FPSRec: Football Players Scouting Recommendation System based on Generative AI

Antonio Maria Rinaldi, Antonio Romano, Cristiano Russo and Cristian Tommasino
Department of Electrical Engineering and Information Technologies, University of Naples Federico II, Naples, Italy
antoniomaria.rinaldi@unina.it, antonio.romano5@unina.it, cristiano.russo@unina.it, cristian.tommasino@unina.it

2024 IEEE International Conference on Big Data (BigData) | 979-8-3503-6248-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/BigData62323.2024.10825692

***Abstract*—Player scouting in soccer is witnessing a surge of interest from the research community. Traditional scouting methods are often limited by subjectivity and biases in evaluation. Moreover, the lack of structured data and models hinders the progress of the field. To overcome these limitations, we introduce a novel player recommendation system which integrates similarity techniques and generative artificial intelligence. It aims to support player recruitment by providing a data-driven and inclusive approach. The novelty of our work lies in its use of advanced machine learning and artificial intelligence to accurately predict player potential and performance by similarity measures, thereby mitigating the influence of subjective biases that often affect talent identification. Our contributions represent a significant advancement in the field of sports analytics and talent identification, offering a more equitable and efficient approach to scouting and recruitment. The results obtained underscore the effectiveness of the proposed system, demonstrating the transformative potential of artificial intelligence in revolutionizing talent scouting.**

***Index Terms*—Recommendation Systems, Player Scouting, Generative AI, Football Analytics, Sports Analytics, Data-driven Scouting.**

I. Introduction

In the highly competitive world of elite sports, particularly in the football industry, the process of player scouting and recruitment is a critical factor that can significantly influence the success of a team. This process involves substantial investments, long-term contracts, and a high degree of uncertainty regarding the performance and return on investment. To date, the lack of a data semantification process is a major limitation. Such process consists in structuring data within multimedia knowledge graphs [1] in order to enable the development of innovative and more sophisticated techniques. There are many successful examples of exploiting such techniques in other fields, such as medicine [2], cognitive robotics [3], [4], and industry [5].

Traditionally, player recruitment has relied heavily on subjective human observation and manual data analysis. However, this traditional approach has several limitations, including time-consuming processes and the consideration of only a limited number of factors. With the rise of big data, the ability to harness large-scale datasets has transformed numerous sectors, providing deeper insights and more robust decision-making frameworks. In sports, and specifically in football, big data analytics has begun to reshape player evaluation and recruitment, allowing for more nuanced and data-driven approaches.

In light of these challenges, we introduce a novel recommendation system, named *FPSRec*, which aims to support the player scouting process by integrating generative AI and similarity techniques. Our proposed system bridges the gap between traditional human-centric scouting and data-driven approaches, thereby enhancing the efficiency and effectiveness of player recruitment.

FPSRec leverages advanced machine learning algorithms to analyze a vast array of player data, including performance statistics, physical attributes, and historical injury records. By doing so, it can identify patterns and correlations that may not be apparent through manual analysis. Furthermore, our system incorporates similarity techniques to compare prospective recruits with current players, providing a more nuanced understanding of how a new player might fit into existing team dynamics. While our dataset may not be as extensive as those used in other domains, the scalability of our approach ensures that it can adapt and grow as more data becomes available, maintaining its relevance and effectiveness over time. Moreover, FPSRec is designed to be user-friendly and intuitive, enabling scouts and coaches to make informed decisions quickly. It provides clear, actionable insights that can guide recruitment strategies and player development programs. As such, our system does not replace human expertise but rather enhances it, combining the best of human judgment and AI-driven insights.

The rest of the paper is structured as follows: Section II discusses existing literature about the development of tools and systems for player scouting. Section III delves into the technical details of FPSRec and describes its architecture and components; while in Section IV we evaluate the system capabilities and performances, by carrying on several experiments and comparison for each component. Lastly, we draw the conclusions in Section V.

II. Related Work

In this section we provide a brief literature overview related to the field of talent scouting and recruitment in sports, particularly soccer, which has seen significant advancements in recent years.

Piggott et al. [6] discuss modern approaches to scouting and recruitment in football, providing an overview of current techniques and strategies. In [7] the authors explore the use of data-driven techniques for player recruitment in football, offering a new perspective on the talent selection process.

The perceptions and biases of recruiters play a crucial role in talent identification and scouting, particularly in sports. Larkin et al. [8] provide a detailed overview of this aspect by examining recruiters' perceptions of the key attributes for player recruitment in youth soccer. The study [9] explores how expert coaches evaluate players based on their subjective talent criteria in top-level youth soccer.

The study [10] addresses methodological issues in soccer talent identification research, providing a critical overview of current research techniques. The same authors [11] explored how soccer scouts identify talented players, offering a detailed overview of the scouting process.

Recruiters often rely on their subjective judgment and "eye for talent" to evaluate potential athletes. This involves assessing not only the athletes' current skills and performance but also their potential for future development. The authors of [12] discuss talent identification and the "practical sense" of top-level soccer coaches. However, these perceptions can be influenced by various biases. For instance, recruiters may favor athletes who share similar backgrounds or characteristics with successful players they have encountered in the past. They might also be influenced by stereotypes related to race, position, or gender. For instance, Bell et al. [13] examine biased recruiting and racial and positional stereotyping in girls' basketball scouting reports. These biases can affect the accuracy and fairness of their reports, potentially overlooking talented individuals who do not fit their preconceived notions. Therefore, it is essential to acknowledge and address these biases to ensure a more objective and inclusive talent identification process. Recommendation systems have become an integral part of many online platforms, providing personalized suggestions to users based on their preferences and behavior [14]–[16]. These systems use a variety of techniques, including collaborative filtering, content-based filtering, and hybrid methods, to predict user preferences [14], [15]. Recent developments in the field have seen the incorporation of advanced machine learning and artificial intelligence techniques, improving the accuracy and effectiveness of these systems [14], [15]. Despite the widespread use of recommendation systems in areas such as e-commerce and entertainment, their potential in sports, particularly in player scouting and talent identification, remains largely untapped. The application of recommendation systems in this context could revolutionize the way talents are identified and recruited, providing a data-driven approach to complement traditional scouting methods.

For example, Smart coach [17] is a hybrid recommendation system for young football athletes, designed to facilitate the interaction between members of a club technical staff and their young athletes, reinforcing the young person counselling, and their potential as an athlete.

Our main contribution is the development of a recommendation system for player scouting which integrates similarity techniques and generative artificial intelligence. This system aims to provide support for the scouting process by providing a data-driven, objective, and inclusive approach to player scouting and recruitment. By integrating advanced machine learning and artificial intelligence techniques, our system can accurately predict player potential and performance, complementing traditional scouting methods and mitigating the influence of subjective biases. This represents a significant step forward in the field of sports analytics and talent identification.

III. FPSRec Architecture

In this section we describe the architecture of our proposed system for recommending professional football players, named *FPSRec*. The architecture is depicted in Figure 1 and consists of four macro-components: *Data Management*, *Similar Players*, *Scouter AI* and *GUI*. The design principles are that of a classical information retrieval system. Specifically, our objective is to create a system that, starting from a player that a team is looking for, calculates the most similar players to the indicated one. From these, the system exploits generative AI to recommend the player best suited to the team's characteristics by generating a report.

In the following, we provide details about each component, method, and technology used.

A. Data Management

In the first component, we include all the aspects related with the management of data. Thus, data collection, data preprocessing and cleaning, data storage, etc. are part of it.

1) Data collection and cleaning: The data utilized for this project was sourced from the Internet, searching websites offering football statistics and information. The source maintains a database encompassing over 200,000 players and teams, tracking an array of statistics. Specifically, for the purposes of this paper, we opted to gather statistics from this source concerning all players participating in the five major European leagues during the 2022-2023 season. The data have been cleaned to improve their quality and to use them effectively. Duplicate rows were identified and removed to prevent redundancy, ensuring the dataset remained concise. Additionally, any missing or incomplete data entries were addressed using appropriate techniques. To maintain uniformity and facilitate subsequent computations, data type conversions were performed, ensuring all values adhered to a consistent numeric format.

2) Feature Selection: After cleaning the data, features were selected in order to identify and retain only the most relevant attributes. This was achieved by calculating and comparing the variances between the features in the dataset. Features with a variance below a predefined threshold were omitted, resulting in a more focused and efficient dataset. In addition, other columns that in context may be unnecessary for the construction of the recommendation system are eliminated. The collected statistics encompass various performance measures crucial for designing the recommendation system. We collected 121 features for 2889 Players.

3) Search server: We use *Apache Solr* to store and index the data. In detail, the cleaned data is transformed into Solr documents and indexed into a *Solr core*. This way, the system is able to implement basic search engine functionalities for

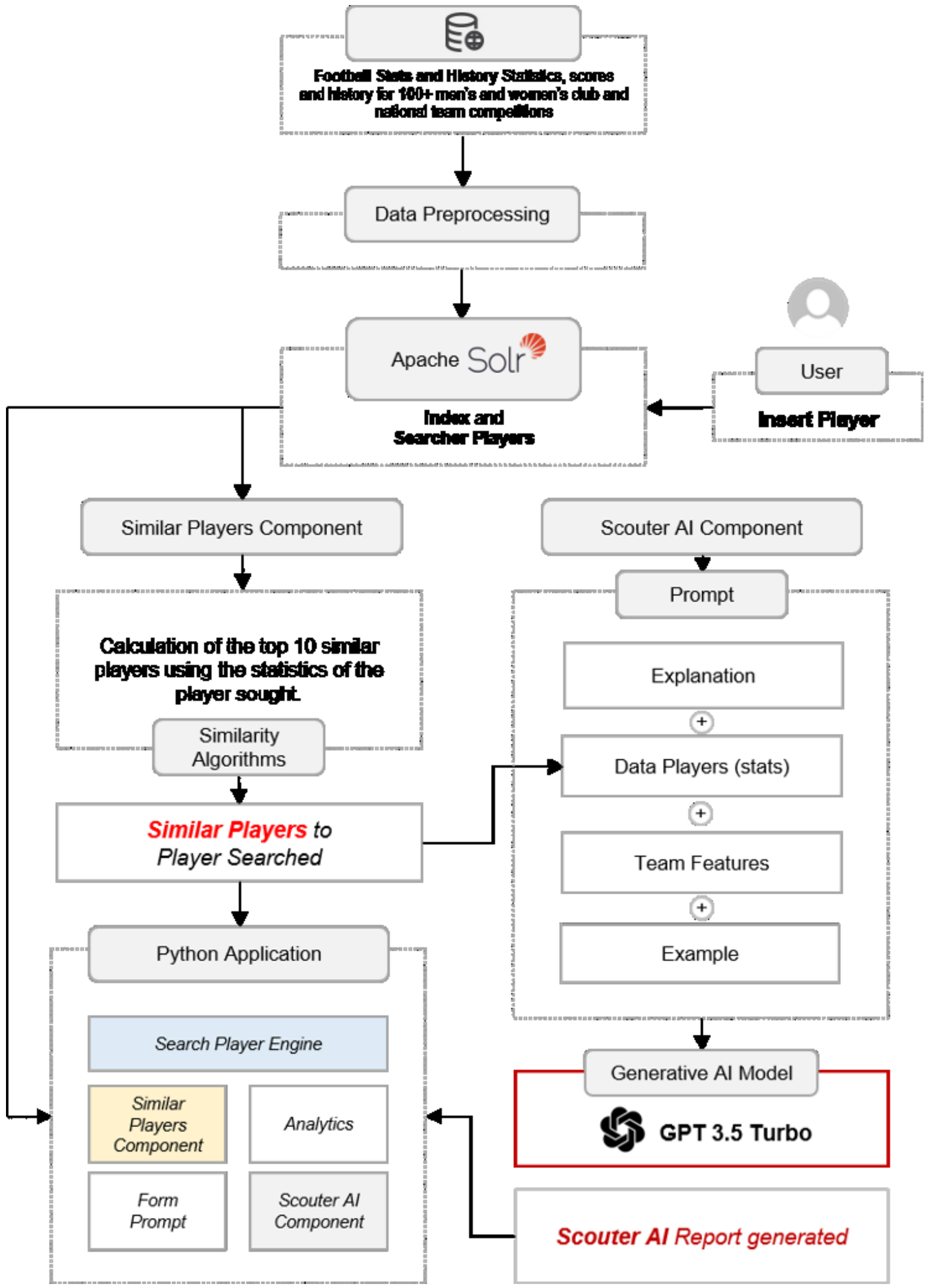


Fig. 1: Overall FPSRec architecture

finding the player and his characteristics. To improve user experience and friendliness of the system, we implemented a suggestion query functionality with auto-completion in the graphical user interface (GUI component).

B. Similar Players Component

With the data preprocessed as described in Section III-A, we proceeded to the design phase of the recommendation system. In our case, that is a football player recommendation system, each player is represented as a vector in which the dimensions represent various player statistics (e.g. goals scored, assists, age, playing time, etc.). The *Similar Players* component employs traditional similarity techniques, such as cosine similarity, to compute players' similarity. To determine the most suitable method for comparing football players and assessing their similarity, we considered two distinct approaches: cosine similarity and clustering with cosine similarity. *MinMaxScaler* normalisation was used to ensure that features with different scales or very large variations have comparable values.

In detail, for the second technique we used *K-means* clustering. We divided the data into K clusters where each cluster is represented by its centroid, that is the mean of the players in that cluster. In order to classify the players into groups, we need a way to compute the (dis)similarity between each pair of players. The result of this computation is known as a dissimilarity or distance matrix. Also in this case, we used the cosine similarity. To select the optimal K, we used the classical *Elbow Method*. The results are shown in Figure 2. As we can notice, the best value for K is 4, which corresponds to the knee of the curve.

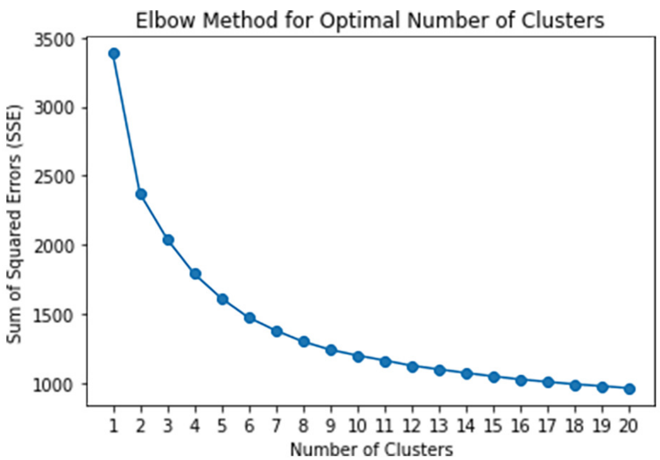


Fig. 2: Results of Elbow Method

Next, the K-Means model is trained on the characteristics of the players, and each player is assigned a cluster label based on its similarity to the other players in the same cluster.

Figure 3 illustrates in detail the *Similar Players* component. Notably, it emphasizes the evaluation of algorithms to determine the most suitable one. If the evaluation results do not meet expectations, adjustments are made to the feature selection during the data preparation phase of the component. This iterative process ensures that the chosen algorithm aligns with the desired outcomes, optimizing the *Similar Players* functionality for player analysis.

To test the goodness of clustering technique used, we used the *Silhouette Score*. Among the tested numbers of clusters, the highest Silhouette Score is achieved when using 4 clusters. A Silhouette Score of 0.274 suggests that the clustering with 4 clusters is relatively better in terms of both cohesion within clusters and separation between clusters compared to other numbers of clusters. Given that the results are close to zero, it indicates that the distance between clusters is not significant. This could suggest that the clustering solution is not optimal. In other words, the results obtained with this clustering solution may not significantly improve compared to a situation where clustering was not applied. More details about the evaluation of this component are provided in Section IV-A.

C. Scouter AI component

The *Scouter AI* component leverages Large Language models (LLMs) to generate a comprehensive textual report using different techniques, including zero-shot and one-shot learning.

Figure 4 offers an overview of this system's component.

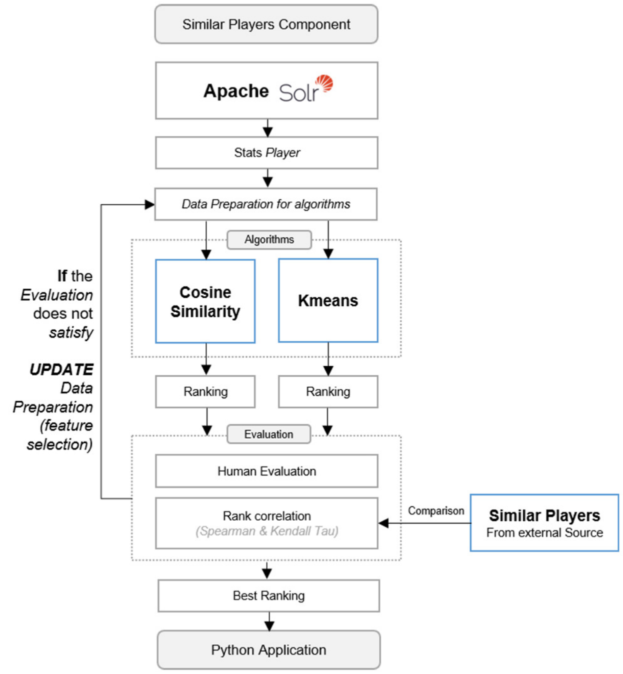


Fig. 3: Similar Players Component

In the present case, GPT-3.5 has been employed to generate a player Scouting recommendation report. This report is generated based on an understanding of the team's characteristics and identifying players who are similar to an ideal player. The decision to use the GPT 3.5 model was based on a detailed analysis of the results obtained during comparative tests with other generative models. From the tests we carried out, as shown in Table I, it emerged that the GPT 3.5 Turbo model was able to generate reports in line with expectations, while ensuring an average execution time of only 24 seconds. This was very beneficial in terms of system efficiency and responsiveness.

TABLE I: Comparison of LLMs performances on our task. The asterisk mark indicates that the model was run locally.

GenAI Models	Average Exec. times (s)	Costs	Results
GPT 3.5 Turbo	24	$0.002 / 1K token	Report generated in line with expectations
LLAMA2/replicate	67	$0.000225 / sec	Report generated in line with expectations but sometimes exceeds token limit
LLAMA2/offline*	486	FREE	Good reports but too much waiting time
Cohere	75	$0.4 / 1M Tokens	Good reports but exceeds token limit for each execution

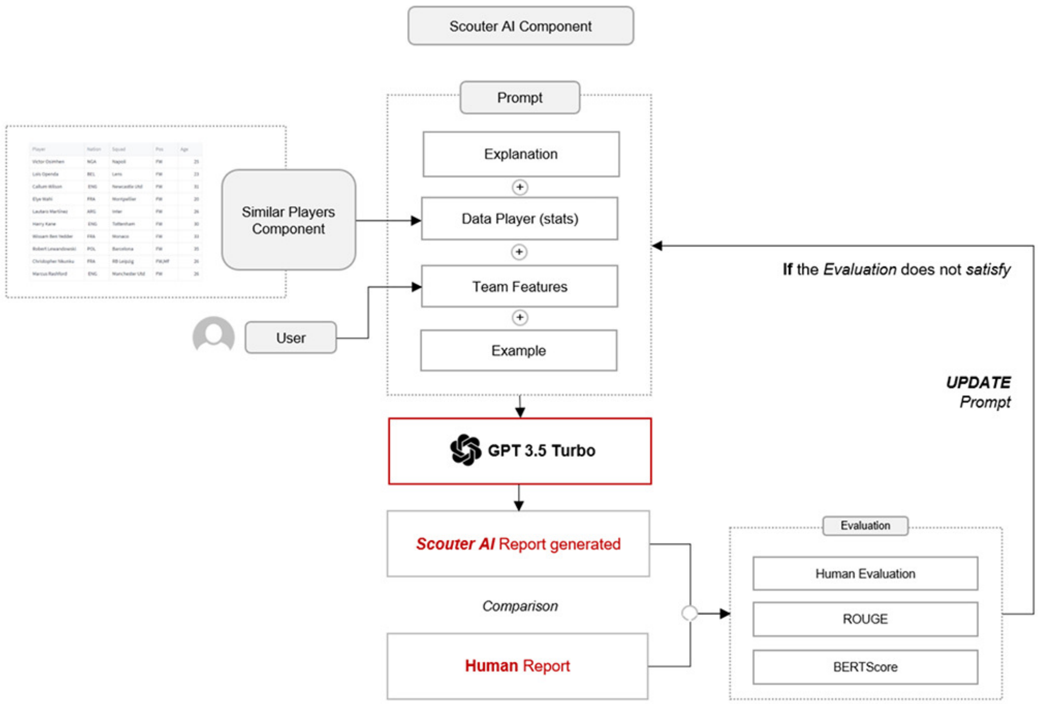


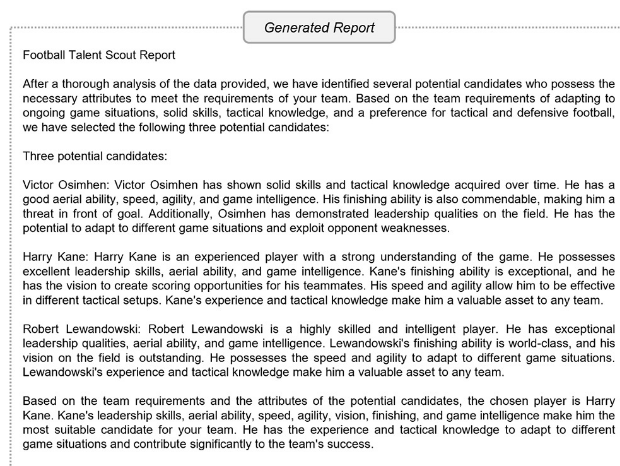
Fig. 4: Scouter AI Component

1) Prompting Techniques: We illustrate here the use of prompt fed to the chosen generative model, that is GPT 3.5, to generate a report. To obtain better results from the generative model, we have investigated the use of various prompting techniques. In particular, we use the following modular structure:

$$P = concat(E, D, F, S) \tag{1}$$

where P represents the Prompt; E represents the explanation of what is wanted from the model; D is the data that the model has to process and summarise, in this case the top 10 similar players; F stands for Feature, that is the team characteristics to be taken into account in the report to be generated; S stands for Show the type of the task, that is the demonstration of an example of output useful for the model.

Several techniques were obtained corresponding to different combination of elements, such as: Zero Shot Prompting ($P = concat(E, F)$), [18]; One Shot Prompting ($P = concat(E, F, S)$) [19]; Augmented Generation One Shot (AGOS) Prompting. The latter technique is proposed by the authors and corresponds to the inclusion of all four elements for building the prompt, as shown in equation 1. In particular, AGOS make use of multiple examples of input (the data D) to generate a more accurate response. Furthermore, by incorporating a "System Role" into the prompt, we introduce a combined AGOS+System Role approach. This addition enables the model to better understand user requests and guides it more precisely in formulating responses. The inclusion of the System Role enhances the model's ability to adhere


Generated Report

Football Talent Scout Report

After a thorough analysis of the data provided, we have identified several potential candidates who possess the necessary attributes to meet the requirements of your team. Based on the team requirements of adapting to ongoing game situations, solid skills, tactical knowledge, and a preference for tactical and defensive football, we have selected the following three potential candidates:

Three potential candidates:

Victor Osimhen: Victor Osimhen has shown solid skills and tactical knowledge acquired over time. He has a good aerial ability, speed, agility, and game intelligence. His finishing ability is also commendable, making him a threat in front of goal. Additionally, Osimhen has demonstrated leadership qualities on the field. He has the potential to adapt to different game situations and exploit opponent weaknesses.

Harry Kane: Harry Kane is an experienced player with a strong understanding of the game. He possesses excellent leadership skills, aerial ability, and game intelligence. Kane's finishing ability is exceptional, and he has the vision to create scoring opportunities for his teammates. His speed and agility allow him to be effective in different tactical setups. Kane's experience and tactical knowledge make him a valuable asset to any team.

Robert Lewandowski: Robert Lewandowski is a highly skilled and intelligent player. He has exceptional leadership qualities, aerial ability, and game intelligence. Lewandowski's finishing ability is world-class, and his vision on the field is outstanding. He possesses the speed and agility to adapt to different game situations. Lewandowski's experience and tactical knowledge make him a valuable asset to any team.

Based on the team requirements and the attributes of the potential candidates, the chosen player is Harry Kane. Kane's leadership skills, aerial ability, speed, agility, vision, finishing, and game intelligence make him the most suitable candidate for your team. He has the experience and tactical knowledge to adapt to different game situations and contribute significantly to the team's success.

Fig. 5: An example of report generated by generative AI model GPT 3.5

to specified guidelines and improves the overall quality and relevance of its outputs. A deep and extensive evaluation of the different techniques is provided in Section IV-B, while Figure 5 shows an example of scouting report generated by GPT 3.5 model.

D. GUI

For the realisation of the user interface (UI), we employed the Streamlit framework. The latter was customised to meet the specific requirements of our system, including the integration of a Solr-based search engine to identify players, the implementation of a component dedicated to the evaluation of similar players, and the use of generative AI for the generation of scout reports. Figure 6 illustrates the user interface of the application created with the streamlit framework.

E. Technologies Utilized

This study was entirely developed in Python (version 3.10.2), leveraging several key libraries, each serving specific functions:

- **LangChain** (version 0.0.275): This library facilitated interaction with OpenAI and management of the generative AI model, GPT-3.5. It enabled the system to generate detailed reports based on specific inputs.
- **Scikit Learn**: Utilized for calculating the Cosine Similarity between players and executing the K-Means algorithm, aiding in the division of players into clusters based on their characteristics.
- **Pandas** (version 1.4.3): Employed for data manipulation tasks, including the organization and analysis of football player data.
- **Scipy** (version 1.9.3): Utilized for computing the correlation between players' rankings, enabling a deeper evaluation of their performance.
- **RougeScore** (version 0.1.2): Employed for assessing the quality of generated reports based on metrics comparing them with reference texts.
- **BERTScore** (version 0.3.13): This metric was used to compute the semantic similarity between the text generated by the AI model and the reference text, providing an indication of generation quality.
- **Streamlit** (version 1.23.1): The Streamlit framework was adapted to the project's requirements to create an intuitive user interface. It facilitated the implementation of a player search engine with automatic suggestion functionality.

Furthermore, a simple player search engine was designed and implemented using **Apache Solr** (version 8.11.2), with interaction in Python managed through the **SolrClient** (version 0.3.1) library. Solr was employed for efficient indexing and searching of player data, thereby enhancing the overall user experience within the system.

FPSRec provides several advantages and implications. It streamlines the scouting process, reducing subjectivity and manual effort. It provides a comprehensive, data-driven evaluation of potential recruits, aiding the decision-making process. This innovative architecture allows to integrate traditional scouting methods with advanced machine learning techniques, providing a comprehensive, objective, and data-driven approach to player scouting and recruitment.

IV. System Evaluation

A. Similar Component evaluation

To assess the results obtained from the two similarity approaches considered, we employed Rank Correlation and Human Evaluation. Specifically, Rank Correlation was evaluated based on some "Top-10 Similar Players" rankings provided by FBRef[1] website.

Here, we provide an example of how we test our system and to demonstrate the evaluation process. The user submits a query, i.e. "Erling Haaland", that is, we aim to find players that are most similar to the forward of Manchester City. Table II shows the two rankings obtained from the two similarity approaches we considered:

1) Human Evaluation: By observing the results, we can immediately notice evident differences between the rankings obtained from the two approaches. The domain expert asserts that the ranking evaluated by cosine similarity has proven superior to that of K-means. Indeed, the results appear to be more "realistic" in the case of cosine similarity. Expanding the evaluations to include additional players tested further confirms the superiority of cosine similarity over clustering.

2) Rank Correlation Evaluation: We are now interested in assessing the disparities between the two approaches employed and a ranking derived from a reliable source like FBRef. Table III shows the ranking provided by FBRef to the same query.

To accomplish this, we utilized traditional rank correlation metrics, in particular Spearman's and Kendall's Tau correlation coefficients to measure the "agreement" between the rankings.

A more comprehensive comparison of similarity techniques with FBref ranking, based on 3 queries, is reported in Table IV.

From the results obtained it is possible to state that cosine similarity is the best solution, among the two considered approaches, to construct the recommendation system as the values of the respective coefficients were better. In fact, there are more negative correlations in the K-means out of n evaluations than in the other solution. In addition, the use of Spearman is considered as it is necessary to evaluate the differences between the relative positions of the players and not only the exact positions, as is the case with Kendall Tau.

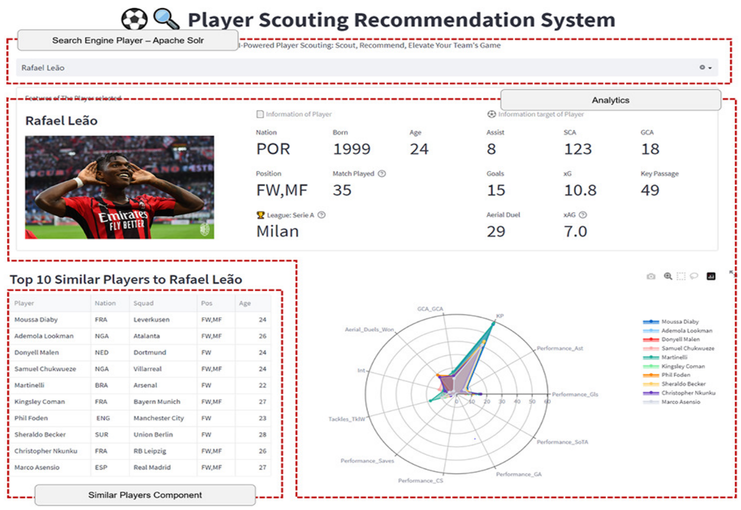
B. Prompting techniques evaluation

To evaluate the reports generated by the models for each technique used to construct the prompts, we employ human evaluation and automated methods to evaluate the quality of generated text.

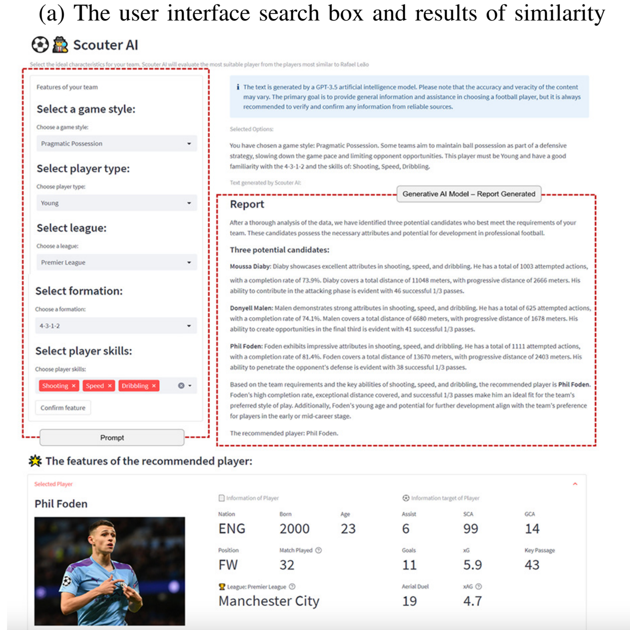
1) Human Evaluation: The human evaluation process involves domain experts who examine and evaluate the texts generated by the text generation model according to several criteria, including completeness, clarity and relevance of content. Experts assign scores according to these criteria, and the sum of the scores reflects the overall quality of the report generated by the model using each prompting technique. Table V shows the results obtained employing human evaluation.

2) Metrics-based Evaluation: Human evaluation is a valuable method for assessing generative models outputs, however it suffers from subjectivity and it is prone to bias. Different human evaluators may have varying opinions, and the evaluation criteria may lack consistency. Additionally, human evaluation

[1] https://fbref.com/



(a) The user interface search box and results of similarity



(b) Scout report AI generated shown to the user

Fig. 6: The user interface of the Python application

TABLE II: Top-10 similar players ranking provided by K-means (left) and Cosine Similarity (right) for Erling Haaland

Player	Nation	Squad
Wissam Ben Yedder	FRA	Monaco
Harry Kane	ENG	Tottenham
Robert Lewandowski	POL	Barcelona
Lautaro Martínez	ARG	Inter
Christopher Nkunku	FRA	RB Leipzig
Loïs Openda	BEL	Lens
Victor Osimhen	NGA	Napoli
Marcus Rashford	ENG	Manchester Utd
Elye Wahi	FRA	Montpellier

Player	Nation	Squad
Victor Osimhen	NGA	Napoli
Loïs Openda	BEL	Lens
Callum Wilson	ENG	Newcastle Utd
Elye Wahi	FRA	Montpellier
Lautaro Martínez	ARG	Inter
Harry Kane	ENG	Tottenham
Wissam Ben Yedder	FRA	Monaco
Robert Lewandowski	POL	Barcelona
Christopher Nkunku	FRA	RB Leipzig

TABLE III: Top-10 similar players ranking provided by FBRef for Erling Haaland

Player	Nation	Squad
Callum Wilson	ENG	Newcastle United
Dušan Vlahović	SRB	Juventus
Loïs Openda	BEL	RB Leipzig
Robert Lewandowski	POL	Barcelona
Cyle Larin	CAN	Mallorca
Terem Moffi	NGA	Nice
Folarin Balogun	USA	Monaco
Marcus Thuram	FRA	Internazionale
Victor Osimhen	NGA	Napoli
Ollie Watkins	ENG	Aston Villa

TABLE IV: Comparison of similarity techniques

Players	Cosine Similarity		Kmeans	
	Kendall Tau	Spearman	Kendall Tau	Spearman
Rafael Leao	0.022	0.03	0.288	0.333
Victor Osimhen	0.288	0.309	-0.022	0.042
Erling Haaland	0.333	0.454	-0.288	-0.321

TABLE V: Human rating

Model	Prompt	Human Evaluation	Human Rating
GPT 3.5	Zero Shot	Completeness	1/5
		Clarity	3/5
		Relevance of Content	2/5
	One Shot	Completeness	3/5
		Clarity	4/5
		Relevance of Content	3/5
	Augmented One Shot	Completeness	4/5
		Clarity	4/5
		Relevance of Content	4/5
	Augmented One Shot + Role	Completeness	4/5
		Clarity	4/5
		Relevance of Content	4/5

can be time-consuming and expensive, especially for large-scale evaluations. In spite of these observations, we decided to evaluate the prompting techniques using metrics, such as Rouge [20] and BERTScore [21]. These metrics provide an objective assessment of the consistency and accuracy of the generated reports by comparing the generated text with a human reference text. Table VI shows the evaluation results obtained.

It is important to emphasize that the results were obtained through a meticulous process involving the generation of three reports for each prompt and their comparison with summaries created by the author, followed by the averaging of scores. First and foremost, it is evident that the addition of content to the prompts has led to an improvement in performance, aligning with expectations. This suggests that providing the model with a greater amount of information can positively influence the quality of the generated responses. However, ROUGE was chosen, as the main focus is on the ability of the generative model to cover most of the relevant information present in the human reference text (recall). From the results of ROUGE, it is evident that Rouge-1 scored higher than ROUGE-2 and ROUGE-L, as the higher specificity of the latter two metrics may make it more challenging to achieve high scores. However, ROUGE-2 and ROUGE-L simultaneously represent more precise measures of consistency and relevance in the generated text and assess the exact overlap between human reference and generated report. Therefore, BERTScore is used to assess the semantic similarity between the generated text and the reference text. The high scores obtained with BERTScore, with an F1-Score of around 90%, indicate that the system is able to generate text that is semantically similar to the human reference text, affirming the effectiveness of generation by the GPT 3.5 model. However, both ROUGE and BERTScore provide information on the text generation capabilities of the GPT 3.5 model but do not measure the informative quality and fidelity of the generated text [22]. In light of these analyses, the choice to adopt the Augmented Generated One Shot Prompt with System Role for the final application appears to be a well-founded decision, considering the overall performance achieved.

V. Conclusions and Future Work

The proposed system offers several advantages, including decision support for professional football teams, similarity-based recommendations, and report generation. Considering its embryonic stage of development, preliminary results have shown great potential. In fact, these features can streamline the player selection process and provide detailed insights for evaluation. However, the system also has its limitations. It relies on data updated within the last year, which may limit the accuracy of recommendations in a rapidly changing context. The quality of recommendations is also dependent on the completeness and accuracy of input data. Furthermore, the AI may not consider intangible factors such as team chemistry or player personalities, that may be relevant to purchasing decisions. Technological limitations of the current AI models and the need for human verification further underscore the challenges of the system.

The project also faces limitations related to the dataset used, the potential for inaccuracies in report generation, and constraints of the AI model, including token limitation and costs. The use of a dataset limited to the year 2022/2023 may not fully reflect a player's entire football career, potentially influencing the recommendations. The report generation by a generative AI model can lead to hallucinations or unreliable information. The model's token limitation may result in a limited length of reports generated, and the costs associated with using generative AI models such as GPT 3.5 can incur significant costs, potentially limiting the system's accessibility to other users. While the system offers valuable features that

TABLE VI: Report generation evaluation results

Model	Prompt	Evaluation Metrics	Rouge			BERTScore
			Rouge-1	Rouge-2	Rouge-L	
GPT 3.5	Zero Shot	Precision	0.606	0.235	0.266	0.883
		Recall	0.508	0.191	0.221	0.876
		F1-Measure	0.544	0.207	0.238	0.879
	One Shot	Precision	0.613	0.266	0.301	0.897
		Recall	0.587	0.254	0.298	0.892
		F1-Measure	0.591	0.256	0.295	0.895
	Augmented One Shot	Precision	0.616	0.263	0.329	0.905
		Recall	0.611	0.249	0.327	0.903
		F1-Measure	0.611	0.250	0.326	0.904
	Augmented One Shot + Role	Precision	0.630	0.324	0.378	0.912
		Recall	0.663	0.328	0.328	0.908
		F1-Measure	0.639	0.324	0.381	0.910

can aid in player selection and evaluation, it is essential to be aware of its limitations and the need for thorough human evaluation before making final player acquisition decisions. In this context, it is essential to engage with experts in the football industry to deepen the understanding of the needs and expectations of teams and coaches, allowing for further customization of the system to meet their specific demands. Future improvements could focus on addressing these limitations to enhance the system's accuracy and reliability. Indeed, our future research direction is aimed at several key areas. Firstly, we plan to enrich the dataset by integrating data from multiple sources [23] including data from several past years, which will provide a more comprehensive view of players' careers and enhance the precision of player similarity evaluations. Secondly, the project intends to advance its generative AI techniques by exploring and incorporating novel models to reduce hallucinations and improve the consistency of the generated reports. This might include the adoption of more recent and sophisticated models such as GPT-4, LLAMA3, or Google Gemini. Thirdly, there is an ongoing effort to optimize prompts for generative AI models to ensure the coherence and pertinence of the generated reports, which could involve discovering new prompt patterns that more effectively steer the text generation process. Lastly, we aim to include multi-modal interfaces to enable visual query posing [24] by users to extend the capabilities of the system.

Acknowledgments

This study was partially supported by the PNRR MUR project PE0000013-FAIR.

Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] M. Montanaro, A. M. Rinaldi, C. Russo, and C. Tommasino, "Using knowledge graphs for audio retrieval: a case study on copyright infringement detection," *World Wide Web*, vol. 27, no. 4, p. 37, 2024.

[2] Q. Wang, M. Li, X. Wang, N. Parulian, G. Han, J. Ma, J. Tu, Y. Lin, H. Zhang, W. Liu *et al.*, "Covid-19 literature knowledge graph construction and drug repurposing report generation," *arXiv preprint arXiv:2007.00576*, 2020.

[3] C. Russo, K. Madani, A. M. Rinaldi, and C. Sabourin, "Multimedia knowledge design and processing for personal robots," in *Proceedings of the 11th International Conference on Management of Digital EcoSystems*, 2019, pp. 316–323.

[4] C. Russo, K. Madani, and A. M. Rinaldi, "An unsupervised approach for knowledge construction applied to personal robots," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 1, pp. 6–15, 2020.

[5] Y. Li and B. Starly, "Building a knowledge graph to enrich chatgpt responses in manufacturing service discovery," *Journal of Industrial Information Integration*, vol. 40, p. 100612, 2024.

[6] D. Piggott and B. Muir, "Modern approaches to scouting and recruitment," in *Science and Soccer*. Routledge, 2023, pp. 382–394.

[7] A. P. Singh and M. Suguna, "Data-driven player recruitment in football," in *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*. IEEE, 2023, pp. 844–850.

[8] P. Larkin and D. O'Connor, "Talent identification and recruitment in youth soccer: Recruiter's perceptions of the key attributes for player recruitment," *PLOS one*, vol. 12, no. 4, p. e0175716, 2017.

[9] N. Jokuschies, V. Gut, and A. Conzelmann, "Systematizing coaches' 'eye for talent': Player assessments based on expert coaches' subjective talent criteria in top-level youth soccer," *International Journal of Sports Science & Coaching*, vol. 12, no. 5, pp. 565–576, 2017.

[10] T. L. Bergkamp, A. S. M. Niessen, R. J. den Hartigh, W. G. Frencken, and R. R. Meijer, "Methodological issues in soccer talent identification research," *Sports Medicine*, vol. 49, pp. 1317–1335, 2019.

[11] T. L. Bergkamp, W. G. Frencken, A. S. M. Niessen, R. R. Meijer, and R. J. den Hartigh, "How soccer scouts identify talented players," *European Journal of Sport Science*, vol. 22, no. 7, pp. 994–1004, 2022.

[12] M. K. Christensen, "“an eye for talent”: Talent identification and the “practical sense” of top-level soccer coaches," *Sociology of sport journal*, vol. 26, no. 3, pp. 365–382, 2009.

[13] T. R. Bell and S. E. Rosado, "Hoopgurlz's biased recruiting: Racial and positional stereotyping in girls' basketball scouting reports," *Feminist Media Studies*, pp. 1–19, 2023.

[14] Y. Li, K. Liu, R. Satapathy, S. Wang, and E. Cambria, "Recent developments in recommender systems: A survey," *IEEE Computational Intelligence Magazine*, vol. 19, no. 2, pp. 78–95, 2024.

[15] K. Pasrija and K. Mittal, "A survey of recent advances in recommendation systems," *IOT with Smart Systems: Proceedings of ICTIS 2021, Volume 2*, pp. 589–597, 2022.

[16] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-based systems*, vol. 46, pp. 109–132, 2013.

[17] P. Matos, J. Rocha, R. Gonçalves, A. Almeida, F. Santos, D. Abreu, and C. Martins, "Smart coach—a recommendation system for young football athletes," in *Ambient Intelligence–Software and Applications–,*

10th International Symposium on Ambient Intelligence. Springer, 2020, pp. 171–178.

[18] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, "Finetuned language models are zero-shot learners," *arXiv preprint arXiv:2109.01652*, 2021.

[19] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[20] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Text summarization branches out*, 2004, pp. 74–81.

[21] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating text generation with bert," *arXiv preprint arXiv:1904.09675*, 2019.

[22] D. Deutsch and D. Roth, "Understanding the extent to which summarization evaluation metrics measure the information quality of summaries," *arXiv preprint arXiv:2010.12495*, 2020.

[23] A. M. Rinaldi, C. Russo, and K. Madani, "A semantic matching strategy for very large knowledge bases integration," *International Journal of Information Technology and Web Engineering (IJITWE)*, vol. 15, no. 2, pp. 1–29, 2020.

[24] A. M. Rinaldi, C. Russo, and C. Tommasino, "Visual query posing in multimedia web document retrieval," in *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*. IEEE, 2021, pp. 415–420.