A hierarchical Bayesian-LLM framework for multi-objective optimization in Fantasy Premier League

Tuan Thi[1]

[1][V+] FPL League

Fantasy Premier League (FPL) presents a complex multi-objective optimization problem requiring sequential decision-making under uncertainty. We present a novel hierarchical framework combining Bayesian statistical modeling, genetic algorithms, and Large Language Model (LLM) analysis for optimal team selection. Our approach integrates a modified Bradley-Terry model with uncertainty quantification, role-specific performance weights, and real-time validation through LLM agents. Testing on the 2024/25 season data (668 active players, 52 optimized teams), our system achieved 336.5-338.2 projected points over 5 gameweeks, representing a 10.8% improvement over baseline strategies. The framework successfully identified and corrected for player transfers (removing 2 departed players), enforced strict squad constraints (15 players: 2 GK, 5 DEF, 5 MID, 3 FWD), and optimized captain selection, with Mohamed Salah (9.78 expected points) consistently selected. Real-world deployment yielded progression from rank 609,310 to 19,601 (top 0.2%) over two seasons. Our results demonstrate the effectiveness of combining traditional optimization with modern AI techniques for constrained portfolio problems.

Introduction

Fantasy Premier League (FPL) engages over 11 million participants globally in a constrained portfolio optimization problem[1]. Players must construct a 15-player squad within a £100m budget, adhering to position requirements (2 goalkeepers, 5 defenders, 5 midfielders, 3 forwards) and team diversity constraints (maximum 3 players per club). Weekly decisions involve selecting 11 starters, designating a captain for double points, and managing transfers within a 4-point penalty system.

This creates a multi-stage stochastic optimization problem with several challenges: (i) player performance uncertainty due to injuries, rotation, and form fluctuations; (ii) dynamic pricing based on ownership and performance; (iii) fixture difficulty variations; (iv) interdependent team selection constraints; and (v) long-term versus short-term trade-offs in transfer strategy.

Previous approaches have addressed subsets of this problem. Matthews et al.[2] applied integer programming for single-gameweek optimization. Bonomo et al.[3] used genetic algorithms but ignored prediction uncertainty. Recent work by Joseph et al.[4] incorporated machine learning for player scoring but treated team selection as a separate problem.

We present an integrated framework that addresses these limitations through: (1) a hierarchical Bradley-Terry model with Bayesian uncertainty quantification for player performance prediction; (2) multi-objective genetic optimization respecting all FPL constraints; (3) LLM-based validation and tactical analysis; and (4) real-time data integration for dynamic adjustments.

Methods

Mathematical formulation

Let $\mathcal{P} = \{p_1, p_2, ..., p_n\}$ denote the set of n available players. Each player p_i has attributes: cost $c_i \in \mathbb{R}^+$, position $r_i \in \{\text{GK, DEF, MID, FWD}\}$, team $t_i \in \mathcal{T}$, and predicted score $s_i \in \mathbb{R}^+$.

The optimization objective over horizon H gameweeks is:

$$\max_{\mathbf{X},\mathbf{Y},\mathbf{C}} \sum_{h=1}^{H} \sum_{i=1}^{n} s_{i,h} \cdot (y_{i,h} + c_{i,h}) - 4 \cdot \tau_h \tag{1}$$

Subject to constraints:

$$\sum_{i=1}^{n} c_i \cdot x_i \leq 100 \qquad \text{(budget)} \tag{2}$$

$$\sum_{i=1}^{n} x_i = 15 \qquad \text{(squad size)} \tag{3}$$

$$\sum_{i:r_i=r} x_i = q_r, \quad \forall r \qquad \text{(position requirements)} \tag{4}$$

$$\sum_{i:t_i=t} x_i \leq 3, \quad \forall t \qquad \text{(team limit)} \tag{5}$$

$$\sum_{i=1}^{n} y_{i,h} = 11, \quad \forall h \qquad \text{(starting XI)} \tag{6}$$

$$y_{i,h} \leq x_i, \quad \forall i, h \qquad \text{(selection feasibility)} \tag{7}$$

$$\sum_{i=1}^{n} c_{i,h} = 1, \quad \forall h \qquad \text{(one captain)} \tag{8}$$

Where $x_i \in \{0, 1\}$ indicates squad membership, $y_{i,h} \in \{0, 1\}$ indicates starting selection in gameweek h, $c_{i,h} \in \{0, 1\}$ indicates captaincy, τ_h is the number of transfers, and q_r represents position quotas.

Hierarchical Bradley-Terry model

We model player performance using a two-level hierarchy. At the individual level, the probability that player i outperforms player j in a head-to-head comparison is:

$$P(i > j|\boldsymbol{\theta}) = \frac{\exp(\theta_i)}{\exp(\theta_i) + \exp(\theta_j)} \tag{9}$$

where θ_i represents player i's latent ability. We extend this with contextual factors:

$$\theta_i = \mu_i + \beta_{t_i} + \gamma_{r_i} + \alpha \cdot \mathbb{I}[\text{home}] + \epsilon_i \tag{10}$$

Here $\mu_i \sim \mathcal{N}(0, \sigma_\mu^2)$ is the baseline ability, β_{t_i} captures team strength, γ_{r_i} represents position-specific effects, α is home advantage, and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ models individual variation.

We employ variational Bayesian inference to estimate posterior distributions:

$$q(\boldsymbol{\theta}) = \prod_{i=1}^{n} \mathcal{N}(\theta_i | m_i, v_i) \tag{11}$$

This provides uncertainty quantification crucial for risk assessment. Players with high variance v_i represent higher-risk selections.

Multi-objective genetic algorithm

We implement a genetic algorithm with population size $N_p = 500$ evolved over $G = 100$ generations. The fitness function combines multiple objectives:

$$F(\mathbf{x}) = w_1 \cdot S(\mathbf{x}) + w_2 \cdot D(\mathbf{x}) - w_3 \cdot R(\mathbf{x}) + w_4 \cdot V(\mathbf{x}) \tag{12}$$

where $S(\mathbf{x})$ is expected score, $D(\mathbf{x})$ measures formation diversity, $R(\mathbf{x})$ quantifies risk, and $V(\mathbf{x})$ represents value efficiency. Weights are set as $\mathbf{w} = [0.5, 0.2, 0.2, 0.1]$ based on parameter tuning.

LLM validation and analysis

Selected teams undergo validation through a Claude-3.5-Sonnet LLM agent configured with:

- Constraint verification (squad rules, budget, captain selection)
- Tactical coherence assessment

- Player compatibility analysis
- Risk stratification
- Injury and rotation monitoring via web scraping

The LLM corrects violations and provides qualitative insights complementing quantitative metrics.

Results

Model performance

Testing on 2024/25 season data (38 gameweeks, 668 active players after removing 2 transferred players), our Bradley-Terry model achieved strong predictive performance. Key player rankings showed clear differentiation:

- Mohamed Salah: $\theta = 2.281 \pm 0.067$, projected 9.78 points/game
- Cole Palmer: $\theta = 1.826 \pm 0.099$, projected 6.22 points/game
- Bryan Mbeumo: $\theta = 1.732 \pm 0.084$, projected 5.99 points/game

Position-specific weights derived from historical data: Goalkeepers (1.15), Defenders (1.08), Midfielders (1.12), Forwards (0.98), reflecting scoring patterns.

Team optimization

The genetic algorithm generated 52 valid teams meeting all constraints. Top teams showed consistent patterns:
Formation distribution: 4-5-1 (73%), 4-4-2 (19%), 3-5-2 (8%)
Budget utilization: Mean £99.7m (range £98.5-100.0m)
Captain selection: Mohamed Salah (100% after LLM validation)
Key players appearing in ¿80% of top teams:

- GK: Matz Sels (Nottingham Forest, £5.0m, 4.37 points)
- DEF: Joško Gvardiol (Man City, £6.0m, 4.60 points)
- MID: Mohamed Salah (Liverpool, £14.5m, 9.78 points), Cole Palmer (Chelsea, £10.5m, 6.22 points)
- FWD: Joël Piroe (Leeds, £5.5m, 3.98 points)

LLM validation impact

The LLM agent identified and corrected critical issues:

1. **Player eligibility:** Removed Joe Hodge (transferred to CD Tondela) and Luis Díaz (departed Liverpool)
2. **Captain optimization:** Corrected 31% of teams incorrectly captaining Chris Wood over Mohamed Salah
3. **Formation compliance:** Fixed 12 teams violating position requirements
4. **Risk assessment:** Classified teams as low (42%), medium (46%), high (12%) risk

Performance projections

Final validated teams showed strong expected performance:
Compared to baseline strategies:

- Template team (top 6 clubs only): 305 points (+10.8% improvement)
- Previous season's top team carried forward: 298 points (+13.4% improvement)
- Random valid team: 287 points (+17.8% improvement)

Team	Formation	Budget	GW1 Score	5GW Score	Confidence
1	4-5-1	£100.0m	65.3	336.5	85%
2	4-5-1	£100.0m	65.6	337.9	82%
3	4-5-1	£100.0m	65.7	338.2	80%

Table 1: Top 3 teams after validation

Real-world deployment

Historical deployment (2022/23 and 2023/24 seasons) demonstrated practical effectiveness:

Season 1: Rank 609,310 → 152,847 (top 2.3%)

Season 2: Rank 152,847 → 19,601 (top 0.2%)

Key success factors included consistent captain selection (92% accuracy), effective transfer timing (average -2.1 points/transfer vs. -4.0 baseline), and risk management during fixture swings.

Discussion

Our framework successfully addresses the multi-faceted challenges of FPL optimization through methodological innovations:

Uncertainty quantification: The Bayesian approach provides actionable risk metrics. High-variance players like Darwin Núñez ($v = 0.152$) were correctly identified as rotation risks and excluded from final teams.

Constraint satisfaction: The genetic algorithm reliably produced valid teams, with LLM validation catching edge cases like departed players that pure optimization missed.

Multi-horizon planning: Balancing immediate returns (GW1) with medium-term performance (5GW) yielded more stable strategies than single-week optimization.

Human-AI collaboration: LLM analysis provided tactical insights (”Nottingham Forest defensive double-up exploits favorable fixtures”) that enhanced purely statistical decisions.

Limitations and future work

Several areas warrant further investigation:

1. **Dynamic pricing:** Our model assumes static prices, but FPL implements weekly adjustments based on transfers. Incorporating price prediction could improve long-term planning.
2. **Chip strategy:** Special chips (Triple Captain, Bench Boost, Free Hit, Wildcard) offer significant scoring opportunities. Optimal timing remains an open problem.
3. **Psychological factors:** Ownership percentages influence point swings through effective rank changes. Differential strategies require balancing template coverage with unique selections.
4. **Computational efficiency:** Full season simulation remains computationally intensive. Approximation methods could enable real-time optimization.

Conclusion

We presented a comprehensive framework combining Bayesian statistics, evolutionary computation, and large language models for Fantasy Premier League optimization. The approach achieved 336.5-338.2 projected points, representing 10.8% improvement over baseline strategies. Real-world deployment yielded top 0.2% finishes, validating practical effectiveness.

The methodology extends beyond FPL to general portfolio optimization under constraints, sequential decision-making with uncertainty, and human-AI collaborative systems. As fantasy sports grow in sophistication and participation, such frameworks become increasingly valuable for both recreational and professional applications.

Data availability

All code and data are available at `https://github.com/[repository]`. Player statistics sourced from official FPL API (`https://fantasy.premierleague.com/api/`).

References

1. Fantasy Premier League. *Official Statistics 2024/25*. Available at: `https://fantasy.premierleague.com`
2. Matthews, T., Ramchurn, S. D. & Chalkiadakis, G. Competing with humans at Fantasy Football: Team formation in large partially-observable domains. *Proc. AAAI Conf. Artif. Intell.* **26**, 1394–1400 (2012).
3. Bonomo, F., Durán, G. & Marenco, J. Mathematical programming as a tool for virtual soccer coaches. *OR Spectrum* **36**, 771–793 (2014).
4. Joseph, A., Fenton, N. E. & Neil, M. Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems* **19**, 544–553 (2006).
5. Bradley, R. A. & Terry, M. E. Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika* **39**, 324–345 (1952).
6. Davidson, R. R. On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *J. Am. Stat. Assoc.* **65**, 317–328 (1970).
7. Cattelan, M. Models for paired comparison data: A review with emphasis on dependent data. *Stat. Sci.* **27**, 412–433 (2012).
8. Hvattum, L. M. & Arntzen, H. Using ELO ratings for match result prediction in association football. *Int. J. Forecast.* **26**, 460–470 (2010).
9. Dixon, M. J. & Coles, S. G. Modelling association football scores and inefficiencies in the football betting market. *Appl. Stat.* **46**, 265–280 (1997).
10. Maher, M. J. Modelling association football scores. *Stat. Neerl.* **36**, 109–118 (1982).

Acknowledgements

We thank the FPL community for insights and the Claude AI system for validation support.

Author contributions

All authors contributed equally to methodology development, implementation, and manuscript preparation.

Competing interests

The authors declare no competing interests.

Supplementary information

Supplementary information including detailed algorithms, additional figures, and complete season results is available online.