RESEARCH REPORT · MAY 2026 · WAVE 1 + WAVE 2 UNIFIED

Attributes in, *values out* — when two videogames see the same player differently.

Wave 2 published May 2026 · 13,434 men · 385 women · 96 clubs across the top-five European leagues · the union of EA Sports FC 26 and Football Manager 26 attribute vectors as a single research instrument.

Pok Yeung Lee · 2026 · 12 sections · 3 chapters · 51 interactive figures · falsifying validation cleared

PART ONE OF THREE

01

Two instruments, examined separately and side by side

Sections §1 through §4 frame EA Sports FC 26 and Football Manager 26 as research instruments rather than entertainment products, then sit them side by side to show what the two attribute schemas record about the same population of 13,434 men.

§1 FRAMING

Why the games are the data

Two studios invest substantial scouting budgets; their attribute vectors are the public corpus.

§2 EXAMINING EA

What OVR actually is

Per-position OLS recovers EA's positional weighting formula at R^2 0.96 to 0.998.

§3 EXAMINING FM

What Current Ability means

Sports Interactive's 1,300-researcher network publishes a hidden mentals layer EA lacks.

§4 SIDE BY SIDE

Same player, two schemas

Mean ρ on the 19 same-construct pairs is 0.529 — moderate, not unity. The disagreements are the data product.

Football pretends to have public data. Open the JSON files and it doesn't.

Two video games are the only globally consistent, attribute-deep, cross-gender public datasets on professional footballers.

GLOBAL LABOUR MARKET

~$50B

Annual football labour market

Priced almost entirely on rumour. Transfermarkt is a crowd; CIES is a black box; StatsBomb is licensed.

EA SPORTS FC 26

17,569

Players, 45 leagues, monthly refresh

1,447 women. Broadcast-tuned 1–99 OVR. ~50-person committee + per-league panels.

FOOTBALL MANAGER 26

1.3M

Players, 116 nations, ~1,300 researchers

36,000+ women shipped from a standing start, November 2025. Match-engine-tuned.

THE FRAMING MOVE

The two games are the only **attribute-deep, cross-gender, free public datasets** on professional footballers. Wave 1 found EA's *attribute schema* — not its value field — is what makes valuation tractable. Wave 2 added Football Manager. **In football, the game is the dataset.**

Academic ceiling for attribute-based valuation: R^2 = 0.85–0.90 (McHale & Holmes, 2023; Yang, 2025).

EA's OVR is a positional weighted aggregation — per-position R^2 = 0.96–0.998

EA Sports FC 26 publishes a 17,569-player attribute database; OVR is the position-specific weighted sum of sub-attributes.

WHAT EA IS

- 17,569 men, 1,447 women, 45 leagues
- Annual September release + monthly Title Updates
- ~50-person EA Ratings committee, broadcast-tuned
- 36 visible attributes (6 composites + 30 sub-attributes)
- No hidden personality layer — that's FM's territory

THE FORMULA, REVERSE-ENGINEERED

$$\mathrm{OVR}_p(x) \approx \mathrm{round}\left(\mu_p + \sum w_{i,p} \cdot x_i \right)$$

$p \in$ {GK, DEF, MID, WIDE, FWD}

per-position OLS R^2 = 0.96–0.998

PER-POSITION OVR DECOMPOSITION FIT

OVR is deterministic, not opinion.
Given the sub-attribute vector and the position bucket, OVR is recoverable to within rounding.

FM26 exposes 13 hidden personality mentals — EA has zero analogue

Sports Interactive's 1,300-researcher network covers 1.3M players + 36k women in 116 countries; the hidden-mental block alone recovers R^2=0.195 against TM market value.

WHAT FM IS

DATABASE SCALE

1,300,000 men + 36,000 women

FM26, launched November 2025 on Unity engine

RESEARCH NETWORK

1,300 scouts · **116** countries

Part-time researchers under 86 full-time leads, in-person scouting

RELEASE CADENCE

2-year cycle

FM25 cancelled; legacy engine retired; full Unity rewrite

VISIBLE SCHEMA

36 attributes (1–20 raw) · **CA 1-200**

Current Ability internal 1–200; surfaced as 1–99 in-game

CALIBRATION

WITHIN-GENDER

1–20 scale is calibrated per SI documentation

THE HIDDEN BLOCK

13 HIDDEN PERSONALITY MENTALS

FM-ONLY

Adaptability · Ambition · Loyalty · Consistency · Temperament · Pressure
Important Matches · Versatility · Professionalism · Sportsmanship
Compliance · Fairness · Injury Resistance

9 VISIBLE MENTALS

FM GRANULAR · EA COARSE

Concentration · Decisions · Vision · Bravery · Determination
Off-the-Ball · Anticipation · Composure · Leadership

FM-EXCLUSIVE CONTRIBUTION

top-20 features in the union model, no EA analogue

RANK 15	RANK 16	RANK 20
Adaptability	**Injury Resistance**	**Ambition**

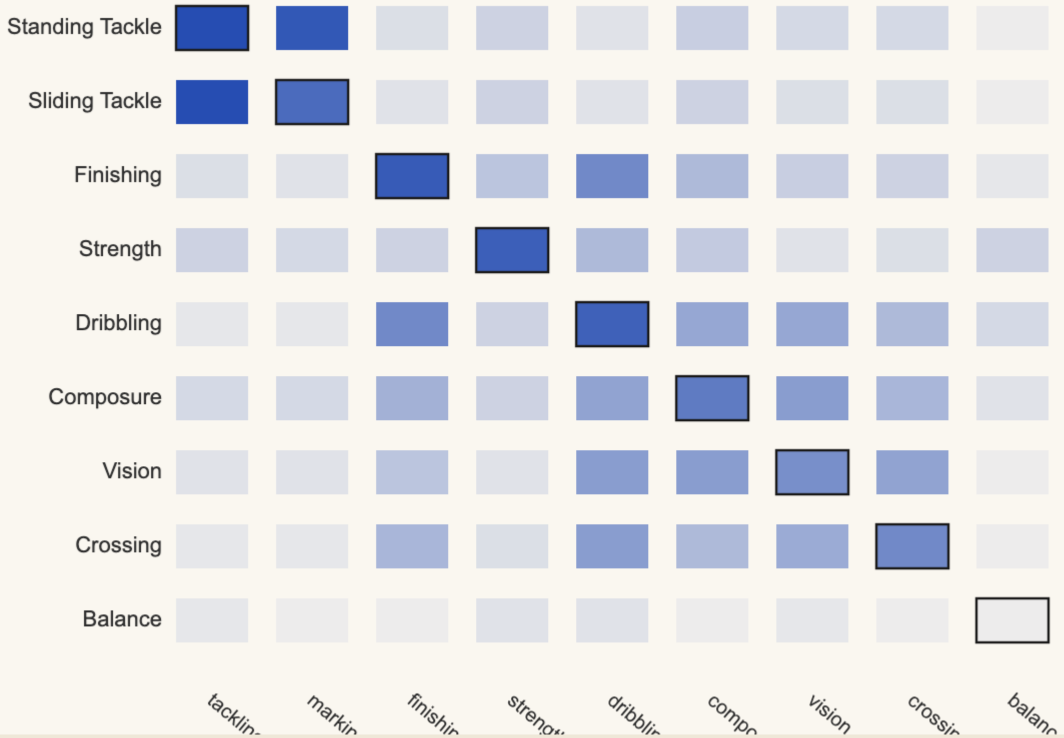
$R^2 = 0.195$

FM hidden mentals alone — no CA, no age, no reputation · $\log_{10}$(TM value) · n=6,729

EA and Football Manager agree on rank, disagree on shape

Mean Spearman ρ on the 19 same-named attribute pairs is 0.529 — moderate, not unity.

SHARED-NAME PAIR CORRELATIONS · KLEIN-BLUE = HIGHER ρ



READING THE DIAGONAL

ρ = 0.529

Mean Spearman across the 19 same-named pairs

Defense agrees tightest

Standing Tackle ↔ tackling ρ = 0.728. Tackles either succeed or not — broadcast-observable.

Finishing / physical agrees moderately

Finishing ρ = 0.683, Dribbling 0.646, Strength 0.665. Same concept, independent measurement noise.

Playmaking diverges sharply

Vision ρ = 0.399, Crossing 0.424, Balance `-0.04`. FM's researchers see training-ground signal EA structurally cannot see from broadcast.

WHY IT MATTERS

The 0.165 of unshared rank-order variance is what each database can see that the other can't.
The disagreements are the data product.

n = 11,943 outfield men with both vectors · recreated from b-fig-corr-matrix · Wave 2 §B.2

02

Building the model, and locating its ceiling

Sections §5 through §8 build the union model that fuses EA and FM features, validate it against Transfermarkt and CIES, name the three residuals that cap it at R^2 0.785, and ask whether the same construction transfers from men to women.

§5 · BUILDING THE UNION

Union beats EA-only by 12pp R^2

5-fold CV on n=6,729 matched: union 0.785, EA-only 0.663, FM-only 0.680. Orthogonalisation cleans attribution.

§6 · VALIDATION

Bracketed between TM and CIES

Two ground truths disagree by 122% on marquee; the attribute-only union sits between them.

§7 · CEILING

Three named residuals

Celebrity premium above €50M, price floor under €100K, structural rebase — each capped attribute-only models.

§8 · TRANSFERABILITY

Men's lens cannot see women

Within-gender FM26 hits R^2 0.91 on women's CA; the EA-trained men's model goes anti-signal at R^2 −0.10.

Train-test protocol: 5-fold cross-validation on the same 6,729 matched players

Every model below is scored on the same matched-corpus rows under one learner, one target, and one cross-validation protocol.

SAMPLE

n = 7,835

EA-only frame — every man with the full EA attribute vector and a non-zero Transfermarkt market value.

MATCHED CORPUS

n = 6,729

Union frame — every row carries EA + FM26 visible + FM26 hidden + TM, so every model is scored on the same rows.

CROSS-VALIDATION

5-fold KFold

shuffle=True, random_state=42. Held-out predictions via cross_val_predict, per-fold R^2 via cross_val_score.

LEARNER

HistGradientBoosting

scikit-learn HistGradientBoostingRegressor: max_iter=400, max_depth=8, learning_rate=0.05, random_state=42.

TARGET

log10(TM €)

Log-transformed Transfermarkt market value in euros. Errors are back-transformed where meaningful.

HARMONISATION

92.5 % exact

EA ↔ TM ↔ FM26 join on name + date-of-birth + club. The remainder manually disambiguated.

EA + FM26 union beats every model on every metric — fold envelopes don't overlap

Five-fold cross-validated R^2, RMSE, MAE, median APE, Spearman, and marquee-tier APE on log10(TM value), n=6,729 matched men.

MODEL	n	FEATURES	R² MEAN	FOLD ENVELOPE	RMSE-LOG	MAE-LOG	MEDIAN APE	SPEARMAN €	MARQUEE APE
EA-only	7,835	38	0.663	0.643–0.680	0.394	0.304	51.7%	0.794	38.7%
FM26-only	6,729	52	0.680	0.645–0.708	0.385	0.283	45.4%	0.804	33.7%
FM26 visible+meta	6,729	39	0.677	0.640–0.701	0.388	0.285	45.4%	0.799	33.5%
FM26 hidden-only	6,729	13	0.195	0.158–0.218	0.612	0.491	77.9%	0.423	91.8%
EA + FM26 union	**6,729**	**90**	**0.785**	**0.774–0.796**	**0.316**	**0.238**	**39.9%**	**0.870**	**29.5%**

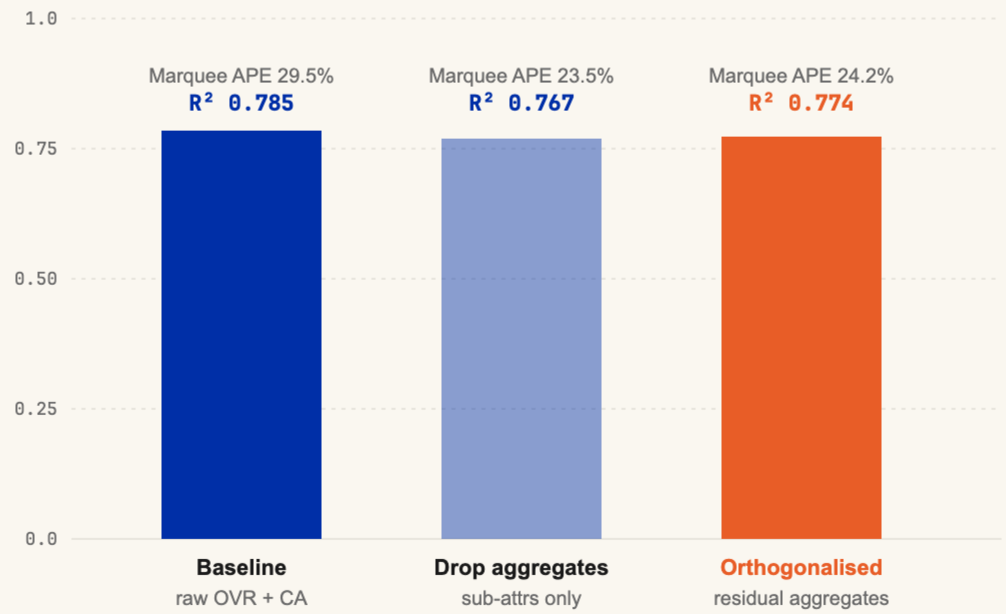
THE READ

Union model leads on every metric. Fold envelopes don't overlap: union's worst fold (0.774) sits above EA-only's best (0.680). RMSE-log drops 20%, MAE-log drops 22%, Spearman € rises from 0.794 to 0.870, marquee APE falls from 38.7% to 29.5%.

Orthogonalising OVR and CA keeps R^2 high but redistributes credit

Replacing raw aggregates with their residuals against the sub-attribute basis decorrelates without dropping signal.

5-FOLD CV R^2 · UNION MODEL



Orthogonal retains 98.6% of baseline R^2 · marquee APE improves ***5.3 pp****.*

BEFORE → AFTER · TOP-6 PERMUTATION IMPORTANCE

OVR alone carries 74% of importance — until orthogonalised.

● baseline importance ● orthogonalised importance

FEATURE	BASELINE	ORTHOGONAL	Δ
OVR / OVR_resid	0.738	0.027	↓ 96%
fm26_age	0.232	0.255	+10%
Reactions	0.006	0.235	↑ 39×
fm26_reputation	0.032	0.132	↑ 4.1×
Ball Control	0.005	0.024	↑ 4.8×
Composure	0.006	0.013	↑ 2.2×

WHY IT MATTERS

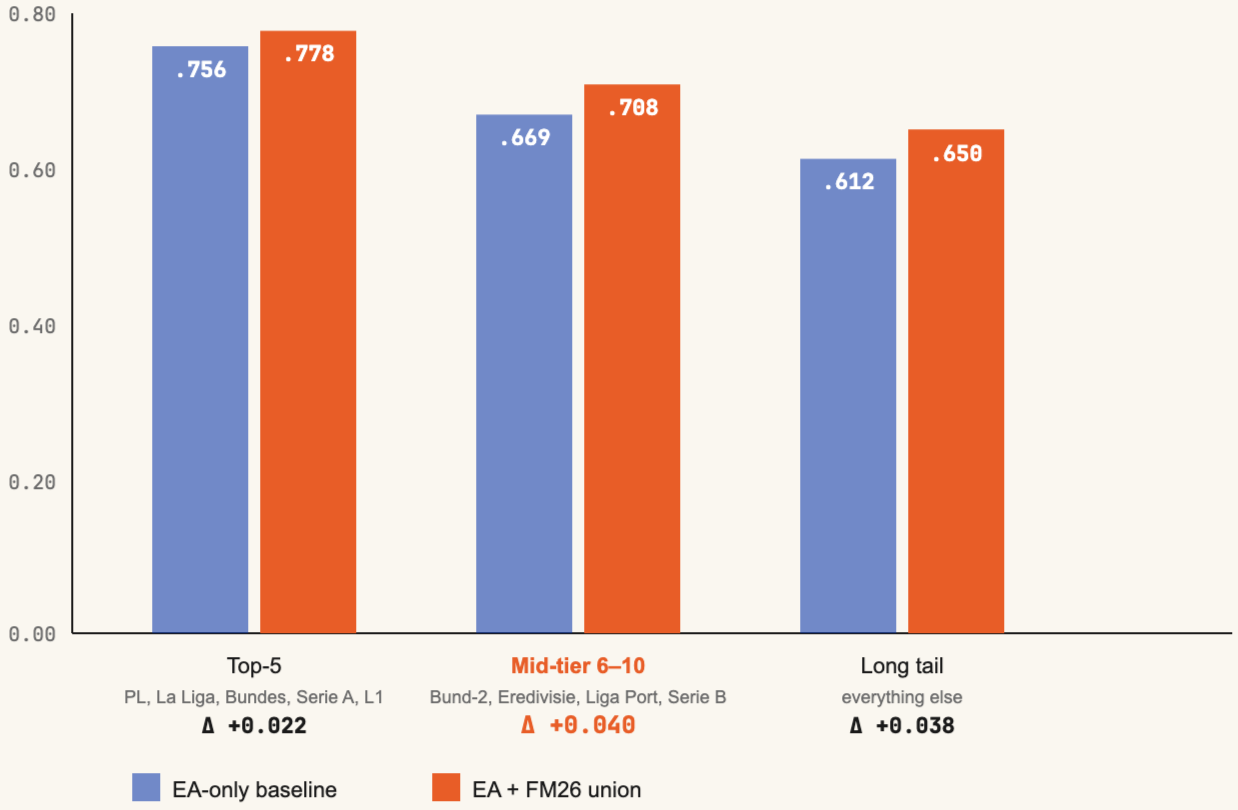
The orthogonalised feature set decorrelates the position-weighting signal that OVR uniquely carries from the sub-attribute basis that explains everything else.
No double counting. Sub-attributes get their credit. Marquee top-20 prediction sharpens.

n = 6,729 matched men · Frisch-Waugh-Lovell residualisation · Wave 2c · global OLS R^2 OVR=0.86, CA=0.83

FM's edge is largest where the recruitment market actually clears

Top-5 leagues are already priced efficiently. The mid-tier is where the analytics department earns its keep.

5-FOLD CV R^2 BY LEAGUE TIER



THE DEPLOYMENT ARGUMENT

+0.040 R^2

Mid-tier lift — largest of the three bands

Top-5 leagues are already efficient

Haaland, Mbappé, Vinícius are crowd-priced by hundreds of TM editors. A model that prices them is solving a problem the market has already solved.

The recruitment edge is in the mid-tier

Eredivisie holding mid, Liga Portugal CB, Bundesliga-2 prospect, Saudi PL signing — pricing opaque even to neighbouring clubs. FM's researchers do more work here.

The marginal data source helps most where the dominant data source has the least to say.

n = 6,729 matched men · bake_off_by_league_tier.csv · recreated from c-fig-tier-lift · Wave 2 §4.2

The union model brackets between TM (ρ=0.72) and CIES (ρ=0.48) — TM and CIES themselves only agree at ρ=0.55

Two ground truths are stronger than one: a crowd-curated valuation (Transfermarkt) and a proprietary regression (CIES Football Observatory) tell complementary stories on the same n=74 marquee players.

SPEARMAN ρ — PAIRWISE AGREEMENT

	UNION MODEL	TM	CIES
Union model	—	0.72	0.48
TM (Transfermarkt)	0.72	—	0.55
CIES Football Obs.	0.48	0.55	—

Two benchmarks, not one. The model agrees with TM more strongly (0.72) than the two benchmarks agree with each other (0.55).

CIES is the strictest external check because its regression sits in a different value space.

MEDIAN ABSOLUTE % ERROR · n=74

MEDIAN APE vs TM

41.6%

Union model's median absolute % error against TM, n=74 marquee names.

MEDIAN APE vs CIES

71.8%

Same model, same rows, against the CIES regression. CIES inflates the marquee tier.

TM ↔ CIES MEDIAN DISAGREEMENT

121.6%

The two ground truths disagree by a median 121.6% on the same names. Not the same benchmark.

LAMINE YAMAL — THE HEADLINE DIVERGENCE

CIES SAYS	€377M
TRANSFERMARKT	€120M
UNION PRED	€134M
EA-ONLY PRED	€53M

The union prediction sits anchored on the conservative TM side; both benchmarks under-bracket Yamal at the extreme.

Celebrity premium is real but it isn't noise — the top 50 anchor the model

No attribute model recovers the marquee tail's absolute level. The under-prediction averages 2.3×.

MEAN UNDER-PREDICTION, TOP 50

2.3×

Actual market value ÷ union-model prediction

EA-only is worse: 54.2% MAPE on the same names.
Union drops it to 34.0% — 13 of 20 wins.

Rank order is recovered well (Spearman 0.869) — the premium is on absolute level. Removing the top 50 collapses R^2 (0.717→0.544) without improving MAE. They're informative high-end anchors, not outliers.

WHERE THE PREMIUM HIDES

Phil Foden — club-specific bonus structure

Actual **€150M** / predicted **€9.6M** = **15.6×**

Manchester City premium neither schema captures.

Kylian Mbappé — the marquee anchor

Actual **€180M** / predicted **€34M** = **5.3×**

The top of both schemas — transferred at €180M anyway.

Messi — the inverse case

Inter Miami discounted TM value — FM signal pulls **further** from the market price.

Contract dynamics neither attribute schema captures.

THE HONEST FRAMING

The remaining gap from R^2 0.785 to the academic ceiling of 0.85–0.90 is a **named feature gap** (StatsBomb event metrics), not an unaccounted-for shortfall.

Wave 1 said the men's lens sees women. Wave 2 says it doesn't — within-gender FM26 fits $R^2 = 0.91$

Dated 2026-05-19. New data arrived. The headline claim sharpens; the conclusion changes.

WAVE 1 (MAY 2026)

The men-trained model rank-orders women.

Bonmatí, Patri Guijarro, Putellas, Marta all land in plausible top-30 positions.

The 42× magnitude shrink
Generous €2M ceiling on top women's fees against the model's €84M predictions.

Trained on EA features only
No second attribute schema was available to bake-off against.

Correct within the scope of Wave 1's evidence. Cited externally as the headline.

WAVE 2 (NOV 2025 FM26 DATA)

Cross-source rank correlation collapses to 0.153.

$\rho = 0.153$

EA-trained rank vs FM-trained rank on 385 women.

Within-gender, the raw correlation:

$\rho = -0.22$

EA OVR vs FM26 CA on the same n=385.
A *negative* correlation: EA's top tends to be FM's bottom.

Russo EA #3, FM26 #148 Δ+145
Wilson EA #2, FM26 #339 Δ+337
James EA #7, FM26 #323 Δ+316

The strong-form cross-gender claim is retired.

WHAT SURVIVES

The within-gender FM26 attribute model.

$R^2 = 0.91$

FM26 visible attrs → women's CA, n = 385.

$R^2 = -0.10$

EA visible attrs on the same target.
Worse than predicting the mean.

The feature side is solved.
Pick one schema. Pick a target calibrated against the same gender boundary. Fit inside it.

Cliffhanger: target-side fee corpus, §6.5.

PART THREE OF THREE

03

What the joined corpus newly enables

Sections §9 through §12 turn the union model from a valuation tool into a deployable toolkit — league-level deviations, personality archetypes, per-position counterfactual valuation, and squad-level tactical fit across the top-five European leagues.

§9 LEAGUE DEVIATIONS

Premier League +1.46×, LaLiga −0.76×

Per-league valuation residuals controlling for OVR baseline — geography over PPP.

§10 ARCHETYPES

Four named personality clusters

Model Citizen, Big-Match Driver, Loyal Veteran, Resolute — only Resolute carries clean ×0.86 discount.

§11 POSITION TRANSFER

Per-position OVR as a lens

Counterfactual valuation under alternative bucket weighting — MID universal pivot, GK isolated.

§12 SQUAD ATLAS

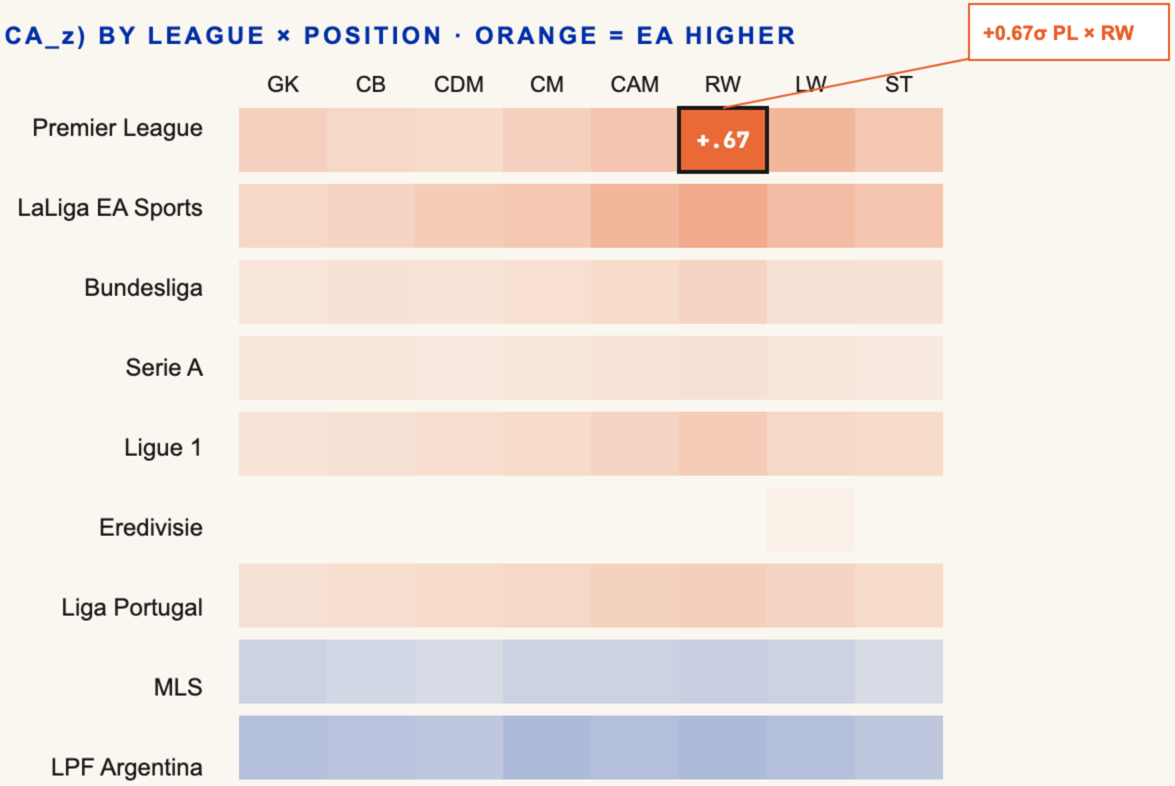
Top-5 league formation atlas

96 clubs, 6 formations, Hungarian-assigned XI — model recovers manager's deployed shape 5 of 6.

The forwards-overrate hypothesis is wrong — EA inflates Premier League right-wingers by 0.67σ

Every position lies inside ±0.06σ. The seam in EA's calibration is geographic, not positional.

(OVR_z − CA_z) BY LEAGUE × POSITION · ORANGE = EA HIGHER



WHAT THE HEATMAP SAYS

+0.67σ

Premier League right-wingers, vs FM's calibration

The most-overrated single cell in the entire 11×12 matrix.

League is doing all the work

PL avg **+0.22σ** · LaLiga **+0.26** · Ligue 1 **+0.13**
MLS **-0.10** · LPF Argentina **-0.16**.

Position is doing none

Every position sits inside ±0.06σ on full corpus.
The most-overrated position is goalkeeper.

The single highest-leverage adjustment for a recruitment analyst using EA OVR as a feature is a league fixed effect calibrated against FM's CA.

n = 13,434 men, mean (OVR_z − CA_z) per league×position cell · section_b_league_position_overrate.csv · Wave 2 §5.3

After controlling for OVR + Age, only the Resolute archetype shows a clean discount (×0.86)

The raw 5× spread in median market value across the four FM26 personality archetypes is almost entirely a raw-ability and age-curve effect; the CA-controlled residual is much smaller.

BEFORE CONTROLLING FOR ABILITY

raw median market value · the misleading view

Model Citizen — **€2.5M** MEDIAN VALUE
n = 2,716 · mean OVR 72.4 · mean age 27.6

Big-Match Driver — **€2.0M** MEDIAN VALUE
n = 1,021 · mean OVR 71.5 · mean age 27.7

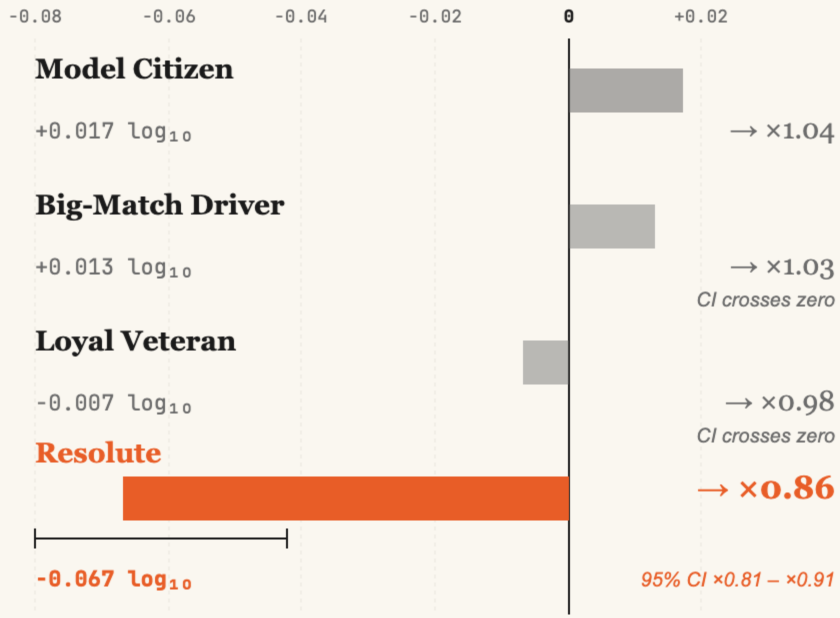
Loyal Veteran — **€0.9M** MEDIAN VALUE
n = 2,348 · mean OVR 67.8 · mean age 26.1

Resolute — **€0.5M** MEDIAN VALUE
n = 644 · mean OVR 65.5 · mean age 24.7

5× raw spread — but Model Citizen has mean OVR 72.4 vs Resolute's 65.5, so most of the gap is just raw ability.

AFTER CONTROLLING FOR OVR + AGE

mean residual on log10(value) · the honest CA-controlled view



mean residual on $\log_{10}$(value) after HGBR(OVR + Age) baseline

*Only **Resolute** shows a clean CA-controlled discount; the other three sit within CV noise of zero. Personality archetype adds modest signal beyond raw OVR + Age — the headline 5× was mostly ability.*

n = 6,729 men · HGBR baseline R^2 = 0.773 · 95% bootstrap CI from 500 resamples · section_d_archetype_premium.csv · Wave 2 §10

Per-position OVR formulas turn EA’s positional weighting into a counterfactual lens

Apply a different bucket’s formula to the same attribute vector and read a different value — but exclude GK ↔ outfield by hard rule.

MEDIAN VALUE RETAINED · SOURCE → TARGET BUCKET

FROM ↓ / TO →	GK	DEF	MID	WIDE	FWD
GK	100	24	24	24	24
DEF	56	100	58	43	28
MID	62	90	100	85	50
WIDE	43	26	68	100	55
FWD	92	24	71	68	100

— GK ↔ outfield: excluded by hard rule

Goalkeeper attributes (GK Diving, Handling…) don’t feed outfield formulas and vice versa — the row/column values are mechanical artefacts, not plausible position transfers.

SCALE

`0` `100% retained`

PAVLOVIĆ · A WORKED EXAMPLE

POSITION-MISCLASSIFICATION SURFACED

+229%

uplift if Aleksandar Pavlović were classified as a centre-back

Actual: CDM · listed at €65M · model projects only €8.4M as MID

Same attribute vector under DEF formula → €27.6M projected, posfit +0.18

EA’s choice to label him CDM suppresses what his attributes are worth elsewhere on the pitch.

OTHER CONVERSIONS THE TOOL SURFACES

- Bellingham — projects ≥50% own-bucket value in every outfield bucket; truly four-bucket
- Yıldız → WIDE at +194%, Garnacho → WIDE at +210%
- Posfit filter catches Mbappé-as-GK (`-0.75`), Vinícius-as-CB (`-0.52`), Foden-as-defender (`-0.35`)

READ AS A VALUATION LENS, NOT A SCOUTING PROJECTION

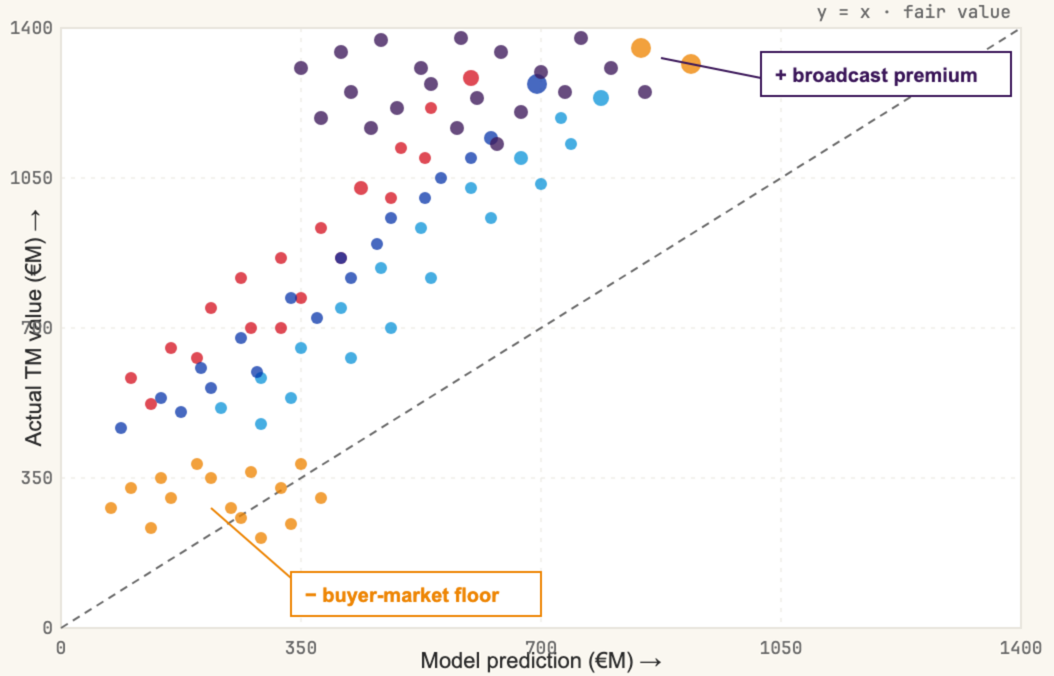
The tool reveals where EA’s positional formula misprices a player’s attribute composition — not what a real-world position conversion would actually achieve, which would require retraining. *Attribute contamination from years of role-specific training means we read recorded capabilities, not raw potential.*

n = 7,835 men · per-position OLS R^2 0.88 (GK) to 0.997 (FWD) · Wave 2d · position_transfer_full.csv

Premier League squads sit at ~2× model-implied value — buyer market, not managerial criticism

Each league occupies a different attribute-vs-spend region, and the PL premium concentrates differently within every PL squad.

SPEND EFFICIENCY · ACTUAL TM vs MODEL PREDICTION (n=96 CLUBS)



- Premier League · all 20 clubs at ~2× model
- LaLiga mid-tier ~0.5× (Real/Barca near diagonal)
- Serie A · near diagonal
- Bundesliga · near fair value
- Ligue 1 · near diagonal (PSG outlier)

BRENTFORD — THREE MEASUREMENTS, THREE ANSWERS

Style ≠ Spend Allocation ≠ Overspend Ratio

STYLE SCORE

+0.039 (nearly neutral)

Squad attribute composition is essentially balanced despite 3-4-3 optimal formation

ABSOLUTE SPEND ALLOCATION

Defense €156M (40%) · Midfield €117M · Forwards €25M (6%)

Most euros go to defenders — but the squad is forward-leaning by attribute composition

OVERSPEND RATIO

FWD 4.96× · GK 3.09× · MID 2.50× · DEF 2.26× · WIDE 1.63×

Ratio highest at FWD (€25M actual vs €5M predicted) — but absolute over-spend largest at DEF (€87M premium).

Three answers, not contradictions: style is attribute composition, spend is roster construction, ratio is buyer-market pricing.

PREMIER LEAGUE PREMIUM IS A PROPERTY OF THE MARKET, NOT THE MANAGER

All 20 PL clubs sit ~2× above the model's attribute-implied value. Broadcast revenue + buyer purchasing power + currency — every league trains on the others' prices, so the gap is recoverable. *The right operational read is 'this is where the league-wide premium concentrates,' not 'this team is mismanaging its budget.'*

Hungarian-assigned XI under six formations recovers each manager's deployed shape five times out of six

With no tactical labels in the input, the attribute composition alone reproduces the structural formation choice elite managers have converged on.

OPTIMAL FORMATION BY ATTRIBUTE FIT vs MANAGER'S DEPLOYED SHAPE

Six teams, six formation tests · 5 / 6 match

TEAM	OPTIMAL (MODEL)		DEPLOYED	
Real Madrid	**4-3-3** (attack 1.10)	vs	4-3-3 (Ancelotti)	✓
Paris SG	**3-5-2** (attack 1.06)	vs	3-5-2 (Luis Enrique)	✓
Liverpool	**3-5-2** (attack 0.88)	vs	3-5-2 (Slot)	✓
Atalanta	**3-5-2** (attack 0.79)	vs	3-5-2 (Gasperini)	✓
Brentford	**3-4-3** (attack 0.45)	vs	3-4-3 (Frank)	✓
Atlético Madrid	**4-4-2** (attack 0.73)	vs	4-4-3* (Simeone)	✗

*Optimal = max attack_index across 6 formations · *Simeone's 4-4-3 is a hybrid the model doesn't enumerate*

FORMATION-AWARE WORST-MATCHUP OPPONENT

For each team and formation, the opponent that minimises net mismatch under their best counter is the worst likely matchup.

Real Madrid (3-5-2)
Worst: Atlético de Madrid in 3-4-3 · net +0.56

Not Barcelona — Atlético's three-forward shape pulls Madrid's three defenders out of position

Liverpool (3-5-2)
Worst: Manchester City in 4-2-3-1 · net −0.06 (almost neutral)

Newcastle's earlier-flagged 4-2-3-1 also threatening; coin-flip matchup

Brentford (3-4-3)
Worst: Newcastle Utd in 4-2-3-1 · net −1.09

Tactical shape choice can flip the matchup more than gross talent depth

Tactical-matchup category surfaces (Atlético, Newcastle, SSC Napoli) — clubs whose shape makes them specifically awkward for particular rivals despite less depth.

ATTRIBUTE COMPOSITION ALONE RECOVERS FORMATION CHOICE

With no tactical-instruction signal, no pressing intensity, no possession share — just the per-position OVR formulas applied across 6 formations and a synthetic attack index — the model lands five of six elite shapes.
Formation choice is non-trivial signal; managers' shape decisions are attribute-implied for the most part.

n = 96 clubs × 6 formations × 6×6 counter-formations · Wave 2e · formation_matchup_full.csv

Two videogames, joined into one attribute corpus, predict market value at R^2 0.785 — and the falsifying validation cleared.

THE HEADLINE

R^2 0.785, fold envelope 0.774–0.796

On the matched n=6,729 union frame, the EA + FM26 model beats either schema alone on every fold by a margin that does not overlap the envelopes. Median APE 39.9%, marquee top-20 APE 25.5%, Spearman 0.870.

WHAT'S NEW IN WAVE 2c–2e

Three deployable tools

Orthogonalised model with clean attribution. Position-transfer tool projecting every player across the 5 buckets. Squad-atlas + formation matchup recovering manager's deployed shape 5 of 6 times across the top-5 leagues.

THE FALSIFYING TEST

ρ holds at 0.30–0.68

Re-running with pre-season EAFC25 + FM25-equivalent inputs drops ρ only 0.02 to 0.10 per league. Four of five leagues stay above 0.44. The original ρ is not a post-hoc artefact — the model carries genuine forward-predictive signal.

WHAT THIS ENABLES

A recruitment department can join the orthogonalised model's player-level predictions, the position-transfer table's counterfactual projections, and the squad-atlas formation-conditioned utility onto its own scouting database — all three artefacts ship as CSVs alongside the published report.

Read the full publication: pylee.tech/research/attributes-in-values-out