

LubZero: An active molecular generation framework for multi-objective properties of lubricant molecules

Luyao Bao^{1,2}, Rui Zhou², Xiangmeng Jia³✉, Feng Zhou¹✉, Meirong Cai¹, Weimin Liu¹

¹ State Key Laboratory of Solid Lubrication, Lanzhou Institute of Chemical Physics, Chinese Academy of Sciences, Lanzhou, 730000, China.

² Shandong Laboratory of Advanced Materials and Green Manufacturing at YanTai, YanTai, 264006, China.

Corresponding author. atomrq@gmail.com, zhouf@licp.cas.cn

Received: February 27, 2026; Revised: July 27, 2026; Accepted: July 30, 2026

© The Author(s) 2026.

Abstract: The screening of lubricant candidates with tailored multi-property profiles is a foundational challenge in tribology, driven by increasing demands for energy efficiency and operational safety. However, the vast discrete chemical space and the sparse-reward nature of molecular design often hinder efficient optimization. In this work, we present LubZero, a graph-based molecular generation framework that adapts the AlphaZero algorithm for the de novo design of lubricant ester materials. LubZero reformulates fragment-based molecular assembly as a single-player combinatorial game, integrating a Graph Isomorphism Network with Edge features (GINE) and Monte Carlo Tree Search (MCTS) to enable principled forward planning. Crucially, we implement a high-throughput parallel self-play architecture that utilizes multi-process inference on a shared GPU, significantly enhancing exploration efficiency and hardware resource utilization compared to serial implementations. This framework enables LubZero to navigate complex property trade-offs efficiently within a constrained molecular-design protocol. We applied LubZero to the multi-objective optimization of ester lubricants, specifically targeting high flash points and low pour points. The results show that LubZero identifies high-ranking surrogate-predicted screening candidates that extend the predicted flash-point/pour-point trade-off space beyond the region covered by the available experimental data.

By integrating artificial intelligence planning with a scalable parallel computing strategy, LubZero offers a surrogate-guided workflow for screening and prioritizing candidate lubricant molecules under explicitly defined objective and applicability-domain constraints.

Keywords: Lubricants, Molecular design, Reinforcement learning

1 Introduction

Lubricating oils constitute a foundational component of modern mechanical systems, functioning as dynamic interfacial media that reduce friction, suppress wear, dissipate heat, and prevent corrosion through the formation of protective tribofilms.[1,2] Their performance critically determines energy efficiency, equipment durability, and system reliability across automotive, aerospace, marine, manufacturing, and renewable energy sectors.[3-5]

Over the past decades, lubricant technology has transitioned from conventional mineral oils toward synthetic base stocks such as polyalphaolefins (PAOs), esters, and polyalkylene glycols (PAGs), combined with increasingly sophisticated additive systems including antioxidants, dispersants, friction modifiers, and anti-wear agents.[6-8] Modern lubricant development must simultaneously optimize viscosity index, oxidative and thermal stability, low-temperature fluidity, volatility, tribological robustness, and environmental compatibility under increasingly stringent regulatory frameworks. The growing compositional complexity of base-oil and additive systems has exposed the limitations of empirical trial-and-error strategies, thereby motivating the integration of predictive modeling and data-driven methodologies into lubricant science.

Recent reviews and perspectives have shown that artificial intelligence (AI) and machine learning (ML) are becoming important tools for tribology and friction-system research.[9-14] These studies describe a broad transition toward tribo-informatics, where experimental, simulation, image, acoustic, vibration, and sensor-derived data are used to model friction, wear, lubrication regimes, elastohydrodynamic film thickness, surface damage, and component condition. For example, Marian et al. used ML models to predict elastohydrodynamic lubrication film-thickness parameters, demonstrating the utility of data-driven models for contact-scale lubrication analysis.[15] Shah et al. further applied ensemble deep learning to wear-particle image analysis, showing how AI can support condition monitoring and wear-state identification in lubrication and tribological systems.[16] Such data-driven approaches are particularly valuable for tribological systems because their behavior is often multiscale, history-dependent, and controlled by nonlinear interactions among surface chemistry, contact mechanics, lubricant composition, temperature, load, and sliding conditions.

In lubricant research specifically, ML has been applied to oil condition monitoring, predictive maintenance, lubricant-material databases, and the prediction of key physicochemical and performance-related properties such as viscosity, viscosity index, pour point, flash point, oxidation stability, and tribological response.[17-20] For lubricant-specific applications, Malaguti et al. developed a supervised ML model for determining lubricant oil operating conditions,[17] whereas Zhou et al. proposed the Lubrication Brain framework to connect lubricant molecular structures with property prediction and molecular design.[21] These advances indicate that AI/ML can reduce empirical screening cost and accelerate the identification of promising lubricant candidates, while also highlighting the need for chemically meaningful representations, reliable surrogate models, and explicit applicability domain assessment.

Despite these advances, lubricant-oriented molecular generation remains challenging. Lubricant performance arises from nonlinear interactions among base oils and multiple additive classes, available experimental datasets are often sparse and noisy, and candidate molecules must satisfy coupled constraints related to thermal safety, low-temperature fluidity, molecular size, functionality, and practical feasibility. Algorithmically, diffusion models require extensive sampling steps, GFlowNets are sensitive to reward shaping, and conventional reinforcement-learning methods suffer from sparse rewards and long-horizon credit assignment issues in discrete chemical spaces.[22-31] The AlphaZero framework integrates neural-network policy/value estimation with Monte Carlo Tree Search (MCTS), enabling forward planning and long-horizon reasoning in high-dimensional combinatorial spaces.[32,33] Building upon this paradigm, we introduce LubZero, a graph-based AlphaZero framework for lubricant-oriented molecular generation. LubZero reformulates de novo lubricant molecule design as a sequential fragment assembly process over chemically meaningful building blocks. A Graph Isomorphism Network with Edge features (GINE)[34] is employed for policy and value estimation; multi-objective optimization[35] is implemented via linear scalarization; and RDKit-based validation,[36] task-specific structural filtering, and MCTS exploration based on the PUCT rule[38] are integrated to balance chemical validity, search efficiency, and multi-property optimization.

2 Methods

2.1 Feature Engineering and Surrogate Model Training

The experimental dataset for the synthetic ester lubricants used in this study was sourced from the work of Zhou et al.[21] Molecular descriptors and Morgan fingerprints (radius = 2, bit length = 2048) were computed using the RDKit library.[36] To interpret the model's decision-making process, feature importance analysis was conducted via the SHAP (SHapley Additive exPlanations)

method.[39] Model training and hyperparameter optimization were implemented using the Scikit-learn framework.[40] For the parity plots in Fig. 2, the 75-molecule dataset was partitioned into training, validation, and test subsets according to an 8:1:1 ratio; the validation and test predictions are shown as non-training points to avoid overinterpreting a very small isolated test subset. As an additional split-independent robustness check, five-fold cross-validation was performed for the principal surrogate-model configurations considered in this study, and the fold-averaged mean $\pm$ standard deviation values are summarized in SI Table 2. Finally, the chemical space distribution of the molecular structures was visualized in 2D using t-distributed Stochastic Neighbor Embedding (t-SNE).[41] In SI Table 2, for flash-point prediction, the XGBoost model with RFE descriptors achieved $R^2 = 0.867 \pm 0.037$, MAE = 9.19 ± 1.89 K, and RMSE = 12.16 ± 2.81 K, whereas the KRR model with PCA fingerprints achieved $R^2 = 0.630 \pm 0.280$, MAE = 14.22 ± 5.08 K, and RMSE = 19.77 ± 10.58 K. For pour-point prediction, the XGBoost model with RFE descriptors achieved $R^2 = 0.794 \pm 0.054$, MAE = 18.18 ± 2.33 K, and RMSE = 22.34 ± 3.51 K, whereas the KRR model with PCA fingerprints achieved $R^2 = 0.729 \pm 0.216$, MAE = 19.38 ± 5.11 K, and RMSE = 23.13 ± 6.08 K.

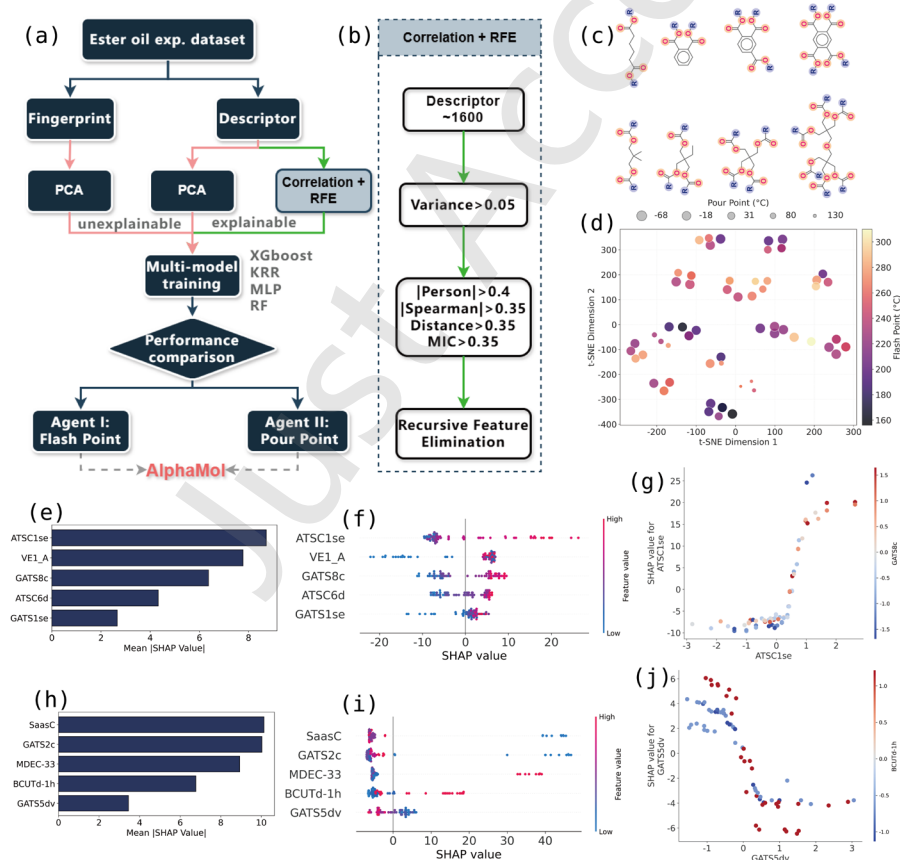


Fig. 1. Framework of feature engineering and surrogate model selection. (a) Schematic workflow of the technical approach; (b) Flowchart of the interpretable feature engineering process; (c) Representative chemical structures of esters in the experimental dataset; (d) Chemical space

distribution of molecular structures, where the color scale and point size represent the flash point and pour point, respectively; (e–g) Feature importance analysis for flash point and (h–j) pour point; (e, h) Ranking of feature importance; (f, i) SHAP value beeswarm distributions for the screened descriptors (color indicates the descriptor feature value, from low to high); (g, j) Individual contributions (SHAP values) of key descriptors to the flash point and pour point.

Automated feature engineering was employed to train surrogate models capable of rapidly predicting the flash and pour points of ester lubricants, thereby establishing a critical evaluation foundation for the subsequent LubZero molecular generation model. By systematically benchmarking various combinations of features and algorithms, we constructed accurate surrogate models to support the efficient assessment of molecular designs within this framework.

The overall technical strategy, illustrated in Fig. 1(a), encompasses three feature dimensionality reduction schemes combined with four classical machine learning models. This systematic comparison was used to compare candidate surrogate models under the fixed train/validation/test workflow while also providing physical interpretations of structure-property relationships. Specifically, three distinct pathways were explored: Path a relies on molecular descriptors and is subdivided into two reduction methods: (a1) Principal Component Analysis (PCA), which enhances training efficiency but offers limited interpretability; and (a2) Correlation Coefficient Screening combined with Recursive Feature Elimination (RFE), which compresses dimensions while preserving interpretability, albeit with a potential minor trade-off in predictive accuracy. Path b utilizes Morgan molecular fingerprints directly as feature inputs, followed by PCA-based dimensionality reduction.

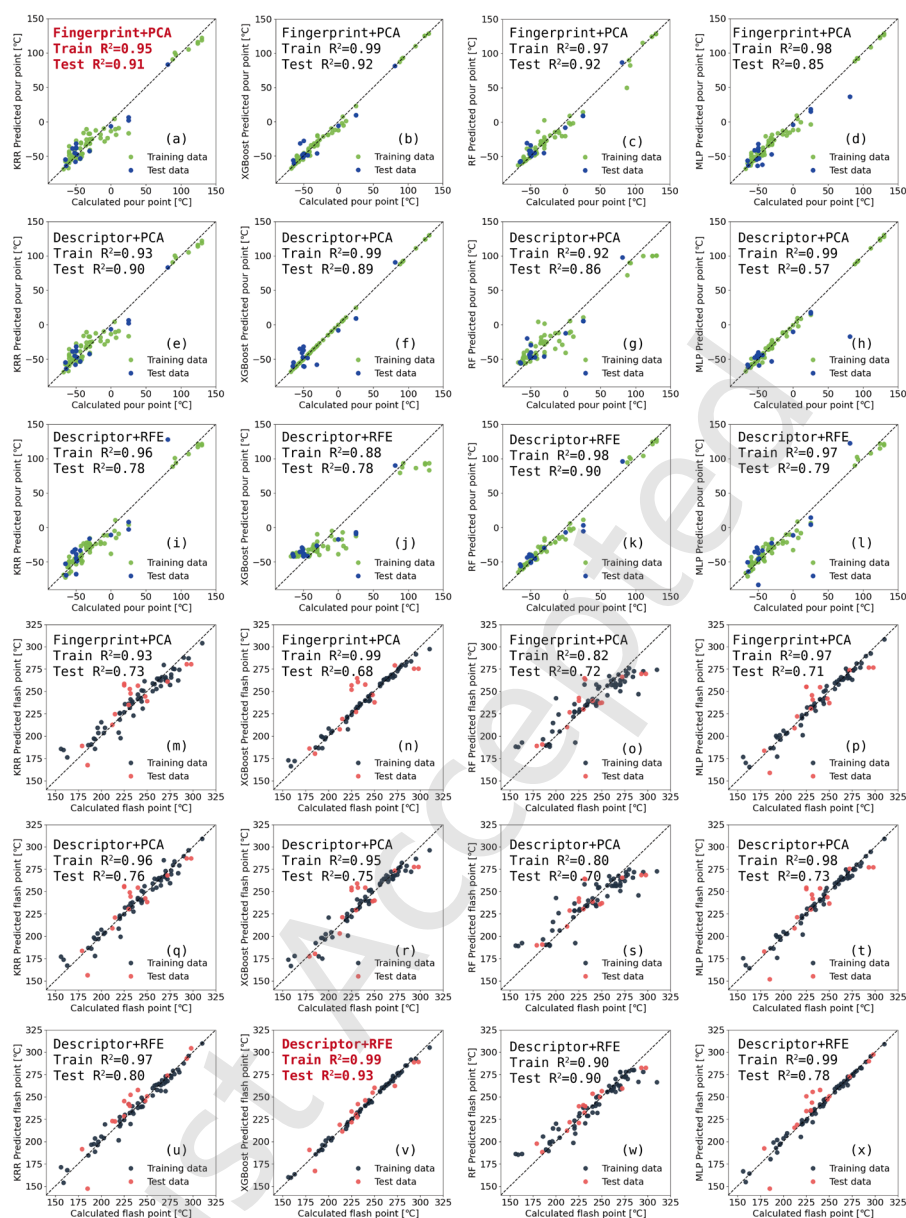
Ultimately, these three feature sets were evaluated using four representative machine learning algorithms—Random Forest (RF), Multi-layer Perceptron (MLP), Kernel Ridge Regression (KRR), and XGBoost. The surrogate configuration used for the reported LubZero generation run was selected independently for flash and pour points from this fixed-split model-selection workflow. Notably, the hierarchical dimensionality reduction strategy adopted in Path a2 (Fig. 1(b)) successfully enhanced model interpretability while retaining critical physicochemical descriptors.

The experimental dataset comprises the chemical structures and corresponding flash and pour points of 75 ester lubricant molecules; representative structures are shown in Fig. 1(c). Fig. 1(d) further illustrates the structural diversity of these esters, providing initial coverage of the ester chemical space for surrogate modeling. A distinct trade-off is observed between the pour point and flash point, highlighting the scarcity of ester molecules that simultaneously satisfy the industrial requirements of a low pour point and a high flash point.

Feature importance analysis for the flash point (Fig. 1(e-g); see SI Fig. 1) reveals that descriptors ATSC1se, VE1_A, GATS8c, ATSC6d, and GATS1se are the most significant governing factors (SI Table 1). These descriptors collectively reflect the dominant roles of molecular backbone scale, structural compactness, and polarity distribution in modulating volatilization and ignition behavior. As shown in Fig. 1(f), most descriptors exhibit a positive correlation with the flash point (red indicates higher contribution). Both ATSC1se and GATS1se are associated with short-range atomic interactions weighted by electronegativity, suggesting that the distribution pattern of polar functional groups (e.g., ester groups) within the carbon skeleton can inhibit molecular escape into the gas phase by enhancing intermolecular forces, thereby increasing the flash point (Fig. 1(g)). VE1_A describes overall molecular connectivity and density; higher values typically correspond to structures with large volumes, compact configurations, and low vapor pressures—characteristic of high-flash-point esters. GATS8c and ATSC6d reflect the synergistic effects of charge distribution and σ -electron structure at medium-to-long ranges, indicating that the flash point depends not only on local functional group properties but also on carbon chain length, branching degree, and overall skeletal stability. In summary, synthetic ester molecules with high flash points tend to possess large skeletal dimensions, appropriately distributed polar groups, and stable electronic structures.

Feature importance analysis for the pour point (Fig. 1(h-j); see SI Fig. 2) identifies SaasC, GATS2c, MDEC-33, BCUTd-1h, and GATS5dv as key determinants (SI Table 1). These descriptors highlight how the carbon skeleton scale, branching patterns, and electronic distribution regulate low-temperature flow behavior. SaasC is closely related to overall skeletal complexity and generally increases with carbon chain length, branching, or the inclusion of cyclic structures. Such features enhance intermolecular Van der Waals interactions and restrict segmental motion, leading to an elevated pour point. MDEC-33 describes the spatial distribution of tertiary carbon atoms, indicating that the arrangement of branching centers significantly impacts molecular packing efficiency and conformational freezing at low temperatures. GATS2c and GATS5dv (SI Fig. 2(e,h)) reflect correlations in charge and valence electron distribution at short and medium-to-long ranges, respectively, suggesting that non-uniform polarity distribution enhances intermolecular attraction, causing the system to enter a kinetically frozen state earlier. BCUTd-1h characterizes molecular rigidity via σ -electron distribution and skeletal connectivity; a more stable skeleton with hindered low-temperature conformational rearrangement corresponds to a higher pour point. Overall, the pour point is co-regulated by skeletal structural features and electronic distribution, essentially reflecting differences in molecular conformational mobility under low-temperature conditions.

Fig. 2. Performance comparison of surrogate models trained using three distinct feature



dimensionality reduction strategies. Evaluation of (a–l) pour point and (m–x) flash point predictions. The feature engineering strategies include: (a–d, m–p) molecular fingerprints with PCA; (e–h, q–t) molecular descriptors with PCA; and (i–l, u–x) molecular descriptors with Recursive Feature Elimination (RFE). The specific algorithms employed are: (a, e, i, m, q, u) Kernel Ridge Regression (KRR); (b, f, j, n, r, v) XGBoost; (c, g, k, o, s, w) Random Forest (RF); and (d, h, l, p, t, x) Multi-layer Perceptron (MLP). In these panels, validation and held-out test samples are grouped in the plotted 'Test data' category as non-training points.

The comparative model performance is presented in Fig. 2. Based on the fixed 8:1:1 split used for the reported generation study, the pour-point reward surrogate was a Kernel Ridge Regression (KRR)

model trained on PCA-reduced molecular fingerprints, and the flash-point reward surrogate was an XGBoost model trained on RFE-reduced molecular descriptors. The five-fold CV results in SI Table 2 are reported as a separate split-robustness analysis; because they show stronger average pour-point performance for XGBoost/RFE than for KRR/PCA, the KRR/PCA pour-point model should be interpreted as the fixed-split surrogate used in the reported generation experiments rather than as a generally superior pour-point predictor.

2.2 The LubZero Framework

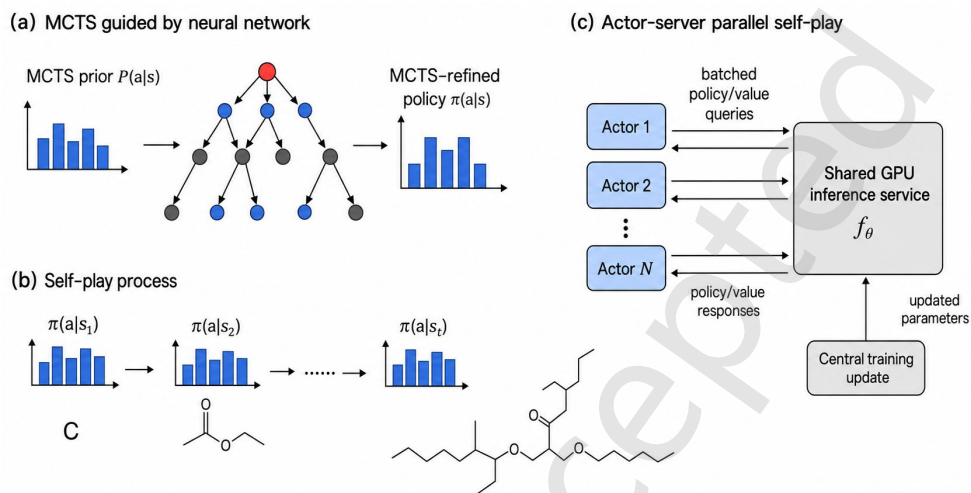


Fig. 3. Schematic illustration of the LubZero algorithm. (a) Diagram of the MCTS search process, where the red circle represents the root node, blue circles denote the first-step search results, and the final output is the MCTS-refined policy distribution $\pi(a|s)$, which guides subsequent action selection. (b) Illustration of a complete self-play trajectory starting from a carbon atom to construct a molecule. (c) Depiction of the actor-server parallel self-play architecture, where multiple CPU actors send policy/value queries to a shared GPU inference service for batched neural-network evaluation. The details are described in the text.

LubZero adapts the AlphaZero framework[32,33] to the domain of de novo molecular design by casting the fragment-based assembly process as a single-player combinatorial game. Each state s is represented as an undirected molecular graph $G = (V, E)$, where V and E denote the sets of heavy-atom nodes and chemical-bond edges, respectively. Node features are encoded as 20-dimensional vectors capturing essential atomic properties, including atom type (one-hot encoding for C, H, O, N), atomic degree (0-5), formal charge (-1, 0, +1), aromaticity, implicit hydrogen count (0-3), and ring membership. Edge features are 2-dimensional vectors encoding bond orders (single, double, triple, or aromatic) and the absolute electronegativity difference between bonded atoms. To ensure consistent tensor shapes for single-atom states, a self-loop with zero-valued edge features is applied.

Molecular generation initiates from a fixed methane (CH_4) seed and proceeds through a sequence of discrete actions. In each step, the agent selects a valence-available attachment site and appends one of the predefined building blocks, such as a single carbon atom, an ester group, a benzene ring, or an ether linkage. Each building block is defined with explicit connection vectors. Following each operation, the molecular structure is validated via the RDKit[36] sanitization protocol; any violation of chemical valency or stability triggers immediate trajectory termination with a failure terminal reward ($z_{\text{fail}} = -1$).

The agent is guided by a GINE-based policy-value network f_θ . The architecture comprises five GINE convolutional layers, each integrated with batch normalization and ReLU activation functions. Global mean pooling is employed to extract a graph-level latent embedding. The network terminates in two specialized heads: a value head, consisting of a two-layer multilayer perceptron that outputs a scalar $v_\theta(s)$ in $[-1, 1]$ via a tanh activation to estimate terminal rewards; and a policy head that produces raw logits for each node across the building blocks. After chemically invalid actions are masked, a softmax over the flattened action space yields $p_\theta(a|s)$, the network policy probability for each allowed action a . This distribution initializes the MCTS prior $P(a|s)$.

Action selection during both inference and self-play is directed by Monte Carlo Tree Search (MCTS) as standardized in the AlphaZero framework, which builds molecules through iterative planning and evaluation. As illustrated in Fig. 3(a), the MCTS search process begins at the root node (red circle), representing the current molecular state s . During each simulation, the tree expands by selecting actions via the PUCT rule, with blue circles denoting the first-step search results. At each expanded leaf, f_θ provides $p_\theta(a|s)$ and $v_\theta(s)$. MCTS uses $P(a|s)$ during selection and backpropagates $v_\theta(s)$ to update action values. The accumulated visit counts define the MCTS-refined target policy $\pi(a|s)$, which guides selection of the next building block.

Each search consists of multiple simulations comprising three phases. First, in the selection phase, the algorithm recursively chooses actions that maximize the PUCT selection rule[32,38]:

$$a^* = \arg \max_a \left[Q(s, a) + c_{\text{PUCT}} P(a | s) \frac{\sqrt{\sum_b N(s, b)}}{1 + N(s, a)} \right] \quad (1)$$

where a^* is the selected action, $Q(s, a)$ is the mean action value, $N(s, a)$ is the visit count, $c_{\text{PUCT}} = 1.0$ is the exploration coefficient, and the sum over b spans all allowed actions. Second, during expansion, the policy-value network evaluates each leaf node. Third, during backup, the resulting value estimate is propagated to update Q and N . At the root, Dirichlet exploration modifies the prior as $P'(a|s) = (1 - \varepsilon)P(a|s) + \varepsilon\eta_a$, where $\eta \sim \text{Dir}(\alpha)$, η_a is the component for action a , $\varepsilon = 0.25$, and $\alpha = 0.3$.

LubZero is trained via an iterative parallel self-play loop, where MCTS-guided actions form complete molecular trajectories. As shown in Fig. 3(b), a single self-play trajectory starts from a single carbon atom (derived from the CH₄ seed) and iteratively applies MCTS-selected fragment additions, such as appending ester groups or rings, until a terminal molecule is formed, collecting states s_t , refined policies π_t , and final rewards z for training.

To scale self-play efficiently, LubZero uses an actor-server parallelization strategy rather than simply increasing the number of CPU threads in a serial search. Multiple independent actor processes execute MCTS-guided molecular assembly in parallel, with each actor exploring different trajectories and random seeds under the same chemical constraints. During tree search, the policy/value queries from these actors are sent through request/response queues to a shared GPU inference service. This service batches queries from different actors according to the maximum batch size and waiting time, so the GPU evaluates many state-action queries efficiently while CPU-side actors continue legal-action enumeration, RDKit validity checks, fragment filtering, and MCTS bookkeeping. This design decouples CPU-side tree search from GPU-side neural-network inference, avoids repeatedly loading separate model copies for each actor, reduces GPU idle time, and converts many serial rollouts into concurrent self-play trajectories. Upon completion of the assigned molecular quota, candidates are filtered, deduplicated, reranked by the surrogate objectives, and the collected state-policy-reward triples are aggregated into a replay buffer. The central neural-network parameters θ are then optimized by minimizing the joint loss function

$$\mathcal{L}(\theta) = (z - v_{\theta}(s))^2 - \sum_a \pi(a | s) \log p_{\theta}(a | s) + c \|\theta\|_2^2 \quad (2)$$

where z is the terminal reward; θ denotes all trainable neural-network weights and biases; $v_{\theta}(s)$ is the θ -parameterized scalar value prediction for state s ; $\pi(a|s)$ is the scalar MCTS-refined target probability for allowed action a ; $p_{\theta}(a|s)$ is the corresponding θ -parameterized scalar network policy probability; and c is the L2 regularization coefficient. The sum runs over all allowed actions, and $\|\theta\|_2^2$ denotes the squared Euclidean (L2) norm of θ . The Adam optimizer is used with gradient clipping, and the updated central parameters are used by the shared inference service for the next epoch of parallel self-play. To address the inherent complexity of materials discovery, LubZero supports general multi-objective optimization through linear scalarization, where the terminal reward z is calculated as the weighted sum of normalized objective scores, with w_i and r_i denoting the weight and normalized score for objective i , respectively. To sample the Pareto frontier, parallel training runs are executed across a grid of weight vectors that sweep the dominant weight while preserving the self-improvement dynamics of the AlphaZero algorithm. In the lubricant-design case studied in

this work, SA score was not used as a direct optimization objective; the task-specific reward was restricted to the predicted flash point and pour point, while chemical feasibility was controlled through RDKit validity checks and lubricant-oriented structural constraints.

3 Results and Discussion

3.1 Toy Validation and Learning Dynamics

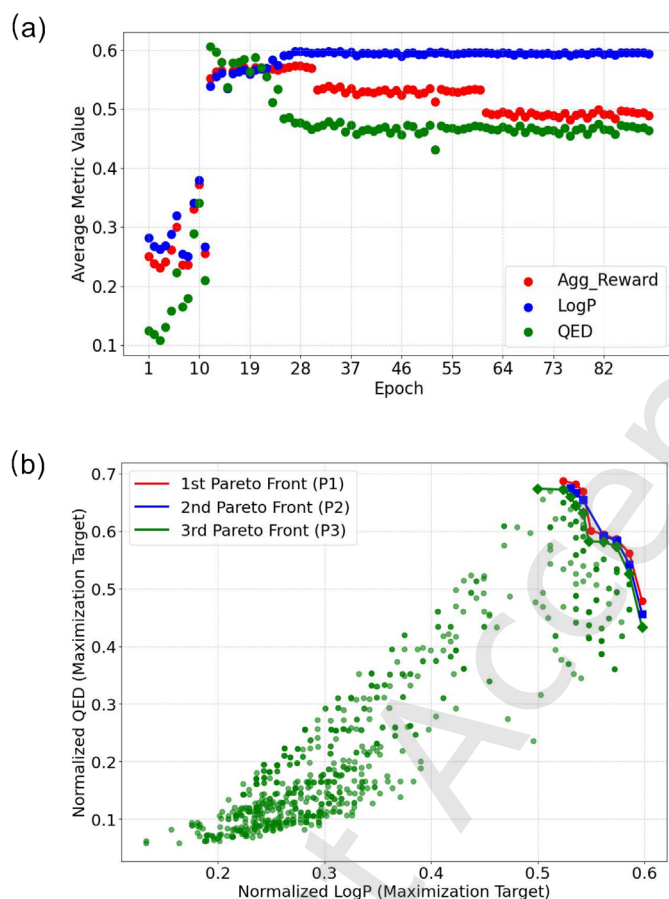


Fig. 4. Multi-objective optimization results for ester-based materials. (a) Average values of logP, QED, and the aggregated reward as a function of training epochs. (b) Sequential Pareto frontiers (P1, P2, P3) showing the distribution of generated molecules and their trajectory toward the optimal property trade-off region. P3, P2, and P1 denote earlier-, intermediate-, and later-stage Pareto frontiers, respectively; the shift toward the upper-right region indicates movement toward higher joint normalized logP and QED values.

To evaluate the multi-objective optimization performance and convergence behavior of LubZero, we conducted a toy validation experiment using a restricted action space of five basic building blocks. Molecular construction starts from a methane seed and proceeds for up to 12 sequential fragment additions. The task simultaneously maximizes logP and drug-likeness (QED) for ester-based compounds. Both properties were linearly normalized to [0, 1]: QED according to its native bounds,

and logP over [0, 15], with values below 0 set to 0 and values above 15 capped at 1.0. To obtain comprehensive coverage of the trade-off surface, independent self-play runs were performed across a grid of fixed scalarization weights ranging from 0.1 to 0.9 (step size 0.1). Each run followed a 10-epoch improvement cycle, with Monte Carlo Tree Search (MCTS) used at each decision step for forward planning. The aggregated results across weight settings were used for subsequent analysis.

As shown in Fig. 4(a), the average normalized logP, QED, and scalarized reward increase rapidly within the first ~30 training epochs, rising from ~0.2–0.3 to ~0.5–0.6, and then plateau. The reward stabilizes near 0.55, with logP around 0.58 and QED around 0.48, indicating efficient early discovery of high-performing assembly sequences and stable convergence of the policy–value network within the 12-step search horizon. The multi-objective trade-off surface is illustrated in Fig. 4(b) through progressive Pareto frontiers in the normalized logP–QED space. The initial generated molecules occupy the lower-left region, while LubZero progressively expands the results toward higher joint values. Ultimately, the Pareto frontiers (P1, P2, P3) reach the upper-right quadrant, with normalized logP up to ~0.6 and QED up to ~0.7. This outward shift demonstrates the model's ability to iteratively identify increasingly diverse non-dominated solutions and systematically explore distinct regions of the trade-off landscape. Specifically, P3 represents the early exploratory frontier, P2 represents the intermediate frontier, and P1 represents the later-stage frontier; this ordering clarifies that the apparent trajectory proceeds from the lower-left region toward the upper-right region with improved joint objective values.

3.2 Multi-Objective Optimization of Ester-Based Polymer Lubricants

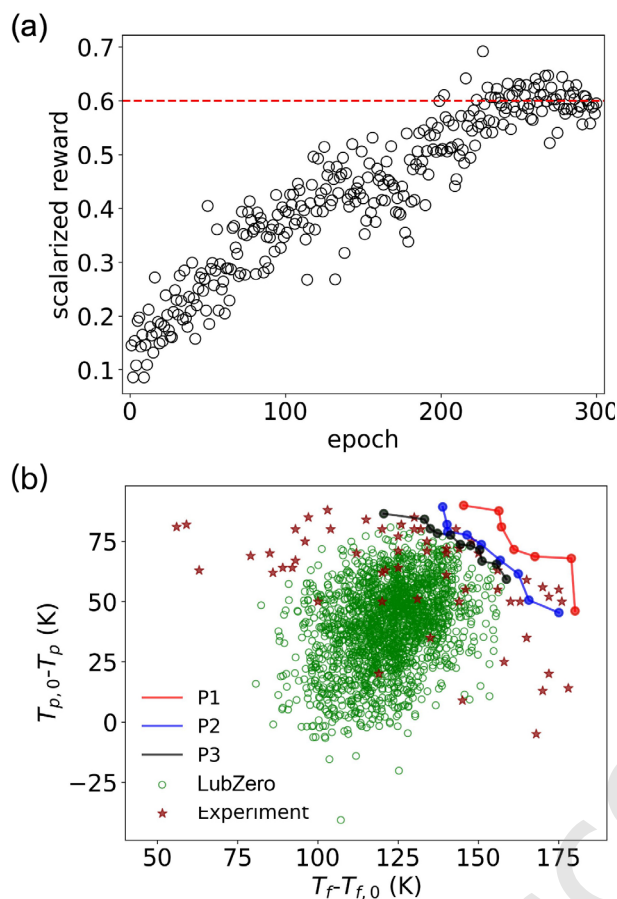


Fig. 5. Multi-objective optimization results for ester-based lubricant materials. (a) Average scalarized reward as a function of training epochs across independent self-play runs. The reward was computed from internally normalized flash-point and inverted pour-point improvements. (b) Progressive Pareto frontiers (P1 in red, P2 in blue, P3 in black) shown in shifted-temperature-improvement coordinates, where $\Delta T_f = T_f - T_{f0}$ and $\Delta T_p = T_{p0} - T_p$ are reported in K. Higher values therefore indicate higher flash point and lower pour point, respectively. Generated molecules (green circles) and experimental benchmark data points (red stars) are shown in the same shifted-temperature space.

To evaluate LubZero in a lubricant-oriented molecular design task, we applied the framework to the targeted optimization of ester-based lubricants, where high flash point (ensuring thermal safety and reduced fire risk) and low pour point (enabling reliable low-temperature performance) are critical requirements. Using four building blocks (single carbon atoms, benzene rings, ester groups, and ether linkages), molecular assembly started from a methane seed and was limited to a maximum of 15 sequential addition steps, with 500 MCTS simulations guiding each action selection. The dual objectives were to maximize flash point and minimize pour point. Property predictions were obtained from dedicated machine learning models trained on experimental lubricant datasets. To

systematically explore the trade-off between these competing objectives, we performed independent self-play runs across a grid of fixed scalarization weights. The dominant objective weight was varied from 0.1 to 0.9 in increments of 0.1, while preserving the reinforcement learning dynamics within each run. The aggregated results across weight settings were used to construct the final Pareto front and identify surrogate-predicted screening candidates spanning different regions of the trade-off landscape.

Flash point and pour point values were first converted to shifted-temperature improvements relative to the selected baselines: $\Delta T_f = T_f - T_{f0}$ for flash point and $\Delta T_p = T_{p0} - T_p$ for pour point, with $T_{f0} = 100$ K and $T_{p0} = 293$ K. Thus, larger values consistently indicate better performance in both dimensions. For reward calculation, these two shifted quantities were then linearly normalized to the range $[0, 1]$, and the terminal reward was computed as a weighted linear scalarization of the normalized objectives. The synthetic accessibility score metric[37] was therefore not directly optimized in the lubricant reward. Instead, the generated candidates were treated as property-optimized and structurally filtered screening leads, with feasibility controlled by RDKit sanitization, ester-like/lubricant-like composition requirements, and limits on excessive ring or aromatic-ring formation. Final synthetic feasibility should be confirmed by retrosynthetic analysis or experimental validation. Viscosity was not included in the reward function in this proof-of-concept lubricant case study. Therefore, the present results should be interpreted as optimization of thermal safety and low-temperature flow indicators rather than complete lubricant-performance optimization. Incorporating viscosity as an additional objective is technically supported by the scalarization framework and will be pursued when a sufficiently validated viscosity surrogate is available.

As presented in Fig. 5(a), the average scalarized reward, which represents the linear combination of normalized flash point and inverted pour point, exhibits distinct learning phases across independent training runs. In the initial stages (epochs 0-50), the reward starts from near-random baseline values (~ 0.2 - 0.3), indicative of exploratory policies that yield suboptimal molecules with moderate property balances. This phase is characterized by a sharp ascent, driven by LubZero's MCTS-guided look-ahead and self-play refinement, which rapidly identifies promising assembly pathways and boosts the reward to ~ 0.6 - 0.7 . During the intermediate epochs (50-200), the improvement continues at a moderated pace with minor fluctuations, reflecting the agent's progressive distillation of complex trade-offs between thermal stability (high flash point) and low-temperature fluidity (low pour point). Finally, in the later stages (beyond epoch 200), the reward stabilizes at a high plateau around 0.65 - 0.7 , with reduced variance, demonstrating robust convergence and the framework's ability to consistently generate high-quality trajectories despite the

sparse terminal rewards and competing objectives. This phased dynamics indicates efficient learning within the present constrained generation protocol through integrated planning and policy iteration.

The evolving trade-off landscape is shown in Fig. 5(b), where progressive Pareto frontiers were constructed via weight-sweeping. Because the plotted axes are shifted-temperature improvements, movement toward the upper-right region corresponds to simultaneously increasing flash point and decreasing pour point after baseline shifting. The final frontier P1 achieves the broadest coverage of this favorable region, whereas earlier frontiers P2 and P3 are more compact and suboptimal, reflecting intermediate and initial exploratory policies. The generated molecules (green circles) form a broad distribution that is systematically traced through Pareto analysis, illustrating LubZero's capacity for diverse exploration. Relative to the experimental data points (red stars), several LubZero-generated candidates from P1 are predicted to extend beyond the experimental cluster in shifted-temperature space, with ΔT_f reaching approximately 175 K and ΔT_p reaching approximately 75 K. These shifted values correspond to predicted absolute temperatures of approximately $T_f = 275$ K and $T_p = 218$ K under the stated baselines. This result suggests that LubZero can identify candidate structures that expand the surrogate-predicted trade-off space while remaining subject to the screening-level limitations of the surrogate models. This predicted shift may result from the strategic integration of aromatic rings for enhanced thermal resistance and aliphatic or ether segments for improved fluidity. Overall, Fig. 5(b) shows a progressive shift of the Pareto frontier toward the favorable upper-right region. Within the surrogate-defined design space, this shift indicates improved trade-offs between predicted flash point and predicted pour point.

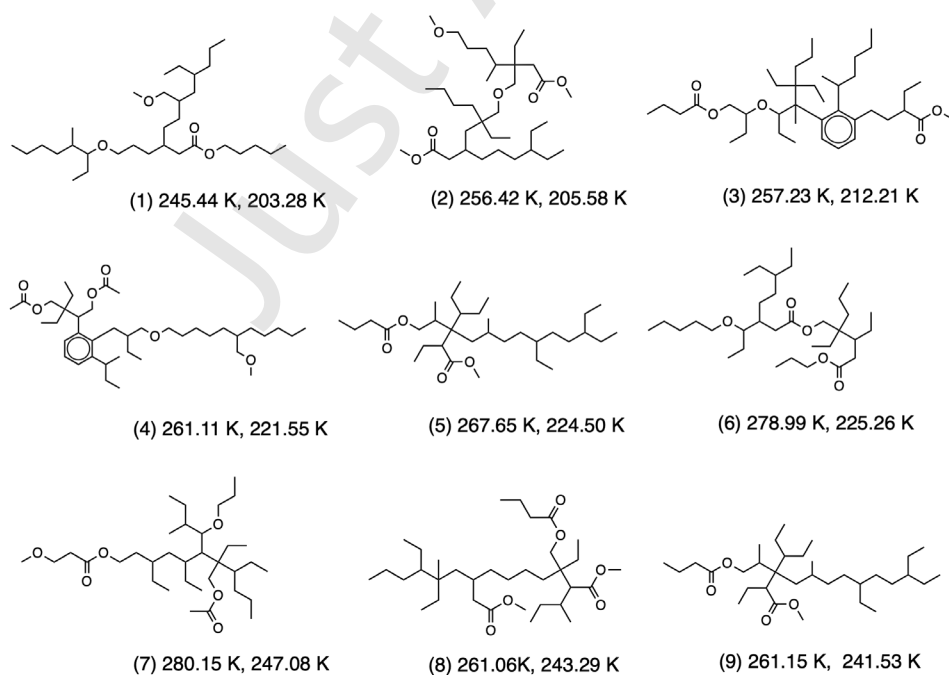


Fig. 6. Nine diverse ester-based lubricant molecules generated by LubZero. Surrogate-predicted flash points and pour points are reported below each molecule in K to illustrate the property profiles identified by the computational framework.

Representative optimized structures from the final frontier are shown in Fig. 6, comprising nine diverse ester-based molecules with aromatic rings for thermal stability, ester and ether linkages for functionality, and flexible aliphatic segments for enhanced low-temperature flow. The predicted flash and pour points are reported and indicate favorable thermal-safety and low-temperature-flow profiles relative to the selected screening objectives. Through MCTS-guided self-play, LubZero identifies high-ranking, structurally filtered screening candidates that balance predicted flash point and pour point, while final synthetic feasibility and measured properties remain to be confirmed experimentally or by higher-fidelity validation. Accordingly, Fig. 6 visualizes the surrogate-predicted property profiles of the generated candidates and illustrates LubZero's ability to identify high-scoring structures within the learned design space.

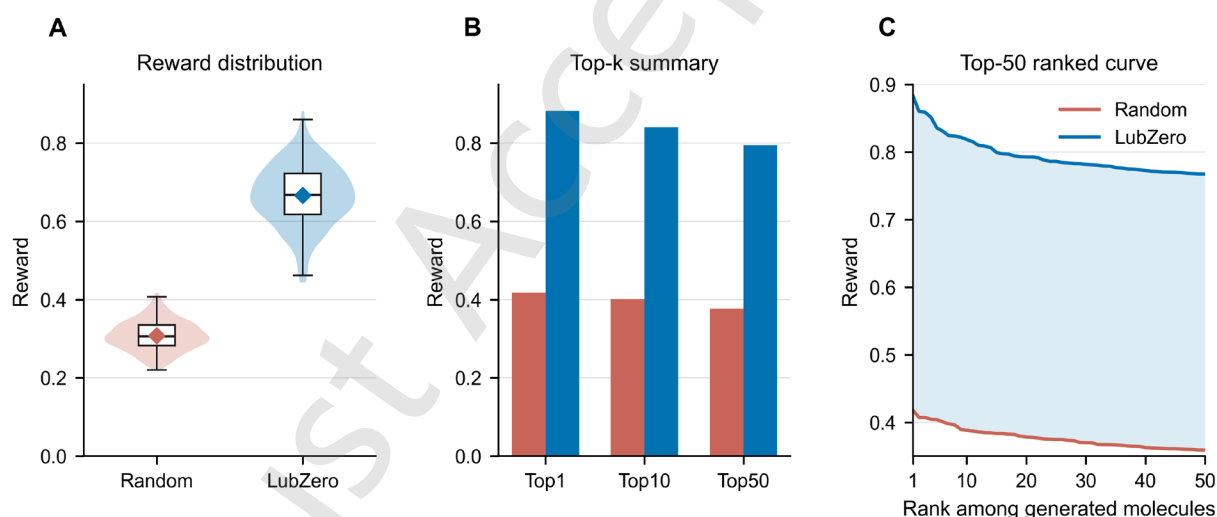


Fig. 7. Baseline evaluation between Random and LubZero generation results. (a) Reward distribution of generated molecules; (b) Top-k reward summary; (c) Top-50 ranked reward curve.

In a separate baseline evaluation, the final LubZero policy was compared with a Random generation baseline under the same generation and ranking protocol. This baseline test focuses on the reward distribution and top-ranked candidates rather than epoch-wise training averages. Under this constrained protocol, LubZero shows a clear upward shift in the reward distribution and consistently higher Top1, Top10, and Top50 rewards than the Random baseline, indicating that the learned policy improves generated-candidate quality relative to random exploration. This comparison should not be

interpreted as evidence of superiority over untested baselines such as genetic algorithms or standard reinforcement-learning variants.

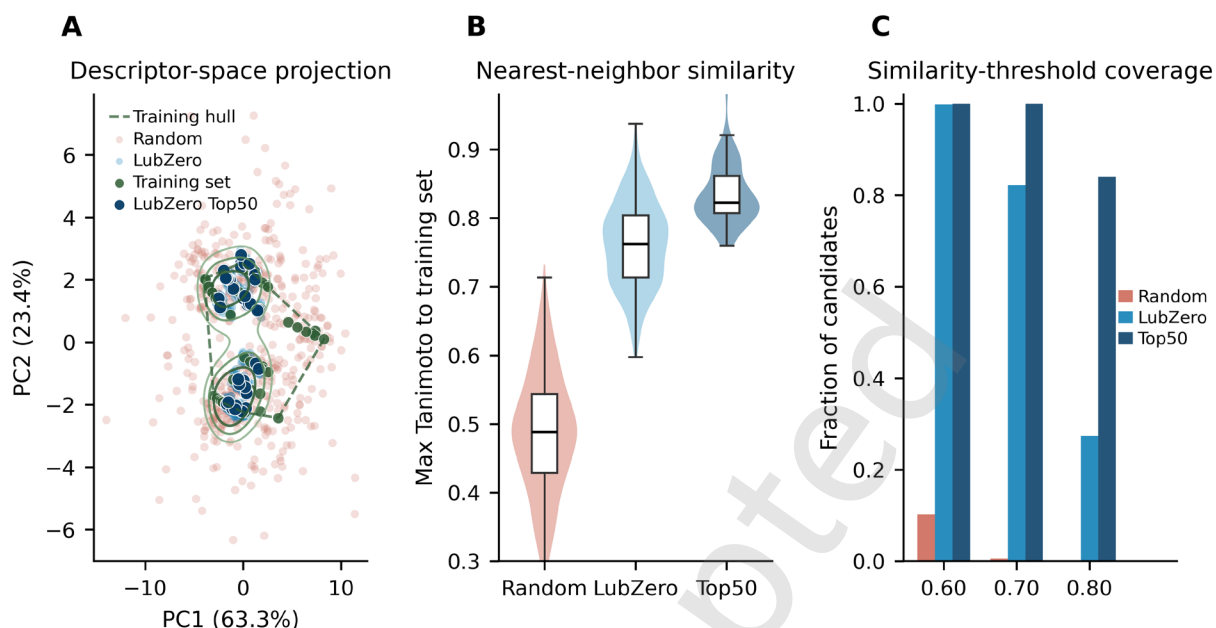


Fig. 8. Applicability-domain proximity diagnostic for surrogate-guided screening. (a) Descriptor-space PCA projection with training-density contours and the training convex hull. Random candidates are broadly dispersed with partial overlap, whereas LubZero and LubZero Top50 candidates concentrate within the training-density regions. (b) Maximum Morgan-fingerprint Tanimoto similarity to the training set. (c) Similarity-threshold coverage at 0.60, 0.70, and 0.80.

To further clarify the applicability-domain issue raised for the surrogate models, we added a descriptor- and fingerprint-based proximity diagnostic. Standardized RDKit molecular descriptors were projected into the first two principal components using the 75 ester molecules as the reference training set, and nearest-neighbor Morgan-fingerprint Tanimoto similarity was used to quantify structural proximity to the training data.

As shown in Fig. 8, the Random baseline is broadly dispersed in descriptor space and shows only partial overlap with the training-set density contours, whereas LubZero candidates, especially the Top50 subset, are concentrated within the high-density training descriptor regions. Consistently, the 500 candidates generated by the final trained LubZero policy for baseline evaluation reach a mean/median maximum Tanimoto similarity of 0.761/0.762, and LubZero Top50 reaches 0.835/0.836; the Top50 coverage at similarity thresholds of 0.60, 0.70, and 0.80 is 100%, 100%, and 76%, respectively. In contrast, the Random baseline shows weaker proximity to the training set (mean/median 0.485/0.488; 10.2%, 0.6%, and 0% coverage at the same thresholds). This behavior is

consistent with the ester-like and lubricant-like constraints imposed during LubZero generation, which restrict the search space toward molecular motifs represented in the ester training set rather than allowing unconstrained random exploration. These results provide an applicability-domain proximity diagnostic, but they do not by themselves prove surrogate generalization or replace experimental validation.

The rationality of the predicted molecules was further assessed by combining chemical validity, structural filtering, applicability-domain proximity analysis, and inspection of top-ranked structures. All retained candidates passed RDKit sanitization and lubricant-oriented structural constraints, including ester-like/lubricant-like requirements and restrictions on excessive ring or aromatic-ring formation. Moreover, the applicability-domain analysis in Fig. 8 indicates that the top-ranked LubZero candidates are closer to the descriptor and fingerprint space of the experimental ester lubricant dataset than the Random baseline. The representative high-ranking structures contain chemically interpretable ester/ether motifs and hydrocarbon skeletons, indicating that these high-reward candidates are plausible ester-lubricant screening leads rather than arbitrary graph artifacts. Final confirmation will still require experimental synthesis or higher-fidelity property validation.

4 Conclusions

In summary, we have developed LubZero, a specialized molecular generation framework that leverages the AlphaZero paradigm for surrogate-guided screening of ester lubricant candidates. By integrating Monte Carlo Tree Search (MCTS) with a GINE-based policy-value network, LubZero addresses sparse rewards and complex multi-property trade-offs within the constrained discrete chemical space studied here.

A key technical contribution of this work is the actor-server parallel self-play architecture. The acceleration is not merely a thread-count increase: parallel CPU actor processes perform MCTS search and chemical validity checks, whereas a shared GPU inference service batches policy/value evaluations from all actors. This separation improves GPU utilization, avoids redundant model loading across processes, and reduces wall-clock generation time for expensive MCTS-guided molecular searches. The resulting high-throughput pipeline enables LubZero to explore the constrained lubricant chemical space more efficiently while retaining the same molecular validity and structural filters used in the serial generation protocol. A controlled single-GPU timing benchmark is reported in SI Table 3: using the same trained LubZero policy-value model, MCTS = 200, and lubricant-like constraints, the one-actor serial setting required 3112.7 s to produce 24 accepted unique candidates (27.8 candidates per hour), whereas the 24-actor actor-server setting

required 181.9 s (475.0 candidates per hour), corresponding to a 17.11-fold throughput increase and a 94.2% reduction in wall-clock time.

Our results demonstrate that LubZero learns chemically intuitive assembly strategies through iterative self-improvement. The framework mapped a surrogate-predicted Pareto frontier for flash point and pour point, identifying high-ranking, structurally filtered screening candidates that extend the predicted trade-off space beyond the region covered by the available experimental data. These generated structures, featuring combinations of aromatic cores and flexible ether/aliphatic chains, provide candidate motifs for subsequent high-fidelity validation, synthesis planning, and experimental assessment.

The Random-baseline comparison supports LubZero's advantage over random generation under the constrained generation and ranking protocol evaluated here. While this study focused on ester materials and did not optimize all lubricant properties, the framework's modular design allows extension to other performance objectives, such as oxidative stability or biodegradability. Ultimately, the high-throughput, parallelized nature of LubZero provides a scalable platform for surrogate-guided screening and prioritization of functional-fluid candidates, with final utility depending on broader surrogate validation and experimental confirmation.

Supporting Information

Supplemental Figures: Detailed SHAP (SHapley Additive exPlanations) beeswarm plots and individual dependency analysis for flash point (SI Figure 1) and pour point (SI Figure 2), providing extended interpretability of the molecular descriptors.

Supplemental Tables: Comprehensive list and definitions of the top-ranked molecular descriptors (ATSC1se, SaasC, etc.) and their specific physicochemical significance in ester-based lubricants (SI Table 1), five-fold cross-validation mean $\pm$ standard deviation values for the surrogate-model configurations (SI Table 2), and single-GPU actor-timing comparison for serial and actor-server parallel generation (SI Table 3).

Competing interests

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

Funding

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDC0180401), and Natural Science Foundation of Shandong Province (Grant No. ZR2024QB379).

Author contributions

The individual contributions of authors to the manuscript should be specified in this section. The Journal recommends authors to use CRediT statements (the Contribution Roles Taxonomy).

References

- [1] Holmberg, K., & Erdemir, A. Influence of tribology on global energy consumption, costs and emissions. *Friction* 5(3): 263-284 (2017).
- [2] Holmberg, K., & Erdemir, A. The impact of tribology on energy use and CO₂ emission globally and in combustion engine and electric cars. *Tribology International* 135: 389-396 (2019).
- [3] Mobarak, H., Mohamad, E. N., Masjuki, H. H., Kalam, M. A., Al Mahmud, K., Habibullah, M., et al. The prospects of biolubricants as alternatives in automotive applications. *Renewable and sustainable energy reviews* 33: 34-43 (2014).
- [4] Wilińska, I., & Wilkanowicz, S. Advancements in Environmentally Friendly Lubricant Technologies: Towards Sustainable Performance and Efficiency. *Energies* 18(15): 4006 (2025).
- [5] Htwe, Y. Z. N., Al-Janabi, A. S., Wadzer, Y., & Mamat, H. Review of tribological properties of nanoparticle-based lubricants and their hybrids and composites. *Friction* 12(4): 569-590 (2024).
- [6] Ray, S., Rao, P. V., & Choudary, N. V. Poly-alpha-olefin-based synthetic lubricants: a short review on various synthetic routes. *Lubrication Science* 24(1): 23-44 (2012).
- [7] Liu, Z., Xue, W., Shi, J., Han, S., & Yan, J. Recent advances in polyalkylene glycol base oil. *Research on Chemical Intermediates* 50(4): 1515-1539 (2024).
- [8] Cañellas, G., Emeric, A., Combarros, M., Navarro, A., Beltran, L., Vilaseca, M., et al. Tribological performance of esters, friction modifier and antiwear additives for electric vehicle applications. *Lubricants* 11(3): 109 (2023).
- [9] Rosenkranz, A., Marian, M., Profito, F. J., Aragon, N., & Shah, R. The use of artificial intelligence in tribology-a perspective. *Lubricants* 9(1): 2 (2021).
- [10] Marian, M., & Tremmel, S. Current trends and applications of machine learning in tribology-a review. *Lubricants* 9(9): 86 (2021).
- [11] Tremmel, S., & Marian, M. Machine learning in tribology-more than buzzwords? *Lubricants* 10(4): 68 (2022).
- [12] Paturi, U. M. R., Palakurthy, S. T., & Reddy, N. S. The role of machine learning in tribology: a systematic review. *Archives of Computational Methods in Engineering* 30(2): 1345-1397 (2023).

- [13] Sose, A. T., Joshi, S. Y., Kunche, L. K., Wang, F., & Deshmukh, S. A. A review of recent advances and applications of machine learning in tribology. *Physical Chemistry Chemical Physics* 25(6): 4408-4443 (2023).
- [14] Yin, N., Yang, P., Liu, S., Pan, S., & Zhang, Z. AI for tribology: Present and future. *Friction* 12(6): 1060-1097 (2024).
- [15] Marian, M., Mursak, J., Bartz, M., Profito, F. J., Rosenkranz, A., & Wartzack, S. Predicting EHL film thickness parameters by machine learning approaches. *Friction* 11(6): 992-1013 (2023).
- [16] Shah, R., Sridharan, N. V., Mahanta, T. K., Muniyappa, A., Vaithiyanathan, S., Ramteke, S. M., & Marian, M. Ensemble deep learning for wear particle image analysis. *Lubricants* 11(11): 461 (2023).
- [17] Malaguti, R., Lourenço, N., & Silva, C. A supervised machine learning model for determining lubricant oil operating conditions. *Expert Systems* 40(5): e13116 (2023).
- [18] Shah, R., Marussich, K., Mittal, V., & Rosenkranz, A. Artificial Intelligence in Lubricant Research-Advances in Monitoring and Predictive Maintenance. *Lubricants* 14(2): 72 (2026).
- [19] Jia, D., Duan, H., Zhan, S., Jin, Y., Cheng, B., & Li, J. Design and development of lubricating material database and research on performance prediction method of machine learning. *Scientific Reports* 9(1): 20277 (2019).
- [20] Rahman, M. H., Shahriar, S., & Menezes, P. L. Recent progress of machine learning algorithms for the oil and lubricant industry. *Lubricants* 11(7): 289 (2023).
- [21] Zhou, R., Ma, R., Bao, L., Cai, M., Zhou, F., Li, W., et al. Lubrication Brain-A machine learning framework of lubrication oil molecule design. *Tribology International* 183: 108381 (2023).
- [22] Bengio, Y., Lahlou, S., Deleu, T., Hu, E. J., Tiwari, M., & Bengio, E. Gflownet foundations. *Journal of Machine Learning Research* 24(210): 1-55 (2023).
- [23] Nica, A. C., Jain, M., Bengio, E., Liu, C.-H., Korablyov, M., Bronstein, M. M., et al. Evaluating generalization in gflownets for molecule design. In *ICLR2022 Machine Learning for Drug Discovery*, 2022.
- [24] Qian, H., Lin, C., Zhao, D., Tu, S., & Xu, L. AlphaDrug: protein target specific de novo molecular generation. *PNAS nexus* 1(4): pgac227 (2022).
- [25] Popova, M., Isayev, O., & Tropsha, A. Deep reinforcement learning for de novo drug design. *Science advances* 4(7): eaap7885 (2018).
- [26] You, J., Liu, B., Ying, Z., Pande, V., & Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. *Advances in neural information processing systems* 31 (2018).

- [27] Zhou, Z., Kearnes, S., Li, L., Zare, R. N., & Riley, P. Optimization of molecules via deep reinforcement learning. *Scientific Reports* 9(1): 10752 (2019).
- [28] Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., Bekas, C., et al. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science* 5(9): 1572-1583 (2019).
- [29] Edwards, C., Lai, T., Ros, K., Honke, G., Cho, K., & Ji, H. Translation between molecules and natural language. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022 (pp. 375-413).
- [30] Vignac, C., Krawczuk, I., Siraudin, A., Wang, B., Cevher, V., & Frossard, P. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734* (2022).
- [31] SV, S. S., Law, J. N., Tripp, C. E., Duplyakin, D., Skordilis, E., Biagioni, D., et al. Multi-objective goal-directed optimization of de novo stable organic radicals for aqueous redox flow batteries. *Nature Machine Intelligence* 4(8): 720-730 (2022).
- [32] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. Mastering the game of go without human knowledge. *Nature* 550(7676): 354-359 (2017).
- [33] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* 362(6419): 1140-1144 (2018).
- [34] Xu, K., Hu, W., Leskovec, J., & Jegelka, S. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018).
- [35] Ehrgott, M., Köksalan, M., Kadziński, M., & Deb, K. Fifty years of multi-objective optimization and decision-making: From mathematical programming to evolutionary computation. *European Journal of Operational Research* 330(1): 1-25 (2026).
- [36] Landrum, G. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum* 8(31.10): 5281 (2013).
- [37] Ertl, P., & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics* 1(1): 8 (2009).
- [38] Rosin, C. D. Multi-armed bandits with episode context. *Annals of Mathematics and Artificial Intelligence* 61(3): 203-230 (2011).
- [39] Mangalathu, S., Hwang, S.-H., & Jeon, J.-S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures* 219: 110927 (2020).

- [40] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12: 2825-2830 (2011).
- [41] Van der Maaten, L., & Hinton, G. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9(11) (2008).



Dr. Luyao Bao is an Associate Professor at the Lanzhou Institute of Chemical Physics, Chinese Academy of Sciences. He received his bachelor's, master's, and doctoral degrees from Northwestern Polytechnical University. His research primarily focuses on molecular simulation engineering and the artificial intelligence design of high-performance lubricant additives and synthetic lubrication materials.



Dr. Rui Zhou earned his Ph.D. from the University of Chinese Academy of Sciences (UCAS) and is currently a Postdoctoral Fellow at the Lanzhou Institute of Chemical Physics (LICP), CAS. His research focuses on ester lubricant systems and their additives, where he integrates high-throughput computing with machine learning to facilitate the rational design and experimental validation of novel molecules.



Dr. Xiangmeng Jia received his Ph.D. from Jilin University in 2021, pursued postdoctoral research at South China University of Technology starting in 2022. Since 2024, he has been assistant researcher at the Shandong Laboratory of Advanced Materials and Green Manufacturing at Yantai. He has been focusing on theoretical modeling and simulation of functional polymers.



Dr. Feng Zhou is a Full Professor in Lanzhou Institute of Chemical Physics, Chinese Academy of Sciences, China, and head of the State Key Laboratory of Solid Lubrication. He gained his PhD in 2004 and spent three years (2005–2008) in the Department of Chemistry, University of Cambridge as a Research Associate. He has published more than 300 journal papers, which have received more than 18 000 citations and have a high-index of 72. His research interests include bioinspired tribology, biomimic surfaces/interfaces of soft matter, drag-reduction, anti-biofouling, and boundary lubrication. He has gained a number of awards including the “Outstanding Youth Award” of the International Society of Bionic Engineering, 2013, and National Award for Natural Sciences (Second Class), 2015. He serves as an Editorial Board Member of Tribology International, Friction, Journal Fiber Bioengineering and Informatics, Coatings etc.