

Player-Specific Policy Adaptation in Elite Chess with MoE-LoRA and Limited Search

Loris Sogliuzzo^[0009-0004-3577-4951], Aloïs Rautureau^[0009-0007-8507-2167],
and Éric Piette^[0000-0001-8355-636X]

ICTEAM, UCLouvain, Louvain-la-Neuve, Belgium
`loris.sogliuzzo@ensta.fr`

Abstract. The historical trajectory of chess artificial intelligence has predominantly focused on superhuman playing strength, yielding policy models whose decision paths optimize purely for game-theoretic value rather than fidelity to human behavior. While recent human-centric neural architectures successfully model population-wide behavior across Elo rating intervals, personalizing policies to capture the individual styles of elite players and World Champions remains an open challenge. Furthermore, standard evaluations rely on deterministic top-1 move accuracy, which fails to capture the natural stochasticity and strategic variance inherent in human play.

This paper investigates player-specific behavioral modeling across a cohort of 14 elite chess players whose international careers began in the 20th century, encompassing six World Champions and eight World Championship contenders or Candidates-level elite players. Because fine-tuning player embeddings over a frozen backbone gives limited control over decision boundaries, a parameter-efficient Mixture of Experts with Low-Rank Adaptation framework is introduced, applying dynamically gated residual adjustments directly to the network logit space. To curb repetition loops and mitigate tactical blunders without collapsing human stylistic variance, this architecture is coupled with budget-constrained Monte Carlo Tree Search regularized by top- p Nucleus Pruning. Behavioral alignment is evaluated strictly on recurring board positions with verified empirical depth of at least five observations per player to avoid single-observation Dirac targets. Experimental results on recurring positions show that this framework improves top-1 move accuracy from 43.61% to 61.95%, reduces the tactical gap Δ_{CPL} from 10.49 to 2.03, and minimizes state-averaged Jensen-Shannon Divergence to 0.2350 among neural models when regularized by search.

Keywords: Human-AI Alignment · Behavioral Modeling · Policy Adaptation · Mixture of Experts · Monte Carlo Tree Search · Jensen-Shannon Divergence.

1 Introduction

The development of artificial intelligence in chess has long been benchmarked by competitive supremacy, from Deep Blue’s historical victory over Garry Kas-

parov [1] to AlphaZero achieving superhuman mastery through self-play reinforcement learning [20]. However, superhuman engines operate under pure objective optimization, frequently selecting moves that diverge from human intuition and heuristics [4]. Consequently, a parallel research domain has emerged to develop human-aligned models capable of replicating human decision patterns across diverse skill levels [10, 23].

While models such as Maia-2 [23] capture population-wide behavior across continuous rating dimensions, modeling individual playing styles remains largely unaddressed for world-class elite players. Recent personalization frameworks, notably Maia4All [24] and individualized modeling studies [11], predominantly target amateur and intermediate players. Modeling elite players and World Champions introduces distinct challenges, as these competitors operate near the human skill ceiling where Elo rating conditioning no longer suffices to differentiate individual decision boundaries [12], while exhibiting sharply divergent strategic signatures that range from dynamic tactical sacrifices to subtle positional maneuvering.

A fundamental difficulty lies in the adaptation mechanism itself. Prior work established that fine-tuning input player embeddings on a frozen neural backbone provides insufficient capacity to displace shared decision boundaries [22]. Furthermore, standard behavioral evaluations rely on top-1 move prediction accuracy [10]. This metric enforces a deterministic assumption that overlooks valid alternative moves and penalizes strategic diversity. In addition, evaluating full career game records suffers from severe empirical support sparsity, as most board configurations occur only once in a player’s history, collapsing empirical targets into single-action Dirac vectors.

This work presents a comprehensive behavior generation and stylometric evaluation framework tailored to a historical cohort of 14 elite 20th-century players. The contributions of this study are threefold. First, a parameter-efficient Mixture of Experts [19] with Low-Rank Adaptation [7] adapter operating directly on the logit space is introduced to overcome the limitations of input embedding fine-tuning. Second, a regularized search mechanism, combining budget-constrained MCTS [3] with top- p Nucleus Pruning [6], is integrated to reduce the risk of tactical loop pathologies and opening blunders while preserving human behavioral alignment. Third, a support-constrained evaluation benchmark requiring at least five observations per position is formalized, eliminating single-action Dirac collapses and demonstrating that MoE-LoRA provides substantial behavioral and distributional alignment across historical elite players. While a preliminary study explored player-specific embedding adaptation and search for historical chess champions [22], the present work substantially expands upon this foundation by introducing a modular logit-level MoE-LoRA architecture and formalizing an exact-state distributional evaluation on recurring positions.

2 Related Work

The pursuit of human-aligned artificial intelligence in board games has evolved from rating-stratified prediction to fine-grained behavioral modeling. The foun-

dational Maia framework [10] demonstrated that convolutional neural networks trained on human games categorized by discrete Elo rating brackets could predict human decisions more accurately than depth-limited superhuman engines such as Stockfish. Building upon this paradigm, Maia-2 [23] introduced a unified ResNet-ViT architecture equipped with Skill-Aware Attention, treating board evaluations as Markov Decision Processes conditioned simultaneously on active turn and opponent ratings. Recent foundational advances, such as Chessformer [13], have further expanded unified transformer architectures across massive multi-task chess modeling benchmarks. At the individual level, McIlroy-Young et al. [11] examined personalized move prediction on online platforms, while Maia4All [24] explored few-shot adaptation for intermediate players via Prototype Matching Networks. Beyond classification, complementary approaches have investigated disentangled latent representations of playing style [2] and generative stylometry [14, 12]. However, existing methods focus primarily on population-wide rating intervals or low-tier personalization, leaving the challenge of distinguishing elite players operating at identical skill ceilings largely unaddressed.

In parallel, parameter-efficient fine-tuning and modular routing have emerged as dominant solutions for adapting large pre-trained representations without full retraining. Low-Rank Adaptation, denoted LoRA, [7] enables efficient task transfer by constraining weight updates to low-rank decomposition matrices. Combining LoRA with Mixture of Experts, denoted MoE, routing mechanisms [19, 5] provides dynamic specialization, allowing gated combinations of modular sub-networks conditioned on contextual embeddings. While recent autoregressive language modeling approaches such as Mixture of Masters [5] route discrete PGN move tokens through specialized language heads, the present work addresses vision-based spatial policy-value backbones such as Maia-2 by learning residual logit-space expert mixtures directly over spatial board encodings. Within behavioral game playing, unconstrained policy networks frequently suffer from tactical blind spots, cyclic piece shuttling, and premature repetitions [21]. Integrating shallow, regularized tree search, such as Monte Carlo Tree Search, denoted MCTS, [3] constrained by top- p Nucleus Pruning [6], restores tactical soundness while preventing the search from converging toward engine-like play [25, 16, 8].

3 Cohort Characterization and Benchmark Design

Historical tournament game records were compiled from the public archives of Chessgames¹. The candidate pool was defined according to two central methodological criteria. First, the investigation specifically targets elite mastery, focusing exclusively on World Champions and Candidates-level players to examine decision-making at the highest competitive tier, where Elo rating ceases to differentiate individual playing styles [12]. Second, an adequate volume of recorded games is required to provide sufficient empirical depth for statistical training and evaluation, setting a practical threshold of at least 2,000 recorded classical games. This volume helps mitigate empirical support sparsity across player actions,

¹ <https://www.chessgames.com>, accessed February 2026

Table 1: Dataset characterization across the cohort of 14 historical elite players.

Player	Cohort Category	Career Span	Total Games	Positions	White (%)
A. Alekhine	World Champion	1905–1946	2,297	87,700	67.8%
M. Tal	World Champion	1949–1992	3,231	103,679	52.5%
T. Petrosian	World Champion	1942–1983	2,211	68,403	51.5%
A. Karpov	World Champion	1961–2022	4,500	153,999	54.0%
G. Kasparov	World Champion	1973–2025	3,098	96,202	55.2%
V. Anand	World Champion	1984–2026	4,528	165,823	51.0%
A. Beliavsky	Candidates Finalist	1967–2020	2,834	108,668	50.6%
B. Larsen	Candidates Finalist	1948–2008	2,896	115,589	50.4%
J. Timman	Candidates Finalist	1964–2019	4,166	143,462	49.9%
L. Portisch	Candidates Finalist	1952–2014	3,447	117,977	50.4%
N. Short	World Championship Match	1974–2025	3,256	118,327	51.7%
U. Andersson	World Elite / Candidate	1967–2023	3,062	100,619	50.1%
V. Ivanchuk	Candidates Finalist	1983–2025	4,349	163,249	50.1%
V. Korchnoi	World Championship Match	1945–2015	5,332	195,682	51.1%
Total	14 Elite Players	—	49,207	1,739,379	52.2%

preventing distorted probability distributions during the stylometric evaluation via Jensen-Shannon Divergence. This data sufficiency requirement naturally excludes 19th-century masters whose tournament archives remain too sparse. Furthermore, 21st-century active players were deliberately excluded in order to focus on historical preservation, capturing distinct classic human heuristics prior to the widespread influence of modern neural chess engines on contemporary elite player preparation. Positional records were systematically verified to filter out incomplete games, corrupted move sequences, and notation anomalies, compiling all publicly available games on Chessgames for each of the 14 elite players.

As summarized in Table 1, the cohort comprises six World Champions and eight World Championship contenders or Candidates-level elite players, totaling 49,207 games and 1,739,379 state-action observations. Rather than deduplicating board states, the dataset intentionally retains all move-decision instances across each master’s recorded career, which is essential to preserve move frequencies and model faithful empirical action distributions. All models are evaluated using 5-fold cross-validation partitioned at the game level. Folds are assigned once per game over the union of cohort games, so the 2,748 games contested between two cohort members fall in the same fold for both sides, preventing the shared trajectory from leaking into the training pool via the opponent.

Evaluating probability distributions directly over raw career histories suffers from severe empirical support sparsity. As quantified in Table 2, an average of $96.71\% \pm 0.49\%$ of unique board configurations appear exactly once in a player’s recorded career, accounting for 81.39% of all recorded decision instances. Evaluating models on singleton positions provides little statistical reliability, as an isolated observation carries insufficient information to distinguish deliberate strategic preferences from accidental choices. Furthermore, when evaluating move distributions, single-visit configurations collapse empirical targets into degenerate

Table 2: Empirical board state frequency distribution across the player cohort. Unique state values report the inter-player mean $\pm$ sample standard deviation, while total moves reflect the aggregate share of move decisions.

	Occurrences per Player	Unique States (%)	Total Moves (%)
1 (Singleton)		$96.71 \pm 0.49\%$	81.39%
2		$1.66 \pm 0.25\%$	2.79%
3–5		$1.02 \pm 0.16\%$	3.07%
6–10		$0.33 \pm 0.06\%$	2.06%
> 10		$0.28 \pm 0.04\%$	10.69%

Dirac delta distributions, reducing probabilistic divergence metrics to trivial binary matches.

To ensure that move accuracy, tactical gap, and distributional alignment are measured on positions with sufficient empirical support, evaluation is restricted to recurring positions with verified empirical support ($n \geq 5$). For each target player p , the qualified evaluation set $\mathcal{S}_p^{(n \geq 5)}$ gathers all board positions visited at least five times throughout the player’s career,

$$\mathcal{S}_p^{(n \geq 5)} = \{s \in \mathcal{S}_p \mid N(s, p) \geq 5\}, \quad (1)$$

where $N(s, p)$ denotes the total number of visits to position s by player p . The threshold $n = 5$ is a pragmatic minimum-support convention. Below five observations the empirical distribution can take at most four distinct count patterns and remains dominated by sampling noise, while higher thresholds rapidly shrink the evaluable set, as only 0.77% of unique states qualify at $n \geq 5$.

4 Behavior Generation Pipeline

4.1 Base Architecture and Limitations of Embedding Fine-Tuning

The pre-trained Maia-2 model [23] serves as the underlying backbone. Board states are represented as spatial tensors $\mathbf{s} \in \mathbb{R}^{8 \times 8 \times 18}$ encoding piece positions, active turn, castling availability, and en-passant squares. A ResNet convolutional stem coupled with Vision Transformer spatial self-attention computes baseline action logits $\mathbf{z}_{\text{base}}$ and scalar position evaluations $v(s)$, as illustrated in Fig. 1.

As established in preliminary investigations [22], fine-tuning player embeddings at the input level while keeping the neural backbone frozen only partially alters decision boundaries. Modifying input embeddings only adjusts early feature maps, leaving the deep convolutional kernels and self-attention projections fixed. The resulting gain in top-1 accuracy remains modest.

4.2 Mixture of Experts LoRA Module

To perform direct logit-level intervention without retraining the full network, the PlayerMoEAdapter module is formulated, as shown in Fig. 2.

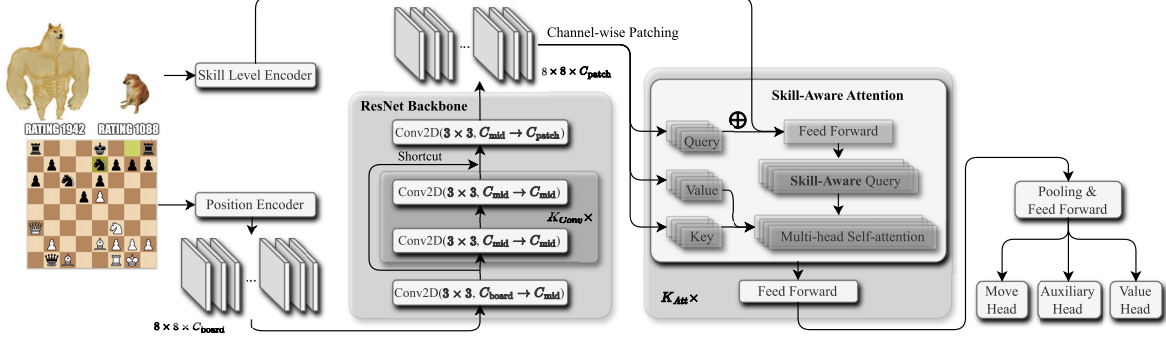


Fig. 1: Architecture of the baseline Maia-2 model, adapted from Tang et al. [23].

An intermediate representation $\mathbf{v}$ is extracted from the final normalization layer (`last_ln`) of Maia-2. The adapter maintains $N = 8$ parallel rank- r ($r = 8$) LoRA expert blocks E_i and a gating router network that are shared across all 14 players. Player-specific personalization is governed exclusively through low-dimensional routing embeddings $\mathbf{e}_p \in \mathbb{R}^{32}$. For target player p , $\mathbf{e}_p$ is concatenated with $\mathbf{v}$. The gating router network with a 128-dimensional hidden layer computes softmax routing weights $\mathbf{g} \in \mathbb{R}^N$ via

$$\mathbf{g} = \text{softmax}(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1[\mathbf{v}; \mathbf{e}_p] + \mathbf{b}_1) + \mathbf{b}_2) \quad (2)$$

where $\mathbf{W}_1 \in \mathbb{R}^{128 \times (\dim(\mathbf{v}) + 32)}$ and $\mathbf{W}_2 \in \mathbb{R}^{8 \times 128}$. Each expert applies down-projection $\mathbf{W}_{\text{down}}^{(i)} \in \mathbb{R}^{r \times \dim(\mathbf{v})}$ initialized from a normal distribution ($\sigma = 0.02$) and up-projection $\mathbf{W}_{\text{up}}^{(i)} \in \mathbb{R}^{d_{\text{out}} \times r}$ initialized to zero, where d_{out} corresponds to the output dimension of Maia-2's move policy head, scaled by $\alpha = 1.0$ as

$$E_i(\mathbf{v}) = (\mathbf{W}_{\text{up}}^{(i)} \mathbf{W}_{\text{down}}^{(i)} \mathbf{v}) \cdot \frac{\alpha}{r} \quad (3)$$

The final logits combine base policy logits with the dynamically weighted expert residual, formulated as

$$\mathbf{z}_{\text{final}} = \mathbf{z}_{\text{base}} + \sum_{i=1}^N g_i E_i(\mathbf{v}) \quad (4)$$

Training is conducted jointly across all cohort players within each fold with a batch size of 64 over 20 epochs using AdamW with a learning rate decaying from 5×10^{-3} to 10^{-5} via Cosine Annealing, weight decay of 10^{-2} , label smoothing of 0.05, and a gradient clipping norm of 1.0. Experiments were executed on a workstation equipped with an AMD Ryzen 7 5700X 8-core processor, 32 GB RAM, and an NVIDIA GeForce RTX 2060 GPU.

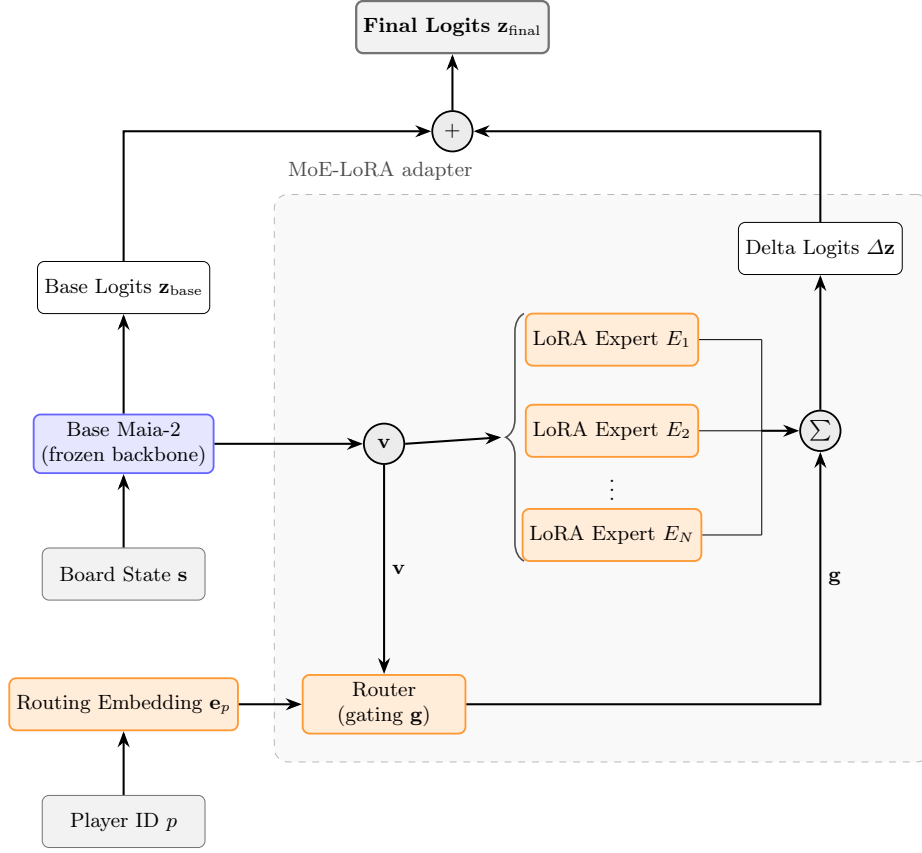


Fig. 2: Architecture of the parameter-efficient MoE-LoRA adapter module.

4.3 Tactical Search and Nucleus Regularization

Direct policy generation occasionally exhibits tactical failure modes, such as cyclic piece shuttling leading to premature threefold repetition draws, illustrated by moves such as 1. Nf3 Nf6 followed by 2. Ng1 Ng8, or immediate mate-in-1 blunders. To enforce tactical safety while preserving stylistic variance, a regularized search mechanism is integrated into the pipeline.

Budget-constrained MCTS performs tree search guided by the PUCT selection formula,

$$a^* = \arg \max_a \left(Q(s, a) + c_{\text{puct}} \cdot P(s, a) \cdot \frac{\sqrt{N(s)}}{1 + N(s, a)} \right) \quad (5)$$

with exploration constant $c_{\text{puct}} = 1.5$ and a shallow simulation budget strictly capped at $N = 50$ visits per decision. Restricting lookahead depth prevents drifting toward superhuman calculation while acting as a safeguard against immediate blunders, tactical oversights, and cyclical draw repetitions. Self-play

blunder and repetition rates were not measured directly. Search changes the benchmark metrics only slightly (Table 3). The search budget is therefore kept tight enough to preserve human-like strategic variance while maintaining sufficient visit density across candidate moves, yielding enough sample counts to construct a well-formed action probability distribution directly from normalized node visits. The Maia-2 baseline is conditioned on the top rating bin (≥ 2000 Elo) for both sides. Subject variation comes only from the adapters.

Before expanding candidate actions in search routines, legal moves are sorted in descending order by prior probability and truncated to the minimal subset K satisfying

$$\sum_{i=1}^K P(s, a_i) \geq p \quad \text{with } p = 0.95 \quad (6)$$

Nucleus Pruning strips noisy tail actions, preventing implausible node expansions during search. All MCTS variants evaluated here apply this truncation inside the search expansion. Variants labeled “N. + MCTS” additionally pre-filter the policy head, so the nucleus is applied twice. The second pass operates on the re-normalized distribution and is largely redundant, which is why the MCTS and N. + MCTS rows of Table 3 are near-identical.

5 Experimental Results and Discussion

5.1 Evaluation Protocol and Metric Formalization

Evaluating player-specific behavioral alignment requires assessing deterministic move fidelity, tactical consistency, and probabilistic calibration over candidate actions simultaneously. For each player p , let $D_p^{(n \geq 5)} = \{(s, a) \in D_p \mid s \in \mathcal{S}_p^{(n \geq 5)}\}$ denote the evaluation subset of recorded move decisions occurring on recurring board states.

Deterministic accuracy is measured via Top-1 Move Accuracy evaluated strictly over instances of recurring choice points,

$$\text{Acc} = \frac{1}{|D_p^{(n \geq 5)}|} \sum_{(s, a) \in D_p^{(n \geq 5)}} \mathbb{I}(a_{\text{model}} = a) \quad (7)$$

Tactical deviation is measured via the Mean Centipawn Loss Gap (Δ_{CPL}) against Stockfish evaluations V_{best} at search depth 10 on the same recurring subset,

$$\Delta_{\text{CPL}} = \left| \frac{1}{|D_p^{(n \geq 5)}|} \sum_{(s, a) \in D_p^{(n \geq 5)}} \text{CPL}(a) - \frac{1}{|D_p^{(n \geq 5)}|} \sum_{(s, a) \in D_p^{(n \geq 5)}} \text{CPL}(a_{\text{model}}) \right| \quad (8)$$

where $\text{CPL}(a) = \max(0, V_{\text{best}} - V(a))$. Unlike raw centipawn loss, which penalizes human-like heuristic concessions and rewards superhuman play, the gap Δ_{CPL} assesses mean tactical proximity, where a gap near zero indicates that the model

matches the average error magnitude of the target player. Stockfish depth-10 point estimates carry small run-to-run variation, of order 1 cp on cohort means, from engine transposition-table state across sequential analyses, so only differences of several centipawns should be interpreted. Since the evaluated recurring positions lie predominantly in openings, where most reasonable moves have near-zero centipawn loss with Stockfish at depth 10, small gaps are partly expected. The Δ_{CPL} metric compares means only and cannot establish that the model reproduces the player’s individual blunders, nor rule out memorization of frequent lines.

To evaluate probabilistic calibration beyond deterministic choices, the Jensen-Shannon Divergence [9] is evaluated pointwise across recurring positions and aggregated as an empirical mean. For a given board state s , let $P = \pi_{\text{player}}(\cdot \mid s)$ and $Q = \pi_{\text{model}}(\cdot \mid s)$ denote the empirical human and predicted action distributions over legal moves, with mixture $M = \frac{1}{2}(P + Q)$. The pointwise divergence is computed using base-2 logarithms,

$$\text{JSD}(P \parallel Q) = \frac{1}{2}D_{\text{KL}}(P \parallel M) + \frac{1}{2}D_{\text{KL}}(Q \parallel M) \quad (9)$$

From an information-theoretic standpoint, $\text{JSD} \in [0, 1]$ quantifies the shared information entropy between distributions, bounded between 0 bits for identical probabilistic preferences across all candidate moves and 1 bit for strictly disjoint candidate move sets. The overall model divergence is computed as the mean divergence across qualified states,

$$\overline{\text{JSD}} = \frac{1}{|\mathcal{S}_p^{(n \geq 5)}|} \sum_{s \in \mathcal{S}_p^{(n \geq 5)}} \text{JSD}(\pi_{\text{player}}(\cdot \mid s) \parallel \pi_{\text{model}}(\cdot \mid s)) \quad (10)$$

This state-averaged score serves as an empirical divergence index summarizing stylistic divergence across recurring positions. Whenever a board state s appears across multiple cross-validation partitions, predicted distributions are averaged out-of-fold prior to computing divergence against the empirical human move distribution.

5.2 Performance Summary and Algorithmic Comparisons

Table 3 summarizes the experimental results across all eight candidate configurations evaluated over the cohort of 14 historical elite players.

In the unconditioned setting, the raw Maia-2 Baseline achieves a top-1 move accuracy of 43.61%, a tactical gap Δ_{CPL} of 10.49, and a mean JSD of 0.3424 bit. Because the baseline lacks any subject-specific routing, it outputs an identical, population-averaged prior across all players. Applying Nucleus Pruning directly to the policy head leaves top-1 move accuracy unchanged (43.61%) while marginally reducing divergence to 0.3337 bit. This slight improvement largely stems from the finite nature of empirical distributions. Because ground-truth human frequencies are derived from limited sample counts, unobserved legal moves carry zero empirical mass, whereas the unconstrained model distributes residual probability

Table 3: Performance summary across model variants evaluated on recurring board states ($n \geq 5$). The per-player book predicts each position’s most frequent training-fold move. Maia-2 FT fine-tunes only player input embeddings. The reported $\pm$ intervals represent the population standard deviation across the cohort of $N = 14$ elite players, capturing the performance dispersion across distinct elite player profiles.

Model Variant	Move Acc (%)	Δ_{CPL}	JSD ($n \geq 5$)
<i>Reference</i>			
Per-player book	73.27% $\pm$ 5.35%	1.94 $\pm$ 0.94	0.0816 $\pm$ 0.0125
<i>Direct Policy Variants</i>			
Maia-2 Baseline	43.61% $\pm$ 2.84%	10.49 $\pm$ 1.38	0.3424 $\pm$ 0.0153
Maia-2 FT	46.10% $\pm$ 2.88%	2.32 $\pm$ 1.69	0.3663 $\pm$ 0.0179
Maia-2 Nucleus	43.61% $\pm$ 2.84%	10.49 $\pm$ 1.38	0.3337 $\pm$ 0.0154
Maia-2 MoE-LoRA	61.95% $\pm$ 4.56%	2.03 $\pm$ 1.58	0.2640 $\pm$ 0.0176
<i>MCTS Search Variants</i>			
Maia-2 MCTS	44.63% $\pm$ 3.46%	9.85 $\pm$ 1.47	0.3209 $\pm$ 0.0149
Maia-2 N. + MCTS	44.65% $\pm$ 3.43%	9.83 $\pm$ 1.46	0.3210 $\pm$ 0.0150
Maia-2 MoE-LoRA N. + MCTS	62.20% $\pm$ 5.33%	2.64 $\pm$ 1.71	0.2350 $\pm$ 0.0170

across the entire legal action space, artificially inflating divergence. Truncating low-probability moves via nucleus filtering mechanically alleviates this support discrepancy.

Integrating the parameter-efficient MoE-LoRA adapter induces a substantial performance leap across all three dimensions. Top-1 move prediction accuracy increases by +18.34 percentage points to 61.95% $\pm$ 4.56%, demonstrating that direct logit-level intervention successfully overcomes the rigidity of the underlying visual backbone. Concurrently, Δ_{CPL} drops by more than a factor of five, falling from 10.49 to 2.03 $\pm$ 1.58. Similarly, the average JSD decreases from 0.3424 to 0.2640 bit, representing a reduction of more than 22% in unshared informational divergence. These gains confirm that routing low-rank residual adapters conditionally through player identity vectors enables the model to accurately capture individual opening preferences and tactical predispositions.

Two additional references put these gains in perspective. First, fine-tuning only the player input embeddings while keeping the backbone frozen reaches 46.10% $\pm$ 2.88% accuracy with a mean JSD of 0.3663 bit and a gap of 2.32. This improves accuracy and tactical proximity over the unconditioned baseline but does not reduce its divergence, and remains far below MoE-LoRA on all three metrics, supporting logit-level adaptation. Second, a per-player opening book predicting each position’s most frequent training-fold move achieves 73.27% $\pm$ 5.35% top-1 accuracy with a mean JSD of 0.0816 bit on the same recurring set, outperforming all neural variants. This is expected. On opening-dominated recurring positions the empirical distributions are highly concentrated, so memorizing the repertoire

beats modeling style. The neural models’ value lies instead in generalizing to positions the book has never seen, where frequency lookup has no prediction to offer.

Applying shallow MCTS ($N = 50$) to the baseline yields minor gains, increasing move accuracy slightly from 43.61% to 44.63% and reducing JSD to 0.3209 bit. However, combining MCTS with MoE-LoRA and Nucleus Pruning achieves the lowest informational divergence across all tested neural architectures, reaching an optimal JSD of 0.2350 ± 0.0170 bit, while maintaining high move accuracy (62.20%) and a tight tactical gap (2.64).

5.3 Trade-Off Between Direct Policy and Search Regularization

Comparing direct MoE-LoRA with its search-regularized counterpart shows that search improves both deterministic accuracy and distributional alignment. MoE-LoRA N. + MCTS achieves the highest top-1 move accuracy at 62.20% and the lowest average JSD at 0.2350 bit among neural variants, at the cost of a slightly larger tactical gap.

Converting tree visit counts into an action policy introduces slight variations compared to raw network logits, resulting in a marginal 0.25% variation in top-1 selection. The two configurations retain comparable gaps Δ_{CPL} of 2.03 and 2.64, indicating that tactical fidelity relative to human master play remains stable whether moves are sampled directly or guided by limited search.

5.4 Tactical Calibration and Centipawn Gap Dynamics

The tactical gap Δ_{CPL} provides valuable insights into the qualitative nature of the synthesized moves. Superhuman engines optimized for zero centipawn loss diverge systematically from human intuition, selecting cold, defensive maneuvers that human players avoid. Conversely, unconstrained human-like policy networks occasionally succumb to acute tactical oversights.

The empirical baseline exhibits a gap Δ_{CPL} of 10.49, reflecting a moderate divergence between the average tactical quality of human choices and model predictions. Under MoE-LoRA conditioning, this gap collapses to 2.03 in direct policy mode and 2.64 under search. This convergence shows that the adapter aligns its average evaluation level with the tactical competence of elite players rather than drifting toward uncharacteristic amateur blunders or superhuman engine calculations. It does not by itself demonstrate reproduction of player-specific blunder patterns.

5.5 Methodological Scope and Full-Career Generalization

The restriction of evaluation to recurring positions with $n \geq 5$ avoids single-observation targets and keeps the empirical distributions multi-observation, in the limited sense defined in Section 3. However, this filtering imposes a fundamental structural constraint that must be highlighted.

As established in Table 2, recurring positions encountered at least five times represent merely 0.77% of unique board configurations in an elite player’s career on average, although they account for 13.55% of total move instances. These recurring states concentrate predominantly in early opening lines, established theoretical variations, and common early middlegame pawn structures. Consequently, while the experimental results show that MoE-LoRA captures player-specific opening repertoires and recurring strategic motifs better than any other neural variant, they do not yet prove that the model can faithfully reproduce an elite player’s individualized decision-making throughout an entire, unique game.

In novel middlegame complications and deep endgames, unique positions ($n = 1$) predominate. In such singleton states, empirical distributions collapse into single-action targets, so distribution-aware metrics such as JSD remain formally defined but their human side is estimated from a single observation and cannot be trusted. Behavioral evaluation on rare states is therefore left to deterministic accuracy and to future metrics tailored to empirical sparsity, such as sequence-level game trajectory modeling or overlap measures like the Bhattacharyya and Hellinger distances. Testing the learned patterns on rare states is needed to check whether they transfer beyond the recurring set, and is left for future work.

Regarding overfitting, the adapter adds roughly 2×10^5 trainable parameters to the frozen backbone, comprising 8 rank-8 experts, a router, and 14 routing embeddings. Training pools each fold draw on the full career histories, about 1.4M positions per fold, and only evaluation is restricted to the recurring subset. Generalization is assessed by strictly out-of-fold evaluation under game-level partitioning, with label smoothing of 0.05 and weight decay of 10^{-2} as regularizers. No early stopping or train/validation gap analysis was performed, so residual overfitting to frequent opening lines cannot be excluded, which is consistent with the strong opening-book baseline discussed below.

6 Conclusion and Future Work

This paper presented player-specific behavioral modeling for elite chess players. To overcome the limitations of input embedding fine-tuning on frozen backbones, the parameter-efficient MoE-LoRA adapter was introduced, increasing move prediction accuracy on recurring positions from 43.61% to 61.95% while cutting the tactical gap Δ_{CPL} from 10.49 to 2.03. Coupling MoE-LoRA with budget-constrained MCTS regularized by Nucleus Pruning achieved the closest distributional alignment, reaching a minimum average JSD of 0.2350 bit. Furthermore, an empirical support threshold ($n \geq 5$) was formalized, avoiding single-observation targets and enabling distributional assessments on recurring positions.

A primary methodological limitation of this study lies in its restriction to recurring positions ($n \geq 5$). While filtering out singleton configurations is necessary to obtain multi-observation empirical targets, recurring board states represent only a fraction of an elite player’s lifetime moves, namely 0.77% of unique

states and 13.55% of instances, concentrated predominantly in early opening branches and common middlegame structures. Recurrent sets are identified on full career histories. Only the game-level train/test partition itself is strictly out-of-fold. A per-player opening book built from training-fold frequencies outperforms all neural variants on this recurring set with 73.27% top-1 accuracy, confirming that much of the predictability there is repertoire memorization. The adapted policy’s advantage lies in generalizing beyond seen positions. Consequently, while the adapted policy captures stylistic choices on frequent choice points, these results do not yet establish that the model faithfully reproduces an individual elite player’s decision process across an entire game, particularly throughout unique middlegame complications and complex endgames.

Extending behavioral evaluation to entire careers without discarding singleton positions is essential future work. Evaluating low-support states will require metrics tailored to empirical sparsity, such as sequence-level game trajectory modeling or overlap measures like the Bhattacharyya and Hellinger distances, which quantify the probability that two stochastic policies produce identical move sequences without suffering from divergence explosions on single-visit observations.

In parallel, extending this framework to low-resource regimes can be pursued by integrating the prototype matching mechanisms established in Maia4All [24]. Initializing expert routing parameters through weighted projections over informative prototype embeddings would enable adaptation to modern elite players, historical competitors with sparse archives, or online practitioners with restricted game histories.

The present methodology aggregated player records across entire careers into single static profiles. Because competitive chess players undergo marked strategic shifts throughout their careers, segmenting historical corpora into chronological career phases, corresponding to early formative periods, peak championship reigns, and late career stages, will allow the training of temporally conditioned models to track longitudinal stylistic evolution.

Evaluating the empirical playing strength of the adapted agents through round-robin tournament simulations and calibrated Elo benchmarking will verify whether the generated policies reproduce historical competitive dynamics, connecting with human-like general game playing frameworks such as Cogni-Play [17]. A complementary cross-player check, attributing held-out positions to the correct player among the cohort, would test whether the adapted policies capture individual style beyond opening repertoire memorization. In parallel, conducting blinded perceptual studies with titled human players against various rating levels will evaluate whether human practitioners subjectively identify player-specific characteristics in interactive play, drawing upon human-agent interaction evaluation paradigms [18].

All source code is openly available at <https://github.com/SogliuzzoL/ChessBehaviors>.

Acknowledgments. This research was conducted within the ICTEAM Institute at UCLouvain. This work is affiliated with the European COST Action CA22145 GameTable [15].

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Campbell, M., Hoane Jr, A.J., Hsu, F.h.: Deep blue. *Artificial intelligence* **134**(1-2), 57–83 (2002)
2. Carlson, J.: Matilda: Engine-agnostic search with human policy guidance (2026)
3. Coulom, R.: Efficient selectivity and backup operators in monte-carlo tree search. In: *International conference on computers and games*. pp. 72–83. Springer (2006)
4. Eisma, Y.B., Koerts, R., de Winter, J.: Turing tests in chess: An experiment revealing the role of human subjectivity. *Computers in Human Behavior Reports* **16**, 100496 (2024)
5. Frisoni, G., Molfetta, L., Freddi, D., Moro, G.: Mixture of masters: Sparse chess language models with player routing. *arXiv preprint arXiv:2602.04447* (2026)
6. Holtzman, A., Buys, J., Du, L., Forbes, M., Choi, Y.: The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751* (2019)
7. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021)
8. Jacob, A.P., Wu, D.J., Farina, G., Lerer, A., Hu, H., Bakhtin, A., Andreas, J., Brown, N.: Modeling strong and human-like gameplay with kl-regularized search. In: *International Conference on Machine Learning*. pp. 9695–9728. PMLR (2022)
9. Lin, J.: Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory* **37**(1), 145–151 (1991)
10. McIlroy-Young, R., Sen, S., Kleinberg, J., Anderson, A.: Aligning superhuman ai with human behavior: Chess as a model system. In: *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. pp. 1677–1687 (2020)
11. McIlroy-Young, R., Wang, R., Sen, S., Kleinberg, J., Anderson, A.: Learning models of individual behavior in chess. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. pp. 1253–1263 (2022)
12. McIlroy-Young, R., Wang, Y., Sen, S., Kleinberg, J., Anderson, A.: Detecting individual decision-making style: Exploring behavioral stylometry in chess. *Advances in Neural Information Processing Systems* **34**, 24482–24497 (2021)
13. Monroe, D., Eilender, G., Chalmers, P., Tang, Z., Anderson, A.: Chessformer: A unified architecture for chess modeling. In: *The Fourteenth International Conference on Learning Representations* (2026)
14. Omi, N., Caccia, L., Sarkar, A., Ash, J.T., Sen, S.: Generative modeling of individual behavior at scale. *arXiv preprint arXiv:2502.14998* (2025)
15. Piette, É., Crist, W., Soemers, D.J., Rougetet, L., Courts, S., Penn, T., Morenville, A.: Gametable cost action: Kickoff report. *ICGA Journal* **46**(1), 11–27 (2024)
16. Rautureau, A., Cazenave, T., Piette, É.: Generalized rapid action value estimation in memory-constrained environments. *arXiv preprint arXiv:2602.23318* (2026)

17. Rautureau, A., Piette, É.: Cogniplay: a work-in-progress human-like model for general game playing. arXiv preprint arXiv:2507.05868 (2025)
18. Rosemarin, H., Rosenfeld, A.: Playing chess at a human desired level and style. In: Proceedings of the 7th international conference on human-agent interaction. pp. 76–80 (2019)
19. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538 (2017)
20. Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al.: Mastering chess and shogi by self-play with a general reinforcement learning algorithm. arXiv preprint arXiv:1712.01815 (2017)
21. Soemers, D.J., Piette, E., Stephenson, M., Browne, C.: Manipulating the distributions of experience used for self-play learning in expert iteration. In: 2020 IEEE Conference on Games (CoG). pp. 245–252. IEEE (2020)
22. Sogliuzzo, L., Rautureau, A., Piette, E.: Toward modeling player-specific chess behaviors. arXiv preprint arXiv:2605.11893 (2026)
23. Tang, Z., Jiao, D., McIlroy-Young, R., Kleinberg, J., Sen, S., Anderson, A.: Maia-2: A unified model for human-ai alignment in chess. *Advances in Neural Information Processing Systems* **37**, 20919–20944 (2024)
24. Tang, Z., Jiao, D., Xue, E., McIlroy-Young, R., Kleinberg, J., Sen, S., Anderson, A.: Learning to imitate with less: Efficient individual behavior modeling in chess. *Transactions on Machine Learning Research* (2026)
25. Zhang, Y., Jacob, A.P., Lai, V., Fried, D., Ippolito, D.: Human-aligned chess with a bit of search. In: The Thirteenth International Conference on Learning Representations (2025)