A Survey of Large Models in Sports

Anonymous ACL submission

Abstract

Sports have witnessed growing global enthusiasm in recent years, serving as a vital force for physical health, cultural exchange, social connection, and economic growth. The rapid advancement large models, particularly (multimodal) large language models (M)LLMs, has demonstrated transformative potential to reshape sports understanding, analysis, and interaction across diverse domains. This paper presents a comprehensive survey of large models in sports, including (i) an overview of tasks and applications across different participant groups; (ii) a detailed analysis of sports-related datasets and benchmarks; and (iii) a critical discussion of current challenges and future directions. Our goal is to establish a foundation for advancing research and practical development of large-model-driven sports intelligence. An open-source GitHub repository is maintained as a continuously updated resource: https://anonymous.4open.science/r/Anonym202B.

1 Introduction

In recent years, the global enthusiasm for sports has continued to rise, with more and more people actively participating in it, and the sports industry has also flourished. To further drive this development, modern sports increasingly rely on massive data support (Hutchins, 2016), while the introduction of Artificial Intelligence (AI) has greatly accelerated this trend (Zhou et al., 2025a).

Early interdisciplinary research in sports and AI focused on natural language processing and computer vision, with applications in tasks such as sports data processing (Cossich et al., 2023) and video analysis (Naik et al., 2022). Entering the era of large models, especially the rapid development of Large Language Models (LLMs) and Multimodal Large Language Models (MLLMs), such as GPT-4 (Achiam et al., 2023) and Google Gemini (Team et al., 2023), has brought new opportunities and challenges to the sports field, as shown in Figure 1. These models not only generate language effectively but also process multiple data modalities, enabling broader applications in sports. Tasks that were previously difficult–such as designing athlete training plans (Skerik et al., 2018), selecting coaching strategies (Bunker and Susnjak, 2022), and generating sports game summarization (Huang et al., 2020)–have been greatly enhanced by large models. Moreover, leveraging their vast knowledge bases, these models can generate more comprehensive and personalized content (Lin et al., 2024b). The number of papers on large models in sports has grown rapidly, from just 1 in 2020 to 78 in 2024, and continues to increase in 2025 (see Figure 4 in the Appendix).

A growing body of review literature has examined the use of AI and deep learning in sports (Zhou et al., 2025a; Zhao et al., 2025). The most relevant work on large models includes studies on their applications to exercise recommendations (Lai et al., 2025), sports science and medicine (Connor and O'Neill, 2023; Naughton et al., 2024), and the sports industry (Wang et al., 2024d), along with surveys of datasets for language and multimodal models (Xia et al., 2024b). However, these studies are still limited in scope, lacking comprehensive coverage of the diverse sports-related tasks and datasets where large models can be applied.

In this study, we used the aforementioned review papers as the starting point for multiple rounds of forward and backward snowballing on Google Scholar, using "sports" and the names of representative large models as keywords, and manually selected relevant studies. As a result, we collected more than **240** academic papers addressing large model applications in sports and provide here a comprehensive and in-depth survey.

We first systematically categorize and summarize existing applications of large models in sports across 6 key groups (§2). Then, we review and conduct an in-depth analysis of the relevant datasets

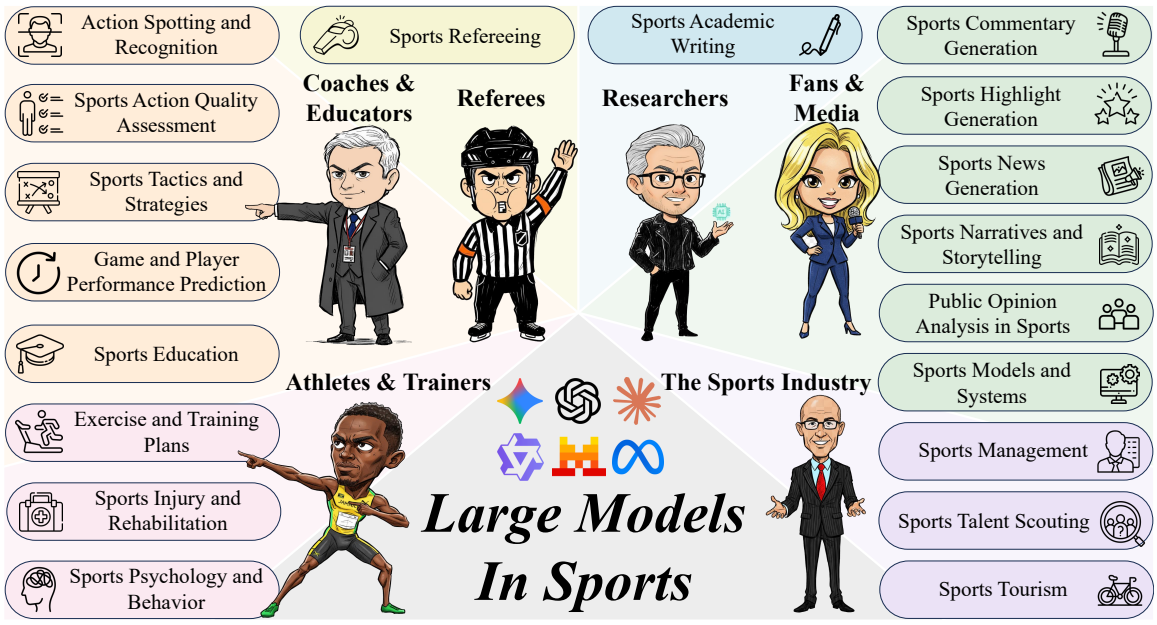


Figure 1: Large models have shown powerful applications across 6 sports stakeholder groups: athletes and trainers, coaches and educators, referees, researchers, fans and social media, and the sports industry, enabling diverse tasks.

and benchmarks (§3). Subsequently, we discuss the current challenges in this field and, finally, outline the potential opportunities and future directions (§4).

2 Large Model Applications in Sports

The fast growth of large models has brought big chances for their use in sports. This section categorizes and introduces large model applications in sports across 6 groups: athletes and trainers, coaches and educators, referees, fans and social media, researchers, and the sports industry. The taxonomy of the applications is shown in Figure 2.

2.1 Applications for Athletes and Trainers

2.1.1 Exercise and Training Plans

Large models help athletes and trainers create exercise prescriptions, translating sports science into practice to improve performance (Phillips and Kennedy, 2012; Wackerhage and Schoenfeld, 2021). Recent AI coaches powered by LLMs can make generating effective training plans easier. Many works have used LLMs to generate exercise prescriptions for various health conditions and fitness goals (Cavazzotto et al., 2024; Cosentino et al., 2024; Puce et al., 2025; Papini et al., 2025a; Masagca, 2025; Lederman et al., 2025), including weight management (Saraç et al., 2025), resistance and jump training (Washif et al., 2024; Havers et al., 2025; Pajo et al., 2025), upper body and core training (CANUZAKOV et al., 2025; Erol and Arıkan, 2024), and nutritional strategies for ultra-endurance sports (Puce et al., 2024; Solomon and Laye, 2025). LLMs can also help trainers develop fitness programs for specific patient populations, including obese people (Li et al., 2025c; Philuek et al., 2025), those with chronic diseases (Xu et al., 2024; Onan et al., 2025; Akrimi et al., 2025), and those with epilepsy (Rocha-Silva et al., 2024, 2025). In terms of methods, some studies employ digital twins with multimodal outputs (Vahdati et al., 2025), behavioral science theories (Hegde et al., 2024; Jörke et al., 2025; Dindorf et al., 2025), and Retrieval-Augmented Generation (RAG) technology (Zhang et al., 2025b; Ko et al., 2025); in terms of effectiveness, some emphasize the importance of personalization and contextual understanding (Dergaa et al., 2024; Zhu et al., 2024; Han, 2025), focus on acceptance, trust, and quality (Düking et al., 2024; Wachholz et al., 2025), and foster user self-reflection (Li et al., 2025g). In addition, some studies have designed AI coaches for specific sports such as boxing (Bullard et al., 2025) and table tennis (Ma et al., 2025a,b). Some of these require large models to provide expert-level evaluation and feedback on basketball (Seino et al., 2025) and football (Ashutosh et al., 2025) players (i.e., **Expert Commentary Generation**), while MAAIG (Yeh et al., 2023) in figure skating and CoachMe (Yeh et al., 2025) in boxing generate sports instructions automatically from 3D skeletal data (i.e., **Motion Instruction Generation**).

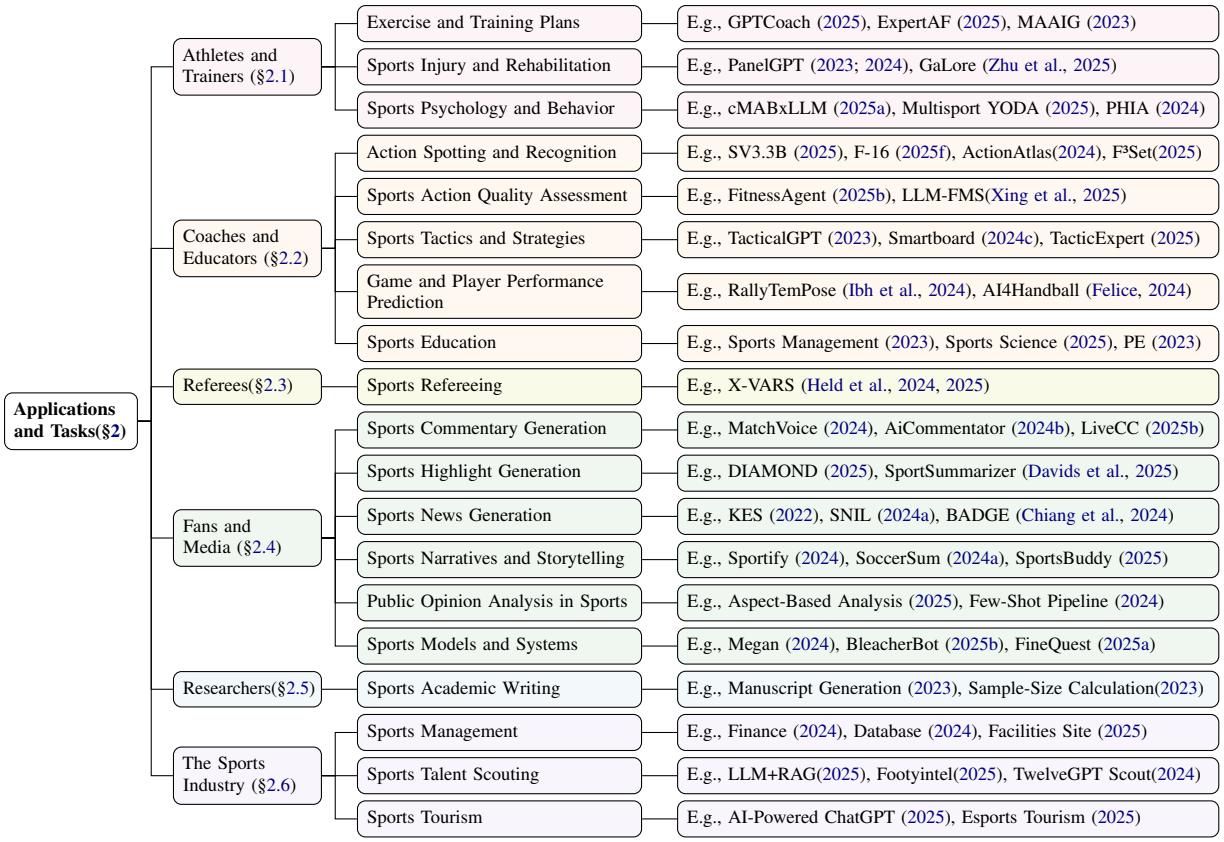


Figure 2: Taxonomy of applications, tasks, and approaches of large models in sports.

2.1.2 Sports Injury and Rehabilitation

Large models can assist athletes and trainers throughout the entire process of sports injury prevention, diagnosis, treatment, and rehabilitation. Diagnosing and treating sports injuries requires extensive interdisciplinary knowledge, and LLMs possess a broad understanding of sports injuries (Hasnain et al., 2023; Lotfi and Madani, 2024), including orthopedics (Fayed et al., 2023) and sports rehabilitation (McBee et al., 2023, 2024). Specifically, they can assist in providing preventive advice (Zhu et al., 2025), identifying and labeling information (Brogly et al., 2025), supporting diagnostic imaging (Lotfi and Madani, 2024) and data processing (Musat et al., 2024), assisting in clinical decision-making (Saglam et al., 2025) and surgical treatment (Cheng et al., 2023), and enabling patient outcome prediction (Ahsan, 2023) and medical oversight.

2.1.3 Sports Psychology and Behavior

Sports psychology, a key part of sports medicine, supports athletes through behavioral interventions to enhance training and performance. LLMs have demonstrated initial potential in this field, answering sports-related questions (Vandelanotte et al., 2023), assessing cognitive abilities (Zuccolotto, 2025), summarizing theories (Oliver and Guiller, 2025), and integrating with wearables for real-time monitoring and behavioral modeling (Merrill et al., 2024; Ferrara, 2024; Imran et al., 2024; Ji et al., 2024). LLMs can also provide assistance in specialized areas such as exercise addiction (Szabo, 2023) and sports and exercise psychology (Oliver and Guiller, 2025), and can play a role in behavioral interventions by increasing motivation to participate in sports activities (Song et al., 2025a) and providing sleep education (Masur et al., 2025).

2.2 Applications for Coaches and Educators

2.2.1 Action Spotting and Recognition

Action spotting and recognition in sports involves temporally localizing and classifying fine-grained player movements or events to provide reliable match facts for downstream analytics (Zhao et al., 2025). Most approaches employ MLLMs for direct action spotting or recognition. Some work leverage keyframe sampling (Kodathala et al., 2025), contrastive pretraining (Shin et al., 2025), or domain adaptation (Jiang et al., 2025b) for soccer. Others fine-tune MLLMs for single-action and rally-sequence recognition in tennis (Teo, 2025), while high-frame-rate modeling has been ex-

plored for gymnastics and diving (Li et al., 2025f). Scene-level classification has also been studied in rugby (Nonaka et al., 2024). Beyond multimodal inputs, textual signals such as commentary are also used for soccer action spotting (Chakraborty et al., 2025). Benchmarks like ActionAtlas (Salehi et al., 2024) and F³Set (Liu et al., 2025) evaluate MLLMs' fine-grained recognition capabilities.

2.2.2 Sports Action Quality Assessment

Sports action quality assessment evaluates the execution of athletic movements (Zhou et al., 2024), primarily to provide training feedback for coaches, and in some cases to support referees with objective scoring (Zhao et al., 2025). Recent work fine-tunes large multimodal models for personalized fitness evaluation (Dibenedetto et al., 2025), designs unified agent frameworks for open-set and user-specific assessments (Tang et al., 2025b), and builds fine-grained datasets with LLM-based evaluation for functional movement screening (Xing et al., 2025). In figure skating, assessment frameworks quantify technical and program scores, supporting both training feedback and referee judging (Wang et al., 2025a).

2.2.3 Sports Tactics and Strategies

Sports tactics and strategy analysis models and interprets on-field plays to uncover tactical structures and game dynamics, with large models automating data analysis and pattern discovery to provide coaches with strategic insights. Early work was text-centric, converting structured or non-linguistic data into natural language for tactical modeling, such as fine-tuning LLMs on event sequences (Caron and Müller, 2023), transforming cycling commentary into structured tactical variables with graph representations (Janssens et al., 2024), or parsing play-by-play logs into spatial spray charts (Michielssen et al., 2024). Other studies examine LLMs' analytical reasoning, including computing team scores from NBA/NFL play-by-play data (Hu et al., 2024c) and aggregating basketball narratives for score inference (Hu et al., 2024d). Recent approaches integrate diverse structural and spatial information for richer tactical reasoning, including sketch-based LLM agents for interactive tactic design (Liu et al., 2024c), graph LLMs for zero-shot basketball generalization (Lingrui et al., 2025), and multi-agent LLM systems combining video detection with statistical inference to support tactical decision-making (Zhang et al., 2025a).

2.2.4 Game and Player Performance Prediction

Game and player performance prediction leverages historical, contextual, or multimodal data to forecast match outcomes and individual behaviors, providing coaches with insights for strategy, preparation, and performance optimization (Xia et al., 2024b). Early studies used BERT-based (Devlin et al., 2019) models to predict player actions or traits, such as forecasting badminton strokes from skeleton poses (Ibh et al., 2024) or NBA players' performance deviations from pre-game interviews (Oved et al., 2020). More recent work leverages LLMs and diverse data sources for game outcomes prediction, including predicting basketball game results via in-context learning on social media data (Sprint, 2024), explaining handball match outcomes through feature attribution summarization (Felice, 2024), and fusing features from multimodal pre-match reports to enhance cricket score predictions (Bhatnagar and Bhatnagar, 2025).

2.2.5 Sports Education

Recent applications of large models in sports education demonstrate their diverse utility for educators and teachers. Some works apply LLMs to assist educators with lesson planning (Genç, 2023; Wang and Wang, 2024), interactive activities (Cui et al., 2025), formative feedback (Keiper et al., 2023), and autonomous assignment design (Kauppinen, 2024). Others emphasize curricular support through automated visualization, synthetic dataset creation (Fazackerley et al., 2025), and sensing-driven feedback reports (Gao et al., 2025b) . Meanwhile, student-centered applications address personalized exercise planning, engagement, and mental health support (Zhang and Liu, 2024), complemented by surveys examining inclusion, trust, and acceptance of ChatGPT in educational practice (Chang et al., 2025).

2.3 Applications for Referees

2.3.1 Sports Refereeing

Large models can also help sports refereeing by supporting decision-making and improving fairness and transparency. X-VARS (Held et al., 2024) introduces an explainable video assistant referee system by fine-tuning MLLMs on expert-annotated foul data and rationales, showing clear benefits for referees through improved decision accuracy, consistency, and trust (Held et al., 2025).

2.4 Applications for Fans and Social Media

2.4.1 Sports Commentary Generation

Sports commentary generation produces natural-language narratives that combine factual event descriptions with tactical analysis and contextual insights. Unlike video captioning, it delivers timely and emotionally resonant narration aligned with the flow of the game (Ge et al., 2024). Recent studies have leveraged LLMs to automatically produce sports commentary, offering fans an enhanced viewing experience. Most approaches adopt an agentic framework, where LLMs are prompted with extracted match information to generate commentary. Such information includes detected key events (Andrews et al., 2024b,a; Sameer et al., 2025; Pavlovich et al., 2023), player and ball tracking data (Vijayakumar et al., 2025; Andrews et al., 2024b,a), player background information (Mori et al., 2025; Xi et al., 2025b), audio signals (Gautam et al., 2022), and external knowledge (Li et al., 2025d). Beyond agentic frameworks, recent work has explored end-to-end training to improve commentary quality. Some studies fine-tune MLLMs (Wang and Yoshinaga, 2024; Jiang et al., 2025b; Cook and Karakuş, 2024; Baughman et al., 2024), while others design and train novel architectures for improved temporal alignment between videos and commentaries (Zhang et al., 2024a; Rao et al., 2024; You et al., 2025). In addition, several efforts target complementary directions, such as constructing sports commentary benchmarks (Chen et al., 2025b; Ge et al., 2024), enabling real-time streaming (Chen et al., 2025b; Ding et al., 2025; Yu and Chai, 2025), supporting multilingual commentary (Sameer et al., 2025), generating personalized commentary (Andrews et al., 2024a), and advancing commercial applications (Baughman et al., 2024).

2.4.2 Sports Highlight Generation

Sports highlight generation aims to automatically identify and compile the most significant moments from a full-length match into concise video or textual summaries. Most research employs MLLMs for highlight generation via key event detection, including action spotting (Banu et al., 2025), multimodal fusion with textual encoding (Davids et al., 2025), and event classification using commentary transcripts or role-play prompting (Sattar et al., 2023; Kang et al., 2025). Other works use MLLMs for summary and caption generation to support social media highlights (Midoglu et al., 2024), or for personalized highlight generation via simulated watch histories and preference descriptions (Lee et al., 2025).

2.4.3 Sports News Generation

Sports news generation automatically produces factual match reports suitable for publication, summarizing key outcomes, events, and statistics in a standard journalistic style. Early work focused on summarizing text commentary, either using LLMs to select and rewrite key segments (Wang et al., 2022, 2021) or to extract salient events (Sarkar et al., 2024). More recent studies extend this to structured data, employing chain-of-thought prompting for statistical tables (Chiang et al., 2025) or leveraging CSV inputs to generate game reports (Chiang et al., 2024). Beyond summarization, Cheng et al. (2024a) proposes an insight-driven approach where high-level user queries guide LLMs to construct narrative episodes enriched with data visualizations.

2.4.4 Sports Narrative and Storytelling

Sports narrative and storytelling creates long-form, multimodal stories about matches or players. It weaves together event sequences and contextual details to produce engaging narratives that captivate fans on social media. Recent research generates factually consistent highlight narrations through prompt-engineered LLMs (Sarfati et al., 2023), or leverages multimodal embedded visualizations and personalized narratives with LLMs to aid tactic understanding (Lee et al., 2024; Lin et al., 2025). Other works adapt narrative generation to platform-specific contexts, producing personalized reports and posts for large-scale fan interaction (Sarkhoosh et al., 2024a; Baughman et al., 2024).

2.4.5 Public Opinion Analysis in Sports

Public opinion and sentiment analysis in sports involves detecting, classifying, and measuring public attitudes toward sporting events or related issues. LLMs have been applied to identify key themes (Qian et al., 2024), perform fine-grained sentiment and stance detection via aspect-based analysis (Qian et al., 2025) and few-shot prompting (Rauchegger et al., 2024), and examine user acceptance of AI tools using social science frameworks (Argan and Dinç, 2025).

2.4.6 Sports Models and Systems

Unlike task-specific research, work on sports models and systems targets general-purpose capabilities and infrastructures to enhance users' overall sports experience, including (1) chatbots and models that enable interactive dialogue and general sport understanding, and (2) retrieval systems that provide precise and reliable access to sports knowledge.

Chatbots and Sports Models. LLMs have been applied to build sports chatbots for interactive dialogue (Priya et al., 2024) or co-viewing experiences (Kim et al., 2025b), with dialogue state tracking for sports-specific conversations (Song et al., 2025b). General sports models, especially in soccer, are developed via fine-tuning (Unlu, 2023; Gautam et al., 2025; Jiang et al., 2025b; Rao et al., 2025b), knowledge graphs (Chen et al., 2025a), or multi-agent LLM systems (Rao et al., 2025a).

Search Engines and Retrieval Systems. Interactive agents combine LLMs with retrieval, including search engines with offline query understanding and online decision-making (Karat et al., 2025), and RAG systems querying sports knowledge from natural language (Schilling et al., 2024; Strand et al., 2024a,b; Sepasdar et al., 2024a,b; Wang et al., 2025b). Extensions include online information retrieval for cricket (Wickramasinghe, 2025) and fine-grained sports video retrieval for gymnastics and diving (Gupta et al., 2025).

2.5 Applications for Researchers

2.5.1 Sports Academic Writing

Large models can aid sports researchers in academic writing, as shown in sports science (Latzel and Glauner, 2024) and sports medicine (Hakam et al., 2024). For instance, ChatGPT can generate outlines, abstracts, and provide grammar and style suggestions (Latzel and Glauner, 2024). However, caution is needed due to risks in content accuracy (Dergaa et al., 2023a), reference reliability (Anderson et al., 2023), calculation accuracy (Methnani et al., 2023), and originality.

2.6 Applications for the Sports Industry

2.6.1 Sports Management

Large models are increasingly applied in sports management. In financial management, they can conduct interviews, extract key themes, and develop tailored strategies (Haghparast et al., 2024); in database management, they structure and analyze club data to improve efficiency (Merilehto, 2024); and in facility management, they support human-computer dialogue to acquire site selection knowledge (Salimi Beni et al., 2025). They are also used to explore future industry scenarios and support data-driven decisions (Haghparast et al., 2025).

2.6.2 Sports Talent Scouting

Sports talent scouting helps clubs identify, evaluate, and predict player potential, and large models can make this process more objective and data-driven (Mateus et al., 2024). Recent research uses LLMs to analyze player data (Raskar et al., 2025), convert unstructured scouting reports into searchable knowledge, and combine them with RAG to optimize information integration (Martire and Ragazzi, 2025).

2.6.3 Sports Tourism

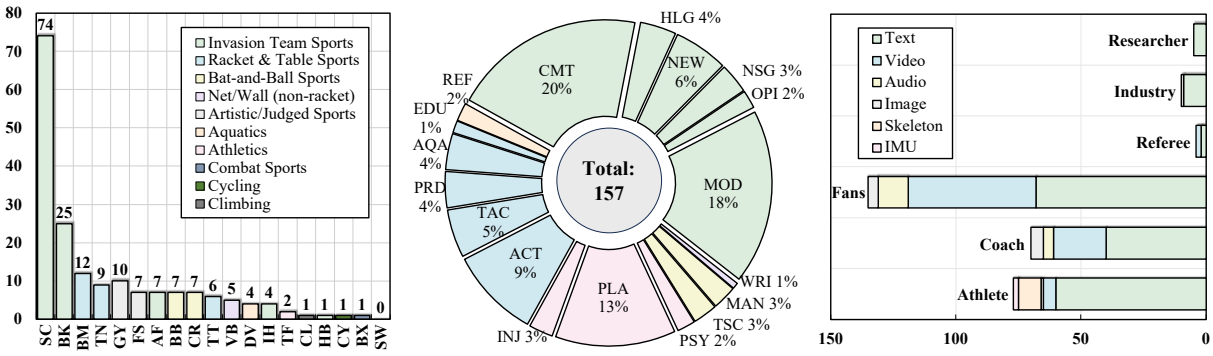
Sports and tourism are closely linked, and large models can enhance intelligence and personalization in sports tourism, offering virtual guides, information assistants, fan engagement, community building, and operational efficiency (Memon et al., 2025). They also show strong potential in esports tourism (Yenisoy and Silik, 2025).

3 Datasets for Large Models in Sports

In this section, we categorize existing datasets for large models in sports into two main types with distinct objectives and design orientations: **task-specific datasets** and **sports understanding datasets**. We will first introduce these two types of datasets and then analyze their characteristics from multiple perspectives, including sport type, task distribution, and modality composition. More details are provided in Appendix A.

3.1 Task-Specific Datasets

Task-specific datasets are developed to address concrete applications of large models across various sports contexts, corresponding to the 6 stakeholder groups and 19 tasks introduced in §2. They typically provide fine-grained annotations and evaluation metrics, supporting model training, fine-tuning, and task-specific performance assessment. Serving as a bridge between large models' general capabilities and practical sports applications, these datasets enable the development of customized systems, foster reproducible evaluation, and advance research on large model deployment in real-world sports scenarios.



(a) Analysis of sports types (b) Analysis of task distribution (c) Analysis of modality composition

Figure 3: Data analysis from 3 different perspectives. Sports and tasks abbreviations are listed in Table 5.

Benchmark	Sports	Input Modal	# Video	# QA
BIG-bench-SU (2023)	SC, BK, etc.	text	-	986
Sports-QA (2024b)	GY, VB, etc.	video, text	5967	94073
SportQA (2024a)	TN, AF, etc.	text	-	70592
SPORTU (2025)	BB, IH, etc.	video, text	1701	12948
Sports-3K-QA (2025b)	49 Sports	video, text	412	1174
FSBench (2025a)	FS	video, text	783	4000
FBBench (2025b)	BM	video, text	2563	2563
Gym-QA (2025a)	GY	video, text	6031	27469
Diving-QA (2025a)	DV	video, text	∼100	1055

Table 1: Overview of specialized sports understanding benchmarks related to large models. Sports abbreviations are listed in Table 5.

3.2 Sports Understanding Datasets

Sports are inherently fast-paced, diverse, and strategically complex, posing unique challenges to the sports understanding capabilities of large models (Xia et al., 2024a). Beyond task-specific datasets, numerous studies have developed sports understanding datasets to evaluate and improve models' general comprehension and reasoning in sports-related contexts. These datasets can be broadly classified into two categories: **(1) datasets specifically designed for sports understanding**, with Table 1 presenting an overview of relevant benchmarks; and **(2) general video understanding datasets that include sports content**, covering tasks such as video captioning (Wang et al., 2024c), multi-view understanding (Grauman et al., 2024), and fine-grained analysis (Liu et al., 2024b).

3.3 Comprehensive Analysis of Sports Datasets

We analyzed the datasets used in studies from sports-focused studies related to large models, across sport types, tasks, and modalities.

Dataset Distribution by Sport Type. Table 3a shows the number of datasets by major sport categories. The distribution is uneven, concentrated in popular sports such as invasion team sports, racket and table sports, and bat-and-ball Sports, with soccer leading at 74 datasets. Some sports, like cycling and boxing, have only one dataset, highlighting coverage imbalance. In addition, fitness-related datasets are relatively numerous (21), and esports datasets total 3, reflecting growing interest in esports.

Dataset Distribution by Task. Table 3b shows that among the 6 sports stakeholder groups, data are relatively abundant for athletes and trainers, coaches and educators, and fans and social media, reflecting stronger coverage of common sports participants. Academically, tasks such as sports commentary generation, training plan generation, and action recognition receive more attention. In contrast, data for referees, sports researchers, and the sports industry remain scarce, with researcher-relevant datasets accounting for only 1%, highlighting the need to address these gaps.

Dataset Distribution by Modality. Table 3c shows that in sports research with large-scale models, video and text are the most common modalities. Some studies also highlight the importance of audio (Xie et al., 2025). In contrast, sensor data such as IMUs and skeletal data are relatively rare, appearing only in athlete-focused studies. Exploring these less common modalities could potentially strengthen large models' understanding of sports.

4 Discussion

4.1 Challenges

Large models show great potential in sports by aiding various stakeholders. However, their adoption in sports faces challenges that need to be addressed to fully realize their benefits.

Data Coverage and Scarcity. Developing large models for sports faces significant challenges due

to data limitations. These include: *(1) narrow sports coverage*, with datasets focusing on a few major sports while others are neglected; *(2) shallow data depth*, as most data comes from elite competitions, underrepresenting Paralympics, secondary leagues, and amateur events; *(3) limited data modalities*, relying mainly on broadcast videos and textual data, while sensor-based and physiological data are rarely used; *(4) high data acquisition costs*, due to the need for expert annotations for detailed data like sports tactics, making large-scale data creation difficult; and *(5) lack of standardization*, with inconsistent annotation practices across sports, fragmenting resources and hindering fair benchmarking.

Limits of Current Modeling Capabilities. Current large models also face intrinsic limitations in sports applications. These include: *(1) hallucination and factual reliability*, with models generating inaccurate statistics (Huang et al., 2025) or misinterpreting match dynamics (Qiu, 2024; Oliver and Guiller, 2025; Hu et al., 2024d; Sepasdar et al., 2024a); *(2) insufficient long-video understanding*, as models struggle with reasoning over extended sequences (Zou et al., 2024; Wang et al., 2025c), making it hard to capture tactical developments across a match (You et al., 2025); *(3) weak generalization*, with models trained on professional leagues failing to transfer to amateur or youth sports (Qiu, 2024; Oliver and Guiller, 2025; Hu et al., 2024d; Sepasdar et al., 2024a).

Bias, Fairness, and Ethics. The deployment of large models in sports raises concerns about bias, fairness, and ethical responsibility, as flawed decisions in athlete selection, performance evaluation, medical support, and refereeing can have significant consequences. These challenges include: *(1) fairness and bias* due to imbalanced training data that can produce discriminatory outputs and amplify inequalities (Biester, 2025; Kim et al., 2025a); *(2) transparency and explainability* needed for high-stakes decisions to maintain trust (Dergaa et al., 2023b; Kim et al., 2025a); *(3) privacy and data ethics*, as sensitive data like biometric and health information lack adequate protection mechanisms (Papini et al., 2025b; Kim et al., 2025a); and *(4) accountability and responsibility*, with unclear liability among stakeholders (Kim et al., 2025a).

4.2 Future Directions

Large models have demonstrated strong capabilities in sports. Despite challenges, their deployment offers unique opportunities, and we outline future directions to address limitations and unlock their full potential.

Efficient Modeling. A key priority is developing models that provide real-time insights under strict latency constraints (Niu et al., 2025). Sports require responses during live play or halftime (Chen et al., 2025b), making lightweight edge-side LLMs/MLLMs with streaming I/O essential for fast, responsive analysis.

Expert-Level Modeling. Future models should deliver expert-level sports analysis. Tactical insights must go beyond shallow statistics, and commentary should be interactive and knowledge-rich, comparable to professional broadcasters. Achieving this requires domain priors and tool-using capabilities (Rao et al., 2025a).

Emerging Technologies and Modalities. Expanding beyond video–text pipelines, integrating sensors, wearables, and physiological data can enhance reasoning (Merrill et al., 2024; Ferrara, 2024). Robotics (Ma et al., 2025c) and mixed reality (Sawan et al., 2020; Cheng et al., 2024b) enable embodied training and immersive simulation.

Collaborative Human–AI Systems. Large models should augment human expertise, supporting interactive workflows where coaches, athletes, and fans refine tactics, prevent injuries, and enhance fan experiences (Zhu-Tian et al., 2022).

Practicability. Progress depends on deployable tools. For elite sports, collaboration with clubs, companies and media can deliver specialized analytics[1]. Extending applications to schools, community programs, and consumer fitness ensures impact from elite performance to grassroots participation (He et al., 2024; Ma and Tran-Nguyen, 2024).

5 Conclusion

This survey reviews the emerging landscape of large models in sports, summarizing applications across 6 participant groups, analyzing relevant datasets and benchmarks, and highlighting key challenges. Based on these findings, we outline future research directions and emphasize the vast opportunities in this rapidly evolving field. We hope this work serves as a foundation for advancing large-model-driven sports intelligence and provides a practical resource for research and development.

[1] https://www.hudl.com/en_gb/products/wyscout, https://www.statsperform.com/opta-analytics/

Limitations

Although this survey strives to provide a comprehensive overview of large models in sports, several limitations remain. Firstly, given the rapid development of this field, our survey may not be able to timely reflect the latest progress before and after the survey. Secondly, as some studies span multiple application domains, minor overlaps are inevitable; we categorize each work based on its primary research focus while cross-referencing related sections when appropriate. Thirdly, our analysis primarily centers on academic research, and the discussion of commercial systems or industrial applications remains limited. Despite these limitations, this survey provides a valuable and timely overview of the field, offering a solid reference for subsequent research and development.

References

Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, and 1 others. 2024. Phi-3 technical report: A highly capable language model locally on your phone. *arXiv preprint arXiv:2404.14219*.

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, and 1 others. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Mohammad Ahsan. 2023. Chatbot generative pre-trained transformer and artificial intelligence in sports physical therapy and rehabilitation. *Saudi Journal of Sports Medicine*, 23(2):61–62.

Samir Akrimi, Leon Schwensfeier, Peter Düking, Thorsten Kreutz, and Christian Brinkmann. 2025. Chatgpt-4o-generated exercise plans for patients with type 2 diabetes mellitus—assessment of their safety and other quality criteria by coaching experts. *Sports*, 13(4):92.

Nash Anderson, Daniel L Belavy, Stephen M Perle, Sharief Hendricks, Luiz Hespanhol, Evert Verhagen, and Aamir R Memon. 2023. Ai did not write this manuscript, or did it? can we trick the ai text detector into generated texts? the potential future of chatgpt and ai in sports & exercise medicine manuscript generation.

Peter Andrews, Oda Elise Nordberg, Njål Borch, Frode Guribye, and Morten Fjeld. 2024a. Designing for automated sports commentary systems. In *Proceedings of the 2024 ACM International Conference on Interactive Media Experiences*, pages 75–93.

Peter Andrews, Oda Elise Nordberg, Stephanie Zubicueta Portales, Njål Borch, Frode Guribye, Kazuyuki Fujita, and Morten Fjeld. 2024b. Aicommentator: A multimodal conversational agent for embedded visualization in football viewing. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, pages 14–34.

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge Luis Reyes-Ortiz, and 1 others. 2013. A public domain dataset for human activity recognition using smartphones. In *Esann*, volume 3, pages 3–4.

Anthropic. 2024. Claude 3 opus model card. https://www.anthropic.com/claude-3-model-card. Accessed: October 5, 2025.

Anthropic. 2025. Claude 3.5 sonnet. https://www.anthropic.com/news/claude-3-5-sonnet. Accessed: October 5, 2025.

Metin Argan and Halime Dinç. 2025. Investigating the factors influencing adoption intentions of chatgpt for sport events. *SPORMETRE Beden Eğitimi ve Spor Bilimleri Dergisi*, 23(2):77–97.

Kumar Ashutosh, Tushar Nagarajan, Georgios Pavlakos, Kris Kitani, and Kristen Grauman. 2025. Expertaf: Expert actionable feedback from video. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 13582–13594.

Kar-Weng Ban, John See, Junaidi Abdullah, and Yuen Peng Loh. 2022. Badmintondb: A badminton dataset for player-specific match analysis and prediction. In *Proceedings of the 5th international ACM workshop on multimedia content analysis in sports*, pages 47–54.

Sameena Banu and 1 others. 2025. Survey paper on ai based sports highlight generation for social media. *Journal of Scientific Research and Technology*, pages 30–38.

Aaron Baughman, Eduardo Morales, Rahul Agarwal, Gozde Akay, Rogerio Feris, Tony Johnson, Stephen Hammer, and Leonid Karlinsky. 2024. Large scale generative ai text applied to sports and music. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4784–4792.

Mohit Bhatnagar and Manya Bhatnagar. 2025. Analyzing key factors influencing ipl cricket scores using explainability and multimodal data. *Journal of Quantitative Analysis in Sports*, 21(3):253–267.

Laura Biester. 2025. Sports and women's sports: Gender bias in text generation with olympic data. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 195–205.

Sidney Black, Stella Biderman, Eric Hallahan, Quentin Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, and 1 others. 2022. Gpt-neox-20b: An open-source autoregressive language model. In *Proceedings of BigScience Episode# 5–Workshop on Challenges & Perspectives in Creating Large Language Models*, pages 95–136.

Chris Brogly, Saif Rjaibi, Charlotte Liang, Erica Lam, Edward Wang, Sarah Paleczny, Adam Levitan, and Michael D Cusimano. 2025. Evaluation of the phi-3-mini slm for identification of texts related to medicine, health, and sports injuries. In *2025 IEEE 4th International Conference on Computing and Machine Intelligence (ICMI)*, pages 1–5. IEEE.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and 1 others. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Enya Bullard, Nibraas Khan, and Nilanjan Sarkar. 2025. Enhancing athletic performance through ai: An iterative prompt engineering approach for llm-based coaching feedback. In *International Conference on Human-Computer Interaction*, pages 251–255. Springer.

Rory Bunker and Teo Susnjak. 2022. The application of machine learning techniques for predicting match results in team sport: A review. *Journal of Artificial Intelligence Research*, 73:1285–1322.

Kanat CANUZAKOV, Ciparkül ABDIRAHMANOVA, Bilal DEMİRHAN, and 1 others. 2025. Digital coaches: an alternative to expert coaches for men's fitness goals.

Matthew Caron and Oliver Müller. 2023. Tacticalgpt: uncovering the potential of llms for predicting tactical decisions in professional football. In *StatsBomb Conference*, pages 1–11.

Timothy Gustavo Cavazzotto, Diego Bessa Dantas, and Marcos Roberto Queiroga. 2024. Chatgpt and exercise prescription: Human vs. machine or human plus machine? *Journal of Sport and Health Science*, 13(5):661–662.

Ritabrata Chakraborty, Rajatsubhra Chakraborty, Avijit Dasgupta, and Sandeep Chaurasia. 2025. Do we need large vlms for spotting soccer actions? *arXiv preprint arXiv:2506.17144*.

S Chan Chang, R Walmsley, J Gershuny, T Harms, E Thomas, K Milton, P Kelly, C Foster, A Wong, N Gray, and 1 others. 2021. Capture-24: Activity tracker dataset for human activity recognition.

Shu-Hao Chang, Su-Yen Chen, and Chin-Han Chang. 2025. Exploring opportunities and challenges toward chatgpt for inclusion in sport education. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 37:100572.

Haodong Chen, Haojian Huang, XinXiang Yin, and Dian Shao. 2025a. Finequest: Adaptive knowledge-assisted sports video understanding via agent-of-thoughts reasoning. *arXiv preprint arXiv:2509.11796*.

Joya Chen, Zhaoyang Lv, Shiwei Wu, Kevin Qinghong Lin, Chenan Song, Difei Gao, Jia-Wei Liu, Ziteng Gao, Dongxing Mao, and Mike Zheng Shou. 2024a. Videollm-online: Online video large language model for streaming video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18407–18418.

Joya Chen, Ziyun Zeng, Yiqi Lin, Wei Li, Zejun Ma, and Mike Zheng Shou. 2025b. Livecc: Learning video llm with streaming speech transcription at scale. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 29083–29095.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde De Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, and 1 others. 2021. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*.

Wei-Yi Chen, Yi-Ling Lin, Yu-An Su, Wei-Hsin Yeh, and Lun-Wei Ku. 2024b. Yourskatingcoach: A figure skating video benchmark for fine-grained element analysis. *arXiv preprint arXiv:2410.20427*.

Yukang Chen, Fuzhao Xue, Dacheng Li, Qinghao Hu, Ligeng Zhu, Xiuyu Li, Yunhao Fang, Haotian Tang, Shang Yang, Zhijian Liu, and 1 others. 2025c. Longvila: Scaling long-context visual language models for long videos. In *The Thirteenth International Conference on Learning Representations*.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, and 1 others. 2024c. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24185–24198.

Kunming Cheng, Qiang Guo, Yongbin He, Yanqiu Lu, Ruijie Xie, Cheng Li, and Haiyang Wu. 2023. Artificial intelligence in sports medicine: could gpt-4 make human doctors obsolete? *Annals of Biomedical Engineering*, 51(8):1658–1662.

Liqi Cheng, Dazhen Deng, Xiao Xie, Rihong Qiu, Mingliang Xu, and Yingcai Wu. 2024a. Snil: generating sports news from insights with large language models. *IEEE Transactions on Visualization and Computer Graphics*.

Liqi Cheng, Hanze Jia, Lingyun Yu, Yihong Wu, Shuainan Ye, Dazhen Deng, Hui Zhang, Xiao Xie, and Yingcai Wu. 2024b. Viscourt: In-situ guidance for interactive tactic training in mixed reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pages 1–14.

Zesen Cheng, Sicong Leng, Hang Zhang, Yifei Xin, Xin Li, Guanzheng Chen, Yongxin Zhu, Wenqi Zhang, Ziyang Luo, Deli Zhao, and 1 others. 2024c. Videollama 2: Advancing spatial-temporal modeling and audio understanding in video-llms. *arXiv preprint arXiv:2406.07476*.

Zixu Cheng, Jian Hu, Ziquan Liu, Chenyang Si, Wei Li, and Shaogang Gong. 2025. V-star: Benchmarking video-llms on video spatio-temporal reasoning. *arXiv preprint arXiv:2503.11495*.

Shang-Hsuan Chiang, Lin-Wei Chao, Kuang-Da Wang, Chih-Chuan Wang, and Wen-Chih Peng. 2024. Badge: Badminton report generation and evaluation with llm. *arXiv preprint arXiv:2406.18116*.

Shang-Hsuan Chiang, Tsan-Tsung Yang, Kuang-Da Wang, Wei-Yao Wang, An-Zi Yen, and Wen-Chih Peng. 2025. Tree-of-report: Table-to-text generation for sports game reports with tree-structured prompting. In *ACL 2025 Student Research Workshop*.

Wei-Lin Chiang, Zhuohan Li, Ziqing Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, and 1 others. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. *See https://vicuna. lmsys. org (accessed 14 April 2023)*, 2(3):6.

Mark Connor and Michael O'Neill. 2023. Large language models in sport science & medicine: Opportunities, risks and considerations. *arXiv preprint arXiv:2305.03851*.

Alec Cook and Oktay Karakuş. 2024. Llm-commentator: Novel fine-tuning strategies of large language models for automatic commentary generation using football event data. *Knowledge-Based Systems*, 300:112219.

Justin Cosentino, Anastasiya Belyaeva, Xin Liu, Zhun Yang, Yun Liu, Shyam A Tailor, Tim Althoff, John B Hernandez, Yossi Matias, Greg Corrado, and 1 others. 2024. Towards a personal health large language model. In *Advancements In Medical Foundation Models: Explainability, Robustness, Security, and Beyond*.

Victor RA Cossich, Dave Carlgren, Robert John Holash, and Larry Katz. 2023. Technological breakthroughs in sport: Current practice and future potential of artificial intelligence, virtual reality, augmented reality, and modern data visualization in performance analysis. *Applied Sciences*, 13(23):12965.

Bin Cui, Wei Jiao, Shuying Gui, Yang Li, and Qun Fang. 2025. Innovating physical education with artificial intelligence: a potential approach. *Frontiers in Psychology*, 16:1490966.

D Minola Davids, A Arul Edwin Raj, and C Seldev Christopher. 2025. Sportsummarizer: A unified multimodal fusion transformer for context-aware sports video summarization. *Neurocomputing*, page 131011.

Adrien Deliege, Anthony Cioppa, Silvio Giancola, Meisam J Seikavandi, Jacob V Dueholm, Kamal Nasrollahi, Bernard Ghanem, Thomas B Moeslund, and Marc Van Droogenbroeck. 2021. Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4508–4519.

Ismail Dergaa, Karim Chamari, Piotr Zmijewski, and Helmi Ben Saad. 2023a. From human writing to artificial intelligence generated text: examining the prospects and potential threats of chatgpt in academic writing. *Biology of sport*, 40(2):615–622.

Ismail Dergaa, Karim Chamari, Piotr Zmijewski, and Helmi Ben Saad. 2023b. From human writing to artificial intelligence generated text: examining the prospects and potential threats of chatgpt in academic writing. *Biology of sport*, 40(2):615–622.

Ismail Dergaa, Helmi Ben Saad, Abdelfatteh El Omri, Jordan Glenn, Cain Clark, Jad Washif, Noomen Guelmami, Omar Hammouda, Ramzi Al-Horani, Luis Reynoso-Sánchez, and 1 others. 2024. Using artificial intelligence for exercise prescription in personalised health promotion: A critical evaluation of openai's gpt-4 model. *Biology of Sport*, 41(2):221–241.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Gaetano Dibenedetto, Elio Musacchio, Marco Polignano, and Pasquale Lops. 2025. Fine-tuning large multimodal models for fitness action quality assessment. In *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*, pages 39–44.

Carlo Dindorf, Jonas Dully, Eva Bartaguiz, Tessa Menges, Claudia Reidick, Johann-Nikolaus Seibert, and Michael Fröhlich. 2025. Characteristics and perceived suitability of artificial intelligence-driven sports coaches: a pilot study on psychological and perceptual factors. *Frontiers in Sports and Active Living*, 7:1548980.

Xin Ding, Hao Wu, Yifan Yang, Shiqi Jiang, Donglin Bai, Zhibo Chen, and Ting Cao. 2025. Streammind: Unlocking full frame rate streaming video dialogue through event-gated cognition. *arXiv preprint arXiv:2503.06220*.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, and 1 others. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, and 1 others. 2024. The llama 3 herd of models. *arXiv e-prints*, pages arXiv–2407.

Peter Düking, Billy Sperlich, Laura Voigt, Bas Van Hooren, Michele Zanini, and Christoph Zinner. 2024. Chatgpt generated training plans for runners are not rated optimal by coaching experts, but increase in quality with additional input information. *Journal of sports science & medicine*, 23(1):56.

Erkan Erol and Halime Arıkan. 2024. Does chatgpt provide comprehensive and accurate information regarding the effects, types and programming of core exercises? *Turkish Journal of Kinesiology*, 10(3):178–182.

Aly M Fayed, Nacime Salomao Barbachan Mansur, Kepler Alencar de Carvalho, Andrew Behrens, Pieter D'Hooghe, and Cesar de Cesar Netto. 2023. Artificial intelligence and chatgpt in orthopaedics and sports medicine. *Journal of Experimental Orthopaedics*, 10(1):74.

Lewis A Fazackerley, Dimitri Perrin, and Geoffrey M Minett. 2025. Harnessing generative ai in exercise and sports science education: enhancing real-world learning and overcoming traditional barriers in data analysis.

Florian Felice. 2024. Ai for handball: predicting and explaining the 2024 olympic games tournament with deep learning and large language models. *arXiv preprint arXiv:2407.15987*.

Emilio Ferrara. 2024. Large language models for wearable sensor-based human activity recognition, health monitoring, and behavioral modeling: A survey of early trends, datasets, and challenges. *Sensors*, 24(15):5045.

Chaoyou Fu, Yuhan Dai, Yongdong Luo, Lei Li, Shuhuai Ren, Renrui Zhang, Zihan Wang, Chenyu Zhou, Yunhang Shen, Mengdan Zhang, and 1 others. 2025. Video-mme: The first-ever comprehensive evaluation benchmark of multi-modal llms in video analysis. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 24108–24118.

Rong Gao, Xin Liu, Zhuozhao Hu, Bohao Xing, Baiqiang Xia, Zitong Yu, and Heikki Kälviäinen. 2025a. Fsbench: A figure skating benchmark for advancing artistic sports understanding. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 13595–13605.

Xian Gao, Jiacheng Ruan, Jingsheng Gao, Mingye Xie, Zongyun Zhang, Ting Liu, and Yuzhuo Fu. 2025b. From motion signals to insights: A unified framework for student behavior analysis and feedback in physical education classes. *arXiv preprint arXiv:2503.06525*.

Sushant Gautam, Cise Midoglu, Saeed Shafiee Sabet, Dinesh Baniya Kshatri, and Pål Halvorsen. 2022. Soccer game summarization using audio commentary, metadata, and captions. In *Proceedings of the 1st Workshop on User-centric Narrative Summarization of Long Videos*, pages 13–22.

Sushant Gautam, Cise Midoglu, Vajira Thambawita, Michael A Riegler, Pål Halvorsen, and Mubarak Shah. 2025. Soccerchat: Integrating multimodal data for enhanced soccer game understanding. *arXiv preprint arXiv:2505.16630*.

Kuangzhi Ge, Lingjun Chen, Kevin Zhang, Yulin Luo, Tianyu Shi, Liaoyuan Fan, Xiang Li, Guanqun Wang, and Shanghang Zhang. 2024. Scbench: A sports commentary benchmark for video llms. *arXiv preprint arXiv:2412.17637*.

Neşe Genç. 2023. Artificial intelligence in physical education and sports: New horizons with chatgpt. *Akdeniz Spor Bilimleri Dergisi*, 6(1-Cumhuriyet'in 100. Yılı Özel Sayısı):17–32.

Silvio Giancola, Mohieddine Amine, Tarek Dghaily, and Bernard Ghanem. 2018. Soccernet: A scalable dataset for action spotting in soccer videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1711–1721.

Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, and 1 others. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*.

Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh, Vijay Baiyya, Siddhant Bansal, Bikram Boote, and 1 others. 2024. Ego-exo4d: Understanding skilled human activity from first-and third-person perspectives. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19383–19400.

Animesh Gupta, Jay Parmar, Ishan Rajendrakumar Dave, and Mubarak Shah. 2025. From play to replay: Composed video retrieval for temporally fine-grained videos. *arXiv preprint arXiv:2506.05274*.

Mehran Haghparast, Mohamad Soltan Hoseini, and Davood Nasr Esfahani. 2024. A financial management maturity model in sports organizations: A novel approach using artificial intelligence. *Journal of New Studies in Sport Management*.

Mehran Haghparast, Mohammad Soltan Hoseini, and Davood Nasr Esfahani. 2025. Foresight in sports businesses: Exploring emerging scenarios based on ai-language models and financial management strategies. *Sports Business Journal*.

Hassan Tarek Hakam, Robert Prill, Lisa Korte, Bruno Lovreković, Marko Ostojić, Nikolai Ramadanov, and

Felix Muehlensiepen. 2024. Human-written vs ai-generated texts in orthopedic academic literature: comparative qualitative analysis. *JMIR formative research*, 8:e52164.

Yuxin Han. 2025. Intent-aware personalized feedback generation from coach-athlete dialogues in sports training. *Journal of King Saud University Computer and Information Sciences*, 37(6):1–17.

Muhammad Hasnain, Bilal Mehboob, and Shahid Imran. 2023. The role of chatgpt in sports trauma: A mini review on strengths and limits of open ai application. *Discover Artificial Intelligence*, 3(1):40.

Tim Havers, Lukas Masur, Eduard Isenmann, Stephan Geisler, Christoph Zinner, Billy Sperlich, and Peter Düking. 2025. Reproducibility and quality of hypertrophy-related training plans generated by gpt-4 and google gemini as evaluated by coaching experts. *Biology of Sport*, 42(2):289–329.

Quantao He, Haiping Chen, and Xiaohe Mo. 2024. Practical application of interactive ai technology based on visual analysis in professional system of physical education in universities. *Heliyon*, 10(3).

Xuehai He, Weixi Feng, Kaizhi Zheng, Yujie Lu, Wanrong Zhu, Jiachen Li, Yue Fan, Jianfeng Wang, Linjie Li, Zhengyuan Yang, and 1 others. 2025a. Mmworld: Towards multi-discipline multi-faceted world model evaluation in videos. In *The Thirteenth International Conference on Learning Representations*.

Xusheng He, Wei Liu, Shanshan Ma, Qian Liu, Chenghao Ma, and Jianlong Wu. 2025b. Finebadminton: A multi-level dataset for fine-grained badminton video understanding. *arXiv preprint arXiv:2508.07554*.

Yuping He, Yifei Huang, Guo Chen, Baoqi Pei, Jilan Xu, Tong Lu, and Jiangmiao Pang. 2025c. Egoexobench: A benchmark for first-and third-person view video understanding in mllms. *arXiv preprint arXiv:2507.18342*.

Narayan Hegde, Madhurima Vardhan, Deepak Nathani, Emily Rosenzweig, Cathy Speed, Alan Karthikesalingam, and Martin Seneviratne. 2024. Infusing behavior science into large language models for activity coaching. *PLOS Digital Health*, 3(4):e0000431.

Jan Held, Anthony Cioppa, Silvio Giancola, Elaf Almahmoud, Katherine M Collins, Umang Bhatt, Bernard Ghanem, and Marc Van Droogenbroeck. 2025. Enhancing football refereeing with {AI}:{VARS} and {X-VARS} for assisted decision-making. In *MathSport Conference*.

Jan Held, Anthony Cioppa, Silvio Giancola, Abdullah Hamdi, Bernard Ghanem, and Marc Van Droogenbroeck. 2023. Vars: Video assistant referee system for automated soccer decision making from multiple views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5086–5097.

Jan Held, Hani Itani, Anthony Cioppa, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. 2024. X-vars: Introducing explainability in football refereeing with multi-modal large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3267–3279.

Jack Hong, Shilin Yan, Jiayin Cai, Xiaolong Jiang, Yao Hu, and Weidi Xie. 2025a. Worldsense: Evaluating real-world omnimodal understanding for multimodal llms. *arXiv preprint arXiv:2502.04326*.

Wenyi Hong, Yean Cheng, Zhuoyi Yang, Weihan Wang, Lefan Wang, Xiaotao Gu, Shiyu Huang, Yuxiao Dong, and Jie Tang. 2025b. Motionbench: Benchmarking and improving fine-grained video motion understanding for vision language models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 8450–8460.

Shiyu Hu, Xuchen Li, Xuzhao Li, Jing Zhang, Yipei Wang, Xin Zhao, and Kang Hao Cheong. 2024a. Fiova: A multi-annotator benchmark for human-aligned video captioning. *arXiv preprint arXiv:2410.15270*.

Yebowen Hu, Kaiqiang Song, Sangwoo Cho, Xiaoyang Wang, Hassan Foroosh, Dong Yu, and Fei Liu. 2024b. Can large language models do analytical reasoning? *arXiv preprint arXiv:2403.04031*.

Yebowen Hu, Kaiqiang Song, Sangwoo Cho, Xiaoyang Wang, Hassan Foroosh, Dong Yu, and Fei Liu. 2024c. SportsMetrics: Blending text and numerical data to understand information fusion in LLMs. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 267–278, Bangkok, Thailand. Association for Computational Linguistics.

Yebowen Hu, Kaiqiang Song, Sangwoo Cho, Xiaoyang Wang, Wenlin Yao, Hassan Foroosh, Dong Yu, and Fei Liu. 2024d. When reasoning meets information aggregation: A case study with sports narratives. *arXiv preprint arXiv:2406.12084*.

Kuan-Hao Huang, Chen Li, and Kai-Wei Chang. 2020. Generating sports news from live commentary: A chinese dataset for sports game summarization. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 609–615.

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and 1 others. 2025. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55.

Hudl. 2024. Wyscout: Football data and analytics platform. https://wyscout.com. Accessed: October 5, 2025.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Brett Hutchins. 2016. Tales of the digital sublime: Tracing the relationship between big data and professional sport. *Convergence*, 22(5):494–509.

Magnus Ibh, Stella Graßhof, and Dan Witzner Hansen. 2024. A stroke of genius: Predicting the next move in badminton. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3376–3385.

Sheikh Asif Imran, Mohammad Nur Hossain Khan, Subrata Biswas, and Bashima Islam. 2024. Llasa: Large multimodal agent for human activity analysis through wearable sensors. *arXiv preprint arXiv:2406.14498*, 3(4).

Bram Janssens, Matthias Bogaert, and Steven Verstockt. 2024. Large language models on race commentary: Towards granular data in cycling analytics. In *International Workshop on Machine Learning and Data Mining for Sports Analytics*, pages 14–25. Springer.

Pedro Calciolari Jardim, Leonardo Mauro Pereira Moraes, and Cristina Dutra Aguiar. 2023. Qasports: A question answering dataset about sports. In *Dataset Showcase Workshop (DSW)*, pages 1–12. SBC.

Sijie Ji, Xinzhe Zheng, and Chenshu Wu. 2024. Hargpt: Are llms zero-shot human activity recognizers? In *2024 IEEE International Workshop on Foundation Models for Cyber-Physical Systems & Internet of Things (FMSys)*, pages 38–43. IEEE.

Yanli Ji, Lingfeng Ye, Huili Huang, Lijing Mao, Yang Zhou, and Lingling Gao. 2023. Localization-assisted uncertainty score disentanglement network for action quality assessment. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 8590–8597.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.

Dongzhi Jiang, Renrui Zhang, Ziyu Guo, Yanmin Wu, Pengshuo Qiu, Pan Lu, Zehui Chen, Guanglu Song, Peng Gao, Yu Liu, and 1 others. 2025a. Mmsearch: Unveiling the potential of large models as multimodal search engines. In *The Thirteenth International Conference on Learning Representations*.

Tiancheng Jiang, Henry Wang, Md Sirajus Salekin, Parmida Atighehchian, and Shinan Zhang. 2025b. Domain adaptation of vlm for soccer video understanding. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 6111–6121.

Matthew Jörke, Shardul Sapkota, Lyndsea Warkenthien, Niklas Vainio, Paul Schmiedmayer, Emma Brunskill, and James A Landay. 2025. Gptcoach: Towards llm-based physical activity coaching. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–46.

Jeonghun Kang, Soonmok Kwon, Joonseok Lee, and Byung-Hak Kim. 2025. Diamond: An llm-driven agent for context-aware baseball highlight summarization. *arXiv preprint arXiv:2506.02351*.

Ankith Karat, Atishay Tibrewal, Nishka Kotian, Manan Dang, Ravindra Valluri, Antony Ravi Teja Marineni, Sarthak Sahni, Rhea Sundaresan, Ankit Kumar, Aditya Mehndiratta, and 1 others. 2025. A system for triggering sports instant answers on search engines. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 4304–4308.

Antti Kauppinen. 2024. Proactive autonomous assignments as pedagogical responses to the rise of artificial intelligence solutions in sport management teaching practice. *Sport Management Education Journal*, 19(1):54–58.

Margaret C Keiper, Gil Fried, Joshua Lupinek, and Heidi Nordstrom. 2023. Artificial intelligence in sport management education: Playing the ai game with chatgpt. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 33:100456.

Jae-Hak Kim, Janghyeon Kim, Hoseok Kang, and Bo-Young Youn. 2025a. Ethical implications of artificial intelligence in sport: A systematic scoping review. *Journal of Sport and Health Science*, page 101047.

Kyusik Kim, Hyungwoo Song, Jeongwoo Ryu, Changhoon Oh, and Bongwon Suh. 2025b. Bleacherbot: Ai agent as a sports co-viewing partner. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–31.

Kangbeen Ko, Minwoo Oh, Minwoo Seong, and SeungJun Kim. 2025. Legolas: Learning & enhancing golf skills through llm-augmented system. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–10.

Sai Varun Kodathala, Yashwanth Reddy Vutukoori, and Rakesh Vunnam. 2025. Sv3. 3b: A sports video understanding model for action recognition. *arXiv preprint arXiv:2507.17844*.

Fanheng Kong, Jingyuan Zhang, Hongzhi Zhang, Shi Feng, Daling Wang, Linhao Yu, Xingguang Ji, Yu Tian, V. W., and Fuzheng Zhang. 2025a. TUNA: Comprehensive fine-grained temporal understanding evaluation on dense dynamic videos. In *Proceedings of the 63rd Annual Meeting of the Association for*

Computational Linguistics (Volume 1: Long Papers), pages 1810–1839, Vienna, Austria. Association for Computational Linguistics.

Fanqi Kong, Weiqin Zu, Xinyu Chen, Yaodong Yang, Song-Chun Zhu, and Xue Feng. 2025b. Siv-bench: A video benchmark for social interaction understanding and reasoning. *arXiv preprint arXiv:2506.05425*.

Xiangxun Lai, Jiacheng Chen, Yue Lai, Shengqi Huang, Yongdong Cai, Zhifeng Sun, Xueding Wang, Kaijiang Pan, Qi Gao, and Caihua Huang. 2025. Using large language models to enhance exercise recommendations and physical activity in clinical and healthy populations: Scoping review. *JMIR Medical Informatics*, 13(1):e59309.

Richard Latzel and Patrick Glauner. 2024. Artificial intelligence in sport scientific creation and writing process. In *Artificial Intelligence in Sports, Movement, and Health*, pages 15–29. Springer.

Oscar Lederman, Alessandro Llana, James Murray, Robert Stanton, Ritesh Chugh, Darren Haywood, Amanda Burdett, Geoff Warman, Joanne Walker, and Nicolas H Hart. 2025. Promises and perils of generative artificial intelligence: a narrative review informing its ethical and practical applications in clinical exercise physiology. *BMC Sports Science, Medicine and Rehabilitation*, 17(1):131.

Chunggi Lee, Tica Lin, Hanspeter Pfister, and Chen Zhu-Tian. 2024. Sportify: question answering with embedded visualizations and personified narratives for sports video. *IEEE Transactions on Visualization and Computer Graphics*.

Jeongeun Lee, Youngjae Yu, and Dongha Lee. 2025. Hippo-video: Simulating watch histories with large language models for history-driven video highlighting. In *Second Conference on Language Modeling*.

Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Peiyuan Zhang, Yanwei Li, Ziwei Liu, and 1 others. 2025a. Llava-onevision: Easy visual task transfer. *Transactions on Machine Learning Research*.

Chaoyu Li, Sid Padmanabhuni, Maryam S Cheema, Hasti Seifi, and Pooyan Fazli. 2025b. Videoa11y: Method and dataset for accessible video description. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–29.

Feng Li, Renrui Zhang, Hao Zhang, Yuanhan Zhang, Bo Li, Wei Li, Zejun Ma, and Chunyuan Li. 2024a. Llava-next-interleave: Tackling multi-image, video, and 3d in large multimodal models. *arXiv preprint arXiv:2407.07895*.

Guanchong Li, Hansen Li, Yuqin Su, Yun Li, Sijia Jiang, and Guodong Zhang. 2025c. Gpt-4 as a virtual fitness coach: a case study assessing its effectiveness in providing weight loss and fitness guidance. *BMC Public Health*, 25(1):2466.

Haopeng Li, Andong Deng, Qiuhong Ke, Jun Liu, Hossein Rahmani, Yulan Guo, Bernt Schiele, and Chen Chen. 2024b. Sports-qa: A large-scale video question answering benchmark for complex and professional sports. *arXiv preprint arXiv:2401.01505*.

Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*, pages 12888–12900. PMLR.

Xiang Li, Yangfan He, Shuaishuai Zu, Zhengyang Li, Tianyu Shi, Yiting Xie, and Kevin Zhang. 2025d. Multi-modal large language model with rag strategies in soccer commentary generation. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6197–6206. IEEE.

Xuchen Li, Xuzhao Li, Shiyu Hu, Kaiqi Huang, and Wentao Zhang. 2025e. Causalstep: A benchmark for explicit stepwise causal reasoning in videos. *arXiv preprint arXiv:2507.16878*.

Yixuan Li, Changli Tang, Jimin Zhuang, Yudong Yang, Guangzhi Sun, Wei Li, Zejun Ma, and Chao Zhang. 2025f. Improving llm video understanding with 16 frames per second. *arXiv preprint arXiv:2503.13956*.

Yunxin Li, Xinyu Chen, Baotian Hu, Longyue Wang, Haoyuan Shi, and Min Zhang. 2024c. Videovista: A versatile benchmark for video understanding and reasoning. *arXiv preprint arXiv:2406.11303*.

Zong-Ying Li, Yen-Hua Lai, Chi-Yu Lin, Chien-Hsing Chou, and Ping-Hsuan Han. 2025g. Visualizing exercise data from combat exergame for exploring the insight from personal informatics with large language models. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–8.

Bin Lin, Yang Ye, Bin Zhu, Jiaxi Cui, Munan Ning, Peng Jin, and Li Yuan. 2024a. Video-llava: Learning united visual representation by alignment before projection. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5971–5984.

Tica Lin, Ruxun Xiang, Gardenia Liu, Divyanshu Tiwari, Meng-Chia Chiang, Chenjiayi Ye, Hanspeter Pfister, and Chen Zhu-Tian. 2025. Sportsbuddy: Designing and evaluating an ai-powered sports video storytelling tool through real-world deployment. In *2025 IEEE 18th Pacific Visualization Conference (PacificVis)*, pages 214–223. IEEE.

Xudong Lin, Ali Zare, Shiyuan Huang, Ming-Hsuan Yang, Shih-Fu Chang, and Li Zhang. 2024b. Personalized video comment generation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 16806–16820.

Xu Lingrui, Liu Mandi, and Zhang Lei. 2025. Tactic-expert: Spatial-temporal graph language model for basketball tactics. *arXiv preprint arXiv:2503.10722*.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, and 1 others. 2024a. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916.

Ye Liu, Zongyang Ma, Zhongang Qi, Yang Wu, Ying Shan, and Chang W Chen. 2024b. Et bench: Towards open-ended event-level video-language understanding. *Advances in Neural Information Processing Systems*, 37:32076–32110.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Zhaoyu Liu, Kan Jiang, Murong Ma, Zhe Hou, Yun Lin, and Jin Song Dong. 2025. F3set: Towards analyzing fast, frequent, and fine-grained events from videos. *arXiv preprint arXiv:2504.08222*.

Ziao Liu, Xiao Xie, Moqi He, Wenshuo Zhao, Yihong Wu, Liqi Cheng, Hui Zhang, and Yingcai Wu. 2024c. Smartboard: Visual exploration of team tactics with llm agent. *IEEE Transactions on Visualization and Computer Graphics*.

S. LLC. 2025. Sportvu. https://www.stats.com/sportvu-basketball/. Accessed: October 5, 2025.

Nizar Lotfi and Mohamed Madani. 2024. Evaluating the qualitative and quantitative performance of generative ai on knowledge in sports medicine: The case of gpt. In *General Aspects of Applying Generative AI in Higher Education: Opportunities and Challenges*, pages 103–119. Springer.

C. Love. 2015. Bbc football commentary data – webscraping. https://sciolisticramblings.wordpress.com/2015/08/24/bbc-football-commentary-data-webscraping/. Accessed: October 5, 2025.

Ji Ma, Jiale Wu, Haoyu Wang, Yanze Zhang, Xiao Xie, Zheng Zhou, Hui Zhang, Jiachen Wang, and Yingcai Wu. 2025a. T3set: A multimodal dataset with targeted suggestions for llm-based virtual coach in table tennis training. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 5686–5697.

Thanh Ma and Minh-Thu Tran-Nguyen. 2024. Petsai-ext: physical education teaching support with artificial intelligence. *SN Computer Science*, 5(7):855.

Wenlong Ma, Yang Liu, Qing Yi, Xutao Liu, Wei Xing, Rongji Zhao, Huan Liu, and Rongzhi Li. 2025b. Table tennis coaching system based on a multimodal large language model with a table tennis knowledge base. *PloS one*, 20(2):e0317839.

Yuntao Ma, Andrei Cramariuc, Farbod Farshidian, and Marco Hutter. 2025c. Learning coordinated badminton skills for legged manipulators. *Science Robotics*, 10(102):eadu3922.

Muhammad Maaz, Hanoona Rasheed, Salman Khan, and Fahad Khan. 2024. Video-chatgpt: Towards detailed video understanding via large vision and language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12585–12602.

Surbhi Madan, Shreya Ghosh, Lownish Rai Sookha, MA Ganaie, Ramanathan Subramanian, Abhinav Dhall, and Tom Gedeon. 2025. Mip-gaf: A mllm-annotated benchmark for most important person localization and group context understanding. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1467–1476. IEEE.

Felice Antonio Martire and Davide Ragazzi. 2025. Leveraging llms and rag for enhanced football talent scouting. In *International Conference on Advanced Information Systems Engineering*, pages 298–309. Springer.

Ramon Carlo Masagca. 2025. The ai coach: A 5-week ai-generated calisthenics training program on health-related physical fitness components of untrained collegiate students. *Journal of Human Sport and Exercise*, 20(1):39–56.

Lukas Masur, Matthew Driller, Haresh Suppiah, Manuel Matzka, Billy Sperlich, and Peter Düking. 2025. Assessment of recommendations provided to athletes regarding sleep education by gpt-4o and google gemini: Comparative evaluation study. *JMIR Formative Research*, 9(1):e71358.

Nuno Mateus, Eduardo Abade, Diogo Coutinho, Miguel-Ángel Gómez, Carlos Lago Peñas, and Jaime Sampaio. 2024. Empowering the sports scientist with artificial intelligence in training, performance, and health management. *Sensors*, 25(1):139.

Joseph C McBee, Daniel Y Han, Li Liu, Leah Ma, Donald A Adjeroh, Dong Xu, and Gangqing Hu. 2023. Interdisciplinary inquiry via panelgpt: application to explore chatbot application in sports rehabilitation. *medRxiv*.

Joseph C McBee, Daniel Y Han, Li Liu, Leah Ma, Donald A Adjeroh, Dong Xu, and Gangqing Hu. 2024. Assessing chatgpt's competency in addressing interdisciplinary inquiries on chatbot uses in sports rehabilitation: simulation study. *JMIR Medical Education*, 10(1):e51157.

Salman Bashir Memon, Jawaid Ahmed Qureshi, and Samar Batool Shah. 2025. Ai-powered chatgpt in sports tourism: Benefits, challenges, and future prospects. *Redefining Tourism With AI and the Metaverse*, pages 163–188.

Juhani Merilehto. 2024. From pdfs to structured data: Utilizing llm analysis in sports database management. *arXiv preprint arXiv:2410.17619*.

Mike A Merrill, Akshay Paruchuri, Naghmeh Rezaei, Geza Kovacs, Javier Perez, Yun Liu, Erik Schenck, Nova Hammerquist, Jake Sunshine, Shyam Tailor, and 1 others. 2024. Transforming wearable data into health insights using large language model agents. *arXiv preprint arXiv:2406.06464*.

Jabeur Methnani, Imed Latiri, Ismail Dergaa, Karim Chamari, and Helmi Ben Saad. 2023. Chatgpt for sample-size calculation in sports medicine and exercise sciences: A cautionary note. *International Journal of Sports Physiology and Performance*, 18(10):1219–1223.

Senne Michielssen, Adam Maloof, Joe Haumacher, Alexander Dreger, Kyle Bonicki, and Karl Hallgren. 2024. Using large language models to generate baseball spray charts in the absence of numerical data. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, page 17543371241257734.

Cise Midoglu, Saeed Shafiee Sabet, Mehdi Houshmand Sarkhoosh, Mohammad Majidi, Sushant Gautam, Håkon Maric Solberg, Tomas Kupka, and Pål Halvorsen. 2024. Ai-based sports highlight generation for social media. In *Proceedings of the 3rd Mile-High Video Conference*, pages 7–13.

Hassan Mkhallati, Anthony Cioppa, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. 2023. Soccernet-caption: Dense video captioning for soccer broadcasts commentaries. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5074–5085.

Yuichiro Mori, Chikara Tanaka, Aru Maekawa, Satoshi Kosugi, Tatsuya Ishigaki, Kotaro Funakoshi, Hiroya Takamura, and Manabu Okumura. 2025. Live football commentary system providing background information. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 394–404.

Carmina Liana Musat, Claudiu Mereuta, Aurel Nechita, Dana Tutunaru, Andreea Elena Voipan, Daniel Voipan, Elena Mereuta, Tudor Vladimir Gurau, Gabriela Gurău, and Luiza Camelia Nechita. 2024. Diagnostic applications of ai in sports: a comprehensive review of injury risk prediction methods. *Diagnostics*, 14(22):2516.

Arsha Nagrani, Sachit Menon, Ahmet Iscen, Shyamal Buch, Ramin Mehran, Nilpa Jha, Anja Hauth, Yukun Zhu, Carl Vondrick, Mikhail Sirotenko, and 1 others. 2025. Minerva: Evaluating complex video reasoning. *arXiv preprint arXiv:2505.00681*.

Arsha Nagrani, Mingda Zhang, Ramin Mehran, Rachel Hornung, Nitesh Bharadwaj Gundavarapu, Nilpa Jha, Austin Myers, Xingyi Zhou, Boqing Gong, Cordelia Schmid, and 1 others. 2024. Neptune: The long orbit to benchmarking long video understanding. *arXiv preprint arXiv:2412.09582*.

Banoth Thulasya Naik, Mohammad Farukh Hashmi, and Neeraj Dhanraj Bokde. 2022. A comprehensive review of computer vision in sports: Open issues, future trends and research directions. *Applied Sciences*, 12(9):4429.

National Strength and Conditioning Association (NSCA). 2025. Certified strength and conditioning specialist (cscs) exam description. https://www.nsca.com/certification/cscs/certified-strength-and-conditioning-specialist-exam-des Accessed: October 5, 2025.

Mitchell Naughton, Paul M Salmon, Heidi R Compton, and Scott McLean. 2024. Challenges and opportunities of artificial intelligence implementation within sports science and sports medicine teams. *Frontiers in Sports and Active Living*, 6:1332427.

Junbo Niu, Yifei Li, Ziyang Miao, Chunjiang Ge, Yuanhang Zhou, Qihao He, Xiaoyi Dong, Haodong Duan, Shuangrui Ding, Rui Qian, and 1 others. 2025. Ovobench: How far is your video-llms from real-world online video understanding? In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 18902–18913.

Naoki Nonaka, Ryo Fujihira, Toshiki Koshiba, Akira Maeda, and Jun Seita. 2024. Rugby scene classification enhanced by vision language model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3256–3266.

Alex Oliver and Jane Guiller. 2025. Generative ai in sport and exercise psychology: exploring opportunities and overcoming challenges. *Sport and Exercise Psychology Review*, 19(2):36–45.

Dilara Onan, Halime Arıkan, İrem Can, Şahan Güven, Levent Işıkay, and Aynur Ozge. 2025. Examining the ability of artificial intelligence with chatgpt-4.0 to create an exercise program: Case scenario examples" lumbar disc herniation, chronic migraine, and urge urinary incontinence". *Turkish Journal of Kinesiology*, 11(1):28–44.

OpenAI. 2024. Gpt-4v(ision) system card. https://cdn.openai.com/papers/GPTV_System_Card.pdf?utm_source=chatgpt.com. Accessed: October 6, 2025.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Nadav Oved, Amir Feder, and Roi Reichart. 2020. Predicting in-game actions from interviews of nba players. *Computational Linguistics*, 46(3):667–712.

Leonard Sydrick Pajo, Rey Rabuya, Arvin Andacao, Arianne Michael Sim Tuano, and Joseph Lobo. 2025. A 10-week large language model-generated (llm) versus human-made volleyball training program on the jumping performance of collegiate volleyball athletes. *Journal of Physical Education*, 36:e3611.

Natalie M Papini, Megan Meyer, Nikole D Squires, and Dawn Clifford. 2025a. Balancing act: Generative ai tools and scope of practice in health coaching. *American Journal of Health Promotion*, page 08901171251340383.

Natalie M Papini, Megan Meyer, Nikole D Squires, and Dawn Clifford. 2025b. Balancing act: Generative ai tools and scope of practice in health coaching. *American Journal of Health Promotion*, page 08901171251340383.

Paritosh Parmar, Amol Gharat, and Helge Rhodin. 2022. Domain knowledge-informed self-supervised representations for workout form assessment. In *European Conference on Computer Vision*, pages 105–123. Springer.

S. Patel. 2022. An api client package to access the apis for nba.com. https://github.com/swar/nba_api. Accessed: October 5, 2023.

Roman V Pavlovich, Evgeniya A Tsybulko, Konstantin N Zhigunov, Aleksandr V Khelvas, Aleksandr A Gilya-Zetinov, and Illya V Tykhonov. 2023. Soccer artificial intelligence commentary service on the base of video analytic and large language models. In *2023 31st Telecommunications Forum (TELFOR)*, pages 1–4. IEEE.

B. Petti and S. Gilani. 2021. baseballr: The sports dataverse's r package for baseball data. https://billpetti.github.io/baseballr/. Accessed: October 5, 2025.

Edward M Phillips and Mary A Kennedy. 2012. The exercise prescription: a tool to improve physical activity. *Pm&r*, 4(11):818–825.

Pakpoom Philuek, Sarawut Kusump, Thanakan Sathianpoonsook, Charee Jansupom, Pongtawat Sawanyawisuth, Kittisak Sawanyawisuth, and Austtasit Chainarong. 2025. The effects of chat gpt generated exercise program in healthy overweight young adults: A pilot study. *Journal of Human Sport and Exercise*, 20(1):169–179.

M Yagnasri Priya, Samrudhi S Kamble, Sanjivani P Shendre, Swetha Sridhar, and 1 others. 2024. Megana sports chatbot using openai apis and django framework with python. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pages 1–8. IEEE.

Luca Puce, Nicola Luigi Bragazzi, Antonio Curra, and Carlo Trompetto. 2025. Harnessing generative artificial intelligence for exercise and training prescription: Applications and implications in sports and physical activity—a systematic literature review. *Applied Sciences (2076-3417)*, 15(7).

Luca Puce, Halil İbrahim Ceylan, Carlo Trompetto, Filippo Cotellessa, Cristina Schenone, Lucio Marinelli, Piotr Zmijewski, Nicola Bragazzi, and Laura Mori. 2024. Optimizing athletic performance through advanced nutrition strategies: can ai and digital platforms have a role in ultraendurance sports? *Biology of Sport*, 41(4):305–313.

Ratish Puduppully, Li Dong, and Mirella Lapata. 2019. Data-to-text generation with entity modeling. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2023–2035.

Tyreal Yizhou Qian, Weizhe Li, Hua Gong, Chad Seifried, and Chenglong Xu. 2025. Experience is all you need: a large language model application of fine-tuned gpt-3.5 and roberta for aspect-based sentiment analysis of college football stadium reviews. *Sport Management Review*, 28(1):1–25.

Tyreal Yizhou Qian, Bo Yu, Weizhe Li, and Chenglong Xu. 2024. Esports debut as a medal event at 2023 asian games: Exploring public perceptions with bertopic and gpt-4 topic fine-tuning. *arXiv preprint arXiv:2409.18798*.

Yepeng Qiu. 2024. The impact of llm hallucinations on motor skill learning: A case study in badminton. *IEEE Access*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Jiayuan Rao, Zifeng Li, Haoning Wu, Ya Zhang, Yanfeng Wang, and Weidi Xie. 2025a. Multi-agent system for comprehensive soccer understanding. *arXiv preprint arXiv:2505.03735*.

Jiayuan Rao, Haoning Wu, Hao Jiang, Ya Zhang, Yanfeng Wang, and Weidi Xie. 2025b. Towards universal soccer video understanding. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 8384–8394.

Jiayuan Rao, Haoning Wu, Chang Liu, Yanfeng Wang, and Weidi Xie. 2024. Matchtime: Towards automatic soccer game commentary generation. *arXiv preprint arXiv:2406.18530*.

Sandeep Raskar, Manas Thosar, Atharva Dandge, and Pranav Fale. 2025. Footyintel: Creating an ai scout for better talent recognition. *International Journal of Environmental Sciences*, pages 99–106.

Christoph Rauchegger, Sonja Mei Wang, and Pieter Delobelle. 2024. Onelove beyond the field–a few-shot pipeline for topic and sentiment analysis during the fifa world cup in qatar. *arXiv preprint arXiv:2408.02520*.

Weiming Ren, Huan Yang, Jie Min, Cong Wei, and Wenhu Chen. 2025. Vista: Enhancing long-duration and high-resolution video understanding by video spatiotemporal augmentation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 3804–3814.

Rizia Rocha-Silva, Braulio Evangelista de Lima, Thalles Guilarducci Costa, Naiane Silva Morais, Geovana Jose, Douglas Farias Cordeiro, Alexandre Aparecido de Almeida, Glauber Menezes Lopim, Ricardo Borges Viana, Bolivar Saldanha Sousa, and 1 others. 2025. Can people with epilepsy trust ai chatbots for information on physical exercise? *Epilepsy & Behavior*, 163:110193.

Rizia Rocha-Silva, Bráulio Evangelista de Lima, Geovana José, Douglas Farias Cordeiro, Ricardo Borges Viana, Marília Santos Andrade, Rodrigo Luiz Vancini, Thomas Rosemann, Katja Weiss, Beat Knechtle, and 1 others. 2024. The potential of large language model chatbots for application to epilepsy: let's talk about physical exercise. *Epilepsy & Behavior Reports*, 27:100692.

Sönmez Saglam, Veysel Uludag, Zekeriya Okan Karaduman, Mehmet Arıcan, Mücahid Osman Yücel, and Raşit Emin Dalaslan. 2025. Comparative evaluation of artificial intelligence models gpt-4 and gpt-3.5 in clinical decision-making in sports surgery and physiotherapy: a cross-sectional study. *BMC Medical Informatics and Decision Making*, 25(1):163.

Mohammadreza Reza Salehi, Jae Sung Park, Aditya Kusupati, Ranjay Krishna, Yejin Choi, Hanna Hajishirzi, and Ali Farhadi. 2024. Actionatlas: A videoqa benchmark for domain-specialized action recognition. *Advances in Neural Information Processing Systems*, 37:137372–137402.

Elham Salimi Beni, Mina Mostahfezian, Majid Khorvash, and Davood Nasr Esfahani. 2025. Comprehensive site selection model for sports facilities in iran: Leveraging ai language models. *Sport Management Journal*.

N Md Sameer, K Jayavardhan, and Oviya Ramalakshmi Iyyappan. 2025. Enhanced cricket commentary using ai vision and multilingual translation. In *2025 IEEE International Conference on Emerging Technologies and Applications (MPSec ICETA)*, pages 1–6. IEEE.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Hakan Saraç, İsmet Tarık Ulusoy, Janset Alpay, Hasan Ödemiş, and Mustafa Söğüt. 2025. Evaluating the potential role of ai chatbots in designing personalized exercise programs for weight management. *International Journal of Human–Computer Interaction*, pages 1–8.

Noah Sarfati, Ido Yerushalmy, Michael Chertok, and Yosi Keller. 2023. Generating factually consistent sport highlights narrations. In *Proceedings of the 6th International Workshop on Multimedia Content Analysis in Sports*, pages 15–22.

Soham Sarkar, Tadisetty Sai Yashwanth, and Animesh Giri. 2024. Advancing cricket narratives: Ai-enhanced advanced journaling in the ipl using language models. In *2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pages 1–6. IEEE.

Mehdi Houshmand Sarkhoosh, Sushant Gautam, Cise Midoglu, Saeed Shafiee Sabet, and Pål Halvorsen. 2024a. Multimodal ai-based summarization and storytelling for soccer on social media. In *Proceedings of the 15th ACM multimedia systems conference*, pages 485–491.

Mehdi Houshmand Sarkhoosh, Sushant Gautam, Cise Midoglu, Saeed Shafiee Sabet, Thomas Torjusen, and Pål Halvorsen. 2024b. The soccersum dataset for automated detection, segmentation, and tracking of objects on the soccer pitch. In *Proceedings of the 15th ACM Multimedia Systems Conference*, pages 353–359.

Husnain Sattar, Muhammad Shamil Umar, Eeman Ijaz, and Muhammad Umair Arshad. 2023. Multi-modal architecture for cricket highlights generation: Using computer vision and large language model. In *2023 17th International Conference on Open Source Systems and Technologies (ICOSST)*, pages 1–6. IEEE.

Nedal Sawan, Ahmed Eltweri, Caterina De Lucia, Luigi Pio Leonardo Cavaliere, Alessio Faccia, and Narcisa Roxana Moşteanu. 2020. Mixed and augmented reality applications in the sport industry. In *Proceedings of the 2020 2nd International Conference on E-Business and E-commerce Engineering*, pages 55–59.

Alexander Schilling, James Anurathan, Johannes Mühlberger, Felix Gerschner, Manfred Rössle, Andreas Theissler, and Marco Klaiber. 2024. Querying football matches for event data: Towards using large language models. In *International Sports Analytics Conference and Exhibition*, pages 216–227. Springer.

A. Secareanu. 2023. Football events. https://www.kaggle.com/datasets/secareanualin/football-events. Accessed: October 5, 2025.

Tatsuki Seino, Naoki Saito, Takahiro Ogawa, Satoshi Asamizu, and Miki Haseyama. 2025. Expert comment generation considering sports skill level using a large multimodal model with video and spatial-temporal motion features. *Sensors*, 25(2):447.

Zahra Sepasdar, Sushant Gautam, Cise Midoglu, Michael A Riegler, and Pål Halvorsen. 2024a. Enhancing structured-data retrieval with graphrag: Soccer data case study. *arXiv preprint arXiv:2409.17580*.

Zahra Sepasdar, Sushant Gautam, Cise Midoglu, Michael A Riegler, and Pål Halvorsen. 2024b.

Soccer-graphrag: Applications of graphrag in soccer. In *International Workshop on Graph-Based Approaches in Information Retrieval*, pages 1–10. Springer.

Dian Shao, Yue Zhao, Bo Dai, and Dahua Lin. 2020. Finegym: A hierarchical video dataset for fine-grained action understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2616–2625.

Yoonho Shin, Sanghoon Park, Youngsub Han, Byoung-Ki Jeon, Soonyoung Lee, and Byung Jun Kang. 2025. Soccer-clip: Vision language model for soccer action spotting. *IEEE Access*, 13:44354–44365.

Tomas Skerik, Lukas Chrpa, Wolfgang Faber, and Mauro Vallati. 2018. Automated training plan generation for athletes. In *2018 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 3865–3870. IEEE.

Thomas PJ Solomon and Matthew J Laye. 2025. The sports nutrition knowledge of large language model (llm) artificial intelligence (ai) chatbots: An assessment of accuracy, completeness, clarity, quality of evidence, and test-retest reliability. *PloS one*, 20(6):e0325982.

Haochen Song, Dominik Hofer, Rania Islambouli, Laura Hawkins, Ananya Bhattacharjee, Meredith Franklin, and Joseph Jay Williams. 2025a. Investigating the relationship between physical activity and tailored behavior change messaging: Connecting contextual bandit with large language models. *arXiv preprint arXiv:2506.07275*.

Sangmin Song, Juhyoung Park, Juhwan Choi, Junho Lee, Kyohoon Jin, and YoungBin Kim. 2025b. Korean football in-game conversation state tracking dataset for dialogue and turn level evaluation. *Engineering Applications of Artificial Intelligence*, 139:109572.

SportDevs. 2025. Handball api. https://sportdevs.com/handball. Accessed: October 5, 2025.

Sportsvision. 2025. Nsva subset: Basketball video-text dataset. https://huggingface.co/datasets/sportsvision/nsva_subset. Hugging Face.

Gina Sprint. 2024. Social networks and large language models for division i basketball game winner prediction. *IEEE Access*, 12:84774–84784.

Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Shoeb, Abubakar Abid, Adam Fisch, Adam R Brown, Adam Santoro, Aditya Gupta, Adri Garriga-Alonso, and 1 others. 2023. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on machine learning research*.

Aleksander Theo Strand, Sushant Gautam, Cise Midoglu, and Pål Halvorsen. 2024a. Soccer information retrieval via natural queries using soccerrag. In *2024 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–5. IEEE.

Aleksander Theo Strand, Sushant Gautam, Cise Midoglu, and Pål Halvorsen. 2024b. Soccerrag: Multimodal soccer information retrieval via natural queries. In *2024 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–7. IEEE.

Attila Szabo. 2023. Chatgpt is a breakthrough in science and education but fails a test in sports and exercise psychology. *Baltic Journal of Sport and Health Sciences*, 1(128):25–40.

Changli Tang, Yixuan Li, Yudong Yang, Jimin Zhuang, Guangzhi Sun, Wei Li, Zejun Ma, and Chao Zhang. 2025a. video-salmonn 2: Captioning-enhanced audio-visual large language models. *arXiv preprint arXiv:2506.15220*.

Changli Tang, Yixuan Li, Yudong Yang, Jimin Zhuang, Guangzhi Sun, Wei Li, Zujun Ma, and Chao Zhang. 2024. Enhancing multimodal llm for detailed and accurate video captioning using multi-round preference optimization. *arXiv preprint arXiv:2410.06682*.

Zhenhui Tang, Jiahao Li, Ping Guo, Bowen Tian, Qingjun Xing, Xuyang Xing, and Peng Wang. 2025b. Fitnessagent: A unified agent framework for open-set and personalized fitness evaluation. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 12437–12444. IEEE.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, and 1 others. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Mistral AI Team. 2024. Mistral-large-instruct-2411: A 123b parameter dense language model. https://huggingface.co/mistralai/Mistral-Large-Instruct-2411. Accessed: October 5, 2025.

Qwen Team and 1 others. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2:3.

Charlton Teo. 2025. Enhancing sports strategy with video analytics and data mining: Assessing the effectiveness of multimodal llms in tennis video analysis. *arXiv preprint arXiv:2507.02904*.

Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, and 1 others. 2022. Lamda: Language models for dialog applications. *arXiv preprint arXiv:2201.08239*.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, and 1 others. 2023a. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, and 1 others. 2023b. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Eren Unlu. 2023. Footgpt: A large language model development experiment on a minimal setting. *arXiv preprint arXiv:2308.08610*.

Monica Vahdati, Kamran Gholizadeh HamlAbadi, Fedwa Laamarti, and Abdulmotaleb El Saddik. 2025. A multi-agent digital twin framework for ai-driven fitness coaching. In *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences*, pages 380–385.

Corneel Vandelanotte, Stewart Trost, Danya Hodgetts, Tasadduq Imam, Mamunur Rashid, Quyen G To, and Carol Maher. 2023. Increasing physical activity using an just-in-time adaptive digital assistant supported by machine learning: a novel approach for hyper-personalised mhealth interventions. *Journal of Biomedical Informatics*, 144:104435.

Madhurima Vardhan, Narayan Hegde, Srujana Merugu, Shantanu Prabhat, Deepak Nathani, Martin Seneviratne, Nur Muhammad, Pranay Reddy, Sriram Lakshminarasimhan, Rahul Singh, and 1 others. 2022. Walking with pace-personalized and automated coaching engine. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, pages 57–68.

Adarsh Vijayakumar, Amal Toms, and Senthil Vadivu. 2025. Player tracking-integrated soccer game commentary generation. *IJSAT-International Journal on Science and Technology*, 16(2).

Felix Wachholz, Stefano Manno, Daniel Schlachter, Nicole Gamper, and Martin Schnitzer. 2025. Acceptance and trust in ai-generated exercise plans among recreational athletes and quality evaluation by experienced coaches: a pilot study. *BMC Research Notes*, 18(1):112.

Henning Wackerhage and Brad J Schoenfeld. 2021. Personalized, evidence-informed training plans and exercise prescriptions for performance, fitness and health. *Sports Medicine*, 51(9):1805–1813.

Fengshun Wang, Qiurui Wang, and Dan Chen. 2025a. From beats to scores: A multi-modal framework for comprehensive figure skating assessment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5905–5914.

Henry Wang, Sirajus Salekin, Jake Lee, Ross Claytor, Shinan Zhang, and Michael Chi. 2025b. Agentic generative ai for media content discovery at the national football league.

Jiaan Wang, Zhixu Li, Qiang Yang, Jianfeng Qu, Zhigang Chen, Qingsheng Liu, and Guoping Hu. 2021. Sportssum2. 0: Generating high-quality sports news from live text commentary. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3463–3467.

Jiaan Wang, Zhixu Li, Tingyi Zhang, Duo Zheng, Jianfeng Qu, An Liu, Lei Zhao, and Zhigang Chen. 2022. Knowledge enhanced sports game summarization. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 1045–1053.

Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, and 1 others. 2024a. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191*.

Wei-Yao Wang, Yung-Chang Huang, Tsi-Ui Ik, and Wen-Chih Peng. 2023. Shuttleset: A human-annotated stroke-level singles dataset for badminton tactical analysis. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5126–5136.

Weihan Wang, Zehai He, Wenyi Hong, Yean Cheng, Xiaohan Zhang, Ji Qi, Xiaotao Gu, Shiyu Huang, Bin Xu, Yuxiao Dong, and 1 others. 2024b. Lvbench: An extreme long video understanding benchmark. *arXiv preprint arXiv:2406.08035*.

Ye Wang, Ziheng Wang, Boshen Xu, Yang Du, Kejun Lin, Zihan Xiao, Zihao Yue, Jianzhong Ju, Liang Zhang, Dingyi Yang, and 1 others. 2025c. Time-r1: Post-training large vision language model for temporal video grounding. *arXiv preprint arXiv:2503.13377*.

Yi Wang, Yinan He, Yizhuo Li, Kunchang Li, Jiashuo Yu, Xin Ma, Xinhao Li, Guo Chen, Xinyuan Chen, Yaohui Wang, and 1 others. 2024c. Internvid: A large-scale video-text dataset for multimodal understanding and generation. In *The Twelfth International Conference on Learning Representations*.

Youze Wang, Zijun Chen, Ruoyu Chen, Shishen Gu, Wenbo Hu, Jiayang Liu, Yinpeng Dong, Hang Su, Jun Zhu, Meng Wang, and 1 others. 2025d. Understanding and benchmarking the trustworthiness in multimodal llms for video understanding. *arXiv preprint arXiv:2506.12336*.

Yuping Wang and Xinyan Wang. 2024. Artificial intelligence in physical education: comprehensive review and future teacher training strategies. *Frontiers in public health*, 12:1484848.

Zefeng Wang, Yueke Hu, Jiaxi Liu, and Lianxin Hu. 2024d. Impact of chatgpt technology on sports industry. *Journal of New Media and Economics*, 1(4):29–37.

Zihan Wang and Naoki Yoshinaga. 2024. Commentary generation from data records of multiplayer strategy esports game. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for*

Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop), pages 263–271.

Jad Washif, Jeffrey Pagaduan, Carl James, Ismail Dergaa, and Christopher Beaven. 2024. Artificial intelligence in sport: Exploring the potential of using chatgpt in resistance training prescription. *Biology of sport*, 41(2):209–220.

Indika Wickramasinghe. 2025. Assessing the accuracy of large language models in extracting latest cricket information. *Scientific Journal of Sport and Performance*, 4(2):268–284.

Sam Wiseman, Stuart M Shieber, and Alexander M Rush. 2017. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2253–2263.

Zeyu Xi, Ge Shi, Xuefen Li, Junchi Yan, Zun Li, Lifang Wu, Zilin Liu, and Liang Wang. 2025a. A simple yet effective knowledge guided method for entity-aware video captioning on a basketball benchmark. *Neurocomputing*, 619:129177.

Zeyu Xi, Haoying Sun, Yaofei Wu, Junchi Yan, Haoran Zhang, Lifang Wu, Liang Wang, and Changwen Chen. 2025b. Player-centric multimodal prompt generation for large language model based identity-aware basketball video captioning. *arXiv preprint arXiv:2507.20163*.

Haotian Xia, Zhengbang Yang, Yuqing Wang, Rhys Tracy, Yun Zhao, Dongdong Huang, Zezhi Chen, Yan Zhu, Yuan-Fang Wang, and Weining Shen. 2024a. Sportqa: A benchmark for sports understanding in large language models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 5061–5081.

Haotian Xia, Zhengbang Yang, Yun Zhao, Yuqing Wang, Jingxi Li, Rhys Tracy, Zhuangdi Zhu, Yuanfang Wang, Hanjie Chen, and Weining Shen. 2024b. Language and multimodal models in sports: a survey of datasets and applications. *arXiv preprint arXiv:2406.12252*.

Haotian Xia, Zhengbang Yang, Junbo Zou, Rhys Tracy, Yuqing Wang, Chi Lu, Christopher Lai, Yanjun He, Xun Shao, Zhuoqing Xie, and 1 others. 2025. Sportu: A comprehensive sports understanding benchmark for multimodal large language models. In *The Thirteenth International Conference on Learning Representations*.

Jingfei Xia, Mingchen Zhuge, Tiantian Geng, Shun Fan, Yuantai Wei, Zhenyu He, and Feng Zheng. 2023. Skating-mixer: Long-term sport audio-visual modeling with mlps. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2901–2909.

Liuyue Xie, George Z Wei, Avik Kuthiala, Ce Zheng, Ananya Bal, Mosam Dabhi, Liting Wen, Taru Rustagi, Ethan Lai, Sushil Khyalia, and 1 others. 2025. Maverix: Multimodal audio-visual evaluation reasoning index. *arXiv preprint arXiv:2503.21699*.

Qingjun Xing, Xuyang Xing, Ping Guo, Zhenhui Tang, and Yanfei Shen. 2025. Llm-fms: A fine-grained dataset for functional movement screen action quality assessment. *PloS one*, 20(3):e0313707.

Chengming Xu, Yanwei Fu, Bing Zhang, Zitian Chen, Yu-Gang Jiang, and Xiangyang Xue. 2019. Learning to score figure skating sport videos. *IEEE transactions on circuits and systems for video technology*, 30(12):4578–4590.

Jinglin Xu, Yongming Rao, Xumin Yu, Guangyi Chen, Jie Zhou, and Jiwen Lu. 2022. Finediving: A fine-grained dataset for procedure-aware action quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2949–2958.

Yang Xu, Qiankun Liu, Jiaxue Pang, Chunlu Zeng, Xiaoqing Ma, Pengyao Li, Li Ma, Juju Huang, and Hui Xie. 2024. Assessment of personalized exercise prescriptions issued by chatgpt 4.0 and intelligent health promotion systems for patients with hypertension comorbidities based on the transtheoretical model: A comparative analysis. *Journal of Multidisciplinary Healthcare*, pages 5063–5078.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mt5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498.

Shuhang Xun, Sicheng Tao, Jungang Li, Yibo Shi, Zhixin Lin, Zhanhui Zhu, Yibo Yan, Hanqian Li, Linghao Zhang, Shikang Wang, and 1 others. 2025. Rtv-bench: Benchmarking mllm continuous perception, understanding and reasoning through real-time video. *arXiv preprint arXiv:2505.02064*.

Songyuan Yang, Weijiang Yu, Wenjing Yang, Xinwang Liu, Huibin Tan, Long Lan, and Nong Xiao. 2025a. Wildvideo: Benchmarking lmms for understanding video-language interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhoufaran Yang, Yan Shu, Zhifei Yang, Yan Zhang, Yu Li, Keyang Lu, Gangyan Zeng, Shaohui Liu, Yu Zhou, and Nicu Sebe. 2025b. Vidtext: Towards comprehensive evaluation for video text understanding. *arXiv preprint arXiv:2505.22810*.

Wei-Hsin Yeh, Pei Hsin Lin, Yu-An Su, Wen Hsiang Cheng, and Lun-Wei Ku. 2023. Maaig: Motion analysis and instruction generation. In *Proceedings of the 5th ACM International Conference on Multimedia in Asia Workshops*, pages 1–5.

Wei-Hsin Yeh, Yu-An Su, Chih-Ning Chen, Yi-Hsueh Lin, Calvin Ku, Wen-Hsin Chiu, Min-Chun Hu, and Lun-Wei Ku. 2025. Coachme: Decoding sport elements with a reference-based coaching instruction generation model. *arXiv preprint arXiv:2509.11698*.

Caner Yenisoy and Cemal Ersin Silik. 2025. Investigating esports tourism research using artificial intelligence applications: Chatgpt versus zekai. *Tourism and Recreation*, 7(1):54–68.

Han Yi, Yulu Pan, Feihong He, Xinyu Liu, Benjamin Zhang, Oluwatumininu Oguntola, and Gedas Bertasius. 2025. Exact: A video-language benchmark for expert action analysis. *arXiv preprint arXiv:2506.06277*.

Ling You, Wenxuan Huang, Xinni Xie, Xiangyi Wei, Bangyan Li, Shaohui Lin, Yang Li, and Changbo Wang. 2025. Timesoccer: An end-to-end multimodal large language model for soccer commentary generation. *arXiv preprint arXiv:2504.17365*.

Jiashuo Yu, Yue Wu, Meng Chu, Zhifei Ren, Zizheng Huang, Pei Chu, Ruijie Zhang, Yinan He, Qirui Li, Songze Li, and 1 others. 2025. Vrbench: A benchmark for multi-step reasoning in long narrative videos. *arXiv preprint arXiv:2506.10857*.

Keunwoo Peter Yu and Joyce Chai. 2025. Temporally-grounded language generation: A benchmark for real-time vision-language models. *arXiv preprint arXiv:2505.11326*.

Huaying Yuan, Jian Ni, Zheng Liu, Yueze Wang, Junjie Zhou, Zhengyang Liang, Bo Zhao, Zhao Cao, Zhicheng Dou, and Ji-Rong Wen. 2025. Momentseeker: A task-oriented benchmark for long-video moment retrieval. *arXiv preprint arXiv:2502.12558*.

Gabriel Van Zandycke. 2024. Deepsport dataset. https://www.kaggle.com/datasets/gabrielvanzandycke/deepsport-dataset. Accessed: October 5, 2025.

Benhui Zhang, Junyu Gao, and Yuan Yuan. 2024a. A descriptive basketball highlight dataset for automatic commentary generation. In *Proceedings of the 32nd ACM international conference on multimedia*, pages 10316–10325.

Hang Zhang, Xin Li, and Lidong Bing. 2023. Video-llama: An instruction-tuned audio-visual language model for video understanding. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 543–553.

Haotian Zhang, Cristobal Sciutto, Maneesh Agrawala, and Kayvon Fatahalian. 2021. Vid2player: Controllable video sprites that behave and appear like professional tennis players. *ACM Transactions on Graphics (TOG)*, 40(3):1–16.

Jiawen Zhang, Dongliang Han, Shuai Han, Heng Li, Wing-Kai Lam, and Mingyu Zhang. 2025a. Chatmatch: Exploring the potential of hybrid vision–language deep learning approach for the intelligent analysis and inference of racket sports. *Computer Speech & Language*, 89:101694.

Kexin Zhang, Yulin Qin, and Baosheng Qin. 2025b. Rag-llm based evaluation pathway and technological exploration for the scientific validity of mass fitness. In *Of Papers Presented at 2025 6th Asia Sport Science Conference (ASSC)*.

Sheng Zhang, Xin Zhang, Hui Wang, Lixiang Guo, and Shanshan Liu. 2018. Multi-scale attentive interaction networks for chinese medical question answer selection. *IEEE Access*, 6:74061–74071.

Yi-Fan Zhang and Xin-Qiao Liu. 2024. Using chatgpt to promote college students' participation in physical activities and its effect on mental health. *World Journal of Psychiatry*, 14(2):330.

YiFan Zhang, Shanglin Lei, Runqi Qiao, Zhuoma GongQue, Xiaoshuai Song, Guanting Dong, Qiuna Tan, Zhe Wei, Peiqing Yang, Ye Tian, and 1 others. 2024b. Multi-dimensional insights: Benchmarking real-world personalization in large multimodal models. *arXiv preprint arXiv:2412.12606*.

Yuanhan Zhang, Jinming Wu, Wei Li, Bo Li, Zejun Ma, Ziwei Liu, and Chunyuan Li. 2025c. Llava-video: Video instruction tuning with synthetic data. *Transactions on Machine Learning Research*.

Zhonghan Zhao, Wenhao Chai, Shengyu Hao, Wenhao Hu, Guanhong Wang, Shidong Cao, Mingli Song, Jenq-Neng Hwang, and Gaoang Wang. 2025. A survey of deep learning in sports applications: Perception, comprehension, and decision. *IEEE Transactions on Visualization and Computer Graphics*.

Diwei Zhou, Justin WL Keogh, Yingliang Ma, Raymond KY Tong, Abdul R Khan, and Nicholas R Jennings. 2025a. Artificial intelligence in sport: A narrative review of applications, challenges and future trends. *Journal of Sports Sciences*, pages 1–16.

Junjie Zhou, Yan Shu, Bo Zhao, Boya Wu, Zhengyang Liang, Shitao Xiao, Minghao Qin, Xi Yang, Yongping Xiong, Bo Zhang, and 1 others. 2025b. Mlvu: Benchmarking multi-task long video understanding. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 13691–13701.

Kanglei Zhou, Ruizhi Cai, Liyuan Wang, Hubert PH Shum, and Xiaohui Liang. 2024. A comprehensive survey of action quality assessment: Method and benchmark. *arXiv preprint arXiv:2412.11149*.

Zitang Zhou, Ke Mei, Yu Lu, Tianyi Wang, and Fengyun Rao. 2025c. Harmonyset: A comprehensive dataset for understanding video-music semantic alignment and temporal synchronization. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 3152–3162.

Weimo Zhu, Wenguang Geng, Lingling Huang, Xiong Qin, Zezhao Chen, and Hai Yan. 2024. Who could and should give exercise prescription: Physicians, exercise and health scientists, fitness trainers, or chatgpt? *Journal of Sport and Health Science*, 13(3):368–372.

Xinli Zhu, Zhiqiang Gao, and Xu An Wang. 2025. Full-parameter fine-tuning method of llms for sports injury prevention and treatment. *International Journal of Mobile Computing and Multimedia Communications (IJMCMC)*, 16(1):1–14.

Chen Zhu-Tian, Qisen Yang, Xiao Xie, Johanna Beyer, Haijun Xia, Yingcai Wu, and Hanspeter Pfister. 2022. Sporthesia: Augmenting sports videos using natural language. *IEEE transactions on visualization and computer graphics*, 29(1):918–928.

Heqing Zou, Tianze Luo, Guiyang Xie, Fengmao Lv, Guangcong Wang, Junyang Chen, Zhuochen Wang, Hansheng Zhang, Huaijian Zhang, and 1 others. 2024. From seconds to hours: Reviewing multimodal large language models on comprehensive long video understanding. *arXiv preprint arXiv:2409.18938*.

Paola Zuccolotto. 2025. 11 th mathsport international conference 4-6 june 2025.

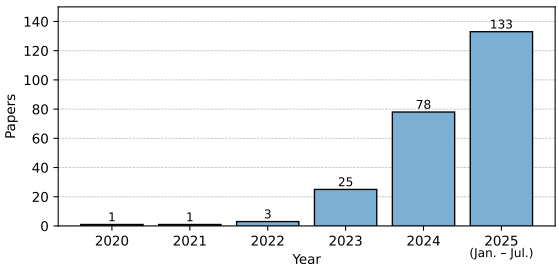


Figure 4: Papers on large models in sports over the years (data for 2025 is up to July).

A More Details on Datasets for Large Models in Sports

In this section, we further provide a comprehensive introduction to the datasets for large models in sports, as an extension of the main discussion in Section 3.

A.1 Task-Specific Datasets

This subsection echoes Section 3.1 of the paper and provides a further overview of these task-specific datasets. Table 2 and Table 3 present the datasets related to specific tasks for 5 sports stakeholder groups: athletes and trainers, coaches and educators, referees, fans and social media, and the sports industry. The information covers dataset names, involved sports types, data modalities, methods used in the papers, corresponding large models, best achieved performance, and links that can be accessed directly.

These datasets are unevenly distributed across tasks. From the perspective of target users, datasets for coaches and fans are relatively abundant, while those for referees, researchers, and the sports industry are relatively scarce. In terms of task types, early tasks in traditional computer vision and natural language processing, such as action spotting and recognition and sports commentary generation, have received more research attention and have richer datasets, whereas tasks like public opinion analysis in sports and sports talent scouting lack open-source data, reflecting an imbalance in scholarly focus across different tasks.

These datasets are unevenly distributed across sports. Popular sports such as soccer, basketball, and badminton receive more attention and have richer datasets, whereas niche sports like track and field, aquatics, and even esports are severely underrepresented. Nevertheless, these underexplored areas hold research value and warrant further expansion and investigation.

These datasets are unevenly distributed across modalities. Video and text are the most common modalities in sports datasets, while audio, sensor data (e.g., IMU), and skeletal data are relatively scarce. This reflects the current research focus on video in the sports domain and also highlights the untapped potential of other modalities.

These datasets are generally used with pre-existing models rather than being used to train or fine-tune models. Most researchers tend to rely on the inherent capabilities of large models, which explains the widespread use of powerful closed-source models such as GPT-4 (Achiam et al., 2023). This trend reflects both the scarcity of sports data and the significant value of constructing dedicated sports datasets and models, emphasizing the need for more attention to the field of large models in sports.

A.2 Sports Understanding Datasets

This subsection echoes Section 3.2 and provides a more detailed overview of datasets related to sports understanding tasks for large models. We cover datasets specifically designed for sports understanding with large models (§A.2.1), general video understanding datasets that include sports content (§A.2.2), and other general-purpose datasets containing sports-related data (§A.2.3). Table 4 presents a comprehensive summary of the first two categories of datasets from multiple perspectives, including dataset name, covered sports types, data source, benchmark availability, input modalities, number and average duration of videos, number of QA pairs, and the availability of open-source links, which can be directly accessed by clicking in the table.

A.2.1 Specialized Sports Understanding Datasets

Recently, numerous datasets have been developed to evaluate and enhance the general sports understanding capabilities of large models.

For LLMs, QASports (Jardim et al., 2023) introduced the first large-scale sports question-answering dataset with rich contextual information and diverse questions for model training and evaluation. The sports understanding subtask in BIG-bench (Srivastava et al., 2023) includes 986 binary-choice questions, primarily testing models' general

understanding of sports activities. SportQA (Xia et al., 2024a) comprises over 70,000 multiple-choice questions across three difficulty levels, enabling a comprehensive evaluation of LLMs' performance in sports understanding. SPORTU-text (Xia et al., 2025) and FSBench-Text (Gao et al., 2025a) assess models' understanding of rules, events, and scenarios in 5 major sports and figure skating, respectively.

For MLLMs, Sports-QA (Li et al., 2024b) is the first dataset specifically designed for sports video question answering, advancing the evaluation of multimodal models in sports video understanding. SPORTU-video (Xia et al., 2025) covers 7 sports and provides systematic video understanding tasks across three difficulty levels, while Sports-3K-QA (Chen et al., 2025b) includes a broader range of 49 different sports.. FSAnno (Gao et al., 2025a) constructs a large-scale, multi-task, multimodal figure skating dataset, while FSBench-Motion (Gao et al., 2025a) extends it by adding motion data and QA pairs, supporting tasks ranging from single-action analysis to full-performance commentary. FineBadminton (He et al., 2025b) is a large-scale badminton video dataset with fine-grained annotations, on which FBBench (He et al., 2025b) evaluates models' fine-grained sports video understanding. Gym-QA and Diving-QA (Chen et al., 2025a), built upon FineGym (Shao et al., 2020) and FineDiving (Xu et al., 2022), respectively, offer new benchmarks for sports video question answering in gymnastics and diving.

A.2.2 General Video Understanding Datasets Featuring Sports

In addition to datasets specifically designed for sports understanding, many general video understanding datasets also include sports content, in which sports constitute an important component.

In addition to large-scale, multi-task, and comprehensive video understanding datasets (Fu et al., 2025; He et al., 2025a; Yang et al., 2025a), some focus on specific capabilities. For example, InternVid (Wang et al., 2024c), FIOVA (Hu et al., 2024a), and VidText (Yang et al., 2025b) are primarily used for video description or subtitle generation, while Ego-Exo4D (Grauman et al., 2024) and EgoExoBench (He et al., 2025c) focus on video understanding from different viewpoints. LVBench (Wang et al., 2024b), MLVU (Zhou et al., 2025b), Neptune (Nagrani et al., 2024), LongVILA_sft (Chen et al., 2025c), and VR-Bench (Yu et al., 2025) are dedicated to long video understanding, while E.T. Bench (Liu et al., 2024b), MotionBench (Hong et al., 2025b), and ExAct (Yi et al., 2025) are used for fine-grained action, skill, or motion understanding. TUNA (Kong et al., 2025a) and VideoA11y-40K (Li et al., 2025b) emphasize temporal information and dynamic video understanding, while VideoVista (Li et al., 2024c), V-STaR (Cheng et al., 2025), MINERVA (Nagrani et al., 2025), VRBench (Yu et al., 2025), and CausalStep (Li et al., 2025e) target video reasoning tasks such as temporal-spatial, multi-step, and causal reasoning.

Furthermore, video-SALMONN-2 (Tang et al., 2024, 2025a), WorldSense (Hong et al., 2025a), HarmonySet (Zhou et al., 2025c), and MAVERIX (Xie et al., 2025) focus on joint understanding of audio and video, demonstrating multimodal capabilities; while OVO-Bench (Niu et al., 2025) and RTV-Bench (Xun et al., 2025) examine the real-time processing capabilities of models. In terms of model capability evaluation, Trust-videoLLMs (Wang et al., 2025d) is used to evaluate the credibility of video understanding, while SIV-Bench (Kong et al., 2025b) studies the understanding of social interaction behaviors in videos.

A.2.3 Other General Datasets Featuring Sports

Moreover, some other types of general datasets also contain sports content. To evaluate the image understanding capabilities of large models, MDI-Benchmark (Zhang et al., 2024b) collected 514 real images and 1,298 question-answer pairs to test basic perception and complex reasoning, and designed sports-related questions for different age groups. MIP-GAF (Madan et al., 2025) constructed a large dataset to examine the understanding of key figures in images, which also includes sports scenes. Furthermore, to assess the ability of large models as multimodal search engines, MMSearch (Jiang et al., 2025a) collected 300 unimodal and multimodal samples, and MomentSeeker (Yuan et al., 2025) constructed a dataset consisting of 268 long videos with an average length of over 1,200 seconds, all of which focus on sports scenes.

Task	Dataset	Sports	Modal	Method	Related Large Model	Performance	Link
				Athletes and Trainers			
PLA	YourSkatingCoach (2024b)	Figure Skating	V, T	MAAIG (2023)	T5 (2020)	22.08 (METEOR)	✗
	PACE (2022)	Fitness	T	Hegde et al. (2024)	LaMDA (2022)	3.78 ± 1.00 (Likert Scale)	✓
	NSCA-CSCS (2025)	Fitness	T	PH-LLM (2024)	Gemini Ultra 1.0 (2023)	88.00 (Acc)	✗
	T3Set (2025a)	Table Tennis	V, M, T	SenseCoach (2025a)	Llama 3.3-70B (2024)	51.64 (P@6-S1L)	✓
	SCD (2025)	Soccer	T	Han (2025)	BERT (2019)	85.64 (BERTScore)	✗
	Custom Dataset (2025b)	Table Tennis	V, I, S, T	Ma et al. (2025b)	GPT-4 (2023)	67.40 (Acc)	✓
	Ego-Exo4D (2024)	SC, BK, CL	V, T	ExpertAF (2025)	Llama 3-8B (2024)	49.60 (METEOR)	✓
	Ego-Exo4D (2024)	Basketball	V, T	Seino et al. (2025)	GPT-4o (2024)	25.60 (METEOR)	✗
	FS (2025)	Figure Skating	S, T	CoachMe (2025)	T5 (2020)	26.5 (BERTScore)	✓
	BX (2025)	Boxing				36.9 (BERTScore)	
INJ	cMedQA2 (2018)	-	T	Zhu et al. (2025)	Qwen2-0.5B (2024)	30.56 (BLEU-4)	✗
	Custom Dataset (2025)	-	T	Saglam et al. (2025)	GPT-4 (2023)	47.80 (Cronbach's α)	✗
	Custom Dataset (2025)	-	T	Brogly et al. (2025)	phi-3-mini (2024)	34.13 (Spearman)	✗
PSY	Custom Dataset (2024)	Fitness	T	PHIA (2024)	Gemini 1.0 Ultra (2023)	84.20 (Acc)	✓
	Capture24 (2021)	Fitness	M	HARGPT (2024)	GPT-4 (2023)	79.50 (F1)	✓
	In-the-Wild (2024)	Fitness	M, T	LLaSA (2024)	Vicuna-7B (2023)	79.95 (Acc)	✓
				Coaches and Educators			
ACT	Custom Dataset (2024)	Rugby	I, T	Nonaka et al. (2024)	LLaVA-7B (2023)	63.10 ± 2.20 (F1)	✗
	ActionAtlas v1.0 (2024)	56 Sports	V	Salehi et al. (2024)	GPT-4o (2024)	42.95 ± 2.91(Acc)	✓
	NSVA Subset (2025)	BK, AF	V, T	SV3.3B (2025)	Llama 3.2-3B (2024)	85.60 ± 5.20 (BERT F1)	✓
	FineTennis (2025)	Tennis	V	Teo (2025)	Video-LLaMA2-7B (2024c)	76.00 (Edit Score)	✓
	SoccerNet-v2 (2021)	Soccer	V	Soccer-CLIP (2025)	ViT-B/32 (2021)	75.70 (t-AmAP)	✗
	Tennis7 (2021)	Tennis	V			93.80 (Acc)	
	F³Set (2025)	TN, BM, TT	V, T	F³ED (2025)	GPT-4 (2023)	75.20 (F1elm)	✓
	Video-MME (2025)	SC,BK,GY,DV	V, T	F-16 (2025f)	LLaVA-OV (2025a)	65.00 (Acc)	✓
	SoccerNet-v2 (2021)	Soccer	T	Chakraborty et al. (2025)	Llama 3.1-8B (2024)	64.50 (mAP)	✗
	SoccerNet-v2 (2021)	Soccer	V, T	Jiang et al. (2025b)	LLaVA-NeXT-Video (2024a)	63.50 (Acc)	✗
	UCI-HAR (2013)	Fitness	M	Gao et al. (2025b)	GPT-4 (2023)	92.30 (Acc)	✗
	SoccerNet (2018)	Soccer	V, A, T	Banu et al. (2025)	Video-LLaMA (2023)	87.00 (F1)	✗
AQA	Fis-V (2019)	Figure Skating	V, A, T	Wang et al. (2025a)	InternVL2 (2024c)	84.00 (Spearman)	✓
	FS1000 (2023)					90.00 (Spearman)	
	FineFS (2023)					76.00 (Spearman)	
	Fitness-AQA (2022)	Fitness	V, T	Dibenedetto et al. (2025)	LLaVA-Video-7B (2025c)	22.82 (mAP)	✓
	FMS (2025b)	Fitness	V, T	FitnessAgent (2025b)	ChatGLM4 (2024)	39.34 (Acc)	✗
	LLM-FMS (2025)	Fitness	V, T	Xing et al. (2025)	-	91.00 (Acc)	
TAC	Custom Dataset (2023)	Soccer	T	TacticalGPT (2023)	GPT-NeoX-20B (2022)	50.00 (Acc)	✗
	STATS SportVU 2025	Basketball	I, T	Smartboard (2024c)	GPT-4V (2024)	-	✗
	Custom Dataset (2021)	Baseball	T	Michielssen et al. (2024)	Curie (2020)	97.00 (Acc)	✓
	SportsMetrics (2024c)	BK, AF	T	Hu et al. (2024b,c)	Gemini-Pro 2023	32.30 (ΔGScore)	✓
	Custom Dataset (2024)	Cycling	T	Janssens et al. (2024)	GPT-4o (2024)	-	✗
	Custom Dataset (2024d)	Basketball	T	SportsGen (2024d)	GPT-4o (2024)	98.41 (DnC-10)	✓
	Custom Dataset (2025a)	Badminton	V, T	ChatMatch (2025a)	GPT-3.5-turbo (2022)	98.84 (Acc)	✗
	Basketball-Instants 2024	Basketball	I, T	TacticExpert (2025)	Vicuna-7B-v1.5 (2023)	83.33 (Macro F1)	✗
PRD	Custom Dataset (2020)	Basketball	T	Oved et al. (2020)	BERT (2019)	58.50 (Acc)	✗
	ShuttleSet (2023)	Badminton	V, T	RallyTemPose (2024)	BERT (2019)	54.30 (Acc)	✓
	BadmintonDB (2022)					62.80 (Acc)	
	Custom Dataset (2024)	Basketball	T	Sprint (2024)	GPT-3.5-turbo (2022)	64.90 (Acc)	✓
	SportDevs (2025)	Handball	T	Felice (2024)	Mistral-7B (2023)	5.20 (RMSE)	✗
	Custom Dataset (2025)	Cricket	V, T	Bhatnagar and Bhatnagar (2025)	GPT-4o mini (2024), etc.	86.30 (F1)	✓
				Referees			
REF	SoccerNet-XFoul (2024)	Soccer	V, T	X-VARS (2024; 2025)	Video-ChatGPT (2024)	3.80 (Likert Scale)	✓

Table 2: Summary of task-specific sports datasets related to large models, including athletes and trainers, coaches and educators, and referees. Task: PLA: exercise and training plans, INJ: sports injury and rehabilitation, PSY: sports psychology and behavior, ACT: action spotting and recognition, AQA: sports action quality assessment, TAC: sports tactics and strategies, PRD: game and player performance prediction, REF: sports refereeing. Sports: SC: soccer, BK: basketball, CL: sports climbing, AF: American football, TN: tennis, BM: badminton, TT: table tennis, GY: gymnastics, DV: diving. Modal: V: video, I: image, A: audio, S: skeleton data, M: IMU data, T: text.

Task	Dataset	Sports	Modal	Method	Related Large Model	Performance	Link
				Fans and Social Media			
CMT	Custom Dataset (2022)	Soccer	V, A, T	Gautam et al. (2022)	GPT-3 (2020)	0.31 (ROUGE-L)	✓
	SN-Caption-test-align (2024)	Soccer	V, T	MatchVoice (2024)	Llama 3 (2024)	42.00 (CIDEr)	✓
	LoL19 (2024)	Esports	T	Wang and Yoshinaga (2024)	Llama 2 13B (2023b)	-4.61 (BARTScore)	✓
	Custom Dataset (2024b)	Soccer	V, T	AiCommentator (2024b)	GPT-3.5-turbo (2022)	0.56 (Cohen's d)	✗
	CommentarySet (2024)	TF,SC,BK,GY,TT,TN	V, T	Ge et al. (2024)	InternVL-Chat-2 (2024c)	5.44 (SCORES)	✗
	Custom Dataset (2015; 2023)	Soccer	T	LLM-Commentator (2024)	LLaMA 7B (2023a)	92.00 (F1)	✓
		Golf			Llama 2 7B (2023b)	99.12 (ROUGE-L)	
	Custom Dataset (2024)	Tennis	V, T	Baughman et al. (2024)	Sandstone 3B (2020)	86.80 (ROUGE-L)	✗
		American Football			Llama 2 7B (2023b)	86.80 (ROUGE-L)	
	BH-Commentary (2024a)	Basketball	V, T	Zhang et al. (2024a)	BERT (2019)	12.19 (CIDEr)	✓
	SoccerNet-Caption (2023)	Soccer	V, T	TimeSoccer (2025)	Llama 2 7B (2023b)	8.30 (CIDEr)	✓
	SoccerNet-V2 (2021)	Soccer	V, T	Jiang et al. (2025b)	Claude 3.5 Sonnet (2025)	2.59 (Likert Scale)	✗
	WyScout (2024)					2.96 (Likert Scale)	
	LFCBI (2025)	Soccer	V, T	Mori et al. (2025)	GPT-4o (2024)	15.50 (MSE)	✓
	SoccerTrack-Commentary (2025)	Soccer	V, I, T	Vijayakumar et al. (2025)	GPT-3 (2020)	33.84 (CIDEr)	✗
	LiveSports-3K-CC (2025b)	49 Sports	V, A, T	LiveCC (2025b)	Qwen2-VL-7B (2024a)	40.08 (Win Rate)	✓
	SoccerNet-v2 (2021)	Soccer	V, A, T	SoccerComment (2025d)	Vicuna-7B-v1.5 (2023)	36.58 (CIDEr)	✗
	NBA-Identity (2025b)	Basketball	V, T	LLM-IAVC (2025b)	Llama 3.2-3B (2024)	105.30 (CIDEr)	✓
	VC-NBA-2022 (2025a)					150.70 (CIDEr)	
	Custom Dataset (2025)	Cricket	V, T	Sameer et al. (2025)	GPT-4o mini (2024), etc.	83.00 (BERT F1)	✗
	SoccerNet-Caption (2023)	Soccer	V, T	StreamMind (2025)	Video-LLaMA2-7B (2024c)	82.04 (ROUGE-L)	✓
	SoccerNet (2018)	Soccer	V, T	VLM-TSI (2025)	VideoLLM-Online (2024a)	39.10 (TRACE)	✓
HLG	CricPulse (2023)	Cricket	V, T	Sattar et al. (2023)	BERT (2019)	97.00 (F1)	✗
	Custom Dataset (2024)	Soccer	V, A	SmartCrop (2024)	GPT-4 (2023)	-	✗
	Custom Dataset (2025)	Baseball	T	DIAMOND (2025)	Mistral-Large (2024)	76.50 (F1)	✗
	HIPPO-Video (2025)	-	V, T	HiPHer (2025)	GPT-4 (2023)	76.60 (mAP)	✓
	Custom Dataset (2025)	Cricket	V, A, T	SportSummarizer (2025)	DistilBERT (2019)	0.93 (HD)	✗
	SoccerNet (2018)	Soccer				0.92 (HD)	
NSG	SportsSum2.0 (2021)	Soccer	T	Wang et al. (2021)	RoBERTa (2019), etc.	47.78 (ROUGE-L)	✓
	SportsSum (2020)	Soccer				47.49 (ROUGE-L)	
	K-SportsSum (2022)	Soccer	T	KES (2022)	mT5 (2021)	47.17 (ROUGE-L)	✓
	SportsSum (2020)	Soccer				47.79 (ROUGE-L)	
	NBA API (2022)	Basketball	T	SNIL (2024a)	GPT-3.5 (2022)	63.00 (Acc)	✓
	ShuttleSet (2023)	Badminton	T	BADGE (2024)	GPT-4 (2023)	8.63/10 (LLM)	✓
	Custom Dataset (2024)	Cricket	T	Sarkar et al. (2024)	Google Gemini (2023)	9.20/10 (ACS)	✗
	RotoWire (2017)	Basketball	T	Tree-of-Report (2025)	GPT-4o mini (2024)	54.92 (CS F1)	✗
	MLB (2019)	Baseball				62.99 (CS F1)	
	ShuttleSet+ (2025)	Badminton				93.94 (CS F1)	
NAR	Custom Dataset (2023)	Soccer	T	Sarfati et al. (2023)	T5-large (2020)	49.04 (ROUGE-L)	✗
	SportsVU (2024)	Basketball	V, T	Sportify (2024)	-	72.22 (Acc)	✓
	SoccerSum (2024a; 2024b)	Soccer	V, A	SoccerSum (2024a)	GPT-4 Turbo (2023)	-	✓
	Custom Dataset (2025)	Basketball	V	SportsBuddy (2025)	GPT-4o (2024)	90.80 (Acc)	✗
OPI	Custom Dataset (2024)	Soccer	T	Rauchegger et al. (2024)	GPT-4-turbo (2023)	70.30 (F1)	✗
	Custom Dataset (2025)	Soccer	T	ABSA (2025)	RoBERTa (2019)	80.00 (F1)	✓
MOD	Custom Dataset (2024)	Soccer	T	Schilling et al. (2024)	GPT-3.5 (2022)	71.40 (Acc)	✗
	SoccerNet (2018)	Soccer	V, A, I	SoccerRAG (2024a; 2024b)	GPT-4 (2023), etc.	80.00 (Acc)	✓
	Custom Dataset (2025)	Soccer	T	Karat et al. (2025)	GPT-4o (2024)	88.25 (Precision)	✗
	Custom Dataset (2025)	Cricket	T	Wickramasinghe (2025)	Copilot (2021)	100.00 (Acc)	✗
	TF-CoVR (2025)	GY, DV	V, T	TF-CoVR-Base (2025)	BLIP (2022)	23.02 (mAP@10)	✓
	KICK (2025b)	Soccer	T	Song et al. (2025b)	GPT-4o (2024)	15.86 (JGA)	✓
	SoccerNet-XFoul (2024)	Soccer	V, T	SoccerChat (2025)	Qwen2-VL-7B (2024a)	6.81/10 (LLM)	✓
	SoccerNet-v2 (2021)					6.42/10 (LLM)	
	SoccerBench (2025a)	Soccer	V, A, T	SoccerAgent (2025a)	DeepSeek-v3 (2024a)	60.9 0(Acc)	✓
	SoccerNet-v2 (2021)					80.10 (Acc)	
	SN-Caption-test-align (2024)	Soccer	V, T	MatchVision (2025b)	Llama 3-8B (2024)	44.18 (CIDEr)	✓
	MVFoul (2023)					44.00 (Acc)	
	SoccerNet-V2 (2021)	Soccer	V, T	Jiang et al. (2025b)	LLaVA-NeXT-Video (2024a)	83.76 (Acc)	✗
	WyScout (2024)					81.83 (Acc)	
	Gym-QA (2025a)	Gymnastics	V, T	FineQuest (2025a)	Video-LLaVA (2024a), etc.	57.00 (Acc)	✗
	Diving-QA (2025a)	Diving	V, T				
	SPORTU (2025)	7 Sports	V, T			73.20 (Acc)	
				The Sports Industry			
MAN	Custom Dataset (2024)	-	T	Merilehto (2024)	Claude 3 Opus (2024)	90.28 (Acc)	✗
TSC	Custom Dataset (2025)	Soccer	T	Martire and Ragazzi (2025)	GPT-4o (2024)	3.80 (Likert Scale)	✗
	Custom Dataset (2025)	Soccer	I, T	Footyintel (2025)	-	-	✗

Table 3: Summary of task-specific sports datasets related to large models, including fans and social media, and the sports industry. Task: CMT: sports commentary generation, HLG: sports highlight generation, NSG: sports news generation, NAR: sports narratives and storytelling, OPI: public opinion analysis in sports, MOD: sports models and systems, MAN: sports management, TSC: sports talent scouting. Sports: TF: track and field, SC: soccer, BK: basketball, TN: tennis, TT: table tennis, GY: gymnastics, DV: diving. Modal: V: video, I: image, A: audio, T: text.

Dataset	Sports	Source	Benchmark	Input Modal	# Video	Avg. Length	# QA	Link
Specialized Sports Understanding Datasets								
QASports (2023)	SC, BK, AF	Fandom	✗	text	-	-	∼1500K	✓
BIG-bench-SU (2023)	SC, BK, AF, BB, IH	program	✓	text	-	-	986	✓
Sports-QA (2024b)	SC, BK, GY, VB	exist (dataset)	✓	video, text	5967	20.9s	94073	✓
SportQA-Level-1 (2024a)	-	exist	✓	text	-	-	21385	
SportQA-Level-2 (2024a)	35 Sports	Wikipedia	✓	text	-	-	45685	✓
SportQA-Level-3 (2024a)	SC, BK, TN, AF, TT, VB	expert	✓	text	-	-	3522	
SPORTU-text (2025)	SC, BK, TN, AF, VB	expert	✓	text	-	-	900	✓
SPORTU-video (2025)	SC,BK,BM,TN,BB,VB,IH	competition	✓	video, text	1701	-	12048	
Sports-3K-QA (2025b)	49 Sports	YouTube	✓	video, text	412	-	1174	✓
FSAnno (2025a)			✗	video, text, audio	783		-	
FSBench-Text (2025a)	FS	competition	✓	text	-	∼3.5m	500	✗
FSBench-Motion (2025a)			✓	video, text	783		3500	
FineBadminton (2025b)	BM	YouTube	✗	video, text	3215	12.4s	-	✓
FBBench (2025b)			✓	video, text	2563	-	2563	
Gym-QA (2025a)	GY	exist	✓	video, text	6031	-	27469	✗
Diving-QA (2025a)	DV	exist	✓	video, text	∼100	-	1055	✗
General Video Understanding Datasets								
InternVid (2024c)	-	YouTube	✗	video, text, audio	7.1M	6.4m	N/A	✓
Ego-Exo4D (2024)	SC, BK, CL	manual	✗	video, audio	5035	2.6m	N/A	✓
E.T. Bench (2024b)	SC, BK, TN, CR, etc.	exist	✓	video, text	7002	129s	7289	✓
LVBench (2024b)	BK, etc.	YouTube	✓	video, text	103	4101s	1549	✓
VideoVista (2024c)	SC, etc.	exist	✓	video, text, audio	894	131s	24906	✓
FIOVA (2024a)	BB, etc.	-	✓	video	3002	33.6s	N/A	✓
Neptune (2024)	BK, etc.	exist	✓	video, text, audio	2405	2.5m	3268	✓
videoSALMONN2 (2025a)	BB, etc.	-	✓	video, text, audio	483	51s	N/A	✓
MMWorld (2025a)	SC, BK, GY, VB	exist	✓	video, text, audio	1910	∼105s	6627	✓
LongVILA_sft (2025c)	-	exist	✗	video, text	15292	-	15292	✓
MLVU (2025b)	SC, BK, BM, TT, VB	-	✓	video, text	3102	930s	3102	✓
Video-MME (2025)	SC, BK, etc.	YouTube	✓	video, text, audio	900	1017.9s	2700	✓
MotionBench (2025b)	BB, etc.	exist, syn, web	✓	video, text	5385	<10s	8052	✓
OVO-Bench (2025)	-	exist, YouTube	✓	video, text	644	428.89s	2814	✓
VISTA-400K (2025)	-	exist	✗	video, text	403994	48.6s	∼381K	✓
HRVideoBench (2025)	-	manual	✓	video, text	200	5.4s	200	✓
HarmonySet-train (2025c)	-	YouTube	✗	video, text, audio	44470	31.5s	44470	✓
HarmonySet-MC (2025c)			✓		3858		3858	
VideoA11y-40K (2025b)	-	online	✗	video, audio	40000	-	N/A	✓
TUNA (2025a)	SC, BK, etc.	exist	✓	video, text	1000	14.5s	2000	✓
WorldSense (2025a)	-	exist	✓	video, text, audio	1662	141.1s	3172	✓
V-STaR (2025)	-	exist, YouTube	✓	video, text	2094	110.23s	-	✓
MINERVA (2025)	BK, TN, etc.	YouTube	✓	video, text	223	12m	1515	✓
MAVERIX (2025)	SC, BK, etc.	exist	✓	video, text, audio	700	5.7m	2556	✗
RTV-Bench (2025)	SC, BK, etc.	exist, online	✓	video, text	552	18.2m	4631	✓
VidText (2025b)	SC, BK, BM, TT, SW	exist, YouTube	✓	video, text, audio	939	108.2s	2857	✓
SIV-Bench (2025b)	SC, etc.	YouTube, TikTok	✓	video, text, audio	2792	32.49s	8728	✓
ExAct (2025)	SC, BK, CL	exist	✓	video, text	3521	105s	3521	✓
VRBench (2025)	SC, BK, VB, etc.	YouTube	✓	video, text, audio	960	1.6h	8243	✓
Trust-videoLLMs (2025d)	-	exist, syn, YouTube	✓	video, text, audio	6955	-	-	✓
CausalStep (2025e)	-	exist	✓	video, text	100	430.5s	1852	✗
EgoExoBench (2025c)	BK, etc.	exist	✓	video, text	-	-	7350	✓
WildVideo (2025a)	-	exist	✓	video, text	1318	∼30s	17625	✗

Table 4: Summary of large-model-related datasets specifically for sports understanding and general video understanding with sports content. Sports: SC: Soccer, BK: Basketball, BM: Badminton, TN: Tennis, GY: Gymnastics, FS: Figure skating, AF: American Football, BB: Baseball, CR: Cricket, TT: Table Tennis, VB: Volleyball, DV: Diving, IH: Ice Hockey, CL: Sports Climbing, SW: Swimming. Source: syn: synthesis.

Sports	**Abbr.**	**Task**	**Abbr.**
Soccer	SC	Exercise and Training Plans	PLA
Basketball	BK	Sports Injury and Rehabilitation	INJ
Badminton	BM	Sports Psychology and Behavior	PSY
Tennis	TN	Action Spotting and Recognition	ACT
Gymnastics	GY	Sports Tactics and Strategies	TAC
Figure Skating	FS	Game and Player Performance Prediction	PRD
Rugby & American Football	RG & AF	Sports Action Quality Assessment	AQA
Baseball	BB	Sports Education	EDU
Cricket	CR	Sports Refereeing	REF
Table Tennis	TT	Sports Commentary Generation	CMT
Volleyball	VB	Sports Highlight Generation	HLG
Diving	DV	Sports News Generation	NSG
Ice Hockey	IH	Sports Narratives and Storytelling	NAR
Track and Field	TF	Public Opinion Analysis in Sports	OPI
Sports Climbing	CL	Sports Models and Systems	MOD
Handball	HB	Sports Academic Writing	WRI
Cycling	CY	Sports Management	MAN
Boxing	BX	Sports Talent Scouting	TSC
Swimming	SW	Sports Tourism	TOU

Table 5: Abbreviations for sports and tasks mentioned in this paper.