TacticAI: an AI assistant for football tactics

Zhe Wang[*,1], Petar Veličković[*,1], Daniel Hennes[*,1], Nenad Tomašev[1], Laurel Prince[1], Michael Kaisers[1], Yoram Bachrach[1], Romuald Elie[1], Li Kevin Wenliang[1], Federico Piccinini[1], William Spearman[2], Ian Graham[4], Jerome Connor[1], Yi Yang[1], Adrià Recasens[1], Mina Khan[1], Nathalie Beauguerlange[1], Pablo Sprechmann[1], Pol Moreno[1], Nicolas Heess[1], Michael Bowling[3], Demis Hassabis[1] and Karl Tuyls[1]

[*]Contributed equally to this work, [1]Google DeepMind, [2]Liverpool FC, [3]University of Alberta, [4]Work completed while at Liverpool FC

Identifying key patterns of tactics implemented by rival teams, and developing effective responses, lies at the heart of modern football. However, doing so algorithmically remains an open research challenge. To address this unmet need, we propose TacticAI, an AI football tactics assistant developed and evaluated in close collaboration with domain experts from Liverpool FC. We focus on analysing corner kicks, as they offer coaches the most direct opportunities for interventions and improvements. TacticAI incorporates both a predictive and a generative component, allowing the coaches to effectively sample and explore alternative player setups for each corner kick routine and to select those with the highest predicted likelihood of success. We validate TacticAI on a number of relevant benchmark tasks: predicting receivers and shot attempts and recommending player position adjustments. The utility of TacticAI is validated by a qualitative study conducted with football domain experts at Liverpool FC. We show that TacticAI's model suggestions are not only indistinguishable from real tactics, but also favoured over existing tactics 90% of the time, and that TacticAI offers an effective corner kick retrieval system. TacticAI achieves these results despite the limited availability of gold-standard data, achieving data efficiency through geometric deep learning.

Introduction

Association football, or simply *football* or *soccer*, is a widely popular and highly professionalised sport, in which two teams compete to score goals against each other. As each football team comprises of up to 11 active players at all times and takes place on a very large pitch (also known as a soccer field), scoring goals tends to require a significant degree of strategic team-play. Under the rules codified in the *Laws of the Game* [20], this competition has nurtured an evolution of nuanced strategies and tactics, culminating in modern professional football leagues. In today's play, data-driven insights are a key driver in determining the optimal player setups for each game and developing counter-tactics to maximise the chances of success [38].

When competing at the highest level the margins are incredibly tight, and it is increasingly important to be able to capitalise on any opportunity for creating an advantage on the pitch. To that end, top-tier clubs employ diverse teams of coaches, analysts and experts, tasked with studying and devising (counter-)tactics before each game. Several recent methods attempt to improve tactical coaching and player decision-making through artificial intelligence (AI) tools, using a wide variety of data types from videos to tracking sensors and applying diverse algorithms ranging from simple logistic regression to elaborate neural network architectures. Such methods have been employed to

Corresponding author(s): Zhe Wang (zhewang@google.com), Petar Veličković (petarv@google.com), Karl Tuyls (karltuyls@google.com)
Google DeepMind, 5 New Street Square, London EC4A 3TW

Google DeepMind 2024 年 3 月 22 日

TacticAI: an AI assistant for football tactics

Zhe Wang[*,1], Petar Veličković[*,1], Daniel Hennes[*,1], Nenad Tomašev[1], Laurel Prince[1], Michael Kaisers[1], Yoram Bachrach[1], Romuald Elie[1], Li Kevin Wenliang[1], Federico Piccinini[1], William Spearman[2], Ian Graham[4], Jerome Connor[1], Yi Yang[1], Adrià Recasens[1], Mina Khan[1], Nathalie Beauguerlange[1], Pablo Sprechmann[1], Pol Moreno[1], Nicolas Heess[1], Michael Bowling[3], Demis Hassabis[1] and Karl Tuyls[1]

[*]Contributed equally to this work, [1]Google DeepMind, [2]Liverpool FC, [3]University of Alberta, [4]Work completed while at Liverpool FC

确定竞争对手球队实施的战术关键模式并发展有效的应对策略，是现代足球的核心。然而，算法化地完成这一任务仍是一个开放的研究挑战。为了满足这一未满足的需求，我们提出了TacticAI，这是一个与利物浦足球俱乐部领域的专家紧密合作开发和评估的AI足球战术助手。我们将重点分析角球，因为它们为教练提供了最直接的干预和改进机会。 TacticAI融合了预测和生成两种组件，使教练能够有效地采样和探索每个角球战术的替代球员配置，并选择那些预测成功率最高的方案。我们在若干相关基准任务上验证了TacticAI：预测接球手和射门尝试，并推荐球员位置调整。通过与利物浦足球俱乐部足球领域专家进行的定性研究，验证了TacticAI的实用性。我们展示了TacticAI的模型建议不仅在90%的情况下无法与真实战术区分，而且比现有战术更受欢迎，并且TacticAI提供了一个有效的角球检索系统。尽管高质量标准数据可用性有限，TacticAI仍取得了这些成果，通过几何深度学习实现了数据效率。

Introduction

足球，简称“足球”或“soccer”，是一项广受欢迎且高度职业化的运动，两个队伍在比赛中相互对抗得分。由于每支足球队在任何时候都由最多11名活跃球员组成，并且比赛在一个非常大的场地（也称为足球场）上进行，因此得分往往需要相当程度的战略团队协作。《比赛规则》 [20]规定了比赛规则，这些规则孕育了细致的策略和战术的演变，最终形成了现代职业足球联赛。在今天的比赛中，数据驱动的洞察是确定每场比赛最佳球员配置和开发对策以最大化成功机会的关键因素[38]。

在最高水平的竞争中，差距极其微小，越来越需要能够利用任何机会在场上创造优势。为此，顶级俱乐部雇佣了由各种教练、分析师和专家组成的多样化团队，负责在每场比赛前研究和制定（对策）战术。近年来，一些方法试图通过人工智能（AI）工具改进战术指导和球员决策，使用从视频到追踪传感器的各种数据类型，并应用从简单的逻辑回归到复杂的神经网络架构的各种算法。这些方法被用于帮助从视频中预测射门事件 [16]，从时空数据中预测屏幕外动作 [27]，判断比赛是否进行中或中断 [23]，或者识别球员动作 [2]。

球员在场上的计划执行是高度动态且不完美的，这取决于包括球员的体力和疲劳、球员移动和定位的变化、天气、场地状况以及对阵球队的应对等多方面因素。相比之下，定位球提供了一个控制结果的机会，因为短暂的比赛中断允许球员根据一项经过练习和预先约定的模式重新定位，并有意向球门发起进攻。这样的定位球包括任意球、角球、球门球、界外球和点球 [38]。

Corresponding author(s): Zhe Wang (zhewang@google.com), Petar Veličković (petarv@google.com), Karl Tuyls (karltuyls@google.com)
Google DeepMind, 5 New Street Square, London EC4A 3TW
© 2024 Google DeepMind. All rights reserved

help predict shot events from videos [16], forecast off-screen movement from spatio-temporal data [27], determine whether a match is in-play or interrupted [23], or identify player actions [2].

The execution of agreed-upon plans by players on the pitch is highly dynamic and imperfect, depending on numerous factors including player fitness and fatigue, variations in player movement and positioning, weather, the state of the pitch, and the reaction of the opposing team. In contrast, *set pieces* provide an opportunity to exert more control on the outcome, as the brief interruption in play allows the players to reposition according to one of the practiced and pre-agreed patterns, and make a deliberate attempt towards the goal. Examples of such set pieces include free kicks, corner kicks, goal kicks, throw-ins, and penalties [38].

Among set pieces, *corner kicks* are of particular importance, as an improvement in corner kick execution may substantially modify game outcomes, and they lend themselves to principled, tactical and detailed analysis. This is because corner kicks tend to occur *frequently* in football matches (with ~10 corners on average taking place in each match [33]), they are taken from a fixed, *rigid* position, and they offer an *immediate* opportunity for scoring a goal—no other set piece simultaneously satisfies all of the above. In practice, corner kick routines are determined well ahead of each match, taking into account the strengths and weaknesses of the opposing team and their typical tactical deployment. It is for this reason that we focus on corner kick analysis in particular, and propose **TacticAI**, an AI football assistant for supporting the human expert with set piece analysis, and the development and improvement of corner kick routines.

TacticAI is rooted in learning efficient representations of corner kick tactics from raw, spatio-temporal player tracking data. It makes efficient use of this data by representing each corner kick situation as a *graph*—a natural representation for modelling relationships between players (Figure 1-A, Table 2), and these player relationships may be of higher importance than the absolute distances between them on the pitch [1]. Such a graph input is a natural candidate for graph machine learning models [41], which we employ within TacticAI to obtain high-dimensional latent player representations.

Uniquely, TacticAI takes advantage of *geometric deep learning* [5] to explicitly produce player representations that respect several *symmetries* of the football pitch (Figure 1-B). As an illustrative example, we can usually safely assume that under a horizontal or vertical reflection of the pitch state, the game situation is equivalent. Geometric deep learning ensures that TacticAI's player representations will be identically computed under such reflections, such that this symmetry does not have to be learnt from data. This proves to be a valuable addition, as high-quality tracking data is often limited—with only a few hundred matches played each year in every league. We provide an in-depth overview of how we employ geometric deep learning in TacticAI in the Methods section.

From these representations, TacticAI is then able to answer various *predictive* questions about the outcomes of a corner—for example, which player is most likely to make first contact with the

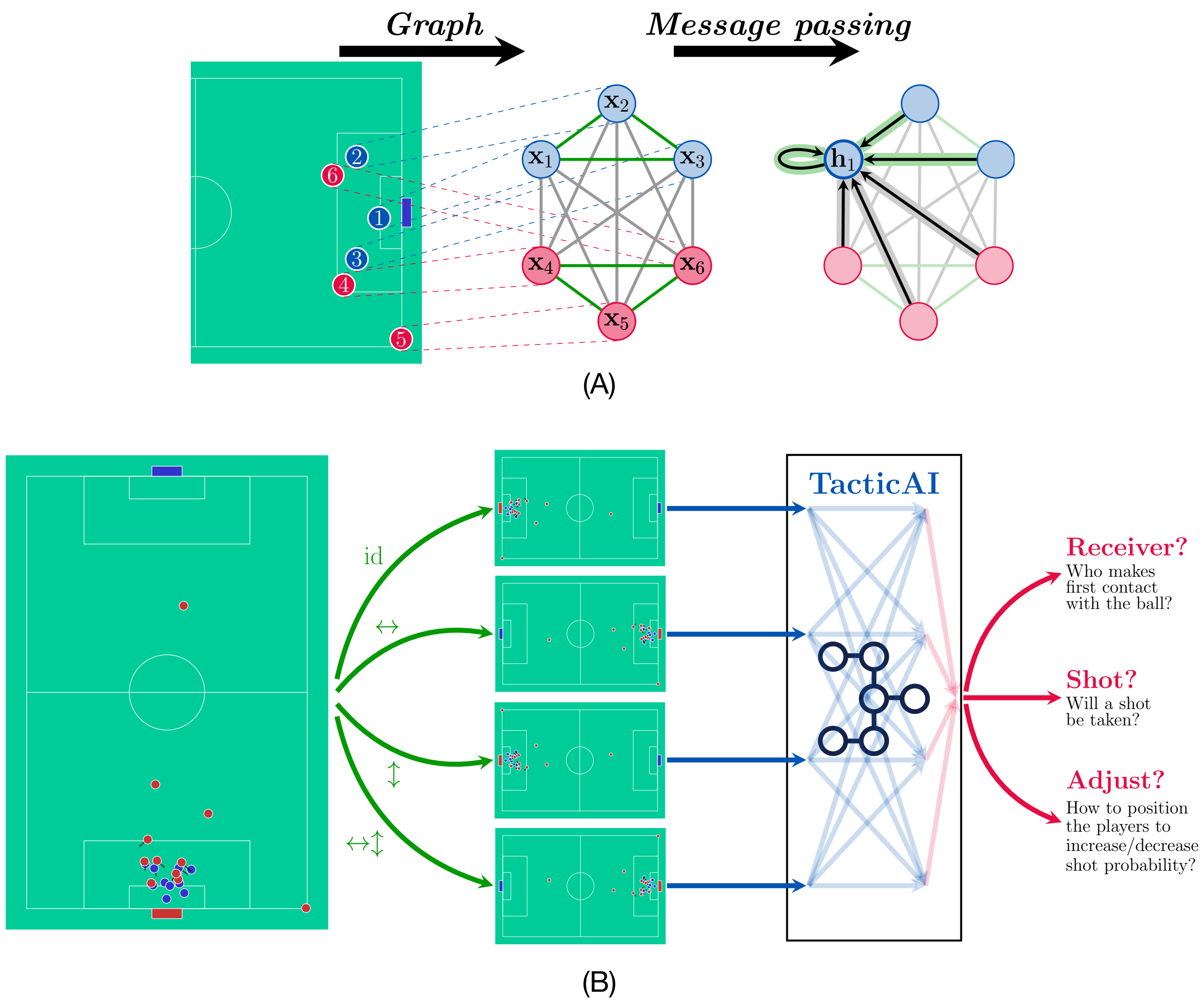


Figure 1 | **对TacticAI的“鸟瞰”概述。 (A)**，如何将角球情况转换为图表示。每个球员被视为图中的一个节点，节点、边和图的特性按照正文的详细说明提取。然后，图神经网络通过执行消息传递在这个图上操作；每个节点的表示通过来自其相邻节点的消息进行更新。 **(B)**，TacticAI如何处理给定的角球。为了确保TacticAI在面对水平或垂直反射时的答案具有鲁棒性，将所有可能的反射组合应用于输入角球，然后这四种视角被输入到TacticAI的核心模型中，它们能够相互交互以计算最终的球员表示—每个“内部蓝色箭头”对应于(A)中的一个消息传递层。一旦计算出球员表示，它们可以用来预测角球的接球者，是否已经射门，以及对球员位置和速度的辅助调整，这些调整会增加或减少射门的概率。

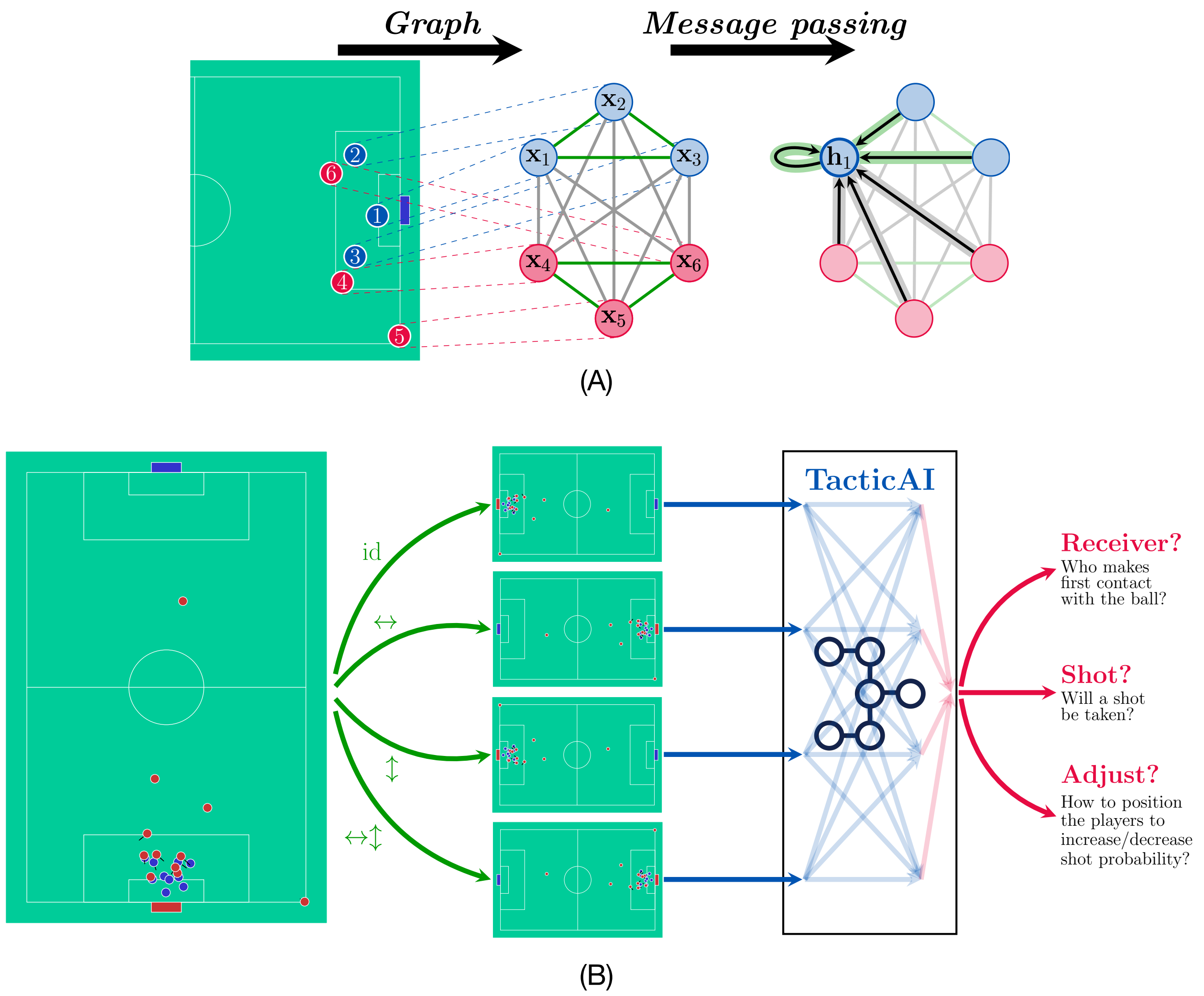


Figure 1 | **A "bird's eye" overview of TacticAI. (A)**, how corner kick situations are converted to a graph representation. Each player is treated as a node in a graph, with node, edge and graph features extracted as detailed in the main text. Then, a graph neural network operates over this graph by performing *message passing*; each node's representation is updated using the messages sent to it from its neighbouring nodes. **(B)**, how TacticAI processes a given corner kick. To ensure that TacticAI's answers are robust in the face of horizontal or vertical reflections, all possible combinations of reflections are applied to the input corner, and these four views are then fed to the core TacticAI model, where they are able to interact with each other to compute the final player representations—each "internal blue arrow" corresponds to a single message passing layer from (A). Once player representations are computed, they can be used to predict the corner's receiver, whether a shot has been taken, as well as assistive adjustments to player positions and velocities, which increase or decrease the probability of a shot being taken.

TacticAI: an AI assistant for football tactics

在所有定位球中，角球尤为重要，因为角球执行的提升可能会显著改变比赛结果，并且它们适合进行原则性、战术性和详细的分析。这是因为角球在足球比赛中往往频繁发生（平均每场比赛有~10个角球 [33]），它们从固定、死板的位置发出，并且提供了立即进球的机会——没有其他定位球能同时满足以上所有条件。在实践中，角球战术会在每场比赛前早早确定，考虑到对手球队的优势和劣势以及他们典型的战术部署。正是出于这个原因，我们特别关注角球分析，并提出了**TacticAI**，这是一个支持人类专家进行定位球分析、角球战术的发展和改进的AI足球助手。

TacticAI从原始的、时空球员追踪数据中学习角球战术的高效表示。它通过将每个角球情况表示为图——一种自然表示球员之间关系的建模方式（图 1-A，表 2），高效利用这些数据，而这些球员关系可能比他们在场上的绝对距离更为重要 [1]。这样的图输入是图机器学习模型 [41]的自然选择，我们在TacticAI中采用这些模型来获得高维潜在球员表示。

独特的是，TacticAI利用几何深度学习 [5]明确地生成尊重足球场几项对称性的球员表示（图 1-B）。作为一个说明性的例子，我们可以通常安全地假设，在场地状态的水平和垂直反射下，游戏情况是等价的。几何深度学习确保了TacticAI的球员表示在这些反射下将完全相同地计算，使得这种对称性无需从数据中学习。这被证明是一个宝贵的补充，因为高质量的追踪数据通常是有限的——每个联赛每年只有几百场比赛。我们在方法部分详细介绍了如何在TacticAI中应用几何深度学习。

从这些表示中，TacticAI能够回答有关角球结果的多种预测性问题——例如，哪个球员最有可能首先接触到球，或者是否会发生射门。TacticAI还可以作为一个检索系统——基于球员表示的相似性挖掘类似的角球情况——以及一个生成性推荐系统，建议调整球员的位置和速度以最大化或最小化估计的射门概率。通过与利物浦FC的领域专家教练和分析人员进行案例研究，进行了几项实验，我们在下一节中呈现的结果提供了清晰的统计证据，证明TacticAI能够立即提供有用、现实和准确的战术建议。

Results and Analysis

为了展示我们方法的多样品质，我们设计了一个包含三个不同预测和生成组件的TacticAI：接球手预测，射门预测，以及通过引导生成的战术推荐，这些也对应于量化评估TacticAI的基准任务。除了利用其预测组件为角球分析提供精确的量化洞察外，TacticAI的预测与生成组件之间的相互作用还允许教练对每个感兴趣的动作采样替代的球员设置，并直接评估这些替代方案可能产生的结果。

我们首先描述我们的定量分析，该分析显示TacticAI的预测组件在预测未参与训练的测试角球中的接球手和射门情况方面是准确的，并且所提出的球员调整并没有与实际情况强烈偏离。然而，这种分析仅间接揭示了TacticAI在部署后会有多有用。我们正面解决这个问题，并与利物浦足球俱乐部的合作伙伴进行了一项全面的案例研究——在研究中，我们直接请人类专家评分员评判TacticAI的预测和球员调整的有用性。以下各节将展开介绍我们所进行的特定结果和分析。

在以下内容中，我们将以理解我们的评估所必需的最低程度描述TacticAI的组件。我们将对TacticAI组件的详细描述推迟到方法部分。

ball, or whether a shot will take place. TacticAI can also be used as a *retrieval* system—for mining similar corner kick situations based on the similarity of player representations—and a generative *recommendation* system, suggesting adjustments to player positions and velocities to maximise or minimise estimated shot probability. Through several experiments within a case study with domain expert coaches and analysts from Liverpool FC, the results of which we present in the next section, we obtain clear statistical evidence that TacticAI readily provides useful, realistic and accurate tactical suggestions.

Results and Analysis

To demonstrate the diverse qualities of our approach, we design TacticAI with three distinct predictive and generative components: *receiver prediction*, *shot prediction*, and tactic recommendation through *guided generation*, which also correspond to the benchmark tasks for quantitatively evaluating TacticAI. In addition to providing accurate quantitative insights for corner kick analysis with its predictive components, the interplay between TacticAI's predictive and generative components allows coaches to sample alternative player setups for each routine of interest, and directly evaluate the possible outcomes of such alternatives.

We will first describe our quantitative analysis, which demonstrates that TacticAI's predictive components are accurate at predicting corner kick receivers and shot situations on held-out test corners, and that the proposed player adjustments do not strongly deviate from ground-truth situations. However, such an analysis only gives an indirect insight into how *useful* TacticAI would be once deployed. We tackle this question of utility head-on, and conduct a comprehensive case study in collaboration with our partners at Liverpool FC—where we directly ask human expert raters to judge the utility of TacticAI's predictions and player adjustments. The following sections expand on the specific results and analysis we have performed.

In what follows, we will describe TacticAI's components at a minimal level necessary to understand our evaluation. We defer detailed descriptions of TacticAI's components to the Methods section.

Benchmarking TacticAI

We evaluate the three components of TacticAI on a relevant benchmark dataset of corner kicks. Our dataset consists of $7,176$ corner kicks from the 2020–2021 Premier League seasons, which we randomly shuffle and split into a training (80%) and a test set (20%). As previously mentioned, TacticAI operates on *graphs*. Accordingly, we represent each corner kick situation as a graph, where each node corresponds to a player. The features associated with each node encode the movements (velocities and positions) and simple profiles (heights and weights) of on-pitch players at the timestamp when the corresponding corner kick was being taken by the attacking kicker (see the Methods section), and no information of ball movement was encoded. The graphs are *fully connected*; that is, for every

TacticAI: an AI assistant for football tactics

Benchmark Task	Graph Learning Task	Node Features	Edge Features	Global Features
Receiver Prediction	Node classification	Player positions Player velocities Player weights Player heights Ball possession	Teammate or opponent	None
Shot Prediction	Graph classification	Same as above	Same as above	Receiver ID
Guided Generation	Node regression	Same as above	Same as above	Shot indicator Receiver ID

Table 1 | 相应任务中使用的特征摘要。

对TacticAI策略的基准测试

我们在一个相关的角球基准数据集上评估了TacticAI的三个组成部分。我们的数据集包含了2020–2021英超赛季的$7,176$次角球，我们将这些数据随机打乱并分为训练集（80%）和测试集（20%）。如前所述，TacticAI在图上操作。因此，我们将每个角球情境表示为一个图，其中的每个节点对应一个球员。与每个节点相关的特征编码了在进攻踢球手执行角球时的戳记时刻场上球员的动作（速度和位置）和简单信息（身高和体重）（详见方法部分），并没有编码球移动的信息。这些图是完全连接的；也就是说，对于每对球员，我们都会在图中包含连接他们的边。这些边中的每一个都编码了一个二进制特征，指示这两名球员是否属于对方球队。对于每个任务，我们生成了相关的节点/边/图特征数据集及其对应标签（表 1 和 2，详见方法部分）。然后分别用对应的角球图对这些组件进行训练。特别是，我们只使用最小特征集来构建角球图，没有编码球的移动，也没有将球员之间的距离显式地编码到图中。对于所有基准任务，我们使用了一致的训练-测试划分，因为这使得我们不仅可以对各个组件进行基准测试，还可以测试它们之间的交互作用。

通过几何深度学习精确预测接球手和射门

TacticAI的一个关键预测模型是在场上的22名球员中预测接球手。接球手被定义为角球开出后第一个触球的球员。在我们的评估中，所有方法都使用了相同的一组特征（见表 1中的接球手预测条目和方法部分）。我们利用接球手预测任务来基准测试几种不同的TacticAI基础模型。我们表现最佳的模型—在经过$50,000$次训练步骤后，在top-3测试准确度上达到了0.782 ± 0.039—是一个深度图注意力网络 [4, 43]，通过使用D_2群卷积 [7]实现了几何深度学习 [5]。我们通过一个详细的消融研究来补充这个结果，验证了我们选择的基础架构和群卷积在接球手预测任务上带来了显著的改进（S.Table 2，见方法部分中的消融研究）。考虑到角球接球手的预测是一项极具挑战性的任务，有许多我们的模型无法看到因素—包括疲劳和体能水平以及实际的球轨迹—我们认为TacticAI的top-3准确度反映了高度的预测能力，并在后续研究中保持基础TacticAI架构不变。除了这个使用评估数据集的定量评估之外，我们还通过一个由人类评分者参与的案例研究来评估TacticAI接球预测组件的性能。更多细节请见案例研究部分。

Benchmark Task	Graph Learning Task	Node Features	Edge Features	Global Features
Receiver Prediction	Node classification	Player positions Player velocities Player weights Player heights Ball possession	Teammate or opponent	None
Shot Prediction	Graph classification	Same as above	Same as above	Receiver ID
Guided Generation	Node regression	Same as above	Same as above	Shot indicator Receiver ID

Table 1 | **Summary of the features used in the corresponding tasks.**

Feature	Feature Type	Explanation
Player positions	Node	XY-positions of 22 players on the pitch.
Player velocities	Node	XY-velocities of 22 players on the pitch.
Player weights	Node	Weights of 22 players.
Player heights	Node	Heights of 22 players.
Ball possession	Node	Binary indicator to indicate whether this player is possessing the ball.
Teammate or opponent	Edge	One-hot encoding to indicate the relationship between two players.
Receiver ID	Global	One-hot encoding to indicate the index of the receiver.
Shot indicator	Global	Binary indicator to indicate if there was a threatening shot attempt.

Table 2 | **Summary of the details of the features used to construct graphs.**

pair of players, we will include the edge connecting them in the graph. Each of these edges encodes a binary feature, indicating whether the two players are on opposing teams or not. For each task, we generated the relevant dataset of node/edge/graph features and corresponding labels (Table 1 and 2, see the Methods section). The components were then trained separately with their corresponding corner kick graphs. In particular, we only employ a minimal set of features to construct the corner kick graphs, without encoding the movements of the ball nor explicitly encoding the distances between players into the graphs. We used a consistent training-test split for all benchmark tasks, as this made it possible to benchmark not only the individual components, but also their interactions.

TacticAI: an AI assistant for football tactics

Feature	Feature Type	Explanation
Player positions	Node	XY-positions of 22 players on the pitch.
Player velocities	Node	XY-velocities of 22 players on the pitch.
Player weights	Node	Weights of 22 players.
Player heights	Node	Heights of 22 players.
Ball possession	Node	Binary indicator to indicate whether this player is possessing the ball.
Teammate or opponent	Edge	One-hot encoding to indicate the relationship between two players.
Receiver ID	Global	One-hot encoding to indicate the index of the receiver.
Shot indicator	Global	Binary indicator to indicate if there was a threatening shot attempt.

Table 2 | **构建图表所用特征详细概要。**

对于射门预测，我们观察到，直接重用基础的TacticAI架构来直接预测射门事件—即直接建模无条件的概率$\mathbb{P}(\text{shot})$—是具有挑战性的，仅得到了一个测试F_1得分[1]为0.52 ± 0.03。然而，鉴于我们已经拥有一个强大的接球手预测器，我们决定使用它的输出来为我们提供关于是否发生射门的额外洞察。因此，我们选择将射门概率分解为

$$\mathbb{P}(\text{shot}) = \sum_{i \in \text{players}} \mathbb{P}(\text{shot}|\text{receiver} = i)\mathbb{P}(\text{receiver} = i) \tag{1}$$

其中 $\mathbb{P}(\text{receiver})$ 是由TacticAI的接球手预测系统计算出的概率，而 $\mathbb{P}(\text{shot}|\text{receiver})$ 模型表示的是在“特定球员首次接触到球后”的条件射门概率。这是通过在相应的角球中提供一个额外的全局特征来指示接球手来实现的（见表 1），而其余的网络架构与接球手预测的架构保持一致（S.图 2，参见方法部分）。在训练时，我们将地面真实接球手作为模型的输入——在推断时，我们尝试每一个可能的接球手，使用TacticAI接球手预测器给出的概率来权衡他们的贡献，如方程 1 所示。这种两阶段的方法使得射门预测的最终测试F_1分数达到了 0.64 ± 0.02，它编码了比无条件射门预测器更多的信号，特别是考虑到与预测射门事件相关联的许多不可观测因素。

此外，我们还观察到，即使只是通过预测接球者，而没有明确分类角落的其他显著特征，TacticAI也能学习到数据的一般化表示。具体来说，具有相似战术模式的团队配置在TacticAI的潜在空间中往往聚集在一起（图2）。然而，在原始输入空间中没有观察到明确的聚类（S.图 1）。这表明TacticAI可以作为一个有用的角球检索系统，我们将在案例研究部分介绍这一假设的评估。

使用分类条件生成模型控制战术细化

配备了能够将角球与其各种结果（例如接球者和射门事件）强烈关联的组件，我们可以探索TacticAI在调整战术上的应用，以增加或减少某些结果发生的可能性。

[1] F_1得分是精确度和召回率的调和平均值,它通常用于处理不平衡数据集的二分类问题。

Accurate receiver and shot prediction through geometric deep learning

One of TacticAI's key predictive models forecasts the receiver out of the 22 on-pitch players. The receiver is defined as the first player touching the ball after the corner was taken. In our evaluation, all methods used the same set of features (see the Receiver Prediction entry in Table 1 and the Methods section). We leveraged the receiver prediction task to benchmark several different TacticAI base models. Our best performing model—achieving 0.782 ± 0.039 in top-3 test accuracy after $50,000$ training steps—was a deep graph attention network [4, 43], leveraging geometric deep learning [5] through the use of D_2 group convolutions [7]. We supplement this result with a detailed ablation study, verifying that both our choice of base architecture and group convolution yielded significant improvements on the receiver prediction task (S.Table 2, see the Ablation Study in the Methods section). Considering that corner kick receiver prediction is a highly challenging task with many factors that are unseen by our model—including fatigue and fitness levels, and actual ball trajectory—we consider TacticAI's top-3 accuracy to reflect a high level of predictive power, and keep the base TacticAI architecture fixed for subsequent studies. In addition to this quantitative evaluation with the evaluation dataset, we also evaluate the performance of TacticAI's receiver prediction component in a case study with human raters. Please see the Case Study section for more details.

For shot prediction, we observe that reusing the base TacticAI architecture to directly predict shot events—i.e., directly modelling the unconditional probability $\mathbb{P}(\text{shot})$—proved challenging, only yielding a test F_1 score[1] of 0.52 ± 0.03. However, given that we already have a potent receiver predictor, we decided to use its output to give us additional insight into whether or not a shot had been taken. Hence, we opted to decompose the probability of taking a shot as

$$\mathbb{P}(\text{shot}) = \sum_{i \in \text{players}} \mathbb{P}(\text{shot}|\text{receiver} = i)\mathbb{P}(\text{receiver} = i) \tag{1}$$

where $\mathbb{P}(\text{receiver})$ are the probabilities computed by TacticAI's receiver prediction system, and $\mathbb{P}(\text{shot}|\text{receiver})$ models the conditional shot probability *after a specific player made first contact with the ball*. This was implemented through providing an additional global feature to indicate the receiver in the corresponding corner kick (Table 1) while the architecture otherwise remained the same as that of receiver prediction (S.Figure 2, see the Methods section). At training time, we feed the ground-truth receiver as input to the model—at inference time, we attempt every possible receiver, weighing their contributions using the probabilities given by TacticAI's receiver predictor, as per Equation 1. This two-phased approach yielded a final test F_1 score of 0.64 ± 0.02 for shot prediction, which encodes significantly more signal than the unconditional shot predictor, especially considering the many unobservables associated with predicting shot events.

Moreover, we also observe that, even just through predicting the receivers, without explicitly classifying any other salient features of corners, TacticAI learned generalisable representations of

[1]The F_1 score is the harmonic mean of the precision and recall, and it is commonly used in binary classification problems over imbalanced datasets.

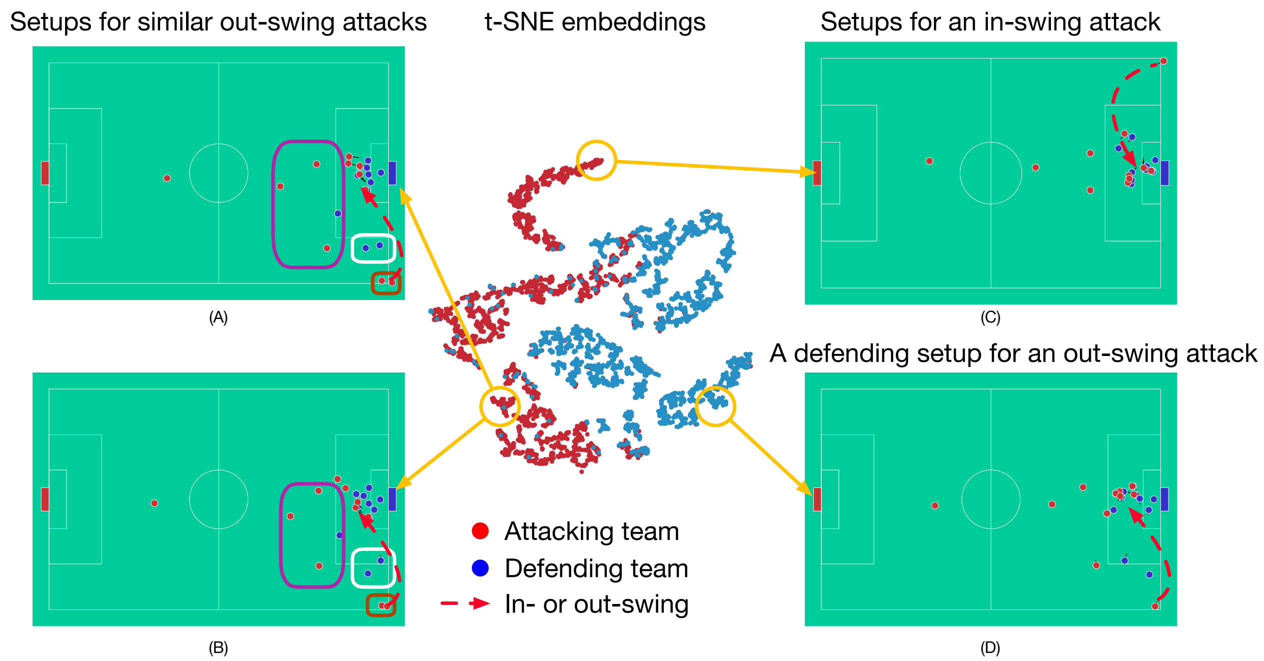


Figure 2 | **由TacticAI塑造的潜在空间中的角球表示。** 我们使用t-SNE可视化了$1,024$次角球中进攻和防守队的潜在表示。一次角球样本中的潜在队伍嵌入是同一进攻队伍(**(A)**, **(B)**和**(C)**)或防守队伍**(D)**的潜在球员表示的平均值。在给定参考角球样本**(A)**的情况下，我们根据潜在空间中其表示的最短距离检索另一个角球样本**(B)**。我们观察到**(A)**和**(B)**都是外摆角球，并且它们的进攻战术具有相似的模式，这些模式用具有相同颜色的矩形突出显示，尽管它们在球员的绝对位置和速度方面存在差异。与此同时，内摆进攻**(C)**的潜在表示在潜在空间中与**(A)**和**(B)**相距甚远。红色箭头仅用于展示内摆与外摆角球之间的区别，而非实际的球轨迹。

具体而言，我们的目标是针对两支队伍中的一支，调整球员的动作，包括他们的位置和速度，以最大化或最小化在初始角球设置条件下射门事件的发生概率，初始角球设置包括两支队伍球员的动作，以及他们的身高和体重。特别地，虽然在现实场景中两支队伍可能会同时对彼此的动作做出反应，但在我们的研究中，我们专注于对球员动作进行适度调整，这有助于发现那些没有正确响应战术的球员。由于这个原因，我们通过仅为一个队伍生成调整，而将另一个队伍保持固定来简化战术细化的过程。我们通过一个自编码目标来训练这个任务的模型：我们将地面真实射门结果（一个二元指标）作为额外的图级别特征输入到TacticAI的模型中（表1），并使其学习重构输入球员坐标的概率分布（图1(B)，也见方法部分）。这个基于自编码器的生成模型是一个独立组件，与TacticAI的预测系统分离。这三种系统共享编码器架构（不共享参数），但使用不同的解码器（见方法部分）。在推理时，我们可以将一个期望的射门结果作为给定角球设置输入，然后使用这个概率分布为某一队伍的球员采样新的位置和速度。这个设置原则上允许灵活的下游使用，使得人类教练可以通过生成针对他们感兴趣的具体结果的调整来优化角球设置——例如，增加攻击队的射门概率，降低防守队的射门概率（图3），或者增加特定前锋接球的机会。

我们首先定量评估生成的调整策略，方法是验证它们与真实角球相比是否容易被一个分类器区分。为此，我们合成一个包含200个角球样本及其相应条件生成调整的数据集。具体来说，对于没有射门事件的角球，我们将射门事件特征设置为1，为进攻方生成调整策略；反之，对于发生了射门事件的角球，我们则为防守方生成调整策略。我们发现，真实样本和生成样本不能被多层感知

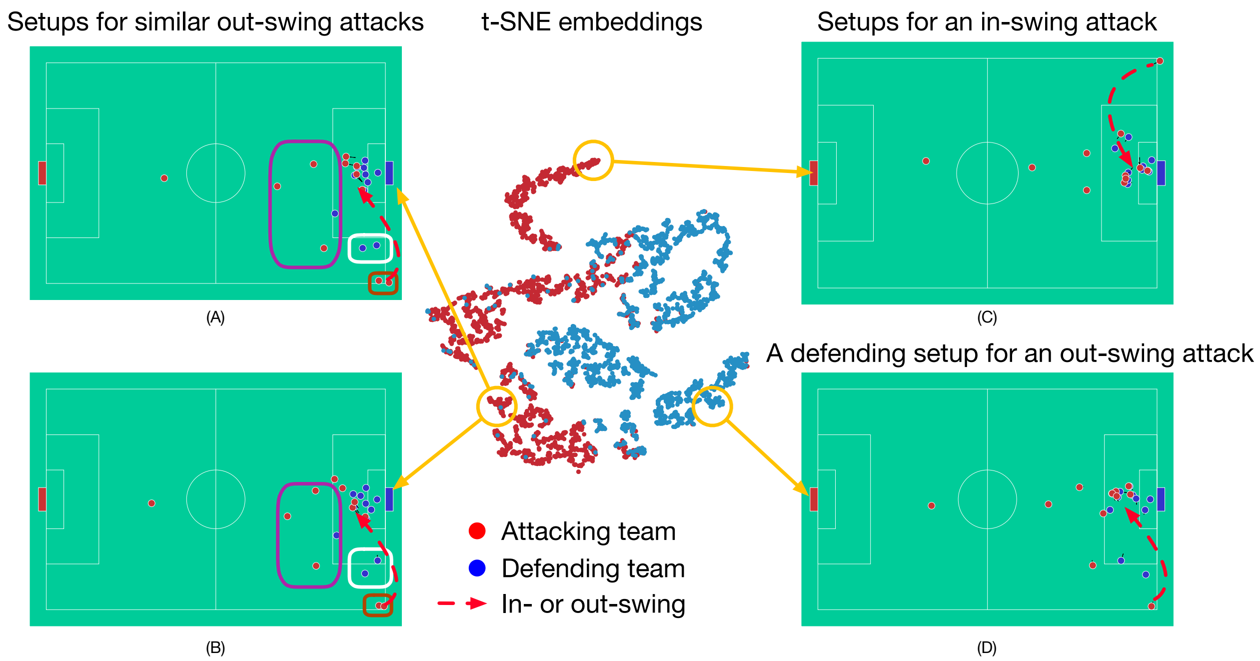


Figure 2 | **Corner kicks represented in the latent space shaped by TacticAI.** We visualise the latent representations of attacking and defending teams in 1,024 corner kicks using *t*-SNE. A latent team embedding in one corner kick sample is the mean of the latent player representations on the same attacking (**(A)**, **(B)** and **(C)**) or defending **(D)** team. Given the reference corner kick sample **(A)**, we retrieve another corner kick sample **(B)** with respect to the closest distance of their representations in the latent space. We observe that **(A)** and **(B)** are both out-swing corner kicks and share similar patterns of their attacking tactics, which are highlighted with rectangles having the same colours, although they bear differences with respect to the absolute positions and velocities of the players. All the while, the latent representation of an in-swing attack **(C)** is distant from both **(A)** and **(B)** in the latent space. The red arrows are only used to demonstrate the difference between in- and out-swing corner kicks, not the actual ball trajectories.

the data. Specifically, team setups with similar tactical patterns tend to cluster together in TacticAI's latent space (Figure 2). However, no clear clusters are observed in the raw input space (S.Figure 1). This indicates that TacticAI can be leveraged as a useful corner kick retrieval system, and we will present our evaluation of this hypothesis in the Case Study section.

Controlled tactic refinement using class-conditional generative models

Equipped with components that are able to potently relate corner kicks with their various outcomes (e.g. receivers and shot events), we can explore the use of TacticAI to suggest *adjustments* of tactics, in order to amplify or reduce the likelihood of certain outcomes.

Specifically, we aim to produce adjustments to the movements of players on one of the two teams, including their positions and velocities, which would maximise or minimise the probability of a shot event, conditioned on the initial corner setup, consisting of the movements of players on both teams, and their heights and weights. In particular, although in real-world scenarios both teams may react simultaneously to the movements of each other, in our study, we focus on moderate adjustments to

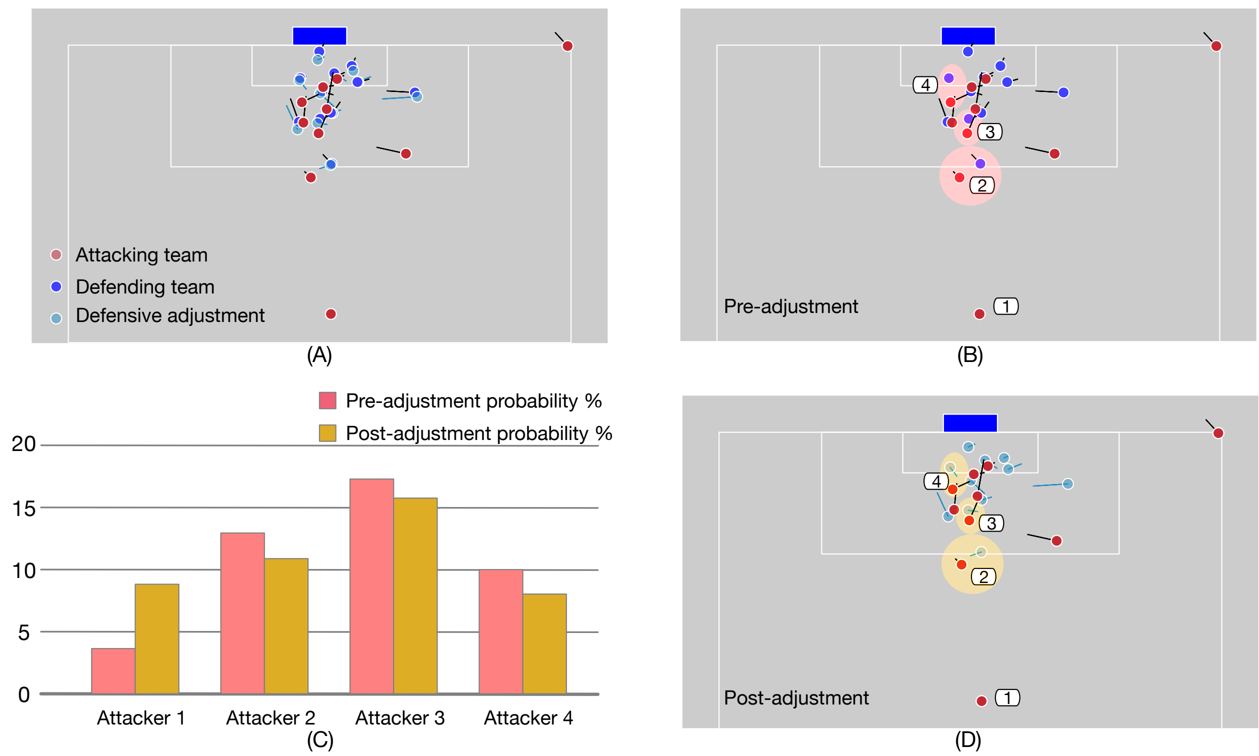


Figure 3 | **使用TacticAI细化角球战术的示例。** TacticAI使人类教练能够以识别关键球员以及提供考虑所有球员的时空协调战术建议的方式，重新设计角球战术，从而帮助攻击方或防守方最大化积极结果的可能性。正如本例所示，对于一个现实中已经尝试射门的角球，TacticAI可以通过调整防守球员的位置生成一个战术调整后的设置，从而降低射门概率。建议的防守球员位置导致攻击球员2–5的接球概率降低（见底行），而距离球门较远的攻击球员1的接球概率则有所增加。该模型能够生成多个此类场景。教练可以通过视觉检查不同的选项，并额外参考TacticAI对所展示战术的定量分析。

player movements, which help to detect players that are not responding to a tactic properly. Due to this reason, we simplify the process of tactic refinement through generating the adjustments for only one team while keeping the other fixed. The way we train a model for this task is through an *auto-encoding* objective: we feed the ground-truth shot outcome (a binary indicator) as an additional graph-level feature to TacticAI's model (Table 1), and then have it learn to reconstruct a probability distribution of the input player coordinates (Figure 1(B), also see Methods). This autoencoder-based generative model is an individual component which separates from TacticAI's predictive systems. All three systems share the encoder architecture (without sharing parameters), but use different decoders (see the Methods section). At inference time, we can instead feed in a *desired* shot outcome for the given corner setup, and then sample new positions and velocities for players on one team using this probability distribution. This setup, in principle, allows for flexible downstream use, as human coaches can optimise corner kick setups through generating adjustments conditioned on the specific outcomes of their interest—e.g., increasing shot probability for the attacking team, decreasing it for the defending team (Figure 3) or amplifying the chance that a particular striker receives the ball.

We first evaluate the generated adjustments quantitatively, by verifying how distinguishable they are from real corner kicks using a classifier. To do this, we synthesised a dataset consisting of 200 corner kick samples and their corresponding conditionally-generated adjustments. Specifically, for corners without a shot event, we generated adjustments for the attacking team by setting the shot event feature to 1, and vice-versa for the defending team when a shot event did happen. We found that the real and generated samples were not distinguishable by an MLP classifier, with an F_1 score of 0.53 ± 0.05, indicating random chance level accuracy. We also evaluated the realism of the adjustments in a qualitative case study, which we will present in the following section.

In addition, we leveraged our TacticAI shot predictor to estimate whether the proposed adjustments were salient. We did this by analysing 100 corner kick samples in which threatening shots occurred, and then, for each sample, generated one defensive refinement through setting the shot event feature to 0. We observed that the average shot probability significantly decreased, from 0.75 ± 0.14 for ground-truth corners to 0.69 ± 0.16 for adjustments ($z = 2.62, p < 0.001$). This observation was consistent when testing for attacking team refinements (shot probability increased from 0.18 ± 0.16 to 0.31 ± 0.26 ($z = -4.46, p < 0.001$)). Moving beyond this result, we also asked human raters to assess the utility of TacticAI's proposed adjustments within our case study, which we detail next.

Case study with expert raters

Although quantitative evaluation with well-defined benchmark datasets was critical for the technical development of TacticAI, the ultimate test of TacticAI as a football tactic *assistant* is its practical downstream utility being recognised by *professionals* in the industry. To this end, we evaluated TacticAI through a case study with our partners at Liverpool FC (LFC). Specifically, we invited a group

器(MLP)分类器区分，F_1得分为0.53 ± 0.05，这表明分类器的准确率仅停留在随机猜测的水平。我们还通过一个定性案例研究评估了调整策略的真实性，我们将在下一节中介绍该案例。

此外，我们还利用我们的TacticAI射门预测器来估计所提出的调整策略是否显著。我们通过分析100个发生威胁射门的角球样本，然后为每个样本通过设置射门事件特征为0生成一个防守上的改进。我们观察到，平均射门概率显著下降，从真实角球的0.75 ± 0.14降至调整策略的0.69 ± 0.16 ($z = 2.62, p < 0.001$)。当测试进攻方的改进时，这一观察结果是一致的（射门概率从0.18 ± 0.16增加到0.31 ± 0.26 ($z = -4.46, p < 0.001$)）。在这一结果之外，我们还请人类评分者评估TacticAI在案例研究中提出的调整策略的实用性，接下来将详细介绍。

与专家评分者的案例研究

虽然使用定义明确的基准数据集进行定量评估对TacticAI的技术发展至关重要，但作为足球战术*助手*的TacticAI最终要接受*业界专业人士*对其实际下游效用的认可。为此，我们通过与利物浦足球俱乐部（LFC）的合作伙伴进行案例研究来评估TacticAI。具体来说，我们邀请了一组五位足球专家：三位数据科学家，一位视频分析师和一位教练助理。他们每个人在案例研究中完成了四项任务，从多个角度评估了TacticAI组件的实用性。这些包括：1) TacticAI生成的调整策略的真实性，2) TacticAI接球预测的可信度，3) TacticAI嵌入用于检索相似角球的有效性，以及4) TacticAI推荐调整策略的有用性。我们在本文中提供了研究结果的概览，感兴趣的读者可以参考S.Figures 3, 4和5以及补充部分的案例研究设计（S1.2）以获取更多详细信息。

我们首先同时评估了TacticAI生成的调整后角球的真实性及其接球预测的可信度。在审查一个由50个角球样本组成的集合时，我们首先要求评分者判断给定的样本是真实的还是由TacticAI生成的，然后他们需要确定角球样本中最可能的接球者（S.Figure 3）。

在分类真实和生成样本的任务中，我们首先发现，评分者在分类真实与生成样本时的平均F_1得分仅为0.60 ± 0.04，个人F_1得分（$F_1^A = 0.54, F_1^B = 0.64, F_1^C = 0.65, F_1^D = 0.62, F_1^E = 0.56$），这表明在许多情况下，评分者无法将TacticAI的调整策略与真实角球区分开来。

前面的评估主要关注于分析现实感检测性能在评分者之间的表现。我们还进行了一项研究，该研究分析现实感检测在样本之间的表现。具体来说，我们为每个样本分配评分——如果样本被人类评分者认为是真实的，则赋值为+1，否则为0——并计算了五个评分者对每个样本的平均评分。重要的是，通过研究评分的分布，我们发现真实角和生成角的平均评分之间没有显著差异（$z = 0.34, p > 0.05$）（图 4(A)）。因此，人类评分者对真实和生成样本赋予的平均评分在统计上是不可区分的。

对于识别接球手这一任务，如果评分者确定的至少一个接球手出现在TacticAI的前三名预测中，我们将TacticAI对接球手的预测评分视为+1，否则为0。人类评分者的平均前三准确率为0.79 ± 0.18；具体来说，真实样本为0.82 ± 0.17，生成样本为0.77 ± 0.21。这些分数与TacticAI在预测保留测试角中的接球手的准确率非常接近，从而验证了我们的定量研究。此外，在对接球手预测样本进行平均评分后，我们发现真实样本和生成样本在预测接球手的平均评分上没有统计上的显著差异（$z = 0.968, p > 0.05$）（图 4(B)）。这表明TacticAI在预测真实角和TacticAI生成调整的接球手方面表现同

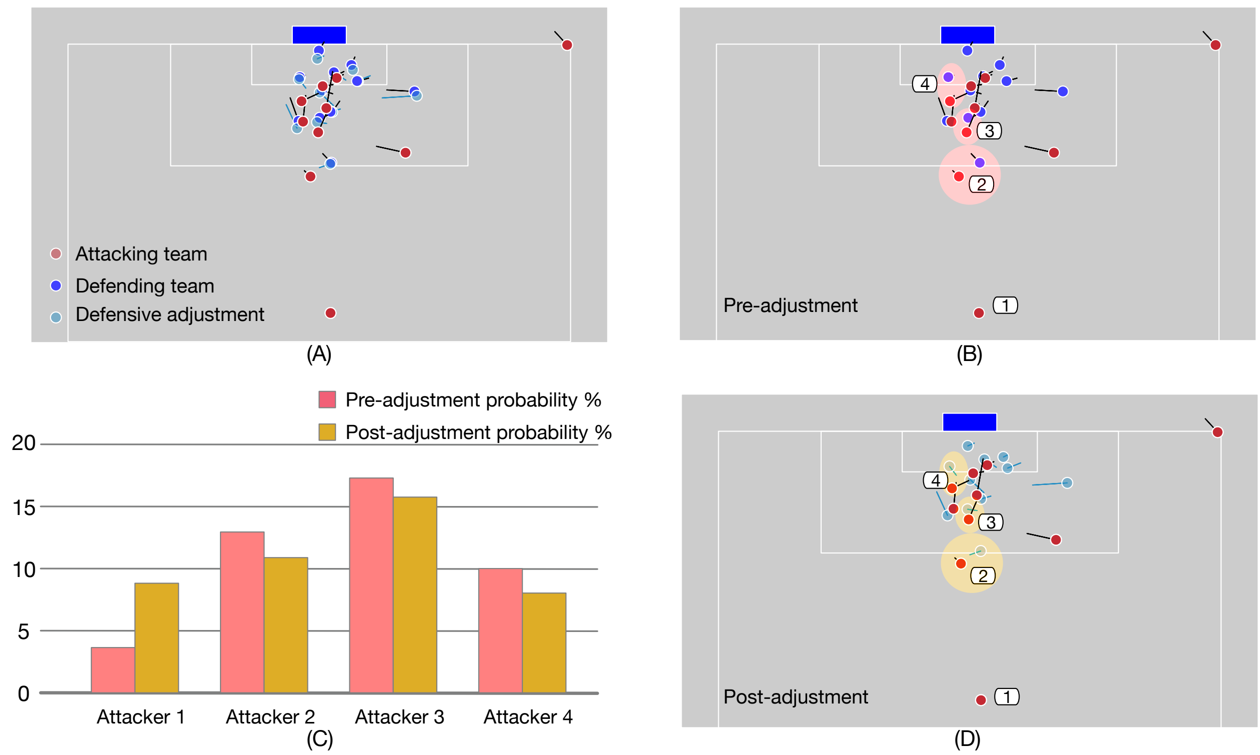


Figure 3 | **Example of refining a corner kick tactic with TacticAI.** TacticAI makes it possible for human coaches to redesign corner kick tactics in ways that help maximise the probability of a positive outcome for either the attacking or the defending team by identifying key players, as well as by providing temporally coordinated tactic recommendations which take all players into consideration. As demonstrated in this example, for a corner kick in which there was a shot attempt in reality, TacticAI can generate a tactically-adjusted setting in which the shot probability has been reduced, by adjusting the positioning of the defenders. The suggested defender positions result in reduced receiver probability for attacking players 2–5 (see bottom row), while the receiver probability of Attacker 1, who is distant from the goalpost, has been increased. The model is capable of generating multiple such scenarios. Coaches can inspect the different options visually and additionally consult TacticAI's quantitative analysis of the presented tactics.

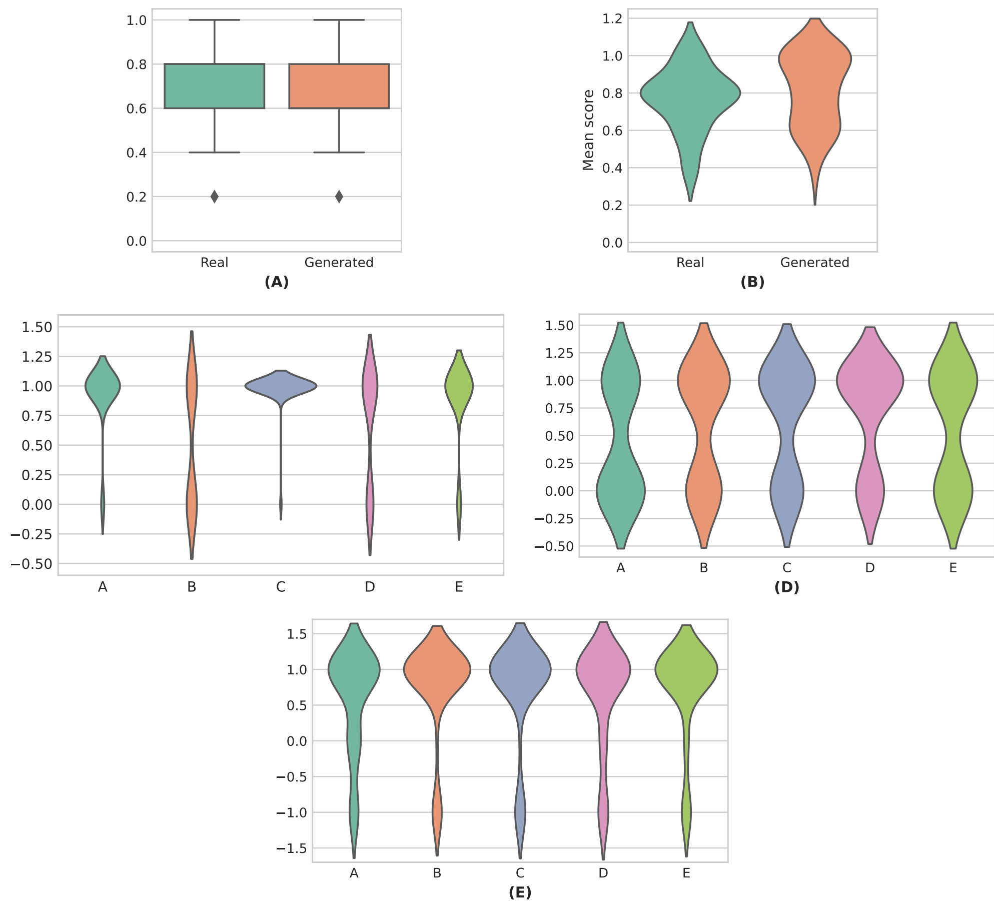


Figure 4 | **案例研究任务统计分析**。在任务1中，我们从两个方面检验了真实角球样本与TacticAI生成的合成样本之间的统计差异。（**A**）两种样本评分的分布；（**B**）使用这些样本进行接球手预测的前三名准确率分布。两种情况下均未观察到统计差异（（A）($z = 0.34, p > 0.05$)，以及（B）($z = 0.968, p > 0.05$)）。另一方面，对于任务2（接球手预测），我们观察到不同评分者之间评分的统计学显著差异，形成了三个清晰的聚类（**C**）。具体来说，评分者A和E的评分相似（$z = 1.41, p > 0.05$），评分者B和D的评分方式也类似（$z = 2.53, p > 0.05$），而评分者C与所有其他评分者的反应不同。这表明人类评分者在角球感知方面具有很好的多样性。在任务3中，关于显著战略设置下检索到的相似角球，不同评分者的评分分布没有显著差异（**D**），这表明在TacticAI检索相似角球的能力方面，大家达成了高度一致（$F_{1,4} = 1.01, p > 0.1$）。最后，在任务4中，我们比较了人类评分者对战略改进的评分（**E**），评分者对TacticAI推荐的战略改进的总体有效性也达成了共识（$F_{1,4} = 0.45, p > 0.05$）。

of five football experts: three data scientists, one video analyst and one coaching assistant. Each of them completed four tasks in the case study, which evaluated the utility of TacticAI's components from several perspectives. These include: 1) realism of TacticAI's generated adjustments, 2) plausibility of TacticAI's receiver predictions, 3) effectiveness of TacticAI's embeddings for retrieving similar corners, and 4) usefulness of TacticAI's recommended adjustments. We provide an overview of our study's results here, and refer the interested reader to S.Figures 3, 4 and 5 and the supplementary Case Study Design section (S1.2) for additional details.

We first simultaneously evaluated the realism of the adjusted corner kicks generated by TacticAI, and the plausibility of its receiver predictions. Going through a collection of 50 corner kick samples, we first asked the raters to classify whether a given sample was real or generated by TacticAI, then they were asked to identify the most likely receivers in the corner kick sample (S.Figure 3).

On the task of classifying real and generated samples, first we found that the raters' average F_1 score of classifying the real vs. generated samples was only 0.60 ± 0.04, with individual F_1 scores ($F_1^A = 0.54, F_1^B = 0.64, F_1^C = 0.65, F_1^D = 0.62, F_1^E = 0.56$), indicating that the raters were, in many situations, unable to distinguish TacticAI's adjustments from real corners.

The previous evaluation focused on analysing realism detection performance *across raters*. We also conduct a study which analyses realism detection *across samples*. Specifically, we assigned ratings for each sample—assigning +1 to a sample if it was identified as real by a human rater, and 0 otherwise—and computed the average rating for each sample across the five raters. Importantly, by studying the distribution of ratings, we found that there was no significant difference between the average ratings assigned to real and generated corners ($z = 0.34, p > 0.05$) (Figure 4(A)). Hence, the real and generated samples were assigned statistically indistinguishable average ratings by human raters.

For the task of identifying receivers, we rated TacticAI's predictions with respect to a rater as +1 if at least one of the receivers identified by the rater appeared in TacticAI's top-3 predictions, and 0 otherwise. The average top-3 accuracy among the human raters was 0.79 ± 0.18; specifically, 0.82 ± 0.17 for the real samples, and 0.77 ± 0.21 for the generated ones. These scores closely line up with the accuracy of TacticAI on predicting receivers for held-out test corners, validating our quantitative study. Further, after averaging the ratings for receiver prediction sample-wise, we found no statistically significant difference between the average ratings of predicting receivers over the real and generated samples ($z = 0.968, p > 0.05$) (Figure 4(B)). This indicates that TacticAI was equally performant in predicting the receivers of real corners and TacticAI-generated adjustments, and hence may be leveraged for this purpose even in simulated scenarios.

There is a notably high variance in the average receiver prediction rating of TacticAI. We hypothese that this is due to the fact that different raters may choose to focus on different salient features when evaluating the likely receivers (or even the *amount* of likely receivers). We set out to

TacticAI: an AI assistant for football tactics

样出色，因此甚至可以在模拟场景中为此目的所用。

TacticAI在接球手预测的平均评分上存在明显的高方差。我们假设这是因为不同的评分者可能会在评估可能的接球手（甚至可能接球手的数量）时选择关注不同的显著特征。为了验证这一假设，我们通过运行单因素方差分析（ANOVA）及后续的Tukey测试，来测试人类评分者的预测之间的成对相似性。我们发现五个评分者的预测分布存在显著差异（$F_{1,4} = 14.46, p < 0.001$），形成三个集群（图 4(C)）。这一结果表明，根据他们在LFC的各种职称，不同的评分者可能会经常使用非常不同的线索来建议可能的接球手。TacticAI在这样一个环境下仍能保持高的前三准确率，这表明它能够捕捉到角球策略的显著模式，这些模式与人类评分者的偏好广泛一致。我们将在第三项任务——识别相似角球中进一步测试这一假设。对于第三个任务，我们要求人类评分员判断50对角球的相似性。每对都是由一个参考角球和一个检索角球组成，其中检索角球是根据它们在TacticAI潜在空间表示中的最近邻选择的，或者作为特征级别的启发式，即它们在角球数据集中的原始特征的余弦相似性（S.图 4）。如果评分员认为案例中呈现的角球有用地相似，则对这对的评分记为+1，否则这对的得分记为0。我们首先计算了每位评分员回忆起他们认为基线或TacticAI检索的对为有用相似的情况的召回率（参见补充材料中的案例研究设计部分（S1.2）的任务3）。对于TacticAI检索，所有评分员的平均召回率为0.63 ± 0.09，对于基线系统，召回率为0.33 ± 0.10。其次，我们通过平均每个参考检索对的评分来评估两种方法的统计差异，发现TacticAI检索的平均评分显著高于基线方法检索的平均评分（$z = 2.34, p < 0.05$）。这两个结果表明，TacticAI在挖掘相似角球方面显著优于特征空间基线方法。这表明TacticAI能够从角球中提取出输入数据本身难以提取出的显著特征，强化了它作为从可用数据中发现对手球队战术的有效工具。最后，我们观察到这项任务在TacticAI检索对上展示了很高的评分员间一致性（$F_{1,4} = 1.01, p > 0.1$）（图 4(D)），这表明人类评分员在评估TacticAI性能方面基本达成一致。

最后，我们评估了TacticAI对球员调整建议的实际效用。具体来说，每位评分员被提供了50项战术改进及其相应的真实角球设置—参见S.图 5，以及补充材料中的案例研究设计部分 S1.2。然后要求评分员对每项改进进行评分，认为显著改善战术的记为+1，显著变糟的记为−1，或者没有显著差异的记为0。我们计算了每位评分员分配的平均评分（这为我们提供了每位评分员在$[-1, 1]$范围内的一个值）。所有五位评分员的这些值的平均为0.7 ± 0.1。此外，对于50个情况中的45个（90%），人类评分员平均认为TacticAI的建议是更有利的（通过多数投票）。这两个结果都表明，TacticAI的建议对于下游足球俱乐部从业者来说是显著且有用的，我们通过统计测试来验证这一点。我们对观察到的正面评分进行了统计显著性检验。首先，对于每一种50种情况，我们计算了五个评分者对其的平均评分，然后进行t检验，以评估平均评分是否显著大于零。实际上，统计检验表明，TacticAI推荐的战术调整总体上是建设性的（$t_{49}^{\text{avg}} = 9.20, p < 0.001$）。其次，我们验证了五位评分者中的每一位都认为TacticAI的建议是建设性的，我们对他们的各自评分分别进行了t检验。对于所有五位评分者，他们的平均评分被发现在统计上显著高于零（$t_{49}^A = 5.84, p^A < 0.001; t_{49}^B = 7.30, p^B < 0.001; t_{49}^C = 7.00, p^C < 0.001; t_{49}^D = 6.04, p^D < 0.001; t_{49}^E = 7.88, p^E < 0.001$）。此外，他们的评分之间也具有较高的一致性（$F_{1,4} = 0.45, p > 0.05$）（图 4(E)），这表明人类专家普遍认可其实用性，尽管他们来自不同的背景。

将所有这些结果综合在一起，我们发现TacticAI在角球预测、检索和战术调整方面具有强大的

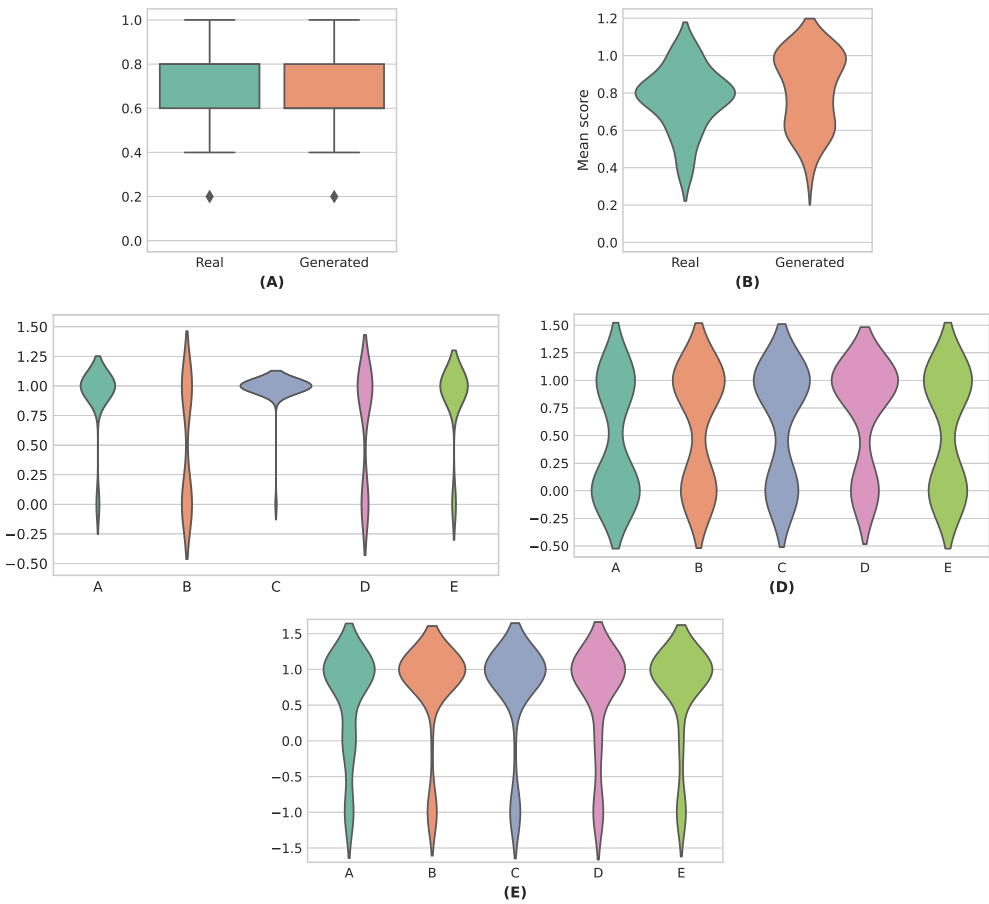


Figure 4 | **Statistical analysis for the case study tasks**. In task 1, we tested the statistical difference between the real corner kick samples and the synthetic ones generated by TacticAI from two aspects. (**A**) the distributions of the ratings of the two types of samples, and (**B**) the distributions of the top-3 accuracy of receiver prediction using those samples. No statistical difference was observed in either cases ((A)($z = 0.34, p > 0.05$), and (B)($z = 0.968, p > 0.05$)). On the other hand, for task 2 (receiver prediction), we observed a statistically significant difference between the ratings of different raters, with three clear clusters emerging (**C**). Specifically, Raters A and E had similar ratings ($z = 1.41, p > 0.05$), and Raters B and D also rated in similar ways ($z = 2.53, p > 0.05$), while Rater C responded differently from all other raters. This suggests a good level of variety of the human raters with respect to their perceptions of corner kicks. In task 3 of identifying similar corners retrieved in terms of salient strategic setups, there was no significant difference among the distributions of the ratings by different raters (**D**), suggesting a high level of agreement on the usefulness of TacticAI's capability of retrieving similar corners ($F_{1,4} = 1.01, p > 0.1$). Finally, in task 4, we compared the ratings of the strategic refinements by the human raters (**E**), and the raters also agreed on the general effectiveness of the refinements recommended by TacticAI ($F_{1,4} = 0.45, p > 0.05$).

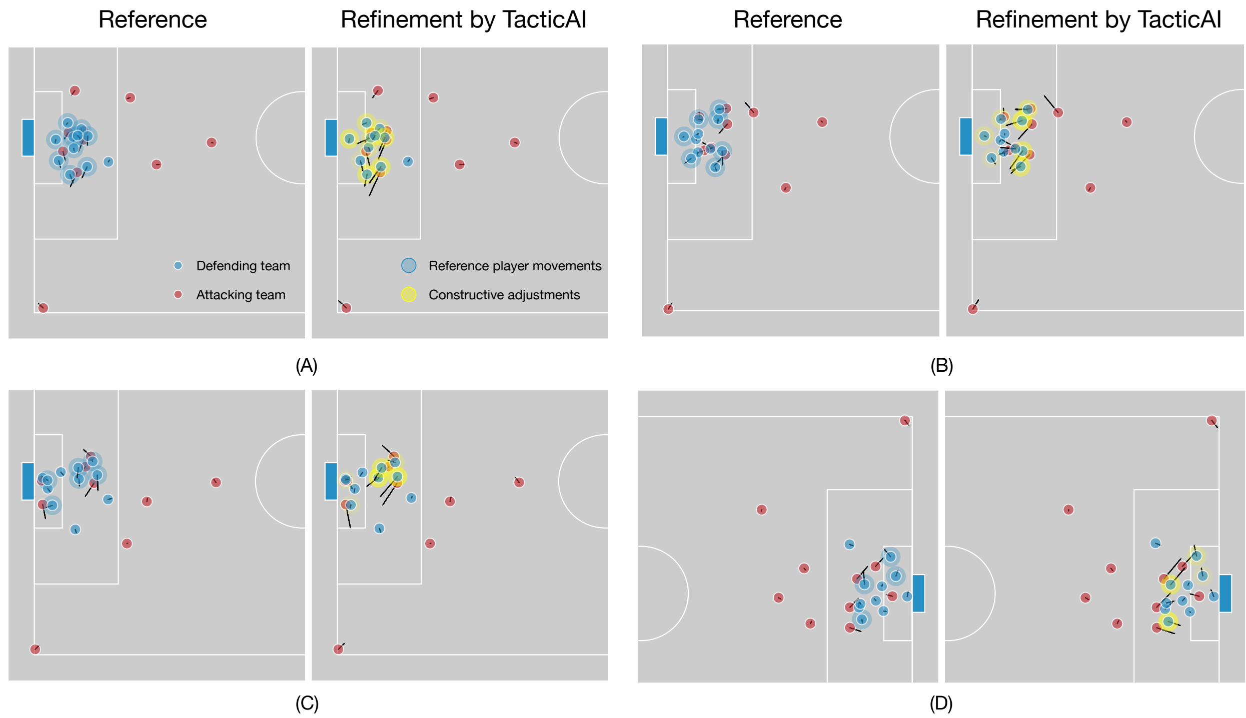


Figure 5 | **TacticAI推荐的战略改进示例**。这些示例是从我们与人类专家的案例研究中选取的，用以展示TacticAI建议给防守角球的球队做出的战术调整的广泛性。黄色圆点的密度与人类专家认为相应的改变是有建设性的次数相吻合。TacticAI不是优化某一个特定球员的移动，而是可以在一步生成中推荐多个球员的改进，通过建议更好的位置来阻挡对方球员，或者更好的方向来更高效地追踪他们。以下是一些专家评分者的具体评论。在(**A**)中，根据评分者的意见，TacticAI为几名后卫提出了更有利的站位，并为其他几名球员改善了追踪跑位—此外，守门员的位置更深，这也是有益的。在(**B**)中，TacticAI建议离角球最远的后卫进行更有效的协防跑位，这被一致认为是有用的，还有几名其他后卫的站位也更合适。在(**C**)中，TacticAI推荐了一组位于禁区中部的后卫进行改进的协防跑位，这被我们的评分者一致认为是显著的。而在(**D**)中，TacticAI为两名中后卫建议了实质上更好的追踪跑位，并为另外两名在球门区的后卫提出了更好的站位。

validate this hypothesis by testing the pair-wise similarity of the predictions by the human raters through running a one-away analysis of variance (ANOVA), followed by a Tukey's test. We found that the distributions of the five raters' predictions were significantly different ($F_{1,4} = 14.46$, $p < 0.001$) forming three clusters (Figure 4(C)). This result indicates that different human raters—as suggested by their various titles at LFC—may often use very different leads when suggesting plausible receivers. The fact that TacticAI manages to retain a high top-3 accuracy in such a setting suggests that it was able to capture the salient patterns of corner kick strategies, which broadly align with human raters' preferences. We will further test this hypothesis in the third task—of identifying similar corners.

For the third task, we asked the human raters to judge 50 pairs of corners for their similarity. Each pair consisted of a reference corner and a retrieved corner, where the retrieved corner was chosen either as the nearest-neighbour of the reference in terms of their TacticAI latent space representations, or—as a feature-level heuristic—the cosine similarities of their raw features (S.Figure 4) in our corner kick dataset. We score the raters' judgement of a pair as +1 if they considered the corners presented in the case to be usefully similar, otherwise the pair is scored with 0. We first computed, for each rater, the recall with which they have judged a baseline- or TacticAI-retrieved pair as usefully similar (see Task 3 of the Case Study Design section (S1.2) in the Supplementary Materials). For TacticAI retrievals, the average recall across all raters was 0.63 ± 0.09, and for the baseline system, the recall was 0.33 ± 0.10. Secondly, we assess the statistical difference between the results of the two methods by averaging the ratings for each reference-retrieval pair, finding that the average rating of TacticAI retrievals is significantly higher than the average rating of baseline method retrievals ($z = 2.34$, $p < 0.05$). These two results suggest that TacticAI significantly outperforms the feature-space baseline as a method for mining similar corners. This indicates that TacticAI is able to extract salient features from corners which are not trivial to extract from the input data alone, reinforcing it as a potent tool for discovering opposing team tactics from available data. Finally, we observed that this task exhibited a high level of inter-rater agreement for TacticAI-retrieved pairs ($F_{1,4} = 1.01$, $p > 0.1$) (Figure 4(D)), suggesting that human raters were largely in agreement with respect to their assessment of TacticAI's performance.

Finally, we evaluated TacticAI's player adjustment recommendations for their practical utility. Specifically, each rater was given 50 tactical refinements together with the corresponding real corner kick setups—see S.Figure 5, and the Case Study Design section in the Supplementary Materials S1.2. The raters were then asked to rate each refinement as saliently improving the tactics (+1), saliently making them worse (−1), or offering no salient differences (0). We calculated the average rating assigned by each of the raters (giving us a value in the range $[-1, 1]$ for each rater). The average of these values across all five raters was 0.7 ± 0.1. Further, for 45 of the 50 situations (90%), the human raters found TacticAI's suggestion to be favourable on average (by majority voting). Both of these results indicate that TacticAI's recommendations are salient and useful to a downstream football club practitioner, and we set out to validate this with statistical tests.

We performed statistical significance testing of the observed positive ratings. First, for each of

TacticAI: an AI assistant for football tactics

组成部分。为了说明TacticAI所提供的显著推荐类型，在图5中，我们展示了四个具有高度评分者间一致性的示例。

Discussion and Conclusions

我们展示了一个人工智能足球战术助手，并通过与利物浦足球俱乐部的专家人类评分员进行的全面案例研究，提供了其有效性的统计证据。首先，TacticAI能够准确预测角球开出后的首个接球手以及角球直接导致射门的概率。其次，TacticAI已经显示出能够生成可信的战术变化，显著改善比赛结果，同时这些变化在领域专家看来与真实场景无异。最后，系统的潜在球员表示是检索类似定位球战术的强大手段，使教练能够分析过去成功的相关战术和反战术。

足球战略建模的更广泛范围之前已从各种个别角度进行探讨，例如传球预测 [18, 19, 30]，射门预测 [16]或角球战术分类 [33]。然而，据我们所知，我们的工作通过结合并评估角球的*预测和生成*建模来进行战术发展，从而脱颖而出。它还以其应用几何深度学习的方法而突出，这允许有效地融入足球场的各种对称性，以提高数据效率。我们的方法融入了最小的领域知识，并且不依赖于复杂的特征工程——尽管其分解设计自然允许在特征可用时采用更复杂的特征工程方法。

我们的方法论需要执行角球时所有球员的位置和速度估计以及随后的事件。这里，我们从高质量的追踪和事件数据中获取这些信息，追踪提供商的数据可用性限于顶级联赛。基于广播视频的球员追踪将大大增加数据的覆盖范围和训练数据，但也可能导致模型输入出现更多噪声。虽然图注意力机制（GATs）可以使我们执行对模型结果贡献最显著因素的内部检查，但我们的方法并未明确建模外部（偶然性）不确定性，这对于足球分析师来说将是宝贵的上下文。

虽然我们方法的实证研究主要集中在足球中的角球上，但它很容易泛化到其他定位球（如界外球，这也同样受益于相似性检索、传球和/或射门预测）和其他具有暂停比赛情况的团队运动。TacticAI的学习表示和整体架构还为未来研究奠定了基础，以整合一个自然语言界面，使助手能够进行基于领域的对话，以检索感兴趣的具体情况，对给定的战术变体进行预测，进行比较和对比，并通过交互式过程指导战术建议的提取。因此，我们相信TacticAI为下一代足球人工智能助手奠定了基础。

Methods

我们设计了一套名为TacticAI的几何深度学习流程，本节将进一步展开介绍。我们将标记的足球时空数据转换为图表示，并在作为分类或回归的基准任务上进行训练和评估。这些步骤将按顺序呈现，随后是所采用的计算架构的细节。

原始角球数据 原始数据集由2020–2021英超赛季收集的9,693次角球组成，由利物浦足球俱乐部提供。这个数据集由四个独立的数据源组成：

- 空间-时间轨迹框架（“追踪数据”），追踪每场比赛中所有场上球员和球的位置，每秒25帧。

the 50 situations, we averaged its ratings across all five raters, and then ran a t-test to assess whether the mean rating was significantly larger than zero. Indeed, the statistical test indicated that the tactical adjustments recommended by TacticAI were constructive overall ($t_{49}^{\text{avg}} = 9.20, p < 0.001$). Secondly, we verified that each of the five raters individually found TacticAI's recommendations to be constructive, running a t-test on each of their ratings individually. For all of the five raters, their average ratings were found to be above zero with statistical significance ($t_{49}^{A} = 5.84, p^{A} < 0.001; t_{49}^{B} = 7.30, p^{B} < 0.001; t_{49}^{C} = 7.00, p^{C} < 0.001; t_{49}^{D} = 6.04, p^{D} < 0.001; t_{49}^{E} = 7.88, p^{E} < 0.001$). In addition, their ratings also shared a high-level of inter-agreement ($F_{1,4} = 0.45, p > 0.05$) (Figure 4(E)), suggesting a level of practical usefulness that is generally recognised by human experts, even though they represent different backgrounds.

Taking all of these results together, we find TacticAI to possess strong components for prediction, retrieval, and tactical adjustments on corner kicks. To illustrate the kinds of salient recommendations by TacticAI, in Figure 5 we present four examples with high degree of inter-rater agreement.

Discussion and Conclusions

We have demonstrated an AI assistant for football tactics, and provided statistical evidence of its efficacy through a comprehensive case study with expert human raters from Liverpool FC. First, TacticAI is able to accurately predict the first receiver after a corner kick is taken as well as the probability of a shot as the direct result of the corner. Second, TacticAI has been shown to produce plausible tactical variations that improve outcomes in a salient way, while being indistinguishable from real scenarios by domain experts. And finally, the system's latent player representations are a powerful means to retrieve similar set piece tactics, allowing coaches to analyse relevant tactics and counter-tactics that have been successful in the past.

The broader scope of strategy modelling in football has previously been addressed from various individual angles, such as pass prediction [18, 19, 30], shot prediction [16] or corner kick tactical classification [33]. However, to the best of our knowledge, our work stands out by combining and evaluating *predictive and generative modelling* of corner kicks for tactic development. It also stands out in its method of applying geometric deep learning, allowing for efficiently incorporating various symmetries of the football pitch for improved data efficiency. Our method incorporates minimal domain knowledge, and does not rely on intricate feature engineering—though its factorised design naturally allows for more intricate feature engineering approaches, when such features are available.

Our methodology requires the position and velocity estimates of all players at the time of execution of the corner and subsequent events. Here, we derive these from high-quality tracking and event data, with data availability from tracking providers limited to top leagues. Player tracking based on broadcast video would increase the reach and training data substantially, but would also likely result in noisier model inputs. While the attention mechanism of GATs would allow us to perform

TacticAI: an AI assistant for football tactics

除了球员位置信息，通过滤波处理位置数据还可以得到他们的速度。对于每个角球，我们仅使用踢球的那个帧作为输入信息。

- 事件流数据，它标注了在对应追踪帧中发生的事件或动作（例如，传球、射门和进球）。
- 对应比赛的阵容数据，记录了球员的个人信息，包括他们的身高、体重和角色。
- 游戏数据，其中包含游戏天数、体育场信息以及场地长度和宽度（以米为单位）。

图形表示与构建 我们假设我们获得了一个输入图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$，其中包含一组节点 $\mathcal{V}$ 和边 $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$。在足球比赛的背景下，我们将 $\mathcal{V}$ 视为当前在场的二十二名球员的集合，对于两个队伍都是如此，并且我们设置 $\mathcal{E} = \mathcal{V} \times \mathcal{V}$；即，我们假设所有球员对都有互动的潜力。进一步的分析，利用更具体的 $\mathcal{E}$ 选择，将是未来工作中的一个有趣的研究方向。

此外，我们假设图已经被适当地特征化。具体来说，我们提供了一个节点特征矩阵，$\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times k}$，一个边特征张量，$\mathbf{E} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}| \times l}$，以及一个图特征向量，$\mathbf{g} \in \mathbb{R}^{m}$。这些对象的适当条目为我们提供了每个节点、边和图的输入特征。例如，$\mathbf{x}_u \in \mathbb{R}^k$ 将提供个体球员 $u \in V$ 的属性，如位置、身高和体重，而 $\mathbf{e}_{uv} \in \mathbb{R}^l$ 将提供特定球员对 $(u, v) \in \mathcal{E}$ 的属性，如他们的距离，以及他们是否属于同一队伍。图特征向量 $\mathbf{g}$ 可用于存储对角球感兴趣的全球属性，如比赛时间、当前比分或球的位置。关于图神经网络如何处理此类输入的简化可视化，请参考图 1(A)。

为了构建输入图，我们首先将四个数据源根据其比赛ID和时间戳对齐，并过滤掉 2,517 个无效角球，由于缺失数据导致对齐失败，例如，缺失跟踪帧或事件标签。此过滤产生了 7,176 个用于训练和评估的有效角球。我们在表 2 中总结了用于构建输入图的确切信息。特别是，除了球员身高（以厘米（cm）为单位）和体重（以千克（kg）为单位）之外，模型中的球员是匿名的。在缺少球员资料的情况下，我们将他们的身高和体重分别设置为180cm和75kg作为默认值。在数据预处理过程中，共有385次此类情况，总计 $213,246\ (= 22 \times 9,693)$。我们将身高和体重缩小了100倍。此外，对于每个角球，我们将场上球员的位置零中心，并在 $10\text{m} \times 10\text{m}$ 的场地上进行归一化，相应地调整他们的速度。在缺少场地尺寸的情况下，我们使用标准的场地尺寸 $110\text{m} \times 63\text{m}$ 作为默认值。

我们在表 1 中总结了特征分组。不同基准任务中实际使用的特征可能有所不同，我们将在下一节中详细介绍。为了专注于建模进攻和防守队伍的高级战术，除了一个球权二进制指标外，没有使用球的移动信息，无论是位置还是速度，来构建输入图。

基准任务构建 TacticAI 包括三个预测和生成模型，这也对应于本研究中实现的三个基准任务。具体来说，1) 接球手预测，2) 有威胁射门预测，以及3) 团队位置和速度引导生成（表 1）。所有基准任务的图使用了相同的节点和边的特征空间，只在全局特征上有所不同。对于所有三个任务，我们的模型首先将节点特征转换为一个潜在节点特征矩阵，$\mathbf{H} = f_{\mathcal{G}}(\mathbf{X}, \mathbf{E}, \mathbf{g})$，从中我们可以回答关于个体球员的问题——在这种情况下，我们学会了在$\mathbf{h}_u$向量（$\mathbf{H}$的行）上相关的分类器或回归器；或者关于全局事件（例如，射门）发生的问题——在这种情况下，我们在聚合的球员向量$\sum_u \mathbf{h}_u$上进行分类或回归。在这两种情况下，分类器都是通过在适当选择的损失函数上使用随机梯度下降进行训练的，例如分类交叉熵用于分类器，均方误差用于回归器。

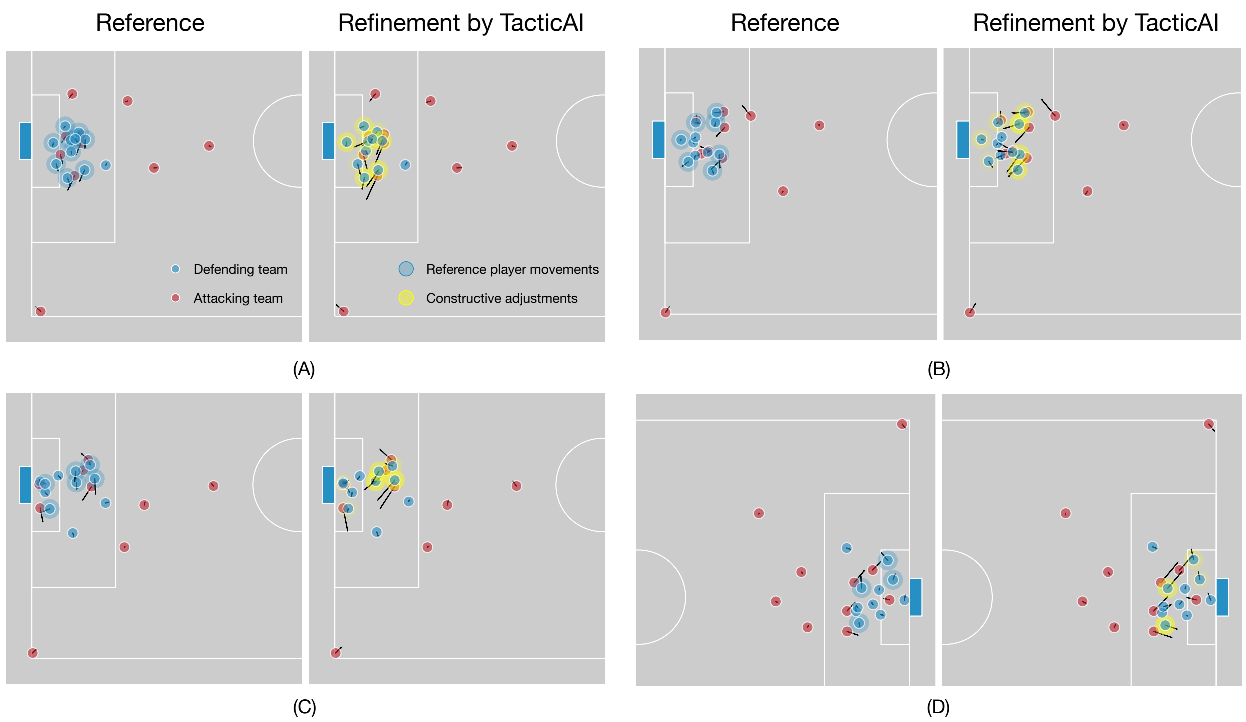


Figure 5 | **Examples of the tactic refinements recommended by TacticAI.** These examples are selected from our case study with human experts, to illustrate the breadth of tactical adjustments that TacticAI suggests to teams defending a corner. The density of the yellow circles coincides with the number of times that the corresponding change is recognised as constructive by the human experts. Instead of optimising the movement of one specific player, TacticAI can recommend the improvements for multiple players in one generation step through suggesting better positions to block the opposing players, or better orientations to track them more efficiently. Some specific comments from expert raters follow. In (**A**), according to raters, TacticAI suggests more favourable positions for several defenders, and improved tracking runs for several others—further, the goalkeeper is positioned more deeply, which is also beneficial. In (**B**), TacticAI suggests that the defenders furthest away from the corner make improved covering runs, which was unanimously deemed useful, with several other defenders also positioned more favourably. In (**C**), TacticAI recommends improved covering runs for a central group of defenders in the penalty box, which was unanimously considered salient by our raters. And in (**D**), TacticAI suggests substantially better tracking runs for two central defenders, along with a better positioning for two other defenders in the goal area.

对于不同的任务，我们从事件流数据或追踪数据中提取相应的真实标签。具体来说，1) 我们将接球预测建模为节点分类任务，并将角球开出后第一个触球的球员标记为目标节点。这名球员可能是进攻或防守球员。2) 射门预测被建模为图分类。特别是，如果进攻队的下一次触球动作是直接角球、进球、空中球、击中门柱、射门被守门员扑救或未命中目标，我们就将其视为射门。这产生了1,736个标记为射门被采取的角球，以及5,440个标记为未采取射门的角球。3) 对于指导生成的球员位置和速度，不需要额外的标签，因为该模型依赖于自我监督的重构目标。

整个数据集通过随机抽样以80 : 20的比例分为训练和评估集，并且相同的分割用于所有任务。

图神经网络 TacticAI的核心模型是图神经网络(GNN) [41]，它通过在节点的邻域内反复组合来在图上计算潜在表示。在这里，我们将节点u的邻域$\mathcal{N}_u$定义为所有一阶邻居的集合，即$\mathcal{N}_u = \{v \mid (v,u) \in \mathcal{E}\}$。然后，单个GNN层通过在相邻节点之间传递消息来转换节点特征 [13]，遵循相关工作 [5]的记法，以及CLRS-30基准基线 [42]的实现：

$$\mathbf{h}_u^{(t)} = \phi\left(\mathbf{h}_u^{(t-1)}, \bigoplus_{v\in\mathcal{N}_u} \psi\left(\mathbf{h}_u^{(t-1)}, \mathbf{h}_v^{(t-1)}, \mathbf{e}_{vu}, \mathbf{g}\right)\right) \tag{2}$$

其中 $\psi : \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^{k'}$ 和 $\phi : \mathbb{R}^k \times \mathbb{R}^{k'} \to \mathbb{R}^{k'}$ 是两个可学习函数（例如多层感知机），$\mathbf{h}_u^{(t)}$ 是节点 u 经过 t 层 GNN 后的特征，$\bigoplus$ 是任意置换不变的聚合器，如求和、最大值或平均值。按定义，我们设置 $\mathbf{h}_u^{(0)} = \mathbf{x}_u$，并将方程 2 迭代 T 步，其中 T 是一个超参数。然后，我们令 $\mathbf{H} = f_{\mathcal{G}}(\mathbf{X}, \mathbf{E}, \mathbf{g}) = \mathbf{H}^{(T)}$ 成为 GNN 最终输出的节点嵌入。

众所周知，方程 2 非常通用：它可以用来表达如 Transformers [39] 等流行模型作为一个特例，并且据称 *所有* 离散深度学习模型都可以用这种形式表达 [40, 3]。这使得 GNN 成为了比较各种建模足球比赛中玩家间互动方法的一个完美框架。

不同的 ψ，ϕ 和 $\bigoplus$ 选择会产生不同的架构。在我们的案例中，我们使用一个可以分解为 *注意力机制* 的消息函数，$a : \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$：

$$\mathbf{h}_u^{(t)} = \phi\left(\mathbf{h}_u^{(t-1)}, \bigoplus_{v\in\mathcal{N}_u} a\left(\mathbf{h}_u^{(t-1)}, \mathbf{h}_v^{(t-1)}, \mathbf{e}_{vu}, \mathbf{g}\right) \psi\left(\mathbf{h}_v^{(t-1)}\right)\right) \tag{3}$$

产生了图注意力网络（GAT）架构 [43]。在我们的工作中，特别地，我们采用了GATv2 [4]所提出的一个两层多层感知机来进行注意力机制：

$$a\left(\mathbf{h}_u^{(t-1)}, \mathbf{h}_v^{(t-1)}, \mathbf{e}_{vu}, \mathbf{g}\right) = \underset{v\in\mathcal{N}_u}{\mathrm{softmax}}\ \mathbf{a}^\top \mathrm{LeakyReLU}\left(\mathbf{W}_1\mathbf{h}_u^{(t-1)} + \mathbf{W}_2\mathbf{h}_v^{(t-1)} + \mathbf{W}_e\mathbf{e}_{vu} + \mathbf{W}_g\mathbf{g}\right) \tag{4}$$

其中 $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{k\times h}$，$\mathbf{W}_e \in \mathbb{R}^{l\times h}$，$\mathbf{W}_g \in \mathbb{R}^{m\times h}$ 和 $\mathbf{a} \in \mathbb{R}^h$ 是注意力机制的可学习参数，而 LeakyReLU 是泄漏修正线性激活函数。这种机制计算每对连接节点 (u, v) 的交互系数（单个标量值），然后使用 softmax 函数对所有 u 的邻居进行归一化。

通过早期的实验，我们已经确定 GAT 能够匹配更通用选择的 ψ（如 MPNN [13]）的性能，同时具有更好的可扩展性。因此，在本研究中，我们将重点放在 GAT 模型上。更多细节可以在消融研究部分找到。

introspection of the most salient factors contributing to the model outcome, our method does not explicitly model exogenous (aleatoric) uncertainty, which would be valuable context for the football analyst.

While the empirical study of our method's efficacy has been focused on corner kicks in association football, it readily generalises to other set pieces (such as throw-ins, which similarly benefit from similarity retrieval, pass and/or shot prediction) and other team sports with suspended play situations. The learned representations and overall framing of TacticAI also lays the ground for future research to integrate a natural language interface that enables domain-grounded conversations with the assistant, with the aim to retrieve particular situations of interest, make predictions for a given tactical variant, compare and contrast, and guide through an interactive process to derive tactical suggestions. It is thus our belief that TacticAI lays the groundwork for the next-generation AI assistant for football.

Methods

We devised TacticAI as a geometric deep learning pipeline, further expanded in this section. We process labelled spatio-temporal football data into graph representations, and train and evaluate on benchmarking tasks cast as classification or regression. These steps are presented in sequence, followed by details on the employed computational architecture.

Raw corner kick data The raw dataset consisted of $9,693$ corner kicks collected from 2020–2021 Premier League seasons, which was provided by Liverpool FC. This dataset is comprised of four separate data sources:

- Spatio-temporal trajectory frames ("tracking data"), which tracked all on-pitch players and the ball, for each match, at 25 frames per second. In addition to player positions, their velocities are derived from position data through filtering. For each corner kick, we only used the frame in which the kick is being taken as input information.
- Event stream data, which annotated the events or actions (e.g., passes, shots and goals) that have occurred in the corresponding tracking frames.
- Line-up data for the corresponding games, which recorded the players' profiles, including their heights, weights and roles.
- Game data, which contained the game days, stadium information, and pitch length and width in meters.

Graph representation and construction We assumed that we were provided with an input graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a set of *nodes* $\mathcal{V}$ and *edges* $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Within the context of football games, we took $\mathcal{V}$ to be the set of twenty-two players currently on the pitch for both teams, and we set $\mathcal{E} = \mathcal{V} \times \mathcal{V}$;

几何深度学习 尽管方程 2 具有强大的能力，但在其完全通用性下使用往往容易过拟合，考虑到 ψ 和 ϕ 中包含的大量参数。在足球分析领域，这个问题尤为严重，因为黄金标准数据通常非常稀缺——例如，在英格兰超级联赛中，每个赛季只进行几百场比赛。

为了解决这个问题，我们可以利用足球比赛产生的数据的巨大*规律性*。策略等效的游戏状态也称为*转置*，自20世纪60年代以来，计算上已经利用了对称性，如通过不同的移动序列达到相同的棋盘位置 [17]。类似地，游戏旋转和平移可能会产生等效的战略情况 [32]。使用*几何深度学习*（**GDL**） [5] 的蓝图，我们可以设计专门的 GNN 架构，利用这种规律性。

也就是说，几何深度学习是一种在神经网络中导出数学约束的一般性方法，使得当输入以某些方式变换时，它们的行为可以预测。在许多重要的情况下，这些约束可以直接解决，从而直接指导神经网络架构设计。关于在3D旋转对称下的点云的全面示例，请参见 Fuchs 等人的工作 [12]。

为了在高级别阐明 GDL 框架的几个方面，让我们假设存在一个输入数据变换的群组（*对称性*），$\mathfrak{G}$，在这些变换下真实标签保持不变。具体来说，如果我们让 $y(\mathbf{X}, \mathbf{E}, \mathbf{g})$ 是用 $\mathbf{X}, \mathbf{E}, \mathbf{g}$ 进行图特征化的标签，那么对于 $\mathfrak{G}$ 中的每个变换 $\mathfrak{g}$，以下性质成立：

$$y(\mathfrak{g}(\mathbf{X}), \mathfrak{g}(\mathbf{E}), \mathfrak{g}(\mathbf{g})) = y(\mathbf{X}, \mathbf{E}, \mathbf{g}) \tag{5}$$

这种条件也被称为$\mathfrak{G}$-不变性。

值得注意的是，如果我们设置对称群$\mathfrak{G}$为$S_{|V|}$，即$|V|$个对象的置换群，那么也可以从图扩散视角推导出GNNs。由于方程2的设计，其输出将不依赖于输入图中节点的确切排列。

帧平均 对于任意预测器$f_{\mathcal{G}}(\mathbf{X}, \mathbf{E}, \mathbf{g})$来说，实现$\mathfrak{G}$-不变性的一种简单机制是在所有$\mathfrak{G}$-变换后的输入上进行*帧平均*：

$$f_{\mathcal{G}}^{\text{inv}}(\mathbf{X}, \mathbf{E}, \mathbf{g}) = \frac{1}{|\mathfrak{G}|} \sum_{\mathfrak{g} \in \mathfrak{G}} f_{\mathcal{G}}(\mathfrak{g}(\mathbf{X}), \mathfrak{g}(\mathbf{E}), \mathfrak{g}(\mathbf{g})) \tag{6}$$

这确保了特定输入的所有$\mathfrak{G}$-变换版本（也称为该输入的*轨道*）将具有完全相同的输出，满足方程5。这种方法的变体也已被应用于AlphaGo架构中 [36]，以编码围棋棋盘的对称性。

在我们的具体实现中，我们设置$\mathfrak{G} = D_2 = \{\text{id}, \leftrightarrow, \updownarrow, \leftrightarrow\updownarrow\}$，即二面体群。利用$D_2$的不变性可以编码*象限对称性*。$D_2$群中的每个元素都编码了输入足球场垂直或水平反射的存在。在这些变换下，假设场地完全对称，因此可以安全地假设许多预测（如哪个球员接角球或从角球射门）保持不变。例如，在方程6中计算变换特征，$\leftrightarrow$ $(\mathbf{X})$水平反射$\mathbf{X}$中所有球员的位置特征（例如，球员的坐标），并取其速度的x轴分量的负值。

群卷积 虽然方程6中的帧平均方法是一种强大的方式，可以将GNN限制为尊重输入对称性，但有人认为它失去了不同$\mathfrak{G}$-变换视图在计算过程中相互交互的机会。对于像D_2这样的小群，可以假设一种更细粒度的方法，在方程2的单个GNN层上操作，我们很快将其写作$\mathbf{H}^{(t)} = g_{\mathcal{G}}(\mathbf{H}^{(t-1)}, \mathbf{E}, \mathbf{g})$。我们需要一个尊重对称性的GNN层满足的条件如下，对于所有变换$\mathfrak{g} \in \mathfrak{G}$：

$$g_{\mathcal{G}}(\mathfrak{g}(\mathbf{H}^{(t-1)}), \mathfrak{g}(\mathbf{E}), \mathfrak{g}(\mathbf{g})) = \mathfrak{g}(g_{\mathcal{G}}(\mathbf{H}^{(t-1)}, \mathbf{E}, \mathbf{g})) \tag{7}$$

that is, we assumed all pairs of players have the potential to interact. Further analyses, leveraging more specific choices of $\mathcal{E}$, would be an interesting avenue for future work.

Additionally, we assume that the graph is appropriately featurised. Specifically, we provide a *node feature matrix*, $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}|\times k}$, an *edge feature tensor*, $\mathbf{E} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|\times l}$, and a *graph feature vector*, $\mathbf{g} \in \mathbb{R}^{m}$. The appropriate entries of these objects provide us with the input features for each node, edge, and the graph. For example, $\mathbf{x}_u \in \mathbb{R}^k$ would provide attributes of an individual player $u \in V$, such as position, height and weight, and $\mathbf{e}_{uv} \in \mathbb{R}^l$ would provide the attributes of a particular pair of players $(u, v) \in \mathcal{E}$, such as their distance, and whether they belong to the same team. The graph feature vector, $\mathbf{g}$, can be used to store global attributes of interest to the corner kick, such as the game time, current score, or ball position. For a simplified visualisation of how a graph neural network would process such an input, refer to Figure 1(A).

To construct the input graphs, we first aligned the four data sources with respect to their game IDs and timestamps and filtered out 2,517 invalid corner kicks, for which the alignment failed due to missing data, e.g., missing tracking frames or event labels. This filtering yielded 7,176 valid corner kicks for training and evaluation. We summarised the exact information that was used to construct the input graphs in Table 2. In particular, other than player heights (measured in centimeters (cm)) and weights (measured in kilograms (kg)), the players were anonymous in the model. For the cases in which the player profiles were missing, we set their heights and weights to 180cm and 75kg, respectively, as defaults. In total, we had 385 such occurrences out of a total of 213,246 (= 22×9,693) during data preprocessing. We downscaled the heights and weights by a factor of 100. Moreover, for each corner kick, we zero-centered the positions of on-pitch players and normalised them onto a 10m × 10m pitch, and their velocities were re-scaled accordingly. For the cases in which the pitch dimensions were missing, we used a standard pitch dimension of 110m × 63m as default.

We summarised the grouping of the features in Table 1. The actual features used in different benchmark tasks may differ, and we will describe this in more details in the next section. To focus on modelling the high-level tactics played by the attacking and defending teams, other than a binary indicator for ball possession, no information of ball movement, neither positions nor velocities, was used to construct the input graphs.

Benchmark tasks construction TacticAI consists of three predictive and generative models, which also corresponded to three benchmark tasks implemented in this study. Specifically, 1) Receiver prediction, 2) Threatening shot prediction, and 3) Guided generation of team positions and velocities (Table 1). The graphs of all the benchmark tasks used the same feature space of nodes and edges, differing only in the global features.

For all three tasks, our models first transform the node features to a *latent node feature matrix*, $\mathbf{H} = f_{\mathcal{G}}(\mathbf{X}, \mathbf{E}, \mathbf{g})$, from which we could answer queries: either about individual players—in which

即无论是将 $\mathfrak{g}$ 应用于函数 $g_{\mathcal{G}}$ 的输入还是输出，最终的答案都是相同的。这种条件也被称为 $\mathfrak{G}$-等变性质，它最近被证明是开发针对生物化学数据的强大GNNs的一种有效范式 [12, 31]。

为了满足 D_2-等变性质，我们采用了*群卷积方法* [7]。在其中，允许输入的视图直接与其$\mathfrak{G}$-变换的变体进行交互，这种方式与网格卷积非常相似（实际上，网格卷积是群卷积的一个特例，将 $\mathfrak{G}$ 设置为平移群）。我们使用 $\mathbf{H}_{\mathfrak{g}}^{(t)}$ 来表示在第 t 层的潜在节点特征的 $\mathfrak{g}$-变换视图。为了简洁起见，省略了 $\mathbf{E}$ 和 $\mathbf{g}$ 输入，并使用我们之前设计的 $g_{\mathcal{G}}$ 层作为一个构建块，可以执行如下群卷积：

$$\mathbf{H}_{\mathfrak{g}}^{(t)} = g_{\mathcal{G}}^{\text{equiv}}(\mathbf{H}_{\mathfrak{g}}^{(t-1)}) = \frac{1}{|\mathfrak{G}|}\sum_{\mathfrak{h}\in\mathfrak{G}} g_{\mathcal{G}}\left(\mathbf{H}_{\mathfrak{h}}^{(t-1)} \| \mathbf{H}_{\mathfrak{g}^{-1}\mathfrak{h}}^{(t-1)}\right) \tag{8}$$

在这里，$\|$ 是连接操作，将两个节点特征矩阵按列连接起来；$\mathfrak{g}^{-1}$ 是 $\mathfrak{g}$ 的逆变换（必须存在，因为 $\mathfrak{G}$ 是一个群）；而 $\mathfrak{g}^{-1}\mathfrak{h}$ 是两个变换的复合。

实际上，方程 8 暗示我们的 D_2-等变 GNN 需要为当前输入的每个 $\mathfrak{G}$-变换维护一个节点特征矩阵 $\mathbf{H}_{\mathfrak{g}}^{(t)}$，并通过在所有通过应用变换 $\mathfrak{h}$ 相关联的一对中调用 $g_{\mathcal{G}}$ 来重新组合这些视图。请注意，所有的反射都是自逆的，因此在 D_2 中，$\mathfrak{g} = \mathfrak{g}^{-1}$。

网络架构 尽管我们执行的三项基准任务在全局特征上对模型的可用性有细微差别，但为它们设计的神经网络模型都具有相同的 *编码器-解码器* 架构。编码器在所有任务中结构相同，而解码器模型则针对每个基准任务适当地产生形状正确的输出。

给定一个输入图，TacticAI 的模型首先生成所有相关的 D_2-变换版本，通过适当地反射玩家坐标和速度。我们将原始输入图称为 *身份视图*，其余三个 D_2-变换的图称为 *反射视图*。

一旦准备好视图，我们应用四个群卷积层（方程 8），使用 GATv2 基础模型（方程 3–4）。每个 GATv2 层有八个注意力头，并为每个玩家计算四个潜在特征。因此，在完成四个群卷积之后，我们有 $\mathbf{H} \in \mathbb{R}^{4\times22\times4}$ 的表示，其中第一个维度对应于四个视图（$\mathbf{H}_{\text{id}}, \mathbf{H}_{\leftrightarrow}, \mathbf{H}_{\updownarrow}, \mathbf{H}_{\leftrightarrow\updownarrow} \in \mathbb{R}^{22\times4}$），第二个维度对应于玩家（每个队有十一个），第三个维度对应于该特定视图中每个玩家节点的 4-维潜在向量。这个表示如何被解码器使用取决于特定的下游任务：

- 对于接收器预测，这是一个完全不变的函数（即反射不会改变接收器），我们对所有视角进行简单的帧平均处理，得到如下结果：

$$\mathbf{H}^{\text{node}} = \frac{\mathbf{H}_{\text{id}} + \mathbf{H}_{\leftrightarrow} + \mathbf{H}_{\updownarrow} + \mathbf{H}_{\leftrightarrow\updownarrow}}{4} \tag{9}$$

  然后在对 $\mathbf{H}^{\text{node}} \in \mathbb{R}^{22\times4}$ 的行上进行节点分类器的学习。我们进一步使用一个线性层将 $\mathbf{H}^{\text{node}}$ 解码为一个 logit 向量 $\mathbf{O} \in \mathbb{R}^{22}$，在此之后计算相应的 softmax 交叉熵损失。
- 对于投篮预测，这同样是完全不变性的（即反射不会改变投篮的概率），我们可以进一步将所有球员的帧平均特征进行平均，以获得全局图表示：

$$\mathbf{h}^{\text{graph}} = \frac{1}{22}\sum_{u=1}^{22} \mathbf{h}_u^{\text{node}} \tag{10}$$

case we learned a relevant classifier or regressor over the $\mathbf{h}_u$ vectors (the rows of $\mathbf{H}$)—or about the occurrence of a global event (e.g. shot taken)—in which case we classified or regressed over the aggregated player vectors, $\sum_u \mathbf{h}_u$. In both cases, the classifiers were trained using stochastic gradient descent over an appropriately chosen loss function, such as categorical cross-entropy for classifiers, and mean squared error for regressors.

For different tasks, we extracted the corresponding ground-truth labels from either the event stream data or the tracking data. Specifically, 1) We modelled receiver prediction as a node classification task, and labelled the first player to touch the ball after the the corner was taken as the target node. This player could be either an attacking or defensive player. 2) Shot prediction was modelled as graph classification. In particular, we considered a next-ball-touch action by the attacking team as a shot if it was a direct corner, a goal, an aerial, hit on the goalposts, a shot attempt saved by the goalkeeper, or missing target. This yielded $1,736$ corners labelled as a shot being taken, and $5,440$ corners labelled as a shot not being taken. 3) For guided generation of player position and velocities, no additional label was needed, as this model relied on a self-supervised reconstruction objective.

The entire dataset was split into training and evaluation sets with a $80:20$ ratio through random sampling, and the same splits were used for all tasks.

Graph neural networks The central model of TacticAI is the *graph neural network* (**GNN**) [41], which computes latent representations on a graph by repeatedly combining them within each node's neighbourhood. Here we define a node's neighbourhood, $\mathcal{N}_u$, as the set of all first-order neighbours of node u, that is, $\mathcal{N}_u = \{v \mid (v, u) \in \mathcal{E}\}$. A single GNN layer then transforms the node features by passing messages between neighbouring nodes [13], following the notation of related work [5], and the implementation of the CLRS-30 benchmark baselines [42]:

$$\mathbf{h}_u^{(t)} = \phi\left(\mathbf{h}_u^{(t-1)}, \bigoplus_{v \in \mathcal{N}_u} \psi\left(\mathbf{h}_u^{(t-1)}, \mathbf{h}_v^{(t-1)}, \mathbf{e}_{vu}, \mathbf{g}\right)\right) \tag{2}$$

where $\psi : \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^{k'}$ and $\phi : \mathbb{R}^k \times \mathbb{R}^{k'} \to \mathbb{R}^{k'}$ are two learnable functions (e.g. multilayer perceptrons), $\mathbf{h}_u^{(t)}$ are the features of node u after t GNN layers, and $\bigoplus$ is any permutation-invariant aggregator, such as sum, max, or average. By definition, we set $\mathbf{h}_u^{(0)} = \mathbf{x}_u$, and iterate Equation 2 for T steps, where T is a hyperparameter. Then, we let $\mathbf{H} = f_{\mathcal{G}}(\mathbf{X}, \mathbf{E}, \mathbf{g}) = \mathbf{H}^{(T)}$ be the final node embeddings coming out of the GNN.

It is well known that Equation 2 is remarkably general; it can be used to express popular models such as Transformers [39] as a special case, and it has been argued that *all* discrete deep learning models can be expressed in this form [40, 3]. This makes GNNs a perfect framework for benchmarking various approaches to modelling player-player interactions in the context of football.

Different choices of ψ, ϕ and $\bigoplus$ yield different architectures. In our case, we utilise a message

然后我们在 $\mathbf{h}^{\text{graph}} \in \mathbb{R}^4$ 上学习一个二分类器。具体来说，我们将隐藏向量通过一个线性层解码为一个单独的logit，并计算与对应标签的sigmoid二进制交叉熵损失。

- 在引导生成（位置/速度调整）方面，我们根据人类教练感兴趣的一个特定结果生成玩家的位置和速度，这是通过对隐藏特征矩阵的行进行预测得出的。例如，模型可能会调整防守布局以降低进攻队的射门概率。现在的模型输出是等变的而不是不变的——适当反映球场的特征可以反映出预测的位置和速度向量。因此，我们不能执行帧平均，而只取身份视图的特征，$\mathbf{H}_{\text{id}} \in \mathbb{R}^{22\times4}$。从这个潜在特征矩阵中，我们可以从每一行学习到一个条件分布，该分布模拟了相应玩家的位置或速度。为此，我们使用变分自编码器（VAE [22]）来扩展骨干编码器。具体来说，对于$\mathbf{H}_{\text{id}}$的第u行，$\mathbf{h}_u$，我们首先将其潜在嵌入映射到二维高斯分布$\mathcal{N}(\mu_u|\sigma_u)$的参数上，然后从这个分布中抽取坐标和速度。在训练时，我们可以使用重参数化技巧 [22]高效地通过此采样操作传播梯度：为每个玩家从单位高斯分布中抽取一个随机值$\epsilon_u \sim \mathcal{N}(0,1)$，然后将$\mu_u + \sigma_u\epsilon_u$视为此玩家的样本。以下内容为了简洁起见，我们省略了边缘特征。对于每个角球样本$\mathbf{X}$及其对应的结果$\mathbf{o}$（例如，表示射门事件的二进制值），我们将标准的VAE损失 [22]扩展到我们这种结果条件引导生成的案例中。

$$\mathcal{L}(\theta, \phi) = -\mathbb{E}_{\mathbf{h}_u \sim q_\theta(\mathbf{h}_u|\mathbf{X},\mathbf{o})}[\log p_\phi(\mathbf{x}_u|\mathbf{h}_u, \mathbf{o})] + \mathbb{KL}(q_\theta(\mathbf{h}_u|\mathbf{X}, \mathbf{o}) \| p(\mathbf{h}_u|\mathbf{o})) \tag{11}$$

其中 $\mathbf{h}_u$ 是对应于 $\mathbf{H}_{\text{id}}$ 的第 u 行的玩家嵌入，而 $\mathbb{KL}$ 是Kullback-Leibler (KL) 散度。具体来说，第一个项是真实玩家输入 $\mathbf{x}_u$ 与从 $\mathbf{h}_u$ 使用解码器 p_ϕ 解码重构的样本之间的生成损失。使用KL 项，潜在嵌入 $\mathbf{h}_u$ 的分布被正则化为趋向于 $p(\mathbf{h}_u|\mathbf{o})$，在我们的情况下这是一个多元高斯分布。

可以在S.Figure 2中概括TacticAI采用的通用编码器-解码器等变架构的完整高级摘要。在下一节中，我们将提供实证证据来证明这些架构决策的合理性。这将通过在我们预测基准（接球预测和射门预测）上进行有针对性的消融研究来完成。

消融研究 正如在结果和分析部分简要描述的那样，我们利用接球预测任务来评估各种基础模型架构，并直接定量评估在此背景下几何深度学习的贡献。我们已经看到，原始角球数据可以通过几何深度学习更好地表示，在潜在空间中产生可分离的簇，这些簇可能对应于不同的进攻或防守策略（图 2）。此外，我们假设这些表示还可以在接球预测任务上获得更好的性能。因此，我们在这个任务上对深度学习的几个设计选择进行消融研究：

- *因子化的图表示有帮助吗？* 为了评估这一点，我们将之与一个不利用图表示的卷积神经网络（CNN [9]）基线进行了比较。
- *图结构是否有帮助？* 为了评估这一点，我们将与一个Deep Sets [44]基线进行了比较，该基线仅在不考虑邻接信息的情况下单独对每个节点进行建模——等价于，将每个邻域$\mathcal{N}_u$设置为单元素集合$\{u\}$。
- *注意力机制图神经网络是一个好策略吗？* 为了评估这一点，我们将之与一种消息传递神经网络（[14, MPNN]）基线进行比较，该基线使用了来自方程 2中的完全势能图神经网络层，而不是GATv2。

function that factorises into an *attentional mechanism*, $a : \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$:

$$\mathbf{h}_u^{(t)} = \phi\left(\mathbf{h}_u^{(t-1)}, \bigoplus_{v\in\mathcal{N}_u} a\left(\mathbf{h}_u^{(t-1)}, \mathbf{h}_v^{(t-1)}, \mathbf{e}_{vu}, \mathbf{g}\right)\psi\left(\mathbf{h}_v^{(t-1)}\right)\right) \tag{3}$$

yielding the graph attention network (GAT) architecture [43]. In our work, specifically, we use a two-layer multilayer perceptron for the attentional mechanism, as proposed by GATv2 [4]:

$$a\left(\mathbf{h}_u^{(t-1)}, \mathbf{h}_v^{(t-1)}, \mathbf{e}_{vu}, \mathbf{g}\right) = \underset{v\in\mathcal{N}_u}{\mathrm{softmax}}\ \mathbf{a}^\top \mathrm{LeakyReLU}\left(\mathbf{W}_1\mathbf{h}_u^{(t-1)} + \mathbf{W}_2\mathbf{h}_v^{(t-1)} + \mathbf{W}_e\mathbf{e}_{vu} + \mathbf{W}_g\mathbf{g}\right) \tag{4}$$

where $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{k\times h}$, $\mathbf{W}_e \in \mathbb{R}^{l\times h}$, $\mathbf{W}_g \in \mathbb{R}^{m\times h}$ and $\mathbf{a} \in \mathbb{R}^h$ are the learnable parameters of the attentional mechanism, and LeakyReLU is the leaky rectified linear activation function. This mechanism computes coefficients of interaction (a single scalar value) for each pair of connected nodes (u, v), which are then normalised across all neighbours of u using the softmax function.

Through early-stage experimentation we have ascertained that GATs are capable of matching the performance of more generic choices of ψ (such as the MPNN [13]) while being more scalable. Hence, we focus our study on the GAT model in this work. More details can be found in the Ablation Study section.

Geometric deep learning In spite of the power of Equation 2, using it in its full generality is often prone to overfitting, given the large number of parameters contained in ψ and ϕ. This problem is exacerbated in the football analytics domain, where gold-standard data is generally very scarce—for example, in the English Premier League, only a few hundred games are played every season.

In order to tackle this issue, we can exploit the immense *regularity* of data arising from football games. Strategically equivalent game states are also called *transpositions*, and symmetries such as arriving at the same chess position through different move sequences have been exploited computationally since the 1960s [17]. Similarly, game rotations and reflections may yield equivalent strategic situations [32]. Using the blueprint of *geometric deep learning* (**GDL**) [5], we can design specialised GNN architectures that exploit this regularity.

That is, geometric deep learning is a generic methodology for deriving mathematical constraints on neural networks, such that they will behave predictably when inputs are transformed in certain ways. In several important cases, these constraints can be directly resolved, directly informing neural network architecture design. For a comprehensive example for point clouds under 3D rotational symmetry, see Fuchs *et al.* [12].

To elucidate several aspects of the GDL framework on a high level, let us assume that there exists a group of input data transformations (*symmetries*), $\mathfrak{G}$, under which the ground-truth label remains unchanged. Specifically, if we let $y(\mathbf{X}, \mathbf{E}, \mathbf{g})$ be the label given to the graph featurised with $\mathbf{X}, \mathbf{E}, \mathbf{g}$, then for every transformation $\mathfrak{g} \in \mathfrak{G}$, the following property holds:

$$y(\mathfrak{g}(\mathbf{X}), \mathfrak{g}(\mathbf{E}), \mathfrak{g}(\mathbf{g})) = y(\mathbf{X}, \mathbf{E}, \mathbf{g}) \tag{5}$$

- *考虑对称性是否有帮助？* 为了评估这一点，我们将带有几何GATv2基线的模型与不使用D_2群卷积但使用D_2帧平均的模型，以及根本不显式利用D_2对称性任何方面的模型进行了比较。

这些模型每个都以固定的预算训练了50,000个训练步骤。所训练模型的测试前k接收器预测准确度在S.Table 2中提供。正如在结果与分析中已经讨论的，使用完整的图结构以及直接考虑反射对称性具有明显优势。此外，与GATv2相比，使用MPNN层会导致轻微过拟合，这说明了注意力机制GNN在这项任务上在表达性与数据效率之间取得了良好的平衡。我们的分析突显了图表示学习与几何深度学习在从追踪数据中进行足球分析时所具有的定量优势。我们还提供了针对射门预测任务的简要消融研究，见S.Table 3。

训练细节 我们独立训练每个TacticAI模型，使用NVIDIA Tesla P100 GPUs。为了最小化过拟合，每个模型的学习目标都通过网络参数的L_2范数惩罚进行正则化。在训练过程中，我们使用Adam随机梯度下降优化器 [21]来优化正则化损失。我们在S.Table 1中总结了每个TacticAI模型的超参数。

Data availability

在当前研究中生成和/或分析的数据集由于许可限制而不公开可用。然而，合理请求下，对应作者可提供数据提供者的联系方式。

Acknowledgements

我们衷心感谢James French、Timothy Waskett、Hans Leitert和Benjamin Hervey在分析TacticAI输出方面所做的广泛努力。此外，我们还感谢Kevin McKee、Sherjil Ozair和Beatrice Bevilacqua在技术讨论中的宝贵意见，以及Marc Lanctôt和Satinder Singh在提交论文前对论文的审阅。

Author contributions

Z.W.、D.H.*、L.P.和K.T.协调和组织了导致这篇论文的研究工作。P.V.和Z.W.开发了核心TacticAI模型。Z.W.、W.S.和I.G.准备了用于训练和评估这些模型的英超联赛角球数据集。P.V.、Z.W.、D.H.*和N.T.与人类专家设计了案例分析，并对结果进行了定性的评估和分析。所有作者都参与了论文的撰写、技术讨论以及对最终手稿的反馈。

Supplementary Materials

Section S1.1	Literature Review
Section S1.2	Details of Case Study Design
Section S1.3	Supplementary Figures and Tables

*D.H.指的是Daniel Hennes。

This condition is also referred to as $\mathfrak{G}$-invariance.

It is worth noting that GNNs may also be derived using a GDL perspective, if we set the symmetry group $\mathfrak{G}$ to $S_{|V|}$, the permutation group of $|V|$ objects. Owing to the design of Equation 2, its outputs will not be dependent on the exact permutation of nodes in the input graph.

Frame averaging A simple mechanism to enforce $\mathfrak{G}$-invariance, given any predictor $f_{\mathcal{G}}(\mathbf{X}, \mathbf{E}, \mathbf{g})$, performs *frame averaging* across all $\mathfrak{G}$-transformed inputs:

$$f_{\mathcal{G}}^{\text{inv}}(\mathbf{X}, \mathbf{E}, \mathbf{g}) = \frac{1}{|\mathfrak{G}|} \sum_{\mathfrak{g} \in \mathfrak{G}} f_{\mathcal{G}}(\mathfrak{g}(\mathbf{X}), \mathfrak{g}(\mathbf{E}), \mathfrak{g}(\mathbf{g})) \tag{6}$$

This ensures that all $\mathfrak{G}$-transformed versions of a particular input (also known as that input's *orbit*) will have exactly the same output, satisfying Equation 5. A variant of this approach has also been applied in the AlphaGo architecture [36] to encode symmetries of a Go board.

In our specific implementation, we set $\mathfrak{G} = D_2 = \{\text{id}, \leftrightarrow, \updownarrow, \leftrightarrow\updownarrow\}$, the dihedral group. Exploiting D_2-invariance allows us to encode *quadrant symmetries*. Each element of the D_2 group encodes the presence of vertical or horizontal reflections of the input football pitch. Under these transformations, the pitch is assumed completely symmetric, and hence many predictions, such as which player receives the corner kick, or takes a shot from it, can be safely assumed unchanged. As an example for how to compute transformed features in Equation 6, $\leftrightarrow(\mathbf{X})$ horizontally reflects all positional features of players in $\mathbf{X}$ (e.g. the coordinates of the player), and negates the x-axis component of their velocity.

Group convolutions While the frame averaging approach of Equation 6 is a powerful way to restrict GNNs to respect input symmetries, it arguably misses an opportunity for the different $\mathfrak{G}$-transformed views to interact while their computations are being performed. For small groups such as D_2, a more fine-grained approach can be assumed, operating over a single GNN layer in Equation 2, which we will write shortly as $\mathbf{H}^{(t)} = g_{\mathcal{G}}(\mathbf{H}^{(t-1)}, \mathbf{E}, \mathbf{g})$. The condition that we need a symmetry-respecting GNN layer to respect is as follows, for all transformations $\mathfrak{g} \in \mathfrak{G}$:

$$g_{\mathcal{G}}(\mathfrak{g}(\mathbf{H}^{(t-1)}), \mathfrak{g}(\mathbf{E}), \mathfrak{g}(\mathbf{g})) = \mathfrak{g}(g_{\mathcal{G}}(\mathbf{H}^{(t-1)}, \mathbf{E}, \mathbf{g})) \tag{7}$$

that is, it does not matter if we apply $\mathfrak{g}$ to the input or the output of the function $g_{\mathcal{G}}$—the final answer is the same. This condition is also referred to as $\mathfrak{G}$-equivariance, and it has recently proved to be a potent paradigm for developing powerful GNNs over biochemical data [12, 31].

To satisfy D_2-equivariance, we apply the *group convolution* approach [7]. Therein, views of the input are allowed to directly interact with their $\mathfrak{G}$-transformed variants, in a manner very similar to grid convolutions (which is, indeed, a special case of group convolutions, setting $\mathfrak{G}$ to be the translation group). We use $\mathbf{H}_{\mathfrak{g}}^{(t)}$ to denote the $\mathfrak{g}$-transformed view of the latent node features at layer t. Omitting $\mathbf{E}$ and $\mathbf{g}$ inputs for brevity, and using our previously designed layer $g_{\mathcal{G}}$ as a building block,

S1. Supplementary Material

S1.1. Literature Review

在以下内容中，我们总结了与足球角球战术或传球预测具有相同目标的最先进技术，或者更广泛地说，制定了可能转移至本研究考虑背景的团队运动策略建模技术。由于我们这里关注的是游戏内战术的方法，因此体育分析研究涉及球员转会和比赛统计的范围之外。

早期的研究已经在识别游戏的关键属性方面产生了大量研究，这些属性来源于各种数据，并利用这些属性提供战术分析和建议 [8, 10, 15, 25, 28, 29]。一个最近的例子是应用随机森林和AdaBoost处理诸如球员位置等时间连续的时空数据的系统，以确定比赛是否正在进行或已中断，称为“比赛中状态” [23]，类似的方法也已被提出用于美式足球 [26]。

其他方法利用游戏数据来刻画球员的行为，比如他们的传球风格 [6]。此类信息可以后续用于可能需要输入球员历史行为表示的下游应用程序。同样，传球预测系统也可能使用游戏数据来预测在给定游戏状态下传球的可能性 [18]。

更接近相关的研究线路分析足球视频以确定射门事件概率，这可以作为球员表现的代理，通过检查时空关系和预测不确定性 [16]。该方法使用图卷积循环神经网络处理视频以捕捉潜在特征，并利用贝叶斯神经网络预测射门概率。类似的其他方法应用机器学习和可解释AI技术进行战术分析以及预测防守战术的成功 [11]。

从球员轨迹中准确识别角球战术的关键模式是足球分析中下游任务的关键前提，包括评估战术的有效性、提高其表现和设计对抗战术。通过利用深度学习，以前的研究表明，游戏内事件的结果，包括射门 [37]和传球 [9, 19]，是可以使用球员轨迹的时空追踪数据来预测的。此外，有效的防守战术模式也可以从追踪数据中恢复 [37]。然而，由于足球比赛中的随机性和复杂动态，追踪数据可能相当噪声，这使得从中挖掘战术模式变得具有挑战性。

还有其他学习技术用于建模团队运动的例子。生成关系推理网络（GRIN） [24]也遵循VAE框架来学习场景中每个代理的解纠缠表示。它通过图卷积网络（GCN）将每个实体编码成一个潜在表示，包括一个代理内意图和一个代理间关系。代理间关系通过图注意力层进行消息传递，其结果与代理内意图结合形成重构。尽管没有群不变层，但实验表明，这种方法可以从篮球球员中发现解纠缠因素，并且对这些因素的干预在预测的球员轨迹中产生了可解释的变化。

DeepHoops [34]使用NBA游戏的时空追踪数据，提出了一种深度学习架构，估计所有微动作对篮球比赛结果的影响。特别是，它允许评估单个无球事件对控球成功的贡献。

在[35]中，应用了一种树搜索方法（核回归UCT）来发现自我对抗中的冰壶策略。作者表明，模型的反事实行为可以用来教导人类非平凡的冰壶概念，正如在用户研究中所证明的。

we can perform a group convolution as follows:

$$\mathbf{H}_{\mathfrak{g}}^{(t)} = g_{\mathcal{G}}^{\text{equiv}}(\mathbf{H}_{\mathfrak{g}}^{(t-1)}) = \frac{1}{|\mathfrak{G}|}\sum_{\mathfrak{h}\in\mathfrak{G}} g_{\mathcal{G}}\left(\mathbf{H}_{\mathfrak{h}}^{(t-1)} \| \mathbf{H}_{\mathfrak{g}^{-1}\mathfrak{h}}^{(t-1)}\right) \tag{8}$$

Here, $\|$ is the concatenation operation, joining the two node feature matrices column-wise; $\mathfrak{g}^{-1}$ is the inverse transformation to $\mathfrak{g}$ (which must exist as $\mathfrak{G}$ is a group); and $\mathfrak{g}^{-1}\mathfrak{h}$ is composition of the two transformations.

Effectively, Equation 8 implies our D_2-equivariant GNN needs to maintain a node feature matrix $\mathbf{H}_{\mathfrak{g}}^{(t)}$ for every $\mathfrak{G}$-transformation of the current input, and these views are recombined by invoking $g_{\mathcal{G}}$ on all pairs related together by applying a transformation $\mathfrak{h}$. Note that all reflections are self-inverses, hence, in D_2, $\mathfrak{g} = \mathfrak{g}^{-1}$.

Network architectures While the three benchmark tasks we are performing have minor differences in the global features available to the model, the neural network models designed for them all have the same *encoder-decoder* architecture. The encoder has the same structure in all tasks, while the decoder model is tailored to produce appropriately-shaped outputs for each benchmark task.

Given an input graph, TacticAI's model first generates all relevant D_2-transformed versions of it, by appropriately reflecting the player coordinates and velocities. We refer to the original input graph as the *identity view*, and the remaining three D_2-transformed graphs as *reflected views*.

Once the views are prepared, we apply four group convolutional layers (Equation 8) with a GATv2 base model (Equations 3–4). Each GATv2 layer has eight attention heads, and computes four latent features overall per player. Accordingly, once the four group convolutions are performed, we have a representation of $\mathbf{H} \in \mathbb{R}^{4\times 22\times 4}$, where the first dimension corresponds to the four views ($\mathbf{H}_{\text{id}}, \mathbf{H}_{\leftrightarrow}, \mathbf{H}_{\updownarrow}, \mathbf{H}_{\leftrightarrow\updownarrow} \in \mathbb{R}^{22\times 4}$), the second dimension corresponds to the players (eleven on each team), and the third corresponds to the 4-dimensional latent vector for each player node in this particular view. How this representation is used by the decoder depends on the specific downstream task:

- For receiver prediction, which is a fully *invariant* function (i.e. reflections do not change the receiver), we perform simple frame averaging across all views, arriving at

$$\mathbf{H}^{\text{node}} = \frac{\mathbf{H}_{\text{id}} + \mathbf{H}_{\leftrightarrow} + \mathbf{H}_{\updownarrow} + \mathbf{H}_{\leftrightarrow\updownarrow}}{4} \tag{9}$$

  and then learn a node-wise classifier over the rows of $\mathbf{H}^{\text{node}} \in \mathbb{R}^{22\times 4}$. We further decode $\mathbf{H}^{\text{node}}$ into a logit vector $\mathbf{O} \in \mathbb{R}^{22}$ with a linear layer before computing the corresponding softmax cross entropy loss.
- For shot prediction, which is once again fully *invariant* (i.e. reflections do not change the probability of a shot), we can further average the frame-averaged features across all players to

S1.2. Case Study Design

正如在结果部分先前描述的，我们设计了一项深入的案例研究来评估TacticAI的预测和生成能力的优势。我们设计了四个具体的任务，提供给与利物浦足球俱乐部有关联的五位专家。为了防止我们的评分者产生偏见，我们在进行案例研究之前严格地将这些任务的所有细节对他们保密。此外，对于所有展示的角球情况，我们没有透露球队或球员的身份——只提供了角球开出时刻球员的位置和速度，以及他们是进攻队还是防守队。

任务1:TacticAI能否生成真实的战术调整？ 为了回答这个问题，我们从英超联赛的50个角球数据集开始，这些数据集是从TacticAI的训练集中剔除的（S.Figure 3）。对于这些角球的一个子集，我们应用了TacticAI的生成头部来提出替代战术，以最小化或最大化射门概率，并用TacticAI的建议替换原始角球。然后我们要求评分者判断这些角球是真实的还是生成的。在任务完成后与评分者的讨论中，以及观察他们的评论时，我们注意到为了检测真实性，采用了多种策略：

- 大多数评分者选择根据检测场景中是否有明显不寻常之处来判断真实性，例如一个让对手玩家处于明显空旷的策略。
- 一个评分者采用了另一种策略，即根据情况是否可能合理发生来判断现实主义，即使其具有明显次优的战术设置。这种方法最终将大多数例子标记为真实。
- 专门从事角球分析的评分员更倾向于采用“检索”策略，即如果他们能够从记忆中回忆起确切的设置，他们将把这个角球评为真实。

尽管存在这些多种方法，正如在结果部分所讨论的，所有的评分者通常都对哪些样本是真实的感到困惑，并且他们一致强调了这个任务有多么具有挑战性。

任务2:TacticAI能否预测合理的接球者？ 为了回答这个问题，我们重新使用了与任务1相同的50个角球，并同时询问我们的评分者，进攻球队的哪位球员最有可能首先接触到球（S.图 3）。允许评分者提供尽可能少或尽可能多的球员—我们将他们的提议与TacticAI推荐的前三位接球者进行交叉参考。通常，我们发现不同评分者提出的球员数量存在显著差异—有些人几乎在所有研究的情况下建议3–5个接球者，而其他人大多数情况下不想建议超过一个接球者。正如在结果部分所讨论的，尽管这种方法存在多样性，但TacticAI的前三位预测与评分者的估计大体一致，并且TacticAI在真实和生成的设置上的准确性没有显著差异，这表明TacticAI可以稳健地用于预测性分析真实的和模拟的角球。正如几位评分者所强调的，这样一个系统将对足球分析师非常有用：接球者预测器可以用来指示哪些球员可能需要在防守上特别关注，以及评估已经进行的比赛中角球的发展是否如预期的那样。

任务3:TacticAI可以用来检索有用的相似角球吗？ 为了回答这个问题，我们利用了来自英超联赛的50个参考角球，这些角球是从TacticAI的训练集中预留出来的（S.图 4）。对于这些每个角球，我们检索了一个与之最接近的英超联赛角球—要么在TacticAI的图级嵌入空间中，要么在输入空间中

get a global graph representation:

$$\mathbf{h}^{\text{graph}} = \frac{1}{22}\sum_{u=1}^{22}\mathbf{h}_u^{\text{node}} \tag{10}$$

and then learn a binary classifier over $\mathbf{h}^{\text{graph}} \in \mathbb{R}^4$. Specifically, we decode the hidden vector into a single logit with a linear layer, and the compute the sigmoid binary cross entropy loss with the corresponding label.

- For guided generation (position/velocity adjustments), we generate the player positions and velocities with respect to a particular outcome of interest for the human coaches, predicted over the rows of the hidden feature matrix. For example, the model may adjust the defensive setup to decrease the shot probability by the attacking team. The model output is now *equivariant* rather than invariant—reflecting the pitch appropriately reflects the predicted positions and velocity vectors. As such, we cannot perform frame averaging, and take only the *identity view*'s features, $\mathbf{H}_{\text{id}} \in \mathbb{R}^{22\times 4}$. From this latent feature matrix, we can then learn a conditional distribution from each row, which models the positions or velocities of the corresponding player. To do this, we extend the backbone encoder with a Variational Autoencoder (VAE [22]). Specifically, for the u-th row of $\mathbf{H}_{\text{id}}$, $\mathbf{h}_u$, we first map its latent embedding to the parameters of a two-dimensional Gaussian distribution $\mathcal{N}(\mu_u|\sigma_u)$, and then sample the coordinates and velocities from this distribution. At training time, we can efficiently propagate gradients through this sampling operation using the reparameterisation trick [22]: sample a random value $\epsilon_u \sim \mathcal{N}(0,1)$ for each player from the unit Gaussian distribution, and then treat $\mu_u + \sigma_u\epsilon_u$ as the sample for this player. In what follows, we omit edge features for brevity. For each corner kick sample $\mathbf{X}$ with the corresponding outcome $\mathbf{o}$ (e.g. a binary value indicating a shot event), we extend the standard VAE loss [22] to our case of outcome-conditional guided generation as

$$\mathcal{L}(\theta,\phi) = -\mathbb{E}_{\mathbf{h}_u\sim q_\theta(\mathbf{h}_u|\mathbf{X},\mathbf{o})}[\log p_\phi(\mathbf{x}_u|\mathbf{h}_u,\mathbf{o})] + \mathbb{KL}(q_\theta(\mathbf{h}_u|\mathbf{X},\mathbf{o})\|p(\mathbf{h}_u|\mathbf{o})) \tag{11}$$

where $\mathbf{h}_u$ is the player embedding corresponding to the u-th row of $\mathbf{H}_{\text{id}}$, and $\mathbb{KL}$ is Kullback-Leibler (KL) divergence. Specifically, the first term is the generation loss between the real player input $\mathbf{x}_u$ and the reconstructed sample decoded from $\mathbf{h}_u$ with the decoder p_ϕ. Using the KL term, the distribution of the latent embedding $\mathbf{h}_u$ is regularised towards $p(\mathbf{h}_u|\mathbf{o})$, which is a multivariate Gaussian in our case.

A complete high-level summary of the generic encoder-decoder equivariant architecture employed by TacticAI can be summarised in S.Figure 2. In the following section, we will provide empirical evidence for justifying these architectural decisions. This will be done through targeted ablation studies on our predictive benchmarks (receiver prediction and shot prediction).

Ablation Study As described briefly in the Results and Analysis section, we leveraged the receiver prediction task as a way to evaluate various base model architectures, and directly quantitatively

原始球员特征的距离上。这个决定使我们不仅可以测量TacticAI的嵌入是否可以用来挖掘相似的角球，还可以评估它们是否比纯粹基于位置更简单的挖掘启发式方法更有优势。我们让评分者判断他们是否认为检索到的角球对参考角球有用地相似；即，并排展示这两个角球是否会被认为是有用的。评分者再次采用了多种方法来决定两个角球的哪些特征足够显著以至于有用。评分者提到的常见考虑因素包括：角球是内旋还是外旋，是否存在短角球选项，区域防守与盯人防守方式，计算18码区内球员的数量，守门员的位置，以及判断两个角球是否可能有相同的进攻或防守队伍，或者来自同一场比赛。与之前的任务类似—从结果部分的讨论我们可以得出结论，尽管评分者考虑了显著特征的广泛多样性，但他们普遍认为TacticAI的检索优于位置启发式。他们还强调了这种检索系统的高效用性：分析师和教练可以使用它来发现和准备其他队伍的常见战术，以及发掘他们可能忽略的特定角球的新想法和变化。还强调说这里采用的“top-1”检索研究可能过于严格。虽然为了使评分者工作量可控，有必要在这里只关注一个检索到的角球，但实际上，分析师可能希望对每个参考角球分析多达10–20个检索到的角球。鉴于TacticAI在top-1情况下已经有一个有利的分数，我们预计如果将其应用于实际使用，其效用将进一步提升。

任务4:TacticAI能否生成有用的战术调整？ 为了回答这个问题，我们再次收集了英超联赛50个保留角球的数据集（S.Figure 5）。对于每一个角球，我们都使用了TacticAI的引导生成模型来提出对防守队球员位置和速度的调整——保持攻击队不变——以降低预测的射门概率。我们选择关注防守队以匹配现实世界的工作量：通常，角球设置由攻击队决定，之后防守队需要以最佳方式对此设置做出回应。对于这些角球，我们让评分者并排查看两个版本（原始版本和TacticAI调整版本）。我们指出左侧的角球是参考角球，要求评分者判断考虑到右侧调整的情况下，防守队的球员是否处于更好的或更差的位置。我们还要求评分者判断右侧的角球总体上对于防守队来说是更好还是更差的情况。为了控制偏见，我们随机选择了一部分角球设置，交换两个版本，使TacticAI的建议成为参考角球。评分者关注建议中的各种显著特征，包括但不限于：防守者是否更好地追踪攻击者的跑动，他们最好是移动还是静止，以及守门员的位置。有两种情况调整被认为是“可能有用”的，这取决于调整球员的特点。例如，某些跑动只有由身体能力或体能高的防守者执行才被认为可能成功。由于我们在这里训练的TacticAI变体无法获取除身高和体重之外的球员特征，这样的细微差别超出了本研究的范围，我们不认为它们是显著的观察结果。总体而言，如结果部分所述，评分者压倒性地支持TacticAI的建议，显示出很高的评分者间一致性。我们在图5中说明了四个TacticAI建议被认为最显著的情况；在这些情况中，有十个防守球员被认为有用地调整了位置，这几乎是整个防守队！此外，这种位置调整系统被我们的评分者一致认为是一个分析师或教练可以利用的高度有用的工具。即使是与参考角球并排查看这些建议的机会也被认为是有用的，因为它促使评分者立即考虑战略变化。由于引导生成系统故意设计为提供对球员坐标和速度的细微修改，评分者认为这样的系统最有用于帮助检测那些可能没有遵循教练指示、忽视战术比赛的球员。这可以导致教练针对性的干预——要么利用一个倾向于忽视指示的对方球员所暴露的弱点，要么提高对本队球员的教练效果，如果发现他们忽视既定的战术的话。结合任务1的积极结果，以及建议的调整通常很难与真实的角球情况区分开来的发现，我们得出任务4的问题得到了肯定的回答。

assess the contributions of geometric deep learning in this context. We already see that the raw corner kick data can be better represented through geometric deep learning, yielding separable clusters in the latent space which could correspond to different attacking or defending tactics (Figure 2). In addition, we hypothesise that these representations can also yield better performance on the task of receiver prediction. Accordingly, we ablate several design choices using deep learning on this task:

- *Does a factorised graph representation help?* To assess this, we compare against a convolutional neural network (CNN [9]) baseline, which does not leverage a graph representation.
- *Does a graph structure help?* To assess this, we compare against a Deep Sets [44] baseline, which only models each node in isolation without considering adjacency information—equivalently, setting each neighbourhood $\mathcal{N}_u$ to a singleton set $\{u\}$.
- *Are attentional GNNs a good strategy?* To assess this, we compare against a message passing neural network [14, MPNN] baseline, which uses the fully potent GNN layer from Equation 2 instead of the GATv2.
- *Does accounting for symmetries help?* To assess this, we compare our geometric GATv2 baseline against one which does not utilise D_2 group convolutions but utilises D_2 frame averaging, and one which does not explicitly utilise any aspect of D_2 symmetries at all.

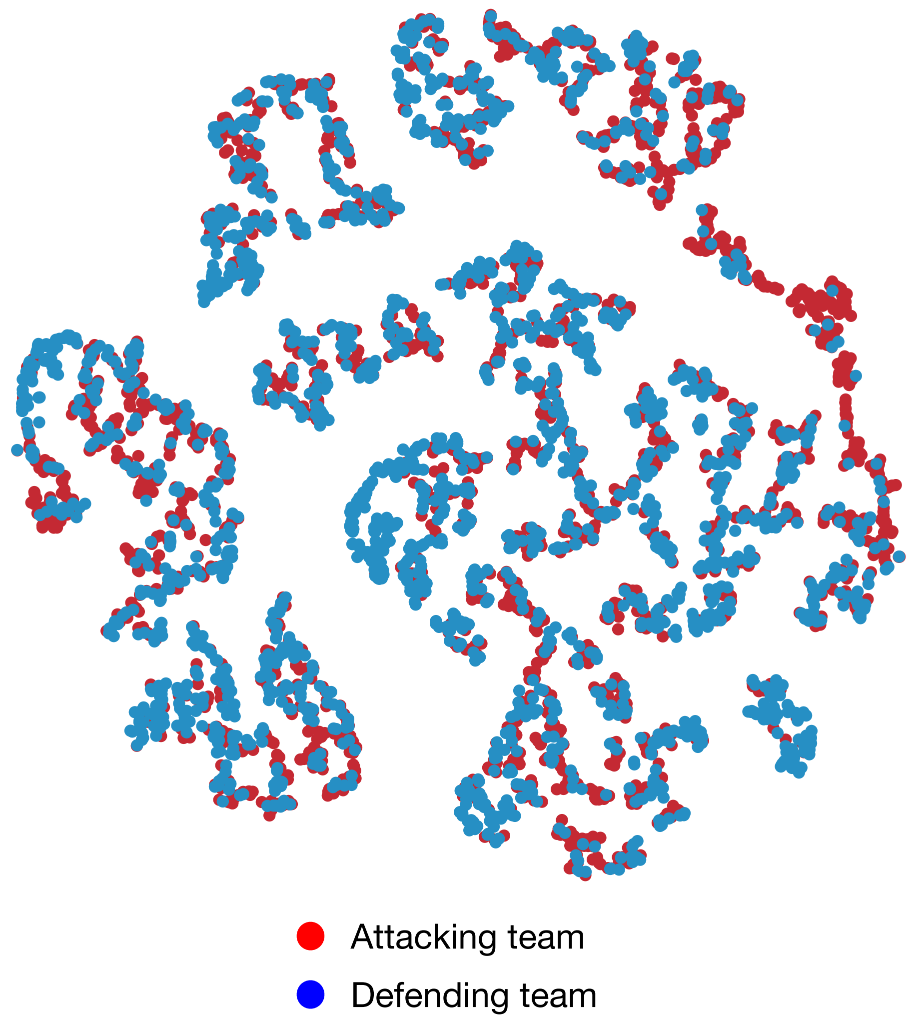
Each of these models have been trained for a fixed budget of $50,000$ training steps. The test top-k receiver prediction accuracies of the trained models are provided in S.Table 2. As already discussed in Results and Analysis, there is a clear advantage to using a full graph structure, as well as directly accounting for reflection symmetry. Further, the usage of the MPNN layer leads to slight overfitting compared to the GATv2, illustrating how attentional GNNs strike a good balance of expressivity and data efficiency for this task. Our analysis highlights the quantitative benefits of both graph representation learning and geometric deep learning for football analytics from tracking data. We also provide a brief ablation study for the shot prediction task in S.Table 3.

Training Details We train each of TacticAI's models in isolation, using NVIDIA Tesla P100 GPUs. To minimise overfitting, each model's learning objective is regularised with an L_2 norm penalty with respect to the network parameters. During training, we use the Adam stochastic gradient descent optimiser [21] over the regularised loss. We summarise the hyperparameters of each TacticAI model in S.Table 1.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to licensing restrictions. However, contact details of the data providers are available from the corresponding authors on reasonable request.

S1.3. Supplementary Tables and Figures



补充图 1 | *t*-**SNE嵌入原始输入特征。** 对于在图 2中*t*-SNE可视化中使用的同一组角球样本，我们使用*t*-SNE可视化它们的原始输入特征。攻击和防守设置的*t*-SNE嵌入并不容易明确分离。

Acknowledgements

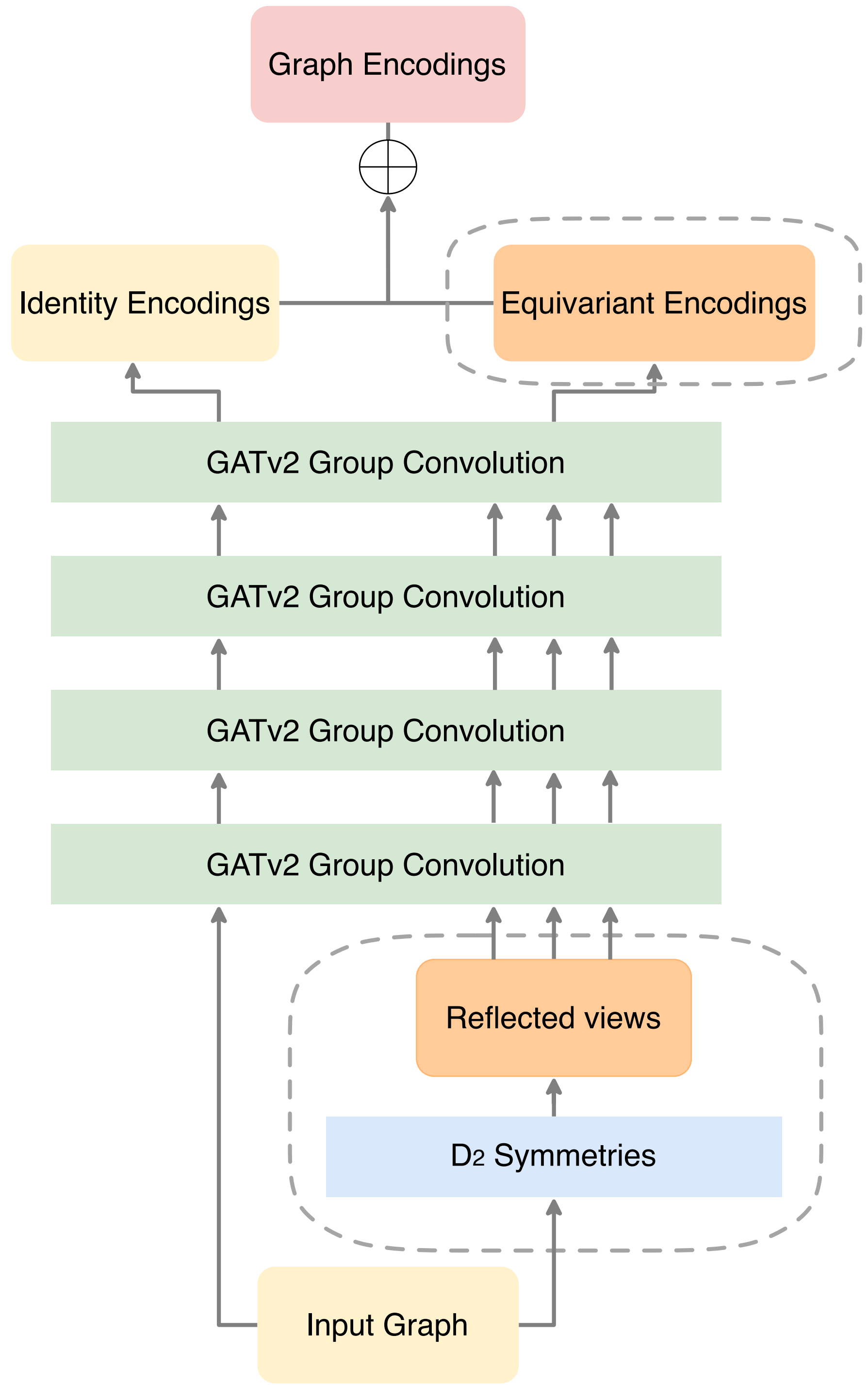
We gratefully acknowledge the support of James French, Timothy Waskett, Hans Leitert and Benjamin Hervey for their extensive efforts in analysing TacticAI's outputs. Further, we are thankful to Kevin McKee, Sherjil Ozair and Beatrice Bevilacqua for useful technical discussions, and Marc Lanctôt and Satinder Singh for reviewing the paper prior to submission.

Author contributions

Z.W., D.H.*, L.P. and K.T. coordinated and organised the research effort leading to this paper. P.V. and Z.W. developed the core TacticAI models. Z.W., W.S. and I.G. prepared the Premier League corner kick dataset used for training and evaluating these models. P.V., Z.W., D.H.* and N.T. designed the case study with human experts and performed the qualitative evaluation and analysis of its outcomes. All authors contributed to writing the paper, technical discussions and providing feedback on the final manuscript.

Supplementary Materials

*D.H. stands for Daniel Hennes.



补充图 2 | **TacticAI预测和生成组件模型的编码器架构。** 对于每个输入图，我们首先生成三个D_2-反射视图。其次，通过一系列四个GATv2 [4]组卷积层处理它们。实际层数可能会因不同任务而异（S.Table 1）。最后，我们通过加权和的方式将输入图（身份编码）及其对应反射视图（等变编码）的编码进行聚合，以获得最终的图编码。具体来说，对于接收器和射门预测任务（这些任务是完全不变的），我们取编码的平均值，对于引导生成任务（这是等变的），我们将身份编码的权重设置为1.0，并将等变编码置零。图编码会根据相应任务进一步处理（请参见方法部分中的网络架构）。

S1. Supplementary Material

S1.1. Literature Review

In the following, we summarise the closest state of the art that pursues the same objective of football corner kick tactics or pass prediction, or which more broadly devises team-sport strategy modelling techniques that could be transferred to the setting considered in this work. Sports analytics research that addresses player transfers and match statistics is out of scope, as we here focus on methods for in-game tactics.

Earlier work has generated a significant body of research on identifying key properties of a game given diverse sources of data, and on using these to provide tactical analysis and suggestions [8, 10, 15, 25, 28, 29]. One recent example is a system for determining whether a match is in-play or interrupted, called "in-game status'' , by applying random forests and AdaBoost to process time-continuous spatio-temporal data such as player positions [23], with similar methods having been proposed for American football [26].

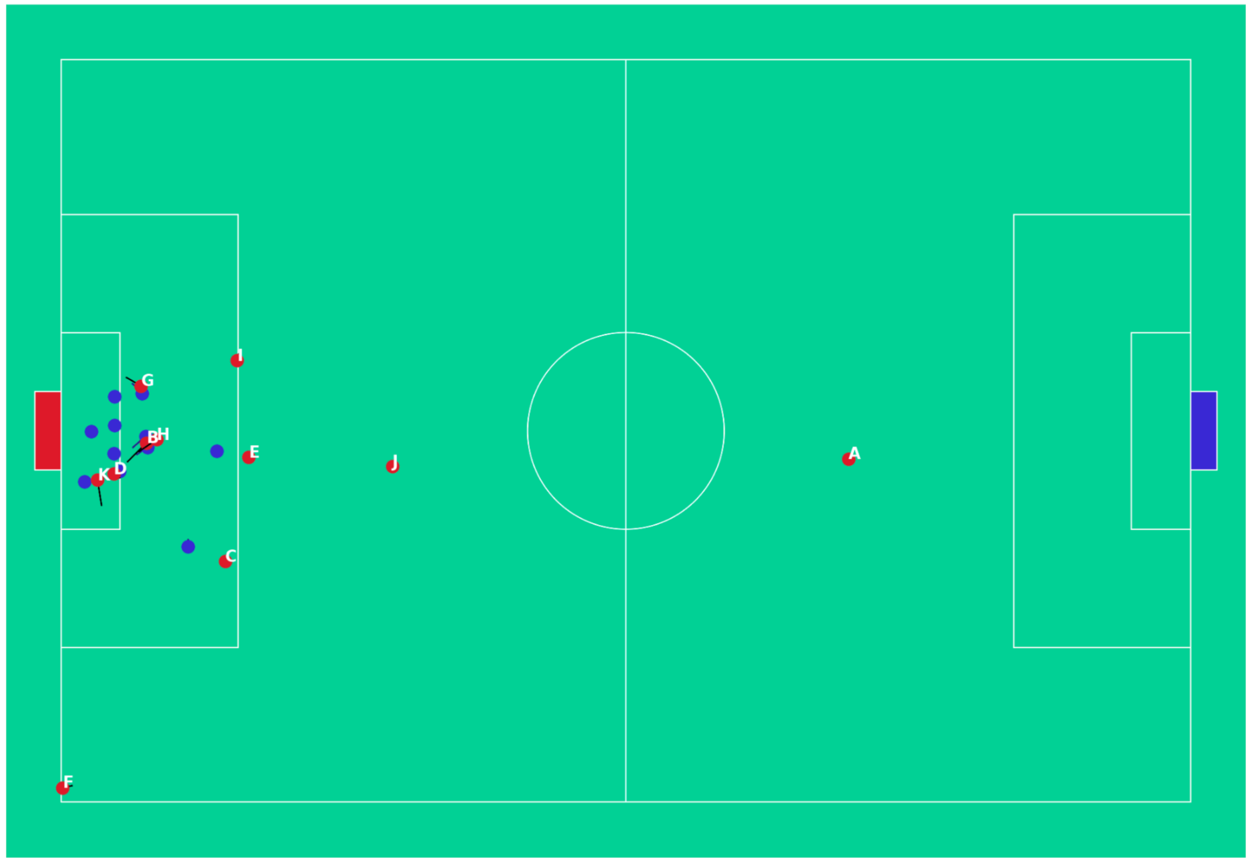
Other methods leverage game data to characterize players behaviour, such as their passing style [6]. Such information can later feed downstream applications that may take as input a representation of a player's historical behaviour. Similarly, pass prediction systems may use game data to predict the probability of a pass in a given game state [18].

A more closely related research line analyzes soccer videos to determine shoot event probabilities, which may serve as a proxy for player performance, by examining spatio-temporal relations and prediction uncertainty [16]. The method processes videos using a graph convolutional recurrent neural network to capture latent features, and leverages Bayesian neural networks to predict the shoot probabilities. Similar other methods apply machine learning and explainable AI techniques for tactical analysis and for predicting the success of defensive play [11].

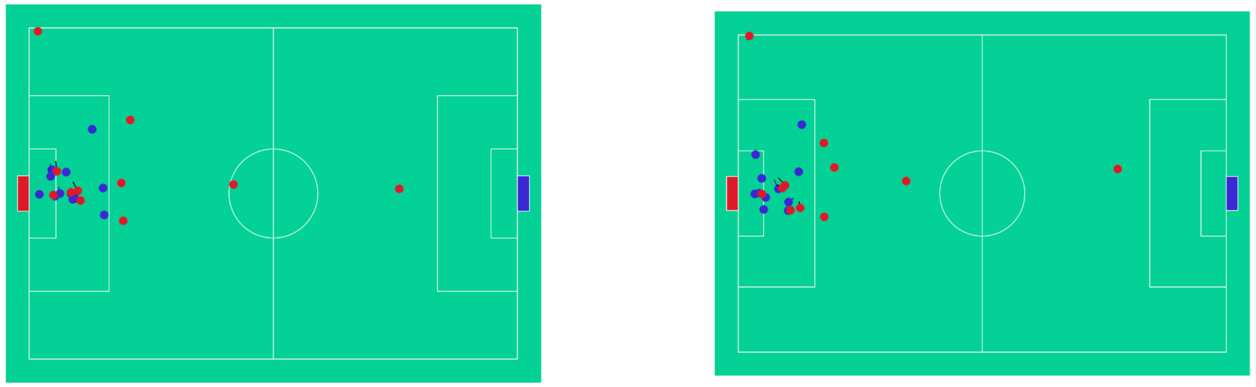
Accurately identifying the salient patterns of corner kick tactics from player trajectories is the key prerequisite for downstream tasks in football analytics, including evaluating the effectiveness of the tactics, improving their performance and designing counter-tactics. Through leveraging deep learning, previous research demonstrated that the outcomes of in-game events, including shots [37] and passes [9, 19], are predictable using the spatio-temporal tracking data of player trajectories. Moreover, the patterns of effective defensive tactics can also be recovered from tracking data [37]. Nevertheless, tracking data could be quite noisy because of the randomness and complex dynamics in football games, rendering mining the tactic patterns from it challenging.

There are also other examples of learning techniques for modelling team sports. The Generative Relational Inference Network (GRIN) [24] also follows the VAE framework to learn disentangled representations for each agent in the scene. It encodes each entity into a latent representation, comprising an intra-agent intention and an inter-agent relation, through a graph convolution network

TacticAI: an AI assistant for football tactics



补充图 3 | **案例研究中的任务1和任务2样本。**在此任务中，每个角球样本以2D“鸟瞰”视角呈现，其中蓝色和红色点分别代表防守和进攻球员，附在每个点上的箭头长度表示他们的速度。在执行任务期间，对于每个样本，我们首先请人类专家识别该样本是真实的角球还是由TacticAI生成的。接下来，他们被要求识别进攻队中最有可能接球的球员。



补充图 4 | **案例研究任务3的样本。** 在这项任务中，我们在每个任务样本中展示了一对角球，其中一个是作为参考角球的代表，另一个是作为检索到的角球，被视为在TacticAI潜在空间或原始特征空间中距离最近的样本。要求人类评分者根据显著模式判断这些对是否相似。

(GCN). The inter-agent relations under go message passing through graph attention layers, the result of which is combined with intra-agent intention to form the reconstruction. Although there is no group-invariant layers, experiments show that this method could discover disentangled factors from basketball players, and interventions on these factors produced interpretable variations in the predicted player trajectories.

Using NBA games spacio-temporal tracking data, DeepHoops [34] presents a Deep Learning architecture that estimates the impact of all micro-actions in on the outcome of Basket Ball plays. In particular, it allows to evaluate the contribution of individual off-ball events to the success of a possession.

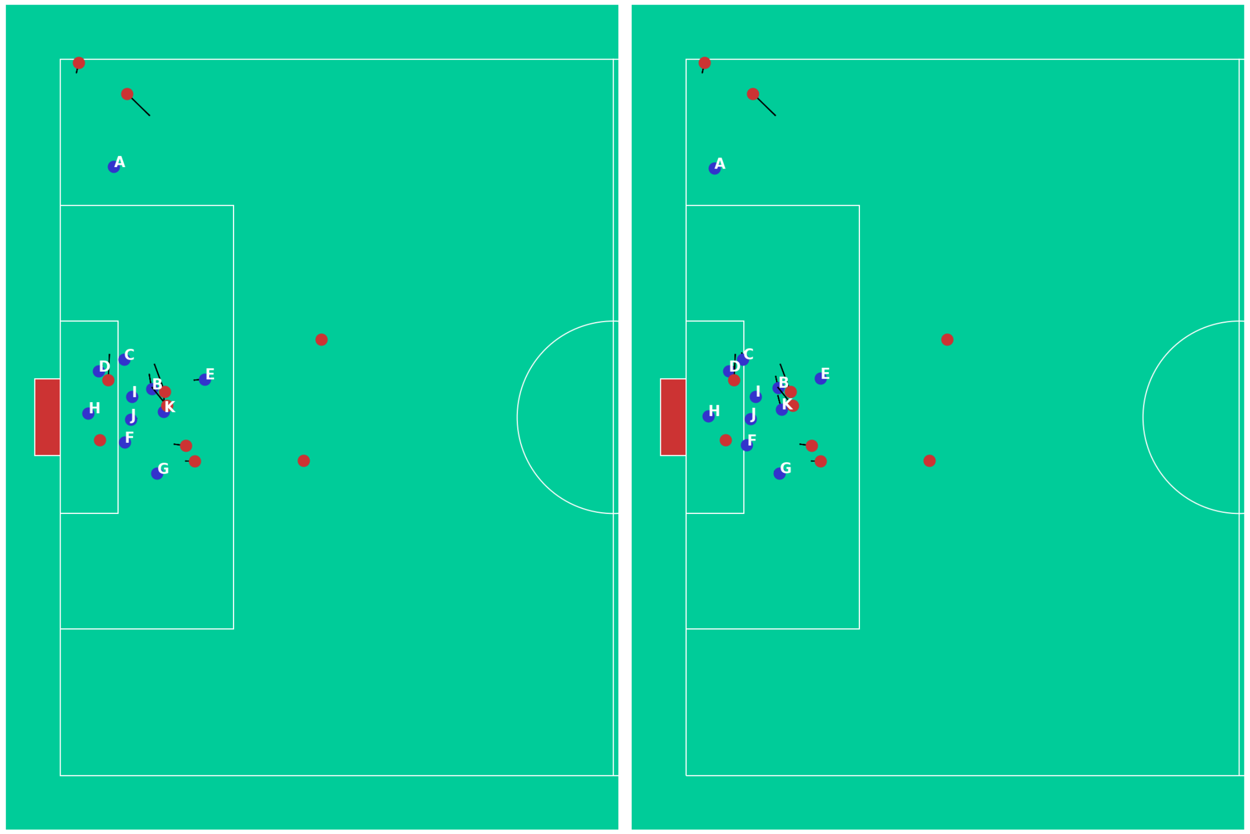
In [35] a tree-search method (Kernel Regression UCT) has been applied to discover curling strategies in self-play. The authors show that counterfactual justifications of actions taken by the model can be used to teach humans non-trivial concepts of curling as demonstrated in a user-study.

S1.2. Case Study Design

As previously described in the Results section, we designed an in-depth case study to evaluate the benefits of TacticAI's predictive and generative capabilities. We designed four specific tasks which we offered to five experts affiliated with Liverpool FC. To avoid biasing our raters, we have kept all details of these tasks strictly hidden from them prior to conducting the case study. Further, for all corner situations shown we did not reveal the team or player identities—only the players' positions and velocities at the moment the corner kick was taken, and whether they are on the attacking or defensive team.

Task 1: Does TacticAI generate *realistic* tactical adjustments? To answer this question, we started from a dataset of 50 corners from the Premier League, which have been held-out from TacticAI's training set (S.Figure 3). For a subset of those corners, we have applied TacticAI's generative head to propose alternate tactics for either minimising or maximising shot probability, and used TacticAI's suggestion in place of the original corner. We then asked our raters to judge whether each of these corners was real or generated. Through discussions with the raters after the task was done, and observing their comments, we noted a variety of strategies were employed to detect realism:

- Most raters chose to decide realism based on detecting whether there was something markedly unusual in the setup, for example a strategy that left opposing players in clear space.
- An alternate strategy, employed by one rater, was to decide realism based on whether the situation could have plausibly happened, even if it had a clearly suboptimal tactical setup. This approach ended up tagging most of the examples as real.
- Raters who specialise in corner kick analysis preferred a "retrieval" based strategy, wherein they would rate a corner as real if they were able to recall from memory that exact setup.



补充图 5 | **案例研究任务4的样本。** 在这项任务中，人类评分员被提供成对的角球样本，每对样本包括一个参考角球样本和由TacticAI生成的调整版本。然后，评分员需要识别调整版本中是否有显著的改进，并就这些改进给出具体的说明。对于每对样本中的调整，如果评分员确认整体调整是有建设性的，我们将其评为+1；如果没有明显差异，则评为0；如果是有破坏性的，则评为−1。

In spite of this variety of approaches, and as discussed in the Results section, all of the raters were generally confused as to which samples were real, and they unanimously highlighted how challenging they found this task.

Task 2: Does TacticAI predict *plausible* receivers? To answer this question, we reused the same 50 corners leveraged for Task 1, and simultaneously asked our raters which of the attacking team's players were most likely to make first contact with the ball (S.Figure 3). The raters were allowed to provide as little or as many players as they wanted to—and we cross-referenced their proposals with TacticAI's top three suggested receivers. Generally, we found significant variability in the amount of players proposed by different raters—some opted to suggest 3–5 receivers on almost all situations studied, whereas others did not want to suggest more than one receiver on most occasions. As discussed in the Results section, in spite of this diversity of approach, the top-3 predictions made by TacticAI largely agreed with the rater estimates, and there was no significant difference between TacticAI's accuracy over real and generated setups, indicating that TacticAI could be robustly used to predictively analyse both real and simulated corner kicks. And, as highlighted by several of the raters, such a system would prove useful to a football analyst: a receiver predictor can be used to indicate which players may need to be especially targeted defensively, as well as to evaluate whether corners in already-played games have unfolded in a manner which is expected.

Task 3: Can TacticAI be used to retrieve usefully *similar* corners? To answer this question, we leveraged 50 reference corners from the Premier League, held-out from TacticAI's training set (S.Figure 4). For each of these corners, we retrieved one other Premier League corner that was closest to it—either in TacticAI's graph-level embedding space, or in terms of the distance of the raw player features in the input space. This decision allowed us to not only measure whether TacticAI's embeddings can be used to mine for similar corners, but also to assess whether they provide an edge over a simpler mining heuristic which is purely position based. We asked our raters whether they find the retrieved corner to be usefully similar to the reference corner; that is, whether showing these two corners side by side would be judged as useful. The raters once again assumed a diverse set of approaches for deciding which features of the two corners were salient enough to be useful. Some of the common considerations mentioned by the raters included: whether the corner is in- or out-swinging, existence of a short corner option, zonal vs. man-marking approach, counting the number of players in the 18-yard box, positioning of the goalkeeper, and assessing whether the two corners likely had the same attacking or defending team, or had come from the same game. Similarly as for the previous tasks—from the discussion in the Results section we can conclude that, despite the wide variety of salient features considered, the raters generally found TacticAI's retrievals to be superior to the positional heuristic. They also highlighted the high utility of such a retrieval system: it can be used by analysts and coaches to discover and prepare for other teams' common routines, as well as to discover new ideas and variations on a particular corner that they might otherwise have

Hyperparameter	**Receiver Prediction**	**Shot Prediction**	**Guided Generation**
Batch size	256	128	128
Learning rate	1e-4	1e-4	5e-5
L^2 weight decay	1e-4	0.	1e-4
Adam optimiser β_1	0.9	0.9	0.9
Adam optimiser β_2	0.999	0.999	0.999
Adam optimiser ϵ	1e-8	1e-8	1e-8
# Graph attention layers	4	2	2
Random seed	42	42	42

Table 1 | **训练TacticAI各组件的超参数。** 我们列出了用于获得TacticAI最佳性能组件模型的超参数。这些模型是根据它们的评估损失进行选择的。特别是，对于引导生成，攻击和防御生成被训练为两个模型，并且共享同一组超参数。

Model	**Average Top-3 Accuracy**
CNN [9]	0.364 ± 0.031
Deep Sets [44]	0.713 ± 0.022
MPNN [14]	0.723 ± 0.017
GATv2 [43, 4]	0.748 ± 0.021
GATv2 + D_2 frame averaging	0.780 ± 0.011
GATv2 + D_2 group convolution [7]	**0.782** ± 0.039

Table 2 | **接收器预测的消融结果。** 我们使用前3名准确率作为评价指标。首先，我们发现使用Deep Sets [44]的图表示方法优于没有利用图表示的卷积神经网络（CNN [9]）。其次，不同于仅使用Deep Sets [44]在孤立模式下对每个球员节点建模而不考虑相邻信息，我们为每对球员节点添加一条边来表示它们是否在同一队，并使用各种图神经网络（GNNs）（MPNN [14] 和 GATv2 [4]）处理得到的完全连接图。这使得使用GATv2 [4]进行前3名预测的准确率又提高了0.035。最后，为GATv2配备D_2群卷积，我们为接收器预测构建了表现最佳的模型。

overlooked. It has also been stressed that the "top-1" retrieval study employed here is likely too strict. While it was necessary to restrict attention to only one retrieved corner here in order to keep the raters' workload manageable, in practice, an analyst may wish to analyse up to 10–20 retrieved corners for every reference corner. Given that TacticAI already has a favourable score in the top-1 regime, we anticipate its utility to only compound if deployed for real-world use.

Task 4: Does TacticAI generate *useful* tactical adjustments? To answer this question, we once again gathered a dataset of 50 held-out corners from the Premier League (S.Figure 5). For each of them, we have used TacticAI's guided generative model to propose adjustments to the defending team's player positions and velocities—leaving the attacking team unchanged—such that the predicted shot probability is reduced. We chose to focus on the defending team to match a real-world workload: typically, a corner setup is dictated by the attacking team, after which the defensive team needs to respond to this setup in an optimal way. For each of these corners, we showed the raters the two versions (original and TacticAI-adjusted) side-by-side. We stated that the corner on the left-hand side is a reference corner, asking the raters to judge which of the defending team's players are in a better or worse position, considering the adjustments on the right-hand side. We also asked the raters to judge whether the right-hand side corner is, overall, a better or worse situation than the left-hand side for the defensive team. To control for bias, we randomly select a subset of corner setups for which we reverse the two versions, making the TacticAI suggestion the reference corner. The raters focused on a variety of salient features in the suggestions, including but not limited to: whether defenders are better at tracking an attacker's run, whether it is better for them to move or stand still, and the positioning of the goalkeeper. There have been two situations where an adjustment was judged as "possibly useful", depending on the adjusted player's profile. For example, certain runs were only deemed likely to succeed if performed by defenders of high physical capability or fitness. As the TacticAI variant we trained here did not have access to player features beyond height and weight, such nuances were out of scope for this study, and we did not consider them to be salient observations. Overall, as outlined in the Results section, the raters were overwhelmingly in favour of TacticAI's suggestions, demonstrating a high degree of inter-rater agreement. We illustrate four of the most salient situations where TacticAI's suggestions were deemed significant in Figure 5; in one of these situations, **ten** defensive players were considered usefully adjusted, which is nearly the entire defending team! Further, this kind of position-adjustment system has been deemed a highly useful tool for an analyst or coach to leverage by all of our raters. Even the opportunity to view such suggestions side-by-side with a reference corner was considered to be useful, as it prompted the rater to immediately consider strategic variations. Since the guided generation system was deliberately designed to offer subtle modifications to player coordinates and velocities, the raters deemed such a system most useful to help detect players that have likely not been following the instructions of their coaches, and neglected the tactical play. This can then lead to targeted interventions by the coaches—either to exploit a weakness posed by an opposing player that tends to ignore instructions,

Model	**Average F_1 Score**
GATv2 [39, 4] (unconditional)	0.521 ± 0.027
GATv2 + receiver conditional	0.677 ± 0.036
GATv2 + receiver conditional + D_2 group convolution [7]	$\mathbf{0.712} \pm 0.011$

Table 3 | **射门预测的消融结果。** 我们使用F_1得分作为评价指标，因为用于开发射门预测组件模型的数据集是失衡的，正面到负面的比例为0.21。我们没有直接预测成功射门尝试的无条件概率，而是预测角球接球手的射门概率条件，这使得F_1得分从0.512 ± 0.027提升到0.712 ± 0.011。

References

[1] Duarte Araújo and Keith Davids. Team Synergies in Sport: Theory and Measures. *Frontiers in Psychology*, 7, 2016.

[2] Moez Baccouche, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt. Action classification in soccer videos with long short-term memory recurrent neural networks. In *International Conference on Artificial Neural Networks*, pages 154–159. Springer, 2010.

[3] Aseem Baranwal, Aukosh Jagannath, and Kimon Fountoulakis. Optimality of Message-Passing Architectures for Sparse Graphs. *arXiv preprint arXiv:2305.10391*, 2023.

[4] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021.

[5] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, 2021.

[6] Hyeonah Cho, Hyunyoung Ryu, and Minseok Song. Pass2vec: Analyzing soccer players' passing style using deep learning. *International Journal of Sports Science & Coaching*, 17(2):355–365, 2022.

[7] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

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

[9] Javier Fernández and Luke Bornn. SoccerMap: A Deep Learning Architecture for Visually-Interpretable Analysis in Soccer. *CoRR*, abs/2010.10202, 2020.

or to improve coaching of the team's player if they are found to be neglecting the determined tactics. Coupled with the positive outcome of Task 1, and the finding that the suggested adjustments are generally hard to distinguish from real corner situations, we conclude Task 4's question to be answered in the affirmative.

[10] Javier Fernández, Luke Bornn, and Dan Cervone. Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In *13th MIT Sloan Sports Analytics Conference*, 2019.

[11] Leander Forcher, Tobias Beckmann, Oliver Wohak, Christian Romeike, Ferdinand Graf, and Stefan Altmann. Prediction of defensive success in elite soccer using machine learning-Tactical analysis of defensive play using tracking data and explainable AI. *Science and Medicine in Football*, pages 1–16, 2023.

[12] Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. SE(3)-transformers: 3D roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems*, 33:1970–1981, 2020.

[13] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

[14] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural Message Passing for Quantum Chemistry. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 06–11 Aug 2017.

[15] FR Goes, LA Meerhoff, MJO Bueno, DM Rodrigues, FA Moura, MS Brink, MT Elferink-Gemser, AJ Knobbe, SA Cunha, RS Torres, et al. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science*, 21(4):481–496, 2021.

[16] Ryota Goka, Yuya Moroto, Keisuke Maeda, Takahiro Ogawa, and Miki Haseyama. Prediction of Shooting Events in Soccer Videos Using Complete Bipartite Graphs and Players' Spatial-Temporal Relations. *Sensors*, 23(9):4506, 2023.

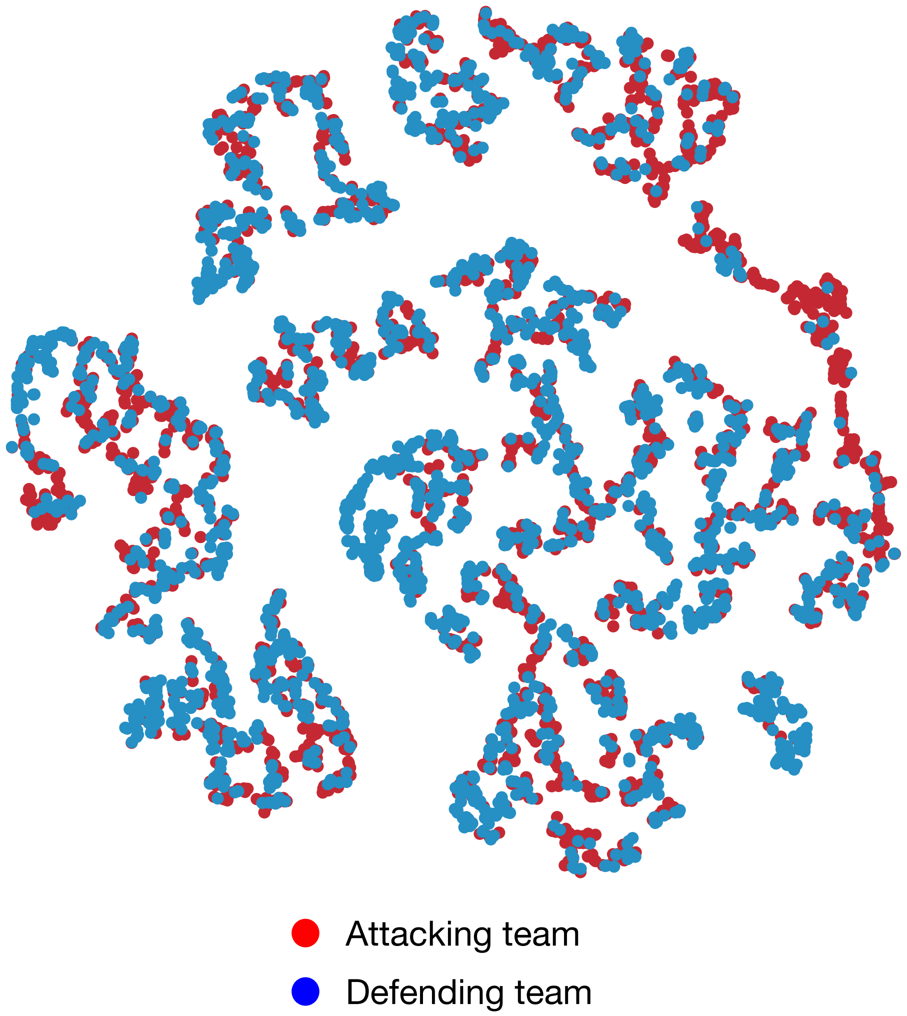
[17] Richard D Greenblatt, Donald E Eastlake III, and Stephen D Crocker. The Greenblatt chess program. In *Proceedings of the November 14-16, 1967, fall joint computer conference*, pages 801–810, 1967.

[18] Yutaro Honda, Rei Kawakami, Ryota Yoshihashi, Kenta Kato, and Takeshi Naemura. Pass Receiver Prediction in Soccer using Video and Players' Trajectories. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3502–3511, 2022.

[19] Ondřej Hubáček, Gustav Sourek, and Filip Železný. Deep Learning from Spatial Relations for Soccer Pass Prediction. In *MLSA@PKDD/ECML*, 2018.

[20] The International Football Association Board (IFAB). Laws of the Game, 2023.

S1.3. Supplementary Tables and Figures



Supplementary Figure 1 | ***t*-SNE embeddings of raw input features.** For the same set of corner kick samples used in the *t*-SNE visualisation in Figure 2, we visualise their raw input features with *t*-SNE. The *t*-SNE embeddings of the attacking and defending setups are not clearly separable.

[21] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In Yoshua Bengio and Yann LeCun, editors, *ICLR (Poster)*, 2015.

[22] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[23] Steffen Lang, Raphael Wild, Alexander Isenko, and Daniel Link. Predicting the in-game status in soccer with machine learning using spatiotemporal player tracking data. *Scientific Reports*, 12(1):16291, 2022.

[24] Longyuan Li, Jian Yao, Li Wenliang, Tong He, Tianjun Xiao, Junchi Yan, David Wipf, and Zheng Zhang. GRIN: Generative Relation and Intention Network for Multi-agent Trajectory Prediction. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 27107–27118. Curran Associates, Inc., 2021.

[25] Benedict Low, Diogo Coutinho, Bruno Gonçalves, Robert Rein, Daniel Memmert, and Jaime Sampaio. A systematic review of collective tactical behaviours in football using positional data. *Sports Medicine*, 50:343–385, 2020.

[26] Jacob Newman, Andrew Sumsion, Shad Torrie, and Dah-Jye Lee. Automated Pre-Play Analysis of American Football Formations Using Deep Learning. *Electronics*, 12(3):726, 2023.

[27] Shayegan Omidshafiei, Daniel Hennes, Marta Garnelo, Zhe Wang, Adria Recasens, Eugene Tarassov, Yi Yang, Romuald Elie, Jerome T Connor, Paul Muller, et al. Multiagent off-screen behavior prediction in football. *Scientific reports*, 12(1):8638, 2022.

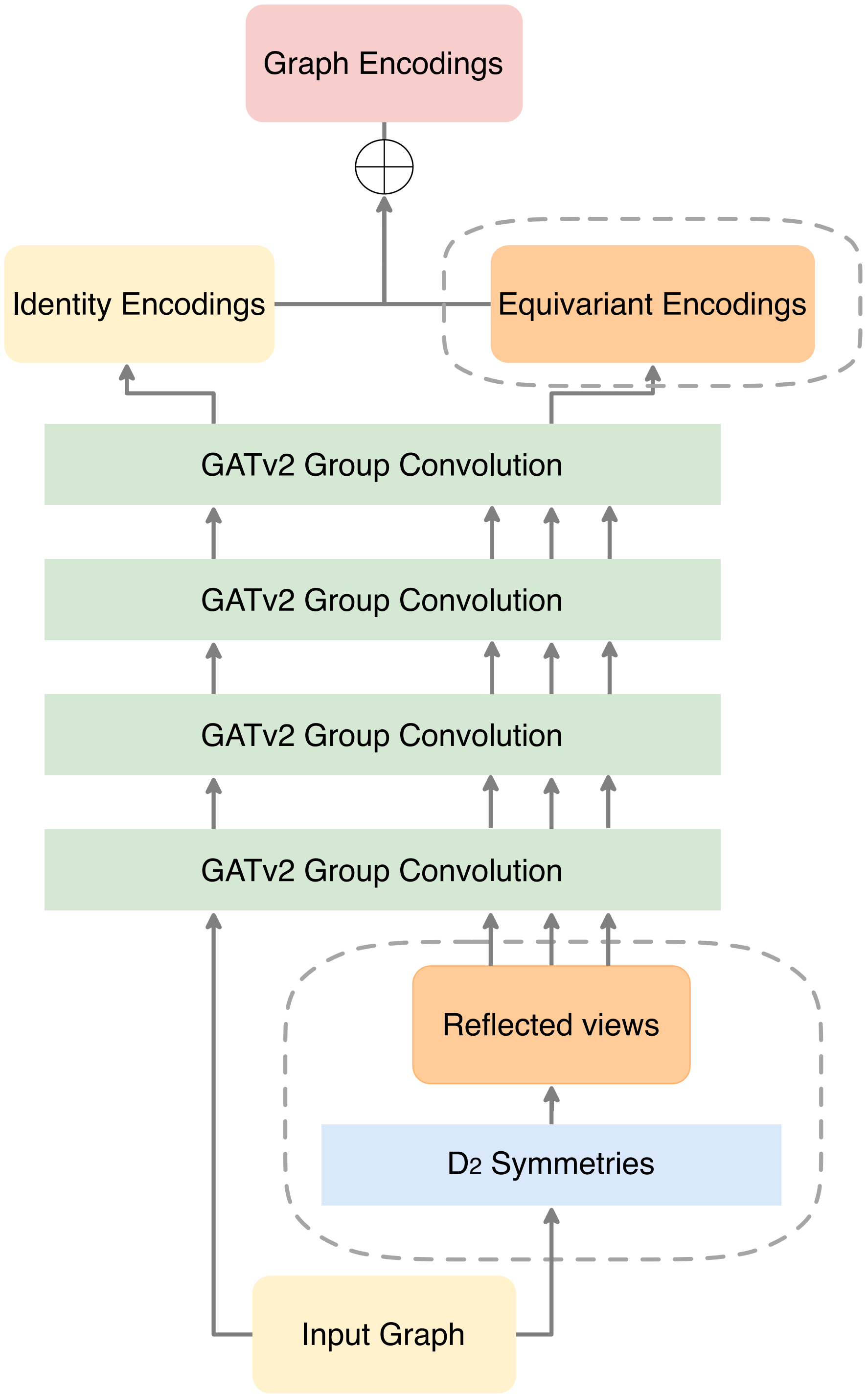
[28] Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. *Scientific data*, 6(1):236, 2019.

[29] Robert Rein and Daniel Memmert. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, 5(1):1–13, 2016.

[30] Samriddha Sanyal. Who will receive the ball? Predicting pass recipient in soccer videos. *Journal of Visual Communication and Image Representation*, 78:103190, 2021.

[31] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021.

[32] Martin Schijf, L Victor Allis, and Jos WHM Uiterwijk. Proof-number search and transpositions. *ICGA Journal*, 17(2):63–74, 1994.



Supplementary Figure 2 | **The encoder architecture of TacticAI's predictive and generative component models.** For each input graph, we first generate three D_2-reflected views. Secondly, we process them via a sequence of four GATv2 [4] group convolution layers. The actual number of layers may be different in different tasks (S.Table 1). Finally, we aggregate the encodings of the input graph (identity encodings) and the encodings of its corresponding reflected views (equivariant encodings) through a weighted sum to obtain the final graph encodings. Specifically, for the receiver and shot prediction tasks (which are fully invariant), we take the mean of the encodings, and for the guided generation task (which is equivariant), we set the weight of the identity encodings to 1.0 and zero out the equivariant encodings. The graph encodings are further processed with respect to the corresponding tasks (see Network Architectures in the Methods section.)

[33] Laurie Shaw and Sudarshan Gopaladesikan. Routine inspection: A playbook for corner kicks. In *Machine Learning and Data Mining for Sports Analytics: 7th International Workshop, MLSA 2020, Co-located with ECML/PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings 7*, pages 3–16. Springer, 2020.

[34] Anthony Sicilia, Konstantinos Pelechrinis, and Kirk Goldsberry. Deephoops: Evaluating micro-actions in basketball using deep feature representations of spatio-temporal data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2096–2104, 2019.

[35] Cleyton R Silva, Michael Bowling, and Levi HS Lelis. Teaching people by justifying tree search decisions: An empirical study in curling. *Journal of Artificial Intelligence Research*, 72:1083–1102, 2021.

[36] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of Go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

[37] Michael Stöckl, Thomas Seidl, Daniel Marley, and Paul Power. Making Offensive Play Predictable -Using a Graph Convolutional Network to Understand Defensive Performance in Soccer. 04 2021.

[38] Karl Tuyls, Shayegan Omidshafiei, Paul Muller, Zhe Wang, Jerome Connor, Daniel Hennes, Ian Graham, William Spearman, Tim Waskett, Dafydd Steel, et al. Game Plan: What AI can do for Football, and What Football can do for AI. *Journal of Artificial Intelligence Research*, 71:41–88, 2021.

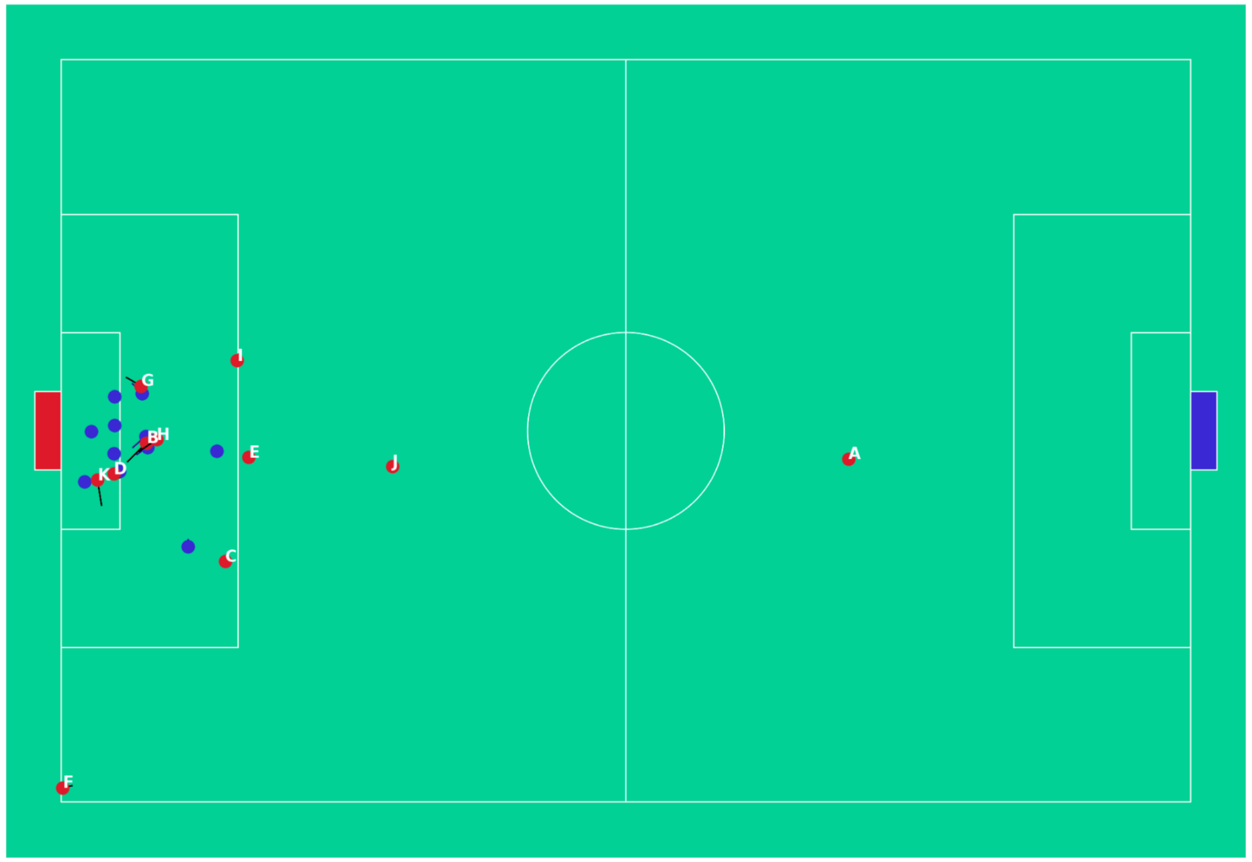
[39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[40] Petar Veličković. Message passing all the way up. *arXiv preprint arXiv:2202.11097*, 2022.

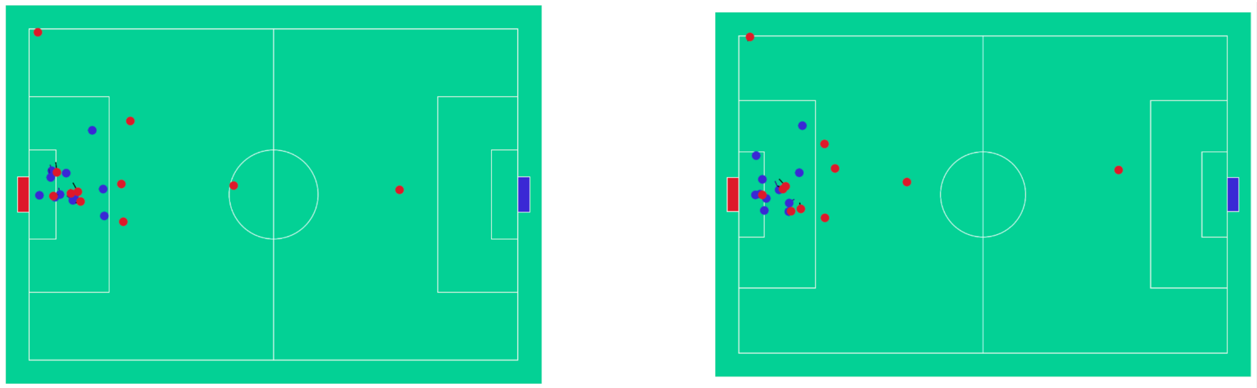
[41] Petar Veličković. Everything is connected: Graph neural networks. *Current Opinion in Structural Biology*, 79:102538, 2023.

[42] Petar Veličković, Adrià Puigdomènech Badia, David Budden, Razvan Pascanu, Andrea Banino, Misha Dashevskiy, Raia Hadsell, and Charles Blundell. The CLRS algorithmic reasoning benchmark. In *International Conference on Machine Learning*, pages 22084–22102. PMLR, 2022.

[43] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *6th International Conference on Learning Representations*, 2017.

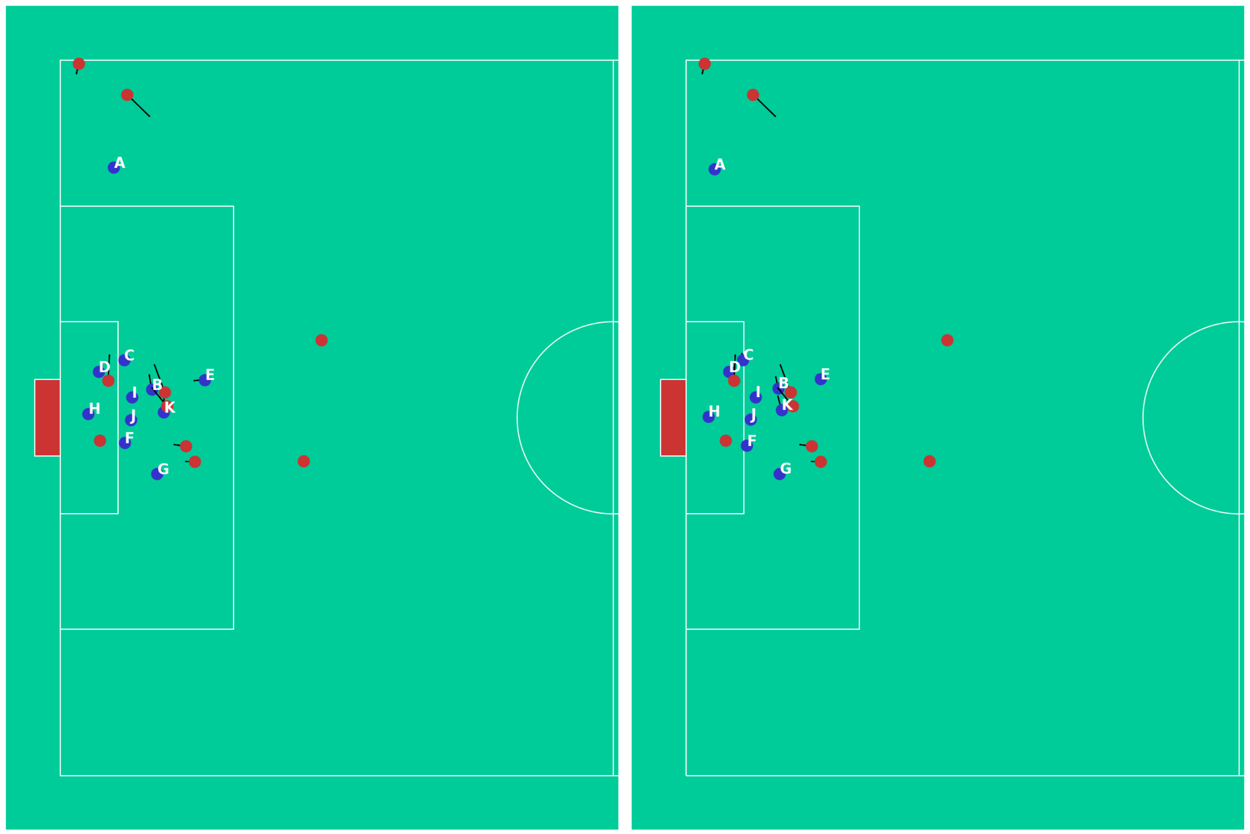


Supplementary Figure 3 | **Sample of Task 1 and 2 in the case study.** Each corner kick sample in this task is presented in a 2D "bird's eye" view, in which blue and red dots represent defending and attacking players, respectively, and the length of the arrow attached to each dot indicates their velocity. During the task, for each sample, we first ask the human experts to identify whether the sample is a real corner kick or a generated one by TacticAI. Next, they are asked to identify the most likely receivers among the attacking team's players.



Supplementary Figure 4 | **Sample of Task 3 in the case study.** In this task, we present a pair of corner kicks in each task sample, one of which represents the reference corner kick, and the other represents a retrieved corner, taken as the closest sample in distance, either in the TacticAI latent space or in the raw feature space. The human raters are asked to rate whether the pairs are similar with respect salient patterns.

[44] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep Sets. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.



Supplementary Figure 5 | **Sample of Task 4 in the case study.** In this task, the human raters are given pairs of corner kick samples, each of which consists of a reference corner kick sample and its adjustment generated by TacticAI. Then, the raters are asked to identify whether there are salient improvements in the adjustment and also to give specific account to those improvements as well. For the adjustment in each pair, we rate it as +1 if the rater confirms that overall adjustments are constructive, 0 if no distinct difference, or −1 if destructive.

Hyperparameter	Receiver Prediction	Shot Prediction	Guided Generation
Batch size	256	128	128
Learning rate	1e-4	1e-4	5e-5
L^2 weight decay	1e-4	0.	1e-4
Adam optimiser β_1	0.9	0.9	0.9
Adam optimiser β_2	0.999	0.999	0.999
Adam optimiser ϵ	1e-8	1e-8	1e-8
# Graph attention layers	4	2	2
Random seed	42	42	42

Table 1 | **Hyperparameters for training TacticAI's components.** We list the hyperparameters with which we obtain TacticAI's best performing component models. The models are selected according to their evaluation losses. In particular, for guided generation, attacking and defensive generations are trained as two models, and share a same set of hyperparameters.

Model	Average Top-3 Accuracy
CNN [9]	0.364 ± 0.031
Deep Sets [44]	0.713 ± 0.022
MPNN [14]	0.723 ± 0.017
GATv2 [43, 4]	0.748 ± 0.021
GATv2 + D_2 frame averaging	0.780 ± 0.011
GATv2 + D_2 group convolution [7]	**0.782** ± 0.039

Table 2 | **Ablation results for receiver prediction.** We use top-3 accuracy as the metric. First, we see that a graph representation with Deep Sets [44] outperforms a convolutional neural network(CNN [9]), which does not leverage a graph representation. Secondly, instead of modelling each player node in isolation without considering adjacency information with Deep Sets [44], we augment each pair of player nodes with an edge to indicate whether they are on the same team, and process the resulting fully connected graphs with various GNNs (MPNN [14] and GATv2 [4]). This yields another performance gain of 0.035 in Top-3 prediction accuracy with GATv2 [4]. Finally, equipping the GATv2 with a D_2 group convolution yields our best performing model for receiver prediction.

Model	Average F_1 Score
GATv2 [39, 4] (unconditional)	0.521 ± 0.027
GATv2 + receiver conditional	0.677 ± 0.036
GATv2 + receiver conditional + D_2 group convolution [7]	$\mathbf{0.712} \pm 0.011$

Table 3 | **Ablation results for shot prediction.** We use the F_1 score as the metric, because the dataset used to develop the shot prediction component model is imbalanced with a positive-to-negative ratio of 0.21. Instead of directly predicting the unconditional probability of a successful shot attempt, we predict the shot probability conditioned on the receiver of the corner kick, which yields an improvement from 0.512 ± 0.027 to 0.712 ± 0.011 in F_1 score.

References

[1] Duarte Araújo and Keith Davids. Team Synergies in Sport: Theory and Measures. *Frontiers in Psychology*, 7, 2016.

[2] Moez Baccouche, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt. Action classification in soccer videos with long short-term memory recurrent neural networks. In *International Conference on Artificial Neural Networks*, pages 154–159. Springer, 2010.

[3] Aseem Baranwal, Aukosh Jagannath, and Kimon Fountoulakis. Optimality of Message-Passing Architectures for Sparse Graphs. *arXiv preprint arXiv:2305.10391*, 2023.

[4] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021.

[5] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, 2021.

[6] Hyeonah Cho, Hyunyoung Ryu, and Minseok Song. Pass2vec: Analyzing soccer players' passing style using deep learning. *International Journal of Sports Science & Coaching*, 17(2):355–365, 2022.

[7] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

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

[9] Javier Fernández and Luke Bornn. SoccerMap: A Deep Learning Architecture for Visually-Interpretable Analysis in Soccer. *CoRR*, abs/2010.10202, 2020.

[10] Javier Fernández, Luke Bornn, and Dan Cervone. Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In *13th MIT Sloan Sports Analytics Conference*, 2019.

[11] Leander Forcher, Tobias Beckmann, Oliver Wohak, Christian Romeike, Ferdinand Graf, and Stefan Altmann. Prediction of defensive success in elite soccer using machine learning-Tactical analysis of defensive play using tracking data and explainable AI. *Science and Medicine in Football*, pages 1–16, 2023.

[12] Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. SE(3)-transformers: 3D roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems*, 33:1970–1981, 2020.

[13] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

[14] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural Message Passing for Quantum Chemistry. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 06–11 Aug 2017.

[15] FR Goes, LA Meerhoff, MJO Bueno, DM Rodrigues, FA Moura, MS Brink, MT Elferink-Gemser, AJ Knobbe, SA Cunha, RS Torres, et al. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science*, 21(4):481–496, 2021.

[16] Ryota Goka, Yuya Moroto, Keisuke Maeda, Takahiro Ogawa, and Miki Haseyama. Prediction of Shooting Events in Soccer Videos Using Complete Bipartite Graphs and Players' Spatial-Temporal Relations. *Sensors*, 23(9):4506, 2023.

[17] Richard D Greenblatt, Donald E Eastlake III, and Stephen D Crocker. The Greenblatt chess program. In *Proceedings of the November 14-16, 1967, fall joint computer conference*, pages 801–810, 1967.

[18] Yutaro Honda, Rei Kawakami, Ryota Yoshihashi, Kenta Kato, and Takeshi Naemura. Pass Receiver Prediction in Soccer using Video and Players' Trajectories. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3502–3511, 2022.

[19] Ondřej Hubáček, Gustav Sourek, and Filip Železný. Deep Learning from Spatial Relations for Soccer Pass Prediction. In *MLSA@PKDD/ECML*, 2018.

[20] The International Football Association Board (IFAB). Laws of the Game, 2023.

[21] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In Yoshua Bengio and Yann LeCun, editors, *ICLR (Poster)*, 2015.

[22] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[23] Steffen Lang, Raphael Wild, Alexander Isenko, and Daniel Link. Predicting the in-game status in soccer with machine learning using spatiotemporal player tracking data. *Scientific Reports*, 12(1):16291, 2022.

[24] Longyuan Li, Jian Yao, Li Wenliang, Tong He, Tianjun Xiao, Junchi Yan, David Wipf, and Zheng Zhang. GRIN: Generative Relation and Intention Network for Multi-agent Trajectory Prediction. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 27107–27118. Curran Associates, Inc., 2021.

[25] Benedict Low, Diogo Coutinho, Bruno Gonçalves, Robert Rein, Daniel Memmert, and Jaime Sampaio. A systematic review of collective tactical behaviours in football using positional data. *Sports Medicine*, 50:343–385, 2020.

[26] Jacob Newman, Andrew Sumsion, Shad Torrie, and Dah-Jye Lee. Automated Pre-Play Analysis of American Football Formations Using Deep Learning. *Electronics*, 12(3):726, 2023.

[27] Shayegan Omidshafiei, Daniel Hennes, Marta Garnelo, Zhe Wang, Adria Recasens, Eugene Tarassov, Yi Yang, Romuald Elie, Jerome T Connor, Paul Muller, et al. Multiagent off-screen behavior prediction in football. *Scientific reports*, 12(1):8638, 2022.

[28] Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. *Scientific data*, 6(1):236, 2019.

[29] Robert Rein and Daniel Memmert. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, 5(1):1–13, 2016.

[30] Samriddha Sanyal. Who will receive the ball? Predicting pass recipient in soccer videos. *Journal of Visual Communication and Image Representation*, 78:103190, 2021.

[31] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021.

[32] Martin Schijf, L Victor Allis, and Jos WHM Uiterwijk. Proof-number search and transpositions. *ICGA Journal*, 17(2):63–74, 1994.

[33] Laurie Shaw and Sudarshan Gopaladesikan. Routine inspection: A playbook for corner kicks. In *Machine Learning and Data Mining for Sports Analytics: 7th International Workshop, MLSA 2020, Co-located with ECML/PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings 7*, pages 3–16. Springer, 2020.

[34] Anthony Sicilia, Konstantinos Pelechrinis, and Kirk Goldsberry. Deephoops: Evaluating micro-actions in basketball using deep feature representations of spatio-temporal data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2096–2104, 2019.

[35] Cleyton R Silva, Michael Bowling, and Levi HS Lelis. Teaching people by justifying tree search decisions: An empirical study in curling. *Journal of Artificial Intelligence Research*, 72:1083–1102, 2021.

[36] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of Go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

[37] Michael Stöckl, Thomas Seidl, Daniel Marley, and Paul Power. Making Offensive Play Predictable -Using a Graph Convolutional Network to Understand Defensive Performance in Soccer. 04 2021.

[38] Karl Tuyls, Shayegan Omidshafiei, Paul Muller, Zhe Wang, Jerome Connor, Daniel Hennes, Ian Graham, William Spearman, Tim Waskett, Dafydd Steel, et al. Game Plan: What AI can do for Football, and What Football can do for AI. *Journal of Artificial Intelligence Research*, 71:41–88, 2021.

[39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[40] Petar Veličković. Message passing all the way up. *arXiv preprint arXiv:2202.11097*, 2022.

[41] Petar Veličković. Everything is connected: Graph neural networks. *Current Opinion in Structural Biology*, 79:102538, 2023.

[42] Petar Veličković, Adrià Puigdomènech Badia, David Budden, Razvan Pascanu, Andrea Banino, Misha Dashevskiy, Raia Hadsell, and Charles Blundell. The CLRS algorithmic reasoning benchmark. In *International Conference on Machine Learning*, pages 22084–22102. PMLR, 2022.

[43] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *6th International Conference on Learning Representations*, 2017.

[44] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep Sets. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.