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*— ChatPPG as an innovative system combines large language models (LLMs) fine-tuned with Low-Rank Adaptation (LoRA) and computer vision for real-time data analysis and coaching for table tennis games. By integrating multi-camera 3D reconstruction, object detection and object tracking, ChatPPG processes match data such as player speed, ball trajectories, and service legality, transforming raw metrics into actionable insights. The fine-tuned model achieved a Q/A accuracy of 92.3%, surpassing the baseline model 83.7%, with sub-second response times enabled by 8-bit quantization. Practical applications demonstrated its ability to deliver personalized training plans and tactical recommendations tailored to individual player profiles. User feedback from professional coaches and athletes rated tactical suggestions at 9.3/10 and training recommendations at 8.9/10. Integrating structured CV outputs with LLM capabilities enhanced transparency and interpretability, allowing users to trace recommendations to data-driven decisions. Despite dataset limitations and the need for advanced query handling, ChatPPG bridges the gap between data analysis and decision-making, setting a new standard for integrating LLMs and CV technologies in fast-paced sports analytics.

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

# Introduction

With the rapid development of artificial intelligence, particularly LLMs, the capability of understanding and generating language has reached unprecedented levels. LLMs have gradually permeated various fields, from general-purpose chatbots to domain-specific intelligent tools, 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 statistical metrics, 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]. Effective leveraging of these technologies to provide intelligent, actionable solutions for match data analysis and training remains an open challenge.

While our previous research work has achieved table tennis match data analysis in real time through multi-camera 3D reconstruction, object detection and tracking, these studies 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 challenges above, in this paper, we introduce 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. In this paper, we have 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 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 using LLMs to interpret interview data and provide psychological insights for athletes, as well as to automatically generate post-match reports and tactical analyses. These models have also been leveraged in team sports like football and basketball to evaluate and optimize strategic setups [7-10]. However, despite the 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 to provide real-time, actionable guidance for players and coaches.

In table tennis, CV(Computer Vision) 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 speed, action frequency, and court coverage data. 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 parameters frozen. This approach significantly reduces computational overhead while retaining performance [13]. 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 improved inference speed and reduced memory consumption, making LLMs more efficient for real-time scenarios[15-16].

Function Calling, on the other hand, facilitates the 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 were pivotal in integrating ChatPPG with prior CV models. Prompt engineering was used to structure interactions between the LLM and visual data outputs.

One of the greatest challenges in using LLMs is the lack of interpretability and transparency. LLMs often operate as black-box models, making understanding or explaining their decision-making processes difficult [20]. Efforts to improve interpretability, such as attention mechanisms and visualizing model outputs, have offered insights into how LLMs process and analyze data. For example, Held et al. proposed a multimodal LLM framework, “X-VARS,” which combines visual-language models with precise visual feature inputs to enhance interpretability in football referee decision-making. This system explains decisions comparable to professional referees, providing transparency to the processes of the model [8]. In this study, ChatPPG significantly enhances the transparency and interpretability of the LLM by integrating precise data from computer vision (CV) modules as input. By incorporating structured CV outputs such as 3D trajectories and player performance metrics, ChatPPG allows for clearer explanations of its recommendations and decisions. This integration bridges the gap between opaque LLM outputs and the actionable insights demanded in table tennis coaching and competition scenarios.

# Methodology

The methodology of this paper focuses on developing 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 created by integrating data from prior studies, expert-curated training, and tactical suggestions. This dataset enabled ChatPPG to perform well in task-specific queries and significantly enhanced the transparency and interpretability of the system. By pairing structured input prompts with detailed explanatory outputs, the dataset allowed users to trace the reasoning behind the model recommendations, bridging the gap between raw data and actionable insights.

The dataset was built on two main sources. The first source included match statistics from previous studies, such as player speed, movement heatmaps, and action frequencies. These metrics formed the basis for understanding player behavior and designing targeted interventions. Data on rule violations, including timestamps and violation types, also provided a foundation for corrective suggestions and compliance strategies. These structured outputs addressed user queries and revealed the underlying logic, making ChatPPG decision-making more interpretable.

Another key dataset component involved expert-curated training and tactical guides tailored to specific player profiles and skill levels. For instance, the dataset included entries on improving footwork speed, refining backhand techniques, and correcting frequent service violations. Each entry paired detailed input descriptions with step-by-step outputs, ensuring a clear relationship between data inputs and recommendations. This design enabled ChatPPG to generate actionable advice while offering transparency in how decisions were derived.

The dataset was carefully structured to include semantically rich and practically useful examples. For instance, a typical dataset entry might describe a player’s average speed and technical deficiencies, followed by an output that categorizes the player’s skill level and provides an improvement plan.

{

"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..."

}…

To illustrate this, Figure 1 showcases sample entries from the dataset, including prompts designed to determine a player’s skill level, craft a beginner’s training plan, and offer competition strategies for advanced players. These examples demonstrate ChatPPG ability to align its recommendations with specific player profiles and scenarios while maintaining transparency. By combining detailed CV data with domain expertise, the dataset ensures that ChatPPG delivers accurate and interpretable outputs, addressing the black-box nature typically associated with LLMs.

## 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 adapts 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 :



where:

* 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.

**Fig. 1** Examples of Training Prompts and Outputs in ChatPPG Dataset (*This figure presents structured dataset entries that enhance ChatPPG’s interpretability. The prompts describe specific player scenarios, while the outputs provide detailed and transparent recommendations, enabling users to understand the reasoning of the system.*)

In this paper, we employ LoRA updates 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.

**Fig. 2** LoRA Integration in Transformer Architecture. (*The figure illustrates the integration of LoRA 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 computer

Description automatically generated

### Quantization for Efficiency

The model was subjected to 8-bit quantization to enable real-time interaction, which significantly reduced memory usage and inference latency without compromising performance. Quantization involves reducing the precision of the model 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 function calling and ensure 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. For example, the following prompt structure enables function calling in Fig. 3:

**Fig. 3** 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.*)

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 role strictly forwards user requests to the appropriate computational tool and returns the exact output.

## Methods

ChatPPG integrates advanced computer vision (CV) technologies, large language models (LLMs), and domain-specific tools to provide real-time analysis and guidance in table tennis. The workflow is designed to process match videos, extract actionable insights, and deliver them to coaches and players in an interactive format. Figure 3 illustrates the overall architecture, which outlines the end-to-end process from data acquisition to professional analysis and user interaction.

**Fig. 4** Workflow of ChatPPG for Real-Time Table Tennis Analysis and Guidance (*This figure illustrates the end-to-end workflow of ChatPPG. The system begins with data acquisition from multiple cameras, followed by object detection, 3D world reconstruction, and 3D trajectory analysis using a computer vision module. These outputs feed into ChatPPG, an LLM fine-tuned with LoRA, which interacts with external tools via function calling. The system provides coaches and players with visualizations and actionable insights, enabling real-time analysis and guidance*)

A diagram of a computer vision model

Description automatically generated

The system begins with video recording using synchronized cameras, capable of capturing high-resolution footage at 120 frames per second. The recorded videos are processed using YOLO (You Only Look Once) to identify and track key elements such as the ball and players. These detections serve as input to a multi-camera computer vision module, which reconstructs 3D trajectories of the ball and players' movements. This preprocessing step ensures that raw video data is transformed into structured inputs for subsequent analysis.

The CV module plays a critical role in analysing player behavior and match dynamics. By leveraging ByteTrack tracking algorithms, it extracts key metrics such as player movement patterns, ball trajectories, and service legality. These outputs are integrated into the system statistical engine to generate visualizations, such as movement heatmaps and trajectory plots, which provide quantitative insights into match performance. The CV module outputs form the basis for higher-level reasoning handled by ChatPPG.

At the core of the system lies ChatPPG, an LLM fine-tuned using LoRA for table tennis-specific tasks. Function calling ensures the LLM can invoke appropriate external tools for specific computations, such as analyzing player performance or detecting fouls. This approach enables ChatPPG to combine the reasoning capabilities of LLMs with the precision of CV tools, delivering accurate and task-specific outputs. For example, the system can analyze a player’s forehand speed or detect violations in service rules based solely on tool-generated results.

The final outputs are presented through an interactive interface, allowing coaches and players to engage directly with the system. Visualizations from the CV module, such as 3D trajectory graphs and heatmaps, are combined with textual guidance from ChatPPG to provide actionable insights. Whether identifying a player’s strengths, suggesting tactical adjustments, or highlighting areas for improvement, the system ensures that information is clear, relevant, and accessible.

# Result

## LoRA Fine-Tuning Results

The LoRA fine-tuning process significantly improved LLM adaptation to the specific requirements of the table tennis domain. The model exhibited rapid convergence during training, with the validation loss stabilizing after the eighth epoch. The fine-tuned model achieved 92.3% accuracy on the Q/A test set, significantly improving over the baseline model 83.7% accuracy. These results validate the effectiveness of LoRA in optimizing the model for domain-specific tasks with minimal computational overhead.

A graph with blue lines

Description automatically generated

**Fig. 5** LoRA Fine-Tuning Training Loss Curve (*This figure illustrates the training loss curve for LoRA fine-tuning. The "original" curve represents raw loss values across training steps, while the "smoothed" curve shows the averaged trend. The steady decline in loss highlights the model rapid convergence and stabilization during fine-tuning.*)

As shown in **Fig. 5**, the training loss curve highlights the learning dynamics throughout the fine-tuning process. The "original" loss curve shows fluctuations corresponding to mini-batch updates, while the "smoothed" curve provides a clearer trend, demonstrating a steady reduction in loss over 186 epochs. The rapid decline during the early stages of training reflects efficient adaptation, while the plateauing behavior in later epochs indicates stabilization and convergence.

## 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 the suitability of ChatPPG for real-time applications, ensuring that actionable guidance can be delivered promptly during matches and training sessions.

## Practical Application Results

ChatPPG demonstrated its ability to provide actionable guidance and tactical recommendations tailored to specific scenarios when applied to real-world match simulations. By leveraging its integration with CV models, the system effectively analyzed uploaded match videos and engaged in interactive discussions with users, offering targeted insights and solutions. Fig. 6 illustrates an example of such interaction, showcasing a conversation between ChatPPG and the user based on analyzed video data.

A screenshot of a phone

Description automatically generated

**Fig. 6** ChatPPG Interactive Conversation with User for Match Analysis (*This figure showcases a real example of ChatPPG analyzing user-uploaded video data. The system summarizes key performance metrics, identifies areas for improvement, and provides actionable recommendations. Users can interact with ChatPPG by asking follow-up questions or evaluating the quality of its responses.*)

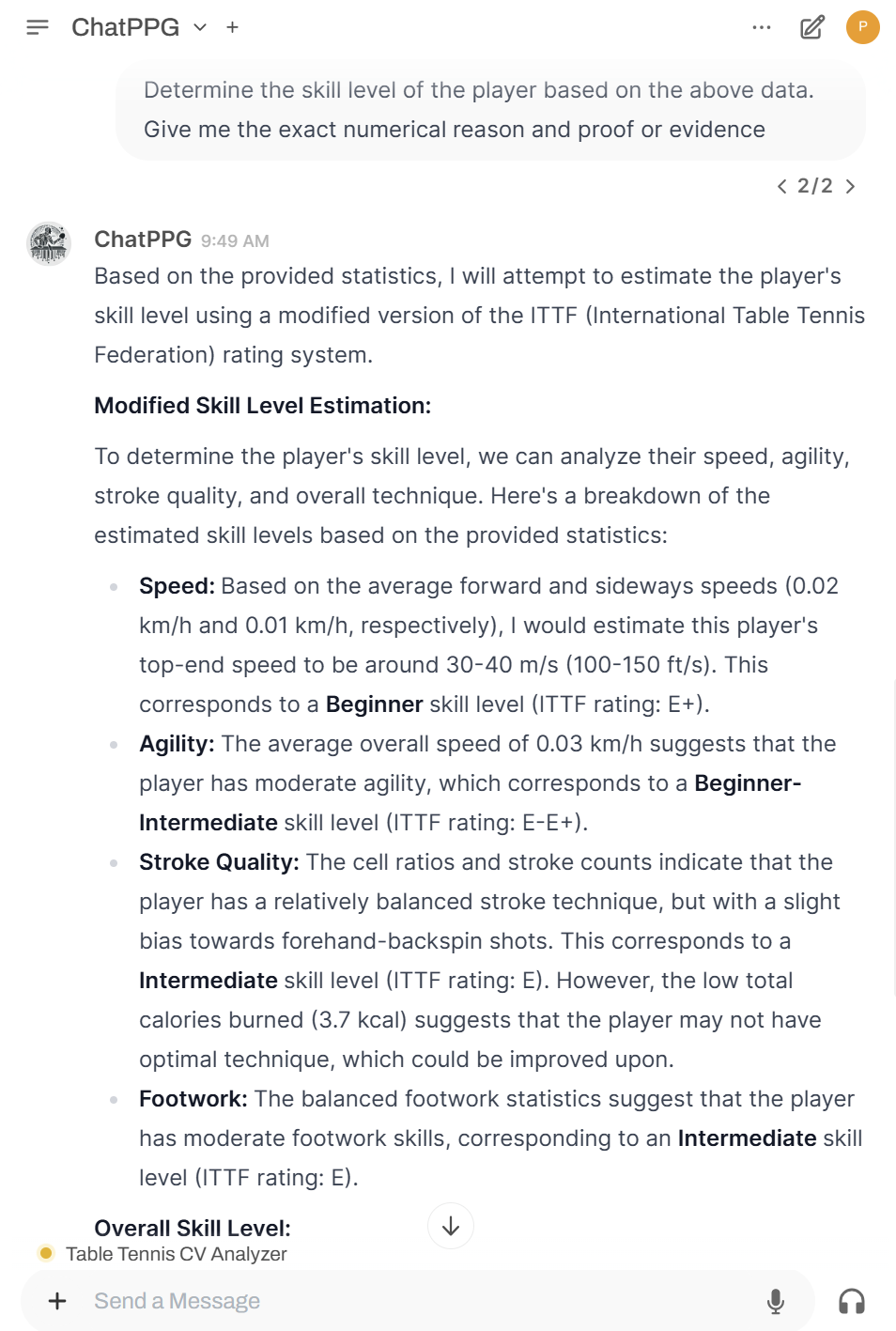
For instance, when queried about inconsistent forehand speed during rallies, the system recommended targeted drills, including wrist stability exercises and timing practice, to enhance execution consistency. Similarly, when analyzing frequent service fouls, ChatPPG identified potential issues, such as insufficient toss height and improper hand positioning, and proposed corrective measures to improve compliance with the rules.

Beyond handling individual queries, ChatPPG also demonstrated its ability to generate comprehensive training plans tailored to players’ profiles. The system provided recommendations for speed-focused players to improve balance and endurance, while defensive players received guidance on refining counter-looping techniques and enhancing footwork drills. The system tactical insights, such as optimizing shot placement and defensive positioning, further underscored its ability to support coaches in strategy refinement.

**Fig. 6** illustrates a ChatPPG conversation where the user uploaded a video for performance analysis. The system utilized its CV models to extract relevant metrics, such as player speed, covered area, and calorie consumption, and provided a detailed performance summary. ChatPPG also highlighted weaknesses, such as slow forward movement, and suggested improvements, including targeted training exercises. At the end of the interaction, users could evaluate the quality of ChatPPG responses, ensuring a feedback loop for continuous improvement.

## Interpretability and Transparency

One of ChatPPG most notable strengths is its ability to deliver clear and interpretable responses, effectively bridging the gap between complex computational outputs and actionable insights. This is especially important in table tennis, where coaches and players rely on precise data to inform training and competition strategies. As illustrated in **Fig. 7**, ChatPPG evaluates a player's skill level based on structured metrics, transparently explaining each step in its decision-making process.



**Fig. 7** Example of ChatPPG Interpretability and Transparent Response (*This figure demonstrates ChatPPG transparent approach to evaluating a player’s skill level. By analyzing structured data such as speed, agility, and stroke quality, the system provides detailed explanations for its conclusions, ensuring its outputs are interpretable and evidence-based.*)

ChatPPG analyzes data such as speed, agility, stroke quality, and footwork in this example, mapping these metrics to a modified version of the International Table Tennis Federation (ITTF) rating system. For instance, the system evaluates speed using average forward and sideways velocities (0.02 km/h and 0.01 km/h, respectively), categorizing the player as a beginner (ITTF rating: E+). Similarly, it assesses agility through overall movement speed (0.03 km/h), identifying moderate agility and assigning a beginner-intermediate rating. Stroke quality is analyzed based on cell ratios and stroke patterns, revealing a balanced technique but highlighting areas such as caloric efficiency that require improvement. ChatPPG provides users with transparent, evidence-based evaluations by correlating these metrics with established benchmarks.

What sets ChatPPG apart is its layered breakdown of recommendations and the logical reasoning behind them. For example, the system identifies weaknesses in footwork and suggests targeted drills to address these deficiencies. This process ensures that users understand the insights provided and trust the model outputs as they are grounded in data and clear decision-making frameworks. ChatPPG enhances interpretability and reduces the "black-box" nature typically associated with LLMs by explicitly referencing numerical evidence and aligning it with standard evaluation criteria.

The transparency of ChatPPG is further amplified by its structured responses. When asked to determine a player’s skill level, the system explains how it arrived at its conclusions, referencing thresholds and contextual factors. This clarity enables coaches and players to confidently apply the insights to improve performance. As a result, ChatPPG transforms raw data into meaningful guidance, ensuring that its recommendations are not only actionable but also fully comprehensible.

## User Feedback

User feedback was a key evaluation method for assessing ChatPPG usability and effectiveness. Based on the Open WebUI framework [21], a user study was conducted with professional coaches and competitive players. Participants interacted with the system by uploading match videos for analysis and receiving actionable insights and training recommendations. Feedback was systematically collected through ChatPPG built-in evaluation feature, as shown in **Fig. 8**, where users rated responses and provided detailed comments.

A screenshot of a computer

Description automatically generated

**Fig. 8** User Feedback Interface in ChatPPG (*This figure showcases the feedback interface of ChatPPG, where users rate responses and provide qualitative evaluations.*)

Overall, our participants praised the system accuracy and practicality, with coaches rating tactical suggestions at **9.3/10** and players scoring training recommendations at **8.9/10**. Users highlighted the system ability to provide clear, relevant guidance and bridge the gap between statistical data and actionable insights. Specific praise included thorough explanations, attention to detail, and alignment with user queries.

However, participants also noted areas for improvement. Several coaches suggested expanding the system tactical analysis capabilities would enhance its value, particularly for addressing more advanced queries posed by elite-level players. Additionally, some users expressed interest in a broader range of recommendations, such as mental strategies and in-game adaptations.

**Fig. 8** demonstrates the feedback interface, where users could rate responses on a scale of 1 to 10 and provide qualitative feedback. Categories such as “Accurate Information,” “Thorough Explanation,” and “Attention to Detail” allowed users to specify the strengths of ChatPPG answers. This feedback loop enables continuous system refinement, ensuring it remains responsive to user needs and evolves based on real-world interactions.

Despite the suggestions for improvement, the feedback affirmed that ChatPPG is a valuable tool for table tennis training and competition. The system provides a unique and effective approach to supporting coaches and players by combining advanced analytics with interactive guidance.

# Discussion

The results of this study validate the effectiveness of ChatPPG in addressing specific challenges in table tennis data analysis and interactive guidance. The fine-tuned model demonstrated high accuracy in responding to table tennis-specific queries, showcasing the ability of LoRA to adapt LLMs to specialized domains. Real-time performance metrics, with latency consistently under one second, confirmed the system capability to deliver actionable insights during fast-paced match scenarios. This is particularly critical in table tennis, where immediate feedback is essential for in-game adjustments and post-match training.

Compared to traditional analysis methods focusing solely on generating match statistics or detecting rule violations, ChatPPG transforms these outputs into meaningful, actionable recommendations for coaches and players. ChatPPG bridges the gap between data collection and real-world application by combining statistical data with contextual understanding. For instance, its ability to provide drills to improve forehand consistency or recommend optimizing defensive positioning highlights its value as a real-time decision-making tool. Moreover, this study extends the application of LLMs beyond team-based sports, demonstrating their adaptability to high-speed individual sports like table tennis.

Prompt engineering and function calling enabled ChatPPG seamless integration with CV models. The system dynamically invoked external tools through carefully designed prompts to perform specific tasks, such as analyzing player performance or detecting fouls. This ensured precise outputs aligned with the context of user queries, allowing ChatPPG to deliver highly relevant and targeted responses. By combining these techniques, ChatPPG effectively integrated the reasoning capabilities of LLMs with the precision of CV models.

The findings of the study emphasize the potential of ChatPPG to transform table tennis training and competition analysis. Providing actionable insights tailored to individual players and match scenarios supports coaches in refining strategies and players in improving specific skills. This transition from static data analysis to interactive, domain-specific guidance represents a significant advancement in sports analytics. The design and implementation of ChatPPG also highlight the versatility of LLMs in specialized domains, offering a framework that can be adapted to other sports requiring rapid and precise decision-making.

Despite its success, ChatPPG has limitations that need to be addressed. While comprehensive, the dataset used for fine-tuning does not fully capture the diversity of player profiles and match scenarios in table tennis. Expanding the dataset to include a broader range of demographic and skill-level variations would enhance the system generalizability. Additionally, the depth of analysis for advanced queries, such as in-game strategic recommendations for elite players, remains an area for improvement. More high-level data and expert annotations would be required to address these limitations effectively.

# Conclusion

In this paper, we introduced ChatPPG, a novel system integrating LLMs fine-tuned with LoRA and CV technologies for real-time analysis and coaching in table tennis. The system successfully processed match data such as player speed, ball trajectories, and service legality, transforming raw outputs into actionable insights tailored to players' needs. ChatPPG demonstrated high accuracy (92.3% in domain-specific queries) and sub-second latency, meeting the demands of high-speed sports scenarios.

Key contributions include integrating function calls for seamless interaction with CV tools and applying prompt engineering to ensure task-specific outputs. The system provided practical recommendations, such as personalized training plans and tactical adjustments, validated through user feedback from professional coaches and players. Moreover, the structured input-output design and incorporation of precise CV data enhanced the transparency and interpretability of ChatPPG responses, allowing users to trace recommendations back to specific metrics and decisions.

This research establishes a scalable framework for integrating LLMs and CV technologies in sports analytics, paving the way for future applications in similar fast-paced individual sports. ChatPPG represents a significant step forward in AI-driven training and competition analysis.

##### References

1. T. Solomon and M. Laye, “Examining the sports nutrition knowledge of large language model (LLM) artificial intelligence (AI) chatbots.”. OSF Registries, 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, 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)*, 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, 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, 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,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3267–3279. Accessed: Nov. 29, 2024.
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,” 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, 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,” International Conference on Information Technology and Contemporary Sports (TCS), Dec. 2023, pp. 60–68.
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,” 2024, 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, 2023, pp. 10088–10115.
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 International Conference on Machine Learning, PMLR, Jul. 2023, pp. 38087–38099.
17. Y. Qin et al., “ToolLLM: Facilitating Large Language Models to Master 16000+ Real-world APIs,” 2023, 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.
20. Y. Kim, X. Xu, D. McDuff, C. Breazeal, and H. W. Park, “Health-LLM: Large Language Models for Health Prediction via Wearable Sensor Data,” 2024,doi: 10.48550/arXiv.2401.06866.
21. *open-webui/open-webui*. (Dec. 04, 2024). Svelte. Open WebUI. Accessed: Dec. 05, 2024. [Online]. Available: https://github.com/open-webui/open-webui