PERSISTENT PREFERENCE INFERENCE FOR QUERY-EFFICIENT INTERACTIVE MULTI-OBJECTIVE OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Interactive multi-objective optimization uses decision-maker feedback to steer search. However, repeated consultations become costly as candidate populations evolve during optimization. We ask whether earlier comparisons can reduce the decision-maker feedback required to select preferred candidates from subsequent populations. We introduce PPL, a persistent preference inference framework that uses linear programming to certify pairwise relations over a feasible preference region maintained across consultations and compatible with historical comparisons. Only comparisons that cannot be resolved using previously elicited feedback require new queries. We prove an expected $\mathcal{O}(m \log N)$ upper bound on the number of explicit queries required by PPL to infer a preference-aligned ranking. We further evaluate its practical query efficiency by constructing preference-aligned model leaderboards from pairwise comparison data on Chatbot Arena. In this setting, PPL achieves ranking quality comparable to the baseline with 98.9% fewer target-user queries. When integrated into multi-objective optimization to guide search, PPL improves preference-region alignment over competing interactive methods under matched cumulative query budgets across 36 benchmark configurations and an RfaH protein sequence design task. These improvements are achieved while resolving more than 99.3% of processed pairwise comparisons through inference from previously elicited preference feedback. The results demonstrate that reusing preference information across evolving populations can substantially reduce interaction without compromising preference-guided search.

1 INTRODUCTION

Balancing multiple objectives is a central challenge in applications ranging from architected material design (Peng et al., 2023) and enzyme engineering (Ding et al., 2024) to the alignment of foundation models with human preferences (Yang et al., 2024). Multi-objective optimization (MO) provides a systematic framework for exploring the resulting trade-offs, while the choice among alternative solutions depends on application requirements and decision-maker (DM) preferences. Beyond approximating the Pareto front, an important practical goal is therefore to identify solutions that reflect the DM's priorities (Li et al., 2020). Interactive MO supports this goal by eliciting preferences during search and directing computation toward decision-relevant regions. Because each preference query may impose substantial cognitive burden on the DM Low et al. (2024), an effective method must deliver useful solutions within a limited cumulative query budget. This is particularly challenging in preference-based evolutionary MO (PBEMO), where the candidate population changes from one generation to the next (Thiele et al., 2009; Huang et al., 2024). Earlier judgments often concern candidates that soon disappear, while each new generation introduces previously unseen alternatives. Yet earlier judgments may still reveal preference information that can inform comparisons among newly generated alternatives. This leads to the central question: *how can feedback about past comparisons reduce the additional queries needed to select the preferred solutions from later populations?*

Existing PBEMO methods partially answer this question. Many fit a value, utility, or ranking model to accumulated comparisons and use the fitted model to guide later search (Branke et al., 2015; Li et al., 2019; 2023). This preserves historical feedback, but subsequent selection depends on how well the fitted model generalizes across populations continually reshaped by evolutionary search.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Compatibility-based methods, notably robust ordinal regression, instead retain a set of consistent models and identify relations shared by all of them (Corrente et al., 2013). Within this line of research, active approaches select informative comparisons among a fixed set of alternatives (Branke et al., 2017). Related geometric active-ranking methods infer Euclidean reference-point rankings from previously elicited comparisons, querying only unresolved relations (Jamieson & Nowak, 2011). In evolutionary optimization, compatible models have also been used to guide changing populations (Branke et al., 2009; Tomczyk & Kadzinski, 2020; 2026). Building on these developments, we focus on retaining a feasible region of compatible aspiration points across successive consultations, so that feedback about candidates no longer present can inform current ranking decisions under a shared cumulative query budget.

To address this gap, we use a *latent aspiration-point preference model*, in which candidates are ranked by Euclidean distance to an unknown target point in objective space. This model captures preferences expressed through proximity to a target rather than global trade-off weights. Normalization is fixed across consultations, keeping historical comparisons and new candidates in a common coordinate system. Pairwise comparisons cancel the quadratic terms in the unknown aspiration point and thus induce linear halfspace constraints. Their intersection defines a convex preference-consistent region, which is retained across consultations and refined with newly elicited feedback. Building on this representation, we propose `PPL`, a preference-consistent active ranking framework. A linear-programming (LP) test certifies pairwise relations over the maintained region, while transitive propagation resolves further comparisons without feedback. Only unresolved comparisons require DM queries, subject to the remaining budget. Alongside the ranking, a soft-margin LP estimates an aspiration point for preference-aware evolutionary selection. Under a realizable aspiration-point model and an informative-query assumption, we establish an expected $\mathcal{O}(m \log N)$ upper bound on explicit queries, where m is the number of objectives and N is the candidate-set size.

We instantiate `PPL` in three representative EMO frameworks, namely NSGA-II (Deb et al., 2002b), MOEA/D (Zhang & Li, 2007), and IBEA (Zitzler & Künzli, 2004), and evaluate the resulting algorithms under same query budgets. Our experiments span standard multi-objective benchmarks and the computational bio-sequence design case, RfaH protein sequence design, using controlled simulated preference feedback. Across these settings, `PPL` substantially improves preference-guided search performance when integrated with NSGA-II, IBEA and MOEA/D. Diagnostic controls and ablation studies examine the contributions of relation inference, cross-consultation memory, and preference-guided population updates, and quantify sensitivity to preference-model mismatch and the associated computational overhead. We further evaluate `PPL` independently as a preference-learning module, measuring its identification accuracy and explicit-query efficiency outside the end-to-end EMO loop. To examine whether the underlying preference-inference mechanism transfers beyond evolutionary optimization, we additionally evaluate it on Chatbot Arena data. In this setting, `PPL` achieves an area under the learning curve (AULC) comparable to that of the strongest baseline while requiring 98.9% fewer explicit target-user queries.

2 Problem Setting

We consider an interactive PBEMO setting with two interacting components: candidate sets that evolve throughout the optimization process and pairwise preference feedback from DMs on these evolving candidates. Throughout, we distinguish decision-space variables from their objective-space representations: $\mathbf{x} \in \Omega \subset \mathbb{R}^n$ denotes an n-dimensional decision vector, while $\phi = \mathbf{F}(\mathbf{x}) \in \mathbb{R}^m$ denotes its m-dimensional objective vector.

Evolving candidate sets. We consider minimizing m potentially conflicting objectives,

$$\min_{\mathbf{x}\in\Omega} \mathbf{F}(\mathbf{x}) := \big(f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\big)^\top, \qquad \Omega \subseteq \mathbb{R}^n, \tag{1}$$

where $\mathbf{x}$ is an n-dimensional decision vector. Let T be the maximum number of generations, and let $\mathcal{T}_{\text{int}} = \{g_1 < \cdots < g_J\} \subseteq \{0, \ldots, T-1\}$ be the set of interaction generations. At generation t, the underlying EMO algorithm supplies N_t indexed candidates. At generation t, let U_t be the indexed pool of individuals supplied by the host EMO algorithm. The consultation set contains the distinct objective vectors represented in that pool: $\Phi_t := \{\mathbf{F}(\mathbf{x}) : \mathbf{x} \in U_t\} = \{\phi_i^{(t)}\}_{i=1}^{N_t}$, and $N_t := |\Phi_t|$. We retain a map $\kappa_t : U_t \to [N_t]$ satisfying $\phi^{(t)}_{\kappa_t(\mathbf{x})} = \mathbf{F}(\mathbf{x})$. Individuals with identical objective

Algorithm 1 NSGA-II with PPL-guided preference feedback

Require: Population size N_{pop}, generations T, interaction set $\mathcal{T}_{\text{int}}$, budgets $\{B_t\}_{t\in\mathcal{T}_{\text{int}}}$, oracle Query
Ensure: Final population R_T
1: Initialize R_0 with $|R_0| = N_{\text{pop}}$; $\mathcal{C} \leftarrow \emptyset$; $\widehat{\mathbf{z}}_{-1} \leftarrow \varnothing$
2: **for** $t = 0$ to $T-1$ **do**
3: $U_t \leftarrow \texttt{GenerateAndCombine}(R_t)$; $(\Phi_t, \kappa_t) \leftarrow \texttt{DistinctObjectives}(U_t)$
4: $\widehat{\sigma}_t \leftarrow \varnothing$; $\widehat{\mathbf{z}}_t \leftarrow \widehat{\mathbf{z}}_{t-1}$
5: **if** $t \in \mathcal{T}_{\text{int}}$ **then** $(\widehat{\sigma}_t, \mathcal{C}, M_t) \leftarrow \texttt{PPL-Rank}(\Phi_t, \mathcal{C}, B_t, \text{Query})$; $\widehat{\mathbf{z}}_t \leftarrow \texttt{SoftEstimate}(\mathcal{C})$
6: **end if**
7: $R_{t+1} \leftarrow \texttt{PreferenceAwareSelect}(U_t, \kappa_t, \widehat{\sigma}_t, \widehat{\mathbf{z}}_t, N_{\text{pop}})$
8: **end for**
9: **return** R_T

vectors share one consultation candidate but remain distinct in environmental selection. Within a fixed consultation set, we omit the superscript t, write $N := N_t$, and let $[N] := \{1, \dots, N\}$.

Persistent pairwise feedback. Let $\succ_\star$ denote the DM's latent preference relation over objective vectors in original coordinates. A query between two candidates with $\phi_i \neq \phi_j$ returns a binary response $o_{ij} \in \{+1, -1\}$. For a pair on which the DM has a strict preference, feedback is $o_{ij} = \begin{cases} +1, & \phi_i \succ_\star \phi_j, \\ -1, & \phi_j \succ_\star \phi_i. \end{cases}$ This preference is distinct from Pareto dominance, which is determined by component-wise objective values, whereas o_{ij} reflects the DM's latent trade-off among objectives. Section 3.1 specifies the persistent aspiration-point model. The hard-mode correctness guarantee relies on the realizability and no-ties conditions in Assumption 4.1. Outside this regime, any supplied binary labels still define an execution of the algorithm, but the hard-mode correctness guarantee no longer applies.

3 Persistent Preference Inference

We propose PPL, which reuses previously queried comparisons to infer preference relations and seeks to reduce explicit queries. At interaction generation t, given the current consultation set Φ_t, PPL produces two preference outputs: a ranking $\widehat{\sigma}_t$ constructed through hard-mode certification and transitive propagation, with fallback inference when necessary, and an aspiration-point estimate $\widehat{\mathbf{z}}_t$ obtained by minimizing average soft-margin loss, followed by anchor-based and lexicographic tie-breaking. The ranking guides selection among the current candidates, whereas the aspiration-point estimate provides a reusable preference signal for subsequent generations. For concreteness, Algorithm 1 illustrates the integration of PPL with NSGA-II. The corresponding integrations with MOEA/D and IBEA are described in Appendix B. We distinguish the population size N_{pop} from the number N of distinct objective vectors in a consultation. Each interaction generation has an integer explicit-query budget $B_t \geq 0$.

3.1 Persistent region

We use a persistent aspiration-point model to certify pairwise relations during the hard phase. All objective vectors used by the persistent model are expressed in their original coordinates, with fixed units across consultations. Aspiration points are expressed in the same coordinate system. Let $Z \subset \mathbb{R}^m$ be a nonempty compact polytope of admissible aspiration points. In the model-compatible setting, the DM's preferences are induced by a fixed aspiration point $\mathbf{z}^\star \in Z$. Its distinct objective vectors form Φ_t,

$$h_{\phi,\phi'}(\mathbf{z}) := \frac{\|\phi' - \mathbf{z}\|_2^2 - \|\phi - \mathbf{z}\|_2^2}{2\|\phi - \phi'\|_2}. \tag{2}$$

The sign of $h_{\phi,\phi'}(\mathbf{z})$ determines the preference induced by $\mathbf{z}$: $h_{\phi,\phi'}(\mathbf{z}) > 0$ means that ϕ is closer to $\mathbf{z}$ than ϕ'. In particular, under the fixed aspiration-point model, $\phi \succ_\star \phi'$ means $h_{\phi,\phi'}(\mathbf{z}^\star) > 0$.

Let $\mathcal{C}$ contain the explicit pairwise answers retained from previous consultations and obtained during the current consultation. An observation is written as (ϕ, ϕ', o), where $o \in \{+1, -1\}$ and $o = +1$ means that ϕ was preferred to ϕ'. Define the compatible aspiration-point region $Z(\mathcal{C}) := \{\mathbf{z} \in Z : o\, h_{\phi,\phi'}(\mathbf{z}) \geq 0,\ \forall(\phi, \phi', o) \in \mathcal{C}\}$. In particular, $Z(\emptyset) = Z$. Since each observation contributes a closed halfspace, $Z(\mathcal{C})$ is a compact polytope whenever it is nonempty, and it can

Algorithm 2 PPL-Rank: hard certification, propagation, and fallback

Require: Distinct objective vectors $\Phi = \{\phi_i\}_{i=1}^N$, queried observations $\mathcal{C}$, query budget B, oracle Query
Ensure: Ranking $\widehat{\sigma}$, updated observations $\mathcal{C}$, mode $M \in \{\text{HARD}, \text{FALLBACK}\}$

$b \leftarrow B$
if $Z(\mathcal{C}) = \emptyset$ **then**
  return $(\texttt{FallbackRank}(\Phi, \mathcal{C}), \mathcal{C}, \text{FALLBACK})$
end if
$\mathbf{Q} \leftarrow \mathbf{0}_{N\times N}$, $\rho \leftarrow$ random permutation of $[N]$, $\mathcal{S} \leftarrow ()$
for $r = 1$ to N **do**
  $j \leftarrow \rho(r)$
  while $\exists k \in \mathcal{S} : \mathbf{Q}_{kj} = 0$ **do**
    $\mathcal{U} \leftarrow \{k \in \mathcal{S} : \mathbf{Q}_{kj} = 0\}$, $i \leftarrow \text{Median}_{\mathcal{S}}(\mathcal{U})$
    Compute α_{ij}^+ and α_{ij}^- using equation 3
    if $(\alpha_{ij}^+, \alpha_{ij}^-) = (1, 0)$ **then** $o \leftarrow +1$
    else if $(\alpha_{ij}^+, \alpha_{ij}^-) = (0, 1)$ **then** $o \leftarrow -1$
    else if $(\alpha_{ij}^+, \alpha_{ij}^-) = (1, 1)$ and $b > 0$ **then** $o \leftarrow \text{Query}(\phi_i, \phi_j)$; $\mathcal{C} \leftarrow \mathcal{C} \cup \{(\phi_i, \phi_j, o)\}$; $b \leftarrow b - 1$
    else return $(\texttt{FallbackRank}(\Phi, \mathcal{C}), \mathcal{C}, \text{FALLBACK})$
    end if
    $\mathbf{Q}_{ij} \leftarrow o$, $\mathbf{Q}_{ji} \leftarrow -o$; $\mathbf{Q} \leftarrow \text{Propagate}(\mathbf{Q}, \mathcal{S}, i, j, o)$
  end while
  Insert j into the unique position in $\mathcal{S}$ consistent with $\mathbf{Q}$
end for
return $(\mathcal{S}, \mathcal{C}, \text{HARD})$

only contract as queried observations are added. The region is a weak compatibility representation: a point on an observation's indifference hyperplane satisfies its constraint without inducing a strict preference. Consequently, nonemptiness alone does not establish strict realizability of the feedback. The region $Z(\mathcal{C})$ represents the persistent preference state. Crucially, $\mathcal{C}$ stores only explicitly queried comparisons. Certified and transitively implied relations are not appended to $\mathcal{C}$; they are maintained separately within each PPL call.

3.2 HARD-MODE CERTIFICATION WITH TRANSITIVE PROPAGATION

For a fixed consultation set $\Phi = \{\phi_i\}_{i=1}^N$, PPL constructs a ranking $\widehat{\sigma} = (\widehat{\sigma}(1), \ldots, \widehat{\sigma}(N))$. For any ranking σ, let $\text{pos}_\sigma(i)$ denote the position of item i. We write $\phi_i \succ_\sigma \phi_j$ when $\text{pos}_\sigma(i) < \text{pos}_\sigma(j)$. The ranking is built by incremental insertion. The procedure maintains a local relation matrix $\mathbf{Q} \in \{-1, 0, +1\}^{N\times N}$, a random insertion order ρ over $[N] = \{1, \ldots, N\}$, and an ordered chain $\mathcal{S}$ of previously inserted candidates. Here, $\mathbf{Q}_{ij} = +1$ places i ahead of j, $\mathbf{Q}_{ij} = -1$ records the reverse ordering, and $\mathbf{Q}_{ij} = 0$ denotes an unresolved relation. The diagonal remains zero. Selecting the median unresolved candidate in chain order implements binary-search insertion.

Hard LP certification. Given a pivot i and the current target j, the procedure checks the feasibility of both orientations:

$$\begin{aligned} \alpha_{ij}^+ &:= \mathbf{1}\{Z\big(\mathcal{C} \cup \{(\phi_i, \phi_j, +1)\}\big) \neq \emptyset\}, \\ \alpha_{ij}^- &:= \mathbf{1}\{Z\big(\mathcal{C} \cup \{(\phi_i, \phi_j, -1)\}\big) \neq \emptyset\}. \end{aligned} \tag{3}$$

Both tests are LP feasibility problems by equation 2. If $(\alpha_{ij}^+, \alpha_{ij}^-) = (1, 0)$, every point in $Z(\mathcal{C})$ satisfies $h_{\phi_i,\phi_j}(\mathbf{z}) > 0$, so the procedure certifies $o = +1$. The case $(0, 1)$ similarly certifies $o = -1$. These certifications consume no explicit queries. If $(\alpha_{ij}^+, \alpha_{ij}^-) = (1, 1)$, the pair is unresolved by the hard test and is queried if budget remains. Because the feasibility constraints are non-strict, this case can occur when the feasible region merely touches the indifference hyperplane; it need not imply that both strict orientations are realizable. If $(\alpha_{ij}^+, \alpha_{ij}^-) = (0, 0)$, then $Z(\mathcal{C}) = \emptyset$, and fallback inference is required. Conversely, a nonempty region always admits at least one orientation.

Here, b is the remaining budget in the current consultation. Exhausting it does not prevent further certification or propagation: fallback is triggered only when an explicit query is required. Under exact feasibility checks, adding a label in the $(1, 1)$ case preserves nonemptiness, since either label selects a halfspace already known to intersect the current region. Thus, an initially nonempty region cannot

become empty through these query updates alone. The empty-region branch handles an incompatible incoming history; nonemptiness is not a general detector of noisy or contradictory strict feedback.

Transitive propagation. Once the direct relation $o \in \{+1, -1\}$ between i and j has been determined, it is propagated along the ordered chain: $\mathcal{T}_{ij}(o) := \begin{cases} \{k \in \mathcal{S} : \mathbf{Q}_{ki} = +1\}, & o = +1, \\ \{k \in \mathcal{S} : \mathbf{Q}_{ik} = +1\}, & o = -1. \end{cases}$ For every unresolved $k \in \mathcal{T}_{ij}(o)$, set $\mathbf{Q}_{kj} \leftarrow o$ and $\mathbf{Q}_{jk} \leftarrow -o$. If i precedes j, all predecessors of i also precede j; if j precedes i, all successors of i also follow j. This resolves one branch of the unresolved interval at zero query cost. Relations recorded in $\mathbf{Q}$ remain local and are not added to $\mathcal{C}$.

3.3 Fallback and EMO Integration

Soft-margin aspiration-point estimation. After each consultation, regardless of its termination mode, we compute a soft estimate from the retained explicitly queried observations. The fit uses the comparison score $h_{\phi,\phi'}$ defined in equation 2. Fix a fitting margin $\tau > 0$ and an anchor $\mathbf{z}_{\mathrm{a}} \in \mathbb{R}^m$, both unchanged across consultations.

Let $n_{\mathcal{C}} := |\mathcal{C}|$. For each observation $e = (\phi_e, \phi'_e, o_e) \in \mathcal{C}$, introduce a separate slack variable ξ_e. We first solve

$$L^\star(\mathcal{C}) := \min_{\mathbf{z} \in Z,\, \boldsymbol{\xi} \geq 0} \frac{1}{\max\{1, n_{\mathcal{C}}\}} \sum_{e \in \mathcal{C}} \xi_e \quad \text{s.t.} \quad o_e h_{\phi_e,\phi'_e}(\mathbf{z}) \geq \tau - \xi_e,\ \forall e \in \mathcal{C}. \tag{4}$$

Let $\mathcal{M}(\mathcal{C})$ denote the set of aspiration points occurring in optimal solutions of equation 4. Among these points, we minimize the ℓ_1 distance to the fixed anchor and resolve any remaining multiplicity lexicographically:

$$\widehat{\mathbf{z}}_{\mathrm{S}} := \mathrm{lexmin} \left(\arg \min_{\mathbf{z} \in \mathcal{M}(\mathcal{C})} \|\mathbf{z} - \mathbf{z}_{\mathrm{a}}\|_1 \right). \tag{5}$$

Thus, proximity to the anchor is used only to select among aspiration points with minimum average margin loss; it is not traded against that loss. The slack variables make the fitting problem feasible even when $Z(\mathcal{C})$ is empty. Since Z is nonempty and compact, the estimate is well defined. Both fitting and anchor-based selection can be solved by linear programming, using the standard linearization of the ℓ_1 norm; the remaining lexicographic selection can also be implemented by successive LPs. When $\mathcal{C} = \emptyset$, the first-stage objective is zero and $\mathcal{M}(\emptyset) = Z$, so the estimate is the lexicographically smallest point in Z minimizing its ℓ_1 distance to the anchor. The fixed anchor $\mathbf{z}_{\mathrm{a}}$ is specified in the original objective coordinates. The positive margin τ is a fitting parameter, distinct from the zero threshold used to define the hard compatible region $Z(\mathcal{C})$. It does not impose an additional margin assumption on the DM in the hard-mode analysis. We write $\texttt{SoftEstimate}(\mathcal{C}) = \widehat{\mathbf{z}}_{\mathrm{S}}$.

Fallback ranking. Fallback is invoked when a required query cannot be issued because $b = 0$, or when the incoming compatible region is empty. If $Z(\mathcal{C}) \neq \emptyset$, let $\widehat{\mathbf{z}}_{\mathrm{H}}$ be its Euclidean Chebyshev center: a center of a largest Euclidean ball contained in $Z(\mathcal{C})$. Among multiple centers, select the lexicographically smallest one. This convention also specifies a center when the region has lower dimension and its maximum inscribed-ball radius is zero. Define $\widehat{\mathbf{z}}_{\mathrm{F}} := \begin{cases} \widehat{\mathbf{z}}_{\mathrm{H}}, & Z(\mathcal{C}) \neq \emptyset, \\ \widehat{\mathbf{z}}_{\mathrm{S}}, & Z(\mathcal{C}) = \emptyset. \end{cases}$ When $Z(\mathcal{C}) = \emptyset$, $\texttt{FallbackRank}$ evaluates $\texttt{SoftEstimate}(\mathcal{C})$ internally to obtain $\widehat{\mathbf{z}}_{\mathrm{S}}$. The routine $\texttt{FallbackRank}(\Phi, \mathcal{C})$ sorts the entire current consultation set by increasing $\|\phi_i - \widehat{\mathbf{z}}_{\mathrm{F}}\|_2^2$, resolving equal distances by a fixed deterministic order. It replaces the partially constructed chain, returns a complete ranking, and issues no further queries in that call. Fallback relations are estimates rather than hard certificates and are not appended to $\mathcal{C}$. The fallback point $\widehat{\mathbf{z}}_{\mathrm{F}}$ is used only to construct the fallback ranking. The continuous output passed to the host EMO algorithm is always the soft estimate $\widehat{\mathbf{z}}_t = \widehat{\mathbf{z}}_{\mathrm{S}}$, defined by equation 4 and equation 5, as specified in Algorithm 1. Between interaction generations, this estimate is retained, while the ranking is reset because it refers to the consultation set on which it was computed.

Preference-aware environmental selection. NSGA-II uses the combined pool U_t of parents and offspring. Its distinct objective vectors form Φ_t, and κ_t associates each individual with its consultation representative. Environmental selection retains nondominated sorting and copies fronts that fit within the next-generation population budget. If a front overflows the remaining capacity, a fresh ranking

$\widehat{\sigma}_t$, when available, orders individuals by the positions of their representatives $\kappa_t(\mathbf{x})$. Individuals sharing a representative are resolved by a fixed deterministic order. When no fresh ranking is available but an aspiration-point estimate has been retained, individuals are prioritized by the score $s_t(\phi) := -\|\phi - \widehat{\mathbf{z}}_t\|_2^2$, with crowding distance used to resolve score ties and a fixed deterministic order used for any remaining ties. Before the first interaction, selection follows standard NSGA-II. In MOEA/D, the soft estimate guides weight-vector adaptation toward the preferred region; in IBEA, distances to this estimate modify the ϵ-indicator among competing nondominated solutions. Further integration details are given in Appendix B.

4 Conditional Query Complexity

We count only explicit calls to Query during one consultation, up to hard-mode completion or the switch to fallback. LP certification and transitive propagation require no DM feedback. Fix $\Phi = \{\phi_i\}_{i=1}^N$ with $N \geq 2$ distinct vectors, let $\mathcal{C}_0$ be the incoming queried history, let B be the consultation's query budget, and let Q be the number of new queries. The candidate set and incoming history are fixed; the expectation below is over the insertion order. Write $h_{ij} := h_{\phi_i,\phi_j}$ using equation 2. The proof is given in Appendix D.

A strict ranking σ is realizable if some $\mathbf{z} \in Z$ satisfies $\phi_i \succ_\sigma \phi_j \iff h_{ij}(\mathbf{z}) > 0,\ i \neq j$. Let $\mathcal{L}(\Phi)$ be the set of such rankings and $N_{\mathrm{rank}} := |\mathcal{L}(\Phi)|$.

Assumption 4.1 (Persistent realizability and no ties)**.** There exists a fixed $\mathbf{z}^\star \in Z$ such that $o\, h_{\phi,\phi'}(\mathbf{z}^\star) > 0$ for all $(\phi, \phi', o) \in \mathcal{C}_0$, and, for every current pair $i \neq j$, $\mathrm{Query}(\phi_i, \phi_j) = \mathrm{sign}\big(h_{ij}(\mathbf{z}^\star)\big)$ with $h_{ij}(\mathbf{z}^\star) \neq 0$.

Definition 4.2 (Consistent ranking set)**.** For the current queried history $\mathcal{C}$, define $\Sigma(\mathcal{C}) := \{\sigma \in \mathcal{L}(\Phi) : \text{some } \mathbf{z} \in Z(\mathcal{C}) \text{ induces } \sigma\}$. Under Assumption 4.1, the ground-truth ranking $\sigma^\star$ belongs to $\Sigma(\mathcal{C})$ throughout the consultation.

Assumption 4.3 (Conditional η-splitting)**.** There exists a fixed $\eta \in (0, 1/2]$ such that, for every explicitly queried pair (ϕ_i, ϕ_j), the history $\mathcal{C}$ immediately before the query satisfies $\max\{|\Sigma_{ij}^+(\mathcal{C})|, |\Sigma_{ij}^-(\mathcal{C})|\} \leq (1-\eta)|\Sigma(\mathcal{C})|$, where $\Sigma_{ij}^+(\mathcal{C})$ contains rankings in $\Sigma(\mathcal{C})$ placing i before j, and $\Sigma_{ij}^-(\mathcal{C})$ those placing j before i. For randomized insertion, this condition holds almost surely at every issued query.

The analysis uses a version-space splitting condition (Mussmann & Liang, 2018). It is an assumption on the actual queries, not a consequence of choosing the median candidate in the insertion chain.

Theorem 4.4 (Conditional explicit-query bound)**.** *Define $A_N := \sum_{r=2}^N \lceil \log_2 r \rceil$. For every execution of Algorithm 2, $Q \leq \min\{B, A_N\}$. Under Assumptions 4.1 and 4.3, almost surely, $Q \leq \min\Big\{B,\ A_N,\ \Big\lfloor \frac{\ln|\Sigma(\mathcal{C}_0)|}{-\ln(1-\eta)} \Big\rfloor\Big\}$. Moreover, with $P := \binom{N}{2}$,*

$$N_{\mathrm{rank}} \leq \sum_{r=0}^{\min\{m,P\}} \binom{P}{r} \leq (P+1)^m. \tag{6}$$

Consequently, $\mathbb{E}[Q] \leq \min\Big\{B,\ A_N,\ \frac{m\ln(P+1)}{-\ln(1-\eta)}\Big\} = \mathcal{O}\Big(\frac{m\ln N}{\eta}\Big)$. The last order bound becomes $\mathcal{O}(m \ln N)$ when η is bounded below independently of N and m. Separately, under Assumption 4.1, any hard-mode completion returns the true full ranking $\sigma^\star$; this correctness statement does not require the splitting assumption.

Non-strict LP constraints can leave a pair unresolved solely because of tie points; such queries need not shrink $\Sigma(\mathcal{C})$ and are excluded by the splitting assumption. Certification and propagation resolve relations without additional queries. The theorem bounds interaction cost, not computational cost or fallback accuracy; exact ranking recovery additionally requires hard-mode completion. The total number of DM queries over an optimization run is the sum of the per-consultation query counts; the corresponding accounting is given in Appendix D.

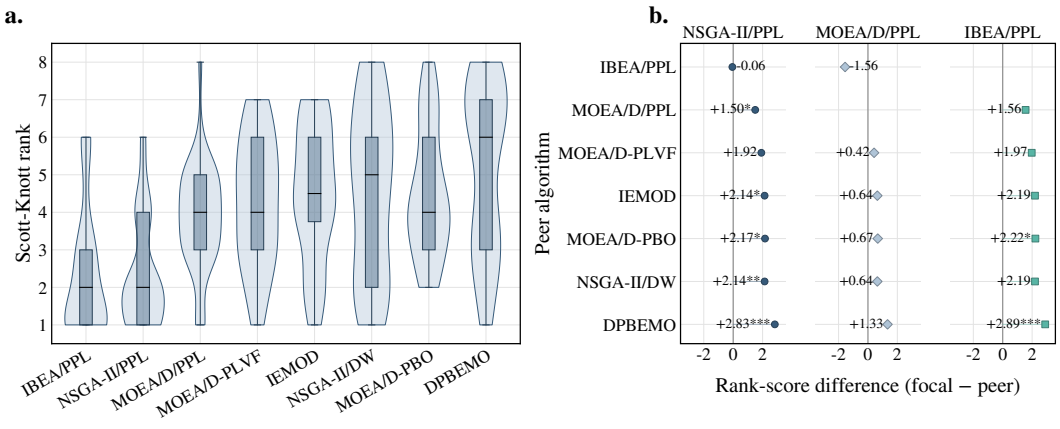

Figure 1: End-to-end comparison on 36 benchmark configurations using $\mathcal{D}(S)$; lower values are better. Each algorithm uses 20 random seeds and at most 500 pairwise preference queries per run. **(a)** Distributions of Scott–Knott group ranks across algorithms; lower ranks are better. **(b)** Mean rank advantages of each PPL integration over peer algorithms; positive values favor PPL. Asterisks mark significant differences under Holm-corrected paired Wilcoxon signed-rank tests.

5 Experimental Setup

We evaluate PPL both in end-to-end interactive MO and as a standalone preference-inference module. For the former, we integrate it with NSGA-II, MOEA/D, and IBEA, representing dominance-, decomposition-, and indicator-based search, respectively. Appendix F provides complete problem definitions, parameter settings, implementation details, and additional results.

End-to-End Evaluation in Interactive EMO. Benchmarks and comparison methods. We use the ZDT, mDTLZ, minus-DTLZ, DTLZ, and WFG families, covering diverse Pareto-front geometries in low- and many-objective settings, together with RfaH protein sequence design Hong & Kortemme (2024). Scalable problems use $m \in \{3, 5, 8, 10\}$ objectives. We compare the three PPL-enhanced algorithms with five preference-aware methods: MOEA/D-PLVF Li et al. (2019), IEMO/D Tomczyk & Kadzinski (2020), DPBEMO Huang et al. (2024), MOEA/D-PBO González et al. (2017), and NSGA-II/DW Thekumparampil et al. (2025). **Preference feedback.** All experiments use controlled simulated DMs whose ground-truth preferences are accessible to the algorithms only through pairwise queries. The default uses a fixed aspiration-point model, while stochastic-response settings assess robustness to feedback noise. All methods share matched DM parameters, initial populations, random seeds, and cumulative pairwise-query budgets. **Evaluation metrics and statistics.** For the final population $S = \{\boldsymbol{\phi}_i\}_{i=1}^N$ and ground-truth aspiration point $\mathbf{z}^\star$, we use two minimized preference-proximity metrics. The mean aspiration-point distance is $\bar{\epsilon}(S) = \frac{1}{N}\sum_{i=1}^N \|\boldsymbol{\phi}_i - \mathbf{z}^\star\|_2$. The Gaussian-averaged aspiration-point distance is $\mathcal{D}(S) = \mathbb{E}_{\mathbf{z}\sim\mathcal{N}(\mathbf{z}^\star,\sigma^2_{\text{metric}}I_m)}[\frac{1}{N}\sum_{i=1}^N \|\boldsymbol{\phi}_i - \mathbf{z}\|_2]$, with $\sigma_{\text{metric}} = 0.1$. Here, σ_{metric} is an evaluation parameter, not DM feedback noise. Gaussian reference samples are shared across algorithms within each problem configuration. Each configuration is evaluated over 20 independent runs paired by matched run index. Statistical comparisons use two-sided Wilcoxon signed-rank tests with Holm correction at $\alpha = 0.05$ (Wilcoxon, 1945; Holm, 1979) and Scott–Knott clustering (Mittas & Angelis, 2013). Appendix F provides complete metric and statistical protocols.

Standalone Evaluation of the PPL Module. Outside the EMO loop, we evaluate PPL on 44 personalized preference-learning tasks derived from 29,206 English-language comparisons among 20 large language models in Chatbot Arena Chiang et al. (2024), and on 240 controlled, linearly realizable complete-oracle scenarios. Baselines include active Bradley–Terry Bradley & Terry (1952), GPPL with BALD Chu & Ghahramani (2005); Houlsby et al. (2011), D-optimal Feature-BTL Saha & Rajkumar (2024); Thekumparampil et al. (2025), AMPLe Oh et al. (2025), GURO Bergström et al. (2024), DopeWolfe Thekumparampil et al. (2025), ASAP Mikhailiuk et al. (2020), and Multi-sort Maystre & Grossglauser (2017). On Chatbot Arena, all methods query only historically observed

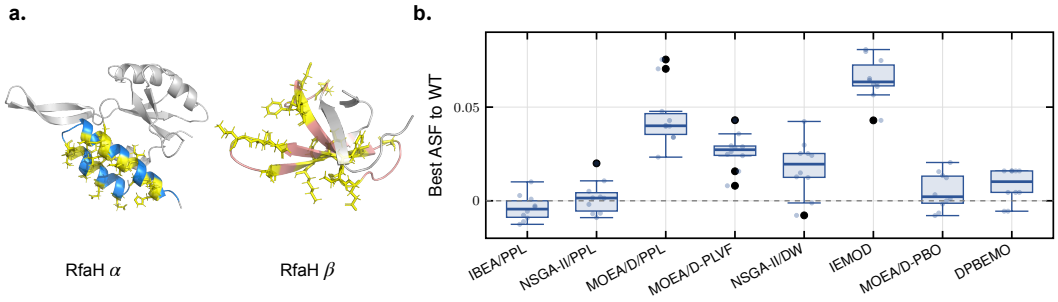

Figure 2: RfaH protein sequence design under a matched preference-query budget. **(a)** RfaH conformational states. The CTD (residues 119–154) is shown in marine (α state) or salmon (β state); yellow sticks mark the 36 redesigned positions. **(b)** Best ASF relative to wild type across 20 independent runs per algorithm. Dots show individual runs; boxes summarize their distributions.

pairs. We report normalized AULC for held-out pairwise accuracy, endpoint pairwise accuracy, and raw target-user query counts. Appendix F details replay protocols, baseline configurations, matched controls, and complete-oracle experiments.

6 Experimental Evaluation

We evaluate PPL through three analyses: end-to-end optimization under matched cumulative query budgets, component ablations examining feedback reuse and preference-guided selection, and standalone preference learning using logged preferences and a complete pairwise oracle to assess ranking quality and query efficiency independently of evolutionary search.

6.1 End-to-End Effectiveness under Matched Query Budgets

We evaluate whether embedding PPL steers EMO search toward preference-relevant regions under a fixed cumulative query budget. We consider 36 controlled benchmark configurations and an RfaH protein sequence-design task. In both settings, methods are compared using the preference-alignment metrics and statistical protocol described above. **Controlled multi-objective benchmarks.** Figure 1 summarizes the comparison. NSGA-II/PPL, MOEA/D/PPL, and IBEA/PPL were assigned to the best Scott–Knott group in 12, 3, and 17 configurations, respectively. Relative to MOEA/D-PBO, their median $\mathcal{D}(S)$ values decreased by 52.4%, 24.6%, and 50.6%, with Holm-corrected Wilcoxon win/tie/loss records of $30/4/2$, $21/6/9$, and $30/4/2$. Gains were most consistent for NSGA-II and IBEA, while MOEA/D showed greater variation across configurations. Results for the other preference-alignment metrics are reported in Appendix F.3. **RfaH protein sequence design.** In Figure 2, NSGA-II/PPL achieved a mean best-ASF of 0.0016, compared with 0.0082 for DPBEMO and 0.0178 for NSGA-II/DW. IBEA/PPL obtained the lowest mean (-0.0038) and median (-0.0044), whereas MOEA/D/PPL was less competitive on this task. Together, the benchmark and RfaH results show that the gains from PPL vary across host algorithms and problem settings.

6.2 Sources of the End-to-End Improvement

To identify the sources of the improvement, we ablate three components of NSGA-II/PPL on the benchmark subset with aligned paired seeds: -Trans removes transitive propagation, -Memory discards preference constraints after each consultation, and -PrefSel disables preference-guided population selection. Figure 3 shows the resulting degradation L_v, explicit-query multiplier, and preference-direction cosine similarity. We also examine how pairwise relations are resolved. **Transitive relation propagation.** Removing propagation reduces the transitivity-resolved fraction from 89.5% to 0%, with a corresponding increase in LP resolution. Explicit queries increase only slightly, and degradation remains small ($L_{\text{-Trans}} = 0.026$). Transitivity therefore primarily reduces redundant LP certification by reusing graph-implied relations, with little effect on preference accuracy or final population quality. **Persistent preference memory.** Discarding earlier constraints increases explicit queries by approximately one order of magnitude, reduces preference-direction cosine similarity

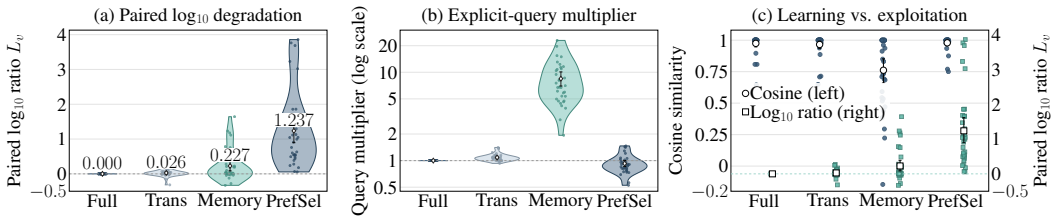

Figure 3: Component ablations of NSGA-II/PPL. **(a)** Paired degradation $L_v = \log_{10}(\bar{\epsilon}_v/\bar{\epsilon}_{\text{Full}})$; larger values are worse. **(b)** Explicit-query multiplier relative to the full method. **(c)** Preference-direction cosine similarity versus final degradation.

from approximately 0.97 to 0.76, and increases final preference-region error. Persistent memory thus amortizes feedback across consultations: reusable constraints prevent repeated queries and stabilize preference estimation as populations evolve. **Preference-guided selection.** Disabling preference-guided selection does not increase explicit queries, and preference estimation remains accurate, with $\cos(\hat{\mathbf{z}}, \mathbf{z}^\star) \approx 0.98$. Nevertheless, degradation rises to $L_{\text{-PREFSEL}} = 1.237$. Accurate preference estimation must therefore guide population selection to produce search pressure toward the target region. **Query reuse.** Across the tested feedback-noise levels, LP certification and transitive propagation together resolve more than 99.3% of recorded pair-resolution attempts without explicit DM queries. Further analyses of query scaling, feedback noise, alternative preference oracles, and regularization are reported in Appendix H and Appendix J.

6.3 Standalone Preference-Learning Validation

We evaluate PPL independently of the EMO loop using logged real-world preferences and a complete pairwise oracle. **Evaluation on logged preference data.** On Chatbot Arena personalized replay (2,200 fold runs), PPL achieves normalized AULC 0.7643 versus 0.7645 for volume-active AMPLe (Holm-adjusted $p = 0.855$), using 98.91% fewer target-user queries under their respective native pipelines (1,560 versus 142,560). Target-query-free capability initialization already yields pairwise accuracy 0.7645, and 55% of PPL runs issue no queries. Relative to the matched variant without certification-based stopping, PPL reduces queries by 98.49% with an AULC decrease of 0.0011, supporting query avoidance given the fixed initialization. Comparisons at common query caps of 0–4 are reported in Appendix I. **Controlled evaluation with complete preference feedback.** We additionally evaluate 240 linearly realizable scenarios with every pairwise outcome available on demand, removing logged-feedback coverage limitations. PPL achieves the highest normalized AULC (0.908968) and endpoint pairwise accuracy (0.942687) among the compared methods. Its AULC exceeds D-optimal Feature-BTL by 0.095888, and its endpoint accuracy exceeds active BT by 0.029099, while requiring 2.15% fewer queries than active BT. Together, these evaluations demonstrate substantial query savings with nearly unchanged ranking quality on logged data and strong ranking recovery under complete-oracle evaluation, providing evidence for the preference-inference and certification mechanisms independently of the host optimizer.

7 Conclusion

In this paper, we presented PPL, an interaction-efficient framework for persistent preference inference over evolving candidate populations. By representing pairwise feedback as halfspace constraints over a compatible aspiration-point region, PPL reuses historical feedback and combines LP-based certification with transitive inference to reduce unnecessary preference queries. The framework supports budgeted preference-aligned ranking and integrates with dominance-, decomposition-, and indicator-based evolutionary search. Under the stated realizability and uniform-splitting conditions, we establish an expected $\mathcal{O}(m \log N)$ upper bound on explicit queries during the hard phase of each consultation. Experiments on synthetic benchmarks, bio-sequence design tasks, and standalone preference-learning problems demonstrate improved preference-guided optimization and substantial query savings. Future work will further investigate the applicability and robustness of the framework under broader preference and optimization settings.

AI USE STATEMENT

We used generative AI tools to assist with editing the manuscript, organizing mathematical notation and experimental data, and checking consistency across the preference model, algorithm descriptions, code, and theoretical claims. AI assistance also supported revisions to the code and explanatory text. The authors reviewed and verified the AI-assisted content and take responsibility for the accuracy and integrity of the submitted work.

ETHICS STATEMENT

The proposed method is intended to support preference-guided multi-objective optimization. Its inferred rankings depend on the supplied feedback and modeling assumptions. The suitability of selected solutions for their intended applications requires task-specific evaluation beyond the inferred preferences.

REPRODUCIBILITY STATEMENT

The preference model and ranking procedure are described in Section 3. Section 4 states the assumptions and theoretical guarantees, with proofs provided in Appendix D. Details of the integrations with the host evolutionary algorithms are given in Appendix B, and the experimental settings are described in Appendix F.

REFERENCES

Sasan Amini, Antonio Candelieri, and Inneke Van Nieuwenhuyse. Decision maker preferences in surrogate-based multi-objective optimization: A survey. In *Learning and Intelligent Optimization*, volume 15744 of *Lecture Notes in Computer Science*, pp. 167–182. Springer Cham, 2026. doi: 10.1007/978-3-032-09156-7_12.

Raúl Astudillo and Peter I. Frazier. Multi-attribute bayesian optimization with interactive preference learning. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 4496–4507. PMLR, 2020. URL https://proceedings.mlr.press/v108/astudillo20a.html.

Herman Bergström, Emil Carlsson, Devdatt P. Dubhashi, and Fredrik D. Johansson. Active preference learning for ordering items in- and out-of-sample. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/8443219a991f068c34d9491ad68ffa94-Abstract-Conference.html.

Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952. ISSN 00063444, 14643510.

Jürgen Branke, Salvatore Greco, Roman Slowinski, and Piotr Zielniewicz. Interactive evolutionary multiobjective optimization using robust ordinal regression. In *EMO'09: Proc. of the 5th International Conference Evolutionary Multi-Criterion Optimization*, volume 5467 of *Lecture Notes in Computer Science*, pp. 554–568. Springer, 2009. doi: 10.1007/978-3-642-01020-0_43. URL https://doi.org/10.1007/978-3-642-01020-0_43.

Jürgen Branke, Salvatore Greco, Roman Slowinski, and Piotr Zielniewicz. Learning value functions in interactive evolutionary multiobjective optimization. *IEEE Trans. Evol. Comput.*, 19(1):88–102, 2015. doi: 10.1109/TEVC.2014.2303783. URL https://doi.org/10.1109/TEVC.2014.2303783.

Jürgen Branke, Salvatore Corrente, Salvatore Greco, and Walter J. Gutjahr. Efficient pairwise preference elicitation allowing for indifference. *Comput. Oper. Res.*, 88:175–186, 2017. doi: 10.1016/J.COR.2017.06.020. URL https://doi.org/10.1016/j.cor.2017.06.020.

Björn M. Burmann, Stefan H. Knauer, Anastasia Sevostyanova, Kristian Schweimer, Rachel A. Mooney, Robert Landick, Irina Artsimovitch, and Paul Rösch. An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell*, 150(2):291–303, 2012. doi: 10.1016/j.cell.2012.05.042.

Lisha Chen, A. F. M. Saif, Yanning Shen, and Tianyi Chen. FERERO: A flexible framework for preference-guided multi-objective learning. In *NeurIPS'24: Proc. of 38th Annual Conference on Neural Information Processing Systems*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/21a33eba893ca3890e395651b38810df-Abstract-Conference.html.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, Michael I. Jordan, Joseph E. Gonzalez, and Ion Stoica. Chatbot arena: An open platform for evaluating llms by human preference. In Ruslan Salakhutdinov, Zico Kolter, Katherine A. Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 8359–8388. PMLR / OpenReview.net, 2024. URL https://proceedings.mlr.press/v235/chiang24b.html.

Wei Chu and Zoubin Ghahramani. Preference learning with gaussian processes. In Luc De Raedt and Stefan Wrobel (eds.), *Machine Learning, Proceedings of the Twenty-Second International Conference (ICML 2005), Bonn, Germany, August 7-11, 2005*, ACM International Conference Proceeding Series, pp. 137–144. ACM, 2005. doi: 10.1145/1102351.1102369. URL https://doi.org/10.1145/1102351.1102369.

Carlos A. Coello Coello. Handling preferences in evolutionary multiobjective optimization: a survey. In Ali M. S. Zalzala (ed.), *CEC'00: Proc. of the 2000 Congress on Evolutionary Computation*, pp. 30–37. IEEE, 2000. doi: 10.1109/CEC.2000.870272. URL https://doi.org/10.1109/CEC.2000.870272.

Salvatore Corrente, Salvatore Greco, Milosz Kadzinski, and Roman Slowinski. Robust ordinal regression in preference learning and ranking. *Mach. Learn.*, 93(2-3):381–422, 2013. doi: 10.1007/S10994-013-5365-4. URL https://doi.org/10.1007/s10994-013-5365-4.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Alexis Courbet, Rob J. de Haas, Neville Bethel, Philip J. Y. Leung, Timothy F. Huddy, Samuel Pellock, Doug Tischer, Fred Chan, Brian Koepnick, Hang Nguyen, Alex Kang, Banumathi Sankaran, Asim K. Bera, Neil P. King, and David Baker. Robust deep learning–based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022. doi: 10.1126/science.add2187.

K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable multi-objective optimization test problems. In *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, volume 1, pp. 825–830 vol.1, 2002a. doi: 10.1109/CEC.2002.1007032.

Kalyanmoy Deb, Samir Agrawal, Amrit Pratap, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.*, 6(2):182–197, 2002b. doi: 10.1109/4235.996017. URL https://doi.org/10.1109/4235.996017.

Kerr Ding, Michael Chin, Yunlong Zhao, Wei Huang, Binh Khanh Mai, Huanan Wang, Peng Liu, Yang Yang, and Yunan Luo. Machine learning-guided co-optimization of fitness and diversity facilitates combinatorial library design in enzyme engineering. *Nat. Commun.*, 15:6392, 2024. doi: 10.1038/s41467-024-50698-y. URL https://doi.org/10.1038/s41467-024-50698-y.

Matthias Ehrgott. *Multicriteria Optimization (2. ed.)*. Springer, 2005. ISBN 978-3-540-21398-7. doi: 10.1007/3-540-27659-9. URL https://doi.org/10.1007/3-540-27659-9.

Matthias Ehrgott, Murat K"oksalan, Miłosz Kadziński, and Kalyanmoy Deb. Fifty years of multiobjective optimization and decision-making: From mathematical programming to evolutionary computation. *European Journal of Operational Research*, 330(1):1–25, 2026. doi: 10.1016/j.ejor.2025.06.012.

Javier González, Zhenwen Dai, Andreas C. Damianou, and Neil D. Lawrence. Preferential bayesian optimization. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, Proceedings of Machine Learning Research, pp. 1282–1291. PMLR, 2017. URL http://proceedings.mlr.press/v70/gonzalez17a.html.

Reinhard Heckel, Nihar B. Shah, Kannan Ramchandran, and Martin J. Wainwright. Active ranking from pairwise comparisons and when parametric assumptions do not help. *The Annals of Statistics*, 2016.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

Lu Hong and Tanja Kortemme. An integrative approach to protein sequence design through multiobjective optimization. *PLOS Computational Biology*, 20(7):e1011953, 2024. doi: 10.1371/journal.pcbi.1011953.

Neil Houlsby, Ferenc Huszar, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *CoRR*, abs/1112.5745, 2011. URL http://arxiv.org/abs/1112.5745.

Tianmeng Hu, Biao Luo, and Ke Li. Human-in-the-loop policy optimization for preference-based multi-objective reinforcement learning, 2026. URL https://openreview.net/forum?id=PRHNKeaZpP.

Tian Huang, Shengbo Wang, and Ke Li. Direct preference-based evolutionary multi-objective optimization with dueling bandits. In *NeurIPS'24: Proc. of 38th Annual Conference on Neural Information Processing Systems*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/dd0ef9308bd10c964bd14c0000438460-Abstract-Conference.html.

Simon Huband, Philip Hingston, Luigi Barone, and Lyndon While. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation*, 10 (5):477–506, 2006.

Joshua Hang Sai Ip, Ankush Chakrabarty, Ali Mesbah, and Diego Romeres. User preference meets pareto-optimality in multi-objective bayesian optimization. In Toby Walsh, Julie Shah, and Zico Kolter (eds.), *Thirty-Ninth AAAI Conference on Artificial Intelligence, Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence, Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2025, Philadelphia, PA, USA, February 25 - March 4, 2025*, pp. 20246–20254. AAAI Press, 2025. doi: 10.1609/AAAI.V39I19.34230. URL https://doi.org/10.1609/aaai.v39i19.34230.

Hisao Ishibuchi, Yu Setoguchi, Hiroyuki Masuda, and Yusuke Nojima. Performance of decomposition-based many-objective algorithms strongly depends on pareto front shapes. *IEEE Transactions on Evolutionary Computation*, 21(2):169–190, 2016.

Kevin G. Jamieson and Robert D. Nowak. Active ranking using pairwise comparisons. In *Advances in Neural Information Processing Systems 24 (NeurIPS 2011)*, pp. 2240–2248, 2011.

Ke Li, Renzhi Chen, Dragan A. Savic, and Xin Yao. Interactive decomposition multiobjective optimization via progressively learned value functions. *IEEE Trans. Fuzzy Syst.*, 27(5):849–860, 2019. doi: 10.1109/TFUZZ.2018.2880700. URL https://doi.org/10.1109/TFUZZ.2018.2880700.

Ke Li, Minhui Liao, Kalyanmoy Deb, Geyong Min, and Xin Yao. Does preference always help? A holistic study on preference-based evolutionary multiobjective optimization using reference points. *IEEE Trans. Evol. Comput.*, 24(6):1078–1096, 2020. doi: 10.1109/TEVC.2020.2987559. URL https://doi.org/10.1109/TEVC.2020.2987559.

Ke Li, Guiyu Lai, and Xin Yao. Interactive evolutionary multiobjective optimization via learning to rank. *IEEE Trans. Evol. Comput.*, 27(4):749–763, 2023. doi: 10.1109/TEVC.2023.3234269. URL https://doi.org/10.1109/TEVC.2023.3234269.

Zhiyuan (Jerry) Lin, Raul Astudillo, Peter I. Frazier, and Eytan Bakshy. Preference exploration for efficient bayesian optimization with multiple outcomes. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera (eds.), *International Conference on Artificial Intelligence and Statistics, AISTATS 2022, 28-30 March 2022, Virtual Event*, Proceedings of Machine Learning Research, pp. 4235–4258. PMLR, 2022. URL https://proceedings.mlr.press/v151/jerry-lin22a.html.

Patrick Löffler, Samuel Schmitz, Enrico Hupfeld, Reinhard Sterner, and Rainer Merkl. Rosetta:MSF: a modular framework for multi-state computational protein design. *PLOS Computational Biology*, 13(6):e1005600, 2017. doi: 10.1371/journal.pcbi.1005600. URL https://doi.org/10.1371/journal.pcbi.1005600.

Andre KY Low, Flore Mekki-Berrada, Abhishek Gupta, Aleksandr Ostudin, Jiaxun Xie, Eleonore Vissol-Gaudin, Yee-Fun Lim, Qianxiao Li, Yew Soon Ong, Saif A Khan, et al. Evolution-guided bayesian optimization for constrained multi-objective optimization in self-driving labs. *npj Computational Materials*, 10(1):104, 2024.

Lucas Maystre and Matthias Grossglauser. Just sort it! A simple and effective approach to active preference learning. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pp. 2344–2353. PMLR, 2017. URL http://proceedings.mlr.press/v70/maystre17a.html.

Kaisa Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston, 1999. doi: 10.1007/978-1-4615-5563-6.

Aliaksei Mikhailiuk, Clifford Wilmot, María Pérez-Ortiz, Dingcheng Yue, and Rafal K. Mantiuk. Active sampling for pairwise comparisons via approximate message passing and information gain maximization. In *25th International Conference on Pattern Recognition, ICPR 2020, Virtual Event / Milan, Italy, January 10-15, 2021*, pp. 2559–2566. IEEE, 2020. URL https://doi.org/10.1109/ICPR48806.2021.9412676.

Nikolaos Mittas and Lefteris Angelis. Ranking and clustering software cost estimation models through a multiple comparisons algorithm. *IEEE Trans. Software Eng.*, 39(4):537–551, 2013. doi: 10.1109/TSE.2012.45. URL https://doi.org/10.1109/TSE.2012.45.

Stephen Mussmann and Percy Liang. Generalized binary search for split-neighborly problems. In *AISTATS'18: Proc. of 2018 International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, pp. 1561–1569. PMLR, 2018. URL http://proceedings.mlr.press/v84/mussmann18a.html.

Kaname Narukawa, Yu Setoguchi, Yuki Tanigaki, Markus Olhofer, Bernhard Sendhoff, and Hisao Ishibuchi. Preference representation using gaussian functions on a hyperplane in evolutionary multi-objective optimization. *Soft Comput.*, 20(7):2733–2757, 2016. doi: 10.1007/S00500-015-1674-9. URL https://doi.org/10.1007/s00500-015-1674-9.

Minhyeon Oh, Seungjoon Lee, and Jungseul Ok. Comparison-based active preference learning for multi-dimensional personalization. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2025, Vienna, Austria, July 27 - August 1, 2025*, pp. 33145–33166. Association for Computational Linguistics, 2025. doi: 10.18653/V1/2025.ACL-LONG.1590. URL https://doi.org/10.18653/v1/2025.acl-long.1590.

Ryota Ozaki, Kazuki Ishikawa, Youhei Kanzaki, Shion Takeno, Ichiro Takeuchi, and Masayuki Karasuyama. Multi-objective bayesian optimization with active preference learning. In Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (eds.), *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, pp. 14490–14498. AAAI Press, 2024. doi: 10.1609/AAAI.V38I13.29364. URL https://doi.org/10.1609/aaai.v38i13.29364.

Ji Won Park, Natasa Tagasovska, Michael Maser, Stephen Ra, and Kyunghyun Cho. Botied: Multi-objective bayesian optimization with tied multivariate ranks. In Ruslan Salakhutdinov, Zico Kolter, Katherine A. Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, Proceedings of Machine Learning Research, pp. 39813–39833. PMLR / OpenReview.net, 2024. URL https://proceedings.mlr.press/v235/park24k.html.

Bo Peng, Ye Wei, Yu Qin, Jiabao Dai, Yue Li, Aobo Liu, Yun Tian, Liuliu Han, Yufeng Zheng, and Peng Wen. Machine learning-enabled constrained multi-objective design of architected materials. *Nat. Commun.*, 14:6630, 2023. doi: 10.1038/s41467-023-42415-y. URL https://doi.org/10.1038/s41467-023-42415-y.

Mathieu Reymond, Eugenio Bargiacchi, and Ann Nowé. Pareto conditioned networks. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '22, pp. 1110–1118, Richland, SC, 2022. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450392136.

Aadirupa Saha and Arun Rajkumar. A graph theoretic approach for preference learning with feature information. In Negar Kiyavash and Joris M. Mooij (eds.), *Uncertainty in Artificial Intelligence, 15-19 July 2024, Universitat Pompeu Fabra, Barcelona, Spain*, volume 244 of *Proceedings of Machine Learning Research*, pp. 3138–3158. PMLR, 2024. URL https://proceedings.mlr.press/v244/saha24a.html.

L. Ben Said, S. Bechikh, and K. Ghedira. The r-dominance: A new dominance relation for interactive evolutionary multicriteria decision making. *IEEE Transactions on Evolutionary Computation*, 14 (3):355–369, 2010. doi: 10.1109/TEVC.2010.2041060. URL https://doi.org/10.1109/TEVC.2010.2041060.

Kiran Koshy Thekumparampil, Gaurush Hiranandani, Kousha Kalantari, Shoham Sabach, and Branislav Kveton. Comparing few to rank many: Active human preference learning using randomized frank-Wolfe method. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 59355–59376. PMLR, 2025.

Lothar Thiele, Kaisa Miettinen, Pekka J. Korhonen, and Julián Molina Luque. A preference-based evolutionary algorithm for multi-objective optimization. *Evol. Comput.*, 17(3):411–436, 2009. doi: 10.1162/EVCO.2009.17.3.411. URL https://doi.org/10.1162/evco.2009.17.3.411.

Louis L. Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273–286, 1927. doi: 10.1037/h0070288.

Michal K. Tomczyk and Milosz Kadzinski. Decomposition-based interactive evolutionary algorithm for multiple objective optimization. *IEEE Trans. Evol. Comput.*, 24(2):320–334, 2020. doi: 10.1109/TEVC.2019.2915767. URL https://doi.org/10.1109/TEVC.2019.2915767.

Michal K. Tomczyk and Milosz Kadzinski. Efficient preference learning algorithm for interactive evolutionary multi-objective optimization. *Swarm Evol. Comput.*, 100:102254, 2026. doi: 10.1016/J.SWEVO.2025.102254. URL https://doi.org/10.1016/j.swevo.2025.102254.

Zhenkun Wang, Yew-Soon Ong, and Hisao Ishibuchi. On scalable multiobjective test problems with hardly dominated boundaries. *IEEE Trans. Evol. Comput.*, 23(2):217–231, 2019. doi: 10.1109/TEVC.2018.2844286. URL https://doi.org/10.1109/TEVC.2018.2844286.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

Rui Yang, Xiaoman Pan, Feng Luo, Shuang Qiu, Han Zhong, Dong Yu, and Jianshu Chen. Rewards-in-Context: Multi-objective alignment of foundation models with dynamic preference adjustment. In *Proceedings of the 41st International Conference on Machine Learning, ICML 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 56276–56297. PMLR, 2024. URL https://proceedings.mlr.press/v235/yang24q.html.

Yucheng Yang, Tianyi Zhou, Mykola Pechenizkiy, and Meng Fang. Preference controllable reinforcement learning with advanced multi-objective optimization. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 71612–71647. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/yang25ax.html.

Jun Yi, Junren Bai, Haibo He, Jun Peng, and Dedong Tang. ar-MOEA: A novel preference-based dominance relation for evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 23(5):788–802, 2019. doi: 10.1109/TEVC.2018.2884133. URL https://doi.org/10.1109/TEVC.2018.2884133.

Qingfu Zhang and Hui Li. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. Evol. Comput.*, 11(6):712–731, 2007. doi: 10.1109/TEVC.2007.892759. URL https://doi.org/10.1109/TEVC.2007.892759.

Eckart Zitzler and Simon Künzli. Indicator-based selection in multiobjective search. In Xin Yao, Edmund K. Burke, José Antonio Lozano, Jim Smith, Juan Julián Merelo Guervós, John A. Bullinaria, Jonathan E. Rowe, Peter Tiño, Ata Kabán, and Hans-Paul Schwefel (eds.), *Parallel Problem Solving from Nature - PPSN VIII, 8th International Conference, Birmingham, UK, September 18-22, 2004, Proceedings*, volume 3242 of *Lecture Notes in Computer Science*, pp. 832–842. Springer, 2004. doi: 10.1007/978-3-540-30217-9_84. URL https://doi.org/10.1007/978-3-540-30217-9_84.

Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evol. Comput.*, 8(2):173–195, 2000. doi: 10.1162/106365600568202. URL https://doi.org/10.1162/106365600568202.

Philipp Konrad Zuber, Kristian Schweimer, Paul Rösch, Irina Artsimovitch, and Stefan H. Knauer. Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. *Nature Communications*, 10:702, 2019. doi: 10.1038/s41467-019-08567-6.

APPENDIX

This appendix contains supplementary material supporting the main paper. It provides additional theoretical details and complete proofs omitted from the main text, together with extended empirical results for the PBEMO instantiation of PARC. We also summarize the experimental settings used throughout the study and discuss limitations and broader implications of the proposed approach.

A RELATED WORK

The integration of DM preferences into multi-objective optimization has evolved from classical *a priori* and *a posteriori* paradigms to interactive learning frameworks that couple preference elicitation with search. This section reviews prior work along two axes most relevant to PPL: first, the optimization framework in which preference information is retained and reused; and second, the query mechanism used to elicit informative comparisons. We position PPL within multi-objective optimization, with particular emphasis on PBEMO. We then discuss active ranking and consistency-based preference elicitation, which are closest to the certification mechanism used in PPL.

A.1 PREFERENCE-BASED MULTI-OBJECTIVE OPTIMIZATION

Preference information can enter multi-objective optimization through several search frameworks, ranging from classical scalarization to Bayesian optimization, preference-conditioned learning, and evolutionary search.

Classical MCDM and mathematical programming. Classical multi-criteria decision making (MCDM) Miettinen (1999) and mathematical-programming Ehrgott (2005) approaches encode preferences by reformulating the vector-valued problem into one or more preference-parameterized subproblems, such as weighted, ϵ-constraint, goal-programming, value-function, reference-point, or achievement-scalarizing formulations Miettinen (1999); Ehrgott (2005). These methods provide rigorous mechanisms for preference articulation and remain foundational in multi-objective optimization. However, preference information is typically consumed through the current scalarized or reformulated subproblem. Historical judgments may influence updated aspiration levels, reference points, or navigation trajectories, but they are typically not retained as explicit reusable comparison constraints across interaction rounds. Moreover, diverse trade-off alternatives often require repeated preference-parameter changes and repeated solves, which can become costly in large-scale or high-dimensional settings Ehrgott et al. (2026).

Preference-aware Bayesian optimization. For computationally expensive black-box problems, preference-aware multi-objective Bayesian optimization (MOBO) typically uses objective surrogates together with decision-maker preference information to steer the search toward preferred regions of the Pareto front Amini et al. (2026). Astudillo & Frazier (2020) models the decision maker's unknown utility from pairwise comparisons over attribute vectors within a multi-attribute Bayesian optimization framework. Lin et al. (2022) further alternates interactive pairwise preference elicitation with Bayesian optimization under a learned utility-outcome model, while Ozaki et al. (2024) incorporates uncertainty in both the objective functions and a Bayesian preference model into an active-preference acquisition rule. In these methods, earlier preference feedback affects later acquisition decisions through posterior updates of the learned utility or preference model; this statistical reuse is different from preserving the elicited pairwise comparisons themselves as explicit, persistent constraints. Moreover, these approaches are primarily designed for sequential, evaluation-budgeted BO, whereas evolutionary search operates over a changing population of candidate solutions. Classical MOBO may allocate substantial effort to approximating the whole Pareto front, and utility-driven preference MOBO can yield dominated solutions when non-dominance is not explicitly enforced Ip et al. (2025). Pareto-aware acquisition designs such as BOtied show that dominance structure can be built into MOBO acquisitions, but they do not by themselves provide a population-level memory of pairwise preference constraints Park et al. (2024).

Preference-conditioned learning and policy optimization. A broader learning-based line recasts preference-guided multi-objective optimization as preference-conditioned model or policy learning.

For example, Chen et al. (2024) formulates preference-guided multi-objective learning as constrained vector optimization, representing relative preferences by cone-induced partial orders and absolute preferences by linear objective constraints. Preference-aware multi-objective reinforcement learning either conditions policies directly on preferences Yang et al. (2025) or learns a latent preference model from interactive feedback to guide policy optimization Hu et al. (2026). Pareto-conditioned networks (PCN) learn a single network conditioned on desired return vectors, enabling a posteriori trade-off selection over a family of non-dominated policies Reymond et al. (2022). These approaches are well suited to differentiable learning or sequential decision problems where the output is a preference-guided model, a preferred policy, or a reusable family of Pareto-efficient policies. They are less directly aligned with our setting, where the optimizer maintains an explicitly evolving population and repeatedly consults the DM over transient candidate sets. When historical preference information is retained in these methods, it is typically absorbed into a reward, utility, or conditioning model rather than maintained as explicit pairwise constraints that can certify later comparisons.

Preference-based evolutionary multi-objective optimization. PBEMO is closest to our setting because the search is conducted over an evolving population of candidate solutions Coello (2000); Ehrgott et al. (2026). Population-based methods maintain multiple trade-off candidates simultaneously, support repeated consultation during search, and allow preference information to shape selection as the candidate distribution changes across generations Thiele et al. (2009); Li et al. (2019; 2023); Huang et al. (2024). Existing PBEMO methods can be broadly grouped by how they use elicited preferences.

- *Model-mediated strategies* fit an explicit preference model from elicited information, such as a value function, ordinal-regression model, Gaussian preference model, or learning-to-rank surrogate Branke et al. (2015); Narukawa et al. (2016); Li et al. (2019; 2023). The learned model is then used to rank candidates or induce preference-guided selection pressure in later generations. This enables historical feedback to be reused through model updates, but the resulting guidance can be sensitive to model misspecification and evolutionary distribution shift: preferences are learned from generation-specific candidate sets whose distribution is continually reshaped by selection, crossover, and mutation.
- *Direct-comparison strategies* inject preference information more directly into the search, for example through reference-point-based dominance relations, preference-biased indicators, decomposition or reference-vector adaptation, or online comparison-feedback mechanisms Said et al. (2010); Yi et al. (2019); Li et al. (2020); Huang et al. (2024). These methods are less tied to a single global utility form and can be more robust when the DM's preferences are difficult to represent by one parametric score. However, the retained information is usually a current dominance relation, reference signal, empirical score, confidence statistic, or population-local ordering. Since the population itself changes across generations, earlier pairwise judgments are rarely preserved as explicit constraints that can be reused to certify relations among later candidates.

Thus, within PBEMO, the limitation is not the absence of pairwise elicitation or population-based search. Rather, historical comparisons are usually either absorbed into a fitted preference model or reduced to local ranking and selection decisions. In contrast, `PPL` maintains explicitly queried pairwise comparisons as durable constraints over a changing candidate population. This persistent constraint representation enables later comparisons to be certified or propagated without treating each consultation round as an isolated preference-learning problem.

A.2 Active ranking and preference elicitation

Active ranking studies how to recover an underlying ordering of items by adaptively selecting informative pairwise queries, thereby reducing sample complexity Heckel et al. (2016). Query-selection strategies can be broadly divided by how they characterize uncertainty in preference relations Thekumparampil et al. (2025).

Probabilistic uncertainty-based methods. One family adopts a probabilistic perspective, typically assuming a stochastic choice model such as Bradley-Terry Bradley & Terry (1952), Thurstone Thurstone (1927), or Gaussian-process preference learning Chu & Ghahramani (2005). Queries are

selected to maximize expected information gain, reduce posterior entropy, or otherwise improve uncertainty reduction under the assumed probabilistic model. These methods can be effective when the model is well specified, but their efficiency is tightly coupled to the assumed stochastic choice model and may degrade under model misspecification or the rapid distribution shifts induced by evolutionary search.

Consistency-based and geometric methods. A second family, closer to PPL's certification mechanism, takes a geometric and consistency-based view Jamieson & Nowak (2011). Rather than maintaining a posterior over utilities, these methods treat preference observations as constraints in a feasible parameter space Corrente et al. (2013). A query is informative only when its outcome cannot be logically inferred from the current constraint set. This perspective is naturally compatible with certification: if one hypothetical answer is inconsistent with the feasible region, the opposite answer is already implied and need not be queried.

Most active ranking methods, however, are designed for static item sets, where the goal is to identify a fixed total order. PBEMO violates this static-item assumption: candidates are transient, the population distribution changes across generations, and useful preference information must remain reusable even after earlier candidates disappear. Applying consistency-based active ranking in this setting therefore requires more than a static query strategy; it requires a persistent preference state that can be carried across consultations and applied to newly generated candidates.

Our perspective. PPL adapts consistency-based active ranking to the non-stationary candidate sets produced by evolutionary search. Rather than pursuing a fixed global ranking over a static item set, PPL maintains historical queried comparisons as reusable constraints, uses LP feasibility checks to certify relations implied by those constraints, and propagates resolved relations along an active sorted chain. This converts active ranking from a static elicitation procedure into a reusable preference-state mechanism for PBEMO. Crucially, our framework separates certification from estimation. The LP-based certification step acts as a hard model-consistency test, and the corresponding query-reduction guarantee is stated under a realizable regime in which the retained queried comparisons admit at least one compatible preference model. This assumption is necessary for certified pairwise inferences to have a precise logical meaning. In contrast, the aspiration-point estimate used for downstream environmental selection is obtained from a soft-margin formulation. Because slack variables penalize, rather than forbid, violations of the unit-margin inequalities, noisy or weakly separated queried comparisons can still be incorporated into a stable preference signal. Thus, the method does not claim to certify preferences under arbitrary human inconsistency; rather, it provides hard certificates under model-consistency while leveraging a soft-margin estimate to improve the robustness of the downstream selection guidance.

B EMO-Specific Integration Details

This section provides the implementation details omitted from Section 3.3. The persistent preference-inference mechanism is shared by all three host EMO algorithms. At consultation c, the inference module maintains the explicit comparison history $\mathcal{C}$ and produces a current aspiration-point estimate $\widehat{\mathbf{z}}_c$. When $\mathcal{Z}_\epsilon(\mathcal{C})$ is nonempty, $\widehat{\mathbf{z}}_c = \widehat{\mathbf{z}}_{\mathrm{H}}$ is the deterministically selected Chebyshev center of the compatible region. When $\mathcal{Z}_\epsilon(\mathcal{C})$ is empty, $\widehat{\mathbf{z}}_c = \widehat{\mathbf{z}}_{\mathrm{S}}$ is obtained from the soft inference problem. Only explicit DM responses are appended to $\mathcal{C}$; relations inferred from either point estimate are used only to guide the host EMO algorithm and are never treated as additional preference evidence.

B.1 NSGA-II integration

For NSGA-II, PPL is integrated into environmental selection. At generation t, the parent and offspring populations form the indexed pool U_t. Its distinct objective vectors constitute the consultation set Φ_t, and κ_t maps each individual to its representative in Φ_t. Individuals with identical objective vectors share one consultation candidate but remain distinct in environmental selection. At an interaction generation, PPL-Rank processes Φ_t using the incoming history $\mathcal{C}$ and the explicit-query budget B_t, and returns a ranking $\widehat{\sigma}_t$, the updated history, and a termination mode. Only newly queried comparisons are appended to $\mathcal{C}$. After the consultation, the soft estimate $\widehat{\mathbf{z}}_t = \texttt{SoftEstimate}(\mathcal{C})$ is

computed regardless of the termination mode and retained between consultations. The ranking is used only for the consultation set on which it was constructed and is reset at noninteraction generations.

Environmental selection retains nondominated sorting and first accepts complete fronts that fit within the next-generation population size N_{pop}. If the next front exceeds the remaining capacity, a fresh ranking $\widehat{\sigma}_t$, when available, orders its individuals by $\mathrm{pos}_{\widehat{\sigma}_t}(\kappa_t(\mathbf{x}))$. Individuals sharing a representative are resolved by a fixed deterministic order. When no fresh ranking is available but an aspiration-point estimate has been retained, individuals are prioritized by

$$s_t(\phi) = -\|\phi - \widehat{\mathbf{z}}_t\|_2^2,$$

with larger scores preferred, crowding distance used to resolve score ties, and a fixed deterministic order used for any remaining ties. Here, $\phi = \mathbf{F}(\mathbf{x})$ and the aspiration point are both expressed in the original objective coordinates. Before the first consultation, selection follows standard NSGA-II.

The score and the ranking induced by it are used only for selection and are not appended to $\mathcal{C}$. Mating selection and variation retain the standard NSGA-II mechanisms.

B.2 MOEA/D INTEGRATION

For MOEA/D, `PPL` uses the accumulated explicit feedback to update an aspiration-point estimate that guides the adaptation of decomposition directions. The returned ranking $\widehat{\sigma}_t$ is not used directly for selection. Only the comparisons explicitly queried during the consultation update the persistent history $\mathcal{C}$; certified, propagated, and fallback relations are not added to it.

Let $\Lambda^0 = \{\boldsymbol{\lambda}_i^0\}_{i=1}^{N_{\mathrm{pop}}}$ denote the baseline decomposition weight vectors, and let $\Lambda = \{\boldsymbol{\lambda}_i\}_{i=1}^{N_{\mathrm{pop}}}$ denote the current weight vectors. The neighborhood collection $\mathcal{B} = \{\mathcal{B}_i\}_{i=1}^{N_{\mathrm{pop}}}$ is built from the current weights. The ideal point $\mathbf{z}^{\mathrm{ideal}}$ records the componentwise best objective values observed so far and is used by the MOEA/D scalarizing functions. It is distinct from the aspiration-point estimate and may change during evolution.

Before the first consultation, MOEA/D uses Λ^0 and the corresponding neighborhoods. At an interaction generation, the current population supplies the indexed consultation pool $U_t = R_t$. The routine `DistinctObjectives` constructs Φ_t from this pool, and `PPL-Rank` processes Φ_t with budget B_t. The soft estimate $\widehat{\mathbf{z}}_t$ is then computed from the updated history, regardless of whether the ranking routine terminates in hard or fallback mode.

The procedure `AdaptDecomposition` takes the baseline weights, the current population, the soft estimate, and the current ideal point as inputs. It adapts the decomposition directions toward the preferred region, rebuilds the neighborhoods, and reassociates the current solutions with the adapted subproblems while retaining N_{pop} population entries. Between consultations, the aspiration-point estimate and decomposition structure are retained.

Evolution then follows the MOEA/D variation and neighborhood-replacement mechanisms using the current weights, neighborhoods, and ideal point. Neither the adapted weights nor relations implied by the decomposition are added to $\mathcal{C}$. All objective vectors retained for persistent preference inference remain in the original coordinates. The exact weight-vector adaptation rule and the MOEA/D-specific experimental parameters are specified in Appendix F.

B.3 IBEA INTEGRATION

For IBEA, `PPL` incorporates the retained soft aspiration-point estimate into indicator-based fitness assignment. The returned ranking $\widehat{\sigma}_t$ is not used directly for environmental selection. The persistent history $\mathcal{C}$ contains only explicitly queried comparisons, which are used by `SoftEstimate` after every consultation, regardless of the termination mode.

Let $\mathcal{F}(S)$ denote the fitness values associated with the individuals in an indexed population S. Before the first consultation, fitness is computed using the standard normalized ϵ-indicator. Once a soft estimate is available, a reference-distance correction is applied to mutually nondominated pairs to favor candidates closer to $\widehat{\mathbf{z}}_t$. The distances to the aspiration point and the persistent comparison model use the original objective coordinates; normalization within the underlying IBEA indicator does not change the coordinates of the observations stored in $\mathcal{C}$.

Algorithm 3 MOEA/D with PPL-guided preference feedback

Require: Population size N_{pop}, generations T, interaction generations $\mathcal{T}_{\text{int}}$, budgets $\{B_t\}_{t\in\mathcal{T}_{\text{int}}}$, baseline weights Λ^0, oracle Query
Ensure: Final population R_T

Initialize and evaluate R_0 with $|R_0| = N_{\text{pop}}$
$\mathcal{C} \leftarrow \emptyset$; $\widehat{\mathbf{z}}_{-1} \leftarrow \varnothing$
$\Lambda \leftarrow \Lambda^0$; build neighborhoods $\mathcal{B}$ from Λ
$\mathbf{z}^{\text{ideal}} \leftarrow \texttt{InitializeIdealPoint}(R_0)$
for $t = 0$ to $T-1$ **do**
  $\widehat{\mathbf{z}}_t \leftarrow \widehat{\mathbf{z}}_{t-1}$
  if $t \in \mathcal{T}_{\text{int}}$ **then**
    $U_t \leftarrow R_t$; $(\Phi_t, \kappa_t) \leftarrow \texttt{DistinctObjectives}(U_t)$
    $(\widehat{\sigma}_t, \mathcal{C}, M_t) \leftarrow \texttt{PPL-Rank}(\Phi_t, \mathcal{C}, B_t, \text{Query})$
    $\widehat{\mathbf{z}}_t \leftarrow \texttt{SoftEstimate}(\mathcal{C})$
    $(\Lambda, \mathcal{B}, R_t) \leftarrow \texttt{AdaptDecomposition}(\Lambda^0, R_t, \widehat{\mathbf{z}}_t, \mathbf{z}^{\text{ideal}})$
  end if
  for $i = 1$ to N_{pop} **do**
    $\mathbf{x} \leftarrow \texttt{VariationFromNeighborhood}(R_t, \mathcal{B}_i)$
    $\texttt{Evaluate}(\mathbf{x})$
    $\mathbf{z}^{\text{ideal}} \leftarrow \min\big(\mathbf{z}^{\text{ideal}}, \mathbf{F}(\mathbf{x})\big)$ ▷ componentwise minimum
    $R_t \leftarrow \texttt{UpdateSubproblems}(R_t, \mathbf{x}, \Lambda, \mathcal{B}_i, \mathbf{z}^{\text{ideal}})$
  end for
  $R_{t+1} \leftarrow R_t$
end for
return R_T

Consultation takes place before mating. At an interaction generation, the current population supplies the indexed consultation pool $U_t = R_t$, and its distinct objective vectors form Φ_t. The routine PPL-Rank processes Φ_t with budget B_t, after which $\widehat{\mathbf{z}}_t$ is computed from the updated history. The fitness values $\mathcal{F}(R_t)$ are then recomputed using this estimate, so the updated preference information affects fitness-based tournament mating in the same generation.

Let O_t denote the offspring population and $V_t = R_t \uplus O_t$ the combined parent–offspring pool used for environmental selection, where $\uplus$ preserves individual identities and multiplicities. After offspring evaluation, IBEA-Fitness assigns fitness to V_t using the retained estimate, if available. Environmental selection repeatedly removes the individual with the worst fitness and updates the remaining fitness values until N_{pop} individuals remain. The surviving individuals and their updated fitness values are retained for the next generation. Between consultations, the aspiration-point estimate remains unchanged.

The tournament-mating, variation, and iterative environmental-selection mechanisms retain their IBEA structure; the preference adjustment enters through fitness assignment. Neither adjusted indicator values nor relations implied by them are appended to $\mathcal{C}$. The exact reference-distance correction and the IBEA-specific experimental parameters are specified in Appendix F.

C FALLBACK INFERENCE

Fallback is triggered when an unresolved comparison requires a new DM query but the remaining consultation-level budget satisfies $b = 0$, or when the incoming compatible region $Z(\mathcal{C})$ is empty. Under exact LP feasibility checks, hard-mode query updates preserve nonemptiness: a query is issued only when both response halfspaces intersect the current compatible region. The empty-region branch therefore handles an incompatible incoming history.

If $Z(\mathcal{C})$ is nonempty, fallback uses $\widehat{\mathbf{z}}_{\text{H}}$, the lexicographically smallest Euclidean Chebyshev center of that region. Otherwise, it uses $\widehat{\mathbf{z}}_{\text{S}}$, obtained by minimizing the average soft-margin slack, followed by anchor-based and lexicographic tie-breaking as described in Section 3.3. Denote the selected fallback point by $\widehat{\mathbf{z}}_{\text{F}}$.

Algorithm 4 IBEA with PPL-guided preference feedback

Require: Population size N_{pop}, generations T, interaction generations $\mathcal{T}_{\mathrm{int}}$, budgets $\{B_t\}_{t\in\mathcal{T}_{\mathrm{int}}}$, oracle Query
Ensure: Final population R_T
Initialize and evaluate R_0 with $|R_0| = N_{\mathrm{pop}}$
$\mathcal{C} \leftarrow \emptyset$; $\widehat{\mathbf{z}}_{-1} \leftarrow \varnothing$
$\mathcal{F}(R_0) \leftarrow \texttt{IBEA-Fitness}(R_0, \varnothing)$
for $t = 0$ to $T-1$ **do**
  $\widehat{\mathbf{z}}_t \leftarrow \widehat{\mathbf{z}}_{t-1}$
  if $t \in \mathcal{T}_{\mathrm{int}}$ **then**
    $U_t \leftarrow R_t$; $(\Phi_t, \kappa_t) \leftarrow \texttt{DistinctObjectives}(U_t)$
    $(\widehat{\sigma}_t, \mathcal{C}, M_t) \leftarrow \texttt{PPL-Rank}(\Phi_t, \mathcal{C}, B_t, \mathrm{Query})$
    $\widehat{\mathbf{z}}_t \leftarrow \texttt{SoftEstimate}(\mathcal{C})$
    $\mathcal{F}(R_t) \leftarrow \texttt{IBEA-Fitness}(R_t, \widehat{\mathbf{z}}_t)$
  end if
  $O_t \leftarrow \texttt{Variation}(\texttt{FitnessTournamentMating}(R_t, \mathcal{F}(R_t)))$
  $\texttt{Evaluate}(O_t)$
  $V_t \leftarrow R_t \uplus O_t$
  $\mathcal{F}(V_t) \leftarrow \texttt{IBEA-Fitness}(V_t, \widehat{\mathbf{z}}_t)$
  $(R_{t+1}, \mathcal{F}(R_{t+1})) \leftarrow \texttt{EnvironmentalSelection}(V_t, \mathcal{F}(V_t), N_{\mathrm{pop}})$
end for
return R_T

The fallback routine sorts the entire current consultation set by increasing $\|\phi_i - \widehat{\mathbf{z}}_{\mathrm{F}}\|_2^2$, resolving equal distances by a fixed deterministic order. This complete ranking replaces the partially constructed chain, and the routine returns without issuing further queries in the current consultation.

Relations predicted by the fallback point are estimates rather than LP certificates. They are neither appended to the persistent history $\mathcal{C}$ nor recorded as hard-mode relations in the local matrix $\mathbf{Q}$. This separation prevents predicted relations from being reused as explicit feedback or hard certificates. Regardless of the termination mode, the continuous preference estimate passed to the host EMO algorithm is always $\widehat{\mathbf{z}}_t = \widehat{\mathbf{z}}_{\mathrm{S}}$; the fallback point $\widehat{\mathbf{z}}_{\mathrm{F}}$ is used only to construct the fallback ranking.

D Theoretical Analysis

We prove Theorem 4.4 for the complete-ranking routine in Algorithm 2, using h_{ij}, $Z(\mathcal{C})$, and $\Sigma(\mathcal{C})$ as defined in the main text. The consultation set Φ, incoming history $\mathcal{C}_0$, and ground-truth point $\mathbf{z}^\star$ are fixed. Randomness refers only to the algorithm's insertion order. Fallback issues no further queries; a budget-induced switch therefore only truncates the sequence analyzed below.

D.1 Realizability, certification, and ranking updates

Lemma D.1 (Ground-truth preservation and hard-mode correctness)**.** *Under Assumption 4.1, every queried history reached during the consultation satisfies*

$$\mathbf{z}^\star \in Z(\mathcal{C}), \qquad \sigma^\star \in \Sigma(\mathcal{C}). \tag{7}$$

Every certified relation is correct. If the routine completes in HARD *mode, its output is $\sigma^\star$.*

Proof. The incoming constraints are satisfied by $\mathbf{z}^\star$, and each new queried answer is its nonzero comparison sign. Thus, adding queried constraints never removes $\mathbf{z}^\star$, which induces the same strict ranking $\sigma^\star$ throughout the call. If $(\alpha_{ij}^+, \alpha_{ij}^-) = (1, 0)$, no point in $Z(\mathcal{C})$ has $h_{ij}(\mathbf{z}) \le 0$, so $h_{ij}(\mathbf{z}^\star) > 0$. The reverse case is analogous, and $(0, 0)$ is impossible. Direct answers are correct by assumption. Transitive propagation preserves their correctness because the ground-truth distance ordering is transitive. Starting with an empty chain, each completed insertion therefore preserves its true order. Hard-mode completion yields the entire ground-truth ranking. □

Lemma D.2 (Exact update of the consistent ranking set)**.** *For a queried pair from the current Φ and a label $o \in \{+1, -1\}$,*

$$\Sigma\big(\mathcal{C} \cup \{(\phi_i, \phi_j, o)\}\big) = \Sigma_{ij}^{o}(\mathcal{C}). \tag{8}$$

Algorithm 5 Detailed PPL-Rank with explicit transitive propagation

Require: Distinct objective vectors $\Phi = \{\phi_i\}_{i=1}^N$, incoming queried history $\mathcal{C}$, explicit-query budget B, oracle Query
Ensure: Ranking $\widehat{\sigma}$, updated history $\mathcal{C}$, mode $M \in \{\text{HARD}, \text{FALLBACK}\}$

$b \leftarrow B$
if $Z(\mathcal{C}) = \emptyset$ **then**
  return $(\texttt{FallbackRank}(\Phi, \mathcal{C}), \mathcal{C}, \text{FALLBACK})$
end if
$\mathbf{Q} \leftarrow \mathbf{0}_{N\times N}$; $\rho \leftarrow$ random permutation of $[N]$; $\mathcal{S} \leftarrow ()$
for $r = 1$ to N **do**
  $j \leftarrow \rho(r)$
  while $\exists k \in \mathcal{S} : \mathbf{Q}_{kj} = 0$ **do**
    $\mathcal{U} \leftarrow \{k \in \mathcal{S} : \mathbf{Q}_{kj} = 0\}$
    $i \leftarrow \text{Median}_{\mathcal{S}}(\mathcal{U})$
    Compute α_{ij}^+ and α_{ij}^- using equation 3
    if $(\alpha_{ij}^+, \alpha_{ij}^-) = (1, 0)$ **then**
      $o \leftarrow +1$
    else if $(\alpha_{ij}^+, \alpha_{ij}^-) = (0, 1)$ **then**
      $o \leftarrow -1$
    else if $(\alpha_{ij}^+, \alpha_{ij}^-) = (1, 1)$ and $b > 0$ **then**
      $o \leftarrow \text{Query}(\phi_i, \phi_j)$
      $\mathcal{C} \leftarrow \mathcal{C} \cup \{(\phi_i, \phi_j, o)\}$; $b \leftarrow b - 1$
    else
      return $(\texttt{FallbackRank}(\Phi, \mathcal{C}), \mathcal{C}, \text{FALLBACK})$
    end if
    $\mathbf{Q}_{ij} \leftarrow o$; $\mathbf{Q}_{ji} \leftarrow -o$
    if $o = +1$ **then**
      $\mathcal{T}_{ij}(o) \leftarrow \{k \in \mathcal{S} : \mathbf{Q}_{ki} = +1\}$
    else
      $\mathcal{T}_{ij}(o) \leftarrow \{k \in \mathcal{S} : \mathbf{Q}_{ik} = +1\}$
    end if
    for all $k \in \mathcal{T}_{ij}(o)$ such that $\mathbf{Q}_{kj} = 0$ **do**
      $\mathbf{Q}_{kj} \leftarrow o$; $\mathbf{Q}_{jk} \leftarrow -o$
    end for
  end while
  Insert j into the unique position in $\mathcal{S}$ consistent with $\mathbf{Q}$
end for
return $(\mathcal{S}, \mathcal{C}, \text{HARD})$

Proof. If σ is induced by a point $\mathbf{z}$ in the updated region, it also belongs to $\Sigma(\mathcal{C})$. Because σ is strict on the current candidates, its inducing point satisfies $h_{ij}(\mathbf{z}) \neq 0$. The added weak inequality $o\, h_{ij}(\mathbf{z}) \geq 0$ therefore implies $o\, h_{ij}(\mathbf{z}) > 0$, giving the required orientation. Conversely, a witness $\mathbf{z} \in Z(\mathcal{C})$ inducing a ranking in $\Sigma_{ij}^o(\mathcal{C})$ already satisfies $o\, h_{ij}(\mathbf{z}) > 0$, so it remains a witness after the update. This proves both inclusions. Historical comparisons may involve vectors outside Φ; the queried pair in this identity belongs to Φ. □

Why LP ambiguity is insufficient. Weak feasibility of both orientations does not imply that both sets $\Sigma_{ij}^+(\mathcal{C})$ and $\Sigma_{ij}^-(\mathcal{C})$ are nonempty. For example, take $m = 1$, $\phi_1 = 0$, $\phi_2 = 1$, $Z = [0, 1]$, $\mathbf{z}^\star = 1/4$, and $\mathcal{C}_0 = \{(\phi_1, \phi_2, +1)\}$. Then $h_{12}(z) = 1/2 - z$ and $Z(\mathcal{C}_0) = [0, 1/2]$. Only the strict ranking $\phi_1 \succ \phi_2$ is realizable in this region, yet both LP orientations are feasible because the reverse weak inequality is satisfied at $z = 1/2$. With an initially empty local relation matrix and a positive budget, the routine queries this pair again without reducing $|\Sigma(\mathcal{C}_0)| = 1$. This realizable example violates Assumption 4.3; it shows why that assumption cannot be deduced from realizability or LP ambiguity.

D.2 Counting realizable rankings

Lemma D.3 (Hyperplane-arrangement bound)**.** *For $P = \binom{N}{2}$, inequality equation 6 holds without a general-position or full-dimensional-domain assumption.*

Proof. Each equation $h_{ij}(\mathbf{z}) = 0$ defines a hyperplane in $\mathbb{R}^m$, since the objective vectors are distinct. There are at most P such hyperplanes. Their complement is partitioned into cells on which all comparison signs are constant. Every strictly realizable ranking has a witness outside all these hyperplanes and hence in one cell. A cell induces only one ranking, so the number of rankings realized in Z cannot exceed the number of cells in the whole space.

Let $R(P, m)$ be the maximum number of cells formed by at most P hyperplanes in $\mathbb{R}^m$. Adding one hyperplane splits at most as many existing cells as are formed by the preceding hyperplanes restricted to it. Therefore,

$$R(P, m) \le R(P-1, m) + R(P-1, m-1), \tag{9}$$

with $R(0, m) = R(P, 0) = 1$. Induction using Pascal's identity gives

$$R(P, m) \le \sum_{r=0}^{\min\{m,P\}} \binom{P}{r}. \tag{10}$$

Finally, this sum counts subsets of $[P]$ of size at most m. Encoding each subset by its increasing list of indices padded with zeros to length m injects these subsets into $\{0, \dots, P\}^m$, proving the upper bound $(P+1)^m$. Degenerate arrangements or restrictions to Z can only reduce the number of realizable rankings. □

D.3 Bounding the actual number of queries

Lemma D.4 (Budget and comparison-attempt bounds)**.** *For every execution of Algorithm 2, the number A of direct comparison attempts satisfies*

$$A \le A_N := \sum_{r=2}^{N} \lceil \log_2 r \rceil, \qquad Q \le \min\{B, A\}.$$

These bounds do not require realizability or splitting.

Proof. During insertion into a chain, unresolved candidates form a contiguous interval. Initially, that interval is the entire chain. Resolving its median and propagating the result removes the median and one side, leaving another contiguous interval. From an interval of size u, a successful attempt leaves at most $\lfloor u/2 \rfloor$ unresolved candidates.

Inserting the rth candidate starts with $r-1$ unresolved candidates and therefore needs at most $\lceil \log_2 r \rceil$ direct comparison attempts. An attempt that triggers fallback only truncates this search, so the same bound applies. Summing over insertions gives $A \le A_N$. At most one oracle call occurs per attempt, and each call decrements the nonnegative remaining budget by one. Hence $Q \le A$ and $Q \le B$. □

Proof of Theorem 4.4. The unconditional budget and comparison-count bounds follow from Lemma D.4. Fix any execution satisfying the two assumptions, and let Q be its actual number of explicit queries. Write $\mathcal{C}_q$ for the queried history after the qth query, for $0 \le q \le Q$, and set $S_q := |\Sigma(\mathcal{C}_q)|$. Ground-truth preservation gives $S_q \ge 1$ at every such index. For each actual query, Lemma D.2 and the splitting assumption give

$$S_{q+1} \le (1-\eta) S_q, \qquad 0 \le q < Q. \tag{11}$$

Iterating along this finite execution yields

$$1 \le S_Q \le (1-\eta)^Q S_0. \tag{12}$$

Taking logarithms, using that Q is an integer, and combining the budget and comparison-count bounds gives

$$Q \le \min\left\{ B,\ A_N,\ \left\lfloor \frac{\ln S_0}{-\ln(1-\eta)} \right\rfloor \right\}. \tag{13}$$

The argument bounds the queries actually issued; it does not identify algorithmic termination with the first time $S_q = 1$. It also allows early fallback with $S_Q > 1$. Since the assumptions hold almost surely

for randomized insertion, the same deterministic upper bound applies to $\mathbb{E}[Q]$. Finally, $S_0 \le N_{\mathrm{rank}}$, Lemma D.3, and $-\ln(1-\eta) \ge \eta$ imply

$$\begin{aligned}\mathbb{E}[Q] &\le \min\left\{B,\ A_N,\ \frac{m\ln(1+\binom{N}{2})}{-\ln(1-\eta)}\right\} \\ &\le \frac{m\ln(1+\binom{N}{2})}{\eta} = \mathcal{O}\left(\frac{m\ln N}{\eta}\right).\end{aligned} \tag{14}$$

Hard-mode ranking correctness follows separately from Lemma D.1, without the splitting assumption. □

The result is conditional on splitting at every issued query, including any query made after the ranking set has become a singleton. At a singleton, splitting is impossible, so a run that issues another query does not satisfy the assumption. Without this assumption, the median-insertion routine still has the general query bound $Q \le \min\{B, \sum_{r=2}^{N}\lceil \log_2 r\rceil\}$, but the logarithmic ranking-count bound is not established. Certification and propagation leave $\mathcal{C}$, and therefore $\Sigma(\mathcal{C})$, unchanged: their benefit is to resolve comparisons without acquiring new labels.

Interaction across consultations. Let Q_t denote the number of explicit queries issued at interaction generation t, counted until hard-mode completion or the switch to fallback. In the present routine, neither `SoftEstimate` nor `FallbackRank` issues additional queries. Therefore, the total number of DM queries over the run is

$$Q_{\mathrm{run}} := \sum_{t\in\mathcal{T}_{\mathrm{int}}} Q_t.$$

For each interaction generation t, the hard-mode query count is bounded by its consultation budget, i.e., $Q_t \le B_t$. Consequently,

$$Q_{\mathrm{run}} \le \sum_{t\in\mathcal{T}_{\mathrm{int}}} B_t.$$

This budget inequality holds without the realizability or splitting assumptions. The conditional bound in Theorem 4.4 applies separately to each consultation for which its assumptions hold, using that consultation's candidate set, incoming history, and budget. Since candidate sets may change across consultations, these per-consultation logarithmic bounds do not by themselves establish a run-wide logarithmic bound independent of the number of consultations.

Table 1: Population size in different problems

Problem	m	N
ZDT	2	100
DTLZ	3	64
	5	128
	8	224
	10	288
WFG	3	64
Inverse RNA	2	100

E EMPIRICAL ANALYSIS

In this section, we show the results of our proposed method running on ZDT, DTLZ, mDTLZ, minusDTLZ1, and WFG test suites. The performance metrics are shown in Table 7, Table 9, Table 8.

- **Low-objective benchmarks.** We consider the classical two-objective ZDT problems (ZDT1–ZDT4 and ZDT6) Zitzler et al. (2000) and the three-objective modified DTLZ problems (mDTLZ1– mDTLZ4) Wang et al. (2019).

Table 2: Maximum generation in different problems

Problem	Generation
ZDT	250
DTLZ1	$500 + 50(m-2)$
DTLZ2	$200 + 50(m-2)$
DTLZ3	$1000 + 50(m-2)$
DTLZ4	$200 + 50(m-2)$
WFG	$1000 + 50(m-2)$
Inverse RNA	250

- **Scalable benchmark families.** We use DTLZ1–DTLZ6 Deb et al. (2002a) and the inverted minusDTLZ1–minusDTLZ4 variants Ishibuchi et al. (2016) with $m \in \{3, 5, 8, 10\}$ objectives.
- **WFG benchmarks.** We include WFG1–WFG7 Huband et al. (2006) to test performance across varied Pareto-front geometries and problem characteristics, including nonseparability, multimodality, deceptiveness, and parameter-dependent bias.
- **Bio-sequence design.** We consider an RfaH protein sequence design problem adapted from the integrative protein sequence design benchmark Hong & Kortemme (2024). The RfaH results are shown in Figure 2.

F Experimental Details

F.1 Experimental Protocol and Compared Methods

We evaluate PPL from two complementary perspectives. First, we integrate PPL with three representative multi-objective evolutionary algorithms, namely NSGA-II, MOEA/D, and IBEA, which rely on dominance, decomposition, and performance indicators, respectively. The resulting preference-guided algorithms are evaluated in end-to-end interactive multi-objective optimization loops. Second, we evaluate PPL as a standalone active ranking module to examine its ranking accuracy, query efficiency, and the contribution of each algorithmic component.

F.2 Peer Algorithms

We compare the three PPL-enhanced EMO algorithms with five preference-aware comparison methods: (i) **MOEA/D-PLVF** Li et al. (2019), which progressively approximates the DM's value function from periodically elicited evaluations and uses the learned model to adapt the decomposition-based search; (ii) **IEMO/D** Tomczyk & Kadzinski (2020), which uses pairwise comparisons to identify compatible preference models and incorporates sampled instances of these models as search directions; (iii) **DPBEMO** Huang et al. (2024), which employs preference-based clustering and stochastic dueling bandits to guide environmental selection in NSGA-II; (iv) **MOEA/D-PBO** González et al. (2017), a comparison variant of D-PBEMO that uses preferential Bayesian optimization as its consultation module and MOEA/D as its optimization module; and (v) **NSGA-II/DW** Thekumparampil et al. (2025), an adaptation developed in this study that incorporates the active ranking strategy from "Comparing Few to Rank Many" into the environmental selection of NSGA-II. At each interaction generation, the combined parent–offspring pool is used as the consultation set; DopeWolfe actively selects a limited number of comparison queries, estimates a latent preference scoring model, and uses the resulting scores to rank candidate solutions during selection. This baseline is included because it is also an active ranking method that infers an ordering over many candidate solutions from few preference queries, thereby providing a direct comparison with the ranking component of PPL. All methods are evaluated with same budget of consulting the DM for fair comparisons.

F.3 Simulated Decision Makers

All end-to-end experiments use simulated DMs whose ground-truth preference models are controlled by the experimenter but hidden from the optimization algorithms. The algorithms can access these preferences only through explicit pairwise comparison queries. We use a fixed aspiration-point model as the default setting and a stochastic variant to evaluate robustness to noisy feedback.

Fixed aspiration-point preference. In the default setting, the DM is governed by a fixed aspiration point $\mathbf{z}^\star$, expressed in the same objective space as the candidate solutions. For a queried pair of candidates (ϕ_i, ϕ_j), the DM prefers the candidate closer to $\mathbf{z}^\star$:

$$\phi_i \succ_\star \phi_j \quad \Longleftrightarrow \quad \|\phi_i - \mathbf{z}^\star\|_2 < \|\phi_j - \mathbf{z}^\star\|_2 . \tag{15}$$

This setting is consistent with the aspiration-point preference model used by `PPL` and therefore serves as the default model-compatible scenario. The aspiration point is fixed throughout an independent run and is not revealed to the optimization algorithm.

Stochastic preference. To model random variability in pairwise judgments, we perturb the ground-truth aspiration point before generating each response. For each queried pair, a temporary reference point $\mathbf{z}$ is sampled according to

$$\mathbf{z} \sim \mathcal{N}\left(\mathbf{z}^\star, \sigma^2 \mathbf{I}\right), \tag{16}$$

and the oracle returns $\phi_i \succ \phi_j$ whenever

$$\|\phi_i - \mathbf{z}\|_2 < \|\phi_j - \mathbf{z}\|_2 . \tag{17}$$

The parameter σ controls the level of stochasticity. When $\sigma = 0$, this model reduces to the fixed aspiration-point oracle. For $\sigma > 0$, repeated comparisons involving nearby candidates may produce different responses, thereby introducing preference noise relative to the fixed-point model assumed by the exact inference procedure.

For each problem instance and independent run, all competing algorithms use the same ground-truth DM parameters and cumulative pairwise-query budget. The same initial population and matched random seeds are used whenever applicable, so that differences in optimization performance are attributable to the preference-learning and search mechanisms rather than to changes in the underlying experimental conditions. The aspiration-point values and uncertainty parameters used for individual benchmark configurations are reported in the corresponding parameter tables.

F.4 Benchmark Problems

Synthetic benchmarks

To evaluate the performance of NSGA-II/`PPL` across a range of problem scales and characteristics, we consider both synthetic benchmarks and the Bio-sequence design applications.

- **Low-dimensional problems.** The classical two-objective ZDT suite (ZDT1–ZDT4, ZDT6) Zitzler et al. (2000) is adopted to assess performance under diverse Pareto front geometries (e.g., convex and discontinuous). We also include the modified DTLZ problems (mDTLZ1–mDTLZ4) Wang et al. (2019) with 3 objectives to further evaluate algorithm behavior in structured yet low-dimensional settings.
- **Scalable high-dimensional problems.** To assess scalability, we include the DTLZ family (DTLZ1–DTLZ6) Deb et al. (2002a), and inverted variants (minusDTLZ1–minusDTLZ4) Ishibuchi et al. (2016), covering objective dimensionalities $m \in \{3, 5, 8, 10, 15, 20\}$. These problems pose varying challenges in terms of convergence, diversity, and degeneracy.
- **Degenerate problems.** The WFG1–WFG7 Huband et al. (2006) suite is used to evaluate performance under complex characteristics such as parameter dependence, non-separability, and degenerate Pareto sets.
- **Bio-sequence design application.** We further examine the algorithm on a real-world biomolecular inverse-design task, an RfaH protein sequence design problem adapted from

the integrative protein sequence design benchmark Hong & Kortemme (2024). The bio-sequence design task is included to assess whether the proposed optimizer can handle discrete, model-coupled search spaces that arise in computational biomolecular design. The RfaH task represents protein sequence design over a 20-letter amino-acid alphabet, where a single sequence must be evaluated against multiple structural states.

BIO-SEQUENCE DESIGN BENCHMARK

This section introduces the bio-sequence design benchmark problem used in our experiments, namely an RfaH protein sequence design problem adapted from the integrative protein sequence design benchmark Hong & Kortemme (2024). The task represents amino-acid-sequence design in a computational biomolecular design setting. It is naturally formulated as a bi-objective optimization problem. In this work, however, we solve it in a preference-guided manner by specifying a reference point in the objective space, so that the search is focused on a practically relevant region of interest rather than on a uniform approximation of the entire Pareto front. Further, we summarize the key parameter settings in our experiments.

RfaH is a representative fold-switching protein whose C-terminal domain can adopt two distinct conformational states: an N-terminal-domain-bound all-α state, denoted as RfaH$*\alpha$, and a dissociated all-β state, denoted as RfaH$*\beta$ Burmann et al. (2012); Zuber et al. (2019). Following the RfaH multistate design benchmark in Hong & Kortemme (2024), we formulate this task as a two-objective discrete protein sequence design problem, where a single amino-acid sequence must be compatible with both structural states.

Let $\mathcal{A} = \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y\}$ be the alphabet of the 20 standard amino acids. The designable residue set is $\mathcal{D} = \{119, 120, \ldots, 154\}$, and therefore the number of decision variables is $L = |\mathcal{D}| = 36$. A candidate solution is a 36-residue amino-acid sequence

$$\mathbf{x} = (x_r) * r \in \mathcal{D} = x * 119x_{120} \ldots x_{154}, \quad x_r \in \mathcal{A}.$$

Thus, the discrete search space is $\mathcal{X} = \mathcal{A}^{36}$, whose size is 20^{36}. All non-designable residues are fixed to the wild-type RfaH sequence, while the residues in $\mathcal{D}$ are optimized by the algorithm.

Let $\mathcal{B}^{\alpha}$ and $\mathcal{B}^{\beta}$ denote the backbone structures of the RfaH$*\alpha$ and RfaH$*\beta$ states, respectively. For each state $s \in \{\alpha, \beta\}$, a full state-specific sequence $X^s(\mathbf{x})$ is constructed by inserting the candidate sequence $\mathbf{x}$ into the designable positions and keeping the remaining residues unchanged:

$$X^s * r(\mathbf{x}) = \begin{cases} x_r, & r \in \mathcal{D}, \\ x^{\mathrm{WT}}r, & r \notin \mathcal{D}, \end{cases} \quad r \in \Omega_s,$$

where $x^{\mathrm{WT}}r$ is the wild-type residue at position r, and Ω_s is the set of residues present in the structural template of state s. In the benchmark implementation, the RfaHα template contains residues 1–155 with the unresolved loop residues 98–117 omitted, while the RfaHβ template contains residues 108–162 Hong & Kortemme (2024). The same candidate residues x_r are tied across the two states. That is, residue r in RfaH$*\alpha$ and residue r in RfaH$_\beta$ must use the same amino acid for every $r \in \mathcal{D}$. This tied-sequence constraint distinguishes the problem from two independent single-state design tasks: the optimizer must find one sequence that balances the sequence preferences of both conformations.

The objective functions are derived from ProteinMPNN Dauparas et al. (2022), a structure-conditioned protein sequence design model. For a given state $s \in \{\alpha, \beta\}$, ProteinMPNN assigns a likelihood to the full sequence $X^s(\mathbf{x})$ conditioned on the backbone $\mathcal{B}^s$. We use the state-specific negative log-likelihood score as the objective:

$$\ell_s(\mathbf{x}) = -\frac{1}{|\Omega_s|} \log p_\theta(X^s(\mathbf{x}) \mid \mathcal{B}^s),$$

where p_θ denotes the probability model parameterized by ProteinMPNN. The normalization by $|\Omega_s|$ gives an average per-residue negative log-likelihood. A smaller value of $\ell_s(\mathbf{x})$ indicates that the candidate sequence is more compatible with the backbone structure of state s under the ProteinMPNN model.

Since RfaH has two structural states, the two minimization objectives are defined as $f_1(\mathbf{x}) = \ell_\alpha(\mathbf{x})$ and $f_2(\mathbf{x}) = \ell_\beta(\mathbf{x})$. Thus, the RfaH protein sequence design problem is formulated as

$$\min_{\mathbf{x}\in\mathcal{A}^{36}} \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}))\,.$$

The first objective favors sequences that are likely to fold into or stabilize the RfaH_α state, whereas the second objective favors sequences that are likely to fold into or stabilize the RfaH_β state. These two objectives can be conflicting because the same residue may experience different secondary-structure environments, solvent accessibility, and local interaction patterns in the two states. Therefore, the goal of the underlying benchmark is to approximate a set of Pareto-optimal sequences representing different trade-offs between compatibility with the RfaH_α and RfaH_β conformations.

Preference-guided setting. The task above remains unchanged as a bi-objective optimization problem; only the experimental decision-making setting is modified. In our experiments, we solve its preference-guided variant by specifying a reference point in the objective space. This choice is well motivated for bio-sequence design. In multistate protein design, the design goal is application-specific and often combines state-specific scores through positive/negative design considerations or weighting factors Löffler et al. (2017). For RfaH specifically, the functional cycle is governed by reversible switching between the autoinhibitory α state and the translation-active β state Zuber et al. (2019), making state preference or state balance a biologically meaningful design consideration.

Accordingly, instead of approximating the whole Pareto front uniformly, we focus the search on a practically relevant region of interest. Let $\tilde{\mathbf{F}}(\mathbf{x}) = (\tilde{f}_1(\mathbf{x}), \tilde{f}_2(\mathbf{x}))$ denote the objective vector after the same scaling or preprocessing used by the preference-based solver, and let $\mathbf{r}$ be a user-specified reference point in that transformed objective space. The optimizer is then guided toward Pareto-optimal solutions close to $\mathbf{r}$.

F.5 Standalone Preference-Inference Evaluation

We evaluate standalone `PPL` on 44 personalized ranking tasks derived from 29,206 English-language Chatbot Arena comparisons among 20 large language models (LLMs) Chiang et al. (2024).

Evaluation Protocol. The logged-preference experiments use five-fold retrospective replay with ten random seeds per task, yielding $44 \times 5 \times 10 = 2{,}200$ fold runs. Acquisition is restricted to pairs available in the historical pool, and one query consumes one raw target-user preference event. We additionally construct 240 linearly realizable complete-oracle scenarios with known ground-truth rankings; all item pairs are queryable, and responses at the nonzero noise level follow the Bradley–Terry model Bradley & Terry (1952). We report the normalized area under the held-out pairwise accuracy query fraction curve (AULC), endpoint pairwise accuracy (PairAcc), and raw target-user query counts. For the logged replay, we aggregate out-of-fold predictions across five folds and ten seeds within each user, then apply paired two-sided Wilcoxon signed-rank tests with Holm correction.

External Preference-Learning Baselines. The comparison pool comprises the following baselines:

(i) **Classical pairwise learners:** an actively queried Bradley–Terry model (active BT) Bradley & Terry (1952), and Gaussian process preference learning (GPPL) with Bayesian active learning by disagreement (BALD) Chu & Ghahramani (2005); Houlsby et al. (2011).

(ii) **Feature-aware and multi-dimensional learners:** a feature-aware Bradley–Terry–Luce model (Feature-BTL) Saha & Rajkumar (2024) equipped with D-optimal acquisition Thekumparampil et al. (2025), denoted D-optimal Feature-BTL, and Comparison-based Active Preference Learning for Multi-dimensional Personalization (AMPLe), a Bayesian multi-dimensional preference learner with volume-based acquisition Oh et al. (2025).

(iii) **External active-ranking algorithms:** the contextual logistic method GURO Bergström et al. (2024); DopeWolfe, which uses randomized Frank–Wolfe optimization for D-optimal query design Thekumparampil et al. (2025); Active Sampling for Pairwise Comparisons via Approximate Message Passing and Information Gain Maximization (ASAP) Mikhailiuk et al. (2020); and Multisort, which aggregates noisy rankings produced by randomized Quicksort Maystre & Grossglauser (2017).

For Chatbot Arena, all acquisition policies are restricted to pairs available in the historical pool; Multisort is therefore evaluated in a graph-constrained form. The 240-scenario complete-oracle scaling suite uses active BT, D-optimal Feature-BTL, and GURO alongside PPL and its matched controls. To isolate the effect of robust certification, we additionally evaluate a matched PPL variant that disables certification-based stopping while retaining the same initialization, inference, and acquisition procedures. The main text highlights AMPLe on Chatbot Arena and D-optimal Feature-BTL and active BT in the complete-oracle comparison; detailed results for the remaining applicable baselines are provided in the appendix.

F.6 Evaluation Metrics and Statistical Analysis for End-to-End EMO

We evaluate the final populations using two population-level preference-proximity measures: the mean aspiration-point distance $\bar{\epsilon}$ and the Gaussian-averaged aspiration-point distance $\mathcal{D}$. Both measures are minimized.

Objective-space representation. Let $S = \{\boldsymbol{\phi}_i\}_{i=1}^N$ denote the objective vectors of the final population, where $\boldsymbol{\phi}_i \in \mathbb{R}^m$, N is the population size, and m is the number of objectives. Let $\mathbf{z}^\star$ denote the designated aspiration point. Distances are computed using the objective values supplied to the metric routines, without additional normalization within these routines. The aspiration point and reference samples are expressed on the same objective scales as the population.

Mean aspiration-point distance. The mean aspiration-point distance is defined as

$$\bar{\epsilon}(S) = \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{\phi}_i - \mathbf{z}^\star\|_2 . \tag{18}$$

This measure summarizes the average proximity of all individuals in the final population to the designated aspiration point. Each population member contributes equally to the reported value.

Gaussian-averaged aspiration-point distance. To evaluate population proximity over reference locations around $\mathbf{z}^\star$, we define

$$\mathcal{D}(S) = \mathbb{E}_{\mathbf{z}\sim Q_{\text{metric}}}\left[\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\phi}_i - \mathbf{z}\|_2\right], \tag{19}$$

where

$$Q_{\text{metric}} = \mathcal{N}\left(\mathbf{z}^\star, \sigma^2_{\text{metric}} I_m\right), \qquad \sigma_{\text{metric}} = 0.1, \tag{20}$$

and I_m is the m-dimensional identity matrix. Thus, σ_{metric} specifies the reference-point perturbation scale in each objective coordinate.

We approximate the expectation using $K = 100$ independent reference samples, $\mathbf{z}^{(k)} \sim Q_{\text{metric}}$, yielding

$$\widehat{\mathcal{D}}(S) = \frac{1}{KN}\sum_{k=1}^{K}\sum_{i=1}^{N}\left\|\boldsymbol{\phi}_i - \mathbf{z}^{(k)}\right\|_2 . \tag{21}$$

For each sampled reference point, the distance is averaged over the entire final population; these population averages are then averaged over the reference samples. The reported numerical values of $\mathcal{D}$ are computed using this estimator.

Reference samples are generated with metric seed 42. The same reference sample set is used across algorithms within each problem configuration, ensuring that comparisons use identical evaluation references.

Relationship to the feedback models. The evaluation distribution Q_{metric} is specified separately from the simulated DM models in Section F.3. Its scale $\sigma_{\text{metric}} = 0.1$ is fixed and does not vary with the stochastic DM's response-noise parameter σ. Both metrics retain a common evaluation target centered at $\mathbf{z}^\star$. Thus, the stochastic experiments assess population proximity to this target under noisy feedback.

Independent runs and statistical comparisons. Each end-to-end experimental configuration is evaluated over 20 independent optimization runs. Algorithms are paired by matched run index under the matched experimental conditions described in Section F.3. Each run contributes one value of $\bar{\epsilon}$ and one value of $\widehat{\mathcal{D}}$. The independent optimization run is the unit of statistical analysis; population members and Gaussian reference samples are used to compute these run-level values.

Pairwise comparisons use two-sided Wilcoxon signed-rank tests (Wilcoxon, 1945) on matched run-level metric values, with Holm correction at $\alpha = 0.05$ (Holm, 1979). Scott–Knott clustering (Mittas & Angelis, 2013) is additionally used to summarize the relative performance of the compared algorithms.

F.7 Parameter Settings

The parameter settings based on the NSGA-II/PPL algorithm framework are as follows:

- The evolutionary generations between two consecutive interactions are set to $\tau = 20$;
- Index of convergence $\eta = 10^{-5}$;
- The elitism preservation coefficient is set to $\nu = 0.5$;
- Crossover probability and distribution index of simulated binary crossover operation: $p_c = 1.0$, $\eta_c = 30$;
- Mutation probability and distribution index of polynomial mutation operation: $p_m = 1/n$, $\eta_m = 20$.
- Population size and iteration number settings can be referenced in Table 1, Table 2;
- The reference points of different test problems in different dimensions are provided in Table 3.

F.8 Implementation and Compute Resources

To facilitate reproducibility, we report the computational environment and resource consumption as follows:

- **Hardware:** All experiments were conducted on a workstation equipped with an **Intel(R) Xeon(R) Gold 6230 CPU @ 2.10GHz** and 128 GB of RAM. No GPU acceleration was required as the algorithm is computationally efficient on CPUs.
- **Software Environment:** The PBEMO framework and the PPL algorithm are implemented in **C++** and compiled with g++ (optimization level -O3). The experiments were performed on Ubuntu 22.04.
- **Execution Time:** For a typical problem instance (e.g., DTLZ series with $M = 3, D = 7, N = 64$), a single run takes approximately **8.5 seconds** on a single core.
- **Total Compute:** Considering all test problems (DTLZ1-7), multiple seeds, and comparative baselines, the total compute time for the results reported in this paper is approximately **10–15 CPU hours** (including hyperparameter tuning and preliminary testing).

G Diagnostic control: beyond active ranking alone

The overall comparison does not yet rule out a simpler explanation: NSGA-II/PPL may improve performance merely because it introduces active preference ranking into NSGA-II. We use NSGA-II-DW as a diagnostic control, since it also embeds active ranking in environmental selection but lacks PPL's ambiguity certification, cross-iteration memory, and reusable preference constraints.

We therefore isolate the PPL–DW comparison under the same interaction budget and statistical protocol. For each controlled benchmark problem, we compute the paired log-ratio $L_{\mathrm{DW},p}$ under $\bar{\epsilon}(S)$ and apply Holm-corrected paired Wilcoxon signed-rank tests over repeated runs. NSGA-II/PPL remains favorable across most benchmark families. Consistently, the quality–interaction frontier in Fig. 4 shows thatPPLreaches and maintains lower expected distance under comparable explicit DM query budgets, whereas NSGA-II-DW exhibits unstable jumps as the interaction budget increases.

Table 3: The settings of the aspiration point $\mathbf{z}^*$ for different benchmark problems (represented in the objective space).

problem	m	**reference point**
ZDT1	2	$(0.5,\ 0.4)^\top$
ZDT2	2	$(0.2,\ 0.8)^\top$
ZDT3	2	$(0.15,\ 0.4)^\top$
ZDT4	2	$(0.3,\ 0.4)^\top$
ZDT6	2	$(0.67,\ 0.55)^\top$
WFG1	3	$(1.4,\ 1.2,\ 2.2)^\top$
WFG2	3	$(0.4,\ 0.14,\ 3.0)^\top$
WFG3	3	$(0.5,\ 0.9,\ 3.0)^\top$
WFG4	3	$(1.4,\ 2.1,\ 2.8)^\top$
WFG5	3	$(1.4,\ 2.1,\ 2.8)^\top$
WFG6	3	$(1.4,\ 2.1,\ 2.8)^\top$
WFG7	3	$(1.4,\ 2.1,\ 2.8)^\top$
DTLZ1	3	$(0.3,\ 0.3,\ 0.2)^\top$
	5	$(0.2,\ 0.1,\ 0.1,\ 0.3,\ 0.4)^\top$
	8	$(0.1,\ 0.2,\ 0.1,\ 0.4,\ 0.4,\ 0.1,\ 0.3,\ 0.1)^\top$
	10	$(0.05,\ 0.1,\ 0.1,\ 0.05,\ 0.1,\ 0.2,\ 0.08,\ 0.03,\ 0.3,\ 0.1)^\top$
DTLZ2	3	$(0.7,\ 0.8,\ 0.5)^\top$
	5	$(0.7,\ 0.6,\ 0.3,\ 0.8,\ 0.5)^\top$
	8	$(0.6,\ 0.5,\ 0.75,\ 0.2,\ 0.3,\ 0.55,\ 0.7,\ 0.6)^\top$
	10	$(0.3,\ 0.3,\ 0.3,\ 0.1,\ 0.3,\ 0.55,\ 0.35,\ 0.35,\ 0.25,\ 0.45)^\top$
DTLZ3	3	$(0.55,\ 0.55,\ 0.55)^\top$
	5	$(0.7,\ 0.6,\ 0.3,\ 0.8,\ 0.5)^\top$
	8	$(0.6,\ 0.5,\ 0.75,\ 0.2,\ 0.3,\ 0.55,\ 0.7,\ 0.6)^\top$
	10	$(0.3,\ 0.3,\ 0.3,\ 0.1,\ 0.3,\ 0.55,\ 0.35,\ 0.35,\ 0.25,\ 0.45)^\top$
DTLZ4	3	$(0.55,\ 0.55,\ 0.55)^\top$
	5	$(0.7,\ 0.6,\ 0.3,\ 0.8,\ 0.5)^\top$
	8	$(0.6,\ 0.5,\ 0.75,\ 0.2,\ 0.3,\ 0.55,\ 0.7,\ 0.6)^\top$
	10	$(0.3,\ 0.3,\ 0.3,\ 0.1,\ 0.3,\ 0.55,\ 0.35,\ 0.35,\ 0.25,\ 0.45)^\top$
DTLZ5	3	$(0.56,\ 0.56,\ 0.56)^\top$
	5	$(0.12,\ 0.12,\ 0.17,\ 0.24,\ 0.7)^\top$
	8	$(0.04,\ 0.04,\ 0.0566,\ 0.8,\ 0.113,\ 0.16,\ 0.2263,\ 0.68)^\top$
	10	$(0,\ 0,\ 0,\ 0,\ 0.0096,\ 0.027,\ 0.082,\ 0.25,\ 0.75,\ 0.08)^\top$
DTLZ6	3	$(0.56,\ 0.56,\ 0.56)^\top$
	5	$(0.12,\ 0.12,\ 0.17,\ 0.24,\ 0.7)^\top$
	8	$(0.04,\ 0.04,\ 0.0566,\ 0.8,\ 0.113,\ 0.16,\ 0.2263,\ 0.68)^\top$
	10	$(0,\ 0,\ 0,\ 0,\ 0.0096,\ 0.027,\ 0.082,\ 0.25,\ 0.75,\ 0.08)^\top$

Therefore, the advantage is not explained by active ranking alone. Rather, it comes from converting sparse comparisons into reusable preference information and then injecting that information into selection. The next analysis decomposes which`PPL`components are responsible for this conversion.

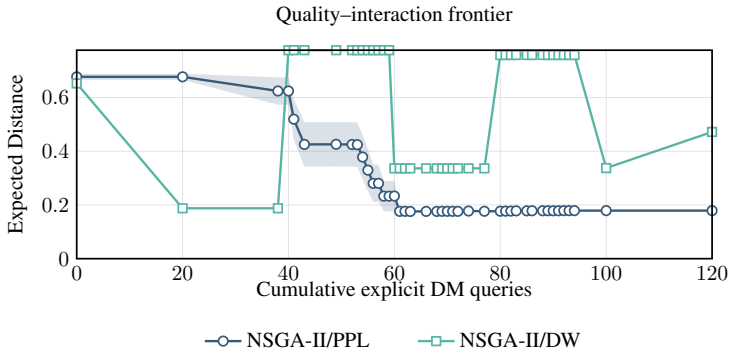

Figure 4: Quality–interaction frontier comparing NSGA-II-PPL and NSGA-II-DW. The x-axis gives cumulative explicit DM queries, and the y-axis gives $\bar{\epsilon}(S)$. Although both methods use active ranking,`PPL`achieves a lower and more stable expected distance under comparable interaction budgets, indicating that the improvement is not attributable to active ranking alone.

H Empirical Query Efficiency and Scaling

We examine explicit-query demand and relation-resolution mechanisms as the candidate-set size N and objective dimensionality m vary. Explicit queries incur interaction cost, whereas LP certification and transitive propagation are performed internally. **Resolution of pairwise relations.** Across the tested feedback-noise levels, explicit feedback accounts for 0.6%–0.7% of recorded pair-resolution attempts; LP certification and transitive propagation account for more than 99.3% (Figure 6(b)). The denominator is the number of recorded pair-resolution attempts. **Variation with candidate-set size.** With $m = 3$, we vary $N \in \{64, 128, 224, 288\}$ on DTLZ2 and WFG1 (Figure 6(c)). Explicit-query counts range from approximately 85 to 97 across these four settings. **Variation with objective dimensionality.** With $N = 64$, we vary $m \in \{3, 5, 8, 10\}$ (Figure 6(d)). The plotted explicit-query fraction increases across the tested dimensions. Together, these measurements characterize query demand and internal resolution over the examined candidate sizes and objective dimensions.

I Personalized-Replay Query Accounting and Common-Budget Comparison

Initialization and query accounting. The Chatbot Arena personalized-replay evaluation comprises 44 users, 5 folds per user, and 10 replay seeds per fold, yielding $2{,}200$ fold runs. `PPL`, its variant without certification-based stopping, and volume-active AMPLe share a capability-informed initialization constructed without target-user queries. Query totals include initialization and replay, with initialization contributing zero queries for all three methods. Each query counts one raw target-user preference event requested during replay.

Zero-query behavior and native pipelines. The shared initialization achieves held-out pairwise accuracy 0.764496 before any target-user query. `PPL` issues no target-user query in $1{,}210$ runs (55.0%). Across all runs, `PPL` issues $1{,}560$ queries (0.709 per run), compared with $103{,}329$ for the variant without certification-based stopping and $142{,}560$ for volume-active AMPLe under their respective native pipelines. These totals correspond to 98.49% and 98.91% fewer queries for `PPL`, respectively. Its normalized AULC is 0.764274, compared with 0.764496 for AMPLe (difference -0.000222; Holm-adjusted $p = 0.855$). Relative to the matched variant without certification-based stopping, its AULC is lower by 0.001053.

Table 4: Held-out pairwise accuracy on Chatbot Arena personalized replay at common absolute target-user query caps. The matched control disables certification-based stopping. Rankings from runs that stop early are carried forward.

Query cap b	PPL	PPL w/o cert. stopping	Volume-active AMPLe
0	0.764496	0.764496	0.764496
1	0.765004	0.764346	0.764496
2	0.764350	0.764716	0.764496
3	0.764266	0.764716	0.764496
4	0.764266	0.764716	0.764496

Common absolute query caps. Table 4 reports held-out pairwise accuracy at common absolute target-user query caps, using the accuracy aggregation defined in the replay evaluation protocol. For runs that stop before reaching a cap, the terminal ranking is carried forward; no values are interpolated. A cap specifies the maximum permitted queries, not the number every run must use. All three methods share the same accuracy at cap zero, and PPL's accuracy changes little across caps 1–4.

Native-pipeline totals measure actual query use under each method's stopping policy, whereas the common-cap comparison evaluates accuracy under the same maximum query allowance. The matched ablation supports certification-based stopping as a contributor to query avoidance given the fixed capability representation, while much of the observed ranking quality is already available at initialization.

J Robustness and Empirical Scope

We assess the empirical scope of PPL under varying feedback noise, an alternative preference oracle, and ambiguity regularization. **Robustness to noisy feedback.** Across the tested noise levels, preference-alignment metrics remain comparatively stable (Figure 6(a)–(b)), while internal inference continues to resolve most relations at the low explicit-query rates reported above. This behavior suggests that persistent constraints and soft-consistency handling support gradual degradation as feedback becomes less reliable. **Alternative preference oracle.** Beyond the default aspiration-point oracle, we test a linear-utility oracle, $u(\boldsymbol{\phi}) = \mathbf{z}^\top \boldsymbol{\phi}$. On three-objective DTLZ2, the attained utility lies within 0.004 of the theoretical optimum 0.1747, while the aspiration-point setting maintains close target-region alignment. These results demonstrate empirical usefulness beyond the default oracle, without establishing oracle-independent guarantees. **Sensitivity to regularization.** Varying λ produces only small differences in $\mathcal{D}(S)$ and $\bar{\epsilon}(S)$, with the best values distributed across neighboring settings (Table 6). We therefore use $\lambda = 0$, reducing ambiguity certification to a feasibility-based test and avoiding additional tuning without sacrificing observed optimization quality. Together, these results characterize PPL under the tested noise levels, preference oracles, and regularization settings.

K Ablation study on the core design choices

We first quantified the contribution of the main components of our method, and evaluated three ablated variants in a paired-seed manner: removing transitive propagation in ranking (-Trans), removing cross-iteration retention of ambiguous queried comparisons (-Memory), and removing preference-guided candidate ordering based on the LP-estimated utility vector (-PrefSel). Since the scale of the final accuracy metric $\bar{\epsilon}(S)$ differs substantially across benchmark suites, we report the raw final $\bar{\epsilon}(S)$ suite-wise in Table 5, rather than collapsing the results into a single overall raw mean. For the paired-seed outcome comparison, we use the paired degradation $L_v = \log_{10}(\bar{\epsilon}_v(S)/\bar{\epsilon}_{\mathrm{Full}}(S))$, which directly measures the degradation of each ablated variant relative to the Full model on the same paired runs. To make the analysis mechanistic rather than purely outcome-based, we also report explicit-query multipliers relative to the Full model, the fraction of pairwise relations resolved by LP certification, transitive propagation, and explicit DM queries, and the quality of the learned preference vector.

Overall, the Full model provides the best trade-off between final optimization performance and interaction efficiency under the paired-degradation analysis. Removing transitive propagation produced the clearest change in the internal relation-resolution mechanism. The transitivity-resolved fraction dropped from 89.5% under the Full model to 0% under -Trans, while the LP-certification share increased correspondingly from approximately 10.0% to 99.5%. By contrast, the explicit-DM-query share remained essentially unchanged at 0.5%, and the explicit-query multiplier stayed close to one. Consistent with this mechanism-level change, the paired degradation L_v remains close to zero. The raw suite-wise final errors are also mixed rather than uniformly worse: Full versus -Trans gives 0.1066 versus 0.1326 on ZDT, 1.1067 versus 1.0496 on DTLZ, and 0.4247 versus 0.4147 on WFG. These results indicate that transitive propagation mainly avoids redundant LP certification and graph-level resolution work, rather than directly improving final optimization quality or substantially reducing explicit DM queries.

Removing cross-iteration memory produced a different behavioral pattern. Under -Memory, ambiguous queried constraints were no longer retained across active-ranking episodes, so the algorithm could not accumulate reusable supervision over generations. As a result, explicit queries increased by roughly an order of magnitude relative to the Full model, and the explicit-DM-query resolution share increased from 0.5% to 1.6%. The learned preference direction also became less stable: the cosine similarity to the ground-truth preference direction dropped from approximately 0.97 under the Full model to approximately 0.76 under -Memory. This deterioration in preference learning was reflected in the raw final errors, which increased to 0.1774 on ZDT, 10.7842 on DTLZ, and 1.4079 on WFG. These results show that the main role of memory is to preserve scarce user comparisons and reuse them as constraints in later LP-based certification, thereby improving long-horizon interaction efficiency and stabilizing the learned preference model.

By contrast, removing preference-guided candidate ordering had a smaller effect on the interaction-side statistics, but a much clearer effect on the final optimization outcome. In the -PrefSel variant, the explicit-query multiplier remained close to that of the Full model, and the explicit-DM-query resolution share was 0.4%. Active ranking still produced an accurate estimate of the preference vector, with $\cos(\hat{w}, w^\star) \approx 0.98$. However, because this estimate was no longer injected into environmental selection, the algorithm learned the user preference without converting it into actual selection pressure. Consequently, the raw final errors degraded substantially to 0.4266 on ZDT, 230.1968 on DTLZ, and 1.8492 on WFG. This suggests that accurate preference learning alone is insufficient: the learned preference model must also be exploited during candidate selection to produce optimization pressure.

Taken together, these three components play complementary roles at different levels of the algorithm. Transitive propagation reduces redundant internal relation resolution, cross-iteration memory amortizes sparse user feedback across generations, and preference-guided selection converts the retained preference model into final population movement. Their combination is therefore necessary for the Full model to maintain both low interaction cost and strong final solution quality. The suite-wise raw final errors are reported in Table 5, and the corresponding mechanistic statistics are shown in Figure 3 and Figure 5.

Table 5: Raw final $\bar{\epsilon}(S)$ by benchmark suite on the same common benchmark set used in Table 7. For each problem, we first average over the shared paired seeds, and then report mean $\pm$ standard deviation across problems within the suite. We report suite-wise raw values rather than a single overall raw mean because the scale of $\bar{\epsilon}(S)$ differs substantially across suites in the current experiments.

Variant	ZDT $\downarrow$	DTLZ $\downarrow$	WFG $\downarrow$
Full	0.1066 ± 0.0541	1.1067 ± 2.5316	0.4247 ± 0.6889
-Trans	0.1326 ± 0.0937	1.0496 ± 2.2682	0.4147 ± 0.7005
-Memory	0.1774 ± 0.1894	10.7842 ± 46.5472	1.4079 ± 0.5447
-PrefSel	0.4266 ± 0.1148	230.1968 ± 544.1556	1.8492 ± 0.4443

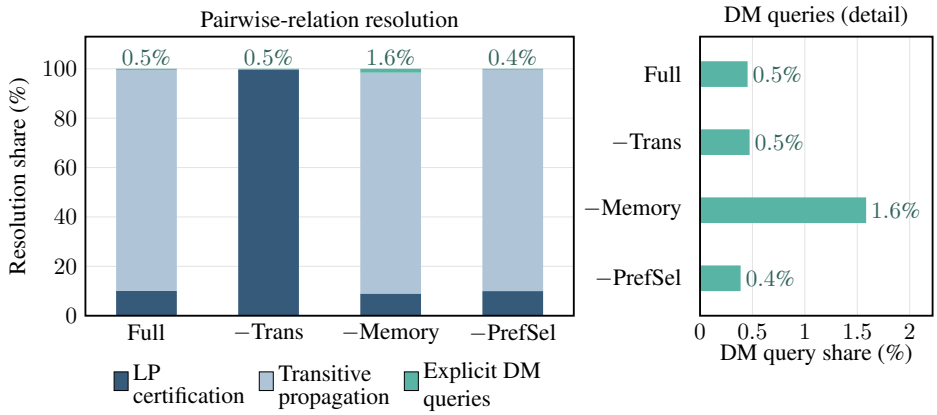

Figure 5: Decomposition of pairwise-relation resolution into LP certification, transitive propagation, and explicit DM queries.

Table 6: Sensitivity analysis of the regularization coefficient λ on DTLZ2. The results are presented as Mean(Std). Best results are highlighted in gray.

λ	$\epsilon^*(S)$	$\hat{\epsilon}(S)$	$\bar{\epsilon}(S)$
0	1.149(1.67e-02)	1.426(1.14e-02)	1.413(1.84e-02)
1e-5	1.149(2.26e-02)	1.432(1.30e-02)	1.418(8.55e-03)
1e-3	1.143(1.45e-02)	1.429(2.10e-02)	1.416(1.63e-02)
1e-1	1.143(2.23e-02)	1.434(1.44e-02)	1.421(1.29e-02)
5e-1	1.154(1.82e-02)	1.434(1.52e-02)	1.422(1.46e-02)

L STRESS TESTS: FEEDBACK CONSISTENCY AND CONSULTATION SCALABILITY

We evaluate PPL under two stressors: noisy decision-maker (DM) feedback and expanded consultation spaces (scaling with population N and objectives m). Figure 6 reports noise robustness, measuring performance degradation via $D_\sigma = \log_{10} \frac{\bar{\epsilon}_\sigma(S)}{\bar{\epsilon}_0(S)}$ for noise level σ. Solution quality remains stable across tested σ; LP certification and transitive propagation successfully resolve most relations, limiting explicit DM queries to $0.6 \sim 0.7\%$. Regarding scalability, while total possible comparisons grow as $O(N^2)$, explicit queries scale subquadratically ($Q_{\exp}(N) \approx a \log N + b$ or $Q_{\exp}(N) \approx aN^\beta + b$). The explicit-query fraction $q_N = \frac{Q_{\exp}}{\binom{N}{2}}$ remains small, though increasing m moderately raises query demand due to reduced certifiability in higher dimensions. Overall, PPL maintains strong interaction efficiency even when its certification mechanisms are naturally stressed by noise and dimensionality.

M OPERATING BOUNDARY: ORACLE GEOMETRY AND REGULARIZATION

To verify that PPL is not overfitted to a specific decision-making model, we evaluate its performance on the 3-objective DTLZ2 problem using two distinct preference structures: the default aspiration-point oracle (representing distance-based preference) and a linear-utility oracle (representing weighted-sum preference, defined as $u(\phi) = w^\top \phi$). As summarized in the main text, both oracles drive the population to the target Pareto region with high precision. Specifically, for the aspiration-point model, the final population achieves an average distance to the preferred region within a narrow margin of the theoretical optimum. For the linear-utility case, despite the change in the underlying selection logic, PPL achieves a convergence value of approximately 0.178, which is within 0.004 of the theoretical lower bound (0.1747). This result, previously visualized via boxplot analysis, confirms that our framework's core mechanisms—transitive propagation and LP-based certification—rely solely on the consistency of pairwise ordinal feedback rather than the specific functional form of the

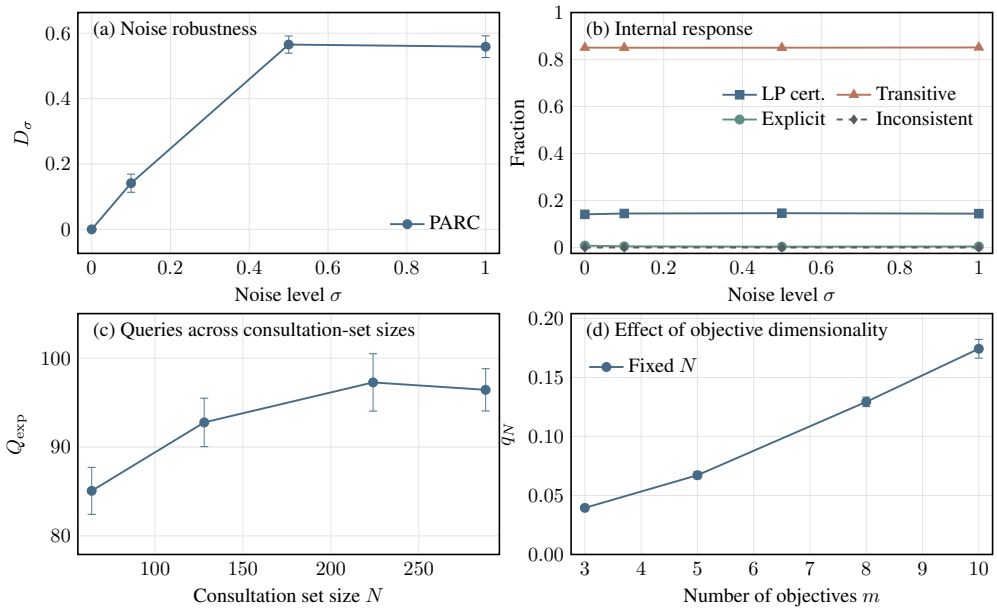

Figure 6: Stress tests of the consultation mechanism. (a) Normalized degradation D_σ under increasing preference noise. (b) Internal response to noisy feedback, including explicit queries, removed inconsistent constraints, and relation-resolution fractions. (c) Scaling of cumulative explicit queries $Q_{\exp}$ with consultation set size N, compared with logarithmic, power-law, and $O(N^2)$ reference trends. (d) Explicit-query fraction $Q_{\exp}/\binom{N}{2}$ and the effect of objective dimensionality m.

decision-maker's utility. Furthermore, we conduct a sensitivity analysis to evaluate the impact of the regularization coefficient λ on PPL. Table 6 reports three preference-proximity metrics: $\epsilon^*(S)$ denotes the distance from the closest solution in S to the designated aspiration point $\mathbf{z}^\star$; $\bar{\epsilon}(S)$ is the mean distance over the population; and $\hat{\epsilon}(S)$ measures the population's proximity to the preferred region around $\mathbf{z}^\star$. On DTLZ2, all three metrics exhibit only minor variations across the tested range of small-to-moderate λ values, with the best mean values attained at different settings. These results suggest that PPL does not require precise tuning of λ within this range. Given its comparable observed optimization quality, we adopt $\lambda = 0$ as the default setting. This choice removes the need to tune an additional hyperparameter and reduces ambiguity certification to a feasibility-based test. Together, these results characterize the performance of `PPL` under the tested noise levels, preference oracles, and regularization settings.

N Further Discussions

N.1 Limitations

The proposed framework adopts a fixed aspiration-point model, under which pairwise comparisons induce affine constraints on the aspiration point. This structure enables persistent feedback reuse and LP-based certification. The hard-mode correctness guarantee requires persistent realizability and the absence of ties, as stated in Assumption 4.1. Fallback returns an estimated ranking without the exact-recovery guarantee of hard-mode completion. The query-complexity result is conditional on the splitting assumption for the comparisons actually queried. In particular, the $\mathcal{O}(m \log N)$ bound applies to a single consultation when the splitting parameter η is bounded below independently of N and m. The algorithm does not guarantee this condition for every candidate set and feedback history. Moreover, the bound concerns explicit DM queries rather than the computational cost of certification and estimation.

Table 7: Mean values of the preference-proximity metric $\mathcal{D}(S)$ for NSGA-II/PPL, IBEA/PPL, MOEAD/PPL, and peer algorithms across benchmark problems. Lower is better; gray shading marks the lowest mean in each row.

PROBLEM	m	DPBEMO	IEMO/D	MOEA/D-PLVF	NSGA-II/PPL	MOEA/D-PBO	NSGA-II/DW	IBEA/PPL	MOEAD/PPL
ZDT1	2	0.098(1.41e-04)	0.111(2.68e-04)	0.165(2.43e-03)	0.127(9.44e-03)	0.089(4.29e-06)	0.127(9.72e-02)	0.074(7.25e-03)	0.163(1.81e-06)
ZDT2	2	0.294(4.76e-02)	0.154(1.93e-06)	0.355(6.61e-02)	0.127(9.44e-03)	0.184(2.23e-03)	0.134(1.23e-03)	0.158(1.51e-06)	0.253(3.79e-04)
ZDT3	2	0.144(2.57e-03)	0.922(1.93e-06)	0.239(2.89e-02)	0.113(1.25e-04)	0.088(4.45e-04)	0.113(1.12e-02)	0.102(1.16e-03)	0.304(8.70e-04)
ZDT4	2	0.098(3.93e-03)	0.115(1.93e-06)	0.208(1.53e-02)	0.284(6.07e-02)	0.238(5.45e-02)	0.284(2.46e-01)	0.060(2.37e-05)	0.216(6.51e-03)
ZDT6	2	0.030(3.43e-04)	0.229(1.93e-06)	0.279(2.75e-02)	0.103(1.19e-02)	0.031(3.12e-04)	0.103(1.09e-01)	0.003(7.65e-07)	0.145(1.29e-04)
WFG1	3	1.227(5.83e-01)	1.813(1.93e-06)	1.360(1.50e-02)	1.027(4.33e-01)	0.979(5.54e-02)	1.027(6.58e-01)	0.240(1.39e-01)	0.529(3.70e-04)
WFG2	3	1.072(7.18e-01)	1.978(1.93e-06)	0.559(3.23e-02)	0.872(5.68e-01)	0.321(3.13e-01)	0.872(7.54e-01)	0.696(2.02e-02)	0.344(5.13e-04)
WFG3	3	1.948(2.78e-01)	0.666(1.93e-06)	0.982(1.66e-01)	0.323(5.84e-01)	0.498(9.02e-02)	1.748(3.58e-01)	1.294(4.30e-02)	1.672(1.32e-01)
WFG5	3	1.084(8.24e-01)	3.183(1.93e-06)	0.687(4.35e-02)	0.217(1.70e-01)	0.884(6.74e-01)	0.884(8.21e-01)	0.052(5.09e-04)	0.883(5.28e-03)
WFG7	3	0.902(8.40e-01)	3.202(1.93e-06)	0.687(4.18e-02)	0.702(6.90e-01)	0.277(2.82e-01)	0.702(8.30e-01)	0.051(8.97e-05)	0.833(2.50e-03)
DTLZ1	3	0.186(3.13e-04)	0.188(1.23e-06)	2.424(8.57e+00)	0.226(1.23e-03)	0.254(1.67e-03)	0.226(3.51e-02)	0.148(5.73e-03)	0.373(2.59e-02)
DTLZ1	5	0.412(7.32e-05)	0.353(1.24e-06)	3.632(7.50e-01)	0.446(8.23e-03)	0.557(6.84e-02)	0.446(9.07e-02)	0.200(9.89e-03)	0.349(2.27e-04)
DTLZ1	8	0.580(1.49e-05)	1.511(3.15e-01)	5.129(6.14e+00)	1.311(1.65e-01)	4.580(1.45e+02)	1.311(4.06e-01)	0.229(2.47e-02)	0.537(6.37e-06)
DTLZ1	10	0.974(1.45e-06)	3.526(1.34e+00)	8.455(7.39e-01)	3.326(1.19e+00)	2.204(8.70e+02)	3.326(1.09e+00)	0.116(1.27e-02)	0.284(5.58e-06)
DTLZ2	3	0.196(2.99e-04)	0.222(1.93e-06)	0.516(8.63e-04)	0.216(1.42e-03)	0.263(5.05e-03)	0.243(3.77e-02)	0.091(7.01e-04)	0.377(1.70e-05)
DTLZ2	5	0.706(3.24e-02)	0.517(1.93e-06)	0.659(3.29e-02)	0.387(9.16e-04)	0.734(5.75e-02)	0.738(3.03e-02)	0.086(9.47e-04)	0.609(6.31e-05)
DTLZ2	8	1.170(3.83e-04)	0.617(1.88e-01)	1.071(8.72e-03)	0.417(3.79e-02)	1.417(4.71e-02)	1.217(1.95e-01)	0.088(1.35e-03)	0.872(4.94e-05)
DTLZ2	10	0.902(1.87e-04)	0.512(1.45e-01)	0.896(4.55e-02)	0.281(1.87e-02)	1.256(1.57e-01)	0.348(1.37e-01)	0.024(7.79e-05)	0.575(6.57e-04)
DTLZ3	3	1.639(1.26e+00)	2.397(1.93e-06)	5.180(7.09e+01)	0.734(3.25e-01)	4.999(3.36e+01)	9.234(4.11e+00)	2.631(4.55e+00)	5.355(5.19e+00)
DTLZ3	5	0.697(1.90e-02)	1.025(1.93e-06)	2.834(6.49e+00)	0.995(3.09e-01)	2.976(2.16e+03)	6.595(2.29e+00)	0.354(4.23e-01)	2.035(1.61e+00)
DTLZ3	8	1.246(2.30e-03)	3.176(7.54e+01)	3.279(6.10e+01)	0.626(2.65e-01)	5.876(2.21e+04)	22.976(8.67e+00)	0.199(8.48e-04)	0.857(4.06e-03)
DTLZ3	10	1.068(4.24e-01)	4.454(1.15e+02)	5.432(7.58e+00)	0.554(2.38e-01)	7.083(4.81e+04)	24.254(1.07e+01)	0.487(8.02e-01)	0.551(9.82e-03)
DTLZ4	3	0.522(1.12e-01)	0.166(1.93e-06)	0.717(3.40e-02)	0.490(9.29e-02)	0.458(1.62e-01)	0.490(3.05e-01)	0.197(5.38e-02)	0.740(6.14e-02)
DTLZ4	5	0.830(3.69e-02)	0.166(1.93e-06)	0.824(3.14e-02)	0.371(3.41e-04)	0.891(2.24e-02)	0.371(1.85e-02)	0.198(6.19e-03)	0.855(2.92e-02)
DTLZ4	8	1.231(4.05e-03)	1.511(1.66e-01)	1.311(1.62e-02)	0.447(1.13e-02)	1.487(1.76e-01)	0.447(1.06e-01)	0.227(3.68e-03)	1.041(9.04e-03)
DTLZ4	10	0.963(1.98e-03)	1.133(1.85e-01)	0.933(3.49e-02)	0.279(1.52e-02)	1.412(3.43e-01)	0.279(1.23e-01)	0.032(2.88e-05)	0.743(1.59e-02)
DTLZ5	3	0.069(3.28e-04)	0.139(1.93e-06)	0.303(6.74e-03)	0.077(3.45e-03)	0.065(1.33e-03)	0.317(3.36e-01)	0.037(2.86e-06)	0.200(5.71e-08)

Continued on next page

Table 7 – continued from previous page

PROBLEM	m	DPBEMO	IEMO/D	MOEA/D-PLVF	NSGA-II/PPL	MOEA/D-PBO	NSGA-II/DW	IBEA/PPL	MOEAD/PPL
DTLZ5	5	0.324(6.59e-04)	0.259(1.93e-06)	0.621(4.59e-02)	0.256(2.85e-03)	0.982(1.03e+00)	0.256(5.34e-02)	0.233(3.03e-06)	0.394(1.54e-03)
DTLZ5	8	0.839(1.83e-03)	1.411(3.66e-01)	1.211(2.16e-01)	0.932(7.14e-02)	2.067(2.97e-01)	0.932(2.67e-01)	0.752(7.28e-06)	0.876(2.90e-04)
DTLZ5	10	0.994(6.50e-04)	1.269(4.24e-01)	1.069(2.74e-01)	1.210(1.68e-02)	2.464(1.16e-01)	1.210(1.29e-01)	1.162(4.32e-04)	1.216(6.32e-02)
DTLZ6	3	0.183(5.81e-03)	0.269(1.93e-06)	1.537(3.64e-03)	0.156(9.92e-04)	1.328(3.25e-02)	0.695(8.52e-02)	0.238(3.83e-03)	0.417(2.26e-03)
DTLZ6	5	0.491(8.96e-03)	1.059(1.93e-06)	1.103(2.19e+00)	1.304(1.89e+00)	9.003(5.26e-01)	1.304(1.37e+00)	0.341(9.32e-04)	0.850(2.26e-03)
DTLZ6	8	2.137(1.95e+00)	3.846(2.77e-01)	3.646(1.27e-01)	0.968(2.06e-03)	9.495(7.24e-01)	2.137(1.40e+00)	2.007(1.37e-02)	1.000(2.87e-03)
DTLZ6	10	1.149(2.68e-03)	1.776(6.39e-01)	1.576(4.89e-01)	1.794(1.42e-01)	10.049(3.37e-01)	3.844(1.12e+00)	1.900(2.66e-01)	1.554(1.16e-01)

Table 8: Mean (standard deviation) of $\bar{\epsilon}(S)$ on the mDTLZ benchmarks for NSGA-II/PPL, IBEA/PPL, MOEAD/PPL, and peer algorithms. Lower is better; gray shading marks the lowest mean in each row.

PROBLEM	m	DPBEMO	IEMO/D	MOEA/D-PLVF	NSGA-II/PPL	MOEA/D-PBO	NSGA-II/DW	IBEA/PPL	MOEAD/PPL
MDTLZ1	3	1.372 (3.25e+02)	0.156 (1.26e-02)	0.161 (6.42e-03)	1.888 (3.91e+00)	0.085 (4.93e-03)	0.208(1.06e-01)	0.192 (3.47e-02)	0.501 (1.87e-01)
MDTLZ2	3	0.134 (2.04e-02)	0.05 (3.76e-04)	0.351 (1.18e-03)	0.239 (1.03e-01)	0.098 (1.35e-04)	0.297(1.22e-01)	0.000264 (1.10e-04)	0.226 (7.37e-03)
MDTLZ3	3	4.798 (7.02e+02)	2.029 (1.76e+00)	6.068 (1.76e+01)	5.747 (6.59e+01)	1.962(2.19e+00)	1.986 (9.79e-01)	1.026 (4.34e-01)	4.227 (2.24e+00)
MDTLZ4	3	0.173 (5.00e-02)	0.209 (7.84e-02)	0.515 (4.45e-02)	0.208 (3.41e-02)	0.214 (6.13e-02)	0.124(8.30e-02)	0.004 (4.84e-03)	0.430 (2.70e-01)

Table 9: Mean values of the preference-proximity metric $\bar{\epsilon}(S)$ for NSGA-II/PPL, IBEA/PPL, MOEAD/PPL, and peer algorithms across the minus-DTLZ benchmarks. Values in parentheses are standard deviations. Lower is better; gray shading marks the lowest mean in each row.

PROBLEM	m	D-PBNSGA-II	IEMO/D	MOEA/D-PLVF	NSGA-II/PPL	MOEA/D-PBO	NSGA-II/DW	IBEA/PPL	MOEAD/PPL
MinusDTLZ1	3	7.375 (1.54e+04)	5.38 (5.14e+03)	5.721 (2.53e+02)	4.226 (2.17e+00)	4.819 (0.00e+00)	8.287 (8.25e+01)	0.013 (5.39e-03)	15.520 (3.91e+00)
MinusDTLZ1	5	4.265 (9.39e+03)	4.838 (3.25e+01)	5.66 (6.55e+03)	3.828 (2.53e+00)	5.314 (0.00e+00)	4.581 (2.51e+01)	0.006816 (4.28e-03)	2.959 (1.64e+00)
MinusDTLZ1	8	6.803 (7.97e+03)	65.03 (4.72e+03)	5.226 (6.80e+02)	4.584 (1.68e+00)	6.782 (4.32e+00)	31.978 (1.38e+01)	0.007946 (6.44e-03)	1.358 (6.34e+00)
MinusDTLZ1	10	5.584 (2.69e+03)	6.584 (3.41e+03)	5.935 (1.86e+03)	4.452 (1.33e+00)	5.223 (3.34e+00)	5.621 (1.30e+01)	0.004538 (2.64e-03)	10.559 (6.46e+00)

Continued on next page

Table 9 – continued from previous page

PROBLEM	m	D-PBNSGA-II	IEMO/D	MOEA/D-PLVF	NSGA-II/PPL	MOEA/D-PBO	NSGA-II/DW	IBEA/PPL	MOEAD/PPL
MinusDTLZ2	3	0.672 (2.71e-01)	1.288 (4.25e-01)	1.289 (1.11e-02)	0.159 (3.36e-02)	1.762 (0.00e+00)	1.111 (5.80e-01)	0.000339 (2.27e-04)	0.734 (2.37e-02)
MinusDTLZ2	5	1.721 (2.15e-01)	3.08 (2.95e-01)	2.385 (9.33e-01)	1.481 (2.18e-01)	1.956 (0.00e+00)	2.128 (5.75e-01)	0.000143 (1.17e-04)	1.406 (2.64e-02)
MinusDTLZ2	8	1.837 (1.72e-01)	1.375 (2.34e-01)	2.436 (1.05e-01)	0.215 (7.71e-03)	1.53 (4.31e-03)	1.482 (4.28e-01)	0.000181 (9.80e-05)	1.467 (6.47e-02)
MinusDTLZ2	10	2.215 (1.77e-01)	2.422 (2.31e-01)	2.861 (8.18e-02)	0.258 (1.83e-02)	1.45 (2.41e-01)	1.232 (3.17e-01)	0.000109 (7.25e-05)	1.540 (3.46e-02)
MinusDTLZ3	3	18.71 (3.04e+05)	16.87 (4.25e+02)	18.905 (1.06e+04)	15.361 (2.46e+02)	17.551 (0.00e+00)	35.555 (3.77e+02)	0.116 (4.07e-02)	4.949 (2.05e+01)
MinusDTLZ3	5	24.306 (1.51e+05)	26.46 (4.42e+05)	29.095 (2.60e+04)	22.804 (5.03e+02)	24.278 (0.00e+00)	27.887 (1.18e+02)	0.072 (2.02e-02)	9.702 (1.20e+01)
MinusDTLZ3	8	28.612 (6.28e+04)	25.612 (3.80e+04)	28.858 (6.12e+03)	24.841 (5.20e+01)	26.42 (2.32e-02)	29.638 (2.79e+02)	0.060 (2.47e-02)	8.621 (3.34e+01)
MinusDTLZ3	10	35.58 (8.31e+04)	34.25 (4.32e+04)	37.339 (3.19e+04)	30.453 (8.52e+01)	34.23 (5.24e+01)	36.336 (1.70e+02)	0.049 (1.61e-02)	9.860 (2.30e+01)
MinusDTLZ4	3	0.394 (1.75e-01)	1.442 (4.07e-01)	1.391 (9.76e-03)	0.167 (7.90e-03)	0.417 (0.00e+00)	0.757 (5.67e-01)	0.005 (3.25e-03)	0.806 (6.00e-03)
MinusDTLZ4	5	1.606 (6.02e-01)	2.995 (8.27e-02)	2.275 (5.60e-01)	1.412 (4.20e-01)	3.033 (0.00e+00)	1.667 (6.09e-01)	0.002 (1.57e-03)	1.482 (8.81e-03)
MinusDTLZ4	8	2.176 (3.59e-01)	2.54 (4.78e-01)	2.748 (1.46e-01)	0.24 (1.18e-02)	1.24 (1.27e-01)	0.888 (2.86e-01)	0.002 (1.42e-03)	1.791 (1.21e-01)
MinusDTLZ4	10	2.658 (2.03e-01)	2.34 (1.45e-01)	2.657 (2.37e-01)	0.34 (1.95e-02)	1.26 (1.56e-01)	1.109 (5.00e-01)	0.002 (1.71e-03)	1.869 (3.01e-01)

O LIST OF NOTATIONS

For convenience, we list the main notations used in this work in Table 10.

Table 10: Summary of the Main Mathematical Notation

Symbol	Description
$\mathbf{x}$, Ω	Decision vector and feasible decision space, respectively.
n, m	Numbers of decision variables and objectives, respectively.
$\mathbf{F}(\mathbf{x})$, ϕ	Objective vector associated with $\mathbf{x}$, with $\phi = \mathbf{F}(\mathbf{x}) \in \mathbb{R}^m$.
t, T, $\mathcal{T}_{\text{int}}$	Generation index, maximum number of generations, and set of interaction generations, respectively.
U_t, κ_t	Indexed pool of individuals supplied by the host EMO algorithm and map from each individual to its consultation representative.
Φ_t, N_t	Set of distinct objective vectors used for consultation and its size $N_t = \lvert\Phi_t\rvert$; written as Φ and N within one consultation.
N_{pop}	Population size of the host EMO algorithm.
$\succ_\star$, o_{ij}	DM's latent preference relation and observed binary response, respectively; $o_{ij} = +1$ reports a preference for ϕ_i over ϕ_j, and $o_{ij} = -1$ reports the reverse.
$\mathcal{C}$, $\mathcal{C}_0$	Retained history of explicitly queried comparisons and incoming history at the start of a consultation, respectively.
$\mathbf{z}$, $\mathbf{z}^\star$	Aspiration point and fixed true aspiration point under the persistent preference model, respectively.
Z, $Z(\mathcal{C})$	Admissible aspiration-point polytope and its subset satisfying $o\, h_{\phi,\phi'}(\mathbf{z}) \geq 0$ for every observation in $\mathcal{C}$, respectively.
$h_{\phi,\phi'}(\mathbf{z})$, $h_{ij}(\mathbf{z})$	Normalized difference in squared distances to $\mathbf{z}$; a positive value favors the first candidate. Here, $h_{ij} = h_{\phi_i,\phi_j}$.
$\mathbf{Q}$	Local pairwise relation matrix: $\mathbf{Q}_{ij} = +1$ places i before j, -1 records the reverse, and 0 denotes an unresolved relation for $i \neq j$.
α_{ij}^+, α_{ij}^-	LP feasibility indicators for the constraints $h_{ij}(\mathbf{z}) \geq 0$ and $h_{ij}(\mathbf{z}) \leq 0$, respectively, over $Z(\mathcal{C})$.
$\widehat{\sigma}_t$, $\sigma^\star$	Output ranking for the consultation at generation t and true ranking under the persistent model, respectively; $\widehat{\sigma}_t$ is written as $\widehat{\sigma}$ within one consultation.
$\widehat{\mathbf{z}}_{\text{S}}$, $\widehat{\mathbf{z}}_t$	Soft-margin aspiration-point estimate and the estimate retained at generation t for EMO guidance, respectively.
$\widehat{\mathbf{z}}_{\text{H}}$	Selected Euclidean Chebyshev center of the nonempty compatible region $Z(\mathcal{C})$.
$\widehat{\mathbf{z}}_{\text{F}}$	Aspiration point used for fallback ranking: $\widehat{\mathbf{z}}_{\text{H}}$ when $Z(\mathcal{C}) \neq \emptyset$, and $\widehat{\mathbf{z}}_{\text{S}}$ otherwise.
τ, ξ_e	Positive fitting margin and nonnegative slack variable for observation e in soft-margin estimation, respectively.
B_t (B), Q	Explicit-query budget at interaction generation t (written as B within one consultation) and number of new explicit queries in that consultation, respectively.
$\mathcal{L}(\Phi)$, N_{rank}	Set of strict rankings realizable by aspiration points in Z and its cardinality $N_{\text{rank}} = \lvert\mathcal{L}(\Phi)\rvert$, respectively.
$\Sigma(\mathcal{C})$	Set of strict rankings of the current candidates induced by some aspiration point in $Z(\mathcal{C})$.
η	Conditional splitting parameter: each explicit query leaves at most a fraction $1 - \eta$ of the current consistent rankings under either response.