Erstellen geschichteter Lichess-Benchmarkkurven für Elo+Chess

Elo+Chess Forschungsnotizen

31. Mai 2026

Abstract

Elo+Chess-Benchmark-Berichte werden aus großen geschichteten Stichproben von erstellt Lichess.org-Spiele, mit dem in diesem Hinweis verwendeten Benchmark-Beispiel 483.974 1+0 Bulletpartien, 479.974 3+0 Blitzpartien, 925.636 10+0 Schnellschachpartien, und 660.540 längere Schnellschachspiele. Für jeden wichtigen Spieltyp wird die Pipeline erweitert Jedes Spiel wird in den Zustand vor und nach dem Zug unterteilt und Hunderte davon berechnet Zwischen- und Spielerspielvariablen, aggregiert diese Variablen zu Coaching-Metriken und mittelt diese Metriken dann nach Lichess Elo-Bucket. Die Die Stichprobe ist bewusst geschichtet und nicht proportional zum Rohspieler Pool: Der Zweck besteht darin, stabile Verhaltenskurven auf jeder Bewertungsebene abzuschätzen. Einschließlich Eimer, die sonst spärlich wären, nicht um das Natürliche zu reproduzieren Bewertungsverteilung. Das Ergebnis ist eine Familie empirischer Benchmark-Kurven: z Bei jeder Metrik können wir fragen, wie oft Spieler auf unterschiedlichen Elo-Stufen ein bestimmtes Ergebnis zeigen Gewohnheit oder Prinzip in echten Spielen. Dieser Hinweis beschreibt die Methode und den Umfang die Berechnung und mehrere Beispiele für Eröffnungsprinzipien wie Rochade, Entwicklung von Nebenstücken, wiederholte Eröffnungsfigurenbewegungen und frühe Damenbewegung. Der Schwerpunkt liegt Methodisch: Elo+Chess geht nicht davon aus, dass Schachprinzipien universell gelten linear über alle Fähigkeitsstufen hinweg. Stattdessen werden Benchmark-Kurven untersucht Bewertungsbereich, und das Produkt konzentriert sich auf den Anfänger- bis frühen Fortgeschrittenenbereich wo die Beziehungen am stärksten und am besten interpretierbar sind.

1 Zweck

Das Ziel des Benchmark-Systems besteht darin, umfassende Schachratschläge in umzuwandeln messbare Größen. Eine Aussage wie „Entwickeln Sie Ihre Nebenstücke“ ist nützlich, aber es wird umsetzbarer, wenn wir sagen können, wie viele Nebenstücke a Der Spieler entwickelt sich normalerweise bis zum 10. Zug, im Vergleich zu seinen Mitspielern Bewertungsgruppe und wie sich diese Gewohnheit im gesamten Bewertungsspektrum ändert.

Elo+Chess erstellt daher eine Metrikbibliothek aus tatsächlichen Spielen. Jede Metrik ist berechnet für jede Spieler-Spiel-Reihe, zusammengefasst nach Spieltyp und Elo-Bucket, und wird im Bericht als Benchmark-Kurve verwendet. Der Bericht kann dann einem Benutzer zugeordnet werden dieselbe Kurve und zeigen an, ob sich der Benutzer über, unter oder in der Nähe des Verhaltens befindet bei vergleichbaren Spielern beobachtet.

2 Quelldaten und Schichtung

Die Benchmark-Quelle ist die öffentliche Spieledatenbank Lichess.org. Elo+Chess verwendet große geschichtete Stichproben nach Spieltyp und Bewertungsgruppe statt einer kleinen Convenience-Probe. Die aktuellen Produktionslaufzeitartefakte decken vier Hauptartefakte ab Spielkategorien:

- 1+0 Kugel,
- 3+0 Blitz,
- 10+0 schnell,
- Schnellschachspiele, die länger als 10 Minuten dauern.

Das Stratifizierungsziel ist die Breite über alle Bewertungsstufen hinweg. Ein proportionaler Zufall Beispiel wäre für die Beschreibung des gesamten Spielerpools effizient, würde es aber tun für die Erstellung von Benchmark-Kurven ineffizient sein. Die gebräuchlichsten Bewertungsbereiche würden die Stichprobe dominieren, während Heckschaufeln mit geringer Stützung zu wenige hätten Spiele für stabile metrische Durchschnittswerte. Elo+Chess probiert daher nach durchschnittlichem Spiel Elo-Eimer. In der aktuellen Clean-Sample-Konstruktion stehen Rohspiele an erster Stelle gefiltert nach gültig bewerteten Spielen mit bekannten Spielern, bekannten Bewertungen, gültigen Ergebnissen, und bekannte Zeitkontrollen. Spiele werden dann nach Spielern mit genügend Spielen gefiltert in dieser Geschwindigkeit und Stabilität genug Bewertungen, um ein kohärentes Fähigkeitsniveau darzustellen: Für jeden Spieler berechnet Elo+Chess die Geschwindigkeitsspanne des Spielers innerhalb der Geschwindigkeit ($\max(\text{Elo}) - \min(\text{Elo})$) im gesamten Quellpool und bleibt erhalten Spiele nur, wenn beide Spieler mindestens die konfigurierte Mindestspielanzahl haben und nicht mehr als 300 Elo-Punkte innerhalb der Geschwindigkeitsdrift. Der Driftfilter ist nicht gemeint um normale Verbesserungen oder Abweichungen zu beseitigen; es reduziert Fälle, in denen ein Spieler Spiele würden wesentlich unterschiedliche Fähigkeitsniveaus vermischen, wie z. B. vorläufige Konten, wiederkehrende Spieler oder Konten, die sich im Laufe der Zeit schnell durch mehrere Buckets bewegt haben der abgetastete Zeitraum. Schließlich werden die teilnahmeberechtigten Spiele innerhalb jedes durchschnittlichen Elo eingestuft Bucket durch einen deterministischen Hash der Spiel-ID und des Startwerts sowie einer begrenzten Anzahl von Spiele werden aus jedem Bucket beibehalten.

Dieses Design macht die Benchmark-Kurven viel nützlicher. Jeder Elo-Eimer trägt genügend Beobachtungen bei, um den Durchschnittswert jeder Metrik abzuschätzen, und seltene Eimer mit hoher oder niedriger Bewertung werden nicht von der natürlichen Form übertönt der Aktivspielerverteilung. Die von Elo+Chess verwendeten Benchmark-Kurven sind dann auf den Bereich beschränkt, der für das Produkt am relevantesten ist, etwa Anfänger durch frühes fortgeschrittenes Spiel. In den Lichess-Skalenberichten dieser primäre Bereich ist 600–1600. Wenn ein Benutzer den Bericht auf der Chess.com-Skala anzeigt, wird der entsprechende Rating-Buckets werden demselben zugrunde liegenden Lichess zugeordnet Benchmark-Buckets mithilfe der separaten plattformübergreifenden Bewertungszuordnungsmethode beschrieben im Begleitpapier: Estimating Cross-Platform Chess Rating Mappings with Modale Regression.

3 Bewertungsverteilungen für aktive Spieler

Schauen wir uns die Verteilung der Elo-Wertungen der Lichess.org-Spieler am Ende an März 2026, nach Hauptspieltyp, sehen wir die folgenden Zählungen und Perzentile. Bei dieser Verteilungsan-

sicht wird jeder Spieler einmal pro genauem Spieltyp gezählt. unter Verwendung der letzten beobachteten Wertung dieses Spielers im entsprechenden Raw Zeitkontrollreihen, nachdem in genau diesem Spiel mindestens fünf Spiele erforderlich waren Typ.[1]

Game type	Players	Median	75th pct.	80th pct.	% $\leq$1600	Exact 1500	1500 share
1+0 bullet	1,221,146	1,479	1,858	1,951	58.9%	4,756	0.4%
3+0 blitz	1,144,045	1,542	1,845	1,912	54.9%	2,315	0.2%
10+0 rapid	1,679,401	1,381	1,691	1,765	68.3%	3,472	0.2%
$>$10 rapid	668,255	1,400	1,686	1,753	68.2%	7,969	1.2%

Table 1: Neueste beobachtete Lichess-Bewertungsverteilung für aktive Spieler nach genauem Schema Spieltyp, nachdem mindestens fünf Spiele in genau diesem Spieltyp erforderlich sind.

Diese Verteilung sollte nicht mit der geschichteten Benchmark-Stichprobe verwechselt werden. Die Verteilung beschreibt den im Rohdatenbestand beobachteten aktiven Lichess-Spielerpool scannen. Die Benchmark-Stichprobe wird dann gezielt geschichtet, so dass Elo-Buckets entstehen ausreichend Unterstützung für stabile metrische Kurven haben.

Um eine Lichess.org Elo-Wertung für einen bestimmten Spieltyp einer zuzuordnen äquivalente Chess.com-Bewertung, Elo+Chess verwendet das spieltypspezifische Modal Im Begleitdokument geschätzte Regressionsgleichungen Elo+Chess rating-mapping paper:

$$\begin{aligned}
\widehat{R}_{\text{Chess.com,bullet}} &= -531.71 + 0.9871 R_{\text{Lichess}}, \\
\widehat{R}_{\text{Chess.com,blitz}} &= -551.54 + 1.0853 R_{\text{Lichess}}, \\
\widehat{R}_{\text{Chess.com,10minrapid}} &= -495.04 + 1.0741 R_{\text{Lichess}}, \\
\widehat{R}_{\text{Chess.com,}>\text{10rapid}} &= -369.02 + 0.9374 R_{\text{Lichess}}.
\end{aligned}$$

Diese Gleichungen bilden Bewertungswerte ab, nicht Perzentilränge. Perzentilränge sind definiert innerhalb eines bestimmten Spielerpools und der Spieler Chess.com und Lichess Pools haben unterschiedliche Formen. In Tabelle 2 für die 10-Minuten-Schnellspieltyp, die erste Spalte gibt einen kleinen Ausschnitt der möglichen Möglichkeiten Chess.com Elo-Bewertungswerte. Die zweite Spalte gibt das entsprechende an Chess.com-Perzentilrang, angezeigt auf Chess.com zum 31. Mai 2026. A 71,7 % Der Rang auf Lichess.org für 10-Minuten-Schnellschach entspricht einer Lichess Elo-Wertung von 1.644[2] und dieser Wert wird im angezeigt dritte Spalte. Wenn wir die Zuordnungsgleichung Chess.com–Lichess anwenden, 1.644 auf Lichess entspricht etwa 1.271 auf Chess.com. Gehe von der Schätzung aus Die Zuordnungsgleichung ist sinnvoll, die Lücke $+522$ zwischen der Chess.com-Bewertung in Spalte 1 und die abgebildete Bewertung in Spalte 4 lassen sich durch eine große Zahl erklären von Gelegenheitsspielern im Chess.com-Spielerpool, die den Chess.com aufblähen Prozentrang. Mit anderen Worten, im 72. Perzentil auf Lichess.org zu sein eine größere Leistung, als auf Chess.com im 72. Perzentil zu sein, weil Perzentilränge lassen sich nicht sauber von einem System auf das andere übertragen. Tatsächlicher Elo Werte können mithilfe von Gleichungen zugeordnet werden, die anhand der Punktezahlen desselben Spielers geschätzt werden die beiden Systeme.

[1] We apply a five-game minimum because lightly active accounts disproportionately remain at the 1500 starting/provisional anchor, creating an artificial spike in all-player latest-rating distributions. Figures 1–4 Zeigt die Verteilungen mit und ohne diesen Filter an.

[2] Calculated from the cumulative distribution function of the ZXQPROT3ZXQ 10+0 rapid latest-rating sample after requiring at least five games in that exact game type; see Table 1 and Figure 3.

Chess.com rating	Chess.com pct.	Same-pct. Lichess	Mapped equiv.	Gap
749	71.7%	1,644	1,271	+522
1,012	86.2%	1,871	1,515	+503
1,358	95.4%	2,095	1,755	+397
2,383	99.9%	2,557	2,251	-132

Table 2: Anschaulicher Vergleich zwischen Chess.com 10-Minuten-Schnellperzentil Punkte und die gleichen Perzentilpunkte im aktiven Lichess 10+0 Rapid Verteilung, nachdem mindestens fünf 10+0-Spiele erforderlich waren. „ZxQPROT2ZXQ zugeordnet Äquiv." wendet den aktuellen Elo+Chess an 10+0 schnelle Konvertierung der Bewertungsskala in die Lichess-Bewertung mit dem gleichen Perzentil. Die Die Zeilen mit dem höchsten Ende sind enthalten, um die Skalenunterscheidung zu zeigen, aber die angepassten Die Bewertungslinie ist im Haupt-Coaching-Bereich am zuverlässigsten.

Der Grund für die Lücken in Tabelle 2 liegt in der Zuordnung schätzt äquivalente Spielstärkewerte, während ein Perzentil beschreibt Position innerhalb der eigenen Population einer Plattform. Ein Spieler mit 71,7. Perzentil Chess.com ist daher nicht automatisch ein 71,7-Perzentil-Spieler auf Lichess. und umgekehrt. Die beobachteten Zahlen stimmen mit denen der Öffentlichkeit von Chess.com überein Rapid-Pool mit einer viel größeren Masse an Spielern mit niedrigerer Bewertung oder Gelegenheitsspielern, während der Der aktive Lichess-Pool mit exakter Kontrolle konzentriert sich stärker auf erfahrene Wiederholungen Spieler. Der entscheidende Punkt ist nicht, dass eine Perzentilskala „richtig" ist und die other ist „falsch"; Es liegt daran, dass der Perzentilrang nicht plattformübergreifend übertragbar ist es sei denn, die zugrunde liegenden Spielerverteilungen haben die gleiche Form.

Note on the 1500 spike. The visible jump near 1500 ist in den Daten vorhanden und wird nicht durch Glättung erzeugt.

Im spätesten beobachtete Bewertungsverteilung, genau 1500 Bewertungen machen 13,1

Bullet-Spieler, 5,6

17,6

Start- oder vorläufiger Anker im Bewertungssystem.

Der Anstieg um genau 1.500 ist größtenteils ein Artefakt schwach aktiver Konten. Zu Überprüfen Sie dies. Wir haben die gleichen neuesten Bewertungsverteilungen nach Bedarf neu berechnet Jeder Spieler muss mindestens fünf Spiele in der genauen Spielart haben. Die Spitze fast verschwindet unter diesem Filter: Genau 1500 Bewertungen fallen von 5,6–17,6
Spieler in der All-Player-Verteilung auf 0,2–1,2 % in den fünf oder mehr Spielen Verteilung in Tabelle 1 angegeben.

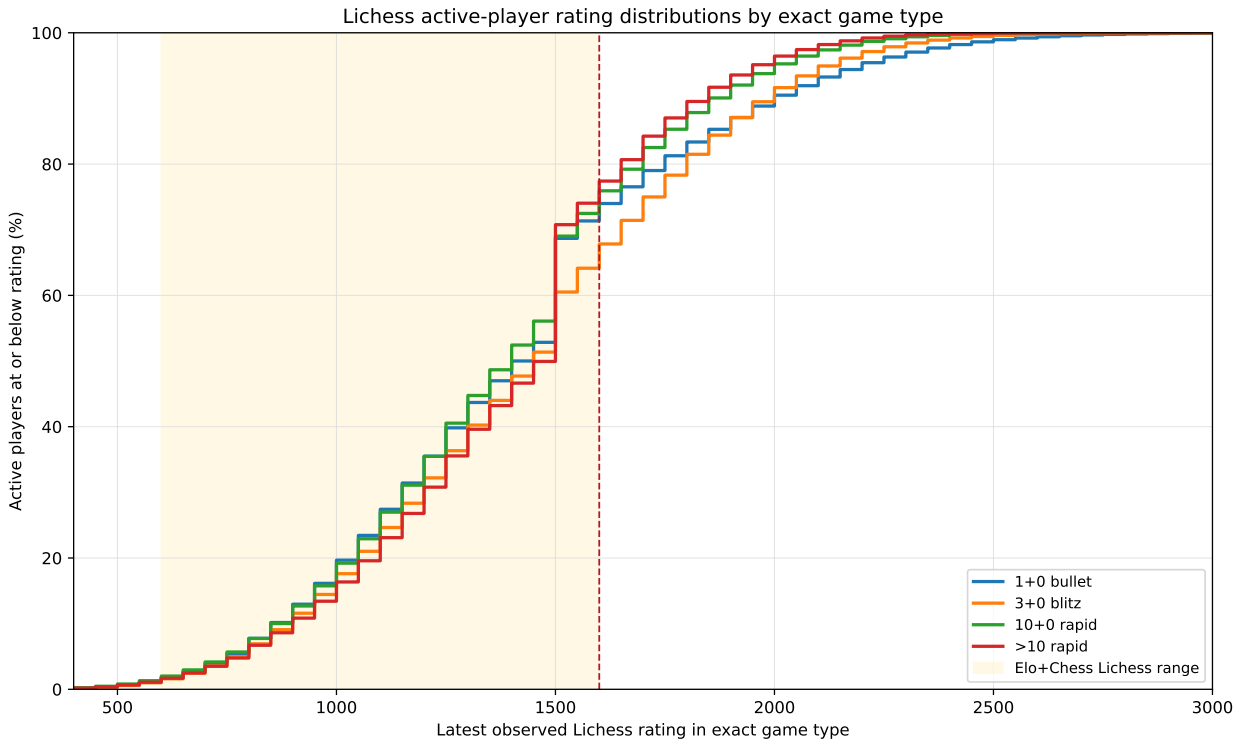


Figure 1: Kumulative Bewertungsverteilungen aktiver Spieler nach genauem Spieltyp. Die Das gelbe Band markiert den Bereich 600–1600 Lichess, der vom Haupt-Elo+Chess verwendet wird Coaching-Benchmarks.

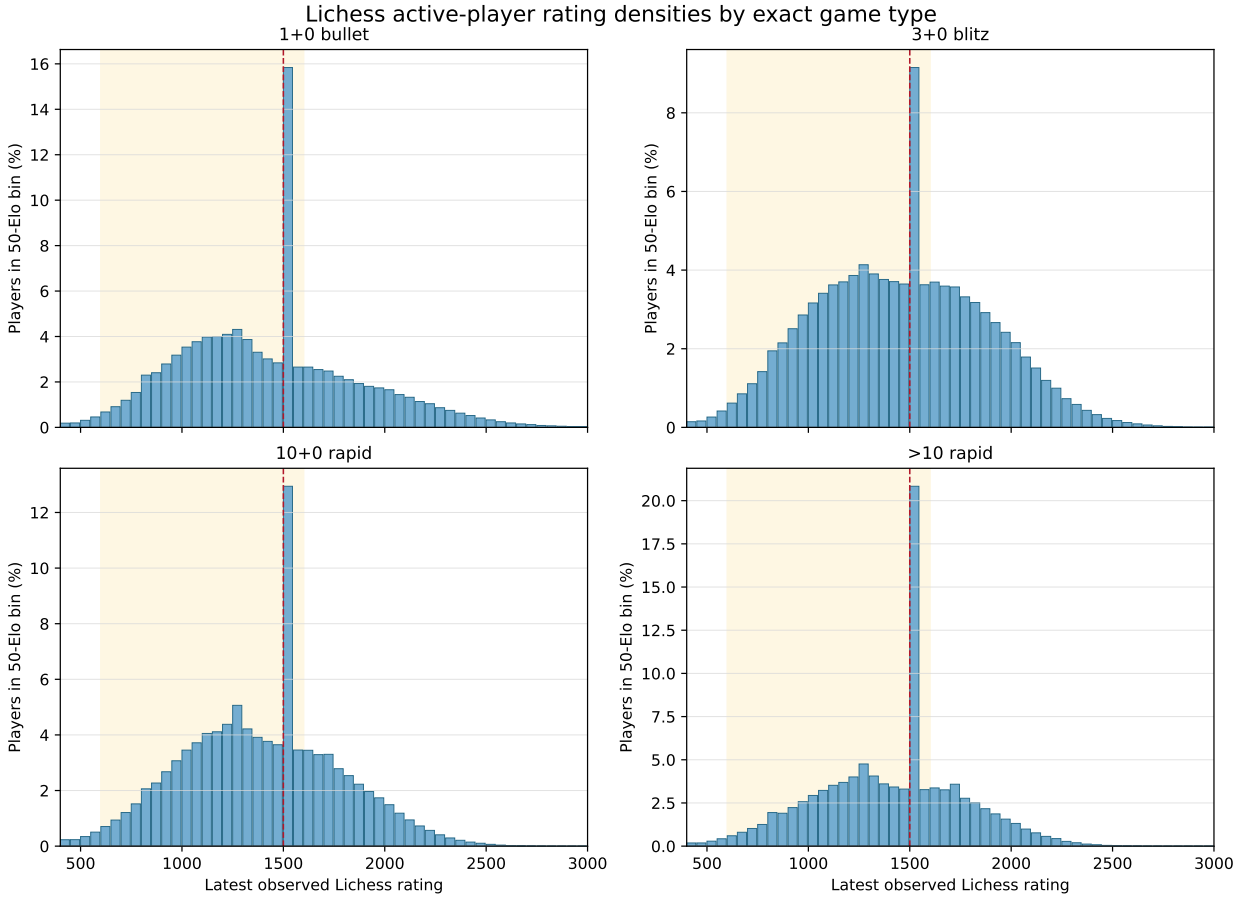


Figure 2: Dichteansicht derselben Bewertungsverteilungen für aktive Spieler. Die Die gestrichelte Linie markiert 1500 Lichess, wo sich die Rohverteilungen mit der neuesten Bewertung befinden ein großer Massenpunkt mit exakter Bewertung. Das gelbe Band markiert den 600–1600 Lichess Bereich, der von den wichtigsten Elo+Chess-Metrikberichten verwendet wird. Die Spitze ist am deutlichsten sichtbar die All-Player-Dichte, da viele wenig aktive Konten genau bei bleiben 1500; Figure 4 wiederholt das Dichtediagramm nachdem mindestens fünf Spiele in der genauen Spielart erforderlich waren, was weitgehend entfällt dieses Artefakt.

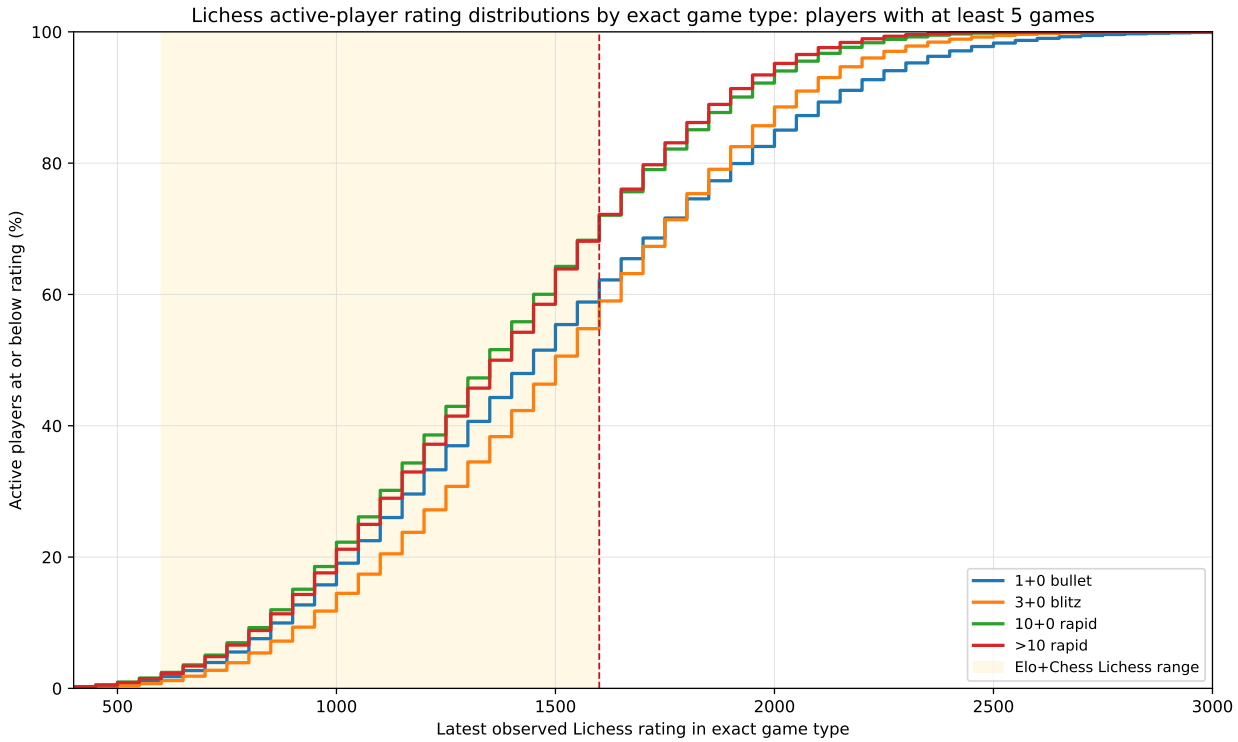


Figure 3: Kumulierte Bewertungsverteilungen aktiver Spieler nach Anforderung von at mindestens fünf Spiele in der genauen Spielart. Der 1500-Sprung ist stark reduziert.

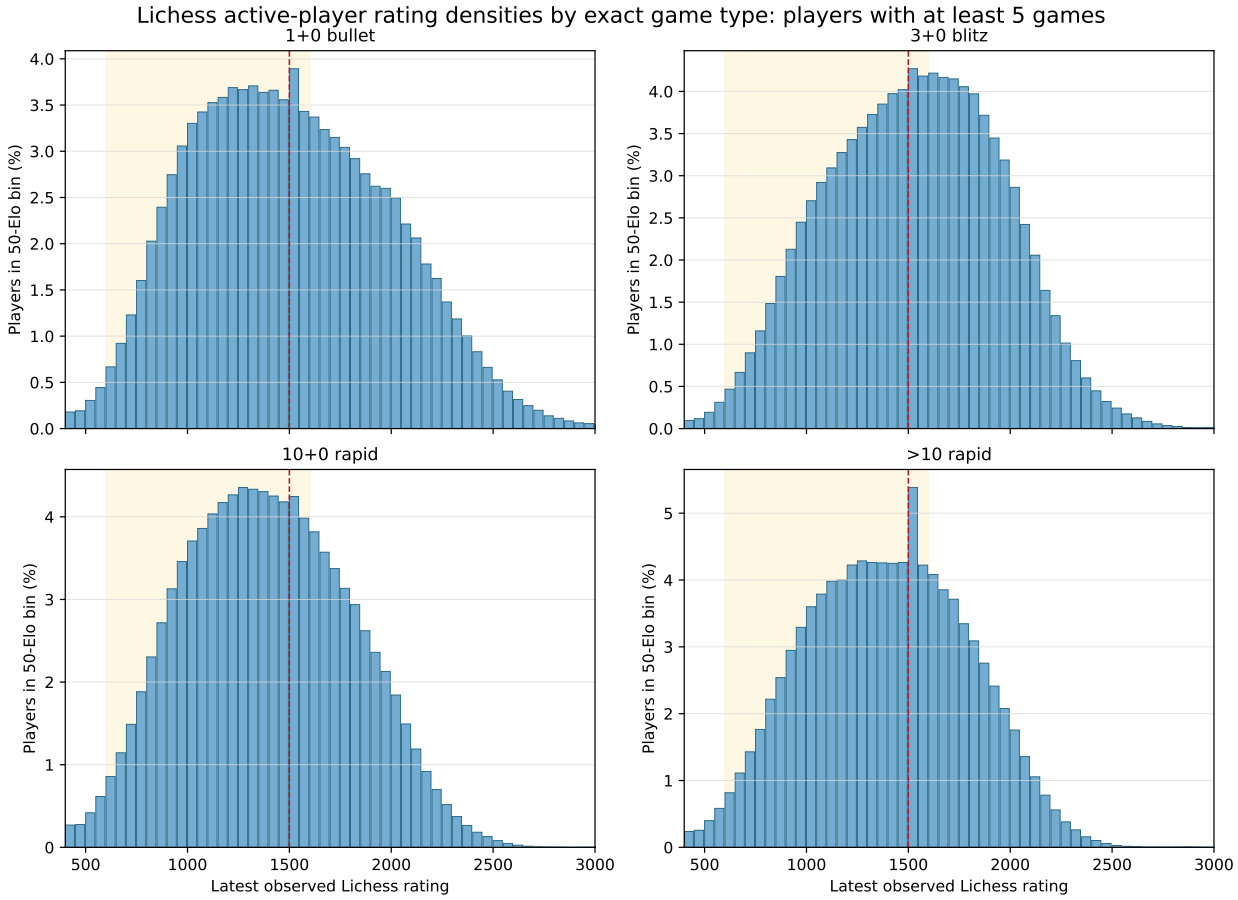


Figure 4: Dichteansicht nach mindestens fünf Spielen im genauen Spiel Typ. Diese Ansicht bestätigt, dass der große 1500-Massenpunkt im All-Player liegt Die Verteilung erfolgt hauptsächlich durch weniger aktive Konten. Das gelbe Band markiert der Bereich 600–1600 Lichess, der von den wichtigsten Elo+Chess-Metrikberichten verwendet wird.

4 Maßstab der aktuellen Benchmark-Artefakte

Der Rechenaufwand ist bewusst groß. Es handelt sich nicht um einen Profilkratzer, sondern um einen kleinen Satz heruntergeladener PGNs oder eine handverlesene Sammlung lehrreicher Beispiele. Der erste Lichess-Pass deckte vollständige monatliche Dateien ab Januar 2025 ab bis März 2026. Die unstratifizierte Scantabelle enthält 1.382.178.087 Spiele und 2.764.356.174 Spieler-Spielbewertungszeilen. In diesen Zeilen 6.623.438 eindeutig Spieler-Benutzernamen erscheinen mindestens einmal.

Speed	Games in raw scan	Player-game rows	Unique players
Bullet	519,375,423	1,038,750,846	3,077,635
Blitz	839,823,863	1,679,647,726	5,987,634
Rapid	22,978,801	45,957,602	1,535,865
All scanned speeds	1,382,178,087	2,764,356,174	6,623,438

Table 3: Ungeschichtete Lichess-Scanskala vor der Benchmark-Probenahme. Der Scan umfasst vollständige monatliche Spielverlaufsdateien von Januar 2025 bis März 2026.

Die Produktions-Benchmark-Artefakte sind viel kleiner als beim Rohscan, weil Sie werden für die metrische Konstruktion ausgewählt. Sie sind immer noch groß: der Maßstab Stage ist eine Move-State-Erweiterungs- und Aggregationspipeline über Hunderte von Tausende Spiele pro Spieltyp.

Game type	Games	Player-game rows	Unique players	Move states	Feature columns
1+0 bullet	483,974	967,948	119,732	22,985,272	277
3+0 blitz	479,974	959,948	129,398	28,138,904	277
10+0 rapid	925,636	1,851,272	229,235	57,554,339	277
$>$10 rapid	660,540	1,321,080	133,661	38,580,214	277

Table 4: Maßstab der stratifizierten Benchmark-Artefakte nach Ausrichtung auf die Gemeinsames 277-Spalten-Laufzeit-Feature-Schema. A „Spieler-Spiel-Reihe“ bedeutet ein Spiel aus der Perspektive eines Spielers, also die meisten Spiele Erzeuge zwei Spieler-Spielreihen.

Allein für 3+0-Blitz werden im aktuellen Produktions-Benchmark 479.974 Partien verwendet mit 129.398 einzelnen Spielern und 28.138.904 verschiedenen Zugzustandsreihen. Die Die ausgerichtete Blitz-Spieler-Spiel-Feature-Tabelle enthält 277 abgeleitete Variablen. Die Schnelle Feature-Build-Artefakte sind in der Move-State-Phase noch umfassender: die vollständige Die während der Feature-Konstruktion verwendete Rapid-Move-State-Tabelle enthält 168 Verschieben Sie Statusspalten vor der Aggregation zu einer Spieler-Spiel-Funktion mit 277 Spalten Tisch. Die Frontend-Laufzeit-Benchmark-Artefakte für alle vier Spieltypen sind dann auf das gleiche 277-Spalten-Spieler-Spiel-Feature-Schema ausgerichtet, einschließlich der erweiterte Taktik-Taxonomie-Breakout-Variablen. Aus diesen Feature-Tabellen werden die Der Bericht wählt die Coaching-Kennzahlen aus, die am besten interpretierbar sind, und zeigt sie an.

Diese Zählungen sind wichtig für die Interpretation der Ergebnisse. Die Benchmark-Kurven sind keine Schätzungen aus einigen hundert manuell überprüften Spielen. Stattdessen sie stammen aus zig Millionen bewerteten Board-Staaten und mehr als vier Millionen Spielerspielreihen in den vier aktuellen Produktionsspielkategorien.

5 Pipeline-Übersicht

Die Benchmark-Baupipeline besteht aus sechs Hauptschritten:

1. Wählen Sie bewertete Lichess-Spiele im Zielspieltyp und in den Bewertungsschichten aus.
2. Analysieren Sie jeden Spielzug Zug für Zug und speichern Sie den Vor- und Nachzug Vorstandsstatus.
3. Berechnen Sie zwischenliegende Zugzustandsvariablen vom Brett, dem Zug, die Spielerperspektive und die Gegnerperspektive.
4. Aggregieren Sie Move-State-Variablen zu Player-Game-Feature-Variablen.
5. Wandeln Sie Spieler-Spiel-Variablen mit einem Clear in Berichtsmetriken um Richtung der Interpretation: höher ist normalerweise besser, niedriger ist normalerweise besser oder beschreibend.
6. Durchschnitten Sie jede Metrik nach Elo-Bucket und überprüfen Sie das Ergebnis manuell Beziehung.

Diese Struktur ist wichtig, da viele Berichtsmetriken nicht direkt sind sichtbar in einem PGN-Header. Sie erfordern eine Rekonstruktion der Position, eine Bestimmung welche Figuren von welchen Startfeldern aus gezogen sind, Vergleich vor dem Zug und Material und Struktur nach dem Umzug und anschließende Aggregation des Ergebnisses daraus Bewegen Sie die Ebene auf die Spielerspielebene.

6 Zwischenvariablen

Die Zwischenvariablen sollen beschreiben, was sich auf der Platine geändert hat und warum es aus der Sicht des Berichtsspielers wichtig ist. Beispiele hierfür sind:

- Zugnummer, Spielerzugnummer, Zugseite, Spielerfarbe und Spieler Farbe;
- Von-Quadrat, Bis-Quadrat, beweglicher Figurentyp, gefangener Figurentyp und ob der Zug eine Eroberung, ein Schach, eine Burg, eine Beförderung oder ein Matt war;
- Materialbilanz vor und nach dem Umzug;
- ob eine Nebenfigur ihr Startfeld um eine bestimmte Zugzahl verlassen hat;
- ob die gleiche Figur bei aufeinanderfolgenden Spielerzügen bewegt wurde Öffnung;
- ob die Dame vor Zug 10 gezogen ist;
- Rochadeseite, Rochadezugnummer und ob Türme verbunden wurden nach der Rochade;
- Lose Figuren, hängende Figuren, Doppelbauern, isolierte Bauern, Freibauern Bauern und rückständige Bauern;
- Turmreihenstatus, offene Reihen, halboffene Reihen und Königssicherheit Merkmale;
- taktische Gelegenheiten, ausgeführte Taktiken, Materialdrucktaktiken, garantierte Folgegewinne und realisierte Gewinne.

Der genaue Zwischensatz variiert je nach Bauphase. Laufzeitberichtsdatenbanken bleiben erhalten Nur die Spalten, die für die schnelle Bereitstellung von Berichten erforderlich sind, während der vollständige Funktionsaufbau gewährleistet ist Zu den Artefakten gehören breitere Move-State-Tabellen und Feature-Patches, die zum Erstellen verwendet werden die letzten Spielerspiel-Feature-Zeilen.

7 Beispiele für Öffnungsprinzip-Metriken

Öffnungsprinzipien sind nützliche Beispiele, da sie für Benutzer einfach umzusetzen sind verständlich und leicht mit Bewegungszustandsberechnungen zu verbinden.

7.1 Minor Pieces Developed by Move 10

Diese Metrik zählt, wie viele der vier Springer und Läufer eines Spielers noch übrig sind ihre Heimatfelder bis zum 10. Zug des Spielers. Die Zwischenrechnung Verwendet die Farbe des Läufers, den Typ der beweglichen Figur, das Startfeld und den Zug des Spielers Nummer. Für Weiß sind die Startfelder b1, g1, c1 und f1. Für Schwarz sind sie es b8, g8, c8 und f8. Für jedes Spielerspiel zählt Elo+Chess das jeweilige Zuhause Felder zwischen Springern und Läufern, die je Spielerzug 10 bewegt wurden.

Dabei handelt es sich um eine Metrik, bei der „höher ist normalerweise besser“ gilt. Es wird nicht behauptet, dass jeder Minderjährige Die Figur muss sich immer um 10 Züge bewegen, aber bei sich entwickelnden Spielern ist das empirisch Der Zusammenhang ist klar: Spieler, die höhere Elo-Buckets erreichen, neigen dazu, sich zu entwickeln weitere kleinere Stücke früher.

7.2 Repeated Piece Moves in the Opening

Diese Metrik zählt die zusätzlichen frühen Züge, die aufgewendet wurden, um stattdessen die gleiche Figur noch einmal zu bewegen etwas Neues zu entwickeln. Die Zwischenberechnung verwendet die vorherige Zielfeld, das aktuelle Ausgangsfeld und die Zugnummer des Spielers. Wenn Die gleiche Figur wurde beim vorherigen Spielerzug bewegt und wird erneut innerhalb des Zuges bewegt Nach den ersten 10 Spielerzügen erhöht sich die Anzahl der wiederholten Züge.

Dabei handelt es sich um eine Metrik, bei der niedriger normalerweise besser ist. Es misst verschwendete Eröffnungstempi in a einfacher Weg. Einige wiederholte Bewegungen sind taktisch gerechtfertigt, aber im Maßstab Die Bucket-Kurve spiegelt die allgemeine Tendenz wider: stärkere Spieler am Ziel Levels erfordern zu Beginn weniger Züge, um dieselbe Figur wiederholt zu bewegen.

7.3 Queen Touched Before Move 10

Diese Metrik zählt die Züge der Dame vor dem 10. Zug des Spielers. Die Zwischenstufe Bei der Berechnung werden der Typ der beweglichen Figur und die Zugnummer des Spielers verwendet. Wenn das bewegliche Stück ist die Königin und die Zugzahl des Spielers beträgt höchstens 10, die Anzahl erhöht sich.

Dies gilt auch für den Zielbereich Elo+Chess. Früh Damezüge können Material gewinnen oder Zugeständnisse erzwingen, aber viele sich entwickelnde Spieler Bringen Sie die Königin zu früh zum Vorschein, verlieren Sie Tempo durch Angriffe auf die Königin und verzögern Sie Entwicklung von Kleinfiguren und Königssicherheit.

7.4 Castling and King Safety

Rochademetriken verwenden Rochadeflaggen, Rochadeseite, Rochadezugnummer und der Brettzustand nach der Rochade. Beispiele hierfür sind, ob der Spieler rochierte, ob der Spieler bis zum 10. Zug rochierte, ob der Spieler am Königsflügel rochierte oder am Damenflügel, ob die Türme nach der Rochade verbunden waren und ob die Türme abgelegt wurden waren nach der Rochade offen oder halboffen.

Diese Kennzahlen werden nicht alle auf die gleiche Weise interpretiert. Fertigstellung der Rochade und Rochaden nach 10 Zügen sind im Zielbereich im Allgemeinen höher, desto besser. Die Rochade an der Königinseite oder die Rochade in offene Dateien können stärker vom Kontext abhängig sein. Deshalb kodiert Elo+Chess nicht einfach eine moralische Regel fest; es baut Eimer Kurven und prüft, welche Beziehungen stabil genug für ein Coaching sind.

8 Schaufelkurven und manuelle Inspektion

Nachdem die Spieler-Spiel-Metriken berechnet wurden, gruppiert Elo+Chess sie nach Lichess Elo Bucket und berechnet Bucket-Durchschnitte. Der Bucket-Durchschnitt ist der empirische Rohwert Benchmark-Punkt für diese Metrik. Im Bericht wird dann eine Trendlinie verwendet Weisen Sie Benutzerwerten Noten im Elo-Stil zu.

Die Kurven werden vor der Verwendung manuell überprüft. Eine Metrik ist nützlicher wenn die Bucket-Mittel eine kohärente Beziehung über den Bewertungsbereich hinweg zeigen durch das Produkt bedient. Eine Metrik ist weniger nützlich, wenn sie verrauscht, flach oder hoch ist nicht monoton oder durch Sonderfälle bedingt, die einem Benutzer schwer zu erklären sind.

Dieser manuelle Inspektionsschritt ist besonders wichtig oberhalb des Elo+Chess-Obermaterials Abschaltung. Stärkere Spieler wissen, wann sie Eröffnungsprinzipien brechen müssen und tun dies auch effektiv. Oberhalb des Anfänger- bis frühen Fortgeschrittenenbereichs ist die Beziehung Die Beziehung zwischen einfachen Gewohnheiten und Bewertungen verflacht oft oder verändert ihre Form. Das Produkt betont daher den Bereich, in dem die empirischen Beziehungen stark sind und pädagogisch nützlich.

9 Beispielhafte Öffnungskurven für den gesamten Bereich

Abbildung 5 zeigt vier 3+0-Blitz-Eröffnungsprinzipkurven. Graue Punkte zeigen den breiteren Bewertungsbereich Lichess. Blaue Punkte und der blaue Trend Die Linie zeigt den Bereich 600–1600 Lichess, der vom Haupttrainer Elo+Chess verwendet wird Maßstäbe. Die gestrichelte rote Linie markiert den oberen Grenzwert.

Das gleiche Muster ist im gesamten 10+0-Schnellbereich und im längeren Schnellbereich sichtbar Artefakte. Die zentrale Modellierungsentscheidung bleibt unverändert: empirisch nutzen Stabiler Bereich für Coaching-Ansprüche und vermeiden Sie es, so zu tun, als wäre es eine einfache Gewohnheitskurve bleibt für immer gleichermaßen informativ.

Das gleiche Muster erscheint numerisch in Tabelle 5. Pisten sind wird als metrische Wertänderung pro 100 Elo-Punkte angegeben. Rochade, Nebenstück Entwicklung, wiederholte Figurenbewegung und frühe Damenbewegung sind alle klar erkennbar Bewegung zwischen 600 und 1600. Oberhalb von 1600 sind die Steigungen dort deutlich geringer Beispiele, was mit der Vorstellung übereinstimmt, dass höher bewertete Spieler scheitern Grundprinzipien gezielter und erfolgreicher umzusetzen. Ein zweiter Mechanismus ist Sättigung. Für einige Gewohnheiten erreicht die Metrik

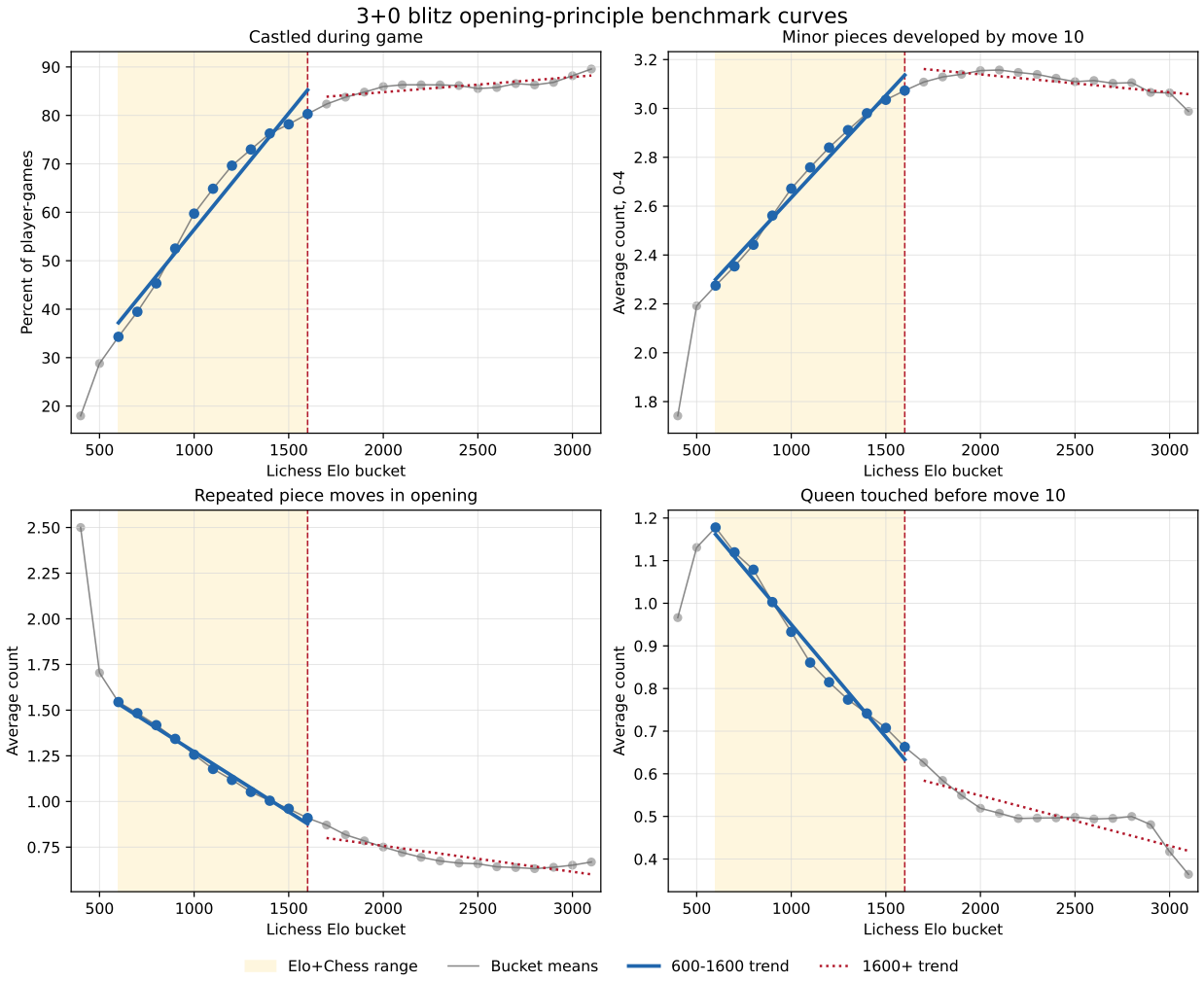


Figure 5: Ausgewählte Eröffnungsprinzip-Benchmarkkurven für 3+0-Blitz. Die Beziehungen sind im Elo+Chess-Coaching-Bereich am stärksten und nützlichsten. Oberhalb des oberen Cutoffs verflachen mehrere Beziehungen erheblich.

eine praktische Obergrenze oder eine Grenze Spot: Starke Spieler haben die Gewohnheit bereits gelernt, es gibt also keinen Grund dafür Die Kurve geht weiter steil nach oben. Mehr ist danach auch nicht immer besser dieser Punkt. Die Daten glätten daher beide, da stärkere Spieler ausscheiden können Regeln in konkreten Positionen und weil viele Grundgewohnheiten bereits erreicht sind ihren nützlichen Betriebsbereich.

10 Warum der obere Grenzwert wichtig ist

Der Zielbenutzer Elo+Chess ist ein Anfänger bis zu einem fortgeschrittenen Spieler. Für diese Grundgewohnheiten eines Spielers sind stark mit der Wertung verbunden: Figuren entwickeln, Vermeiden Sie wiederholte frühe Figurenzüge, vermeiden Sie unnötige frühe Damenzüge und Burgen Verbinden Sie rechtzeitig Türme, verwalten Sie das Material und vermeiden Sie es, Teile lose zu lassen oder hängend.

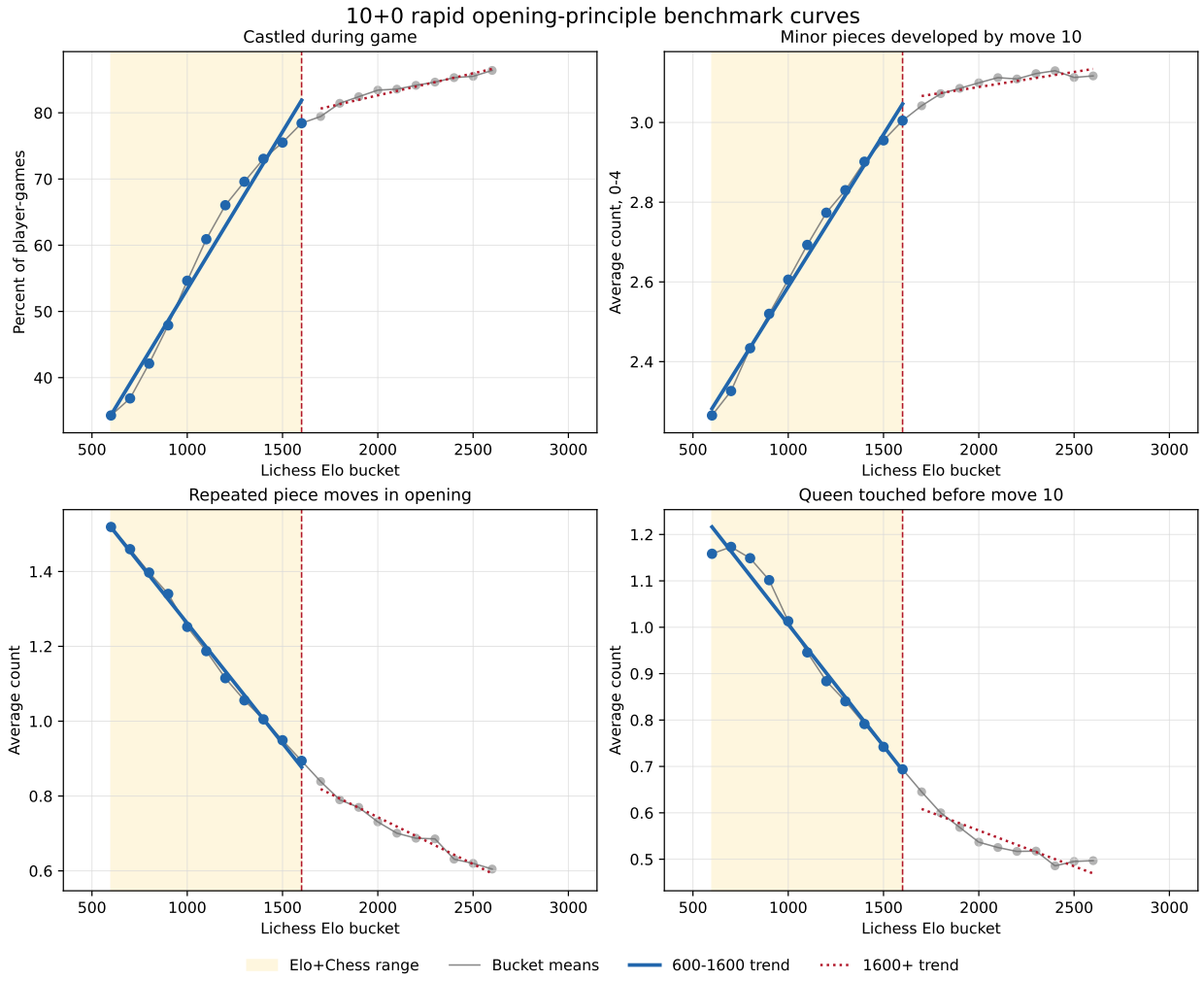


Figure 6: Ausgewählte Eröffnungsprinzip-Benchmarkkurven für 10+0 Schnellschachspiele. Die Beziehungen sind im Elo+Chess-Coaching-Bereich am stärksten und nützlichsten. Oberhalb des oberen Cutoffs verflachen mehrere Beziehungen erheblich.

Bei höheren Bewertungen werden dieselben Oberflächengewohnheiten weniger diagnostisch. Ein starker Der Spieler kann die Rochade verzögern, weil das Zentrum geschlossen ist. Bewegen Sie die Königin früher weil die Stellung eine konkrete Taktik enthält, oder dieselbe Figur zweimal bewegen weil es Tempo gewinnt oder ein strukturelles Zugeständnis erzwingt. Die rohe Gewohnheit immer noch hat Bedeutung, aber die Beziehung zwischen der Gewohnheit und der Bewertung wird stärker bedingt. Aus diesem Grund verwendet Elo+Chess einen oberen Cutoff für sein Haupt-Coaching Ansprüche und warum Benutzer oberhalb dieses Bereichs vor der engen Beziehung gewarnt werden zwischen prinzipiellen Gewohnheiten und Bewertungsbrüchen.

11 Zweiseitige taktische Metriken und Perzentilziele

Einige Metriken haben die entgegengesetzte Form zu den Beispielen mit dem Eröffnungsprinzip. Gabeln sind ein nützliches Beispiel, da das Ereignis von Natur aus zweiseitig ist. Ein Spieler, der

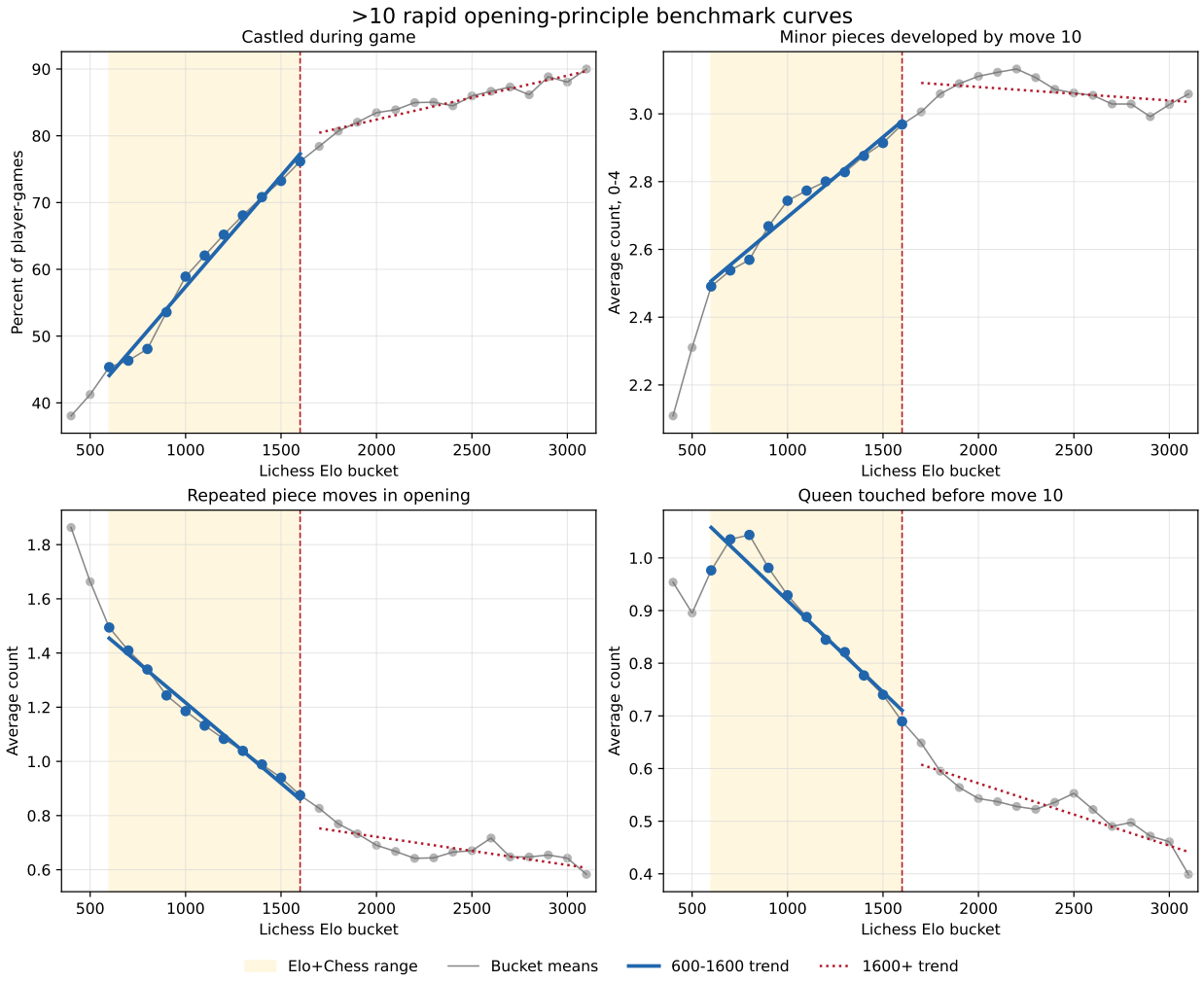


Figure 7: Ausgewählte Eröffnungsprinzip-Benchmark-Kurven für längere Schnellpartien.

Findet und konvertiert mehr nützliche Forks als andere, tut etwas Gutes, aber Die Populationsrate der konvertierten Gabeln sinkt mit der Bewertung, da sie stärker ist Gegner lassen weniger Fork-Möglichkeiten zu und verteidigen sich gegen taktische Bedrohungen aktiver.

Abbildung 8 zeigt diesen Effekt. Beide umgebauten Gabeln von der Reporter und umgewandelte Forks des Gegners werden seltener Ratings steigen. Das bedeutet nicht, dass die Ausführung von Forks schlecht ist. Es bedeutet das Rohereignis Die Rate wird teilweise durch die Fehlerrate des Gegners und die Spielstruktur gesteuert. Für Metriken Auf diese Weise verwendet Elo+Chess daher Peer-History-Perzentile innerhalb des den eigenen Elo-Bucket des Spielers, anstatt sich nur auf einen globalen Bucket-Mean-Trend zu verlassen.

Um diese Perzentilziele festzulegen, befragt Elo+Chess Spieler nach Spieltyp und Elo-Bucket verwendet dann die letzten Qualifikationsspiele für jeden ausgewählten Spieler. Die Der aktuelle Zielsatz verwendet 11 Eimer von 600 bis 1600, 200 Spieler pro Eimer und bis zu 50 aktuelle Qualifikationsspiele pro Spieler für jede Produktion Spieltyp. Dies ergibt ungefähr 2.200 Spielerbeispiele pro Spieltyp 110.000 Spieler-Spiel-Zeilen pro Spieltyp für Perzentil innerhalb des Buckets Verteilungen.

Figure 9 veranschaulicht das resultierende Ziel Verteilung für die umgerechnete Fork-Rate des

Game type	Metric	600–1600 slope per 100 Elo	1600+ slope per 100 Elo
1+0 bullet	Castled during game	4.137	0.550
	Minor pieces developed by move 10	0.063	0.009
	Repeated piece moves in opening	-0.051	-0.026
	Queen touched before move 10	-0.040	-0.011
3+0 blitz	Castled during game	4.804	0.313
	Minor pieces developed by move 10	0.084	-0.007
	Repeated piece moves in opening	-0.066	-0.014
	Queen touched before move 10	-0.053	-0.012
10+0 rapid	Castled during game	4.753	0.665
	Minor pieces developed by move 10	0.076	0.008
	Repeated piece moves in opening	-0.064	-0.025
	Queen touched before move 10	-0.052	-0.015
>10 minute rapid	Castled during game	3.320	0.659
	Minor pieces developed by move 10	0.047	-0.004
	Repeated piece moves in opening	-0.059	-0.010
	Queen touched before move 10	-0.035	-0.012

Table 5: Ungefähre Eimerkurvensteigungen für die vier beispielhaften Öffnungsprinzipien Metriken für verschiedene Produktionsspieltypen.

Game type	Elo buckets	Players per bucket	Games per player	Player samples
1+0 bullet	11	200	49–50	2200
3+0 blitz	11	200	27–50	2200
10+0 rapid	11	200	37–50	2200
>10 minute rapid	11	200	47–50	2200

Table 6: Aktuelle Peer-History-Perzentil-Zielsätze. Das sind Stichproben aus der Spielerhistorie, die zur Schätzung der Verteilungen innerhalb des Buckets verwendet werden Metriken im Perzentilstil.

gesampelten Spielers im 3+0-Blitz. Jede Box ist eine Verteilung innerhalb des Buckets über die ausgewählten Spieler, kein Bucket-Mittelwert über alle Spieler hinweg einzelne Spiele. In der Berichts-Benutzeroberfläche wird der Wert eines Benutzers mit dem relevanten Wert verglichen Feld für den zugeordneten Elo-Bucket des Benutzers. Dies ist die Grundlage für Aussagen wie ob die konvertierte Fork-Rate eines Spielers im Vergleich zu seinen Mitspielern hoch oder niedrig ist gleichen Niveau, auch wenn die Ereignisrate der Gesamtbevölkerung mit der Bewertung abnimmt.

12 Replikationskosten

Die Methode ist grundsätzlich reproduzierbar, aber nicht leichtgewichtig. Wiederaufbau Die Benchmark-Kurven erfordern:

- Zugriff auf große monatliche Spielverlaufsdateien Lichess;
- geschichtete Filterung nach Spieltyp, Bewertungsgruppe und Teilnahmeberechtigung Regeln;
- ein Parser für rechtliche Schritte und ein Board-State-Evaluator, der erweitert werden kann Dutzende Millionen Umzugszustände;

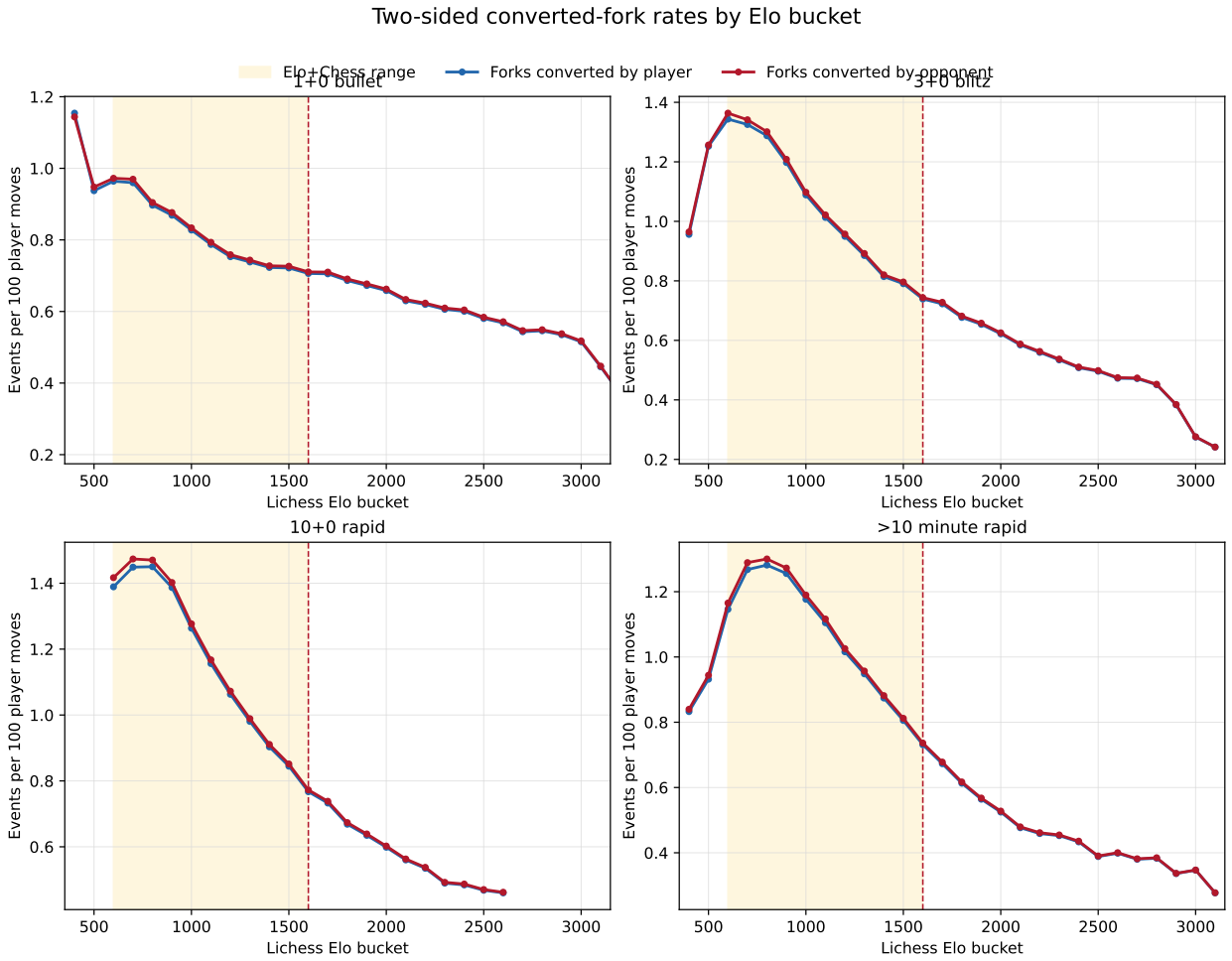


Figure 8: Zweiseitig umgerechnete Gabelzinken durch Elo-Schaufel. Konvertierte Fork-Ereignisse fallen bei höheren Bewertungen, weil beide Seiten besser verteidigen und weniger saubere Gabeln haben Konvertierungen sind möglich.

- Feature-Logik für Eröffnung, Material, Bauernstruktur, Königssicherung, Turm Aktivität, Taktik, Zeitmanagement und Spielergebnisse;
- Aggregation vom Bewegungsstatus zu Spieler-Spielfunktionen;
- Zusammenfassungen auf Bucket-Ebene für jede Berichtsmetrik;
- Manuelle Überprüfung jeder Benchmark-Kurve, bevor sie verwendet wird ein Coachingbericht.

Für 3+0-Blitz bedeutet das fast eine halbe Million Partien, also fast eine Million Spieler-Spiel-Zeilen und mehr als 28 Millionen Bewegungs-Zustands-Zeilen. Über die vier Benchmark-Spieltypen enthalten die aktuellen Benchmark-Artefakte mehr als 2,55 Millionen Spiele, mehr als 5,10 Millionen Spieler-Spielreihen und mehr als 147 Millionen Move-State-Zeilen. Die Skala ist von zentraler Bedeutung für den Wert des Benchmark: Der Bericht zeigt Benutzern, was tatsächlich in gespielten Spielen passiert von Spielern, die ihrem Niveau nahekommen, anstatt sich nur auf allgemeine Schachparolen zu verlassen.

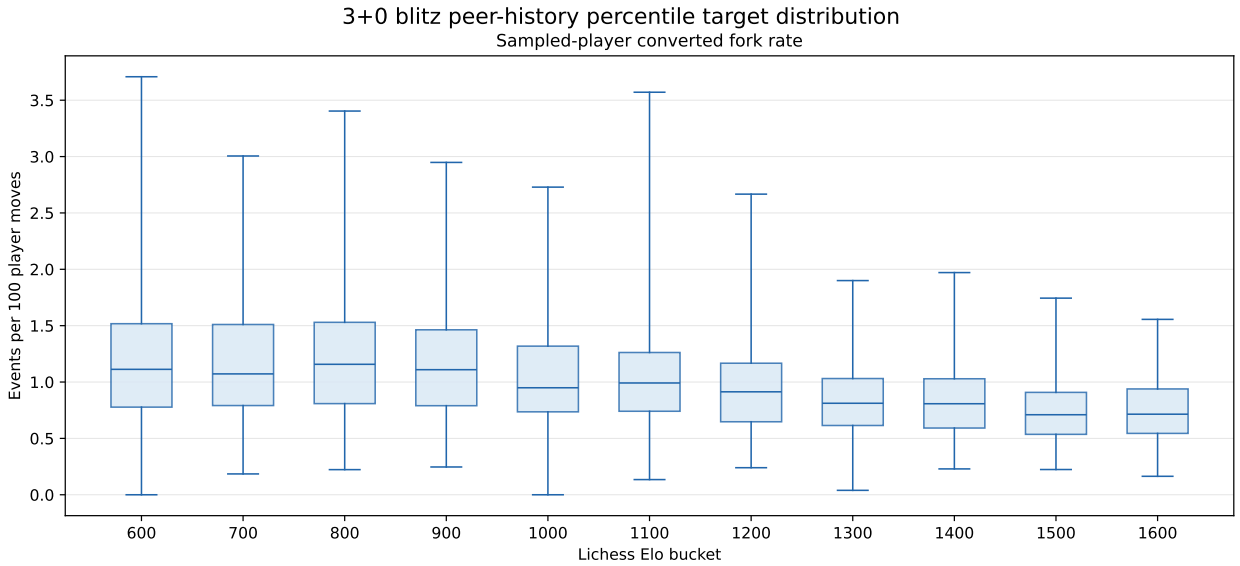


Figure 9: Beispiel für eine 3+0-Blitz-Peer-History-Zielverteilung für einen gesampelten Spieler umgerechnete Gabelrate. Perzentile werden innerhalb des Elo-Buckets berechnet, also im Bericht vergleicht einen Spieler mit Gleichaltrigen, die weitgehend ähnlichen taktischen Möglichkeiten gegenüberstehen Tarife.

13 Einschränkungen

Bei den Benchmark-Kurven handelt es sich um empirische Beschreibungen, nicht um Kausalbeweise. Ein Spieler erhält keine Bewertung, indem er lediglich einen Bucket-Mittelwert anpasst. Metriken sind es auch unvollständige Zusammenfassungen komplexer Stellungen: Ein früher Damenzug kann ausgezeichnet sein in einer Position und rücksichtslos in einer anderen. Der Zweck der Kurven besteht darin Identifizieren Sie wiederkehrende Gewohnheiten und vergleichen Sie sie mit dem Verhalten von Gleichaltrigen, nicht mit dem Verhalten von Gleichaltrigen Ersetzen Sie die Berechnung oder das Coaching-Urteil.

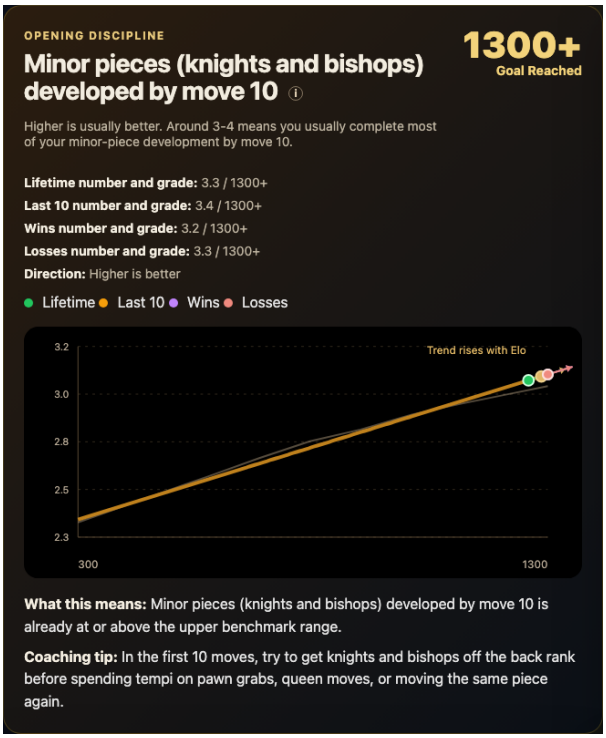
Bei den Kurven handelt es sich ebenfalls um Kurven im Lichess-Maßstab. Wenn Benutzer Chess.com-Spiel mitbringen Historien ordnet Elo+Chess den relevanten Rating-Bucket der Lichess-Benchmark zu Skalierung mithilfe des separaten plattformübergreifenden Zuordnungsmodells. Diese Zuordnungsebene ist beschrieben im begleitenden Forschungsbericht Elo+Chess zu Lichess-to-Chess.com Bewertungsumrechnung: Estimating Cross-Platform Chess Rating Mappings with Modale Regression. Diese Zuordnung sollte nicht als erhaltendes Perzentil verstanden werden Rang. Es bewahrt eine geschätzte Bewertungsäquivalenz zwischen den beiden Bewertungen Systeme; Perzentilrang bleibt plattformspezifisch, da der aktive Spieler Verteilungen sind unterschiedlich.

14 Grafiken zum Produktionsbericht

Der Produktionsbericht wandelt die Benchmark-Artefakte in Panels auf Metrikebene um Das kann gelesen werden, ohne die vollständige Build-Pipeline zu kennen. Figuren 10 and 11 zeigen drei Beispiele aus dem Live-Berichtsdesign Elo+Chess.

Die Öffnungsprinzip-Panels in Abbildung 10 Verwenden Sie die übliche Benchmark-Kurvendarstellung. Jede Karte beginnt mit dem Metrikfamilie und Metrikname, dann gibt es eine kurze Interpretationsregel. Die Der große Grad in der oberen rechten Ecke ist absichtlich vom Rohteil getrennt Metrikwert: Er beantwortet die Frage des Benutzers: „Wie bewertet diese Gewohnheit?“ im Vergleich zum Benchmark?”, während in den Zeilen darunter die Messwerte erhalten bleiben. Die vier Zusammenfassungszeilen zeigen „Lebenszeit“, „Letzte 10“, „Siege“ und „Verluste“. Das macht das Panel-Diagnose und nicht nur dekorativ. Wenn Lifetime und Last 10 übereinstimmen, Die Gewohnheit ist wahrscheinlich stabil. Wenn Gewinne und Verluste voneinander abweichen, ist die Kennzahl möglicherweise unterschiedlich ergebnisabhängig oder an den Spielstand gebunden und nicht eine einfache Gewohnheit.

Der Diagrammbereich verwendet einen zurückhaltenden dunklen Hintergrund, eine goldene Trendlinie und Farben Markierungen für die vier Scheiben. In der Trendanmerkung wird explizit angegeben, ob Die Benchmark-Beziehung steigt oder fällt mit Elo. Das ist wichtig, weil Bei einigen Kennzahlen gilt das Prinzip „je höher desto besser“, z. B. die Entwicklung kleinerer Teile, während Bei anderen gilt: „Niedriger ist besser“, wie zum Beispiel die frühe Königinnenbewegung. Die Karte daher fordert den Benutzer nicht auf, allein aus der Neigung auf die Richtung zu schließen. Das Finale Die Zeilen „Was das bedeutet“ und „Coaching-Tipp“ übersetzen das Benchmark-Ergebnis in einen konkreten Schachunterricht unter Wahrung des empirischen Kontextes umzuwandeln.



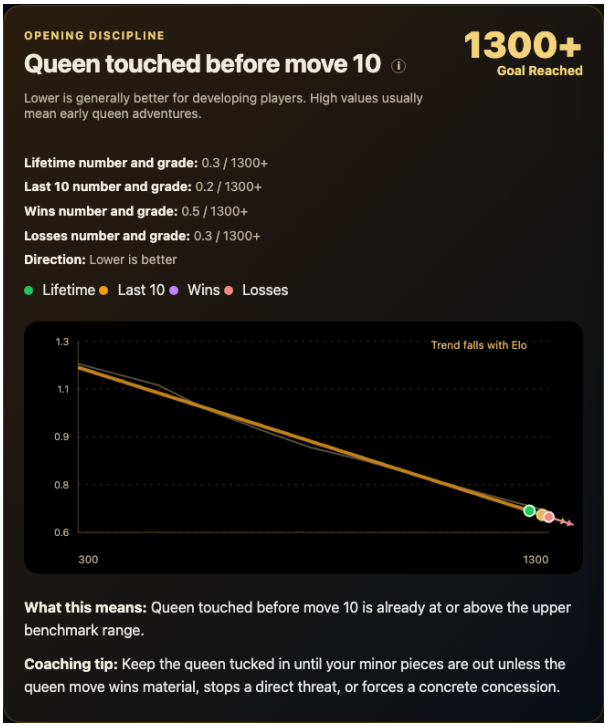


Figure 10: Herstellung von Paneelen nach dem Öffnungsprinzip. Die gleiche visuelle Grammatik behandelt a „höher ist besser“-Metrik, Nebenstückentwicklung und eine niedrigere ist besser-Metrik, frühe Königinnenbewegung. Beide Panels zeigen die Lebensdauer, die letzten 10, Siege und Verluste So kann der Bericht stabile Gewohnheiten von aktuellen oder ergebnisabhängigen Gewohnheiten trennen Muster.

Figure 11 zeigt den Perzentilstil Darstellung, die für Metriken verwendet wird, deren Rohrate nicht gut durch a dargestellt wird einziger globaler Trend. Winning Forks sind ein taktisches

Beispiel: die Bevölkerungsrate der umgewandelten Gabeln ändert sich sowohl mit den Fähigkeiten des Spielers als auch mit der Fehlerrate des Gegners. Die Die Karte vergleicht daher den Wert des Benutzers mit der Historie anderer Spieler in der zugeordneter Elo-Eimer. Der Boxplot zeigt die Peer-Verteilung, einschließlich der Interquartilbereich, Median und Whiskers. Die Markierung des Benutzers wird auf dem gezeichnet gleiche Skala und sowohl mit Rohwert als auch Perzentilrang beschriftet. Angrenzend Vergleichs-Buckets bleiben in stummgeschalteter Form sichtbar, sodass der Benutzer sehen kann, dass die Der Peer-Bucket ist lokal und nicht universell.

Dieses Design macht Perzentilmetriken überprüfbar. Der Benutzer sieht den Rohwert, das Perzentil, der zum Vergleich verwendete Peer-Bucket und die übersetzte Chess.com-zu-Lichess-Bucket-Zuordnung. Das gleiche Leben, die letzten 10, Siege und Verlustzeilen bleiben erhalten, sodass taktische Perzentilwerte weiterhin überprüft werden können für Aktualitätseffekte und Ergebnisabhängigkeit.

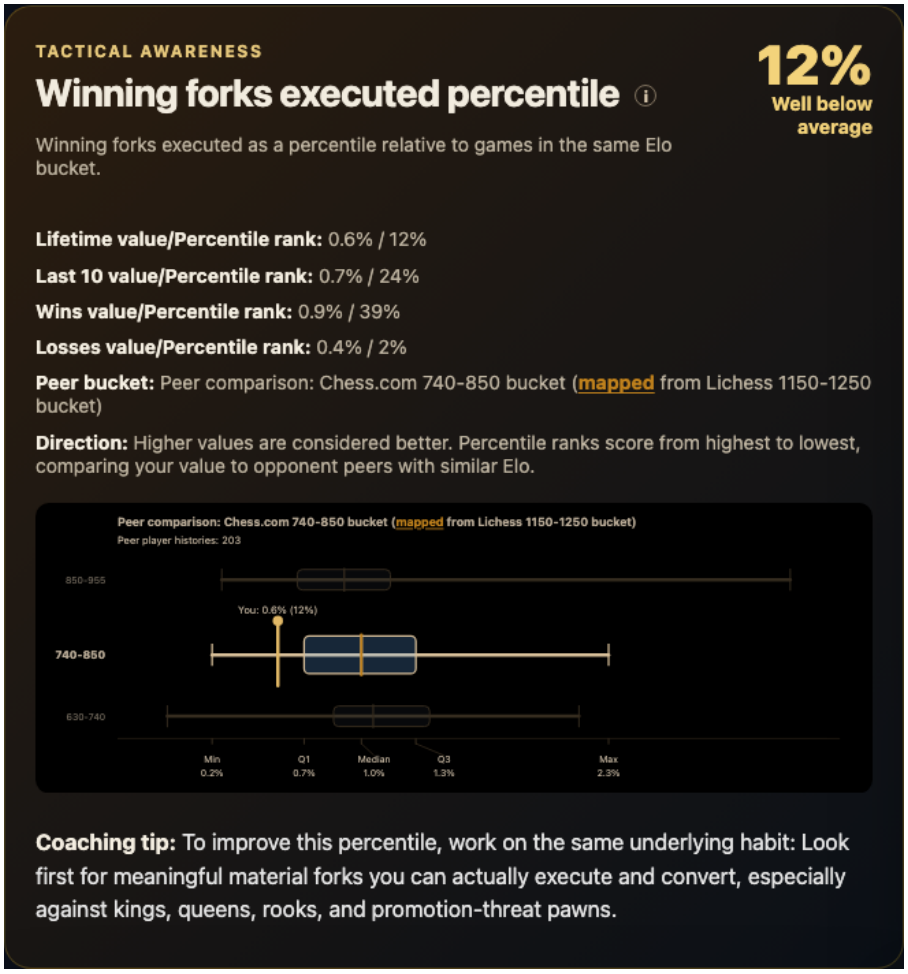


Figure 11: Produktionsperzentil-Panel für Gewinngabeln ausgeführt. Der Boxplot zeigt die im Bericht verwendete Gleiche-Bucket-Peer-Verteilung, während das Gold Der Marker zeigt den Rohwert und den Perzentilrang des Benutzers an.

15 Abschluss

Elo+Chess-Benchmark-Berichte basieren auf einer umfangreichen Move-State-Berechnung geschichtete Lichess-Proben. Jede Metrik beginnt als eine Reihe konkreter Board-Zustände und Bewegungszustandsvariablen werden zu einem Spieler-Spiel-Feature und werden dann gemittelt von Elo Bucket, um eine empirische Benchmark-Kurve zu bilden. Das Öffnungsprinzip Beispiele zeigen, warum dieser Ansatz nützlich ist: Einfache Schachgewohnheiten haben starke, sichtbare Beziehungen mit der Bewertung im Zielbereich, während diese Beziehungen werden oberhalb der Obergrenze oft flacher oder bedingter. Das ist der zentrale Grund, warum Elo+Chess seine Coaching-Ansprüche auf das Rating konzentriert Bereich, in dem die Daten sie am deutlichsten unterstützen.