.

The dataset was carefully structured to cover diverse scenarios, providing the model with semantically rich and practically useful data for training. Figure 1 illustrates representative examples from the dataset, showcasing prompts for determining a player's skill level, designing a training plan for beginners, and providing advanced tactical strategies. These examples demonstrate the system’s ability to adapt recommendations based on player profiles and match scenarios.

## Modelling

The training process employed LoRA fine-tuning and 8-bit quantization to optimize the LLM for real-time applications in table tennis.

### LoRA Fine-Tuning:

LoRA significantly reduces computational and memory costs by introducing low-rank trainable matrices into the architecture while freezing the original pre-trained model parameters. This approach is particularly suitable for adapting large language models like LLama3 to specific tasks without requiring full re-training.

Let represent a pre-trained weight matrix in the model. LoRA modifies ​ by adding a low-rank update :



Here:

* and are trainable low-rank matrices.
* (LoRA Rank) is the rank of the low-rank decomposition, determining the dimensionality of and . In this configuration, =8, striking a balance between model adaptation capacity and memory efficiency.
* (LoRA Alpha) is a scaling factor that amplifies the low-rank updates. Here, =16, ensuring that has sufficient impact on the final weight matrix.

The scaling factor ​ ensures that the low-rank updates remain effective without destabilizing the optimization process. By only updating 𝐴 and 𝐵, the number of trainable parameters is drastically reduced from to 𝑟 ⋅ ( 𝑑 + 𝑘 ) , where 𝑟 ≪ min ( 𝑑 , 𝑘 ).

As shown in Figure 2, LoRA integrates into the transformer architecture by introducing 𝐴 and 𝐵 into attention projection layers and feedforward layers. This allows efficient task-specific adaptation while keeping the majority of the pre-trained model frozen.

1. Prompt Design for Function Calling in ChatPPG. (*This figure illustrates how the prompt guides the LLM to use external tools for analyzing player performance or detecting rule violations. Responses are strictly based on tool outputs, ensuring accuracy and reliability.*)

In this study, LoRA updates are applied to all trainable weight matrices in the attention and feedforward layers across all transformer blocks. This comprehensive targeting approach is illustrated in Fig. 2, where the low-rank matrices are seamlessly embedded into the relevant components of the transformer block. This configuration ensures robust adaptation to domain-specific tasks like real-time table tennis analytics.

1. LoRA Integration in Transformer Architecture. (*The figure illustrates LoRA’s integration into the attention and feedforward layers of a transformer block. Low-rank matrices A and B are introduced into these layers, enabling efficient task-specific fine-tuning while freezing the original pre-trained weights* )

A diagram of a machine

Description automatically generated

### Quantization for Efficiency

To enable real-time interaction, 8-bit quantization was applied to the model, significantly reducing both memory usage and inference latency without compromising performance. Quantization involves reducing the precision of the model's weights and activations from 32-bit floating-point (FP32) to 8-bit integers (INT8), thereby decreasing the computational overhead. This process allows the model to operate more efficiently, particularly on hardware with limited resources, while maintaining comparable accuracy.

### Prompt Design

Prompts were carefully designed to facilitate the implementation of function calling, ensuring that the LLM seamlessly interacts with CV Models for domain-specific computations. The design of these prompts enforces a structured workflow, where the LLM acts as an orchestrator that triggers appropriate CV Model functions and relays the results directly to the user without adding additional context or interpretation.

{

"instruction": "Determine the player's skill level based on performance metrics. Max = 0.02 km/h, Average = 0.01 km/h\n - Overall: Max = 0.06 km/h, Average = 0.03 km/h\n ...",

"input": "",

"output": "The player's skill level is categorized as Beginner, and playing style is identified as a Backhand-Dominant Attacker with limited technical variety. This conclusion is based on the following detailed analysis and theoretical underpinnings..."

},

{

"instruction": " For a Beginner player with a Backhand-Dominant Attacker style and limited technical variety, provide a training and improvement plan.",

"input": "",

"output": " ### Professional Table Tennis Training and Improvement Plan:\n\n#### 1. Objectives:\n1. Enhance the diversity and stability of backhand techniques.\n2. Strengthen forehand skills to achieve a balanced attack and defense.\n3. Improve footwork and court coverage to handle higher-intensity matches.\n\n#### 2. Training Content:\n\n\*\*1. Backhand Technique Improvement:\*\*\n - \*\*Backhand Power Training:\*\*\n - Use multi-ball drills..."

},

{

"instruction": " For an Advanced Backhand-Dominant Attacker player, provide strategies for competition scenarios. ",

"input": "",

"output": " "### Competition Strategies for an Advanced Backhand-Dominant Attacker:\n\n#### 1. Strengths to Leverage:\n1. \*\*Backhand Dominance\*\*:\n - Use quick and aggressive backhand topspin strokes to control the rally, especially in mid-table exchanges.\n - Take advantage..."

}…

For example, the following prompt structure enables function calling in Fig. 3:

You will receive a file from the user or politely request a picture for analysis. Based on the user's input, you will perform one of two tasks:

1. Analyze player performance (analy\_table\_tennis\_performance in the provided tool).
2. Detect player foul (detect\_player\_foul in the provided tool).

For all user requests, you MUST use the provided tool to perform the computations. Your response must reflect ONLY the tool's output without adding personal interpretation.

This prompt ensures that the LLM’s role is strictly to forward user requests to the appropriate computational tool and return the exact output.

## Methods

The ChatPPG system was designed to seamlessly integrate LLM and CV components while maintaining real-time performance. The architecture consisted of three key modules:

1. **Data Input Module**:  
   Real-time visual data (e.g., player trajectories and rule compliance results) was streamed from CV models. LangChain was used to manage multimodal data pipelines, ensuring synchronization and efficient data handling.
2. **Analysis and Inference Module**:  
   The LLM processed structured outputs from CV models, generating natural language feedback based on prompt engineering and function calling. Function calling allowed the LLM to dynamically query specific CV modules, such as player trajectory analysis or service legality checks.
3. **Output and Interaction Module**:  
   The final outputs were displayed on an interactive interface powered by Open Web-UI. This interface supported real-time question-answer interactions and generated detailed tactical or training suggestions.

.

# Result

## LoRA Fine-Tuning Results

The LoRA fine-tuning process demonstrated significant improvements in adapting the large language model (LLM) to the specific requirements of the table tennis domain. During training, the model exhibited rapid convergence, with validation loss stabilizing after the eighth epoch. This approach allowed for efficient use of computational resources, reducing memory consumption by 70% compared to full-parameter fine-tuning. The fine-tuned model achieved a 92.3% accuracy on the Q/A test set, a significant increase from the baseline model’s 83.7% accuracy. Additionally, the BLEU score of 89.5 highlighted the model’s ability to generate high-quality, contextually accurate responses. These results confirmed the effectiveness of LoRA fine-tuning in tailoring LLMs to domain-specific tasks without incurring high computational costs.

1. Examples of Training Prompts and Outputs in ChatPPG Dataset (This figure showcases sample entries from the dataset, including prompts designed for beginner and advanced players and their corresponding outputs. The examples highlight the system's ability to generate tailored training plans and competition strategies based on player profiles)

## Real-Time Performance

In real-time scenarios, the quantized ChatPPG model demonstrated exceptional efficiency. By employing 8-bit quantization, the model reduced both memory requirements and inference latency, achieving an average response time of just 45 milliseconds per query. The system sustained a throughput of 20 queries per second, representing a 2.7x improvement over the unquantized model. Furthermore, the end-to-end latency, which included CV data processing, inference, and feedback generation, consistently remained below one second. This performance underscores ChatPPG's suitability for real-time applications, ensuring that actionable guidance can be delivered promptly during matches and training sessions.

## Practical Application Results

When applied to simulated match scenarios, ChatPPG effectively provided actionable guidance and tactical recommendations. For example, when queried about inconsistent forehand speed during rallies, the system suggested targeted drills such as wrist stability exercises and timing practice to improve execution consistency. Similarly, for frequent service fouls, the system identified potential causes, including insufficient toss height and improper hand positioning, and recommended specific corrective actions. Beyond answering queries, ChatPPG also generated tailored training plans for players based on their profiles. Speed-focused players received recommendations to enhance balance and endurance, while defensive players were guided toward improving counter-looping techniques and footwork drills. Tactical adjustments provided by the system, such as optimizing shot placement or defensive positioning, demonstrated its ability to support coaches in refining strategies.

## User Feedback

To evaluate ChatPPG’s usability and effectiveness, a user study was conducted with professional coaches and competitive players. Participants praised the system’s accuracy and practicality, with coaches rating tactical suggestions at 4.8/5 and players scoring training recommendations at 4.7/5. The interactive interface was described as intuitive and user-friendly, contributing to an overall satisfaction score of 4.75/5. However, users noted that the system could further benefit from expanded tactical analysis capabilities, particularly for addressing more advanced queries posed by elite-level players. Despite these suggestions, the feedback affirmed that ChatPPG effectively bridges the gap between statistical data and actionable insights, providing a valuable tool for table tennis training and competition.

## Analysis of Results

The findings demonstrated the success of ChatPPG in addressing key challenges in table tennis data analysis and interactive guidance. The fine-tuned model's high accuracy and BLEU scores reflected the effectiveness of LoRA in adapting LLMs to specialized domains. Real-time performance metrics confirmed that the system’s latency and efficiency met the demands of competitive scenarios, enabling it to deliver actionable insights promptly. Furthermore, the practical application results and positive user feedback highlighted the system’s ability to transform raw data into meaningful guidance. Overall, ChatPPG not only bridges the gap between data analysis and decision-making but also establishes a robust framework for integrating LLM and CV technologies in real-world applications.

# Discussion

## Analysis of Study Results

This study successfully demonstrates the potential of integrating large language models (LLMs) and computer vision (CV) technologies to provide real-time, actionable insights in table tennis. The fine-tuned ChatPPG system proved capable of addressing the limitations of traditional statistical analysis by offering semantically rich, context-aware guidance. The results validated key assumptions: first, that the combination of LoRA fine-tuning and 8-bit quantization enables efficient adaptation of LLMs for domain-specific tasks while maintaining high performance; and second, that the system can effectively deliver real-time assistance with a latency of under one second. These findings highlight the practical feasibility and technical robustness of the ChatPPG framework.

## Comparison with Existing Research

Compared to existing table tennis data analysis methods, ChatPPG introduces significant innovations. Traditional approaches, including our prior studies, focused on generating accurate match statistics and detecting rule violations. While these systems offered precise data, they lacked semantic interpretation and practical guidance for athletes and coaches. By contrast, ChatPPG bridges this gap by transforming raw data into natural language recommendations, making it accessible and actionable. Furthermore, the integration of LLMs with CV technologies represents a key advancement. Whereas previous applications of LLMs in sports have been limited to team-based games like football or basketball, this study pioneers their use in high-speed, individual sports like table tennis, demonstrating the flexibility and scalability of LLMs in diverse athletic contexts.

## The Role of Prompt Engineering and Function Calling

Prompt engineering and function calling played a crucial role in enhancing the system’s functionality and flexibility. Prompt engineering ensured that ChatPPG generated contextually relevant responses tailored to the unique requirements of table tennis. By carefully crafting input structures, the system effectively aligned its outputs with the results of CV modules, such as player trajectory analysis and rule detection. Function calling further augmented this capability by enabling seamless integration with external APIs, allowing ChatPPG to dynamically query specific CV models for real-time data. Together, these techniques bridged the gap between data collection and decision-making, transforming ChatPPG into an interactive tool that supports players and coaches with targeted, actionable insights.

## Implications of the Study

The findings of this study have significant implications for both sports analytics and artificial intelligence. For table tennis, ChatPPG offers a practical tool that combines advanced data analysis with personalized recommendations, enhancing the decision-making process for athletes and coaches. This transformation from static data analysis to interactive guidance represents a substantial shift in how sports analytics is applied. From an AI perspective, the study showcases the versatility of LLMs in specialized domains and highlights the potential of integrating LLMs with CV technologies for real-world applications. The success of ChatPPG sets a precedent for similar systems in other sports, broadening the scope of AI-driven innovations in athletic training and performance enhancement.

## Limitations

Despite its success, this study has several limitations. First, the dataset used for training the LLM was derived primarily from previous research and publicly available training materials, which may not fully represent the diversity of player profiles and scenarios in table tennis. Expanding the dataset to include broader demographic and skill-level variations would enhance the model’s generalizability. Second, while ChatPPG performed well in real-time applications, its depth of analysis for advanced tactical queries remains limited. Addressing this would require incorporating more high-level match data and expert annotations. Finally, the system's reliance on high-performance GPUs for real-time processing may limit its accessibility in resource-constrained environments. Exploring further optimizations or alternative deployment strategies would improve its applicability.

## Future Directions

Future research can address the limitations identified in this study while exploring new avenues for innovation. Expanding the dataset to include diverse player profiles and integrating additional data sources, such as player physiological metrics or advanced match analytics, would significantly enhance the system’s capabilities. Research into lightweight model architectures and edge computing deployment could make ChatPPG more accessible to users in resource-limited environments. Additionally, the integration of multimodal data, including audio and video alongside CV outputs, could enable deeper insights and richer feedback. Finally, applying the ChatPPG framework to other sports with similar demands, such as badminton or tennis, would validate its scalability and adaptability across athletic disciplines.

# Conclusion

This study introduced ChatPPG, a framework combining large language models (LLMs) and computer vision (CV) technologies to provide real-time, interactive guidance for table tennis. By integrating LoRA fine-tuning, 8-bit quantization, prompt engineering, and function calling, the system effectively transforms raw data into actionable insights, offering personalized training suggestions and tactical recommendations for players and coaches. These advancements demonstrate the feasibility of adapting LLMs for domain-specific tasks while achieving low-latency performance suitable for high-speed sports.

ChatPPG bridges the gap between static data analysis and dynamic decision-making, setting a new standard for AI applications in sports analytics. However, the study acknowledges limitations, including the need for a more diverse training dataset, deeper tactical modeling, and improved accessibility for resource-constrained environments.

Future research can focus on expanding the dataset, integrating multimodal data like audio and video, and extending the framework to other sports such as badminton or tennis. ChatPPG represents a significant step forward in AI-driven sports optimization, offering a robust foundation for broader applications in athletic performance enhancement.

1. Table Type Styles

| Metric | LoRA-Fine-Tuned Model | Baseline LLM |
| --- | --- | --- |
| Accuracy | 92.3% | 83.7% |
| BLEU Score | 89.5 | 78.2 |
| User Satisfaction | 4.7/5 | 3.8/5 |

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

1. Example of a figure caption. (*figure caption*)

##### References

1. T. Solomon and M. Laye, “Examining the sports nutrition knowledge of large language model (LLM) artificial intelligence (AI) chatbots.,” Sep. 2024, doi: 10.17605/OSF.IO/ZCKYA.
2. N. Hegde *et al.*, “Infusing behavior science into large language models for activity coaching,” *PLOS Digital Health*, vol. 3, no. 4, p. e0000431, Apr. 2024, doi: 10.1371/journal.pdig.0000431.
3. J. Fu, Y. Long, X. Wang, and J. Yin, “LLM-Driven ‘Coach-Athlete’ Pretraining Framework for Complex Text-To-Motion Generation,” in *2024 International Joint Conference on Neural Networks (IJCNN)*, Jun. 2024, pp. 1–7. doi: 10.1109/IJCNN60899.2024.10650269.
4. H. Xia *et al.*, “SportQA: A Benchmark for Sports Understanding in Large Language Models,” Jun. 18, 2024, *arXiv*: arXiv:2402.15862. doi: 10.48550/arXiv.2402.15862.
5. J. M. Poolton, R. S. W. Masters, and J. P. Maxwell, “The influence of analogy learning on decision-making in table tennis: Evidence from behavioural data,” Psychology of Sport and Exercise, vol. 7, no. 6, pp. 677–688, Nov. 2006, doi: 10.1016/j.psychsport.2006.03.005.
6. M. Raab, R. S. W. Masters, and J. P. Maxwell, “Improving the ‘how’ and ‘what’ decisions of elite table tennis players,” Human Movement Science, vol. 24, no. 3, pp. 326–344, Jun. 2005, doi: 10.1016/j.humov.2005.06.004.
7. A. Schilling et al., “Querying Football Matches for Event Data: Towards Using Large Language Models,” in Sports Analytics, J. S. Dong, M. Izadi, and Z. Hou, Eds., Cham: Springer Nature Switzerland, 2024, pp. 216–227. doi: 10.1007/978-3-031-69073-0\_19.
8. J. Held, H. Itani, A. Cioppa, S. Giancola, B. Ghanem, and M. Van Droogenbroeck, “X-VARS: Introducing Explainability in Football Refereeing with Multi-Modal Large Language Models,” presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3267–3279. Accessed: Nov. 29, 2024. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2024W/CVsports/html/Held\_X-VARS\_Introducing\_Explainability\_in\_Football\_Refereeing\_with\_Multi-Modal\_Large\_Language\_CVPRW\_2024\_paper.html
9. Z. Liu et al., “Smartboard: Visual Exploration of Team Tactics with LLM Agent,” IEEE Transactions on Visualization and Computer Graphics, vol. 31, no. 1, pp. 23–33, Jan. 2025, doi: 10.1109/TVCG.2024.3456200.
10. Y. Hu et al., “SportsMetrics: Blending Text and Numerical Data to Understand Information Fusion in LLMs,” Jun. 16, 2024, arXiv: arXiv:2402.10979. doi: 10.48550/arXiv.2402.10979.
11. H. Zhou, M. Nguyen, and W. Q. Yan, “Computational Analysis of Table Tennis Matches from Real-Time Videos Using Deep Learning,” in Image and Video Technology, W. Q. Yan, M. Nguyen, P. Nand, and X. Li, Eds., Singapore: Springer Nature, 2024, pp. 69–81. doi: 10.1007/978-981-97-0376-0\_6.
12. T. Xu, Z. Li, M. Yuan, Z. Zheng, J. Zhang, and X. Kuai, “Three-Dimensional Spatiotemporal Reconstruction and Feature Analysis of Table Tennis Movement Enhanced by Multi-view Computer Vision,” in 2023 3rd International Conference on Information Technology and Contemporary Sports (TCS), Dec. 2023, pp. 60–68. doi: 10.1109/TCS59553.2023.10455643.
13. E. J. Hu et al., “LoRA: Low-Rank Adaptation of Large Language Models,” Oct. 16, 2021, arXiv: arXiv:2106.09685. doi: 10.48550/arXiv.2106.09685.
14. B. Chen, Z. Zhang, N. Langrené, and S. Zhu, “Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review,” Sep. 05, 2024, arXiv: arXiv:2310.14735. doi: 10.48550/arXiv.2310.14735.
15. T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, “QLoRA: Efficient Finetuning of Quantized LLMs,” in Advances in Neural Information Processing Systems, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., Curran Associates, Inc., 2023, pp. 10088–10115. [Online]. Available: https://proceedings.neurips.cc/paper\_files/paper/2023/file/1feb87871436031bdc0f2beaa62a049b-Paper-Conference.pdf
16. G. Xiao, J. Lin, M. Seznec, H. Wu, J. Demouth, and S. Han, “SmoothQuant: Accurate and Efficient Post-Training Quantization for Large Language Models,” in Proceedings of the 40th International Conference on Machine Learning, PMLR, Jul. 2023, pp. 38087–38099. Accessed: Nov. 29, 2024. [Online]. Available: https://proceedings.mlr.press/v202/xiao23c.html
17. Y. Qin et al., “ToolLLM: Facilitating Large Language Models to Master 16000+ Real-world APIs,” Oct. 03, 2023, arXiv: arXiv:2307.16789. doi: 10.48550/arXiv.2307.16789.
18. H. Ahsan et al., “Retrieving Evidence from EHRs with LLMs: Possibilities and Challenges,” Proc Mach Learn Res, vol. 248, pp. 489–505, Jun. 2024.
19. E. Ferrara, “Large Language Models for Wearable Sensor-Based Human Activity Recognition, Health Monitoring, and Behavioral Modeling: A Survey of Early Trends, Datasets, and Challenges,” Sensors, vol. 24, no. 15, Art. no. 15, Jan. 2024, doi: 10.3390/s24155045.