

DEVELOPMENT OF AN INTELLIGENT SPORTS ANALYTICS SYSTEM BASED ON MATCH TRIGGER ANALYSIS AND THE APPLICATION OF ARTIFICIAL INTELLIGENCE METHODS

Kaliyeva A. T., Martyntsov N.V.^a

Astana University College, Astana, Kazakhstan

Corresponding author: Martyntsov N.V. Nikolay.Martyntsov@astanait.edu.kz

Abstract. This article discusses the development of a sports analysis system based on match data processing and the application of artificial intelligence methods. A software solution is proposed that allows statistical data to be uploaded in Excel format and automatically identifies key factors influencing the course of a match, such as winning streaks, draws, and changes in game dynamics. Various artificial intelligence models are examined and a comparative analysis of their applicability to sports analysis tasks is conducted, taking into account the specifics of trigger data. Based on the identified triggers, a structured set of features is generated and sent to the artificial intelligence model for further analysis. A language model is used to interpret the obtained data and generate analytical conclusions and recommendations for players. The results of the study show that the proposed approach automates match analysis, improves the objectivity of performance evaluation, and reduces data processing time. The developed system can be used to optimize training and identify weaknesses in players' performance. The article also presents the mathematical models and formulas used to identify triggers, and justifies the choice of these methods in terms of analytical effectiveness.

Keywords: sports analytics, artificial intelligence, match analysis, triggers, Excel, data processing.

Introduction. In the context of the modern world, the need for tools that enable rapid and accurate data processing and extraction of relevant information is becoming increasingly evident [1, 2]. In practice, match analysis is often conducted manually or using simple data processing methods. This approach is time-consuming and does not always reveal important trends in the game, as confirmed in studies on the development of training methods [3]. Furthermore, the results of the analysis can be influenced by the subjective opinion of the specialist, which reduces its accuracy.

This requires the development of a system capable of automatically processing match data, identifying key factors, and using artificial intelligence methods for further analysis [4]. The proposed approach aims to simplify the analysis process, increase its objectivity, and reduce its time.

The aim of the study is to develop a sports analytics system using trigger-based artificial intelligence to automate statistical data processing. Table tennis is characterized by a fast pace, rapidly changing game situations, and a large number of short rallies. Furthermore, winning streaks, changes in tempo, and consistency in player performance play a significant role in table tennis; therefore, the use of automated analysis allows for more accurate match evaluation and the identification of patterns in play.

Materials and methods. Online resources on sports match data processing and game analysis methods were utilized. A table tennis match analysis expert was also consulted to identify factors that trigger critical game situations and recurring game patterns.

Particular attention was paid to the selection of the artificial intelligence model, as not all language models are capable of accurately analyzing sports data and interpreting game progress. LM Studio software [5] was used to develop these models, allowing for the processing of language models locally on the computer and their adaptation based on specific data.

When developing sports analysis systems, match processing is divided into two main types: individual analysis and head-to-head analysis.

Individual analysis focuses on the statistics of a specific player, independent of other players in the game. This analysis examines individual performance indicators, consistency of play, win and loss streaks, and performance changes in various game situations.

1. *Single-player analysis.* The analysis used metrics based on match statistics, wins and losses, match schedule, and post-holiday changes in results. These metrics were chosen because they allow us to determine a player's consistency, mental toughness, and the impact of external factors on their overall performance.[6]

A basic list of triggers for single-player analysis that are most commonly encountered by players:

- Top-performing players
- High percentage of defeats
- A losing streak
- Post-holiday performance decline
- Time-of-day analysis

Top-performing players. This trigger identifies players with the highest performance. The primary metric used is win rate over the analyzed period. A player is considered highly effective if their win rate is at least 70% across all matches. This enables the identification of strong and consistent players.

$$W_r = \frac{W}{M} * 100\%$$

W – number of victories.

M – number of matches.

High percentage of defeats. This trigger identifies players with a high loss rate exceeding 50%. It also helps detect inconsistencies in their results and potential issues with game preparation.

$$L_r = \frac{L}{M} * 100\%$$

L – number of defeats.

M – number of matches.

Series of defeats. This trigger is used to analyze long losing streaks. Its formula identifies the current and longest losing streak. Such patterns are often observed in players who are mentally unprepared or experiencing a loss of form.

$$S_l = \sum_{i=1}^n l_i$$

l_i – consecutive defeat.

S_l – length of losing streaks.

n – number of matches in the series.

Post-holiday performance decline. This trigger was developed to analyze changes in performance after a rest period and determine the impact of breaks on an athlete's performance.

$$P_h = \frac{L_h}{M_h}$$

L_h – defeats after the holidays.

M_h – number of matches after the holidays.

Time-of-day analysis. The "Time-of-day analysis" trigger allows the system to evaluate the impact of time of day on a player's performance during a match. This enables the identification of periods during which player performance declines or improves.

$$T_p = \frac{L_t}{M_t} * 100\%$$

L_t – defeats in a certain period of time.

M_t – number of matches in a given period.

Based on the received data, the system forms a structured set of features, which are then used by the artificial intelligence model to generate analytical conclusions and recommendations.

2. Head-to-head analysis. The analysis was based on metrics such as losing streaks, player behavior in tied-score situations, first-set evaluations, typical final results, and unusual game situations. These metrics were chosen because they allow for a clear analysis of athletes' mental resilience, their ability to maintain an advantage, and the specific characteristics of their performance against specific opponents.

A losing streak against a specific opponent

Losses in close match endings

Recurring match scores

Anomalous sets

The impact of the first set

Behavior when the score is 1:1

Behavior with a 2:0 lead

Losing streaks against a specific opponent. The "Loss Streak vs. Opponent" metric helps identify repeated losses to a specific player. This metric allows for the accurate identification of strong opponents and an analysis of an athlete's ability to counter a specific playing style.

$$S_h = \sum_{i=1}^n l_i$$

S_h – a losing streak against a specific opponent.

l_i – individual loss.

n – number of matches in the series.

Losses in close match endings. This trigger is used to analyze matches lost by minimal sets. It helps determine a player's ability to maintain consistency during key moments of the match.

$$C_l = \frac{L_c}{L_t} * 100\%$$

L_c – number of defeats in different endings.

L_t – total number of defeats.

Recurring match scores. Recurring game scenarios help identify the most common outcomes in matches between two players. Analyzing these recurring scenarios allows us to discover patterns in the matchup and dominant game tendencies.

$$F_s = \max (score_i)$$

$score_i$ – specific match score.

F_s – most frequently repeated set.

Anomalous sets. This trigger indicates anomalous situations in sets, such as major losses or wins. Such results are very rare, so they indicate dominance of one of the opponents.

$$A_s = \sum_{i=1}^n a_i$$

a_i – anomalous set.

A_s – number of anomalous sets.

Impact of the first set. The first set often has a profound impact on a player's mental state and confidence. By identifying a pattern between a win or loss in the first round and the final result of the game, we can predict the outcome in advance using this trigger. In this case, there are four possible outcomes: win one set but lose the game; win one set and win the game; lose one set and lose the game; or lose one set but win the game.

$$F_w = \frac{W_{1:0}}{M_{1:0}} * 100\%$$

$W_{1:0}$ – victories after winning the first set.

$M_{1:0}$ – matches where the player led 1:0.

Behavior when the score is 1:1. A 1-1 score is a key moment in a match. Therefore, using this trigger, we can detect a player's tendency, or more precisely, their behavior after a 1-1 score. If the percentage is greater than 80%, it indicates that the player is losing ground in crucial moments. This could be caused by poor mental preparation or excessive anxiety.

$$D_b = \frac{W_d}{M_d} * 100\%$$

W_d – number of successful scenarios after a 1:1 score.

M_d – total number of matches with a score of 1:1.

Behavior with a 2:0 lead. This trigger allows the system to identify a player's behavior when they have a high chance of winning a match. The player may either maintain the advantage and win with a score of 3:0, 3:1, or 3:2, or lose the match despite the initial advantage. If they win, this indicates that the player is tenacious and can turn the game around in a stressful situation.

$$L_{2:0} = \frac{M_{lose}}{M_{2:0}} * 100\%$$

M_{lose} – matches lost after a score of 2:0.

$M_{2:0}$ – total number of matches with a 2:0 advantage.

Head-to-head analysis allows us to identify recurring patterns between specific opponents and analyze player behavior in key game situations. The resulting triggers help us understand athletes' mental resilience, their ability to maintain an advantage, and their effectiveness against specific opponents. The analysis results are used by a language model to generate analytical reports and recommendations.

Research results and their discussion. The system used artificial intelligence language models to generate analytical inferences. Selecting the appropriate model was a crucial step, as different models analyze statistical data and game situations differently [7]. During development, several models were tested, accessible through the LM Studio program:

- openai/gpt-oss-20b;
- DeepSeek-R1-0528-Qwen3-8B;
- google/gemma-2-2b-it;
- qwen2.5-3b-instruct.

When comparing the models, the following were taken into account:

- accuracy;
- data processing speed;
- stability of responses;
- quality of comparisons between players;
- the ability of the model to find recurring game scenarios.

During testing, the DeepSeek-R1-0528-Qwen3-8B model showed the best results when analyzing pairs of players (Figure 1).[8]

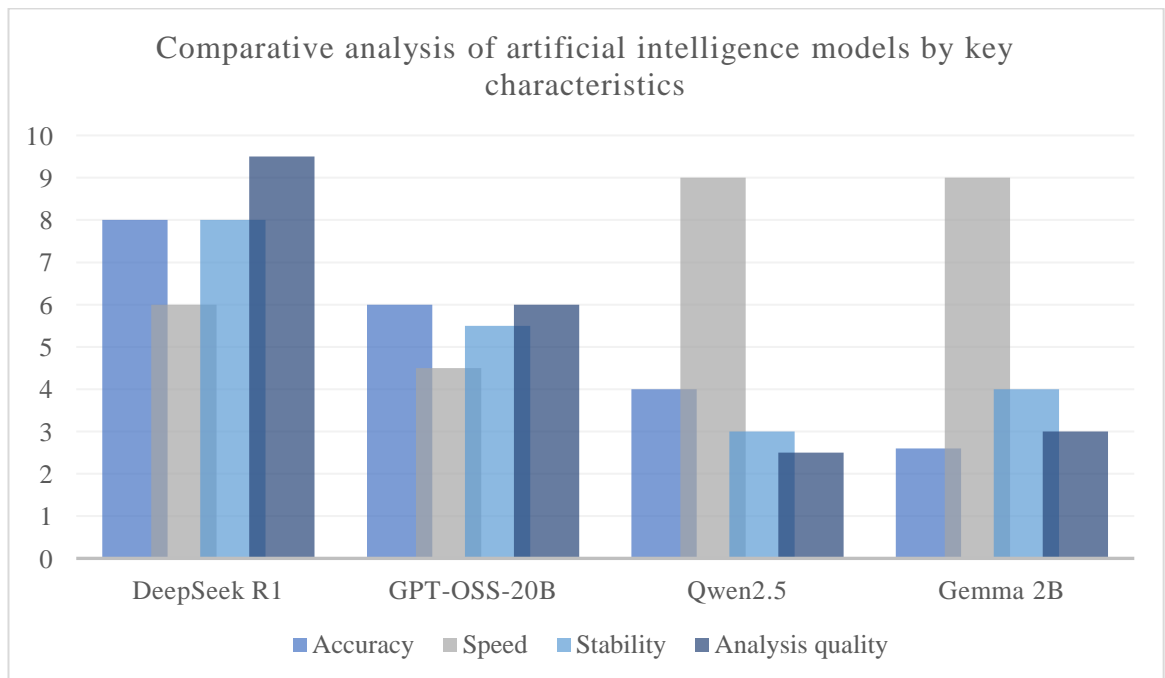


Figure 1 - Comparative analysis of artificial intelligence models by key characteristics

It proved more effective at analyzing matches between specific opponents, accurately identifying patterns, and drawing logical conclusions. Furthermore, the model demonstrated the ability to perform in-depth analysis using match history data and evidence-based reasoning. (Table 1)

Table 1 – A comparative analysis of artificial intelligence models in sports match analysis

Model	Advantages	Disadvantages
R1 DeepSeek	Good comparative analysis. Provides evidence.	Requires high computational resources.
GPT-OSS-20B	Fast response generation, has the ability to make a comparative analysis.	Slower response generation and occasional analytical inaccuracies.
Qwen 2.5	Fast generation.	Limited analytical capabilities and lower accuracy.
Gemma 2B	Low resource consumption, fast response generation.	Low accuracy, less stable output, and grammatical errors. Generated excessive irrelevant information.

The model also demonstrated good performance and high stability when running repeated queries. This is especially important, as sports analysis requires minimal random errors to draw conclusions.

The openai/gpt-oss-20b model also performed well, particularly when analyzing individual players and generating detailed recommendations. However, it had grammatical errors when generating responses in Russian and sometimes became confused in its conclusions, providing incorrect information.

The Google/gemma-2-2b-it and qwen2.5-3b-instruct models [9, 10] demonstrated low analysis accuracy and inconsistent results; therefore, they were not included in the final version of the system.

As a result, the DeepSeek-R1-0528-Qwen3-8B model was chosen as the main model for pairwise analysis, since it provided a good balance between speed, accuracy of analysis and quality of analytical conclusions.

The final program and its architecture were developed using React JS, with the server side using Fast API. Player data and analysis history are stored in PostgreSQL 17. (Figure 2)

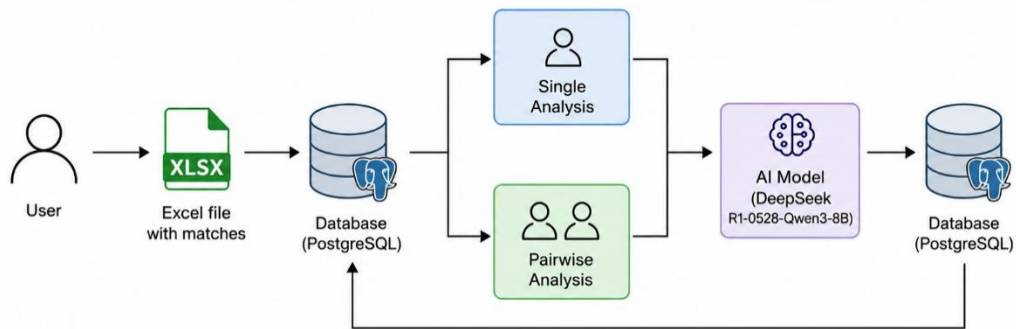


Figure 2 - Sports analytics system architecture

The user uploads an Excel file with match data, after which the information is stored in a PostgreSQL database. The system then performs individual or head-to-head match analysis based on developed triggers. The obtained results are fed to the DeepSeek-R1-0528-Qwen3-8B artificial intelligence model to generate analytical conclusions and recommendations. Once the analysis is complete, the final data is saved back to the database for further use and display in the system.

Conclusion. This study developed a table tennis analysis system based on trigger analysis and artificial intelligence. The system processes match statistics, identifies recurring game situations, and generates analytical insights to evaluate athlete performance. The dataset included more than 1,500 table tennis matches collected during 2024–2025.

The study identified key triggers for individual and head-to-head player analysis. Using these triggers, it was possible to accurately determine the consistency of athletes' performance,

psychological factors, specific characteristics of play against certain opponents, and other trends that influence game outcomes.

The study also tested various artificial intelligence models for sports data analysis. The results showed that the DeepSeek-R1-0528-Qwen3-8B model was the most effective for head-to-head player analysis and drawing conclusions.

The results demonstrate that the use of artificial intelligence and automated analysis reduces data processing time, limits the impact of human error, and improves the objectivity of sports analysis. The developed system can be used to analyze matches, athlete training, and develop analytical tools for sports. The proposed approach demonstrates the potential of artificial intelligence methods for automating sports analytics and supporting coaching decision-making processes.

МАТЧТАРДЫ ТРИГГЕРЛІК ТАЛДАУ ЖӘНЕ ЖАСАНДЫ ИНТЕЛЛЕКТ ӘДІСТЕРІН ҚОЛДАНУ НЕГІЗІНДЕ СПОРТТЫҚ АНАЛИТИКАНЫҢ ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕСІН ӘЗІРЛЕУ

Қалиева А. Т., Мартынцов Н. В.^а

Astana University College, Астана қ., Қазақстан

Хат алмасу үшін автор: Мартынцов Н. В. Nikolay.Martyntsov@astanait.edu.kz

Андатпа. Бұл мақалада матч деректерін өңдеу және жасанды интеллект әдістерін қолдану негізінде спорттық талдау жүйесін әзірлеу қарастырылады. Excel форматындағы статистикалық деректерді жүктеуге және жеңіс сериялары, тең аяқталған матчтар мен ойын динамикасының өзгерістері сияқты матч барысына әсер ететін негізгі факторларды автоматты түрде анықтауға мүмкіндік беретін бағдарламалық шешім ұсынылады. Жасанды интеллекттің әртүрлі модельдері қарастырылып, триггерлік деректердің ерекшелігін ескере отырып, олардың спорттық талдау міндеттеріне қолданылуына салыстырмалы талдау жүргізіледі. Анықталған триггерлер негізінде жасанды интеллект моделіне одан әрі талдау үшін жіберілетін белгілердің құрылымдық жиыны қалыптастырылады. Тілдік модель алынған деректерді интерпретациялау және ойыншыларға арналған аналитикалық қорытындылар мен ұсынымдар генерациялау үшін қолданылады. Зерттеу нәтижелері ұсынылған тәсілдің матчтарды талдауды автоматтандыратынын, нәтижелерді бағалаудың объективтілігін арттыратынын және деректерді өңдеу уақытын қысқартатынын көрсетеді. Әзірленген жүйе жаттығуларды оңтайландыру және ойыншылардың әлсіз жақтарын анықтау үшін пайдаланылуы мүмкін. Мақалада сонымен қатар триггерлерді анықтауға арналған математикалық модельдер мен формулалар ұсынылып, бұл әдістердің таңдауы талдау тиімділігі тұрғысынан негізделеді.

Түйінді сөздер: спорттық аналитика, жасанды интеллект, матчтарды талдау, триггерлер, Excel, деректерді өңдеу.

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ СПОРТИВНОЙ АНАЛИТИКИ НА ОСНОВЕ ТРИГГЕРНОГО АНАЛИЗА МАТЧЕЙ И ПРИМЕНЕНИЯ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Калиева А. Т., Мартынцов Н. В.^а

Astana University College, г. Астана, Казахстан

Автор для корреспонденции: Мартынцов Н. В. Nikolay.Martyntsov@astanait.edu.kz

Аннотация. В данной статье рассматривается разработка системы спортивного анализа, основанной на обработке данных матчей и применении методов искусственного интеллекта. Предлагается программное решение, позволяющее загружать статистические данные в формате Excel и автоматически определять ключевые факторы, влияющие на ход матча, такие как победные серии, ничьи и изменения в динамике игры. Рассматриваются различные модели искусственного интеллекта и проводится сравнительный анализ их применимости к задачам спортивного анализа с учетом специфики триггерных данных. На основе выявленных триггеров генерируется структурированный набор характеристик, который отправляется в модель искусственного интеллекта для дальнейшего анализа. Языковая модель используется для интерпретации полученных данных и генерации аналитических выводов и рекомендаций для игроков. Результаты исследования показывают, что предложенный подход автоматизирует анализ матчей, повышает объективность оценки результатов и сокращает время обработки данных. Разработанная система может быть использована для оптимизации тренировок и выявления слабых сторон игроков. В статье также представлены математические модели и формулы, используемые для выявления триггеров, и обосновывается выбор этих методов с точки зрения эффективности анализа.

Ключевые слова: спортивная аналитика, искусственный интеллект, анализ матчей, триггеры, Excel, обработка данных.

References

- 1 Mateus N., Abade E., Coutinho D., Gómez M.-Á., Sampaio J. Empowering the Sports Scientist with Artificial Intelligence in Training, Performance, and Health Management // *Sensors*. – 2025. – Vol. 25. – Art. 139. – DOI: 10.3390/s25010139. – URL: <https://www.mdpi.com/1424-8220/25/1/139> (accessed: 01.05.2026).
- 2 Ghosh I. et al. Sports analytics review: artificial intelligence applications, emerging technologies, and algorithmic perspective // *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. – 2023. – Vol. 13, No. 5. – Art. e1496. – DOI: 10.1002/widm.1496.
- 3 Dzholiev I.M.O., Saparov B.M., Obnosov V.A., Mishin A.S. Development of innovative training methods for improving the professional skills of young basketball players [in Russian] // *Management of Education*. – 2025. – Vol. 15, No. 1-1. – P. 81–96. – DOI: 10.25726/d0195-2862-7353-e.
- 4 Manuylenko E.V., Tashchiyan A.A., Sozaeva A.S. The use of artificial intelligence technologies in sports [in Russian] // *Bulletin of Rostov State University of Economics (RINH)*. – Rostov-on-Don, 2023.
- 5 LM Studio Documentation: official LM Studio website. – URL: <https://lmstudio.ai/docs/app> (accessed: 01.05.2026).
- 6 Lei J., Kang T., Cao Y., Ren S. Capturing Momentum: Tennis Match Analysis Using Machine Learning and Time Series Theory // *arXiv preprint arXiv:2404.13300*. – 2024. – URL: <https://arxiv.org/abs/2404.13300> (accessed: 01.05.2026).
- 7 Xia H., Yang Z., Wang Y. et al. SportQA: A Benchmark for Sports Understanding in Large Language Models // *Proceedings of the 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. – 2024. – URL: <https://arxiv.org/abs/2402.15862> (accessed: 01.05.2026).
- 8 DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning // *arXiv preprint arXiv:2501.12948*. – 2025. – URL: <https://arxiv.org/abs/2501.12948> (accessed: 10.05.2026).
- 9 Qwen2.5 Technical Blog: official Qwen AI blog. – URL: <https://qwen.ai/blog?id=qwen2.5> (accessed: 19.04.2026).

10 Qwen2.5: A Party of Foundation Models // arXiv preprint arXiv:2408.00118. – 2024. – URL: <https://arxiv.org/abs/2408.00118> (accessed: 25.04.2026).

Авторлар туралы ақпарат // Информация об авторах // Information about the Authors

Мартынцов Николай Викторович – магистр, старший преподаватель, ТОО «Astana IT University», г. Астана, Казахстан

Мартынцов Николай Викторович – магистр, аға оқытушы, «Astana IT University» ЖШС, Астана қ., Қазақстан

Martyntsov Nikolay Viktorovich – Master’s Degree, Senior Lecturer, Astana IT University LLP, Astana, Kazakhstan

e-mail: Nikolay.Martyntsov@astanait.edu.kz

ORCID iD: 0009-0000-4580-6911

Калиева Айдана Талгатовна – студентка 3 курса по специальности «Программное обеспечение», ТОО «Колледж Astana IT University», г. Астана, Казахстан

Қалиева Айдана Талгатқызы – «Бағдарламалық жасақтама» мамандығының 3-курс студенті, «Astana IT University» Колледжі ЖШС, Астана қ., Қазақстан

Kaliyeva Aydana Talgatovna – 3rd-year Software Engineering student, Astana IT University College LLP, Astana, Kazakhstan

e-mail: aidanakalieva576@gmail.com

ORCID iD: 0009-0009-0244-7269