FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Efficient Fine-Tuning of Large Events Models for Player-Team Fit Estimation in Soccer Analytics

João Tomás Teixeira



Master in Informatics and Computing Engineering

Supervisor: Prof. João Pedro Mendes Moreira

Second Supervisor: Prof. Tiago Neves

July 21, 2026

© João Tomás Teixeira, 2026

Efficient Fine-Tuning of Large Events Models for Player-Team Fit Estimation in Soccer Analytics

João Tomás Teixeira

Master in Informatics and Computing Engineering

Approved in oral examination by the committee:

President: Prof. André Restivo
Referee: Prof. Carlos Ferreira
Supervisor: Prof. João Moreira

July 21, 2026

Resumo

As transferências de jogadores são decisões fundamentais no futebol moderno, tanto pelo seu impacto desportivo como pelo seu peso financeiro. Apesar do crescimento da análise de dados no futebol, estimar se um jogador se adapta a uma nova equipa continua a ser um problema difícil. O desempenho de um jogador depende do contexto em que atua, incluindo o modelo de jogo, os colegas, os adversários, o papel tático e a forma como as ações se desenvolvem ao longo do jogo. Por essa razão, estatísticas e modelos tradicionais de força de equipa são úteis como referência, mas não conseguem representar totalmente a relação entre comportamento individual, identidade coletiva e dinâmica sequencial do jogo.

Esta dissertação propõe e avalia uma estrutura metodológica para estimar o ajuste jogador-equipa no futebol. A abordagem combina dados de eventos, perfis táticos de jogadores e equipas, embeddings, um Large Events Model condicionado por embeddings, adaptação específica por equipa através de Low-Rank Adaptation (LoRA), e simulação de jogos e cenários de transferência. O objetivo não é prever o sucesso de uma transferência como um resultado fixo. Em vez disso, a estrutura metodológica procura gerar sinais interpretáveis sobre compatibilidade tática, impacto esperado e alterações no comportamento coletivo após a introdução de um jogador numa nova equipa.

A avaliação considera diferentes níveis de análise. Primeiro, o modelo base é testado através da simulação de jogos completos e do cálculo de pontos esperados. Depois, adaptadores LoRA específicos por equipa são treinados e avaliados em métricas de validação ao nível dos eventos, eficiência computacional, reprodução de resultados de época e métricas de estilo tático. Por fim, são simulados cenários de transferência para analisar alterações em pontos esperados, golos esperados, perfil tático e deslocamento no espaço de embeddings.

Os resultados mostram que a adaptação LoRA consegue reduzir a perda de validação e capturar sinais persistentes de identidade coletiva. Nas simulações de época, o modelo identifica de forma razoável a hierarquia das equipas mais fortes, mas comprime os pontos previstos em direção à média da liga. Assim, equipas dominantes tendem a ser subestimadas e equipas mais fracas tendem a ser sobrestimadas. As simulações de transferência produzem indicações úteis e interpretáveis, especialmente quando analisadas em conjunto com métricas táticas e movimento no espaço de embeddings. No entanto, os resultados também mostram que o modelo deve ser entendido como uma ferramenta de apoio à decisão, e não como um modelo determinístico de previsão de transferências.

Palavras-chave: Análise de dados de futebol • Large Events Models • Low-Rank Adaptation • Ajuste jogador-equipa • Simulação de transferências

Abstract

Player transfers are critical decisions in modern football because they carry both sporting and financial consequences. Although data analysis has become increasingly important in football, estimating whether a player will fit a new team remains difficult. Player performance depends on context, including the tactical system, teammates, opponents, assigned role, and the way actions unfold throughout a match. For this reason, aggregate statistics and traditional team-strength models are useful as references, but they cannot fully represent the relationship between individual behaviour, collective identity, and the sequential dynamics of the game.

This dissertation proposes and evaluates a framework for estimating player-team fit in football. The approach combines event data, player and team behavioural profiles, embeddings, an embedding-conditioned Large Events Model, team-specific adaptation through Low-Rank Adaptation (LoRA), and simulation of matches and transfer scenarios. The goal is not to predict transfer success as a fixed outcome. Instead, the framework aims to generate interpretable signals about tactical compatibility, expected impact, and changes in collective behaviour after introducing a player into a new team.

The evaluation considers different levels of analysis. First, the base model is tested through full-match simulation and expected points calculation. Then, team-specific LoRA adapters are trained and evaluated using event-level validation loss, computational efficiency, season outcome reproduction, and tactical style metrics. Finally, transfer scenarios are simulated to analyse changes in expected points, expected goals, tactical profile, and movement within the embedding space.

The results show that LoRA adaptation can reduce validation loss and capture stable patterns of team identity. In season simulations, the model identifies the hierarchy of the strongest teams reasonably well, but it compresses predicted points toward the league average. As a result, dominant teams tend to be underestimated and weaker teams tend to be overestimated. Transfer simulations produce useful and interpretable signals, especially when analysed together with tactical metrics and embedding-space movement. However, the results also show that the framework should be understood as a decision-support tool, rather than a deterministic model for predicting transfer success.

Keywords: Football analytics • Large Events Models • Low-Rank Adaptation • Player-team fit • Transfer simulation

The United Nations Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework for achieving a better and more sustainable future for all. It includes 17 goals (https://sdgs.un.org/goals) to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice.

This dissertation is mainly aligned with SDG 9 and SDG 8. The contribution is indirect, since the work does not address sustainability through a direct social or environmental intervention. Instead, it contributes through engineering research, technological innovation, and data-driven decision support. The proposed framework applies artificial intelligence methods to football analytics, with the goal of improving player-team fit estimation and supporting more informed recruitment decisions.

The specific Sustainable Development Goals mentioned have the following names:

SDG 9 Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation

SDG 8 Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

SDG	Target	Contribution	Performance Indicators and Metrics
9	9.5	The dissertation contributes to sports analytics research by developing a LEM-based framework for player-team fit estimation.	Validation loss reduction after LoRA adaptation; comparison with baseline models; number of successful transfer simulation scenarios; computational efficiency metrics.
8	8.2	The framework supports more informed recruitment decisions in football. This can help clubs reduce uncertainty in transfer decisions and allocate financial resources more effectively.	Expected points variation after simulated transfers; ranking and outcome error against real season data.

Acknowledgements

I would like to express my gratitude to my supervisor for his availability, guidance, and for all the knowledge he shared throughout the development of this dissertation.

I am also grateful to my co-supervisor for his interest in this work, for his support, and for the valuable knowledge and perspectives he shared during its development.

I would like to thank my colleagues who accompanied me over these years. Their friendship and shared experiences made this academic path richer and more memorable.

Finally, to my family, especially my parents and my sister, I am grateful for all their support, patience, and encouragement throughout this journey.

João Tomás Teixeira

Declaration of honour

I declare that this dissertation is my own work and has not previously been submitted or assessed at this or any other institution. References to other authors (statements, ideas, thoughts), as well as the use of published works, strictly comply with the rules of attribution and are duly identified in the text and in the bibliographic references, in accordance with the applicable referencing standards. I am aware that the practice of plagiarism or self-plagiarism constitutes an academic offence, as established in the U.Porto Code of Ethics.
I also declare that I used Artificial Intelligence tools to support the writing and language revision of parts of this dissertation. The use of these tools and their outputs was carefully reviewed, edited, and checked by me to avoid errors, inaccuracies, unfounded attributions, bias, or misrepresentation. I take full responsibility for the final content, arguments, interpretations, and authorship of this dissertation.

João Tomás Teixeira



Contents

Acknowledgements v

1 Introduction 1
- 1.1 Context . . . 1
- 1.2 Motivation . . . 2
- 1.3 Problem . . . 3
- 1.4 Research Questions . . . 3

2 Literature Review 5
- 2.1 Introduction . . . 5
- 2.2 Review Methodology . . . 5
  - 2.2.1 Type of Review . . . 5
  - 2.2.2 Search Strategy . . . 6
  - 2.2.3 Screening and Selection Process . . . 6
- 2.3 Football Analytics and Event Data . . . 7
  - 2.3.1 Evolution from Traditional Statistics . . . 7
  - 2.3.2 Event-Based Data in Football . . . 7
  - 2.3.3 Modeling Football as Sequences of Actions . . . 7
- 2.4 Transfer Evaluation and Player-Team Fit . . . 8
- 2.5 Model Adaptation and Fine-Tuning Methods . . . 9
  - 2.5.1 Overview and Context in Football Analytics . . . 9
  - 2.5.2 Theoretical Foundations of Model Adaptation . . . 10
- 2.6 Summary and Identified Research Gap . . . 12

3 State of the Art 13
- 3.1 Introduction . . . 13
- 3.2 State of the Art Review . . . 14
  - 3.2.1 Traditional Statistical and Team-Strength Models . . . 14
  - 3.2.2 Large Events Models . . . 14
  - 3.2.3 Player–Team Fit Estimation . . . 15
  - 3.2.4 Parameter-Efficient Fine-Tuning . . . 19
- 3.3 Positioning of this Dissertation . . . 22

4 Methodological Approach 24
- 4.1 Overview of the Proposed Framework . . . 24
- 4.2 Main Study Assumptions . . . 25
- 4.3 Data Preparation and Event Normalization . . . 26
- 4.4 LEM Event Representation . . . 26

4.5 Behavioural Profiles and Embeddings . . . 28
4.5.1 Player Behavioural Profiles . . . 28
4.5.2 Player Embeddings . . . 29
4.5.3 Team Embeddings . . . 30
4.6 Embedding-Conditioned Base LEM . . . 31
4.7 LoRA Team Adaptation . . . 32
4.8 Simulation Procedure and Expected Points Calculation . . . 33
4.8.1 Sequential Event Generation . . . 33
4.8.2 Dynamic Embedding Assignment . . . 33
4.8.3 Match Termination . . . 34
4.8.4 Expected Points Calculation . . . 34
4.9 Transfer Simulation . . . 34
4.10 Baseline Models . . . 35
4.10.1 Elo Baseline . . . 35
4.10.2 Player-Aware Expected Points Baseline . . . 36
4.10.3 Base LEM Baseline . . . 36
4.11 Summary . . . 37

5 Experimental Design **38**
5.1 Overview . . . 38
5.2 Experimental Scenarios . . . 38
5.2.1 Base LEM Evaluation . . . 39
5.2.2 LoRA Team Adaptation Evaluation . . . 40
5.2.3 Transfer Simulation Scenarios . . . 40
5.3 Evaluation Metrics . . . 40
5.3.1 Event-Level Metrics . . . 41
5.3.2 Match Outcome Metrics . . . 41
5.3.3 Tactical Style Metrics . . . 41
5.3.4 Embedding-Space Fit Metrics . . . 42
5.3.5 Baseline Comparison Metrics . . . 42
5.3.6 Computational Efficiency Metrics . . . 43
5.4 Baseline Comparison Design . . . 43
5.4.1 Elo Baseline . . . 43
5.4.2 Player-Aware Expected Points Baseline . . . 43
5.4.3 Base LEM Baseline . . . 44
5.5 Validation Strategy and Experimental Limitations . . . 44
5.6 Summary . . . 45

6 Results and Discussion **46**
6.1 Overview . . . 46
6.2 Base LEM Simulation Results . . . 46
6.2.1 Initial Simulation Behaviour . . . 46
6.2.2 Training Corrections . . . 47
6.2.3 Interpretation of Base LEM Results . . . 47
6.3 LoRA Team Adaptation Results . . . 48
6.3.1 Event-Level Validation Loss . . . 48
6.3.2 Adaptation Capacity and Training Configuration . . . 49
6.3.3 Computational Efficiency . . . 50
6.3.4 Interpretation of LoRA Adaptation Results . . . 51

6.4 Full-Season Team Simulation Results . . . 51
6.4.1 Outcome-Level Results . . . 51
6.4.2 Comparison with Outcome Baselines . . . 53
6.4.3 Tactical Style Reproduction . . . 56
6.4.4 Interpretation of Full-Season Simulation Results . . . 58
6.5 Transfer Simulation Results . . . 58
6.5.1 Transfer Scenarios and Fit Indicators . . . 59
6.5.2 Interpretation for Recruitment Decision Support . . . 68
6.6 Discussion of Main Findings and Limitations . . . 69
6.6.1 Summary of Main Findings . . . 69
6.6.2 Limitations Observed in the Results . . . 70
6.6.3 Implications for Player-Team Fit Estimation . . . 71

7 Conclusion **72**
7.1 Summary of the Work . . . 72
7.2 Answers to the Research Questions . . . 72
7.3 Contributions . . . 75
7.4 Limitations . . . 75
7.5 Future Work . . . 76
7.6 Final Remarks . . . 77

References **78**

A Supplementary Figures **80**
A.1 Supplementary VAE Representations . . . 80
A.2 LoRA Adapter Training Diagnostics . . . 84
A.3 Supplementary Full-Season Simulation Tables . . . 86

List of Figures

2.1 Schematic of a match sequence evaluated using the VAEP framework, Source: [2]. 10

3.1 The diagram shows how LoRA adjusts the pretrained weight matrices by introducing two small matrices, A and B, of rank r that capture the update while keeping the original parameters fixed. Source: [14]. 21

4.1 Overview of the proposed pipeline for player–team fit estimation. 25
4.2 Projection of player embeddings into two dimensions using PCA. 30

6.1 Validation cross-entropy improvement from LoRA adaptation by team. Higher values indicate a larger reduction in validation loss compared with the non-adapted base LEM. 49
6.2 Real season points versus simulated expected points for the base LEM and the LoRA-adapted LEM. The dashed line represents perfect agreement between real and simulated points. 52
6.3 Aggregate outcome-level error comparison between the base LEM and the LoRA-adapted LEM. Lower values indicate closer agreement with real season values. . 52
6.4 2022/2023 season outcome reference for the Low-Rank Adaptation (LoRA)-adapted LEM simulation. 54
6.5 Elo rating table for the 2022/2023 season. 55
6.6 Player-aware expected points table for the 2022/2023 season. 56
6.7 Tactical style error comparison between the base LEM and the LoRA-adapted LEM. Lower style MAE indicates closer tactical profile reproduction. 57
6.8 Tactical metric improvement from the base LEM to the LoRA-adapted LEM. Positive values mean that LoRA reduced the error compared with the base model. . . 58
6.9 Projection of the simulated Gyökeres to Sporting CP transfer in team embedding Principal Component Analysis (PCA) space. The arrow shows the movement from the original Sporting CP embedding to the simulated post-transfer embedding. . . 60
6.10 Change in simulated expected points, goals scored, and goals conceded after the Gyökeres to Sporting transfer scenario. Values represent the difference between the post-transfer simulation and the baseline Sporting simulation. 61
6.11 Relative tactical change after the simulated transfer of Gyökeres to Sporting. The bars show the percentage change in each tactical metric compared with the baseline Sporting simulation. 62
6.12 Relative tactical change radar for the simulated Gyökeres to Sporting CP transfer. The baseline Sporting CP profile is indexed to 100, and the transfer profile is shown as a percentage of that baseline. 63
6.13 Projection of the simulated Ugarte to Benfica transfer in team embedding PCA space. 64

6.14 Change in simulated expected points, goals scored, and goals conceded after the Ugarte to Benfica transfer scenario. 64
6.15 Relative tactical change after the simulated transfer of Ugarte to Benfica. 65
6.16 Average simulated action-location shift for the Ugarte to Benfica scenario. The markers compare Benfica's baseline average action location with the simulated post-transfer location. 66
6.17 Projection of the simulated Vinícius Jr. to Porto transfer in team embedding PCA space. 67
6.18 Change in simulated expected points, goals scored, and goals conceded after the Vinícius Jr. to Porto transfer scenario. 67
6.19 Relative tactical change after the simulated transfer of Vinícius Jr. to Porto. . . . 68

A.1 VAE player identity map for the 2022/2023 data. 81
A.2 Projection of the simulated Vinícius Júnior to Porto transfer in the team VAE latent space. 82
A.3 Projection of the simulated Manuel Ugarte to Benfica transfer in the team VAE latent space. 83
A.4 Projection of the simulated Viktor Gyökeres to Sporting CP transfer in the team VAE latent space. 84
A.5 Training loss evolution for the Benfica LoRA adaptation under different epoch limits. 85
A.6 Validation loss evolution for the Benfica LoRA adaptation under different epoch limits. 85
A.7 Complete predicted league table generated by the base LEM for the 2022/2023 Liga Portugal season. 86
A.8 Complete predicted league table generated by the LoRA-adapted LEM for the 2022/2023 Liga Portugal season. 86

List of Tables

3.1 Comparison of Modeling Approaches in Football Analytics 16

4.1 Core event variables used in the LEM representation. 27
4.2 Context variables used in the LEM representation. 27
4.3 Examples of metric groups used to describe playing style. 29

5.1 Research questions and supporting evidence 39

6.1 Computational efficiency summary for each adapter. 50
6.2 Embedding-space fit indicators for the simulated transfer scenarios. 60

List of Acronyms

xT Expected Threat

LEMs Large Events Models

LLMs Large Language Models

LoRA Low-Rank Adaptation

PCA Principal Component Analysis

PEFT Parameter-Efficient Fine-Tuning

SDG Sustainable Development Goal

VAE Variational Autoencoder

VAEP Valuing Actions by Estimating Probabilities

Chapter 1

Introduction

Player transfers are among football's most critical and financially significant decisions, with clubs investing millions in the hope that a player's skills and style will integrate effectively into a new tactical system. However, predicting this fit is highly challenging. Traditional methods, including descriptive statistics and scouting, cannot fully capture the complex ways in which a player's actions mesh with team dynamics and tactical frameworks. This creates uncertainty that often results in costly transfer misjudgments, exposing clubs to both sporting and financial risks [11] [4].

Advances in football analytics, particularly the availability of detailed event data, have enabled new approaches to this problem. These datasets, recording sequences of passes, shots, and duels, can be interpreted as a "language of football." Large Events Models (LEMs), based on machine learning architectures, can learn from millions of such event sequences to simulate matches and evaluate player performance in context. However, these models are generally trained as general-purpose tools and are not optimized for specific questions, such as player-team fit.

This thesis develops and evaluates a simulation-based framework for estimating player-team fit in football. The framework combines event-level modeling, player and team behavioural embeddings, parameter-efficient team adaptation through Low-Rank Adaptation (LoRA), and Monte Carlo match simulation. It examines whether the proposed framework can capture team tactical identity, reproduce realistic match and style metrics, and provide useful player-team fit signals when compared with simpler baselines.

1.1 Context

Modern football analytics is undergoing a transformation driven by the availability of detailed event-level data. Each match can now be represented as a sequence of structured actions — passes, shots, duels, recoveries — each annotated with time, location, player identity, and contextual features such as pressure or possession phase. This data allows analysts to move beyond summary statistics like goals or assists and instead capture how the game unfolds. For example, a single attacking sequence might include a chain of short passes in midfield, a through ball into space, and

a cutback leading to a shot. Represented as a series of timestamped events, it becomes possible to study not just outcomes but sequential interactions: how positioning, tempo, and decision-making combine to create scoring opportunities.

LEMs build on the success of Large Language Models (LLMs), treating football matches as sequences of structured events (passes, shots, or duels), similar to words in a sentence. Using sequence-learning architectures, LEMs learn the contextual dependencies between actions, capturing how early events influence later ones within a possession or match. Trained on millions of sequences from multiple leagues, they develop general representations of football dynamics that can simulate matches, reveal tactical patterns, and evaluate player performance.

This work explores how fine-tuned LEMs can move from describing past events to forecasting how a player's actions might integrate into a new tactical environment, producing interpretable signals that can support club decision-making.

1.2 Motivation

The football transfer market has evolved into one of the most financially significant ecosystems in global sport. In 2024 alone, clubs worldwide spent over 10 billion euros on player transfers, with top European leagues accounting for more than half of that total [15]. These transactions are not mere sporting decisions, they are strategic business investments that can determine a club's competitive and financial trajectory. A single transfer can reshape commercial revenues, fan engagement, and brand value, but it can also cause severe financial strain when expectations are not met.

Yet, despite the scale of these investments, evaluating whether a player will truly fit a new club remains one of football's greatest uncertainties. A player who excels in one environment may fail to adapt in another due to tactical mismatches, differing roles, or cultural factors within the team. When transfers go wrong, clubs face not only sporting setbacks but also substantial financial losses through high wages and depreciated market value.

This challenge is therefore both a sporting and economic imperative. With the rapid growth of large-scale event data and the emergence of LEMs, the opportunity now exists to approach this problem in a more structured and context-aware way. By leveraging these models to capture the underlying patterns of play and context-specific performance, it becomes possible to simulate how a player's style might align with a team's tactical identity.

The motivation for this thesis lies in this convergence: the business need to reduce uncertainty in transfer decisions, and the technological potential of advanced machine learning models to provide data-driven evidence. Bridging these perspectives offers a path toward more sustainable recruitment strategies and a deeper understanding of how performance, context, and economics intertwine in modern football.

1.3 Problem

Player performance in football is highly dependent on context. A player's output is shaped not only by individual ability, but also by the team's tactical system, the surrounding teammates, the opponent's behaviour, the player's role, and the sequences of events in which the player is involved. Traditional statistics and team-strength models can provide useful baselines, but they have limitations. They can describe past performance and estimate broad team quality, but they struggle to represent how actions unfold over time and how well a player's behaviour fits a specific tactical environment [13].

In this dissertation, player-team fit is defined as the degree to which a player's behavioural profile can be introduced into a team's tactical context while producing coherent outcome and style signals. For this reason, fit is not treated as a binary label, where a player either fits or does not fit. Instead, it is interpreted through changes in expected points, goals, tactical style metrics, and movement within the embedding space. This definition also recognises that fit can have different meanings depending on the recruitment objective: continuity, where the incoming player preserves the existing tactical identity of the team and reduces disruption; a compatible upgrade, where the player improves expected outcome or style indicators while remaining aligned with the team's current way of playing; or a transformational signing, where the player changes the team's tactical profile in a deliberate direction. This last case may be valuable when a club wants to evolve its playing model, but it also carries higher uncertainty because the player is not simply replacing an existing profile. This requires a modeling approach that can represent both team context and event-sequence behaviour.

LEMs offer a promising way to address this problem because they model football as a sequence of structured events. In this setting, all actions are treated as connected events rather than isolated statistics. However, a general LEM trained on broad football data does not automatically learn the tactical identity of a specific team or the behavioural compatibility between a player and that team. Fully retraining the model for each team is also not ideal, since team-level datasets are small and can lead to overfitting.

This dissertation addresses this gap by developing and evaluating a framework for player-team fit estimation. The framework combines behavioural player and team embeddings, an embedding-conditioned LEM, LoRA-based team adaptation, and Monte Carlo match simulation. The goal is not to predict transfer success as a fixed outcome. Instead, the work evaluates whether the framework can reproduce team-specific tactical tendencies and provide useful player-team fit signals when compared with simpler baselines, such as Elo and player-aware expected points models.

1.4 Research Questions

Building upon these motivations and challenges, this thesis addresses the following research questions:

- How can pre-trained LEMs be efficiently fine-tuned to capture team-specific football contexts?
- What trade-off does LoRA provide between adaptation quality and computational efficiency when fine-tuning LEMs for team-specific simulation?
- Can player and team behavioural embeddings derived from event-level data capture tactical identity in football?
- How can transfer simulations provide interpretable player-team fit signals for recruitment decision support?

These research questions are evaluated through the methodology and experiments developed in the following chapters. The first question is assessed by adapting the base LEM to individual teams and comparing event-level validation loss before and after team-specific adaptation. The second question is evaluated by comparing adaptation quality with computational efficiency indicators. The third question is addressed through player and team embeddings, tactical style metrics, and embedding-space analysis. Finally, the fourth question is evaluated through simulated transfer scenarios, where changes in expected points, goals, tactical profile, and embedding-space position are used as interpretable fit signals.

Chapter 2

Literature Review

2.1 Introduction

This chapter reviews the main contributions and gaps in the fields of football analytics and machine learning for sports. It explores how researchers have used event-level data to analyze player behavior, tactical systems, and performance prediction. Special attention is given to LEMs, which treat football as a sequence of structured actions and enable context-aware performance assessment.

The chapter also outlines the methodology used to collect, screen, and select the literature included in this review. This process ensures transparency and allows other researchers to replicate the same search and obtain comparable results.

2.2 Review Methodology

This section describes how the literature used in this thesis was identified and selected. The review followed a structured approach, instead of a full systematic literature review. This distinction is important because the objective was not to exhaustively cover all publications in football analytics or machine learning. Instead, the goal was to identify the most relevant work for the development of this dissertation, with particular focus on event-based football analytics, LEMs, model adaptation, and parameter-efficient fine-tuning.

2.2.1 Type of Review

This study follows a structured literature review approach with two distinct components. The first component focused on LLMs, examining their architectures, training methodologies, and applications across various domains. The goal was to identify, evaluate, and summarize existing research on LEMs and their sequential prediction capabilities. For the football analytics component, rather than conducting a separate systematic review, this research relied predominantly on the work by Tiago Mendes-Neves, "Estimating Player Performance in Different Contexts Using Fine-tuned Large Events Models," and its cited references. This foundational paper introduces

LEMs, an innovative application of LLM principles to football analytics, which predict sequences of match events rather than words, enabling player performance evaluation across different team contexts. By building upon this established framework and its reference literature, the study leverages proven methodologies for event-based modeling and machine learning approaches to player assessment.

2.2.2 Search Strategy

The search strategy combined database search and citation tracking. Scopus was used as the primary academic database, supplemented by Google Scholar and arXiv when relevant. The database search focused mainly on LLMs, sequential modeling, transfer learning, domain adaptation, and parameter-efficient fine-tuning. Core keywords included large language models, event-based models, attention mechanisms, sequential modeling, pre-training, fine-tuning, transfer learning, domain adaptation, LoRA, and parameter-efficient fine-tuning. Boolean operators were used to combine and refine search terms, and the search covered publications approximately between 2010 and 2025.

For the football analytics component, the search was narrower and more problem-driven. The starting point was the work by Mendes-Neves et al. (2024) [13]. From this paper, relevant references were selected based on their connection to event data, action valuation, player evaluation, team context, and sequence-based modeling in football. Additional sources were then identified through citation tracking and targeted searches for specific concepts such as expected threat, Valuing Actions by Estimating Probabilities (VAEP), and Elo models.

2.2.3 Screening and Selection Process

After collecting the initial set of papers from the search on LLMs, duplicates were removed and all titles and abstracts were reviewed to exclude studies not related to transformer architectures, attention mechanisms, sequential modeling, or pre-training methodologies. The remaining papers were then examined to confirm their relevance and quality, focusing on works that presented original research on LLM architectures, training strategies, fine-tuning approaches, or applications to sequential prediction tasks.

Only studies that presented original research, applied quantitative or machine learning methods to language modeling, or provided useful insights into event-based models and their adaptation to different domains were included. Papers that were opinion-based or lacked methodological detail were excluded. For the football analytics component, no separate screening process was conducted. Instead, this research built directly upon the work by Mendes-Neves et al. (2024) [13], utilizing its cited references and identifying additional relevant studies through backward and forward citation tracking.

2.3 Football Analytics and Event Data

Event-level data has transformed football analysis by shifting the paradigm from aggregate statistics to granular, action-by-action understanding of match dynamics. Modern football analytics now captures every on-ball action with precise timestamps, spatial coordinates, and contextual attributes, enabling analysts to decode the sequential nature of the game rather than summarizing outcomes.

2.3.1 Evolution from Traditional Statistics

Traditional football analysis relied on summary statistics such as goals, assists, possession percentages, and shot counts. These metrics provided limited insight into the complex tactical dynamics and causal relationships between actions. The advent of event-based datasets marked a paradigm shift, allowing researchers to move beyond "what happened" to understand "how and why it happened" [9].

2.3.2 Event-Based Data in Football

In football analytics, events represent discrete, observable on-ball actions that occur during a match, capturing specific moments of play with detailed contextual information. Events include actions such as passes, shots, tackles, dribbles, fouls, substitutions, and cards, with each event characterized by its type, timestamp, involved player(s), spatial coordinates on the field, and outcome (e.g., whether a pass was successful or a shot was on target). Modern event-data collection processes are increasingly automated using computer vision and machine learning algorithms that analyze video footage to detect and classify events [19], though human verification remains standard for guaranteeing data quality.

Unlike tracking data, which captures continuous player positioning throughout a match, event data focuses on the moments of high significance, the discrete actions that define tactical decisions and performance outcomes. This structured representation enables comprehensive match analysis, performance evaluation, tactical assessment, and player comparison across different contexts. For this project, the Wyscout Open Access Dataset is used, which provides spatio-temporal event data from major European leagues and international competitions.

2.3.3 Modeling Football as Sequences of Actions

While event-based datasets capture discrete actions such as passes, shots, or tackles, their true analytical potential emerges when these events are interpreted as part of a broader sequence of play. Football is inherently sequential, every decision affects what comes next, and understanding these dependencies allows models to capture the game's underlying structure rather than isolated moments.

Early modeling efforts approached this by grouping consecutive actions into possession sequences, describing how teams progressed the ball and changed game states. Frameworks such as

the Expected Threat (xT) model applied this logic by valuing each on-ball action according to how it increased or decreased the probability of scoring, quantifying the impact of movement across pitch zones [17].

Building on this, action-based models like VAEP evaluated each action's contribution to a team's likelihood of scoring or conceding within a short time frame. These models integrated contextual features such as location, action type, and match situation to represent both risk and reward in decision-making [13].

2.4 Transfer Evaluation and Player-Team Fit

Transfer evaluation and player-team fit remain difficult problems because player performance depends heavily on context. A player's output is shaped by more than individual ability, it also depends on the tactical system, the role assigned by the coach, the quality and behaviour of teammates, the opponent's structure, and the types of actions the player is asked to perform. A player who performs well in one team may not produce the same level of performance in another team if the new environment changes these conditions.

Traditional approaches usually evaluate players through aggregate indicators such as goals, assists, pass completion, shot volume, defensive actions, or market-related variables. These measures are useful because they are easy to interpret and can describe past performance. Regression-based models can also identify links between player output and contextual factors, such as teammate quality or opponent strength [1]. However, these approaches treat performance as a set of summary values. They do not fully show how a player's actions emerge within the flow of the game.

This creates a challenge for player-team fit analysis. Aggregate statistics can indicate what a player has done, but they say less about how that performance was produced. They do not capture how actions connect across possessions, how teammates create or restrict passing options, or how tactical structure changes the value of the same behaviour. For example, a progressive pass may carry a different meaning in a possession-based team than in a direct team that attacks space quickly. In this sense, fit should be understood not only as expected output, but also as behavioural and tactical compatibility between the player and the team.

These limitations motivate the use of event-based and sequence-based approaches. By representing football as a sequence of connected actions, these methods can capture aspects of performance that aggregate models miss. They provide a way to study how players act within tactical systems, how teams create recurring patterns of play, and how player behaviour may change across contexts. This makes them relevant for player-team fit analysis, where the central question is not only whether a player is good, but whether the player's style can work inside a specific team environment.

2.5 Model Adaptation and Fine-Tuning Methods

This section draws primarily from the work of Mendes-Neves et al. (2024), [13], which introduces an innovative application of LEMs to football analytics. Building on this article and its direct references, this overview examines how model adaptation is currently approached in football analytics, explores existing methodologies for adjusting models to specific contexts, and identifies the limitations that motivate the need for more sophisticated approaches to player-team fit evaluation.

2.5.1 Overview and Context in Football Analytics

Contemporary football analytics has developed various methodologies to adjust models to specific contexts, though most remain context-agnostic. Possession value frameworks, such as the VAEP model, treat player evaluation as a machine learning challenge by calculating the probability of an action leading to a goal within subsequent actions, as illustrated in Figure 2.1. Similarly, the xT metric assigns values to different pitch areas based on scoring likelihood from those zones.

However, regression-based player rating systems encounter substantial practical difficulties, including low explanatory power, statistically insignificant coefficients, and "wrong" signs in estimated relationships. The fundamental limitation is that regression-based weightings reflect statistical predictive power rather than game importance from a coaching perspective [6]. Furthermore, these models tend to attribute the highest weights to end-of-sequence variables like shots and goals, obscuring the contributions of build-up play.

Mendes-Neves et al. introduced an innovative application of LEMs that addresses these contextual limitations through fine-tuning techniques. Analogous to LLMs in natural language processing, LEMs learn the "language" of football by predicting variables for subsequent events rather than words, enabling match simulation and player performance prediction across different team contexts. The methodology involves training a foundational LEMs on broad datasets (Ligue 1 and Bundesliga data for example) and subsequently fine-tuning these models with team-specific data to reflect tactical dynamics.

The fine-tuning process creates context-specific models by reshaping training data to include only relevant subsets: team-specific behavior, player-specific contributions, player additions to new teams, or player replacements accounting for outgoing squad members. Applied to the 2017-2018 Premier League season using the WyScout dataset, this approach successfully evaluated hypothetical transfers, such as analyzing the potential impact of Cristiano Ronaldo or Lionel Messi joining different Premier League teams. The results demonstrated that player impact varies substantially based on team dynamics and existing squad composition, for instance, Messi performed better than Ronaldo for Watford due to the quality of players he would replace (Roberto Pereyra or André Carrillo), while the opposite occurred at Leicester where Ronaldo would replace weaker alternatives.

	TIME	PLAYER	ACTION	P_{scores}	VALUE
1	92m4s	S. Busquets	pass	0.03	0.00
2	92m6s	L. Messi	pass	0.02	- 0.01
3	92m8s	S. Busquets	pass	0.03	+ 0.01
4	92m11s	L. Messi	take on	0.08	+ 0.05
5	92m12s	L. Messi	pass	0.17	+ 0.09
6	92m14s	A. Vidal	shot	1.00	+ 0.83

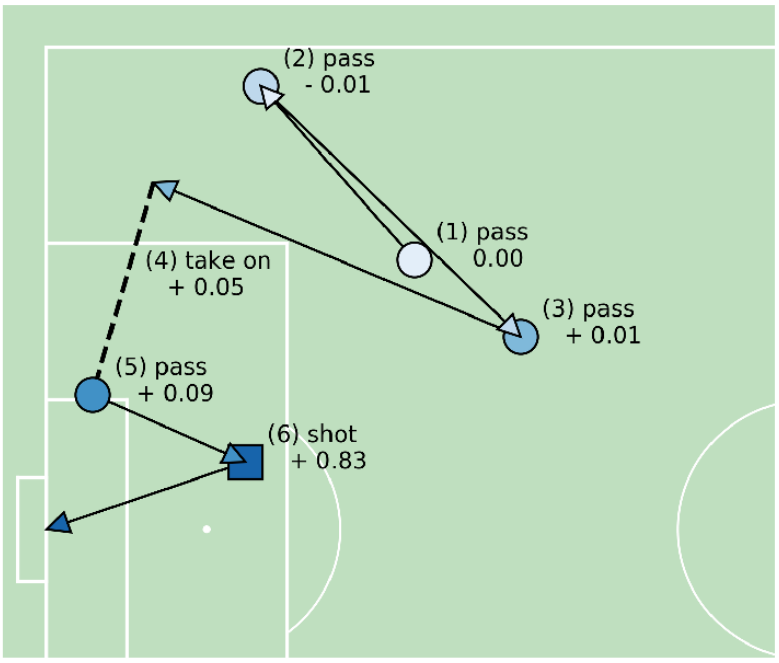


Figure 2.1: Schematic of a match sequence evaluated using the VAEP framework, Source: [2].

2.5.2 Theoretical Foundations of Model Adaptation

Model adaptation builds upon the principles of **transfer learning**, a framework designed to reuse knowledge acquired in one task or domain (the *source*) to improve learning performance in another (the *target*) [21]. This paradigm is especially relevant when labeled data is abundant in the source domain but scarce or costly to obtain in the target domain, a situation commonly encountered in football analytics, where extensive league-wide data exists, but team-specific data is limited.

Within transfer learning, three main categories are distinguished:

- **Inductive transfer learning**, where the model learns to apply knowledge from a source task to a different but related target task.
- **Transductive transfer learning**, where the learning task remains the same but the model adapts to a different data domain.

- **Unsupervised transfer learning**, where both the source and target tasks are unsupervised, and the goal is to transfer learned representations or structures between domains.

Within transductive transfer learning, a relevant subset is **domain adaptation**, which assumes that the source and target domains share the same feature space (e.g., event representations) but have different data distributions [21]. In football analytics, this translates to adapting models trained on league-wide data (source domain) to a specific team context (target domain), where tactical style, player profiles, and match dynamics diverge.

Two main approaches dominate research in domain adaptation:

1. **Adversarial adaptation**, which uses a discriminator network to promote domain-invariant feature representations.

2. **Distribution alignment**, which explicitly minimizes statistical distances (e.g., Maximum Mean Discrepancy) between source and target feature spaces.

The central theoretical challenge is avoiding **negative transfer**, which occurs when adaptation harms performance relative to training solely on target data. This risk increases when the divergence between domains is large or when the target dataset is small [20].

Parameter-Efficient Fine-Tuning

To mitigate these limitations in large-scale models, **parameter-efficient fine-tuning (PEFT)** methods have become prominent. Instead of updating all parameters of a pre-trained model, PEFT techniques modify only a small subset, dramatically reducing computational and memory requirements. Among these, LoRA is particularly effective, it decomposes weight updates into low-rank matrices that can be merged with the base model at inference time, enabling efficient adaptation without latency penalties [7]. LoRA's efficiency makes it well suited to football analytics, where models must adapt to multiple teams with limited available data.

Technical Challenges

Although they offer benefits, model adaptation techniques face technical challenges that must be managed to ensure successful deployment in football analytics contexts.

Overfitting represents a primary concern, particularly when fine-tuning on small team-specific datasets where models may achieve high performance on training data while failing to generalize to unseen match scenarios. This challenge is exacerbated in football analytics, where teams may play fewer than 40 matches per season, limiting the available data for adaptation.

Data scarcity imposes a constraint on the fine-tuning process, as many practical applications involve limited labeled data in the target domain. In football analytics, this manifests as insufficient match event data for specific team contexts, particularly for newly promoted teams or those with limited historical data coverage.

Computational cost continues to pose a challenge despite parameter-efficient methods, as fine-tuning still requires specialized hardware and careful resource management, therefore the balance between model expressivity (number of trainable parameters) and computational feasibility must be optimized based on available resources and task requirements.

2.6 Summary and Identified Research Gap

The introduction of LEMs has demonstrated that treating football matches as sequences of structured events enables context-aware performance prediction and match simulation. Besides, fine-tuning methodologies, particularly parameter-efficient techniques such as LoRA, have shown effectiveness in adapting pre-trained models to team-specific contexts with limited computational overhead. Studies have successfully applied these models to evaluate hypothetical transfers, revealing that player impact varies based on tactical systems and existing squad composition [13]. Furthermore, research in transfer learning and domain adaptation has established theoretical frameworks for knowledge transfer across different distributions while minimizing negative transfer risks. These advances collectively represent significant progress in moving from descriptive statistics toward sequential, context-aware player evaluation.

However, existing work still leaves an open gap in how player-team fit can be estimated as an interpretable and operational decision-support problem. Current approaches show that event-sequence models can simulate player impact in different contexts, but they do not fully connect team-specific adaptation, player and team behavioural embeddings, outcome simulation, tactical style reproduction, and comparison with simpler baselines within a single evaluation framework. As a result, it remains unclear whether efficient fine-tuning of LEMs can provide useful fit signals beyond broad team-strength or aggregate player-aware models.

Building on this gap, this dissertation investigates how LoRA can adapt a general event-sequence model to team-specific contexts while using limited team data. It also evaluates whether behavioural embeddings and simulation outputs can capture tactical identity and support player-team fit analysis. The scientific gap addressed by this work is therefore the lack of an integrated, parameter-efficient, and interpretable framework for estimating player-team fit from event-level football data, linking sequence modeling, team adaptation, embedding-based representation, and simulation-based evaluation.

Chapter 3

State of the Art

This chapter reviews the existing research supporting this dissertation. It analyzes how football analytics evolved from traditional statistics to event-based approaches and explains how modern datasets enable sequential modeling of the game. It then examines current methods for evaluating player-team fit and highlights their limitations, particularly when dealing with the contextual nature of football performance. The chapter also explores recent advances in model adaptation and parameter-efficient fine-tuning, focusing on their relevance for LEMs. By combining insights from these areas, the chapter identifies the gaps that motivate this research and lays the foundation for the methodology developed in the following sections.

3.1 Introduction

Understanding player performance and predicting how a player will adapt to a new team requires a solid view of the current state of football analytics and sequential modeling. Research in this area has evolved quickly, moving from simple descriptive indicators to detailed event-level datasets and advanced machine learning architectures. These developments made it possible to model football as a sequence of actions rather than a collection of isolated statistics, providing a deeper understanding of tactical behaviors and player contributions.

For this dissertation, these advances serve a practical purpose. The goal is not only to review the evolution of event-based models and LEMs but to demonstrate how their core ideas can be leveraged and adapted to estimate player–team fit. Existing methods provide valuable foundations, yet they fall short when the task requires understanding how a player's action patterns interact with a specific tactical environment. Recent research on LEMs shows that event sequences can capture context and dependencies between actions, but these models still need efficient adaptation methods to reflect team-specific behaviour.

This chapter therefore focuses on the aspects of the literature most relevant for the methodology developed later in this work. It examines how sequential modeling can represent player

behaviour in context, where traditional approaches face limits, and why parameter-efficient fine-tuning methods such as LoRA offer a suitable way to adapt general LEMs to team-specific analysis. This framing supports the broader goal of estimating player-team fit through tactical, behavioural, and outcome-oriented signals, rather than treating transfer evaluation as a single prediction task.

3.2 State of the Art Review

3.2.1 Traditional Statistical and Team-Strength Models

Traditional football analytics often starts with aggregate statistics. These include goals, assists, shots, pass completion, defensive actions, duels, and other summary measures. These statistics are useful because they are easy to read and compare. A scout, coach, or analyst can look at them and quickly understand part of a player's profile. However, they only show a summary of what happened. They do not show how the actions happened, how they were connected, or how the team's style shaped them.

Team-strength models offer another simple way to estimate performance. Elo rating systems are a common example. Elo uses past match results to estimate how strong a team is. After each match, the rating changes based on whether the result was better or worse than expected [8]. This makes Elo useful as a baseline for match outcome estimation. It gives a clear view of team strength over time. However, Elo works at team level, it does not explain player behaviour, tactical style, event sequences, or how a player may fit into a different team.

A stronger baseline can add player-level information to expected points estimation. In this type of model, player data is first converted into team-level indicators. These indicators can then be used to estimate the probability of a win, draw, or loss [16]. This makes the model more detailed than Elo because it includes squad information, not only past team results. It can also support simple counterfactual analysis, such as estimating how expected points would change if a player were removed or replaced. Even so, the model remains based on summary data. It does not show how players interact during possessions or how actions build on each other during a match.

3.2.2 Large Events Models

The emergence of LEMs is linked to the availability of detailed event-level football data and to advances in sequence models originally developed for natural language processing. Early approaches such as Expected Threat (xT) and VAEP demonstrated that football actions can be valued by modeling how sequences unfold, but these frameworks still relied on hand-engineered features and limited context awareness. The introduction of multi-layer perceptron models brought a step-change, allowing event sequences to be interpreted similarly to natural language, where each action plays a role analogous to a token in a sentence. This shift enables models to learn dependencies between passes, carries, duels, and shots across entire possessions, as shown by Mendes-Neves et al [13].

However, general-purpose LEMs are not enough on their own for player-team fit analysis. Football performance depends on context, and a player's contribution depends on the tactical system, surrounding teammates, and the team's preferred build-up and pressing patterns. Traditional regression-based methods struggle with this complexity, often producing unstable coefficients and limited explanatory power [6], while descriptive metrics fail to capture sequential dependencies in play [1]. Action-value models such as VAEP move closer to contextual evaluation, but they still do not simulate how a player's behaviour may change inside a new tactical environment [2]. A comparative overview of these modeling paradigms and their respective limitations is provided in Table 3.1.

These limitations make LEMs relevant for player-team fit analysis because they model football as a sequence of connected actions. Unlike aggregate statistics, LEMs can represent how each action depends on previous events, match context, and team behaviour. This allows player and team performance to be studied as part of a wider event structure, rather than as isolated outputs. However, a general LEM is not automatically adapted to a specific team. A model trained on broad football data may learn general event patterns, but it may not fully capture the tactical identity of a target team or the behavioural fit between a player and that team. This creates the need for adaptation methods that can specialize the model to team-specific contexts while keeping the general football knowledge learned during pre-training.

Because team-level datasets are small, this thesis employs parameter-efficient fine-tuning techniques, particularly LoRA [7], to adapt pretrained LEMs without overfitting. LoRA introduces low-rank adapters that modify only a subset of parameters, enabling rapid specialization while preserving general football knowledge. This directly addresses the risk of negative transfer noted in transfer learning research [20]. The result is a scalable and computationally efficient framework: a single foundational LEM can be adapted to multiple teams and scenarios with minimal cost, making it suitable for real-world recruitment processes.

In summary, LEMs in this thesis move beyond descriptive analysis and become practical tools for context-aware transfer forecasting. By combining broad pre-training, team-specific specialization, and efficient fine-tuning, the thesis uses LEMs as a tool to study player-team fit in a more contextual way. This approach builds directly on recent advances in sequential football modeling [13] and offers a richer foundation for analysing player-team fit than traditional statistical or rule-based methods.

3.2.3 Player–Team Fit Estimation

In this thesis, player–team fit is approached as a problem of behavioural adaptation, rather than a comparison of isolated performance indicators. The central idea is that a player's contribution is not fixed; it evolves when the tactical system, role expectations, and interaction patterns around the player change.

Most existing methods quantify what a player has done in the past, but they do not estimate how the player will behave when placed inside a new team's structure. These approaches also assume that a player's historical action patterns are portable across teams, which rarely holds

Table 3.1: Comparison of Modeling Approaches in Football Analytics

Approach	**What It Models**	**Strengths**	**Limitations**	**Relevance to This Thesis**
Traditional Statistics	Aggregated player or team outputs (e.g., goals, assists).	Simple to compute; widely used in scouting.	Ignores sequence, context, and tactical interactions.	Used only as a baseline; insufficient for transfer prediction.
xT / VAEP Models	Action value within possession sequences.	Captures contribution to scoring or conceding; context-aware.	Limited modeling of long-range dependencies; not team-specific.	Provides key conceptual foundation for action valuation.
Sequence Models	Dependencies between consecutive events across possessions.	Learns flow of play; captures temporal structure.	Not specialised for team contexts; needs large data to train.	Precursor to Large Events Models.
Large Events Models (LEMs)	Full event sequences with rich context.	High expressive power; scalable; captures tactical dynamics.	General-purpose; needs adaptation for specific teams or players.	Core modeling framework used in this thesis.
Fine-Tuned LEMs (LoRA)	Player behaviour embedded into team-specific tactical context.	Efficient adaptation; avoids overfitting; preserves base knowledge.	Depends on quality and size of team-level data.	Central method for estimating player–team fit in this thesis.

when team tempo, build-up strategy, or pressing behaviour differ. This thesis adopts the opposite assumption: fit emerges from the interaction between a player's behavioural tendencies and the team's tactical identity. As soon as these two components are combined, new event sequences arise. The same player may take fewer risks in a system that circulates the ball more slowly, progress vertically more often in a direct team, or reduce dribbling frequency when surrounded by teammates positioned closer to the ball. These behavioural changes cannot be inferred from aggregated performance metrics or similarity-based matching.

By framing fit as a dynamic transformation of event-level behaviour, this thesis establishes the conceptual foundation for using a sequence-based model capable of simulating how a player's decisions evolve inside a new tactical environment. This shift is essential because it aligns with the way football behaviour is generated, through sequences of interdependent actions shaped by

system constraints. This perspective directly motivates the fine-tuning strategy and simulation-based approach developed later in the thesis.

Capturing Team Tactical Identity

A pre-trained LEM provides a broad understanding of football dynamics, following the approach introduced in Mendes-Neves et al., 2024 [13]. Instead of modeling isolated statistics, LEMs treat football as a sequence of interconnected events and learn how actions unfold across different tactical contexts. This gives the model a general representation of match behaviour that can later be specialised for team-specific analysis.

However, general-purpose LEMs alone cannot capture the specific behavioral signature of an individual team. Each team exhibits unique tactical patterns shaped by coaching philosophy, player profiles, and strategic priorities, these are characteristics that manifest in event-level tendencies but remain hidden within league-wide models. To represent a specific team's identity, this thesis fine-tunes the pre-trained LEM using LoRA, a parameter-efficient method that injects trainable low-rank matrices into the model's architecture without retraining all parameters [7]. LoRA enables efficient adaptation with limited team-specific data, preserving the foundational football knowledge while embedding team-specific dynamics.

Fine-tuning with team-specific event sequences embeds the team's behavioral signature into the model, including:

- **Build-up structure and tempo** — How quickly the team advances possession and the typical number of passes per sequence [18];
- **Direction and intensity of progression** — Whether the team favors vertical penetration, wide circulation, or cautious horizontal play;
- **Spacing tendencies reflected in event outcomes** — Patterns of spatial occupation that influence passing options, dribbling frequency, and positional density;
- **Risk tolerance** — Propensity to attempt through balls, carry the ball into pressure, or recycle possession.

This process creates a specialized LEM variant that learns recurring event patterns associated with the team. Once adapted, the model can simulate how any player would behave under that team's tactical constraints, translating individual action tendencies into team-specific context. This capability is essential for estimating player-team fit, as it enables the evaluation of whether a player's natural style aligns with or conflicts with the receiving team's embedded tactical identity.

Representing Player Behaviour

While capturing team tactical identity provides the structural framework for this work, understanding individual player behaviour is equally essential. This thesis builds a player representation based on action sequences, not aggregated statistics. Traditional player evaluation relies on

summary metrics such as goals, assists, or pass completion rates, which obscure the sequential dependencies and contextual decision-making patterns that define how a player actually contributes to team play [13]. In contrast, sequence-based representations preserve the temporal and spatial structure of player actions, enabling models to capture behavioural tendencies rather than isolated outcomes.

Player embeddings in this framework are derived from event-level behavioural profiles computed across multiple matches. These embeddings capture recurring tendencies in how a player participates in play, including:

- **Passing behaviour** — pass accuracy, long-pass tendency, crossing tendency, and involvement in possession sequences;
- **Progression and ball-carrying** — carry frequency, progressive action rate, forward progress per action, and directness;
- **Action profile and spatial tendencies** — shot involvement, dribble share, duel participation, recovery actions, wide or central action share, final-third involvement, and attacking-half activity.

These embeddings provide a summary of how a player typically behaves across different match contexts. Rather than describing what a player has achieved, they encode how the player acts. This distinction is important because player-team fit depends not on past output, but on whether a player's action tendencies align with the receiving team's tactical system.

When combined with the team-adapted LEM, player embeddings drive the simulation of how behaviour shifts within the new structure. The fine-tuned team model imposes tactical constraints—tempo, spacing, progression patterns, and positional roles, that reshape how the player's embedded tendencies manifest in event sequences. For instance, a player with a high-risk passing profile may reduce long-ball attempts when inserted into a possession-dominant team that prioritizes ball circulation, or conversely, a progressive dribbler may increase vertical carries in a counter-attacking system that rewards rapid transitions. This interaction between player embeddings and team-specific constraints enables the LEM to generate context-aware predictions of expected performance, forming the basis for estimating player-team fit.

Simulating Fit Signals

The team-adapted LEM and the player embedding are combined to simulate event sequences conditioned on the player and target-team representations. This simulation process represents the core mechanism through which player-team fit is quantified. Unlike traditional approaches that match historical statistics or compute similarity scores between players, this framework estimates how a player's behavioural tendencies interact with the receiving team's tactical constraints to produce new sequences of actions.

These projected sequences are then converted into contextual performance indicators, not only expected points variation but also encompassing metrics that better reflect how a team plays, such

as duel rate, passing volume and accuracy, and shot creation. Expected points are computed as a weighted expectation over match outcomes, as defined in Equation 3.1.

$$\text{xPts} = 3 \cdot p(\text{win}) + 1 \cdot p(\text{draw}) + 0 \cdot p(\text{loss}) \tag{3.1}$$

Expected points frameworks have been broadly adopted in sports analytics, particularly in American football, where they quantify how game-state variables translate into scoring probabilities. In football, the analogous approach evaluates how a player's actions shape the team's overall performance profile by estimating the change in key outcomes attributable to their presence. By aggregating simulated event sequences, the LEM computes the expected league position or points total, as well as projected on-ball contributions, for the team with the player integrated into the squad, comparing these against a baseline scenario without the player. This metric captures not only direct contributions such as goals and assists but also indirect effects such as improved build-up play or reduced defensive vulnerability [13].

The model, therefore, quantifies fit not by matching features but by forecasting behaviour in the new tactical environment, which fills a gap in the state of the art. Traditional regression-based methods rely on aggregated statistics that cannot represent sequential dependencies or contextual adaptation [6]. Action-value models like VAEP capture individual action quality but do not simulate how those actions change when embedded in a different team structure [2]. Similarity-matching approaches identify comparable players but do not include a mechanism to forecast performance under new tactical rules. By contrast, this thesis provides a framework that models the interaction between player tendencies and team identity, producing predictions grounded in simulated match data. This shift from descriptive comparison to predictive simulation enables clubs to evaluate transfer targets with greater precision and mitigate the financial and sporting risks associated with player recruitment.

3.2.4 Parameter-Efficient Fine-Tuning

Parameter-efficient fine-tuning (PEFT) plays an important role in this thesis because team-level datasets are small, and fully fine-tuning a Large Events Model would risk overfitting and high computational cost. The objective is to adapt a general LEM, trained on a diversity of matches, into a team-specialised model capable of simulating how a player behaves inside that specific tactical context. This requires a method that can integrate new patterns (e.g., the team's preferred passing structure) without distorting the general football knowledge learned during pre-training.

The approach taken in this thesis is inspired by the framework introduced in Mendes-Neves et al., 2024 [13], which demonstrates that small updates to a pre-trained LEM are sufficient to encode tactical signatures of individual teams. Building on this, the thesis evaluates Parameter-Efficient Fine-Tuning (PEFT) methods that allow adapting only a small subset of model parameters while keeping the original model intact.

Motivation for Parameter-Efficient Adaptation

Each team provides limited event data per season (typically a few thousand events across 30-40 matches) which proves insufficient for full-model fine-tuning without overfitting. As noted in the Literature Review, teams play fewer than 40 matches annually, constraining the available data for adaptation and exacerbating the risk of models memorizing specific sequences rather than learning generalizable tactical patterns. Full fine-tuning of large LEMs on such sparse datasets would distort the pre-trained model's broad football knowledge, leading to negative transfer where performance degrades compared to the base model.

Tactical identity manifests in repeating event patterns that can be captured through lightweight parameter updates. PEFT methods like LoRA address this by injecting low-rank adapters that modify just a small portion of the parameters [7], enabling the model to encode team-specific tendencies while preserving generalization from pre-training. This approach aligns with the LEM framework established earlier, where team-specific fine-tuning embeds behavioral signatures like build-up tempo and risk tolerance without requiring extensive retraining.

PEFT makes it possible to create multiple team-specific adapters without retraining the full LEM each time. This is useful in a player-team fit setting because each adapter can represent a different team context while preserving the general football knowledge learned by the base model. In this work, LoRA provides a scalable way to adapt the model to team-specific event patterns and compare how these adapted models differ in their simulated outcomes and style metrics.

LoRA

The primary method used in this thesis is LoRA, introduced in the work of Hu et al. [7]. Building on the motivation for parameter-efficient adaptation outlined above, LoRA adds learnable low-rank matrices to selected linear layers of the pre-trained LEM, allowing the model to learn team-specific behaviour using only a small number of added parameters (see Figure 3.1). This approach addresses the data scarcity challenge identified earlier, where full fine-tuning on limited team event data risks overfitting while preserving the base model's multi-league generalization.

In the context of this thesis, LoRA enables the model to internalise tactical characteristics manifest in event sequences, extending the team behavioral signatures from the Capturing Team Tactical Identity subsection. Specifically, the low-rank adapters capture:

- **Passing behaviour:** Differences in passing volume, pass accuracy, long-pass use, and circulation patterns.
- **Event distribution:** High-tempo teams produce shorter and more frequent event chains whereas low-tempo teams generate longer preparatory sequences.
- **Spatial tendencies:** Differences in where teams act on the pitch, such as central areas, wide areas, the attacking half, or the final third.

- **Duel and defensive involvement:** Identifiable through duel rates, defensive actions and recoveries.

By training only these small LoRA matrices on team-specific sequences, the model adapts its event predictions to the target team's style of play. The forward pass becomes

$$h = W_0 x + \Delta W x = W_0 x + BAx,$$

where W_0 are frozen pre-trained weights and B, A are low-rank matrices ($r \ll \min(d_{in}, d_{out})$, where d_{in} and d_{out} denote the dimensionalities of the input and output vector of the linear layer, respectively, and r denotes the rank of the low-rank adaptation, chosen to be much smaller than $\min(d_{in}, d_{out})$ in order to reduce the number of trainable parameters compared to the full weight matrix) trained via backpropagation [7]. This formulation can be efficient at inference, as adapters merge with the base model, enabling rapid switching between team variants for the multi-scenario simulations described earlier.

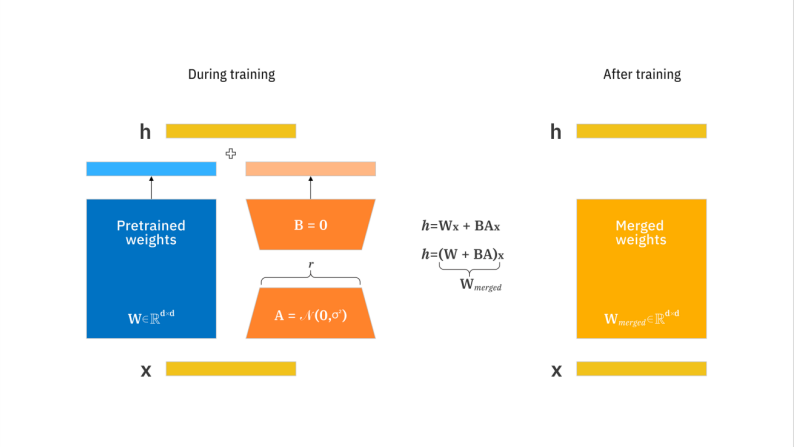


Figure 3.1: The diagram shows how LoRA adjusts the pretrained weight matrices by introducing two small matrices, A and B, of rank r that capture the update while keeping the original parameters fixed. Source: [14].

Other PEFT Options

Although LoRA serves as the primary technique, other PEFT methods provide useful points of comparison at a conceptual level. These alternatives follow the same general idea: adapting a pre-trained model to a specific context while updating only a small number of parameters.

Infused Adapter by Inhibiting and Amplifying Inner Activations (IA^3)-style multiplicative gates adapts a pre-trained model by inserting small multiplicative scaling vectors that amplify or inhibit internal activations of selected layers rather than adding new trainable layers. By training only a tiny set of gating parameters, the model can adapt to a new team while preserving broad event-level knowledge, reducing the risk of overfitting to limited match sequences [5].

Bias-only tuning (BitFit) adapts a model by updating only the bias terms of each layer, keeping all weight matrices frozen. This makes it one of the most lightweight PEFT approaches to personalize the LEM to a target team, allowing the incorporation of minimal tactical adjustments while maintaining low parameter cost. Although less expressive than low-rank updates, bias-only tuning is more suitable for situations where the available team-specific data cover only a small number of matches per season [3].

Even though these approaches provide useful alternatives, the thesis gives priority to LoRA. LoRA offers a stronger balance between adaptation capacity and parameter efficiency. In the implemented model, it adapts selected linear layers of the MLP-based LEM while keeping the base model fixed. As a result, the thesis focuses on low-rank adaptation rather than broad experimentation with multiple PEFT families.

Concluding Remarks on PEFT

LoRA is useful because it adapts selected linear layers through low-rank updates, without updating the full model. This allows the adapted LEM to learn team-specific event patterns from limited data while keeping the broader football knowledge of the base model. These patterns may reflect differences in how teams circulate the ball, progress play, create shots, or engage in duels. Rather than encoding fixed tactical rules, the adapters learn from the team-specific event sequences used during training.

LoRA's modularity supports the transfer fit evaluation pipeline in Player-Team Fit Estimation (3.2.3). Team-specific adapters can be swapped instantaneously, enabling simulations of Player X entering Team A, Team B, or Team C using the backbone LEM, without retraining overhead or catastrophic forgetting. This scalability supports the multi-scenario analyses motivated earlier, such as comparing adaptation across possession-oriented, counter-attacking, or wide-build-up systems, while addressing the overfitting risks from limited per-team data.

3.3 Positioning of this Dissertation

The research gap identified in Chapter 2 shows the need to study player-team fit as a context-dependent problem. A player's value cannot be understood only through past output, as it also depends on how the player behaves inside a team structure, how actions connect over time, and how the team's tactical identity shapes those actions.

The state-of-the-art review in this chapter clarifies where this dissertation is positioned. Aggregate statistics, Elo ratings, and player-aware expected points models provide useful baselines. They help estimate team strength and expected outcomes in a simple way. However, they do not model event sequences, tactical behaviour, or how a player may fit into a new team context. Action-value models such as xT and VAEP move closer to contextual player evaluation given that they value actions based on where and when they occur. In this work, they are used as conceptual background, not as implemented baselines.

Large Events Models provide a stronger foundation for this problem because they model football as a sequence of connected actions. This makes them better suited to study behaviour in context. However, a general LEM does not automatically represent the tactical identity of a specific team and therefore needs adaptation before it can be used to analyse team-specific behaviour.

This dissertation is positioned at this point. It develops and evaluates an embedding-conditioned LEM adapted through LoRA for player-team fit estimation. The aim is not to produce a final prediction of transfer success. Instead, the goal is to assess whether the adapted model can provide valuable signals about team style, tactical identity, and behavioural compatibility between a player and a target team.

Chapter 4

Methodological Approach

This chapter describes the methodology used to build and evaluate the proposed player-team fit framework. It presents the data representation, embedding construction, model adaptation, simulation process, and baseline comparisons that support the experimental work developed in the following chapters.

4.1 Overview of the Proposed Framework

The proposed framework estimates player-team fit through a sequence-based and behaviour-oriented pipeline. The process starts with raw event-level football data and converts it into a format that can be used by a Large Events Model. Match events are first cleaned and normalized so that actions such as passes, carries, duels, shots, and cards follow a consistent representation. These events are then transformed into LEM inputs using event variables, match context variables, and previous-event context. This allows each action to be modeled as part of a sequence, rather than as an isolated observation.

In parallel, the framework builds behavioural profiles for players and teams. Player profiles are created from event-level indicators such as passing behaviour, carry and dribble involvement, shot share, duel activity, spatial tendencies, and progression metrics. These profiles are then transformed into player embeddings. Team profiles are built from the same behavioural logic, through weighted aggregation of player embeddings. This creates team embeddings that represent the behavioural identity of each team.

These embeddings are then included in the LEM training data. For each event, the model receives information about the acting team, the opposing team, and the acting player through their respective embeddings. This creates an embedding-conditioned base LEM that can learn general football event dynamics while also using behavioural context. In practice, the model is trained to predict the next event variable from the current match state, previous events, and the relevant player and team representations.

The next stage adapts the base model to specific teams using LoRA. The base LEM remains fixed, while small team-specific LoRA adapters are trained on the event sequences involving each

target team. This allows the model to learn team-specific event patterns without retraining all model parameters. The adapted models are then used in Monte Carlo match simulations to generate possible match sequences and estimate style-based indicators, such as expected points, goals, event distributions and tactical style metrics.

The framework also includes an embedding-space transfer simulation. This step estimates how a team's behavioural identity may change when one player is replaced by another. The incoming player embedding is added to the target team profile, while the outgoing player embedding is removed. The resulting simulated team embedding is then compared with the original team embedding to measure the size and direction of the behavioural shift.

Finally, the proposed framework is evaluated against simpler baselines, including Elo and player-aware expected points models. These baselines provide interpretable reference points for team strength and outcome estimation, but they do not model event sequences or tactical behaviour.

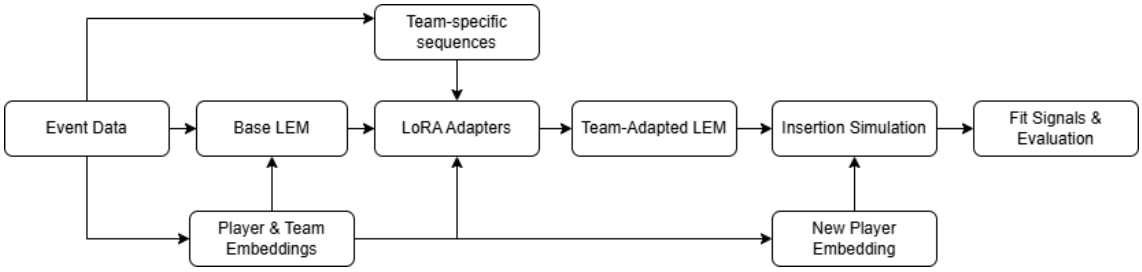


Figure 4.1: Overview of the proposed pipeline for player–team fit estimation.

4.2 Main Study Assumptions

This study relies on a set of methodological assumptions that define how the results should be interpreted. These assumptions do not invalidate the analysis, but they set the boundaries of what the proposed approach can and cannot claim.

First, on-ball event data are used as an approximation of behavioural patterns. These data describe recorded actions on the ball, such as passes, carries, duels, shots, and interruptions. However, they do not capture off-ball movement, team formations, pressing structures, or direct tactical instructions from the coach. Consequently, the model observes behavioural traces rather than the tactical organisation of a team.

Second, player and team embeddings are interpreted as compact representations of behavioural tendencies. They summarize recurring patterns in the event data, but they should not be read as complete descriptions of tactical identity. In the same way, team identity is approximated through event-weighted averages of player embeddings. This assumes that players with more recorded actions contribute more strongly to the observed team profile.

Third, the LoRA adapters are assumed to capture team-specific deviations from the base LEM without retraining the full model. This makes adaptation more efficient, but it also means that the adapted model remains constrained by the base model's learned representation of football events.

Finally, the simulation outputs are interpreted as estimates of tendencies and not predictions of real match results. The same-season evaluation is used as a diagnostic test of whether the model can reproduce observed team patterns under controlled conditions, not as proof of future generalization. Similarly, transfer simulations estimate behavioural compatibility and identity shift in embedding space. They do not establish real-world causal effects, since actual transfer outcomes also depend on coaching decisions, role usage, injuries, and broader squad context.

4.3 Data Preparation and Event Normalization

The first stage of the methodology converts raw event-level data into a consistent structure for sequence modeling. Each event file is loaded, cleaned, and enriched with season and competition metadata. The event table is then sorted within each match using the match period, minute, second, timestamp, and event identifier. This chronological order is important because the LEM learns from the relation between previous and future events. If the event order is wrong, the sequence no longer represents the real flow of the match.

Event types are normalized from the original provider fields into the fixed LEM vocabulary. This step reduces ambiguity and guarantees that the model predicts consistent event categories. For example, generic shot events are converted into right-foot shots, left-foot shots, or headers when body-part information is available. Passes are separated into regular passes, long passes, and crosses, while duel events are split into defensive duels, offensive duels, aerial duels, and loose-ball duels. Other relevant actions, such as carries, dribbles, saves, and cards, are also mapped into the event vocabulary used by the model.

The pipeline also computes a success variable for each event. For passes, success is based on pass accuracy. For shots, success indicates whether the shot resulted in a goal. For selected duel and possession events, success is based on whether possession was retained, recovered, or progressed. Carries are treated as successful actions by default because they represent completed ball movement in the event data. This success variable gives the model a simple outcome signal for each action, while keeping the representation compact.

Lastly, the variable (h) identifies the acting side in the match. When home-team information is available, this variable can reflect the home and away distinction. When this information is not available, the first team appearing in the match sequence is assigned one side and the opponent is assigned the other. For this reason, (h) should be interpreted as a stable side indicator rather than always as a true home/away flag. This distinction is important because the model needs a consistent way to separate the two teams in a match, but the available data does not always provide a reliable home-team label.

4.4 LEM Event Representation

After normalization, each match is represented as a sequence of LEM events. Each event is described through six core variables: (h), (e), (x), (y), (t), and (a). Together, these variables describe

which side acted, what action occurred, where the action happened, how much time passed since the previous action, and whether the action was successful. Table 4.1 summarizes these variables.

Variable	Meaning
h	Acting side/team flag
e	Event type
x	Event x-coordinate
y	Event y-coordinate
t	Time difference from previous event
a	Event success/outcome flag

Table 4.1: Core event variables used in the LEM representation.

The event variables are combined with match context variables that describe the current state of the game. These variables include the match period, the current time, the score, and the number of cards for each side. Table 4.2 summarizes these context variables. This context helps the model distinguish similar events that happen in different match states. For example, a pass in the first minute of a 0–0 match does not have the same context as a pass in stoppage time when one side is leading.

Variable	Meaning
p	Match period
m	Minute
s	Second
hg	Goals for side $(h = 1)$
ag	Goals for side $(h = 0)$
hr	Red cards for side $(h = 1)$
ar	Red cards for side $(h = 0)$
hy	Yellow cards for side $(h = 1)$
ay	Yellow cards for side $(h = 0)$

Table 4.2: Context variables used in the LEM representation.

The names (hg/ag), (hr/ar), and (hy/ay) are inherited from the original LEM notation. In this implementation, these variables are interpreted according to the two sides defined by (h). Therefore, (hg), (hr), and (hy) refer to the goals, red cards, and yellow cards for side (h=1), while (ag), (ar), and (ay) refer to the same quantities for side (h=0).

To allow the model to learn short-term sequential dependencies, each prediction row also includes a fixed number of previous events. This implementation uses a sequence length of three. As a result, each row contains the current match context and the previous three events, represented with prefixes such as (c1), (c2), and (c3). For example, $(c1_h)$, $(c1_e)$, $(c1_x)$, $(c1_y)$, $(c1_t)$, and $(c1_a)$ describe the most recent previous event, while (c2) and (c3) describe older events.

This representation gives the model a compact view of the recent match history. The choice of three previous events reflects a trade-off between contextual information and model complexity. A short sequence is sufficient to capture local event dependencies, such as a pass followed by a carry, a duel after a loose ball, or a shot after a final-third action. However, it does not represent long possessions, full attacking phases, or broader tactical structures that unfold over many actions. Therefore, the sequence length should be understood as a local-context design choice rather than a complete representation of the match phase.

4.5 Behavioural Profiles and Embeddings

The framework uses behavioural profiles and embeddings to represent players and teams in a numerical form. Player profiles describe how each player takes part in match events, while team profiles describe patterns in how a team plays. These profiles are then converted into embeddings, which the model can use as input features and which also support the analysis of player-team fit.

4.5.1 Player Behavioural Profiles

Player behavioural profiles are built by aggregating each player's event-level actions across matches. The goal is to describe how a player tends to take part in play, not only what final outputs the player produces. Goals and assists are useful, but they do not show whether a player progresses the ball, plays wide, attacks the final third, takes shots, or contributes through duels and recoveries.

Before the embedding step, the framework excludes player-team-season profiles with fewer than 500 recorded events. This threshold is used as a practical reliability filter, not as a theoretically optimal cut-off. Profiles below this level may be strongly affected by limited minutes, or substitute appearances, making behavioural metrics less stable. The threshold therefore reduces noise in the embedding space while still retaining enough players for team-level aggregation.

Each profile combines metrics from several parts of the game, including passing, progression, attacking involvement, defensive and duel activity, spatial tendencies, and possession behaviour. Table 4.3 summarizes these metric groups and provides examples of the features included in each one.

The selected metrics were chosen because they describe observable aspects of player behaviour that are relevant for player-team fit. Passing and possession metrics capture how the player participates in circulation and ball retention. Progression metrics describe how the player helps move play forward. Attacking metrics capture direct involvement in chance creation and shooting situations. Defensive and duel metrics represent contribution without stable possession. Spatial metrics describe where the player tends to act on the pitch. Together, these groups provide a balanced description of how a player contributes to different phases of play, while remaining interpretable and directly derived from event data.

Metric group	Examples
Passing	Pass accuracy, long-pass percentage, cross percentage
Progression	Carry share, progressive action rate, forward progress per action
Attacking	Shot share, final-third action share
Defensive/duels	Duel share, defensive duel share, recovery share
Spatial	Wide action share, central action share, attacking-half action share
Possession	Passes per possession, shots per possession, directness

Table 4.3: Examples of metric groups used to describe playing style.

4.5.2 Player Embeddings

Player embeddings are constructed from the behavioural profiles described above. In this work, the term embedding refers to a standardized behavioural vector derived from profile metrics, not to a latent representation learned directly by a neural network. This distinction is important because the embedding remains interpretable: each dimension comes from a known behavioural metric.

The embedding process starts by selecting the behavioural metrics from each player profile. Invalid values, such as infinite values, are replaced, and missing values are imputed using the median value of each metric. The resulting feature matrix is then standardized with z-score normalization using `StandardScaler`. Each standardized metric is then stored as one embedding component.

The resulting player embedding has 20 dimensions. The labels emb_0 to emb_19 are storage names assigned after standardization. Each dimension corresponds to one standardized behavioural metric, following the fixed order used during embedding construction: pass accuracy, long-pass percentage, cross percentage of passes, carry share, dribble share, shot share, duel share, defensive duel share, offensive duel share, aerial duel share, recovery share, progressive action rate, forward progress per action, wide action share, central action share, final-third action share, attacking-half action share, passes per possession, shots per possession, and directness measured as forward progress per possession. Therefore, players with similar standardized behavioural profiles should have similar embedding vectors, while players with different tendencies should be positioned further apart in the embedding space.

To inspect whether these standardized behavioural vectors capture meaningful differences between players, the 20-dimensional player embeddings are projected into two dimensions using Principal Component Analysis (PCA). Figure 4.2 shows this projection. Each point represents one player, and players with similar behavioural profiles appear closer together in the map. The highlighted players help illustrate how different player types occupy different regions of the embedding space. As a supplementary check, Appendix A.1 also presents a Variational Autoencoder projection of the player representations.

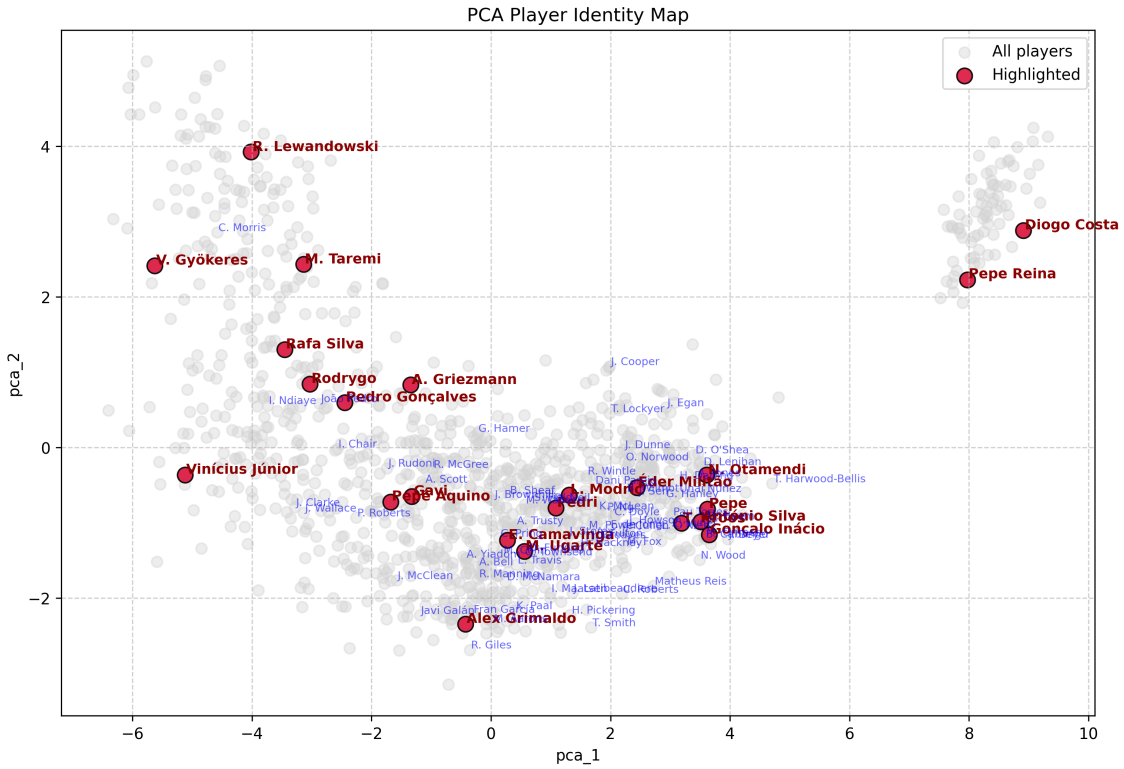


Figure 4.2: Projection of player embeddings into two dimensions using PCA.

4.5.3 Team Embeddings

Team embeddings are built by aggregating the player embeddings of all players who belong to the same team, season, and competition. The aggregation uses event-count weights, so players with more involvement in the event data contribute more to the final team representation.

For a team T with players $i = 1, \ldots, n$, the team embedding is computed as an event-weighted average of the player embeddings:

$$\mathbf{z}_T = \sum_{i=1}^{n} w_i \mathbf{z}_i, \qquad w_i = \frac{N_i}{\sum_{j=1}^{n} N_j}$$

where $\mathbf{z}_T$ is the team embedding, $\mathbf{z}_i$ is the embedding of player i, and N_i is the number of events associated with player i. The weight w_i therefore represents the share of the team's total events attributed to that player.

This design uses event counts as a proxy for each player's observed involvement in the team's behaviour. A simple unweighted average would give the same influence to a regular starter and to a low-minute substitute, even if their contribution to the event data is very different. Event-weighted aggregation reduces this problem by giving more weight to players who appear more often in the team's recorded actions. For example, a midfielder involved in many possessions will shape the aggregate team vector more than a substitute with few recorded actions. Therefore, the team embedding represents the behavioural identity most visible in the event data, instead of an

equal average of all registered players.

However, the team embedding should not be read as a complete tactical model. The aggregation does not explicitly represent player positions, tactical roles, formations, or coaching instructions. The team embedding is therefore best understood as an aggregate behavioural identity built from player-level event involvement. These embeddings are then used in the next stage of the framework, where team, opponent, and player embeddings are added to the LEM input so that event prediction can be conditioned on behavioural context.

4.6 Embedding-Conditioned Base LEM

The embedding-conditioned base LEM adds behavioural context to each training example, extending the event representation described earlier. Each row keeps the original LEM variables, including the current event variables, match-context variables, and previous-event context. In addition, the row includes three embedding blocks: the acting team embedding, the opponent team embedding, and the acting player embedding. This allows the model to predict events using not only the recent sequence of play, but also the behavioural profile of the team performing the action, the opposing team, and the player involved in the event.

The model input can therefore be divided into four parts. The first part contains the current event and match-context variables. The second part contains the previous event sequence, which gives the model short-term memory of the actions that happened before the current prediction. The third part contains team-level embeddings, with one block for the acting team and one block for the opponent. The fourth part contains the player embedding for the acting player. Together, these inputs allow the model to combine local match context with behavioural information regarding the teams and player involved in the game.

During the tabular conversion stage, the acting-team, opponent-team, and acting-player embeddings are merged into each LEM row. The implementation creates separate embedding columns for the acting team, the opponent, and the player that are later used by the base model as embedding features. Missing embeddings are rare, because most teams and players in the analysed competitions have enough event data to construct behavioural profiles. When an embedding is missing, the pipeline uses fallback values based on available team-level averages or, when these are unavailable, global averages. This avoids dropping valid event rows and keeps the input structure consistent across matches, teams, and players. However, these fallback values also reduce specificity, since replacing a missing profile with an average representation can pull the corresponding event rows closer to the mean behaviour of the dataset.

The base model is trained as a multi-layer perceptron with embedding-conditioned inputs. The architecture uses three hidden layers with 196 units each and dropout regularization with a dropout rate of 0.25. The model is trained to predict the LEM event variables sequentially. These variables include the acting-side indicator, event type, spatial coordinates, time difference, and success outcome of the event.

Cross-entropy loss is used for training because each prediction target is treated as a categorical value. This is important for the LEM setup, where each stage predicts one valid class from a fixed set of possible values. For example, event type is predicted from the event vocabulary, spatial variables are predicted from discretized pitch coordinates, and the success variable is predicted as a binary outcome. This formulation trains the model to learn categorical token distributions, which are later used during simulation to sample complete event sequences.

The resulting base model provides the general event-sequence foundation for the rest of the framework. The next stage adapts this base model to specific teams through LoRA, allowing the framework to preserve general event knowledge while learning team-specific event patterns.

4.7 LoRA Team Adaptation

After training the embedding-conditioned base LEM, the framework adapts the model to specific teams using LoRA [7]. This stage seeks to specialize the general event model to the behavioural patterns of a target team without retraining the full network. The base model keeps the broad event knowledge learned during pre-training, while the adapter learns team-specific adjustments from the target team data.

The adaptation process starts by loading the trained `MLPWithEmbeddings` model. The backbone parameters are then frozen, so the original model weights are not updated during team adaptation. LoRA is applied to selected MLP linear layers by replacing them with `LoRALinear` modules. The original linear weights remain frozen, and only the low-rank LoRA matrices introduced in these modules are trainable during adaptation.

The LoRA configuration was selected after preliminary adaptation experiments, including a Benfica stress test discussed later in Chapter 6. This test was used to assess whether a lighter adapter had enough capacity to adapt the base model to a distinctive team context. Based on these preliminary experiment, the final configuration used rank ($r = 32$), scaling parameter ($\alpha = 64$), dropout (0.05), and adapters applied to layers ([0, 4, 8]). These layers correspond to the main hidden linear blocks of the MLP, so adapting them allows the model to adjust intermediate event representations while keeping the base model and output structure fixed. This setup provides more adaptation capacity than the initial lighter adapter, but still preserves the parameter-efficient nature of the approach because only the low-rank LoRA matrices are updated.

For each target team, the training data is filtered to include all matches involving that team. This includes both actions performed by the target team and actions performed by its opponents in matches involving that team. This allows the adapter to capture the team's own behavioural patterns, as well as the patterns associated with how opponents play against it.

The filtered team data is split by match into training and validation subsets. This avoids mixing events from the same match across both sets and gives a cleaner validation signal. Validation loss then measures how well the adapter predicts unseen event rows from matches in the same team context. The training objective remains categorical cross-entropy over the predicted LEM token, matching the objective used for the embedding-conditioned base model.

This process produces one LoRA adapter per team. Since each adapter can be loaded with the same frozen base LEM, the framework can scale across multiple team contexts. The adapted model can then be used for simulation, allowing the framework to compare match outcomes and style metrics under different team-specific versions of the base model.

4.8 Simulation Procedure and Expected Points Calculation

After team-specific adaptation, the adapted LEM is used to simulate match sequences. The simulation follows the same staged structure used during training. At each step, the model generates the acting side, event type, spatial location, time difference, and event outcome. The generated sequence then provides the basis for estimating goals, win/draw/loss probabilities, expected points, and style metrics.

4.8.1 Sequential Event Generation

Each simulation starts from a generic kickoff-like context. This initial state gives the model a valid starting point, with a neutral match state and a recent previous event that resembles the beginning of play. From this point, the model generates the match one event at a time.

Each event is generated in stages, following the LEM variable order (h, e, x, y, t, a). The model first predicts the acting side, then the event type, spatial coordinates, time difference, and success outcome. Before each variable is predicted, the unknown variables of the current event are masked. This matches the training setup, where the model learns to predict one token while the remaining current-event tokens are hidden.

After all six variables are sampled, the generated event is added to the simulated sequence. The event then becomes part of the recent event context used in the next prediction step. This rolling context allows the simulation to build a match as a chain of connected actions, where each new event depends on the current match state and the events generated before it.

4.8.2 Dynamic Embedding Assignment

The simulation must preserve the same embedding convention used during training. In the training data, the team embedding represents the side performing the current event, the opponent embedding represents the other side, and the player embedding represents the acting player. During simulation, the acting side can change from one generated event to the next, so these embeddings must be assigned from the correct perspective at each step.

After the acting side (h) is sampled, the simulator updates the embedding inputs before predicting the remaining event variables. When (h=1), the team embedding is assigned to side 1, while the opponent embedding is assigned to side 0. When (h=0), the assignment is reversed. The player embedding follows the same acting-side logic.

This step is mainly an implementation detail, but it is important for consistency. Without dynamic assignment, the model could receive a team embedding that does not match the side performing the event, creating an artificial mismatch between the event variables and the behavioural context. Dynamic assignment avoids this issue and keeps the generated sequence aligned with the input structure used during training.

4.8.3 Match Termination

The simulation includes simple control rules to keep generated matches within realistic boundaries. The event vocabulary contains specific tokens for `first_half_end` and `game_end`. During simulation, these tokens are encouraged according to the simulated match time, so that the first half and the full match end at plausible moments. A maximum number of simulation steps is also used as a safeguard in case a generated sequence does not naturally produce an ending token.

4.8.4 Expected Points Calculation

Each match context is simulated several times using Monte Carlo sampling [10]. Every simulation produces a final scoreline, which is then used to estimate the probability of each match outcome. The share of simulated matches won gives (P(win)), the share of simulated draws gives (P(draw)), and the share of simulated losses gives (P(loss)).

Expected points are computed from these outcome probabilities as:

$$xPts = 3P(\text{win}) + P(\text{draw}) \tag{4.1}$$

This value represents the average number of points expected from the simulated match context. A win is worth three points, a draw is worth one point, and a loss contributes zero points. When the evaluation covers a full season, match-level expected points are summed to obtain a season-level estimate.

Although the simulations can condition on the team and opponent embeddings, they still start from a standardized kickoff-like state rather than from the full real fixture context. Therefore, the resulting xPts values should be interpreted as model-based estimates of expected outcome tendency, rather than direct predictions of the actual match result.

4.9 Transfer Simulation

The transfer simulation estimates how a player replacement may change the behavioural identity of a target team. This step uses the player and team embeddings described earlier. The goal is to measure whether the incoming player is close to the target team's behavioural profile and how much the team profile would move after the replacement.

The outgoing player embedding is then removed from the team representation using a weight based on that player's event involvement. The incoming player embedding is added using the same type of event-count weight. This produces a simulated post-transfer team embedding:

$$\mathbf{z}'_T = \mathbf{z}_T - w_{\text{out}}\mathbf{z}_{\text{out}} + w_{\text{in}}\mathbf{z}_{\text{in}} \tag{4.2}$$

where $\mathbf{z}'_T$ is the simulated post-transfer team embedding, $\mathbf{z}_T$ is the original team embedding, $\mathbf{z}_{\text{out}}$ is the outgoing player embedding, $\mathbf{z}_{\text{in}}$ is the incoming player embedding, and w_{out} and w_{in} are event-count weights. These weights reflect each player's share of the team's event volume.

The simulation then computes a variety of fit indicators. Incoming fit before the replacement is measured by comparing the incoming player embedding with the original team embedding. Incoming fit after the replacement compares the incoming player embedding with the simulated post-transfer team embedding. The outgoing player is also compared with the original team embedding, which provides a reference point for the player being replaced.

The framework also measures how much the team changes after the simulated replacement. Cosine similarity between the original and simulated team embeddings gives the team identity similarity before and after the transfer, while the embedding shift size measures the distance between both team representations. A small shift suggests that the incoming player fits the current behavioural identity of the team more closely, while a larger shift suggests a stronger change in the team profile.

When useful, the original and simulated team embeddings can also be projected into a two-dimensional Principal Component Analysis (PCA) space. This projection does not change the fit calculation, but it helps visualize the direction of the team identity shift relative to other teams. The PCA view is therefore used as an interpretation tool, not as the main measurement method.

This approach should be interpreted as an embedding-based player-team fit signal. It can indicate behavioural compatibility and the size of the expected identity shift in the team representation, but it does not account for all factors that shape a real transfer, such as tactical role, coaching decisions, injuries, adaptation time, or squad dynamics.

4.10 Baseline Models

The proposed framework is compared with simpler baselines to assess whether event-sequence modeling, behavioural embeddings, and team-specific adaptation provide useful information beyond traditional outcome-oriented methods. These baselines also give interpretable reference points: one based on team strength, one based on player-aware aggregate features, and one based on the non-adapted LEM.

4.10.1 Elo Baseline

The Elo baseline estimates team strength from historical match results. Each team receives a rating that is updated after every match according to the difference between the expected result and the observed result. Teams gain rating points when they perform better than expected and lose rating points when they perform worse than expected.

This baseline provides a simple and interpretable measure of team strength over time. However, Elo works only at team level. It does not model event sequences, tactical style, player behaviour, or player-team fit. For this reason, Elo is used as a basic outcome-oriented reference point rather than as a direct fit model.

4.10.2 Player-Aware Expected Points Baseline

The player-aware expected points baseline estimates match outcomes using player information from the previous season. For each team, historical player metrics are computed from the training season, covering shooting, passing, duels, defensive actions and ball involvement. These player metrics are then aggregated into team-level strength indicators for each match.

In the implemented cross-season setup, the model is trained on the previous season and evaluated on the full 2022/23 season. During evaluation, players with previous-season data are represented by their historical profiles. However, some players do not have available profiles from the training season. In those cases, the model assigns fallback profiles based on the previous-season team average when available, or on the global previous-season player average otherwise. This keeps the evaluation season complete while preserving temporal separation between training and testing.

Next, each team-match row is represented by the difference between the team indicators and the opponent indicators. A multinomial logistic regression model then estimates the probabilities of a win, draw, and loss. Expected points are computed using the same rule as in the simulation framework:

$$xPts = 3P(\text{win}) + P(\text{draw}) \tag{4.3}$$

This baseline is stronger than a pure team-strength model because it includes squad-level information derived from player histories, rather than relying only on past team results. However, the model remains aggregate as it does not simulate event sequences, model tactical behaviour during the match, or represent how a player's actions may change inside a different team context. In addition, teams with many players lacking previous-season profiles rely more on fallback representations, which makes their estimates less reliable.

Therefore, this baseline provides a useful comparison point, but not a direct substitute for a context-aware event simulation model.

4.10.3 Base LEM Baseline

The non-adapted base LEM is also used as a baseline. This model contains the football event dynamics learned during base training and uses the same embedding-conditioned input structure, but it does not include team-specific LoRA adaptation.

Comparing the base LEM with the LoRA-adapted version helps isolate the effect of team-specific adaptation. If the adapted model better reproduces team style, event distributions, or

simulated outcome signals, the difference can be linked to the information learned by the LoRA adapter rather than to the base model alone.

4.11 Summary

This chapter described the methodological framework used to estimate player-team fit. The framework starts with raw event-level data, which is cleaned, normalized, and converted into a LEM-compatible representation. This representation combines event variables, match context, and recent event history so that the model can learn football as a sequence of connected actions.

Next, the methodology explained how player and team behaviour are represented. Player behavioural profiles capture how each player tends to contribute on the pitch, combining indicators related to passing, progression, attacking involvement, duels, recoveries, possession behaviour, and spatial tendencies. These profiles are converted into standardized player embeddings, which are then aggregated into team embeddings using event-count weights.

The framework also includes an embedding-conditioned base LEM and a LoRA-based team adaptation process. The base model learns general football event dynamics while using player, team, and opponent embeddings as behavioural context. LoRA then adapts this model to specific teams by training small low-rank matrices while keeping the base model fixed.

Finally, the chapter presented the simulation procedure, the expected points calculation, the embedding-based transfer simulation, and the baseline models used for comparison. Together, these components define the implemented framework evaluated in the following chapters.

Chapter 5

Experimental Design

After defining the methodology, the next step is to test if the framework behaves as intended. This chapter sets out the experimental design used to evaluate the model, moving from event-token prediction to team-level simulation and embedding-space fit analysis. In doing so, it defines the structure needed to interpret the results without treating them as exact match or transfer predictions.

5.1 Overview

This chapter describes the experimental design used to evaluate the proposed framework. Chapter 4 presented how the framework is built, while this chapter explains how it is tested. The evaluation tests how well the embedding-conditioned LEM, LoRA team adapters, and transfer simulation produce signals for outcomes, playing style, and player-team fit.

The experiments are organized around four levels of analysis. First, event-level evaluation measures how well the base and adapted models predict LEM tokens. Second, match and season simulations assess whether the generated outcomes and tactical style metrics are close to real team behaviour. Third, embedding-space transfer analysis examines whether player replacements produce interpretable shifts in team identity. Fourth, computational efficiency metrics assess whether team-specific adaptation can be performed with limited trainable parameters and viable training time.

To make the link between the research questions and the experimental evidence explicit, Table 5.1 summarizes how each research question is evaluated.

5.2 Experimental Scenarios

The evaluation uses various experimental scenarios to test different parts of the framework. These scenarios are not designed to produce a single deterministic prediction of transfer success. Instead, they assess whether the models can adapt to team-specific contexts, reproduce outcome and style indicators, and generate interpretable player-team fit signals.

Table 5.1: Research questions and supporting evidence

Research question	Metrics / evidence
RQ1: How can pre-trained LEMs be efficiently fine-tuned to capture team-specific football contexts?	Validation cross-entropy and same-team validation.
RQ2: What trade-off does LoRA provide between adaptation quality and computational efficiency?	Validation loss and number of trainable parameters.
RQ3: Can player and team behavioural embeddings derived from event-level data capture tactical identity in football?	Tactical style error, PCA and VAE analysis, and identity shift.
RQ4: How can transfer simulations provide interpretable player-team fit signals for re- cruitment decision support?	Similarity scores, embedding shift, tactical deltas, and expected-points deltas.

5.2.1 Base LEM Evaluation

First, the non-adapted embedding-conditioned base LEM is evaluated. This model includes event, context, player, team, and opponent information, but does not use any team-specific LoRA adapter. The goal is to assess how well the base model generates complete match sequences and produces plausible outcome distributions before team adaptation is added.

In this scenario, Monte Carlo simulations start from a kickoff-like context using average team and player embeddings as neutral behavioural inputs. The results provide a reference point for later comparisons with the LoRA-adapted models.

Training Objective Correction

During the experimental design, the training objective of the base LEM was reviewed to ensure that it matched the structure of the prediction task. The LEM predicts one token at a time for each event variable. For example, the event type variable corresponds to one class among 32 possible event classes, while the spatial and temporal variables correspond to one class among 101 possible token values. Therefore, each prediction stage is a categorical classification problem.

For this reason, categorical cross-entropy was used as the training loss, with targets represented as integer class indices. This objective is appropriate because it trains the model to assign probability mass across mutually exclusive classes. At each prediction stage, only one token is correct, and the model must learn a probability distribution over all possible alternatives.

This differs from binary cross-entropy with one-hot encoded labels. Binary cross-entropy treats each output dimension as an independent binary prediction. This setup is suitable for multi-label problems, where several labels may be correct at the same time, but it does not match categorical token prediction. In the LEM setting, the model should not independently decide whether each possible token is present or absent. Instead, it should choose one token from a fixed vocabulary.

Aligning the loss function with the categorical structure of the task is important because the simulator samples events from probability distributions over tokens. If the model is not trained to learn these distributions directly, the generated matches may show poor calibration, especially in low-frequency but important outcomes such as goals. Hence, the final experiments use categorical cross-entropy as a methodological correction before model evaluation.

5.2.2 LoRA Team Adaptation Evaluation

This scenario evaluates two connected aspects of the framework: whether LoRA improves event-level prediction in team-specific contexts, and whether the adapted models can reproduce observed team behaviour in simulation. For each target team, the adaptation dataset includes matches involving that team, covering both its own actions and opponent actions in those matches.

The team-specific data is split at match level into training and validation subsets. In this implementation, an 80/20 split is used, with match identifiers shuffled using a random seed before splitting. This prevents events from the same match from appearing in both sets and gives a cleaner view of how the adapter performs on unseen matches from the same team context.

After adaptation, the same team-specific model is evaluated through same-season simulation. Each LoRA adapter is trained using event data from the 2022/2023 season and then used to simulate that same season. The simulated outcomes and tactical style indicators are compared with the real season values to show how the adapted model reproduces real team outcomes and tactical style when there is no cross-season generalization challenge.

Together, event-level validation and same-season simulation test whether the adapter improves token prediction and if this improvement translates into realistic team-level outcome and style signals.

5.2.3 Transfer Simulation Scenarios

Transfer simulation scenarios evaluate player-team fit in embedding space. Each scenario defines a target team, an incoming player, and an outgoing player. The original team embedding is then modified by removing the outgoing player's weighted contribution and adding the incoming player's weighted contribution, producing a simulated post-transfer team embedding.

The simulated embedding is then compared with the original team embedding and the incoming player embedding to estimate behavioural fit and team identity change. PCA and Variational Autoencoder (VAE) projections are used only to visualize the direction of the transfer movement, while the main fit indicators are computed in the full embedding space.

5.3 Evaluation Metrics

The evaluation uses metrics at five dimensions: event-level prediction, match outcomes, tactical style reproduction, embedding-space fit, and computational efficiency. This broader view is needed because player-team fit cannot be judged only through points or goals. For example, a

model may produce plausible expected points while missing how a team plays. Therefore, the metrics are designed to evaluate both outcome quality and behavioural realism.

5.3.1 Event-Level Metrics

Event-level performance is evaluated using cross-entropy loss [12]. For a target token y and predicted class probabilities $\hat{p}$, the loss is defined as:

$$\mathscr{L}_{\mathrm{CE}} = -\log \hat{p}_y \tag{5.1}$$

where $\hat{p}_y$ is the probability assigned to the correct token. Lower cross-entropy indicates better token prediction because the model assigns higher probability to the observed event variable.

In this work, cross-entropy is used during base LEM training and LoRA adaptation. Training loss monitors how well the base model learns general event dynamics, while validation cross-entropy measures how well the adapted model predicts unseen event tokens from the same team context.

5.3.2 Match Outcome Metrics

Match outcome metrics are computed from Monte Carlo scorelines. For a set of N simulated matches, the outcome probabilities are defined as:

$$P(\mathrm{win}) = \frac{N_{\mathrm{win}}}{N}, \qquad P(\mathrm{draw}) = \frac{N_{\mathrm{draw}}}{N}, \qquad P(\mathrm{loss}) = \frac{N_{\mathrm{loss}}}{N} \tag{5.2}$$

where N_{win}, N_{draw}, and N_{loss} are the numbers of simulated wins, draws, and losses. Expected points are then calculated as:

$$\mathrm{xPts} = 3P(\mathrm{win}) + P(\mathrm{draw}) \tag{5.3}$$

At season level, match-level expected points are summed to obtain a season-level estimate. The same aggregation is applied to simulated goals for and goals against. These values are compared with real points, goals scored, goals conceded, and league ranking. Lower points error and closer ranking alignment indicate better outcome-level performance.

5.3.3 Tactical Style Metrics

Tactical style metrics compare real and simulated event patterns. The metrics are grouped into event-rate indicators, spatial indicators, and possession-style indicators. Event-rate indicators include pass rate, shot rate, long-pass rate, cross rate, carry rate, dribble rate, and duel rates. Spatial indicators include final-third action rate, attacking-half action rate, wide action rate, central action rate, and mean event location. Possession-style indicators include controlled action rate, shot-to-pass ratio, long-pass share of passes, cross share of passes, and aerial share of duels.

For each style metric k, the absolute error between the real and simulated value is computed as:

$$E_k = \left| M_k^{\text{real}} - M_k^{\text{sim}} \right| \tag{5.4}$$

where M_k^{real} is the observed value and M_k^{sim} is the simulated value for metric k. These errors can then be averaged across all K style metrics to obtain a style profile error:

$$E_{\text{style}} = \frac{1}{K} \sum_{k=1}^{K} E_k \tag{5.5}$$

Lower style profile error indicates that the simulated event patterns are closer to the real tactical profile of the team.

5.3.4 Embedding-Space Fit Metrics

Embedding-space fit is evaluated using cosine similarity and embedding distance. Cosine similarity measures the alignment between two embedding vectors:

$$\text{cosine}(\mathbf{z}_a, \mathbf{z}_b) = \frac{\mathbf{z}_a \cdot \mathbf{z}_b}{\|\mathbf{z}_a\| \|\mathbf{z}_b\|} \tag{5.6}$$

This metric is used to compare the incoming player embedding with the target team embedding, and to compare the original team embedding with the simulated post-transfer team embedding. Higher cosine similarity indicates closer behavioural alignment.

Team identity shift is measured as the distance between the original and simulated team embeddings:

$$D_{\text{shift}} = \|\mathbf{z}'_T - \mathbf{z}_T\| \tag{5.7}$$

where $\mathbf{z}_T$ is the original team embedding and $\mathbf{z}'_T$ is the simulated post-transfer team embedding. A smaller shift suggests that the incoming player preserves the current team profile, while a larger shift indicates a stronger behavioural change. PCA and VAE projections are used only as visual diagnostics; the numerical fit indicators are computed in the embedding space.

5.3.5 Baseline Comparison Metrics

Baseline comparison uses the metrics supported by each model type. Elo and player-aware expected points are compared through outcome-level indicators. The non-adapted base LEM and the LoRA-adapted LEM are also evaluated through outcome-level indicators, but because they generate event sequences, they can additionally be compared using tactical style metrics.

For points-based comparison, the absolute points error is computed as:

$$E_{\text{points}} = \left| \text{Pts}^{\text{real}} - \text{Pts}^{\text{pred}} \right| \tag{5.8}$$

where Pts^{real} is the observed points total and Pts^{pred} is the predicted or simulated points estimate. Lower values indicate closer agreement with the real season outcome. Ranking consistency is also used to compare whether the models preserve the relative order of teams.

5.3.6 Computational Efficiency Metrics

Computational efficiency is evaluated to make sure the adaptation strategy is practical when applied to multiple teams. This is important because the framework uses one shared base LEM and then adapts it to team-specific contexts, instead of retraining the full model for each team.

Training time measures how long it takes to adapt the base model to one team. In this work, this includes only the LoRA fine-tuning stage. Base model training is excluded because the base LEM is trained once and reused across all team-specific scenarios.

The number of trainable parameters measures how many model parameters are updated during adaptation. In this setup, the original base model parameters remain frozen, while only the LoRA parameters are trained. As a result, this metric shows how much lighter the adaptation process is compared with full fine-tuning.

Together, training time and trainable parameter count are used to evaluate the trade-off between adaptation quality and computational cost.

5.4 Baseline Comparison Design

The baseline comparison tests whether the proposed LEM-based framework adds value beyond simpler outcome models. Each baseline represents a different level of information: team results, player-aware aggregate features, general event-sequence modeling, and team-adapted event-sequence modeling.

5.4.1 Elo Baseline

The Elo baseline is used as the simplest team-strength reference. Since it relies only on historical match outcomes, it provides a useful lower-complexity benchmark for assessing if the framework captures more than team quality.

5.4.2 Player-Aware Expected Points Baseline

The player-aware expected points baseline is used as a stronger statistical reference. This model includes squad-level information, so it helps test whether the LEM-based approach adds value beyond player-informed aggregate outcome prediction.

5.4.3 Base LEM Baseline

The base LEM baseline is used to separate the effect of event-sequence modeling from the effect of team-specific adaptation. Comparing the non-adapted base LEM with the LoRA-adapted LEM shows if the model benefits from the additional adaptation.

5.5 Validation Strategy and Experimental Limitations

The validation strategy combines event-level validation with simulation-based evaluation. During LoRA adaptation, the team-specific dataset is split at match level into training and validation subsets. This prevents events from the same match from appearing in both sets and allows validation cross-entropy to measure token prediction on unseen matches from the same team context. The adapted model is also compared with the non-adapted base LEM on the same validation data, which helps assess whether the adapter improves team-specific event prediction.

Beyond event-level validation, the framework is evaluated through simulation. Same-season simulations test whether the adapted model can reproduce team behaviour from the adaptation season and is compared with real team values using simulated outcomes and style indicators.

However, the simulation setup has some limits. Simulated matches start from a standardized kickoff context rather than from the full real fixture state. Hence, the generated results should not be interpreted as predictions of individual match outcomes. They are better understood as model-based estimates of team outcome tendency and behavioural style under the adapted model.

A second limitation concerns match specificity. Although the simulations can include team and opponent embeddings, they do not fully represent fixture-specific factors such as tactical plans, fatigue, home advantage, injuries, or match importance. As a result, season-level point estimates may become compressed toward the league average, especially when some teams depend on factors not represented by the simulation context.

Another limitation concerns player representation during simulation. Player-level sampling is not implemented, so player embeddings act as aggregate behavioural context rather than dynamically selected individual players for each event. This limits the model's ability to simulate detailed player decision-making inside match sequences.

Finally, the transfer analysis is embedding-based. This analysis estimates behavioural compatibility by comparing player and team embeddings and by measuring the simulated movement of the team representation after a replacement. However, it does not model the full causal effect of a real transfer. Tactical role, coaching decisions, adaptation time, injuries, squad hierarchy, and psychological or contractual factors remain outside the scope of the model.

Together, these limitations define how the results should be read. The experiments evaluate whether the framework captures useful event-level, team-style, and embedding-space signals for player-team fit analysis. They do not claim to predict exact match results, final league tables, or real transfer success with full accuracy.

5.6 Summary

Overall, the experimental design evaluates the framework from four connected perspectives: event-level prediction, team-level simulation, embedding-space transfer analysis, and computational efficiency. Together, these perspectives make it possible to determine if the model produces practical outcome signals while also preserving tactical patterns.

The evaluation also compares the framework with simpler baselines and states the main limitations before presenting the results. The next step is to evaluate whether the framework performs better than these references and whether its signals are reliable enough to support player-team fit analysis.

Chapter 6

Results and Discussion

6.1 Overview

After defining the methodology and evaluation design, this chapter presents the experimental results. The analysis follows the main components of the framework, moving from base LEM simulation to LoRA team adaptation, full-season simulation, and transfer simulation.

Instead of treating the outputs as transfer predictions, the discussion focuses on the signals produced by the framework. Therefore, expected points and goals are analysed together with event-level loss, tactical style metrics, and embedding-space movement. This makes it possible to evaluate not only if the model produces plausible outcomes, but also whether it captures behavioural patterns for player-team fit estimation.

Finally, the results are interpreted in light of the limitations defined in the experimental design, including the simplified simulation context, aggregate embeddings, and embedding-based nature of the transfer analysis.

6.2 Base LEM Simulation Results

Before analysing the simulation outputs, it is important to define the data scope of the base model. The embedding-conditioned base LEM was trained on event data from the Austrian league, Bundesliga, Championship, LaLiga, Ligue 1, Liga Portugal, Liga Portugal 2, and the Danish league, covering the seasons from 2019/2020 to 2021/2022.

These competitions were selected to combine stronger European leagues with leagues closer to the Portuguese context. This mix gives the base model exposure to high-level football patterns while also keeping part of the training data closer to the league used most often in the later team adaptation and transfer simulation experiments.

6.2.1 Initial Simulation Behaviour

The first stage evaluated whether the base LEM could generate complete match sequences. After converting the event data into the LEM format, the model was trained for five epochs with the

event variables (h,e,x,y,t,a), the context variables (p,m,s,hg,ag,hr,ar,hy,ay), and a sequence length of three previous events.

At this stage, the main objective was to confirm that the simulation pipeline worked end-to-end. The model generated 2000 complete matches per run, and the simulated matches terminated correctly with `game_end` events. In addition, the win, draw, and loss distributions were coherent, showing that the event generation loop and expected points calculation were functional.

The initial expected points estimate was around 1.39. To check stability, the simulation was repeated with three random seeds. The resulting expected points showed low variance across runs, which suggested that the simulation behaviour was stable enough to support the next experiments.

6.2.2 Training Corrections

The first simulations revealed a training issue. The model generated a realistic number of shots, with around 20 shots per match, but produced only about 0.10 goals per match. In contrast, the real data contained roughly 2.6 goals per match. Therefore, the model was reaching shooting situations at a plausible rate, but was not assigning enough probability to successful shots.

Further analysis showed that shot location was not the main cause. Simulated shots had similar spatial distributions to real shots, which suggested that the problem was closer to the scoring probability learned by the model. The main issue was linked to the probability assigned to event outcomes.

Following this diagnostic step, the training objective was aligned with the categorical nature of the LEM prediction task, as described in Chapter 5. The final base LEM simulations reported in this section use this corrected training setup. After this correction, the simulated outputs became more realistic. Across three seeds, expected points moved to approximately 1.62–1.66, while simulated goal averages reached around 1.5 goals for one side and 1.1 for the other, corresponding to a total close to 2.6 goals per match.

6.2.3 Interpretation of Base LEM Results

The base LEM results show that the general simulation pipeline is technically viable, while also clarifying the role of the base model in the full framework. The corrected model can generate complete match sequences, produce plausible score distributions, and support expected points calculation. However, at this stage, the model should be read as a general football event generator, not as a team-specific or player-team fit model.

This distinction matters because the base LEM was trained across several leagues and seasons. As a result, it learns broad event regularities, such as how actions tend to follow each other, how matches progress over time, and how scoring events relate to the surrounding context. However, this broad training objective also limits its ability to reproduce the tactical identity of a specific team without adaptation. Hence, the base model provides a stable foundation that can later be specialized through team-specific LoRA adapters.

In the evaluation, the base model serves two connected roles. First, it validates the LEM representation and simulation procedure. Second, it provides a reference point for measuring the effect of team-specific adaptation. If the LoRA-adapted models improve event-level validation, outcome estimates, or tactical style reproduction compared with the base model, those gains can be linked more directly to the adaptation process rather than to the simulation pipeline alone.

6.3 LoRA Team Adaptation Results

After validating the base simulation pipeline, the next stage evaluates whether LoRA adapters can specialize the embedding-conditioned base LEM to team-specific contexts. For this adaptation stage, the 2022/2023 Liga Portugal season was used as the main dataset, since it is also the main league context analysed in the full-season simulations.

For each team, the adaptation dataset included all matches involving that team. This means that the adapter was trained not only on the team's own actions, but also on opponent actions in those matches. As a result, each adapter was exposed to the full match context in which the team played, including attacking, defensive, and transitional patterns. The analysis begins with event-level validation loss and then examines whether this adaptation improves season-level simulation outputs and tactical style reproduction.

6.3.1 Event-Level Validation Loss

Validation cross-entropy was used as the main event-level metric during LoRA adaptation. This metric measures how well the adapted model predicts unseen event tokens from matches in the same team context. Lower values indicate better token prediction, meaning that the model assigns higher probability to the observed event variables.

Overall, the results show that LoRA improved validation loss for all 18 Liga Portugal teams. The average validation loss decreased from 1.645 for the non-adapted base model to 1.625 for the LoRA-adapted model. This corresponds to an absolute reduction of 0.020 points, or approximately 1.2%. Therefore, the improvement is positive and consistent across teams, but modest in magnitude. This suggests that LoRA helps the model adapt to team-specific event patterns, although the base model already captures much of the general event structure before adaptation. Figure 6.1 shows the validation cross-entropy improvement for each team.

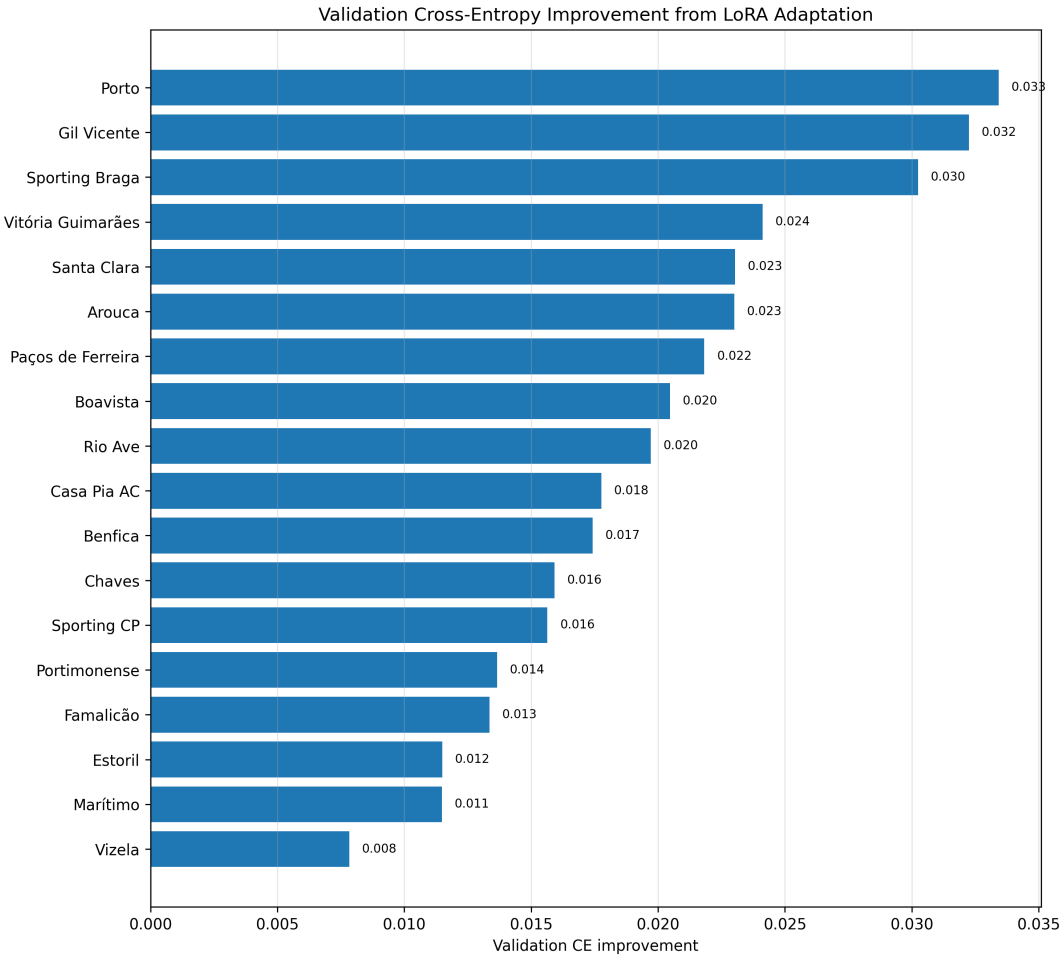


Figure 6.1: Validation cross-entropy improvement from LoRA adaptation by team. Higher values indicate a larger reduction in validation loss compared with the non-adapted base LEM.

The strongest validation improvements were observed for Porto, Gil Vicente, and Sporting Braga, with reductions above 0.030 cross-entropy points. Even the smallest improvement, observed for Vizela, remained positive. Therefore, the adapted models consistently improved event-token prediction compared with the general base LEM.

These results suggest that team-specific adaptation helps the model capture recurring event patterns that are not fully represented by the base model alone. However, the size of the improvement remains moderate. This means that LoRA refines the base model and specializes it to each team context, but still preserves most of the general football knowledge learned during base training.

6.3.2 Adaptation Capacity and Training Configuration

The initial adaptation experiments showed that the LoRA configuration influenced both validation loss and simulation quality. In early runs, some teams still appeared underfitted because training

and validation losses were decreasing when training stopped. This suggested that the adapters had not yet captured enough of the team-specific event distribution.

To investigate this, Benfica was used as a reference case. As the strongest team in the 2022/2023 Liga Portugal season, Benfica provided a useful stress test because the model needed to reproduce not only tactical style, but also a high level of outcome dominance. The supplementary diagnostics in Appendix A.2 show that both training and validation loss decreased during LoRA adaptation, with most of the improvement occurring before epoch 10. Validation loss continued to improve slightly until epoch 40, although the additional gains after the early epochs were small.

Based on these experiments, the final setup used a stronger adaptation configuration, with rank $r = 32$, $\alpha = 64$, dropout (0.05), and 10 training epochs. The adapters were applied to selected MLP linear layers, while the base model remained frozen. As a result, the framework increased team-specific adaptation capacity without losing the parameter-efficient nature of the approach, since only the LoRA matrices were updated during training.

6.3.3 Computational Efficiency

The computational efficiency results show that LoRA provided a lightweight alternative to full model fine-tuning. As shown in Table 6.1, each adapter updated only a subset of the adapted model parameters, while the remaining parameters from the base model stayed frozen. For each adapter, the adapted model contained 151,057 parameters in total, but, only 34,336 parameters were trainable during team-specific adaptation, while 116,721 parameters remained frozen. This means that the adapter updated 22.7% of the adapted model parameters, or approximately 29.4% of the parameters that would be updated under full base model fine-tuning. As a result, the setup is efficient when compared with full fine-tuning, although it should not be interpreted as extremely small in absolute terms. The adapter still updates a meaningful share of the model, which reflects the need for enough adaptation capacity to capture team-specific event patterns.

This reduction is important because the framework requires one adaptation per team. Training a separate full model for each club would increase computational cost, storage requirements, and experimental complexity. In contrast, the LoRA setup keeps one shared embedding-conditioned base LEM and stores a compact adapter for each team. This way, team-specific adaptation becomes more scalable, since each club can be represented by a small set of additional parameters instead of a fully retrained model.

Table 6.1: Computational efficiency summary for each adapter.

Metric	**Value**
Total adapted model parameters	151,057
Frozen base model parameters	116,721
Trainable LoRA parameters	34,336
Trainable share of adapted model	22.7%
Trainable parameters relative to full base model	29.4%

6.3.4 Interpretation of LoRA Adaptation Results

The event-level results show that LoRA can specialize the embedding-conditioned base LEM to team-specific contexts without retraining the full model. The consistent reduction in validation loss indicates that the adapters learn additional structure from team-specific event sequences while keeping the football dynamics learned by the base model.

Still, validation loss solely measures token prediction. A lower loss means that the adapted model assigns higher probability to observed event variables, but this does not automatically guarantee better season-level simulation. Expected points, goal totals, and tactical style metrics depend on how many generated events interact over a full simulated match.

Thus, the next stage examines whether the event-level gains from LoRA also translate into improved simulated outcomes and tactical profiles.

6.4 Full-Season Team Simulation Results

After evaluating LoRA at event level, the adapted models were tested through same-season simulation. In this experiment, each team adapter was trained with event data from the 2022/2023 Liga Portugal season and then used to simulate that same season.

This setup removes the cross-season generalization challenge, where squad changes and tactical evolution can affect the comparison. Hence, the results provide a diagnostic view of whether the adapted models can reproduce observed team outcomes and tactical behaviour under favourable evaluation conditions.

6.4.1 Outcome-Level Results

Outcome-level performance was evaluated by comparing real season points with simulated expected points. Figure 6.2 compares the real points totals with the simulated estimates from the base LEM and the LoRA-adapted LEM. Points closer to the diagonal line indicate closer agreement between the simulation and the real season outcome.

The LoRA-adapted model generally moved the simulated estimates closer to the real values than the non-adapted base LEM. This suggests that team-specific adaptation improves the outcome signal produced by the simulation. However, the simulated point totals remained compressed toward the middle of the table. Stronger teams were still underestimated, while weaker teams tended to be overestimated.

Figure 6.3 summarizes this effect by comparing aggregate outcome-level errors between the base LEM and the LoRA-adapted LEM. The results show that adaptation improves some outcome indicators, especially points and ranking error, but does not fully correct goal calibration. The complete predicted league tables for both models are provided in Appendix A.3, where the full team-level values and errors are reported.

This pattern indicates that the model captures part of the relative team-strength signal, but does not fully reproduce outcome dominance. As such, the simulation is more reliable for comparing general outcome tendencies than for predicting exact point totals and outcomes.

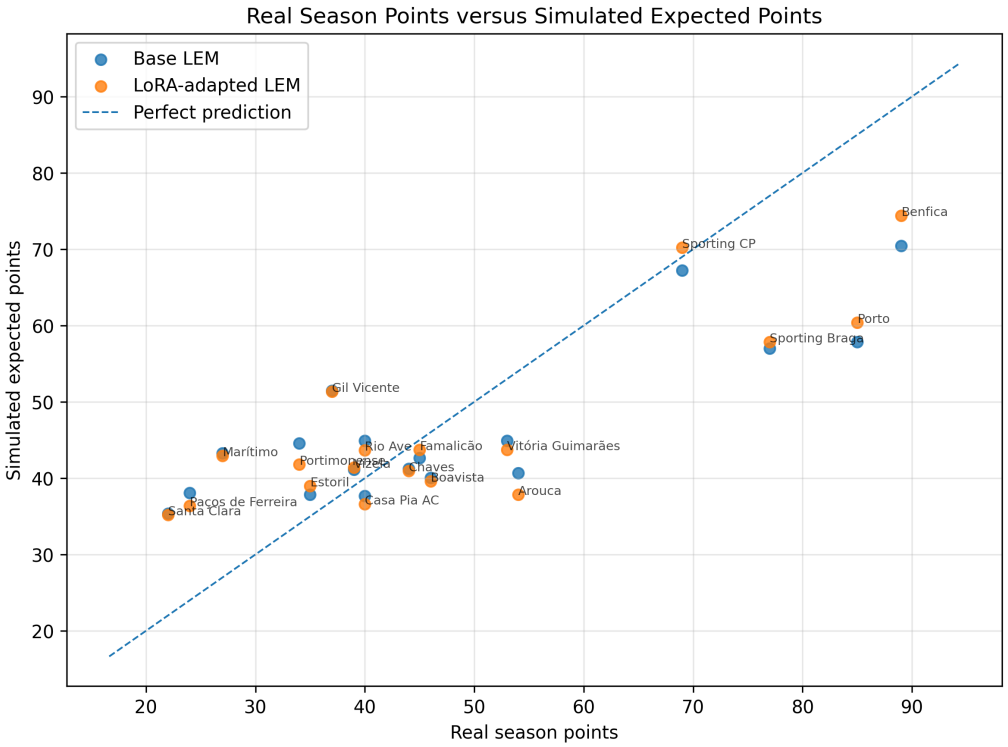


Figure 6.2: Real season points versus simulated expected points for the base LEM and the LoRA-adapted LEM. The dashed line represents perfect agreement between real and simulated points.

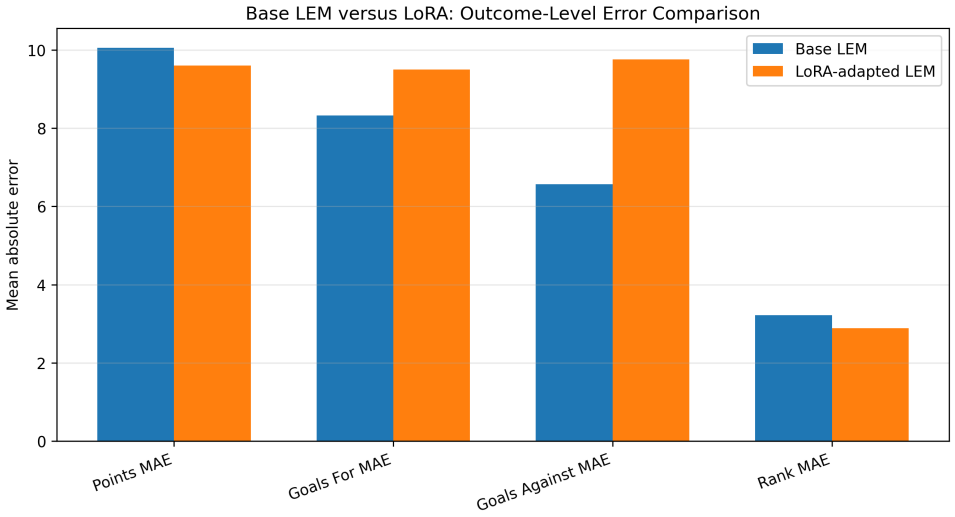


Figure 6.3: Aggregate outcome-level error comparison between the base LEM and the LoRA-adapted LEM. Lower values indicate closer agreement with real season values.

6.4.2 Comparison with Outcome Baselines

This comparison is limited to outcome-level behaviour, so the focus is on points, goals scored, goals conceded, goal difference, and ranking signals. The internal logic of the Elo and player-aware baselines was already described in the experimental design, consequently the discussion here focuses only on what their results reveal when compared with the LoRA-adapted LEM.

Figure 6.4 shows that the LoRA-adapted LEM captures the broad hierarchy of the league. Benfica, Sporting CP, Porto, and Sporting Braga are identified as the four strongest teams, which suggests that the adapted model retains a robust team-strength signal. Still, the simulated points remain compressed compared with the real table. Benfica, Porto, and Sporting Braga are underestimated, while some weaker teams, such as Marítimo, Paços de Ferreira, and Santa Clara, receive higher expected points than their real season totals.

This compression is partly explained by the simulation setup. Matches start from a neutral kickoff-like context, with limited fixture-specific information. As a result, the model does not fully condition each simulation on the specific opponent, home and away effects, player availability, line-ups, or match-specific tactical conditions. In addition, the current setup does not explicitly sample individual players during simulation, so player-level differences are represented only through aggregate embeddings. This reduces the model's ability to reproduce the repeated dominance of elite teams across a full season.

The goal columns help explain these point errors. For the strongest teams, the model often predicts a lower goal difference than observed in reality. Porto is the clearest example, with a predicted goal difference far below the real value. This happens because the model both underestimates goals scored and overestimates goals conceded. In contrast, Sporting is much closer to its real season outcome, both in points and in goal difference. These patterns show that the adaptation helps identify relative strength, but still struggles to reproduce the scale of dominance shown by the best teams.

Predicted vs Real Season Results - Primeira Liga 2022/23

Rank	Team	Pred Pts	Real Pts	Pred GF	Real GF	Pred GA	Real GA	Pred GD	Real GD
1	Benfica	74.4	89	71.4	78	26.6	19	44.8	59
2	Sporting CP	70.2	69	67.4	66	29.8	31	37.6	35
3	Porto	60.4	85	58.6	71	37.7	18	20.9	53
4	Sporting Braga	57.9	77	57.8	72	40.5	30	17.3	42
5	Gil Vicente	51.4	37	36.6	31	28.9	41	7.7	-10
6	Vitória Guimarães	43.7	53	36.2	34	40.1	38	-3.9	-4
7	Famalicão	43.7	45	34.9	37	38.3	45	-3.4	-8
8	Rio Ave	43.7	40	26.1	34	28.1	40	-2.0	-6
9	Marítimo	42.9	27	49.2	31	56.6	62	-7.3	-31
10	Portimonense	41.8	34	23.2	25	27.2	47	-4.0	-22
11	Vizela	41.4	39	50.5	30	61.2	37	-10.6	-7
12	Chaves	40.9	44	25.6	34	31.3	40	-5.7	-6
13	Boavista	39.6	46	43.8	42	56.3	52	-12.5	-10
14	Estoril	39.0	35	48.5	33	63.2	47	-14.8	-14
15	Arouca	37.9	54	28.9	36	39.9	36	-11.1	0
16	Casa Pia AC	36.6	40	25.5	29	38.4	38	-12.9	-9
17	Paços de Ferreira	36.4	24	52.4	25	73.5	58	-21.2	-33
18	Santa Clara	35.2	22	40.4	26	60.2	55	-19.8	-29

Figure 6.4: 2022/2023 season outcome reference for the LoRA-adapted LEM simulation.

Figure 6.5 provides the simplest team-strength reference. The Elo ranking places Porto, Benfica, Sporting Braga, and Sporting as the four strongest teams, which is consistent with the real competitive structure of the season. It also ranks Arouca highly, which reflects their strong real season and contrasts with the LoRA-adapted LEM, where Arouca is underestimated.

This confirms the value of Elo as a compact league-strength baseline. Since it is directly updated from match results, Elo is well suited to pure ranking comparison and can capture final competitive strength with few assumptions. However, its explanatory value is limited because it does not show if performance comes from attacking execution, defensive output, event behaviour, or tactical style. In contrast, the LoRA-adapted LEM is less direct as a ranking model, but it links outcome signals to generated event patterns and tactical behaviour. For this reason, Elo is the stronger reference for ranking alone, while the LoRA-adapted LEM is more useful for the broader player-team fit objective of this dissertation.

Final Elo Ratings — Primeira Liga 2022/23

Rank	Team	Elo Rating
1	Porto	1665.0
2	Benfica	1662.9
3	Sporting Braga	1611.1
4	Sporting CP	1596.7
5	Arouca	1536.3
6	Vitória Guimarães	1520.0
7	Famalicão	1500.5
8	Boavista	1498.5
9	Chaves	1490.5
10	Rio Ave	1469.3
11	Vizela	1461.4
12	Casa Pia AC	1454.2
13	Gil Vicente	1453.9
14	Estoril	1435.6
15	Portimonense	1423.1
16	Marítimo	1420.1
17	Paços de Ferreira	1412.3
18	Santa Clara	1388.6

Figure 6.5: Elo rating table for the 2022/2023 season.

Finally, Figure 6.6 shows that adding player-level information creates a useful but unstable outcome signal. The baseline identifies Sporting, Benfica, and Sporting Braga as strong teams, but it also produces large deviations for several clubs. Porto is strongly underestimated despite having a high known-share value, while Santa Clara is overestimated.

The known-share column indicates the proportion of player information available from the previous season for each team. Low known-share values help explain why some estimates are less reliable, especially for teams such as Rio Ave, Chaves, and Casa Pia. That said, the Porto result shows that even high player coverage does not guarantee accurate prediction. This demonstrates that aggregate player information alone is not enough to reproduce team outcomes when tactical context and event dynamics are not explicitly modeled.

Cross-Season Player-Aware Expected Points Baseline

Rank	Team	Matches	Actual Points	Predicted xPts	Error	GF	GA	GD	Known Share
1	Sporting CP	34	69	83.54	14.54	66	31	35	0.77
2	Benfica	34	89	66.885	-22.115	78	19	59	0.572
3	Sporting Braga	34	77	65.743	-11.257	72	30	42	0.721
4	Rio Ave	34	40	60.738	20.738	34	40	-6	0.109
5	Chaves	34	44	59.367	15.367	34	40	-6	0.108
6	Casa Pia AC	34	40	56.98	16.98	29	38	-9	0.14
7	Vitória Guimarães	34	53	53.072	0.072	34	38	-4	0.67
8	Santa Clara	34	22	52.706	30.706	26	55	-29	0.386
9	Famalicão	34	45	51.648	6.648	37	45	-8	0.608
10	Gil Vicente	34	37	46.651	9.651	31	41	-10	0.809
11	Porto	34	85	45.021	-39.979	71	18	53	0.957
12	Arouca	34	54	37.498	-16.502	36	36	0	0.565
13	Paços de Ferreira	34	24	32.508	8.508	25	58	-33	0.665
14	Estoril	34	35	30.833	-4.167	33	47	-14	0.476
15	Vizela	34	39	25.478	-13.522	30	37	-7	0.777
16	Boavista	34	46	24.525	-21.475	42	52	-10	0.688
17	Portimonense	34	34	24.185	-9.815	25	47	-22	0.574
18	Marítimo	34	27	20.112	-6.888	31	62	-31	0.671

Figure 6.6: Player-aware expected points table for the 2022/2023 season.

To conclude, the three approaches capture different parts of the problem. Elo provides the simplest reference for final team strength, because it is directly based on match outcomes. The player-aware baseline adds squad information, but its results show that aggregate player indicators can still produce large point errors when they are not connected to event sequences or tactical context. The LoRA-adapted LEM is not a perfect league-table predictor, since its point totals remain compressed and some teams are misestimated. However, it offers richer diagnostic information because its outcomes are generated from simulated event sequences.

This distinction is important for the interpretation of the framework. If the objective is only to rank teams after a season, Elo is the cleaner baseline. If the objective is to include player information in an aggregate outcome model, the player-aware baseline is informative but sensitive to coverage and squad representation. In contrast, the LoRA-adapted LEM adds value because it links outcome estimates with tactical style, and later transfer simulation signals. Hence, its contribution is not that it replaces simpler baselines as a pure points predictor, but that it produces outcome signals together with behavioural information that can support player-team fit analysis.

6.4.3 Tactical Style Reproduction

Tactical style was evaluated by comparing real and simulated event-distribution metrics. Figure 6.7 compares the style profile error of the base LEM and the LoRA-adapted LEM. Lower values indicate that the simulated tactical profile is closer to the observed team behaviour.

The results show that LoRA improves tactical style reproduction more distinctly than outcome prediction. This is expected because tactical style metrics are based on repeated event patterns, such as where teams act on the pitch, how often they keep controlled possession, and how they progress the ball. These behaviours appear many times across a season, giving the adapter more

examples from which to learn team-specific tendencies. In contrast, outcome metrics depend more on rarer and more volatile events, such as goals, finishing efficiency, red cards, defensive errors, and game-state effects, which change match results but are harder to learn from limited team-specific data.

Figure 6.8 shows that the improvement is not equally distributed across all tactical dimensions. The strongest gains appear in metrics linked to spatial occupation and possession control. This indicates that the adapter learns stable and frequent team behaviours more effectively, since patterns such as field position, width, centrality, and controlled action rhythm are repeated across many match sequences.

That said, some duel-related and directness-related metrics become worse after adaptation. These behaviours are more dependent on match state, opponent pressure, physical matchups, and specific game situations. Since the simulation uses a simplified match context and does not fully model player-level duels or tactical roles, these dimensions are harder to reproduce. Therefore, LoRA improves the tactical dimensions that are more stable in the data, while less stable or more context-dependent behaviours remain harder to capture.

In general, the adapted LEM provides a tangible event-level and style-level benefit, but only a partial outcome-level benefit. The adapters help the model reproduce team-specific tactical behaviour, especially spatial tendencies and controlled possession patterns. However, the remaining compression in expected points and the mixed goal errors show that team adaptation alone is not enough to produce exact result predictions. For this reason, the adapted model is more reliable as a source of tactical and behavioural signals than as a direct league-table forecasting model.

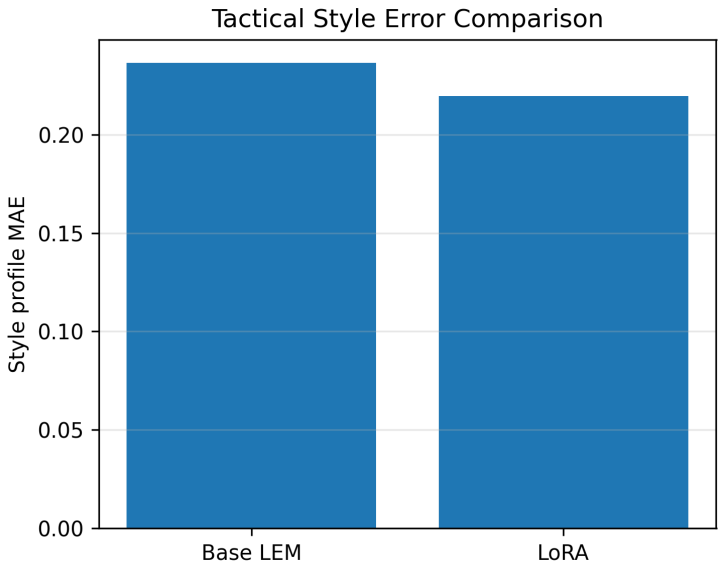


Figure 6.7: Tactical style error comparison between the base LEM and the LoRA-adapted LEM. Lower style MAE indicates closer tactical profile reproduction.

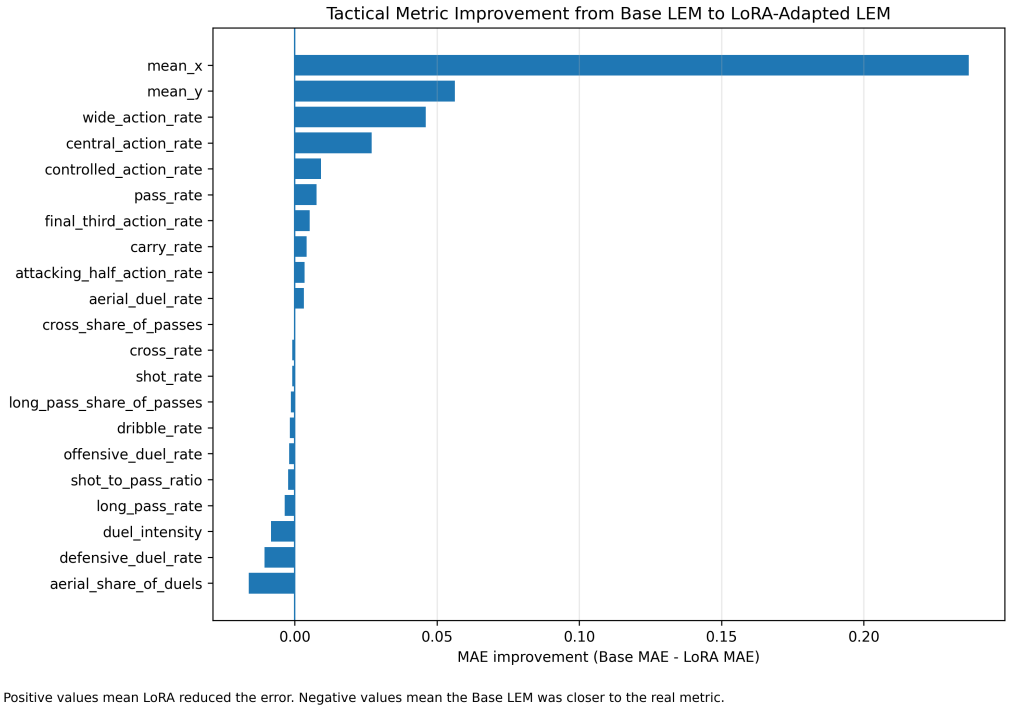


Figure 6.8: Tactical metric improvement from the base LEM to the LoRA-adapted LEM. Positive values mean that LoRA reduced the error compared with the base model.

6.4.4 Interpretation of Full-Season Simulation Results

The full-season results show that the framework reproduces behavioural and tactical patterns more effectively than exact outcome totals. This is expected because match outcomes depend on factors that are only partly represented in the simulation, such as opponent-specific plans, home advantage, injuries, substitutions, and match-state dynamics.

In addition, the persistence of point compression suggests that the simulation setup remains conservative. The standardized kickoff-like starting context and the calibrated event-generation rules help produce stable match sequences, but they also reduce the separation between dominant and weaker teams. This explains why the adapted model can improve tactical style metrics while underestimating the strongest teams and overestimating some weaker teams.

For the purpose of this thesis, this result remains useful. Player-team fit is not assessed only through predicted points. A framework that captures tactical identity and behavioural compatibility can still provide recruitment-relevant signals, even when exact outcome prediction remains limited.

6.5 Transfer Simulation Results

After evaluating the framework at event and team-simulation level, the final experimental component analyses player-team fit through embedding-space transfer simulation. This experiment

estimates how replacing one player with another changes the behavioural representation of a target team.

Three transfer scenarios are analysed: Gyökeres to Sporting, Ugarte to Benfica, and Vinícius Jr. to Porto. These cases were selected because they represent different player profiles and different types of potential tactical impact. Therefore, they provide a practical way to test if the embedding-based simulation can produce interpretable fit signals across different transfer contexts.

Each scenario is evaluated through numerical embedding-space indicators, behavioural delta metrics, and two projection-based visualisations: PCA and VAE. The objective is not to predict whether a transfer would succeed in real football. Instead, the analysis focuses on whether the framework can describe behavioural compatibility and tactical identity movement in a way that is useful for player-team fit interpretation.

6.5.1 Transfer Scenarios and Fit Indicators

The transfer analysis starts with three scenarios selected to represent different types of player-team fit. The first scenario simulates Gyökeres moving to Sporting and replacing Paulinho. This case was selected because it is an intra-league transfer, involves a player with a compatible attacking profile, and also happened in real football. Therefore, it provides a useful reference case, since the simulated signal can be interpreted alongside a known real-world outcome.

The second scenario simulates Ugarte moving to Benfica and replacing Florentino. This case represents a transfer between two dominant Liga Portugal teams, Sporting and Benfica. Since both teams operate near the top of the league but have different tactical identities, this scenario tests whether the framework can identify compatibility and identity movement in a high-level domestic context.

The third scenario simulates Vinícius Jr. moving to Porto. Unlike the previous two cases, this involves a player whose behavioural profile differs more strongly from Porto's team context in that season. For this reason, it provides a more disruptive transfer case, where the framework is expected to show worse indicators.

Each scenario is evaluated using embedding-space fit indicators. Cosine similarity measures the behavioural alignment between the incoming player and the target team, while team identity similarity compares the original team embedding with the simulated post-transfer embedding. In addition, the shift size measures how much the team representation moves after the replacement.

These indicators should be interpreted together. A high player-team similarity with a small team shift suggests a continuity signing, where the incoming player fits the existing identity. In contrast, a larger shift suggests a more transformative signing, where the player may still be valuable but would change the behavioural profile of the team more strongly.

Table 6.2 summarizes the main embedding-space indicators for the three simulated transfer scenarios. Incoming fit, outgoing fit, fit gap, and team similarity are based on cosine similarity. Incoming fit compares the incoming player embedding with the original target team embedding, while outgoing fit compares the replaced player embedding with the same team embedding. The

fit gap is calculated as incoming fit minus outgoing fit. Team similarity compares the original team embedding with the simulated post-transfer team embedding.

Additional VAE-based visualisations of the transfer simulations are provided in Appendix A.1. These figures are used as a nonlinear alternative to the PCA projections and provide a supplementary view of the identity shifts observed in the simulated transfer scenarios.

Table 6.2: Embedding-space fit indicators for the simulated transfer scenarios.

Scenario	Replacement	Incoming fit	Outgoing fit	Fit gap	Team sim.
Gyökeres to Sporting	Gyökeres for Paulinho	-0.248	-0.244	-0.003	0.907
Ugarte to Benfica	Ugarte for Florentino	0.189	0.693	-0.504	0.994
Vinícius Jr. to Porto	Vinícius Jr. for Galeno	-0.228	-0.166	-0.062	0.906

Gyökeres to Sporting

Figure 6.9 shows the projected movement of Sporting CP in the team embedding PCA space after the simulated replacement. The simulated transfer moves Sporting CP away from its original point, but it remains in the same broad region of the embedding space. This indicates that the transfer changes Sporting's behavioural identity in a visible way without fully breaking the original team profile. In practical terms, the model reads Gyökeres as a player who can alter Sporting's attacking behaviour while still fitting within the team's broader identity.

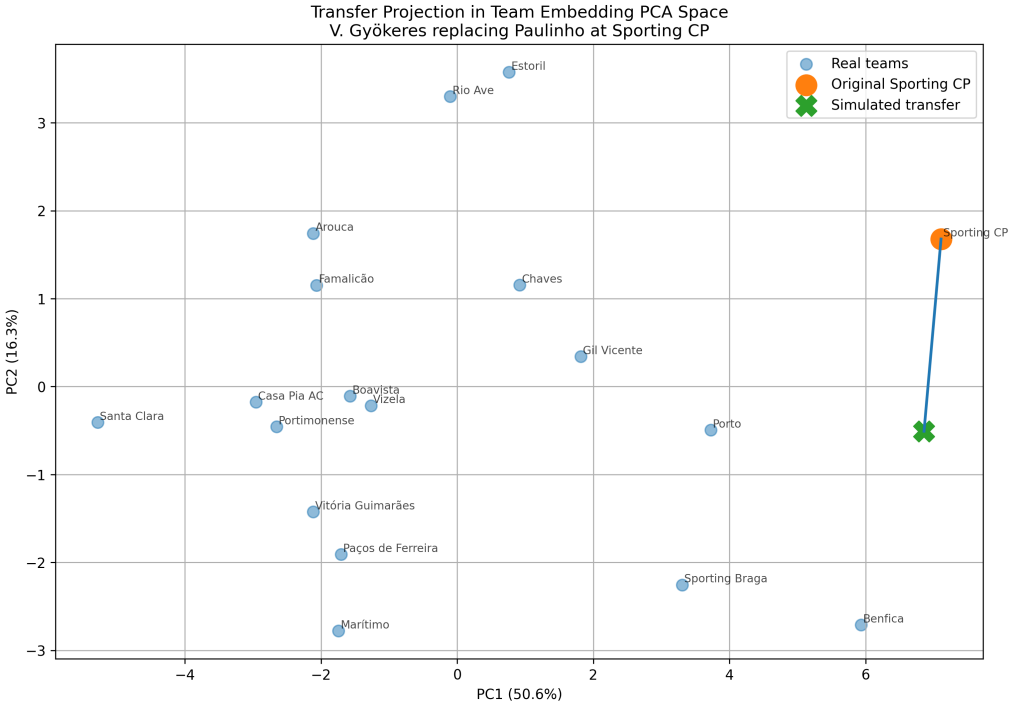


Figure 6.9: Projection of the simulated Gyökeres to Sporting CP transfer in team embedding PCA space. The arrow shows the movement from the original Sporting CP embedding to the simulated post-transfer embedding.

The outcome-level changes also support this interpretation. As shown in Figure 6.10, the simulated transfer increases Sporting's predicted expected points from 70.25 to 73.60, a gain of 3.35 points. Predicted goals scored also increase by 4.89 goals, while predicted goals conceded decrease by 1.66 goals. Hence, the model estimates a positive impact not only in attacking output, but also in the overall balance of the team.

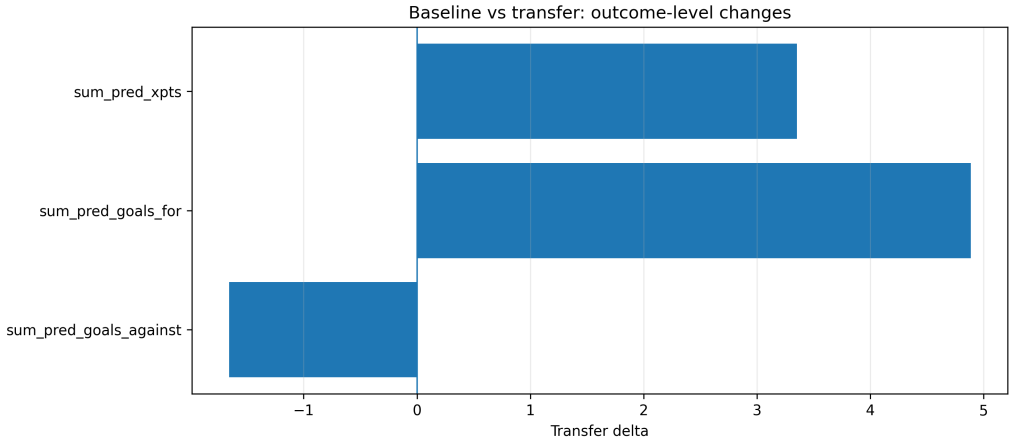


Figure 6.10: Change in simulated expected points, goals scored, and goals conceded after the Gyökeres to Sporting transfer scenario. Values represent the difference between the post-transfer simulation and the baseline Sporting simulation.

Figure 6.11 shows the tactical changes behind this outcome movement. The largest increases appear in shot-to-pass ratio, shot rate, final-third action rate, and attacking-half action rate. This means that the simulated Sporting CP becomes more direct in the final phase of attack, with more frequent shooting actions and more activity in advanced zones. At the same time, pass rate, controlled action rate, and long-pass share decrease slightly, suggesting that the transfer shifts the team away from slower circulation and toward a more goal-oriented attacking profile.

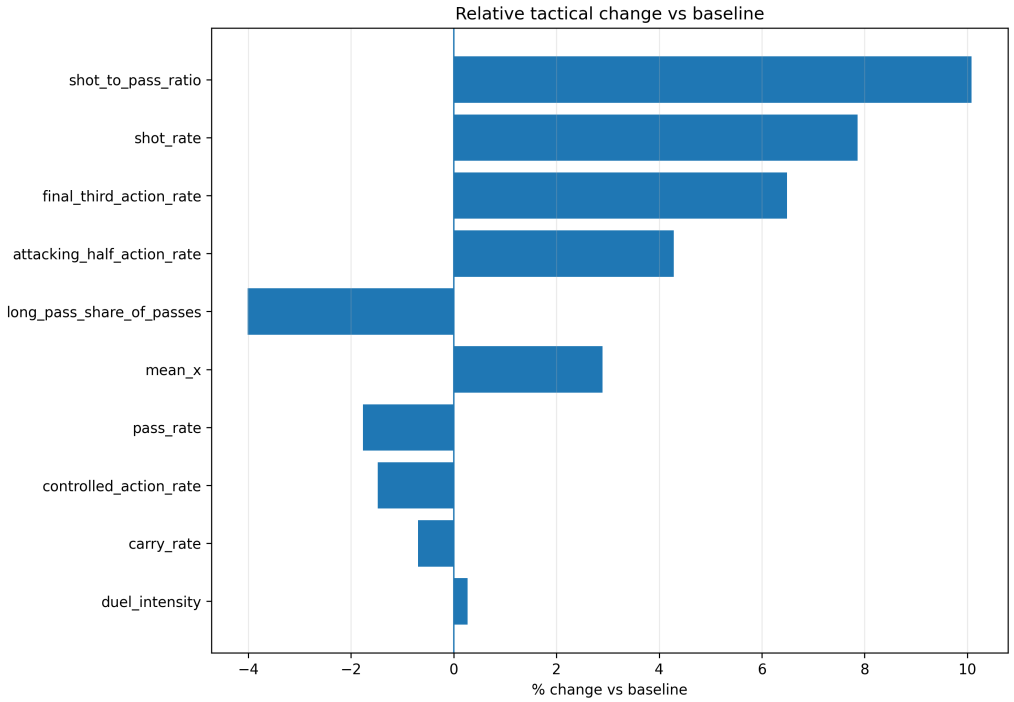


Figure 6.11: Relative tactical change after the simulated transfer of Gyökeres to Sporting. The bars show the percentage change in each tactical metric compared with the baseline Sporting simulation.

The radar plot in Figure 6.12 gives the same result in a more compact view. The post-transfer profile expands mainly around shooting and advanced-territory metrics, while remaining close to the baseline in duel intensity, carry rate, and central possession indicators. This reinforces the idea that the simulated transfer does not transform Sporting CP across every tactical dimension. Instead, it mainly increases the team's attacking presence and shot production.

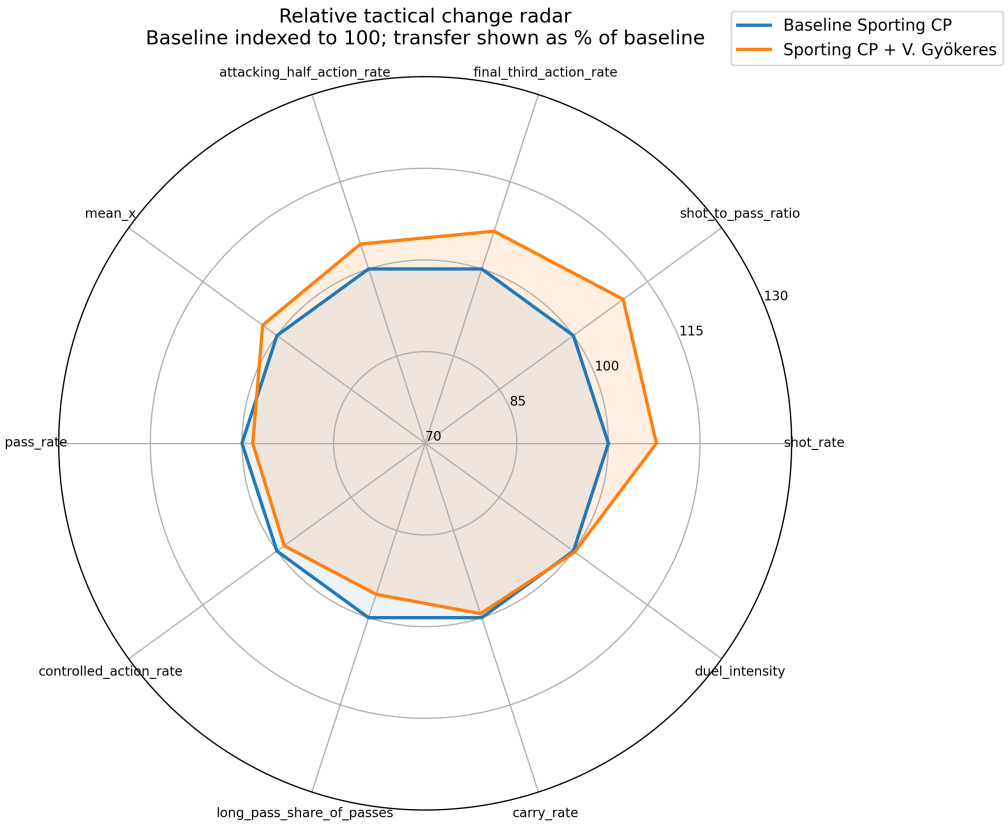


Figure 6.12: Relative tactical change radar for the simulated Gyökeres to Sporting CP transfer. The baseline Sporting CP profile is indexed to 100, and the transfer profile is shown as a percentage of that baseline.

Overall, the Gyökeres to Sporting CP scenario produces a coherent fit signal. The player increases Sporting's attacking output, raises expected points, and moves the team identity in a clear but not excessive direction. For this reason, the simulation reads the transfer as a positive and compatible change. In the transfer scenarios strategic profile, this scenario is best interpreted as a compatible upgrade: the player improves the expected outcome and attacking profile while remaining aligned with the team's tactical identity.

Ugarte to Benfica

The second transfer scenario simulates Ugarte joining Benfica and replacing Florentino. Unlike the Gyökeres scenario, the expected effect is not a large tactical transformation. Instead, this scenario tests whether the framework can identify a similar player-team fit, where the incoming player preserves most of the target team's behavioural identity.

Figure 6.13 shows the projected movement of Benfica in the team embedding PCA space after the simulated replacement. The movement is small and remains close to the original Benfica point. This indicates that Ugarte changes the team representation only slightly, which is consistent with the idea of a compatible replacement rather than a disruptive transfer.

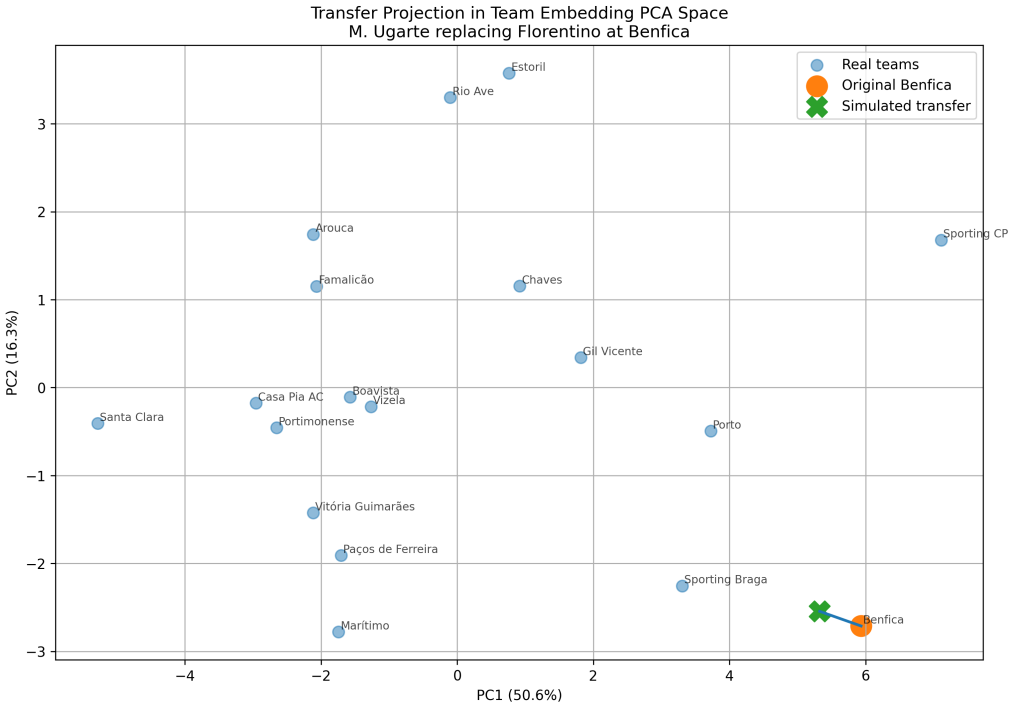


Figure 6.13: Projection of the simulated Ugarte to Benfica transfer in team embedding PCA space.

The outcome-level changes also indicate a limited effect. As shown in Figure 6.14, the simulated transfer reduces Benfica's predicted expected points by only 0.40 points. Predicted goals scored decrease by 0.46 goals, while predicted goals conceded increase by 0.12 goals. These changes are small at season level, which suggests that the replacement keeps Benfica's simulated performance almost unchanged.

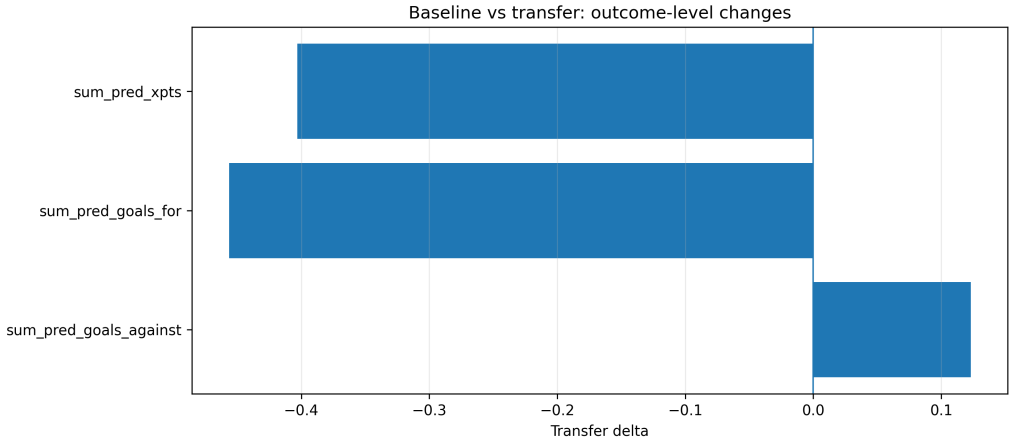


Figure 6.14: Change in simulated expected points, goals scored, and goals conceded after the Ugarte to Benfica transfer scenario.

Figure 6.15 confirms the same pattern at tactical level. Most relative changes are close to zero, with only small decreases in shot rate, final-third action rate, attacking-half action rate, and

shot-to-pass ratio. Duel intensity increases slightly, which is consistent with Ugarte's more duel-oriented profile. However, the size of the changes remains limited, so the simulated transfer does not meaningfully alter Benfica's tactical identity.

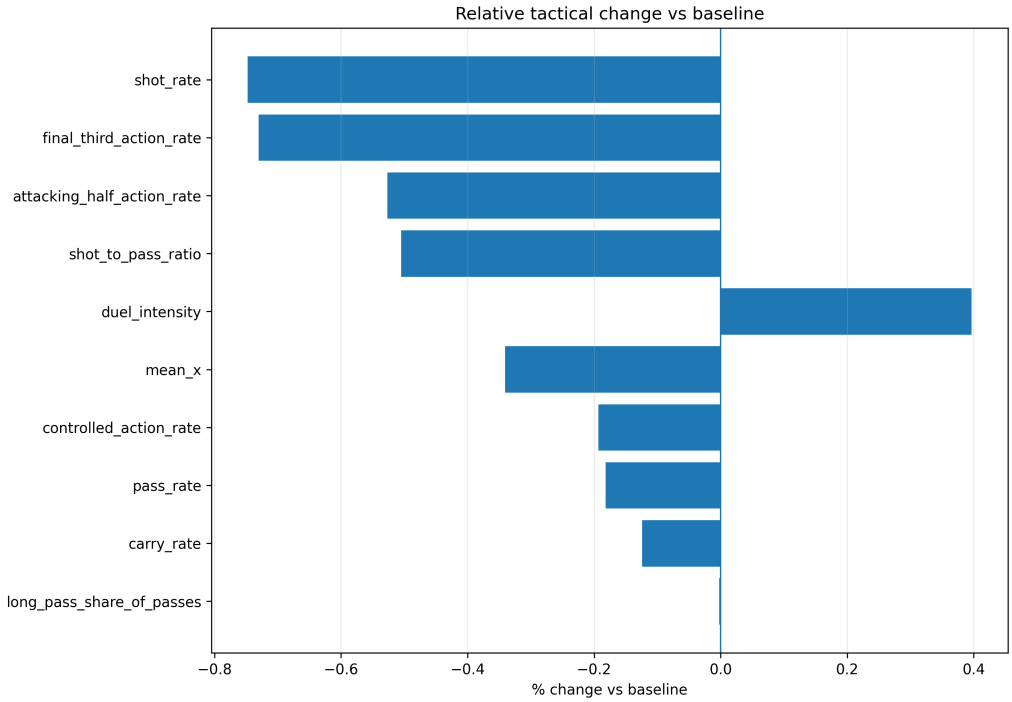


Figure 6.15: Relative tactical change after the simulated transfer of Ugarte to Benfica.

The average action-location plot in Figure 6.16 provides another view of this stability. The baseline and post-transfer average action locations are almost overlapping, with a small backward movement in mean (x). This supports the interpretation that the transfer does not move Benfica into a new spatial profile. Instead, the team's average territorial behaviour remains almost the same.

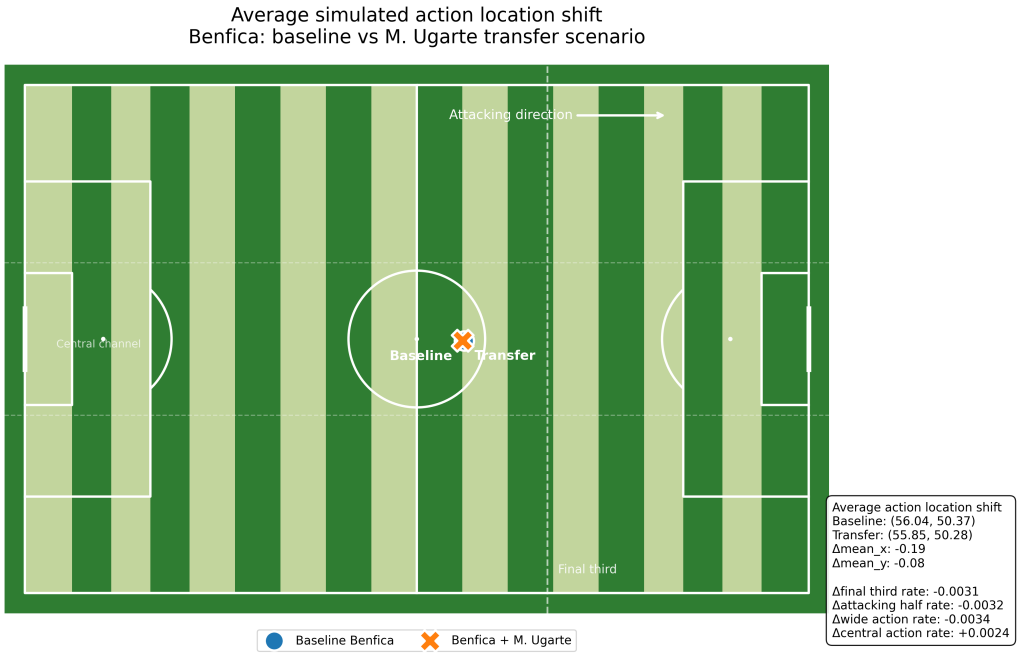


Figure 6.16: Average simulated action-location shift for the Ugarte to Benfica scenario. The markers compare Benfica's baseline average action location with the simulated post-transfer location.

In sum, the Ugarte to Benfica scenario produces a stable fit signal. The simulated team identity remains close to the original Benfica profile, the tactical deltas are small, and the outcome-level changes are almost neutral. Therefore, the framework interprets this transfer as a low-disruption replacement: a profile that preserves continuity, creates limited tactical change, and fits the existing team identity without substantially altering the expected outcome.

Vinícius Jr. to Porto

The third transfer scenario simulates Vinícius Júnior joining Porto and replacing Galeno. In contrast with the previous two cases, this scenario was selected as a more challenging fit. The objective was not to model a like-for-like replacement, but to test what happens when the incoming player does not align with the target team's existing playing style.

Figure 6.17 shows the movement of Porto in the team embedding PCA space after the simulated transfer. The post-transfer point moves away from the original Porto position, which suggests a more visible identity shift than in the Ugarte to Benfica scenario. This does not mean that the transfer is necessarily bad in outcome terms, but it does indicate that the team would no longer behave in quite the same way.

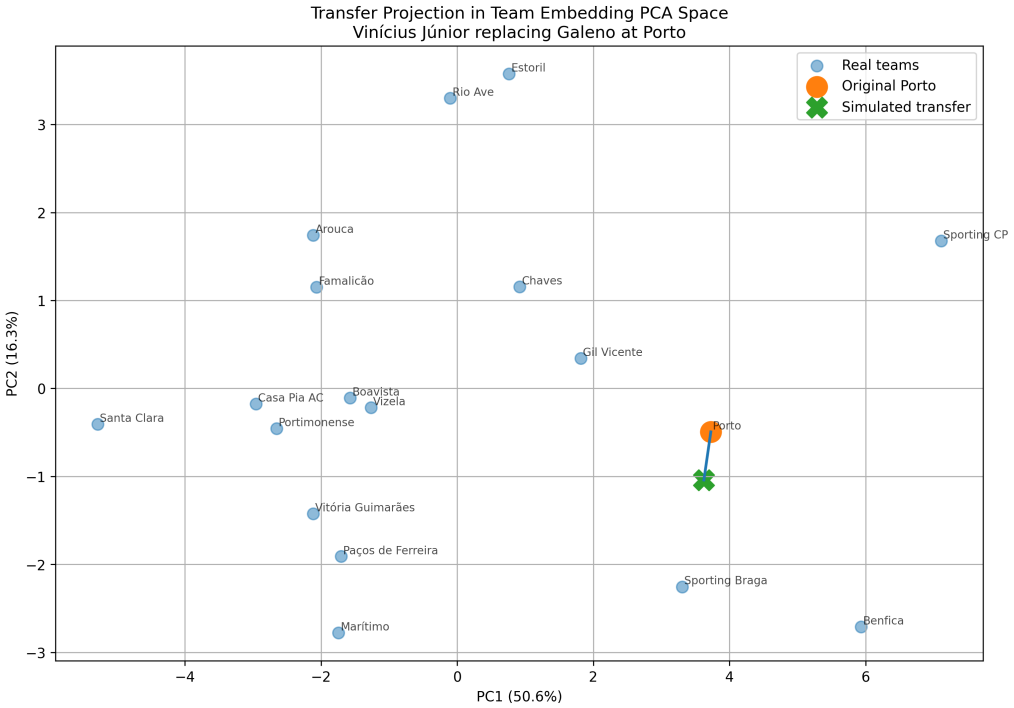


Figure 6.17: Projection of the simulated Vinícius Jr. to Porto transfer in team embedding PCA space.

The outcome-level changes in Figure 6.18 are interesting because they are directionally positive. The simulated transfer increases expected points and goals scored, while also reducing goals conceded. At first sight, this could suggest an attractive transfer. Nevertheless, this is exactly why this scenario is useful. The outcome signal alone points in a favourable direction, yet the behavioural evidence suggests a more complex picture.

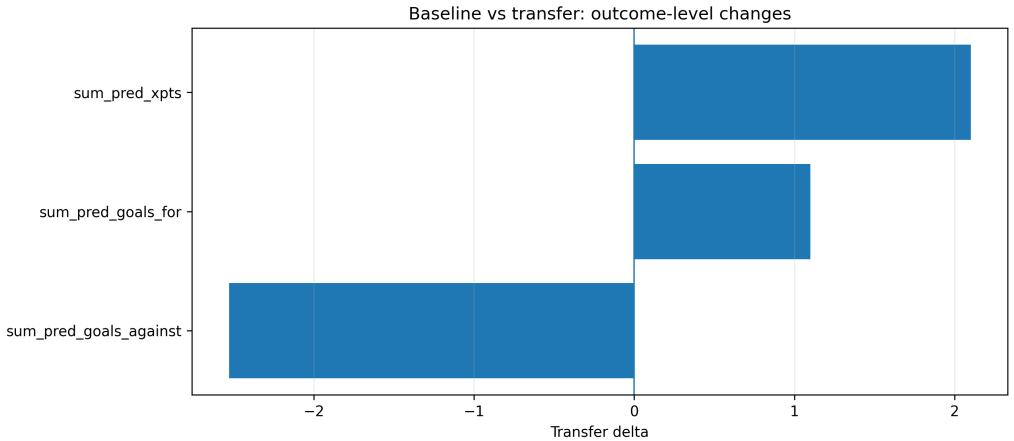


Figure 6.18: Change in simulated expected points, goals scored, and goals conceded after the Vinícius Jr. to Porto transfer scenario.

Figure 6.19 helps explain that complexity. The simulated team becomes more aggressive in

the final third, with increases in shot-related and territorial attacking metrics, while the share of long passes decreases. At the same time, pass rate and controlled action rate fall. In other words, the transfer pushes Porto toward a more direct and individually driven attacking profile, rather than preserving the team's previous balance. This makes sense for a player like Vinícius Júnior, whose profile is strongly associated with ball carrying, attacking threat, and final-third impact, but it also means that the change is not a natural continuation of Porto's original style.

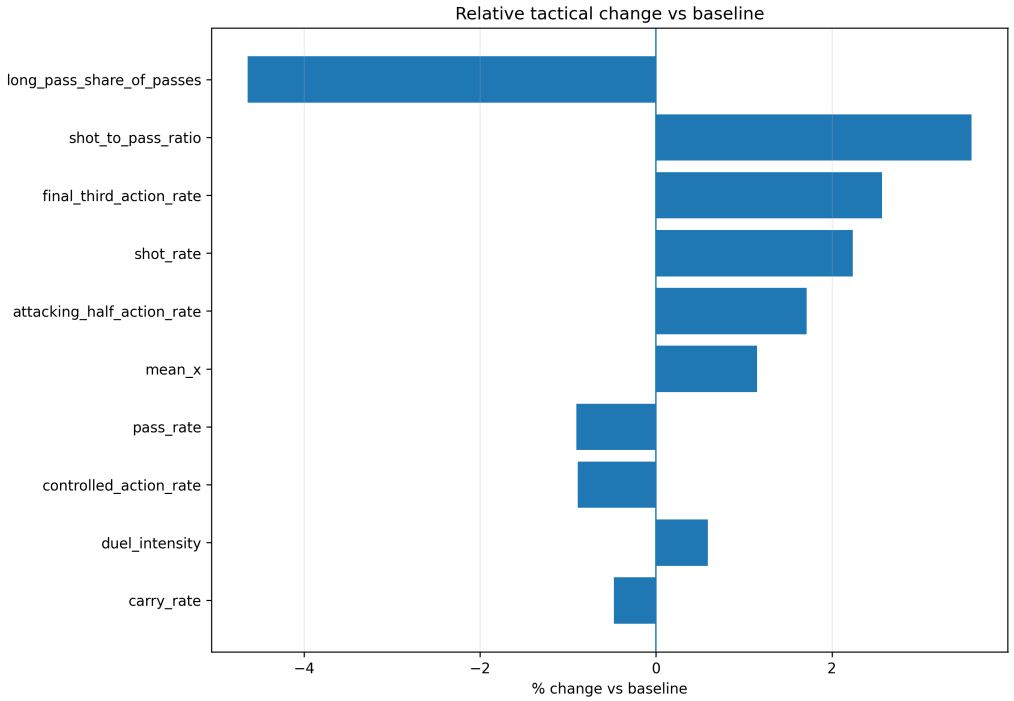


Figure 6.19: Relative tactical change after the simulated transfer of Vinícius Jr. to Porto.

For this reason, the Vinícius Júnior to Porto scenario should be read as a case of limited stylistic fit but potentially high-impact change. Unlike Gyökeres to Sporting, which reinforces an already compatible style, or Ugarte to Benfica, which works as a low-disruption replacement, this transfer creates a stronger tension between fit and upside. This scenario is best interpreted as a high-risk tactical change: the player may improve some simulated outcome signals, but also moves the team further away from its original tactical identity. In that sense, the scenario shows why player-team fit should not be judged from expected points alone.

6.5.2 Interpretation for Recruitment Decision Support

The transfer simulation results show that the framework can support recruitment analysis by separating behavioural compatibility from tactical transformation. A high similarity score suggests that the incoming player is close to the target team's current behavioural identity. However, the team identity shift shows whether the transfer would preserve that identity or move the team toward a different profile.

This distinction matters because recruitment decisions can have different objectives. In some cases, the goal is continuity: replacing an outgoing player with limited tactical disruption. In other cases, the goal is transformation: changing how the team attacks, defends, progresses the ball, or occupies space. Consequently, the framework is useful because it does not only ask whether a player fits. It also shows how the team representation changes after the replacement.

The three transfer scenarios illustrate these different recruitment signals. Gyökeres to Sporting suggests a positive attacking fit, where the incoming player increases goal threat while still remaining compatible with the general team identity. Ugarte to Benfica suggests a continuity profile, where the transfer changes very little and mostly preserves the existing midfield balance. In contrast, Vinícius Júnior to Porto indicates a more disruptive attacking transformation, where some outcome indicators improve but the tactical identity moves further from the original team profile.

Rather than acting as definitive transfer predictions, these results provide empirical evidence to inform recruitment discussions. Specifically, the framework categorizes potential transfers into three strategic profiles: low-disruption replacements, compatible upgrades, or high-risk tactical changes. However, the final decision would still need scouting, tactical role analysis, injury history, and financial considerations.

6.6 Discussion of Main Findings and Limitations

This section brings together the main findings from the experimental results and discusses what they mean for player-team fit estimation.

6.6.1 Summary of Main Findings

Overall, the results show that the proposed framework is technically viable and can generate useful signals for player-team fit analysis. The first important finding came from the base LEM simulation. The initial model produced realistic shot volume, but unrealistically low goal totals. Further analysis showed that the problem was caused by using binary cross-entropy with one-hot labels for a categorical token prediction task. After replacing this with categorical cross-entropy and integer targets, the simulated goals and expected points became much more realistic.

Then, the LoRA adaptation experiments showed that the base model can be specialized to team-specific contexts without retraining the full model. The adapters reduced validation cross-entropy across the Liga Portugal teams, which indicates that they captured additional event-level structure from match sequences. At the same time, the base model remained frozen, so the general football knowledge learned during base training was preserved.

However, the strongest improvements appeared in tactical style reproduction instead of correct outcome prediction. The LoRA-adapted models improved some behavioural metrics, especially those related to possession structure, spatial occupation, and field position. In contrast, expected points and goal totals remained more difficult to reproduce precisely. This suggests that the framework is better at capturing how teams play than at predicting exact season outcomes.

At outcome level, the model preserved part of the league hierarchy, especially among stronger teams. Nevertheless, predicted points remained compressed toward the league average, as strong teams were generally underestimated, while weaker teams were often overestimated. Thus, the simulation should be interpreted as a source of relative team-strength and behavioural signals, not as a league table forecasting model.

The transfer experiments showed that the framework can describe different types of recruitment scenarios. Gyökeres to Sporting illustrated a positive attacking fit, Ugarte to Benfica showed a low disruption continuity profile, and Vinícius Jr. to Porto illustrated a more unpredictable transformation case.

Finally, the computational setup supports the practical motivation of the thesis. The base model is trained once and reused, while team adapters provide specialization with a limited number of trainable parameters. As a result, the framework offers a scalable way to adapt a general event model to multiple team contexts, with the strongest value appearing in tactical and behavioural analysis rather than match prediction.

6.6.2 Limitations Observed in the Results

The results also reveal a number of limitations that affect how the framework should be interpreted. The most visible limitation is the compression of simulated outcomes. Even after LoRA adaptation, predicted points remained closer to the league average than the real table. As a result, the model captured part of the team hierarchy, but did not fully reproduce the separation between dominant, mid-table, and weaker teams.

This effect was especially clear for top teams. Although the model correctly categorized Benfica, Porto, and Sporting Braga as high-performing clubs, it consistently underpredicted their simulated point totals. Hence, the model seems to capture relative strength better than outcome dominance. This indicates that repeated tactical patterns are easier for the simulation to reproduce than the finishing efficiency, match control, and opponent-specific advantages that allow better teams to accumulate high point totals.

Another limitation comes from the simulation setup itself. Matches start from a standardized kickoff context, meaning that each simulated match begins from the same neutral initial state, with a generic starting event and no match history already encoded. This creates a controlled starting point and helps the simulator generate stable match sequences, but it also removes fixture specific information such as the opponent's tactical plan, home advantage, lineup choices, and early match pressure. As a consequence, strong teams may lose part of the advantage they usually create through opponent dominance, while weaker teams may benefit from the neutral starting context. Therefore, the generated results should not be read as strict match predictions, but as estimates of team tendency under a controlled simulation context.

In addition, player-level event sampling is not explicit during match simulation. Although the model uses player and team embeddings, the simulator does not dynamically select the exact acting player for every generated event. This means that player embeddings act as a broader squad-level behavioural context rather than as a fully explicit player decision process.

The transfer simulation has a related limitation, as it is based on embedding-space replacement, not on a causal model of real transfer impact. Therefore, the results estimate behavioural compatibility and team identity movement, but they do not account for factors such as tactical role assignment, adaptation time, injuries, or coaching decisions.

6.6.3 Implications for Player-Team Fit Estimation

The main implication of these results is that the framework works best as a contextual decision-support tool. Its value comes from combining several signals that describe different aspects of fit: expected outcomes, tactical style reproduction, embedding-space similarity, and the movement of team identity after a simulated replacement.

This combination is important because player-team fit is not captured by one metric alone. Expected points and goal deltas indicate whether the simulated transfer improves the model's outcome tendency. Tactical deltas show how the team's playing profile changes, including possession structure, shot behaviour, spatial occupation, and directness. Embedding-space similarity then provides a compact measure of behavioural compatibility between the player and the target team. Finally, team identity movement shows whether the transfer preserves the existing tactical profile or pushes the team toward a different one.

Accordingly, the framework can help distinguish between different recruitment cases. A player with high similarity and a small team shift may represent a low disruption replacement. A player with positive outcome deltas and moderate identity movement may represent a compatible upgrade. In contrast, a player with larger tactical movement may represent a more revolutionary and potentially riskier signing, even if some simulated outcome indicators improve.

For this reason, the model is most useful as an additional layer in recruitment analysis. It does not replace scouting, tactical judgement, or financial analysis. Instead, it provides structured evidence about how a player's behavioural profile relates to a target team's tactical identity and how that identity may change after the transfer.

Chapter 7

Conclusion

This chapter concludes the dissertation by summarizing the work developed, answering the research questions, and highlighting the main contributions of the proposed framework. It also presents the main limitations and outlines possible directions for future work. Finally, it closes with a brief reflection on the value of using event-level modeling, behavioural embeddings, and efficient adaptation methods for player-team fit estimation.

7.1 Summary of the Work

This dissertation developed and evaluated a framework for estimating player-team fit in football using event-level data. The work started from the idea that player performance should not be analysed only through isolated statistics, since a player's behaviour depends on the tactical context, the surrounding team, and the sequence of actions in which the player participates.

To address this problem, the framework combined a number of components. First, raw football events were converted into a LEM-compatible representation. Then, player and team behavioural profiles were built and transformed into embeddings. These embeddings were used to condition a base Large Events Model, which was later adapted to team-specific contexts through LoRA. As a result, the same base model could be reused while small team adapters captured more specific tactical patterns.

The framework was evaluated through event-level validation, full-season team simulation, comparison with simpler baselines, and embedding-based transfer simulations. In general, the work showed how an adapted event-sequence model can provide practical signals about team behaviour, tactical style, and player-team compatibility.

7.2 Answers to the Research Questions

RQ1: How can pre-trained LEMs be efficiently fine-tuned to capture team-specific football contexts?

Pre-trained LEMs can be efficiently fine-tuned by using a shared embedding-conditioned base model and adapting it with team-specific LoRA adapters. In this setup, the base LEM keeps the football event knowledge learned from broader training data, while each adapter learns the event distribution of a specific team context.

This approach is efficient because the full model does not need to be retrained for every team. Instead, the base model remains frozen and only the low-rank adapter parameters are updated using matches involving the target team. As a result, the model can specialize to team-specific patterns while keeping the adaptation process lightweight.

The experimental results support this approach. After LoRA adaptation, validation cross-entropy decreased compared with the non-adapted base LEM for all 18 Liga Portugal teams, showing that the adapters improved event-token prediction in team-specific contexts. However, this evidence should be interpreted with caution because the validation setup is based on same-season team data. Therefore, it shows that the adapters learn team-specific event patterns within the evaluated season, but it does not prove cross-season generalization.

Overall, LoRA provides a practical way to adapt pre-trained LEMs to individual teams without losing the broader dynamics captured by the base model. The results show consistent team-level adaptation, while also indicating that stronger temporal validation would be needed to confirm whether these learned patterns remain stable across future seasons.

RQ2: What trade-off does LoRA provide between adaptation quality and computational efficiency?

LoRA provides a useful trade-off by improving team-specific adaptation while avoiding full model retraining. In this work, the final configuration used rank $r = 32$, $\alpha = 64$ and dropout (0.05), applied to selected MLP linear layers. This increased the adapter capacity enough to capture more team-specific behaviour, while keeping the base LEM frozen.

Consequently, each team could be represented through a reduced set of trainable adapter parameters instead of a separate full model. The computational results support this efficiency claim, since each adapter trained 34,336 parameters out of 151,057 parameters in the adapted model. This corresponds to 22.7% of the adapted model parameters, or approximately 29.4% of the parameters that would be updated under full base model fine-tuning. Therefore, LoRA reduces the number of trainable parameters compared with full fine-tuning. However, this reduction should be interpreted with nuance: 22.7% is still not negligible for a relatively small MLP-based model, so the adapter is efficient but not extremely tiny in absolute terms.

This makes the approach more scalable, especially because team-level datasets contain a limited number of matches and full fine-tuning could increase overfitting risk. Nevertheless, the results also show that the benefits are not equal across all evaluation dimensions. LoRA improved event-level validation loss and helped reproduce tactical style metrics more clearly, especially those linked to possession structure and spatial behaviour. In contrast, the improvement in exact outcome prediction was more limited, since expected points and goal totals remained compressed toward the league average.

Thus, LoRA is most valuable in this framework as an efficient method for learning team-specific behavioural and tactical signals. It improves adaptation quality at a lower computational cost than full fine-tuning, but it does not fully solve the harder problem of precise season outcome prediction.

RQ3: Can player and team behavioural embeddings derived from event-level data capture tactical identity in football?

Yes, the results indicate that player and team behavioural embeddings can capture tactical identity to a certain extent. In this model, player embeddings are built from standardized behavioural metrics derived from event-level data, including indicators related to passing, shooting, duels, progression, spatial occupation, and directness. Team embeddings are then constructed as event-weighted averages of the player embeddings, so players with greater involvement contribute more strongly to the team representation.

This design allows the embeddings to summarize how players and teams tend to behave, instead of only describing their final outputs. The evidence supports this interpretation in two ways. First, the LoRA-adapted model improved tactical-style reproduction, especially for repeated event and spatial patterns. Second, the embedding-space transfer simulations produced interpretable team-identity movements, making it possible to observe whether a simulated transfer preserved the original team profile, moved it slightly, or pushed it toward a different behavioural identity.

However, the embeddings should not be interpreted as a complete representation of tactical identity. They are aggregate behavioural representations and do not fully encode formations, tactical roles, or coaching instructions. Hence, they capture relevant tactical tendencies, especially repeated event and spatial patterns, but they remain a simplified representation of team identity.

RQ4: How can transfer simulations provide interpretable player-team fit signals for recruitment decision support?

Embedding-based transfer simulations can support recruitment decisions by showing two complementary aspects of player-team fit: behavioural compatibility and tactical change. In this framework, player-team similarity indicates how close the incoming player is to the target team's current behavioural identity. At the same time, team identity movement shows how much the team representation changes after replacing one player with another.

This distinction is important because not all transfers have the same purpose. Some signings are intended to maintain continuity and replace an outgoing player with limited tactical disruption. Others are expected to change the team's profile, for example by increasing attacking output or progression. The three transfer scenarios analysed in this dissertation support this interpretation: Gyökeres to Sporting illustrates a compatible upgrade, Ugarte to Benfica represents a low-disruption replacement, and Vinícius Júnior to Porto shows a high-risk tactical change.

The framework helps interpret these different cases through embedding similarity, expected points variation, goal variation, tactical metric changes, and movement in team identity space. However, these simulations should be interpreted as illustrative decision-support signals, not as causal evidence of real transfer success. They estimate how the model expects the team profile

and outcome indicators to change under a simulated replacement, but they do not account for all real-world factors that shape transfers.

7.3 Contributions

This dissertation makes five main contributions in the examination of player-team fit in football analytics.

First, it proposes a framework that combines behavioural embeddings with an embedding-conditioned Large Events Model for football event simulation. This connects player and team behavioural representations with sequential event generation, allowing the model to analyse football performance beyond aggregate statistics.

Second, it implements a LoRA-based adaptation approach for team-specific football contexts. The method freezes the base LEM and trains small team-specific adapters for each team. Therefore, the approach supports efficient specialization while preserving the event knowledge learned by the base model.

Third, it defines a player and team embedding construction process based on event-level behavioural profiles. Player embeddings summarize repeated behavioural tendencies from event data, while team embeddings are built from event-weighted player representations. As a result, the framework delivers compact representations of player behaviour and team behavioural identity.

Fourth, it introduces an embedding-based transfer simulation mechanism. This mechanism estimates how a target team representation changes when an outgoing player is replaced by an incoming player. The resulting team identity movement provides an interpretable signal for analysing whether a transfer preserves, adjusts, or strongly changes the team's tactical profile.

Finally, it presents an evaluation design that combines multiple complementary perspectives. The framework is assessed through event-level validation loss, outcome metrics, tactical style metrics, transfer fit indicators, baseline comparisons, and computational efficiency analysis. Together, these evaluation dimensions offer a more comprehensive view of player-team fit than outcome prediction alone.

7.4 Limitations

Although the framework produced useful behavioural and tactical signals, some limitations remain.

The main season-level evaluation was performed on the same season used for team-specific adaptation. This choice was useful as a diagnostic setting because it removed part of the cross-season generalization challenge and allowed the study to test whether the adapted models could reproduce team behaviour under favourable conditions. Despite this, same-season evaluation does not prove that the framework generalizes to future seasons, different squads, new coaches, or changing tactical contexts.

Another limitation concerns the absence of explicit player sampling during match simulation. The model uses player and team embeddings as behavioural context, but the simulator does not dynamically select the acting player at each generated event. Therefore, player embeddings influence the generated sequence as contextual information, but the simulation does not reproduce detailed player-level decision-making inside each match sequence.

The embeddings themselves also simplify the football context. Player embeddings summarize repeated event-level tendencies, while team embeddings are built from weighted combinations of player representations. This makes the framework compact and easier to interpret, but it also means that individual variation, tactical roles, and within-team structure are reduced to aggregate behavioural signals.

The available data introduces an additional constraint because it is limited to on-ball information. As a result, the framework does not capture off-ball movement, defensive positioning away from the ball, pressing triggers, team shape, formations, or explicit tactical instructions. For this reason, the embeddings should be interpreted as representations of observable behavioural tendencies, not as complete descriptions of tactical identity.

The match simulation also starts from a neutral kickoff-like context. This creates a stable and controlled starting point for Monte Carlo simulation, but it removes match-specific information such as lineups, fatigue, home advantage, opponent strategy, injuries, and match importance. As a result, the generated outcomes should be interpreted as controlled estimates of team tendency, not as exact fixture predictions.

The transfer simulations should not be read as causal validation of real transfer success. The framework estimates how a team representation and its simulated outcome indicators change after replacing one player with another. However, real transfer performance also depends on factors outside the model, including adaptation time, coaching decisions, squad hierarchy, injuries, psychology, contractual context, and the player's actual tactical role after joining the club.

The number of transfer scenarios is also limited. The selected cases illustrate different types of recruitment situations, including compatible upgrades, low-disruption replacements, and more transformational profiles. Nevertheless, a larger set of real and hypothetical transfers would be needed to test if the same interpretation patterns hold across positions, leagues, team styles, and recruitment objectives.

Together, these limitations show that the framework should be understood as a decision-support approach, not as a deterministic prediction system. Its main value lies in producing interpretable behavioural, tactical, and embedding-space signals that can complement scouting and recruitment analysis.

7.5 Future Work

The framework developed in this dissertation can be extended in several directions.

A natural next step is to evaluate the adapted models across future seasons and different competitions. The same-season evaluation used in this dissertation was useful as a diagnostic setting,

but future work should test whether the learned team-specific patterns remain stable when squads, coaches, opponents, and tactical contexts change. This would provide a stronger view of generalization and would help separate short-term reproduction from more persistent team identity.

Another important extension is event generation with explicit player sampling. At the moment, player embeddings provide behavioural context, but the simulator does not select the acting player at each generated event. Adding this step would make the simulation more player-specific and would allow the model to represent how different players participate within the same team structure.

Moreover, finishing and goal calibration should be improved. The corrected training objective produced more realistic goal totals, but the model still compressed outcome predictions. Future work could include more detailed shot-quality modeling, improved goal probability calibration, and post-simulation calibration methods to better reflect the difference between dominant and weaker teams.

The embedding construction process could also be extended with role-aware and position-aware representations. Future embeddings could move beyond aggregate behavioural profiles by distinguishing between tactical roles, positions, and zones of responsibility. This would be especially useful when comparing players who operate in different systems or perform different functions within the same nominal position.

A further improvement would be to include off-ball and tracking data. Event data captures what happens on the ball, but many tactical behaviours depend on movement, spacing, pressing, defensive shape, and positioning away from the ball. Adding this information would make the representation of tactical identity more complete and would reduce the gap between behavioural embeddings and real tactical roles.

Finally, future work should evaluate a larger set of transfer scenarios, including real transfers with known post-transfer outcomes. This would not prove causality by itself, but it would allow the framework to be tested against a broader range of cases, positions, leagues, and recruitment objectives. In parallel, LoRA could be compared with other parameter-efficient fine-tuning methods and with continuous fine-tuning strategies to assess whether alternative adaptation methods improve team-specific behaviour, reduce outcome compression, or provide better stability across seasons.

7.6 Final Remarks

This dissertation showed that player-team fit can be studied through behavioural and tactical signals, rather than solely through past performance. Although the proposed framework still has important limitations, it demonstrates that event-level models, behavioural embeddings, and efficient team adaptation can be combined to support a more contextual view of recruitment analysis. Overall, the work provides a foundation for future systems that help clubs evaluate not only whether a player is good, but how that player may fit, preserve, or change a specific team identity.

References

[1] Koen van Arem, Floris Goes-Smit, and Jakob Söhl. "Forecasting the Future Development in Quality and Value of Professional Football Players". In: *Applied Sciences* 15 (2025). ISSN: 2076-3417. DOI: 10.3390/app15168916.

[2] Tom Decroos et al. "Actions Speak Louder than Goals: Valuing Player Actions in Soccer". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '19. Association for Computing Machinery, 2019. ISBN: 9781450362016. DOI: 10.1145/3292500.3330758.

[3] Nigel Doering et al. *Empirical Analysis of Efficient Fine-Tuning Methods for Large Pre-Trained Language Models*. 2024.

[4] footballytics. *Summary of Ian Graham's Presentation at StatsBomb Conference 2021*. Transcription of a presentation by Ian Graham at the StatsBomb Conference 2021. 2021. URL: https://www.footballytics.ch/post/analyticssetup-opportunities-challenges (visited on 10/13/2025).

[5] Avelina Asada Hadji-Kyriacou and Ognjen Arandjelovic. *Context-PEFT: Efficient Multi-Modal, Multi-Task Fine-Tuning*. 2023.

[6] Danny Hill, James Skinner, and Anna Grosman. "A review of football player metrics and valuation methods: a typological framework of football player valuations". In: *Managing Sport and Leisure* (2025). DOI: 10.1080/23750472.2025.2459727.

[7] Edward Hu et al. "LoRA: Low-Rank Adaptation of Large Language Models". In: *ICLR 2022 - 10th International Conference on Learning Representations* (2022).

[8] Stephanie Kovalchik. "Extension of the Elo rating system to margin of victory". In: *International Journal of Forecasting* 36 (2020). ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2020.01.006.

[9] Pavlina Kröckel and Freimut Bodendorf. "Process Mining of Football Event Data: A Novel Approach for Tactical Insights Into the Game". In: *Frontiers in Artificial Intelligence* 3 (2020). DOI: 10.3389/frai.2020.00047.

[10] Jiří Lahvička. "Using Monte Carlo Simulation to Calculate Match Importance: The Case of English Premier League". In: *Journal of Sports Economics* 16.4 (2015). DOI: 10.1177/1527002513490172.

[11] Rob Mackenzie. "The Changing Face of Scouting: The power and limitations of data". In: *Sky Sports* (2017).

[12] Anqi Mao, Mehryar Mohri, and Yutao Zhong. "Cross-Entropy Loss Functions: Theoretical Analysis and Applications". In: *Proceedings of Machine Learning Research* 202 (2023).

[13] Tiago Mendes-Neves, Luís Meireles, and João Mendes-Moreira. *Estimating Player Performance in Different Contexts Using Fine-tuned Large Events Models*. 2024. URL: https://arxiv.org/abs/2402.06815.

[14] Joshua Noble. *What is LoRA (low-rank adaptation)?* URL: https://www.ibm.com/think/topics/lora.

[15] CIES Football Observatory. "Global economic analysis of the transfer market (2015-2024)". In: *CIES Football Observatory* (2024).

[16] Konstantinos Pelechrinis and Wayne Winston. "A Skellam regression model for quantifying positional value in soccer". In: *Journal of Quantitative Analysis in Sports* 17 (2021). DOI: 10.1515/jqas-2019-0122.

[17] Statsbomb. *What is Expected Threat (xT)? Possession Value models explained.* URL: https://statsbomb.com/soccer-metrics/possession-value-models-explained/.

[18] José Teixeira et al. "Mapping football tactical behavior and collective dynamics with artificial intelligence: a systematic review". In: *Frontiers in Sports and Active Living* 7 (2025). DOI: 10.3389/fspor.2025.1569155.

[19] Ferran Vidal-Codina et al. "Automatic event detection in football using tracking data". In: *Sports Engineering* 25 (2022). DOI: 10.1007/s12283-022-00381-6.

[20] Wen Zhang et al. "A Survey on Negative Transfer". In: *IEEE/CAA Journal of Automatica Sinica* 10.2 (2023). ISSN: 2329-9274. DOI: 10.1109/jas.2022.106004.

[21] Fuzhen Zhuang et al. "A Comprehensive Survey on Transfer Learning". In: *Proceedings of the IEEE* (2021). DOI: 10.1109/JPROC.2020.3004555.

Appendix A

Supplementary Figures

A.1 Supplementary VAE Representations

This section presents supplementary representations based on a Variational Autoencoder latent space. These figures are included as an alternative to the PCA projections used in the main analysis. While PCA provides a linear two-dimensional projection of the embedding space, the VAE learns a nonlinear latent representation. Therefore, these figures provide an additional view of player and team identity, helping to verify whether the observed transfer-related movements are still visible under a different approach.

Figure A.1 shows the player identity map in the VAE latent space. The grey points represent the full player population, while the highlighted points help illustrate how different player types occupy different regions.

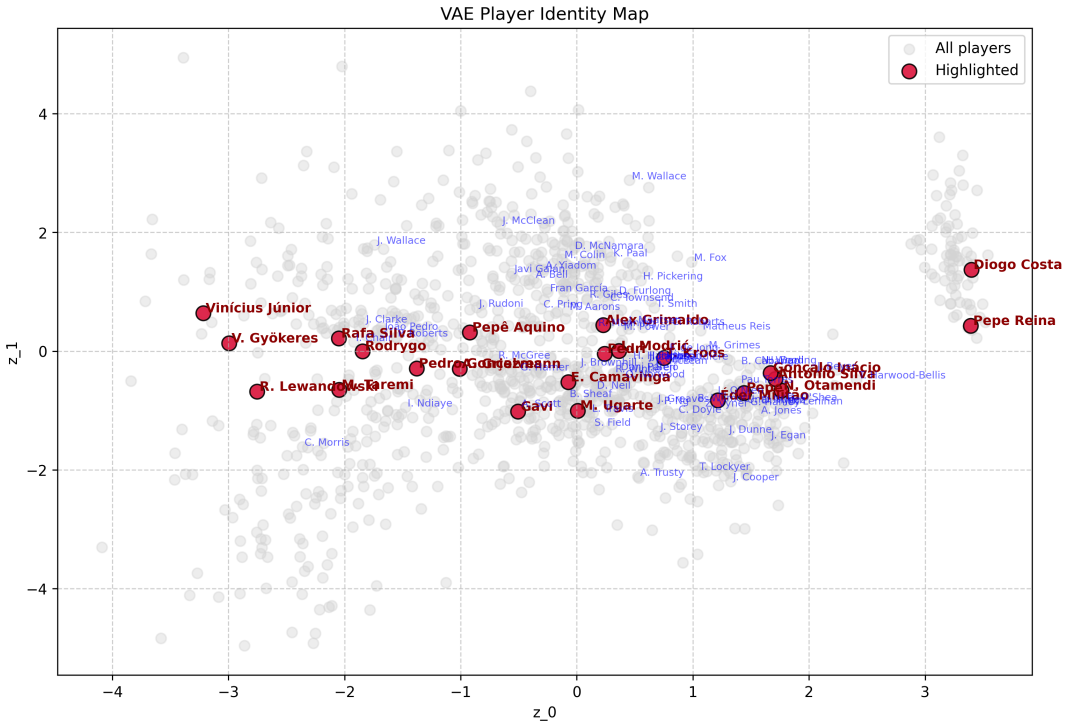


Figure A.1: VAE player identity map for the 2022/2023 data.

Figures A.2, A.3, and A.4 show the projection of simulated transfer scenarios in the team VAE latent space. In each figure, the blue points represent real teams, the orange point represents the original target team, and the green marker represents the simulated post-transfer team representation. The movement between the original and simulated points provides an alternative visual signal of how much the transfer changes the team identity in the learned latent space.

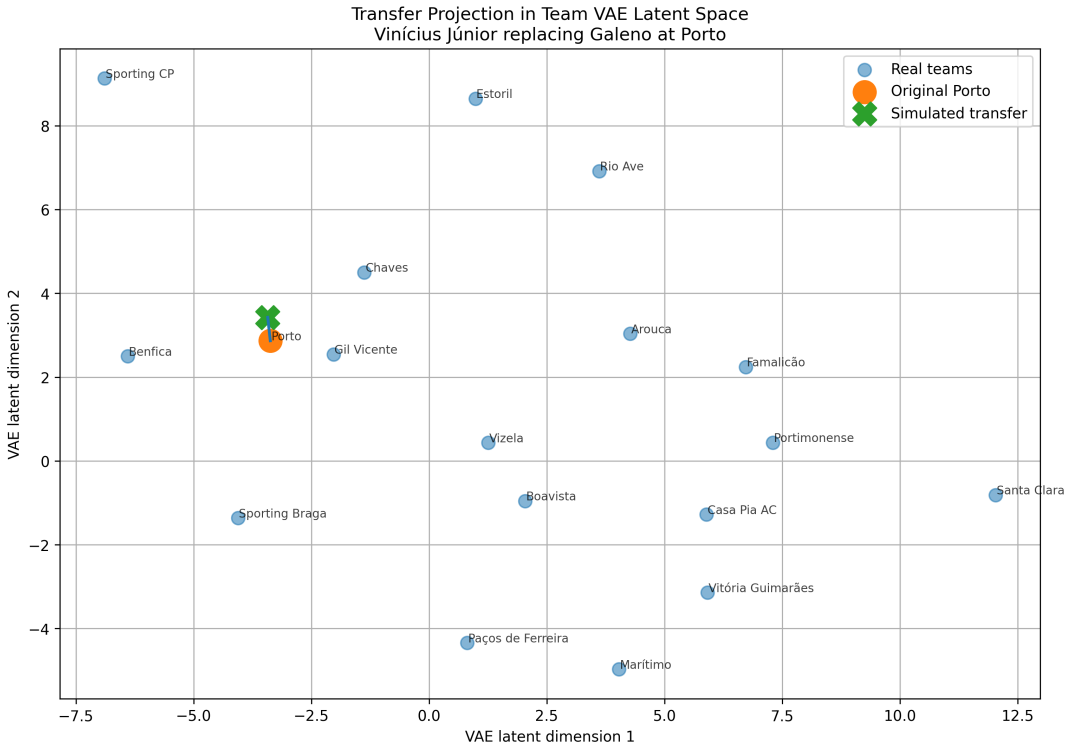


Figure A.2: Projection of the simulated Vinícius Júnior to Porto transfer in the team VAE latent space.

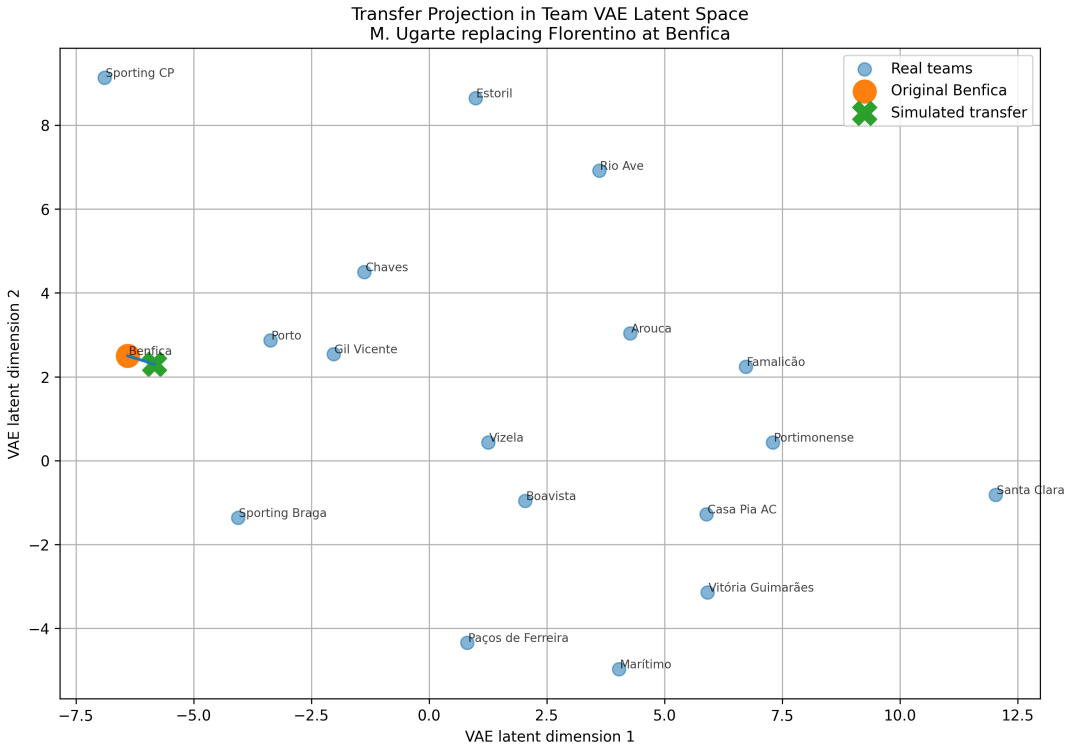


Figure A.3: Projection of the simulated Manuel Ugarte to Benfica transfer in the team VAE latent space.

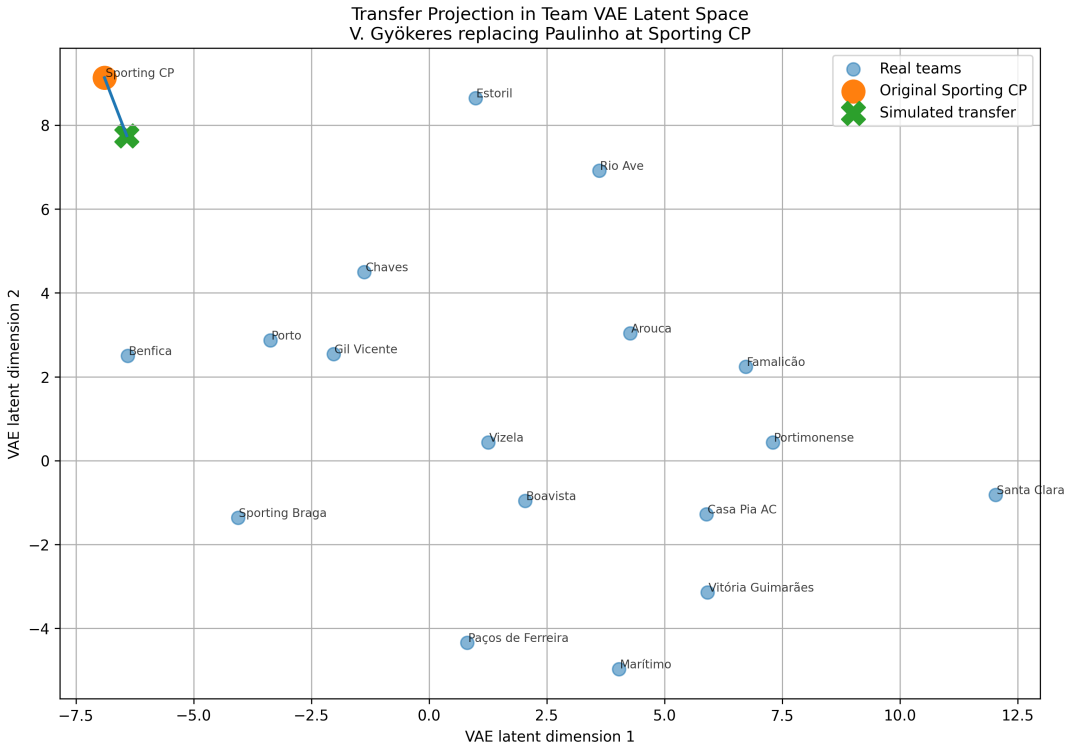


Figure A.4: Projection of the simulated Viktor Gyökeres to Sporting CP transfer in the team VAE latent space.

A.2 LoRA Adapter Training Diagnostics

This section presents supplementary diagnostics used to inspect the LoRA adaptation configuration. Benfica was used as a representative case to compare different training lengths and verify whether longer adaptation improved the event-level objective.

Figure A.5 shows the training loss evolution across the tested epoch limits. The loss drops quickly in the first epochs and then stabilizes, indicating that the adapter captures most of the team signal early in training. Figure A.6 shows the corresponding validation loss. The validation curve follows a similar decreasing trend, with no sign of overfitting. Most of the improvement occurs before epoch 10, while the gains after that point are smaller.

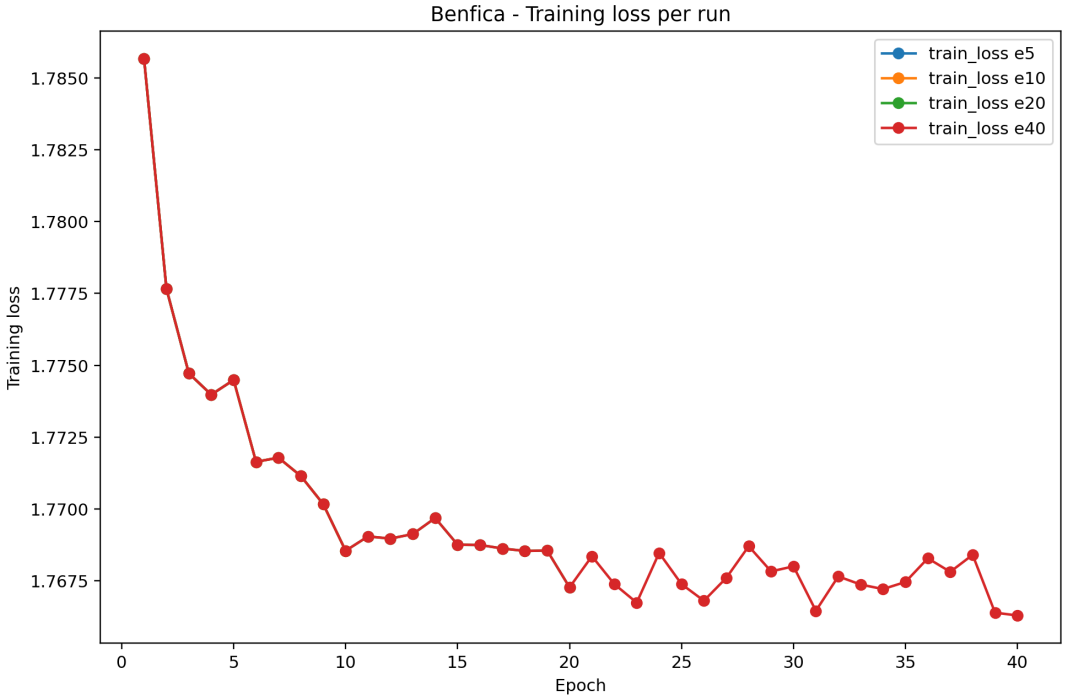


Figure A.5: Training loss evolution for the Benfica LoRA adaptation under different epoch limits.

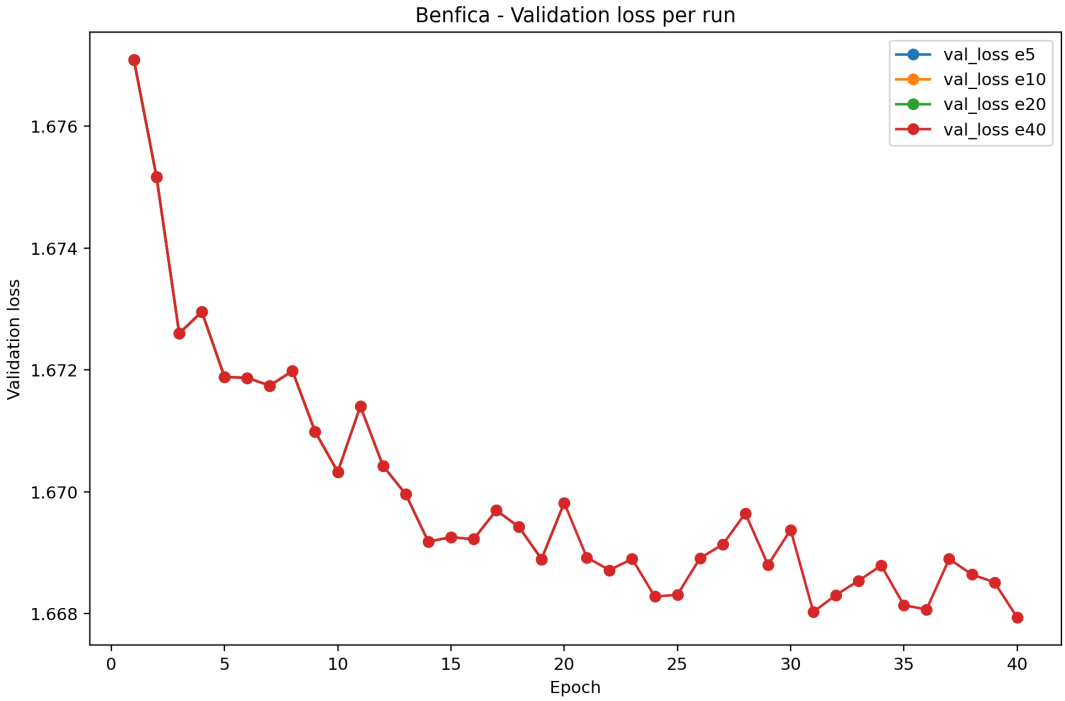


Figure A.6: Validation loss evolution for the Benfica LoRA adaptation under different epoch limits.

A.3 Supplementary Full-Season Simulation Tables

This section presents the complete predicted league tables for the base LEM and the LoRA-adapted LEM. These tables provide the detailed team-level values behind the aggregate outcome comparison discussed in Section 6.4. They are included in the appendix because the main chapter focuses on the overall interpretation of the results, while these tables report the full set of predicted and real values for each team.

Base LEM Predicted League Table

Pred Rank	Team	Pred xPts	Pred GF	Pred GA	Pred GD	Real Rank	Real Pts	Real GF	Real GA	Real GD	Rank Error	Pts Error	GF Error	GA Error	Style MAE
1	Benfica	70.43	58.34	23.38	34.96	1	89	78.0	19.0	59.0	0	18.57	19.66	4.38	0.16
2	Sporting CP	67.21	55.79	25.69	30.1	4	69	66.0	31.0	35.0	2	1.79	10.21	5.31	0.25
3	Porto	57.87	47.29	31.06	16.23	2	85	71.0	18.0	53.0	1	27.13	23.71	13.06	0.16
4	Sporting Braga	56.99	46.72	31.9	14.82	3	77	72.0	30.0	42.0	1	20.01	25.28	1.9	0.32
5	Gil Vicente	51.47	42.15	35.05	7.1	13	37	31.0	41.0	-10.0	8	14.47	11.15	5.95	0.26
6	Vitória Guimarães	44.9	37.47	39.69	-2.22	6	53	34.0	38.0	-4.0	0	8.1	3.47	1.69	0.22
7	Rio Ave	44.89	37.35	39.54	-2.19	10	40	34.0	40.0	-6.0	3	4.89	3.35	0.46	0.23
8	Portimonense	44.58	37.07	39.82	-2.76	15	34	25.0	47.0	-22.0	7	10.58	12.07	7.18	0.24
9	Marítimo	43.25	36.62	41.55	-4.93	16	27	31.0	62.0	-31.0	7	16.25	5.62	20.45	0.17
10	Famalicão	42.66	35.72	41.1	-5.38	8	45	37.0	45.0	-8.0	2	2.34	1.28	3.9	0.25
11	Chaves	41.24	35.69	43.3	-7.61	9	44	34.0	40.0	-6.0	2	2.76	1.69	3.3	0.22
12	Vizela	41.12	34.96	42.57	-7.62	12	39	30.0	37.0	-7.0	0	2.12	4.96	5.57	0.28
13	Arouca	40.68	34.14	42.28	-8.14	5	54	36.0	36.0	0.0	8	13.32	1.86	6.28	0.27
14	Boavista	40.03	33.55	42.77	-9.22	7	46	42.0	52.0	-10.0	7	5.97	8.45	9.23	0.21
15	Paços de Ferreira	38.09	33.05	45.34	-12.29	17	24	25.0	58.0	-33.0	2	14.09	8.05	12.66	0.23
16	Estoril	37.87	32.84	45.33	-12.49	14	35	33.0	47.0	-14.0	2	2.87	0.16	1.67	0.2
17	Casa Pia AC	37.71	32.7	45.45	-12.75	11	40	29.0	38.0	-9.0	6	2.29	3.7	7.45	0.32
18	Santa Clara	35.36	31.13	47.29	-16.16	18	22	26.0	55.0	-29.0	0	13.36	5.13	7.71	0.25

Figure A.7: Complete predicted league table generated by the base LEM for the 2022/2023 Liga Portugal season.

LoRA-Adapted LEM Predicted League Table

Pred Rank	Team	Pred xPts	Pred GF	Pred GA	Pred GD	Real Rank	Real Pts	Real GF	Real GA	Real GD	Rank Error	Pts Error	GF Error	GA Error	Style MAE
1	Benfica	74.38	71.42	26.58	44.84	1	89	78.0	19.0	59.0	0	14.62	6.58	7.58	0.23
2	Sporting CP	70.25	67.39	29.8	37.59	4	69	66.0	31.0	35.0	2	1.25	1.39	1.2	0.25
3	Porto	60.41	58.64	37.74	20.89	2	85	71.0	18.0	53.0	1	24.59	12.36	19.74	0.19
4	Sporting Braga	57.86	57.83	40.53	17.3	3	77	72.0	30.0	42.0	1	19.14	14.17	10.53	0.29
5	Gil Vicente	51.38	36.59	28.87	7.73	13	37	31.0	41.0	-10.0	8	14.38	5.59	12.13	0.24
6	Vitória Guimarães	43.74	36.21	40.06	-3.85	6	53	34.0	38.0	-4.0	0	9.26	2.21	2.06	0.18
7	Famalicão	43.72	34.93	38.3	-3.37	8	45	37.0	45.0	-8.0	1	1.28	2.07	6.7	0.25
8	Rio Ave	43.66	26.05	28.09	-2.03	10	40	34.0	40.0	-6.0	2	3.66	7.95	11.91	0.23
9	Marítimo	42.95	49.23	56.58	-7.34	16	27	31.0	62.0	-31.0	7	15.95	18.23	5.42	0.13
10	Portimonense	41.78	23.21	27.19	-3.97	15	34	25.0	47.0	-22.0	5	7.78	1.79	19.81	0.2
11	Vizela	41.42	50.55	61.18	-10.63	12	39	30.0	37.0	-7.0	1	2.42	20.55	24.18	0.26
12	Chaves	40.93	25.58	31.28	-5.7	9	44	34.0	40.0	-6.0	3	3.07	8.42	8.72	0.18
13	Boavista	39.61	43.79	56.3	-12.51	7	46	42.0	52.0	-10.0	6	6.39	1.79	4.3	0.17
14	Estoril	38.98	48.46	63.23	-14.78	14	35	33.0	47.0	-14.0	0	3.98	15.46	16.23	0.22
15	Arouca	37.87	28.86	39.92	-11.06	5	54	36.0	36.0	0.0	10	16.13	7.14	3.92	0.24
16	Casa Pia AC	36.58	25.5	38.39	-12.89	11	40	29.0	38.0	-9.0	5	3.42	3.5	0.39	0.26
17	Paços de Ferreira	36.4	52.39	73.54	-21.15	17	24	25.0	58.0	-33.0	0	12.4	27.39	15.54	0.2
18	Santa Clara	35.2	40.42	60.2	-19.78	18	22	26.0	55.0	-29.0	0	13.2	14.42	5.2	0.23

Figure A.8: Complete predicted league table generated by the LoRA-adapted LEM for the 2022/2023 Liga Portugal season.