17z: A Spatiotemporal Foundation Model for Football Intelligence

17z Research · lab@17z.gg

v1 · July 2026

Abstract

The information that decides a football match lives in the fine spatiotemporal structure of how twenty-two players and a ball move through space and time — a resolution too high and a tempo too fast for any staff to hold in their heads. A generation of football analytics addressed this with narrow, hand-built models, one metric at a time. The frontier has since moved to *learned*, symmetry-aware models of the game, but that capability remains locked inside a handful of super-clubs with proprietary tracking data and in-house research labs. We present **17z**, a spatiotemporal foundation model for football: a single transformer over player-and-ball tracking data, with the geometric symmetries of the pitch built into its architecture, pretrained on simulation and open data and specialized through lightweight heads for structured concept extraction, state valuation, multi-agent action valuation, and embedding-based retrieval. Because pretraining requires no proprietary data, a club adopts the model by connecting its own tracking on its own infrastructure — the data never leaves the building — collapsing the cold-start that has kept advanced football intelligence out of reach for all but the richest clubs. This paper sets out the model, the training regime, the equivariance argument for sample efficiency, the bring-your-own-data deployment model, and our evaluation program. It is a design report: empirical results are reported in subsequent versions.

1 Motivation

Football is decided in detail. The difference between a chance created and a chance conceded is often a half-second earlier trigger, a body angle, a run into space that no highlight reel records. This detail is not noise to be averaged away; it *is* the game. Measured at the resolution modern tracking systems now capture — the position of every player and the ball, roughly twenty-five times a second — a single match contains on the order of a hundred thousand frames of spatial configuration. No coaching staff can hold that in working memory, and the summary statistics that fit on a page discard most of what matters.

The first decade of football analytics coped by building a separate model for each question: expected goals for shot quality, possession-value surfaces for territory, bespoke pipelines for pressing or set pieces. Each model learned the game from scratch and captured a thin slice of it. The result is useful but fragmented, and it leaves most of the spatiotemporal signal on the floor.

A second shift is now underway. Rather than hand-engineering one metric at a time, the strongest clubs are training models that learn a general representation of play and reason over it. The clearest public example is TacticAI, a tactical assistant built by a national-scale AI lab in partnership with an elite club, which models corner kicks as graphs and exploits the reflective symmetry of the pitch to generalize from limited data.[4] The direction is unambiguous. The problem is access: this capability sits behind proprietary tracking contracts and research budgets that only a few clubs command.

That gap is structural, and it produces a cold-start that has stalled every independent attempt to close it. A model of football needs tracking data to be built; clubs will not share tracking data until a model has proven itself. 17z is designed to break that loop. Our thesis is that a foundation model can be *pretrained on non-proprietary data* — simulation and the limited open tracking that exists — and then *specialized on a club's own data, inside the club's own environment*. The intelligence becomes infrastructure a club plugs into, rather than a service it must first feed.

2 Background

Possession value. The idea of assigning a value to a game state — how likely the current situation is to lead to a goal — runs through modern analytics. Expected Threat (xT), popularized in its now-standard form by Singh,[1] lays a value surface over the pitch; the underlying formulation traces to earlier work by Rudd.[2] Frameworks such as VAEP extend valuation from locations to individual on-ball actions.[3] These remain the reference points for any learned value head, and we treat reproducing them on open data as a basic soundness check (§6).

Learned models of play. TacticAI demonstrates that geometric deep learning — encoding the pitch's symmetries directly into the network — yields models that are both more accurate and more data-efficient, and whose suggestions domain experts judged favorably against real tactics.[4] Sequence and representation models over tracking, such as multi-entity transformers for multi-agent trajectories,[5] and early "foundation model for soccer" efforts,[6] point toward a single pretrained backbone serving many tasks. 17z sits in this lineage and pushes it toward full spatiotemporal tracking and a deployment model built for clubs that are not at the frontier.

Tracking data. Optical and broadcast tracking record all players and the ball at 25–30 Hz. Formats are fragmented across providers, and standardization efforts are only now emerging.[7] This fragmentation is not incidental to our design: normalizing heterogeneous provider formats into one canonical representation is a first-class component of the system (§5).

3 The 17z Model

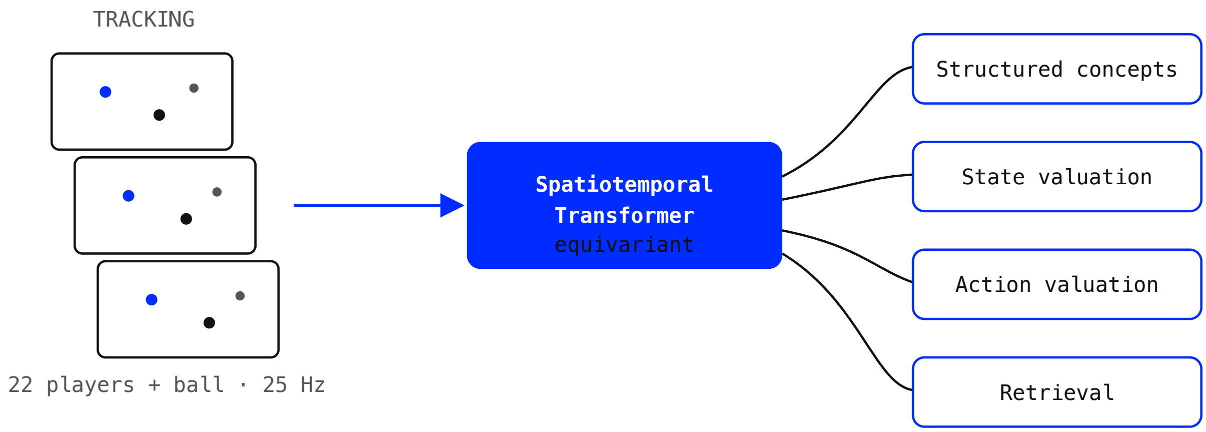

Figure 1: One backbone, many tasks. Tracking sequences are encoded by a spatiotemporal transformer with built-in geometric equivariances; lightweight heads specialize it for concepts, valuation, and retrieval.

3.1 Input representation

The model's view of a match is not video but geometry. Each player and the ball is an entity carrying position, velocity, and role context; a short window of play is a sequence of these configurations over time. We refer to this as a *spatiotemporal* representation — position *and* its evolution — because momentum, runs, and the build of pressure are legible only across frames, never in a single one. Entities are tokenized so that attention can operate jointly over players and over time.

3.2 Architecture and equivariance

The backbone is a transformer whose attention learns, for each entity, which other entities and which moments are relevant — the off-ball run that decides a play rather than the twenty-one other bodies on the pitch. Crucially, the pitch has symmetries: a move down the left is the mirror of the same move down the right; a phase of play is unchanged under swapping attacking direction or the reflection of the field. We constrain the network so that these transformations of the input produce correspondingly transformed outputs, following the group-equivariant approach that TacticAI applied to corner kicks, where predictions are forced to agree across all reflections of a situation.[4] A model without this prior must learn each mirrored situation independently; a model with it does not.

3.3 Pretraining on simulation and open data

Football's supervised data is scarce relative to the space of configurations a model must understand, and the richest tracking is precisely what is not publicly available. We therefore pretrain the backbone with self-supervised objectives — masked-trajectory and next-state prediction over unlabeled sequences — on two sources. *Simulation* supplies the volume: unlimited full twenty-two-agent spatiotemporal sequences that no public corpus can match. *Open tracking data*, though limited, grounds the model against how real players actually move, mitigating the sim-to-real gap that pure simulation would introduce. The objective is a backbone that has learned the structure of the game before it has seen a single club's data.

3.4 Task heads

1. **Structured concepts.** Interpretable tactical descriptions — pressing lines, overloads, formation and shape — extracted from the learned representation into the language a coach already uses.
2. **State valuation.** A learned value of the current game state, in the lineage of Expected Threat: how much this exact configuration favors one side. Reproducing xT-style behavior on open data is our first calibration target.
3. **Multi-agent action valuation.** The value of individual decisions — pass, carry, or shot — across all players, providing the basis for player evaluation and recruitment.
4. **Embedding-based retrieval.** Any moment is mapped to a fixed-length embedding such that similar moments lie nearby, turning "find every situation like this" into nearest-neighbor search over an entire season — without manual tagging.

4 Why Equivariance Matters

The equivariance built into §3.2 is not a cosmetic prior; it is the mechanism that makes learning from scarce football data tractable. By construction, a play and its mirror image are represented as the same object, so the model never spends capacity or examples relearning reflected situations. Geometric deep learning has repeatedly shown that respecting a domain's symmetries yields more generalizable models trained on less data, and TacticAI's results on real match data are a direct demonstration in football specifically.[4] Where an unconstrained model would need to see both wings, 17z learns each once — a decisive advantage in a domain where labeled examples are the binding constraint.

5 Deployment: Bring Your Own Data

Because the backbone is pretrained without proprietary data, adoption inverts the usual order. A club does not send us its data so we can build; it connects its data to a model that already works.

The adapter layer. Every provider — among them Second Spectrum, Hawk-Eye, SkillCorner, Sportec/DFL, and StatsBomb — uses a different schema, coordinate system, and frame rate. Normalizing these into one canonical spatiotemporal representation is what makes "just connect your data" real, and it is the component in which much of the system's practical value and defensibility resides.

On-premise inference and privacy. The pretrained model is delivered into the club's own environment. Inference runs there; the club's tracking never leaves its infrastructure. For organizations that guard their data as a competitive asset, this is not a feature but a precondition.

Calibration. The representation and retrieval capabilities transfer immediately. The valuation heads sharpen with a light fine-tune on the club's own realized outcomes, so that "how dangerous is this moment" is calibrated to that club's context rather than a generic prior.

6 Evaluation Program

This paper is a design report; the following is the program by which we are validating the model, not a claim of completed results.

1. **Value-head soundness.** Reproduce Expected Threat behavior on open tracking data and compare against the published xT surface as a sanity check.
2. **Retrieval quality.** Human-rated similarity of nearest-neighbor moments against held-out queries, measured for precision at small k.
3. **Valuation calibration.** Calibration of state- and action-value predictions against realized outcomes on held-out matches.
4. **Ablations.** Isolate the contribution of geometric equivariance and of simulation pretraining by removing each in turn, measuring the change in data efficiency and downstream accuracy.

Results from this program are reported in subsequent versions of this document.

7 Roadmap

We are building the system in order of evidence rather than breadth. First, the equivariant backbone and the retrieval head, trained on simulation and open data — the most demonstrable capability, and the one that requires no labeled outcomes to be useful. Second, the adapter layer that lets a real club connect its own data without it leaving the building. Third, the valuation heads, designed to work out of the box and improve with a club's outcomes. Fourth, the structured-concept layer and the surfaces through which a non-technical staff uses the model daily.

8 Conclusion

The frontier of football intelligence is a single learned model of the game, made efficient by the pitch's own symmetries. What has kept that frontier inside a few clubs is not the idea but the data. By pretraining on simulation and open data and deploying where the data already lives, 17z aims to make frontier football intelligence something any club can plug into — and to do the research in the open.

References

1. Singh, K. *Introducing Expected Threat (xT)*. 2018. karun.in/blog/expected-threat.html
2. Rudd, S. *A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains*. NESSIS, 2011.
3. Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. *Actions Speak Louder than Goals: Valuing Player Actions in Soccer*. Proc. KDD, 2019.
4. Wang, Z., Veličković, P., Hennes, D., et al. *TacticAI: an AI assistant for football tactics*. Nature Communications 15, 2024. nature.com/articles/s41467-024-45965-x
5. Alcorn, M. A., & Nguyen, A. *baller2vec: A Multi-Entity Transformer for Multi-Agent Spatiotemporal Modeling*. 2021.
6. Mendes-Neves, T., et al. *Towards a foundation large events model for soccer*. Machine Learning (Springer), 2024. link.springer.com/article/10.1007/s10994-024-06606-y
7. Bassek, M., et al. *An integrated dataset of spatiotemporal and event data in elite soccer*. Scientific Data, 2025. pmc.ncbi.nlm.nih.gov/articles/PMC11787359